THESE DE DOCTORAT DE L'UNIVERSITE DE LA ROCHELLE

Ecole doctorale: Sciences pour l'Environnement Gay Lussac
Spécialité: Aspect Moléculaire et Cellulaire de la Biologie

Présentée par:
Adriana IGLESIAS VASQUEZ

Pour obtenir le grade de:
Docteur de l'Université de la Rochelle

DETECTER LA DIVERGENCE RECENTE A PARTIR DE L'INFERENCE PHYLOGENETIQUE ET DE LA PHYLOGEOGRAPHIE STATISTIQUE: CAS D'UN COMPLEXE D'ESPECES CHEZ UN OISEAU MARIN TROPICAL

Soutenue le 09 Septembre 2016

Devant le jury composé de:

Vincent Bretagnolle	Directeur de Recherche, CEBC-CNRS, *Directeur de thèse*
Jean-François Renno	Directeur de Recherche, IRD, *Rapporteur*
Pierre-André Crochet	Directeur de Recherche, CEFE-CNRS, *Rapporteur*
Andreanna Welch	Chercheur Associé, Université de Durham, *Examinatrice*
Pascale Garcia-Meunier	Professeur, Université de la Rochelle, *Président*

UNIVERSITY OF LA ROCHELLE

Thesis submitted for the degree of
Doctor of Philophy

Doctoral School: Environmental Sciences Gay Lussac
Speciality: Molecular and Cell Aspects of Biology

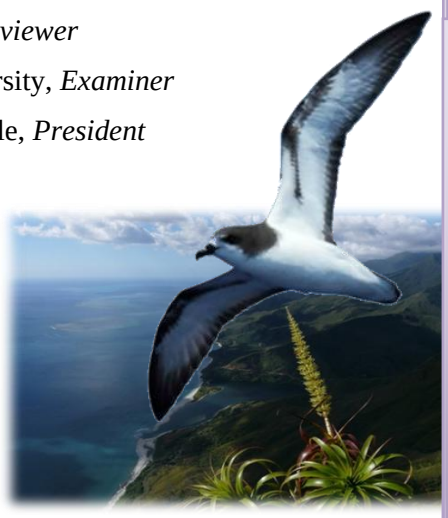
Adriana IGLESIAS VASQUEZ

PATTERNS OF EARLY DIVERGENCE IN A PETREL COMPLEX USING MULTILOCUS PHYLOGENETIC INFERENCE AND STATISTICAL PHYLOGEOGRAPHY

Defended on September 09th, 2016

Jury composition:

Vincent Bretagnolle	Senior Researcher, CEBC-CNRS, *Thesis advisor*
Jean-François Renno	Senior Researcher, IRD, *Reviewer*
Pierre-André Crochet	Senior Researcher, CEFE-CNRS, *Reviewer*
Andreanna Welch	Research Associate, Durham University, *Examiner*
Pascale Garcia-Meunier	Professeor, University of La Rochelle, *President*



"No one ought to feel surprise at much remaining as yet unexplained in regard to the origin of species and varieties if he makes due allowance for our profound ignorance in regard to the mutual relations of the many beings which live around us."

~ Charles Darwin

RESUME

De nombreuses études en phylogéographie, spéciation et délimitation des espèces restreignent leur attention à un sujet limité: la réciprocité monophylétique afin d'élucider les mécanismes et les processus qui donnent origine à la différentiation des populations. D'ailleurs, la réciprocité monophylétique ne fournit pas un critère idéal sur des aspects encore plus précis de la divergence évolutive. De plus, à des niveaux de différenciation récents, des phénomènes tels que la lignée incomplète de tri et le flux génétique continu prédominent. Dans cette optique, l'étude des étapes intermédiaires de divergence contribuera à augmenter notre compréhension de la spéciation géographique, la délimitation des espèces et le meilleur établissement de priorités de conservation. Dans cette thèse, j'ai eu comme objective étudier la structuration génétique et l'histoire démographique d'un complexe de morpho-espèces contraignantes, *Pterodroma leucoptera: P. l. leucoptera, P. l. caledonica, P. l. brevipes,* à l'échelle de son aire de distribution geographique, qui s'étend de l'archipel du Vanuatu à la Polynésie française pour *P. l. brevipes*, de la Nouvelle Calédonie pour *P. l. caledonica*, à l'Australie pour *P. l. leucoptera*. D'ailleurs, grâce aux introns nucléaires (Papier 2), des niveaux de polyphylie entre les populations *caledonica* et *leucoptera* ont été identifié. Le temps de divergence révélé par notre analyse statistique (*BEAST: 30 000 ans BP aprox; IM: 70 00 ans BP), la taille de la population effective (caledonica ~52 384 Ne; leucoptera ~11 814 Ne) et le temps de génération (20 ans) signalent la présence d'un flux génétique actuel plutôt que la rétention d'une diversité de la population ancestrale ce qui pourra expliquer l'absence de structuration génétique entre ces clades. En ce qui concerne les populations du Pacifique du Sud, *brevipes* taxa, (Papier 1), les marqueurs nucléaires et mitochondriaux n'ont pas fourni une différentiation claire entre ces colonies, ce qui suggère que l'isolement reproductif entre eux n'est pas complet. Cependant, la phylogénie et les réseaux d'haplotypes ont révélé une diversité génétique considérable au sein du groupe, mais qui représente une structuration génétique faible entre eux et suggérant la présence du flux génétique continu pour certaines colonies et l'existence d'une divergence très récente pour certains d'entre eux (par exemple *P. b. magnificens*). Dans l'ensemble, notre travail est une autre preuve de la prééminence des marqueurs mitochondriaux sur les sites nucléaires pour éclaircir la structuration génétique aux premiers stades de la différenciation (voir chapitre III, par exemple au sud-ouest des colonies pacifiques contre les populations de l'Australie et de la Nouvelle-Calédonie). Cependant, à des niveaux plus superficiels de divergence (les populations australiennes contre calédoniennes) ces marqueurs n'ont pas détectés de différenciation et place notre étude comme un autre fascinant mais obscur et mal compris exemple du mécanisme de différenciation des pétrels qui peuvent présenter des différences morphologiques et même écologiques, mais pas de différenciation génétique significative. Compte tenu des différences dans l'écologie et de l'état de conservation, et malgré l'absence de différenciation génétique et morphologique, nous recommandons donc que les populations australiennes et calédoniennes devraient être protégés avec le même effort pour préserver le potentiel évolutif de ces lignées. En particulier, comme une source de variabilité en mesure de maintenir la diversité génétique de cette espèce, le taxon caledonica demande plus d'efforts de conservation.

Mot clés: Phylogéographie, *Pterodroma*, différentiation des populations, histoire démographique.

ABSTRACT

Many studies of phylogeography, speciation, and species limits restrict their focus to a narrow issue: gene tree monophyly in order to unravel the mechanisms and processes that shape population differentiation. Furthermore, reciprocal monophyly does not provide an ideal touchstone criterion of any aspect of evolutionary divergence. At shallow time depths, stochastic phenomena such as incomplete lineage sorting and continuous gene flow predominate. In this light, studying intermediate stages of divergence will increase our understanding of geographical speciation, species limits, and conservation priorities. In this thesis, I aimed to assess the species boundaries and to understand the demographic history of a contentious morpho-species complex, *Pterodroma leucoptera,* geographically distributed through the South-West Pacific Ocean: *Pterodroma leucoptera leucoptera* (from Australia, hereafter leucoptera)*, P. l. caledonica* (from New Caledonia, hereafter caledonica)*, P. l. brevipes* (from Vanuatu to French Polynesia, hereafter brevipes). Results obtained by the nuclear introns markers (Paper 2) identified levels of polyphyly between leucoptera and caledonica populations. Moreover, the statistical analysis, such as the theoretical splitting time (*BEAST: 30 000 years BP approx; IM: 70 00 years BP), the effective population size (leucoptera ~11 814N_e; caledonica ~52 384 N_e) and the generation time (20 years), supported by current gene flow rather than the retention of an ancestral population diversity, to explain the non-genetic structuration between these two populations. In regards to the brevipes taxa, (Paper 1), neither mitochondrial nor nuclear introns provided a clear distinction among their colonies, suggesting that reproductive isolation among them is not complete. However, phylogeny and networks revealed considerable genetic diversity within the group but representing a weak genetic structuring among them suggesting gene flow still happens for some colonies and that there is a very incipient divergence for some of them (e.g. between *P. brevipes magnificens* from Vanuatu and the other brevipes populations). Overall, this work is another proof of the preeminence of mtDNA markers over the nuclear ones to unravel genetic structuration at early stages of differentiation (See Chapter III e.g. brevipes colonies against leucoptera and caledonica populations). However, at more shallow levels of divergence (leucoptera against caledonica populations), these markers did not detect signals of differentiation placing our study as another compelling though obscure and poorly understood example of the mechanism of differentiation of petrels being able to portray morphological and even ecological differences but not significant genetic differentiation. Given differences in ecology and conservation status, and despite the absence of neutral genetic and morphological differentiation, we thus recommend that Australian and Caledonian should be protected with equal effort commitment to preserve the evolutionary potential of these lineages. In particular, as a source of variability able to maintain the genetic diversity of this species, taxon caledonica warrants more conservation effort.

Keywords: Phylogeography, Pterodroma, early stages of differentiation, demographic evolutionary history.

This study is dedicated to

my father who inspired me and taught me

to seek knowledge,

Manuel IGLESIAS

"I also think about all the young females in science who can stand on our shoulders, because we will be providing a ladder for them not pulling it up as so many before us have done."

~Laura Boykin, computational biologist

ACKNOWLEDGEMENTS

My first thanks to Vincent BRETAGNOLLE, advisor of this study, for giving me the opportunity to do science, to trust in my person and to allow me to bloom my scientific side. Also thanks, for his precious time, his patience, and his comments which have improved this study.

My deepest thanks to the French government, especially to the Poitou-Charentes and Deux Sevres department that sponsored this work.

I am grateful to Vicki FRIESEN and Tim BIRT for sharing their scientific knowledge and to host me in their Lab and also for showing me that Science and Life can go together.

Many Thanks to Alice CIBOIS for all her advice through many numbers of emails that enlightened my dark moments of doubts. I appreciate them a lot!!!

I would also like to thank Dr. AUSTIN and Dr. IBARGUCHI for their kindly disposition and advises related to museum samples. Many thanks to DANO Stephanie, and RIBOUT Cecile laboratory assistants of the Molecular laboratory of the CEBC.

I am extremely grateful to Dr. WELCH who contributed a great deal to the development of this study sharing their genetics knowledge.

I would also like to thank four anonymous reviewers for helpful comments on the manuscript, Chapter V. I would like to extend my thanks to all the co-authors involved in this research: SHIRIHAI H, PRIDDEL D, CARLILE N, RIBOUT C, GANGLOFF B, DANNO S, CIBOIS A, FRIESEN V your comments and feedback were of enormous help to overcome this manuscript. I am grateful to Sarah SAMADI and her advice that improved in a big deal the final version of this manuscript.

I am also very grateful to all the jury members for accepting to read my manuscript and to contribute to my scientific formation: Dr. Welch and Dr. Garcia. Special thanks, to Dr. Renno who literally saved this Ph.D. I will be always in debt with you !!!

I would like to acknowledge all of my native English friends who kindly accepted to review this manuscript: Sau Ping Yi, Evy Reategui, Lisa SZta, Jacquie Grace, Conchur O'Miachain, Marco Girardello. Especially thanks to my Kiwi friend Elizabeth Brown for giving me a lasting memory of her kindness and generosity reviewing not only the English but also the grammar of this Manuscript.

My thanks also go to the administrative staff of the CEBC: BONNET Delphine ROUBY Evelyse, MARIANI, Andre, PHYILIP Arnaud LACALLE Martine, Mme Nina, and DUCHEMIN Xavier. L'atelier Team DENNIS Simon, DENNIS Patrice, THIBURCE Christian. My Big Big Big thanks to our *Chers Cuistots:* COIFFET Cristoph and VIOLLET Arnaud for all the delicious meals that without a doubt make Happier our days at the Lab.

Totally grateful to the CEBC heads: BONNET Xavier and GUINET Cristophe for all the facilities and support to a foreign student. Thank you so much. I am very thankful to the CEBC-CNRS because literally, this institute was my HOME during FIVE years.

I am also grateful to the Université de la Rochelle staff: Paco BUSTAMANTE Jennifer DE LA CORTE GOMEZ and Isabelle HIRSCH for their always kindness HELP and administrative advice. M. GUICHARD for their time and his good will to teach me YMIR.

I am very thankful with all the *Stagiaires* of the CEBC that made really awesome all this time at CHYSE. I won't mention each of them because I will need two or three pages but each of you were really special in this European and Scientific Adventure, each of you taught me not only your French culture but also many many many things that help me to become a stronger BIRD thanks for all the moments they will last with me my entire life. *Merci les loulous!*

I am deeply thankful to my "*Banshee Boys*" you were kind of angels during this last time: Etienne Debedest and Pierre Bouët du Portal. THANK YOU guys!!! *Et maintenant c'est Bukkkk….*!

I would also like to thank, my *Ashton Kutcher Zumba Trainer*: Thomas LEBRAULT, I do not know what I would have become without sports and dance during these 3 years!!!! Thanks for that two hours per week that helped me to empty my brain and to gain a new energy!!!

I am deeply thankful to Sandra MURGUET, her conferences and meditations came at the right moment just when I was asking for some influential and spiritual empowerment. Thanks for transmitting me all your energy and security. I would like to extend my thanks to Dominique GUINNET for the yoga time, that helped me to re-equilibrate my body and soul.

To my Peruvian friends for sending me a little bit of my culture through the WhatsApp's, skypes and Facetime: Javi Velasquez, Alfonso Lopez, Gabriel Rodriguez, Adriana Lopez, Annete Zevallos. Special thanks To "las Rosarinas Promo 98" who with their jokes made easier the hardest days: Ruth Saldias, Melissa Palacios, Karla Tuesta, Priscylla Pfnening, Lita Onnely, Melina Ramirez, Silvia Mestanza, Luz Dary Nunez, Milagros Murrieta.....My deepest THANKS to Albert who helped me to built the ladders that brought me where I am now, thanks to my well remembered and beloved friend.

My deepest Thanks to my MUM, my BROTHER, and all my family for always being THERE. And Thanks, GOD and NATURE for allow me to feel you and give me all that I have and all that I am. And finally, but one of the most important things: MUSIC without it I wouldn't handle those days.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

A
AIC: Akaike information criterion
AMNH: American Museum of Natural History of New York.
AMOVA: Analysis of Molecular Variance
APE: Analysis of Phylogenetic and Evolution

B
bp : base pairs

C
°C: Centigrades
CEBC: Centre d'études Biologiques de Chize
CO1: Cytochrome Oxidase 1
CSDE 5F: Cold Shock domain-containing protein E1-CSDE-1. Forward sequence
CSDE 6R: Cold Shock domain-containing protein E1-CSDE-1. Reverse sequence
Cytb: Cytochrome b

D
DNA: Deoxyribonucleic acid
dNTP: Deosynucleotide Triphosphates

E

F
F1B: Name of forward sequence CO1

G

H
H16025: Name of reverse sequence of Cytochrome b

I
i.e.: That is to say; in other words
IRF2F: Interferon regulatory Factor 2. Forward sequence
IRF23R: Interferon regulatory Factor 2. Forward sequence
J

K

L

M
mDNA: Mitochondrial Deoxyribonucleic acid
MEGA: Molecular Evolutionary Genetics Analyses
mg: Milligrams
MgCl2: Magnesium chloride
ML : Maximum Likelihood

mM: Millimolars
min: Minutes
MTI: Musée de Tahiti et des îles

N
N: Sample size
Ne: Effective population size
nuDNA: Nuclear Deoxyribonucleic acid

O

P
PAX 20 F: PAX Interacting Protein Forward sequence
PAX 21R: PAX Interacting Protein Reverse sequence
PCR: Polymerase Chain Reaction

Q

R
R1B: Name of Reverse sequence CO1

S
SOP: Societé d'Ornithologie de Polynésie-MANU

T
TPM 1: Tropomosyn 1
Trop 6aF: Tropomosyn 1. Forward sequence
Trop 6bR: Tropomosyn 1. Forward sequence

U
ul: Microliters
uM: Micromolar
ug: Micrograms

LIST OF FIGURES

CHAPTER III

CHAPTER IV

LIST OF TABLES

LIST OF APPENDIX FIGURES

Appendix A.

Appendix B

LIST OF APPENDIX TABLES

Appendix A

Appendix B

"Patterns of early divergence in a Petrel complex using multilocus phylogenetic inference and statistical phylogeography"

"The major ideas in science start out as working hypotheses and they have to survive test and controversy before they are accepted by the scientific community"

~S. Stearns

CHAPTER I

INTRODUCTION

"Knowledge is limitless,
Imagination encircles the WORLD"

~Albert Einstein

CHAPTER I. INTRODUCTION

Since its inception, the theory of evolution was conceived to understand life and the evolutionary forces that forge and maintain biological diversity on earth (Streelman and Danley 2003). Evolution may be described as the change in the relative frequencies of alleles in a population over time from one generation to the next (Futuyma 2005). In addition, as a dynamic process of genetic change, evolution through natural selection produces adaptive characters and species (Carson and Templeton 1984). Thus, the study of evolution embraces two evolutionary patterns, micro and macro evolution, acting at two different levels and timescales (De Queiroz 1998; Stearns and Hoekstra 2005). Indeed, since the recognition that macroevolutionary patterns are governed by the principles of microevolution, several studies have focused on microevolutionary mechanisms to gain insight into macroevolutionary processes, such as speciation (Dobzhansky 1937; Erwin 2000; Simons 2002; Wright et al. 2009; Oliver 2013; Chevin 2016).

Microevolutionary processes regroup changes or variation at the population level, which occur rapidly in terms of time (Stearns and Hoekstra 2005). Populations evolve via adaptation of individuals through natural selection, or via neutral evolution through random genetic drift (Ridley 1994; Reznick and Ricklefs 2009); or both. Whereas natural selection increases beneficial alleles frequency, genetic drift refers to the tendency of any allele to vary randomly in frequency over time (Bull 2006). Character changes within populations occur on various traits, from DNA sequences to morphology and behavior, and traits may follow different evolutionary divergence times or trajectories (Ball and Avise 1992; Greenberg et al. 1998; De Queiroz 2007).

On the other hand, macroevolutionary processes involve the origin of new species and divisions of the taxonomic hierarchy above the species level (Reznick and Ricklefs 2009). The main theory of speciation based on geographic criteria and proposed by Mayr (1963), suggests that speciation in order to occur, the two populations must be isolated by a barrier, which may be physical (model known as allopatric speciation) or ecological (model known as sympatric speciation) (Turelli et al. 2001). Furthermore, during the buildup of these barriers, gene flow between populations is partially restricted, and population genetic differentiation can be progressively accumulated until new species form (Matessi et al. 2002; Albert and Schulter 2005). Indeed, despite more than a century of research on the process of species formation, major questions remain. For instance, how reproductive barriers evolve, how much gene flow restriction is needed for speciation, which state of the speciation process should be considered to assess species boundaries (Morris-Pocock 2012; Loenen 2013; Pavlova et al. 2014; Pereira and Wake 2015). In this view, it was suggested that focusing on early stages of the speciation process may help to understand the drivers of population divergence, speciation, species persistence and the influence of gene flow during the process of species formation (Coyne and Orr 2004; Shafer and Wolf 2013; Lexer et al. 2014).

De Queiroz (1998) illustrated the speciation process by means of a splitting lineage. He made an emphasis to the interval of time where the divergence between emergent lineages is recent and named this as a temporal "grey zone". Such a zone is characterized by distinct loci evolving at different evolutionary rates, with

divergence entailing a competition between unifying and diversifying genetic processes (Hey and Nielsen 2004). Moreover, shared polymorphism may be found within the grey zone (see Clark 1997; Gaggiotti 2011; Silva et al. 2015) for definition and examples), which might be the result of incomplete lineage sorting (persistence of ancestral polymorphism) or ongoing gene flow (through current migration) between young species since reproductive barriers are not completed (Broquet and Petit 2009). In the context of shared polymorphism due to the persistence of ancestral polymorphism, migration is expected to be low or negligible (Slatkin 1987; Welch et al. 2011; Collevatti et al. 2013). Molecular taxonomy is the science of describing, naming and classifying organisms by means of molecular tools (Yang 2014). Indeed, one major challenge in molecular taxonomy is to distinguish species that have low levels of genetic divergence, either because of an ancestral polymorphism signal or because the species continue to exchange genes (Petit and Excoffier 2009).

My thesis was conceived within a molecular taxonomy context associated with phylogeography and coalescent-based evolutionary demography, to study the taxonomy and the evolutionary population size changes of a controversial systematic species complex, *Pterodroma leucoptera* (Hindwood and Serventy 1941). The Introduction section of the thesis has been divided into two parts. PART I, introduces the state of the art of speciation in the literature, including the various modes of speciation, the gradual evolution of the species concept over the last decades and its implications for conservation management, then presenting the current methods used to assess species boundaries and introducing the integrative taxonomy framework, as a way to conceal different types of data and methodologies to delimit and describe taxa. Finally, this first section introduces the phylogeographic and the evolutionary demography approaches as complementary tools to taxonomy. PART 2, describes the biological system and the aims of the thesis.

PART I: State of the art and theoretical context

I. 1. Speciation

The past few decades have seen a resurgence of interest in speciation, which is reflected by the wide literature available (Gavrilets 2003; Butlin et al. 2008; Schluter 2009; Sobel et al. 2009; Feder et al. 2012; Seehausen et al. 2014; Edwards et al. 2015; Lavretsky et al. 2015; Rabosky 2015; Servedio 2015; Chevin 2016). Although the basic framework for understanding speciation was laid down in the early days of evolutionary biology, recent years have seen a change in emphasis. Most notably, greater attention is now given to the role of selection as a cause of divergence (Panhuis et al. 2001; Schluter 2001; Ryan et al. 2007; Turelli et al. 2014; Pyron et al. 2015; Wellborn and Langerhans 2015).

Speciation is defined as the evolution of reproductive isolation within an ancestral species resulting into two or more descendant species (Futuyma 2005). Under a larger scope, speciation is viewed as a process modeled by evolutionary forces, the latter acting either against or in favor of divergence (Albert and Schulter 2005). The former includes migration or gene flow, which tends to homogenize populations, and thus limits the divergence between populations (Slatkin 1985; Bolnick and Nosil 2007). The latter includes natural selection, mutation and random genetic drift, which allow new alleles to be fixed at population or subpopulations levels, which could potentially lead to divergence (Schluter 2001; Nosil et al. 2003;

Schluter 2009). New approaches to understanding speciation have been introduced, many made feasible by the abundance of molecular markers and coalescent-based estimates of divergence and demographic parameters analysis (Drummond et al. 2003; Heled and Drummond 2008; Seehausen et al. 2014; Edwards et al. 2015). For instance, genes that determine species differences can be located and in some cases isolated (Orr 2001; Seehausen et al. 2014). Despite the great progress made in our understanding of speciation mechanisms, major questions and controversies remain, including: which are the mechanisms that drive or constrain the evolution of reproductive isolation? what conditions does geographic isolation constitute a reproductive isolating barrier? When is speciation non-ecological? How mechanisms of habitat preference evolve and promote divergence with gene flow, How long it takes for a new species to form and what factors affect this duration? etc. (Sobel et al. 2009; Butlin et al. 2012; Berner and Thibert-Plante 2015) (Figure 1.1).

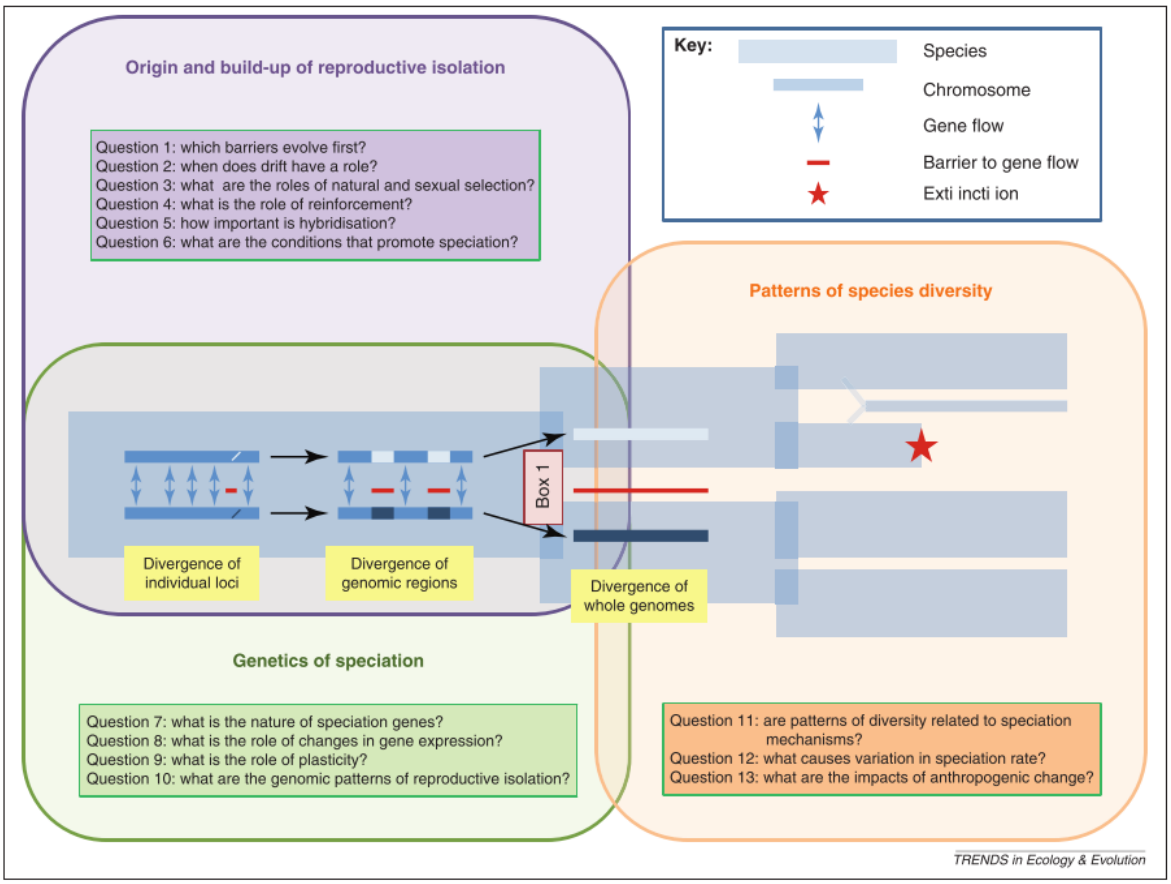


Figure 1.1 Current outstanding questions in relation to the process of speciation and diversification in sexual eukaryotes. From Butlin et al. (2012).

I.1. 1. **Modes of speciation**

Speciation models have been classified historically by several criteria, for instance: (i) the geographical arrangement of populations undergoing the process of speciation (allopatric, sympatric or parapatric), a model that focuses on the inhibitory effects of gene flow on the evolution of reproductive isolation (Mayr 1942a), ii) the

genetic and causes of the evolution of the barriers, (Table 1.1). New insights into this process have suggested a new classification based on mechanisms that drive the evolution of reproductive isolation (Schluter 2000; Schluter 2001; Via 2001) categorizing speciation into ecological and non-ecological (Table 1.2). Below, I summarize the research framework developed for studying the “biology of speciation" until present.

Table 1. 1. Modes of speciation classified by several criteria: the geographic origin of reproductive barriers, genetic bases of the barriers, and the causes of evolution of the barriers. From Futuyma (2005).

MODES OF SPECIATION
I**. Classified by geographic origin of reproductive barriers**
A. Allopatric speciation
1. Vicariance
2. Peripatric speciation
B. Parapatric speciation
C. Sympatric speciation
II. **Classified by genetic and causal bases**
A. Genetic divergence (allele substitutions)
1. Genetic drift
2. Peak shift (peripatric speciation)
3. Natural selection
a. Ecological selection
i. For reproductive isolation
ii. Of reproductive barriers
iii. Of pleiotropic genes
b. Sexual selection
B. Cytoplasmic incompatibility
C. Cytological divergence
1. Polyploidy
2. Chromosome rearrangement
D. Recombinational speciation

I.1.1. Models of speciation based on the geographic basis of isolation

Under the geographic schema, speciation may occur in three kinds of geographic scenarios (Figure 1.2). The allopatry speciation model has become the most supported mode of speciation (Coyne and Orr 2004; Fitzpatrick et al. 2009; Grant and Grant 2009; Silva et al. 2015). Allopatric speciation is defined as the evolution of reproductive barriers in populations that are prevented by a geographic

barrier from exchanging genes at more than a negligible rate. Thus physical barriers can lead to a genetic isolation between populations and generate new species. An early example of this scenario was provided by Darwin himself on finches (Darwin 1859) where the geographic isolation between different species is now confirmed to have led to several speciation events, such as species radiation (Grant and Grant 2002; Grant and Grant 2009). Frequently, a subdivision of this model is made, allopatric speciation by vicariance (divergence of two large populations, Figure 1.2A) and by peripatric condition (divergence of a small population from a widely distributed ancestral form, Figure 1.2B) (Futuyma 2005). The latter category of speciation was first called by Mayr (1954) the founder effect speciation. This hypothesis was based on the observation, in birds and other animals characterized by isolated populations with the peripheral location to the distribution of a probable parent species. Frequently, these populations are highly divergent, to the point of being classified as different species or even genera. Mayr views considered that genetic change should be very rapid in localized populations founded by a few individuals and cut off from gene exchange with the main group of the species (Futuyma 2005). A second category under the geographic criteria is called parapatric speciation (Figure 1.2C), where neighboring populations with moderate gene flow, then diverge and become reproductively isolated. In addition, parapatric speciation undoubtedly occurs and may be even common, but it is very difficult to demonstrate that it provides a better explanation than allopatric speciation (Coyne and Orr 2004; Butlin et al. 2008; Bank et al. 2012; Seehausen et al. 2014). Finally, the most debated mode of speciation under the geographic criteria is speciation by sympatry: sympatric speciation occurs without the existence of a geographical barrier; indeed, divergent populations coexist in the same geographical space (Bush 1994; Bolnick and Fitzpatrick 2007). Thus sympatric speciation assumes the presence of a non-physical barrier restraining the gene flow between conspecific organisms/forms at the population's core, which leads to an *in situ* divergence of the population (Savolainen et al. 2006). Despite extended scepticism about species origin under sympatric scenarios, several studies have now supported the existence of this type of speciation (Diehl and Bush 1989; Turner and Burrows 1995; Roe and Lydeard 1998; Turelli et al. 2001; Friesen et al. 2007; Turelli et al. 2014; Berner and Thibert-Plante 2015; Gabirot et al. 2015; Ruskey and Taylor 2016).

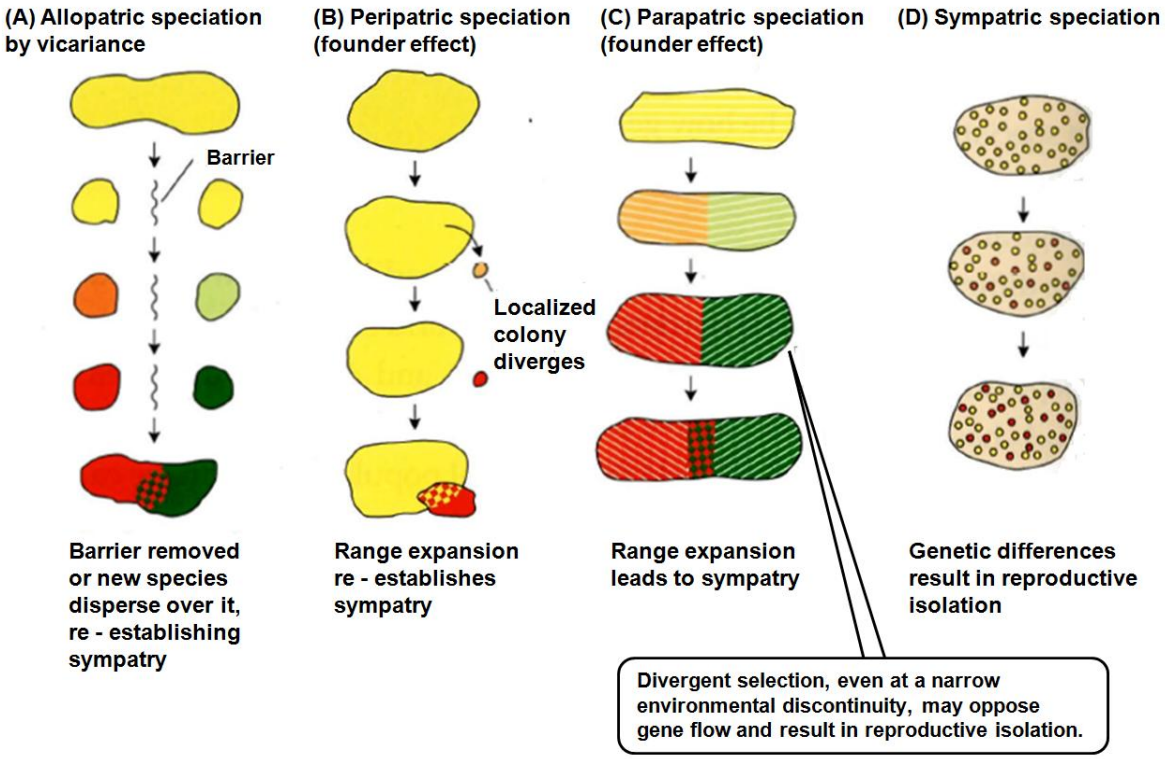


Figure 1.2. Models of speciation under the geographic speciation model. (A) Allopatric speciation. (B) The peripatric, or founder effect, model of allopatric speciation (C) Parapatric speciation (D) Sympatric speciation. From Futuyma 2005.

This classification has however been considered unsatisfactory and in some cases, obsolete (Butlin et al. 2008; Fitzpatrick et al. 2009). As for instance, its focus on the extremes of the speciation continuum, subordinating other dimensions on which speciation processes may vary. In this view, it has been claimed that this classification ignores the fact that speciation is a continuous process that commonly has phases in different spatial contexts (Butlin et al. 2008). Moreover, from a theoretical perspective, allopatric and sympatric speciation are considered as the ends of continuum of gene flow levels among diverging populations (rate exchange (m), allopatry m=0; sympatry m=0.5 (Gavrilets 2004). However, parapatric conditions where gene exchange considers larger migration rate ($0<m<0.5$) is not well studied and has received little attention. Furthermore, considering the duration of the speciation process, maintenance of the extreme conditions of allopatry and sympatry has been argued to be less plausible (Butlin et al. 2008). Discomfort with the geographical classification of speciation has led researchers to explore and reconsider other methods of categorization. Two additional axes have been contemplated: the forces driving the divergence between populations which lead to reproductive isolation and the genetic basis of isolation (Wu 2001; Butlin et al. 2008).

I.1.1.2. Models of speciation based on the mechanisms of divergence and the genetics basis of isolation

One of the first attempts to focus on the mechanisms or processes of speciation based on population genetics was made by Templeton (1981). However, he strictly focused on genetics, considering only one aspect of a complex process of

speciation. A more recent classification by Kirkpatrick et al (2002) considers five requirements for speciation: i) the existence of a disruptive selection, ii) a prezygotic isolating mechanism iii) a link that transmits the force of selection to the isolation mechanism, iv) a genetic basis for the increase in isolation and v) something to initiate divergence (including a period of allopatry). Another attempt under the processes-based classification has been made by Dieckmann et al (2004). They called this classification “speciation routes", emphasizing that speciation is a journey that takes time, during which conditions can change. They consider three axes: spatial, mating and ecological differentiation. Favorable points of this classification are that a) it takes into account the interactions among driving forces, and between those forces and the developing differentiation, b) the possibility for different phases of speciation to occur in different contexts (such as initial divergence in allopatry followed by reinforcement in parapatry). However, it lacks representation of the different genetic mechanisms that might underlie differentiation.

Another categorization of speciation which is re-gaining momentum over the last years is the ecologically-based divergent criteria (Schluter 2001; Via 2001; Via 2002; Rundle and Nosil 2005; Langerhans et al. 2007; Schluter 2009; Turelli et al. 2014; Pyron et al. 2015; Wellborn and Langerhans 2015). The idea that speciation is the result of the microevolutionary process of ecologically-based divergent selection has its roots in the modern evolutionary synthesis (e.g. Mayr 1942; 1947; 1963 Dobzhansky 1951). The current usage of ecology divergent criteria classifies speciation as ecological and non-ecological (Schluter 2001; Sobel et al. 2009). There are several definitions of ecological speciation, while too many uncertainties when defining non-ecological speciation (Sobel et al. 2009). In their more concise forms ecological speciation is defined as the case in which divergent selection leads to reproductive isolation, and speciation is driven by natural selection, by uniform selection, by polyploidy (mutation order), while "non-ecological" is the speciation promoted only by genetic drift (Schluter 2009; Sobel et al. 2009).

Table 1.2. Modes of speciation categorization based on the mechanisms that drive the evolution of reproductive isolation. From (Schluter 2001; Sobel et al. 2009)

Type Schluter 2001	Mode of speciation	Mechanism of initial divergence	Initial form of reproductive isolation	Proximate basis of reduced hybrid fitness	Examples of the roles of natural selection	Example of roles of sexual selection	Type Sobel et al (2009)
Ecological	By natural selection	Divergent natural selection	Pre-zygotic or post-zygotic	Ecological selection, genetic incompatibility and sexual incompatibility	Initial: Drive in phenotypic traits. Final: Reinforcement	Amplify divergence of mate preferences initiated by natural selection. Reinforcement	***Ecological***
Non - ecological	By divergence under uniform selection	Different advantageous mutations occur in separate populations experiencing similar selection pressures	Pre-zygotic or post-zygotic	Genetic incompatibility and sexual incompatibility	Initial: Drive fixation incompatible mutations in different populations Final: Reinforcement	Drive fixation of alternative incompatible mutations in different populations. Reinforcement	
	By polyploidy	Post-zygotic	Genetic drift	Genetic incompatibility	Initial: none; or promotes further genetic divergence. final: reinforcement	Reinforcement	
	By drift	Pre-zygotic or post-zygotic	Pre-zygotic or post-zygotic	Genetic incompatibility and sexual incompatibility	Initial: None; or opposes divergence. Final: Reinforcement caused by drift	Amplify differences in mate preferences. reinforcement	***Non-ecological***

I.1.1.2.1. Ecological speciation

Ecological speciation is defined as the process by which barriers to gene flow evolve between populations as a result of ecologically-based divergent selection, arising from differences between ecological environments (Schluter 2001; Via 2002). It predicts that reproductive isolation should evolve between populations adapting to contrasting environments but not between populations adapting to similar environments**.** Divergent selection is ecological when it arises as a consequence of the interaction of individuals with their environment during resource acquisition. For example, selection can arise from an individual's quest to obtain food and other nutrients, attract pollinators, or avoid predators. It can also arise from their interaction with other organisms in their attempt to achieve these goals (e.g. competition) (Sobel et al. 2009; Nosil et al. 2009). Moreover, selection is divergent when it acts in contrasting directions in the two populations and we include here the special case in which selection favors opposite, usually extreme, phenotypes within a single population (termed disruptive selection), as occurs during sympatric speciation (Nosil et al. 2009). Ecological speciation covers other models in which chance events play a central role, including speciation by polyploidization, hybridization, genetic drift, and founder events/ population bottlenecks. An alternative definition of ecological speciation would restrict it to situations in which barriers to gene flow are ecological in nature (Table 3) (Schluter 2001, Sobel et al 2009).

I.1.1.2.1.1. Mechanisms of ecological speciation (Schluter 2001)

Ecological speciation is a concept that unites speciation process in which reproductive isolation evolves ultimately as a consequence of divergent (including disruptive) selection on traits between environments. "Environment" refers to biotic and abiotic elements of habitat (e.g. climate, resources and physical structure) as well as to interactions with other species (e.g. resource competition, predation, mutualism and various forms of interspecific inference). A diversity of evolutionary processes might be involved. Ecological speciation might occur in allopatry or in sympatry, lead to premating isolation, post-mating isolation or the combination of both. It includes several (but not all) modes of speciation involving sexual selection). Ecological speciation may indirectly result as a consequence of natural selection on morphological, physiological or behavioral traits, or may include direct selection on premating isolation (reinforcement). Distinguishing the ways in which divergent selection has led to reproductive isolation is among the greatest challenges of the empirical study of ecological speciation.

I.1.1.2.1.2. Markers and time frames to understand ecological speciation

Most molecular markers used in population genetic studies are expected to evolve neutrally and should therefore only be subjected to drift mechanisms. This implies that their evolution is determined only by the primary mutation rate, which can be considered constant over time and for which reasonable estimates exist for different markers (microsatellites, mitochondrial DNA sequences, nuclear introns, SNPS (Single nuclear polymorphism). Not only the level of polymorphism of the molecular markers is important to estimate variability, but also the nature of the markers. Indeed, to study neutral variation the markers need to be neutral, but to estimate adaptive variation -adaptation to different conditions, adaptive markers are needed (e.g. morphological traits subjected to positive or negative selection). The

differential behavior of these two types of markers can be used to make specific predictions about the patterns that are to be expected after an initial split of populations. Ecological or adaptive speciation modes imply that population subdivisions arise on the basis of selection, rather than neutral drift. Thus even with small selection coefficients, they are likely to occur within relatively short time frames, probably measuring thousands of generations at most. On this basis, Dieckmann et al. (2004) proposed the following generalized phases of further differentiation after the initial splitting has occurred under ecological-sympatry conditions.

Phase 1. Under this phase, traits experience disruptive selection thus allowing the utilization of alternative niches. The divergence that results in these traits is coupled to an increasing degree of assortative mating. In a natural population, two differently adapted types that mate assortatively would be expected to emerge. At this early stage, the expectation is also that most of the polymorphic alleles between the populations would still be shared because the populations would not have experienced a significant bottleneck, and gene flow may still occur, at least with respect to genes and chromosomal regions that are not involved directly in the differential adaptation. This phase may last less than 100 generations and may thus be detectable only in special circumstances and remains largely to be explored.

Phase 2. Within 100 to 1000 generations, the morphotypes, and the assortative mating tendency will become more pronounced and reach the final state. The resultant strong reduction in gene flow means that the neutral alleles in the two subgroups increasingly become subject to independent drift, which produces different relative frequencies of the alleles. This phase can be inferred reliably from the study of the degree of the genetic subdivision with highly variable molecular markers, such as microsatellites.

Phase 3. Within 1000 to 10 000 generations, no further change with respect to morphotype differentiation or assortative mating pattern is expected. However, significant molecular differences will now have built up. While only allele frequency changes occur in phase 2, phase 3 is characterized by fixation and lineage sorting of neutral alleles (related to genotypic clustering, but generally referring to the grouping of taxa in gene genealogies, which can range from polyphyletic trough reciprocally monophyletic relationships (Nosil et al. 2009)). Furthermore, new mutations will arise as a single mutational step away from pre-existing alleles, which can be used as diagnostic markers. Molecular phylogenetic reconstructions methods by means of neutral markers become applicable at this stage.

Phase 4. After possibly longer separation times, of up to millions of generations after the initial split, prediction of the further evolution of the involved adaptive characters is no longer possible. Additional adaptations might occur, but relative stasis with respect to the initial adaptations is also possible. However, there will now be a clear molecular distinction with respect to allele types and frequencies. Many population-specific alleles will have evolved, and differ by multiple mutational steps from alleles that existed previously. The accumulation of many mutations may also have led to postzygotic isolation, and species status will be generally acknowledged.

I.1.1.2.2. Non- ecological speciation

Sobel et al (2009) suggested that under the non-ecological criteria, the mechanisms that may lead to the reproductive isolation of speciation is genetic drift. However, they point that factors that influence the magnitude of genetic drift, such as variation in population size or mating success may often have an ecological basis. Non-ecological speciation may also include models in which selection is involved, but is non-ecological and/or is not divergent between environments. Examples include certain models of speciation by sexual selection. One argument against speciation by genetic drift is that most of the traits that have the potential to be involved in reproductive isolation are likely to be subjected to natural selection, and it will be difficult for drift to alter such adaptive traits unless drastic population bottlenecks are involved (Devaux and Lande 2008). An additional argument against genetic drift as a driver of speciation is the time required for it to cause complete reproductive isolation. Theoretical treatments show that isolation by genetic drift alone is slow, and would be overshadowed by even extremely weak selection (Nei 1976; Nei et al. 1983). Reproductive isolation by drift takes much longer to evolve than by selection, and it must evolve to completion before it can result in new species. These conditions raise considerable doubt that genetic drift alone is a significant mechanism of speciation.

Table 1.3. Reproductive isolation caused by ecological factors. From Sobel et al 2009.

Reproductive barrier	System	Role of ecology	References
Habitat isolation	*Lucania* (killifish)	*L.goodie* and *L. parva* display reduced survival to adulthood when reared at nonnative salinity levels. Natural distributions along salinity gradients generally correspond to fitness differentiation.	Fuller et al 2007
	Drosophila (fruit fly)	*D. santomea* and *D. yakuba* inhabit distinct habitats based on ecological conditions associated with elevation. Both species exhibit reduced survival to adulthood and fertility when reared at nonnative temperatures, and each species chooses its native temperature range when placed on a temperature cline.	Matute et al 2009
	Mimulus (mon-keyflower)	*M. lewisii* and *M. cardinalis* show considerable allopatric separation based primarily on differences in altitude inhabited. Reciprocal transplants demonstrate that each species is most fit in its native range.	Ramsey et al 2003; Angert and Schemske 2005.
Temporal isolation	*Inurois* (geometrid moth)	Populations of *I. punctigera* in colder climates show divergence	Yamamoto and Sota 2009.

		for early or late flight periods conferring temporal reproductive isolation.	
	Mimulus	Inland and coastal forms of *M. guttatus* experience selection for different growth and flowering times resulting in flowering phenology with little overlap.	Lowry et al. 2008
Sexual and pollinator isolation	*Drosophila*	Divergent artificial selection in laboratory populations of *D. serrata* results in assortative mating.	Rundle et al. 2005
	Gasterosteus (stickleback fish)	Anadromous and freshwater *G. aculeatus* experience divergent selection for body size and assortative mating is based on this trait.	McKinnon et al. 2004
	Mimulus	In sympatry, *M. cardinalis* and *M. lewisii* experience almost complete reproductive isolation due to pollinator preference for floral traits.	Ramsey et al. 2003
Gametic isolation	*Echinometra (sea urchin)*	Lineage specific positive selection on bindin, a gamete recognition protein, was detected in E. lucunter,which experience a strong block to fertilization by sperm of its Neotropical congeners.	McCartney and Lessios 2004 Ramsey

	Mimulus	Divergent pollinator-selected style lengths in *Mimulus cardinalis* and *M. lewisii* lead to differentiated pollen tube lengths, reducing the amount of expected hybridization in mixed pollinations.	Ramsey et al. 2003
Intrinsic postzygotic isolation	*Drosophila*	Adaptive divergence in nuclear pore proteins causes lethality in hybrids of *D. melanogaster* and *D. simulans*.	Presgraves et al. 2003
	Drosophila	Hybrid incompatibility and sterility between *D. melanogaster* and sibling species *D. simulans*, *D. mauritiana*, and *D. sechellia* involves the *Hmr* gene that exhibits signature of positive selection.	Barbash et al 2003.
	Mimulus	Hybrid inviability between *M. guttatus* populations on and off copper mine tailings is linked to two genes for copper tolerance.	Christie and Macnair 1987
Extrinsic postzygotic isolation	Gasterosteus	Divergent natural selection causes low fitness in *G. aculeatus* benthic–limnetic hybrids, despite the absence of intrinsic postzygotic isolation.	Hatfield and Schluter 1999; Rundle 2002
	Sylvia (warbler)		
	Heliconius (butterfly)		

I.2. **Species delimitation**

Identifying species-level in biological diversity has experienced a shift and huge progress with the development of new tools and easier acquisition of the genetic data (Fujita et al. 2012; Hsieh et al. 2013; Satler et al. 2013; Andrews et al. 2014; Lavretsky et al. 2015). However, despite progress in data and methods, taxonomists and systematicians now face new issues such as inaccuracy and incongruence of the different methods. For instance, Carstens et al. (2013) showed that fewer than 30% of the studies reviewed have made taxonomic recommendations and only 25% described new species. Indeed the time required for enough evolutionary changes to appear and allow two distinct lineages to be recognized is not instantaneous, which makes difficult to set species boundaries. In addition, the formulation of a species concept and the development of approaches for delimiting species are two different tasks (De Queiroz 2005; 2007). In this view, systematists must choose both species concept and a criterion to apply this species concept to the data, which has been confounding as pointed by de Queiroz (2007) who distinguished between species conceptualization concepts and species delimitation. Hohenegger (2012) similarly separated explanatory and operational concepts. In this ax I will present the species concepts developed through the last decades, its implication into conservation policies to then give an overview of the current methods for delimiting species, while concluding with the holistic approach called Integrative taxonomy.

I.2.1.**Species concepts**

Despite “species” is one of the most important units in biology, and commonly used as a reference of biological information, the definition of the term is still controversial. A whole discipline, taxonomy, is dedicated to the delimitation of species, which requires a species concept (Hausdorf 2011). Species concepts not only define what a species is, but also define what speciation is. Thus, research hypotheses and studies focusing on the conditions and factors resulting in speciation, ultimately depend on the species concept. Species are fundamental units when studying evolutionary biology, ecology, and conservation research, since an inaccurate understanding of species diversity may lead to errors in analyses that use species as units (e.g. phylogenetic community structure analyses), and may hinder conservation efforts (Wiens 2007). Even in legislation, species is one of the most often used biological units, not only with respect to conservation but also with regard to agricultural or medical issues. However, the question of what defines a species is contentious, as evidenced by the large number of species concepts developed by evolutionary biologists (Table 1.4; adapted from De Queiroz (2007). In fact, there is surprisingly little agreement about the species concept in spite of the effort to unify them. Actually, most species concepts are incompatible with one another (e.g.; phylogenetic vs biological species concept), in that they lead to the recognition of different levels of species taxa (Cracraft 1983; Zink 1996). De Queiroz (1998) summarized speciation by means of a bifurcating lineage into two new forms through time (Figure 1.3). His figure depicts the time at which splitting lineages acquire different properties, which are used as species criteria: phenetic differences, reciprocally monophyletic, ecologically differences, etc. His diagnosis is rather convincing, and one may wonder why there is still a problem, since describing the problem is the first step in solving it. The species problem is reflected by the fact that

each species concept identifies the specific level of divergence through the continuum of speciation (Figure 1.3). Thus some closely related species might be placed in the De Queiroz (1998) "grey zone" where the speciation process is ongoing and where different types of characters and criteria will not provide the same identification. Species concept has to be formulated in such a way that the earliest stages of speciation are also covered, which is the uneasy part of the divergence continuum. After divergence, differentiation of the resulting species continues, thus it is easier to distinguish later stages of the speciation processes.

Table 1.4. Alternative contemporary species concepts adapted from de Queiroz (2007)

Species Concept (Traditional name)	**Distinctive Properties (Species criteria)**	**Advocates Proposing Explicit (Species definitions)**
Biological	Potential interbreeding/intrinsic reproductive isolation	Wright (1940); Mayr (1942, 1963); Dobzhanski (1950)
Isolation	Isolating mechanisms	Mayr (1942, 1963); Dobzhansky (1970)
Recognition	Compatible mate recognition and fertilization systems	Paterson (1978, 1985)
Evolutionary	Unitary evolutionary role, tendencies, and fate	Simpson (1951, 1961); Wiley (1978; 1981)
Ecological	Distintct adaptative zone (niche)	Van Valen (1976)
Cohesion	Intrinsic cohesion mechanisms	Templeton (1989)
Phylogenetic	Association with Phylogenetic Systematics (Cladistics)	
Hennigian	Species bounded at both ends by cladogenetic (lineage splitting)	Hennig (1966); Ridley (1989)
Monophyletic (Apomorphic)	Monophyly (as evidenced by apomorphies= derived character states	Rosen (1979); Mishler (1985)
Diagnosable	Diagnosability (possession of fixed character state differences)	Cracraft (1983); Nixon and Wheeler (1990)
Genealogical	Exclusive coalescence of alleles for multiple loci	Baum and Shaw (1995)
Phenetic	Phenetic cluster (group of similar organisms separated by gaps from other such groups)	Michener (1970); Sneath and Sokal (1973)
Genotypic Cluster	Deficit of genotypic intermediates (heterozygotes) at multiple loci	Mallet (1995)

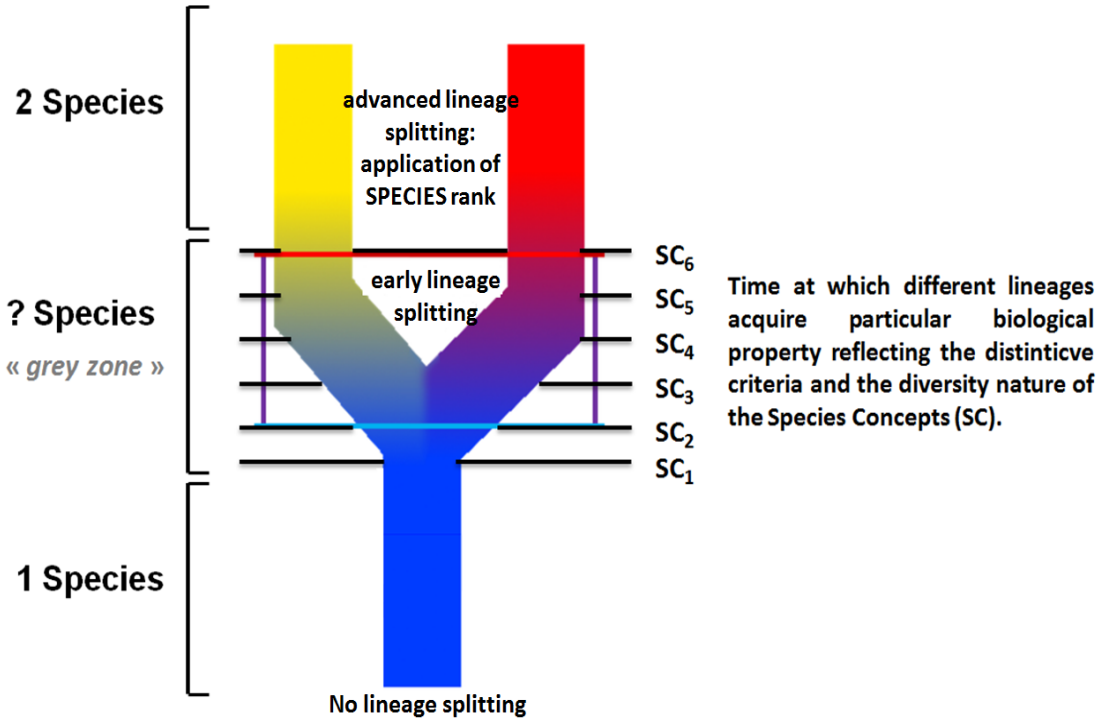


Figure 1.3 . Lineage divergence and alternative species criteria. The diagram represents the process of lineage divergence through a cladogenetic event. The numbered horizontal lines (SC_1-SC_6) represent the times at which the daughter lineages acquire different properties relative to each other (e.g., when they become phenetically distinguishable, diagnosable, reciprocally monophyletic, reproductively incompatible, ecologically distinct, and so forth). The species problem results from disagreements about with of these properties are necessary (defining) properties of the species category (species criteria). The entire set of properties defines a zone, "grey zone" in which there will be disagreement about the number of species among authors adopting properties as their species criteria. From De Queiroz (1998)

Over the past decades, the two most used species concepts were the biological species concept (BSC) and the phylogenetic species concept (PSC), and most researchers favor one or the other, despite these concepts and their interpretations have been adapted with recent theoretical and technical progress, such as the presence of gene flow during the diversification processes. Mayr (1942) proposed the most influential species concept, the biological species concept. He described species as "groups of actually or potentially interbreeding natural populations, which are reproductively isolated from other such groups". Despite wide acceptation, four main points were and are still debated. First, the BSC applies only to biparental reproducing organisms. Second, the finding that reproductive barriers are semipermeable to gene flow and that species can differentiate despite ongoing interbreeding contradicts or at least smoothes the BSC. Third, strictly allopatric populations are difficult to classify under the BSC, since populations are not in contact with each other making the level of interbreeding impossible to evaluate, which limits the applicability of BSC. Finally, hybridization is not considered in the BSC, because under its strict interpretation the production of viable hybrids questions

the separation of the two taxa as different species. Some authors proposed a relaxed interpretation of the BSC (Cracraft 1997; Coyne and Orr 2004). For instance, gene flow between species could be tolerated as long as it is minor. The main alternative species concept is the phylogenetic species concept (PSC), which has the main advantage over the BSC to be applied to uniparental organisms, fossils, and allopatric populations. According to the diagnosable version of the PSC, a species is “a diagnosable cluster of individuals within which there is a parental pattern of ancestry and descent (Eldredge and Cracraft 1980). More recently, the necessity for clusters to be monophyletic was added (Donoghue 1985), as well as the necessity to share at least one unique derived character (e.g. morphological, behavioral or genetic character). Hybridization is also covered under the PSC concept, taking the two parent taxa as separate species so long as there is some level of character discontinuity between them. The PSC can be easily used in DNA barcoding/ DNA taxonomy studies (e.g., Hare et al. 2007; Monaghan et al. 2009). However, one of the main shortcomings of the PSC is its negation of reproductive isolation as criteria. Both BSC and PSC share the necessity of a threshold for discriminating between species, which indeed may be subjective (Agapow et al. 2004).Since reproductive isolation is build up gradually over time, the PSC will lead to accepting divergence at early stages of the speciation process than the BSC does. Biologist are forced to decide what level of difference between two taxa makes each worthy of species-level classification (Hendry et al. 2000).

De Queiroz (2005) attempted to clarify the "*species concept problem" by* proposing the unified species concepts, "the general lineage concept". His idea was to retain the element that was common to all contemporary species concepts eliminating the conflicts between the rival species concepts. He recognized that contemporary species concepts are diverse, being based on important biological properties, necessary for considering lineages to be species (e.g., intrinsic reproductive isolation, monophyly, diagnosability). Nevertheless, all share the fundamental idea that species are segments of lineages at the population level of biological organization. Under his view, there are no secondary species criteria, it is only necessary to assess lineage separation. Therefore lineages do not have to be phenetically distinguishable or diagnosable or monophyletic, or reproductively isolated, or ecologically divergent or anything else. They only have to be evolving separately from other lineages. Thus, undifferentiated and undiagnosable lineages belong to the same species. It is important to note that lineages under this concept are not clades nor monophyletic group, as in the phylogenetic species concept for instance (Table 1.4). Here lineages connect ancestry and descent, where members share more recent common ancestors with recently diverged side branches than they do with earlier members of the same lineage (Figure 1.4). Moreover under this view, a population is an organizational level above that of the organism, rather than the usual sense of a reproductive community of sexual organisms. De Queiroz approach has been criticized, however. For instance, he did not specify any properties of this “organizational level above that of the organism,” leaving the exact definition of population and, thus, lineage (and species) unclear (Hausdorf 2011). Similarly, how “subpopulations” are defined or how populations are “connected” by limited gene flow are open questions. Besides, the notion of “separately evolving” units is doubtlessly the core of all modern species concepts. However, to be a useful criterion in a species concept, it has to be specified what “separately evolving” means (Hausdorf 2011). In particular, how populations still connected by limited gene flow should be

classified? Overall, lineage-based species concepts, through certainly presenting advances, provided limited progress toward a generally applicable species concept.

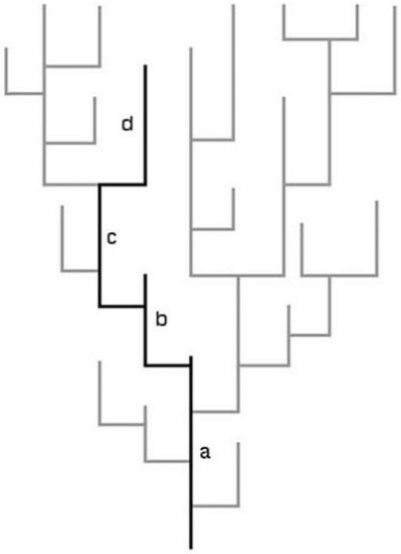


Figure 1.4 . Representation of a species (population) level lineage and its component species within a branching tree. The lineage of interest, represented by the darker lines, is made up of a series of ancestral and descendant species, labeled with the letters **a** through **d.** Note that lineage is not monophyletic in that some of its latest members (d) share more recent common ancestors with earlier members (a) than with recently diverged branches (gray lines) Form De Queiroz (2005).

One of the most important new insights with regard to the species concept was to identify that reproductive barriers are semi-permeable to gene flow and that species can differentiate despite ongoing interbreeding (Coyne and Orr 2004; Mallet 2005; Mallet 2008).More recent attempts to solve the species concept problem are worthy to mention: Wu (2001), developed the genic species concept, and Samadi and Barberousse (2006) the internodal species concept. Wu (2001) defined species as “groups that are differentially adapted and, upon contact, are not able to share genes controlling these adaptive characters, by direct exchanges or through intermediate hybrid populations.” Although it was originally formulated for biparental species, this concept is in principle also applicable to monoparental reproducing species. This concept has been criticized because it considers only differential adaptation originated by mutations in genes (Britton-Davidian 2001; Noor 2002). Other processes that might also cause the formation of species such as genetic drift should not be excluded a priori (Rundle et al. 2000; Noor 2002). In the same attempt to clarify the species concept problem, Samadi and Barberousse (2006, 2009) assimilated previous ideas related to the internodal conception of species mentioned by Simpson (1951; 1961); Hennig (1966); De Queiroz (1998) to propose a rigorous definition of the species concept. They consider an internodal species concept that formalizes the definitive divergence of evolutionary lineages. This internodal species concept represents a broader version of the Hennigian phylogenetic concept (Table 1.4) where a species disappears when splitting into new lineages (Figure 1.5) and differs from the Lineage Species Concept of De Queiroz in that it defines more precise boundaries to species (Samadi and Barberousse 2006). In a broad sense, under the internodal species concept, each speciation event gives rise to two new species originating from the ancestral one, the ancestral one differing alike from the

two new species. "'A species is thus a branch segment of the tree of life delimited either by two nodes or branching points (i.e. by two speciation events) or by a node and the end of a branch (i.e. an extinction event)'' (Figure 1.5). However Velasco (2008), pointed some pitfalls of the Samadi and Barberousse species concept. Specifically, it does not allow for the possibility of interspecies hybridization. Velasco also argues that the proposal is unworkable in practice, and also theoretically problematic because it entails that in many lineages, speciation events are taking place every few generations. Stamos (2003) introduced the 'biosimilarity concept', whereas Hausdorf (2011) sought a general solution of the species problem by introducing the 'differential fitness concept'. Richards (2010) offered a tripartite grouping of all concepts into laborious, metaphysical and scrutinizing operational groups.

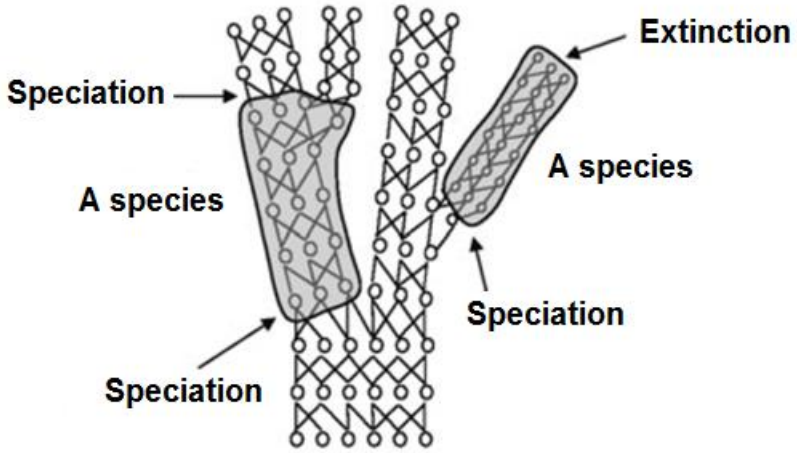


Figure 1.5. Internodal Species concept. Species are subsets of organisms connected to one another through the global genealogical network, existing either between two successive speciations or between a speciation and an extinction From Samadi and Barberousse (2006).

I.2.2. Species concepts and conservation

Trying to assess species designation can be contentious, as species are conceived as the primary units for conservation. However, it is becoming increasingly apparent that conservation should also recognize intraspecific diversity to conserve the evolutionary legacy of species (Waples 1991; Fraser and Bernatchez 2001). Intraspecific diversity can arise at the early speciation stage. If species delineation is based on properties which were acquired at the later stages, the approach may result in false negatives, that is, species in early stages of speciation being ignored. Conversely, if the properties on which species are defined are acquired prior to a speciation event, then false positives may occur. For instance, The PSC recognizes differences between two subsets of a population at an earlier evolutionary stage than

the BSC. As a result, when evaluated using the PSC, biological species are often split into smaller groups of recognizable phylogenetic species. Agapow et al. (2004) found that when groups of organisms were classified by both the PSC and other methods, under the PSC the average number of species in a group increased by 48.7%. This is called taxonomic inflation and it can have serious implications for conservation biology (Isaac et al 2004).The increased number of new species directly increases the number of endangered species. Under this reclassification resulted from the PSC species concept gives the false impression of a change in the conservation status of a particular group when in fact there has been none (Agapow et al. 2004).

Approaches for defining conservation units were thus developed to try to relate taxonomic status with conservation. The Evolutionarily Significant Unit (ESU) concept was proposed by Waples (1991), in which an ESU is a population or group of populations, that shows substantial reproductive isolation from other populations, and constitutes an important component of the evolutionary legacy of the species. In terms of conservation, it can be defined as a population unit that merits separate management and has priority for conservation (Ryder 1986). Despite this, there are appealing points to this concept (e. g. the use of diverse biological and even environmental information, to discriminate units for protection). Some authors criticized the objectivity of the approach (Moritz et al 1995) and its applicability to a broad taxonomic spectrum (Pennock and Dimmick 1997), while others noted that the definition contains subjective wording such as "substantially", and "important" that may be difficult to implement (Moritz et al 1995). The alternative approach by Moritz (1994), adopts the criteria that ESUs should exhibit reciprocal monophyly for mtDNA haplotypes and significant genetic differentiation at nuclear loci. However, one of the main caveats noted by several authors, is the criteria of reciprocal monophyly: 1. which may be feasible by several methods, that yields the most likely phylogeny in all situations (Wapples 1995) 2. which not always infer historical isolation (Crandall et al 2000); 3. and may be the most important, the fact that reciprocal monophyly can not represent recently founded populations or quickly radiated species assemblages that deserve protection based on heritable phenotypic or ecological diversity (Taylor 1999). 4. Finally, mitochondrial DNA evolves slower in certain taxa such plants (40-100 times slower than animal mtDNA) (Palmer 1992) making it limited for phylogeographic studies in plants (Coates 2000). Moritz, 1994 reviewed his precedent postulate, specifically the genetic applications that reveal a varying amount of phylogenetic separation and sometimes less separation than reciprocal monophyly. To avoid this issue, a second category, management units (MU), i.e. populations showing significant differentiation at nuclear loci, but not necessarily reciprocal monophyly at mtDNA. The MU was conceived as a level of conservation unit below, that of the larger ESU, that had statistically significant divergence in allele frequencies (nuclear or mitochondrial), no matter the phylogenetic differentiation of the alleles. The focus of the MU is on contemporary population structuring and short-term monitoring. This approach enables conservation biologists to apply molecular genetics, while at the same time being able to avoid the dilemma of determining how much genetic variation is enough for warranting protection to a given population(s) of a species (Fraser and Bernatchez 2002). To avoid this issue, Moritz (1999), proposed a two steps approach for defining conservation units. But he did not give guidelines as to how apply this step precisely.

Crandall et al. (2000) introduced a new framework for defining conservation units. Considering both reproductive isolation and adaptive divergence by accepting hypotheses of genetic exchangeability (reproductive isolation) and ecological exchangeability (essentially adaptive divergence). Moreover, this criteria is considered on both a contemporary and historical time scale. Other pros on it are i) it recognizes the importance of adaptive distinctiveness in populations ii) combines genetics and ecological principles ii) it tries to prioritize objectively conservation value across a broad taxonomic spectrum, including invertebrates and plants. However, critics are made related to the ESU designation it forces the continuous distribution of genetic diversity into just two categories (ESU or not). Finally, the Designatable Unit (DU) described by Green 2005 is closely related to Waples' ESU definition but additionally takes extinction risk into account.

Table 1.5. Evolutionary significant unit (ESU) criteria since Ryder (1986) and species concepts, from Fraser and Bernatchez (2001).

Author (s)	**Criteria**
Evolutionary significant units	
Ryder (1986)	Subsets of the more inclusive entity species, which possess genetic attributes significant for the present and future generations of the species in question.
Wapples (1991)	A population or group of populations that: (i) is substantially reproductively isolated from other conspecific population units; and (ii) represents an important component of the evolutionary legacy of the species
Dizon et al (1992)	Populations or groups of populations demonstrating significant divergence in allele frequencies
Avise (1994)	Sets of populations derived from consistently congruent gene phylogenies
Moritz (1994)	Populations that: (i) are reciprocal monophyletic for mtDNA alleles; and (ii) demonstrate significant divergence of allele frequencies at nuclear loci
Vogler and DeSalle 1994	Groups that are diagnosed by characters which cluster individuals or populations to the exclusion of other such clusters
Crandall et al 2000	Abandon the term ESU for more holistic concept of species, consisting of populations with varying levels of gene flow evolving through drift and selection.
Fraser and Bernatchez 2001	A lineage demonstrating highly restricted gene flow from other such lineages within the higher organizational level

	(lineage) of the species
Most related species concept to ESUS	
Biological Species Concept (Mayr and Ashlock 1991)	A species is a group of interbreeding natural populations that are reproductively isolated from other such groups
Evolutionary Species Concept (Simpson 1961; Wiley 1978)	An entity composed of organisms that maintain its identity from other such lineages and has its own independent evolutionary tendencies and historical fate
Phylogenetic Species Concept (Cracraft 1983)	The smallest diagnosable cluster of individuals organisms which there is a parental pattern of ancestry and descent
General Lineage Concept of Species (de Queiroz 1998)	Species are segments of population-level lineages

Most evolutionary biologist now agrees that species are evolving as separate lineages of populations or metapopulations (Padial et al. 2010). However, discordance remains about where, along the divergence continuum, separate lineages should be recognized as distinct species. After 250 years of the predominance of comparative morphology in species discovery, new methods (mainly molecular) are now used in taxonomy and systematics, and more recently, "integrative taxonomy" has emerged (Dayrat 2005; Will et al. 2005; Valdecasas et al. 2008; Schlick-Steiner et al. 2010).

I.2.3. Methods for delimiting species

After giving a broad overview of the different concepts and criteria for the current understanding of species, the question arises in which way species can be delimited. Below, I present the methods and approaches for delimiting species covering genetic, genomic and morphological approaches to finish with the emergent integrative taxonomy criteria for identifying and delimiting species (Carstens et al. 2013b; Pante et al. 2014a; Seehausen et al. 2014; Pante et al. 2014b; Pompanon and Samadi 2015).

I.2.3.1. Methods for species delimitation applied in molecular genetics and coalescent-based analyses

DNA-based species delimitation is a relatively new approach for evaluating whether populations of organisms are isolated from one another on an "evolutionary timescale" thus supporting the hypothesis that they may be distinct species (Yang and Rannala 2010). Most current species delimitation methods require *priori* assignments of individuals to populations (Table 1.6, Carstens 2013). The delimitation methods then determine the level of support for different nodes (splits) as defining distinct species rather than simply populations which have some admixture, or a relatively short history of complete isolation. Current research using DNA sequences follow a three steps process to delimit and to assess species delimitation: (1) Population assignment; (2) The inference of phylogenetic relationships; (3) The evaluation of statistical support for species delimitation. The division of the

delimitation process into these three steps is done for pragmatic reasons. It is simply not computationally feasible to jointly infer population assignment, phylogeny and species delimitation using currently available approaches (Yang and Rannala 2014). Below, I describe the ruling methods using genetic sequence data following Carstens (2013); Heled and Drummond (2010); Yang (2015).

Table 1.6. Species delimitation methods using genetic sequence data and coalescence-based analyses. From Carstens (2013); Heled and Drummond (2010); Yang (2015).

I.1. Genetic clustering approaches

Operates by identifying the population assignments and level of clustering that minimizes Hardy-Weinberg disequilibrium (Pritchard et al. 2000; Huelsenbeck et al. 2011). While clustering algorithms do not explicitly model population parameters, such as migration and rates or population divergence, they are flexible in terms of the data that are required and have been applied to a wide range of systems (Olivieri et al. 2008; Rodríguez et al. 2013) . Some shortcomings are signaled as the clustering algorithms to cluster samples does not provide a history of phylogenetic divergence because the population structure is inferred without consideration of historical patterns of diversification, and there is not a clear correspondence between a given level of clustering and branching pattern of a species tree Jackson and Austin (2010). These limitations made that most of the investigations take clustering approaches as a starting point for species delimitation complementing with methods that explicitly model population divergence.

Structurama Huelsenbeck et al. (2011)

Implements the clustering algorithms first described by Pritchard et al (2000) for their program Structure that clusters samples into populations by minimizing Hardy- Weinberg disequilibrium for a given partitioning level. Nearly any type of genetic data can be input into structurama, and the program can assign individuals to a population with or without the admixture.

Gaussian Clustering (Hausdorf and Hennig 2010)

Groups samples into populations using genotypic data by searching for clusters that can be attributed to being a mixture of normal allele frequency distributions. Like Structurama this method is flexible in terms of the data that can be analyzed. As in other clustering approaches, temporal divergence among putative groups is not explicitly estimated.

The general mixed Yule-coalescent model (GMYC) Monaghan et al. (2006)

Takes an ultrametric genealogy estimated from a single genetic locus as input. The method attempts to model the transition point between cladogenesis and allele coalescence by utilizing the assumption that the former will occur at a rate far lower than the later. This results in a shift in the rate of branching of the genealogy that reflect the transition between species-level processes (such as speciation and extinction) and population-level processes (allele coalescence). Reid & Carstens (2012) proposed a version of the GMYC that accounts for phylogenetic uncertainty gene tree estimates using a Bayesian analysis. Both implementations of the GMYC are likely to delimit well-supported clades of haplotypes as independent lineages and as such may be prone to over delimitation.

Joint demography and assignment (JDA) Choi and Hey (2011)

Methods for jointly estimating population assignment along with the parameters of an isolation-with-migration model. Joint demography and assignment (JDA) is applicable to an island or two population models, while joint demography and assignment of population tree (JDAP) is applicable to more than two diverging populations. Each takes sequence data as input and is implemented within IMa2 Hey and Nielsen (2007).

O'Meara's heuristic method O'Meara (2010)

This method takes gene trees from multiple loci as input and operates under a similar assumption to the GMYC (namely that allelic coalescence occurs more rapidly than speciation). Provided that this assumption is true, the longest branches of gene trees are likely to represent species- level differences, and thus, congruence across loci is indicative of both the species tree and the population assignments

II. Methods requiring a priori population assignment

II.1. Multispecies coalescent model (MSC)

Compared with traditional phylogenetic analyses, which assumes that the same tree underlies all gene loci, the MSC accounts for the coalescent process in both the modern and ancestral species and the resultant gene tree conflicts. Thus, a reliable estimation of the species phylogeny is possible even if the information at every locus is weak so that the gene tree is highly uncertain (Heled and Drummond 2010). Multispecies coalescent models offer an important shift over the conventional methods of phylogenetic inference related to what constitutes an operational taxonomic unit (OTU) and how genetic data can be used to delimit lineages. First, rather than using a single individual or several representative individuals as exemplars, the OTUs are explicitly evolutionary lineages with multiple samples contained within each lineage. Secondly, the species tree paradigm enables the relationships among lineages and the membership of individual samples in these lineages to be evaluated in a rigorous statistical framework (Edwards 2009). These models for species delimitation compare the probability of trees with differing numbers of OTUs to identify optimal partitions of the data (e.g. spedeSTEM, BPP). Satler et al. (2013)Salter et al 2013 extend this strategy to its maximum extent by calculating the probability of the phylogeny that treats individual samples as putative lineages. The putative lineages are then sequentially collapsed on the basis of which samples are most closely related, the probability of the species tree is recalculated, and information theory (Burnham & Anderson 2002) is used to identify the optimal model of lineage composition. Thus, spedeSTEM discovery can be used to simultaneously delimit evolutionary lineages and assign samples to these lineages.

Bayesian Phylogenetics and Phylogeography (BPP) Yang (2015)

The Bayesian Phylogenetics and Phylogeography is a Bayesian Markov chain Monte Carlo (MCMC) program for analyzing DNA sequence alignments under the multispecies coalescent model (MSC) (Rannala and Yang, 2003; see also Takahata et al., 1995; Yang, 2002). The method takes sequence data as input and also requires the user to define the topology of the species tree. Given this information, the algorithm implemented in BPP then traverses the parameter space to compute the posterior probability of the proposed nodes of the species tree. While inaccurately specified guide trees can lead to false-positive delimitations, the accuracy of BPP does not generally appear dependent on its ability to estimate gene trees.

Star-BEAST (Bayesian evolutionary analysis by sampling trees) Heled and Drummond (2010)

This method provides a joint inference of a species tree topology (ancestral relationships), divergence times (the times ancestral species separated into two species), population sizes (population sizes history for each species) and gene trees from multiple genes sampled from multiple individuals across a set of closely related species. This software is an extension of previous BEAST , Bayesian phylogenetic inference (Drummond and Rambaut 2007). The *BEAST approach considers species not necessarily as a taxonomic rank but designates any group of individuals that after some divergence time, have no history of breeding with individuals outside that group. A species tree defines barriers for gene flow, and so the term is a catch-all for taxonomic rank, subspecies or any diverging population structure.

SpedeSTEM (Carstens and Dewey 2010)

This method was developed to test species boundaries in a system with subspecies taxonomy. Under this approach, the probability of the gene trees given the species tree for all hierarchical permutations of lineage grouping is computed. The accuracy of this approach is dependent on the quality of the gene tree estimates.

I.2.3.2. Methods for species delimitation based on next generation sequencing and speciation genomics

Speciation genomics is defined as the field of speciation research that addresses the influence of genomic properties on the evolution of reproductive barriers and the signatures of speciation processes that are observable in genomic patterns (for example, processes of diversity and divergence). Its aim is a conceptual and methodological integration of genomic approaches with other empirical and theoretical speciation research.

Table 1.7. Next generation sequencing and genomic tools for studying speciation.

Next generation sequencing and genomic tools

Patterns of genomic divergence

Several methods can be used to investigate genome-wide divergence along the "speciation continuum". These methods include genome scans using single-nucleotide polymorphism (SNP) arrays, restriction-site associated DNA sequencing (RAD-seq) or related genotyping by sequencing (GBS) methods, whole exome or transcriptome sequencing and whole-genome resequencing of population samples. Patterns in genome-wide divergence can be visualized and compared using, for example, F_{st} kernel density plots and Manhattan plots.

Testing for signatures of introgression

Various approaches are available to assess whether the sharing of genetic variants between incipient species is a result of hybridization or incomplete lineage sorting. The ABBA-BABA test is particularly applicable to genome-scale data sets. It relies on the frequencies of two specific patterns of allele sharing among a group of four species.

Identifying signatures of selection

Genome scans can reveal genomic regions that show evidence of divergent selection between incipient species using Fst outlier analyses or related approaches, which can be applied either to individual SNPs or to smoothed average Fst values within windows (that is regions of a defined size) of the genome. The latest methods can account for demographic variation and other sources of variation and make improved use of high density marker information.

Mapping genes that are involved in reproductive isolation

A logical first step in the search for candidate genes that are involved in reproductive isolation is to carry our genome scans of incipient species pairs at several different stages along speciation continuum. A range of genetic mapping tools are available for identifying links between divergent genomic regions and the phenotypic traits that contribute to reproductive isolation. Quantitative trait locus (QTL) mapping is one of such method that is powerful for doing so. In short, a genome-wide set of markers is genotyped in a phenotypically variable population that has know pedigree data, and statistical associations are identified between the genetic markers (in this case, QTLs) and phenotypes of interest (in this case, traits related to reproductive isolation). With functional information on genes that are in the vicinity of a QTL, candidate reproductive isolation genes can be identified.

Admixture mapping

If pedigree data are not available, then it is possible to take advantage of the phenotypic and genetic differences between hybridizing taxa and use admixture as the basis to genetically map phenotypes that contribute to reproductive isolation using samples from wild hybrid populations. Both intrinsic and extrinsic postzygotic barriers involve alleles that are selected against hybrids and various methods can be used to identify such alleles in hybrid zones or in other situations in which admixture occurs. Genomic cline analysis is one such method that can identify candidate reproductive isolation loci with low levels of introgression relative to most of the genome.

Manipulative selection experiments

Both Quantitative trait locus QTL and admixture mapping have an unfortunate bias towards detecting loci of large effect. Alternatively, alleles that affect fitness and reproductive isolation can be located using manipulative selection experiments, which track allelic changes or genome-wide responses. Estimates of these effects can be ascertained by measuring selection and introgression in the wild. So far, few studies have taken this approach and none has measured effects on reproductive isolation.

Gene expression studies

To further investigate the functional importance of candidate loci that are involved in reproductive isolation, expression QTL (eQTL) analysis can be useful. it identifies genomic loci that regulate expression levels of mRNAs. Systematically generated eQTL information can provide insights into the mechanism that underlies reproductive isolation in regions that have been identified through genome-wide association studies, and such eQTL information can help to identify networks of genes and the role of genetic interaction (including epistasis in Bateson-Dobzhansky-Muller incompatibilities) in reproductive isolation

.

From (Pante et al. 2014a; Seehausen et al. 2014).

I.2.3.3. Methods for species delimitation based on morphology

Hohenegger (2014) proposed a method for delimiting species based on frequency distributions of morphological characters considering three aspects, all related in some way to the criterion of homogeneity. The combination of these factors within a three-dimensional coordinate space helps to recognize morphologically environmentally homogeneous groups. In such cases, the consideration of ontogenetic cohesion leads to the recognition and delimitation of species without molecular genetics investigation.

Table 1.8. Aspects based on Hohenegger (2014) method for species delimitation under morphological criteria.

Character homogeneity (traits)
Homogeneity can be tested Homogeneity can be tested in morphology based on a single or few characters, but should comprise, under ideal conditions, the maximum of measurable morphological characters in order to detect diagnostic characters separating several homogeneous groups by descriptive statistical methods (e.g. discriminatory analyses). A multitude of morphological characters have to be treated simultaneously for testing shape homogeneity.
Ontogenic cohesion
Based on some complications on the determination of ontogenetic cohesion, such as the acquisition of similar adult shape by different ontogenetic pathways, morphologies between adults or late ontogenetic stages is insufficient to differentiate species. The ontogenic cohesion method propose comparing morphogenesis (a biological process that causes an organism to develop its shape) between individuals and detecting homogeneities in these functions in order to separate species that are similar to adults, but different in their ontogeny.
Niche homogeneity
Every species is assigned to intervals along environmental gradients. These intervals represent the ecological niches, where the fundamental niche is determined as the species niche in the absence of any interspecific competition, thus representing its physiological capability.

I.2.3.4. Integrative taxonomy

Species identification, species limits and species discovery have been revived within the modern systematics (e.g., Sites and Marshall 2003; 2004; Mace 2004; Dayrat 2005; Weisrock et al. 2010; Camargo and Sites 2013; Carstens et al. 2013; Hotaling et al. 2016). But still, evolutionary biologists have great difficulties in recognizing species at the very early stages of divergence, due to the limited time for differences to evolve (de Queiroz 1999). Identifying lineages in the early stages of

species divergence is, nonetheless, extremely important because the study of these lineages is expected to be the most informative about the speciation process (Wiens 2004). Conversely to when species are old and well-differentiated (the “easy” part of Figure 1.3) where methods and data concur to the same species delimitation, young species in the "*grey zone"* raise conceptual and empirical difficulties. Given that multiple strategies for delimiting species using the same data set co-occur, leading to different conclusions (Wiens and Penkrot 2002; Sites and Marshall 2003; 2004; Marshall et al. 2006; Wolf et al. 2010; Puillandre et al. 2012). Time has come for integration of population biology, phylogenetics, and other evolutionary disciplines into taxonomy (Sites and Marshall 2004; Glaw et al. 2010; Bochkov et al. 2014; Pante et al. 2014a; Pante et al. 2014b; Hotaling et al. 2016; Smith et al. 2016). Indeed, what matters for the study of speciation matters for taxonomy as well, and a better integration of theory and methods appeared necessary. Dayrat (2005) proposed such a methodological framework, known as integrative taxonomy (Figure 3), an approach based on hypotheses to support species delimitation using all possible sources of evidence. Integrative taxonomy consists in analyzing different characters, with different methods, and applying different criteria of species delimitation to propose species hypotheses that are as robust as possible. Although the formal description of Integrative Taxonomy is relatively recent, the practice of combining different kinds of information (e.g., morphological, behavioral, molecular), is indeed not new (e.g., Forister et al. 2008). Dayrat (2005) and other authors already proposed an organized framework for integrating molecular methods and approaches of population genetics within the classical taxonomic procedures (but see Valdecasas et al 2008). Still, disagreements related to the type of characters and the degree of congruence that different characters must show to consider a population or a group of populations as a separate species is hotly debated. Two schools were actually developed, "integration by congruence" and "integration by cumulation" (Figure 1.7).

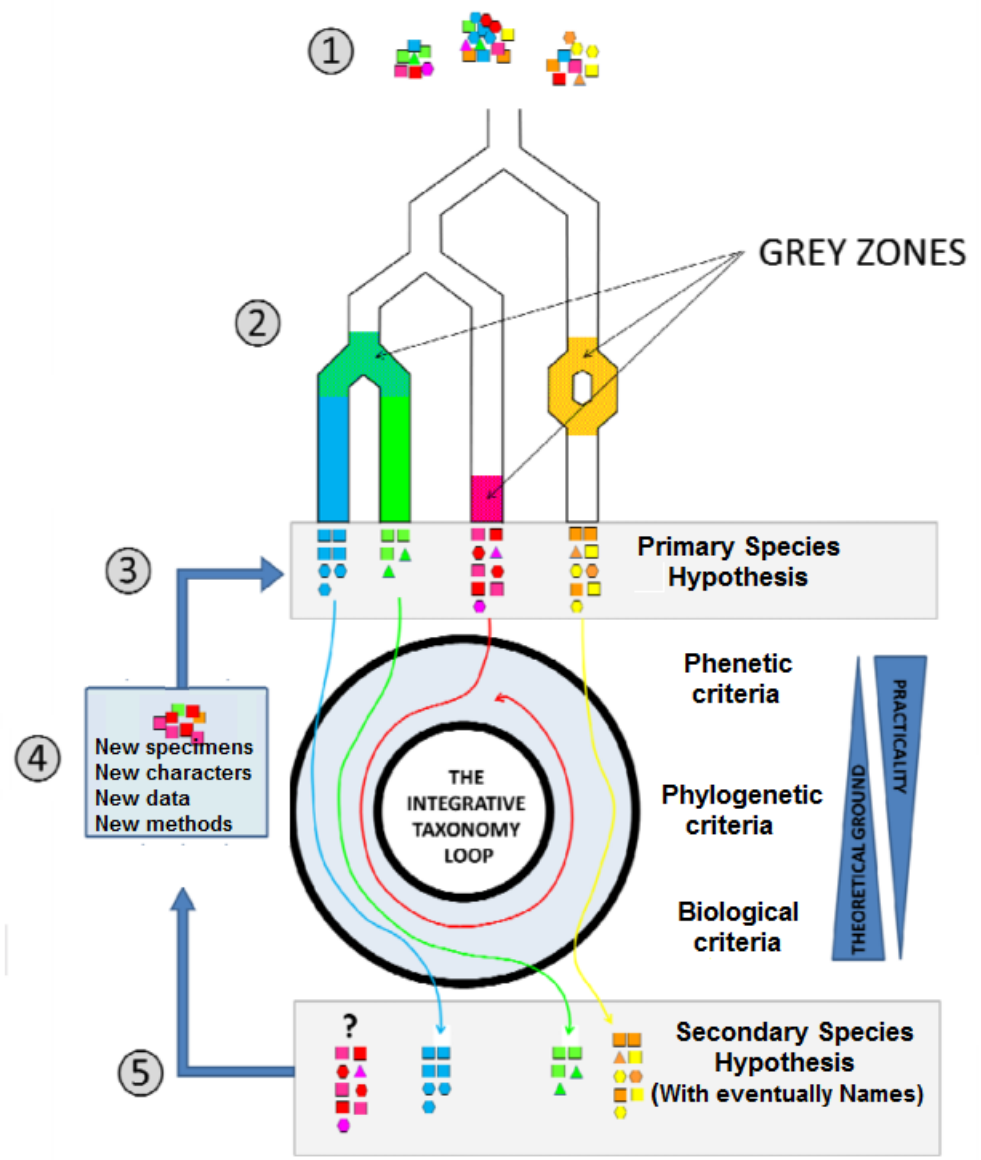


Figure 1.6. The integrative taxonomy loop. The different criteria (see Samadi and Barberousse (2015) are listed right of the loop, with the more theoretically grounded displayed on the bottom, and the more operational (i.e. easy to test practically) on the top. For each kind of criteria and characters (i.e. morphology, behavior, ecology, biochemical, genetic, etc.) and methods (i.e. distances, maximum parsimony, maximum likelihood, population genetics inferences, crossing experiments, observations, etc) may be applied. The different steps are as follows. 1. Population and phylogenetic sampling. 2. Sampled species may be highly differentiated (blue and green); recently diverged species that are still in a "grey zone" with most characters undifferentiated (pink and red); or a single species that went through a temporal split into several temporary lineages (yellow/orange). 3. Primary Species Hypotheses (PSH) are proposed, for example using morphology or a single molecular marker. 4. PSH are engaged in the integrative taxonomy loop and are evaluated, possibly with the addition of new material, using different criteria for species delimitation. The more theoretically-grounded biological criteria can be tested directly using cross experiments, or indirectly with unlinked markers, and complemented with more operational criteria. 5. When possible, taxonomic decisions are taken by turning PSH into Secondary Species Hypotheses – SSH and are named. Some lineages (i.e. the pink/red lineage) may stay in the loop, needing more conclusive data before being turned into SSHs. Most of the

literature and methods for species delimitation focus on species that are currently in the grey zone (cf. (Carstens et al. 2013) even though most delimitation cases fall outside of this range from Pante et al. (2014b).

I.2.3.4.1. Integrative taxonomy by congruence

Integration by congruence is based on the rationale of the phylogenetic species recognition, under the assumption that concordant patterns of divergence among several taxonomic characters indicate full lineage separation. Taxonomists indeed consider highly improbable that a coherent pattern of character concordance will emerge by chance. One major advantage of the congruence approach is that it promotes taxonomic stability: most taxonomists will agree on the validity of a species supported by several character sets, as long as it is clear that they are unlinked and fixed. But the major limitation is the risk of underestimating species numbers because the process of speciation is not always accompanied by character change at all levels, and therefore, there may be a bias toward older species (i.e. those that diverged in the past which will have an increased probability of showing complete gene lineage sorting and reciprocal monophyly for many loci). Thus, recent radiations may be lost by a strict consensus approach (Shaffer and Thomson 2007). For example, in Darwin's finches or some lineages of cichlid fishes, groups of closely related species show striking morphological differences that originated through fast divergent selection associated with ecological transitions but show weak reproductive isolation, low genotypic clustering, and little neutral genetic differentiation (Presgraves 2010).

I.2.3.4.2. Integrative taxonomy by cumulation

The framework of integration by cumulation is based on the assumption that divergences in any of the organismal attributes, that constitute taxonomic characters, can provide evidence for the existence of a species. Thus this approach defends the view that because all taxonomic characters are contingent in existence, the order of appearance, and magnitude of divergence during speciation.The only way for true integration is allowing any source of evidence (even a single one), to form the basis for species existence. Under this approach, congruence is desired, but not considered necessary. A major advantage of this approach is that it does not bind species delimitation to the identification of any particular biological property. Taxonomists can thus select and focus on the most appropriate set of taxonomic characters for each group of organisms. Furthermore, cumulation is probably the most suitable framework to uncover recently diverged species (Shaffer and Thomson 2007). However, limitations have also been warned, e.g. the cumulation approach may possibly overestimate species numbers when using a single line of evidence (e.g. a single locus of mtDNA). For example, the Bornean and Sumatran populations of the clouded leopard *Neofelis nebulosa diardi,* demonstrated the action of genetic drift on reduced populations, isolated over a short period of time resulting in signals of reciprocally monophyletic for some diagnosable characters, with full species status for some authors but not all (Almstrong et al 2008).

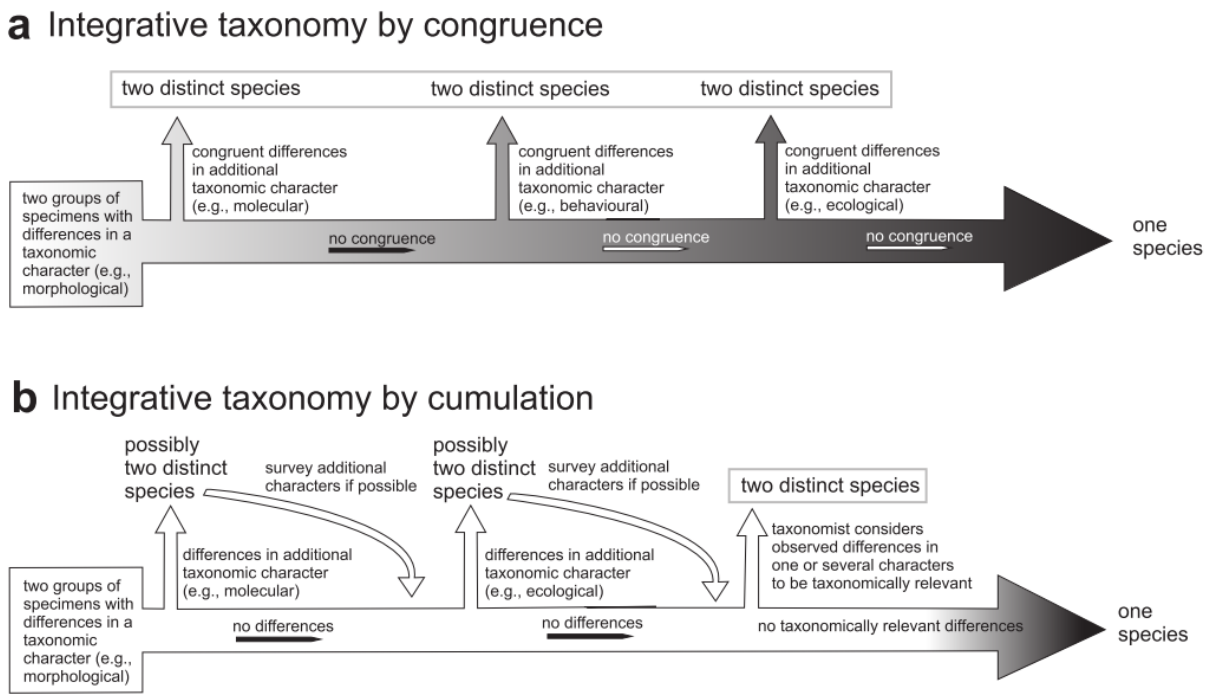


Figure 1.7. Schematic representation of work protocols in taxonomy. Workflow in (a) integrative taxonomy by congruence and (b) by cumulation. Modified from (Padial et al. 2010).

I.3. Phylogeography

Phylogeographic analyses are a common approach in molecular ecology, connecting historical processes in evolution with spatial distributions that traditionally scale over millions of years (Knowles and Maddison 2002). Phylogeography studies the "historical, phylogenetic components of the spatial distribution of gene lineages, especially within and among closely related species" (Avise 2009). It refers to the principles and processes that govern the geographic distributions of genealogical lineages, including those at intraspecific level (Avise 1994). Time and space are jointly considered axes of phylogeography onto which are mapped particular gene genealogies of interest. The aim is to analyze how phylogenetic relationships of genealogical lineages are distributed across their geographical range. Thus phylogeography is a critical juncture among several microevolutionary and macroevolutionary disciplines (Figure 1.8).

1.3.1. Phylogeography and molecular tools

The field of phylogeography rose to prominence in the late 1980s, with the adoption of molecular data, specifically mtDNA, to infer patterns of genetic variation interpreted in a geographical context, that made possible to get insights about different roles of bottlenecks population expansions, vicariance events and gene flow, in structuring genetic variability (Avise 2000; Knowles and Maddison 2002). Basic phylogeographic analyses typically consist in the estimation of a tree using phylogenetic methods, or possibly the estimation of a network (Avise 2009; Nielsen and Beaumont 2009). The branches of the tree are then related to historical events in a geographical or biogeographic context. For example, the emergence of a clade

only existing in a particular area may be interpreted as evidence of a historical event, by which a population, or group of individuals, separated became isolated. In general, a basic premise for phylogeography is that the branches on the tree can be interpreted as evidence of the occurrence of specific historical demographic events in a biogeographical context (Nielsen and Beaumont 2009). Gene trees can be constructed from a variety of different genetic analyses, either through Mitochondrial DNA (mtDNA) or Nuclear DNA (nuDNA) data or from a combination of both (Avise 2009). Ideally, phylogeographic analyses would use gene trees based on a combination of both mtDNA and nuDNA. Over the last two decades mtDNA has been the most used molecular tool by geneticists due to its interesting properties such as a) availability, rapid, easy and inexpensive to sequence, b) non-recombining properties c) high mutation rate, d) small effective population size (Ne) leading the lineage sorting occurring faster compared to the nuDNA (4 Ne), allowing the detection of recent vicariance events (Brown 1985; Avise et al. 1987; Zink and Barrowclough 2008). Nonetheless, mtDNA alone provides no information on the genetic structure of loci, which are biparentally inherited (Moritz et al. 1987; Avise 1998; Avise 2009). Indeed, several studies have revealed that mtDNA information alone, may be incomplete data on populations structures or assessing species boundaries unless it is reinforced by additional evidence such as nuclear gene data (Edwards et al. 2005; Rubinoff and Holland 2005; Welch et al. 2011a). Actually, since all genes in the mtDNA evolve as a single linkage unit, mtDNA will always concur into a single gene tree, which may misrepresent the whole evolution history of the taxa (Ballard and Whitlock 2004). Furthermore, mtDNA reflects just one window of evolution, a matrilineal history that could well differ compared to the overall population history, if the divergence was led by a sex-biased dispersal (Zink and Barrowclough 2008). In addition, it is important to remember that different types of loci have different effective population sizes and therefore have different durations before complete genetic divergence occurs between isolated populations. Indeed mean coalescent (see below for definition), time for biparentally inherited diploid nuclear loci, is generally fourfold longer than the mean coalescent time, for uniparentally inherited haploid mitochondrial loci (Avise 2000; Avise 2009). Therefore, complete lineage sorting for mtDNA occurs four times faster than for nuDNA, and for two recently diverged populations, they may be perceived as reciprocally monophyletic for mitochondrial gene but not for nuclear genes (Zink and Barrowclough 2008; Barrowclough and Zink 2009; Edwards and Bensch 2009). Since additional markers are constantly being developed, especially nuclear loci, multilocus sequence data have led to the availability of multilocus approaches to phylogenetic inference, especially in avian species (Edwards 2009; Kimball et al. 2009). Indeed, evolutionary trees from different genes often have conflicting branching patterns (Nichols 2001).

I.3.2. Statistical phylogeography and coalescent-based demographic parameters modeling

Phylogeographic studies are based on statistical inferences advances in coalescent theory and computational statistics have provided such framework (Slatkin 1987; Griffiths and Tavare 1994; Kuhner et al. 1995; Templeton et al. 1995; Wakeley and Hey 1997; Beerli and Felsenstein 1999; Nielsen and Wakeley 2001; Beaumont et al. 2002). An increasing number of phylogeography studies rely on coalescent-based methods for estimating relevant population parameters such as effective population size, migration rates, divergence time, time elapsed since the onset of a demographic expansion, providing insights into the demography of speciation (Rosenberg and Nordborg 2002; Hey and Nielsen 2004; Carstens and Dewey 2010; Hey 2010; Heled et al. 2013; Ogilvie et al. 2015; Yang 2015). The coalescent theory has flourished thanks to the development of the theory of neutral evolution of nucleotide sequences. To coalesce means to merge or to join. The coalescent theory, also known as Kingman's coalescent, was developed in the early 1980s (Kingman 1982; Hudson 1983; Nei and Tajima 1983) and has been critical in the progress of population genetics. It underlies modern computational methods for the comparative analysis of genetic data, in particular, genomic sequence data, including both polymorphism data from the same species and divergence data from closely related species. Under the phylogeographic/ phylogenetic framework, coalescent refers to tracing the genealogical relationships of a sample of chromosomes (or genes), from a particular genomic region backwards in time over the generations. In that way, the lineages join or coalesce when they meet their common ancestors. In its broader sense, the coalescent is the genealogical process of joining lineages when one traces the genealogy of the sample backwards in time (Yang 2014). In addition, under the coalescent theory, the descent of any set of individuals (DNA sequences) can be traced back, with common ancestry denoting coalescent events. Under the assumption of neutral evolution, the probability in any generation that a DNA lineage will give rise to two distinct daughter lineages is the same for all lineages. Coalescent theory provides a mathematical framework which describes the distribution of gene trees in populations. It can be used as a mathematical tool for deriving theoretical population genetic parameters, and also more directly, to connect demographic models with gene trees. With the advent of novel computational techniques like Markov chain Monte Carlo (MCMC), jointly with bayesian multi-locus genetic sequence data approaches, it has become feasible to estimate parameters of interest. Demographic inference is a well-known example of genealogy-based population genetics that benefited from these advances (Gill et al. 1993; Wakeley and Hey 1997; Drummond et al. 2003; Drummond et al. 2005; Gill et al. 2012). Under this inference, the genealogy of a DNA sample can be traced backwards in time. Every pair of lineages coalesces at the rate 1/2N. If the population size (N) has been changing in the past, lineages will coalesce faster when N is small and more slowly when N is large. Thus the shape of the genealogical tree and the distribution of node ages on the tree will reflect the demographic history of the population. For example, a sample taken from a population that has been expanding rapidly will tend to have a star-shaped genealogy, with coalescent events pushed towards the root. Multi-locus genetic sequence data, by providing information about the genealogical trees, can be used to infer past demographic history of the species. In this way, the number and timings of the bottlenecks, as well as the timing and rate of expansion, can be traced back in time, which are of great interest for the

understanding of the origin and migration history of the species. Bayesian reconstruction methods have also enabled further generalization of the coalescence probability analysis, by removing the necessity to fix the Markov model parameters to obtain ancestral states and the necessity to specify a fixed tree topology with known branch lengths. Bayesian inference, however, requires prior probability distributions to be known for all aspects of the model (Sanmartín et al. 2008). Bayesian MCMC methods also made possible ancestral state reconstruction while simultaneously accounting for both phylogenetic and mapping uncertainties. Phylogeographic analysis has further benefited from integrating spatial, temporal and demographic inference (Ronquist et al. 2004). For instance, simulations of demographic parameters of isolation can test whether the observed data best fit a model of genetic isolation among populations, gene flow between populations, or even isolation of populations in presence of migration (e.g. Migrate-n; IM, IMA, IMA2 (Hey 2007; Beerli and Palczewski 2010; Hey 2011; Gaggiotti 2011), models that can estimate the effective population sizes, the dynamic size changes of population and the exponential growth rate of a population (e.g. BSP, EBSP, FLUCTUATE (Kuhner et al. 1995; Kuhner et al. 1998; Heled and Drummond 2008; Shapiro and Ho 2011).

I.3.2.1. A Special Case: Distinguishing levels of polyphyly as signals of either contemporary gene flow or incomplete lineage *sorting* using statistical phylogeography

Reciprocal monophyly has been treated as a touchstone of significant genetic divergence for determining species limits (Donoghue 1985; Zink and McKitrick 1995), conservation priorities (Moritz 1994), and phylogeographical history (Avise 2000). However, reciprocal monophyly represents just one stage on the continuum of genetic divergence, that populations undergo following isolation, and may have no evolutionary significance (Rosenberg and Nordborg 2002; Baker et al. 2003). Consequently, each stage of this continuum is informative, and as studies include larger sample sizes and more sophisticated analyses, it is useful to evaluate all stages in this process. Yet theoretical models of speciation, including coalescence theory (Neigel and Avise 1986; Avise et al. 1987), predict patterns of non-monophyly at early stages of speciation (see also Patton and Smith 1994; Crisp and Chandler 1996). There are a number of intermediate stages that are seldom explored. For instance. Omland et al. (2006) labeled these as ‘intermediate polyphyly, which is represented in three separate stages (See Table 1.1; Figure 1.8). First, new mutations begin to occur in each of the populations, generally being not shared. Common ancestral haplotypes at internal nodes in networks will be shared, but novel haplotypes will generally not (Castelloe and Templeton 1994). Secondly, after enough time, genetic drift will result in the loss of the ancestral haplotypes in one population, and then eventually both populations, would not share haplotypes between them. Note that at this stage, haplotypes in a given population may have their closest relative in the other population, so the haplotypes can still be completely inter-mixed. Thirdly, after a given time, more and more haplotypes will go extinct, so that most haplotypes will have as their closest relatives other haplotypes from the same population. Phylogenetically, these haplotype groups can still be largely intermixed between the two populations. Like any pattern controlled by stochastic lineage sorting, the order of these stages may sometimes be different, they may not be mutually exclusive, and their durations will not be precisely predictable.

Table1.9. Modified from Omland et al (2006). Typical stages of population divergence (for one locus, e.g. mtDNA). These stages assume divergence caused when a single population is divided by a vicariant event or when a large number of founders disperse across a barrier.

Rampant polyphyly (Haplotypes completely intermixed)

A. Panmixia. Initially both populations have indistinguishable haplotype frequencies.
B. Frequency differences. Haplotype frequency differences, many haplotypes shared.

Intermediate polyphyly (haplotypes start to sort)

Stage1: Neotypy (novel mutation). Only common ancestral haplotypes are shared; new haplotypes are created by mutations and are not shared by population (e.g. Baker et al, 2003; Omland et al 2006)
Stage2: Allotypy (local fixation). No haplotypes shared between populations (Hare et al 2002; Peters et al 2005)
Stage3: Allophyly (deeper fixation). Populations from two groups generally in monophyletic subclades, although samples from two populations are still intermixed across the tree (e.g. McCracken et al 2001; Fig. 4b in Hare et al 2002).

Simple paraphyly (one group a monophyletic subset of the other)

One population forms a monophyletic group that is genetically nested within the other group

Reciprocal monophyly (haplotypes completely sorted)

A. Sample monophyly. All individuals sampled within each population share a more recent common ancestor with individuals in their own population.
B. Group monophyly. Verifying reciprocal monophyly of the entire two populations would require sequencing all individuals in each of the two populations. Thus this yes/no criterion probably has seldom been verified for populations or species.

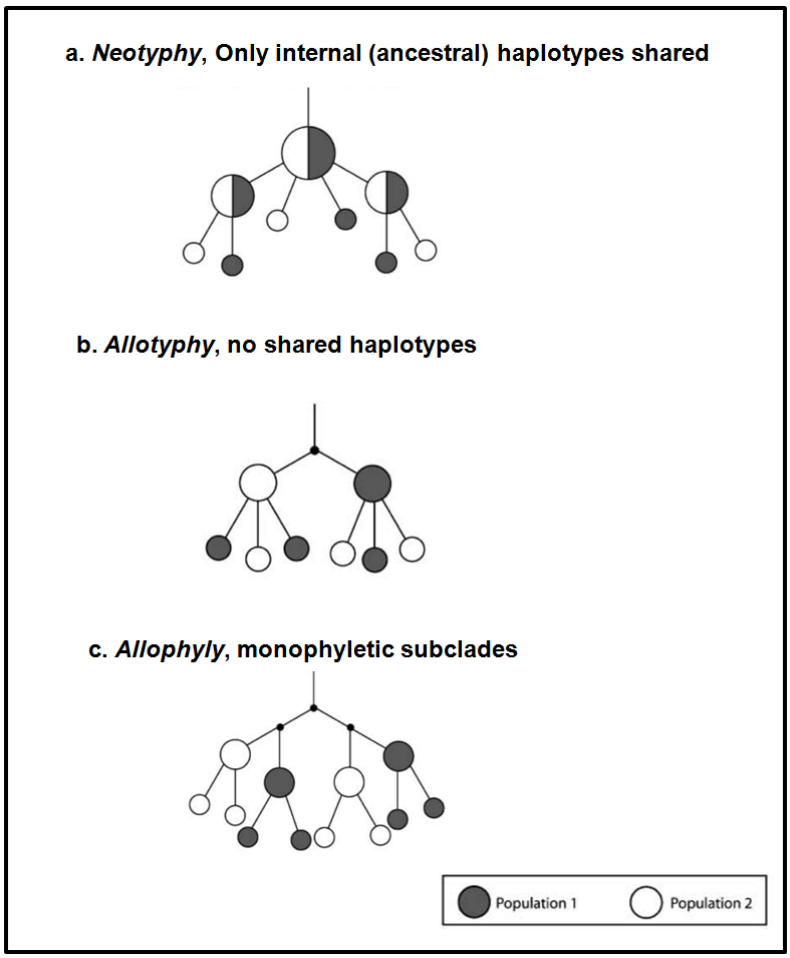


Figure 1.8. Haplotype networks showing the temporal progression in the stages of intermediate polyphyly. These stages occur as lineage sorting proceeds in two populations from panmixia to reciprocal monophyly (see also above Table 1.9). Modified from Omland et al 2006.

Uncovering intermediate patterns of divergence between populations may have a huge significance for our understanding of divergence processes (Weisrock et al. 2010). Why, when and how often divergence occurs at earliest stages of differentiation is a central question concerning speciation within clades. Diverging populations, as living evolutionary laboratories, serve as a good model to look for clues in order to disentangle questions, related to the timing of divergence, such as when the basal split took place, as well as the factors and scenarios leading to divergence. In evolutionary studies of diverging populations, a key question is the role that gene flow plays in the divergence process (Felsenstein 1981; Rice and Hostert 1993; Barton 2001). It has been thought that the divergence process happens without any gene exchange and thus the divergence of populations from the ancestral one was necessarily defined by the cessation of gene flow (Nielsen and Slatkin 2000). However, the recent availability of multi-locus datasets and the development of new population genetic methods have found signals of gene flow during species formation, calling into question the conventional view that gene flow is absent during speciation (Hey 2006). A general explanation of this finding is the fact that divergence can occur at some genes, even if there is gene exchange for other genes (Hey 2006). Such discrepancy has been explained by *heterogeneous genomic*

divergence during the process of population divergence and speciation, during which genetic differentiation accumulates in some regions, while the homogenizing effects of gene flow or inadequate time for random differentiation by genetic drift precludes divergence in other regions (reviewed in Nosil et al. (2009).

Under the selective neutrality scenario, moderate levels of gene flow can prevent divergence. Thus if divergence has occurred despite gene flow, it may suggest that natural selection is driving the divergence process (Wilson et al. 2013). Gene flow can be estimated at different temporal scales, e.g. since initial population divergence or in the past generations (Isolation with migration (IM), (BayesAss approach). The magnitude of these parameters (time of divergence, gene flow, and its timing) can be used to test alternative models of speciation. For example, the allopatric scenario is characterized by the following expectations: no gene flow (i.e., its estimation not significantly different from zero), and gene flow older than divergence time. Different scenarios are expected under sympatric and parapatric speciation: in the former, we expect a wide posterior distribution of gene flow timing placed prior or near the divergence time, and high rates of gene flow. In parapatric scenarios, a wide posterior distribution of gene flow timing placed prior or near the divergence time is also expected, however, the rates of gene flow are lower than those found in sympatric scenarios (Gaggiotti 2011).

The Thesis Objectives

The first part above was a broad overview of the different theoretical issues that I used to conceive this thesis principal emphasis, i.e. how the understanding of the species concepts and the modes of speciation have progressively evolved along with model of ecological speciation by sympatry, and the attempt to conceal concepts and criteria to define species by means of the integrative taxonomy framework. I also introduced the phylogeographic and the evolutionary demography approaches as complementary tools to taxonomy to explain the spatial separation of populations and the demography of speciation. Here I will introduce the aims of this thesis (the next chapter will present the study models). My thesis aimed at using Statistical phylogeography and coalescent-based demographic parameters modeling in order to analyse and infer the sequential and gradual divergence in population differentiation at the very early stages of the speciation process. Below I detail some of the major steps in order to reach these objectives and present the thesis aims.

Aims of the Thesis

The first aim of this thesis was to assess species boundaries within the *P. leucoptera* complex (see below for a detailed presentation of the complex). Following Dieckmann et al (2004), I predict that the geographically separated and morphologically distinct populations are at early stage of ecological speciation. Under this stage, it is expected that most of the polymorphic alleles between the populations would still be shared because the populations would not have experienced a significant bottleneck, and gene flow may still occur, at least with respect to genes and chromosomal regions that are not involved directly in the differential adaptation. Under ecological or adaptive speciation scenarios, population subdivisions arise on the basis of selection, rather than neutral drift. During incipient stages of ecological

speciation, drift does not have enough time to fix different relative frequencies of the alleles, and gene flow may still occur. An alternative prediction falls to the incomplete lineage sorting as a remain of an ancestral polymorphism. Further, because the majority of these morphologically/genetically distinct populations are or have been considered as vulnerable or in critical situation, I tried to assess if they can be recognized as evolutionary Units to conservation. In particular, I will try to analyse whether these different populations could be considered as Evolutionary significant Unit (ESUs), or as Management Units Criteria (MUs). The thesis manuscript is based on two main papers (in press or in preparation) presented as a chapters: Chapter IV and V.

Chapter IV

Overall here I combined mtDNA and nuclear introns and used gene tree or an allele clustering and coalescent species trees to assess the species boundaries within the populations of the *P. leucoptera* complex by means of multilocus, clustering, and coalescence-based species tree inference. I also wanted to disentangle their historical demographic history (time to divergence, migration rate per generation and long-term effective population size) in order to understand how these evolutionary parameters have shaped the process of speciation of *P. leucoptera* complex. I further made use of phylogenetic and phylogeographic analyses to infer the evolutionary population units and assess whether they fit ESUs or MUs criteria.

Chapter V

In this Chapter, I wanted to corroborate or discard the presumption of an ecological speciation process under sympatric scenario mediated by a pre-mating mechanism of isolation (see Introduction: mechanisms of ecological speciation) between New Caledonia and Australian populations of the *P. l leucoptera* complex. These two populations deploy not only ecological (asynchronous breeding time and differences in habitat preferences) and morphological differences but also, no exchange has ever been documented between both populations (though long-term ringing has been done only on Cabbage Tree Island). Additionally, both taxa experienced recent fluctuations in population sizes whether by the introduction of predators or anthropogenic settlement. Furthermore, in this chapter, we wanted to understand how historical events, such as Pleistocene glacial/interglacial cycles impacted demographic fluctuations that could shape the genetic structure of apparently differentiated populations by ecological and morphological traits (regarding parameters such effective population size (Ne), migration, time to divergence). Along this view, we aimed : (i) to clarify the genetic relationships within Caledonica and Australia populations of *Pterodroma leucoptera* complex, and ascertain whether their taxonomic treatment as two subspecies is supported by molecular variation and assuming the neutral evolution of mtDNA for this species complex ii) to test whether its present genetic structure results from demographic fluctuations through time (expansions and bottlenecks) due to Pleistocene climatic oscillations, or from more recent changes likely related to anthropogenic pressure; and iii) to establish time of divergence and past effective population sizes.

REFERENCES

Agapow PM, Bininda-Emonds OR, Crandall, KA, et al (2004) The impact of species concept on biodiversity studies.The quarterly review of biology. 79.2:161-179.

Albert AY, Schulter D (2005) Selection and the origin of species. Current Biology 15:283–288.

Andrews KR, Norton EL, Fernandez-Silva I, et al (2014) Multilocus evidence for globally distributed cryptic species and distinct populations across ocean gyres in a mesopelagic copepod. Molecular Ecology. 1:5462–5479.

Avise JC (1994) Molecular markers, natural history and evolution. Chapmann and Hall Inc. New York.

Avise JC (2000) Phylogeography: the history and formation of species. Harvard university press.

Avise JC (2009) Phylogeography: Retrospect and prospect. Journal of Biogeography 36.1:3–15.

Avise JC (1998) The history and purview of phylogeography: a personal reflection. Molecular Ecology.7.4:371–379.

Avise JC, Arnold J, Ball, RM, et al (1987) Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. Annual Review of Ecology and Systematics.489-255.

Baker J, Lopez-Medrano E, Navarro-Siguenza A, et al (2003) Recent speciation in the Orchard Oriole group: divergence of Icterus spurius and Icterus spurius fuertesi. The Auk. 120.3:848-859.

Ball RM, Avise JC (1992) Mitochondrial DNA phylogeographic differentiation among avian populations and the evolutionary significance of subspecies. The Auk. 626–636

Ballard JWO, Whitlock MC (2004) The incomplete natural history of mitochondria. Molecular Ecology.13:729–744.

Bank C, Bürger R, Hermisson J (2012) The limits to parapatric speciation: Dobzhansky-Muller incompatibilities in a continent-Island model. Genetics. 191:845–863.

Barrowclough GF, Zink RM (2009) Funds enough, and time: mtDNA, nuDNA and the discovery of divergence. Molecular Ecology. 18:2934–2936.

Barton NH (2001) Speciation. Trends in Ecology and Evolution. 16.7:325.

Beaumont MA., Zhang W, Balding DJ (2002) Approximate Bayesian computation in population genetics. Genetics. 162.4:2025–2035.

Beerli P, Felsenstein J (1999) Maximum-likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. Genetics. 152.2:763–773.

Beerli P, Palczewski M (2010) Unified framework to evaluate panmixia and migration direction among multiple sampling locations. Genetics. 185.1:313–326.

Berner D, Thibert-Plante X (2015) How mechanisms of habitat preference evolve and promote divergence with gene flow. Journal of Evolutionary Biology. 28.9:1641–1655.

Bochkov AV., Klimov PB, Hestvik G, Saveljev AP (2014) Integrated Bayesian species delimitation and morphological diagnostics of chorioptic mange mites (Acariformes: Psoroptidae: Chorioptes). Parasitology Research. 113.7:2603–2627.

Bolnick DI, Fitzpatrick BM (2007) Sympatric Speciation: Models and empirical Evidence. Annual Review of Ecology, Evolution, and Systematics. 38:459–487.

Bolnick DI, Nosil P (2007) Natural selection in populations subject to a migration load. Evolution. 61.9:2229–43.
Britton-Davidian J (2001) How do chromosomal changes fit in? Journal of Evolutionary Biology. 14:872–873.
Broquet T, Petit EJ (2009) Molecular Estimation of Dispersal for Ecology and Population Genetics. Annual Review of Ecology, Evolution, and Systematics. 40:193–216.
Brown WM (1985) The mitochondrial genome of animals. Molecular Evolutionary Genetics. Edited by: MacIntyre R.J. 95–130.
Bull LS (2006) Geographical variation in the morphology of the Wedge-tailed Shearwater (Puffinus pacificus). Emu. 106.3:233–243.
Bush GL (1994) Sympatric speciation in animals: new wine in old bottles. Trends in Ecology and Evolution. 9.8:285–288.
Butlin RK, Debelle A, Kerth C, et al (2012) What do we need to know about speciation? Trends in Ecology and Evolution 27.1:27–39.
Butlin RK, Galindo J, Grahame JW (2008) Review. Sympatric, parapatric or allopatric: the most important way to classify speciation? Philosophical Transactions of the Royal Society Lond B: Biological Sciences. 363.1506:2997–3007.
Camargo A, Sites JJ (2013) Species Delimitation: a decade after the renaissance. INTECH Open Acces Publisher. 225–247.
Carson HL, Templeton AR (1984) Genetic revolutions in relation to speciation phenomena: the Founding of new Populations. Annual Review of Ecology Systematics. 15:97–132.
Carstens BC, Brennan RS, Chua V, et al (2013a) Model selection as a tool for phylogeographic inference: an example from the willow Salix melanopsis. Molecular Ecology. 22:4014–28.
Carstens BC, Dewey TA (2010) Species delimitation using a combined coalescent and information-theoretic approach: an example from North American myotis bats. Systematic Biology. 59:400–414.
Carstens BC, Pelletier T a., Reid NM, Satler JD (2013b) How to fail at species delimitation. Molecular Ecology. 22:4369–4383.
Castelloe J, Templeton A (1994) Root probabilities for intraspecific gene trees under neutral coalescent theory. Molecular Phylogenetics and Evolution. 3.2:102–113.
Chevin L (2016) Species selection and random drift in macroevolution. Evolution. 70.3:513-525.
Choi SC, Hey J (2011) Joint inference of population assignment and demographic history. Genetics. 189.2, 561–577.
Clark AG (1997) Neutral behavior of shared polymorphism. Procedings of the National Academy of Science. 94.15:7730–7734.
Coates DJ (2000) Defining conservation units in a rich and frag- mented flora: implications for the management of genetic resources and evolutionary processes in south-west Australian plants. Australian Journal of Botany. 48: 329–339.
Collevatti RG, Telles MPC, Nabout JC, et al (2013) Demographic history and the low genetic diversity in Dipteryx alata (Fabaceae) from Brazilian Neotropical savannas. Heredity (Edinb) 111.2:97–105.
Coyne J, Orr H (2004) Speciation.37. Sunderland, MA: Sinauer Associates.
Cracraft J (1983) Species concepts and speciation analysis. Current Ornithology.

Springer US. 159–187.
Cracraft J (1997) Species concepts in systematics and conservation biology an ornithological viewpoint. Systematics Association Special Volume. 54:325-340.
Crandall KA, Bininda-Emonds OR, Mace GM et al (2000) Considering evolutionary processes in conservation biology. Trends in ecology and evolution. 15.7:290-295.
Crisp M, Chandler G (1996) Paraphyletic species. Telopea 6.4:813–844.
Darwin C (1859) On the origin of species by means of natural selection. John Murray, London, UK.
Dayrat B (2005) Towards integrative taxonomy. Biological Journal Linnean Society 85.3:407–415.
De Queiroz K (1998) The general lineage concept of species, species criteria, and the process of speciation. A conceptual unification and terminological recommendations. Endless forms.Species and speciation. Oxford University Press, Oxford, England, UK. 470 pp.
De Queiroz K (2005) A unified concept of species and its consequences for the future of taxonomy. Proceedings of the California Academy of Sciences. 56.18:196–215.
De Queiroz K (1999) The general lineage concept of species and the defining properties of the species category. In: Wilson, RA, ed., Species. New Interdisciplinary Essays. MIT Press, Cambridge, Massachussetts, USA. 325 pp.
De Queiroz K (2007) Species concepts and species delimitation. Systematic Biology. 56:879–886.
Devaux C, Lande R (2008) Incipient allochronic speciation due to non-selective assortative mating by flowering time, mutation and genetic drift. Proceedings of the Royal Society of London B: Biological Sciences. 275.1652:2723-2732.
Dieckmann, U., Tautz, D., Doebeli, M., Metz JAJ (2004) Adaptive speciation. Epilogue. Cambridge University Press.
Diehl, SR., Bush GL (1989) The role of habitat preference in adaptation and speciation. Speciation and its Consequences. 345–365.
Dobzhansky T (1951) Genetics and the Origin of Species, 3rd edn. Columbia University Press, New York.
Dobzhansky T (1937) Genetics and the Origin of Species, No. 11. Columbia University Press.
Donoghue M (1985) A critique of the biological species concept and recommendations for a phylogenetic alternative. Bryologist. 88:172–181.
Drummond AJ, Rambaut A., Shapiro B, Pybus OG (2005) Bayesian coalescent inference of past population dynamics from molecular sequences. Molecular Biology Evolution. 22.5:1185–1192.
Drummond AJ, Pybus OG, Rambaut A, et al (2003) Measurably evolving populations. Trends in Ecology and Evolution. 18:481–488.
Edwards SV. (2009) Is a new and general theory of molecular systematics emerging? Evolution (NY). 63:1–19.
Edwards SV, Kingan SB, Calkins JD, et al (2005) Speciation in birds: genes, geography, and sexual selection. Proceedings of the National Academy of Sciences,*102* (suppl 1), 6550-6557.
Edwards S, Bensch S (2009) Looking forwards or looking backwards in avian phylogeography? A comment on Zink and Barrowclough 2008. Molecular Ecology 18:2930–3; discussion 2934–6.

Edwards T, Tollis M, Hsieh P, et al (2016) Assessing models of speciation under different biogeographic scenarios; an empirical study using multi-locus and RNA-seq analyses. Ecology and Evolution.6.2: 379-396.
Eldredge N, Cracraft J (1980) Phylogenetic patterns and the evolutionary process. University Press., New York, NY: Columbia.
Erwin DH (2000) Macroevolution is more than repeted rounds of microevolution. Evolution and Development. 2 2:78–84.
Feder JL, Egan SP, Nosil P (2012) The genomics of speciation-with-gene-flow. Trends in Genetics. 28:342–350.
Felsenstein J (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach. Journal of Molecular Evolution. 17:368–76.
Fitzpatrick BM, Fordyce JA, Gavrilets S (2009) Pattern, process and geographic modes of speciation. Journal of Evolutionary Biology 22:2342–7.
Forister ML, Nice CC, Fordyce JA., et al (2008) Considering evolutionary processes in the use of single-locus genetic data for conservation, with examples from the Lepidoptera. Journal of Insect Conservation. 12.1:37–51.
Fraser DJ, Bernatchez L (2001) Adaptive evolutionary conservation: towards a unified concept for defining conservation units. Molecular Ecology. 10:2741–2752.
Friesen VL, Smith AL, Gómez-Díaz E, et al (2007) Sympatric speciation by allochrony in a seabird. Proceedings of the National Academy of Sciences. 104. 47:18589–18594.
Fujita MK, Leaché, AD, Burbrink FT, et al (2012) Coalescent-based species delimitation in an integrative taxonomy. Trends in Ecology and Evolution. 27:480–488.
Futuyma DJ (2005) Evolution. Sinauer Ass. Inc., Massachusets, USA.
Isaac NJ, Mallet J, Mace, GM (2004) Taxonomic inflation: its influence on macroecology and conservation. Trends in Ecology and Evolution,19(9), 464-469.
Gabirot M, Raux L, Dell'ariccia G, Al. E (2015) Chemical labels differ between two closely related shearwater taxa.
Gaggiotti OE (2011) Making inferences about speciation using sophisticated statistical genetics methods: look before you leap. Molecular Ecology. 20:2229–32.
Gavrilets S (2004) Fitness landscapes and the origin of species. Princeton, NJ: Princeton University Press.
Gavrilets S (2003) Perspective: models of speciation: what have we learned in 40 years? Evolution 57:2197–215.
Gill F, Mostrom A, Mack A (1993) Speciation in North American chickadees: I. Patterns of mtDNA genetic divergence. Evolution (NY) 195–212.
Gill M, Lemey P, Faria N, et al (2012) Improving Bayesian Population Dynamics Inference: A Coalescent-Based Model for Multiple Loci. Molecular Biology and Evolution. 30:713–724.
Glaw F, Köhler J, De la Riva I, et al (2010) Zootaxa, Integrative taxonomy of Malagasy treefrogs: combination of molecular genetics, bioacoustics and comparative morphology reveals twelve additional species of Boophis. *Zootaxa*. 2383.1: 82.
Grant BR, Grant PR (2002) Simulating secondary contact in allopatric speciation: an empirical test of premating isolation. Biological Journey of the Linnean Society.

76.4:545–556.
Grant PR, Grant BR (2009) The secondary contact phase of allopatric speciation in Darwin's finches. Proceedings of the National Academy of Sciences. 106. 48:20141–8.
Green DM (2005). Designatable units for status assessment of endangered species. Conservation Biology. 19(6), 1813-1820.
Greenberg R, Cordero PJ, Droege S, Fleischer RC (1998) Morphological adaptation with no mitochondrial DNA differentiation in the coastal plain swamp sparrow. The Auk. 706–712.
Griffiths RC, Tavaré S (1994) Ancestral Inference in population genetics. Statistical Science. 9:307–319.
Hare KM, Daugherty CH, Chapple DG (2007) Comparative phylogeography of three skink species (Oligosoma moco, O. smithi, O. suteri; Reptilia: Scincidae) in northeastern New Zealand. Molecular Phylogenetics and Evolution. 46.1:303–315.
Hausdorf B (2011) Progress toward a general species concept. Evolution 65:923–31. Hausdorf B (2011b) Progress toward a general species concept. Evolution (NY). 65:923–931.
Hausdorf B, Hennig C (2010) Species delimitation using dominant and codominant multilocus markers. Systematic Biology.59:491–503.
Heled J, Bryant D, Drummond AJ (2013) Simulating gene trees under the multispecies coalescent and time-dependent migration. BMC. Evolutionary Biology. 13:44.
Heled J, Drummond AJ (2008) Bayesian inference of population size history from multiple loci. 15:1–15.
Heled J, Drummond AJ (2010) Bayesian inference of species trees from multilocus data. Molecular Biology and Evolution. 27.3:570–80.
Hendry a P, Vamosi SM, Latham SJ, et al (2000) Questioning species realities. Conservation Genetics. 1:67–76.
Hennig W (1966) Phylogenetic systematics. Urbana: University of Illinois Press.
Hey BJ (2011) Documentation for IMa2. 1–60.
Hey J (2010) Isolation with migration models for more than two populations. Molecular Biology and Evolution. 27.4:905–20.
Hey J (2007) Introduction to the IM and IMa computer programs Understanding Markov chain Monte Carlo (MCMC). How Long to Run the Program Metropolis Coupling of Markov chains Assessing the Autocorrelation of Recorded Values Parameter Conversions Cautions , Suggestions. 1–16.
Hey J (2006) Recent advances in assessing gene flow between diverging populations and species. Current Opinion in Genetics and Development. 16:592–596.
Hey J, Nielsen R (2004) Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of Drosophila pseudoobscura and D. persimilis. Genetics. 167:747–60.
Hey J, Nielsen R (2007) Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. Proceedings of the National Academy of sciences. 104.8:2785–2790.
Hindwood K, Serventy D (1941) The Gould petrel of Cabbage Tree Island. Emu.41.1:1-20.
Hohenegger J (2012) Transferability of genomes to the next generation: the

fundamental criterion for the biological species. Zootaxa. 3572:11–17.
Hohenegger J (2014) Species as the basic units in evolution and biodiversity : Recognition of species in the Recent and geological past as exempli fi ed by larger foraminifera. Gondwana Res 25:707–728.
Hotaling S, Foley ME, Lawrence NM, et al (2016) Species discovery and validation in a cryptic radiation of endangered primates: coalescent-based species delimitation in Madagascar's mouse lemurs. Molecular Ecology.
Hsieh Y, Chung J, Wang C, et al (2013) Historical connectivity , contemporary isolation and local adaptation in a widespread but discontinuously distributed species endemic to Taiwan , Rhododendron oldhamii (Ericaceae). Heredity (Edinb) 111:147–156.
Hudson RR (1983) Properties of a neutral allele model with intragenic recombination. Theoretical population biology. 23.2:183–201.
Huelsenbeck JP, Andolfatto P, Huelsenbeck ET (2011) Structurama: Bayesian inference of population structure. Evolutionary Bioinformatics. 7:55–59.
Jackson ND, Austin CC (2010) The combined effects of rivers and refugia generate extreme cryptic fragmentation within the common ground skink (Scincella lateralis). Evolution (NY). 64:2:409–428.
Kimball RT, Braun EL, Barker FK, et al (2009) A well-tested set of primers to amplify regions spread across the avian genome. Molecular Phylogenetics and Evolution 50:654–660.
Kingman JFC (1982) The coalescent. Stochastic Processes and their Applications. 13:235–248.
Kirkpatrick M, Ravigne V (2002) Speciation by Natural and Sexual Selection: Models and Experiments. The American Naturalist. 159(S3):S22–35.
Knowles LL, Maddison WP (2002) Statistical phylogeography. Molecular Ecology. 11.12:2623-2635.
Kuhner MK, Yamato J, Felsenstein J (1998) Maximum likelihood estimation of population growth rates based on the coalescent. Genetics 149:429–434.
Kuhner MK, Yamato J, Felsenstein J (1995) Estimating effective population size and mutation rate from sequence data using Metropolis-Hastings sampling. Genetics 140:1421–1430.
Langerhans RB, Gifford ME, Joseph EO (2007) Ecological speciation in Gambusia fishes. Evolution. 61:2056–74.
Lavretsky P, Dacosta JM, Hernández-Baños BE, et al (2015) Speciation genomics and a role for the Z chromosome in the early stages of divergence between Mexican ducks and mallards. Molecular Ecology. 24:5364–5378.
Lexer C, Wüest RO, Mangili S, et al (2014) Genomics of the divergence continuum in an African plant biodiversity hotspot, I: Drivers of population divergence in Restio capensis (Restionaceae). Molecular Ecology 23.17:4373–4386.
Mace GM (2004) The role of taxonomy in species conservation. Philosophical Transactions of the Royal Society B: Biological Sciences. 359. 1444:711–719.
Mallet J (2005) Hybridization as an invasion of the genome. Trends in Ecology and Evolution. 20.5: 229–237.
Mallet J (2008) Hybridization, ecological races and the nature of species: empirical evidence for the ease of speciation. Philosophical Transactions of the Royal Society of London B: Biological Sciences. 363.1506:2971–2986.
Marshall D, Simon C, Buckley T (2006) Accurate branch length estimation in partitioned Bayesian analyses requires accommodation of among-partition rate

variation and attention to branch length priors. Systematic Biology. 55:993–1003.
Matessi C, Gimelfarb A, Gavrilets S (2002) Long-term buildup of reproductive isolation promoted by disruptive selection: how far does it go? Selection. 2:41–64.
Mayr E (1963) Animal Species and Evolution. Harvard University Press, Cambridge.
Mayr E (1942) Systematics and the origin of species, from the viewpoint of a zoologist. Harvard University Press.
Mayr E (1954) Change of genetic environment and evolution.157-1.
Mayr E (1947) Ecological factors in speciation. Evolution (NY). 1:263– 288.
Monaghan MT, Balke M, Pons J, Vogler AP (2006) Beyond barcodes: complex DNA taxonomy of a South Pacific Island radiation. Proceedings of the Royal Society of London B: Biological Sciences, 273.1588:887–93.
Monaghan MT, Wild R, Elliot M, et al (2009) Accelerated species inventory on Madagascar using coalescent-based models of species delineation. Systematic Biology. 58:298–311.
Moritz C (1994) Defining “evolutionarily significant units” for conservation. Trends in Ecology and Evolution.9.10:373–375.
Moritz C (1999). Conservation units and translocations: Strategies for conserving evolutionary processes. Hereditas. 130:217–228.
Moritz C, Dowling TE, Brown WM (1987) Evolution of animal Mitochondrial DNA: relevance for population biology and systematics. Annual Review Ecology and systematics. 18:269–292.
Morris-Pocock J (2012) Phylogeography of Brown and Red-Footed Boobies: a Complex History of Isolation and Gene Flow.
Nei M (1976) Mathematical models of speciation and genetic distance. Population genetics and ecology. Academic Press, New York. 723–766.
Nei M, Maruyama T, Wu CI (1983) Models of evolution of reproductive isolation. Genetics.103.3:557–579.
Nei M, Tajima F (1983) Maximum likelihood estimation of the number of nucleotide substitutions from restriction sites data. Genetics 105.1:207–217.
Neigel J, Avise J (1986) Phylogenetic relationships of mitochondrial DNA under various demographic models of speciation. Evolutionary processes and theory 515–534.
Nichols R (2001) Gene trees and species trees are not the same. Trends in Ecology and Evolution. 16.7:358–364.
Nielsen R, Beaumont M a. (2009) Statistical inferences in phylogeography. Molecular Ecology.18:1034–1047.
Nielsen R, Slatkin M (2000) Likelihood analysis of ongoing gene flow and historical association. Evolution (NY).54:44–50.
Nielsen R, Wakeley J (2001) Distinguishing migration from isolation: a Markov chain Monte Carlo approach. Genetics. 158:885–896.
Noor MAF (2002) Is the biological species concept showing its age? Trends in Ecology and Evolution.17:153–154.
Nosil P, Crespi BJ, Sandoval CP (2003) Reproductive isolation driven by the combined effects of ecological adaptation and reinforcement. Proceedings of the Royal Society of London. 270.1957:1911–1918.
Nosil P, Harmon LJ, Seehausen O (2009) Ecological explanations for (incomplete) speciation. Trends in Ecology and Evolution. 24.3:145–56.
O’Meara BC (2010) New heuristic methods for joint species delimitation and species

tree inference. Systematic Biology. 59:59–73.
Ogilvie HA, Heled J, Xie D, Drummond AJ (2016) Computational Performance and Statistical Accuracy of *BEAST and Comparisons with Other Methods.
Oliver JC (2013) Microevolutionary processes generate phylogenomic discordance at ancient divergences. Evolution (NY) 67.6:1823–1830.
Olivieri GL, Sousa V, Chikhi L, Radespiel U (2008) From genetic diversity and structure to conservation: Genetic signature of recent population declines in three mouse lemur species (Microcebus spp.). Biol Conservation. 141.5:1257–1271.
Omland KE, Baker JM, Peters JL (2006) Genetic signatures of intermediate divergence: population history of Old and New World Holarctic ravens (Corvus corax). Molecular Ecology 15.3:795–808.
Orr H (2001) The genetics of species differences. Trends in Ecology and Evolution 16:343–350.
Padial J, Miralles A, De la Riva I, Vences M (2010) The integrative future of taxonomy Review. Frontiers in Zoology.7.1:1–14.
Palmer JD (1992) Mitochondrial DNA in plant systematics: applica- tions and limitations. In: Molecular Systematics of Plants. Chapman and Hall (NY).36-49.
Panhuis TM, Butlin R, Zuk M, Tregenza T (2001) Sexual selection and speciation. Trends in Ecology and Evolution. 16.7:364–371.
Pante E, Abdelkrim J, Viricel A, et al (2014a) Use of RAD sequencing for delimiting species. Heredity (Edinb). 114:450–459.
Pante E, Schoelinck C, Pulliandre N (2014b) From integrative taxonomy to species descrption: one step beyond: Species delimitation and description. Systematic Biology. http://hal.archives-ouvertes.fr/hal-01076864.
Patton J, Smith M (1994) Paraphyly, Polyphyly and the nature of species boundaries in pocket gophers (Genus Thomomys). Systematic Biology. 43:11–26.
Pavlova A, Selwood P, Harrisson KA, et al (2014) Integrating phylogeography and morphometrics to assess conservation merits and inform conservation strategies for an endangered subspecies of a common bird species. Biolological Conservation.174:136–146.
Pennock DS, Dimmick WW. (1997). Critique of the evolutionarily significant unit as a definition for “distinct population segments” under the US Endangered Species Act. Conservation Biology.11.3:611-619.
Pereira RJ, Wake DB (2015) Ring species as demonstrations of the continuum of species formation. Molecular Ecology . 24.21:5312–5314.
Petit RJ, Excoffier L (2009) Gene flow and species delimitation. Trends in Ecology and Evolution. 24.7:386–93.
Pompanon F, Samadi S (2015) Next generation sequencing for characterizing biodiversity: promises and challenges. Genetica. 143.2:133–138.
Presgraves DC (2010) The molecular evolutionary basis of species formation. Nature Reviews Genetics. 11.3:175–180.
Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics.155:2:945–959.
Puillandre N, Bouchet P, Boisselier-Dubayle MC, et al (2012) New taxonomy and old collections: integrating DNA barcoding into the collection curation process. Molecular Ecology Resources. 12.3:396–402.
Pyron RA, Costa GC, Patten MA, Burbrink FT (2015) Phylogenetic niche conservatism and the evolutionary basis of ecological speciation. Biological

Reviews. 90.4:1248–1262.
Rabosky DL (2015) Reproductive isolation and the causes of speciation rate variation in nature. Biological Journal of the Linnean Society.13–25.
Reznick DN, Ricklefs RE (2009) Darwin's bridge between microevolution and macroevolution. Nature. 457.7231:837–842.
Rice W, Hostert EE (1993) Laboratory Experiments on speciation. what have we learned in 40 years. Evolution. 47:1637–1653.
Richards RA (2010) The species problem.-A philosophical analysis. Cambridge University Press, Cambridge, UK.
Ridley M (1994) Evolution, Third Edit. Blackwell Publishing company
Rodríguez V, Brown RP, Terrasa B, et al (2013) Multilocus genetic diversity and historical biogeography of the endemic wall lizard from Ibiza and Formentera, Podarcis pityusensis (Squamata: Lacertidae). Molecular Ecology. 22.19:4829–41.
Roe KJ, Lydeard C (1998) Species delineation and the identification of evolutionarily significant units: Lessons from the freshwater mussel genus Potamilus (Bivalvia : Unionidae). Journal of Shellfish Research. 17.5:1359–1363.
Ronquist F, Huelsenbeck JP, Britton T (2004) Bayesian Suepertrees. In Phylogenetics supertrees. 3:193-224.
Rosenberg NA, Nordborg M (2002) Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. Nature Reviews Genetics. 3.5:380–390.
Rubinoff D, Holland B (2005) Between two extremes: mitochondrial DNA is neither the panacea nor the nemesis of phylogenetic and taxonomic Inference. Systematic Biology. 54.6:952–961.
Rundle HD, Nosil P (2005) Ecological speciation. Ecology Letters. 8.3:336–352.
Rundle HD, Nagel L, Boughman JW, et al (2000) Natural selection and parallel speciation in sympatric sticklebacks. Science. 287.5451:306–308.
Ruskey JA, Taylor EB (2016) Morphological and genetic analysis of sympatric dace within the Rhinichthys cataractae species complex : a case of isolation lost. 117.3:547–563.
Ryan PG, Bloomer P, Moloney CL, et al (2007) Ecological speciation in South Atlantic island finches. Science.315.5817:1420–3.
Ryder OA (1986) Species conservation and systematics: the dilemma of subspecies. Trends in Ecology and Evolution.1.1:9–10.
Samadi S, Barberousse A (2015) Species. In Handbook of Evolutionary Thinking in the Sciences. Springer Netherlands. 141–157.
Samadi S, Barberousse A (2009) Species: towards new, well-grounded practices. Biological Journal Linnean Society. 97.1:217–222.
Samadi S, Barberousse A (2006) The tree, the network, and the species. Biological Journal Linnean Society. 89.3:509–521.
Sanmartín I, Van Der Mark P, Ronquist F (2008) Inferring dispersal: A Bayesian approach to phylogeny-based island biogeography, with special reference to the Canary Islands. Journal of Biogeography. 35.3:428–449.
Satler J, Carstens B, Hedin M (2013) Multilocus species delimitation in a complex of morphologically conserved trapdoor spiders (Mygalomorphae, Antrodiaetidae, Aliatypus). Systematic Biology. 62.6:805–823.
Savolainen V, Anstett MC, Lexer C, et al (2006) Sympatric speciation in palms on an oceanic island. Nature. 441.7090:210–213.
Schlick-Steiner BC, Steiner FM, Seifert B, et al (2010) Integrative taxonomy: a

multisource approach to exploring biodiversity. Annual Review Entomology 55:421–38.
Schluter D (2009) Evidence for ecological speciation and its alternative. Science 323.5915:737–741
Schluter D (2001) Ecology and the origin of species. Trends in Ecology and Evolution.16.7:372–380..
Schluter D (2000) The Ecology of Adaptive Radiation. Oxford University Press.
Seehausen O, Butlin RK, Keller I, et al (2014) Genomics and the origin of species. Nature Reviews Genetics. 15.3:176–92.
Servedio MR (2015) Geography, assortative mating, and the effects of sexual selection on speciation with gene flow. Evolutionary Applications. 9.1:91-102.
Shafer A, Wolf JBW (2013) Widespread evidence for incipient ecological speciation: a meta-analysis of isolation-by-ecology. Ecology Letters. 16.7:940–950.
Shaffer HB, Thomson RC (2007) Delimiting species in recent radiations. Systematic Biology. 56.6:896–906.
Shapiro B, Ho SY (2011) Skyline-plot methods for estimating demographic history from nucleotide sequences. Molecular Ecology Resources.11.3:423–434.
Silva MC, Matias R, Wanless RM, et al (2015) Understanding the mechanisms of antitropical divergence in the seabird White-faced Storm-petrel (Procellariiformes: Pelagodroma marina) using a multilocus approach. Molecular Ecology. 24.12:3122–3137.
Simons AM (2002) The continuity of microevolution and macroevolution. Journal of Evolutionary Biology.15.5:688–701.
Simpson G (1961) Principals of animal taxonomy. Columbia University Press, New York.
Simpson G (1951) The species concept. Evolution (NY) 5:285–298.
Sites JW, Marshall JC (2004) Operational criteria for delimiting species. Annual Review of Ecology, Evolution, and Systematics.35:199–227.
Sites JW, Marshall JC (2003) Delimiting species: a renaissance issue in systematic biology. Trends in Ecology and Evolution.18.9:462–470.
Slatkin M (1987) Gene flow and the geographic structure of natural populations. Science. 236.4803:787–792.
Slatkin M (1985) Gene flow in natural populations. Annual Review of Ecology Systematics.16:393–430.
Smith TA, Martin MD, Nguyen M, Mendelson TC (2016) Epigenetic divergence as a potential first step in darter speciation. Molecular Ecology. 25:1883-1894.
Sobel J, Chen G, Watt L, Schemske D (2009) The biology of speciation. Evolution. 64:295–315.
Stamos D (2003) The species problem: Biological species, ontology, and the metaphysics of biology. Lexington Books.
Stearns S, Hoekstra R (2005) Evolution: An introduction. Oxford University Press.
Streelman JT, Danley PD (2003) The stages of vertebrate evolutionary radiation. Trends in Ecology and Evolution.18:126–131.
Taylor EB (1999) Species pairs of north temperate freshwater fishes: evolution, taxonomy, and conservation. Reviews in Fish Biology and Fisheries.9: 299–324.
Templeton AR, Routman E, Phillips CA (1995) Separating population structure from population history: A cladistic analysis of the geographical distribution of mitochondrial DNA haplotypes in the tiger salamander, Ambystoma tigrinum. Genetics.140.2:767–782.

Templeton AR (1981) Mechanisms of speciation a population genetic approach. Annual Review of Ecology and Systematics. 12:23–48.
Turelli M, Lipkowitz JR, Brandvain Y (2014) On the coyne and origin of species: effects of intrinsic postzygotic isolation, ecological differentiation, x chromosome size, and sympatry on drosophila speciation. Evolution (NY).68:1176–1187.
Turelli M, Barton NH, Coyne JA (2001) Theory and speciation. Trends in Ecology and Evolution. 16.7:330–343.
Turner, GF, Burrows MT (1995) A model of sympatric speciation by sexual selection. Proceedings of the Royal Society of London B: Biological Society. 260.1359:287–292.
Valdecasas AG, Williams D, Wheeler QD (2008) “Integrative taxonomy” then and now: A response to Dayrat (2005). Biological Journal of the Linnean Society. 93.1:211–216.
Van-Loenen AL (2013) Reconstructing the Genetic Legacy of Cook’s Petrels (Pterodroma cookii). Project report for the degree B. Sc.
Velasco JD (2008) The internodal species concept: A response to “The tree, the network, and the species.” Biological Journal of the Linnean Society. 93.4:865-869.
Via S (2001) Sympatric speciation in animals: the ugly duckling grows up. Trends in Ecology and Evolution. 16:381–390.
Via S (2002) The ecological genetics of speciation. The American Naturalist 159 (S3):S1–S7.
Wakeley J, Hey J (1997) Estimating ancestral population parameters. Genetics 145.3:847–855.
Waples RS (1991) Pacific salmon, Oncorynchus spp., and the definition of “species” under the Endangered Species Act. Marine Fisheries Review. 53.3:11–22.
Waples, RS (1995) Evolutionarily signicant units and the conservation of biological diversity under the Endangered Species Act. In Evolution and the aquaticecosystem American Fisheries Society, Bethesda, Maryland.8-27.
Weisrock DW, Rasoloarison RM, Fiorentino I, et al (2010) Delimiting species without nuclear monophyly in Madagascar’s mouse lemurs. PLoS One. 5.3:e9883.
Welch AJ, Yoshida AA, Fleischer RC (2011) Mitochondrial and nuclear DNA sequences reveal recent divergence in morphologically indistinguishable petrels. Molecular Ecology. 20.7:1364–77.
Wellborn GA, Langerhans RB (2015) Ecological opportunity and the adaptive diversification of lineages. Ecology and Evolution. 5.1:176–195.
Wiens J (2007) Species delimitation: new approaches for discovering diversity. Systematic Biology. 56.6:875–878.
Wiens JJ (2004) Speciation and ecology revisited : phylogenetic niche conservatism and the origin of species. Evolution. 58.1:193–197.
Wiens JJ, Penkrot TA (2002) Delimiting species using DNA and morphological variation and discordant species limits in spiny lizards (Sceloporus). Systematic Biology. 51.1:69–91.
Will KW, Mishler BD, Wheeler QD (2005) The Perils of DNA barcoding and the need for integrative taxonomy. Systematic Biology. 54.5:844–851.
Wilson RE, Peters JL, McCracken KG (2013) Genetic and phenotypic divergence between low- and high-altitude populations of two recently diverged cinnamon teal subspecies. Evolution (NY). 67.1:170–184.
Wolf JBW, Bayer T, Haubold B, et al (2010) Nucleotide divergence vs. gene

expression differentiation: comparative transcriptome sequencing in natural isolates from the carrion crow and its hybrid zone with the hooded crow. Molecular Ecology. 19(s1),162–175.
Wright SD, Gillman LN, Ross HA, et al (2009) Slower tempo of microevolution in island birds: Implications for conservation biology. Evolution (NY). 63.9:2275–2287.
Wu CI (2001) The genic view of the process of speciation. Journal of Evolutionary Biology. 14.6:851–865.
Yang Z (2014) Molecular evolution: a statistical approach. OUP Oxford.
Yang Z (2015) The BPP program for species tree estimation and species delimitation. Currently Zoology. 61.5:854–865.
Yang Z, Rannala B (2014) Unguided Species Delimitation Using DNA Sequence Data from Multiple Loci. Molecular Biology and Evolution. 31:3125–3135.
Yang Z, Rannala B (2010) Bayesian species delimitation using multilocus sequence data. Procedings National Academy Sciences. USA.107.20:9264–9.
Zink R (1996) Species concepts, speciation, and sexual selection. Journal of Avian Biology. 27.1:1–6.
Zink R, McKitrick M (1995) The debate over species concepts and its implications for ornithology. The Auk. 701–719.
Zink RM, Barrowclough GF (2008) Mitochondrial DNA under siege in avian phylogeography. Molecular Ecology. 17.9:2107–21.

CHAPTER II

THE STUDY MODEL

A case study with the Petrels.

"The first part of knowledge is getting the names right"

Chinese proverb

CHAPTER II. THE STUDY MODEL

Over the past several decades, studies on birds have revealed a great variety of phylogeographic patterns, both in space and time, allowing understanding the temporal durations of behavioral and morphological specializations (Zink 1994; Zink 1996; Avise 2000). Phylogeographic studies have focused on characterizing breeding and overwintering connectivity on broad geographic scales (Kimura et al. 2002) detecting extreme population subdivisions and cryptic species (Baker et al. 1995), hybridization cases with deviation from linearity of molecular evolution (Saetre et al. 2001) and prehistoric genetic diversity decline in endangered species (Paxinos et al. 2002). However, deciphering levels of geographic structure and magnitude of gene flow at the core of avian populations is not an easy task. Birds (specifically males) are often extremely philopatric which constrains contemporary gene flow. In addition, geographic differences in phenotypic characters such as song, plumage, or body size are reported (Gómez-Díaz et al. 2006; Pinet et al. 2009; Bretagnolle and Shirihai 2010; Tennyson et al. 2012). Furthermore, most birds have exceptional dispersal potential by virtue of their capacity for flight and proclivity to migrate (van Bekkum et al. 2006). During the last decades, several studies based on different molecular markers, particularly mtDNA, have examined the genealogical and population history of avian populations from a phylogeographic point of view. The outcomes can vary greatly even among closely related taxa. For instance, *Parus carolinensis* shows highly divergent clades between its west and west of central Alabama populations, probably as a result of two Pleistocene refugia (Gill et al. 1993; Thomson et al. 2014). Conversely, the black-capped chickadee (*P. atricapillus*) and boreal chickadee (*P. hudsonicus*) feature the same haplotype from New York to Alaska, as a possible result of continental expansion from a single source population. In other cases, no phylogeographic structure has been detected, presumably as a result of post-Pleistocene range expansions into formerly glaciated regions (e.g. red-winged blackbird; snow goose, gene flow across, etc). In contrast, many other avian species have been reported to share common haplotypes over large or moderate geographic areas (*Uria lomvia* (Birt-Friesen et al. 1992), song sparrow *Melospiza melodie* (Zink and Dittmann 1993). In addition, failure to detect substantial matrilineal structure should not be dismissed as uninformative. For instance, no differences between geographic populations of Artic-breeding red knot has been detected probably due to a paucity of lineage variation between populations going through a bottleneck in the late Pleistocene, having expanded only within the last 10,000 years (Baker and Marshall 1997). Milot et al. (2000) confirmed the hypothesis of Mengel (1964) in which divergence between the yellow warbler (*Dendroica petetchia*) resulted from the splitting of ancestral contiguous populations into allopatric populations during the Pleistocene glaciations.

II.1. Seabirds

Seabirds are marine organisms that divide their lives between sea and land environments. The mechanisms and processes underpinning their phylogeographic patterns are not well understood. Most studies are conducted in Polar regions with a particular interest in comparing phylogeographic patterns in codistributed marine birds (Beheregaray 2008). Because of their wide dispersal ability, most studies on seabirds have been conducted at the hemispheric scale (Beheregaray 2008). Hence,

climatic oscillations during the Pleistocene have been ascribed a dominant role in sculpting phylogeographic patterns of seabirds (Klicka 1997; Moum and Arnason 2001). Pleistocene conditions impacted genetic differentiation between extant conspecific populations and shaped the phylogeographic variety into the many present day sister species of marine birds (Avise and Walker 1998; Steeves et al. 2005), for instance, in Artic seabirds (Wennerberg and Bensch 2001). In the Brunnich's guillemot *Uria lomvia*, and razorbill *Alica torda*, the colonization from a single south refugia of the North during Pleistocene climate oscillations explained current phylogeography (Friesen et al. 1996; Moum and Arnason 2001). Exactly the same scenario was invoked by Wojczulanis-Jakubas et al. (2014) for genetic structure of the little auk (*Alle alle*).

It has been proposed that isolating barriers influencing divergence processes that shape phylogeographic patterns are not the same at high and low latitudes (reviewed in Steeves et al. (2005). Mitochondrial control region data were used in order to elucidate the process of isolation in the widely distributed *Sula dactylatra*. Results suggested that the predominant pattern of differentiation between Indo-Pacific and Atlantic populations was shaped by physical barriers to gene flow. However, population divergence at the intra-ocean population level seemed to follow non-physical barriers to gene flow such as limited natal dispersal combined with local adaptation and/or genetic drift (Steeves et al. 2005).

II.2. The petrels

The order Procellariiformes alone represents more than one-third of all world seabirds and includes the largest family of marine birds: Procellariidae (82 species in 14 genera) (Warham 1990). Members of this family are highly philopatric, and thus we expect geographic variation across their ranges (Ainley et al. 1983; Baillie and Milne 1988). Moreover, this family shows widely distributed taxa with some species having breeding colonies located throughout the world's oceans, thus they experience a variety of environmental pressures (Smith et al. 2007; Peck et al. 2008). Furthermore, Procellariids are able to spend weeks and in some cases months or years at sea (Schreiber and Burger 2002) where it is believed that the main selective pressures act on foraging ability and predator avoidance (Bretagnolle 1993). In contrast, on land where they spend their time breeding, selective pressures switch to communication ability such as vocalizations cues for mating (Bretagnolle 1996). Indeed, as a consequence, procellariids are in most cases morphologically uniform (smaller species are cryptic species) and show moderate phylogenetic diversity (Austin et al. 2004). Procellariiformes usually inhabit remote oceanic breeding islands habitats, and exhibit a wide distribution, long generation time, high dispersal abilities, and a suite of life history traits that are suggested to affect their genetic population structure at different levels and time stages of differentiation (reviewed in Friesen et al. (2007). Procellaridae is thus a particularly attractive model system to investigate whether shared polymorphism results from incomplete lineage sorting or migration when population differentiation is incipient. In addition, many species are currently endangered (Hatfield et al. 2012) due to anthropogenic threats (IUCN 2015).

II.2.1 Current knowledge in petrel phylogeography

Speciation events related to climatic changes are trendy studies in seabirds phylogeography. For instance, Gómez-Díaz et al. (2006) evidenced that their analysis with clock calibrations match the first speciation event within *Calonectris*

species complex to the Panama Isthmus formation, suggesting a vicariant scenario for the divergence of the Pacific and the Palearctic clades of this complex. They hypothesized that the separation between the Atlantic and Mediterranean clades would have occurred in allopatry by range contraction followed by local adaptation during the Pleistocene. This result highlights the importance of oceanographic boundaries as potentially effective barriers shaping the phylogeographic structure of populations and species of pelagic seabirds. In addition, mitochondrial and nuclear data suggests that glacial and interglacial periods of the Pleistocene have influenced the strikingly deep phylogeographic structure of anti-tropical populations of white-faced storm petrels *Pelagodroma marina* (Silva et al. 2015). Southern and Northern populations of white faced storm-petrels may have diverged in isolation and currently exchange few migrants. The phylogeography of two morphospecies of Giant Petrels, *M. halli* (Northern form) and *M. giganteus* (Southern form) have been resolved by means of mtDNA and microsatellites (Techow et al. 2010). Both markers supported separate species status, although sequence divergence in cytochrome b was only 0.42% (corrected). Evidence of past fragmentation during the Pleistocene with a subsequent secondary contact within Southern Giant Petrels was also suggested. They also found support for a period of past population expansion that corresponded roughly to the timing of speciation and the separation of an ancestral giant petrel population from the fulmar *Fulmarus* clade.

Studies of recently diverged taxa using both mitochondrial and nuclear data set frequently produced gene trees with differing levels of phylogenetic signal as a result of differences in coalescence times. However, as it was mentioned above several factors can lead to this pattern, and it is important to distinguish between them to gain a better understanding of the process of divergence and the factors driving it (Welch et al. 2011). Using coalescent analysis from nuclear and mitochondrial data these authors have found that there was no migration between *Pterodroma phaeopygia* and *P. sandwichensis* over the last 100,000 generations and that they diverged relatively recently, approximately 550,000 years ago, suggesting that recent divergence and incomplete lineage sorting are causing the different patterns of phylogenetic signal in nuclear and mitochondrial data. Similar results were found by Gangloff et al. (2013) for Macaronesian petrel populations (*Pterodroma madeira, P. deserta* and *P. feae*) supporting that very recent divergence and incomplete lineage sorting are underpinning contrasting molecular signals. In various marine organisms, Pleistocene climatic oscillations have influenced population distribution and levels of gene exchanges and genetic variability through the appearance and disappearance of glacial refugia (Pielou 2008; Cheang et al. 2012). In particular, Pleistocene cyclic glacial/interglacial periods have had marked effects on seabirds (e.g. Gómez-Díaz et al. 2006; Techow et al. 2010).

II.2.2. *Pterodroma leucoptera* complex

II.2.2.1. Systematics and species status

Pterodroma genus contains 39 % of all the Procellaridae's species and 24.4% of all the species of Procellariiformes order (Birdlife International 2015.). Despite the longstanding interest in Petrels of the *Pterodroma leucoptera* complex, its taxonomic treatment over the last fifty years has been continuously changing. The *Pterodroma leucoptera* complex was first proposed by Cushman (1929) as a taxon containing two subspecies *P. leucoptera* and *P. brevipes.* At that time, two populations differing in morphological attributes were known: the first in Australia (*P. leucoptera)* and the

second in Vanuatu and Fiji (*P. brevipes)*. Hindwood and Serventy (1941) proposed that the taxon contains two sub-species, *P. leucoptera (*Australia) and *P. brevipes* (Vanuatu and Fiji). However, De Naurois (1978) found a new breeding population in New Caledonia. Imber and Jenkins (1981) named and described this population as another subspecies (*caledonica*), rearranging the *Pterodroma leucoptera* taxa into three subspecies: *leucoptera, caledonica,* and *brevipes.* Later, Sibley and Monroe (1990) elevated *P. brevipes* and *P. leucoptera* species status. In 1990-91, Seitre and Seitre (1991) discovered a new breeding population in the Raivavae, Australes islands, that V. Bretagnolle identified as possibly *caledonica* based on birds caught and photographed by A. Guillemont (in Brooke (2004). In addition, V. Bretagnolle and later, H. Shirihai identified three particular individuals (smaller in size) from the Society archipelago which could be placed either in *P. leucoptera* or in *P. brevipes*. V. Bretagnolle also reported one individual from Rarotonga, Cook island (a specimen held at Wellington, NZ) the identification of which remains uncertain (Bretagnolle and Shirihai 2010). In addition, in 1995, Bretagnolle (and after, Shirihai pers. comm.) identified three particular individuals (smaller in size) from the Society archipelago which could be placed either in *P. leucoptera* or in *P. brevipes*. The same year, Bretagnolle (1995 pers. comm.) reported one individual from Cook Island of which its identification remains uncertain. Based on color differentiations with *P. brevipes,* Bretagnolle and Shirihai (2010) gave described a new population from the northern end of Vanuatu, ascribing to its subspecies status (*P. brevipes magnificens)*. Finally, Bretagnolle and Shirihai (2010) found a new unidentified population from Eromango (although it was known from museum specimens held in AMNH, New York). Currently, the systematics of *P. brevipes* is obscure. All these taxonomic considerations have been exclusively based on morphometrics, coloration, and breeding location, and no study had ever been carried out using genetic markers. There has been no comprehensive evolutionary analysis among the populations and subspecies of the *P. leucoptera* complex that might shed additional light on the nature of differentiation within and among this group. Overall, their variation in color pattern, life history and ecology make the *Pterodroma leucoptera* complex a good candidate to obtain insights regarding how microevolutionary processes operate in a non-model marine bird. All these taxonomic considerations have been exclusively based on morphometrics, coloration (e. g. Figure 2.1; Table 2.1), and breeding location. Until 2010, no study, in particular, had ever been carried out using genetic markers. Gangloff et al. (2010) was the first to attempt to differentiate the *P. leucoptera* species complex using the mitochondrial DNA CO1 gene. His preliminary phylogenetic results showed no genetic differentiation between *P. l. leucoptera* and *P. l. caledonica.* However, both were differentiated from Vanuatu - *P. brevipes* populations and this last population was separated from Cook Island and the Society clade. Moreover, inside the Vanuatu - *P. brevipes* clade, he distinguished two sub-clades: one specifically to Vanuatu - *P. brevipes* and other clade mixing Vanuatu - *P. brevipes* and Fiji - *P. brevipes* populations. Therefore, the systematics within the *P. leucoptera* complex are far from being established, with many uncertainties in taxonomic status of the various populations, and the mechanisms that lead to a genetic differentiation among its populations are completely unknown. Iglesias-V (2012) found different levels of phylogenetic resolution using mitochondrial and nuclear DNA in *Pterodroma leucoptera - brevipes* complex. Unfortunately, these last results could not be analyzed using methods based on coalescent theory to corroborate the hypotheses of recent divergence of *Pterodroma* populations.

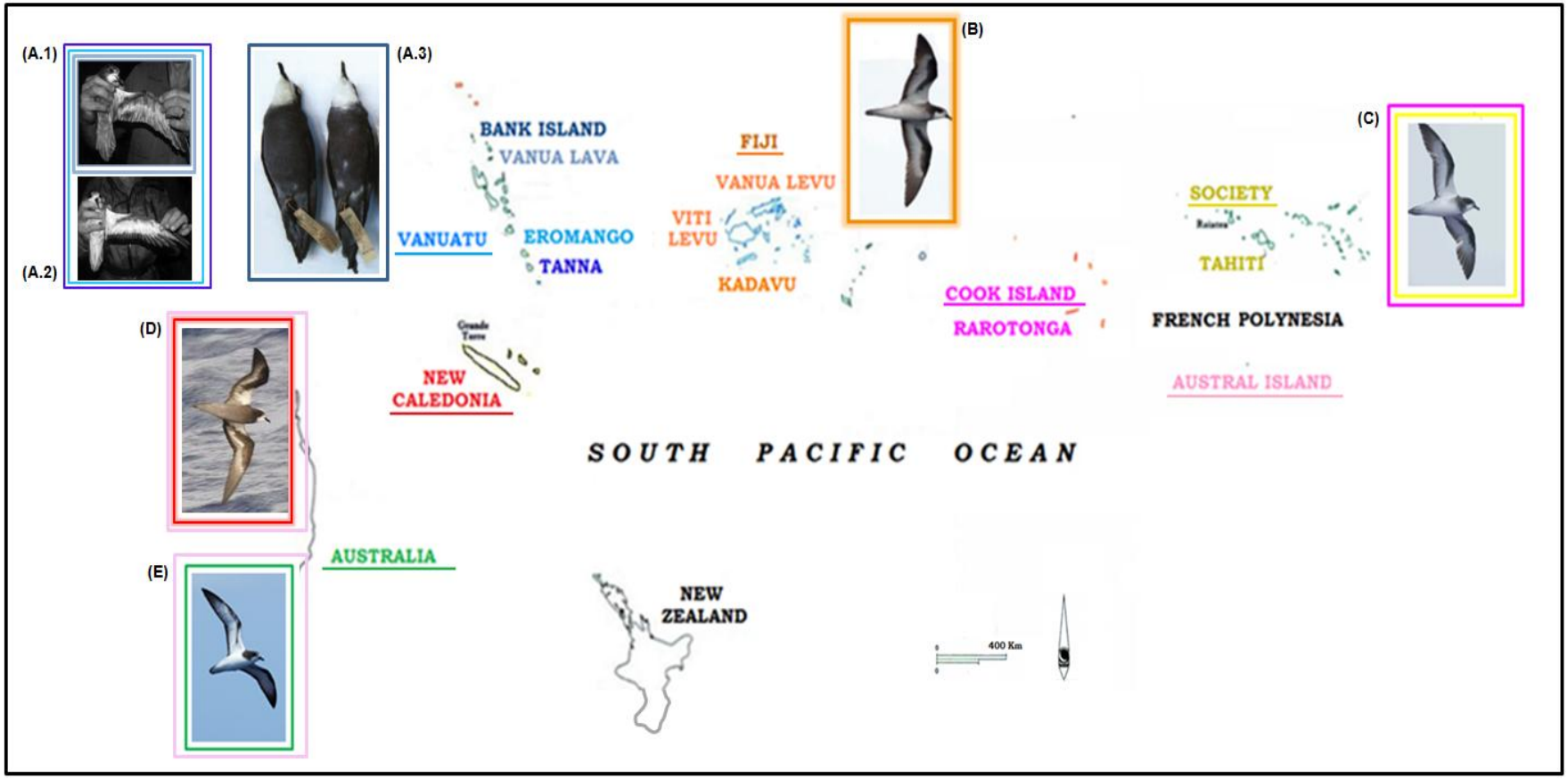


Figure 2.1. Morpho subspecies distribution and key morphological distinctions of *P. leucoptera* complex. **Vanuatu Island** - *P. brevipes* **(A.1), (A.2), (A.3)** (presence of both, dark grey morph and paler morph related to bold coloration in the upper and lower wing (Tennyson et al. 2012, Bretagnolle pers. comment), collier complete: **(A.1)**: Dark grey morph, *P. brevipes* (Tanna, Eromango, Vanua Lava); **(A.2)**: Paler morph, *P. brevipes* (Tanna, Eromango); **(A.3):** Extreme dark grey morph, *P. brevipes magnificens* (Bank Isl.: "Vanua Lava II") (Bretagnolle and Shirihai 2010). **(B): Fiji** - *P. brevipes* (paler morph) from Viti Levu, Vanua Levu, and Kadavu. **(C):** *P. brevipes* from **Society - Tahiti** (paler morph and collier not complete (Bretagnolle pers. comm.), **Cook Isl.- Rarotonga** and **Australes Isl**. (dark grey morph and collier not complete (Bretagnolle pers. comments.) **(D)** *P. l. caledonica* **(E)** *P. l. leucoptera*: dark white lower wing coloration

Table 2.1. Details of the 2 live collared petrels caught on Vanua Lava compared with the type series of *P. brevipes magnificens* (Bretagnolle and Shirihai 2010). Table from (Tennyson et al. 2012)*.*

	7 March	8 March	*P. brevipes* magnificens
Band Number	D-201601	D-201602	
Bill-length (mm)	22.9	23.1	21.5-24.6
Bill-width (mm)	9.5	8.8	9.5-10.4
Bill depth (mm)	9.6	9.7	
Bill depth at gonys (mm)	7.8	7.8	7.1-8.3
Total head-length (mm)	57.3	58.6	
Wing length (mm)	215	217	206-222
Wingspan (cm)	67	70	
Tail-length (mm)	96	96	86-102
Tarsus (mm)	27.4	29.1	25.3-27.8
Mid-Toe (mm)	30.8	33.8	
Weight (g)	123	141	
Plumage score	Dark grey	Smoky or grey peppering	
Moult	All plumage fresh	All plumage fresh	
Iris colour	Brown	Brown	
Leg colour	Grey-blue	Pale blue with pink tinge	
Foot colour	Mostly black inner toe grey-blue with black joints, middle toe half grey-blue & inner third of webbing pink	Mostly black but inner toe blue with black joints, middle toe one-quarter blue & inner third of webbing pink	
Brood Patch	Scattered down	Covered in grey down	
Cloaca size	Small	Moderate	
Sex	Female	Female	

II.2.2.2. Distribution and ecology

The Australian and Caledonia populations have been treated as an endemic subspecies by Imber and Jenkins (1981) because of subtle morphological and coloration differences, though overlap in discriminating criteria exist (Bretagnolle and Shirihai 2010). The Australian subspecies, *P. l. leucoptera* (hereafter *leucoptera*) is now found breeding only on two tiny islets of New South Wales (Cabbage Tree and Boondelbah island: (Carlile et al. (2012), while the New Caledonian subspecies (*caledonica*) is restricted to the south-central chain of New Caledonia (de Naurois 1978, Bretagnolle and Shirihai 2010). Despite birds having been ringed in the two populations (especially on Cabbage Tree Is.), no exchange between the two has ever been documented. The *brevipes* form area of distribution has been reported on several islands of the Pacific Ocean: Gau, Viti Levu, Kadavu (Fiji), Rarotonga (Cook Islands), Vanua Lava, South Vanuatu, Eromango, and Society Islands, Tahiti, and Australes (French Polynesia). Furthermore, the breeding habitats and phenology of this species complex are quite different (Figure 2.2-2.4). The *leucoptera* and *caledonica* forms differ in breeding habitats (rocks under shrubs in *leucoptera* at sea level, steep soil, high in mountains in *caledonica*), breeding phenology (a lag of one month: Priddel et al. (2014) foraging areas during breeding as well as migration and non-breeding areas (Priddel et al. 2014) (Figure 2.5).The *brevipes* form nests on Gau, Fiji, between 100 and 500 m, in burrows on steep, well-forested slopes (MacGillivray 1860; Watling 1986). Young have been found in the nest on Fiji from May to August (MacGillivray 1860; Watling 1986), although on Vanuatu small downy young were found in February (MacGillivray 1860; Brooke 2004). It is perhaps relatively sedentary, tending to remain close to the breeding islands, although some are recorded to have dispersed east between about 10°N and 15°S almost to the Galápagos (Brooke 2004). Its diet is little known but is thought to consist chiefly of cephalopods and fish (MacGillivray 1860).

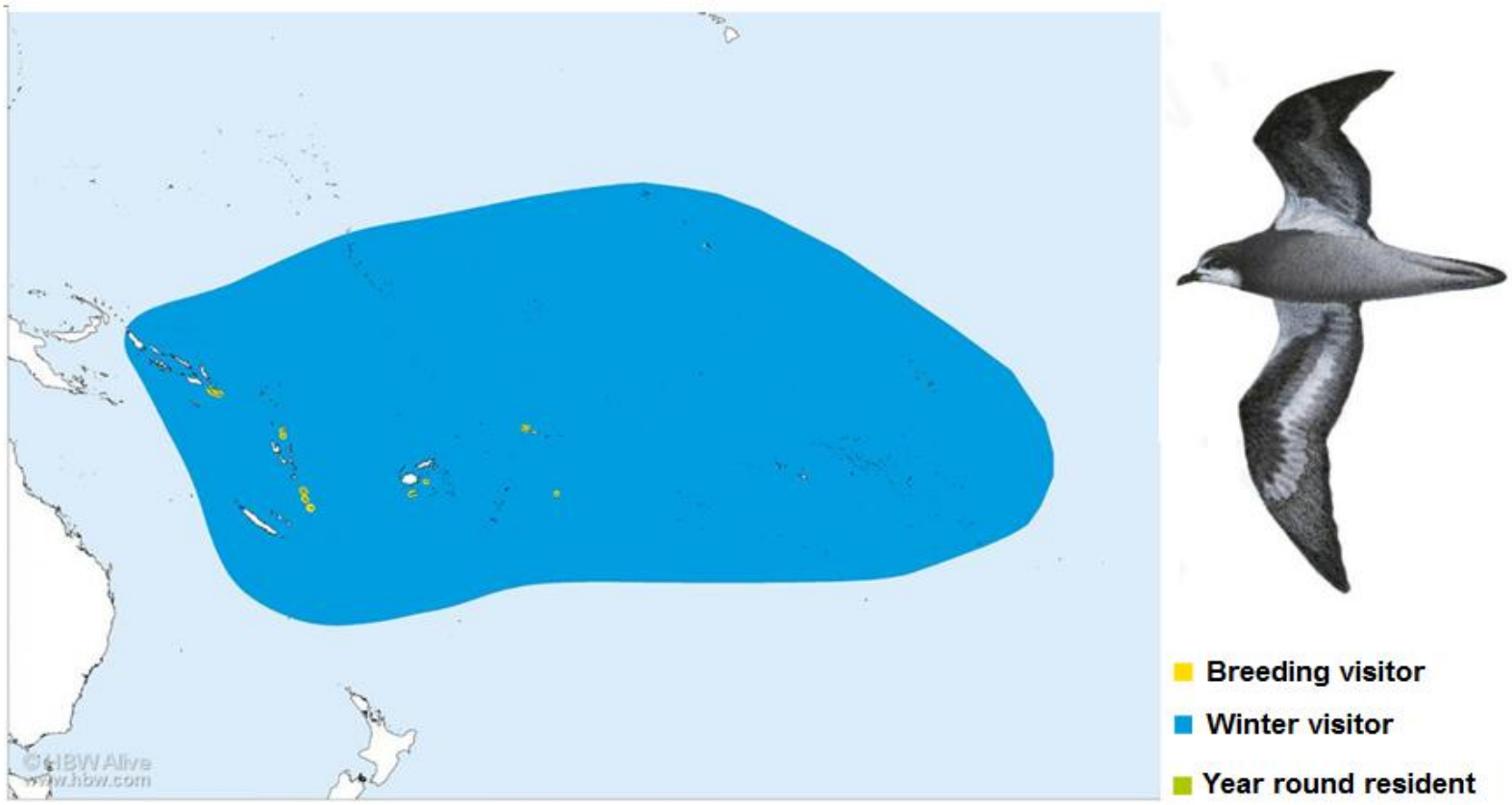


Figure 2.2. Breeding distribution and observations points of South West Pacific ocean colonies of *P.l. brevipes*. Modified from del Hoyo et al. (2016).

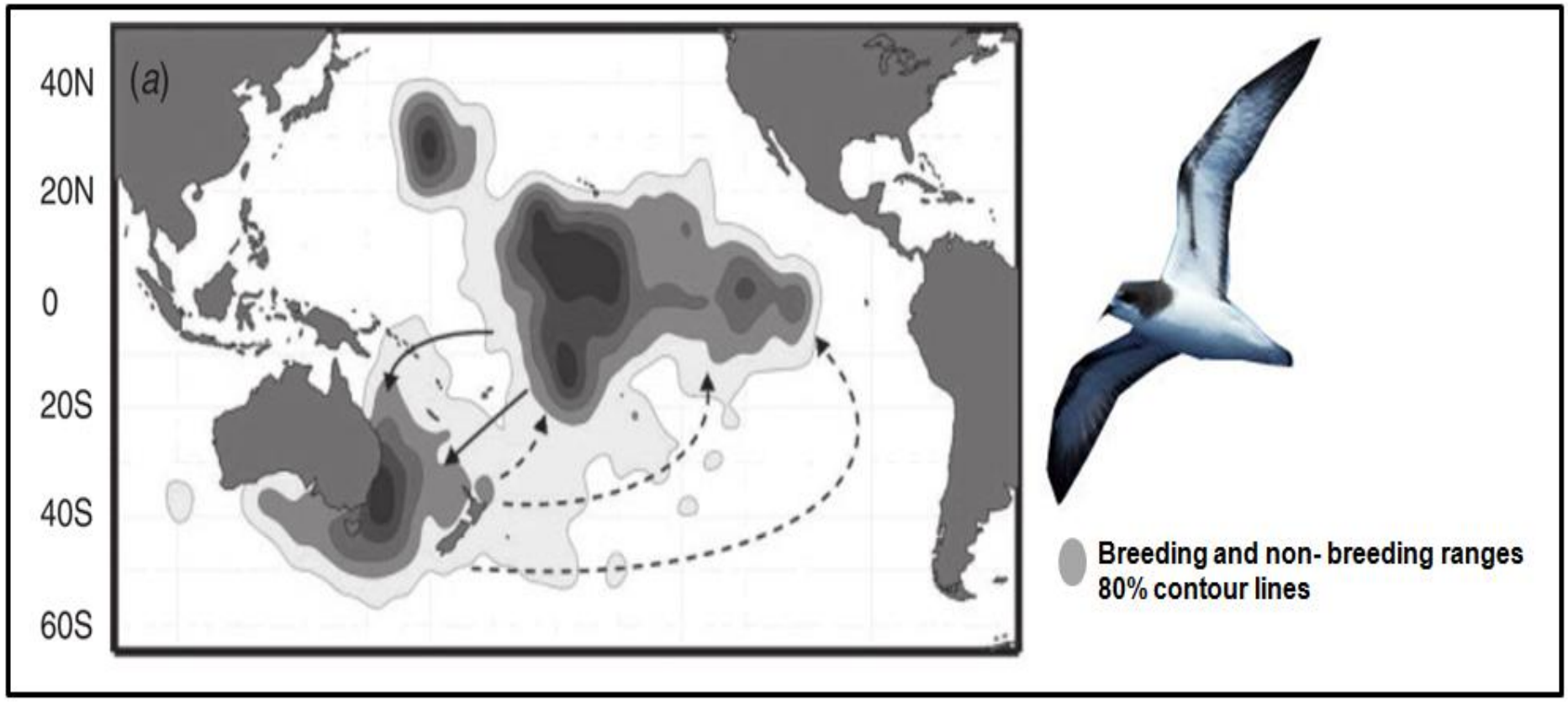


Figure 2.3. Kernel density distributions for *P.l. leucoptera* tracked with geolocators March 2010-February 2011. Shaded polygons represent the 20, 40, 60, 80 and 95% density contours. The approximate post breeding (dashed lines) and pre-breeding migration (solid lines) paths are shown. From Priddel et al 2014.

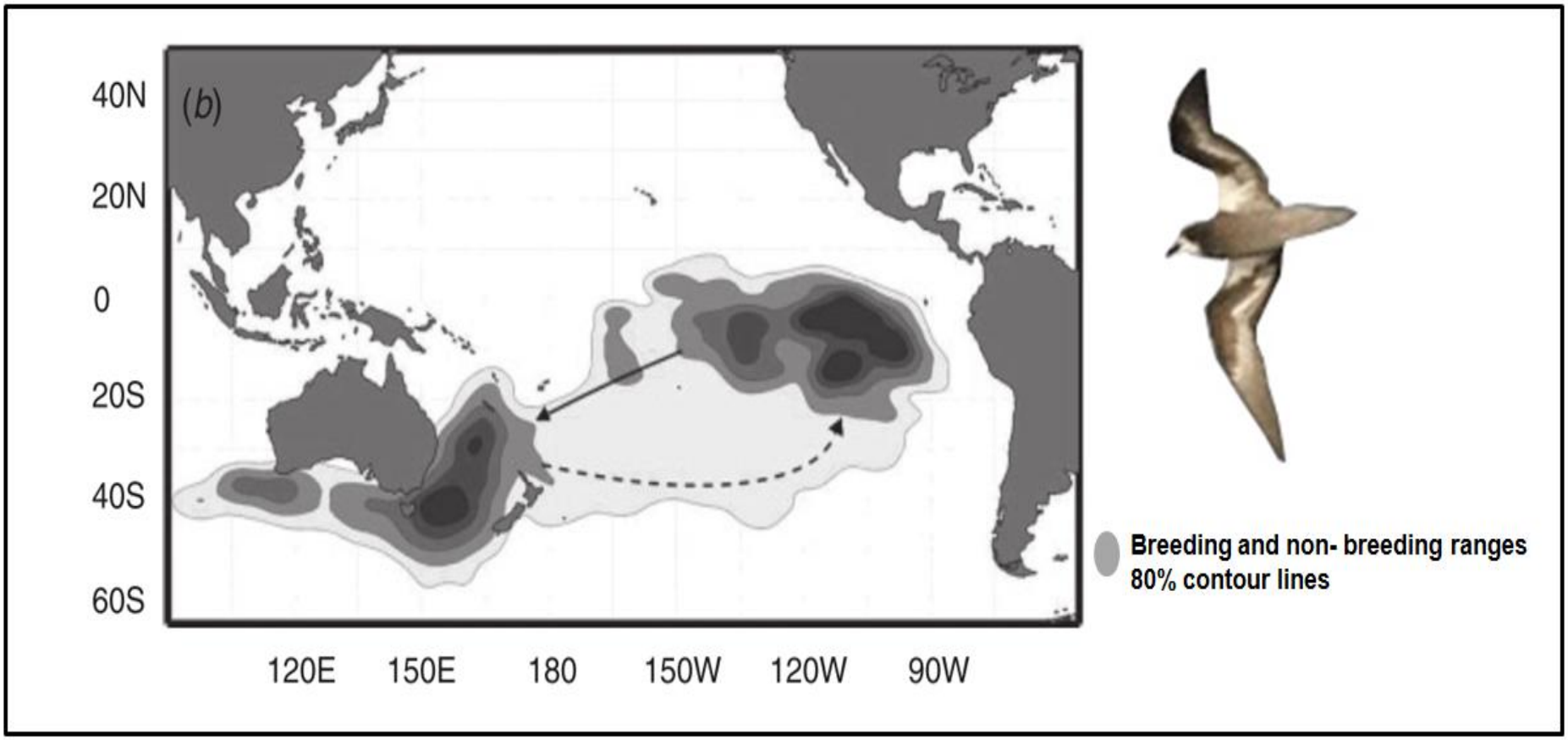


Figure 2.4. Kernel density distributions for *P.l. caledonica* tracked with geolocators January 2010 - January 2011. Shaded polygons represent the 20, 40, 60, 80 and 95% density contours. The approximate post-breeding (dashed lines) and pre-breeding migration (solid lines) paths are shown. From Priddel et al. 2014

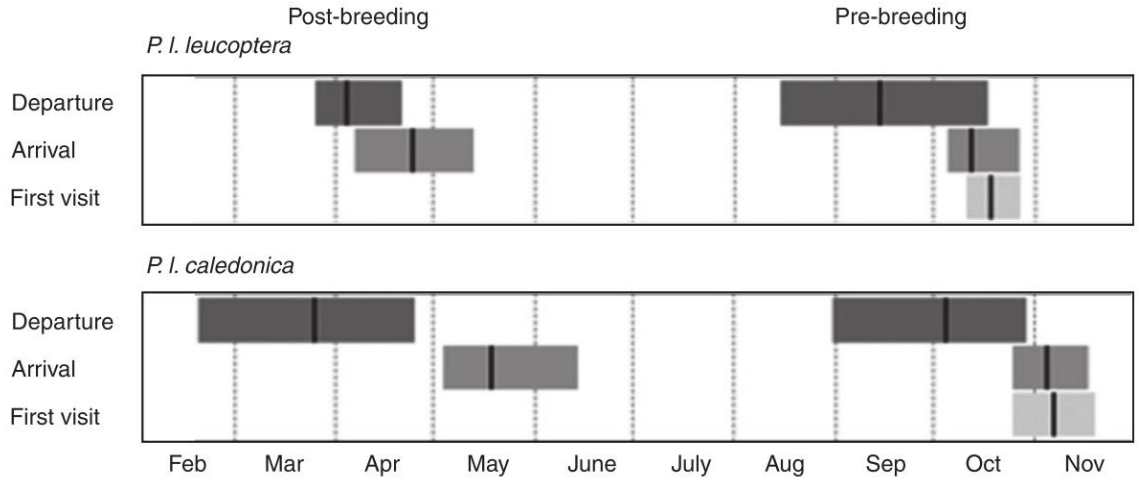


Figure 2.5. Migration schedule of *P.l. leucoptera* and *P.l. caledonica* as revealed by geolocation data. Post-breeding migration begins on departure from the breeding range and concludes on arrival at the non-breeding range. Pre-breeding migration is the return journey. The date of the first visit to the nesting burrow is also shown. Vertical lines indicate means; shaded areas show the range. From Priddel et al. (2014).

II.3.1.3.2. Population threats and conservation status

Caledonica and *leucoptera* taxa experienced recent fluctuations in population sizes: *leucoptera* was numerous when discovered in the eighteenth century Fullagar (1976), but decreased to a low of 1500 individuals in 1992 (c. 200 pairs), and was classified as *vulnerable* (IUCN 2015). More recently the population recovered to 1000 pairs thanks to a restoration program (Priddel and Carlile 1995; Priddel and Carlile 2009). The *caledonica* taxon is also supposed to have decreased in population size because of the recent introduction of predators (brown rats *Rattus rattus*, cats *Felis catus* and pigs *Sus scrofa*) with European settlement c. 300 years ago (IUCN 2015). G. Dutson in press (2005), suggest that the *brevipes* global population numbers 1,000-10,000 individuals, roughly equivalent to 670-6,700 mature individuals. The main threat to this form is thought to be predation by introduced mammals, especially the mongoose *Herpestes auro punctatus* (G. Dutson in press. 2005), but also pigs, cats, dogs and rats. Habitat degradation by goats is also believed to be a problem. Local communities in the Banks Islands are reported to have heavily exploited petrels and especially shearwaters for many generations until the 19th century, but in recent years they have apparently only been infrequently harvested at most (Totterman 2009). On Tanna, small numbers of young birds have been killed as part of a ritual (V. Bretagnolle in press. 2005).

REFERENCES

Ainley DG, LeResche RE, Sladen WJL (1983) Breeding biology of the Adelie penguin.

Austin J, Bretagnolle V, Pasquet E (2004) A global molecular phylogeny of the small Puffinus Shearwaters and implications for systematic of the Little-Audubon's

Shearwater complex. The Auk. 121:847–864.
Avise JC (2000) Phylogeography: the history and formation of species. Harvard university press.
Avise JC, Walker D (1998) Pleistocene phylogeographic effects on avian populations and the speciation process. Proceedings of the Royal Society of London B: Biological Sciences. 265:457–463.
Baillie SR, Milne H (1988) Movements of eiders Somateria mollissima on the east coast of Britain. Ibis (Lond 1859). 321–335.
Baker a J, Daugherty CH, Colbourne R, McLennan JL (1995) Flightless brown kiwis of New Zealand possess extremely subdivided population structure and cryptic species like small mammals. Proceedings of the National Academy of Sciences U S A. 92:8254–8258.
Baker A, Marshall H (1997) Mitochondrial control region sequences as tools for understanding evolution. In: Press A (ed) In: D.P. Mindell (ed.), Avian molecular systematics and evolution. New York. 51–82
Beheregaray LB (2008) Twenty years of phylogeography: The state of the field and the challenges for the Southern Hemisphere. Molecular Ecology. 17:3754–3774.
Birdlife International 2015. Speciesfactsheet:Pterodromaleucoptera. http://www.birdlife.org/datazone/speciesfactsheet.php?id=3887.
Birt-Friesen V, Montevecchi W, Gaston A, Davidson W (1992) Genetic structure of thick-billed murre (Uria lomvia) populations examined using direct sequence analysis of amplified DNA. Evolution (NY). 46:267–272.
Bretagnolle V (1993) Adaptive significance of seabird colouration: the case of Procellariiforms. Am Nat 142:141–173.
Bretagnolle V (1996) Acoustic communication in a group of nonpasserine birds, the petrels. In: (Ed. by Kroodsma, D. E. & Miller EH. (ed) Ecology and Evolution of Acoustic Communication in Birds. Ithaca, NY: Cornell University Press. 160 – 177.
Bretagnolle V, Shirihai H (2010) A new taxon of Collared Petrel Pterodroma brevipes from the Banks Islands , Vanuatu. 130:286–301.
Brooke M (2004) Albatrosses and Petrels across the World. Oxford.
Carlile N, Priddel D, Madeiros J (2012) Establishment of a new, secure colony of Endangered Bermuda Petrel Pterodroma cahow by translocation of near-fledged.
Cheang CC, Tsang LM, Ng WC, et al (2012) Phylogeography of the cold-water barnacle Chthamalus challengeri in the north-western Pacific: effect of past population expansion and contemporary gene flow. Journal of Biogeography. 39.10: 1819-1835.
Cushman R (1929) On the Pterodroma cookii and its allies. American Museum Novitates. 370:1-17.
De Naurois R (1978) Procellariidae reproducteurs en Nouvelle- Calédonie pendant l’été austral. Comptes Rendus de l'Académie des Sciences. Paris 287:269–271.
del Hoyo J, Elliott A, Sargatal J, et al (2016) Handbook of the Birds of the World Alive. In: Lynx Edicions, Barcelona. (retrieved from http://www.hbw.com/ on 22/12/2015.
Friesen VL, Montevecchi WA, Baker AJ, et al (1996) Population differentiation and evolution in the common guillemot Uria aalge. Molecular Ecology. 5:793–805.
Friesen VL, Burg TM, McCoy, KD (2007). Mechanisms of population differentiation in seabirds. Molecular Ecology.*16.*9:1765-1785.

Fullagar PJ (1976). Seabird islands No. 35: Cabbage Tree Island. New South Wales. Australian Bird Bander.*14*: 94-97.
Gangloff, B (2010) Systematics and phylogeography in gadfly petrels (Aves: Procellariiformes) and implications for conservation. Ph.D. Thesis, University of Poitiers, France.
Gangloff B, Zino F, Shirihai H, et al (2013) The evolution of north-east Atlantic gadfly petrels using statistical phylogeography. Molecular Ecology. 22:495–507.
Gill F, Mostrom A, Mack A (1993) Speciation in North American chickadees: I. Patterns of mtDNA genetic divergence. Evolution (NY).195–212.
Gómez-Díaz E, González-Solís J, Peinado MA, Page RDM (2006) Phylogeography of the Calonectris shearwaters using molecular and morphometric data. Molecular phylogenetics and evolution. 41:322–32.
Hatfield JS, Reynolds MH, Seavy NE, Krause CM (2012) Population dynamics of hawaiian seabird colonies vulnerable to sea-level rise. Conservation Biology. 26:667–78.
Hindwood K, Serventy D (1941) The Gould petrel of Cabbage Tree Island. Emu. 4.1:1-20.
Imber MJ, Jenkins JAF (1981) The New Caledonian Petrel. Department og Internal Affairs.
Iglesias-V A (2012) Genetic Structure of Pterodroma leucoptera complex, Hindwood and Serventy (1941). Project report for the degree of Master in Science. Nantes University, France.
IUCN (2015) The IUCN Red List of Threatened Species. Version 2015.1. Downloaded on 01 February 2015. <http://www.iucnredlist.org>.
Kimura M, Clegg SM, Lovette IJ, et al (2002) Phylogeographical approaches to assessing demographic connectivity between breeding and overwintering regions in a Nearctic-Neotropical warbler (Wilsonia pusilla). Molecular Ecology. 11:1605–1616.
Klicka J, Zink RM (1997) The importance of recent ice ages in speciation: a failed paradigm. Science. 80.277:1666–1669.
MacGillivray J (1860) Zoological notes from Aneiteum, New Hebrides. Zoologist. 18:7133–7142.
Mengel R (1964) The probable history of species formation in some northern wood warblers (Parulidae). Living Bird. 3:9–43.
Milot E, Lisle Gibbs H, Hobson KA (2000) Phylogeography and genetic structure of northern populations of the yellow warbler (Dendroica petechia). Molecular Ecology. 9:667–681.
Moum T, Arnason E (2001) Genetic diversity and population history of two related seabird species based on mitochondrial DNA control region sequences. Molecular Ecology. 10:2463–2478.
Paxinos EE, James HF, Olson SL, et al (2002) mtDNA from fossils reveals a radiation of Hawaiian geese recently derived from the Canada goose (Brantacanadensis). Proceedings of the National Academy of Sciences USA. 99:1399–1404.
Peck DR, Bancroft WJ, Congdon BC (2008) Morphological and molecular variation within an ocean basin in wedge-tailed shearwaters (Puffinus pacificus). Marine Biology. 153:1113–1125.
Pielou E (2008) After the ice age: the return of life to glaciated North America.

University of Chicago Press.
Pinet P, Salamolard M, Probst J, et al (2009) Barau's Petrel Pterodroma Baraui: history , biology and conservation of an endangered endemic petrel. 113:107–113.
Priddel D, Carlile N (1995) Mortality of adult Gould 's Petrels Pterodroma leucoptera leucoptera at the nesting site on Cabbage Tree Island , New South Wales. 95:259–264.
Priddel D, Carlile N (2009) Key elements in achieving a successful recovery programme: A discussion illustrated by the Gould's Petrel case study. Ecology Management Restoration. 10:97–102.
Priddel D, Carlile N, Portelli D, et al (2014) Pelagic distribution of Gould ' s Petrel (Pterodroma leucoptera): linking shipboard and onshore observations with remote-tracking data. Emu. 114.4: 360–370.
Saetre GP, Borge T, Lindell J, et al (2001) Speciation, introgressive hybridization and nonlinear rate of molecular evolution in flycatchers. Molecular Ecology. 10:737–749.
Schreiber E, Burger J (2002) Biology of marine birds.CRC Press.
Seitre R, Seitre J (1991) Causes de disparition des oiseaux terrestres de Polynésie française. South Pacific Commission.
Sibley CG, Monroe BL (1990) Distributions and taxonomy of birds of the world. Yale University Press.
Silva MC, Matias R, Wanless RM, et al (2015) Understanding the mechanisms of antitropical divergence in the seabird White-faced Storm-petrel (Procellariiformes: Pelagodroma marina) using a multilocus approach. Molecular Ecology. 24.12:3122–3137.
Smith AL, Monteiro L, Hasegawa O, Friesen VL (2007) Global phylogeography of the band-rumped storm-petrel (Oceanodroma castro; Procellariiformes: Hydrobatidae). Molecular phylogenetics and evolution. 43:755–73.
Steeves TE, Anderson DJ, Friesen VL (2005) The Isthmus of Panama: a major physical barrier to gene flow in a highly mobile pantropical seabird. Journal of Evolutionary Biology. 18:1000–8.
Techow NMSM, O'Ryan C, Phillips RA, et al (2010) Speciation and phylogeography of giant petrels Macronectes. Molecular phylogenetics and evolution 54:472–87.
Tennyson AJD, Miskelly CM, Totterman SL (2012) Observations of collared petrels (Pterodroma brevipes) on Vanua Lava, Vanuatu, and a review of the species' breeding distribution. Notornis. 59:39–48.
Thomson CE, Gilbert JDJ, Brooke MDL (2014) Cytochrome b divergence between avian sister species is linked to generation length and body mass. PLoS One 9:1–6. e85006.
Totterman S (2009) Vanuatu Petrel (Pterodroma occulta) discovered breeding on Vanua Lava, Banks Islands, Vanuatu. Notornis. 56:57–62.
van Bekkum M, Sagar PM, Stahl J-C, Chambers GK (2006) Natal philopatry does not lead to population genetic differentiation in Buller's albatross (Thalassarche bulleri bulleri). Molecular Ecology.15:73–9.
Warham J (1990) The petrels: their ecology and breeding systems. A&C Black.
Watling D (1986) Notes on the collared petrel Pterodroma (leucoptera) brevipes. Bulletin of British Ornithologist's Club.106:63–70.
Welch AJ, Yoshida AA, Fleischer RC (2011) Mitochondrial and nuclear DNA sequences reveal recent divergence in morphologically indistinguishable petrels.

Molecular Ecology. 20.7:1364–77.
Wennerberg L, Bensch S (2001) Geographic variation in the dunlin Calidris alpina as revealed by morphology, mtDNA and microsatellites genetic variation and migration of waders. Doctoral dissertation, PhD. Thesis. University of Lund.
Wojczulanis-Jakubas K, Kilikowska A, Harding AM, et al (2014) Weak population genetic differentiation in the most numerous Arctic seabird, the little auk. Polar Biology. 37.5:621–630
Zink RM (1994) The geography of mitochondrial DNA variation, population structure, hybridization, and species limits in the fox sparrow (Passerella iliaca). Evolution (NY). 96–111.
Zink R (1996) Species concepts, speciation, and sexual selection. Journal of Avian Biology. 27.1–6.
Zink R, Dittmann D (1993) Gene flow, refugia and evolution of gegraphic variation in the song sparrow (Melospiza melodia). Evolution (NY). 47:717–729.

CHAPTER III

GENERAL METHODS & SAMPLES

"Look deep into Nature and then

you will understand everything better"

~Albert Einstein

CHAPTER III. GENERAL METHODS

III.1. Study sites and sample collection

A total of 149 samples representing all the extant and possible origin-extinct populations were collected from eleven islands to this study (Figure 3.1). Because the samples had different origins, they were classified into two categories: Fresh and museum samples (Table 3.1).

Toe pad skin samples (1 mm long) were collected from five museums: the American Museum of Natural History of New York-AMNH, the Te papa Tongarewa Museum of New Zealand, the *Muséum National d'Histoire Naturell*e-MNHN and the *Musée de Tahiti et de Iles*- MTI (Table 3.1, 3.2). These samples were stored in microtubes containing ethanol at 90% and then frozen at -20°C until processing.

Table 3.1. Summary of taxa sampled, collection localities, provenance, and maximum sample size used per locality (N).

MORPHO SUBSPECIES	LOCATION	PROVENANCE	N	COLLECTOR
P. l. leucoptera	**Australia**		**66**	
	Cabbage Tree Island	Fresh	66	**NC**
P. l. caledonica	**New Caledonia**		**45**	
	Noumea	Fresh	45	**VB & PV**
P. brevipes	**Vanuatu Islands**		**27**	
	North Vanuatu		**9**	
	Vanua Lava I	Fresh	6	**VB & HS**
P. b. magnificens	Bank Isl. "Vanua Lava II"	AMNH Museum	3	**BG**
	South Vanuatu		**18**	
	Eromango	Fresh	15	**VB**
	Tanna	AMNH Museum	3	**BG**
	Fiji		**3**	
	Kadavu	AMNH Museum	1	**BG**
	Viti Levu	MTI Museum	2	**HS**
	Cook island		**1**	
	Rarotonga	Museum	1	**VB**
	Society		**6**	
	Tahiti	Fresh	4	**SOP**
		MTI Museum	2	**HS**
	Australes			
	Raivavae	TePapa, Museum	1	**VB**

TOTAL (N)			**149**	
SPECIES OUTGROUP				
Pterodroma oculta	**Bank Islands**	Fresh	**6**	**VB & HS**
Pterodroma madeira	**Madeira**	Genbank	**2**	
Pterodroma deserta	**Bugio**	Genbank	**1**	
Pterodroma feae	Cape verde	Genbank	**1**	
Puffinus lherminieri	**Martinique**	Fresh	**1**	**CP**
Puffinus bailloni	**Seychelles**	Fresh	**1**	**LC**

NC: Nicholas Carlile; **VB:** Vincent Bretagnolle; **BG:** Benoit Gangloff; **PV:** Pascal Villard; **HS:** Hadoram Shirihai; **CP:** Carine Precheur; **LC:** Licia Calabresse; **SOP:** Société d'Ornithologie de Polynésie-MANU; **FZ:** Frank Zino. **AMNH:** American Museum of Natural History of New York; **MTI:** *Musée de Tahiti et des îles*

Table 3.2. List of museum samples with their respective museum of provenance, identification number and sampling location.

MUSEUM	CATALOG NUMBER	ID_{thesis}	LOCATION	MORPHO SEPECIES
AMNH				
	216919	BN92	*Vanua Lava II*	*P. b. magnificens*
	216921	BN94	*Vanua Lava II*	*P. b. magnificens*
	216923	BN95	*Vanua Lava II*	*P. b. magnificens*
	336697	BN99	Tanna	*P. brevipes*
	336707	BO03	Tanna	*P. brevipes*
	336709	BO04	Tanna	*P. brevipes*
	250893	BN97	*Kadavu*	*P. brevipes*
	528336	*BO05*	*Viti Levu*	*P. brevipes*
MNHN				
	1876113	*BW69*	*Viti Levu*	*P. brevipes*
MTI				
	2004-3-12	*BO12*	*Society*	*P. brevipes*
	2004-3-13	*BO13*	*Society*	*P. brevipes*
Te Papa				
	OR.023110	*BN90*	*Rarotonga*	*P. brevipes*
	25377	*BO18*	*Raivavae*	*P. brevipes*

AMNH: American Museum of Natural History of New York; **MNHN:** *Muséum National d'Histoire Naturelle***; MTI:** *Musée de Tahiti et des îles*

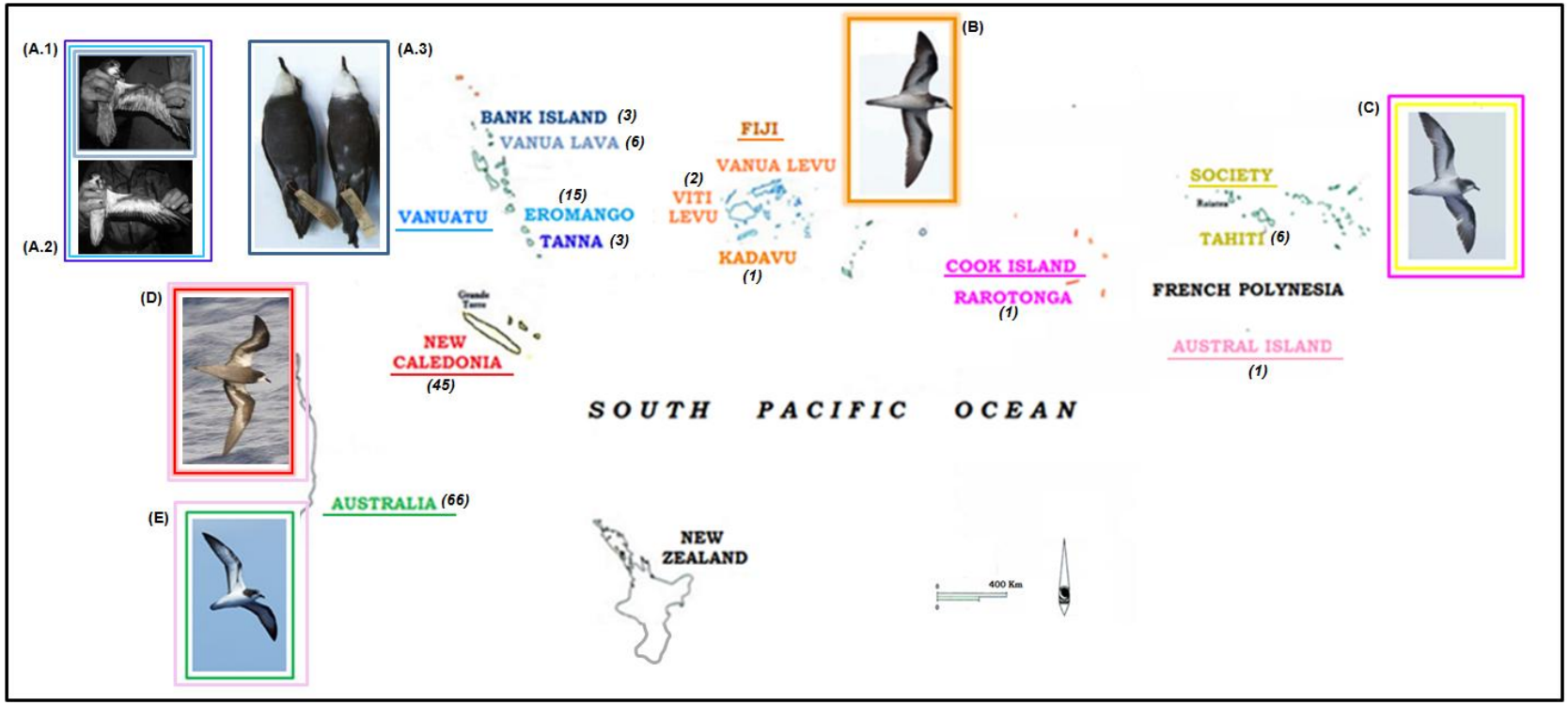


Figure 3.1 Sampling points *P. leucoptera complex.* **Vanuatu Island** - *P. brevipes* **(A.1), (A.2), (A.3)** (presence of both, dark grey morph and paler morph related to bold coloration in the upper and lower wing (Tennyson et al. 2012, Bretagnolle *pers. comment*), collier complete: **(A.1)**: Dark grey morph, *P. brevipes* (Tanna, Eromango, Vanua Lava); **(A.2)**: Paler morph, *P. brevipes* (Tanna, Eromango); **(A.3):** Extreme dark grey morph, *P. brevipes magnificens* (Bank Isl.: "Vanua Lava II") (Bretagnolle and Shirihai 2010). **(B): Fiji** - *P. brevipes* (paler morph) from Viti Levu, Vanua Levu, and Kadavu. **(C):** *P. brevipes* from **Society - Tahiti** (paler morph and collier not complete (Bretagnolle *pers. comments*)), **Cook Isl.- Rarotonga** and **Australes Isl**. (dark grey morph and collier not complete (Bretagnolle *pers. comments*.) **(D)** *P. l. caledonica* **(E)** *P. l. leucoptera:* dark white lower wing coloration. **The number in brackets represents** the total number of individuals collected per population. (See Appendix section for specific individuals used in each chapter).

Table 3.3 Total of individuals per population and morpho-subspecies indicating the type and ID identification used in this thesis.

MORPHO SUBSPECIES	LOCATION	Type	ID
P. l. leucoptera	**Australia**		**66**
	Cabbage Tree Island	Fresh	**66**
		Blood	BH57
		Blood	BH58
		Blood	BH59
		Blood	BH60
		Blood	BH61
		Blood	BH62
		Blood	BH63
		Blood	BH64
		Blood	BH65
		Blood	BH66
		Blood	BH67
		Blood	BH68
		Blood	BH69
		Blood	BH70
		Blood	BH71
		Blood	BH72
		Blood	BH73
		Blood	BH74
		Blood	BH75
		Blood	BH76
		Blood	BH77
		Blood	BH78
		Blood	BH79
		Blood	BH80
		Blood	BH81

Blood	BH82
Blood	BH83
Blood	BH84
Blood	BH85
Blood	BH86
Blood	BH87
Blood	BH88
Blood	BH89
Blood	BH90
Blood	BH91
Blood	BH92
Blood	BH93
Blood	BH94
Blood	BH95
Blood	BH96
Blood	BH97
Blood	BH98
Blood	BH99
Blood	BI01
Blood	BI02
Blood	BI03
Blood	BI04
Blood	BI05
Blood	BI06
Blood	BI07
Blood	BI08
Blood	BI09
Blood	BI10
Blood	BI11
Blood	BI12
Blood	BI13

		Blood	BI14
		Blood	BI15
		Blood	BI16
		Blood	BI17
		Blood	BI18
		Blood	BI19
		Blood	BI20
		Blood	BI21
		Blood	BI22
		Blood	BI23
P. l. caledonica	**New Caledonia**		**45**
	Noumea	Fresh	45
		Feather	BE06
		Feather	BE07
		Feather	BE08
		Feather	BE09
		Feather	BE10
		Feather	BE11
		Feather	BE12
		Feather	BE13
		Feather	BE14
		Feather	BE15
		Feather	BE16
		Feather	BE17
		Feather	BE18
		Feather	BE19
		Feather	BE20
		Feather	BE21
		Feather	BE22
		Feather	BE23
		Feather	BE24
		Feather	BE25
		Feather	BE26
		Feather	BE27
		Feather	BE28
		Feather	BE29
		Feather	BE30
		Feather	BE31
		Feather	BE32
		Feather	BE33
		Feather	BE34
		Feather	BE35

		Feather	BE36
		Feather	BE38
		Feather	BE39
		Feather	BT76
		Feather	BT77
		Feather	BT78
		Feather	BT81
		Feather	BT82
		Feather	BT83
		Feather	BT84
		Feather	BT85
		Feather	BT86
		Feather	BT87
		Feather	BT88
		Feather	BT89
P. brevipes	**Vanuatu Islands**		**27**
	North Vanuatu		**9**
	Vanua Lava I	Fresh	**6**
P.b. magnificens		Blood	BR06
		Blood	BR07
		Blood	BR08
		Blood	BR09
		Blood	BR10
		Blood	BR11
P.b. magnificens	Bank Isl. "Vanua Lava II"	AMNH Museum	**3**
		Toe pad skin	BN92
		Toe pad skin	BN94
		Toe pad skin	BN95
	South Vanuatu		**18**
	Eromango	Fresh	**15**
		Blood	BR12
		Blood	BR13
		Blood	BR14
		Blood	BR15
		Blood	BR16
		Feather	BR17
		Blood	BR18
		Blood	BR19
		Feather	BR20
		Blood	BR21
		Blood	BR22
		Blood	BR23

	Blood	BR24
	Blood	BR25
	Blood	BR26
Tanna	AMNH Museum	**3**
	Toe pad skin	BN99
	Toe pad skin	BO04
	Toe pad skin	BO03
Fiji		**3**
Kadavu	AMNH Museum	1
	Toe pad skin	BN97
Viti Levu	MTI Museum	2
	Toe pad skin	BW69
	Toe pad skin	BO05
Cook island		**1**
Rarotonga	Museum	1
	Toe pad skin	BN90
Society		**6**
Tahiti	Fresh	4
	Blood	BR32
	Blood	BR33
	Blood	BR34
	Feather	BW44
	MTI Museum	2
	Toe pad skin	BO12
	Toe pad skin	BO13
Australes		**1**
Raivavae	TePapa, Museum	1
	Toe pad skin	BO18

III.2. DNA extraction, PCR amplification, and DNA sequencing

III.2.1. Modern DNA

Modern DNA was collected from blood, feathers samples during different fieldwork seasons. Once the samples were collected they were stored in microtubes containing ethanol at 90% and then frozen at -20°C until processing. A small amount of blood was collected using FTAT cards (www.gelilifesciences.com/whatman) and stored at room temperature.

III.2.2. Historic and ancient DNA

The incorporation of ancient DNA (aDNA) in phylogeographic studies brought several challenges, mostly due to the high probability of contamination with more stable modern DNA (Willerslev and Cooper 2005). DNA extracted directly from museum specimens is often fragmented and damaged in several ways (Axelsson et al. 2008; Molak and Ho 2011) and it also almost certainly contains contaminant DNA (Hofreiter et al. 2001). Short DNA fragments (potentially contaminant) must be targeted and processed individually, greatly multiplying the workload needed to collect the same amount of genetic data from better preserved starting material. However, with the development of new aDNA technologies and aDNA criteria reliability (Willerslev and Cooper 2005; Barnett and Larson 2012; Shapiro and Hofreiter 2012), an increasing number of studies have been accomplished, ranging from the comparison of a few ancient samples to modern populations, to population-level analyses of both ancient and modern populations (Gilbert et al. 2005; Leonard 2008; Mccormack et al. 2015; Brown and Blois 2016; Linderholm 2016). This study applied the traditional Sanger sequencing, thus many limitations were expected and found.

III.2.3. DNA extraction

Total genomic DNA was isolated from the *Pterodroma leucoptera* complex and outgroup samples (*Pterodroma oculta, Pterodroma madeira, Puffinus lherminieri, Puffinnus bailloni*) using either the Qiagen DNasy ® Blood and Tissue Extraction Kits (Qiagen, Inc., Valencia, CA). for feather and blood samples or following a phenol-chloroform protocol procedure for toe pad skin museum samples (Sambrook et al. 1989).

III.2.3.1. *Molecular Markers: Mitochondrial and Nuclear DNA markers*

The development of molecular markers has enhanced the understanding of the evolutionary history of many taxa, revealing the genetic differentiation at intra and inter-population levels (Avise 2000; Friesen 2007). Over almost the last two decades, mitochondrial DNA (mtDNA) has been the most used molecular tool by geneticists due to its a) availability, and rapid, easy and inexpensive sequencing, b) non-recombining properties c) high mutation rate, and d) small effective population size (Ne) allowing lineage sorting to occur faster compared to nuclear DNA (nuDNA) (4Ne), thus allowing the detection of recent vicariance events (Brown 1985; Avise et al. 1987; Barrowclough and Zink 2009). However, several studies have revealed that mtDNA information alone may be incomplete when inferring populations structures or assessing species boundaries unless it is reinforced by other evidence such as nuclear gene data (Edwards et al. 2005; Rubinoff and Holland 2005; Welch et al. 2011a). Because all genes in mtDNA evolve as a single linkage unit, mtDNA gene analysis will always concur into a single gene tree which may misrepresent the whole

evolution history of the taxa. Furthermore, mtDNA reflects just matrilineal history, and this one window of evolution, could well differ compared to the overall population history if the divergence was initiated by a sex-biased dispersal (Zink and Barrowclough 2008).

For all the reasons mentioned above, Rubinoff and Holland (2005) suggest that researchers should compare mtDNA gene trees with multiple unlinked nuclear loci in order 1) to get access to both matrilineal and the whole lineage of evolution history and 2) to uncover by the nuclear genes the high variability of the mtDNA information.

III.2.3.1.1. Mitochondrial DNA markers

Mitochondria occupy a central position in the biology of cells and can represent as much as 25% of the volume of the cytoplasm. Mitochondrial DNA normally consist of a circular, double-stranded DNA molecule. Thirty-seven genes are encoded by mtDNA itself (22tRNas and two rRNAs). The additional 13 genes encode for subunits of the codon transport chain where carbohydrates and fats are oxidized to generate carbon dioxide, water, and ATP (Ballard and Whitlock 2004). Indeed, mitochondria are responsible for the majority of ATP production (Figure 3.2). From a molecular view, mitochondrial DNA represents a rapidly evolving locus that facilitates the observation of haplotypes (allele variation) in 50% or more of individuals, providing sufficient variation from which to draw inferences. It is known that the mutation rate of a genetic marker determines the density of observed changes on the sampled coalescent tree and hence its resolution in terms of the observed number of haplotypes or alleles, but does not affect its depth in time. Thus, a mitochondrial coding region may have a lower mutation rate than a non-coding control region (similarly an intron may have a lower rate than an exon) but this only affects the number of changes inferred on the tree, not its temporal scale. The mitochondrial and nuclear genomes differ in many other ways, including ploidy, mode of inheritance, degree of recombination, effective population size, and mutation rate. These differences are important when assessing the evolutionary history of the whole organism and/or population. Some limitations are associated with mtDNA, including maternal inheritance which allows for detection of female-mediated gene flow, only. Furthermore, the rapid rate of mtDNA evolution might leads to recurrent substitution at single base positions (homoplasy or saturation), which can obscure the signal of deeper history. It is expected that molecular markers to study evolution and ecology show substitutions that are neutral to selection (Friesen 2000). Ballard 2004 showed that the distribution of three mitochondrial types was non-random in populations of *Drosophyla simulans*. Therefore, it is recommended not to assume that mtDNA evolves as a strictly neutral marker because changes in the mtDNA sequence can have substantial impacts on the fitness of individuals (Ballard and Whitlock 2004).

The field of molecular evolutionary biology, principally phylogeographic inferences, were successfully developed with the realization that mitochondrial DNA (mtDNA) could be easily obtained from animals and that might be informative at a variety of taxonomic levels. Studies have used many different loci of the mitochondrial genome such as 12S rRNA (Balitzki-Korte et al. 2005; Melton and Holland 2007), 16s rRNA (Mitani et al.; Dubey et al. 2009) , COII (Roe and Sperling 2007; Alessandrini et al. 2008) and others. However, the main locus used in species discrimination is cytochrome b (cytb) (Kocher et al. 1989; Irwin et al. 1991). More

recently the use of cytochrome oxidase subunit I (COI) has increased owing primarily to its adoption by the barcode for life consortium (Kimura et al. 2002; Hebert et al. 2003; Kerr et al. 2007; Ward 2009; Mutanen et al. 2015).

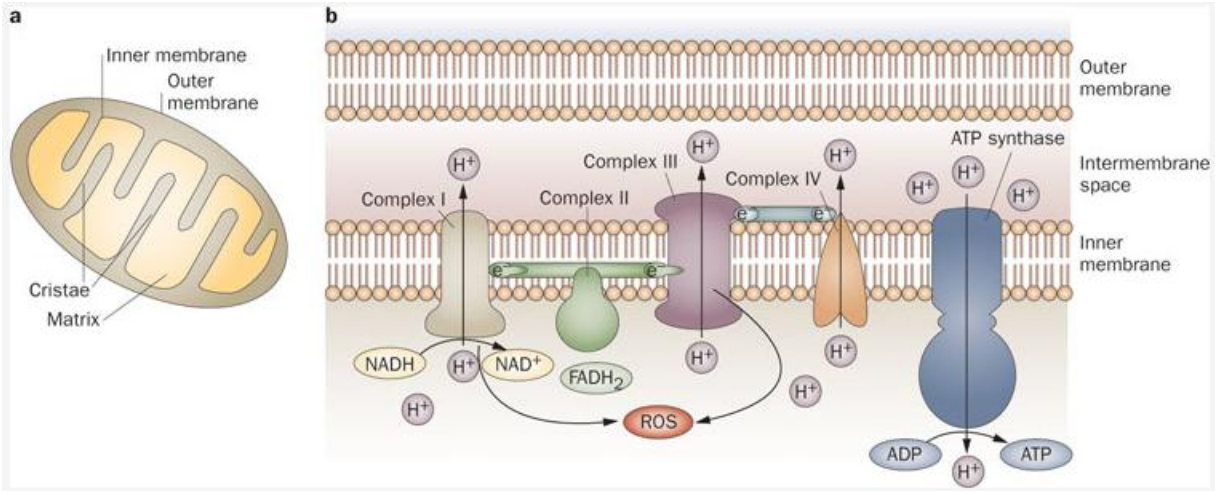


Figure 3.2. a) Mitochondria contain inner and outer membranes composed of phospholipid bilayers and proteins. The outer mitochondrial membrane encloses the entire organelle and contains porins that are involved in the transport of molecules ≤5 kDA. Larger molecules require active transport by mitochondrial membrane transport proteins. The inner mitochondrial membrane is compartmentalized into cristae, which expand the surface of the inner mitochondrial membrane and enhance its capacity to generate ATP.

III.2.3.1.2. Nuclear Introns

Introns are noncoding segments of DNA that interrupt the coding sequences (exons of nuclear genes of all eukaryotes (Figure 3.3). These untranslated gene regions of genomic DNA that are spliced out in the formation of mature RNA molecules and conveniently divided into groups, based on their splicing mechanism (Group I and II introns characterized by different self-splicing mechanisms and are found in some bacterial and organellar genomes (Hausner et al. 2006). Additionally, to them there is the group of introns that require a complex of five RNAs and hundreds of proteins, known as the spliceosome, to facilitate intron excision in the formation of mature messenger RNA molecules (Bonen and Vogel, 2001; Roy and Gilbert, 2006). This group are called the spliceosome introns and are the most common insertions in eukaryotic nuclear pre-mRNa genes. In eukaryotic introns, two types of spliceosome are recognized. The common U2-type splices GT-AG introns, so called because the introns start with 5’ GT and end with 3’ AG dinucleotides, and possess a characteristic pyrimidine-rich region that precedes the 3’ splice site (Stryer, 1988; Senapathy et al. 1990; Friesen, 2000). The second U12- type, splices the vary rare AT-AC introns, that have a number of dinucleotides at the 3’ end (Belshaw and Bensasson, 2006). Introns range in size ~50bp (base pairs) to tens of thousands of bp (Brown et al 1996), and sometimes contain microsatellites or exons for other loci. Their function is uncertain, but they may either provide sites for recombination between the functional domains of protein (De Souza et al 1998) or represent entirely nonfunctional "junk" DNA (Friesen 2000). Empirical data have shown that they can be considered as neutral markers that possess a number of traits that are desirable for molecular phylogenetics (Friesen et al. 1997, 2000). Indeed, the potential of nuclear introns to provide molecular markers for studies of

ecology and evolution results primarily from the fact that they are noncoding and essentially neutral to selection; the substitution rate for introns, therefore, is greater than for other single-copy nuclear DNA (Li and Graur 1991). Besides, because they are flanked by exons, which generally are under stabilizing selection, conserved sites are usually available for general polymerase chain reaction (PCR) primers that will amplify the desired target in a variety of species. This greatly reduces the amount of groundwork required for laboratory analyses compared to the analysis of microsatellites and increases the versatility of introns relative to allozymes, which require high-quality tissue samples. Also, because introns occur in virtually all structural genes and are scattered throughout the genome, they can provide potentially thousands of independent markers (Friesen 2000). Furthermore, sequence variation in introns can be determined directly, in contrast to methods such as protein electrophoresis and the analysis of restriction fragments, which do not detect a high proportion of variation, this enables of both sequence and frequency based statistical analyses. Non-coding introns are now routinely used in molecular systematics as independent markers (Oakley and Philips, 1999, van Oppen et al. 2000), or in concert with other gene partitions in an attempt to recover gene trees that are concordant with species trees in marine birds molecular phylogenetics (Silva et al. 2011; Welch et al. 2011b; Gangloff et al. 2013; Andrews et al. 2014; Silva et al. 2015; Silva et al. 2016). Nonetheless, few limitations have been reported for Introns as a tool for studying ecology and evolution. On a theoretical level, introns are nuclear genes and, therefore, have a larger effective population size than mitochondrial genes. Thus, they are less useful for studying recent population bottlenecks and restrictions in gene flow than is mtDNA (Friesen 2000). Along this view, in the present study, we have used five nuDNA markers and two mtDNA markers (Table 3.4) which were selected following Creer et al (2005) approach that was locating a suitable selection of primers that had a well performance in other studies or primers that have worked well in related taxa. Chosen CO1 and Cytb mtDNA and nuclear introns primers have showed a great ability to detect population substructure in other seabird species (Friesen and Anderson 1997; Prychitko and Moore 1997; Kimball et al. 2009a; Gangloff et al. 2010; Rayner et al. 2010; Welch et al. 2011a; Gangloff et al. 2011). Overlapping primers were designed to flank variable sites identified in an alignment of *Pterodroma* petrel sequences. More specifically, six, fourteen, three and three set of overlapping primers fragments pairs were designed for CO1, Cytb, β-fibint7; PAXIPI respectively using Primer3 (Table 3.4).

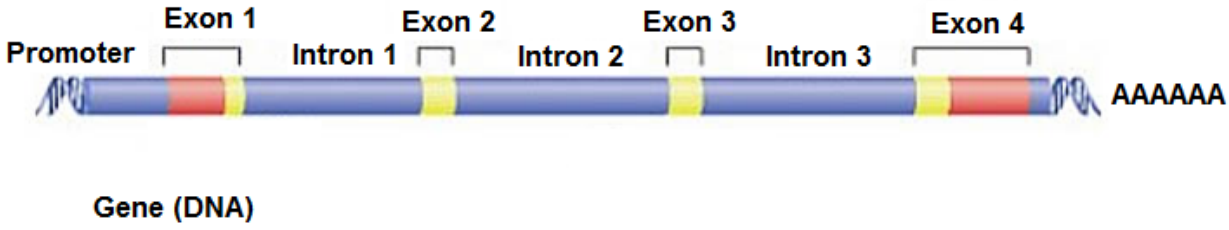


Figure 3.3. The structure of a typical eukaryotic gene, including introns and, exons. Promoter indicates the 5' cap. AAAAAA, represents the poly A tail. II.2.3.2. Primer design.

Table 3.4 Primers used for DNA amplification and sequencing.

GENE	PRIMER NAME	NUCLEOTIDE SEQUENCE (5' --->3')	SOURCE
Nuclear DNA			
PAX Interacting Protein 1-PAXIP1	PAX 20F PAX 21R	CCCTCAGACACTGGATTAYGAATCAT CCAAGGATTCCGAAGCAGTAAG	(Kimball et al. 2009)
	PAX-F-CRI PAX-R-CRI PAX-Fint-CRI	GTGTGGTTTACTAGAAGTTT GGCATTTTACTATCCATCAAG GTAGAATATTAAGCAACAGGA	*This study*
Cold shock domain-containing protein E1-CSDE1	CSDE5F CSDE6R	CTGGTGCTGTAAGTGCTCGTAAC CCAGGCTGTAAGGTTTCTAGGTCAC	(Kimball et al. 2009)
Beta-Fibrinogen intron 7-BFibint7	FIB-BI7U FIB-BI7L	GGAGAAAACAGGACAATGACAATTCAC TCCCCAGTAGTATCTGCCATTAGGGTT	(Prychitko and Moore 1997)
IRF2F	IRF2 2F IRF2 3R	ATGTCTTTGGGTCGGGTTTA GAAACTGGGCAATTCACACA	(Kimball et al. 2009)
	IRF2-F-Pter-CRI IRF2-R-int-CRI	AGGCAAATTAATAACAGCGTAGG TGTTGGGAGTAGAGCACACT	*This study*
TPM1	Trop 6aF Trop 6bR	AATGGCTGCAGAGGATAA TCCTCTTCAAGCTCAGCACA	(Primmer et al. 2002)
Mitochondrial DNA			
Cytochrome Oxidase 1-CO1	F1B R1B	AACCGATGACTATTYT-CAACC TACTACRTGYGARATGATTCC	(Gangloff et al. 2013)

	COI-F1-Pter-CRI	CACAARGATATYGGTACCCT	
	COI-R1-Pter-CRI	CAGTTCATCCTGTACCTGCY	
	COI-F2-Pter-CRI	GACATAGCATTCCCACGTATR	
	COI-R2-Pter-CRI	GTGAGAGTAAAAGTAGGACGG	
	COI-F3-Pter-CRI	GGGCAATCAACTTCATYACAAC	
	COI-R3-Pter-CRI	GAATGTAGACTTCTGGGTGGC	
Cytochrome b - Cytb	L14987	TATTTCTGCTTGATGAAACT	(Jesus et al. 2009)
	H16025	CTAGGGCTCCAATGATGGGGA	
	Cyt.Pter.F22	AGCCATGCACTACACACAGCCG	
	Cytb.Pter.R221	CGAAGGCAGTTGCTATGAGG	
	Cytb.Pter.F177	ATTCTAYTRTGGCTCCTACC	
	Cytb.Pter.R404	RCAAAAGGTAGGAGGAAGT	
	Cytb.Pter.F374	GGGATTCTCAGTAGATAACC	
	Cytb.Pter.R634	GTAAAGTTTTCTGGGTCTCC	
	CytB-F1-Pter-CRI	CTACTAGCCATRCACTACAC	*This study*
	CytB-R1-Pter-CRI	RAATGATATTTGTCCTCAGGG	
	CytB-F2-Pter-CRI	GGCTCCTACCTRTACAAAGAG	
	CytB-R2-Pter-CRI	GTAGGGGTGRAATGGGATT	
	CytB-F3-Pter-CRI	ATYCACCTCACCTTCCTCCA	
	CytB-R3-Pter-CRI	GGAATTGAGCGTAGGATAKCG	
	CytB-F4-Pter-CRI	GGAGACCCAGAAAACTTTACY	
	CytB-R4-Pter-CRI	TCAGAAGAGGAGTTGGGAGA	

III.2.3.3. Mitochondrial and nuclear DNA amplification and sequencing

We amplified a total of seven genes (two mitochondrial DNA and five nuclear introns) (Table 3.3). Polymerase chain reactions were carried out in a total volume of 25 ul utilizing 1X PCR Buffer, 1.5-2.5mM MgCl2; 0.2mM each dNTP, 0.5-0.7uM of each Primer, 0.06 units of GoTaq® Flexi DNA Polymerase (Promega) and 2-6ul of DNA extract. Amplifications were carried out under the following thermo cycle profile: an initial denaturation step at 95 °C for 8 min and then 30-35 cycles of 92 °C for 45 s, a primer-specific annealing temperature between 58 and 65 °C for 45 s and an extension step at 72 °C 30 s, followed by a final extension at 72 °C for 5 min. The amplified genes were sequenced at the enterprise Genoscreen, Lille-France and at the Genome Québec Innovation (McGill University Montreal, QC, Canada).

III.2.3.3.1. Editing and alignment of sequences

All sequence were checked visually, edited and aligned with CLUSTAL X (Thompson et al. 1997), implemented in BioEdit Sequence alignment Editor v.7.2.5 (Hall 1999). For the nuclear introns loci, since two alleles of each locus are amplified simultaneously, heterozygous sites are represented by double peaks (generally reduced in height relative to adjacent homozygous positions) in the resulting chromatograms. Therefore, forward and reverse sequences were visually checked and edited as necessary to ensure that all double peaks were correctly identified and labeled with standard degenerate nucleotide base codes, IUPAC code (Cornish-Bowden 1985). In cases that heterozygous positions could not be determined with confidence at least in one of the two directions ,the removal of this individuals from the dataset was applied. Insertions and deletions (indels) were also removed from the final datasets because they are not considered by some of the sofware packages.

III.2.3.3.2. Mitochondrial origin and intralocus recombination

Mitochondrial origin of mitochondrial concatenated gene sequences was confirmed by translating DNA sequences to check for stop codons and other potential indications of nuclear origin in BioEdit Sequence alignment Editor v. 7.2.5 (Hall 1999). Sequences were checked visually and aligned with CLUSTAL X (Thompson et al. 1997), implemented in BioEdit Sequence alignment Editor. Recombination in nuclear loci was tested with the four-gamete test (Hudson and Kaplan 1985) implemented in DnaSp v.5.10.01 (Librado and Rozas 2009). When the test suggested intralocus recombination, we retained the longest contiguous unrecombined sequence for subsequent analyses. Because the Isolation with Migration analysis requires having known phase for nuclear sequences, we determined the gametic phase using the program PHASE implemented in DnaSp (Stephens et al. 2001) with default parameters and a threshold value of 0.90. Phased haplotypic data were then used in all analyses, individuals thus being represented by two sequences.

III.3. STATISTICAL ANALYSIS

III.3.1. Polymorphism and neutrality tests

In order to describe the genetic diversity of each locus, summary statistics were calculated using DNAsp v. 5.10.01 software (Librado and Rozas 2009). For each locus, the number of segregating sites, the number of haplotypes, the haplotype diversity, and the nucleotide diversity were estimated for all loci. Two tests were used

to assess if genetic variation deviated from neutral expectations due to either a recent population expansion or selection: Tajima's D (Tajima 1989) and Fu's Fs (Fu 1997) both implemented in ARLEQUIN. Tajima's D values >0 suggest either a recent population bottleneck or some form of balancing selection, while D<0 indicates a population expansion or directional selection. These values are considered significant when p<0.05. Fu's Fs tends to be negative when there is an excess of recent mutations (therefore an overabundance of rare alleles), characteristic of a recent population expansion. Positive Fu's Fs values indicate a deficiency of rare alleles, suggesting a population bottleneck or overdominant selection. Following (Fu 1997) Fs values are regarded as significant if p<0.02. which corresponds to the conventional significant p<0.05 for Tajima'sD. Significant negative Tajima's and $\mathbf{F_s}$ indices may also indicate selection and genetic hitchhiking associated with selective sweeps. These analyses were conducted for all genes.

III.3.2. Population genetic structure

III.3.2.1. Haplotype networks and clustering inference

Haplotype frequencies were inferred with DnaSP. Genealogical relationships among haplotypes were reconstructed using a median-joining network (Bandelt et al. 1999) in NETWORK v.4.6.0.0 (http://www.fluxus-engineering.com). A second method used to corroborate or reject genetic population structuration (PAPER1) was BAPS (Bayesian Analysis of Population Structure) (Corander 2003). BAPS assumes that the target population is potentially genetically structured, such that boundaries limiting gene flow exist (or have existed). The extent and shape of such substructuring are typically unknown for natural populations.Analysis were launched with no prior information on geographic location. Replicated runs were performed on different values of k (differentiated clusters) from two to ten.

III.3.2.2. Pairwise comparisons of Φ_{ST} and analysis of molecular variance

The analysis of population genetic structure was also investigated using pairwise comparisons of Φ_{ST}, and through the analysis of molecular variance (AMOVA, Excoffier et al. (1992) as implemented in Arlequin v3.11. Φ_{ST} is a fixation index (Nei 1973) that incorporates both haplotype frequencies and the genetic distances between alleles to detect population genetic structure. Pairwise Φ_{ST} comparisons were calculated among populations using pairwise differences as the genetic distance measure. The significance was tested with 10 000 permutations and considering a significance level of 0.05, after Bonferroni corrections (Rice 1989). The AMOVA, which is also based on analyses of variance of gene frequencies and takes into account the distance between molecular haplotypes, was performed to partition the total sequence variation present in the dataset in different hierarchical levels: between groups, within groups. AMOVA intrinsic calculates pairwise differentiation between populations $\Phi_{ST,}$ a direct analogue of Wright's F_{st} for nucleotide sequence divergence data.

III.3.3. Phylogenetic and evolutionary relationships

Coalescent theory (Kingman 1982; Nei and Tajima 1983) which models genealogies within populations, can be used to investigate probabilities that gene trees have branching patterns (topologies) that differ from a species tree topology. Species tree, a tree of ancestor–descendant relationships for a set of populations, where branch lengths depend on time measured in number of generations and on

effective population sizes. The basic model called "multispecies coalescent", generalizes the Wright-Fisher model of genetic drift (Nordborg 2001; Wakeley 2009), applying it to multiple populations connected by an evolutionary tree. Multispecies coalescent model, thus is applied to gene trees in a species tree; this model is used to assemble separate coalescent processes occurring in populations connected by an evolutionary tree. The coalescent for a single population traces the ancestries of a subset of individual copies of a gene backward in time from the present. The population is assumed to have a constant size and nonoverlapping generations. Each gene is copied from a random parental gene in the previous generation. The coalescent model approximates the process of choosing random parents backward in time when the population size is large relative to the number of sampled lineages (Degnan and Rosenberg 2009). Along this view, *BEAST (Heled and Drummond 2010), an algorithm implemented in the software BEAST v1.6.1 (Drummond and Rambaut 2007), was used to estimate a species tree. This approach considers the genealogies of the different loci included in the analysis, i.e. coestimates multiple gene trees from multiple individuals per species, to estimate the tree of species relationships (Degnan and Rosenberg 2009).

III.3.4. Coalescent methods based in MCMC to estimate effective population sizes, gene flow and population divergence time

The integration of coalescent theory into a statistical framework has led to the growing development of coalescent-based methods that analyze genetic diversity among a sample of DNA sequences to infer population demographic history. The fundamental relationship exploited by these coalescent-based methods is between the distribution of divergence times among individuals and effective population size. This relationship is embodied by the genetic diversity parameter $\theta = 4N_e\mu$, where Ne is the effective population size ($2N_e\theta\mu$ for haploids) and μ is the mutation rate. Early developments in the estimation of θ were based on summary statistics, such as the number of segregating sites among DNA sequences or the mean number of nucleotide differences in all pairwise comparisons (Nei 1987). However, this approach is limited because it does not take into account the genealogical structure of the data (Felsenstein 1992). Recent developments have tried to remedy this using Maximum likelihood estimates incorporating METROPOLIS-HASTINGS sampling and integration across several phylogenies (Kuhner et al. 1995).

III.3.5. Likelihood estimation by Metropolis-Hastings sampling

Rather than using a single estimate of the genealogy for a population sample of DNA sequences, Metropolis-Hastings sampling allows for genealogical uncertainty when estimating (θ) (Kuhner et al. 1995). This is done by not only analysing the phylogenetic tree that best describes the DNa sequence data, but also a sample of those trees that are less probable. However, compared with other methods for inferring population demographic history, this approach is computationally demanding, because algorithms are based in Monte Carlo resampling, Markov Chain Monte Carlo algorithms.

A Markov chain is a series of stochastic events whereby the state of the process at the next time step, t+1, depends on: 1) The current state of the process (e.g., contained in a state matrix) 2) The probability of changing to another state in the next time step (e.g., defined in a transition matrix). Some Markov chains possess a unique equilibrium distribution. Informally, this means that if we start the chain

somewhere in the state space and run the chain long enough, then the chain will settle into an equilibrium distribution that does not depend on the initial condition. We say that the chain becomes stationary.

III.3.6. Stationarity and convergence

Reaching stationarity means that the current state of the chain no longer depends on the initial conditions of the chain. This property implies that if we start the chain from different points in the state space, then we will eventually end up sampling in the same place, no matter where we started.

III.3.6.1. Visual method: Diagnostic plots

One of the easiest ways to determine that stationarity has not been achieved is to simply visualize the state of the chain through "time" (iterations, sometimes called generations).

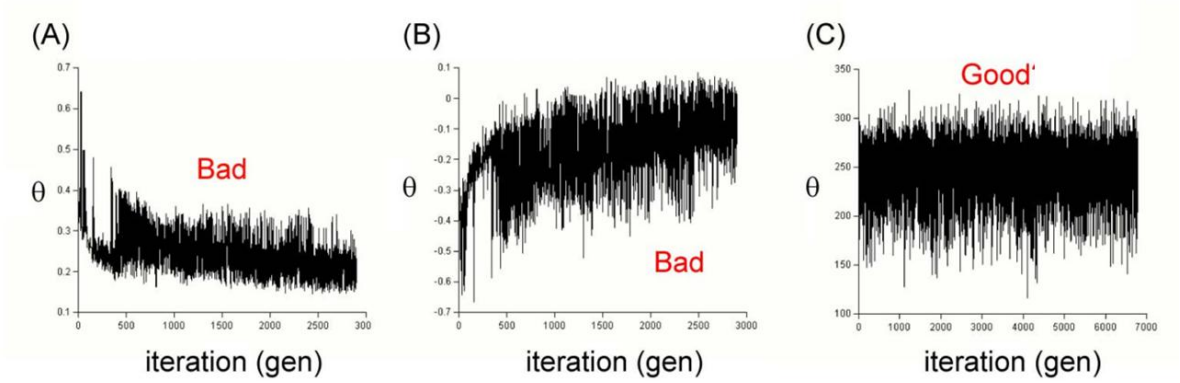


Figure 3.4. Time series plots of model parameter θ as the chain progresses. In plots A and B, iteration zero was the initial starting value of θ. Neither of these chains has reached stationarity as is evident by their trend. In plot C, the burn-in period has been removed, and the chain appears to be stationary.

III.3.7. Coalescent estimates of gene flow, divergence times and population size based on the Isolation with Migration model

Isolation with migration family models (IM, IMa and IMA2) (Hey and Nielsen 2004; Hey 2005, 2007) implements a coalescent-based method that uses MCMC sampling of gene genealogies to estimate population parameters scaled by mutation rate μ, such as divergence time t (t = Tμ, where T is the time since divergence in years), population size θ (θ = 4Neμ (for a nuclear locus) or θ = 2Neμ (for a mitochondrial locus), where Ne is the effective population size) and gene flow m (m = M/μ, where M is the effective immigration rate). This method is particularly appropriate for cases of recent divergences, since it does not assume mutation-drift-migration equilibrium, and thus allele sharing may be due to either gene flow and/or incomplete lineage sorting (Hey and Nielsen 2007). Here patterns of historical and contemporary connectivity between the subspecies of *P. leucoptera* complex were disentangled through a coalescent inference using the Isolation with Migration under Changing Population Size model (IM, Hey and Nielsen 2004; Hey 2005). This model in comparison to the other IM family models implements the the splitting parameter (S) allowing for population size change through time.

III.3.8. Demographic population size changes

The Tajima's D neutrality test assumes equilibrium and the presence of random mating. Therefore, a significant deviation of D from 0 does not necessarily mean that the neutral theory is not applicable (Tajima 1993), but factors/natural forces other than selection (population expansion, bottleneck, heterogeneity of mutation rates) could have been modeling the patterns of DNA polymorphism. Thus this test was used for all population and loci to detect possible demographic changes. The Fu's F_S neutrality test was also performed for all loci and populations (Fu 1997). Although it also tests the neutral mutation hypothesis, it is very sensitive to population demographic expansion, which generally leads to large negative F_S values. In addition, the population demography through time was investigated for each population with the Extended Bayesian Skyline Plots (Heled and Drummond 2008) implemented in BEAST v1.6.1 (Drummond and Rambaut 2007). This approach uses the differences in coalescence times among lineages to estimate population size changes through time. The analysis used all the loci with the same mutation rate and models of substitution that were used in estimating the species tree.

REFERENCES

Alessandrini F, Mazzanti M, Onofri V, et al (2008) MtDNA analysis for genetic identification of forensically important insects. Forensic Sci Int Gen. Sup 1:584–585.

Andrews KR, Norton EL, Fernandez-Silva I (2014) Multilocus evidence for globally distributed cryptic species and distinct populations across ocean gyres in a mesopelagic copepod. 1:5462–5479.

Avise JC (2000) Phylogeography: the history and formation of species. Harvard university press.

Avise JC, Arnold J, Martin Ball R, et al (1987) Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. Annu. Rev. Ecol. Syst. 489-522.

Axelsson E, Willerslev E, Gilbert MTP, Nielsen R (2008) The effect of ancient DNA damage on inferences of demographic histories. Mol Biol Evol. 25:2181–7.

Balitzki-Korte B, Anslinger K, Bartsch C, Rolf B (2005) Species identification by means of pyrosequencing the mitochondrial 12S rRNA gene. Int J LegMed 119:291–294.

Ballard JWO, Whitlock MC (2004) The incomplete natural history of mitochondria. Mol Ecol. 13:729–744.

Bandelt HJ, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol. 16:37–48.

Barnett R, Larson G (2012) A phenol-chloroform protocol for extracting DNA from ancient samples. Ancient DNA: Methods and Protocols. 840:13–19.

Bonen L, Vogel J (2001) The ins and outs of group II introns. Trends. Genet. 17:322–331.

Barrowclough GF, Zink RM (2009) Funds enough, and time: mtDNA, nuDNA and the discovery of divergence. Mol Ecol. 18:2934–2936.

Bretagnolle V, Shirihai H (2010) A new taxon of Collared Petrel Pterodroma brevipes from the Banks Islands , Vanuatu. 130:286–301.

Brown JH, Stevens GC, Kaufman DM (1996) The geographic range: size, shape,

boundaries, and internal structure. Annu. Rev. Ecol. Syst. 27, 597–623.
Brown SK, Blois JL (2016) Ecological insights from ancient DNA. eLS 1–7.
Brown WM (1985) The mitochondrial genome of animals, in: MacIntyre R.J. ed. Molecular Evolutionary Genetics. Plenum Press, New York., pp 95–130.
Corander J (2003) Bayesian analysis of genetic population structure using BAPS.
Cornish-Bowden A (1985) Nomenclature for incompletely specified bases in nucleic acid sequences: Recommendations 1984. Nucleic Acids Res. 13:3021–30.
Creer S, Malhotra A, Thorpe, RS, et al 2005. Targeting optimal introns for phylogenetic analyses in non-model taxa: experimental results in Asian pitvipers. Cladistics. 21, 390–395.
Degnan J, Rosenberg N (2009) Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends Ecol Evol. 24:332–40.
De Souza SJ, Long M, Klein RJ, et al, (1998) Towards a resolution of the introns early/late debate: only phase zero introns are correlated with the structure of ancient proteins. Proceedings of the National Academy of Sciences of the United States of America. 95, 5094–5099.
Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol. 7:214.
Dubey B, Meganathan P, Haque I (2009) Multiplex PCR assay for rapid identification of three endangered snake species of India. Conserv Genet. 10:1861–1864.
Edwards SV, Kingan SB, Calkins JD, et al (2005) Speciation in birds: genes, geography, and sexual selection. Proc Natl Acad Sci USA. 102 Suppl :6550–7.
Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. Genetics 131:479–491.
Felsenstein J (1992) Estimating effective population size from samples of sequences: inefficiency of pairwise and segregating sites as compared to phylogenetic estimates. Genet Res. 59:139–147.
Friesen VL (2000) Introns. In: Blackwell (ed) Molecular methods in ecology. Oxford, pp 274–290
Friesen VL (2007) New roles for molecular genetics in understanding seabird evolution , ecology and conservation. 96:89–96.
Friesen VL, Anderson DJ (1997) Phylogeny and Evolution of the Sulidae (Aves : Pelecaniformes): A Test of Alternative Modes of Speciation. 7:252–260.
Friesen VL, Congdon BC, Walsh HE, Birt TP (1997) Intron variation in marbled murrelets detected using analyses of single-stranded conformational polymorphisms. Mol Ecol. 6:1047–58.
Fu Y-X (1997) Statistical Tests of Neutrality of Mutations Against Population Growth, Hitchhiking and Background Selection.
Gangloff, B (2010) Systematics and phylogeography in gadfly petrels (Aves: Procellariiformes) and implications for conservation. Ph.D. Thesis, University of Poitiers, France.
Gangloff B, Shirihai H, Watling D, et al (2011) The complete phylogeny of Pseudobulweria, the most endangered seabird genus: systematics, species status and conservation implications. Conserv Genet. 13:39–52.
Gangloff B, Zino F, Shirihai H, et al (2013) The evolution of north-east Atlantic gadfly petrels using statistical phylogeography. Mol Ecol. 22:495–507.
Gilbert M, Bandelt H, Hofreiter M, Barnes I (2005) Assessing ancient DNA studies. Trends Ecol Evol. 20:541–544.

Hall T (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. 41:95–98.
Hausner G, R O, Simon D (2006) Origin and evolution of the chloroplast trnK (matK) intron: A model for evolution of group II intron RNA structures. Mol Biol Evol 23:280–391.
Hebert PD, Cywinska A, Ball SL, (2003) Biological identifications through DNA barcodes. Proc Biol Sci. 270: 313-321.
Heled J, Drummond AJ (2010) Bayesian inference of species trees from multilocus data. Mol Biol Evol. 27:570–80.
Heled J, Drummond AJ (2008) Bayesian inference of population size history from multiple loci. 15:1–15.
Hey BJ (2007) Using the IMa Program. 2785–2790. Distributed by the author: http://lifesci.rutgers.edu/~heylab/heylabsoftware.htm.
Hey J (2005) On the number of new world founders: A population genetic portrait of the peopling of the Americas. PLoS Biol. 3:0965–0975.
Hey J, Nielsen R (2004) Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of Drosophila pseudoobscura and D. persimilis. Genetics 167:747–60.
Hey J, Nielsen R (2007) Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. Proc Natl Acad Sci USA. 104:2785–2790.
Hofreiter M, Serre D, Poinar HN, et al (2001) Ancient DNA. Nat Rev Genet. 2:353–9.
Hudson RR, Kaplan NL (1985) Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics. 111:147–64.
Irwin DM, Kocher TD, Wilson AC (1991) Evolution of the cytochrome b gene of mammals. J Mol Evol. 32:128–144.
Li, W.-H. and D. Graur. 1991. Fundamentals of Molecular Evolution. 284pp. Sinauer Associates, Sunderland, Mass.
Jesus J, Menezes D, Gomes S, et al (2009) Phylogenetic relationships of gadfly petrels Pterodroma spp. from the Northeastern Atlantic Ocean: molecular evidence for specific status of Bugio and Cape Verde petrels and implications for conservation. Bird Conserv Int. 19:199.
Kerr KCR, Stoeckle MY, Dove CJ, et al (2007) Comprehensive DNA barcode coverage of North American birds. Mol Ecol Notes. 7:535-543.
Kimball RT, Braun EL, Barker FK, et al (2009) A well-tested set of primers to amplify regions spread across the avian genome. Mol Phylogenet Evol 50:654–60.
Kimura M, Clegg SM, Lovette IJ, et al (2002) Phylogeographical approaches to assessing demographic connectivity between breeding and overwintering regions in a Nearctic-Neotropical warbler (Wilsonia pusilla). Mol Ecol 11:1605–1616.
Kingman JFC (1982) The coalescent. Stoch Process their Appl 13:235–248.
Kocher TD, Thomas WK, Meyer a, et al (1989) Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. Proc Natl Acad Sci USA. 86:6196–6200.
Kuhner MK, Yamato J, Felsenstein J (1995) Estimating effective population size and mutation rate from sequence data using metropolis-hastings sampling. Genetics 140:1421–1430.
Leonard J (2008) Ancient DNA applications for wildlife conservation. Mol Ecol 17:4186–4196.

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451–2.
Linderholm A (2016) Ancient DNA: The next generation-chapter and verse. Biol J Linn Soc. 117:150–160.
Mccormack JE, Tsai WLE, Faircloth BC (2015) Sequence capture of ultraconserved elements from bird museum specimens. Mol Ecol Resour.
Melton T, Holland C (2007) Routine forensic use of the mitochondrial 12S ribosomal RNA gene for species identification. J Forensic Sci. 52:1305–1307.
Mitani T, Akane A, Tokiyasu T, et al (2009) Identification of animal species using the partial sequences in the mitochondrial 16S rRNA gene. Legal Medicine. 11, S449-S450.
Molak M, Ho SYW (2011) Evaluating the impact of post-mortem damage in ancient DNA: A theoretical approach. J Mol Evol. 73:244–255.
Mutanen M, Kekkonen M, Prosser SWJ, et al (2015) One species in eight : DNA barcodes from type specimens resolve a taxonomic quagmire. 967–984.
Nei M (1987) Molecular evolutionary genetics. Columbia university press.
Nei M, Tajima F (1983) Maximum likelihood estimation of the number of nucleotide substitutions from restriction sites data. Genetics. 105:207–217.
Nei M (1973) Analysis of gene diversity in subdivided populations. Proc Nat Acad Sci USA. 70:3321–3323.
Nordborg M (2001) Coalescent theory. In Handbook of Statistical Genetics, Balding, D. Wiley, pp 179–212
Oakley TH, Phillips RB (1999) Phylogeny of Salmonine fi shes based on growth hormone introns: Atlantic (Salmo) and Pacific (Oncorhyn- chus) salmon are not sister taxa. Mol. Phyl. Evol. 11:381–393.
Primmer CR, Borge T, Lindell J, Sætre GP (2002) Single-nucleotide polymorphism characterization in species with limited available sequence information: High nucleotide diversity revealed in the avian genome. Mol Ecol. 11:603–612.
Prychitko TM, Moore WS (1997) The utility of DNA sequences of an intron from the beta-fibrinogen gene in phylogenetic analysis of woodpeckers (Aves: Picidae). Mol Phylogenet Evol. 8:193–204.
Rayner MJ, Carraher CJF, Clout MN, Hauber ME (2010) Mitochondrial DNA analysis reveals genetic structure in two New Zealand Cook's petrel (Pterodroma cookii) populations. Conserv Genet. 11:2073–2077.
Rice WR (1989) Analyzing tables of statistical tests. Evolution. 43: 223–225.
Roe A, Sperling F (2007) Patterns of evolution of mitochondrial cytochrome c oxidase I and II DNA and implications for DNA barcoding. Mol Phylogenet Evol 325–345.
Roy SW, Gilbert, W. 2006. The evolution of spliceosomal introns: patterns, puzzles and progress. Nat. Rev. Genet. 7:211–221.
Rubinoff D, Holland B (2005) Between Two Extremes: Mitochondrial DNA is neither the Panacea nor the Nemesis of Phylogenetic and Taxonomic Inference. Syst Biol. 54:952–961.
Sambrook, Fritsch, Maniatis (1989) Molecular Cloning: A Laboratory Manual, 2nd Edition.
Senapathy P, Sahpiro M, Harris N (1990). Splice junctions, brunch point sites, and exons: sequence statistics, identification, and application to genome project. Methods in Enzymology. 183: 252–278.
Shapiro B, Hofreiter M (2012) Ancient DNA : Methods and protocols. Humana Press, New York.

Silva MC, Duarte MA, Coelho MM (2011) Anonymous nuclear loci in the white-faced storm-petrel Pelagodroma marina and their applicability to other Procellariiform seabirds. J Hered. 102:362–5.
Silva MC, Matias R, Wanless RM, et al (2015) Understanding the mechanisms of antitropical divergence in the seabird White-faced Storm-petrel (Procellariiformes: Pelagodroma marina) using a multilocus approach. Mol Ecol. 24:3122–3137.
Silva MF, Smith AL, Friesen VL, et al (2016) Mechanisms of global diversification in the marine species Madeiran Storm-petrel Oceanodroma castro and Monteiro's Storm-petrel O. monteiroi: Insights from a multi-locus approach. Mol Phylogenet Evol. 98:314–323.
Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. Am J Hum Genet. 68:978–989.
Stryer, L (1988) Biochemistry. Third Edition. W.H. Freeman and Company. New York.
Tajima F (1989) The effect of change in population size on DNA polymorphism. Genetics 123:597–601.
Tajima F (1993) Simple methods for testing the molecular evolutionary clock hypothesis. Genetics. 135:599–607.
Tennyson AJD, Miskelly CM, Totterman SL (2012) Observations of collared petrels (Pterodroma brevipes) on Vanua Lava, Vanuatu, and a review of the species' breeding distribution. Notornis. 59:39–48.
Thompson JD, Gibson TJ, Plewniak F, et al (1997) The CLUSTAL X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25:4876–4882.
Van Oppen MJH, Willis BL, van Vugt, HWJA, et al, (2000) Examination of species boundaries in the Acropora cervicornis group (Scleractina, Cnidaria) using nuclear DNA sequence analyses. Mol. Ecol. 9:1363–1373.
Wakeley J (2009) Coalescent Theory: An introduction. 575: 519.2..
Ward R (2009) DNA barcode divergence among species and genera of birds and fishes. Mol Ecol Resour. 9:1077–1085.
Welch AJ, Yoshida A, Fleischer RC (2011) Mitochondrial and nuclear DNA sequences reveal recent divergence in morphologically indistinguishable petrels. Mol Ecol. 20:1364–77.
Willerslev E, Cooper A (2005) Ancient DNA. Proc Biol Sci 3–16.
Zink RM, Barrowclough GF (2008) Mitochondrial DNA under siege in avian phylogeography. Mol Ecol. 17:2107–21.

CHAPTER IV

Complex phylogeography of a tropical seabird and species limits: coalescent inference using nuclear and mtDNA

by

Iglesias-V., A[A]., Gangloff, B[A,D]., Ribout, C[A]., Shirihai, H[B]., Friesen, V[C]., & Bretagnolle, V[A].

"The most erroneous stories are those we think we know best - and therefore never scrutinize or question."
~ Stephen Jay Gould

Complex phylogeography and species limits of a tropical seabird: coalescent inference using nuclear and mtDNA

Iglesias-V., A[A]., Gangloff, B. [A, D], Ribout, C[A]., Shirihai, H[B]., & Bretagnolle, V[A].

[A]Centre d'Etudes Biologiques de Chize, UMR 7372, CNRS & Université de La Rochelle, 79360 Villiers en Bois, France.
[B] Office Environnemet and Heritage, PO BOX 1967, Hurtsvilee, NSW 2220, Australia.
[D] Service de Systématique Moléculaire, UMS 2700 OMSI MNHN-CNRS, MNHN, Paris, France.

ABSTRACT

The correct delimitation of evolutionary population units, and the delineation of reproductively cohesive populations is vital in phylogeographic and conservation studies. This study consisted in several concatenated phylogenetic reconstructions, clustering analyses and historical demographic approach to review the systematics of a contentious species complex *Pterodroma leucoptera* distributed across the south and west of the Pacific Ocean. We used two mitochondrial genes and five nuclear loci and units of conservation delimitation criteria (e.g. ESUs, MUs) to evaluate *a priori* two: (1) *P. leucoptera* vs (2) *P. brevipes* and eight: (1) *P. leucoptera*; (2) *P. caledonica*; (3) *P. caledonica or leucoptera,* Australes; (4) *P. brevipes*, Vanua Lava I-II, (5) *P. brevipes* Tahiti; (6) *P. brevipes*, Cook Island; (7) *P. brevipes,* Fiji, (8) *P. brevipes*, Eromango-Tanna hypothetical subspecies derived from previous studies. We also attempted to recover historical demographic information such as changes in genetic diversity through time, migration, and time of divergence. Mitochondrial results were congruent with the a priori two population hypothesis while demographic analysis revealed a split of 100 000 years before present. Effective gene flow estimates ($2N_em$) validated the two mitochondrial linages. The nuclear genes recovered a polyphyly stage of differentiation among all populations. Presumably extinct breeding populations of Fiji, Viti Levu, and Rarotonga, did not show genetic diversity loss when comparing with extant samples. We conclude that the *P. leucoptera* complex holds two mitochondrial lineages (*P. brevipes* and *P. leucoptera)*. Because nuclear haplotypes frequencies showed certain levels of differentiation these lineages could be treated as ESUs for conservation management. Regarding the other populations, they should be treated as MUs for conservation policies .

Keywords: *Phylogeography, conservation, historical demography inference, Pterodroma, Petrel.*

INTRODUCTION

The correct delimitation of evolutionary population units and the delineation of reproductively cohesive populations represent key components in phylogeography and conservation studies (Militão et al, 2014; Mace 2004), as inaccurately assigned populations may yield biased demographic parameter estimates (Beerli 2004; Slatkin 2005). In addition, wrongly defined populations units of conservation may not receive adequate protection or conversely acquire unjustified management protection (Avise 2009). Therefore, the accurate assignment of evolutionary units is of relevance for understanding species diversity and designing effective strategies to conserve biodiversity (Sites and Marshall 2003). Over the last century, species delimitation was frequently examined by means of morphological cues, quite often failing to assess the complex limits in cryptic taxa and recently differentiated lineages (Avise 2000; Knowles and Carstens 2007; Wiens 2007). More recently, the development of molecular tools and statistical inference based on coalescent theory have allowed better quantification of the degree of isolation and/or gene flow among taxa (Yang and Rannala 2014). However, the supremacy of mitochondrial genes (mtDNA), markers of choice for over two decades, has been disputed over the last ten years by the often conflicting results emerging from studies using nuclear genes (e.g., Edwards et al. 2005; Zink and Barrowclough 2008; Lee and Edwards 2008; Edwards and Bensch 2009). In addition, it has been argued that phylogenetic inferences based solely on mtDNA should be regarded with caution (Chan and Levin 2005; Zink and Barrowclough 2008). For instance, mtDNA rapid evolutionary rate may lead to recurrent substitutions at single base position (homoplasy) obscuring the signal of deeper history. Another major pitfall can be ascribed to the fact that mitochondrial genome evolves as a single linkage unit (several mtDNA genes sequences yields to a same single gene tree), hence misrepresenting the organismal phylogenetic history, by reflecting only the matrilineal gene history in the case of male sex-biased dispersal (Zink and Barrowclough 2008). Nonetheless, mtDNA is acknowledged for its non-recombining properties, high mutation rate and lower effective population size (N_e) compared to nuclear DNA (nuDNA) (Avise et al. 1987; Zink and Barrowclough 2008; Avise 2009). Therefore lineage sorting will occur faster for mtDNA than for nuDNA (Bensch et al. 2006; Zink and Barrowclough 2008). Moreover, acknowledging that all loci, upon lineage splitting, undergo a transition from an initial state of polyphyly towards monophyly when time since speciation increases (Avise and Ball 1990; Kulikova et al. 2005), taxa might be delimited with a criteria such as reciprocal monophyly assuming no gene flow (Ryder 1986; Moritz 1994a; Knowles and Carstens 2007). However, there is an inherent contradiction between such an exclusive criterion to delimit species, and the actual continuous process of speciation (Crandall et al. 2000). For instance, well-differentiated sister species may have reciprocally monophyletic mtDNA but unsorted alleles at nuclear loci (De Salle and Giddings 1986; Slade et al. 1994; Welch et al. 2011; Gangloff et al. 2013). Indeed, recently diverged species or sister species will tend to remain undiscovered under the overly restrictive reciprocal monophyly criterion (Slade et al. 1994; Paetkau 1999; Degnan and Rosenberg 2009). Consensus has eventually emerged that complex evolutionary or biogeographical histories cannot be solved with mtDNA or one locus alone (Chan and Levin 2005) and that multilocus phylogenetic inference and statistical phylogeography are better recommended (Degnan and Rosenberg 2009; Heled and Drummond 2010a; Hey 2010a). In addition, it was pointed that one might

have to sequence two to three times as many nuclear bases as mitochondrial to uncover equivalent numbers of variable sites and produce a tree with equivalent resolution (Zink and Barrowclough 2008). Multilocus sequence data can provide support for different species delimitation using recently developed theoretical models that combine species phylogenies and gene genealogies via ancestral coalescent processes (Yang and Rannala 2010). Nonetheless, coalescence-based methods are scarcely applied in non-model species phylogenies, especially within the "grey zone" (sensu (De Queiroz 1998, 2007) when population and species are not easily differentiated and the possibility of incomplete lineage sorting is greatest (Degnan and Rosenberg 2009). Different strategies are being used (e.g. Zhou et al. 2007; Belfiore et al. 2008; Leaché et al. 2009; Yang and Rannala 2010; Salicini et al. 2011; Sousa and Hey 2013) showing the diversity of attempts to apply good theory to challenge complicated realities.

In addition, from a conservation viewpoint, the correct delimitation and assessment of evolutionary population units are determinant to the implementation of adequate population protection. In this view, the Evolutionary Significant Units (ESUs, (Ryder 1986) was developed to take into account the molecular inheritance and ensure that evolutionary heritage is conserved (Moritz 1994a; Paetkau 1999). ESUs may be recognized even if phylogeographic structuring of nuclear gene variation is absent (Moritz 1994). Populations that do not show reciprocal monophyly, but have diverged in allele frequency, may also be significant units for conservation, under the Management Units criteria (MUs), regardless of the phylogenetic distinctiveness of the alleles (Moritz 1994a). Therefore the distinction between ESUs and MUs is important since it helps to recognize two types of conservation units, both necessary for management: ESUs, ascribed to historical population structure, mtDNA phylogeny, and long-term conservation needs; and MUs, related to current population structure, allele frequency, and short-term management issues (Moritz 1994b; Palsbøll et al. 2007).

Petrels of the *Pterodroma leucoptera* complex (Genus *Pterodroma,* Hindwood and Serventy 1941) are geographically distributed throughout the southern Pacific Ocean (Figure 4.1). Their taxonomic treatment over the last fifty years has been continuously changing. Hindwood and Serventy (1941) proposed that the taxon contains two sub-species, *Pterodroma leucoptera leucoptera* from Australia and *P. l. brevipes* from Vanuatu and Fiji archipelagos. De Naurois and Rancurel (1978) found a new breeding population in New Caledonia, later described as a third subspecies, *P. l. caledonica* by Imber and Jenkins (1981). Sibley and Monroe (1990) split *P. l. brevipes* from *P. l. leucoptera,* giving it full species status. In 1990-91, Seitre and Seitre (1991) heard the call of an unknown gadfly petrel on Raivavae, Austral Islands, French Polynesia. These birds were identified by VB as potential *P. l. leucoptera* or *P. l. caledonica* based on photographs taken by A. Guillemont (in Brooke 2004). In addition, Bretagnolle and Shirihai (2010), identified three particular individuals (smaller in size) from Society archipelago which could be placed either in *P. leucoptera* or in *P. brevipes* group. V. Bretagnolle also reported a specimen from Rarotonga, Cook island (held at National Museum of New Zealand Te Papa), the identification of which remains uncertain (Bretagnolle and Shirihai 2010). Finally, in 2010 a new taxon from northern Vanuatu was described as *P. brevipes magnificens* by these authors based on colour differentiations, body size and breeding season (Bretagnolle and Shirihai 2010). The same authors also rediscovered a *P. brevipes* breeding population on Eromango island, Vanuatu, that was previously known only

from museum specimens held at the American Museum of Natural History, New York. They suggested that a pale and dark morphs of *P. brevipes magnificens* co-occurred in southern Vanuatu and therefore suggested that identification of birds from this area based on body size and breeding chronology should be regarded with caution and would require further investigation.

In this study, the chosen methodological framework consisted of performing different phylogenetic and clustering analyses based on mtDNA and nuclear sequences, which not only allow to identify and compare genetic clusters but also allowed assessing their robustness. Identified clusters were then used to assign *a priori* "species" in a multi-locus species tree inference to achieve a better resolution of their interrelationships and to assess the reliability of the node in a coalescence-based framework (Belfore et al 2008, Fujita et al 2010). We used the aforesaid approach to assess the species boundaries within the populations of the *P. leucoptera* complex by means of multiloci, clustering, and coalescence-based species tree inference. We tested the *a priori* two (*P. leucoptera* (Australia and Caledonia populations) and *P. brevipes* (South west pacific populations) (Hindwood and Serventy 1941) and eight (1. *P. leucoptera*; 2. *P. caledonica*; 3. *P. caledonica or leucoptera,* Australes; 4. *P. brevipes*, Vanua Lava I-II; 5. *P. brevipes*, Eromango; 6. *P. brevipes* Tahiti; 7. *P. brevipes*, Cook Island; 8. *P. brevipes,* Fiji,) hypothetical subspecies derived from previous studies. We also attempted to disentangle their historical demographic history (divergence time, migration rate per generation and long-term effective population size) to understand how evolutionary parameters have shaped the process of speciation of *P. leucoptera* complex. Coalescent-based demographic inferences were also used for population assignment into ESUs (linked to historical population structure, mtDNA phylogeny and long-term conservation needs) or MUs (related to current population structure, allele frequency, and short-term management issues)

METHODS

Specimens and sites

We sampled the two populations of *P. leucoptera* (Australia and New Caledonia) and virtually all extant populations of *P. brevipes*: Eromango, Tanna, Vanua Lava and Bank Island in Vanuatu, Society Islands and Raivavae in French Polynesia, Cook Island and Fiji (Table 4.1., Figure 4.1). All populations were sampled during either from field expeditions (blood samples) or from museum collections (Table 4.1; Appendix Table A.3). A total of 121 individuals from 11 islands were used in this study (Figure 4.1; Appendix Table A1, A.3). Fresh blood samples from 79 individuals were collected from the tarsal vein and kept in 90% ethanol at ambient temperature, then stored at -20°C. Similarly, 29 feathers and 13 toe pad skin samples birds were collected and stored as mentioned above.

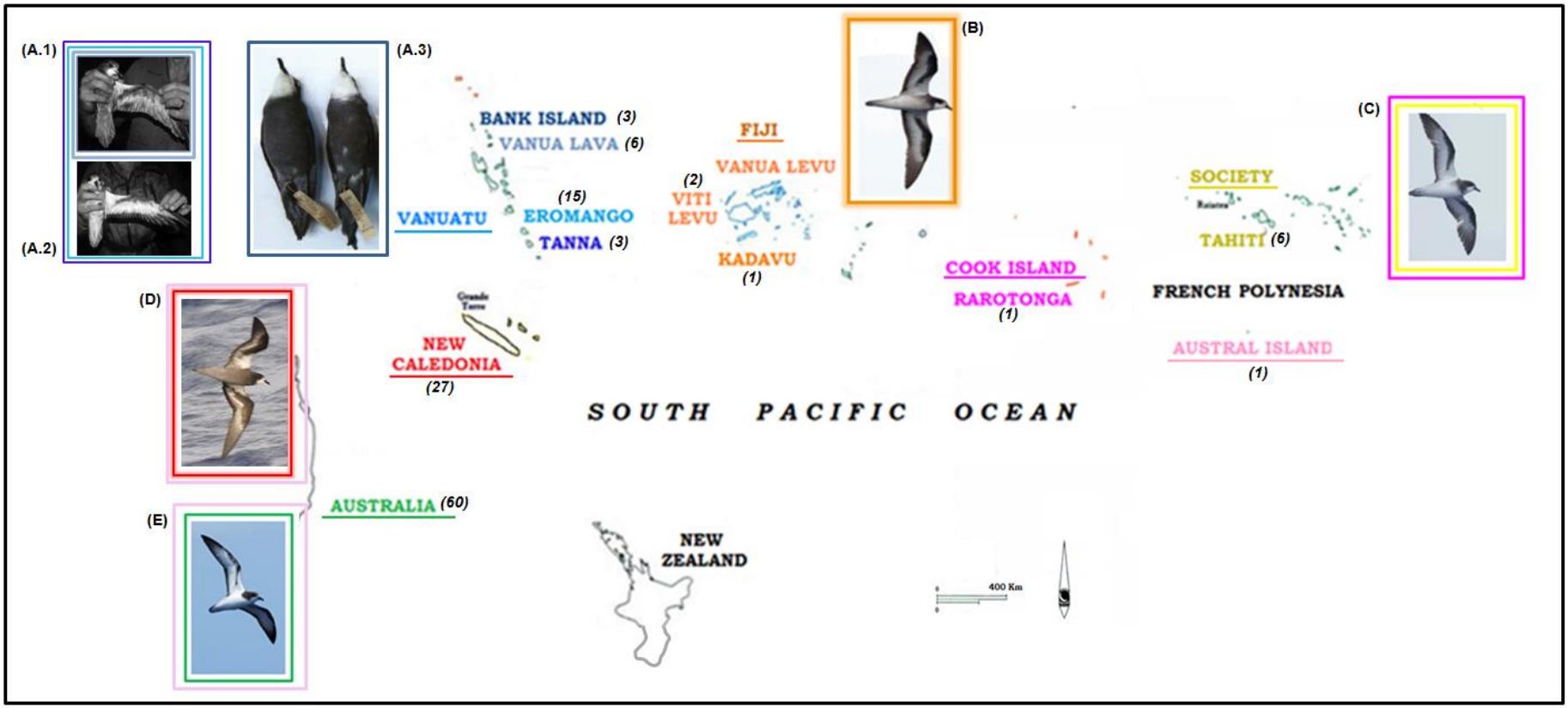


Figure 4.1. Morpho sub-species distribution and key morphological distinctions of *P. leucoptera* complex. **Vanuatu Island** - *P. brevipes* **(A.1), (A.2), (A.3)** (presence of both, dark grey morph and paler morph related to bold coloration in the upper and lower wing (Tennyson et al. 2012, Bretagnolle *pers. comment*), collier complete: **(A.1)**: Dark grey morph, *P. brevipes* (**Tanna, Eromango, Vanua Lava**); **(A.2)**: Paler morph, *P. brevipes* (**Tanna, Eromango**); **(A.3):** Extreme dark grey morph, *P. brevipes magnificens* (**Bank Island: "Vanua Lava II"**) (Bretagnolle and Shirihai, 2010). **(B): Fiji** - *P. brevipes* (paler morph) from Viti Levu, Vanua Levu, and Kadavu. **(C):** *P. brevipes* from **Society - Tahiti** (paler morph and collier not complete (Bretagnolle *pers. comments*), **Cook Isl.- Rarotonga** (dark grey morph; collier not complete (Bretagnolle *pers. comments*.) **(D)** *P. l. caledonica* **(E)** *P. l. leucoptera:* dark white lower wing coloration; **(D)-(E):** *P. l. leucoptera* or *P.l. caledonica* from **Australes** (Bretagnolle *pers. comments*.). **Numbers in brackets ()** show the maximum number of individuals sampled.

Table 4.1. Maximun Number of individuals per population and loci used (See Appendix Table A3 for ID samples). Number in brackets **()** indicates the loci sequence length used.

***mtDNA* CO1+Cytb and Nuclear Introns loci**						
Population	**CO1+Cytb (1310)**	**PAXIPI (424)**	**CSDE (481)**	**βfibin7 (911)**	**IRF2 (635)**	**TPM (463)**
Eromango - Tanna	15	16	15	14	13	10
Vanua Lava I	5	6	6	6	5	5
Vanua Lava II (Bank Isl.)	2	2	-	1	3	-
Society	5	5	5	5	3	2
Cook Island	1	-	-	-	-	-
Fiji_Viti Levu	1	2	-	1	1	-
Fiji_Viti Kadavu	-	1	-	-	1	-
Australes	1	1	-	1	1	-
New Caledonia	17	26	16	15	26	15
Australia	20	56	51	35	45	17
TOTAL (N)	**67**	**115**	**93**	**78**	**98**	**49**

DNA Sequencing

Whole genomic DNA from blood and feathers was extracted using a DNeasy blood and tissue preparation kit (Qiagen Ltd). To minimize the risk of contamination from modern DNA (Sampled at the CEBC-Molecular Biology Laboratory), ancient DNA extractions were performed at Poitiers University, France, that undergoes regular decontamination with UV light-irradiation and hypochloride treatment. Blank DNA extractions and polymerase chain reactions (PCRs) were routinely screened for contamination. Samples were ground to a fine powder using Fastprep machine during 3 repetitions of 30-45 seconds then lysed in 30 *µl* of proteinase K and lysed overnight at 56°C in a thermomixer (Eppendorf) at 9 000rpm. DNA from toe pad skin was extracted using the Phenol-chloroform protocol (Sambrook et al. 1989). GenomiPhi kit (GE Healthcare Biosciences) was used to increase yield and concentration of whole DNA samples. Recovered DNA was purified following manufacturer protocol ExoProStar^{TM} kit (GE Healthcare). Isolated DNA was used to amplify and sequence two mitochondrial DNA genes and five nuclear introns.
We sequenced two mitochondrial loci, Cytochrome Oxidase sub-unit 1 (CO1) and Cytochrome b (Cytb) genes. PCR amplification followed standard procedures and was performed using primers listed in Table Appendix A.1. PCR products were sequenced in both directions. We also sequenced five intron loci: Interacting Protein 1 (*PAXIP1*), Cold shock domain-containing protein E1 (*CSDE1*), Beta-Fibrinogen intron 7 (*BFibint7*), Interferon regulatory factor 2 (IRF2F) and Tropomyosin (TPM). PCR methods were performed as described in Table Appendix A.10, following previous studies applied on seabirds (Primmer et al. 2002; Kimball et al. 2009; Gangloff et al. 2011). PCR products were purified and sequenced using the same PCR primers by Eurofins Scientific (France). Nuclear haplotype sequences were obtained for each PCR product using the software program PHASE v2.1 (Stephens et al. 2001) implemented in DnaSp v.5.10.01 (Librado and Rozas 2009). For each individual, phased haplotypic data (probability cut-off 0.50) were then used in all analyses, individuals thus being represented by two sequences.

Mitochondrial origin of concatenated mitochondrial gene sequences was confirmed by translating these to check for the presence of stop codons to discard indices of their potential nuclear origin in BioEdit Sequence alignment Editor v. 7.2.5 (Hall 1999). Recombination in nuclear loci was tested with the four-gamete test (Hudson and Kaplan 1985) implemented in DnaSp v5. When the test suggested an intralocus recombination we discarded the sites within the recombination block and retained the longest possible continuous unrecombined sequence for subsequent analyses. Mitochondrial DNA sequences were translated at the amino acid level to then be visually evaluated and aligned with CLUSTAL X (Thompson et al. 1997), implemented in BioEdit Sequence alignment Editor v. 7.2.5. (Hall 1999). Furthermore, because the Isolation with Migration analysis required having known phase for nuclear sequences, we determined the gametic phase using the program PHASE (Stephens et al. 2001), implemented in DnaSp v5. Phased haplotypic data of nuclear introns were then used in all analyses individuals thus being represented by two sequences.

Data analyses

Diversity Indices and Population genetic structure

We used DnaSp to estimate variable sites, parsimony informative sites, number of haplotypes, haplotype diversity and nucleotide diversity for each marker. Population genetic differentiation was tested using two approaches implemented in Arlequin v. 3.5.1 (Excoffier and Lischer 2010). First pairwise $\boldsymbol{F_{ST}}$ interpopulation values were inferred at 0.05 significance level determined by 10,000 permutations. Then, the analysis of hierarchical molecular variance (AMOVA) was completed with distances set up with the most appropriate model of sequence evolution following the Akaike information criterion (AIC) carried through the package Analysis of Phylogenetic and Evolution (APE; Paradis et al. (2004) implemented in R v. 2.15 (R Development Core Team 2011). AMOVA intrinsic calculates pairwise differentiation between *a priori* populations (subspecies using) ΦST (with substitution model set up as mentioned before) a direct analogue of Wright's $\boldsymbol{F_{st}}$ for nucleotide sequence divergence.

Phylogenetic gene trees from the combined sequences of CO1 and Cytb were constructed using Bayesian analysis with BEAST v1.6.1(Drummond and Rambaut 2007) and Maximum Likelihood (ML) inference through MEGA v5 (Tamura et al. 2011). Our Beast analysis consisted of two runs of 1 x 10^7 generations sampled every 1000 steps; the first 10% steps were discarded as burn-in. We checked the convergence of the runs and that the effective sample sizes (ESS) were all above 200 by exploring the likelihood plots using TRACER v 1.6 (Rambaut et al. 2014). Nuclear data was also used to build single nuclear gene trees in order to highlight possible differences in the nuclear clustering signals (different levels of polyphyly at the nuclear genes). Single nuclear gene trees were reconstructed using MEGA v5 and clade support was provided by 1000 bootstrap replicates. Besides, the phylogeny of concatenated nuclear loci was inferred in *Beast and analysis were performed using three runs of 2 x 10^7 generations and sampled every 1000 generations; the first 25% were discarded as burn-in. We checked the convergence of the runs and that the effective sample sizes (ESS) were all above 200 by exploring the likelihood plots using TRACER v 1.5.6. Finally, the results obtained from phylogenetic gene trees and clustering analyses were used to define and to assign the individuals to be used as 'species' in a multilocus coalescence-based species tree phylogenetic analysis. Species tree analysis, needs *a priori* information regarding the species delimitation and the species assignment of the individuals in order to reconstruct the topology of the species tree (Heled and Drummond 2010). Nuclear alleles and the mitochondrial sequences in the hierarchical Bayesian analysis *BEAST (Heled and Drummond 2010) implemented in the software BEAST v.1.5.4 (Drummond and Rambaut 2007). The analysis was launched using an uncorrelated lognormal relaxed clock and a yule process species under a coalescent model assuming a constant population over the time period was chosen. Because *BEAST enables a calibration date on the root node and/or a prior on the substitution rate (Drummond and Rambaut 2007) for divergence dating, per lineage mutation rate (*mu*) for both mitochondrial and nuclear genes were calculated by using *mu*= dxy/2T (Nei 1987). where dxy stands for interlineage rate (dxy: 0.0189 s/s/My for concatenated mtDNA (Weir and Schluter 2008) and dxy: 0.0036 s/s/My for nuclear introns (Axelsson et al 2004); and (T) divergence time since two unique lineages diverged (See Appendix Table A.5). All analyses were run three times to check for convergence with 1 x 10^7 generations and sampling every 2x10^3 generations. The

first 30% of the genealogies were discarded as burn-in. Convergence, stationarity, effective sample size for each parameter of interest and the appropriate burn-in were evaluated using the software TRACER v.1.6 (Rambaut et al. 2014).

Two types of data and analytical methods were used to genetically corroborate or reject the 'two-population and eight population hypothesis'. The first method involved using our mtDNA tree to test our population hypotheses by looking for exclusive clusters of haplotypes present in the phylogenetic trees or the haplotype networks that correspond with either or both the two and eight population hypotheses. Note that the criterion of 'exclusivity' of haplotypes for a predefined population can be satisfied either by bifurcation groups in an unrooted tree or clades in a rooted tree (Brower 1999). Haplotype frequencies were inferred with DnaSP. Genealogical relationships among haplotypes were reconstructed using a median-joining network (Bandelt et al. 1999) in NETWORK v.4.6.0.0 (http://www.fluxus-engineering.com). The second method used, program BAPS ("Bayesian Analysis of Population Structure"; (Corander 2003), has advantages over other structuring programs because it can treat individual base positions within a locus as linked loci, while treating different loci as independent. No prior information on geographic location was specified. Replicated runs were performed on different values of k (clusters) from 2 to 10.

Estimation of demographic Parameters

Fu's F_s (Fu 1997) and Tajima's D (Tajima 1989) both implemented in ARLEQUIN v.3. 5 (Excoffier and Lischer 2010) were used to infer whether selection affected the examined loci and to test for signatures of population expansion. Positives F_s values indicate a deficiency of alleles characteristic of a recent population bottleneck or an overdominant selection. Fu's simulations suggest that F_s is more powerful to estimate population expansion than Tajima's D. Following Fu (1997) F_s values are regarded as significant if $p<0.02$ which corresponds to the conventional significant $p<0.05$ for Tajima's D. Non-significant results are consistent with a locus being neutral and having stable population size, whereas a positive D reflects a locus that experienced balancing selection or a population bottleneck (Tajima 1989).

To test which demographic model best explain the demographic history of *a priori* two population hypothesis of *P. leucoptera* complex Constant Size population model and the Extended Bayesian Skyline Plot (EBSP) were performed. Both demographic tests were launched separately using Beast v1.8., we analyzed all the available sequences for both concatenated mitochondrial and nuclear sequences (Appendix Table A.6; A.7). We ran a Markov Chain for 1×10^8 steps discarding the first 2×10^3 as a burn-in. Same per lineage mutation rates used in species tree analysis (see species tree section) were converted into years to use it in the constant size and EBSP analysis (4.87×10^{-9} s/s/y for concatenated mitochondrial DNA and 6.48×10^{-9} s/s/y for nuclear introns analysis. We inspected parameters traces with Tracer V. 1.6 to assess stationarity and to check for high effective sample sizes (ESSs). Simulations with the highest ESS for both constant size and EBSP analyses were retained and used to compare the ratio of the Marginal Likelihood between them by means of the Bayes Factor (BF) implemented in Tracer V. 1.6. The Marginal likelihood ln(BF) range > 5 was retained and used to explain the demographic history of our data.

We used our five nuclear introns and the Bayesian Markov Chain Monte Carlo (MCMC) approach implemented in the program IMa2 (Hey 2010) to test the two and eight population hypothesis and whether or not they fit with the ESU criteria . Furthermore, our IMA2 analyses was used to jointly examine migration and isolation for each divergence event. Six population parameters (scaled by mutation rate u) were estimated: sizes of the ancestral (*Θ*A) and daughter populations (*Θ*1, *Θ*2); population divergence time (t); migration rate/gene/generation (m1), which is defined as the rate genes come into population 1 from population 2 as time moves forward. The program IMa2 requires that the user specifies the population (or species) tree before initiating the analysis. Of the two possible population trees, we chose the tree topology having the relationships Australia, New Caledonia, *P. leucoptera* vs South Pacific islands, *P. brevipes* (which is supported by Sibley and Monroe (1990) and by our mtDNA gene tree (see Figure 4.2). The Hasegawa–Kishino Yano mutation model (HKY) (Hasegawa et al. 1985) was used for the mtDNA and nuclear introns.

Specified priors for divergence and migration rates were indicated following the recommendations of the IMa2 documentation. A geometric heating model with 50 independent heated chains was found to provide good mixing and high repeatability between the first and second halves of the run and between analyses that began from different starting positions. After exploratory runs, we ran the MCMC chain for 20 x 10^7 steps discarding the first 2 x 10^5 steps as burn-in. Conversion of the demographic parameters was done using the per locus mutation rates (substitution/year) obtained by multiplying per lineage mutation rate calcul (similar as used in EBSP) by the number of base pairs of concatenated mtDNA sequence. Finally, a generation time of 20 years was used to convert raw parameter estimates into demographic units .To better understand population assignment results, we used IMa2 effective population size and gene flow estimates to calculate $2N_em$ values between each hypothetical population. The $2N_em$ value is defined as the population migration rate which is the effective rate at which genes come into a population, per generation. Given that the evolutionary consequences of effective gene flow varies between two extreme conditions, genetically isolated populations and a single panmictic one, a criterion for defining separate populations can be based on $2N_em$ values that reflect a departure from panmixia. We, therefore, chose 2Nem ≤25 as our criterion because gene flow, even this large, can be associated with departures from panmixia (Waples and Gaggiotti 2006). For descriptive purposes, we classify $2N_em$ values in terms of their relative strength of genetic isolation: strong when $2N_em$ ≤1 moderate when 1 < $2N_em$ ≤ 5 and weak when 5 < $2N_em$ ≤ 25 (Waples and Gaggiotti 2006).

RESULTS

Molecular variation

The concatenated mitochondrial dataset consisted of 67 sequences with a total length of 1310 base pairs (bp) (714 bp for CO1 and 596 bp for Cytb). Nuclear introns had sequences ranging from 424 bp (*PAXIPI)* to 911 bp (*β*fibin7). Introns, with the highest level of haplotype variability (*PAXIPI*, *Bfibin7, and IRF2f1)*, were concatenated and resulted of 1,872 bp. We discarded other nuclear loci because they didn't show enough variability during this analyse. Overall, we found that mitochondrial loci exhibited higher levels of nucleotide variability (28 variable sites for assembled CO1 and Cytb with 22 different haplotypes) while introns nuclear variability ranged from 0 (*TPM*) to 17 for (*PAXIPI)* and *18 (*βfibint7). Details on

fragment lengths, variability and number of haplotypes for each intron are shown in Table 4.1.

Species trees support the existence of two well-defined lineages

Concatenated mtDNA produced supported topologies (Bayesian posterior probability ≥0.7 and ML Bootstrap ≥60% from the basal split and terminal branches with ML inference; see Figure 4. 2 AB), showing two lineages: one clade grouping all the *P. brevipes* populations and a second clade corresponding to *P. leucoptera*. In addition, mtDNA indicated that *P. leucoptera* and *P. brevipes* diverge by 22 mutational steps with strictly no shared haplotypes, though within *P. brevipes* only 5 haplotypes separated by only one mutation were found. Similarly, nuclear introns portrayed similar clades (Figure 4.3). However, neither the phylogeny based on single introns nor the concatenated nuclear introns (PAXIPI, Bfbint7, and IRF2) supported the monophyly found with mtDNA (either based on Bayesian or ML inference), a difference expected with recently diverged species. We used the two lineages as assigned by the mitochondrial phylogenies in *BEAST analysis. The resulting tree depicted the relationships between the clades that were found by the mitochondrial loci. Branches were well supported (Figure 4.4). BAPS analysis with the complete set of loci using a random number of populations with no prior population assignment also yielded a highest posterior probability for the number of populations *k*=2 (Table 4.2, Appendix, Figure A1). Multilocus BAPS (from mtDNA and nuclear introns) analysis also recovered the same clustering generated by the mitochondrial DNA trees.

Population sub-structuring within species

Population differentiation was tested using both F_{st} and AMOVA separately for each of the two presumed species. With concatenated mtDNA, a significant genetic differentiation was found between the Australia-Caledonia *vs* the South West Pacific island populations ($\boldsymbol{F_{st}}$=0.95 corresponding to 0.018 average sequence divergence (Table 4.4). Statistical power precluded any analysis with the eight *a priori* populations, but $\boldsymbol{F_{st}}$ resulted in highly significant differences, ranging between 0.95 (Society vs New Caledonia) and 0.98 Vanua Lava-Mera Lava vs Australia (Table 4.5. Analogous ***Fst*** obtained by AMOVA standing for signals of interpopulation variation were all highly significant for mitochondrial loci (0.92 $p<0.01$).

Table 4.2. Summary statistics fortwo mitochondrial (CO1 and Cytb) and five nuclear (PAXIPI, CSDE, *βfibint7, IRF2F1, TPM1)* genes , without accounting *a-priori* population definition in *P. leucoptera – P. brevipes* complex. **(N):** Number of sequences; (**SL):** Sequence Length**; (Np): Number of polymorphic sites; (Nh):** Number of haplotypes; (**Hd):** Haplotype diversity; **(π, in %):** nucleotide diversity. Significant values for tests of neutrality: p< 0.05 for Tajima's D and p<0.02 for Fu's Fs. **ET: Eromango-Tanna; Vanua Lava I and II: VL-I , II**

						Diversity Indices			Neutrality Tests		
Gene	**N**	**SL**	**Np (not including GAPS)**	**Parsimony informative sites**	**Nh**	**Hd [SD $^{(x10-2)}$]**	**π% [SD]**	**Tajima's D**	**pvalue**	**Fu's F_s**	**pvalue**
CO1+Cytb	67		43	28	22	0.816 [0.040]	0.001 0.005]	**-3.5019**	0.03	**-14.22**	0.002
PAXIPI	238	424	17	17	6	0.1376 [0.032]	0..002 [0.0012]	**-1.95491**	0.002	-3.37	0.09
CSDE	186	481	15	4	13	0.387 [0.039]	0.00341[0.032]	**-2.06091**	0.0004	**-12.043**	0.00
βfibint7	156	911	21	11	18	0.669 [0.025]	0.001[0.001]	**-2.07566**	0.002	**-14.735**	0.00
IRF2F1	196	635	9	7	10	0.393 [0.0401]	0.00075 [0.05]	**-1.56898**	0.019	**-7.375**	0.002
TPM1	48	463	8	0	4	0.120 [0.044]	0.041 [0.045]	**-1.92641**	0.002	-1.434	0.210

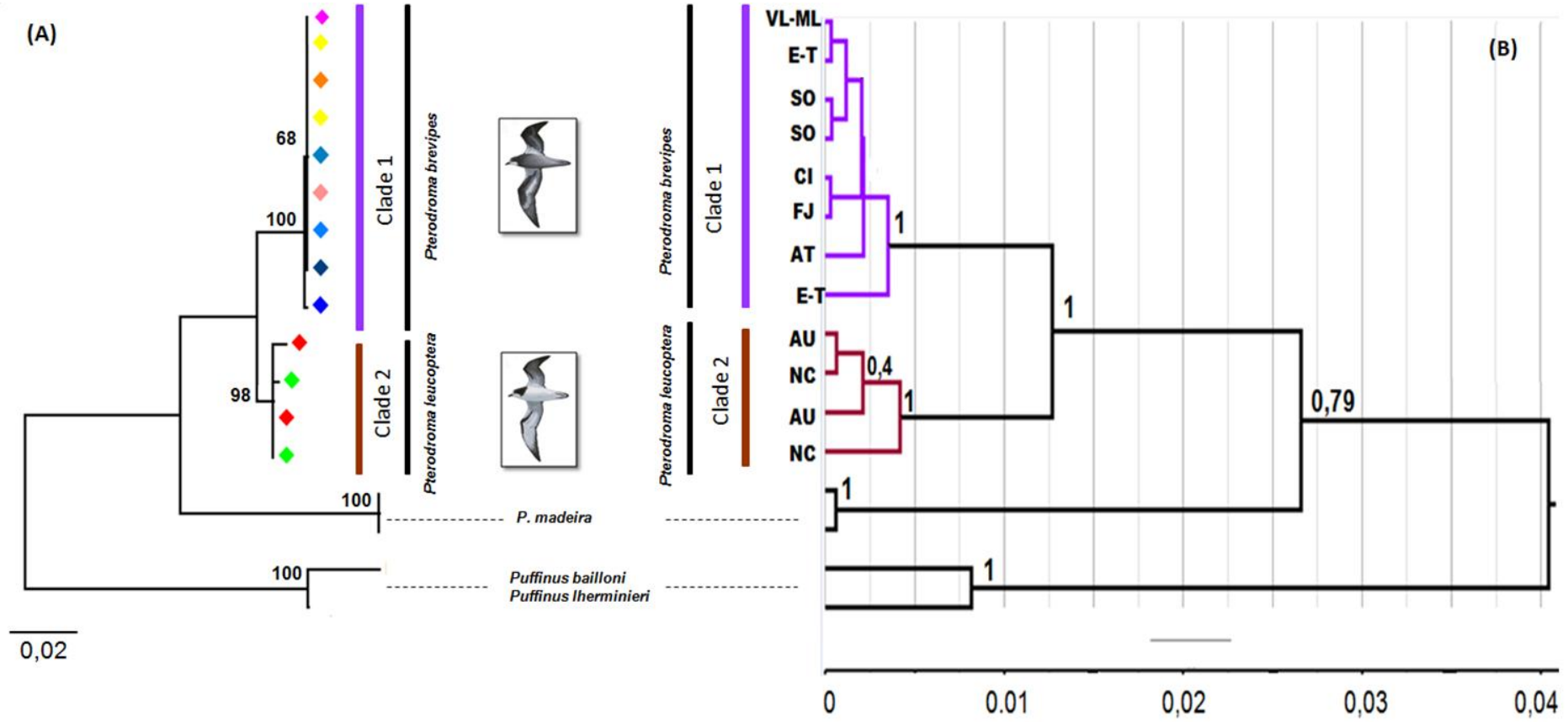


Figure 4.2. Mitochondrial DNA phylogenies for concatenated CO1 and Cytb loci. **(A)** ML tree analysis obtained with MEGA. Numbers at nodes represent bootstrap support. **(B)** Bayesian analysis obtained with BEAST. Numbers above nodes represent Bayesian posterior probabilities. Letters and colors represent a priori populations: (**NC):** New Caledonia; **(AU)**: Australia; **(E-T)**: Eromango-Tanna; **(VL-ML)**: Vanua Lava I. II; **(CI):** Cook Island; **(SO):** Society; (**FJ):** Fiji**; (AT): Australes** (see also Figure 4.1).

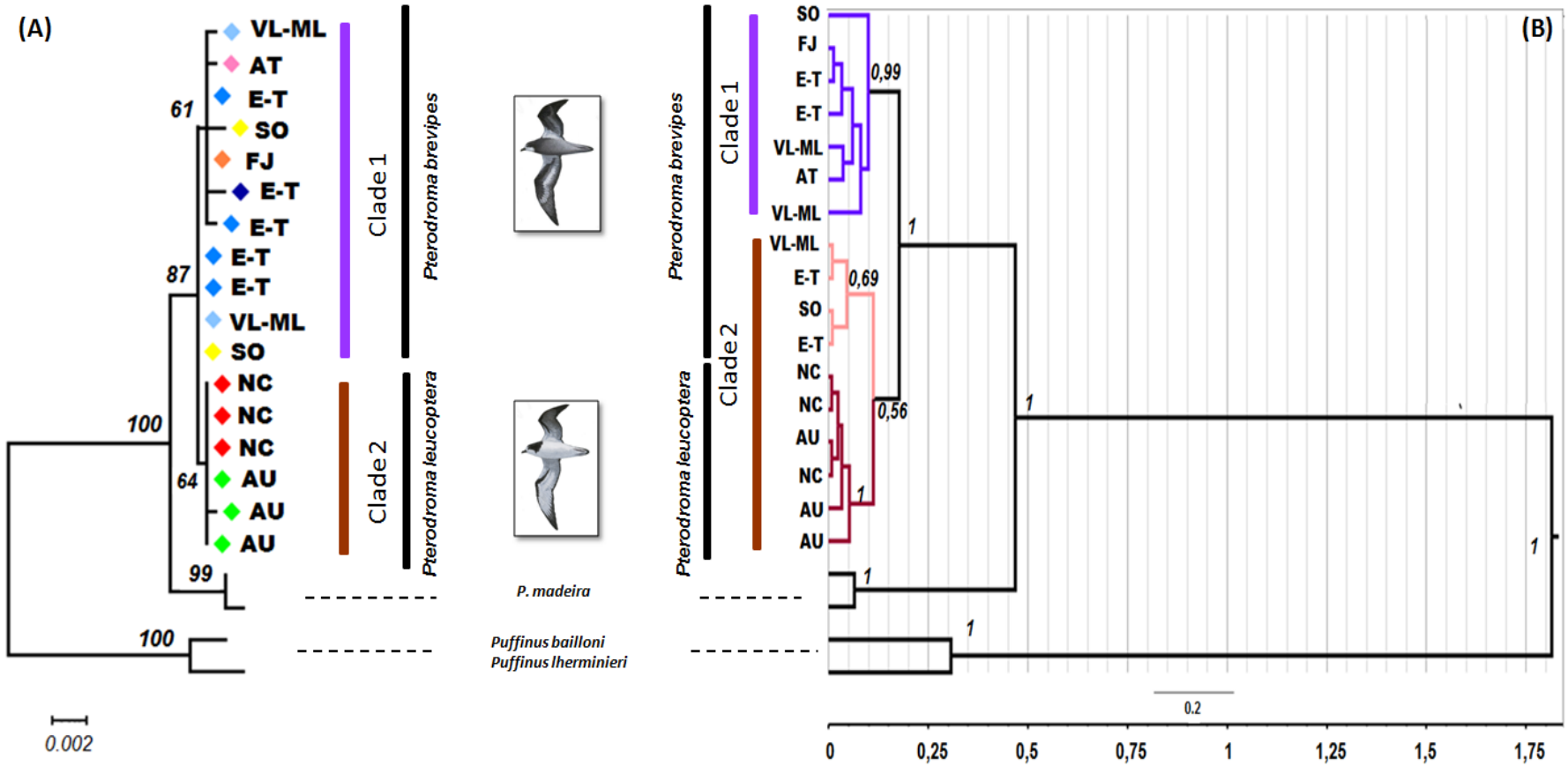
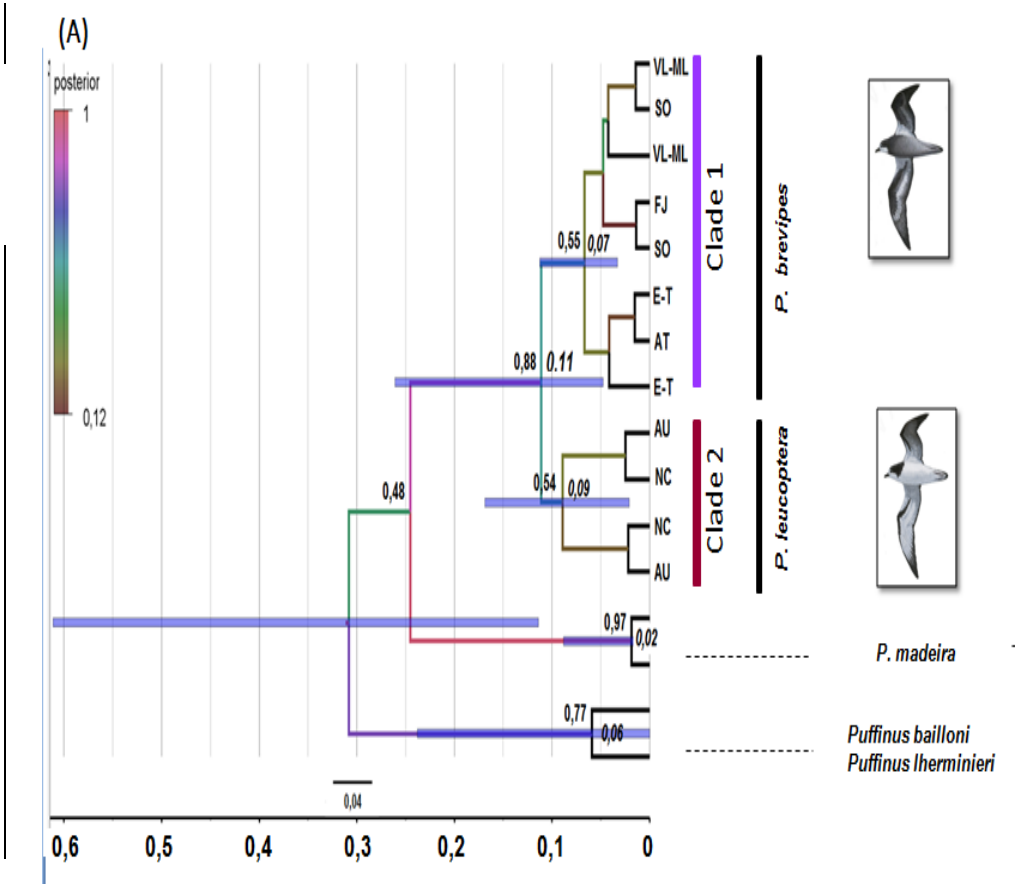


Figure 4.3. Nuclear DNA phylogenies for concatenated PAXIPI, IRF2 βFbin7 loci in *P. leucoptera – P. brevipes* complex**. (A)** ML tree analysis obtained with MEGA. Numbers at nodes represent bootstrap support. **(B)** Bayesian analysis obtained with BEAST. Numbers above nodes represent Bayesian posterior probabilities. Letters and colors represent a priori populations: (**NC):** New Caledonia; **(AU)**: Australia; **(E-T)**: Eromango-Tanna; **(VL-ML)**: Vanua Lava I, II; **(SO):** Society; (**FJ):** Fiji**; (AT): Australes** (see also Figure 4.1).Black bars indicate mtDNA clades.



4.4. Species tree obtained with two mitochondrial genes (concatenated CO1 and Cytb) and five nuclear intron (PAXIP1, CSDE1, βfibint7, IRF2F1, TPM1) in *Beast showing simultaneously the phylogeny and divergence time of *P. leucoptera* complex **(A)** under a strict clock assumption. Divergence times corresponding to the mean posterior estimate in Million years are displayed at each node. Purple bars represent the 95%HPD interval for the divergence time estimates. Numbers on branches represent the posterior probability of each clade Letters represent a priori populations: (**NC):** New Caledonia; **(AU)**: Australia; **(E-T)**: Eromango -Tanna; **(VL-ML)**: Vanua Lava I, II; **(SO):** Society; (**FJ):** Fiji**; (AT): Australes** (see also Figure.3.1). Horizontal scale bar in Million years.

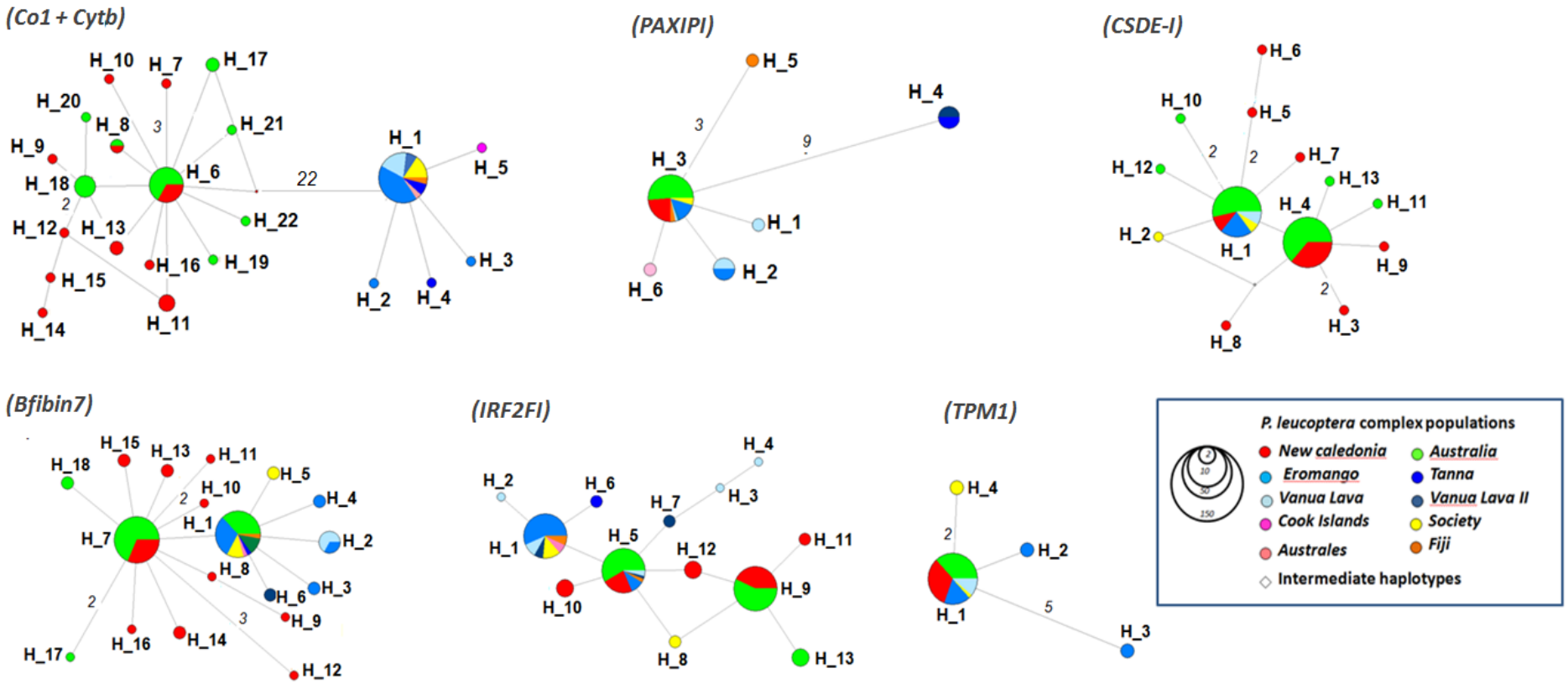


Figure 4.5. Haplotype networks with two concatenated mtDNA CO1 and Cytb and five nuclear introns, *PAXIP1*, *βFibint7, CSDE1, IRF21, TPM1* loci within *P. leucoptera.* Node sizes are proportional to number of individuals found with that haplotype, while color codes refer to sampling site (see Figure 4.1 for sample site locations). Polygon correspond to intermediate (non-sampled) haplotypes. Branch lenghts are not to scale to improve visualization. Number on branches show the number of mutation between alleles, if no numbers indicated, only one mutation step occurred.

Table 4.3. Individual loci for mitochondrial and nuclear markers and population assignment by BAPS analysis**.** Letters stands for our hypothetical a priori populations: (**NC):** New Caledonia; **(AU)**: Australia; **(E-T)**: Eromango-Tanna; **(VL-ML)**: Vanua Lava I. II; **(CI):** Cook Island; **(SO):** Society; (**FJ):** Fiji**; (AT): Australes** (see also Fig. 4.1).

	Haplotypes						**Population Clustering**	
	CO+Cytb	***PAXIPI***	***CSDE-I***	***βfibint7***	***IRF2***	***TPM***	***mtDNA***	***nuDNA***
NC	h6	h3 :h3	h1 :h1	h7:h7	h5:h5	h1 :h1		
	h7			h7:h10	h9:h9			
	h8			h7:h11	h10:h10			
	h9			h7:h12	h11:h11			
	h10			h7:h16	h12:h12			
	h11			h8 :h9				
	h12			h13 :h13				
	h13			h14 :h14				
	h14			h15 :h15				
	h15						I	
	h16							
AU	h6	h3 :h3	h:h1	h1:h1	h5 :h5	h1 :h1		
	h8		h1:h4	h1:h7	h9 :h9			I
	h17		h1:h10	h1:h17	h13 :h13			
	h18		h1:11	h7:h7				
	h19		h1:h12	h18:h18				
	h20		h1:h13					
	h21		h4 :h4					
	h22							
E-T	h1	h2 :h2	h1 :h1	h1 :h1	h1 :h1	h1 :h1		
	h2	h3 :h3		h3 :h3	h1 :h5	h2 :h2		
	h3	h4 :h4		h4 :h4	h5 :h5	h3 :h3		
	h4			h5 :h5	h6 :h6		II	
VL-ML	h1	h1 :h1	h1 :h1		h1:h1			

		h2 :h2			h1:h2	
		h3 :h3			h3:h4	
		h4 :h4			h5:h5	
					h7:h7	
CI	h5				H9 :h9	
SO	h1	h3 :h3	h1 :h1	h1 :h1	h1 :h1	h1 :h1
			h1 :h2	h1 :h5	h8 :h8	h4 :h4
FJ	h1	h3 :h3		h1 :h1	h1 :h1	
		h5 :h5			h5 :h5	
AT	h1	h6 :h6			h1 :h1	

Table 4.4. F_{st} of mDNA between the hypothetical *a priori* two populations in *P. leucoptera – P. brevipes* complex. (D): Number of net nucleotide substitution per site between populations.

		CO1+Cytb	
A priori Two populations	F_{st}	No of fixed differences	Average sequence divergence (Da)
Australia_Caledonia *vs* South Pacific Island	0.95471	23	0.018

Table 4.5. Fst of concatenated *mt*DNA under the hypothetical *a priori* of eight populations. Numbers in brackets **()**: represent the number of individuals used per population.

		Population							
	***A priori* eight populations (Fst)**	Australia	New Caledonia	Eromango-Tanna	VanuaLava-Mera Lava	Society	Cook Island	Raivavae	Fiji
CO1+Cytb	Australia (20)	-							
	New Caledonia (17)	0.097	-						
	Eromango-Tanna (15)	0.97	0.935	-					
	VanuaLava-Mera Lava (7)	0.98	0.945	0.0	-				
	Society (5)	0.98	0.945	0.0	0.0	-			
	Cook Island (1)	ne	ne	ne	ne	ne	-		
	Raivavae (1)	ne	ne	ne	ne	ne		-	
	Fiji (1)	ne	ne	ne	ne	ne			-

		Population							
	***A priori* eight populations (Fst)**	Australia	New Caledonia	Eromango-Tanna	VanuaLava-Mera Lava	Society	Cook Island	Raivavae	Fiji
CO1+Cytb	Australia (20)	-							
	New Caledonia (17)	0.097	-						
	Eromango-Tanna (15)	0.97	0.935	-					
	VanuaLava-Mera Lava (7)	0.98	0.945	0.0	-				
	Society (5)	0.98	0.945	0.0	0.0	-			
	Cook Island (1)	ne	ne	ne	ne	ne	-		
	Raivavae (1)	ne	ne	ne	ne	ne		-	
	Fiji (1)	ne	ne	ne	ne	ne			-

Past population demography

Tajima's D tests indicated that both mtDNA and all but two introns (PAXIPI and TPM1) were significantly different from zero, indicating an excess of rare alleles characteristic of a recent population expansion (Table 4.1). We then used each mitochondrial lineage in order to test for two demographic models, a constant model of population size and the Extended Bayesian Skyline Plot model (EBSP) using *BEAST. EBSP model was retained as the model that best explained the demographic trajectory of the clades (Bayes Factor comparisons were lon (BF) >5 in both cases). Results suggested an exponential growth of the *P. brevipes* clade c.60 ka ybp, yielding a current effective population size of c.33,000 individuals (%HPD 6 003-61492). Conversely *P. l. caledonica* population (Results took from Iglesias-V et al (2016), Thesis-Chapter V) stayed stable until c.70 000 ypb then experienced a major change to reach a current effective population of c. 52,000 individuals (%HPD 20 358 - 97 364), while *P. l.* leucoptera showed steady increase around 10 000 ybp with a current effective population size estimated at c. 12,000 (%HP 3,102- 31,440).

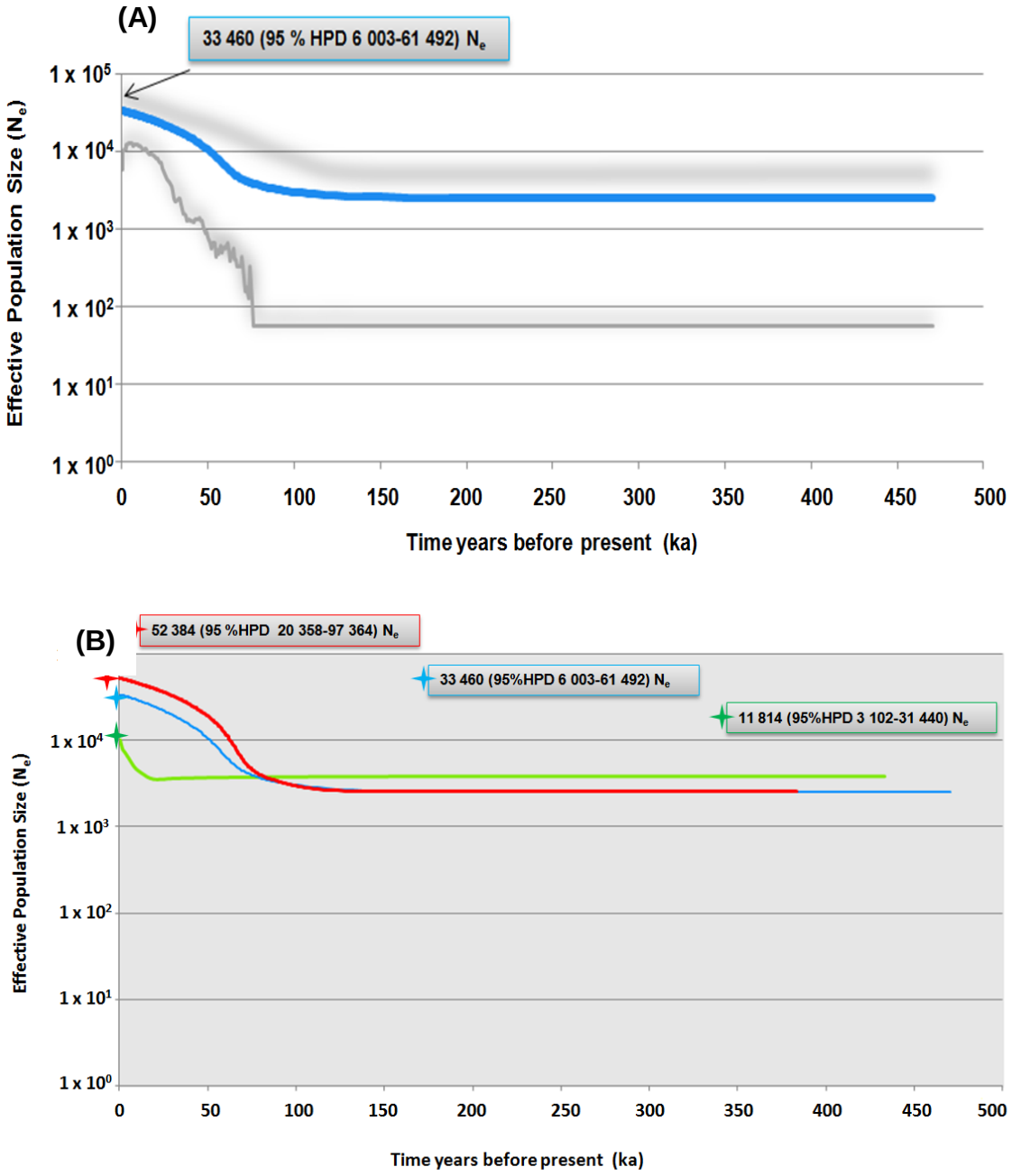


Figure 4.6. Extended Bayesian Skyline Plots of **(A)** *P. brevipes* and **(B)** *P. l. caledonica (red line)*, *P. l. leucoptera (green line)* sub species based on the CO1, Cytb, PAXP1, CSDE1, βFibint7 IRF21 and *TPM1* data sets. The blue, green, and red lines represents the median of the parameter Ne, proportional to the effective population size in thousands (Blue: *P. brevipes;* Red: *P. l caledonica;* Green: *P. l. leucoptera*), while the grey lines in **(A)** are the 95% confidence interval.

DISCUSSION

Species boundaries within the *Pterodroma leucoptera* complex were previously studied using behavioral, morphometric traits and single loci genetic markers(Imber and Jenkins 1981; Sibley and Monroe 1990; Bretagnolle and Shirihai 2010; Tennyson et al. 2012). Here we attempted, for the first time, to review the systematic of this controversial species complex by means of a multi-locus approach and coalescent-based analyses with the maximum sampling effort to date in these taxa.

Population assignment revisited: mtDNA vs. nuclear DNA and Species tree inference

Haplotype networks, gene trees and species trees un-ambiguously supported two mitochochondrial genetic lineages: the Australian- New Caledonian populations on one clade, and the south-west pacific populations on the other. Results were more nuanced for nuclear genes where polyphyly stages of weak differentiation were detected. MtDNA is well known for its high mutation rate and lower effective population size (N_e) compared to nuclear DNA, thus lineage sorting will occur faster for mtDNA than for nuDNA (Avise et al. 1987; Zink and Barrowclough 2008; Avise 2009). When such discrepancy occurs, a multi nuclear loci inference is recommended (Heled and Drummond 2010a; Hey 2010a). Species tree incorporating both mtDNA and nuclear DNA recover signals of high and significant differentiation between these lineages.

Similarly and most probably because of its coalescent nature, nuclear intron genes failed to recover high and significant levels of structuration within *P. brevipes* clade. The Rarotonga specimen (NMNZ, OR.023110) and Viti Levu (Fiji, AMNH 528336), the latter from a presumably extinct population: BirdLife International 2011a; Tennyson et al 2012) were placed into the *P. brevipes* lineage (with 1310 bp for concatenated mtDNA), though Rarotonga bird showed one private haplotype (H5), being separated by only one mutation from the main haplotype group (H1). Unfortunately, no nuclear sequences could be recovered for this individual. In addition, there is on-going debate about whether the Raivavae (Australes) bird (NMNZ 25377) may be a *P. leucoptera* or *P.* brevipes. Both mitochondrial and nuclear data placed the bird into the *P. brevipes* clade, though with a private haplotype (H6) for PAXIPI nuclear intron. Birds from Society were also placed within *P. brevipes* (5 individuals) though they showed higher haplotype variety for nuclear introns than for mitochondrial DNA. Our data for Vanua Lava consisted in 6 individuals, including the ones used by Bretagnolle and Shirihai (2010) representing both museum and fresh samples (Appendix Table A.3 for museum Identification). Haplotypes diversity results showed that *Pterodroma brevipes magnificens* showed high numbers of private haplotypes (IRF2 intron loci for example showed three private haplotypes H2, H3, H4, though separated by only one single mutation step from the main haplotype group). Similarly PAXIPI showed two haplotypes also separated by one mutation transition from the main haplotype. In this regards our results suggest that there is incipient structuration in this population.

Historical demography

Despite robust statistical methods now exist to infer past population history, caution has to be taken in regard to mutation rate used with such analyses (Hope and Pask 1998; García-Moreno et al. 2003) since coalescent estimates of past population changes depend on assumed mutation rate. Potential sources of error include rate

differences between evolutionary lineages and among genetic loci, or difficulty of accounting for saturation (successive changes at single nucleotide sites). Acknowledging such issues, the Cytb Weir and Schluter (2008) inter-lineage molecular rate (1.89% per milion year +-0.35, Procellariiformes) was chosen and converted into a per-lineage mutation rate following Nei (1987), which accounts for both lineages and calibration time, criteria frequently forgotten when calculating this rate (Axelsson et al. 2004; Weir and Schluter 2008). EBSP multi-locus results suggested that only one change occurred in the population trajectory of both clades, split time between *P.l. caledonica* and *P.l. leucoptera* clades assessed by *Beast suggested 110 000 years bp using a Constant clock and 600 000 years bp using a relaxed clock, underlying the high sensibility of the analyse to the priors used (e. g. molecular clock under constant evolution assumes that mutation rates in DNA and protein sequences through time are constant evolving among lineages but not across different region of the genome (Ho and Duchène 2014).

Conclusions and Taxonomic fate

Moritz (1994) defined ESUs-as Evolutionary significant Units that could be recognized as reciprocally monophyletic for mtDNA alleles and show significant divergence of allele frequencies at nuclear loci even if phylogeographic structuring of nuclear gene variation is absent. MUs-Management Units Criteria defined as populations that do not show reciprocal monophyly (either mtDNA or nuDNA), but have diverged in allele frequency. Our data based on multi-locus analysis and coalescent based demographic tests showed two mitochondrial lineages: *P. brevipes* and *P. leucoptera* but no strong signals of differentiation for nuclear intron sequences. Following Moritz (1994) definition we can consider populations of this clade as ESUs. On the other side, within *P. brevipes* populations, neither mtDNA nor nuclear introns genes showed strong signals of differentiation, though some haplotype frequencies differences (that should be taken with caution regarding the barely number of samples per population). In terms of management policies we can consider this populations as Mus for conservation.

Authors contributions

VB, DP and NC collected the samples. VB and AIV conceived the research. CR BG and AIV performed the laboratory work. AIV analyzed the data. AIV and VB wrote the manuscript. All the authors read and approved the final manuscript.

Acknowledgments

The authors would like to thanks T.P. Birt for alignment advices. We are also thankful to G. Ibarguchi for technical advices with museum DNA sampling. We are also grateful to A. Gotscho and S.P. Tseng for exchanges that improved the IM and EBSP analyses. JF. Renno for scientifc exchanges who helped to improve and conceive the manuscript. We thank A. Welch who helped to improve the manuscript, P. Duncan for English corrections on a first draft . For field work in New Caledonia, VB thanks I. Brun, M. Pandolfi and especially L. Renaudet and P. Villard. We also thank M. Guichard and E. Pante who provided critical assistance with the SuperMachine YMIR- *Université de la Rochelle.* We also thanks, D. Pinaud for assistance with Q-Gis software*.* Finally this work was supported by the Région Poitou Charentes and Deux Sevres department, who funded IAIV PHD grant.

REFERENCES

Avise JC (2000) Phylogeography: the history and formation of species. Harvard university press.

Avise JC (2009) Phylogeography: Retrospect and prospect. J Biogeogr 36:3–15.

Avise JC, Arnold J, Martin Ball R, et al (1987) Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics.

Avise JC, Ball RM (1990) Principles of genealogical concordance in species concepts and biological taxonomy. In: Biology O surveys in evolutionary (ed)p 7

Axelsson E, Smith NGC, Sundström H, et al (2004) Male-biased mutation rate and divergence in autosomal, z-linked and w-linked introns of chicken and Turkey. Mol Biol Evol 21:1538–47.

Bandelt HJ, Forster P, Röhl a (1999) Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol 16:37–48.

Beerli P (2004) Effect of unsampled populations on the estimation of population sizes and migration rates between sampled populations. Mol Ecol 13:827–836.

Belfiore NM, Liu L, Moritz C (2008) Multilocus phylogenetics of a rapid radiation in the genus Thomomys (Rodentia: Geomyidae). Syst Biol 57:294–310.

Bensch S, Irwin DE, Irwin JH, et al (2006) Conflicting patterns of mitochondrial and nuclear DNA diversity in Phylloscopus warblers. Mol Ecol 15:161–71.

Bretagnolle V, Shirihai H (2010) A new taxon of Collared Petrel Pterodroma brevipes from the Banks Islands , Vanuatu. 130:286–301.

Brooke M (2004) Albatrosses and Petrels across the World. Oxford.

Brower a V (1999) Delimitation of phylogenetic species with DNA sequences: a critique of Davis and Nixon's population aggregation analysis. Syst Biol 48:199–213.

Chan KMA, Levin SA (2005) Leaky Prezygotic Isolation and Porous Genomes: Rapid Introgression of Maternally Inherited Dna. Evolution (N Y) 59:720–729.

Corander J (2003) Bayesian analysis of genetic population structure using BAPS.

Crandall K a. KA, Bininda-emonds ORPORP, Mace GMGM, Wayne RKRK (2000) Considering evolutionary processes in conservation biology. Trends Ecol Evol 15:290–295.

De Queiroz K (1998) The general lineage concept of species, species criteria, and the process of speciation: a conceptual unification and terminological recommendations. Endless forms: species and speciation.

De Queiroz K (2007) Species concepts and species delimitation. Syst Biol 56:879–886.

De Salle R, Giddings L (1986) Discordance of nuclear and mitochondrial DNA phylogenies in Hawaiian Drosophila. Proc Natl Acad Sci U S A 83:6902–6.

Degnan J, Rosenberg N (2009) Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends Ecol Evol 24:332–40.

Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol 7:214.

Edwards S V, Kingan SB, Calkins JD, et al (2005) Speciation in birds: genes, geography, and sexual selection. Proc Natl Acad Sci U S A 102 Suppl :6550–7.

Edwards S, Bensch S (2009) Looking forwards or looking backwards in avian phylogeography? A comment on Zink and Barrowclough 2008. Mol Ecol 18:2930–3; discussion 2934–6.

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol

Resour 10:564–567.
Fu Y-X (1997) Statistical Tests of Neutrality of Mutations Against Population Growth, Hitchhiking and Background Selection.
Fujita MK, McGuire JA, Donnellan SC, et al, (2010) Diversification and persistence at the arid-monsoonal interface. Australia-wide biogeography of the Bynoe's gecko (Heteronotia binoei; Gekkonidae). Evolution 64, 2293–2314.
Gangloff B, Shirihai H, Watling D, et al (2011) The complete phylogeny of Pseudobulweria, the most endangered seabird genus: systematics, species status and conservation implications. Conserv Genet 13:39–52.
Gangloff B, Zino F, Shirihai H, et al (2013) The evolution of north-east Atlantic gadfly petrels using statistical phylogeography. Mol Ecol 22:495–507.
García-Moreno J, Sorenson MD, Mindell DP (2003) Congruent avian phylogenies inferred from mitochondrial and nuclear DNA sequences. J Mol Evol 57:27–37.
Hall T (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. 41:95–98.
Hasegawa M, Kishino H, Yano TA (1985) Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J Mol Evol 22:160–174.
Heled J, Drummond AJ (2010a) Bayesian inference of SPECIES TREES from multilocus data. Mol Biol Evol 27:570–80.
Heled J, Drummond AJ (2010b) Bayesian Inference of Species Trees from Multilocus Data Research article. 27:570–580.
Hey J (2010a) The divergence of chimpanzee species and subspecies as revealed in multipopulation isolation-with-migration analyses. Mol Biol Evol 27:921–33.
Hey J (2010b) Isolation with migration models for more than two populations. Mol Biol Evol 27:905–20.
Hindwood K, Serventy D (1941) The Gould petrel of Cabbage Tree Island.
Ho S, Duchène S (2014) Molecular-clock methods for estimating evolutionary rates and timescales. Mol Ecol 23:5947–5965.
Hope G, Pask J (1998) Tropical vegetational change in the late Pleistocene of New Caledonia. Palaeogeogr Palaeoclimatol Palaeoecol 142:1–21. doi: http://dx.doi.org/10.1016/S0031-0182(97)00140-5
Hudson RR, Kaplan NL (1985) Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics 111:147–64.
Imber MJ, Jenkins JAF (1981) The New Caledonian Petrel.
Kimball RT, Braun EL, Barker FK, et al (2009) A well-tested set of primers to amplify regions spread across the avian genome. Mol Phylogenet Evol 50:654–660.
Knowles LL, Carstens B (2007) Delimiting Species without Monophyletic Gene Trees. Syst Biol 56:887–895.
Kulikova I V, Drovetski S V., Gibson DD, et al (2005) Phylogeography of the Mallard (Anas Platyrhynchos): Hybridization , Dispersal , and Lineage Sorting Contribute To Complex Geographic Structure. Auk 122:949–965.
Leaché AD, Koo MS, Spencer CL, et al (2009) Quantifying ecological, morphological, and genetic variation to delimit species in the coast horned lizard species complex (Phrynosoma). Proc Natl Acad Sci U S A 106:12418–23.
Lee JY, Edwards S V (2008) Divergence across Australia's Carpentarian barrier: statistical phylogeography of the red-backed fairy wren (Malurus melanocephalus). Evolution 62:3117–34.
Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA

polymorphism data. Bioinformatics 25:1451–2.
Mace GM (2004) The role of taxonomy in species conservation. Philos Trans R Soc Lond B Biol Sci 359:711–719.
Militão T (2014) Comparing Multiple Criteria for Species Identification in Two Recently Diverged Seabirds. 1–27.
Moritz C (1994a) Defining “evolutionarily significant units” for conservation. Trends Ecol Evol 373–375.
Moritz C (1994b) Applications of mitochondrial-DNA analysis in conservation- A Critical Review. Mol Ecol 3:401–411.
Naurois R, Rancurel P (1978) Observations nouvelles sur les Laridae reproducteurs en Nouvelle-Caledonie. 287:495–498.
Nei M (1987) Molecular evolutionary genetics. Columbia university press
Paetkau D (1999) Using genetics to identify intraspecific conservation units: A critique of current methods. Conserv Biol 13:1507–1509.
Palsbøll PJ, Bérubé M, Allendorf FW (2007) Identification of management units using population genetic data. Trends Ecol Evol 22:11–16.
Paradis E, Claude J, Strimmer K (2004) APE: Analyses of phylogenetics and evolution in R language. Bioinformatics 20:289–290.
Primmer CR, Borge T, Lindell J, Sætre GP (2002) Single-nucleotide polymorphism characterization in species with limited available sequence information: High nucleotide diversity revealed in the avian genome. Mol Ecol 11:603–612.
R Development Core Team R (2011) R: A Language and Environment for Statistical Computing. R Found. Stat. Comput. 1:409.
Rambaut A, Suchard M, Xie D, Drummond A (2014) Tracer v1.6. Available from http://beast.bio.ed.ac.uk/Tracer.
Ryder O a. (1986) Species conservation and systematics: the dilemma of subspecies. Trends Ecol Evol 1:9–10.
Salicini I, Ibáñez C, Juste J (2011) Multilocus phylogeny and species delimitation within the Natterer’s bat species complex in the Western Palearctic. Mol Phylogenet Evol 61:888–898.
Sambrook, Fritsch, Maniatis (1989) Molecular Cloning: A Laboratory Manual, 2nd Editio.
Seitre R, Seitre J (1991) Causes de disparition des oiseaux terrestres de Polynésie française. South Pacific Commission.
Sibley C, Monroe B (1990) The ecology of bird communities, processes of birds of the World.
Sites JW, Marshall JC (2003) Delimiting species: A Renaissance issue in systematic biology. Trends Ecol Evol 18:462–470. doi: 10.1016/S0169-5347(03)00184-8
Slade RW, Moritz C, Heideman A (1994) Multiple nuclear gene phylogenies - application to pinnipeds and comparison with a mitochondrial DNA phylogeny. Mol Biol Evol 11:341–356.
Slatkin M (2005) Seeing ghosts: The effect of unsampled populations on migration rates estimated for sampled populations. Mol Ecol 14:67–73.
Sousa V, Hey J (2013) Understanding the origin of species with genome-scale data : modelling gene flow. Nat Publ Gr 14:404–414.
Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. Am J Hum Genet 68:978–989.
Tajima F (1989) The effect of change in population size on DNA polymorphism. Genetics 123:597–601.

Tamura K, Peterson D, Peterson N, et al (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28:2731–9.

Tennyson AJD, Miskelly CM, Totterman SL (2012) Observations of collared petrels (Pterodroma brevipes) on Vanua Lava, Vanuatu, and a review of the species' breeding distribution. Notornis 59:39–48.

Thompson JD, Gibson TJ, Plewniak F, et al (1997) The CLUSTAL X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25:4876–4882.

Waples RS, Gaggiotti O (2006) What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. Mol Ecol 15:1419–39.

Weir JT, Schluter D (2008) Calibrating the avian molecular clock. Mol Ecol 17:2321–8.

Welch AJ, Yoshida A a, Fleischer RC (2011) Mitochondrial and nuclear DNA sequences reveal recent divergence in morphologically indistinguishable petrels. Mol Ecol 20:1364–77.

Wiens J (2007) Species Delimitation: New Approaches for Discovering Diversity. Syst Biol 56:875–878.

Yang Z, Rannala B (2014) Unguided Species Delimitation Using DNA Sequence Data from Multiple Loci. Mol Biol Evol 31:3125–3135.

Yang Z, Rannala B (2010) Bayesian species delimitation using multilocus sequence data. Proc Natl Acad Sci U S A 107:9264–9.

Zhou R, Zeng K, Wu W, et al (2007) Population Genetics of Speciation in Nonmodel Organisms: I. Ancestral Polymorphism in Mangroves. Mol Biol Evol 24:2746–2754.

Zink RM, Barrowclough GF (2008) Mitochondrial DNA under siege in avian phylogeography. Mol Ecol 17:2107–21.

CHAPTER V

Population expansion, current and past gene flow in Gould's Petrel: implications for conservation

(Paper Accepted and Published)

by

Iglesias-V., A[A]., Gangloff, B[A, E]., Ruault, S[A]., Ribout C[A]., Priddel, D[B]., Carlile, N[B]., Friesen, V[C]., Cibois, A[D]. & Bretagnolle, V[A].

"It is not the strongest or the most intelligent who will survive but those who can best manage change."

~Leon C. Megginson

Population expansion, current and past gene flow in Gould's Petrel: implications for conservation

Iglesias-V., A[A]., Gangloff, B[A, E]., Ruault, S[A]., Ribout C[A]., Priddel, D[B]., Carlile, N[B]., Friesen, V.L.[C]., Cibois, A[D]. & Bretagnolle, V[A].

[A]Centre d'Etudes Biologiques de Chizé, UMR 7372, CNRS & Université de La Rochelle, 79360 Villiers en Bois, France.
[B]Office of Environment and Heritage, PO Box 1967, Hurstville, NSW 2220, Australia.
[C]Department of Biology, Queen's University, Kingston, Ontario, Canada K7L3N6.
[D]Department of Mammalogy and Ornithology, Natural History Museum of Geneva, 1211 Geneva 6, Switzerland.
[E]Service de Systématique Moléculaire, UMS 2700 OMSI, MNHN – CNRS, 75005 Paris, France

ABSTRACT

Seabird life-history traits such as long generation time, low annual fecundity and delayed sexual maturation make them more prone to population loss and consequently to extinction; petrels are indeed amongst the most threatened birds. Based on coalescence and multiloci inference this study examines the extent of genetic differentiation of a vulnerable New Caledonia (*caledonica*) and Australia (*leucoptera*) subspecies of *Pterodroma leucoptera* (Gould's Petrel), and whether the genetic relationship between them results from the influence of past events like variation in sea level, or is dominated by contemporary gene flow. Sequences of two mitochondrial genes and five nuclear introns were obtained from 86 individuals from both populations. Haplotype networks were used to infer relationships between the haplotypes of both populations. The demographic history of the *P. leucoptera* complex was studied using neutrality tests and Extended Bayesian Skyline Plots. A weak population differentiation was revealed. The Extended Bayesian Skyline plot suggested a population expansion approximately 80,000 years before present (bp) for *caledonica* and 30,000 years bp for *leucoptera*. The split was dated to 30 000 years bp by means of multilocus inference through *BEAST. Despite genetic similarity of the two taxa, we advocate to consider them as independent units for conservation management, given their strong ecological distinctiveness (foraging distribution, winter distribution, breeding phenology and breeding distribution).

Key words: Phylogeography; Extended Bayesian Skyline plot; Procellariiformes; seabird; Isolation with Migration; conservation genetics.

INTRODUCTION

Anthropogenic impacts on islands, in particular biological invasion and habitat destruction, have been identified as the most important causes of genetic diversity loss, population extinction and, island fauna depletion (Lande 1998; Steadman 2006; Illera et al. 2012). Population extinction is (usually) a long process through which the population first declines in size, then experiences demographic stochasticity and finally becomes extinct (Caughley 1994). Indeed, population genetic theory predicts that when a population declines, genetic diversity is lost as a result of genetic drift and inbreeding depression (Frankham et al. 2002; Berec et al. 2006; Allendorf and Luikart 2009). However, disentangling the relative roles of contemporary and historical processes on the overall genetic diversity and population differentiation, and ultimately, survival of populations, is notoriously difficult (Chiucchi and Gibbs 2010; Henriques et al. 2014).

Maximum likelihood and Bayesian analyses based on coalescent theory, when applied to DNA (site and length polymorphism), may provide insights into demographic history and genetic structure (Rocha et al. 2011; Brown et al. 2012; Zieliński et al. 2013). Coalescent theory assumes that for any two genetic sequences drawn from a population at random, the probability that they coalesce at a given time is a function of population size at that time (Kingman 1982). Coalescent theory also allows estimation of historic gene flow (Beerli and Felsenstein 1999; Hey and Nielsen 2004). This approach has proved useful in conservation genetics to understand patterns of intraspecific diversity, especially when population size and gene flow are key factors for the long-term survival of conservation units (Hey 2005; Rocha et al. 2011; Zieliński et al. 2013).

Pleistocene climatic oscillations have influenced population distribution, gene flow and genetic variability through changes in habitat availability (e.g. Pielou 2008; Cheang et al. 2012). In particular, Pleistocene glacial/interglacial cycles have had marked effects on seabirds due to their extreme characteristics in life-history traits that make them particularly suitable for investigating past and recent population changes using genetic tools. Among seabirds, petrels (Procellariidae) are the most extreme in regard to demographic parameters (as reviewed in Dobson and Jouventin 2007). They usually show large distributions, and have very high dispersal abilities (Shaffer et al. 2006). In addition, they breed on remote oceanic islands and exhibit high natal philopatry, often returning to breed within a few metres of their natal nest (Rabouam et al. 1998; Huyvaert and Anderson 2004). High dispersal ability and philopatry act in opposition: strong dispersal capacities should promote gene flow (van Bekkum et al. 2006), while their strong philopatry should promote genetic differentiation (Burg and Croxall 2001; Dearborn et al. 2003).

Within the petrels, the genus *Pterodroma* (c.35 species) alone accounts for 10% of all seabird species. Gould's Petrel *Pterodroma leucoptera* is a small pelagic gadfly petrel (200-250 g), breeding only on two sites separated by c.1500 km in the southwest Pacific Ocean (Fig. 5.1). Each population has been treated as an endemic subspecies because of subtle differences in morphology and coloration (Imber and Jenkins 1981; Bretagnolle and Shirihai 2010). The Australian subspecies, *P. l. leucoptera* (hereafter *leucoptera*), now breeds only on two small islands in New South Wales (Cabbage Tree Island and Boondelbah Island separated by 1.6 km: Carlile et al. 2003), while the New Caledonian subspecies (*P.l. caledonica,* hereafter *caledonica*) is restricted to the south central chain of New Caledonia (Naurois and Rancurel 1978; Bretagnolle and Shirihai 2010). Several thousands of birds have

been ringed but no exchange has ever been documented; however long-term ringing has been done only on Cabbage Tree Island (Priddel et al. 2014). Subspecies also differ in breeding behavior and habitats (*leucoptera* nests in natural cavities among rock scree close to sea level, *caledonica* excavates soil burrows high in the mountains), breeding phenology (a lag of one month; Priddel et al. 2014), foraging zones while breeding, and migration and non-breeding areas (Priddel et al. 2014). However, the genetic relationship between the two taxa remains mostly unknown. Both taxa experienced recent fluctuations in population sizes: *leucoptera* was numerous when discovered in the eighteenth century (Gould 1865), but decreased to fewer than 1500 individuals in 1992 (c. 200 pairs). Classified as *vulnerable* (IUCN 2015), its population recovered to 1000 pairs thanks to a recent restoration program (Priddel and Carlile 1995; Priddel and Carlile 2009). Although no precise information on population trend exists for *caledonica*, its numbers decreased following the introduction of predators (black rats *Rattus rattus*, cats *Felis catus* and pigs *Sus scrofa*) with European settlement approximately 190 years ago (Miller and Mullette 1985; Armstrong 1992; IUCN 2015).

In this study, seven molecular markers were used (two mitochondrial genes and five nuclear introns), i) to clarify whether taxonomic treatment of the two subspecies is supported by molecular data, ii) to test whether present genetic structure results from demographic fluctuations (expansions and bottlenecks) due to Pleistocene climatic oscillations, or from more recent changes likely related to anthropogenic pressure; and iii) to estimate time of divergence of the two populations and their past effective population sizes.

METHODS

Sampling, DNA purification, gene amplification and sequencing

A total of 86 samples of *Pterodroma leucoptera*, consisting of blood (45) and feathers (41), were collected from the two known breeding localities (Fig. 5.1). Feathers were sampled from live adult birds from the Mt. Dzumac colony, New Caledonia, during the breeding seasons of 2005-2006 and 2008-2009. Similarly, blood samples were collected from Australian birds at the Cabbage Tree Island colony from 2005 to 2008. Blood was collected from veins on the leg or wing using microcapillaries. Samples were stored until analysis in 70% ethanol at -20°C. Total genomic DNA was isolated from blood samples using QiagenDNeasy Blood and Tissue extraction kits (Qiagen, Inc., Valencia, CA). Polymerase Chain Reaction (PCR) was used to amplify two mitochondrial DNA (mtDNA) loci and five nuclear introns using previously published primers (Appendix Table B1). Amplifications were performed in 25µl reactions containing 1.5 µl (30-100 ng/ µl) of the DNA template, 1X QIAGEN Multiplex PCR Master Mix (Qiagen, UK) and 0.8µM of each forward and reverse primer (Appendix Table B1) PCR products were purified and sequenced using the same PCR primers by Eurofins Scientific (France) and Genome Québec Innovation Centre (McGill University, Montreal, QC, Canada).

Mitochondrial origin and intralocus recombination

Mitochondrial origin of the concatenated mtDNA gene sequences was confirmed by translating DNA sequences to check for stop codons and other potential indications of nuclear origin (Ibarguchi et al. 2006) in BioEdit Sequence alignment Editor v. 7.2.5 (Hall 1999). Sequences were checked visually and aligned with CLUSTAL X (Thompson et al. 1997), implemented in BioEdit Sequence alignment Editor.

Recombination in nuclear loci was tested with the four gamete test (Hudson and Kaplan 1985) implemented in DnaSp v.5.10.01 (Librado and Rozas 2009). When the

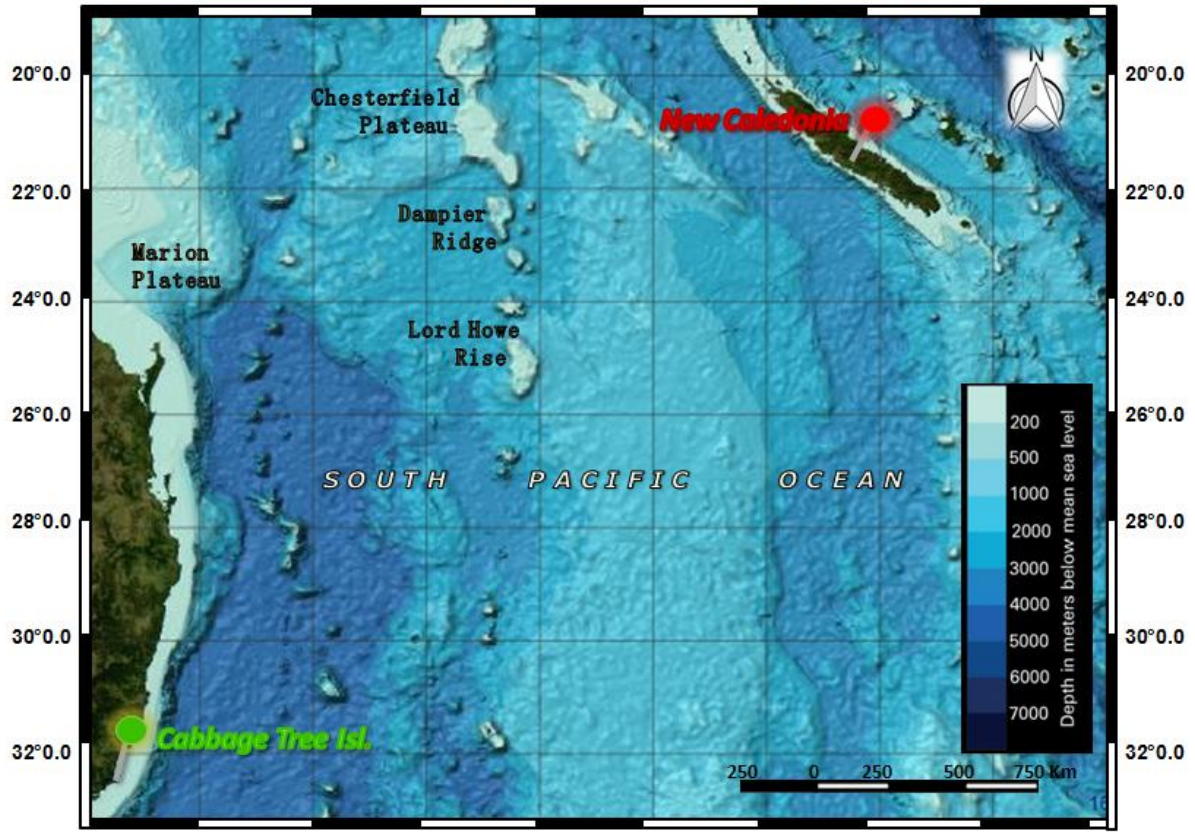


Figure 5.1. Sampled breeding localities of *Pterodroma leucoptera*: Dzumac, New Caledonia; Cabbage Tree Island, Australia. Bathymetry is also shown, with dark grey indicating 100m

test suggested intralocus recombination, we retained the longest contiguous unrecombined sequence for subsequent analyses. Because the Isolation with Migration and BEAST analysis requires having known phase for nuclear sequences, we determined the gametic phase using the program PHASE (Stephens et al. 2001) implemented in DnaSp with default parameters and a threshold value of 0.90. Phased nuclear data were then used in all analyses, individuals thus being represented by two sequences.

Evolutionary relationships, genetic diversity and population differentiation

Haplotype frequencies were inferred with DnaSP. Genealogical relationships among haplotypes were reconstructed using a median-joining network and default parameters settings (Bandelt et al. 1999) in NETWORK v.4.6.0.0 (http://www.fluxus-engineering.com). Concatenated mitochondrial and nuclear sequences were used to estimate general statistics of genetic diversity, including number of polymorphic sites (Np), number of haplotypes (H), number of private haplotypes (PH), haplotype diversity (h; Nei 1987) and nucleotide diversity (π; Nei and Tajima 1983), for each subspecies using DnaSP and Arlequin v. 3.5.1 (Excoffier and Lischer 2010). The proportion of genetic variance accounted for within and between subspecies was estimated using an analysis of molecular variance, AMOVA (Excoffier et al. 1992) in Arlequin, and tested for statistical significance using 10 000 permutations. We

calculated pairwise differentiation between subspecies using Φ_{ST} (with Tamura-Nei substitution model), a direct analogue of Wright's F_{st} for nucleotide sequence divergence.

Population size fluctuations through time

Two tests were used to assess if genetic variation deviated from neutral expectations due to either a recent population expansion or selection: Tajima's D (Tajima 1989) and Fu's F_s (Fu 1997) both implemented in Arlequin. Tajima's D values >0 suggest either a recent population bottleneck or balancing selection, while D<0 indicates a population expansion or directional selection. These values are considered significant when p<0.05. Fu's F_s tends to be negative when there is an excess of recent mutations (therefore an overabundance of rare alleles), characteristic of a recent population expansion. Positive Fu's F_s values indicate a deficiency of rare alleles, suggesting a population bottleneck or overdominant selection. Following (Fu 1997) Fu's F_s values are regarded as significant if p<0.02. which corresponds to the conventional significant p<0.05 for Tajima's D. Significant negative Tajima's D and Fu's F_s indices may also indicate selection and genetic hitchhiking associated with selective sweeps. These analyses were conducted for all genes.

A coalescent-based graphical method, the Extended Bayesian Skyline Plot (EBSP) was carried out in BEAST v. 2.1.3 (Drummond and Rambaut 2007) for both populations independently and pooled together to infer potential historical fluctuations in effective population size (N_e). The EBSP allows simultaneous analysis of data from multiple unlinked loci, taking into account their specific mode of inheritance, thus significantly improving the reliability of demographic inferences over single-locus analyses (Heled and Drummond 2008; Shapiro and Ho 2011). EBSP analysis was run using strict clock models as it is considered a good approximation for intra-population level analyses, and it simplifies the coalescent model, helping convergence. Per lineage mutation rate (mu) was calculated by using $\mu = d_{xy}/2T$ (Nei, 1987), where d_{xy} stands for interlineage divergence and T is the divergence time since two unique lineages diverged (See Appendix Table B3). For concatenated mtDNA, d_{xy}= 0.0189 s/s/My (substitution/site/million years) (Weir and Schluter 2008) and for nuclear introns d_{xy}= 0.0036 s/s/My (Axelsson et al. 2004); All analyses were run three times to check for convergence with 7×10^7 generations, and sampling every 2×10^3 generations. The first 30% of the genealogies were discarded as burn-in. Convergence, stationarity, effective sample size for each parameter of interest and the appropriate burn-in were evaluated using the software TRACER v.1.6 (Rambaut et al. 2014). A generation time of 20 years as indicated by (Gangloff et al. 2013) was used to convert effective population size to individuals using $N_e = \Theta/(4G\mu)$;N_e (effective population size) divided by G (generation time).

Estimation of gene flow and population connectivity

Patterns of historical and contemporary connectivity between the two subspecies were disentangled through a coalescent inference using the Isolation with Migration under Changing Population Size model (IM) (Hey and Nielsen 2004; Hey 2005). The following parameters were calculated between *leucoptera* and *caledonica*: effective population sizes (N_e *caledonica*; N_e *leucoptera*; N_e *ancestral*), population divergence time (T), the splitting parameter (S) allowing for population size change through time, and migration rates (M1, M2), where *M1* indicates the probability of migration per generation migrating from *caledonica* to *leucoptera* forwards in time, and *M2*

indicates migration in the other direction. Concatenated mtDNA and multiloci nuclear DNA were used both jointly and separately to perform the IM analysis. Several preliminary runs were conducted to optimize priors (looking for posterior density curves that rise from zero, peak and then fall to zero within the range for each of the required demographic parameters) following Hey 2009. The final analysis was carried out with the HKY mutation model (Hasegawa et al. 1985) for both nuclear introns and mtDNA, a geometric heating scheme (g1= 0.96 and g2= 0.9), 10 chains, and a chain length of 2 million steps following a 1 million step burn-in. To assess convergence, three separate runs were conducted with different random seed numbers. Effective sample size (ESS) values were monitored to ensure proper mixing of the Markov chain. To convert raw parameter estimates into demographic values, we used the per-locus mutation rates (substitution/year) obtained by multiplying per lineage mutation rate (as used in EBSP) by the number of base pairs of each sequence. The geometric mean of the per-locus mutation rates (μ) was calculated and then used to compute the divergence time by using T = t/μ, expressed in years before present (*t*, is the maximum likelihood estimate of the parameter T). To calculate effective population size (N_e), we used $N_e = \Theta /(4G \mu)$, with a generation time (G) of 20 years (Gangloff et al. 2013). To estimate the population migration rate (M), we used $2N_eM = N_em/2$, where *m* stands for the maximum likelihood estimate of the parameter M. The number of migrants from the ancestral to the *leucoptera* population was calculated as $(1-s)\Theta a$ (where (1-s) represents the size of the *P. leucoptera* population and Θa stands for the effective size of the ancestral population (Hey 2005).

As the divergence time obtained with IM was unreliable (Appendix Table B5; Fig. B.2), we also used *Beast (Heled and Drummond 2010), implemented in BEAST v.1.6.1 (Drummond and Rambaut 2007), which provides simultaneously phylogenies and divergence time estimates. To root the tree four related species were used as outgroups, two of them, *P. brevipes and P. oculta,* were amplified and sequenced in this study while sequences from *P. feae* and *P. madeira* were obtained from genebank (Appendix Table B4). Three runs of $5x10^7$ generations, sampling every 1000 generations with a burn in of 2000 trees were performed and then combined in LogCombiner v.1.6.1 (Heled and Drummond 2010). HKY nucleotide substitution model and a strict clock model with an uncorrelated lognormal distribution were used. Per lineage mutation rate of 4.87×10^{-3} s/s/My was used. For the tree, a Yule process speciation under a coalescent model assuming a constant population over the time period was chosen. Finally, Tracer v.1.6 (Drummond and Rambaut 2007) was used to visualize the results of the runs and to check the effective sample size of each parameter.

RESULTS

We sequenced 1327 base pairs for concatenated CO1 and Cytb (see Table 5.1 for exact numbers of individuals sequenced for each gene) and 500, 481, 924, 637, and 452 base pairs for the introns PAXIPI, CSDE1, βfibint7, IRF2F1 and TPM1 respectively. Of the five nuclear introns, Bfibint7 was the only one presenting signals of recombination, so we kept the longest possible contiguous unrecombined sequence (918 base pairs) for subsequent analyses. As TPM1 did not present any variation, it was withdrawn from any further analysis. The concatenated mtDNA data did not display insertions or deletions, and after translation, no nonsense or stop codons were found. No ambiguous sites were detected, and true mitochondrial origin

of obtained sequences was therefore very likely. DNA was amplified with specific primers designed for *Pterodroma* species (Primmer et al. 2002; Kimball et al. 2009; Jesus et al. 2009; Gangloff et al. 2013) rather than universal primers, which is supposed to reduce the risk of the coamplification of nuclear copies of mitochondrial genes (*numts*) amplification (Sorenson and Quinn 1998; Ibarguchi et al. 2006).

Molecular Variability

Concatenated mitochondrial data contained 18 polymorphic sites, with a total of 17 different haplotypes (13 for *caledonica* and 7 for *leucoptera*; Table 1). Nuclear data exhibited a total of 55 variable sites (7 in PAXIP1, 15 in CSDE1, 26 in βfibint7 and 7 in IRF2F1), leading to 7, 9, 17 and 4 haplotypes respectively (after phasing the nuclear intron sequences). Haplotype and nucleotide diversities tended to be lower in *leucoptera* than in *caledonica* (Table 1).

Population Structure and evolutionary relationships

Pairwise estimates of population (subspecies) differentiation for mtDNA and nuDNA were low but significantly different from 0 between the two subspecies taken together ($p<0.05$ for all global mtDNA and nuDNA except PAXIPI ($p>0.05$). Indeed, >90% of genetic variation was detected at the intra-population level for all loci (Appendix Table B2). No pronounced phylogeographic structure could be detected in the haplotype networks, either for mitochondrial or nuclear markers (Fig. 5.2). Indeed, all networks were characterized by one or more dominant haplotypes shared by the two subspecies. Similarly, most networks showed star-like topologies, with one central prevalent haplotype, and other haplotypes having much lower frequencies, suggesting possible past and rapid population expansion (Slatkin and Hudson 1991; Kulikova et al. 2005). *Pterodroma l. caledonica* showed a higher number of private haplotypes despite a lower sample size (Table 5.1 and Fig. 5.2).

Table 5. 1. Diversity indices and results of tests for deviations from neutrality for two subspecies of *Pterodroma leucoptera.* Number of individuals sequenced (N), number of polymorphic sites (Np); number of haplotypes (Nh); number of private haplotypes (ph); haplotype diversity (Hd), nucleotide diversity (π, in %), average number of nucleotide differences between haplotypes (K). Significant values for tests of neutrality: p<0.05 for Tajima's D and p <0.02 for Fu's F_s are shown in bold.

						Diversity Indices		Neutrality Tests	
Gene	**Subspecies**	**N**	**Np**	**Nh**	**ph**	**Hd [SD $^{(x10-2)}$]**	**π% [SD]**	**Tajima's D**	**Fu's Fs**
CO1-Cytb									
	caledonica	17	13	11	9	0.93 [0.05]	0.16 [0.09]	-1.04	**-5.19**
	leucoptera	20	7	8	6	0.79 [0.07]	0.063 [0.05]	-1.40	**-4.46**
PAXIPI									
	caledonica	26	2	4	2	0.28 [7.0]	0.04 [0.06]	-0.64	-2.12
	leucoptera	56	5	7	5	0.21 [5.2]	0.05 [0.06]	**-1.41**	**-5.81**
CSDE1									
	caledonica	16	10	8	6	0.68 [5.6]	0.08 0.06]	**-1.62**	**-3.08**
	leucoptera	50	5	5	3	0.38[5.0]	0.03 [0.03]	-1.16	-1.83
βfibint7									
	caledonica	15	19	17	13	0.88 [5.1]	0.09 [0.06]	**-1.87**	**-12.9**
	leucoptera	35	7	8	4	0.56 [4.8]	0.031 [0.03]	-1.25	**-3.91**
IRF2F1									
	caledonica	26	4	5	3	0.68 [6.0]	0.25 [0.18]	0.47	-0.24
	leucoptera	45	3	3	1	0.46 [6.0]	0.19 [0.15]	0.86	1.89
TPM1									
	caledonica	15	0	1	0	0	0		
	leucoptera	18	0	1	0	0	0		

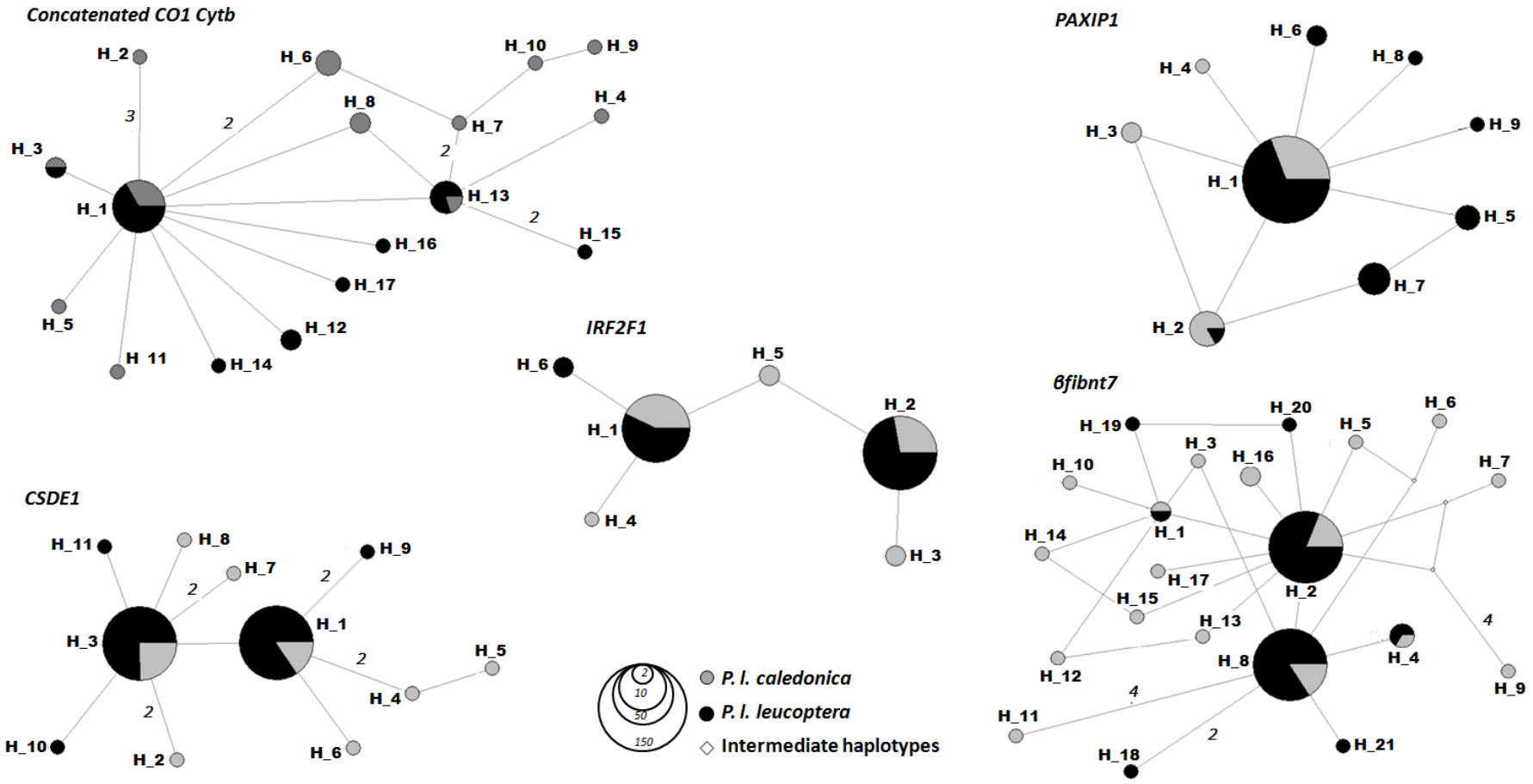


Figure 5.2. Haplotype networks obtained with concatenated mtDNA CO1 and Cytb genes, and five nuclear introns, PAXIP1, βFibint7, CSDE1 and IRF21 within *Pterodroma leucoptera* (TPM1 not shown since there was no variation). Node sizes are proportional to number of individuals found with that haplotype, while color codes refer to sampling site (see **Figure 5.1** for sample site locations). Polygons correspond to intermediate (non-sampled) haplotypes. Branch lengths are not scale to improve visualization. Numbers on branches show the number of mutations between alleles. When no number is indicated, only one mutation step occurred.

Past population history

Most estimates of Tajima's D and Fu's F_s were negative, suggesting a recent population expansion for either *caledonica* or *leucoptera* (see Table 5.1 for all values and associated statistics). Reconstruction of the population size history by means of EBSP suggested an expansion episode for both *caledonica* and *leucoptera* (Fig. 5.3). EBSP further underlined an earlier population increase in *caledonica* (c. 60,000 years bp) than in *leucoptera* (c. 15,000 years bp).

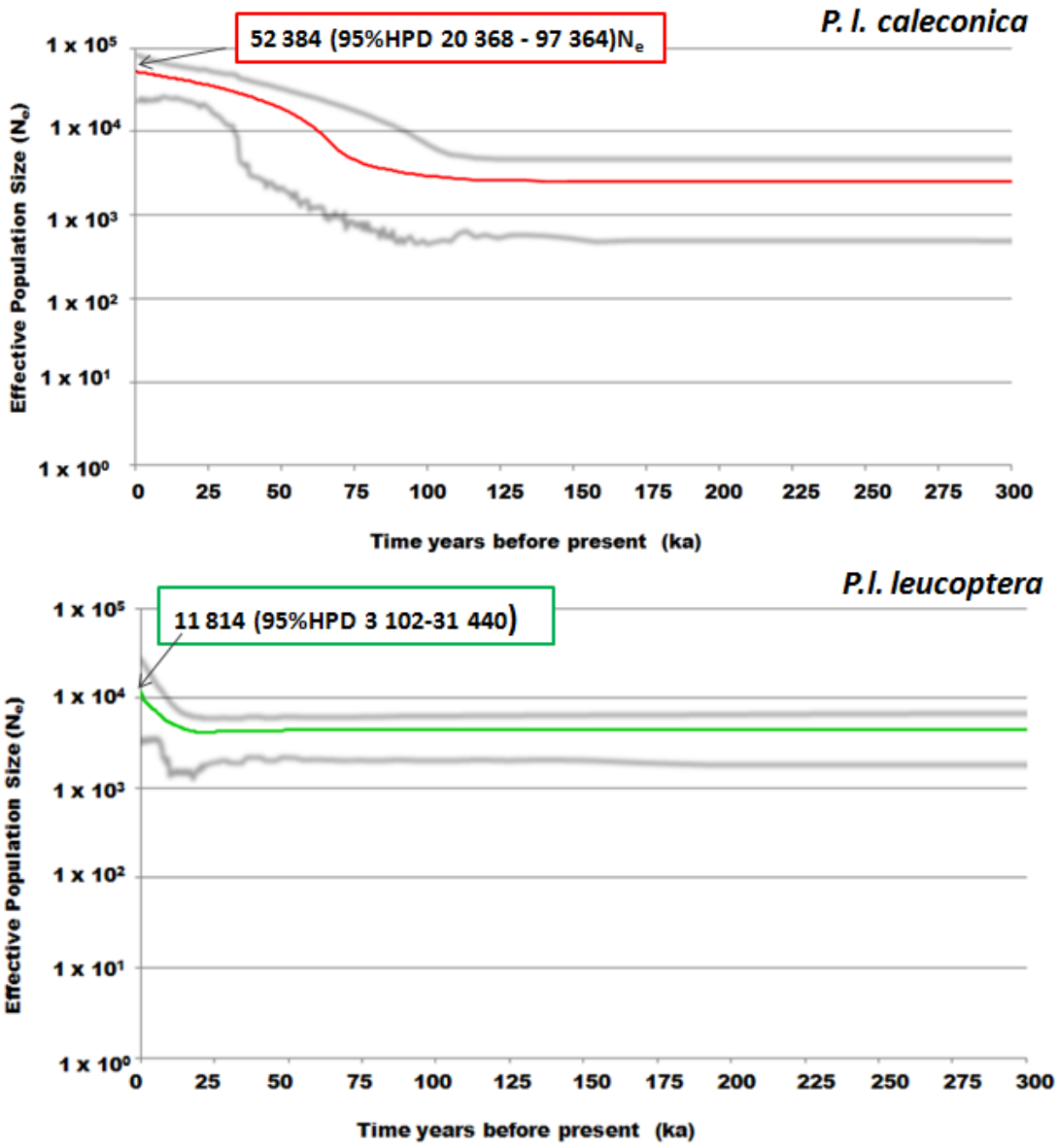


Figure 5.3. Extended Bayesian Skyline Plots of *P. l. caledonica* and *P. l. leucoptera* subspecies based on the CO1, Cytb, PAXP1, CSDE1, βFibint7 and IRF21 data sets. The green and red lines represent the median of the parameter N_e, proportional to the effective population size in thousands, while the grey lines are the 95% confidence interval.

Estimation of gene flow and Population Connectivity

Migration rates, divergence time and present effective population sizes were obtained with IM analysis only for the two mitochondrial loci, since our data failed to find convergence when using all loci. Posterior density curves were acceptable (see Hey 2007) for all but three parameters (Appendix Table B5, Fig. B.2), and therefore, estimates of Θa, T (divergence time) and *f (*number of founder individuals of *leucoptera)* were unreliable and not presented in the results. Estimates of divergence time obtained by BEAST analysis indicated a split event of 30 000 years ago (Appendix Fig. B1) between *P. l. caledonica* and *P. l. leucoptera* .

DISCUSSION

The two *P. leucoptera* subspecies show evidence of recent divergence with presence of ongoing gene flow between them, as indicated by high numbers of shared haplotypes, few private haplotypes, and low indices of differentiation. In the absence of population bottlenecks, higher haplotype and nucleotide diversities are expected in ancestral populations compared to more recent ones (Bisconti et al. 2011). Here, we found greater nucleotide and haplotype diversity in *caledonica*, for five out of the six loci under study, suggesting that *caledonica* may be the ancestral form.

A proposed historical reconstruction

Pterodroma l. caledonica was apparently stable in numbers and remained at low population size during the climate cooling period (and associated lowering sea levels) that extended from 130,000 to 70,000 years bp (Lambeck et al. 2002). We found no sign of population expansion during this first Pleistocene period, despite expansions being documented for other marine species (Crandall et al. 2008; Kochzius et al. 2009). From 70,000 years bp, a very smooth increase in population size is however apparent, possibly in relation to further cooling of water temperature and a presumable increase in ocean productivity: today, the two populations feed in southern seas during breeding, and this is especially marked in *caledonica* (Priddel et al. 2014). We thus suggest that colder, more productive waters, moving slowly northward and therefore closer to breeding localities in New Caledonia, may have allowed this population increase. The two lineages apparently split during the Last Glacial cycle (LG, starting 70,000 years bp). Between 70,000 and 20,000 years bp sea level dropped from c.-80m to -120m, while sea surface temperature decreased sharply (Rohling et al. 1998; Barrows et al. 2007). Several ridges between New Caledonia and Australia, currently underwater (e.g. Chesterfield Ridge; Fig. 5.1), were presumably above sea level, providing many lowland islets. *P. l. caledonica* may have colonized one or several islets halfway between New Caledonia and Australia at this time (see Fig. 1). Divergence hypothetically occurred between the populations, leading to the *leucoptera* lineage. Interestingly, these new islands were lowland islets, a character now found in the breeding habitat of *leucoptera*. Around 18,000 years bp, sea level and sea surface temperatures started to rise. Breeding sites became unavailable for *leucoptera* and the population then presumably colonized or survived only on the few islands off the Australian coast, where they are breeding today.

This proposed historical scenario however relies on accurate estimate of time of divergence as well as population size estimates. However, our IM simulations did not fully converge for parameters such as divergence time and long term effective size. Indeed similar lack of convergence was already found in other petrel studies (Welch et al. 2011), with the posterior distribution of the curves showing two distinct peaks, hence very wide HPD. Low resolution in parameter estimates with IM can result from sensitivity to inadequate sampling of target populations (Jacobsen and Omland 2012), since parameter estimates will also depend on the level of gene flow between unsampled populations. Besides, low numbers of loci coverage was suggested to promote failure convergence of the IM parameters estimates (Hey 2005; Jacobsen and Omland 2012). In our study we also used multi locus coalescence-based approaches implemented in EBSP and *BEAST rather than concatenated sequences analysis (Degnan and Rosenberg 2006; Kubatko and Degnan 2007), the latter accounting for discrepancies between gene and species trees history when studying closely related species (Zink and Barrowclough 2008). Indeed *BEAST accepts shared polymorphisms originated from incomplete lineage sorting but excludes introgression, treating the latter by conducting separate analyses for each marker (Heled and Drummond 2010; Drummond and Bouckaert 2015; Meyer et al. 2016). To conclude, polyphyly and dissimilarities in gene tree topologies were expected since we analysed closely related species or subspecies, as was appointed by Maddison and Knowles 2006.

The accuracy of estimates of divergence times based on evolutionary substitution rates is also increasingly debated (Garcia-Moreno 2004; Lovette 2004; Ho et al. 2007; Shapiro and Ho 2011), since they are affected by base composition, calibration point sensitivity, generation time, metabolic rates and population size (Lovette 2004; Ho et al. 2005). For instance, Nunn and Stanley 1998 used a calibration point older than 12 Myr, causing saturation of mutations in their cytb sequences and consequently underestimating sequence divergence. It is indeed recommended to use calibrations derived from lineages as close as possible to the organism under study (Lovette 2004; Peterson 2006). For these reasons, we used the interlineage molecular rate derived from (Weir and Schluter 2008) and converted it to a per lineage mutation rate. Nevertheless, our estimate of ~ 30 000 years bp divergence time should be regarded cautiously given remaining uncertainties in molecular rate estimates.

Contemporary differentiation and taxonomic consequences

These two populations are currently classified as separate subspecies based on slight morphological and color differentiation (Imber and Jenkins 1981). Some researchers even suggested species status (e.g. (Onley and Scofield 2007) despite strong overlap in measurements and coloration (Bretagnolle and Shirihai 2010). Our haplotype network, based on 7 loci including both nuclear introns and mitochondrial DNA, revealed very low population differentiation between the two taxa based on Φ_{ST} statistics (mitochondrial Φ_{ST}= 0.01, nuclear Φ_{ST} ranging 0.0049-0.0876 (Appendix

Table B2) and no phylogeographic differentiation in haplotype networks. Several petrel studies have now used both nuclear and mtDNA, revealing in general greater resolution in mitochondrial loci and deeper levels of genetic divergence than for nuclear loci (Silva et al. 2011; Welch et al. 2011; Gangloff et al. 2011; Silva et al. 2015). Unlike other petrels (Ovenden et al. 1991; Friesen et al. 2006), our two taxa showed few locally restricted haplotypes. Conversely, these populations differ in their breeding as well as non-breeding ecology (Priddel et al. 2014). Differences in ecological traits despite similarity at neutral molecular markers and few locally restricted haplotypes may suggest recent divergence with on-going gene flow between *caledonica* and *leucoptera*, with lineages currently unsorted but likely in the process of divergence, rather than a remnant of a large ancestral population.

Conservation implications

Conservation management should target demographically independent populations whose population dynamics depend largely on local birth and death rates rather than on migration (Palsbøll et al. 2007). Conservation genetics can help decide whether subspecies or populations within a species should be managed as separate units (Moritz 2002). Furthermore, independently evolving populations are arguably worth conserving even if they are not different species or taxa (Tobias et al. 2010). The present study provides an interesting case, where neutral molecular markers failed to find strong differences between two populations that are traditionally considered separate conservation units given their different ecological requirements and breeding habitats. In addition, these populations currently have different potential fates: *leucoptera* shows a very small but increasing breeding population on safe islands, while *caledonica* shows a much bigger but currently declining population on an unsafe island where predation by invasive species occurs. Overall, *Pterodroma leucoptera* is considered vulnerable (Birdlife International 2015.). Given differences in ecology and conservation status, and despite the weak neutral genetic and morphological differentiation, we thus recommend that both populations should be protected to preserve the evolutionary potential of these lineages. In particular, as a source of variability able to maintain the genetic diversity of this species, taxon *caledonica* warrants more conservation effort.

Authors contributions

VB, DP and NC collected the samples. VB, BG and AI-V conceived the research. SR, BG and AI-V performed the laboratory work. AI-V analyzed the data, with the help of AC. AI-V and VB wrote the manuscript. All the authors read and approved the final manuscript.

Acknowledgments

The authors would like to thank T.P. Birt for alignment advice and primer dessign. We also thank to the Genome Québec Innovation Centre (McGill University Montreal, QC, Canada) for facilitating part of the sequencing of this work. We are also grateful to A. Gotscho and S.P. Tseng for discussions that improved the IM and EBSP analyses. For field work in New Caledonia, VB thanks I. Brun, M. Pandolfi and especially L. Renaudet and P. Villard. We thank A. Welch who helped to improve the manuscript, and P. Duncan for English corrections on a first draft. Thanks also go to the editors and referees for commenting on and correcting the manuscript. We also

thank M. Guichard and E. Pante who provided critical assistance with the SuperMachine YMIR- *Université de la Rochelle.* Finally this work was supported by the Région Poitou Charentes and Deux-Sèvres department, who funded IVA PHD grant.

REFERENCES

Avise JC (2000) Phylogeography: the history and formation of species. Harvard university press.

Barrowclough GF (1983) Biochemical studies of microevolutionary processes.

Beerli P, Felsenstein J (1999) Maximum-likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. Genetics. 152:763–773.

Beerli P, Palczewski M (2010) Unified framework to evaluate panmixia and migration direction among multiple sampling locations. Genetics 185:313–326.

Carlile N, Priddel D, Zino F, et al (2003) A review of four successful recovery programmes for threatened sub-tropical petrels. Mar Ornithol. 31:185–192.

Castillo-Davies C, Mekhedov S, Hartl D, et al (2002) Selection for short introns in highly expressed genes. Nat Genet. 31:415–418.

Crisp M, Chandler G (1996) Paraphyletic species. Telopea 6:813–844.

del Hoyo J, Elliott A, Sargatal J, et al (2016) Handbook of the Birds of the World Alive. In: Lynx Edicions, Barcelona. (retrieved from http://www.hbw.com/ on 22/12/2015.

Edwards S, Beerli P (2000) Perspective: gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. Evolution, 54:1839–1854.

Friesen VL, Congdon BC, Walsh HE, Birt TP (1997) Intron variation in marbled murrelets detected using analyses of single-stranded conformational polymorphisms. Mol Ecol. 6:1047–58.

Friesen VL, Smith a L, Gómez-Díaz E, et al (2007) Sympatric speciation by allochrony in a seabird. Proc Natl Acad Sci USA. 104:18589–94.

Garcia-Moreno J (2004) Is there a universal molecular clock for birds? J Avian Biol. 35:465–468.

Gómez-Díaz E, González-Solís J (2007) Geographic assignment of seabirds to their origin: Combining morphologic, genetic, and biogeochemical analyses. Ecol Appl 17:1484–1498.

Gómez-Díaz E, González-Solís J, Peinado MA, Page RDM (2006) Phylogeography of the Calonectris shearwaters using molecular and morphometric data. Mol Phylogenet Evol. 41:322–32.

Hey BJ (2007a) Using the IMa Program. 2785–2790.

Hey J (2009) Using the IM Program. Softw. Man. 27.

Hey J (2010) Isolation with migration models for more than two populations. Mol Biol Evol 27:905–20.

Hey J (2007b) Introduction to the IM and IMa computer programs Understanding Markov chain Monte Carlo (MCMC). How Long to Run the Program Metropolis Coupling of Markov chains Assessing the Autocorrelation of Recorded Values Parameter Conversions Cautions , Suggestio. 1–16.

Ho SYW, Heupink TH, Rambaut A, Shapiro B (2007) Bayesian Estimation of Sequence Damage in Ancient DNA. Mol Biol Evol. 24:1416–1422.

Ho SYW, Lanfear R, Bromham L, et al (2011) Time-dependent rates of molecular evolution. Mol Ecol 20:3087–3101.
Holsinger K, Weir B (2009) Genetics in geographically structured populations: defining, estimating and interpreting FST. Nat Rev Genet. 10:639–650.
Lambert DM, Ritchie PA, Millar CD, et al (2002) Rates of evolution in ancient DNA from Adelie penguins. Science (80) 295:2270–2273.
Lee JY, Edwards SV (2008) Divergence across Australia's Carpentarian barrier: statistical phylogeography of the red-backed fairy wren (Malurus melanocephalus). Evolution. 62:3117–34.
Lovette I (2004) Mitochondrial dating and mixed support for the "2% rule" in birds. Auk. 121:1–6.
Maddison WP, Knowles LL (2006) Inferring phylogeny despite incomplete lineage sorting. Syst Biol 55:21–30.
Marko PB, Hart MW (2011) The complex analytical landscape of gene flow inference. Trends Ecol Evol. 26:448–456.
Martin AP, Palumbi SR (1993) Body size, metabolic rate, generation time, and the molecular clock. Proc Natl Acad Sci USA. 90:4087–4091.
Nunn G, Stanley S (1998) Corrigendum for: body size effects and rates of cytochrome-b evolution in tube-nosed Seabirds. Mol Biol Evol. 17:1774.
Omland KE, Baker JM, Peters JL (2006) Genetic signatures of intermediate divergence: Population history of Old and New World Holarctic ravens (Corvus corax). Mol Ecol 15:795–808.
Pante E, Puillandre N, Viricel A, et al (2015) Species are hypotheses: Avoid connectivity assessments based on pillars of sand. Mol Ecol 24:525–544.
Patterson SA, Morris-Pocock JA, Friesen VL (2011) A multilocus phylogeny of the Sulidae (Aves: Pelecaniformes). Mol Phylogenet Evol 58:181–91.
Patton J, Smith M (1994) Paraphyly, Polyphyly and the Nature of Species Boundaries in pocket gophers (Genus Thomomys). Syst Biol. 43:11–26.
Paxinos EE, James HF, Olson SL, et al (2002) mtDNA from fossils reveals a radiation of Hawaiian geese recently derived from the Canada goose (Brantacanadensis). Proc Natl Acad Sci USA. 99:1399–1404.
Peck DR, Bancroft WJ, Congdon BC (2008) Morphological and molecular variation within an ocean basin in wedge-tailed shearwaters (Puffinus pacificus). Mar Biol 153:1113–1125.
Pereira S, Baker A (2006) A mitogenomic timescale for birds detects variable phylogenetic rates of molecular evolution and refutes the standard molecular clock. Mol Biol Evol 23:1731–1740.
Peterson AT (2006) Application of molecular clocks in ornithology revisited. J Avian Biol 37:541–544.
Petit RJ, Excoffier L (2009) Gene flow and species delimitation. Trends Ecol Evol 24:386–93.
Priddel D, Carlile N (1995) Mortality of Adult Gould ' s Petrels Pterodroma leucoptera leucoptera at the Nesting Site on Cabbage Tree Island , New South Wales. 95:259–264.
Priddel D, Carlile N, Portelli D, et al (2014) Pelagic distribution of Gould ' s Petrel (Pterodroma leucoptera): linking shipboard and onshore observations with remote-tracking data. 360–370.
Quinzin MC, Mayer F, Elvinger N, et al (2015) Theoretical expectations of the Isolation-Migration model of population evolution for inferring demographic

parameters Theoretical expectations of the Isolation- Migration model of population evolution for inferring demographic parameters. 610–620.
Shapiro B, Ho S (2011) Skyline-plot methods for estimating demographic history from nucleotide sequences. Mol Ecol Resour 423–434.
Shields GF, Wilson AC (1987) Calibration of mitochondrial DNA evolution in geese. J Mol Evol. 24:212–217.
Sousa V, Hey J (2013) Understanding the origin of species with genome-scale data : modelling gene flow. Nat Publ Gr. 14:404–414.
Steeves TE, Anderson DJ, Friesen VL (2005) The Isthmus of Panama: a major physical barrier to gene flow in a highly mobile pantropical seabird. J Evol Biol 18:1000–8.
Stölting KN, Nipper R, Lindtke D, et al (2013) Genomic scan for single nucleotide polymorphisms reveals patterns of divergence and gene flow between ecologically divergent species. Mol Ecol 22:842–855.
Strasburg JL, Rieseberg LH (2011) Interpreting the estimated timing of migration events between hybridizing species. Mol Ecol 20:2353–66.
Wayne RK, Morin PA (2004) Conservation genetics in the new molecular age. Front Ecol Environ 2:89–97.
Weir JT, Schluter D (2008) Calibrating the avian molecular clock. Mol Ecol 17:2321–8.
Weisrock DW, Rasoloarison RM, Fiorentino I, et al (2010) Delimiting species without nuclear monophyly in Madagascar's mouse lemurs. PLoS One.
Welch AJ, Yoshida A, Fleischer RC (2011) Mitochondrial and nuclear DNA sequences reveal recent divergence in morphologically indistinguishable petrels. Mol Ecol 20:1364–77.
Won Y, Hey J (2005) Divergence Population Genetics of Chimpanzees.
Zink RM, Barrowclough GF (2008) Mitochondrial DNA under siege in avian phylogeography. Mol Ecol 17:2107–21.

CHAPTER VI

DISCUSSION and PERSPECTIVES

"We cannot win this battle to save species and environments
without forging an emotional bond

between ourselves and nature as well....

We really must make room for nature in our hearts"

~ Stephen Jay Gould

GENERAL DISCUSSION

Unraveling the mechanisms and processes that shape population differentiation and assessing genetic subdivisions within species is one of the major goals of biogeographical history. In this thesis, I presented the genetic structuration and a first attempt to understand the demographic history of a contentious morpho-species complex, *Pterodroma leucoptera (sensu largo).* Based on my results and their possible limitations, I discuss below the following points: (i) *Genetic differentiation within Pterodroma leucoptera: Current gene flow vs incomplete lineage sorting. (ii) Going backwards to the basics: The importance of a good Calibration mutation rates* (ii) Pleistocene oscillations shaping the fate of seabird genetic structuration (iv) Future directions.

(i) Genetic differentiation within Pterodroma leucoptera: Current gene flow vs incomplete lineage sorting.

Phylogeographic patterns of intraspecific variation may provide insights into the population-level processes responsible for speciation and yield useful information for conservation purposes. However, most studies of phylogeography, speciation, and species limits restrict their focus to a narrow issue: gene tree monophyly. Reciprocal monophyly does not provide an ideal touchstone criterion for all aspects of evolutionary divergence (Omland et al. 2006). There is a continuum of divergence stages as isolated populations go from initial allele frequency differences to well-differentiated species (Patton and Smith 1994; Crisp and Chandler 1996). The challenges associated with inferring evolutionary relationships of recently diverged species or populations differ significantly from those for deep phylogenetic divergence. At shallow time depths, hazards such as incomplete lineage sorting and continuous gene flow predominate (Omland et al. 2006). Although phylogenetic patterns generated by incomplete lineage sorting have been discussed for many years (e.g., Throckmorton, 1965; Farris, 1978; Felsenstein, 1979; Arnold, 1981; Patterson et al 2011, Welch et al 2011; Roy et al 2015, Meyer et al 2016), there remains to develop and assess methods that consider these issues during phylogenetic reconstruction (Maddison and Knowles 2006). Studying intermediate stages of divergence will increase our understanding of geographical and ecological speciation, species limits, and conservation priorities. One of the principal objectives of this thesis was to assess species boundaries within *P. leucoptera* complex to resolve its controversial systematic history and to discern the factors that shaped the ancestral and current differentiation of its populations. In order to undertake this goal, we used both indirect and direct estimates of gene flow. Gene flow estimation has been proved essential for characterizing levels of speciation, connectivity among threatened populations and to uncover processes of local adaptation (Petit and Excoffier 2009; Welch et al. 2011; Pante et al. 2015). Indirect methods such Wright's F_{st} has been used to calculate gene flow under evolutionary equilibrium assumptions (e.g. equilibrium between the counteracting forces of mutation, genetic drift, and gene exchange) well known as the island model (Wright 1969). However, these methods do not consider the genealogies (i.e. history) of alleles and cannot distinguish between allele sharing owing to recurrent gene flow from allele sharing caused by ancestral polymorphism (Marko and Hart 2011). Therefore, we used coalescent-based population genetic methods (Beerli and Felsenstein 1999; Won and Hey 2005; Hey 2007a; Hey 2009; Beerli and Palczewski 2010; Hey 2010). These

methods use the retrospective concept of the coalescent (i.e. the temporal distribution of coalescent events within gene genealogies) to model demographic parameters (gene flow, divergence time or changes in effective population size) that operated in the past to shape observed patterns of genetic variation within and between present day-populations) (Hey 2007b; Hey 2009; Holsinger and Weir 2009; Sousa and Hey 2013). Wright's F_{st} and global Φ_{ST} from mtDNA and nucDNA revealed low but significant values of population differentiation between Australia and New Caledonia lineages (Chapter IV). Indeed > 90% of genetic variation was detected at intra population level for all loci. When analyzing, Australia-New Caledonia (*P. l. leucoptera*) vs South West Pacific island populations (*P. l. brevipes*) we found significant genetic differentiation (F_{st}= 0.95 corresponding to 0.018 average sequence divergence) (Chapter III). Unfortunately, statistical power (scarcely number of individual per population) precluded the analysis of these parameters among South West Pacific island populations (*P. l. brevipes*). On the other side, coalescent-based Isolation with migration under changing populatuion size model (IM) analyses were unable to give reliable estimates of migration rates for these populations, but New Caledonia to Australia with less than 1% suggesting that there might have been and still exist, some migration in accordance with low but significant F_{st}.

Complementary gene and species tree analyses by means of mitochondrial and nuclear markers revealed different stages of differentiation within *P. leucoptera* complex. Levels of polyphyly were found between South-west *P. l. brevipes* populations (Chapter III), and Australia vs New Caledonia populations (Chapter IV). Reciprocal mtDNA monophyly was revealed when comparing *P. l. brevipes* vs *P. l. leucoptera* populations (Chapter III). We discuss below the rationale and details of these findings.

(i.1) MtDNA versus nuclear markers

Evolutionary studies of non-model systems such as avian species have relied in the past on mitochondrial DNA (mtDNA) markers (Friesen et al. 2007). Although mtDNA has been preferred in phylogeographic and phylogenetic studies until recently (Avise 2000), it reflects only the maternal history of a population, and as a single locus, it provides low accuracy estimates of population parameters (Edwards and Beerli 2000). This has led to a newly directed focus toward the nuclear genome, initiated by the use of intron markers (e.g., Friesen et al. (1997) which, however, might not always be neutrally evolving (Castillo-Davies et al. 2002). Microsatellites have also been markers of choice in population genetics due to their high levels of variation (Wayne and Morin 2004). However, the mutational processes underlying this variation are not well understood, making comparisons with other markers in the same species or between species, difficult. Microsatellites are also of limited use in phylogeographical analyses (Zink and Barrowclough 2008). Reciprocal monophyly or even some sort or restricted gene flow at mitochondrial genes and signals of intermediate polyphyly at nuclear loci are commonly expected in recently diverged species or sister species (Peters et al 2012, Gangloff et al 2013, Welch et al 2011). Due to its intrinsic low evolutionary rate nuclear lineages are expected to depict full lineage sorting, only after a long period of mitochondrial lineages divergence (Hare et al 2002, Omland et al 2006, Degnan and Rosemberg 2009). The stochasticity of the coalescence process has been also attributed to explain genealogical differences among loci, even if population share the same demographic histories (Jennings and

Edwards 2005). As aforesaid, mitochondrial and nuclear markers revealed diverse stages of differentiation within *P. leucoptera* complex. Nuclear markers displayed polyphyly states for all population comparisons (haplotype networks, gene trees and species trees), conversely to concatenated mtDNA for which haplotype network resolved the *P. l. brevipes* - *P. l. leucoptera* clades (Chapter III). However unresolved mtDNA haplotype network and gene tree were exhibited within *P. l leucoptera* (Australia vs New Caledonia populations (Chapter IV). Several studies on seabirds have been carried under similar approach (comparing nuclear vs mitochondrial markers and multilocus inference). For instance, seventy genetic differentiation studies at different geographical scopes (e.g. among Oceans, within Oceans, within archipelagoes) conducted on seabirds from the 90s to 2015 were reviewed by Friesen (2015). From them, eleven studies (within archipelagoes) used mtDNA and nuclear introns sequences and other markers (e.g. AFLPs, μsats) (Table 1c supplementary material in Friesen 2015). Four studies evidenced signals of strong differentiation with both markers (*Pterodroma sandwichensis* (Welch et al, 2011, 2012; Wiley et al 2012); *Sula granti* (Levin et al 2012), while five evidenced differences in allele or haplotype frequencies (*Pterodroma phaeopygia* (Friesen et al 2006, Welch et al 2011); *Phalacrochorax harrisi* (Duffie et al 2009); *Fregata minor* (Dearborn et al 2003; Levin et al, 2012), and finally two studies showed no differentiation with either of these markers (*Phoebastria irrorata* (Huyvaert and Parker 2006); *Spheniscus mendiculus* (Nims et al 2008)). Differences in signals of differentiation in almost all the cases were attributed to the coalescent nature of the markers (e.g. mtDNA evolves monophyly 4X faster than the average nuclear allele (Palumbi et al. 2001; also see Hudson and Turelli (2003) rather than sex-biased dispersal or other factors such as phylopatry, ecological adaptation, foraging distribution, etc. Differences of lineage sorting between mitochondrial and nuclear loci have been documented by several studies on seabirds when using multilocus approach (Lee and Edwards 2008; Friesen et al 2006; Welch et al 2012). Comparatively to our resuls, diferent levels of genetic differentiation found in *P. leucoptera* complex, frigatebirds populations (*Fregata manificent*) were studied using matrilineal, mtDNA markers, and biparental inherited markers (μsatellites and nuclear introns). Analyses showed different levels of genetic differentiation depending on the markers and the geographical distinctions. While mtDNA and μsatellites exhibited resolved gene trees when comparing Atlantic and Eastern Pacific (non Galapagos) from Galapagos populations (Global $\Phi_{ST=}$0.90). These same markers showed little structuration within non Galapagos populations ($\Phi_{ST=}$20). Nuclear introns exhibited more diverse levels of differentiation. From three nuclear introns only one haplotype network resolved the Galapagos vs non Galapagos clade. Long term genetic differentiation between the two mtDNA based lineages happened ~22 000 ybp well before the Last Glacial Maximum. Different factors seems to explain this long term split such as complicated behavioral courtship, geographical range isolation and ecological adptative scenarios promoted by ciclically climate changes oscillations marked by fluctuations of trade wind patterns, water nutrient levels, sea level, sea surface temperature and circulation patterns, implying vast changes to marine habitats. Other recent studies include Silva et al (2015), who analyzed six subspecies of White-faced Storm petrel *Pelagodroma marina* across its distribution, or Gangloff et al. (2013) for *Pterodroma,* revealed similar patterns of differentiation (polyphyly levels exhibited by nuclear markers and mtDNA reciprocal monophyly). Regarding direct inferences of gene flow conducted through mtDNA IM analysis showed

migration rates (from New Caledonia to Australia, within *P. l leucoptera* populations) less than 1% suggesting that there might have been, and still exist some migration (Chapter IV). IM analysis was unable to provide reliable estimates of migration from Australia to Caledonia. In addition, the theoretical splitting event between these populations was estimated at c.70 000 years bp by IM analysis, compared to 30 000 years pb estimates by *BEAST. Additionally, the effective population size (Caledonia ~52 384 $N_{e;}$ Australia ~11 814N_e) and assuming their long generation time (20 years) yielded a most probable scenario to explain their non genetic structuration by current gene flow rather the retention of an ancestral population diversity. Indeed, these two populations deploy not only ecological (asynchronous breeding time and differences in habitat preferences) and morphological differences (e.g. Caledonia : mean wing= 228mm vs Australia wing=224mm) but also, no exchange has ever been documented between both populations (though long-term ringing has been done only on Cabbage Tree Island) (Priddel et al. 2014; del Hoyo et al. 2016). Additionally, both taxa experienced recent fluctuations in census population size whether by the introduction of predators or anthropogenic settlement (Priddel and Carlile 1995; Carlile et al. 2003; del Hoyo et al. 2016).

Dieckmann et al (2004) predicted patterns of populations at early stage of ecological speciation. Under this stage, it is expected that most of the polymorphic alleles between the populations would still be shared because the populations would not have experienced a significant bottleneck, and gene flow may still occur, at least with respect to genes and chromosomal regions that are not involved directly in the differential adaptation. As under ecological or adaptive speciation implies that population subdivisions arise on the basis of selection, rather than neutral drift under this incipient stage, drift did not have enough time to fix different relative frequencies of the alleles. IMa2 analyzes based on mtDNA sequences were unable to give reliable estimates of gene flow between, Australia-Caledonia (*P. l. leucoptera*) vs South West Pacific island populations. As commented above these clades showed resolved mtDNA phylogenies while incomplete lineage sorting depicted by nuclear introns gene tree. It was also signaled that Incomplete lineage sorting could be explained either by a recent divergence time (small relative to Ne) or by current migration. When migration does occur, estimating divergence time and migration rates appears to be more difficult (Marko and Hart 2011; Strasburg and Rieseberg 2011; Quinzin et al. 2015). Simulations conducted by the latter authors showed that IM models seems to depend on whether migration among the studied populations is strong. This is, species characterized by a small ability to migrate and separated by a major geographic barrier could presume the absence of migration between them. In that case, relatively good estimates for the divergence time between the two populations can be expected, even more so with recent divergence times (i.e. relative to the population effective size). On the other hand, in the case of a much more mobile organism, as our seabirds, strong migration rates are expected, which will probably preclude the possibility to get a reliable estimate of divergence time, although population size, or even migration rates can probably still be somewhat estimated if larger data sets are used. Our IM analysis was based on only one kind of marker, matrilineal inherited mtDNA. As we assume that Cytb and CO1 mtDNA share the same mutations rate they should concatenated, thus finally we used one loci to make complex demographical inference. This fact was one of the major drawbacks in the analysis since separated nuclear and combined nuclear and mitochondrial IM parameters simulations did not converge. Quinzin et al 2015, signaled the effects of

sample size , number loci have on IM parameters estimates. They found that adding sampled sequences or loci will vary their simulation cases.

In this view, split times events were estimated using coalescent-based *Beast dating the vicariant event between South west-Pacific, *P.l. brevipes* vs Australia and Caledonia , *P. l. leucoptera* at approximately 110 000 ybp (constant clock) and ~ 600 000 ybp (relaxed clock). It is known that tropical seabirds have experienced significant spatio-temporal fluctuations of the available marine nutrients (and thus of their prey), which probably influenced their foraging patterns and posiible local adaptations. This latter in confluence with phylopaptry behaviour and long term genetic difference (~110 000 ybp) and large Ne (33 460 for *P. l. brevipes* and 52 384 for Caledonica and 11814 Ne for Australia) and long generationt time ascribed to these seabirds, and the low evolutionary rate nature of nuclear genes might explain levels of polyphyly at nuclear loci while strong genetic differentiation at mtDNA between these populations. In regards to the south pacific populations within the *brevipes* taxa (Chapter III results), neither mitochondrial nor nuclear introns provided a clear distinction among populations, suggesting that reproductive isolation among them is not complete. However, phylogeny and networks revealed considerable genetic diversity within the group but representing a weak genetic structuring among them suggesting gene flow still happens for some colonies and that there is a very incipient divergence for some of them (e.g. *P. b. magnificens*).

Overall our work thus confirms that mtDNA markers are better able to detect genetic structuration at early stages of differentiation. However, the pattern of mtDNA distribution does not support species designations, either within the *brevipes* clade nor within *leucoptera*. The limited sequence data available from some colonies of the south Pacific populations, and the lack of data from nuclear markers for these populations make it difficult to draw any definitive conclusions. Overall, we suggest to conduct studies based on anonymous loci (Lee and Edwards 2008) or even next generation sequencing (Stölting et al. 2013) to recover more information about early stages of differentiation, for instance, by examining genomic distributions and autocorrelations of polymorphisms that are variable *vs*. those that are fixed among these closely related group.

(ii) Going backwards to the basics: The importance of a good Calibration mutation rates

There is an increasing concern among evolutionary biologist related to the suitability and accuracy when using evolutionary substitution rates to assess divergence dates between lineages and intraspecific divergence events (Garcia-Moreno 2004; Lovette 2004; Ho et al. 2007; Shapiro and Ho 2011). Common and potential sources of errors cover underlying rate differences between evolutionary lineages and among genetic loci. When calculating these rates several effects such as base composition, calibration points, generation time, metabolic rates and population size are usually undermined (Martin and Palumbi 1993; Pereira and Baker 2006). The sensitivity of chosen calibration time points to calculate rates may generate saturation (successive changes at single nucleotide sites) on the sequences thus leading underestimation of the sequence divergence. For instance, Pereira and Baker (2006) in order to assess mitochondrial evolutionary rate in 35 avian vertebrates used a calibration point older than 12 Myr which was discarded

since it led to the saturation of Cytb sequences and consequently to the underestimation of sequence divergence (Weir and Schluter 2008). Another concern is that rates of molecular evolution apparently differ across avian groups (e. g. (Nunn and Stanley 1998) and therefore it is best to employ calibrations derived from lineages related as closely as possible to the organism under study (Lovette 2004; Peterson 2006). However, the 2%-2.1% divergence rate per million years rule clock calibration obtained from mitochondrial RFLPs (Shields and Wilson 1987) and mtDNA sequences (Paxinos et al. 2002) has been widely applied to avian studies without critical assessment of its validity in the groups under investigation. Furthermore, potentially the most serious issue is that this value of 0.02 substitutions per site per millions years (s/s/Myr) rate is a substitution rate derived from interspecific comparisons. Additionally, it was pointed that molecular change within species is theoretically expected to be higher than those observed among species when accounting for the loss of polymorphism at the population level over long time period. An empirical example is the finding of short-term mutation rate of 95% per million years of *Pygoscelis adeliae* penguins compared to the commonly 2% rate used (Lambert et al. 2002). Acknowledging all such issues, the Weir and Schulter (2008) inter-lineage molecular rate estimated covering base composition and different methods for choosing calibration points was selected for this study (1.89% per million years +-0.35) and converted to a per lineage mutation rate. Finally following all this considerations our findings of app. 30 000 years bp (supplementary Fig. 1) should be regarded as an approximation of the time divergence between these two taxa until the issues surrounding the short and long term substitutions can be corrected by rigorous testing of model assumptions and careful selection of calibrating information (Ho et al. 2011).

(iii) Did Pleistocene oscillations shape seabird genetic structuration?

Pleistocene was characterized by several major glaciations and interglacial periods (Knowles and Richards 2005). Fluctuations in average global climate, the repeated advance and retreat of glacial ice sheets and changes in oceanographic conditions have had a major impact on many seabird taxa (Kidd and Friesen 1998; Congdon et al. 2000; De Dinechin et al. 2009, Gangloff et al. 2012; Morris Pocock et al. 2011; Silva et al 2015), although it is not clear what the effects might have been on tropically distributed species (Smith et al. 2007). With respect to caledonica and Australia populations, we found no sign of population expansion during the first Pleistocene period, despite expansions being classically documented for other marine species (Crandall et al. 2008; Kochzius et al. 2009). From 70,000 ybp, a very smooth increase in population size is, however, apparent, possibly in relation to further cooling of water temperature and a presumable increase in ocean productivity: today, the two populations feed in southern seas during breeding, and this is especially marked in caledonica (Priddel et al. 2014). A similar scenario was proposed by Smith et al. (2007), indicating that glacial cycles may have increased breeding habitat and improved foraging opportunities due to lower sea levels and increased occurrence of cold water upwellings, favouring dispersal, and the establishment of new colonies. Consequently, suggesting that interglacial stages restricted dispersal and promoted divergence across populations. We thus suggest that colder, more productive waters, moving slowly northward and therefore closer to breeding localities in New Caledonia, may have allowed this population increase. The

two lineages may have split at the Last Glacial Maximum (LGM, starting 70,000 years bp).

(iv) Future directions

During the last decade, conservation policies and recovering actions has greatly improved the main populations of *P. leucoptera* complex specifically on the west pacific colonies (cabbage tree island) (Carlile et al. 2003), leading to the establishment of protected areas and restoration programs that increased extant breeding sites of this taxa. Despite these achievements and efforts, several conservation actions have yet to be enforced. The research presented in this thesis reveals the existence of early structuration for some colonies of Southwest pacific and the confirmation of mitochondrial lineages between caledonica and leucoptera *vs* brevipes clade.

If I should continue with this project, my first priority would be to obtain DNA samples from presumably extinct populations, well to mention, Cook Island and Fiji populations and to develop a bunch of fine scale markers such as anonymous loci in order to complete the phylogeographic work lead in the present. Secondly, I would like to develop three future axes (1) the recognition of differentiation patterns at shallow levels of speciation in Petrels (already started in this thesis). From this research project, a new subject captured my attention which is to understand how ecological adaptation promotes the speciation of seabirds to enlight this point I would like to identify and compare possible patterns experiencing (2) Adaptive speciation v*ersus* those under neutral selection: Adaptive Evolution *Versus* Neutral Evolution , and finally, (3) Molecular comparative analyses at interspecific level using two procellarid genera, Are patterns of diversity related to speciation mechanisms on seabirds? A macroevolutionary approach.

(1) Patterns of early Divergence in Petrels: searching for evolutionary forces and speciation models

Uncovering patterns of divergence between populations can have huge significance for our understanding of divergence processes (Weisrock et al. 2010). A central question concerning speciation within clades relates to why, when and how often divergence occurs at earliest stages. Diverging populations, as living evolutionary laboratories, are good models to look for clues in order to disentangle questions not only related to the timing of divergence, such as when the basal split took place, but also the factors and scenarios which are leading such current divergence.

To investigate the timing of divergence and whether it occurred with or without gene flow and to get insights into the mechanisms leading the divergence process the following hypotheses will be tested: H0: The splitting among diverging Procellariidae populations reflects an ancient divergence modeled by genetic drift through bottleneck or founder effects scenarios rejecting the model of differentiation with migration and supporting an allopatric pattern of differentiation.H1: the splitting among diverging Procellariidae populations is a contemporaneous process modeled by natural selection in presence of gene flow through environmental adaptation.

Microsatellites, Nuclear and mitochondrial (RPEX, LAM, PAX, CSDE, Bfib, IRF2, TPM, RP40, D_loop and Cytb), and Next generation sequencing

would be usded to asses a demo-genetic model of the divergence process including ancestral population sizes and gene flow between populations at various stages of divergence. Furthermore, we aim to infer whether the mechanism of selection or neutral evolution is driving divergence in the Procellariidae (*P. leucoptera brevipes (increasing individuals per population)* complex, *P. heraldica complex,* and *P. assimilis* complex).

(2) Adaptive Evolution *Versus* Neutral Evolution

If levels of morphological differentiation and diversity are determined by the same neutral mechanisms, concordance between degree of change at the molecular level and the morphological level is expected (Barrowclough 1983). In order to assess the spatial correlation between the geographic variation in morphology and the genetic traits, we would like test the following hypotheses: H0: The spatial co-variation between the two sets of traits is driven by neutral evolution through genetic drift. H1: The congruence between spatial patterns of traits is driven by a mixture of natural selection and genetic drift. Previous studies in petrels have found that non-random mechanism or a combination of selection with the former process underpin variation in petrels. Peck et al. (2008) studied the phenotypic and genetic variation of four colonies of the wedge-tailed shearwater *Puffinus pacificus*, populations using morphological, microsatellite and nuclear introns data. They found that genetic divergence among populations were relatively low and did not correlate with morphological distances, geographic distance, or environmental differences suggesting that that non-random processes such as selection or plasticity boost morphological diversity seen in this and possibly other seabird species. Similarly, (Gómez-Díaz et al. 2006; Gómez-Díaz and González-Solís 2007) combined molecular (mitochondrial DNA), morphological and ecological data and showed that geographic adaptation promotes the patterns of genetic variability between the mediterranean *Calonectris diomedea diomedea* and the Atlantic*, C. d. borealis* Cory's shearwater.

If I should continue this work I would like to use not only nuclear introns (RPEX, LAM, PAX, CSDE, Bfib, IRF2, TPM, RP40) and mitochondrial (CO1 and Cytb) genes, but also to would like to identify and analyze markers under selection through genome sequencing in order to compare genetic variation to morphological traits distances (accounting for geographical distance) in order to assess the process creating diversity in populations of these three species, *Puffinus puffinus*, *P. assimilis* and *P. lherminieri*. These three species differ in the apparent level of morphological differentiation (according to the number of recognized subspecies. The Bank of samples of the CEBC, CNRS count with biometric data on about 1500 specimens from various museums that could be used to undertake this objective.

(3) Molecular comparative analyses at interspecific level using two procellarid genera, Are patterns of diversity related to speciation mechanisms on seabirds? A macroevolutionary approach.

Comparative analyses using a macro evolutionary approach will be used to investigate the evolutionary history of Procellariidae. Two genera, *Pterodroma* and *Puffinus* will be used to get insights about patterns of speciation. Observed differences such as i) geographical distribution, i.e. restricted versus worldwide ocean

ranges ii) geographic variation in morphology (number of subspecies), iii) location of breeding habitats (low (Puffinus) versus high altitude breeding habitats(Pterodroma), as found in these two genera suggest they have followed different evolutionary routes of speciation.

Previous studies using closely related species have investigated the barriers to gene flow leading to formation of new species (Steeves et al. 2005; Patterson et al. 2011). It has been proposed that isolating barriers influencing divergence process are not the same at high and low latitudes (reviewed in Steeves *et al*., 2005). Mitochondrial control region data were used in order to elucidate the process of isolation in the widely distributed *Sula dactylatra*. Results suggested that the predominant pattern of differentiation between Indo-Pacific and Atlantic populations was shaped by physical barriers to gene flow. However, populations divergence at the intra-ocean population level seemed to follow non-physical barriers to gene flow such as limited natal dispersal combined with local adaptation and/or genetic drift (Steeves *et al.,* 2005).

In this context, in order to understand macro evolutionary processes, two genera within the *Procellariidae* family showing radiation (high number of species, about 30 each) and having different observed spatial distributions or altitude and latitude preferences will be used to test the following 2 sets of hypotheses: H0: Evolutionary patterns of differentiation of *Pterodroma* and *Puffinus* are similar regardless of their different ecological patterns. H1: Evolutionary trajectories of *Pterodroma* and *Puffinus* are different, in accordance with their different ecological patterns. H0: The pattern of isolating process is influenced by the latitude of *Pterodroma* and *Puffinus* populations. H1: There is no relation between the latitude and the barriers to gene flow within *Pterodroma* and *Puffinus* genera*.*

These hypotheses will be tested using several types of data. While the contemporaneous DNA can reveal the end result of their ancestors' ancient trajectories, to really understand the dynamics of how modern genetic patterns were created it is also needed to study ancient material. Therefore, phylogenies, including fossils to date phylogenies, will be used as frameworks for understanding the order and numbers of coalescent events and for the possibility of dating such events through the use of molecular clocks (data).

REFERENCES

Avise JC (2000) Phylogeography: the history and formation of species. Harvard university press.

Barrowclough GF (1983) Biochemical studies of microevolutionary processes.

Beerli P, Felsenstein J (1999) Maximum-likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. Genetics 152:763–773. doi: 10.1073/pnas.081068098

Beerli P, Palczewski M (2010) Unified framework to evaluate panmixia and migration direction among multiple sampling locations. Genetics 185:313–326. doi: 10.1534/genetics.109.112532

Carlile N, Priddel D, Zino F, et al (2003) A review of four successful recovery programmes for threatened sub-tropical petrels. Mar Ornithol 31:185–192.

Castillo-Davies C, Mekhedov S, Hartl D, et al (2002) Selection for short introns in highly expressed genes. Nat Genet 31:415–418.

Crisp M, Chandler G (1996) Paraphyletic species. Telopea 6:813–844.

del Hoyo J, Elliott A, Sargatal J, et al (2016) Handbook of the Birds of the World

Alive. In: Lynx Edicions, Barcelona. (retrieved from http://www.hbw.com/ on 22/12/2015.
Edwards S, Beerli P (2000) Perspective: gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. Evolution, 54:1839–1854.
Friesen VL, Congdon BC, Walsh HE, Birt TP (1997) Intron variation in marbled murrelets detected using analyses of single-stranded conformational polymorphisms. Mol Ecol 6:1047–58.
Friesen VL, Smith a L, Gómez-Díaz E, et al (2007) Sympatric speciation by allochrony in a seabird. Proc Natl Acad Sci U S A 104:18589–94. doi: 10.1073/pnas.0700446104
Garcia-Moreno J (2004) Is there a universal molecular clock for birds? J Avian Biol 35:465–468.
Gómez-Díaz E, González-Solís J (2007) Geographic assignment of seabirds to their origin: Combining morphologic, genetic, and biogeochemical analyses. Ecol Appl 17:1484–1498. doi: 10.1890/06-1232.1
Gómez-Díaz E, González-Solís J, Peinado M a, Page RDM (2006) Phylogeography of the Calonectris shearwaters using molecular and morphometric data. Mol Phylogenet Evol 41:322–32. doi: 10.1016/j.ympev.2006.05.006
Hey BJ (2007a) Using the IMa Program. 2785–2790.
Hey J (2009) Using the IM Program. Softw. Man. 27.
Hey J (2010) Isolation with migration models for more than two populations. Mol Biol Evol 27:905–20. doi: 10.1093/molbev/msp296
Hey J (2007b) Introduction to the IM and IMa computer programs Understanding Markov chain Monte Carlo (MCMC). How Long to Run the Program Metropolis Coupling of Markov chains Assessing the Autocorrelation of Recorded Values Parameter Conversions Cautions , Suggestio. 1–16.
Ho SYW, Heupink TH, Rambaut A, Shapiro B (2007) Bayesian Estimation of Sequence Damage in Ancient DNA. Mol Biol Evol 24:1416–1422. doi: 10.1093/molbev/msm062
Ho SYW, Lanfear R, Bromham L, et al (2011) Time-dependent rates of molecular evolution. Mol Ecol 20:3087–3101. doi: 10.1111/j.1365-294X.2011.05178.x
Holsinger K, Weir B (2009) Genetics in geographically structured populations: defining, estimating and interpreting FST. Nat Rev Genet 10:639–650.
Lambert DM, Ritchie PA, Millar CD, et al (2002) Rates of evolution in ancient DNA from Adelie penguins. Science (80-) 295:2270–2273. doi: 10.1126/science.1068105
Lee JY, Edwards S V (2008) Divergence across Australia's Carpentarian barrier: statistical phylogeography of the red-backed fairy wren (Malurus melanocephalus). Evolution 62:3117–34. doi: 10.1111/j.1558-5646.2008.00543.x
Lovette I (2004) Mitochondrial dating and mixed support for the "2% rule" in birds. Auk 121:1–6. doi: Doi 10.1642/0004-8038(2004)121[0001:Mdamsf]2.0.Co;2
Maddison WP, Knowles LL (2006) Inferring phylogeny despite incomplete lineage sorting. Syst Biol 55:21–30. doi: 10.1080/10635150500354928
Marko PB, Hart MW (2011) The complex analytical landscape of gene flow inference. Trends Ecol Evol 26:448–456. doi: 10.1016/j.tree.2011.05.007
Martin AP, Palumbi SR (1993) Body size, metabolic rate, generation time, and the molecular clock. Proc Natl Acad Sci U S A 90:4087–4091. doi:

10.1073/pnas.90.9.4087
Nunn G, Stanley S (1998) Corrigendum for: body size effects and rates of cytochrome-b evolution in tube-nosed Seabirds. Mol Biol Evol 17:1774.
Omland KE, Baker JM, Peters JL (2006) Genetic signatures of intermediate divergence: Population history of Old and New World Holarctic ravens (Corvus corax). Mol Ecol 15:795–808. doi: 10.1111/j.1365-294X.2005.02827.x
Pante E, Puillandre N, Viricel A, et al (2015) Species are hypotheses: Avoid connectivity assessments based on pillars of sand. Mol Ecol 24:525–544. doi: 10.1111/mec.13048
Patterson S a, Morris-Pocock J a, Friesen VL (2011) A multilocus phylogeny of the Sulidae (Aves: Pelecaniformes). Mol Phylogenet Evol 58:181–91. doi: 10.1016/j.ympev.2010.11.021
Patton J, Smith M (1994) Paraphyly, Polyphyly and the Nature of Species Boundaries in pocket gophers (Genus Thomomys). Syst Biol 43:11–26.
Paxinos EE, James HF, Olson SL, et al (2002) mtDNA from fossils reveals a radiation of Hawaiian geese recently derived from the Canada goose (Brantacanadensis). Proc Natl Acad Sci U S A 99:1399–1404. doi: 10.1073/pnas.032166399
Peck DR, Bancroft WJ, Congdon BC (2008) Morphological and molecular variation within an ocean basin in wedge-tailed shearwaters (Puffinus pacificus). Mar Biol 153:1113–1125. doi: 10.1007/s00227-007-0883-x
Pereira S, Baker A (2006) A mitogenomic timescale for birds detects variable phylogenetic rates of molecular evolution and refutes the standard molecular clock. Mol Biol Evol 23:1731–1740. doi: 10.1093/molbev/msl038
Peterson AT (2006) Application of molecular clocks in ornithology revisited. J Avian Biol 37:541–544. doi: 10.1111/j.0908-8857.2006.04029.x
Petit RJ, Excoffier L (2009) Gene flow and species delimitation. Trends Ecol Evol 24:386–93. doi: 10.1016/j.tree.2009.02.011
Priddel D, Carlile N (1995) Mortality of Adult Gould ' s Petrels Pterodroma leucoptera leucoptera at the Nesting Site on Cabbage Tree Island , New South Wales. 95:259–264.
Priddel D, Carlile N, Portelli D, et al (2014) Pelagic distribution of Gould ' s Petrel (Pterodroma leucoptera): linking shipboard and onshore observations with remote-tracking data. 360–370.
Quinzin MC, Mayer F, Elvinger N, et al (2015) Theoretical expectations of the Isolation-Migration model of population evolution for inferring demographic parameters Theoretical expectations of the Isolation- Migration model of population evolution for inferring demographic parameters. 610–620. doi: 10.1111/2041-210X.12347
Shapiro B, Ho S (2011) Skyline-plot methods for estimating demographic history from nucleotide sequences. Mol Ecol Resour 423–434. doi: 10.1111/j.1755-0998.2011.02988.x
Shields GF, Wilson AC (1987) Calibration of mitochondrial DNA evolution in geese. J Mol Evol 24:212–217. doi: 10.1007/BF02111234
Sousa V, Hey J (2013) Understanding the origin of species with genome-scale data : modelling gene flow. Nat Publ Gr 14:404–414. doi: 10.1038/nrg3446
Steeves TE, Anderson DJ, Friesen VL (2005) The Isthmus of Panama: a major physical barrier to gene flow in a highly mobile pantropical seabird. J Evol Biol 18:1000–8. doi: 10.1111/j.1420-9101.2005.00906.x

Stölting KN, Nipper R, Lindtke D, et al (2013) Genomic scan for single nucleotide polymorphisms reveals patterns of divergence and gene flow between ecologically divergent species. Mol Ecol 22:842–855. doi: 10.1111/mec.12011
Strasburg JL, Rieseberg LH (2011) Interpreting the estimated timing of migration events between hybridizing species. Mol Ecol 20:2353–66. doi: 10.1111/j.1365-294X.2011.05048.x
Wayne RK, Morin P a. (2004) Conservation genetics in the new molecular age. Front Ecol Environ 2:89–97. doi: 10.1890/1540-9295(2004)002[0089:CGITNM]2.0.CO;2
Weir JT, Schluter D (2008) Calibrating the avian molecular clock. Mol Ecol 17:2321–8. doi: 10.1111/j.1365-294X.2008.03742.x
Weisrock DW, Rasoloarison RM, Fiorentino I, et al (2010) Delimiting species without nuclear monophyly in Madagascar's mouse lemurs. PLoS One. doi: 10.1371/journal.pone.0009883
Welch AJ, Yoshida A a, Fleischer RC (2011) Mitochondrial and nuclear DNA sequences reveal recent divergence in morphologically indistinguishable petrels. Mol Ecol 20:1364–77. doi: 10.1111/j.1365-294X.2011.05008.x
Won Y, Hey J (2005) Divergence Population Genetics of Chimpanzees. doi: 10.1093/molbev/msi017
Zink RM, Barrowclough GF (2008) Mitochondrial DNA under siege in avian phylogeography. Mol Ecol 17:2107–21. doi: 10.1111/j.1365-294X.2008.03737.x

Appendix A. Supplementary Figures and Tables for Chapter IV

Table A. 1. Primers used for DNA amplification and sequencing

GENE	PRIMER NAME	NUCLEOTIDE SEQUENCE (5' --->3')	SOURCE
Nuclear DNA			
PAX Interacting Protein 1-PAXIP1	PAX 20F	CCCTCAGACACTGGATTAYGAATCAT	(Kimball et al. 2009)
	PAX 21R	CCAAGGATTCCGAAGCAGTAAG	
	PAX-F-CRI	GTGTGGTTTACTAGAAGTTT	*This study*
	PAX-R-CRI	GGCATTTTACTATCCATCAAG	
	PAX-Fint-CRI	GTAGAATATTAAGCAACAGGA	
Cold shock domain-containing protein E1-CSDE1	CSDE5F	CTGGTGCTGTAAGTGCTCGTAAC	(Kimball et al. 2009)
	CSDE6R	CCAGGCTGTAAGGTTTCTAGGTCAC	
Beta-Fibrinogen intron 7- BFibint7	FIB-BI7U	GGAGAAAACAGGACAATGACAATTCAC	(Prychitko and Moore 1997)
	FIB-BI7L	TCCCCAGTAGTATCTGCCATTAGGGTT	
IRF2F	IRF2 2F	ATGTCTTTGGGTCGGGTTTA	Kimbal et al 2009
	IRF2 3R	GAAACTGGGCAATTCACACA	

	IRF2-F-Pter-CRI	AGGCAAATTAATAACAGCGTAGG	*This study*
	IRF2-R-int-CRI	TGTTGGGAGTAGAGCACACT	
TPM1	Trop 6aF	AATGGCTGCAGAGGATAA	(Primmer et al. 2002)
	Trop 6bR	TCCTCTTCAAGCTCAGCACA	
Mitochondrial DNA			
Cytochrome Oxidase 1-CO1	F1B	AACCGATGACTATTYT-CAACC	(Gangloff et al. 2013)
	R1B	TACTACRTGYGARATGATTCC	
	COI-F1-Pter-CRI	CACAARGATATYGGTACCCT	*This study*
	COI-R1-Pter-CRI	CAGTTCATCCTGTACCTGCY	
	COI-F2-Pter-CRI	GACATAGCATTCCCACGTATR	
	COI-R2-Pter-CRI	GTGAGAGTAAAAGTAGGACGG	
	COI-F3-Pter-CRI	GGGCAATCAACTTCATYACAAC	
	COI-R3-Pter-CRI	GAATGTAGACTTCTGGGTGGC	
Cytochrome b - Cytb	L14987	TATTTCTGCTTGATGAAACT	(Jesus et al. 2009)
	H16025	CTAGGGCTCCAATGATGGGGA	

Cyt.Pter.F22	AGCCATGCACTACACACAGCCG	
Cytb.Pter.R221	CGAAGGCAGTTGCTATGAGG	
Cytb.Pter.F177	ATTCTAYTRTGGCTCCTACC	
Cytb.Pter.R404	RCAAAAGGTAGGAGGAAGT	
Cytb.Pter.F374	GGGATTCTCAGTAGATAACC	
Cytb.Pter.R634	GTAAAGTTTTCTGGGTCTCC	
CytB-F1-Pter-CRI	CTACTAGCCATRCACTACAC	*This study*
CytB-R1-Pter-CRI	RAATGATATTTGTCCTCAGGG	
CytB-F2-Pter-CRI	GGCTCCTACCTRTACAAAGAG	
CytB-R2-Pter-CRI	GTAGGGGTGRAATGGGATT	
CytB-F3-Pter-CRI	ATYCACCTCACCTTCCTCCA	
CytB-R3-Pter-CRI	GGAATTGAGCGTAGGATAKCG	
CytB-F4-Pter-CRI	GGAGACCCAGAAAACTTTACY	
CytB-R4-Pter-CRI	TCAGAAGAGGAGTTGGGAGA	

Table A. 2. Number of individuals per population and loci used for the summary statistics inference, without accounting *a priori* population definition. The number in brackets: sequence length per locus.

*mtDNA*CO1+Cytb and Nuclear Introns loci						
Population	**CO1+Cytb (1310)**	**PAXIPI (424)**	**CSDE (481)**	**βfibin7 (911)**	**IRF2 (635)**	**TPM (463)**
Eromango - Tanna	15	16	15	14	13	10
Vanua Lava I	5	6	6	6	5	5
Vanua Lava II (Bank Isl.)	2	2	-	1	3	-
Society	5	5	5	5	3	2
Cook Island	1	-	-	-	-	-
Fiji_Viti Levu	1	2	-	1	1	-
Fiji_Viti Kadavu	-	1	-	-	1	-
Australes	1	1	-	1	1	-
New Caledonia	17	26	16	15	26	15
Australia	20	56	50	35	45	17
TOTAL (N)	**67**	**115**	**92**	**78**	**98**	**49**

Table A.3. Table of total individuals per population and loci used in Chapter IV. Numbers in brackets represent the total number of individuals per locus

Population
Eromango
Tanna
Vanua Lava I
Vanua Lava II
Society
Cook Isl
Fiji
Australes
New Calenia
Australia

CO1+Cytb (1310bp) (67)	PAX (424bp) (115)	CSDE (481 bp) (92)	Bfib (911bp) (78)	IRF2 (635bp) (98)	TPM (463bp) (49)
BR13	BR12	BR12	BR12	BR12	BR12
BR14	BR13	BR13	BR13	BR13	BR12
BR15	BR14	BR14	BR14	BR14	BR13
BR16	BR15	BR15	BR15	BR15	BR13
BR18	BR16	BR16	BR18	BR16	BR14
BR19	BR17	BR17	BR19	BR19	BR14
BR20	BR18	BR18	BR20	BR21	BR15
BR22	BR19	BR19	BR21	BR22	BR15
BR23	BR20	BR20	BR22	BR23	BR16
BR24	BR21	BR21	BR23	BR24	BR16

BR25	BR22	BR22	BR24	BR25	BR18
BR26	BR23	BR23	BR25	BR26	BR18
BN99	BR24	BR24	BR26	BN99-BO03	BR19
BO04	BR25	BR25	BN99	BR06	BR19
BO03	BR26	BR26	BR06	BR08	BR23
BR06	BO04	BR06	BR07	BR09	BR23
BR07	BR06	BR07	BR08	BR10b	BR22
BR08	BR07	BR08	BR09	BR11	BR22
BR09	BR08	BR09	BR10	BN92	BR24
BR10	BR09	BR10	BR11	BN94	BR24
BN94	BR10	BR11	BN95	BN95	BR06
BN95	BR11	BR32	BO12	BO13	BR06
BO12	BN94	BR33	BO13	BW44	BR07
BO13	BN95	BR34	BR32	BO12	BR07
BR33	BO12	BO12	BR33	BW69	BR08
BR34	BO13	BO13	BR34	BN97	BR08
BW44	BR32	BE06	BW69	BO18	BR09
BN90	BR33	BE07	BO18	BE06	BR09
BW69	BR34	BE13	BE06	BE08	BR11
BO18	BO05	BE14	BE09	BE09	BR11
BE09	BW69	BE16	BE12	BE10	BR32
BE12	BN97	BE20	BE13	BE12	BR32
BE14	BO18	BE22	BE14	BE13	BR33
BE15	BE06	BE23	BE15	BE14	BR33
BE17	BE09	BE27	BE17	BE15	BE08
BE20	BE10	BE28	BE21	BE16	BE08
BE21	BE12	BE29	BE26	BE17	BE09
BE23	BE13	BE31	BE23	BE18	BE09
BE24	BE14	BE32	BE27	BE19	BE11
BE27	BE15	BE34	BE28	BE20	BE11

BE28	BE17	BE35	BE29	BE21	BE12
BE29	BE20	BE36	BE31	BE23	BE12
BE30	BE21	BH57	BE35	BE24	BE13
BE32	BE23	BH58	BH60	BE26	BE13
BE33	BE24	BH59	BH61	BE27	BE15
BE34	BE26	BH60	BH62	BE28	BE15
BE36	BE27	BH61	BH63	BE29	BE22
BH65	BE28	BH62	BH64	BE30	BE22
BH59	BE29	BH63	BH65	BE31	BE26
BH62	BE32	BH64	BH66	BE33	BE26
BH66	BE33	BH65	BH67	BE34	BE27
BH67	BE35	BH67	BH69	BE35	BE27
BH68	BE36	BH68	BH70	BE39	BE29
BH69	BE38	BH69	BH71	BH57	BE29
BH70	BE39	BH70	BH72	BH58	BE30
BH74	BT82	BH71	BH73	BH59	BE30
BH76	BT83	BH72	BH74	BH60	BE36
BH77	BT86	BH74	BH75	BH61	BE36
BH79	BT88	BH75	BH76	BH62	BT82
BH80	BH57	BH76	BH77	BH63	BT82
BH83	BH59	BH77	BH79	BH64	BT83
BH84	BH60	BH83	BH82	BH65	BT83
BH90	BH61	BH84	BH83	BH66	BT86
BH91	BH62	BH85	BH84	BH67	BT86
BH96	BH63	BH86	BH85	BH68	BH57
BH97	BH64	BH87	BH86	BH69	BH57
BI03	BH65	BH88	BH87	BH70	BH58
	BH67	BH91	BH88	BH71	BH58
	BH68	BH94	BH89	BH73	BH59
	BH69	BH95	BH91	BH74	BH59

BH70	BH96	BH93	BH75	BH60
BH71	BH97	BH94	BH76	BH60
BH72	BH99	BH95	BH77	BH63
BH73	BI01	BH96	BH79	BH63
BH75	BI02	BH97	BH81	BH74
BH76	BI04	BH98	BH82	BH74
BH77	BI05	BH99	BH83	BH75
BH80	BI06	BI01	BH84	BH75
BH81	BI08		BH86	BH76
BH82	BI09		BH87	BH76
BH83	BI10		BH88	BH77
BH84	BI11		BH94	BH77
BH85	BI12		BH95	BH87
BH86	BI13		BH96	BH87
BH87	BI15		BH98	BH94
BH88	BI16		BH99	BH94
BH90	BI17		BI01	BH95
BH91	BI18		BI03	BH95
BH93	BI20		BI04	BH96
BH94	BI21		BI05	BH96
BH95	BI22		BI06	BH97
BH96	BI23		BI07	BH97
BH98			BI09	BH98
BH99			BI10F	BH98
BI01			BI11	BH99
BI02			BI12	BH99
BI03			BI13	BI02
BI04			BI14	BI02
BI05				
BI06				

	BI07	
	BI08	
	BI09	
	BI10	
	BI11	
	BI12	
	BI13	
	BI14	
	BI16	
	BI17	
	BI18	
	BI20	
	BI21	
	BI22	
	BI23	

Table A.4. List of museum samples with their respective museum of provenance, identification number and sampling location

MUSEUM	CATALOG NUMBER	ID_{thesis}	LOCATION	MORPHO SEPECIES
AMNH				
	216919	BN92	*Vanua Lava II*	*P. b. magnificens*
	216921	BN94	*Vanua Lava II*	*P. b. magnificens*
	216923	BN95	*Vanua Lava II*	*P. b. magnificens*
	336697	BN99	Tanna	*P. brevipes*
	336707	BO03	Tanna	*P. brevipes*
	336709	BO04	Tanna	*P. brevipes*
	250893	BN97	*Kadavu*	*P. brevipes*
	528336	*BO05*	*Viti Levu*	*P. brevipes*
MNHN				
	1876113	*BW69*	*Viti Levu*	*P. brevipes*
MTI				
	2004-3-12	*BO12*	*Society*	*P. brevipes*
	2004-3-13	*BO13*	*Society*	*P. brevipes*
Te Papa				
	OR.023110	*BN90*	*Rarotonga*	*P. brevipes*
	25377	*BO18*	*Raivavae*	*P. brevipes*

AMNH: American Museum of Natural History of New York; **MNHN:** *Muséum National d'Histoire Naturelle***; MTI:** *Musée de Tahiti et des îles*

Table A.5. Per lineage evolutionary mutation rate estimate

	Interlineage mutation rate		Per lineage mutation rate mu = *d*xy/2*T* , as in (Nei 1987)	
Author source	**%s/s/My**	**s/sMy**	**s/s/My**	**s/s/y**
Cytb (Weir and Schluter 2008)	1.89	0.0189	4.87×10^{-3}	4.87×10^{-9}
Nuclear (Axelsson et al. 2004)	0.36	0.0036	6.48×10^{-3}	6.48×10^{-9}

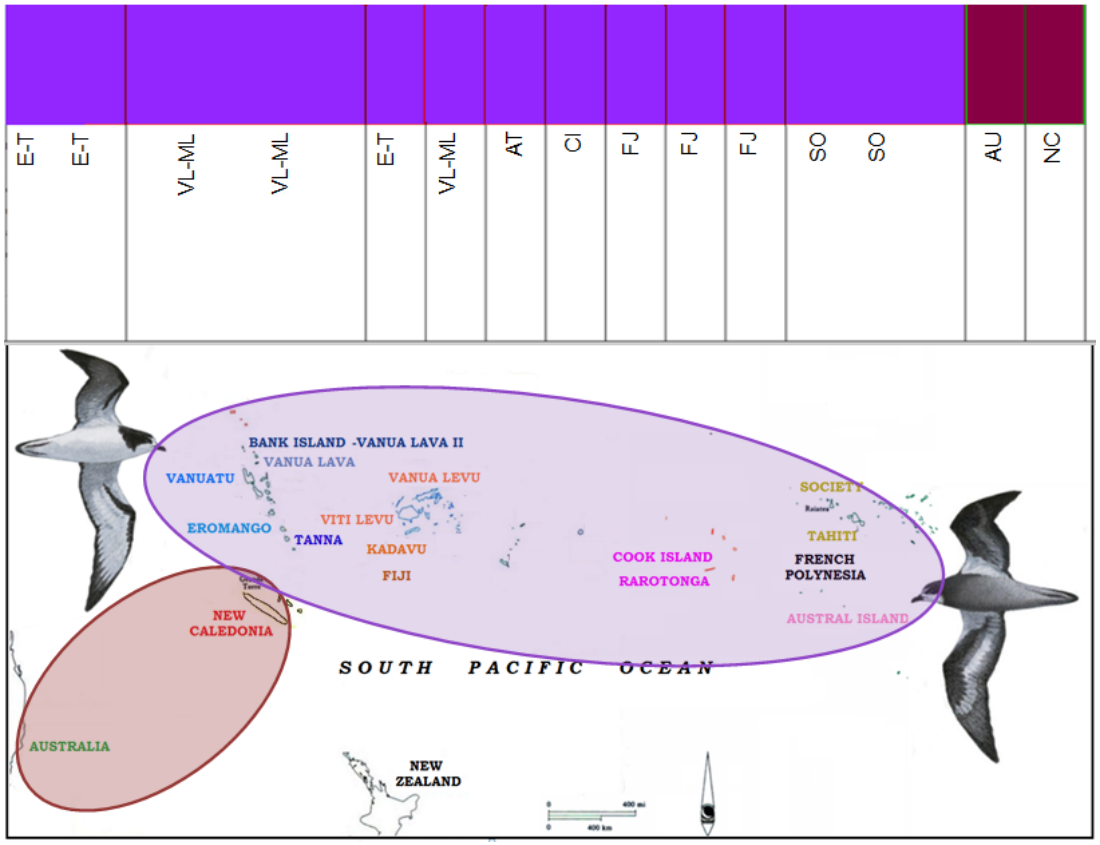


Figure. A1. Map showing the sampling point with their assignment into clusters usin multiloci BAPS using global mitochondria and nuclear introns PAX IRF2 BFIB CSDE TPM loci. BAPS sugestedv two clusters as the most likely value for ***k***. Abbreviations our hypothetical a priori populations: (**NC):** New Caledonia; **(AU)**: Australia; **(E-T)**: Eromango-Tanna; **(VL-ML)**: Vanua Lava I, II; **(SO):** Society; (**FJ):** Fiji**; (AT): Australes** (see also Figure.4.1 in text).

Table A.6. List of individuals per population and loci used in the concatenated mitochondrial DNA phylogenies for CO1 and Cytb found through Maximum Likelihood (MEGA) and Bayesian (BEAST) inferences. Total sequence length of the concatenated mtDNA loci indicated in brackets : **(bp).**

***mtDNA*CO1+Cytb (1165bp)**	
Population	**Individuals (ID)**
Eromango - Tanna	
	BR13
	BR24
	BO04
Vanua Lava I	
	BR06
Vanua Lava II (Bank Isl.)	
	BN94
Society	
	BO12
	BO13
Cook Island	
	BN90
Fiji_Viti Levu	
	BW69
Australes	
	BO18
New Caledonia	
	BE30
	BE09

Australia	
	BH59
	BH65
Outgroups	
Pterodroma madeira	BJ29
	BJ30
Puffinus bailloni	BW49
Puffinus lherminieri	BV01
TOTAL (N)	**18**

Table A.7. List of individuals per population and loci used in the concatenated nuclear introns loci phylogenies for PAXIPI, IRF2 and βfibin7 found through Maximum Likelihood (MEGA) and Bayesian (BEAST) inferences. Total sequence length of the concatenated nuclear introns loci indicated in brackets: **(bp).** Total number of individuals per loci: **TOTAL (N).**

Nuclear Introns loci			
Population	**PAXIPI (394)**	**IRF2 (567)**	**βfibin7 (911)**
Eromango - Tanna			
	BR13	BR13	BR13
	BR12	BR12	BR12
	BR14	BR14	BR14
	BR23	BR23	BR23
Vanua Lava I			
	BR06	BR06	BR06
	BR08	BR08	BR08
	BR11	BR11	BR11
Vanua Lava II (Bank Isl.)			
	BN95	BN95	BN95
Society			
	BO12	BO12	BO12
	BO13	BO13	BO13
Fiji_Viti Levu			
	BW69	BW69	BW69
Australes			
	BO18	BO18	BO18
New Caledonia			
	BE14	BE14	BE14

	BE35	BE35	BE35
	BE27	BE27	BE27
Australia			
	BH57	BH60	BH60
	BH71	BH71	BH71
	BI01	BI01	BI01
Outgroups			
Pterodroma madeira	BJ29	BJ29	BJ29
	BJ30	BJ30	BJ30
Puffinus bailloni	BW49	BW49	BW49
Puffinus lherminieri	BV01	BV01	BV01
TOTAL (N)		**22**	

Table A.8. List of individuals per population and loci used in the *BEAST species tree for two concatenated mitochondrial DNA: CO1 and Cytb, and five nuclear intron loci: PAXIPI, IRF2, βfibin7**,** CSDE1, TPM1. Total sequence length of the concatenated nuclear introns loci indicated in brackets: **(bp)**. Total number of individuals per loci: **TOTAL (N).**

***mtDNA*CO1+Cytb and Nuclear Introns loci**				
Population	**CO1+Cytb (1165)**	**PAXIPI (394)**	**IRF2 (567)**	**βfibin7 (911)**
Eromango - Tanna				
	BR13	BR13	BR13	BR13
	BR24	BR20	BR20	BR21
Vanua Lava I, II				
	BR06	BR06	BR06	BR06
Vanua Lava II (Bank Isl.)				
	BN95	BN95	BN95	BN95
Society				
	BO12	BO12	BO12	BO12
	BO13	BO13	BO13	BO13
Fiji_Viti Levu				
	BW69	BW69	BW69	BW69
Australes				
	BO18	BO18	BO18	BO18
New Caledonia				
	BE30	BE14	BE14	BE14
	BE09	BE35	BE35	BE35
Australia				
	BH59	BH57	BH60	BH60
	BH65	BH71	BH71	BH71
Outgroups				

Pterodroma madeira	BJ29 BJ30	BJ29 BJ30	BJ29 BJ30	BJ29 BJ30
Puffinus bailloni	BW49	BW49	BW49	BW49
Puffinus lherminieri	BV01	BV01	BV01	BV01
TOTAL (N)			**22**	

Table A.9. Number of individuals per population and loci used in the EBSP for *P. brevipes* lineage

	*mtDNA*CO1+Cytb and Nuclear Introns loci					
Population	**CO1+Cytb (1310)**	**PAXIPI (416)**	**IRF2 (620)**	**βfibin7 (911)**	**CSDE (481)**	**TPM (463)**
Eromango - Tanna	15	14	11	12	15	8
Vanua Lava I	5	6	5	6	6	6
Vanua Lava II (Bank Isl.)	2	1	3	-	-	-
Society	5	4	4	4	4	2
Fiji_Viti Levu	1	2	1	1	-	-
Fiji_Kadavu	-	1	1			
Australes	1	1	1	1	1	1
Cook Island	1	-	-	-	-	-
TOTAL (N)	**30**	**29**	**26**	**23**	**26**	**17**

Table A.10 PCR profiles conditions.

Réactive	Conc. initiale	Volumen	Conc. Finale	X	Vol fin
Multiplex MIX	2X	12.5	1X	6	75
Primer.1Forward	20uM	1	0,8	6	6
Primer2Reverse	20uM	1	0,8	6	6
Premix (3 FOIS)		14.5			

	Sample
EAU Ultrapure	9
ADN (ng/ul)	1.5

PRE- PCR	PCR_annealing temperature Depending on the primer	POST- PCR
94°C 15min	94°C 1min 54-57°C 50sec 72°C 50sec	72°C 5min
1cicle	40 cicles	1cicle

REFERENCES

Axelsson E, Smith NGC, Sundström H, et al (2004) Male-biased mutation rate and divergence in autosomal, z-linked and w-linked introns of chicken and Turkey. Mol Biol Evol 21:1538–47.

Gangloff B, Zino F, Shirihai H, et al (2013) The evolution of north-east Atlantic gadfly petrels using statistical phylogeography. Mol Ecol 22:495–507.

Jesus J, Menezes D, Gomes S, et al (2009) Phylogenetic relationships of gadfly petrels Pterodroma spp. from the Northeastern Atlantic Ocean: molecular evidence for specific status of Bugio and Cape Verde petrels and implications for conservation. Bird Conserv Int 19:199.

Kimball RT, Braun EL, Barker FK, et al (2009a) A well-tested set of primers to amplify regions spread across the avian genome. Mol Phylogenet Evol 50:654–60.

Kimball RT, Braun EL, Barker FK, et al (2009b) A well-tested set of primers to amplify regions spread across the avian genome. Mol Phylogenet Evol 50:654–660.

Nei M (1987) Molecular evolutionary genetics. Columbia university press

Primmer CR, Borge T, Lindell J, Sætre GP (2002) Single-nucleotide polymorphism characterization in species with limited available sequence information: High nucleotide diversity revealed in the avian genome. Mol Ecol 11:603–612.

Prychitko TM, Moore WS (1997) The utility of DNA sequences of an intron from the beta-fibrinogen gene in phylogenetic analysis of woodpeckers (Aves: Picidae). Mol Phylogenet Evol 8:193–204. doi: 10.1006/mpev.1997.0420

Weir JT, Schluter D (2008) Calibrating the avian molecular clock. Mol Ecol 17:2321–8.

Appendix B. Supplementary Figures and Tables for CHAPTER V.

Table B1. Primers used for DNA amplification and sequencing

GENE	PRIMER NAME	NUCLEOTIDE SEQUENCE (5' --->3')	SOURCE
Nuclear DNA			
PAX Interacting Protein 1-PAXIP1	PAX 20F	CCCTCAGACACTGGATTAYGAATCAT	Kimball *et al.* 2009
	PAX 21R	CCAAGGATTCCGAAGCAGTAAG	
Cold shock domain-containing protein E1-CSDE1	CSDE5F	CTGGTGCTGTAAGTGCTCGTAAC	Kimball *et al.* 2009
	CSDE6R	CCAGGCTGTAAGGTTTCTAGGTCAC	
Beta-Fibrinogen intron 7- BFibint7	FIB-BI7U	GGAGAAAACAGGACAATGACAATTCAC	
	FIB-BI7L	TCCCCAGTAGTATCTGCCATTAGGGTT	
IRF2F	IRF2 2F	ATGTCTTTGGGTCGGGTTTA	Kimbal *et al.* 2009
	IRF2 3R	GAAACTGGGCAATTCACACA	
TPM1	Trop 6aF	AATGGCTGCAGAGGATAA	Primmer *et al.* 2002
	Trop 6bR	TCCTCTTCAAGCTCAGCACA	
Mitochondrial DNA			

Cytochrome Oxidase 1-CO1	F1B	AACCGATGACTATTYTCAACC	Gangloff *et al.* 2013
	R1B	TACTACRTGYGARATGATTCC	
Cytochrome b - cytb	L14987	TATTTCTGCTTGATGAAACT	Jesus *et al.* 2009
	H16025	CTAGGGCTCCAATGATGGGGA	

Table B2. Estimates of Inter and intrapopulation variance and percentage of genetic variation (AMOVA) between *Pterodroma leucoptera leucoptera* and *P. leucoptera caledonica.* Significant interpopulation variance p values are marked in bold.

		d.f.	Sum of squares	Variance components	Percentage of variation	P values
Co1Cytb						
	Interpopulation	1	2.951	0.1091	10.34	**<0.05**
	Intrapopulation	35	33.163	0.9458	89.66	
PAXIP1						
	Interpopulation	1	0.386	0.0033	2.17	>0.05
	Intrapopulation	162	24.284	0.1499	97.83	
CSDE1						
	Interpopulation	1	1.747	0.02965	8.76	**<0.05**
	Intrapopulation	130	40.170	0.30900	91.24	
βfibint7						
	Interpopulation	1	1.046	0.00906	1.34	**<0.05**
	Intrapopulation	98	65.21	0.66541	98.66	
IRF21						
	Interpopulation	1	1.046	0.0026	0.49	**<0.05**
	Intrapopulation	69	65.21	0.5337	99.51	

Table B3. Per lineage evolutionary mutation rate estimate

	Interlineage mutation rate		Per lineage mutation rate mu = *d*xy/2*T* , as in Nei 1987	
Author source	**%s/s/My**	**s/s/My**	**s/s/My**	**s/s/y**
Cytb (Weir and Schluter 2008)	1.89	0.0189	4.87×10^{-3}	4.87×10^{-9}
Nuclear (Axelsson et al. 2004)	0.36	0.0036	6.48×10^{-3}	6.48×10^{-9}

Table B4. GeneBank accession numbers of the nucleotide sequences of *Pterodroma* species used for the phylogenetic and divergence time reconstruction by means of *Beast under a relaxed clock assumption.

	Loci				
Species	CO1	Cytb	PAXP1	CSDE1	βfibint7
P. deserta	JX674236.1	JX674458.1	JX674798.1	JX674765.1	JX674615.1
P. feae	JX674226.1	JX674446.1	JX674821.1	JX674785.1	JX674749.1

Table B5. Effective population sizes, migration frequencies and divergence times estimated by IM. Present-day effective population sizes (N_e) of the two subspecies, *P. leucoptera caledonica* and *P. l. leucoptera*, and their common ancestor. *T* indicates the divergence time between the two populations in years. *M1* indicates the probability of migration per generation from *P. l. caledonica* to *P. l. leucoptera* forwards in time, and *M2* indicates migration in the other direction. *2N1m1* indicates the effective number of gene migrants into *P. l. leucoptera* from *P. l. caledonica* and *2N2m2* in the opposite direction. This analysis is based on the mitochondrial dataset. High Point indicates the value of the highest posterior probability. 95% Low indicates the value to which 2.5% of the total distribution lies to the left. 95% High indicates the value to which 2.5% of the total distribution lies to the right. All values are averages from the marginal posterior distributions from three runs with different starting seeds.

	Ne caledonica	*Ne leucoptera*	*Ne ancestor*	*T (years)*	*Founders*	*M1*	*M2*	*2N1m1*	*2N2m2*
High Point	108 000	5 326	392	67 000	176	5.5	0.6	96.00	0.5
5% percentile	33 000	2 236	1 960	118 000	1 914	2.4	0.2	13.00	0.1
95% percentile	136 000	29 000	133 000	4 700 00	2 611	10.0	9.0	213.00	42.00

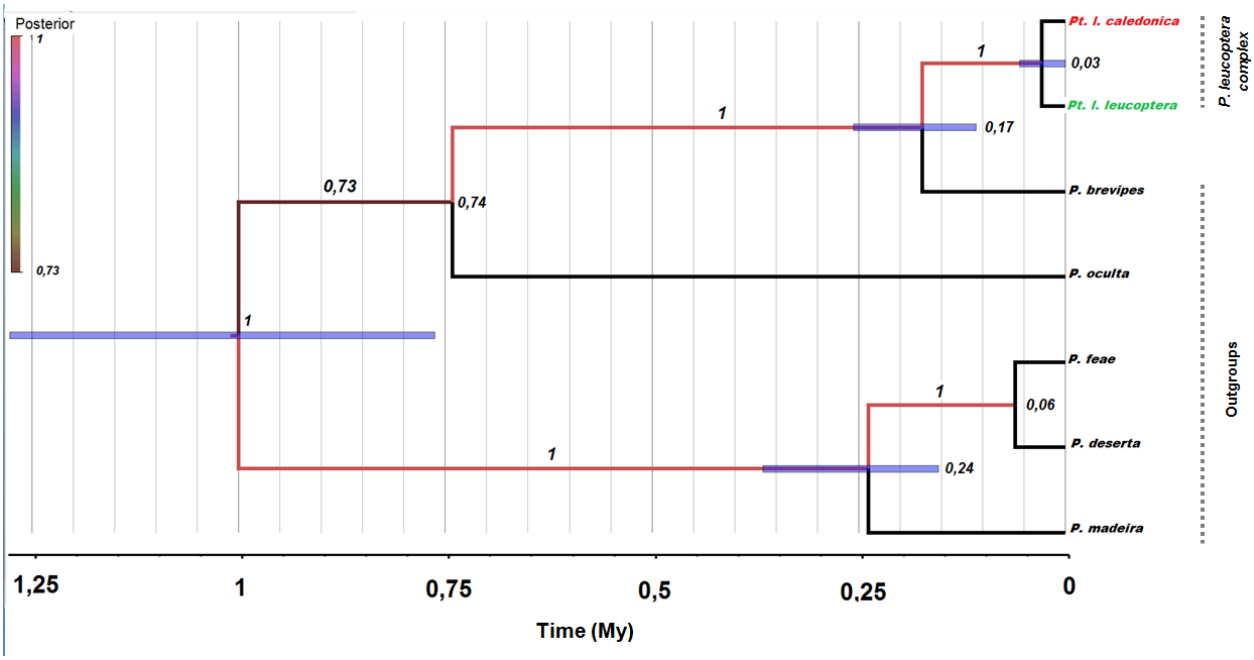


Figure B1.*Beast tree using two mitochondrial genes (concatenated CO1 and Cytb) and four nuclear intron (PAXIP1, CSDE1, βfibint7, IRF2F1) showing simultaneously the phylogeny and divergence time of *P. l. caledonica* and *P. l. leucoptera* sub-species under a relaxed clock assumption. The divergence times expressed in Million years are displayed at each node. Purple bars represent the 95%HPD interval for the divergence time estimates. Numbers on branches represent the posterior probability of each clade.

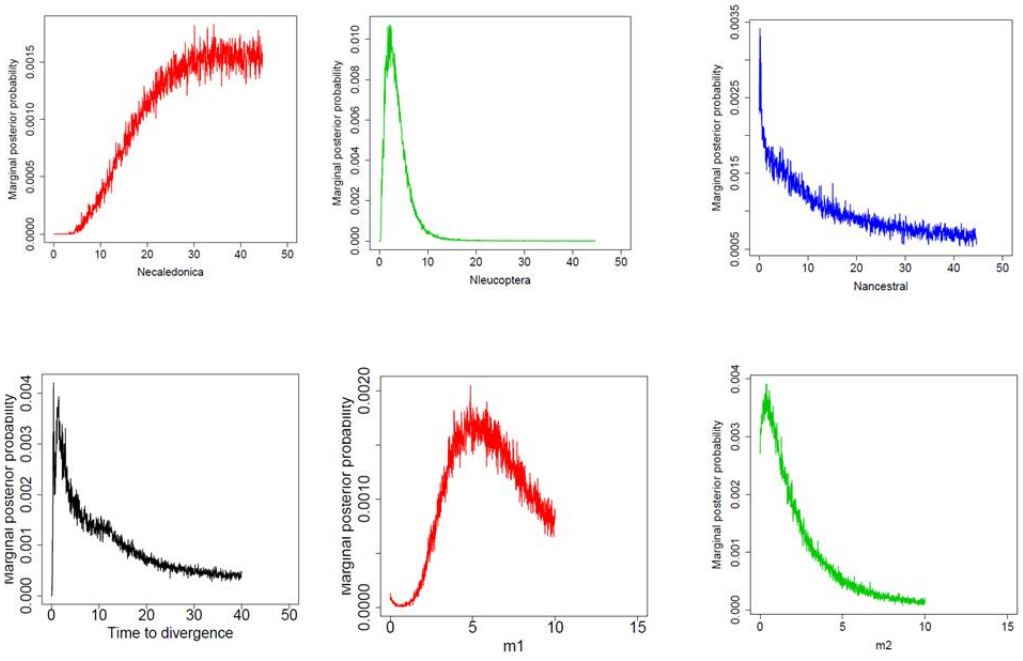


Figure B2. Marginal posterior probability distributions for six parameter estimates obtained for *P. l. caledonica* and *P. l. leucoptera* subspecies using IM (one of three replicates). Parameters were estimated using only CO1 and Cytb. Final long runs consisted of 2 million steps following a 1 million-step burn-in. Ten chains with geometric heating (g1=0.96, g2=0.90) were used in order to ensure good mixing. *m1; m2,* stands for the maximum likelihood estimate of the parameter M (migration rate), numbers are used as a distinctiveness of subspecies where (1: *P. l. caledonica;* 2: *P. l. leucoptera*).

REFERENCES

Axelsson E, Smith NGC, Sundström H, et al (2004) Male-biased mutation rate and divergence in autosomal, z-linked and w-linked introns of chicken and Turkey. Mol Biol Evol 21:1538–47.

Gangloff B, Zino F, Shirihai H, et al (2013) The evolution of north-east Atlantic gadfly petrels using statistical phylogeography. Mol Ecol 22:495–507.

Jesus J, Menezes D, Gomes S, et al (2009) Phylogenetic relationships of gadfly petrels Pterodroma spp. from the Northeastern Atlantic Ocean: molecular evidence for specific status of Bugio and Cape Verde petrels and implications for conservation. Bird Conserv Int 19:199.

Kimball RT, Braun EL, Barker FK, et al (2009) A well-tested set of primers to amplify regions spread across the avian genome. Mol Phylogenet Evol 50:654–60.

Nei M (1987) Molecular evolutionary genetics. Columbia university press

Primmer CR, Borge T, Lindell J, Sætre GP (2002) Single-nucleotide polymorphism characterization in species with limited available sequence information: High nucleotide diversity revealed in the avian genome. Mol Ecol 11:603–612.

Weir JT, Schluter D (2008) Calibrating the avian molecular clock. Mol Ecol 17:2321–8.