Reconstructing the evolutionary history of the seabird order Procellariiformes

LEVY, MAX,ANTONY

Reconstructing the evolutionary history of the seabird order Procellariiformes

Max Levy

Abstract

The seabird order Procellariiformes, comprising the albatrosses, storm-petrels, petrels, and shearwaters, are the truly oceanic birds of the world, spending most of their lives at sea and returning to land only to breed. Despite the constraints of life in the inhospitable marine ecosystem, Procellariiformes are remarkably diverse in morphology, ecology, and life history. The order is amongst the most threatened of all bird groups, facing a range of threats at sea and at breeding colonies. Despite the intriguing diversity presented and imminent conservation concern, a modern phylogenetic hypothesis for the evolution of Procellariiformes has not yet emerged. This thesis aimed to address this by applying a dataset of thousands of genome-wide markers to a taxonomically comprehensive sample set to investigate the evolutionary history and systematics of Procellariiformes. The evolutionary relationships among species, subspecies, and populations were resolved to a fine scale with few areas of conflict, allowing new taxonomic insights. As part of the phylogenetic process, marker selection and filtering techniques have become critical, yet highly debated, steps. With growing evidence of the important roles of sex chromosomes in divergence and speciation in a wide variety of taxa, the novel genome-wide dataset was used to investigate the utility of sex-linked loci in phylogenetic reconstruction of Procellariiformes, with valuable insights for further work employing similar methods. In addition to the nuclear genome, the evolution of the mitochondrial genome in Procellariiformes was investigated, with particular relevance to the rearrangements and duplications that have been identified in birds and other animals. Mitogenomic duplications were common across Procellariiformes, with an ancient origin inferred that offers insights into this trait across bird evolution. The phylogeny presented in this thesis provides a useful resource for defining relationships in conservation management of Procellariiformes, and an evolutionary context for further comparative exploration of this remarkable order of birds.

Reconstructing the evolutionary history of the seabird order Procellariiformes

Max Levy

A thesis submitted for the degree of Doctor of Philosophy

Department of Biosciences
Durham University
2023

Table of Contents

Declaration

Of the samples used in this study a minority had been obtained and sequenced before I started working on this project. Obtaining the majority of samples and conducting labwork and organising the sequencing of these samples was my responsibility. All data processing, analyses, and results presented in this thesis are entirely my own work.

Statement of Copyright

Acknowledgements

I would like to thank my supervisor, Andreanna Welch, and Helen James, Terry Chesser, and Vincent Bretagnolle for their advice and expertise throughout. I also thank Andrea Estandía, Joan Ferrer Obiol, Alma Liss, and all the other members of our lab group. I wish to thank all of the museums and individuals that generously contributed samples that were a critical part of this thesis.

Chapter 1: Introduction

This introductory chapter outlines the approaches and methods used throughout this thesis in sections 1.1-1.3, and introduces the seabird order Procellariiformes that formed the study system that these approaches and methods were applied to in section 1.4. A summary and motivations of this thesis are given in section 1.5, along with summaries of each chapter in the form of abstracts. The background sections here are applied and expanded in the results-based Chapters 2-4. Finally, Chapter 5 provides a review of the preceding chapters, with an emphasis on future directions.

1.1 Next-generation sequencing methods

1.1.1 Whole-genome sequencing

In recent years it has become possible to generate large sequencing datasets of genomic DNA, with a wide variety of applications (Goodwin et al. 2016). Notably, the costs of producing high quality whole-genome sequences have decreased at a rapid pace (Wetterstrand 2022), allowing the generation of large numbers of *de novo* genomes of non-model organisms (Hotaling et al. 2021). Large-scale sequencing projects have been formed to investigate genomic patterns, or with a taxonomic group in focus. For example, the 1000 Genomes Project was completed in 2015, providing information on genetic variation from 2,504 human genomes (Sudmant et al. 2015, The 1000 Genomes Project Consortium 2015). The Vertebrate Genomes Project has the ambitious aim of sequencing the genomes of all vertebrate species (Rhie et al. 2021). More ambitious still, the Earth Biogenome Project aims to sequence the genomes of all eukaryotic species (Lewin et al. 2018). The Genomic Encyclopedia of Bacteria and Archaea project focuses on sequencing genomes of bacteria and archaea (Wu et al. 2009). Although completing reference genomes of all living organisms will take decades, if ever, to achieve, whole-genome sequencing has already provided insights into human health and pathology, variation at the scale of individual sites to large-scale genomic structure, signals of natural selection, gene flow, and further evolutionary processes (Cirulli & Goldstein 2010, Ellegren 2014, Feng et al. 2020, Malinsky et al. 2018, Mukherjee et al. 2017, Sudmant et al. 2015).

1.1.2 Reduced representation sequencing methods

In addition to the streamlined production of whole-genome sequences, reduced representation sequencing using short-read sequencing methods to target predetermined genomic markers have become commonplace (Jones & Good 2016). These approaches have enabled the cost-effective generation of thousands of homologous loci spanning individuals within populations and taxa to investigate evolutionary hypotheses (Davey et a. 2011, Ekblom & Galindo 2010, McCormack et al. 2013b). In a phylogenetic context, this has reduced the reliance on single or few mitochondrial and nuclear markers that have been widely used (Ballard & Whitlock 2004, Burleigh et al. 2015, Deagle et al. 2014, Tobe et al. 2010), and provided significant increases in resolution to resolve the evolutionary history of the tree of life in a phylogenomic era (Delsuc et al. 2005).

A number of different approaches to sequence large numbers of homologous genomic sequences have been introduced. In Restriction site-Associated DNA sequencing (RADseq), DNA is digested using restriction enzymes, size-selected, and sequenced (Baird et al. 2008). These regions are random rather than targeted, generating large numbers of polymorphisms at sites adjacent to digestion sites that can then be used for comparison between individuals. These methods are mostly applicable at shallow taxonomic scales as restriction sites may not be conserved across species (Leaché et al. 2015, Rubin et al. 2012).

Target capture methods leverage conserved homologous genomic regions to design probe sets to capture these regions (Faircloth et al. 2012, Hodges et al. 2007, Lemmon et al. 2012, Mamanova et al. 2010). Fragmented DNA is hybridised with RNA or DNA probes, and targeted DNA is captured, often using PCR amplification, and sequenced (Andermann et al. 2020, Glenn & Faircloth 2016). In contrast to RADseq and other restriction digestion methods (Davey et al. 2011), these sequence capture methods enrich specifically targeted loci, for which the probes are designed. Polymorphisms in obtained sequences adjacent to the conserved probes can provide phylogenetic utility. The most commonly used capture sequencing methods employ Ultraconserved Elements (UCEs; Faircloth et al. 2012) and Anchored Hybrid Enrichment (AHE; Lemmon et al. 2012) and both of these methods target the variable sites around conserved ‘anchor’ sequences (Karin et al. 2020). Further capture approaches targeting Conserved nonexonic elements (CNEEs; Edwards et al. 2017) and Rapidly Evolving Long Exon Capture (RELEC; Karin et al. 2020) have been introduced with similar methodologies.

1.1.3 Ultraconserved Elements

UCEs are short nuclear genomic sequences that are highly conserved between distantly related taxa (Faircloth et al. 2012). They are distributed throughout the genome, and are mostly non-coding and single copy in vertebrates (Bejerano et al. 2004, Van Dam et al. 2021). UCEs have been associated with a broad-range of regulatory functions (Bejerano et al. 2004, Dickel et al 2018, Reneker et al 2012, Sandelin et al. 2004, Warnefors et al 2016, Woolfe et al. 2004), and lack a single known unifying function (Harmston et al. 2013). Strong purifying selection relative to coding regions underlies the highly conserved nature of these regions (Katzman et al 2007). When enriched and sequenced for phylogenetics, UCEs are extremely conserved at the core and show increasing informative variation in flanking sequences (Faircloth et al. 2012).

The highly conserved nature of UCEs can provide probe sets that will successfully enrich genomic loci in taxonomically disparate groups and are notably useful for deep divergences spanning large evolutionary timescales (Bossert & Danforth 2018, Faircloth et al. 2012, Leaché et al. 2015, McCormack et al. 2012, McCormack et al. 2013a). In addition to their utility in evolutionary older clades, UCEs have been successfully applied at shallow scales between closely related species and within species (Andermann et al. 2019, Burress et al. 2018, Ferrer Obiol et al. 2021, Harvey et al. 2016, Smith et al. 2014, Winker et al. 2018). The broad-scale 'Tetrapods' bait set (Faircloth et al. 2012) is frequently used, and further bait sets are now available for a wide variety of taxonomic groups, including Cnidaria (Quatrini et al. 2018), fish (Alfaro et al. 2018, Faircloth et al. 2013), and many arthropod classes and orders (Brantstetter et al. 2017, Faircloth et al. 2015, Faircloth 2017). The ability to produce sequencing data at high coverage across divergent taxa with predesigned probe sets has contributed to the high popularity of UCEs and other sequence capture methods for phylogenetic research (Andermann et al. 2020). In common with other capture sequencing methods, a major benefit of the use of predesigned bait sets that consistently capture the same sets of loci is that meta-analyses of combined datasets from multiple studies to assemble the tree of life will eventually become possible (Lemmon et al. 2012).

1.1.4 Concerns and best practices in next-generation sequencing datasets

Next-generation sequencing has allowed the rapid and cost-effective generation of large phylogenomic datasets spanning hundreds or thousands of loci with high power to answer evolutionary questions. However, resolving the tree of life unambiguously has remained

challenging despite the use of large sequencing datasets (Jeffroy et al. 2006, Philippe et al. 2011, Shen et al. 2017, Walker et al. 2018). Indeed, incongruence at difficult nodes remains, for example at prominent points in the evolution of mammals (Esselstyn et al. 2017), birds (Braun & Kimball 2021), and the ancient animal lineages of sponges and ctenophores (King & Rokas 2017).

The causes of these issues may be evolutionary processes, such as incomplete lineage sorting, introgression, and natural selection (Castoe et al. 2009, Edwards 2009, Maddison 1997), or due to data type effects (Blom et al. 2017, Reddy et al. 2017). With the availability of different classes of markers through restriction-digest and sequence capture approaches, marker choice has become a strong component of phylogenetic methodology. Differing results from different marker types has been a common finding from genome-wide studies (Alda et al. 2021, Braun & Kimball 2021, Reddy et al. 2017). Different markers may be more suitable for specific applications (Betancur-R et al. 2014, Chen et al. 2015, Karin et al. 2020, Salichos & Rokas 2013, Yang 1998) and analysing combined datasets incorporating multiple marker types may increase phylogenetic power to resolve difficult nodes (e.g. Chan et al. 2020, Cloutier et al. 2019, Ferrer Obiol et al. 2021, Harvey et al. 2020, Hutter & Duellman 2023, Singhal et al. 2017). The use of different marker types in combined analyses also raises questions as to the bioinformatic treatment of such datasets. Filtering of various parameters of sequencing datasets that can impact phylogenetic results under different conditions has received a great deal of attention in recent years (e.g. Bossert et al. 2017, Gilbert et al. 2018, Hosner et al. 2016, Mclean et al. 2019, Molloy & Warnow 2018, Portik & Wiens 2021, Simmons et al. 2016, Streicher et al. 2016, Van Dam et al. 2021, White & Braun 2019, Wiens & Morrill 2011) and provided guidance on the best practice treatment of datasets, although without concordance in all areas.

1.2 Phylogenetic methods

The increasing feasibility of generating large sequencing datasets to apply to phylogenetic questions has created challenges for developing effective and computationally tractable analyses methods (Bravo et al. 2019). The use of maximum parsimony, that has been commonly used to infer trees from morphological datasets, is less applicable in an era of large-scale molecular sequencing datasets (Felsenstein 1978, Yang & Rannala 2012). The more computationally intensive model-based approaches of maximum likelihood (ML) and Bayesian inference (BI) have become widely used for phylogenetic inference of molecular sequence datasets (Young & Gillung 2019).

These methods use substitution models of evolution to describe the rates of change of mutations among aligned sequence datasets (Arenas 2015). A multitude of substitution models to account for evolutionary processes have been described (Liò & Goldman 1998). The most complex model is the general time reversible model (GTR; Tavaré 1986) that includes different rates for each possible nucleotide substitution change. The GTR model is often combined with further parameters incorporating the proportion of invariant sites (+I; Shoemaker & Fitch 1989) and the rate of variation across sites (+G; Yang, 1993), and GTR+I+G or GTR+G models are frequently used (Arenas 2015). Although model selection is often considered an important step in phylogenetic inference, the most parameter rich GTR+I+G model very often outperforms less complex models (Abadi et al. 2019; but see Sumner et al. 2012). A GTR+G model may be opted for instead given the inherent difficulties of independently estimating I and G (Yang 2006).

ML phylogenetic inference determines the topology that maximises the probability of observing the data given the parameters (Yang & Rannala 2012). In phylogenetics, BI estimates prior distributions of parameters and the observed data to produce posterior distributions of topologies (Yang & Rannala 2012). Although ML and BI work in alternate ways, they are often used in combination and produce concordant phylogenetic results.

These methods typically use datasets of multiple concatenated loci or genomic regions, that can be partitioned by shared characteristics (Lanfear et al. 2012). In this way, different models can be applied to different partitions, but a single topology and branch lengths encompassing the entire dataset is produced. However, gene trees are not necessarily equivalent to species trees (Degnan & Rosenberg 2006, Nichols 2001), and concatenation of genes with heterogeneous evolutionary histories can lead to well-supported but incorrect topologies (Edwards et al. 2007, Jarvis et al. 2014, Rokas et al. 2003, Song et al. 2012). This has led to the formation of species tree methods under the multispecies coalescent model that can account for incomplete lineage sorting (Edwards 2009, Edwards et al. 2016). These methods do not assume the same topology for all genes (Liu et al. 2015) and can produce summary species trees (e.g. ASTRAL-III; Zhang et al. 2018) or site-based species trees (e.g. SVDQuartets; Chifman & Kubatko 2014). However, conflicting phylogenetic results between concatenated approaches and MSC species tree methods are common in genome-wide datasets (e.g. Chan et al. 2020, Gatesy & Springer 2014, Song et al. 2012), and debates as to which methods to use are ongoing (Gadagkar et al. 2005, Gatesy & Springer 2014, Hutter & Duellman 2023, Knowles 2009, Kubatko & Degnan 2007, Simmons & Gatesy 2015, Springer & Gatesy 2016). The necessary construction of prior gene trees for summary-based MSC inference in particular

causes problems, most notably in the form of gene tree estimation error (Hahn & Nakhleh 2016, Xi et al. 2015). MSC methods that conduct simultaneous estimation of gene trees and the species tree, such as StarBeast (Heled & Drummond 2010), can provide more robust inference but are computationally expensive and not suitable for the large-scale sequencing datasets that are now commonplace (Liu et al. 2015).

Beyond the artefactual impediments of model misspecification and other data driven issues (Reddy et al. 2017), naturally occurring evolutionary processes that create heterogeneous signals in genome-wide datasets, such as incomplete lineage sorting, introgression, horizontal gene transfer, and natural selection (Castoe et al. 2009, Edwards 2009, Hobolth et al. 2011, Knowles et al. 2018, Maddison 1997), distance phylogenetic inference from a fully resolved branching phylogeny (Bravo et al. 2019). Rapid radiations of species, creating short branches between speciation events and little opportunity for informative variation to arise are a particular problem for accurate phylogenetic reconstruction (Alda et al. 2019, Degnan & Rosenberg 2006). Beyond concatenation and species tree methods, phylogenetic networks may be a useful complement regarding these issues in constructing trees (Bapteste et al. 2013, Huson & Bryant 2006). Networks can incorporate reticulate evolutionary patterns that arise through incomplete lineage sorting and introgression, and may represent evolutionary histories more aptly than strictly bifurcating trees at some nodes (Hallström & Janke 2010, Mallett et al. 2016, Suh 2016). Reflecting the conflicting signals inherent to genome-wide datasets, considering different approaches may be useful in testing evolutionary hypotheses.

1.3 Evolutionary history beyond phylogeny

Obtaining a phylogeny opens further possibilities for investigating the evolution, ecology, biodiversity and other diverse characteristics of a group of interest. Phylogenetic comparative methods allow inference into the history of such traits (Cornwell & Nakagawa 2017).

Estimating the divergence times of internal nodes in a phylogeny can be conducted to investigate timelines of evolution (dos Reis et al. 2016). The rates of molecular evolution in sequencing datasets can be used with strict clock models (Zuckerkandl & Pauling 1965), or relaxed clock models that allow variation over time and between lineages to estimate divergence times (Drummond et al. 2006). Fossils with known taxonomic placement and age can be used to calibrate the ages of nodes in the phylogeny (Parham et al. 2012). Estimations of divergence times can be partnered with biogeographical analyses to investigate current distribution patterns and how these have evolved from ancestral ranges over time (Ronquist &

Sanmartín 2011). Combined estimates of diversification ages and where these events occurred can be useful in inferring the potential involvement of past climatic or geological processes (e.g. Davies & Buckley 2011, Dufour et al. 2020, Ferrer Obiol et al. 2022).

Combining phylogenetic relationships with observed traits in organisms can provide insights into the evolutionary history of such characteristics. Using this approach, ancestral state reconstruction is a useful tool for extrapolating back in time the character states of common ancestors of taxa (Joy et al. 2016). This can provide insights into when and how frequently morphological, ecological, or life history traits have evolved over time, and the nature of the ancestor of a group of interest (e.g. Ksepka et al. 2017, Legendre et al. 2016, Mayr 2011, Odom et al. 2014, Schluter et al. 1997). As well as phenotypic and geographic traits, ancestral state reconstruction can be informative in testing genomic changes during evolution, such as ancestral sequences, gene copy number and order (e.g. Blanchette et al. 2004, Damas et al. 2018, O'Connor et al. 2018, Ouzonis 2005).

1.4 Introduction to the seabird order Procellariiformes: albatrosses, storm-petrels, petrels, and shearwaters

Procellariiformes is an order of seabirds containing four families: Diomedeidae (albatrosses), Oceanitidae (Southern storm-petrels), Hydrobatidae (Northern storm-petrels), and Procellariidae (petrels and shearwaters). These four families are represented by 26 genera, containing 138-145 extant species and 185-191 taxa including subspecies (Clements et al. 2022, Gill et al. 2023, HBW & BirdLife International 2022). Procellariiformes are the truly oceanic birds of the world, spending most of their lives at sea and returning to land only to breed (Warham 1990). Procellariiformes species cover vast distances (Weimerskirch et al. 2014) and perform some of the longest known migrations (Shaffer et al. 2006). Despite the apparent constraints of life in the inhospitable marine ecosystem, Procellariiformes are remarkably long-lived, reaching over 30 years in many species (Warham 1996, Wasser & Sherman 2010) and, at over 70 years, a *Phoebastria immutabilis* individual is currently the oldest known wild bird (U.S. Fish & Wildlife Service 2022). As indicated by their high longevity, annual survivorship is high in Procellariiformes adults (Warham 1996). However, Procellariiformes are becoming increasingly threatened by human activity, and face uncertain futures (Clark et al. 2023, Croxall et al. 2012, Dias et al. 2019, Phillips et al. 2023, Rodríguez et al. 2019).

1.4.1 Procellariiformes morphology, flight and feeding

The Procellariiformes are known as 'tubenoses' in English, referring to the diagnostic tubular nostrils unique to the order. In Diomedeidae species the nostrils are separated by the upper ridge of the bill, and in other families the nostrils are merged as a single tube separated by a vertical septum (Brooke 2004). This trait is associated with the highly developed olfactory bulb of the brain in Procellariiformes, which is amongst the largest of all birds (Bang 1966). This confers a very strong sense of smell, and hints at the mechanisms that allow Procellariiformes to find prey in the visually monotonous oceanic landscape (Nevitt 2000). In addition to foraging, smell is important in navigation and nest and social recognition in Procellariiformes (Nevitt 2008).

Procellariiformes also show unique morphology of the digestive tract among birds, with an enlarged proventriculus and no crop (Warham 1996). The proventriculus frequently contains extremely calorically dense stomach oil derived from consumed food, which allows Procellariiformes to transport large amounts of energy over long distance foraging trips (Warham et al. 1976). This oil is also regurgitated and ejected in some species when disturbed, forming a defensive mechanism (Warham et al. 1976).

Plumage of Procellariiformes species is often black, white, and with shades of grey and brown. Many species show a familiar pattern of dark coloured upperparts and pale underparts (Brooke 2004), although there are many deviations including all dark plumages (e.g. *Phoebetria fusca*), all white plumages (e.g. *Pagodroma nivea*), mostly grey plumages (e.g. *Oceanodroma furcatus*), and more intricate patterns (e.g. *Daption capense*). Plumage patterns in Procellariiformes have been associated with feeding ecology and species behaviour at sea (Bretagnolle 1993). Although rare in birds (Galeotti et al. 2003), plumage polymorphism occurs in some Procellariiformes species, often in the form of light and dark morphs and intermediates between these extremes, for example in *Fulmarus glacialis* (Fisher 1952) and *Macronectes giganteus* (Carlos & Voisin 2008).

Despite similar plumage patterns throughout Procellariiformes, morphometric size and structure are highly diverse among species and offer insights into the flight and feeding specialisations found across the order. Storm-petrel species (Hydrobatidae and Oceanitidae) are very small, weighing as little as 20g with wingspans as small as 32cm in *Oceanodroma microsoma* (Onley & Scofield 2007). These species have broad wings of low aspect ratio with high manoeuverability (Warham 1977), and are specifically adapted to surface pattering with frequent flapping in search of food (Brooke 2004).

In contrast, the great wandering albatrosses of the genus *Diomedea* can weigh over 11kg and have wingspans up to 3.5m (Tickell 1968), the largest of all birds (del Hoyo et al. 2020). Diomedeidae and many Procellariidae species have long wings with high aspect ratios that are well adapted for long distance gliding at high speeds (Warham 1977). Diomedeidae species make use of wind speed gradients to glide efficiently in dynamic soaring (Pennycuick 1982). During calm conditions Diomedeidae species often alight on the sea, illustrating the critical importance of efficient gliding in their flight patterns (Warham 1977). Only Diomedeidae and *Macronectes* giant petrels, the largest members of Procellariiformes, possess a shoulder lock in the form of a tendon sheet stretching from the sternum to the humerus, which reduces the muscle power needed to hold the wing fully outstretched, enabling protracted gliding (Pennycuick 1982).

The prion genus *Pachyptila* show a unique adaptation in the form of lamellae along the edge of the bill (Brooke 2004). By forcing water past the lamellae, *Pachyptila* species filter-feed on plankton either when swimming or in low flight over the surface using the feet and wings to maintain balance (Murphy 1936).

The diving-petrel genus *Pelecanoides* is atypical among Procellariiformes species, with short wings of lower aspect ratio that are not well adapted for long distance flight (Warham 1977). Unlike many Procellariiformes species that glide frequently, *Pelecanoides* fly with rapidly beating wings and feed mostly close to breeding colonies (Brooke 2004, Navarro et al. 2015). The wing bones are flattened, in a paddle-like shape that is instead perfectly adapted for wing-propelled diving underwater (Kuroda 1967). Similar to the diving-petrels, the shearwater genera *Ardenna* and *Puffinus* are adapted for diving and exhibit a comparably flattened humerus (Kuroda 1954) that aids in plunge or surface diving and underwater pursuit of prey (Ashmole 1971). *Procellaria* and some Diomedeidae species are also strong divers (Brooke 2004). *Pterodroma* species are very strong fliers, swooping high during gliding, but do not dive and feed mostly at the surface (Brooke 2004, Warham 1977).

A range of food items is consumed at sea by Procellariiformes species, including fish, nudibranchs, cephalopods, amphipods, copepods, and euphausiids (Prince & Morgan 1987, Warham 1996, Brooke 2004, Navarro et al. 2013, Imber 1973, Reid et al. 1997, Ridoux 1994). The *Macronectes* giant petrels are the only Procellariiformes species to obtain a significant part of their diet on land, by scavenging from seal, penguin, and petrel carcasses (Brooke 2004, Ridoux 1994). *Phoebetria* albatross species will also consume penguin and other petrel carcasses (Ridoux 1994). Beyond these rare examples, Procellariiformes obtain their entire diets at sea.

1.4.2 Procellariiformes distributions and breeding

Procellariiformes are colonial breeders, occupying remote islands worldwide and on continental landmasses in some places (Brooke 2004, Warham 1990). The greatest diversity is present in the southern hemisphere, where dense assemblages of species and populations are supported at South Georgia, Gough Island, and Tristan da Cunha in the South Atlantic Ocean, at the Crozet Islands and Kerguelen Islands in the southern Indian Ocean, and in the New Zealand region in the South Pacific Ocean (Chown et al. 1998, Clarke et al. 2012, Cuthbert 2004, Warham 1990, Brooke 2004, Jouventin et al. 1985, Robertson et al. 2021, Richardson 1984, Ryan et al. 1990, Weimerskirch et al. 1989). Within the northern hemisphere, Hawaii has the highest species richness of Procellariiformes, and Baja California, Madeira, the Canary Islands, and the Pacific coast of South America are hotspots of endemism (Bedolla-Guzmán et al. 2019, Chown et al. 1998, Harrison 1990). Contrary to a typical species richness gradient of decreasing towards high latitudes, Procellariiformes density is highest at 37° to 59°S in all three main ocean basins (Chown et al. 1998).

The breeding patterns of Procellariiformes are relatively similar throughout the order (Brooke 2004). Most species return to their colonies only at night, where they nest in burrows or natural crevices (Brooke 2004, Schramm 1986, Warham 1990). Some species nest in a scrape in the ground, which is associated with diurnal colony visits; the opposite is true for hidden, burrow or crevice nesting species that tend to visit nocturnally (Brooke 2004).

Visual mating displays can be spectacular in some albatross species, which visit nests diurnally, while calls are more important in attracting mates for nocturnally visiting species (Brooke 2004). Mating with the same partner year after year is common, and some species form long bonds (Bried et al. 2003, Warham 1990). Fecundity in Procellariiformes is low, with a single egg laid per breeding attempt in all species (Warham 1990). In the large and long-lived albatross species, age at first breeding is high at around 8-12 years (Pickering 1989, Weimerskirch et al. 1987). Breeding frequency varies, including breeding more than once per year in some tropical species, typical annual breeding, and breeding every two years in some species (Warham 1990). For some species and populations of Hydrobatidae storm-petrels, separate populations breeding in the warm and cool seasons at the same island, and often using the same burrows as nests, have been found (Bolton et al. 2008, Friesen et al. 2007, Taylor et al. 2017, Taylor et al. 2019). Natal philopatry is generally strong in seabirds, including Procellariiformes (Coulson 2002). In *Pterodroma baraui* on Reunion Island in the Indian

Ocean, strong philopatry was found between breeding colonies separated by only 5km, resulting in isolated genetic populations at a very fine scale (Danckwerts et al. 2021).

1.4.3 Procellariiformes as a study system

In many cases Procellariiformes represent the extremes of a diverse range of traits, occupying the high latitudes of the Southern Ocean in greatest density, living very long lives in some species, showing variations in size over orders of magnitude, and covering vast distances as part of a pelagic life. Procellariiformes inhabit all the world's oceans, making them appropriate sentinels of marine health (Clark et al. 2023, Dias et al. 2019). New taxa of Procellariiformes have been described in recent years (Bretagnolle et al. 2022, Harrison et al. 2013, Rodríguez, et al. 2020), and recent work has uncovered insights into the breeding biology (Rackete et al. 2021, Danckwerts et al. 2021), movement and spatial ecology (Neves et al. 2023, Receveur et al. 2022), and biogeography and systematics (Ferrer Obiol et al. 2022, Ferrer Obiol et al. 2023) of the group, with wider evolutionary and ecological implications. Procellariiformes present tremendous opportunities as a study system, at a pressing time when their future conservation is increasingly threatened (Dias et al. 2019).

1.5 Summary of this thesis

This thesis implements the described next-generation sequencing methods, specifically the reduced representation markers UCEs, to the intriguing study system of the seabird order Procellariiformes. A well-resolved phylogeny sampling the diversity of a taxonomic group is a necessity for further evolutionary and ecological study of a group of interest. Despite the large amount of attention that birds have received in general biological study, and specifically in systematic study, this was notably lacking for the charismatic and diverse Procellariiformes and formed a significant part of the motivation of this thesis. This was addressed in Chapter 2 of this thesis, for which genome-wide UCEs were used to provide a comprehensive phylogeny of Procellariiformes. This phylogeny represents a significant advancement on what has come before, providing genome-wide data for many taxa for the first time, and including taxa that have not been investigated in a phylogenetic context previously.

The analytical approaches and methods used are important facets to phylogenetic studies. During the process of building the phylogeny, a strong effect of sex-linked UCE loci at fine scales was encountered, leading to further exploration of this. Strong evidence has indicated

the importance of sex chromosomes in evolutionary divergence and speciation, but a relative paucity of information of this importance in a phylogenetic context has emerged, forming a further motivation of this thesis. The phylogeny of Procellariiformes presented in Chapter 2, and the large sequencing dataset underlying it, were used to investigate the importance of subsampling genomic loci of the autosomes and of sex chromosomes in Chapter 3 of this thesis, providing valuable insights for further work using similar methods.

Structural variants have become an increasingly common finding in genome sequencing studies. Within birds and other animals, multiple different gene rearrangements, including extensive duplications, of the mitochondrial genome have been found. With a strongly resolved phylogeny of Procellariiformes, a comparative approach was used to investigate the formation and evolution of mitogenomic duplications in this study group in Chapter 4 of this thesis. Investigating the evolution of the mitogenome, to complement the preceding chapters focusing on the nuclear genome, was a motivation for this chapter. Evidence of an ancient duplication event in the mitogenome with subsequent non-uniform degradation over time was found, with relevance to the evolution of this trait in the wider waterbird clade and among all birds.

Chapter 5 provides a review of the points derived from each of these chapters, the outstanding questions they pose, and the future directions for further research.

The chapters of this thesis demonstrate stages of evolutionary study, from the generation of a novel and comprehensive phylogeny for a taxonomic group of interest, testing the best-practices among the data analyses methods employed, and testing evolutionary scenarios beyond phylogeny that are unattainable in the absence of a resolved tree. More detailed summaries of each chapter, other than Chapter 5, follow below.

1.5.1 Summary of Chapter 2: A novel phylogeny of Procellariiformes using genome-wide markers

Gaps in biodiversity knowledge, including in the evolutionary relationships among lineages and their tree of life, are an impediment to conservation in a worsening environmental crisis. Overcoming local shortfalls in taxonomic clarity across the tree of life should be a high priority in the near future. Procellariiformes, the seabird order containing albatrosses, storm-petrels, petrels and shearwaters, are among the most threatened groups of all birds, and are subject to a range of threats that are set against the uncertain backdrop of climate change. In addition to this, Procellariiformes exhibit an extreme range of diversity despite occupying the challenging oceanic niche. Despite this, a modern phylogenetic hypothesis for the evolution of

Procellariiformes has not yet emerged. The most comprehensive attempt at resolving the phylogenetic relationships of Procellariiformes was previously restricted to a limited proportion of taxa using a single mitochondrial marker, resulting in low resolution. In this chapter, the evolutionary relationships of Procellariiformes were addressed using a novel dataset of thousands of genome-wide UCE loci applied to a taxonomically thorough sample set of Procellariiformes populations. Results were presented for 335 individual birds, representing 89% of species and 85% of all Procellariiformes taxa. Evolutionary relationships were comprehensively resolved to a fine scale, allowing many new taxonomic insights. Several complex nodes of Procellariiformes evolution were identified, including instances of likely introgression and incomplete lineage sorting, and rapid speciation events resulting in multiple taxa. The robust hypothesis of Procellariiformes evolution presented provides useful information of taxon boundaries and relationships for conservation management. Additionally, the results offer a thorough phylogenetic context for further exploratory study of the evolution and ecology of this diverse clade of birds.

1.5.2 Summary of Chapter 3: Investigating the impacts of sex-linked genomic markers on phylogenetic reconstruction in Procellariiformes seabirds

Genomic marker selection and filtering techniques employing phylogenomic subsampling of various parameters aimed at amplifying phylogenetic signal and reducing noise have become critical, yet highly debated, steps in the phylogenetic process. Compared to some frequently tested targets for marker subsampling, the impact of sex-linked markers (i.e. sites or loci on sex chromosomes) has been relatively underexplored. There is growing evidence of the critical roles played by sex chromosomes in divergence and speciation in a wide variety of taxa and sex determination systems, suggesting that sex-linked markers may be a relevant subsample to consider in phylogenetic study. Furthermore, given the differences in inheritance dynamics of sex chromosomes relative to autosomes between sexes, and divergent evolutionary processes including sexually antagonistic polymorphisms that are expected to arise throughout the evolution of sex chromosomes, it is reasonable to anticipate the potential existence of sex-biased effects on phylogenetic reconstruction of both mitochondrial and nuclear markers. In this chapter, the phylogenetic utility and potential sex-biased effects of sex-linked markers relative to autosomal markers were tested in the case of the seabird order Procellariiformes. Sex-linked loci differed statistically from autosomal loci in being consistently shorter, having lower taxon occupancy, greater missing data, and higher numbers of parsimony informative

sites. Phylogenetic trees built from sex-linked loci showed sex-specific clades of females with spuriously long branches and males with comparable branch lengths relative to autosomal topologies, lower values at support metrics, and resolved considerably fewer monophyletic species than autosomal trees. Underlying these patterns were higher missing data in sex-linked loci in females relative to autosomal females and males of both sex-linked and autosomal loci, and a higher frequency of heterozygous sites in sex-linked loci in females than in autosomal loci. These results provide strong evidence for sex-biased effects that mislead phylogenetic inference, and indicate a role for sex-linked loci in determining phylogenetic results in Procellariiformes. Testing these patterns in other groups with potentially heterogeneous sex chromosomes should be a priority, and this class of loci should be considered in future phylogenomic subsampling methods.

1.5.3 Summary of Chapter 4: Investigating the prevalence and evolution of duplications in the mitochondrial genome of the seabird order Procellariiformes

The conserved size, gene content, and gene order of animal mitochondrial genomes (mitogenomes) has led to the suggestion that these characteristics are maintained by selection. However, duplications of genes and rearrangements of the mitochondrial gene order have become a frequent observation in animals, and especially in birds. A considerable number of different rearrangements and duplicated regions spanning from CYTB to the control region have been identified since the unduplicated chicken mitogenome was hypothesised to represent the ancestral avian structure. The subsequent documentation of mitogenome duplications among taxonomically diverse clades of birds has led to the suggestion that it could be considered the default state. However, knowledge of when and how frequently independent mitogenome duplication events have occurred in birds remains lacking, and this trait has seldom been tested at large taxonomic scales in concert with a well-resolved phylogeny. This question was addressed using Procellariiformes, for which there is some evidence for the presence of mitogenome duplications, as well as evidence in the wider waterbird clade encompassing this order. Mitogenomic duplications were detected through elevated sequencing depth coverage in a large-scale capture enrichment sequencing dataset of Procellariiformes, assessing the prevalence and evolutionary history of mitogenomic duplications across this group of birds. Duplications of varying extents were present across the entire tree of Procellariiformes, with repeated degeneration of duplicated regions and, rarely, complete loss of such regions in only two clades containing three taxa each. A variety of

different mitogenome rearrangements were found, including the region spanning ND6 to the control region in most species, a 3' prime segment of CYTB was also duplicated in some species, and for some species only the control region was duplicated. Ancestral state reconstructions indicated that mitogenome duplications were present in the ancestor of Procellariiformes, preceding the radiation of this clade and persisting in this state over a large evolutionary timescale. The loss of various duplicated regions across the tree was consistent with the tandem duplication and random loss model of mitogenome rearrangement. The extremely high prevalence of mitogenomic duplications found in Procellariiformes was unusual among birds, and the persistence over a large timescale poses questions of the possible functions of duplicated regions. These genomic changes may be related to the remarkable flight capabilities and longevity of Procellariiformes, with which duplications have previously been associated.

Chapter 2: A novel phylogeny of Procellariiformes using genome-wide markers

2.1 Introduction

Gaps in biodiversity knowledge are an impediment to conservation in a worsening environmental crisis (Barnosky et al. 2011, Dubois 2003, Wheeler et al. 2004). These gaps can be characterised into a variety of forms (Hortal et al. 2015), including the lack of description of species relative to the true number that exist (the Linnean shortfall; Lomolino 2004), lack of knowledge of species distributions (the Wallacean shortfall; Lomolino 2004), and the persisting lack of knowledge of the evolutionary relationships among lineages and their tree of life, termed the Darwinian shortfall (Diniz-Filho et al. 2013). Despite advances in sequencing and phylogenetic methods (Buermans & den Dunnen 2014, Delsuc et al. 2005), and an increasing frequency of phylogenetic studies, information regarding fine-scale relationships often remains lacking (Diniz-Filho et al. 2013, Engel et al. 2021). Successful conservation requires management at appropriate scales of taxa at a biologically relevant taxonomic rank with known geographic ranges. Overcoming local Darwinian shortfalls by defining the relationships between lineages and clarifying taxonomic ranks, for example as species, subspecies, or other evolutionary units within populations for management is thus a critical component for effective and targeted conservation action (Agapow et al. 2004, Mace 2004, Moritz 1994, Phillimore & Owens 2006, Purvis et al. 2005, Rojas-Soto et al. 2010). With this and the pressing environmental crisis in mind, taxonomic clarity across the tree of life should be a high priority in the near future (Engel et al. 2021, Wheeler et al. 2004). Beyond having a critical impact on conservation efforts, our well-defined biodiversity knowledge gaps are fundamentally interconnected, and hinder our ability to progress potential studies in evolution and ecology at broad and fine scales (Hortal et al. 2015).

Procellariiformes are among the most threatened groups of all birds, with one of the highest proportions of declining populations (Croxall et al. 2012, Phillips et al. 2023). The International Union for the Conservation of Nature (IUCN) Red List accords 66 of the 145 (45.5%) Procellariiformes species they recognise as vulnerable, endangered, or critically endangered (IUCN 2022), considerably higher than in other bird groups (Rodríguez et al. 2019). Procellariiformes are highly adapted to the remote marine environment, where increasing anthropogenic pressures at breeding colonies and at sea are contributing to precarious conservations positions of many species (Dias et al. 2019, Rodríguez et al. 2019). Of the threats faced by Procellariiformes, nest predation by invasive species is the greatest (Dias et al. 2019,

Spatz et al. 2023), and has directly contributed to the extinction of species (BirdLife International 2023). At sea, fishery bycatch (Anderson et al. 2011, Dias et al. 2019, Montevecchi 2023, Žydelis et al. 2013) and plastic waste consumption (Charlton-Howard et al. 2023, Clark et al. 2023, Phillips et al. 2023, Phillips & Waluda 2020, Roman et al. 2019) are significant problems. A backdrop of climate change accompanies these specific threats, and is expected to have large scale effects on seabird populations worldwide (Pistorius et al. 2023). In addition to imminent conservation concern, the Procellariiformes exhibit an extreme range of diversity despite occupying the remarkably challenging oceanic niche. The order encompasses a wide array of morphologies, from the tiny storm-petrels to massive albatrosses, a multitude of flight and feeding specialisations from surface pattering to deep diving, unusually high longevity and low fecundity, and other unique life history and behavioural traits (Brooke 2004, Onley & Scofield 2007, Warham 1990, Warham 1996), representing an ideal system for further evolutionary and ecological study.

Despite this, a comprehensive modern phylogenetic hypothesis for the evolution of Procellariiformes has not yet emerged. The order has a complicated taxonomic history punctuated by frequent revisions, including monographs by Coues (1864a, 1864b, 1866a, 1866b, 1866c), Godman (1907-1910), Loomis (1918), and Alexander (1928). Gregory Mathews conducted much of the early work on procellariiform classification based on external morphological and skeletal characters and, in 1934, he synthesised a wide body of literature into a checklist of the order comprising four families of 51 genera and 87 species (Mathews 1934). This work did not, however, settle debates around Procellariiformes systematics of this era with further revisions to come (e.g. Kuroda 1954, Mathews & Hallstrom 1943, Murphy 1936, Witherby et al. 1940). Mathews himself would come to reject many of his own recommendations, and in 1948 he lumped the Procellariiformes from 51 genera to a mere 12 (Mathews 1948). Unsurprisingly, Mathews' work was controversial (Serventy 1950) and, in 1965, motivated by ongoing issues in Procellariiformes nomenclature and classification, a group of 15 eminent ornithologists published their preferred classification of Procellariiformes (Alexander et al. 1965). This list approached a standardised classification of the order, and provided a benchmark for more recent work into the systematic relationships of the order using genetic methods.

Early molecular work used the mitochondrial gene cytochrome *b* to examine relationships in albatross species (Nunn et al. 1996), shearwaters (Austin 1996, Heidrich et al. 1998), the genera *Bulweria* and *Psuedobulweria* in relation to other Procellariidae genera (Bretagnolle et al. 1998), and in finer-scale species relationships among taxa (Wink et al. 1993a, Wink et al.

1993b, Brooke & Rowe 1996). Encompassing the 14 albatross cytochrome *b* sequences provided by Nunn et al. (1996) and novel sequences for 71 additional species, Nunn & Stanley (1998) produced what remains to date the most comprehensive attempt at resolving the phylogenetic relationships of Procellariiformes. In this phylogeny, the paraphyletic Oceanitidae and Hydrobatidae (considered subfamilies Oceanitinae and Hydrobatinae within the family Hydrobatidae) were sister to Diomedeidae and Procellariidae, with the genus *Pelecanoides* as sister to all other Procellariidae, conforming to previous assumptions that the diving-petrels formed the distinct family Pelecanoididae (Nunn & Stanley 1998).

Several large-scale bird phylogenies in subsequent years examining the deep relationships of Neoaves have inferred discordant family level relationships of Procellariiformes (Hackett et al. 2008, Kuhl et al. 2021, Prum et al. 2014, Reddy et al. 2017) to those found by Nunn & Stanley (1998). More recently, Estandia et al. (under review) constructed a backbone phylogeny of Procellariiformes by sequencing genome-wide ultraconserved elements (UCEs) of 51 species, resolving Diomedeidae as sister to all other families, followed by a linear branching pattern of Oceanitidae, Hydrobatidae, and Procellariidae, with the genus *Pelecanoides* embedded within Procellariidae. This family level topology is concordant with those found by some large-scale studies of birds using genome-wide datasets (Kuhl et al. 2021, Prum et al. 2014, Reddy et al. 2017), but not all (Hackett et al. 2008), and is discordant with those found by Nunn & Stanley (1998).

Discordance in phylogenetic relationships inferred by mitochondrial and nuclear markers is a frequent occurrence in animals (Toews & Brelsford 2012), and is well known in Procellariiformes species (Gangloff et al. 2013, Silva et al. 2015, Torres et al. 2021). Although often highly variable, mitochondrial markers may not represent the true relationships between taxa (Ballard & Whitlock 2004, Edwards & Bensch 2009, Funk & Omland 2003). The reliance on cytochrome *b* alone to infer Procellariiformes relationships is unlikely to have provided robust results, and the discrepancies found at higher-level relationships (Estandia et al. under review, Nunn & Stanley 1998) are likely to be replicated at finer scales. Moreover, reliance on any single mitochondrial or nuclear marker to produce robust phylogenetic relationships is unwise given the presence of genomic processes such as incomplete lineage sorting, introgression, and natural selection that generate heterogeneous histories across genomes (Bravo et al. 2019, Degnan & Rosenberg 2009, Edwards et al. 2016, Rokas et al. 2003). Sequencing entire genomes, or using reduced representation methods to generate homologous sequences for thousands of genomic loci is now commonplace, for example using UCEs (Faircloth et al. 2012), restriction site-associated sequencing (RADseq; Baird et al. 2008), and

anchored hybrid enrichment (Lemmon et al. 2012). Revisiting previous phylogenetic results based on mitochondrial loci alone, or on limited numbers of loci, may provide valuable insights (Kimball et al. 2021).

In addition to discordance with modern genomic studies, the largest phylogenetic study of Procellariiformes sampled 85 species (Nunn & Stanley 1998), lacking a significant proportion of named diversity. Current classifications of Procellariiformes recognise between 119 and 145 species, and 184 to 191 taxa including subspecies, with a total of 199 taxa (Christidis et al. 2018, Clements et al. 2022, Gill et al. 2023, HBW and BirdLife International 2022). Several studies have sampled further taxa by studying the evolutionary relationships of clades within Procellariiformes using limited numbers of mitochondrial loci or mitochondrial and nuclear loci (e.g. Chambers et al. 2009, Gangloff et al. 2012, Masello et al. 2019, Masello et al. 2022, Robertson et al. 2016, Wallace et al. 2017, Welch et al. 2014). More recently, some studies have begun to apply genome-wide markers to finer-scale questions in Procellariiformes evolution. Taylor et al. (2019) used RADseq to investigate the relationships of populations of two Hydrobatidae species. Ferrer Obiol et al. (2021) applied RADseq and UCEs to produce a well resolved phylogeny of shearwaters. Shepherd et al. (2022) used RADseq to examine the fine-scale relationships of two *Pachyptila* species. However, a large number of Procellariiformes taxa have not received further molecular study, and almost all have not yet been examined with modern, genome-wide markers, or using modern phylogenetic inference methods. As a result, known gaps in systematic information remain throughout Procellariiformes. Additionally, the existence of cryptic species and recent descriptions of novel taxa (Abeyrama et al. 2021, Bolton et al. 2008, Bretagnolle et al. 2022, Bretagnolle & Shirihai 2010, Harrison et al. 2013, Pyle et al. 2011, Robertson et al. 2016, Rodríguez, et al. 2020, Ryan et al. 2014, Taylor et al. 2019) indicate that further unknown gaps in systematic knowledge also remain in Procellariiformes. Given the imminent conservation concern, and potential interest as a biological system for further study, addressing the local Darwinian shortfall in Procellariiformes to establish more robust hypotheses of evolution and relationships should be a priority.

In this chapter, these questions are addressed using a novel dataset of thousands of genome-wide UCE loci applied to a taxonomically thorough sample set of Procellariiformes populations. The UCE dataset of 51 Procellariiformes species employed by Estandia et al. (under review) was incorporated and expanded to provide UCE data for 335 samples, representing 89% of species and 85% of all taxa. Evolutionary relationships were comprehensively resolved to the subspecies level, allowing new taxonomic insights and

recommendations. Areas of the Procellariiformes tree with phylogenetic conflict within the dataset where further taxonomic study will prove useful were identified. The robust hypothesis of Procellariiformes evolution presented should provide further understanding of taxon boundaries and relationships useful for conservation management. Additionally, the topology presented offers a useful and comprehensive taxonomic context for further exploratory study of the evolution and ecology of a diverse clade of birds.

2.2 Methods

2.2.1 Taxonomic and sampling schemes for a novel phylogeny of Procellariiformes

A novel phylogenetic study requires samples that are suitable for genetic work that span the diversity of the taxonomic group being studied to a predefined level (e.g. genus, species, subspecies, or population). The major taxonomic classification lists of birds, the IOC World Bird List v13.1 (Gill et al. 2023), The Clements Checklist of Birds of the World v2022 (Clements et al. 2022), Handbook of the Birds of the World and BirdLife International Taxonomic Checklist Version 7 (HBW and BirdLife International 2022), and The Howard and Moore Complete Checklist of the Birds of the World version 4.1 (Christidis et al. 2018) were compared for the Procellariiformes (Appendix 2.1). This provided a baseline taxonomy to target sampling of the order. A further review of literature on the taxonomy, systematics, and population genetics of Procellariiformes supplemented this to determine additional populations of potential interest that lacked a recognised taxonomic rank in the major classifications. This review was also informative in assessing areas of disagreement among the major classifications. Using this review and the checklist comparison, a list of Procellariiformes was formed as a guide for targeting sampling (Appendix 2.2). Although the existence of multiple taxonomic lists with different criteria and species concepts can be a hindrance (Garnett & Christidis 2017, Collar 2018), with a primary aim of this study to encompass a comprehensive range of diversity within Procellariiformes, no single taxonomic list was followed. Despite this, the list used was very similar to the IOC World Bird List v13.1 (Gill et al. 2023), differing only in the recognition of *Pachyptila desolata banksi*, *Pachyptila turtur eatoni*, *Pterodroma leucoptera caledonica* and *Ardenna pacifica chlororhyncha*.

Using the taxonomic list, the following sampling scheme was followed: 1. Aim to include at least two samples of each monotypic species; 2. Aim to include at least one sample of each subspecies of polytypic species; 3. Aim to include further samples of taxa of interest where

evidence exists of further unnamed diversity. Birds collected at known breeding locations were favoured. In taxa with wide breeding ranges, different breeding locations were favoured when sampling more than one individual.

2.2.2 Sample acquisition

Most samples were sourced from collections of museums and ornithological organisations, where the presence of voucher specimens was an important benefit. Sample loans were obtained from the following museums and organisations: 1. USA: American Museum of Natural History (AMNH), University of Kansas Biodiversity Institute and Natural History Museum (KU), Natural History Museum of Los Angeles County (LACM), Louisiana State University Museum of Natural Science (LSUMZ), National Zoological Park (Smithsonian Institution; NZP), San Diego Natural History Museum (SDNHM), University of Alaska Museum (UAM), University of Michigan Museum of Zoology (UMMZ), National Museum of Natural History (USNM), University of Washington Burke Museum (UWBM); 2. Europe: Museu de Ciències Naturals de Barcelona (MCNB), Natural History Museum (UK; NHMUK), Natural History Museum of Oslo (NHMO), Zoological Museum of the University of Copenhagen (ZMUC); 3. Australia: Australian Antarctic Division (AAD), Australian Museum (AM), National Museum of Victoria (NMV), South Australian Museum (SAM); 4. Japan: Yamashina Institute for Ornithology (YIO).

This was supplemented by samples of individual researchers with sample collections relating to geographic locations or taxonomic groups of interest. I collected samples during fieldwork in the Azores, Portugal. The following researchers contributed samples to this study: Alice Cibois and Jean-Claude Thibault (Natural History Museum of Geneva and Muséum national d'Histoire naturelle, Paris), Gary Nunn (formerly American Museum of Natural History), Hallvard Strøm (Norwegian Polar Institute), Heraldo Norambuena (Universidad Santo Tomás, Chile), Jacob González-Solís and Joan Ferrer Obiol (Universitat de Barcelona), Jeremy Austin (formerly Natural History Museum, UK), Kate Huyvaert (Washington State University), Kazuto Kawakami (Forestry and Forest Products Research Institute, Japan), Paul Scofield (Canterbury Museum, New Zealand), Petra Quillfeldt (University of Giessen, Germany), Richard Phillips (British Antarctic Survey), Theresa Burg (University of Lethbridge, Canada), Vincent Bretagnolle (Centre National de la Recherche Scientifique, France), Yuliana Bedolla (Grupo de Ecología y Conservación de Islas, Mexico), Yves Cherel (Centre d'Études Biologiques de Chizé, France).

All samples were collected and transported with the relevant permits. Samples from European Union (EU) member states were imported to the UK under a general authorisation as research and diagnostic samples. Samples from non-EU states were imported with specific import authorisations granted to Dr. Andreanna Welch by Animal and Plant Health Agency, UK.

2.2.3 Preparing samples for sequencing

Samples were prepared for DNA extraction with a protocol depending on their condition and medium. Blood samples suspended in ethanol or other buffers were centrifuged at maximum speed (13,200rpm) for two minutes and approximately 10µL of blood was transferred to nuclease-free 1.5mL Eppendorf tubes. To remove buffers, samples were centrifuged at maximum speed for four minutes and the supernatant was removed. 1mL of nuclease-free distilled water was added and mixed by pipetting, before centrifuging at maximum speed for two minutes and removing the supernatant. This process was repeated once further with a second wash of 1mL nuclease-free distilled water. Samples were left for 30 minutes to one hour in an incubator at 45°C with the lid open to evaporate any trace supernatant remaining. Dried blood samples and freeze-dried blood samples that had not been stored in any buffer were not processed in this way and proceeded immediately to lysis. For blood on Whatman Flinders Technology Associates (FTA) cards, a small subsample of approximately 5mm by 5mm was cut and transferred to a new tube for lysis. For tissue samples in ethanol and other buffers, subsamples of approximately 30mg were cut and transferred to new tubes. These samples were cut into smaller pieces using a scalpel and washed and heated to remove buffers following the same method described for blood samples. Dried muscle tissues and other tissue samples not stored in buffers were cut using a scalpel and proceeded directly to lysis. For feather samples, the calamus up to the superior umbilicus was removed and cut laterally to expose the inner surface for lysis.

Lysis and DNA extraction was carried out using E.Z.N.A. Tissue DNA Kits (Omega Bio-Tek) according to the manufacturer's protocol with minor modifications. Buffer names in this section follow names used in E.Z.N.A. Tissue DNA Kits. For lysis, 250µL of elution buffer (10mM Tris-HCl), 25µL Proteinase K, and 4µL RNase A (10mg/mL) were added to blood samples, vortexed, and left at room temperature for 2 minutes. Then 250µL BL buffer was added, before vortexing and incubation at 55°C overnight until they became clear. To tissue samples, 200µL TL buffer, 25µL proteinase K, and 4µL RNase A (10mg/mL) were added, vortexed, and left at room temperature for 2 minutes, and then incubated at 55°C overnight.

Feather samples were treated with the same conditions as tissue samples for lysis, except Proteinase K volume was increased to 45µL, 35µL 1M dithiothreitol (DTT) was added, and lysis was extended to 3 days following Bayard De Volo et al. (2008), in which these conditions produced higher DNA yields from feather samples.

Following lysis, samples with insoluble undigested material (blood on FTA cards, tissues, and feather samples) were centrifuged at maximum speed for 5 minutes. The supernatant was then transferred to new tubes and pelleted material was discarded. 220µL BL buffer was then added to these samples, they were vortexed and incubated at 70°C for 10 minutes. All forms of blood samples other than on FTA cards that did not leave undigested material after lysis did not undergo this step. 250µL 100% ethanol was added to samples before transfer to mini columns inserted in 2mL collection tubes. Samples were centrifuged at maximum speed for 1 minute and the filtrate was discarded. 500µL HBC buffer diluted with 100% isopropanol was added before centrifuging at maximum speed for 1 minute, and the filtrate was discarded. Mini columns were transferred to a new collection tube and washed twice with 700µL DNA Wash Buffer diluted with 100% ethanol, centrifuged and the filtrate was discarded. Empty tubes were centrifuged at maximum speed for 4 minutes to dry the mini columns, and mini columns were then transferred to nuclease-free 1.5mL Eppendorf tubes. DNA was eluted into a volume of 65µL elution buffer (10mM Tris-HCl), and a second elution step used the same volume of 65µL pipetted back into the mini column to maintain a final volume of 65µL.

Negative lysis and extraction controls that received no input sample material, containing only lysis and extraction reagents, were included with each individual batch of samples processed to check for contamination. Samples received as extracted DNA were incorporated with extracted samples before quality assessment. DNA quantity was assessed using a Qubit 3 Fluorometer (Invitrogen, Inc., Carlsbad, CA, USA) with a Qubit dsDNA high sensitivity assay kit (HS; 0.2-100ng/µL) according to the manufacturer's instructions. For samples that exceeded the range of the HS assay kit a broad range assay kit (2-1000ng/µL) was used to quantify DNA concentration. Sample quality was visualised on 1% agarose gels stained with 2µL 1µg/mL ethidium bromide.

2.2.4 DNA sequencing methods

Samples were normalised such that 35µL DNA in elution buffer contained at least 1µg DNA with a concentration range of 30-100ng/µL. Samples below this concentration threshold were concentrated using a SpeedVac to evaporate elution buffer. Samples with under 1µg of total

DNA were concentrated to under 35µL and made up to 35µL with elution buffer so that all extracted DNA could be used. Normalised DNA samples were loaded to semi-skirted 96-well plates (Thermo Scientific) and sealed with strip caps (Thermo Scientific).

Plates were sent to RAPiD Genomics (Gainesville, FL, USA) for library preparation, capture enrichment and sequencing of UCE loci following methods detailed in Faircloth et al. (2012). Indexed libraries were prepared and pooled prior to being hybridised with probe sets. High quality and moderately degraded samples were enriched with the Tetrapods-UCE-5Kv1 probe set (available at ultraconserved.org) of 5,472 probes targeting 5,060 UCE loci. Heavily degraded samples were enriched with the 'Toepads' UCE-5kv2 probe set, a recently designed UCE probe set for degraded DNA by Brant Faircloth, comprising 86,680 probes targeting 4,434 UCE loci that overlap with the Tetrapods-UCE-5Kv1 probe set. This probe set was designed with extensive tiling of overlapping probes to increase capture efficiency in degraded DNA samples with low fragment lengths. Enriched libraries were sequenced on an Illumina HiSeq 3000 platform generating 100bp or 250bp paired-end reads.

2.2.5 Bioinformatic processing of sequencing reads

Demultiplexed reads in FASTQ format were supplied by RAPiD Genomics. Reads were trimmed for adapter content and low-quality bases using Trimmomatic version 0.39 (Bolger et al. 2014). After trimming, quality per sample was assessed using FastQC version 0.11.9 (Andrews 2010). Trimmed reads were processed using the PHYLUCE pipeline version 1.7.1 (Faircloth 2016). First, trimmed reads were assembled into contigs using SPAdes version 3.15.4 (Bankevich et al. 2012) implemented by the PHYLUCE program phyluce_assembly_assemblo_spades. Contigs were matched to UCE probes in the Tetrapods-UCE-5Kv1 probe set and extracted using the PHYLUCE programs phyluce_assembly_match_contigs_to_probes, phyluce_assembly_get_match_counts, and phyluce_assembly_get_fastas_from_match_counts, to generate a monolithic FASTA file containing UCE-matched contigs for all samples. Individual UCE loci were aligned using MAFFT version 7.475 (Katoh & Standley 2013) implemented by the PHYLUCE program phyluce_align_seqcap_align. Aligned loci were internally trimmed using Gblocks version 0.91b (Castresana 2000) implemented by the PHYLUCE program phyluce_align_get_gblocks_trimmed_alignments_from_untrimmed. Locus names were removed from trimmed alignments using the PHYLUCE program phyluce_align_remove_locus_name_from_files. Only UCE loci with minimum 75% taxon

occupancy were retained using the PHYLUCE program phyluce_align_get_only_loci_with_min_taxa. UCE loci mapping to the Z or W chromosomes were removed (see Chapter 3) to retain only loci on autosomal chromosomes. Summary statistics of UCE loci were computed using the PHYLUCE program phyluce_align_get_align_summary_data. Retained loci were concatenated using the PHYLUCE program phyluce_align_concatenate_alignments to produce a final matrix (hereafter referred to as the 75% matrix).

2.2.6 Phylogenetic inference methods

The 75% matrix was partitioned by UCE locus and partitioning schemes were evaluated by corrected Akaike Information Criterion (AICc) scores using PartitionFinder version 2.1.1 (Lanfear et al. 2017) with the general time reversible (GTR) substitution model with gamma distribution for rate heterogeneity (+G) and the rcluster search algorithm with rcluster-max set to 5000 (Lanfear et al. 2014).

A maximum likelihood (ML) tree was constructed for the partitioned 75% matrix using the best scoring partitioning scheme found by PartitionFinder using RAxML-NG version 1.1.0 (Kozlov et al. 2019) with the GTR+G substitution model for all partitions and 60 ML searches with 30 random starting trees and 30 parsimony starting trees. To assess support for bipartitions present in the best scoring tree, 1000 non-parametric bootstrap replicates were conducted and mapped onto the tree using the RAxML-NG commands --bootstrap and --support. Nodes with bootstrap support values under 70% were collapsed to polytomies using TreeCollapserCL4 (Hodcroft 2021).

Trees were constructed using Bayesian inference (BI) on the partitioned 75% matrix using the multi-threaded MPI hybrid variant of ExaBayes version 1.5.1 (Aberer et al. 2014). Four independent runs were conducted, with random starting trees each with 4 million generations sampling the Markov chain Monte Carlo chains every 500 generations. The consense tool of ExaBayes was used to construct a 50% majority rule consensus tree from the combined 4 independent runs with the first 25% (1 million generations per run) discarded as burn-in. Nodes with posterior probabilities under 0.7 were collapsed to polytomies using TreeCollapserCL4 (Hodcroft 2021). Run parameters were assessed using Tracer version 1.7.2 (Rambaut et al. 2018). The relative Robinson-Foulds distance (Robinson & Foulds 1981) was calculated using the RAxML-NG command --rfdist (Kozlov et al. 2019) to assess similarity between topologies inferred by ML and BI.

To account for heterogeneous evolutionary histories of loci, trees were also built under the multispecies coalescent (MSC) model using SVDQuartets (Chifman & Kubatko 2014) with the concatenated 75% matrix. SVDQuartets uses single-site patterns to infer quartet trees for subsets of four tips and combines quartet trees into an overall tree. To facilitate the evaluation of all possible quartets, the alignment was separated by family (Diomedeidae, Oceanitidae, Hydrobatidae, and Procellariidae) and SVDQuartets was implemented in PAUP* version 4 (Swofford 2002). 100 bootstraps were conducted and mapped onto the tree. Nodes with support values under 70% were collapsed to polytomies using TreeCollapserCL4 (Hodcroft 2021).

2.2.7 Investigating areas of discordance in the Procellariiformes tree

Nodes above the species level with low support in partitioned concatenated ML or BI analyses were investigated in greater detail. For these conflicting nodes, ML bootstrap support, BI posterior probabilities, and MSC bootstrap support were compared for each topology identified in tree searches and each possible alternative species-level quartet. Site concordance factors (sCF) were calculated in IQ-TREE version 2.2.2.3 (Minh et al. 2020a, Minh et al. 2020b) using the concatenated 75% matrix for each possible quartet topology at conflicting nodes. The updated version of sCF based on ML was used with the --scfl option in IQ-TREE to compute 100 replicates (Mo et al. 2023). An alternative measure of nodal support to traditional bootstrapping, sCF is defined as the percentage of decisive sites in an alignment supporting a branch in a tree. As sCF compares the three possible topologies of quartets, values greater than one third indicate support for a particular quartet topology.

To further visualise conflicting signal in the dataset and potentially reticulate evolutionary patterns, phylogenetic networks were constructed for these nodes of interest, and for all taxa. The 75% concatenated matrix was separated by genus and trimmed to retain only parsimony informative sites (PIS) using ClipKIT version 1.4.0 (Steenwyk et al. 2020). Genera with fewer than two sampled taxa or with fewer than four total individuals sampled (*Phoebetria*, *Garrodia*, *Nesofregetta*, *Hydrobates*, *Pagodroma*, *Daption*, *Thalassoica*, *Bulweria*, *Halobaena*, *Aphrodroma*) were grouped with their closest related genus or genera, as inferred from ML and BI trees. Phylogenetic networks were constructed from these trimmed PIS alignments using the NeighborNet algorithm (Bryant & Moulton 2004) implemented in SplitsTree version 5.3.0 (Huson & Bryant 2006).

Finally, information from the ML and BI trees, the MSC tree, phylogenetic networks, and support metrics was used to summarise a consensus topology of Procellariiformes to the subspecies level.

2.3 Results

2.3.1 Sequence data and UCE loci summary statistics

Sequence data was obtained for 335 samples representing 129 species, 166 taxa including subspecies, and 3 outgroup samples (Appendix 2.3). This accounted for 89% of species and 85% of all taxa including subspecies on the Procellariiformes taxonomic list. A mean of 1,410,621 (95% CI: ±158,983) reads were sequenced for each sample. After trimming and assembling reads into contigs and extracting UCE loci, a mean of 4,755.44 (95% CI: ±13.34) UCE loci were recovered for each sample. After aligning these loci and filtering out sex-linked loci (see Chapter 3) and loci with under 75% taxon occupancy, 4,335 UCE loci were retained. This final set of loci contained a minimum of 252 samples in any locus and a mean of 328.63 (95% CI: ±0.39) samples per locus. The mean length per locus was 604.54bp (95% CI: ±4.33). The mean missing data per locus was 2.26% (95% CI: ±0.03). The mean number of PIS per locus was 85.36 (95% CI: ±1.62). The concatenated 75% matrix of 4,335 loci comprised 2,620,685 sites and 370,035 PIS. The total nucleotide content of the 75% matrix was 96.34%, with 3.52% missing data, and 0.14% gaps.

2.3.2 Higher-level relationships of the Procellariiformes tree using ML and BI methods

The best scoring partitioning scheme found by PartitionFinder had 2,273 subsets and was used to construct partitioned phylogenetic trees from the concatenated 75% matrix of 4,335 UCE loci using ML and BI. Topologies inferred by ML and BI using RAxML-NG and ExaBayes respectively were very similar (relative Robinson-Foulds distance 0.027), mainly differing at the shallowest nodes of populations within taxa. Topologies were well resolved at all levels with only minor areas of ambiguity in almost all clades and very few areas of phylogenetic conflict within the dataset identified. Both methods recovered the same branching structure of families and genera (Fig. 2.1). Diomedeidae was sister to all other families. The two storm-petrel families Oceanitidae and Hydrobatidae were paraphyletic, and sister to the Procellariidae. Within Diomedeidae, two sister clades of *Diomedea* and *Phoebastria*, and *Thalassarche* and *Phoebetria* were inferred. *Oceanites* was sister to all other genera of

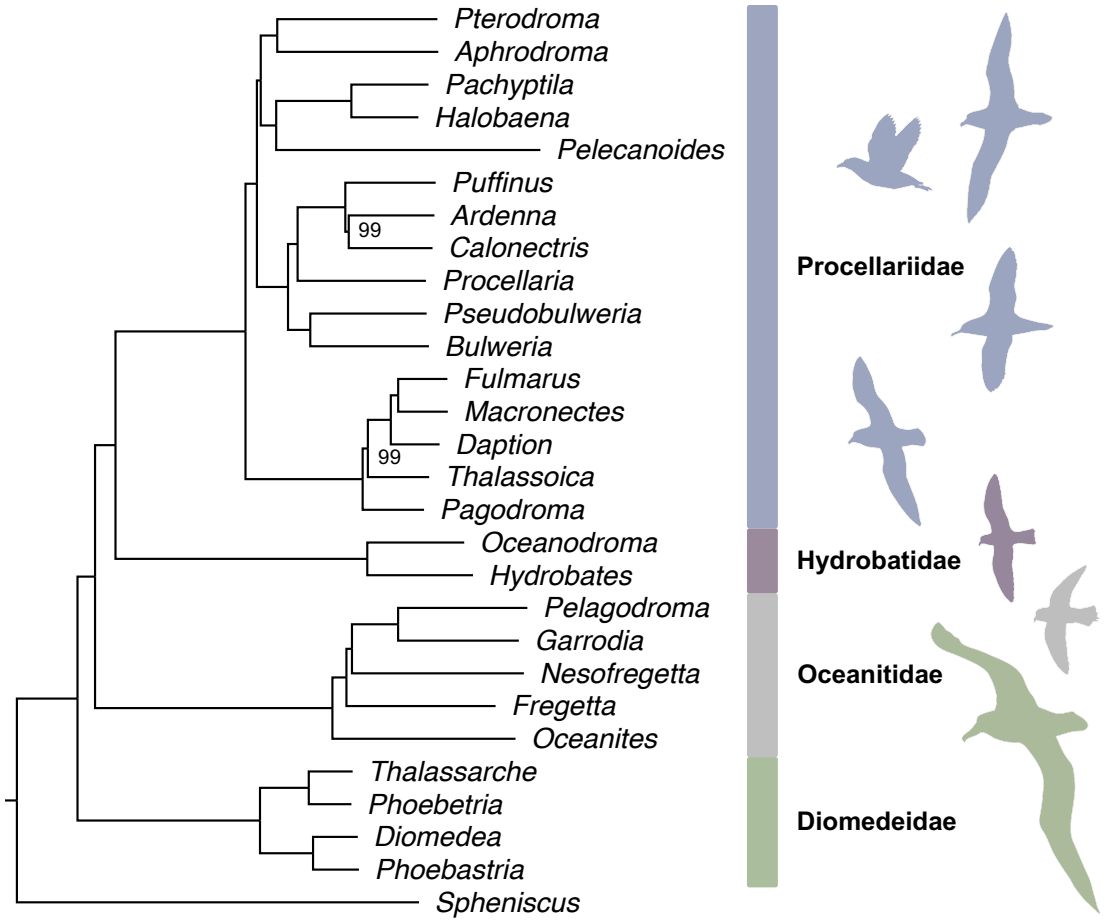


Figure 2.1 Relationships of families and genera of Procellariiformes inferred by all methods. Numbers indicate ML bootstrap values. All nodes without support values indicated were fully supported by ML bootstraps and BI posterior probabilities. Branch lengths are approximated on the full taxon ML tree produced using RAxML-NG. Families are indicated by coloured bars and illustrations show species representative of each family.

Oceanitidae, with a linear branching pattern of *Fregetta* and *Nesofregetta*, and a sister relationship between *Pelagodroma* and *Garrodia*. The two genera of the family Hydrobatidae, *Hydrobates* and *Oceanodroma*, were not monophyletic. Procellariidae included the fulmarine genera of *Pagodroma*, *Thalassoica*, *Daption*, *Macronectes* and *Fulmarus* in a linear branching order as sister to all other genera of the family. A further clade comprised the sister genera *Bulweria* and *Pseudobulweria*, and *Procellaria* as the sister to the shearwaters, within which *Ardenna* and *Calonectris* were sister genera alongside *Puffinus*. The final clade within Procellariidae contained the gadfly petrel genus *Pterodroma* as sister to *Aphrodroma*, and a further monophyletic clade comprising the diving-petrel genus *Pelecanoides* sister to the pair of genera *Pachyptila* and *Halobaena*.

2.3.3 Finer-scale relationships of the Procellariiformes tree using ML and BI methods

ML and BI produced very similar topologies at shallower nodes. Almost all nodes were fully supported by BI posterior probabilities. Bootstrap support values were mostly fully supported, and lower than BI posterior probabilities in some cases. The following sections provide results within families, where support values are mentioned only for nodes that were not fully supported.

2.3.3.1 Diomedeidae

Despite the short branches throughout the family Diomedeidae relative to other families, relationships were almost entirely well resolved and strongly supported (Fig. 2.2). In the genus *Phoebastria*, *P. irrorata* was sister to *P. albatrus* and the sister species *P. nigripes* and *P. immutabilis*.

Within *Diomedea*, *D. epomophora* and *D. sanfordi* were sister to the wandering albatross clade of *D. amsterdamensis*, *D. dabbenena*, *D. exulans* and *D. antipodensis*. In this clade, *D. amsterdamensis* formed the outgroup in ML and BI analyses, with *D. dabbenena*, *D. exulans* and *D. antipodensis* forming a monophyletic clade with 92% bootstrap support and full posterior probability support. Monophyly of two samples of *Diomedea exulans* each from South Georgia (South Atlantic Ocean) and the Kerguelen Islands (Indian Ocean) was found by both ML and BI methods but with low support. ML and BI inferred the same relationships within the species *D. antipodensis* of three sister pairs of individuals, including one paraphyletic pair of the subspecies *D. antipodensis antipodensis* and *D. antipodensis gibsoni*. Nodes within this species were not fully supported in either BI or ML.


T. melanophris 2
0.75
T. melanophris 6
T. melanophris 5
T. melanophris 4
T. melanophris 3
T. melanophris 1
T. impavida 1
87
T. impavida 2
T. impavida 3
T. chrysostoma 1
T. chrysostoma 2
T. cauta cauta 1
T. cauta cauta 2
T. cauta steadi 1
T. cauta steadi 2
T. cauta steadi 3
T. salvini 1
T. salvini 2
T. bulleri bulleri 1
T. bulleri platei 1
T. carteri 1
T. carteri 2
T. chlororhynchos 1
99
T. chlororhynchos 2
Ph. fusca 1
Ph. fusca 2
Ph. palpebrata 1
D. antipodensis antipodensis 1
90
D. antipodensis antipodensis 3
0.88
D. antipodensis antipodensis 2
0.88
*
D. antipodensis gibsoni 1
D. antipodensis gibsoni 2
D. antipodensis gibsoni 3
D. exulans 1
92
D. exulans 2
D. exulans 3
D. exulans 4
D. dabbenena 1
D. dabbenena 2
D. amsterdamensis 1
D. amsterdamensis 2
D. amsterdamensis 3
D. sanfordi 1
D. epomophora 1
P. immutabilis 1
P. immutabilis 2
P. nigripes 1
P. nigripes 2
P. albatrus 1
P. albatrus 2
*
P. irrorata 1
P. irrorata 2
P. irrorata 3
Diomedeidae
0.004

Figure 2.2 ML tree of Diomedeidae based on the partitioned concatenated matrix of 4,335 UCE loci. ML bootstrap support values are given as integers and BI posterior probabilities are given as decimals. Only nodes with bootstrap values <100 or posterior probabilities <1 are indicated. Nodes with support values <70/0.7 are collapsed. Differing relationships inferred by BI are overlaid in red, where an asterisk indicates that ML inferred the same branching pattern but is collapsed due to <70 bootstrap support. The full ML topology of Procellariiformes showing branch lengths between clades is shown (left), with Diomedeidae highlighted in black. Names of genera are abbreviated as follows: *P. Phoebastria*, *D. Diomedea*, *Ph. Phoebetria*, *T. Thalassarche*.

Phoebetria palpebarata and *P. fusca* were sister species, and placed sister to the genus *Thalassarche*. The deepest divergence within *Thalassarche* separated the species pair *T. carteri* and *T. chlororhynchos* from the rest of the genus. *T. bulleri*, *T. salvini*, and *T. cauta* formed a monophyletic clade separated from *T. chrysostoma* and the sister species *T. impavida* and *T. melanophris*. The subspecies *T. cauta cauta* and *T. cauta steadi* were not monophyletic and no strongly supported relationships between individuals were inferred by either ML or BI. Similarly, individuals of *T. melanophris* did not show strongly supported structuring, where BI identified one sister pair of individuals receiving a 0.75 posterior probability value.

2.3.3.2 Oceanitidae

The relationships within Oceanitidae were concordant at all levels between ML and BI topologies and well supported throughout (Fig. 2.3). In the genus *Oceanites*, *O. oceanicus* was paraphyletic with respect to *O. gracilis*. The subspecies *O. oceanicus chilensis* was fully supported in all analyses as the sister to *O. gracilis* rather than to the other subspecies of *O. oceanicus*. The subspecies *O. oceanicus oceanicus* and *O. oceanicus exasperatus* showed strongly supported structure, but not by subspecies identity. An *O. oceanicus oceanicus* individual from the Kerguelen Islands (Indian Ocean) was the deepest divergence from an *O. oceanicus oceanicus* individual from the Falkland Islands and *O. oceanicus exasperatus* individuals from the Antarctic Peninsula and South Shetland Islands.

The genus *Fregetta* diverged between the two species *F. grallaria* and *F. tropica*, with the exception of the paraphyletic subspecies *F. grallaria leucogaster*. Of two individuals of this subspecies collected from St. Paul Island (French Southern and Antarctic Lands, Indian Ocean), one was sister to another individual of *F. grallaria leucogaster* from Gough Island and

the other clustered with all individuals of *F. tropica*. The subspecies *F. grallaria segethi* was sister to all other *F. grallaria* subspecies, including *F. grallaria leucogaster* and the sister pair *F. grallaria grallaria* and *F. grallaria titan*. The species *F. tropica* formed pairs with 90% bootstrap support and 0.75 posterior probabilities, although one of these pairs united two degraded samples with long branches that may have been artefactual.

Nesofregetta was sister to the sister genera *Garrodia* and *Pelagodroma*. Relationships between the subspecies of *Pelagodroma* were fully supported with *P. marina marina* and *P. marina maoriana* sister to the subspecies *P. marina dulciae*, *P. marina hypoleuca* and *P. marina eadesorum*.

2.3.3.3 Hydrobatidae

Relationships within the family Hydrobatidae were almost entirely concordant between ML and BI topologies (Fig. 2.3). With the exception of one node above the species-level with under 100% bootstrap support, the family was well resolved and supported. Hydrobatidae was split into two major clades, largely conforming to a divergence between Atlantic Ocean and Pacific Ocean breeding species. In the majority Atlantic Ocean breeding clade, *Hydrobates pelagicus* was sister to *O. furcatus*, the only species breeding only in the Pacific Ocean of this clade. Within this clade, this species pair was deeply diverged from the species *O. jabejabe*, *O. monteiroi* and *O. castro*, the latter species breeding in both the Atlantic and Pacific Oceans.

Of the larger Hydrobatidae clade, all species were restricted to breeding in the Pacific Ocean with the exception of *O. leucorhous* which breeds in both the Pacific Ocean and Atlantic Ocean. Within this clade, three subclades were inferred. *O. melania*, *O. microsoma* and *O. tethys* formed the deepest diverging subclade. The monophyly of subspecies *O. tethys tethys* and *O. tethys kelsalli* was fully supported in the BI tree, but not in the ML tree (67% bootstrap support). *O. matsudairae* and *O. monorhis* were sister species alongside *O. tristrami* in a further subclade. This subclade was sister to a subclade comprising *O. homochroa*, *O. hornbyi*, *O. markhami*, *O. leucorhous*, *O. cheimomnestes* and *O. socorroensis*. The branching pattern of this subclade with *O. homochroa* as sister to all other species was inferred with 89% bootstrap support in the ML tree and full support in the BI tree. The subspecies *O. leucorhous leucorhous* and *O. leucorhous chapmani* were monophyletic. The monophyly of *O. cheimomnestes* with respect to *O. socorroensis* was fully supported in BI analyses and 90% supported in ML analyses.

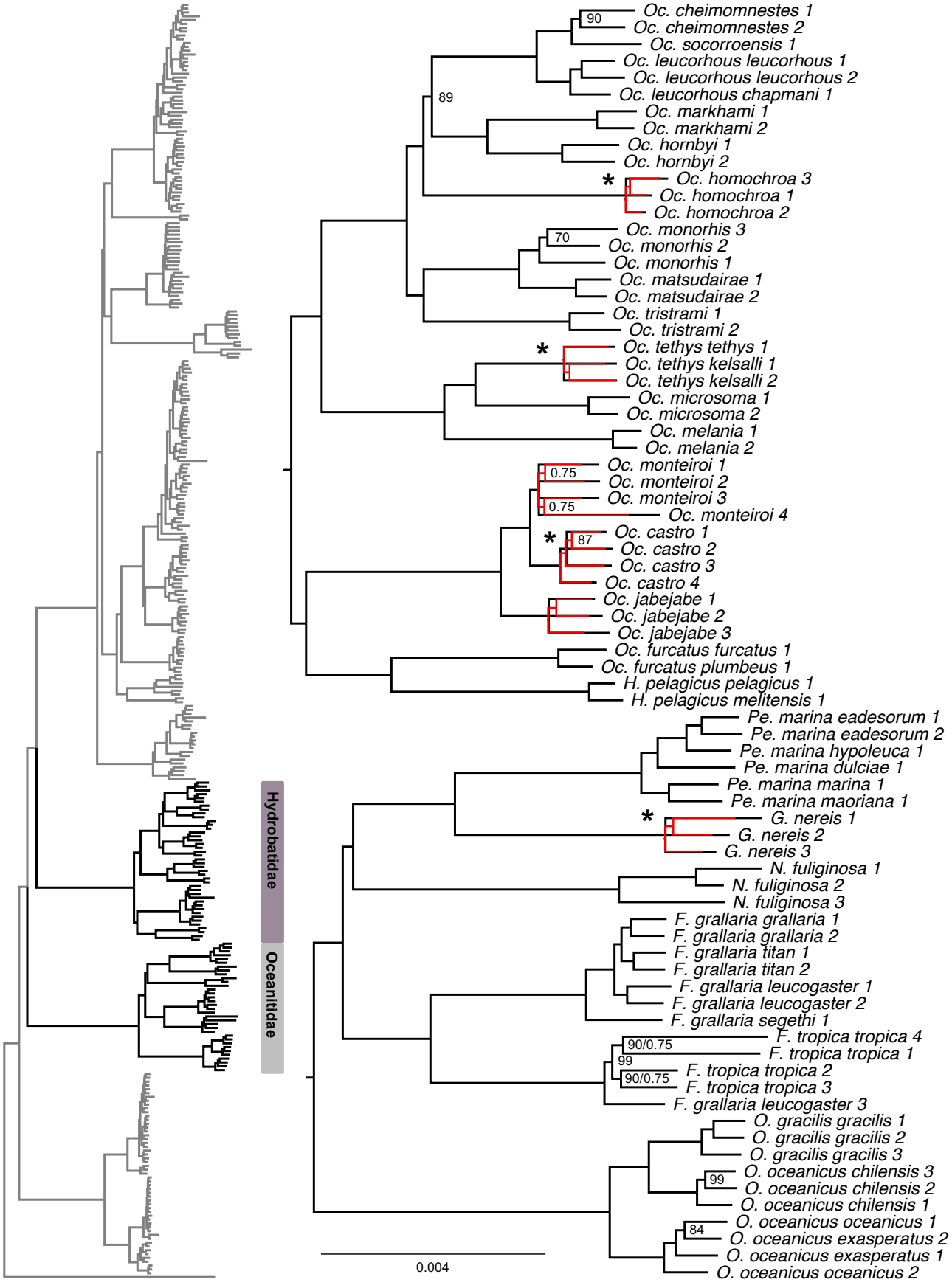

Oc. cheimomnestes 1
Oc. cheimomnestes 2
Oc. socorroensis 1
Oc. leucorhous leucorhous 1
Oc. leucorhous leucorhous 2
Oc. leucorhous chapmani 1
Oc. markhami 1
Oc. markhami 2
Oc. hornbyi 1
Oc. hornbyi 2
Oc. homochroa 3
Oc. homochroa 1
Oc. homochroa 2
Oc. monorhis 3
Oc. monorhis 2
Oc. monorhis 1
Oc. matsudairae 1
Oc. matsudairae 2
Oc. tristrami 1
Oc. tristrami 2
Oc. tethys tethys 1
Oc. tethys kelsalli 1
Oc. tethys kelsalli 2
Oc. microsoma 1
Oc. microsoma 2
Oc. melania 1
Oc. melania 2
Oc. monteiroi 1
Oc. monteiroi 2
Oc. monteiroi 3
Oc. monteiroi 4
Oc. castro 1
Oc. castro 2
Oc. castro 3
Oc. castro 4
Oc. jabejabe 1
Oc. jabejabe 2
Oc. jabejabe 3
Oc. furcatus furcatus 1
Oc. furcatus plumbeus 1
H. pelagicus pelagicus 1
H. pelagicus melitensis 1
Pe. marina eadesorum 1
Pe. marina eadesorum 2
Pe. marina hypoleuca 1
Pe. marina dulciae 1
Pe. marina marina 1
Pe. marina maoriana 1
G. nereis 1
G. nereis 2
G. nereis 3
N. fuliginosa 1
N. fuliginosa 2
N. fuliginosa 3
F. grallaria grallaria 1
F. grallaria grallaria 2
F. grallaria titan 1
F. grallaria titan 2
F. grallaria leucogaster 1
F. grallaria leucogaster 2
F. grallaria segethi 1
F. tropica tropica 4
F. tropica tropica 1
F. tropica tropica 2
F. tropica tropica 3
F. grallaria leucogaster 3
O. gracilis gracilis 1
O. gracilis gracilis 2
O. gracilis gracilis 3
O. oceanicus chilensis 3
O. oceanicus chilensis 2
O. oceanicus chilensis 1
O. oceanicus oceanicus 1
O. oceanicus exasperatus 2
O. oceanicus exasperatus 1
O. oceanicus oceanicus 2
90
89
70
0.75
0.75
87
90/0.75
99
90/0.75
99
84
0.004
Hydrobatidae
Oceanitidae

Figure 2.3 ML tree of Oceanitidae and Hydrobatidae based on the partitioned concatenated matrix of 4,335 UCE loci. ML bootstrap support values are given as integers and BI posterior probabilities are given as decimals. Only nodes with bootstrap values <100 or posterior probabilities <1 are indicated. Nodes with support values <70/0.7 are collapsed. Differing relationships inferred by BI are overlaid in red, where an asterisk indicates that ML inferred the same branching pattern but is collapsed due to <70 bootstrap support. The full ML topology of Procellariiformes showing branch lengths between clades is shown (left), with Oceanitidae and Hydrobatidae highlighted in black. Names of genera are abbreviated as follows: *O. Oceanites*, *F. Fregetta*, *N. Nesofregetta*, *G. Garrodia*, *Pe. Pelagodroma*, *H. Hydrobates*, *Oc. Oceanodroma*.

2.3.3.4 Procellariidae

ML and BI produced almost identical topologies of the family Procellariidae with a single difference at one node within the genus *Pachyptila* (Fig. 2.4). ML produced lower support values at some shallow nodes leading to more collapsed polytomies than in the BI tree.

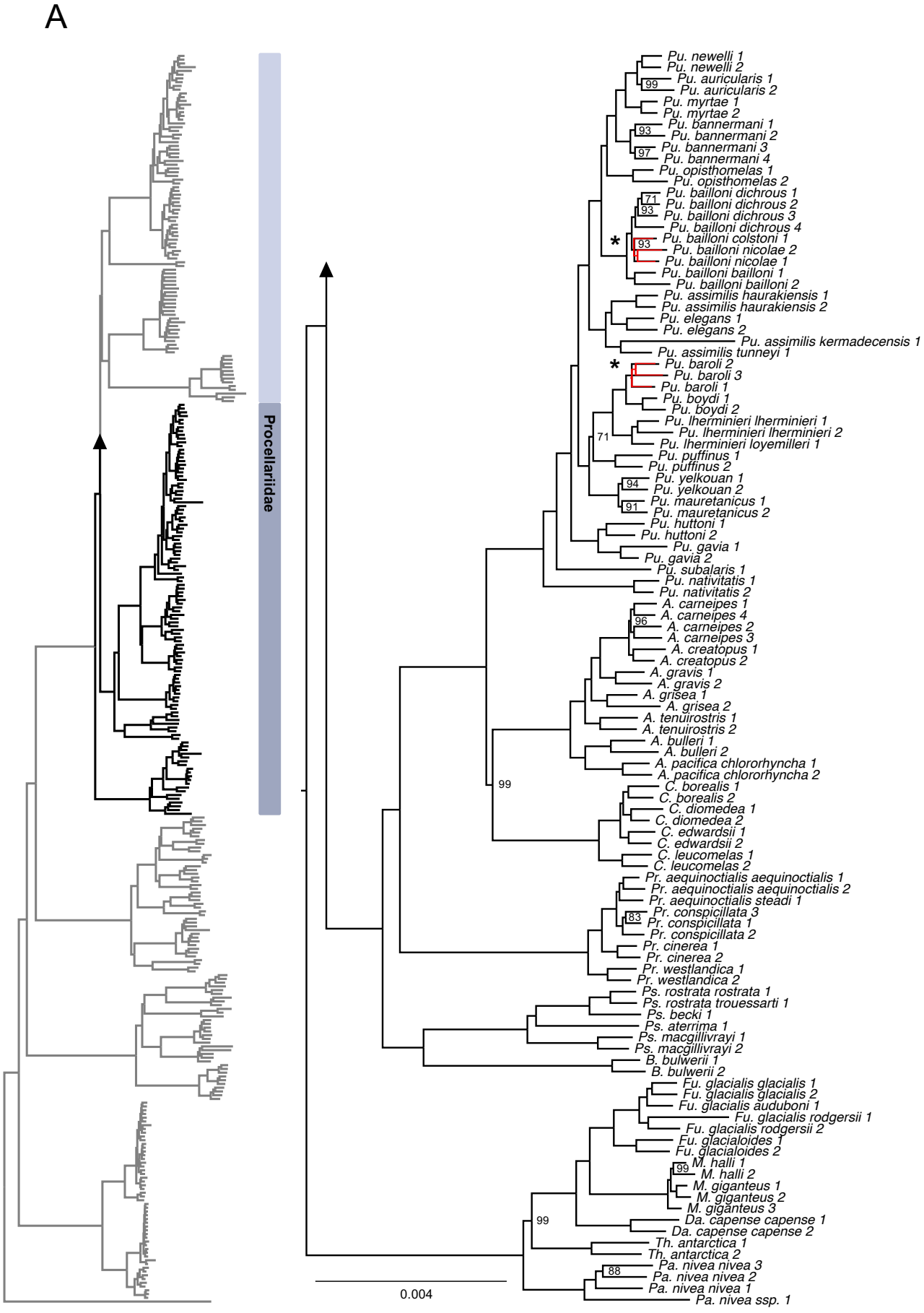
The genus *Pagodroma* was sister to all other fulmarine species. An Antarctic individual of *P. nivea* of unknown subspecies from a mixed colony of small birds (*P. nivea nivea*) and large birds (*P. nivea major*) where hybridisation occurs and there is a size continuum (Jouventin & Viot 1985) was the deepest divergence from three samples of *P. nivea nivea* from the South Atlantic.

In the giant-petrel genus *Macronectes*, two individuals of *M. giganteus* from Gough Island were sister to two individuals of *M. halli* from South Georgia and New Zealand. An additional individual of *M. giganteus* from South Georgia was sister to these pairs.

In the genus *Fulmarus*, *F. glacialoides* and *F. glacialis* were monophyletic sister species. *F. glacialis* showed the deepest segregation between Pacific Ocean (*F. glacialis rodgersii*) and Atlantic Ocean subspecies (*F. glacialis glacialis* and *F. glacialis auduboni*).

Pseudobulweria presented a linear branching pattern of *P. macgillivrayi*, *P. aterrima*, *P. becki* and *P. rostrata*.

Procellaria species diverged in a linear order of *P. westlandica*, *P. cinerea*, *P. conspicillata* and *P. aequinoctialis*. Although only one individual of the subspecies *P. aequinoctialis steadi* was sampled (from New Zealand), it was monophyletic with respect to two individuals of *P. aequinoctialis aequinoctialis* (from the South Atlantic Ocean).


A
Procellariidae
Pu. newelli 1
Pu. newelli 2
Pu. auricularis 1
Pu. auricularis 2
Pu. myrtae 1
Pu. myrtae 2
Pu. bannermani 1
Pu. bannermani 2
Pu. bannermani 3
Pu. bannermani 4
Pu. opisthomelas 1
Pu. opisthomelas 2
Pu. bailloni dichrous 1
Pu. bailloni dichrous 2
Pu. bailloni dichrous 3
Pu. bailloni dichrous 4
Pu. bailloni colstoni 1
Pu. bailloni nicolae 2
Pu. bailloni nicolae 1
Pu. bailloni bailloni 1
Pu. bailloni bailloni 2
Pu. assimilis haurakiensis 1
Pu. assimilis haurakiensis 2
Pu. elegans 1
Pu. elegans 2
Pu. assimilis kermadecensis 1
Pu. assimilis tunneyi 1
Pu. baroli 2
Pu. baroli 3
Pu. baroli 1
Pu. boydi 1
Pu. boydi 2
Pu. lherminieri lherminieri 1
Pu. lherminieri lherminieri 2
Pu. lherminieri loyemilleri 1
Pu. puffinus 1
Pu. puffinus 2
Pu. yelkouan 1
Pu. yelkouan 2
Pu. mauretanicus 1
Pu. mauretanicus 2
Pu. huttoni 1
Pu. huttoni 2
Pu. gavia 1
Pu. gavia 2
Pu. subalaris 1
Pu. nativitatis 1
Pu. nativitatis 2
A. carneipes 1
A. carneipes 4
A. carneipes 2
A. carneipes 3
A. creatopus 1
A. creatopus 2
A. gravis 1
A. gravis 2
A. grisea 1
A. grisea 2
A. tenuirostris 1
A. tenuirostris 2
A. bulleri 1
A. bulleri 2
A. pacifica chlororhyncha 1
A. pacifica chlororhyncha 2
C. borealis 1
C. borealis 2
C. diomedea 1
C. diomedea 2
C. edwardsii 1
C. edwardsii 2
C. leucomelas 1
C. leucomelas 2
Pr. aequinoctialis aequinoctialis 1
Pr. aequinoctialis aequinoctialis 2
Pr. aequinoctialis steadi 1
Pr. conspicillata 3
Pr. conspicillata 1
Pr. conspicillata 2
Pr. cinerea 1
Pr. cinerea 2
Pr. westlandica 1
Pr. westlandica 2
Ps. rostrata rostrata 1
Ps. rostrata trouessarti 1
Ps. becki 1
Ps. aterrima 1
Ps. macgillivrayi 1
Ps. macgillivrayi 2
B. bulwerii 1
B. bulwerii 2
Fu. glacialis glacialis 1
Fu. glacialis glacialis 2
Fu. glacialis auduboni 1
Fu. glacialis rodgersii 1
Fu. glacialis rodgersii 2
Fu. glacialoides 1
Fu. glacialoides 2
M. halli 1
M. halli 2
M. giganteus 1
M. giganteus 2
M. giganteus 3
Da. capense capense 1
Da. capense capense 2
Th. antarctica 1
Th. antarctica 2
Pa. nivea nivea 3
Pa. nivea nivea 2
Pa. nivea nivea 1
Pa. nivea ssp. 1
0.004

B

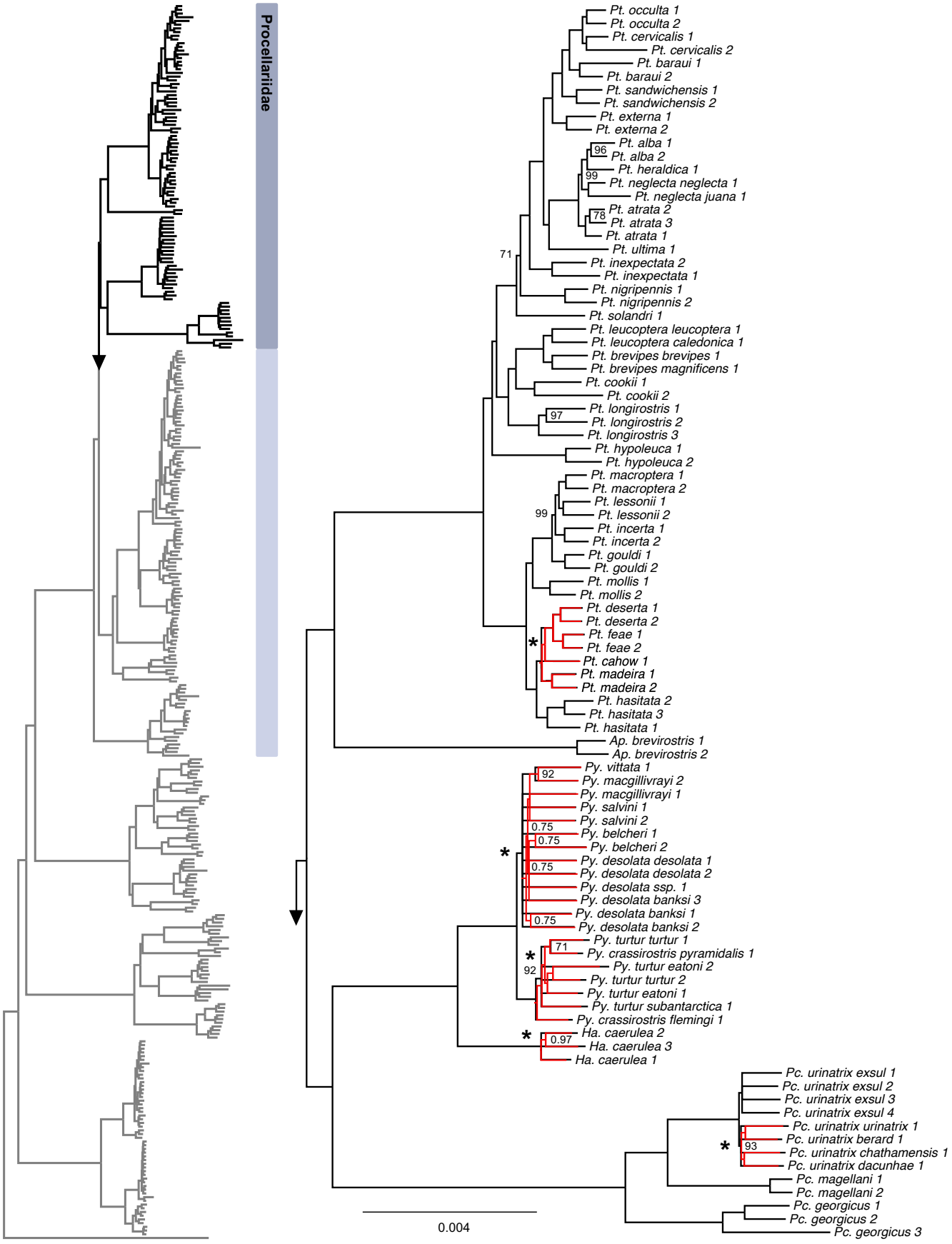

Procellariidae
Pt. occulta 1
Pt. occulta 2
Pt. cervicalis 1
Pt. cervicalis 2
Pt. baraui 1
Pt. baraui 2
Pt. sandwichensis 1
Pt. sandwichensis 2
Pt. externa 1
Pt. externa 2
Pt. alba 1
96
Pt. alba 2
Pt. heraldica 1
99
Pt. neglecta neglecta 1
Pt. neglecta juana 1
Pt. atrata 2
78
Pt. atrata 3
Pt. atrata 1
Pt. ultima 1
71
Pt. inexpectata 2
Pt. inexpectata 1
Pt. nigripennis 1
Pt. nigripennis 2
Pt. solandri 1
Pt. leucoptera leucoptera 1
Pt. leucoptera caledonica 1
Pt. brevipes brevipes 1
Pt. brevipes magnificens 1
Pt. cookii 1
Pt. cookii 2
Pt. longirostris 1
97
Pt. longirostris 2
Pt. longirostris 3
Pt. hypoleuca 1
Pt. hypoleuca 2
Pt. macroptera 1
Pt. macroptera 2
Pt. lessonii 1
99
Pt. lessonii 2
Pt. incerta 1
Pt. incerta 2
Pt. gouldi 1
Pt. gouldi 2
Pt. mollis 1
Pt. mollis 2
Pt. deserta 1
Pt. deserta 2
*
Pt. feae 1
Pt. feae 2
Pt. cahow 1
Pt. madeira 1
Pt. madeira 2
Pt. hasitata 2
Pt. hasitata 3
Pt. hasitata 1
Ap. brevirostris 1
Ap. brevirostris 2
Py. vittata 1
92
Py. macgillivrayi 2
Py. macgillivrayi 1
Py. salvini 1
Py. salvini 2
0.75
Py. belcheri 1
0.75
Py. belcheri 2
*
Py. desolata desolata 1
0.75
Py. desolata desolata 2
Py. desolata ssp. 1
Py. desolata banksi 3
Py. desolata banksi 1
0.75
Py. desolata banksi 2
Py. turtur turtur 1
71
Py. crassirostris pyramidalis 1
*
Py. turtur eatoni 2
92
Py. turtur turtur 2
Py. turtur eatoni 1
Py. turtur subantarctica 1
Py. crassirostris flemingi 1
*
Ha. caerulea 2
0.97
Ha. caerulea 3
Ha. caerulea 1
Pc. urinatrix exsul 1
Pc. urinatrix exsul 2
Pc. urinatrix exsul 3
Pc. urinatrix exsul 4
Pc. urinatrix urinatrix 1
*
Pc. urinatrix berard 1
93
Pc. urinatrix chathamensis 1
Pc. urinatrix dacunhae 1
Pc. magellani 1
Pc. magellani 2
Pc. georgicus 1
Pc. georgicus 2
Pc. georgicus 3
0.004

Figure 2.4 ML tree of the Procellariidae based on the partitioned concatenated matrix of 4,335 UCE loci. Part A shows the genera *Pagodroma*, *Thalassoica*, *Daption*, *Macronectes*, *Fulmarus*, *Bulweria*, *Pseudobulweria*, *Procellaria*, *Calonectris*, *Ardenna*, and *Puffinus*; part B shows the genera *Pelecanoides*, *Halobaena*, *Pachyptila*, *Aphrodroma*, and *Pterodroma*. ML bootstrap support values are given as integers and BI posterior probabilities are given as decimals. Only nodes with bootstrap values <100 or posterior probabilities <1 are indicated. Nodes with support values <70/0.7 are collapsed. Differing relationships inferred by BI are overlaid in red, where an asterisk indicates that ML inferred the same branching pattern but is collapsed due to <70 bootstrap support. The full ML topology of Procellariiformes showing branch lengths between clades is shown (left), with the relevant Procellariidae clade shown highlighted in black. Arrows indicate the connection between the two trees in parts A and B. Names of genera are abbreviated as follows: *Pa. Pagodroma*, *Th. Thalassoica*, *Da. Daption*, *M. Macronectes*, *Fu. Fulmarus*, *B. Bulweria*, *Ps. Pseudobulweria*, *Pr. Procellaria*, *C. Calonectris*, *A. Ardenna*, *Pu. Puffinus*, *Pc. Pelecanoides*, *Ha. Halobaena*, *Py. Pachyptila*, *Ap. Aphrodroma*, *Pt. Pterodroma*.

The genus *Calonectris* showed a deepest divergence of the Pacific Ocean breeding *C. leucomelas* from the North Atlantic Ocean breeding species *C. edwardsii* and the sister pair *C. diomedea* and *C. borealis*.

Within the genus *Ardenna*, *A. bulleri* and *A. pacifica* were sister species and formed the base of the genus. A linear branching pattern followed of *A. tenuirostris*, *A. grisea*, *A. gravis* and the sister pair *A. creatopus* and *A. carneipes*. The latter two species were separated by a very short branch. Species monophyly of *A. carneipes* was fully supported in BI and received a 96% bootstrap value in ML analyses. Within this species, sister pairs of two individuals from the Pacific Ocean and two individuals from the Indian Ocean were most closely related.

Puffinus nativitatis formed the base of the speciose genus *Puffinus*, followed by *P. subalaris*. *P. gavia* and *P. huttoni* were monophyletic and sister to the four further clades of the genus.

The first clade comprised the Caribbean, North Atlantic and Mediterranean species *P. yelkouan*, *P. mauretanicus*, *P. puffinus*, *P. lherminieri*, *P. boydi* and *P. baroli*. The species *P. yelkouan* and *P. mauretanicus* were monophyletic with 94% and 91% bootstrap support respectively. The position of *P. puffinus* as sister to *P. lherminieri*, *P. boydi* and *P. baroli* was supported by 71% bootstrap value in ML and fully supported in BI analyses.

The second clade within *Puffinus* contained the species *P. assimilis* of Australia and New Zealand and *P. elegans* of the subantarctic Atlantic, Indian, and Pacific Oceans as sister to further clades. *P. elegans* was embedded within *P. assimilis* and sister to *P. assimilis haurakiensis*, and a further subspecies pair of *P. assimilis kermadecensis* and *P. assimilis tunneyi*.

Following this, *P. bailloni* of tropical latitudes of the Indian and Pacific Oceans formed a clade. Within this species, *P. bailloni bailloni* was sister to a subclade containing *P. bailloni nicolae*, *P. bailloni colstoni* and *P. bailloni dichrous*. The monophyly of *P. bailloni nicolae* with respect to *P. bailloni colstoni* was fully supported in the BI trees and 42% supported in the ML tree, and monophyly of this trio of individuals of these two subspecies was also not fully supported in ML analysis (93%).

P. bailloni was sister to a further clade comprising a linear branching order of *P. opisthomelas*, *P. bannermani*, *P. myrtae*, and the sister pair *P. auricularis* and *P. newelli*. The monophyly of *P. auricularus* was fully supported in the BI tree and 99% supported in the ML tree.

The genus *Pelecanoide*s was sister to the genera *Halobaena*, *Pachyptila*, *Aphrodroma*, and *Pterodroma* and separated by a long branch relative to other genera. Within *Pelecanoides*, *P. georgicus* was sister to *P. magellani* and *P. urinatrix*. The subspecies *P. urinatrix exsul* was monophyletic with no strongly supported structuring between individuals. In both the ML and BI trees the subspecies *P. urinatrix urinatrix* and *P. urinatrix berard*, and *P. urinatrix dacunhae* and *P. urinatrix chathamensis* formed sister pairs, although with low support in ML analyses (50% and 53% respectively), and the entire clade of four subspecies received 93% bootstrap support.

In contrast to all other groups within the order Procellariiformes, the genus *Pachyptila* was highly atypical in presenting poorly resolved relationships. No species were found to be monophyletic with full support in either ML or BI topologies, and populations of single taxa with more than one sampled individual did not form sister relationships in several cases. Despite this, two deep-level clades were fully supported: *P. turtur* and *P crassirostris*, and *P. desolata*, *P. belcheri*, *P. vittata*, *P. macgillivayi* and *P. salvini*. Within the *P. turtur* and *P. crassirostris* clade, the ML tree inferred the same topology as the fully supported BI tree, albeit with poor support at most nodes. An individual of *P. crassirostris flemingi* from Heard Island in the Indian Ocean was the deepest divergence, followed by *P. turtur subantarctica* from New Zealand. Two samples of *P. turtur eatoni* from St. Paul Island (French Southern and Antarctic Lands, Indian Ocean) were not monophyletic, but were monophyletic with one individual of *P. turtur turtur* from mainland New Zealand. These individuals were sister to a pair of

individuals of *P. turtur turtur* and *P. crassirostris pyramidalis* both from the Chatham Islands, New Zealand.

The larger clade of *Pachyptila* showed further structure, dividing into subclades containing *P. desolata* and *P. belcheri*, and *P. vittata*, *P. macgillivrayi* and *P. salvini* in both ML and BI trees. The species pair of *P. desolata* and *P. belcheri* received 0.75 support in the BI tree and the larger clade was fully supported. Two individuals of *P. desolata banksi* did not conform to this arrangement, and formed the outgroup of the overall clade. Further individuals of *P. desolata banksi*, *P. desolata desolata*, and a *P. desolata* individual from Bouvet Island (Atlantic Ocean) of unknown subspecies inside the main subclade were monophyletic but did not group by subspecies identity and received poor support values at all nodes. The two sampled individuals of *P. belcheri* were monophyletic with 0.75 posterior probability in the BI tree and 53% bootstrap support in the ML tree. In the second subclade, the two sampled individuals of *P. salvini* were monophyletic in BI analyses with a posterior probability of 0.5, but were not monophyletic in ML analyses. One individual of *P. macgillivrayi* was well supported as sister to *P. vittata*, but further individuals of *P. macgillivrayi* and *P. salvini* formed differing paraphyletic relationships with low support in both the ML and BI trees.

The speciose genus *Pterodroma* was well resolved with entirely concordant relationships inferred by the ML and BI trees. Two major clades were inferred. Within the smaller clade, a Caribbean and North Atlantic Ocean subclade of *P. hasitata*, *P. madeira*, *P. cahow*, *P. feae* and *P. deserta* was inferred. *P. cahow* was inferred as closest to the sister pair *P. feae* and *P. deserta* with a short branch and 53% bootstrap support in the ML tree and fully supported in the BI tree. A further subclade comprised a linear branching pattern of *P. mollis*, *P. gouldi*, *P. incerta*, *P. lessonii* and *P. macroptera*. *P. hypoleuca* was at the base of the larger major clade. Within this clade, a linear subclade of *P. longirostris*, *P. cookii*, *P. brevipes* and *P. leucoptera* was inferred. A short branch separating *P. solandri* from *P. nigripennis* and the majority of larger clade was inferred and fully supported in the BI tree and 71% supported in the ML tree. *P. inexpectata* was at the base of two further subclades. *P. ultima*, *P. atrata*, *P. neglecta*, *P. heraldica* and *P. alba* formed a closely related clade with short branch lengths. The monophyly of *P. neglecta*, *P. heraldica* and *P. alba* received a 99% bootstrap support in the ML tree. The monophyly of the two sampled individuals of *P. alba* received a 96% bootstrap support relative to the sampled individual of *P. heraldica*. Both of these nodes were fully supported in the BI tree. Finally, *P. externa*, *P. sandwichensis*, *P. baraui*, *P. cervicalis* and *P. occulta* formed a linear clade with full support at all nodes.

2.3.4 The Procellariiformes tree using MSC methods

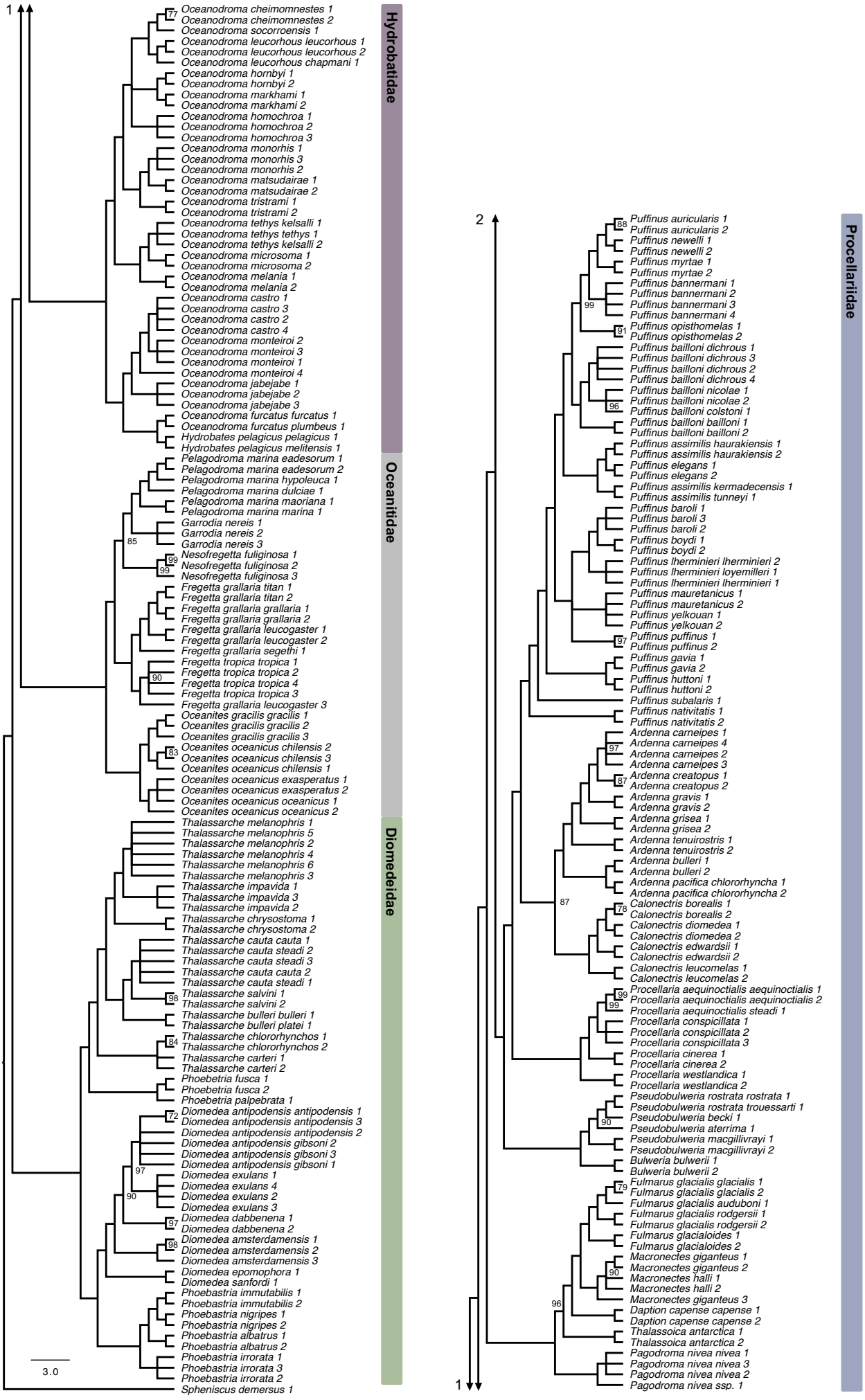
The tree built under the MSC model using SVDQuartets produced identical higher-level relationships to concatenated ML and BI methods (Fig. 2.5). Three generic nodes received lower bootstrap support using this method: *Fregetta*/*Nesofregetta* (85%), *Pagodroma*/*Thalassoica* (96%), *Ardenna*/*Calonectris*/*Puffinus* (87%). Lower-level relationships were very similar to the ML and BI topologies, with differences mainly at the shallowest nodes. Support values were generally lower at shallow nodes in the MSC tree relative to concatenated methods.

Other than minor differences, relationships within Diomedeidae, Oceanitidae, and Hydrobatidae were very similar to concatenated methods. In these families differences were restricted to shallow nodes in almost all cases. As found using concatenated methods, *Diomedea dabbenena* was sister to *D. exulans* and *D. antipodensis* with 90% support.

The sister species *Thalassarche chlororhynchos* and *T. carteri* were each monophyletic, but with low support (84% and 62% respectively).

Within Oceanitidae, *Oceanites oceanicus exasperatus* was monophyletic but with low support (36%). Within Hydrobatidae, one of four sampled individuals of *Oceanodroma monteiroi* was outside of the species clade, and was sister to the clade of *O. castro* and other *O. monteiroi* individuals. This individual presented a long branch in both ML and BI analyses and its non-monophyly is likely artefactual. The subspecies *O. tethys tethys* and *O. tethys kelsalli* were not monophyletic, although the sister pair of individuals inferred had low support of 37%. Relationships of the clade containing *O. homochroa*, *O. markhami*, *O. hornbyi* and *O. leucorhous* differed from concatenated methods, with a low supported bifurcation (31%) of *O. homochroa* as sister to *O. leucorhous*.

As with the preceding families, the MSC tree of Procellariidae showed similar results to concatenated methods with some differences at previously identified conflicting nodes and at the shallowest nodes within taxa, and some lower support values. *Macronectes halli* was monophyletic but had low support (59%) and was collapsed. Within the genus *Pseudobulweria*, the clade containing *P. aterrima*, *P. becki* and *P. rostrata* was not fully supported (90%). *Puffinus puffinus* was poorly supported as sister to a clade containing *P. yelkouan*, *P. mauretanicus*, *P. lherminieri*, *P. boydi* and *P. baroli* (45%), differing from relationships inferred by concatenated methods. The subspecies *P. lherminieri lherminieri* and *P. lherminieri loyemilleri* were not monophyletic, although the sister pair inferred had low support (55%).


1
Oceanodroma cheimomnestes 1
Oceanodroma cheimomnestes 2
Oceanodroma socorroensis 1
Oceanodroma leucorhous leucorhous 1
Oceanodroma leucorhous leucorhous 2
Oceanodroma leucorhous chapmani 1
Oceanodroma hornbyi 1
Oceanodroma hornbyi 2
Oceanodroma markhami 1
Oceanodroma markhami 2
Oceanodroma homochroa 1
Oceanodroma homochroa 2
Oceanodroma homochroa 3
Oceanodroma monorhis 1
Oceanodroma monorhis 3
Oceanodroma monorhis 2
Oceanodroma matsudairae 1
Oceanodroma matsudairae 2
Oceanodroma tristrami 1
Oceanodroma tristrami 2
Oceanodroma tethys kelsalli 1
Oceanodroma tethys tethys 1
Oceanodroma tethys kelsalli 2
Oceanodroma microsoma 1
Oceanodroma microsoma 2
Oceanodroma melania 1
Oceanodroma melania 2
Oceanodroma castro 1
Oceanodroma castro 3
Oceanodroma castro 2
Oceanodroma castro 4
Oceanodroma monteiroi 2
Oceanodroma monteiroi 3
Oceanodroma monteiroi 1
Oceanodroma monteiroi 4
Oceanodroma jabejabe 1
Oceanodroma jabejabe 2
Oceanodroma jabejabe 3
Oceanodroma furcatus furcatus 1
Oceanodroma furcatus plumbeus 1
Hydrobates pelagicus pelagicus 1
Hydrobates pelagicus melitensis 1
Pelagodroma marina eadesorum 1
Pelagodroma marina eadesorum 2
Pelagodroma marina hypoleuca 1
Pelagodroma marina dulciae 1
Pelagodroma marina maoriana 1
Pelagodroma marina marina 1
Garrodia nereis 1
Garrodia nereis 2
Garrodia nereis 3
Nesofregetta fuliginosa 1
Nesofregetta fuliginosa 2
Nesofregetta fuliginosa 3
Fregetta grallaria titan 1
Fregetta grallaria titan 2
Fregetta grallaria grallaria 1
Fregetta grallaria grallaria 2
Fregetta grallaria leucogaster 1
Fregetta grallaria leucogaster 2
Fregetta grallaria segethi 1
Fregetta tropica tropica 1
Fregetta tropica tropica 2
Fregetta tropica tropica 4
Fregetta tropica tropica 3
Fregetta grallaria leucogaster 3
Oceanites gracilis gracilis 1
Oceanites gracilis gracilis 2
Oceanites gracilis gracilis 3
Oceanites oceanicus chilensis 2
Oceanites oceanicus chilensis 3
Oceanites oceanicus chilensis 1
Oceanites oceanicus exasperatus 1
Oceanites oceanicus exasperatus 2
Oceanites oceanicus oceanicus 1
Oceanites oceanicus oceanicus 2
Thalassarche melanophris 1
Thalassarche melanophris 5
Thalassarche melanophris 2
Thalassarche melanophris 4
Thalassarche melanophris 6
Thalassarche melanophris 3
Thalassarche impavida 1
Thalassarche impavida 3
Thalassarche impavida 2
Thalassarche chrysostoma 1
Thalassarche chrysostoma 2
Thalassarche cauta cauta 1
Thalassarche cauta steadi 2
Thalassarche cauta steadi 3
Thalassarche cauta cauta 2
Thalassarche cauta steadi 1
Thalassarche salvini 1
Thalassarche salvini 2
Thalassarche bulleri bulleri 1
Thalassarche bulleri platei 1
Thalassarche chlororhynchos 1
Thalassarche chlororhynchos 2
Thalassarche carteri 1
Thalassarche carteri 2
Phoebetria fusca 1
Phoebetria fusca 2
Phoebetria palpebrata 1
Diomedea antipodensis antipodensis 1
Diomedea antipodensis antipodensis 3
Diomedea antipodensis antipodensis 2
Diomedea antipodensis gibsoni 2
Diomedea antipodensis gibsoni 3
Diomedea antipodensis gibsoni 1
Diomedea exulans 1
Diomedea exulans 4
Diomedea exulans 2
Diomedea exulans 3
Diomedea dabbenena 1
Diomedea dabbenena 2
Diomedea amsterdamensis 1
Diomedea amsterdamensis 2
Diomedea amsterdamensis 3
Diomedea epomophora 1
Diomedea sanfordi 1
Phoebastria immutabilis 1
Phoebastria immutabilis 2
Phoebastria nigripes 1
Phoebastria nigripes 2
Phoebastria albatrus 1
Phoebastria albatrus 2
Phoebastria irrorata 1
Phoebastria irrorata 3
Phoebastria irrorata 2
Spheniscus demersus 1
3.0
Hydrobatidae
Oceanitidae
Diomedeidae
2
Puffinus auricularis 1
Puffinus auricularis 2
Puffinus newelli 1
Puffinus newelli 2
Puffinus myrtae 1
Puffinus myrtae 2
Puffinus bannermani 1
Puffinus bannermani 2
Puffinus bannermani 3
Puffinus bannermani 4
Puffinus opisthomelas 1
Puffinus opisthomelas 2
Puffinus bailloni dichrous 1
Puffinus bailloni dichrous 3
Puffinus bailloni dichrous 2
Puffinus bailloni dichrous 4
Puffinus bailloni nicolae 1
Puffinus bailloni nicolae 2
Puffinus bailloni colstoni 1
Puffinus bailloni bailloni 1
Puffinus bailloni bailloni 2
Puffinus assimilis haurakiensis 1
Puffinus assimilis haurakiensis 2
Puffinus elegans 1
Puffinus elegans 2
Puffinus assimilis kermadecensis 1
Puffinus assimilis tunneyi 1
Puffinus baroli 1
Puffinus baroli 3
Puffinus baroli 2
Puffinus boydi 1
Puffinus boydi 2
Puffinus lherminieri lherminieri 2
Puffinus lherminieri loyemilleri 1
Puffinus lherminieri lherminieri 1
Puffinus mauretanicus 1
Puffinus mauretanicus 2
Puffinus yelkouan 1
Puffinus yelkouan 2
Puffinus puffinus 1
Puffinus puffinus 2
Puffinus gavia 1
Puffinus gavia 2
Puffinus huttoni 1
Puffinus huttoni 2
Puffinus subalaris 1
Puffinus nativitatis 1
Puffinus nativitatis 2
Ardenna carneipes 1
Ardenna carneipes 4
Ardenna carneipes 2
Ardenna carneipes 3
Ardenna creatopus 1
Ardenna creatopus 2
Ardenna gravis 1
Ardenna gravis 2
Ardenna grisea 1
Ardenna grisea 2
Ardenna tenuirostris 1
Ardenna tenuirostris 2
Ardenna bulleri 1
Ardenna bulleri 2
Ardenna pacifica chlororhyncha 1
Ardenna pacifica chlororhyncha 2
Calonectris borealis 1
Calonectris borealis 2
Calonectris diomedea 1
Calonectris diomedea 2
Calonectris edwardsii 1
Calonectris edwardsii 2
Calonectris leucomelas 1
Calonectris leucomelas 2
Procellaria aequinoctialis aequinoctialis 1
Procellaria aequinoctialis aequinoctialis 2
Procellaria aequinoctialis steadi 1
Procellaria conspicillata 1
Procellaria conspicillata 2
Procellaria conspicillata 3
Procellaria cinerea 1
Procellaria cinerea 2
Procellaria westlandica 1
Procellaria westlandica 2
Pseudobulweria rostrata rostrata 1
Pseudobulweria rostrata trouessarti 1
Pseudobulweria becki 1
Pseudobulweria aterrima 1
Pseudobulweria macgillivrayi 1
Pseudobulweria macgillivrayi 2
Bulweria bulwerii 1
Bulweria bulwerii 2
Fulmarus glacialis glacialis 1
Fulmarus glacialis glacialis 2
Fulmarus glacialis auduboni 1
Fulmarus glacialis rodgersii 1
Fulmarus glacialis rodgersii 2
Fulmarus glacialoides 1
Fulmarus glacialoides 2
Macronectes giganteus 1
Macronectes giganteus 2
Macronectes halli 1
Macronectes halli 2
Macronectes giganteus 3
Daption capense capense 1
Daption capense capense 2
Thalassoica antarctica 1
Thalassoica antarctica 2
Pagodroma nivea nivea 1
Pagodroma nivea nivea 3
Pagodroma nivea nivea 2
Pagodroma nivea ssp. 1
Procellariidae

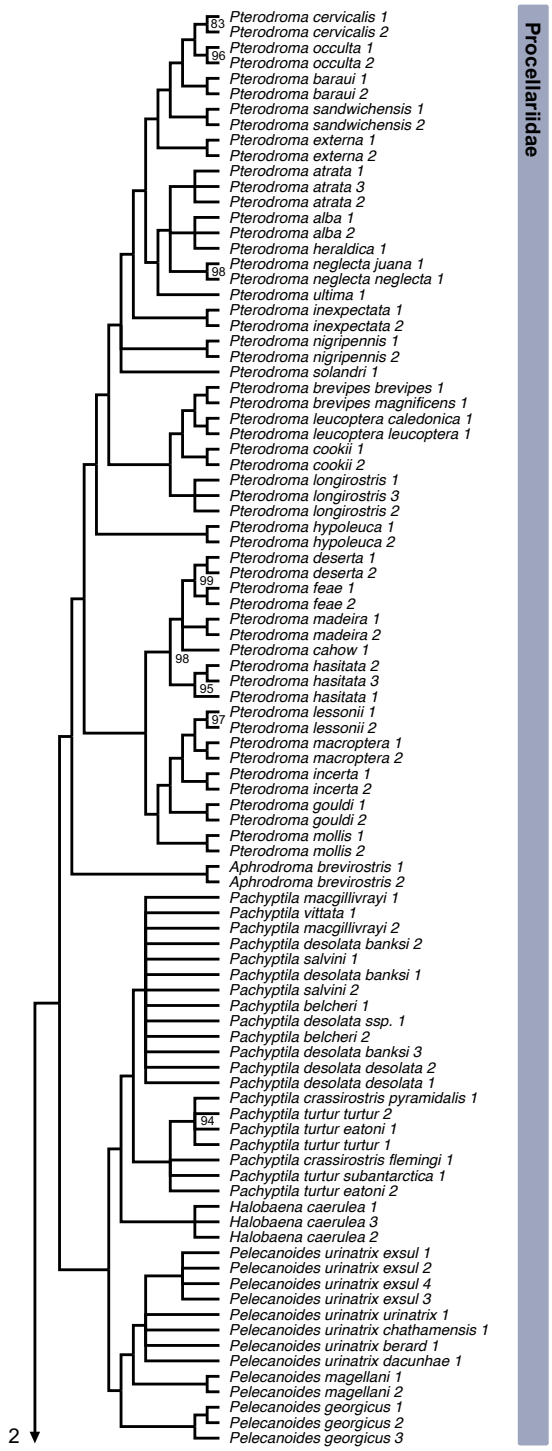


Figure 2.5 Tree constructed under the MSC model based on site patterns of 4,335 UCE loci. Bootstrap support values are given for nodes with <100 support. Nodes with support values <70 are collapsed. Numbered arrows indicate the connection points between sections of the tree. Families are indicated by coloured bars.

P. yelkouan and *P. mauretanicus* were recovered as monophyletic species with low support (41% and 35% respectively).

As in concatenated methods, relationships of the genus *Pachyptila* were poorly resolved and supported. The main divergence of clades containing *P. turtur* and *P. crassirostris*, and *P. desolata*, *P. belcheri*, *P. vittata*, *P. macgillivayi* and *P. salvini* was maintained in the MSC tree. Relationships within these subclades were poorly supported.

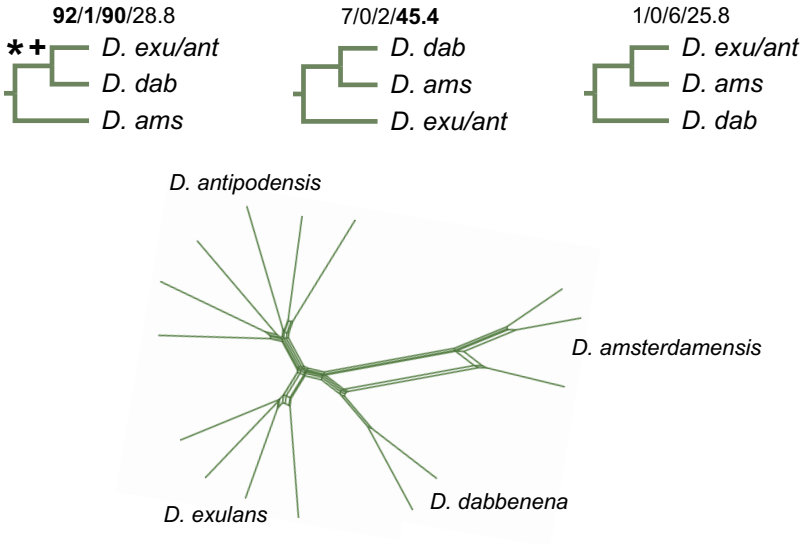
In the genus *Pterodroma*, the relationships of *P. cahow* relative to *P. madeira*, *P. feae*, and *P. deserta* differed to those inferred by concatenated methods. *P. cahow* was inferred as sister to *P. madeira* with 62% support. At a further conflicting node in *Pterodroma*, *P. solandri* and *P. nigripennis* were inferred as sister species with 57% support, differing from the linear pattern of *P. solandri* and *P. nigripennis* inferred by concatenated methods. *P. neglecta* was inferred as sister to *P. atrata* rather than to *P. alba* and *P. heraldica* in concatenated analyses, although with low support (38%). *P. alba* was recovered as monophyletic relative to *P. heraldica*, but with low support (59%).

2.3.5 Evaluating areas of phylogenetic conflict in the Procellariiformes tree

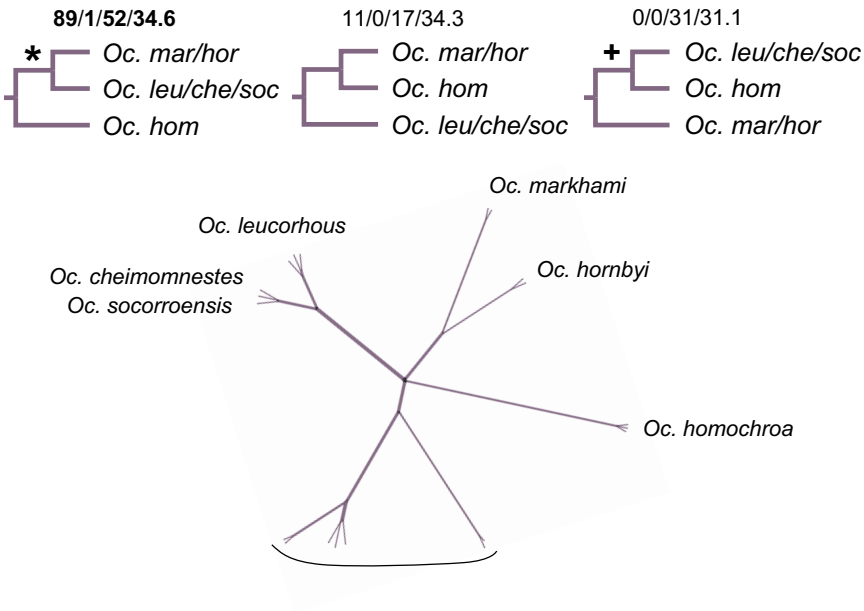
Five nodes above the species level with conflicting phylogenetic signal were identified in ML and BI trees and support metrics. Short internal branches and low ML bootstrap values were common to all nodes with conflicting signal (Figs. 2.2-2.4). For these nodes, each possible topological arrangement was depicted with support metrics (ML bootstraps, BI posterior probabilities, MSC bootstraps, and sCF) alongside phylogenetic networks (Fig. 2.6).

Within the genus *Diomedea*, ML analyses identified *D. amsterdamensis* (DA) as the sister to *D. dabbenena* (DD) and the pair of sister species *D. exulans* and *D. antipodensis* (DEA) with 92% bootstrap support. This topology was inferred by BI methods with full support and the MSC tree method implemented using SVDQuartets with 90% bootstrap support (Fig. 2.6A). Conversely, sCF values for this species complex favoured a sister relationship of DD and DA supported by 45.4% of decisive sites, considerably higher than the 28.8% for the topology inferred by all tree building methods. A phylogenetic network of this clade showed a reticulate branching pattern of one of three sampled individuals of DA with the DD lineage, a strong indicator of gene flow in the ancestry of this DA individual.

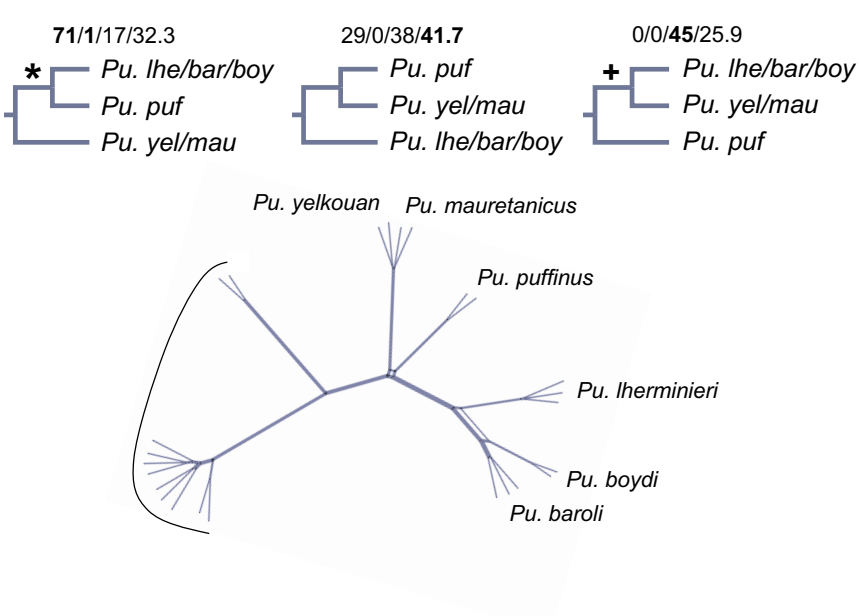
A


92/1/90/28.8
* +
D. exu/ant
D. dab
D. ams
7/0/2/**45.4**
D. dab
D. ams
D. exu/ant
1/0/6/25.8
D. exu/ant
D. ams
D. dab
D. antipodensis
D. amsterdamensis
D. exulans
D. dabbenena

B


89/1/52/34.6
*
Oc. mar/hor
Oc. leu/che/soc
Oc. hom
11/0/17/34.3
Oc. mar/hor
Oc. hom
Oc. leu/che/soc
0/0/31/31.1
+
Oc. leu/che/soc
Oc. hom
Oc. mar/hor
Oc. markhami
Oc. leucorhous
Oc. hornbyi
Oc. cheimomnestes
Oc. socorroenesis
Oc. homochroa

C


71/1/17/32.3
*
Pu. lhe/bar/boy
Pu. puf
Pu. yel/mau
29/0/38/**41.7**
Pu. puf
Pu. yel/mau
Pu. lhe/bar/boy
0/0/**45**/25.9
+
Pu. lhe/bar/boy
Pu. yel/mau
Pu. puf
Pu. yelkouan
Pu. mauretanicus
Pu. puffinus
Pu. lherminieri
Pu. boydi
Pu. baroli

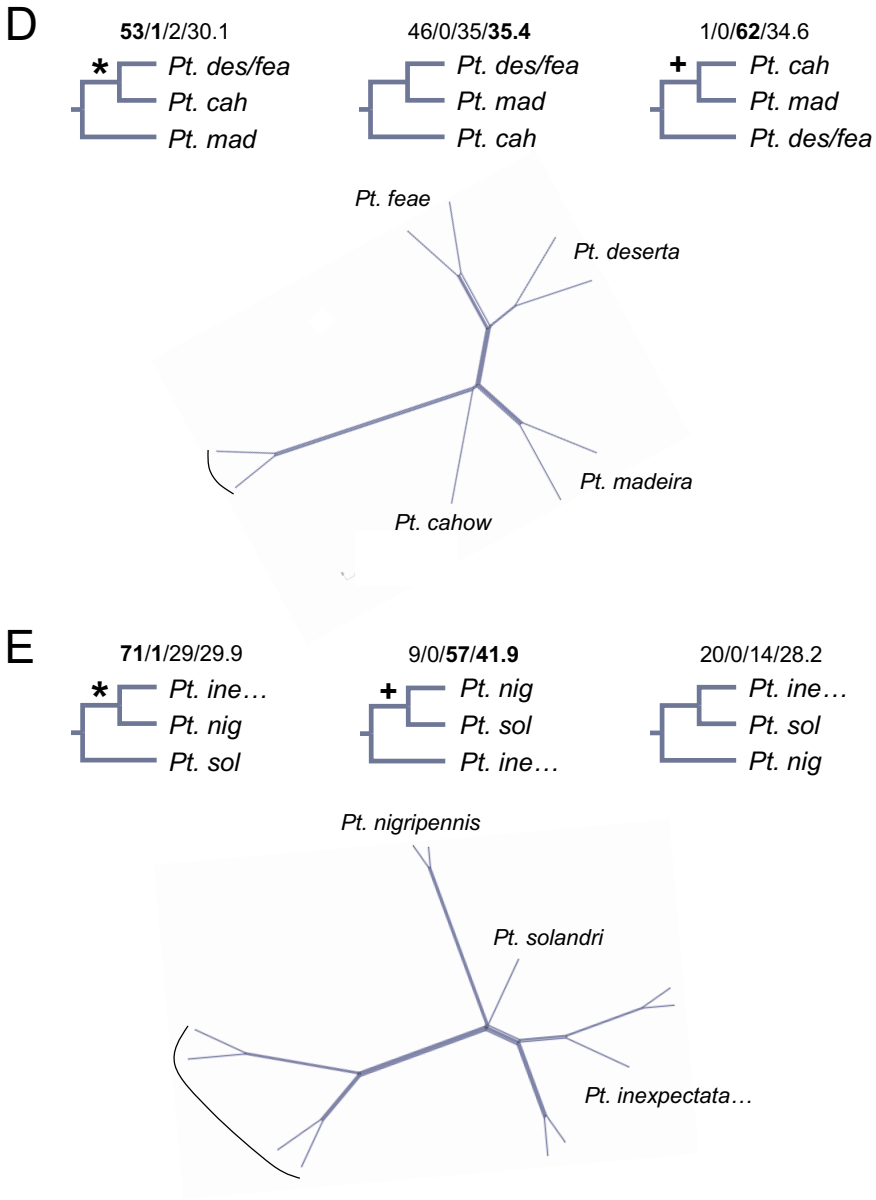


Figure 2.6 Topological arrangements of species relationships at nodes with conflicting phylogenetic signal. Genera are abbreviated as follows: A: *D. Diomedea*; B: *Oc. Oceanodroma*; C: *Pu. Puffinus*; D, E: *Pt. Pterodroma*. An asterisk indicates the topology inferred by ML and BI methods and a plus sign indicates the topology inferred under the MSC model. Support metrics are given above each topology in the following order: ML bootstraps, BI posterior probabilities, MSC bootstraps, sCF. The highest scoring topology for each metric type is in bold. Phylogenetic networks are depicted below topologies for each species complex, with additional outgroup taxa in some cases under curved lines to aid viewing.

The family Hydrobatidae was well resolved by ML and BI analyses with the exception of the branching pattern of a species complex comprising *Oceanodroma homochroa* (OH), the monophyletic species trio *O. leucorhous*, *O. cheimomnestes* and *O. socorroensis* (OLCS), and the pair of sister species *O. markhami* and *O. hornbyi* (OMH; Fig. 2.6B). ML analyses identified OH as the outgroup to a sister relationship of OLCS and OMH with 89% bootstrap support and a short branch relative to other divergences within Hydrobatidae. This topology was fully supported by BI. Most MSC tree bootstrap replicates supported this topology (52%), but the MSC tree instead inferred a sister relationship of OH to OLCS with 31% bootstrap support. sCF values did not indicate strong support for any topology at this node (34.6%/34.3%/31.1%), although marginally higher support for the topology inferred by concatenated ML and BI methods and least support for the topology inferred by the MSC method was apparent. This ambiguity was also present in a phylogenetic network of the species complex, which showed the near absence of a branch between the divergences of OH, OLCS and OMH, indicating a simultaneous speciation event, or a very rapid series of speciation events.

Within the genus *Puffinus*, a clade of *P. puffinus* (PP), the sister pair *P. yelkouan* and *P. mauretanicus* (PYM), and the species trio *P. lherminieri*, *P. boydi* and *P. baroli* (PLBB) was identified. PYM was sister to the monophyletic PP and PLBB with 71% bootstrap support in the ML tree, and full support in the BI tree (Fig. 2.6C). sCF values strongly favoured a sister relationship of PP and PYM (41.7%) relative to the ML and BI topology (32.3%) and the other possible topology of PLBB and PYM as sisters (25.9%). The MSC tree, however, inferred the latter topology, although with low support at 45% that was similar to its support for the topology identified by sCF at 38%. A network of this species group showed a closer genetic distance of PP to PYM, but the absence of a branch between divergences of the three monophyletic clusters.

ML and BI trees identified two areas of phylogenetic conflict in the genus *Pterodroma*. ML analyses inferred *P. madeira* (PM) as sister to a sister relationship of *P. cahow* (PC) and the species pair *P. feae* and *P. deserta* (PFD) with 53% bootstrap support (Fig. 2.6D). The same topology was fully supported in the BI tree. The MSC tree instead inferred a sister relationship of PC and PM with 62% support, and sCF values marginally favoured the third possible topology of PFD and PM as sisters with 35.4%. A phylogenetic network indicated a near simultaneous branching of the three species clusters that likely underlies the short branch in ML and BI trees and the lack of decisive PIS at this node.

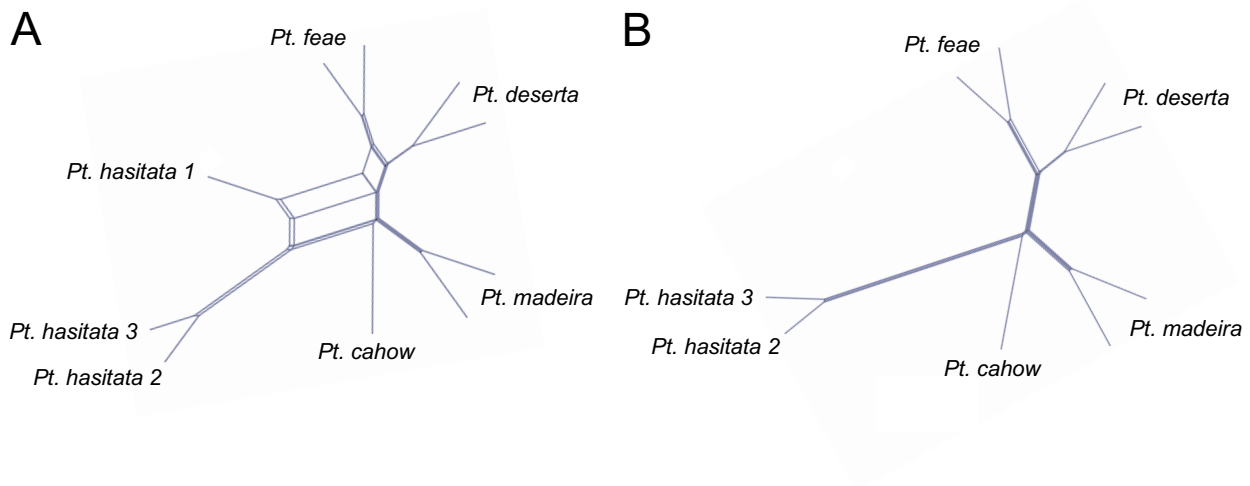


Figure 2.7 Phylogenetic networks of *Pterodroma* (*Pt.*) species. Reticulate patterns are present between *Pt. hasitata 1* and *Pt. feae* (A), and absent when this individual is removed (B).

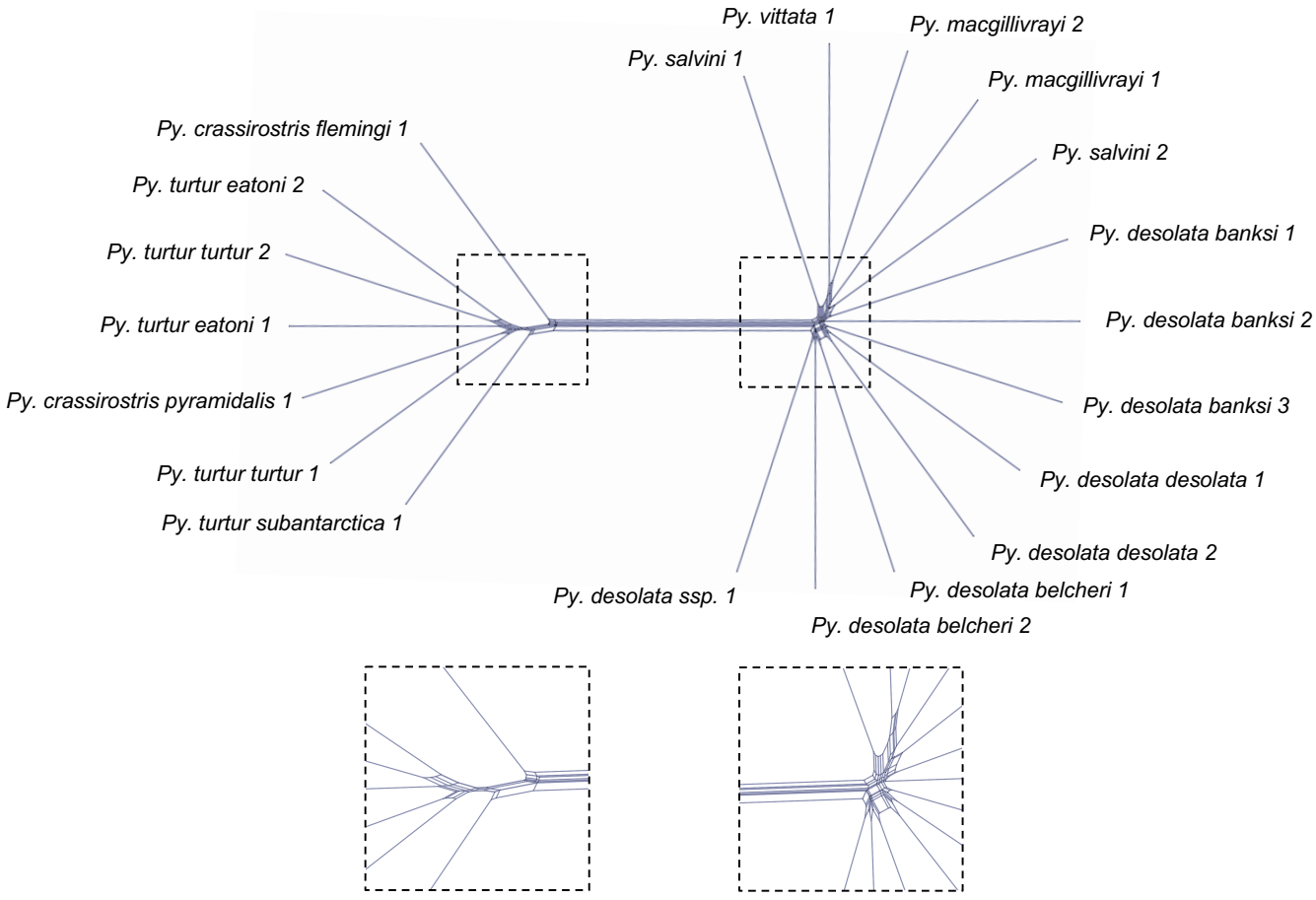


Figure 2.8 Phylogenetic network of the genus *Pachyptila* (*Py.*). Boxes with dashed outlines below show enlarged versions of the boxes within the network.

A larger network of this clade including the closest related species *Pterodroma hasitata* showed reticulate branches between the sampled individual *P. hasitata 1* and the *P. feae* lineage (Fig. 2.7A) that were not present when this individual was removed (Fig. 2.7B), indicating admixture in the ancestry of this individual. This individual was found at a *P. feae* breeding colony in Cape Verde, outside its normal breeding range in the Caribbean.

In *Pterodroma*, a further conflicting node was identified in the branching order of *P. solandri* (PS) as an outgroup to *P. nigripennis* (PN), and further species including *P. inexpectata* (PI), which was inferred by ML with 71% bootstrap support and full support by BI (Fig. 2.6E). In this case, sCF strongly favoured an alternative topology of PS and PN as sisters, supported by 41.9% of decisive sites in the alignment relative to 29.9% for the ML and BI topology, and 28.2% for the other possible topology of PS and PI as sisters. The MSC tree also favoured this topology, albeit with only 57% bootstrap support. A phylogenetic network of this species clade indicated a lack of genetic distance between the divergence of PS and PN, suggesting a simultaneous branching event consistent with a polytomy, or a rapid series of branching events.

The poorly resolved genus *Pachyptila* was examined as a phylogenetic network. This network inferred the same deeper-level structure of two main clades of *P. turtur* and *P crassirostris*, and *P. desolata*, *P. belcheri*, *P. vittata*, *P. macgillivayi* and *P. salvini* that was inferred in the ML and BI trees, and mostly similar relationships within these clades (Fig. 2.8). Reticulate branches were present in both clades at deep levels relative to terminal branch lengths.

A consensus topology of Procellariiformes was summarised from the results of phylogenetic trees inferred by different methods, phylogenetic networks, and support metrics (Fig. 2.9). This tree was identical to the topologies inferred by ML and BI analyses, except that some conflicting and poorly supported nodes were collapsed to polytomies.

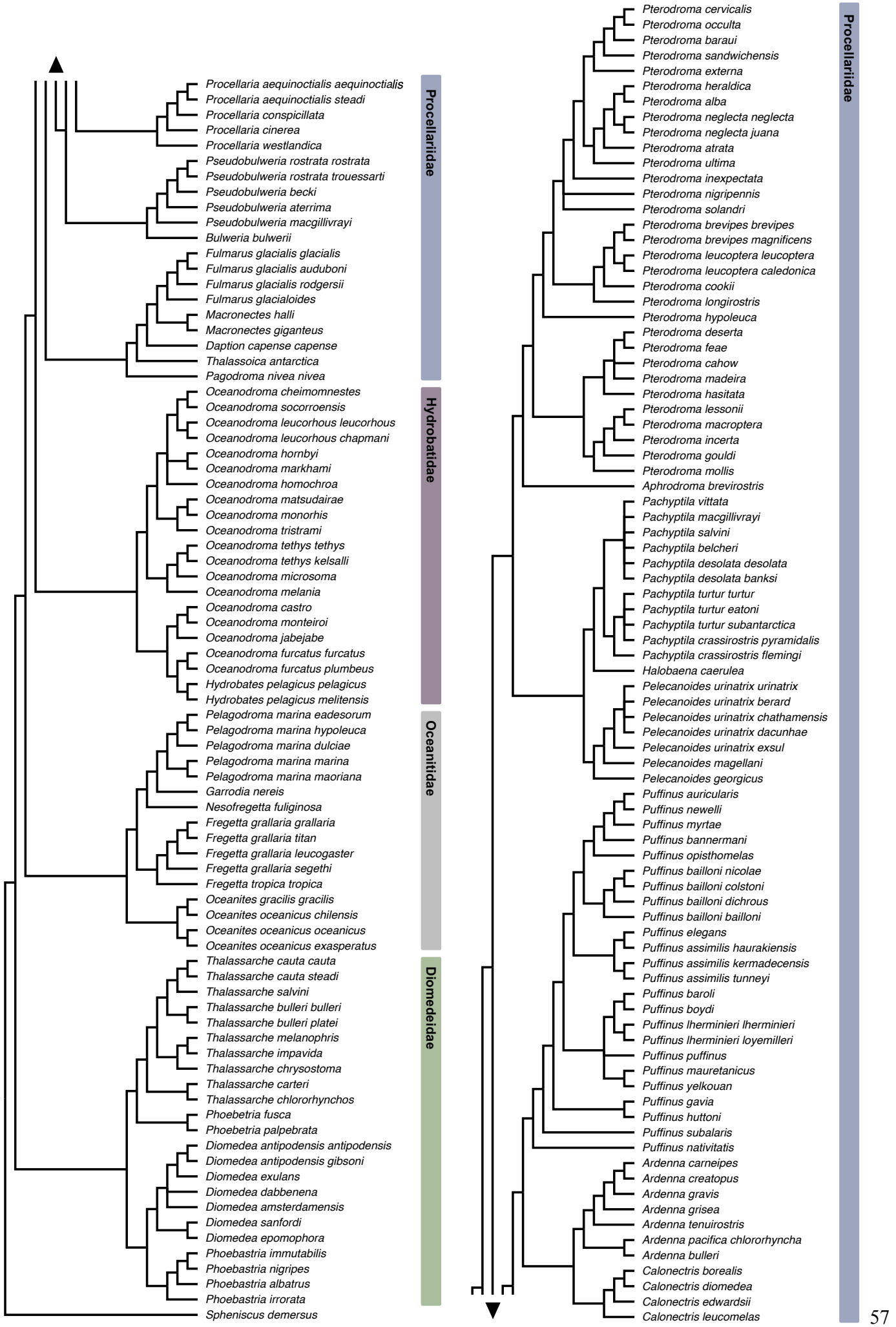

Procellariidae
Hydrobatidae
Oceanitidae
Diomedeidae
Procellariidae
Procellaria aequinoctialis aequinoctialis
Procellaria aequinoctialis steadi
Procellaria conspicillata
Procellaria cinerea
Procellaria westlandica
Pseudobulweria rostrata rostrata
Pseudobulweria rostrata trouessarti
Pseudobulweria becki
Pseudobulweria aterrima
Pseudobulweria macgillivrayi
Bulweria bulwerii
Fulmarus glacialis glacialis
Fulmarus glacialis auduboni
Fulmarus glacialis rodgersii
Fulmarus glacialoides
Macronectes halli
Macronectes giganteus
Daption capense capense
Thalassoica antarctica
Pagodroma nivea nivea
Oceanodroma cheimomnestes
Oceanodroma socorroensis
Oceanodroma leucorhous leucorhous
Oceanodroma leucorhous chapmani
Oceanodroma hornbyi
Oceanodroma markhami
Oceanodroma homochroa
Oceanodroma matsudairae
Oceanodroma monorhis
Oceanodroma tristrami
Oceanodroma tethys tethys
Oceanodroma tethys kelsalli
Oceanodroma microsoma
Oceanodroma melania
Oceanodroma castro
Oceanodroma monteiroi
Oceanodroma jabejabe
Oceanodroma furcatus furcatus
Oceanodroma furcatus plumbeus
Hydrobates pelagicus pelagicus
Hydrobates pelagicus melitensis
Pelagodroma marina eadesorum
Pelagodroma marina hypoleuca
Pelagodroma marina dulciae
Pelagodroma marina marina
Pelagodroma marina maoriana
Garrodia nereis
Nesofregetta fuliginosa
Fregetta grallaria grallaria
Fregetta grallaria titan
Fregetta grallaria leucogaster
Fregetta grallaria segethi
Fregetta tropica tropica
Oceanites gracilis gracilis
Oceanites oceanicus chilensis
Oceanites oceanicus oceanicus
Oceanites oceanicus exasperatus
Thalassarche cauta cauta
Thalassarche cauta steadi
Thalassarche salvini
Thalassarche bulleri bulleri
Thalassarche bulleri platei
Thalassarche melanophris
Thalassarche impavida
Thalassarche chrysostoma
Thalassarche carteri
Thalassarche chlororhynchos
Phoebetria fusca
Phoebetria palpebrata
Diomedea antipodensis antipodensis
Diomedea antipodensis gibsoni
Diomedea exulans
Diomedea dabbenena
Diomedea amsterdamensis
Diomedea sanfordi
Diomedea epomophora
Phoebastria immutabilis
Phoebastria nigripes
Phoebastria albatrus
Phoebastria irrorata
Spheniscus demersus
Procellariidae
Pterodroma cervicalis
Pterodroma occulta
Pterodroma baraui
Pterodroma sandwichensis
Pterodroma externa
Pterodroma heraldica
Pterodroma alba
Pterodroma neglecta neglecta
Pterodroma neglecta juana
Pterodroma atrata
Pterodroma ultima
Pterodroma inexpectata
Pterodroma nigripennis
Pterodroma solandri
Pterodroma brevipes brevipes
Pterodroma brevipes magnificens
Pterodroma leucoptera leucoptera
Pterodroma leucoptera caledonica
Pterodroma cookii
Pterodroma longirostris
Pterodroma hypoleuca
Pterodroma deserta
Pterodroma feae
Pterodroma cahow
Pterodroma madeira
Pterodroma hasitata
Pterodroma lessonii
Pterodroma macroptera
Pterodroma incerta
Pterodroma gouldi
Pterodroma mollis
Aphrodroma brevirostris
Pachyptila vittata
Pachyptila macgillivrayi
Pachyptila salvini
Pachyptila belcheri
Pachyptila desolata desolata
Pachyptila desolata banksi
Pachyptila turtur turtur
Pachyptila turtur eatoni
Pachyptila turtur subantarctica
Pachyptila crassirostris pyramidalis
Pachyptila crassirostris flemingi
Halobaena caerulea
Pelecanoides urinatrix urinatrix
Pelecanoides urinatrix berard
Pelecanoides urinatrix chathamensis
Pelecanoides urinatrix dacunhae
Pelecanoides urinatrix exsul
Pelecanoides magellani
Pelecanoides georgicus
Puffinus auricularis
Puffinus newelli
Puffinus myrtae
Puffinus bannermani
Puffinus opisthomelas
Puffinus bailloni nicolae
Puffinus bailloni colstoni
Puffinus bailloni dichrous
Puffinus bailloni bailloni
Puffinus elegans
Puffinus assimilis haurakiensis
Puffinus assimilis kermadecensis
Puffinus assimilis tunneyi
Puffinus baroli
Puffinus boydi
Puffinus lherminieri lherminieri
Puffinus lherminieri loyemilleri
Puffinus puffinus
Puffinus mauretanicus
Puffinus yelkouan
Puffinus gavia
Puffinus huttoni
Puffinus subalaris
Puffinus nativitatis
Ardenna carneipes
Ardenna creatopus
Ardenna gravis
Ardenna grisea
Ardenna tenuirostris
Ardenna pacifica chlororhyncha
Ardenna bulleri
Calonectris borealis
Calonectris diomedea
Calonectris edwardsii
Calonectris leucomelas

Figure 2.9 Consensus tree of Procellariiformes relationships based on all tree building methods and support metrics. Families are indicated by coloured bars. Arrows indicate the connection between the two tree sections.

2.4 Discussion

This study represents the first time a dataset of genome-wide markers has been applied to a taxonomically thorough sampling of the seabird order Procellariiformes, and provides the most comprehensive phylogenetic tree of this group to date. The combined dataset of 4,335 UCE loci and 370,035 PIS produced concordant and well supported results across analyses in almost all clades. Results were presented for 129 species (89% of all species) and 166 taxa including subspecies (85% of all taxa) of Procellariiformes, elucidating relationships to fine scales for the first time, with the branching order and patterns of species and subspecies well resolved in almost all cases. As a result, a significant gap in biodiversity knowledge, specifically of the Darwinian shortfall of evolutionary relationships among lineages (Diniz-Filho et al. 2013, Hortal et al. 2015), for this remarkable bird order has been addressed.

2.4.1 Conflicting relationships in Procellariiformes

The utility of UCEs to resolve phylogenetic relationships is well known at deep evolutionary scales (Faircloth et al. 2012, McCormack et al. 2013a), and at shallower scales of closely related taxa (Ferrer Obiol et al. 2021, Harvey et al. 2016, Herman et al. 2022, Smith et al. 2014, Winker et al. 2018). Outside of the genus *Pachyptila*, which showed unexpectedly poorly resolved relationships, only subspecies of three species were not inferred as monophyletic: *Diomedea antipodensis antipodensis*/*gibsoni*, *Thalassarche cauta cauta*/*steadi*, *Oceanites oceanicus oceanicus*/*exasperatus*. It is possible that the conserved nature of UCEs did not provide sufficient variation to separate these shallow divergences, or that current nomenclature does not accurately reflect biological divergence in these cases. Other genome-wide markers, notably RADseq that has been particularly applied to shallow phylogenetic questions (e.g. Benestan et al. 2015, Vendrami et al. 2017) may prove useful in elucidating the fine-scale relationships within these species.

In addition to these cases, several nodes provided evidence of conflicting phylogenetic signal within the dataset. These areas of the tree shared a common feature in having short internal branches between nodes in the ML and BI trees, which is associated with phylogenetic conflict

(Degnan & Rosenberg 2006). Short internal branches, with short underlying time periods between speciation events, allow few substitutions that may be phylogenetically informative to arise, and increase the likelihood of incomplete lineage sorting (ILS; Degnan & Rosenberg 2006, Maddison & Knowles 2006, Pamilo & Nei 1988). One of these identified areas, the relationships of several *Puffinus* taxa of the North Atlantic Ocean (Fig. 2.6C), was also identified as a conflicting node with a short internal branch in a genomic study of shearwater species employing UCEs and RADseq loci (Ferrer Obiol et al. 2021), leading the authors to suggest that high ILS had resulted from nearly simultaneous speciation events.

Simultaneous speciation events represent hard polytomies that can not be portrayed as fully bifurcating nodes (Walsh et al. 1999, Suh 2016). At such nodes, it is expected that individual loci or sites within the dataset should not favour a particular topology (Degnan & Rosenberg 2006). Several examples of apparent simultaneous or near simultaneous divergences were identified in phylogenetic networks (Fig. 2.6B-E), and notably two where the signal in the dataset (evaluated by sCF) did not strongly favour a particular quartet topology (Fig. 2.6B, D). These two nodes, the relationships of recently diverged clades of *Oceanodroma* and *Pterodroma* species, provide the strongest evidence for the presence of hard polytomies within Procellariiformes. However, soft polyomties can not be ruled out given the relatively low PIS content of UCE loci, that may not have generated sufficient signal between rapid speciation events (Meiklejohn et al. 2016). Further genomic data may help to elucidate whether these nodes represent resolvable soft polytomies or unresolvable hard polytomies.

The signal of ancient rapid divergences can be obscured by subsequent homoplasy due to saturation (Whitfield & Lockhart 2007). This is a possible explanation for the conflicting relationships inferred for *Pterodroma solandri* relative to *P. nigripennis* and other congeners (Fig. 2.6E). This node was embedded within the *Pterodroma* radiation in the ML and BI trees, showed the absence of a branch between divergences in a phylogenetic network, and sCF values strongly favoured a particular quartet topology.

In addition to ILS, introgression is an important process that can obscure phylogenetic relationships and contribute to discordance (Abbott et al. 2016). The conflicting relationships in *Diomedea* species identified reticulate patterns between one individual of *D. amsterdamensis* and the *D. dabbenena* lineage (Fig. 2.6A), indicating that introgression has contributed to the phylogenetic conflict at this node. In addition to this, short branches were found between all species in the ML and BI trees, and it is unsurprising that the relationships in this clade have proved difficult to disentangle using mitochondrial markers (Burg & Croxall 2004, Rains et al. 2011). Although rare, breeding-site vagrancy and hybridisation has been

recorded in other Diomedeidae species (Phillips et al. 2018). Despite high philopatry (Milot et al. 2008, Rains et al. 2011), long-distance dispersal has been identified as a factor influencing population structure within *D. exulans* (Cooper & Weimerskirch 2003, Milot et al. 2008). A juvenile *D. dabbenena* was recently recorded at a *D. exulans* breeding colony in the Crozet Islands, Indian Ocean (Bond et al. 2021), leading the authors to suggest that dispersal events in *Diomedea* species may be more common than previously thought given the challenging identification of juvenile birds in this genus.

Signals of introgression were also identified among the species of North Atlantic Ocean *Pterodroma*, specifically involving an individual of *P. hasitata* relative to *P. feae* individuals (Fig. 2.7). This individual was found at a *P. feae* breeding colony in Cape Verde, outside the breeding range of *P. hasitata* in the Caribbean. Hybridisation has been documented in *Pterodroma* petrels in a remarkable scenario of three interbreeding species at Round Island in the Indian Ocean (Booth Jones 2017, Brown et al. 2011). Sporadic cases of hybridisation, as evidenced here, have been sparsely documented in *Pterodroma* species (Tennyson et al. 2013).

Conflicting phylogenetic results between concatenated and MSC methods are a common feature of large genome-wide datasets (e.g. Chan et al. 2020, Gatesy & Springer 2014, Song et al. 2012). Although mostly very similar, the tree built under the MSC model using SVDQuartets presented some differences to the almost entirely identical ML and BI trees. The MSC tree produced fewer monophyletic species and subspecies and lower support values than ML and BI trees. MSC methods can accommodate the heterogeneous histories of individual loci due to ILS and other processes (Maddison 1997, Pamilo & Nei 1988), which is likely to occur at rapid radiations. Summary-based MSC methods are commonly used (Bravo et al. 2019), and rely on the prior estimation of individual loci trees. As a result, summary MSC methods are sensitive to gene tree estimation error (Mirarab et al. 2016, Roch & Warnow 2015, Springer & Gatesy 2016), which is a particular problem in short loci with low PIS, such as UCEs (Blom et al. 2017, Hosner et al. 2016, Meiklejohn et al. 2016, Molloy & Warnow 2018). By contrast, concatenated ML methods perform well with large numbers of low signal loci (Molloy & Warnow 2018). Given the potential problems of estimating trees from individual UCE loci, a site-based MSC method that bypasses prior gene tree estimation, SVDQuartets, was used here. ML and BI trees identified four conflicting nodes with short internal branches and near simultaneous divergences in phylogenetic networks (Fig. 2.6B-E), where ILS would be expected to be high and an MSC method incorporating ILS should perform best (Kubatko & Degnan 2007, Roch & Steel 2015). However, in three of these nodes, SVDQuartets inferred a topology that no other method or support metric favoured (Fig. 2.6B-D). SVDQuartets has

been shown to be more accurate than summary-based MSC methods under high ILS, but not concatenated ML even though SVDQuartets would be expected to produce better results in these conditions (Molloy & Warnow 2018). Furthermore, concatenated ML methods have been shown to outperform both site-based and summary-based MSC methods in most ILS and gene tree estimation error conditions (Chou et al. 2015, Molloy & Warnow 2018, Portik & Wiens 2021), and for these reasons the topologies inferred by ML and BI are likely to be more accurate throughout.

2.4.2 Updating Procellariiformes systematics

The dense taxon sampling applied here corroborated the deeper level relationships of Procellariiformes inferred recently by Estandia et al. (under review) with identical topologies of families and genera. These results are considerably different in the arrangement of families and genera to the main previous phylogeny of the order (Nunn & Stanley 1998), and provide updated, robust relationships at these scales (Fig. 2.1).

As expected, given the frequently observed mito-nuclear discordance in Procellariiformes (Gangloff et al. 2013, Silva et al. 2015, Torres et al. 2021), incongruence with previous works that mainly used mitochondrial markers remained frequent at finer scales. The dataset presented here provides an updated view of Procellariiformes systematics at these scales. Systematic information and remaining questions in Procellariiformes systematics within genera are discussed below for each family.

2.4.2.1 Diomedeidae

The phylogeny of Diomedeidae was well resolved and supported, and genome-wide data were used to assess relationships within the family the first time. Within the wandering albatross complex of the genus *Diomedea*, a branching pattern of *D. amsterdamensis*, *D. dabbenena*, *D. exulans*, *D. antipodensis* was best supported (see section 2.4.1). The sister relationship of *D. amsterdamensis* and *D. exulans* previously found using mitochondrial DNA (Chambers et al. 2009, Nunn & Stanley 1998, Rains et al. 2011) was not supported. Monophyly of *D. antipodensis antipodensis* and *D. antipodendsis gibsoni* was not recovered in this study, although this could be due to the relatively low variability of UCEs, given that genetic variation in *Diomedea* species is low (Chambers et al. 2009).

Results presented here support *Phoebastria irrorata* as the deepest divergence within its genus, which has been found by one study using cytochrome *b* (Nunn & Stanley 1998), but not replicated in another cytochrome *b* study (Chambers et al. 2009).

Relationships among *Thalassarche* species are concordant with those previously found (Chambers et al. 2009, Nunn & Stanley 1998). Monophyly of *T. cauta cauta* and *T. cauta steadi* was not recovered in any analyses. Minor plumage and morphometric differences with overlap exist between these taxa (Double et al. 2003), and microsatellites and mitochondrial DNA suggest low differences (Chambers et al. 2009) with some degree of genetic isolation (Abbott & Double 2003a, Abbott & Double 2003b). Fine-scale genomic work would prove useful to more robustly test the appropriate taxonomic ranks of these taxa. Mitochondrial and microsatellite evidence of differentiation of the Falkland Islands population of *T. melanophris* from other populations (Burg & Croxall 2001, Burg et al. 2017) was not replicated here despite sampling individuals from the Falkland Islands, South Georgia, Kerguelen Islands, and New Zealand. A finer-scale dataset of genome-wide markers would prove useful in assessing these populations.

2.4.2.2 Oceanitidae

The well resolved and supported phylogeny of Oceanitidae represents the first time genome-wide data have been used to assess relationships within the family. In the genus *Oceanites*, the paraphyly of subspecies *O. oceanicus oceanicus* and *O. oceanicus exasperatus* was consistent and well supported. One individual of *O. oceanicus oceanicus* from the Kerguelen Islands, Indian Ocean, was sister to a monophyletic clade of three individuals of *O. oceanicus oceanicus* and *O. oceanicus exasperatus* from the South Atlantic Ocean and Southern Ocean (Antarctic Peninsula, South Shetland Islands, and Falkland Islands). The subspecies *O. oceanicus exasperatus* is larger in size than the nominate subspecies, but this may be clinal in nature (Brooke 2004). The currently used subspecies names do not reflect population structure inferred here, and subsuming *O. oceanicus exasperatus* within *O. oceanicus oceanicus* is recommended. The differentiation of Indian Ocean from South Atlantic populations may be more valid, and warrants further population genetic study.

The subspecies *O. oceanicus chilensis* was fully supported in all analyses as sister to *O. gracilis*, rather than the other subspecies of *O. oceanicus*. This taxon is smaller and shows minor plumage differences to *O. oceanicus oceanicus* and *O. oceanicus exasperatus* (Harrison

et al. 2013, Onley & Scofield 2007). It is recommended that this taxon is recognised as the species *Oceanites chilensis*, for which the nomenclature was clarified by Palma et al. (2012).

Relationships among *Fregetta grallaria* subspecies were well resolved in this study. However, the relationships of *F. grallaria* (white-bellied storm-petrel) and *F. tropica* (black-bellied storm-petrel) populations have proved difficult to resolve. The existence of white-bellied forms of the black-bellied storm-petrel (*F. tropica*) that are morphologically similar to the white-bellied storm-petrel (*F. grallaria*) have caused particular confusion (Robertson et al. 2016). Robertson et al. (2016), using cytochrome *b* and a single nuclear marker, concluded that there is a white-bellied *F. tropica* form (*F. tropica melanoleuca*) at Gough Island and Inaccessible Island (South Atlantic Ocean), which was thought to be *F. grallaria leucogaster*, and that there was no definitive evidence that *F. grallaria* had occurred at Gough Island. One *Fregetta* individual from Gough Island was sampled here, and grouped conclusively with *F. grallaria* samples, indicating that *F. grallaria leucogaster* does occur at Gough Island, or at least did when this sample was collected. Therefore, sympatric forms of *F. grallaria leucogaster* and *F. tropica melanoleuca*, both with white bellies, do, or at least did, occur at Gough Island. Furthermore, Robertson et al. (2016) found that *F. grallaria leucogaster* and *F. tropica melanoleuca* (as a white-bellied form) occur sympatrically at Inaccessible Island. Of the three individuals of presumed *F. grallaria leucogaster* sampled here, two were from St. Paul Island, a relict population in the Indian Ocean (Tollu 1984). Intriguingly, one of these individuals placed with the *F. grallaria leucogaster* individual from Gough Island, and the other was sister to the four *F. tropica tropica* individuals sampled (from South Georgia, South Shetland Islands, Bouvet Island, and the Kerguelen Islands). Thus, the situation of sympatric populations of *F. tropica* and *F. grallaria*, both with white bellies, on Gough Island and Inaccessible Island in the South Atlantic Ocean appears to be replicated in the small population on St. Paul Island in the Indian Ocean. This individual, notated as *F. grallaria leucogaster 3* appears to actually represent *F. tropica melanoleuca*. This repeated pattern of sympatric cryptic species now demonstrated in three island populations in two different oceans poses questions of the identity of further *Fregetta* populations worldwide. Additional work is needed to fully disentangle the taxonomy of these species.

Further to this, greater resolution than previously available of the relationships of the subspecies of *F. grallaria* was provided in this study. The subspecies *F. grallaria titan* has been considered a distinct species based on its large size (e.g. Harrison et al. 2021), but was embedded within the *F. grallar*ia clade here and its recognition as a subspecies is supported. Robertson et al. (2016) inferred this subspecies as sister to *F. grallaria segethi*. However, the

results presented here indicate that *F. grallaria segethi* is the deepest divergence within this species.

The relationships of the subspecies of *Pelagodroma marina* were fully supported in all analyses with *P. marina marina* and *P. marina maoriana* as sisters, and a further trio of *P. marina dulciae*, *P. marina hypoleuca*, and *P. marina eadesorum*. Silva et al. (2015) sampled *P. marina hypoleuca*, *P. marina marina* and *P. marina maoriana*, finding a comparable sister relationship between the latter two subspecies. Further sampling of *P. marina eadesorum* and *P. marina dulciae* in this study added further resolution to the relationships within this species, and identified an interesting pattern of diversification. The two North Atlantic subspecies (*P. marina hypoleuca* and *P. marina eadesorum*) shared a common ancestor with the south Australian subspecies *P. marina dulciae*, and the South Atlantic subspecies *P. marina marina* shared a common ancestor with the New Zealand breeding *P. marina maoriana*.

2.4.2.3 Hydrobatidae

All 21 taxa of the family Hydrobatidae were sampled, and relationships were well resolved and supported. This represents the first time genome-wide data has been applied to the relationships within Hydrobatidae, and is a significant upgrade in resolution compared to the most complete previous phylogeny based on cytochrome *b* and five nuclear introns (Wallace et al. 2017), and a partial phylogeny of Pacific Ocean species based on the mitochondrial gene ND1 (Sausner et a. 2016). Deeper divergences of clades within Hydrobatidae, that could not be confidently inferred by Wallace et al. (2017), were inferred here with full support. The sister species *Hydrobates pelagicus* and *Oceanodroma furcatus* were closest related to the species *O. jabejabe*, *O. monteiroi*, and *O. castro*, and were separated from all other species. The larger clade of mostly Pacific Ocean breeding species was split into three groups. Two species trios, *O. melania*, *O. microsoma* and *O. tethys*, and *O. tristrami*, *O. matsudairae* and *O. monorhis* formed monophyletic groups. Finally, *O. homocroa*, the sister species *O. markhami* and *O horbyi*, and the species *O. leucorhous*, *O. cheimomnestes* and *O. socorroensis* formed a monophyletic group.

As previously found, *Hydrobates* and *Oceanodroma* are not monophyletic genera (Nunn & Stanley 1998), leading some authors to subsume all *Oceanodroma* taxa into *Hydrobates*, that has priority (e.g. Carboneras & Bonan 2014, Wallace et al. 2017). The well resolved deeper-level clades found in this study should provide a basis for redefining the genera within Hydrobatidae, for which there is a clear need.

At the species level, this study provides many new insights into Hydrobatidae evolution. A sister relationship between *O. melania* and *O. markhami* with a very short branch length between species was inferred by Wallace et al. (2017). This was likely caused by a misidentified specimen of *O. markhami*, as in the tree inferred here *O. melania* is sister to *O. microsoma* and *O. tethys*, the same species as in the Wallace et al. (2017) tree, but a sister relationship of *O. markhami* with *O. hornbyi* is instead confirmed here. The mitochondrial ND1 tree presented by Sausner et al. (2016) also separated *O. melania* and *O. markhami* into the clades found in this study. A sister relationship between *O. tristrami* and *O. homochroa* was inferred by Wallace et al. (2017); instead a sister relationship of *O. tristrami* to the species pair *O. matsudairae* and *O. monorhis* is confirmed here, and *O. homochroa* forms a monophyletic group with *O. markhami*, *O horbyi*, *O. leucorhous*, *O. cheimomnestes* and *O. socorroensis*.

2.4.2.4 Procellariidae

The phylogeny of Procellariidae was well resolved, and provided genome-wide data for the first time for almost all clades. In the genus *Macronectes*, individuals of *M. giganteus* were paraphyletic with respect to *M. halli*. Two *M. giganteus* individuals from Gough Island were sister to *M. halli* individuals from South Georgia and New Zealand, and a third *M. giganteus* from South Georgia placed outside these pairs. The identity of the giant petrel population breeding at Gough Island has been debated, and they have been given a subspecies name *M. giganteus solanderi* (Salomon & Voisin 2010, Voisin & Bester 1981). A phylogeographic study of *Macronectes* populations using cytochrome *b* and microsatellites found paraphyletic relationships of the two species and significant differences in the Gough population of *M. giganteus* from other populations (Techow et al. 2010). The results presented here suggest that the Gough population is more closely aligned with *M. halli*, and further work is needed to define the population-level species identities and relationships in *Macronectes*.

Relationships of the *Fulmarus glacialis* subspecies were concordant with those found by mitochondrial and nuclear makers (Kerr & Dove 2013) that the deepest divergence was of the Pacific Ocean subspecies *F. glacialis rodgersii* from Atlantic Ocean subspecies *F. glacialis glacialis* and *F. glacialis auduboni*. Although only one individual of *F. glacialis auduboni* was sampled here, the two Atlantic Ocean subspecies were monophyletic, unlike the results of Kerr & Dove (2013).

In the genus *Pseudobulweria* a linear branching pattern of *P. macgillivrayi*, *P. aterrima*, *P. becki* and *P. rostrata* was found. A study of the genus using two mitochondrial markers and one nuclear marker inferred either *P. macgillivrayi*, *P. aterrima*, or a sister relationship of *P. aterrima* and *P. macgillivrayi* as the deepest divergence within the genus in different analyses (Gangloff et al. 2012), and the former topology is confirmed here.

In the genus *Procellaria*, a linear branching pattern of *P. westlandica*, *P. cinerea*, *P. conspicillata* and *P. aequinoctialis* was found, in contrast to a study using cytochrome *b* that found *P. cinerea* as sister to all other congeners (Techow et al. 2009).

Relationships of the shearwater genera *Calonectris* and *Ardenna* were identical to those found by Ferrer Obiol et al. (2021), that used genome-wide UCE and RADseq loci. Within the species *A. carneipes*, an individual from the remote Indian Ocean colony of St. Paul Island was sister to an individual from Western Australia, its closest geographical colony. Individuals from New Zealand and the north Pacific Ocean were sister. The populations of New Zealand and Lord Howe Island have been considered a separate subspecies, *A. carneipes hullianus*, from the populations further west in Australia and St. Paul Island (Hindwood 1945), for which there is some mitochondrial and nuclear genetic evidence (Lombal et al. 2018). Monophyly of these populations in this study supports the recognition of this taxon, but further fine-scale work would be useful to evaluate this.

Relationships of the shearwater genus *Puffinus* were identical to those found by Ferrer Obiol et al. (2021) using UCE and RADseq loci, other than the position of *P. puffinus*, and the addition of several species and subspecies sampled here that are discussed below.

P. subalaris was not sampled by Ferrer Obiol et al. (2021), and has been inferred as sister to *P. nativitatis* using cytochrome *b* (Austin et al. 2004, Pyle et al. 2011). It was unambiguously sister to all *Puffinus* species after *P. nativitatis* here, instead of a sister relationship of these species. In this study, a conflicting node was identified with *P. puffinus* sister to *P. lherminieri*, *P. baroli* and *P. boydi*, with this clade sister to *P. yelkouan* and *P. mauretanicus*. Ferrer Obiol et al. (2021) inferred *P. puffinus* instead as sister to *P. yelkouan* and *P. mauretanicus* in most analyses, and they suggested nearly simultaneous divergences of these lineages (see section 2.4.1).

The sister relationship of *P. assimilis tunneyi* and *P. assimilis kermadecensis* found here was concordant with cytochrome *b* results (Austin et al. 2004, Kawakami et al. 2018). However, *P. assimilis haurakiensis* was sister to the other subspecies of *P. assimilis*, and with *P. elegans* placing outside the *P. assmilis* clade in cytochrome *b* studies (Austin et al. 2004, Kawakami et al. 2018, Pyle et al. 2011). Here, *P. elegans* was sister to *P. assimilis haurakiensis* and

embedded within the species *P. assimilis*. Internodal branch lengths within the *P. elegans* and *P. assimilis* clade were long relative to other *Puffinus* species, and further study of these lineages to clarify taxonomic ranks should be a priority. In the absence of further sampling and sequencing, it is recommended that *P. elegans* is subsumed into *P. assimilis* as the subspecies *P. assimilis elegans*.

A further clade inferred here contained *P. bailloni*, *P. opisthomelas*, *P. bannermani*, *P. myrtae*, *P. auricularis*, and *P. newelli*. Martinez-Gomez et al. (2015), using cytochrome *b*, recommended lumping *P. auricularis* and *P. newelli* after finding non-monophyly of samples of these species. Individuals of these species were monophyletic here, and their recognition as separate species is recommended. *P. myrtae*, breeding only at Rapa Island in the Pacific Ocean, was described as a subspecies of *P. assimilis* (Bourne 1959), and has often been considered conspecific with *P. newelli* after Austin et al. (2004), where a sister relationship was inferred. It was separated from the sister pair *P. newelli* and *P. auricularis* in this study, and its recognition as a species is recommended.

P. bannermani, restricted to the Ogasawara Islands of Japan, was sequenced for the first time by Kawakami et al. (2018) using cytochrome *b*, and the authors recommended its recognition as a species. It formed a monophyletic clade with *P. myrtae*, *P. newelli*, and *P. auricularis* in Kawakami et al. (2018) and in this study, and its recognition as a species is further supported.

In cytochrome *b* studies, the position of *P. opisthomelas* has been either unresolved (Kawakami et al. 2018, Martinez-Gomez et al. 2015, Pyle et al. 2011) or sister to a clade containing *P. newelli*, *P. myrtae*, and *P. bailloni* (Austin et al. 2004). In this study, *P. opistomelas* is confirmed as sister to the clade containing *P. bannermani*, *P. myrtae*, *P. auricularis*, and *P. newelli*, and it was sister to *P. newelli* in the UCE and RADseq phylogeny of Ferrer Obiol et al. (2021) where *P. bannermani*, *P. myrtae*, and *P. auricularis* were not sampled. *P. bailloni* is sister to this clade in this study and in Ferrer Obiol et al. (2021), rather than embedded within it as in Austin et al. (2004). Within *P. bailloni*, *P. bailloni bailloni* was the deepest diverged subspecies, as in genome-wide (Ferrer Obiol et al. 2021) and cytochrome *b* studies (Kawakami et al. 2018, Pyle et al. 2011). In this study *P. bailloni dichrous* was sister to *P. bailloni nicolae* and *P. bailloni colstoni*, concordant with cytochrome *b* studies (Austin et al. 2004, Kawakami et al. 2018, Pyle et al. 2011) and with Ferrer Obiol et al. (2021), where *P. bailloni colstoni* was not sampled. *P. bailloni nicolae* and *P. bailloni colstoni* were not well supported as reciprocally monophyletic in all analyses, and the recognition of *P. bailloni colstoni* remains questionable (Austin et al. 2004, Bretagnolle et al. 2000).

In the genus *Pelecanoides*, *P. georgicus* was inferred as the sister to the species pair *P. magellani* and *P. urinatrix*. In the cytochrome *b* phylogeny presented by Nunn & Stanley (1998), *P. urinatrix* was sister to the species pair *P. georgicus* and *P. magellani*. However, the gene sequences of these species uploaded to GenBank from that study (Nunn & Stanley 1998) have mismatched names (Grosser et al. 2021), and thus the relationships found are concordant. A further cytochrome *b* study found the same species-level relationships (Grosser et al. 2021) as in this study.

The relationships of subspecies within *P. urinatrix* were not entirely well supported. *P. urinatrix exsul* formed a monophyletic clade of individuals from South Georgia, the Kerguelen Islands, and Heard Island without any strongly supported structure. *P. urinatrix urinatrix* and *P. urinatrix berard* were sister subspecies, as were *P. urinatrix chathamensis* and *P. urinatrix dacunhae*, albeit with low support in ML analyses. Using cytochrome *b*, Grosser et al. (2021) similarly found monophyletic clades of *P. urinatrix urinatrix* and *P. urinatrix exsul*, but non-monophyly of other subspecies and no strongly supported relationships. The data here support the recognition of *P. urinatrix exsul* and at least one further subspecies among *P. urinatrix urinatrix*, *P. urinatrix berard*, *P. urinatrix chathamensis* and *P. urinatrix dacunhae*, for which further sampling and sequencing will be required to determine.

The genus *Pachyptila* was atypical in producing poorly resolved relationships at the species and subspecies level. This is, however, typical of the difficulty that resolving the taxonomy of *Pachyptila* has presented (e.g. Bretagnolle et al. 1990, Cox 1980, Falla 1940, Fleming 1941, Harper 1980, Mathews 1912). All species are morphologically similar and differ mainly in bill morphometrics, which are related to foraging behaviour and diet (Harper 1980, Warham 1990). The deepest level divergence inferred in this study was of *P. turtur* and *P. crassirostris* from *P. desolata*, *P. belcheri*, *P. vittata*, *P. macgillivayi* and *P. salvini*. Using 25 microsatellites and cytochrome *b*, Moodley et al. (2015) similarly found the deepest divergence of *P. turtur* from other species (*P. crassirostris* was not sampled), with relationships within the larger clade varying by markers and methods used. Masello et al. (2019) used the same panel of 25 microsatellites and short fragments of mitochondrial and nuclear DNA, finding that *P. vittata* was sister to all other species. Population structure analysis found little to no species-specific structure and high admixture between species, inferring an optimal value of three species among *P. turtur*, *P. desolata*, *P. belcheri*, *P. vittata*, and *P. salvini* (Masello et al. 2019). Masello et al. (2022) provided an updated study that included samples of *P. macgillivrayi* using cytochrome *b*, COI, and the same previously used microsatellite loci. This study found that *P. turtur* was the deepest divergence of the genus (Masello et al. 2022), concordant with results

found in this study. Similar to the previous population structure analysis (Masello et al. 2019), Masello et al. (2022) found high admixture and inferred an optimal value of two species among *P. desolata*, *P. vittata*, *P. salvini*, and *P. macgillivrayi*.

Shepherd et al. (2022) investigated the relationships of *P. turtur* and *P. crassirostris* using mitochondrial cytochrome *b*, COI, and nuclear RADseq data, finding that neither species were monophyletic. *P. crassirostris pyramidalis* of the Chatham Islands was sister to *P. turtur turtur* individuals of the Chatham Islands rather than to other individuals of *P. crassirostris* (Shepherd et al. 2022). Notably, *P. crassirostris pyramidalis* was sister to the only individual of *P. turtur turtur* from the Chatham Islands in this study, and results in this clade are similar to Shepherd et al. (2022). Shepherd et al. (2022) recommended including *P. crassirostris flemingi* from Heard Island (Tennyson & Bartle 2005) under the name *P. turtur eatoni* based on its COI haplotype (cytochrome *b* and RADseq data were not available for this taxon). Conversely, *P. crassirostris flemingi* was fully supported as the deepest divergence from all other individuals of *P. turtur* and *P. crassirostris* in this study, and this population requires further study to determine its affinities. Similar to studies of the larger *Pachyptila* clade (Masello et al. 2019, Masello et a. 2022), population structure analysis of *P. turtur* and *P. crassirostris* populations inferred an optimal value of one or two species, indicating little genetic divergence between species and populations (Shepherd et al. 2022).

Pachyptila species can cover vast distances (Navarro et al. 2015, Quillfeldt et al. 2015), and show evidence of high gene flow, admixture, and migration rates, and a lack of genetic structure between populations in several species (Masello et al. 2019, Masello et al. 2022, Quillfeldt et al. 2017). A hybrid origin of the species *P. salvini* between putative parent species *P. desolata* and *P. vittata* has been proposed based on intermediate morphology and a highly admixed genetic profile (Masello et al. 2019). The traditional delineation of taxa based on bill morphology is not borne out by genetic data (this study, Masello et al. 2019, Masello et al. 2022, Shepherd et al. 2022), suggesting that these traits may be plastic, evolving rapidly in response to ecological conditions. The widely disparate populations with ongoing gene flow may instead represent ecomorphs of a small number of wide-ranging species. Fine-scale sequencing data, particularly of the larger clade of *P. desolata*, *P. belcheri*, *P. vittata*, *P. macgillivayi* and *P. salvini* to supplement the RADseq data produced for *P. turtur* and *P. crassirostris* (Shepherd et al. 2022) would be useful to define accurate taxonomic boundaries in this genus that have been an elusive target.

The relationships of the genus *Pterodroma* have thus far been investigated using only cytochrome *b* (e.g. Nunn & Stanley 1998, Welch et al. 2014, Wood et al. 2017), with limited

taxon sampling and resolution. The results presented here are a significant improvement in resolving the relationships of the genus.

A clade of Caribbean and North Atlantic Ocean species of *P. hasitata*, *P. madeira*, *P. cahow*, *P. feae* and *P. deserta* was inferred, with conflicting placement of *P. cahow* (see section 2.4.1). Studies using cytochrome *b* have found a branching pattern of *P. hasitata*, *P. cahow*, *P. madeira*, *P. deserta*, and *P. feae* with a short branch between the divergence of *P. cahow* and sister species (Jesus et al. 2009, Welch et al. 2014). A phylogenetic network of this clade indicated rapid divergences between these species, and finer-scale sequencing using RADseq may prove useful in resolving these divergences.

This clade was sister to a clade comprising a branching order of *P. mollis*, *P. gouldi*, *P. incerta*, *P. lessonii*, and *P. macroptera*. *P. mollis* placed outside this smaller clade, and sister to this and the preceding clade of North Atlantic Ocean species in cytochrome *b* studies (Jesus et al. 2009, Nunn & Stanley 1998, Welch et al. 2014, Wood et al. 2017). *P. gouldi* was previously considered synonymous with *P. macroptera*, but was not sister to this species in this study, confirming the separation of these species found by Wood et al. (2017) using cytochrome *b*.

The larger clade of *Pterodroma* species inferred here clarified the positions of several species with variable placement in cytochrome *b* studies (Nunn & Stanley 1998, Welch et al. 2014, Wood et al. 2017), and sampled several taxa for the first time. The increase in resolution is such that it is difficult to compare to previous phylogenies of this group. *P. hypoleuca* was clarified as the base of a large clade of species. A clade of *P. longirostris*, *P. cookii*, *P. brevipes*, and *P. leucoptera* was inferred. The position of *P. solandri* was conflicting with respect to *P. nigripennis* and further species (see section 2.4.1), and further study of these species is needed. *P. inexpectata* was sister to two monophyletic clades. The first contained *P. ultima*, *P. atrata*, *P. neglecta*, *P. heraldica*, and *P. alba*. The dark morph populations of *P. heraldica* were split as the species *P. atrata* by Brooke & Rowe (1996) and these species are not sister here, supporting the recognition of *P. atrata*. The second clade contained *P. externa*, *P. sandwichensis*, *P. baraui*, *P. cervicalis*, and *P. occulta*.

2.4.3 Conclusions

In summary, a well resolved and well supported hypothesis of the evolutionary relationships of Procellariiformes based on thousands of genome-wide markers was presented. Discordance between this study and previous works mainly making use of mitochondrial sequences was frequent. The enhanced power of genome-wide markers illustrates the need to revisit previous

work based on mitochondrial markers, or small numbers of mitochondrial or nuclear markers, to provide robust, updated phylogenetic hypotheses. In addition to this, further taxon sampling to fill systematic gaps is a clear benefit, and provides the next step towards a dense and comprehensively resolved tree of life. Despite the need for some fine-scale work to resolve remaining questions in Procellariiformes systematics, the well resolved and well supported consensus topology generated should provide useful information for additional work into the taxonomy of this clade, and comparative context for further evolutionary and ecological study. Clarification of the relationships between taxa should also prove useful for determining future conservation measures in this highly threatened group.

Chapter 3: Investigating the impacts of sex-linked genomic markers on phylogenetic reconstruction in Procellariiformes seabirds

3.1 Introduction

The collection of large-scale sequencing datasets has led to an era of phylogenomics rather than phylogenetics, with advances in power to resolve relationships (Delsuc et al. 2005) and, conversely, in opportunities for incongruence to arise (Bravo et al. 2019, Jeffroy et al. 2006, Philippe et al. 2011). A multitude of processes that can occur at the level of single genes or non-coding regions, such as introgression, incomplete lineage sorting (ILS), natural selection, and horizontal gene transfer generate heterogeneous histories across genomes, and impede our ability to resolve the true tree of life in the form of a dichotomously branching phylogeny (Bravo et al. 2019, Hobolth et al. 2011, Knowles et al. 2018, Mallet et al. 2016). In addition to the problems posed by these naturally occurring processes, artefactual impediments introduced by error in sequence assembly or alignment, model inadequacy, and other facets of phylogenetic methods pose further problems (Blom et al. 2017, Reddy et al. 2017, Shen et al. 2021, Xi et al. 2015), and solutions will require advances across all steps of the tree building process (Bravo et al. 2019). Within our current control, marker selection remains a critical component of tackling phylogenetic questions, for example of coding or non-coding regions, and has been frequently implicated as a determinant of phylogenetic outcomes and a source of incongruence (e.g. Gilbert et al. 2015, Karin et al. 2020, King & Rokas 2017, Rokas et al. 2003).

Birds provide several exceptional examples of the importance of marker selection in determining difficult phylogenetic relationships. Within the Paleognathae, a clade encompassing tinamous and flightless ratites that are sister to all other birds, short internal branches and extensive ILS have contributed to an anomaly zone that has obscured relationships inferred between different classes of markers, including conserved nonexonic elements, introns, and UCEs (Cloutier et al. 2019). Similarly, the higher-level relationships of the Neoaves, containing almost all extant birds, has received a great deal of phylogenetic attention with inconsistent results (e.g. Hackett et al. 2008, Jarvis et al. 2014, Kimball et al. 2013, Kuhl et al. 2021, McCormack et al. 2013a, Prum et al. 2015, Suh 2016), of which a strong component has been attributed to data type effects related to marker selection (Braun & Kimball 2021, Braun et al. 2019, Reddy et al. 2017).

In addition to the choice between distinct classes of phylogenetic markers, further efforts have been directed at best practices within marker types. Empirical studies using newly generated

datasets and reviews based on empirical and simulated datasets have explored the impact of a wide variety of filtering techniques on sites and loci aimed at amplifying phylogenetic signal and reducing noise, including filtering based on thresholds of missing data, informative sites, GC content, gene tree estimation error, genomic characterisation and separation, and other parameters of the alignment process itself (Blom et al. 2017, Bossert et al. 2017, Gilbert et al. 2018, Hosner et al. 2016, Hutter & Duellman 2023, Mclean et al. 2019, Molloy & Warnow 2018, Portik & Wiens 2021, Simmons et al. 2016, Streicher et al. 2016, Van Dam et al. 2021, White & Braun 2019, Wiens & Morrill 2011). In parallel to the discrepancies in phylogenetic results from different classes of markers, inconsistency remains as to whether these processes of phylogenomic subsampling (Mongiardino Koch 2021) based on a multitude of parameters are preferable or not, or preferable in specific contexts.

Compared to some of these common targets for marker subsampling, the impact of sex-linked markers (i.e. sites or loci on sex chromosomes) has been relatively underexplored in the modern era of genome-wide markers.

Sex determination varies greatly across the tree of life, with systems including male heterogamety (XX/XY, where the Y chromosome determines male development) that is widespread in mammals, amphibians, reptiles and invertebrates, and female heterogamety (ZZ/ZW, where the W chromosome determines female development) found in birds, some amphibians, reptiles, and butterflies (Bachtrog et al. 2014). Despite the independent evolutionary origins of male and female heterogametic systems (Fridolfsson et al. 1998), both systems share several notable characters found only in the heterogametic sex (Y and W chromosomes). The majority of the Y and W chromosomes do not recombine, other than the pseudoautosomal regions (PAR) that are shared with their X and Z counterparts (Hinch et al. 2014, Smeds et al. 2014). Due to sustained recombination, the PAR are similar in sequence identity between the sex chromosomes, in contrast to the non-recombining regions that evolve independently in the absence of recombination (Otto et al. 2011). While the X and Z chromosomes have retained almost all genes from the ancestral autosomes they evolved from (Bellott et al. 2010), the Y and W chromosomes have degenerated (Bellott et al. 2017, Charlesworth & Charlesworth 2000), retaining few of their ancestral genes (Soh et al. 2014) and are rich in repetitive sequences (Itoh & Mizuno 2002, Saitoh et al. 1991, Skaletsky et al. 2003).

Within these overall patterns of heterogametic systems, there is strong variation in size and morphology of the bird W chromosome (Rutkowska et al. 2012), in contrast to the conserved Z chromosome (Stiglec et al. 2007). Several ancient Paleaognathae lineages exhibit

homomorphic sex chromosomes of similar length with extensive recombination (del Priori & Pigozzi 2017, Tsuda et al. 2007, Xu, et al. 2019, Yazdi et al. 2020). Among Neognathae species, the chicken represents the other extreme of the spectrum with an almost completely degenerated W chromosome consisting mainly of repetitive regions and few retained genes (Bellott et al. 2017). There is, however, strong variation in the degree of degeneration and extent of the PAR in the W chromosome across Neognathae species (Zhou et al. 2014a). Despite the generalized patterns observed in sex chromosomes across broad taxonomic scales, further study of non-model species may produce further insights into divergent evolutionary trajectories of sex chromosomes at finer scales.

In addition to the heterogeneous histories of sex chromosomes across species, there is further evidence of their faster mutational rates, greater genetic differentiation between species relative to autosomes, and critical roles in divergence and speciation in a wide variety of taxa and sex determination systems (Irwin 2018, Presgraves 2018, Wilson Sayres 2018), suggesting that sex-linked markers may be a relevant subsample to consider in phylogenetic study.

Gene flow can vary between sexes, with greater female dispersal and male philopatry expected in birds, and the reverse dynamic in mammals (Greenwood 1980). As inheritance of the mitogenome is matrilineal, this pattern of sex-biased gene flow has been suggested as a contributing factor to the frequently encountered discordance in phylogenetic relationships inferred by mitochondrial and nuclear markers (Toews & Brelsford 2012). Furthermore, given the differences in inheritance dynamics of sex chromosomes relative to autosomes between sexes, and divergent evolutionary processes including sexually antagonistic polymorphisms that are expected to arise throughout the evolution of sex chromosomes (Irwin 2018), it is reasonable to anticipate the potential existence of sex-biased effects on phylogenetic reconstruction of both mitochondrial and nuclear markers. Indeed, previous work has found paralogous genes present on the bird Z and W chromosomes, termed ‘gametologs’ (non-recombining paralogs; Garcia-Moreno & Mindell 2000), that grouped samples by Z or W chromosome sequences rather than species identity (Berlin & Ellegren 2006, Garcia-Moreno & Mindell 2000, Kahn & Quinn 1999, Nam & Ellegren 2008, Smeds et al. 2015, Suh et al. 2011). These gametologs, persisting on non-degenerated portions of the W chromosome, have shared ancestry between the sex chromosomes but a lack of recombination leading to independent evolutionary histories (Garcia-Moreno & Mindell 2000), clearly illustrating sex-biased influences on sequence identity.

With these potential biases in mind, some phylogenetic studies including closely related species or populations within species have opted to filter out sex-linked sites or loci or to

separate datasets by sex (e.g. Torres et al. 2021, Winker et al. 2018). However, the effect of sex-linked markers on phylogenetic reconstruction at different scales has rarely been empirically tested. A large-scale phylogenetic study of passerine birds using UCEs and exonic loci indicated discordance between topologies based on a complete dataset of all loci versus only sex-linked loci at a comparable level to those found by the complete dataset versus only coding loci (Harvey et al. 2020). A recent paper on the population genetics of the shearwater species *Ardenna pacifica* found that, when including sex-linked loci, populations separated genomically by sex rather than the expected source populations (Herman et al. 2022). When sex-linked loci were removed, autosomal loci resolved the expected separation based on population sampling (Herman et al. 2022). Although not yet widely investigated, these results indicate that sex-linked loci may introduce biases at deep and shallow scales in phylogenetic datasets and are worthy of further investigation.

The primary aims of this chapter were to characterise sex-linked and autosomal markers in terms of phylogenetic utility, to investigate potential biases and influences of sex-linked markers, and to test for sex-biased effects on phylogenetic reconstruction using the dataset of genome-wide UCEs of the seabird order Procellariiformes generated in Chapter 2 of this thesis. With some previous evidence of sex-linked markers influencing phylogenetic results in a single species of this order (Herman et al. 2022), this taxonomically thorough dataset of Procellariiformes represented an ideal opportunity to test for further evidence across a range of evolutionary scales from the level of families to populations within species. UCEs are distributed throughout the nuclear genome, on autosomal and sex chromosomes, and are mostly non-coding and single copy in vertebrates (Bejerano et al. 2004, Van Dam et al. 2021), representing a suitable set of markers for this purpose. Furthermore, UCEs are frequently used to test phylogenetic relationships at deep and shallow evolutionary scales (e.g. Faircloth et al. 2012, McCormack et al. 2013a, McCormack et al. 2016, Smith et al. 2014), and this study is informative on the best practices for future studies in subsampling and processing of sex-linked markers. Potential causes of phylogenetic influences of sex-linked markers, and the wider relevance to other genetic markers and taxonomic groups were considered.

3.2 Methods

3.2.1 Identifying UCE loci on sex chromosomes by reference mapping

The UCE Tetrapods-UCE-5Kv1 probe set of 5,472 probes, targeting 5,060 unique loci, that was used for sequencing of almost all samples included in previous analyses (see Chapter 2), was originally constructed from homologous sequences in the Carolina anole (*Anolis carolinensis*), chicken (*Gallus gallus*), and zebra finch (*Taeniopygia guttata*) genomes (Faircloth et al. 2012). To identify UCE loci in the Tetrapods-UCE-5Kv1 probe set on sex chromosomes, the probes were mapped to the chicken genome assembly GRCg6a Z and W chromosomes (GenBank accessions CM000122.5 and CM000121.5; Warren et al. 2017), and the zebra finch genome assembly bTaeGut1.4.pri Z chromosome and assembly bTaeGut2 W chromosome (GenBank accessions CM012113.2 and CM018260.2; Warren et al. 2010) using BLASTn (Altschul et al. 1997). Matching UCE loci were assessed by percentage identity of sequences and bit scores, and a combined list of matched loci found on either the chicken or zebra finch sex chromosomes was constructed. Loci matched to the sex chromosomes are hereafter referred to as "ZW loci", and all other loci are referred to as "autosomal loci".

3.2.2 Comparisons of ZW and autosomal UCE loci

Aligned and trimmed loci were separated into ZW loci and autosomal loci based on the results of mapping to the chicken and zebra finch sex chromosomes, and loci with at least 75% taxon occupancy were retained for analyses (see Chapter 2 section 2.2.5 for general methods of processing loci). This resulted in 325 retained ZW loci. Three similar subsets of 325 randomly selected autosomal loci with at least 75% taxon occupancy were created for comparison with the ZW loci. The number of loci and informative sites is important in determining phylogenetic power, and using subsets of autosomal loci rather than all autosomal loci provided a control for relative phylogenetic power. Summary statistics of ZW and autosomal loci subsets were computed using the program phyluce_align_get_align_summary_data within the PHYLUCE pipeline version 1.7.1 (Faircloth 2016). Means and 95% confidence intervals of UCE length, taxon occupancy, missing data, and parsimony informative sites (PIS) were compared. The total number and percentage of PIS within alignments was compared.

3.2.3 Phylogenetic reconstruction of ZW and autosomal UCE loci

The ZW and autosomal loci subsets were concatenated using the PHYLUCE program

phyluce_align_concatenate_alignments. Phylogenetic trees were built for each subset using maximum likelihood (ML) implemented in RAxML-NG version 1.1.0 (Kozlov et al. 2019) with the GTR+G substitution model, matrices partitioned by locus, and 20 ML searches with 10 random starting trees and 10 parsimony starting trees.

To compare pairwise differences between the topologies obtained for each subset, the relative Robinson-Foulds distance (Robinson & Foulds 1981) was calculated using the RAxML-NG command --rfdist. To assess phylogenetic support and conflict in the datasets of each subset several metrics were calculated. The percentage of species with at least 2 individuals included forming monophyletic clades was scored for each subset tree. Non-parametric bootstrapping with 200 replicates was conducted for each subset to assess support for bipartitions in the best scoring trees using RAxML-NG. As a further measure of nodal support, 100 replicates of site concordance factors (sCF) were calculated for each subset in IQ-TREE version 2.2.2.3 (Minh et al. 2020a, Minh et al. 2020b) using alignments trimmed to retain only PIS, prepared using ClipKIT version 1.4.0 (Steenwyk et al. 2020). The updated version of sCF based on ML was used with the --scfl option (Mo et al. 2023). sCF calculates the percentage of decisive sites that support a bipartition in the tree. It compares each of the three possible topologies of quartets and values greater than one third confer support for a bipartition.

ML trees were built for individual ZW loci and autosomal subset loci using RAxML-NG with the GTR+G substitution model and 20 ML searches with 10 random starting trees and 10 parsimony starting trees. These individual loci ('gene') trees were used to calculate gene concordance factors (gCF) in relation to the previously generated best scoring concatenated trees for each subset in IQ-TREE version 2.2.2.3 (Minh et al. 2020a, Minh et al. 2020b). For each branch in the concatenated tree gCF calculates the percentage of gene trees containing that branch, where only gene trees that can contain that branch are considered, as missing data precludes all branches from being present in all gene trees.

3.2.4 Sex-dependent factors in phylogenetic reconstruction across ZW and autosomal loci

Phylogenetic reconstruction of ZW loci revealed distinct clades with elevated branch lengths throughout the tree when compared with autosomal loci subsets (see section 3.3.3). It was hypothesised that differing branch lengths were caused by sex-dependent factors in phylogenetic reconstruction. To test this, the dataset was reduced to contain 116 sampled individuals from museum collections with known sex recorded upon collection. Previously

aligned and trimmed loci separated into ZW loci and autosomal loci were used to create subsets with at least 75% taxon occupancy of the samples of known sex. This resulted in 334 ZW loci, and three autosomal subsets with the same number of loci were created for comparison. These sex-reduced alignments were used to construct phylogenetic trees in RAxML-NG with the GTR+G substitution model, alignments partitioned by locus, and 20 ML searches with 10 random starting trees and 10 parsimony starting trees.

Terminal branch lengths of tips were extracted from the trees to test for an association between sex and branch length. Because the terminal branch length for any tip in a tree is partially determined by how closely related the sister tip is (measured as mean number of substitutions per site in RAxML-NG), only species with at least 2 included individuals in the dataset were retained for testing branch lengths. Using this dataset, means and 95% confidence intervals of branch lengths were calculated for females and males of each subset and unpaired t-tests were used to test for significant differences.

3.2.5 Investigating causes underlying patterns of sex-dependent phylogenetic reconstruction

To test potential causes in sex-related branch length disparities, summary metrics were calculated for female and male alignments and individual samples. Because branch lengths are, in part, determined by substitutions and missing data, the total percentage of sites containing SNPs private to individual samples was calculated for female and male alignments of the ZW loci and autosomal loci subsets. The percentage of private sites was scaled to incorporate the different number of female (51) and male (65) samples included. The percentage of missing data of female and male samples was calculated for the ZW loci and autosomal loci subsets.

To test for relative differences in heterozygosity, allele phasing was conducted for each sample in each subset. The approach of *de novo* assembly of reads into contigs reduces the contigs to consensus sequences with a loss of information on the heterozygote sites that occur in diploid organisms (Andermann et al. 2019, Iqbal et al. 2012). The coverage depth of raw reads can be used instead to call variants and recover phased allele sequences. A phasing approach was used to compare females and males between ZW loci and autosomal loci subsets. Trimmed UCE loci for each sample in each subset were used as individual-specific references to map cleaned reads in FASTQ format against. This was done using the PHYLUCE "mapping" workflow, which aligns the reads to the references, marks duplicates, calculates coverage, and outputs BAM files. The PHYLUCE "phasing" workflow was then used, which uses the BAM files produced in the mapping workflow to produce phased allele sequences for each reference UCE

loci for each sample in each subset. The ratio of heterozygote sites of female and male samples were compared between ZW loci and autosomal loci.

3.3 Results

3.3.1 Identifying UCE loci on sex chromosomes by reference mapping

Mapping UCE probes in the Tetrapods-UCE-5Kv1 probe set to the chicken Z chromosome identified 397 matching probes of 375 unique UCE loci. Mapping probes to the zebra finch Z chromosome identified 396 matching probes of 372 unique UCE loci. Five loci that mapped to the chicken Z chromosome were not found on the zebra finch Z chromosome, and 2 loci mapped to the zebra finch Z chromosome were not found on the chicken Z chromosome. Mapping UCE probes to the chicken W chromosome identified 3 matching probes of 3 unique UCE loci, 2 of which also mapped to the chicken Z chromosome and 1 mapped only to the W chromosome. This loci also mapped to the zebra finch W and Z chromosomes. Mapping probes to the zebra finch W chromosome identified 14 probes of 14 unique loci, all of which also mapped to the zebra finch Z chromosome. This provided a total list of 377 unique UCE loci mapped to the ZW chromosomes, representing approximately 7.5% of the 5060 UCE loci in the Tetrapods-UCE-5Kv1 probe set.

3.3.2 ZW and autosomal UCE loci have different statistical profiles

Summary statistics of ZW loci and three similar autosomal loci subsets of 325 loci each with at least 75% taxon occupancy presented different statistical profiles (Fig. 3.1). Mean UCE length was lower in ZW loci (549.2) compared to the autosomal subsets (598.3-606.2). Mean taxon occupancy was lower in ZW loci (94.8%) compared to the autosomal subsets (97.5-98.4%). Mean missing data was higher in ZW loci (2.6%) than autosomal loci (2.2-2.3%). This metric considers only missing data (encoded as '?') for taxa present in trimmed locus alignments, and is therefore largely distinct from the missing data that ZW loci exhibit relative to autosomal loci in terms of UCE length and taxon occupancy. Despite these differences, mean PIS per locus was higher in ZW loci (91.2) than autosomal loci (83.5-88.3), although with overlap in confidence intervals and greater proportional variation within autosomal subsets than in other metrics. Total PIS as a percentage of alignments was higher in ZW loci (16.6%) than autosomal loci (14-14.6%).

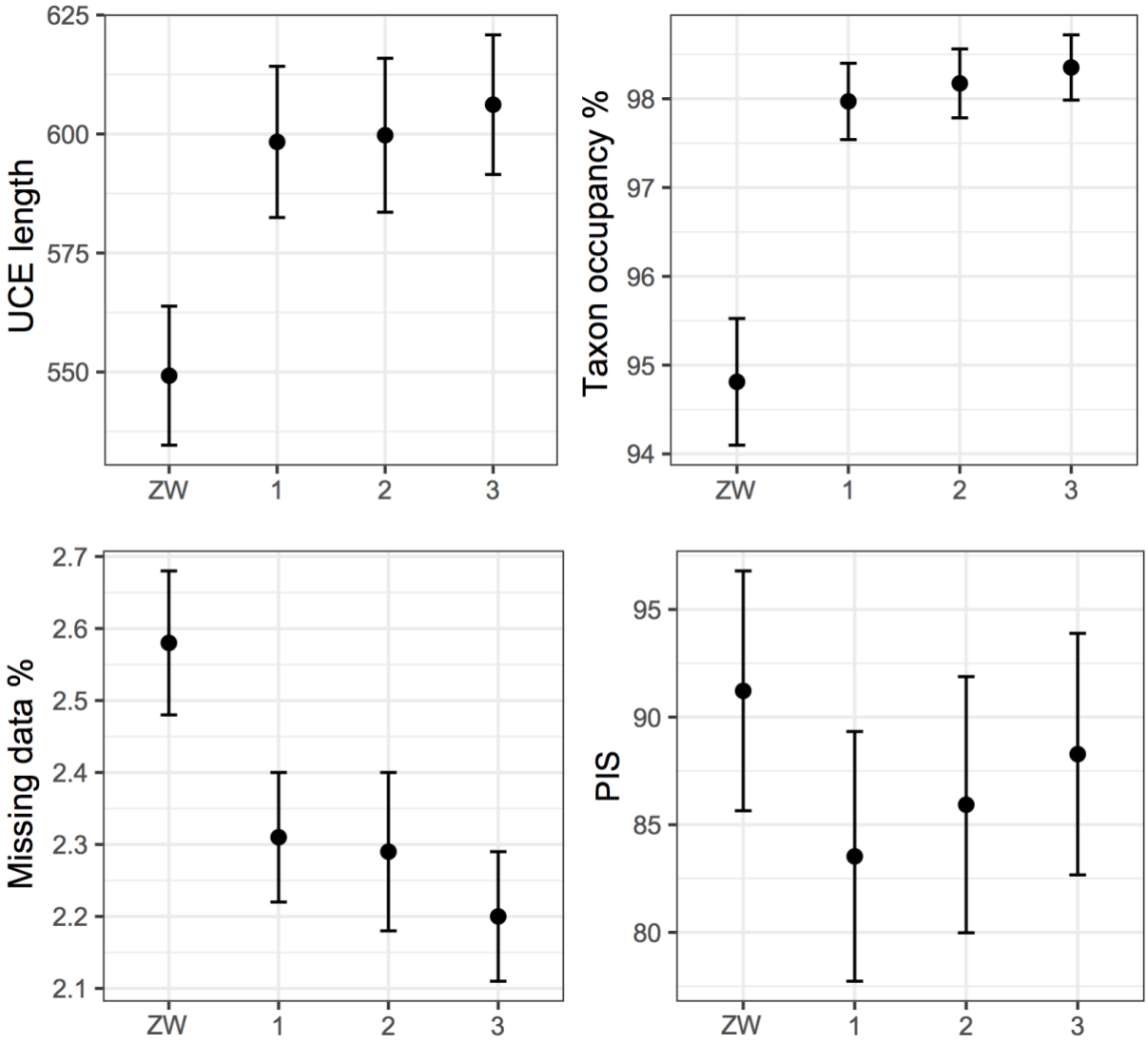


Figure 3.1 Means and 95% confidence intervals of summary statistics of individual UCE loci of ZW loci and three autosomal loci subsets. From left to right: UCE length, percentage of taxon occupancy, percentage of missing data, and parsimony informative sites (PIS).

3.3.3 ZW loci produce lower resolution phylogenetic trees than autosomal loci

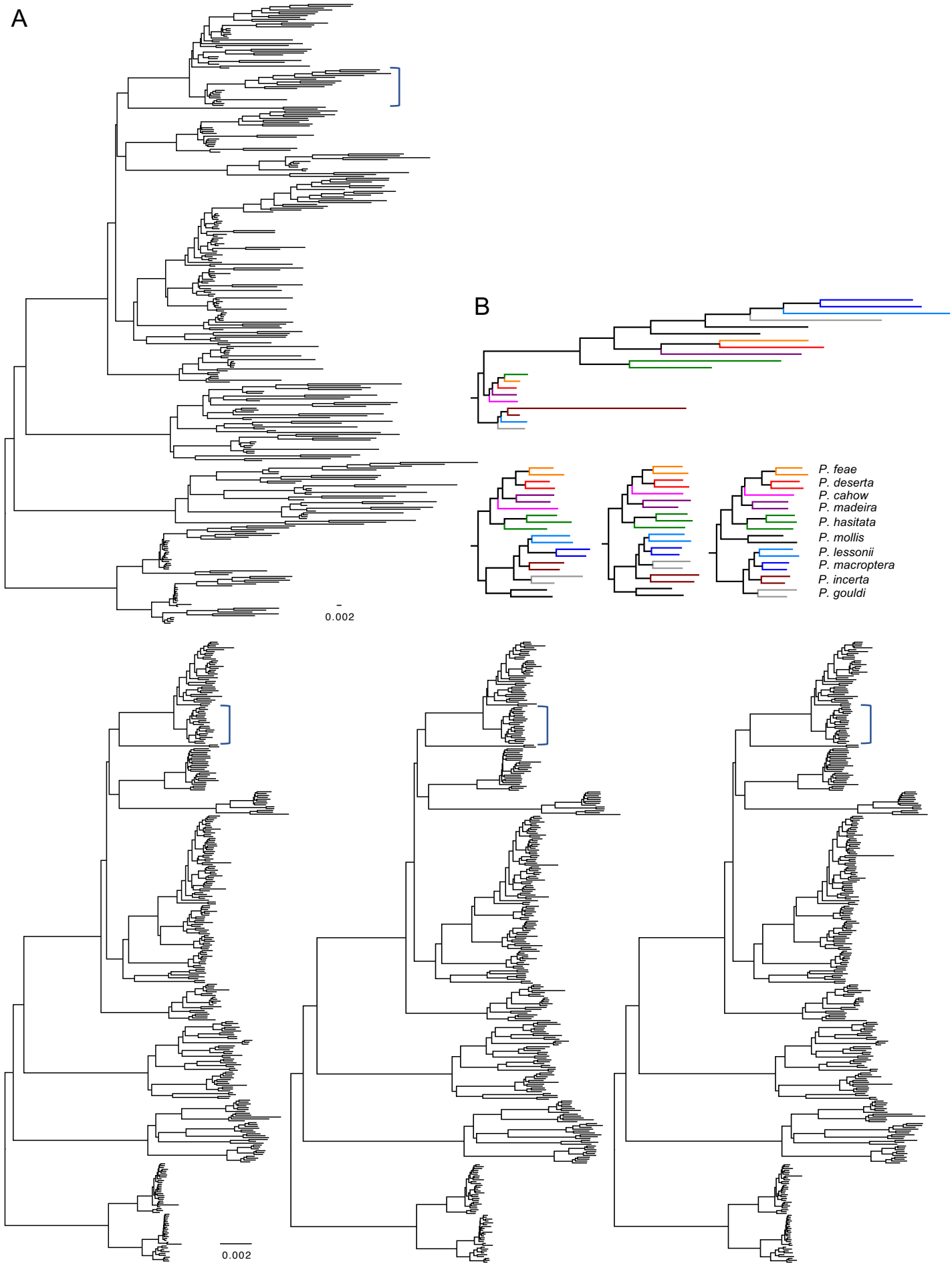
ML trees were built from concatenated ZW loci and autosomal subsets. All trees produced monophyletic families and genera other than the paraphyletic relationship of *Hydrobates* and *Oceanodroma*, a known taxonomic issue (see Chapter 2). The relative Robinson-Foulds distances between the trees (Table 3.1) showed consistently more similar topologies between the autosomal loci subsets (0.22-0.23) than between the ZW and autosomal subsets (0.45-0.46).

Table 3.1 Relative Robinson-Foulds distances between trees inferred from ZW loci and three subsets of autosomal loci.

	ZW	1	2	3
ZW	0	0.45	0.46	0.46
1	0.45	0	0.22	0.23
2	0.46	0.22	0	0.23
3	0.46	0.23	0.23	0

The ZW loci tree exhibited clades separated by branch lengths throughout the tree, which were not present in autosomal trees with relatively consistent branch lengths in clades (Fig. 3.2A). Closer examination of the ZW tree revealed that branch length specific clades frequently did not constitute monophyletic groupings of species, with individuals of species split across branch length specific clades. Autosomal trees did not show the same pattern, for example as visualised in a clade of *Pterodroma* species where the ZW tree did not resolve monoyphyletic species for almost all species, in contrast to autosomal trees (Fig. 3.2B). The overall percentage of species with at least 2 individuals included forming monophyletic clades was largely consistent and higher in trees built from autosomal loci subsets (83.8-88.9%) than the ZW loci tree (59.8%; Figure 3.2C).

Measures of nodal support also indicated differences between the ZW tree and autosomal trees (Fig. 3.2C). Mean bootstrap support per node was lower for the ZW tree (80.9%) compared to autosomal trees (83.3-84.1%) although with overlap in confidence intervals. Concordance of sites in concatenated alignments (sCF) and individual loci (gCF) showed a similar relationship with marginal differences (60% ZW, 61.1-61.8% autosomal for sCF; 13.9% ZW, 16.4-17.6% autosomal for gCF) and overlap in confidence intervals in most cases.


A
0.002
B
P. feae
P. deserta
P. cahow
P. madeira
P. hasitata
P. mollis
P. lessonii
P. macroptera
P. incerta
P. gouldi
0.002

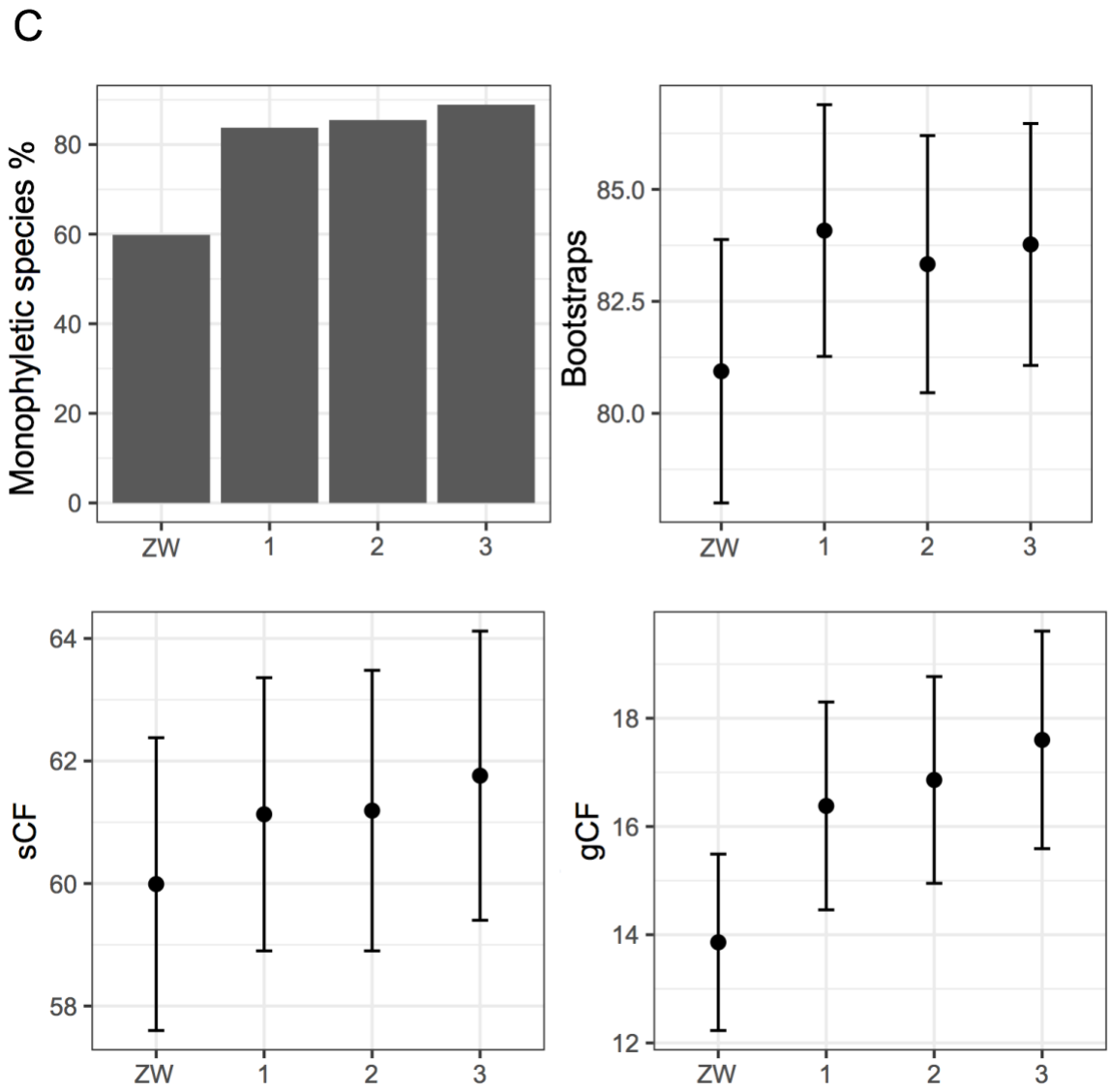


Figure 3.2 A: Maximum likelihood topologies of ZW loci (top) and three randomised subsets of autosomal loci (bottom). Brackets correspond to the enlarged clade shown in B. **B:** Enlarged clade of *Pterodroma* (Procellariidae). Colours indicate species. **C:** Summary statistics of support metrics for ZW loci and autosomal loci subset trees. From top left clockwise: the percentage of monophyletic species recovered for species with ≥2 individuals sampled, means and 95% confidence intervals of bootstrap support per node, sCF per node, and gCF per node.

3.3.4 ZW loci produce sex-dependent tree topologies

To test sex-dependent factors on phylogenetic reconstruction, a reduced dataset of individuals of known sex was created. This resulted in alignments of 334 UCE loci of 116 individual

samples containing 51 females and 65 males for each of the ZW loci and autosomal loci subsets. Trees produced from these alignments displayed similar patterns to the larger dataset of all samples, with branch length specific clades throughout the ZW tree and relatively consistent branch lengths in clades of autosomal loci subset trees (Fig. 3.3). Throughout the ZW loci tree, in almost all cases females formed monophyletic clades with longer branches and males formed monophyletic clades with shorter branches. By contrast, autosomal loci trees did not produce clades grouped by sex. Of the 116 sexed individuals included in these analyses, 3 had a recorded sex that did not match the sex indicated by branch lengths in the ZW loci tree: a female *Phoebastria nigripes*, a male *Oceanodroma melania*, and a female *Ardenna pacifica*.

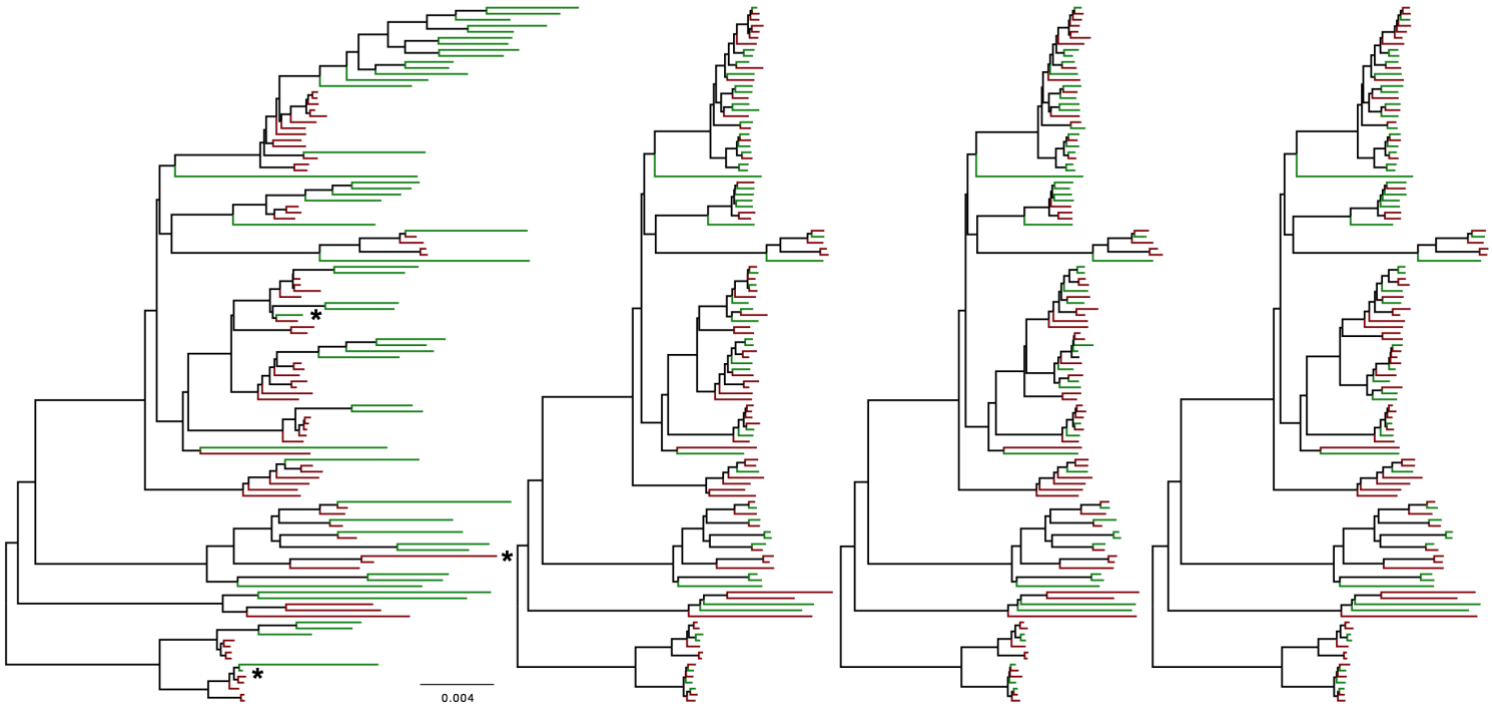


Figure 3.3 Maximum likelihood topologies of ZW loci (left) and three randomised subsets of autosomal loci (right) containing individuals of known sex. Branch colours correspond to sex with females in green and males in brown. Asterisks indicate samples where recorded sex does not match sex indicated by branch lengths.

To further test associations between sex and terminal branch lengths, a further reduced dataset of species with at least 2 individuals of known sex was created, containing 71 individuals of 34 females and 37 males. In the ZW loci, females (mean = 0.0044) had significantly longer branch lengths than males (mean = 0.0008; $t(69) = 10.1$, $p < 0.0001$; Fig. 3.4). Females and males did not have significantly different branch lengths in the three autosomal loci subsets; subset 1: female mean = 0.0006, male mean = 0.0006; $t(69) = 1.1$, $p = 0.2662$; subset 2: female mean = 0.0006, male mean = 0.0006; $t(69) = 1.1$, $p = 0.2901$; subset 3: female mean = 0.0006, male mean = 0.0005; $t(69) = 1.3$, $p = 0.2047$.

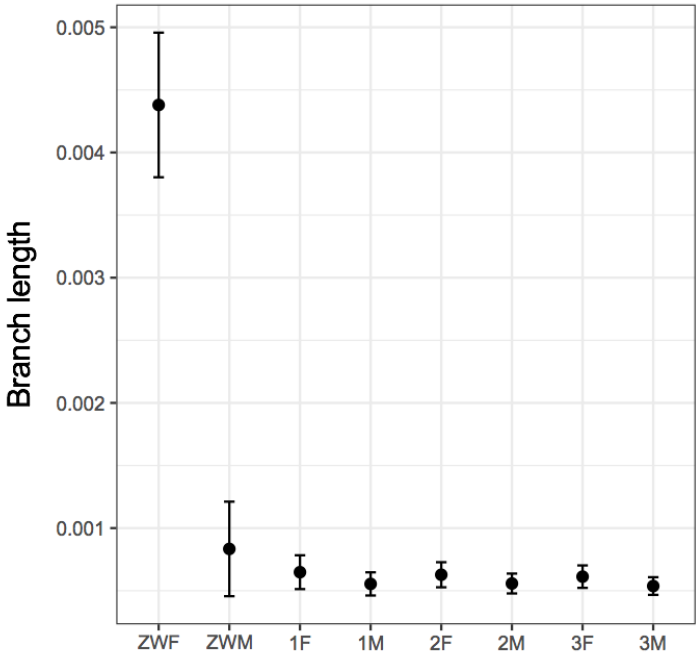


Figure 3.4 Means and 95% confidence intervals of branch lengths of the trees in Fig. 3.3 for females and males of ZW loci and three autosomal loci subsets.

3.3.5 Genomic patterns underlie sex-dependent topologies in ZW loci

To test for potential causes in differences in branch lengths between females and males, the percentage of sites containing SNPs private to individual samples and missing data was calculated for each subset. The difference in total percentage of sites containing SNPs private to individual samples was marginally higher in females compared to males in the ZW loci subset (0.07% in females, 0.04% in males) relative to autosomal loci subsets (0.06% in females

in each subset, 0.04% in males in each subset; Fig. 3.5). The mean percentage of missing data was higher in ZW loci females (11.7%) than males (2.8%), whereas in autosomal loci subsets mean missing data was lower in females (2.8-3.4%) than males (3.5-4.2%; Fig. 3.5).

To test for potential differences in the frequency of heterozygote sites across females and males of ZW loci and autosomal loci subsets, phased allele sequences were produced. Phased allele sequences indicated that females had a higher frequency of heterozygote sites in ZW loci compared to their equivalent samples in autosomal loci (mean 6.98 fold difference), and males had a lower frequency of heterozygote sites in ZW loci compared to their equivalent samples in autosomal loci (mean 0.69 fold difference; Fig. 3.5).

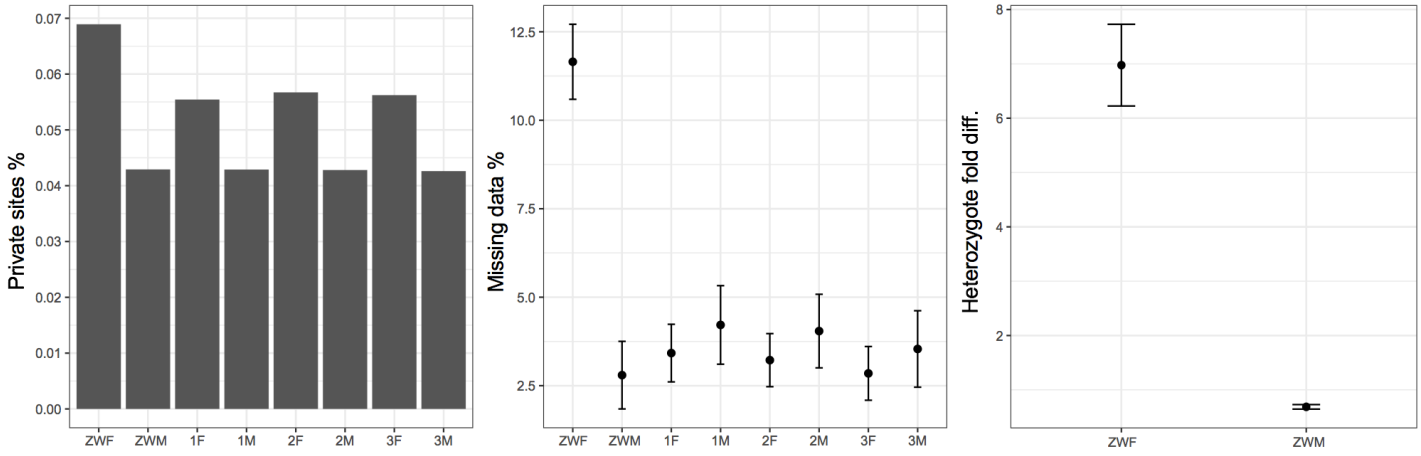


Figure 3.5 Percentage of sites containing private SNPs in alignments of female and male ZW loci and autosomal loci subsets (left), means and 95% confidence intervals of the percentage of missing data of individual samples of female and male ZW loci and autosomal loci subsets (middle), and means and 95% confidence intervals of the fold differences in frequency of heterozygote sites in individual female and male samples between ZW loci and their equivalent samples in autosomal loci subsets (right).

3.4 Discussion

The data presented here indicate a strong role for sex-linked loci in determining phylogenetic results in Procellariiformes. Sex-linked UCEs were statistically divergent from autosomal UCEs and produced unexpected sex-dependent topologies. The differences found between sex-linked and autosomal loci hint at the underlying explanations of these results, that may have

wider impacts on other taxonomic groups and classes of phylogenetic markers. These results represent the first thorough empirical test of the utility of sex-linked markers in genome-wide phylogenetic study, and provide recommendations for treating these markers in future studies.

Mapping UCE probes to the chicken and zebra finch Z and W chromosomes identified 377 sex-linked UCE loci, representing approximately 7.5% of the commonly used Tetrapods-UCE-5Kv1 probe set for studying the phylogenetics of tetrapods. This is similar to the proportion that the Z chromosome comprises of the chicken and zebra finch genome assemblies used for mapping (7.7% and 7.1% respectively). These genome assemblies contained a small percentage of unplaced sequences (1.3% chicken, 0.6% zebra finch), meaning that some sex-linked UCE loci may not have been found by mapping. However, the low percentages of unplaced sequences and of found sex-linked UCEs relative to autosomal UCEs indicates that only a very small number, if any, were not found and this is unlikely to have affected the outcomes of the analyses presented here.

Summary statistics assessing data quality and phylogenetic informativeness presented differing profiles of autosomal and ZW loci. Sex-linked UCEs were shorter, had lower taxon occupancy, and greater missing data. The lower taxon occupancy and greater missing data of ZW loci indicates lower capture success of these loci relative to autosomal loci, and may reflect the lower frequency of the Z chromosome relative to autosomal chromosomes (Wilson Sayres 2018). Despite these differences, ZW loci had higher mean PIS per locus and total PIS in concatenated alignments of subsets. Birds display a trend of lower diversity of sex chromosomes relative to autosomes within populations (Berlin & Elleren 2006, Corl & Ellegren 2012, Irwin 2018, Smeds et al. 2015), but greater diversity of sex chromosomes to autosomes between species (Irwin 2018, Smeds et al. 2015), in keeping with the findings here of greater PIS in ZW than autosomal loci in this diverse set of Procellariiformes species.

UCEs are highly conserved at the core and more variable in flanking sequences (Faircloth et al. 2012). The profile of higher PIS despite shorter ZW loci is anomalous among UCEs, where decreased length usually causes decreased variation and therefore phylogenetic signal (Hosner et al. 2016, Molloy & Warnow 2018, Van Dam et al. 2021). The shorter length of ZW UCEs would suggest a lack of phylogenetic utility, in contrast to their greater PIS that would suggest high phylogenetic utility.

Topologies of trees built from autosomal subsets were consistently different to the topology of the ZW tree, that showed a pattern of clade-specific long branches that did not correspond to species identity. All trees produced expected relationships of monophyletic families and genera, suggesting that the phylogenetic utility of sex-linked loci in Procellariiformes was

limited to deeper nodes, while they produced spurious relationships at the species level and below. In addition to the lower resolution topology, the ZW tree scored worse for all recorded metrics of nodal support. These nodal support metrics refer to the topology constructed from their datasets, and thus do not imply relative topological correctness, rather they illustrate greater phylogenetic conflict and tree sampling variance for ZW loci than autosomal loci.

Trees including individuals of known sex confirmed that clades with long branches were female samples and short branches were male samples. This pattern was almost total: of the 116 sexed individuals, 3 had a recorded sex that did not match the sex indicated by branch lengths. The patterns found in sex-reduced trees are unambiguous between sexes, and it is likely that this small number of individual samples were mistakenly sexed or recorded. Testing the sex of these individuals from voucher specimens would be necessary to confirm this conclusion.

Previous work on increasing phylogenetic utility within markers, including many studies specifically on UCEs, has shown that characterising and subsequent subsampling of individual loci based on typical metrics of PIS, length, and missing data (Hosner et al. 2016, Molloy & Warnow 2018, Streicher et al. 2016, Wiens & Morrill 2011), as well as GC content (Bossert et al. 2017, White & Braun 2019), gene tree estimation error (Blom et al. 2017, Simmons et al. 2016), and intergenic and genic content (Van Dam et al. 2021) can improve results. The results presented here provide evidence that sex-linked loci may be an additional relevant subsample to characterise in phylogenetic studies.

This result is relatively novel. Harvey et al. (2020), applying UCEs and exons to the radiation of the passerine suborder Tyranni, found discordance between trees based on sex-linked loci and their complete dataset of all loci that was similar in scale to the discordance found by comparing coding loci to their complete dataset (Harvey et al. 2020). This is significant as the function of markers as coding or non-coding has been considered a strongly contributing factor to phylogenetic incongruence (Chen et al. 2017, Reddy et al. 2017). Harvey et al. (2020) concluded that these discrepancies could have been caused by the small number of loci present in sex-linked and coding datasets, and was not further investigated. This conclusion was precluded in this study by comparing consistent subsamples of ZW and autosomal loci, suggesting that the differences found are biological in nature rather than artefactual. The effect of sex-linked markers on phylogenetic reconstruction in the modern era of genome-wide studies has received little further attention until now.

Interpreting the potential wider implications of the results presented here requires further understanding of the genetic basis underlying the dynamics of sex chromosome evolution.

After finding that samples of the shearwater species *Ardenna pacifica* grouped genomically by sex rather than population, Herman et al. (2022) hypothesized that sequenced reads from the W chromosome could contaminate paralogous sequence assemblies on the Z chromosome in females, creating spurious alternative alleles and influencing phylogenetic results. Supporting this, Herman et al. (2022) found that females had greater heterozygosity at sex-linked sites than autosomal sites, a pattern that was replicated in this study across diverse Procellariiformes species. As females are heterogametic and have a single copy of the Z chromosome, they should be homozygous at most sex-linked sites if the W chromosome is expected to be mostly degenerated (Charlesworth & Charlesworth 2000). The presence of non-recombining gametologs in bird sex chromosomes, with independent evolutionary trajectories that have accrued sex-specific Z and W alleles (Berlin & Ellegren 2006, Garcia-Moreno & Mindell 2000, Kahn & Quinn 1999, Nam & Ellegren 2008, Smeds et al. 2015) provide a basis for the greater heteroyzgosity seen in female sex-linked loci than autosomal loci in this study and in Herman et al. (2022), and for the sex-dependent tree topologies found in this study. Zhou et al. (2014a), studying the W chromosomes of 17 lineages across birds, uncovered considerable variation in the degree of W chromosome degeneration. Less degenerated W chromosomes may contain more gametologous sequences with the potential to form chimeric contigs assembled from reads from both the Z and W chromosomes in females, and introducing sex-biased results when these contgis/loci are used for phylogenetic reconstruction. Further study of other taxonomic groups, notably the Paleaognathae lineages that exhibit homomorphic, extensively recombining sex chromosomes (del Priori & Pigozzi 2017, Tsuda et al. 2007, Xu et al. 2019, Yazdi et al. 2020), and across other Neognathae species with varying degrees of degenerated W chromosomes may further refine the underlying patterns suspected here of causing sex-linked effects on phylogenetic reconstruction in birds.

In addition to the higher incidence of heterozygote sites found in female sex-linked loci that were expected to cause poor phylogenetic results, a considerably higher proportion of missing data was also found in ZW loci females compared to all other subsets and sexes in this study. Systematically missing data is known to cause problems for phylogenetic inference (Darriba et al. 2016, Simmons 2012). The sex-dependent clades of females with long branches in ZW loci trees is likely the result of a combination of these factors: the incorporation of W chromosome reads into Z chromosome assembled loci creating shared variation in females, and sex-biased missing data creating long branches in females. When combined, these factors may have facilitated the action of long branch attraction in some clades.

The results presented here are unambiguous in that sex-linked loci introduced undesired effects into phylogenetic reconstruction of Procellariiformes seabirds, and that in future studies they could be considered a relevant target for phylogenetic subsampling and filtering from datasets. The degree of these effects may be determined by differing features of the sex chromosomes across lineages and may vary across taxonomic scales. With strong evidence here from Procellariiformes, and some corroborating evidence in a disparate bird lineage within Passeriformes (Harvey et al. 2020), it is possible that the patterns observed here will be consistent across other bird lineages, with the possible exclusion of some Paleognathae lineages that display anomalous sex chromosomes among birds (del Priori & Pigozzi 2017, Tsuda et al. 2007, Xu et al. 2019, Yazdi et al. 2020). Although the female heterogametic system in birds evolved independently from the male heterogametic system in mammals and other lineages (Fridolfsson et al. 1998), they share many features and the results found here may be applicable to other taxonomic groups with varying degeneration of the heterogametic chromosome, for which further work will be needed to determine. Testing for similar effects in other genomic markers beyond UCEs, such as RADseq (Baird et al. 2008) and anchored hybrid enrichment (Lemmon et al. 2012) would provide useful information. As the likely explanations for biases of sex-linked loci on phylogenetic reconstruction are biological in nature, it is unexpected, however, that these methods should be robust to the effects found here.

Chapter 4: Investigating the prevalence and evolution of duplications in the mitochondrial genome of the seabird order Procellariiformes

4.1 Introduction

The animal mitochondrial genome (mitogenome) is considered to be under selection to maintain compactness and gene content and order (Brown et al. 1979, Quinn & Wilson 1993, Rand 1993). Several shared characteristics common to vertebrate mitogenomes provide evidence of such conditions, including the routine 15-20kb size range with few introns or intergenic spacers (Quinn & Wilson 1993), and the strongly conserved genic content of the same set of 13 protein-coding genes, 2 ribosomal RNAs (rRNAs), 22 transfer RNAs (tRNAs), and the control region, a non-coding region with an imprecisely known role in the replication of the mitogenome (Reyes et al. 2005), with few exceptions (Boore 1999, Clayton 1991). Despite the expectations of such a selection regime, duplications of mitochondrial genes and regions have been frequently observed, for example in arthropods (Ogoh & Ohmiya 2007), fish (Lee et al. 2001), snakes (Kumazawa et al. 1998), lizards (Moritz & Brown 1987), birds (Gibb et al. 2007), and other groups (Arndt & Smith 1998).

In contrast to other vertebrates, birds show a different gene order near the control region (Desjardins & Morais 1990). This gene order, from 5' to 3': cytochrome *b* (CYTB), tRNA-Thr, tRNA-Pro, ND6, tRNA-Glu, control region, was first found in the chicken (*Gallus gallus*; Desjardins & Morais 1990) and subsequently in other bird lineages (Mindell et al. 1998) and may represent the ancestral mitochondrial gene order in birds (Desjardins & Morais 1990, Gibb et al. 2007). An alternative, rearranged gene order of CYTB, tRNA-Thr, control region, tRNA-Pro, ND6, tRNA-Glu, non-coding region, was first documented by Mindell et al. (1998) and was found throughout the evolutionary tree of birds, suggesting multiple independent origins (Bensch & Härlid 2000, Gibb et al. 2007). In this gene order the non-coding region varied in length, and in some species contained sequences with high similarity to the control region (Mindell et al. 1998). In some parrot species (Psittaciformes), the non-coding region was found instead to be replaced by a second, intact copy of the control region and a second partially degenerated copy of ND6 was present (Eberhard et al. 2001). A similar pattern was found in species of Accipitriformes and Piciformes, but with no identifiable remnants of ND6 (Gibb et al. 2007). Abbott et al. (2005) identified a similar gene rearrangement in *Thalassarche* albatross species (Procellariiformes) with intact copies of ND6 and the control region, and a duplication

of a short 3' fragment of CYTB, which was partially degraded and partially identical to the functional copy of CYTB.

Gibb et al. (2007) proposed names for these four types of bird mitogenomic arrangements: "ancestral avian" for that found in the chicken with no apparent duplications present (Desjardins & Morais 1990), "remnant CR(2)" for those with a duplicated but degenerated control region found in many orders (Mindell et al. 1998), "duplicate CR" for those with a retained duplicated control region (Eberhard et al. 2001), and "duplicate tThr–CR" for those described in albatrosses with the duplication extending from before the control region (Abbott et al. 2005). Further similar arrangements have been found with an additional conserved duplication of approximately half of CYTB at the 3' end (Cho et al. 2009, Eda et al. 2010, Gibb et al. 2013, Morris-Pocock et al. 2010). Other differences in avian gene order and duplication have been further reported, for example in ruffs (Charadriiformes) where the full duplications were present other than the absence of a second copy of tRNA-Thr (Verkuil et al. 2010). Zhou et al. (2014b) identified two further gene orders with minor differences in herons (Pelecaniformes), and additional gene orders have been documented in parrots (Psittaciformes; Eberhard & Wright 2016) and Passeriformes (Gibb et al. 2015).

The clear homology between duplicated copies of the control region, ND6, fragments of CYTB, and tRNA genes is a common feature of mitogenome rearrangements in birds (Gibb et al. 2013, Urantówka et al. 2020). A tandem duplication and subsequent deletion or degeneration of some stretches of sequence is a likely explanation for the variety of similar yet different mitogenome arrangements that have been observed (Bensch & Härlid 2000, Boore 2000, Macey et al. 1997, Mindell et al. 1998, Moritz & Brown 1986, San Mauro et al. 2006). This process has been known as the tandem duplication and random loss model of mitogenome rearrangement, where one copy of the duplicated genes or control region is usually expected to eventually lose function and gradually degrade or be deleted entirely (Boore 2000). This degeneration following duplication is consistent with the expectation of selection for compact size in mitogenomes. However, the persistence of stable identical or near-identical copies of duplicated regions over large evolutionary timescales poses questions of the possible functions and selection pressures involved. Several mechanisms could explain the tandem duplications that have been observed in animals, for example slipped strand mispairing has been proposed as the mechanism underlying mitogenome duplications in lizards (Fujita et al. 2007). However, the widespread nature of observed duplications are likely to have a multitude of causes (Boore 2000, Mueller & Boore 2005).

The frequent documentation of mitogenome duplications among taxonomically diverse clades of birds has led to the suggestion that it could be considered the default state (Gibb et al. 2013). Indeed, the various configurations of rearrangements are now known to be present in a wide variety of bird lineages across deep evolutionary scales (e.g. Cadahía et al. 2009, Eberhard & Wright 2016, Gibb et al. 2007, Gibb et al. 2013, Gibb et al. 2015, Hanna et al. 2017, Sammler et al. 2011, Urantówka et al. 2018, Urantówka et al. 2020, Verkuil et al. 2010, Wang et al. 2015, Zhou et al. 2014b), while other orders appear to lack any duplications (Urantówka et al. 2018). Although the avian mitogenome lacking any duplication that was originally found in the chicken (Desjardins & Morais 1990) was termed the "ancestral" arrangement in birds (Gibb et al. 2007), newer evidence suggests that mitochondrial duplication was the ancestral state even for the ancient Paleognathae lineages that are sister to all other birds (Urantówka et al. 2020). However, knowledge of when and how frequently independent mitogenome duplication events have occurred in birds remains lacking. Despite the recognition of duplications among highly diverse avian clades, this trait has seldom been tested at large taxonomic scales with dense sampling. In particular, a comprehensive and well-resolved phylogeny is also essential to allow inference of the evolutionary history of mitogenome duplications. A large-scale study of parrots found that the ancestor of this group had a duplicated region of the mitogenome, with subsequent variation in degeneration and loss throughout the tree and further duplications arising in some species (Urantówka et al. 2018). Further studies on Palaeognathae lineages (Urantówka et al. 2020), and on Gruiformes (Akiyama et al. 2017), Accipitriformes and Cathartiformes (Urantówka et al. 2021), and Passeriformes (Caparroz et al. 2018, Mackiewicz et al. 2019) of the Neognathae have offered insights into the evolutionary histories of mitogenome duplications in these groups. Detailed study into the evolution of such traits in further lineages remains lacking.

This study focuses on the Procellariiformes, an order of seabirds comprising four families of 26 genera and approximately 138-145 extant species (Clements et al. 2022, Gill et al. 2023, HBW & BirdLife International 2022), for which a novel, comprehensive phylogeny was generated in Chapter 2 of this thesis. There is some evidence for the presence of mitogenome duplications in Procellariiformes species, for example as found in *Thalassarche* albatross species (Diomedeidae) by Abbott et al. (2005) that comprised intact copies of ND6, the control region, and a short partially degraded 3' fragment of CYTB. Gibb et al. (2007) found similar duplications in a further *Thalassarche* species. Eda et al. (2010) found further duplication in *Phoebastria* species (Diomedeidae), with a similar arrangement to that described by Abbott et al. (2005), with the addition of an identical copy of approximately half of the 3' end of CYTB.

Rains et al. (2011) found duplication in the control region of *Diomedea* species (Diomedeidae), but did not sequence beyond this region. Gibb et al. (2013) found evidence of duplication in *Pelagodroma marina* (Oceanitidae), but full sequences could not be generated. Duplicated copies of the control region have been found in *Oceanodroma castro* (Hydrobatidae; Smith et al. 2007), *Fulmarus glacialis* (Procellariidae; Burg et al. 2014), and *Pterodroma magentae* (Procellariidae; Lawrence et al. 2008) without further evidence as to the gene order or further duplicated regions. Torres et al. (2019) used long-read sequencing to demonstrate intact duplicated copies of ND6 and the control region without any duplicated CYTB fragment in *Puffinus lherminieri* (Procellariidae).

These results indicate that mitogenome duplications may be a common feature in Procellariiformes. Furthermore, mitogenome duplications are frequent in the wider waterbird clade (Aequornithes; Sangster et al. 2022) that contains Procellariiformes (Cho et al. 2009, Gibb et al. 2013, Morris-Pocock et al. 2010, Rodrigues et al. 2017, Urantówka et al. 2020, Zhou et al. 2014b). However, further sampling to test for this trait is needed to determine if an ancient ancestral duplication, or multiple independent duplications across evolutionary time underlie these patterns. Additionally, very few studies in Procellariiformes sequenced the entire duplicated region and the extent of intact or degenerated copies of different genes remains mostly unknown in this order.

Next-generation sequencing libraries have been found to contain off-target mitochondrial sequences often in substantial numbers, for example in libraries enriched for capture of exomes (Picardi & Pesole 2012) and ultraconserved elements (UCEs; Raposo do Amaral et al. 2015). These mitochondrial reads are not intentionally sequenced as part of the probe set that is hybridized with genomic DNA for enrichment, but appear as off-target bycatch. This is likely due to the high ratio of mitogenomic DNA to nuclear DNA in some tissue types (Raposo do Amaral et al. 2015). These mitochondrial sequences are often sufficient to assemble complete, or assumed to be complete, mitogenomes at high coverage depth without the use of mitochondrial specific probes (e.g. Allio et al. 2020, Meiklejohn et al. 2014, Raposo do Amaral et al. 2015, Tamashiro et al. 2019, Wang et al. 2017). For example, Kimball et al. (2021) used off-target reads in UCE sequencing libraries to assemble 47 new mitogenomes of Galliformes species for phylogenetic use.

These off-target mitochondrial sequences in sequence capture datasets of short reads can be used to detect duplications in the mitogenome through elevated coverage of specific regions. If copies of duplicated regions are similar in identity, which is a frequent observation in birds (e.g. Gibb et al. 2013, Morris-Pocock et al. 2010), a higher proportion of reads present in next-

generation sequencing libraries would appear similar to duplicated regions relative to single copy regions. This method has been successfully used to confirm the presence of known mitogenomic duplications in Procellariiformes species (Lounsberry et al. 2015, Torres et al. 2019).

Using the UCE dataset and robustly inferred phylogenetic tree spanning the diversity of Procellariiformes introduced in Chapter 2 of this thesis, this approach was used here in a comparative context to investigate the prevalence and evolution of mitogenomic duplications across this speciose order of birds. First, the presence and frequency of off-target mitochondrial reads and putative mitogenomic assemblies in the previously generated sequencing dataset were assessed, and the factors affecting this considered. Secondly, this dataset was used to test for mitogenome duplications across Procellariiformes at a fine taxonomic scale, that has seldom been possible in other bird clades. With some evidence of mitogenome duplications in Procellariiformes (e.g. Abbott et al. 2005, Torres et al. 2019), this provided an opportunity to investigate the possibility of an ancient duplication and subsequent stasis underlying this clade, or of multiple independent origins across a large evolutionary timescale. Furthermore, the ability to detect the extent of intact, degraded, and lost mitogenomic regions allowed greater clarity as to the nature of mitogenomic duplications throughout Procellariiformes.

4.2 Methods

4.2.1 Testing for off-target mitochondrial sequences in a sequence capture dataset of Procellariiformes

The short read sequence capture dataset targeting UCE loci of 335 Procellariiformes samples described in Chapter 2 was used to test for the presence of off-target mitochondrial reads. Paired-end sequencing reads were trimmed to remove adapter content and low-quality bases using Trimmomatic version 0.39 (Bolger et al. 2014). For each sample, reads were searched against a custom reference database containing 2,158 individual CYTB sequences downloaded from GenBank spanning across the order Procellariiformes using BLASTn (Altschul et al. 1997), and hits were assessed by bit-scores and E-values. Results were compared for tissue or blood samples to evaluate if sample type affected off-target mitochondrial sequence capture. Trimmed paired-end reads were assembled into contigs using SPAdes version 3.15.4 (Bankevich et al. 2012) to test for the presence of assembled mitochondrial contigs. The frequency of putative complete assembled mitochondrial contigs in the size range 15-20kb and putative partially complete assembled mitochondrial contigs in the size range 10kb-14,999bp

were recorded. For 10 randomly selected samples with contigs in each of these size ranges a randomly selected 1kb section was searched against BLASTn (Altschul et al. 1997) to test for mitochondrial or nuclear origins. Mean coverage per base was recorded for each putative complete or partial mitochondrial contig.

4.2.2 Testing for duplicated regions of Procellariiformes mitogenomes

Duplication of the mitogenome was investigated by mapping reads to mitochondrial genomes. Complete and annotated mitogenomes of Procellariiformes species were downloaded from GenBank and used as reference sequences (Table 4.1). Complete and annotated mitogenomes were available for 12 Procellariiformes genera. The novel phylogeny of Procellariiformes presented in Chapter 2 was used to determine which reference to use for specific genera and species across the order. In the family Hydrobates, deep divergences among clades of taxa has suggested the possibility of further genera beyond *Hydrobates* and *Oceanodroma* (see Chapter 2). For this reason, and with the availability of further complete and annotated mitogenomes, *Oceanodroma castro* and *Oceanodroma tristrami* were used as references for mapping reads of their closest related species respectively (Table 4.1). A complete annotated mitogenome of *Spheniscus demersus*, African Penguin of the order Sphenisciformes that is sister to the Procellariiformes (Prum et al. 2015), was available for mapping reads of this species and served as an outgroup. Reference mitogenomes were split into the 13 protein-coding genes (5' to 3': ND1, ND2, CO1, CO2, ATP8, ATP6, CO3, ND3, ND4L, ND4, ND5, CYTB, ND6), two rRNAs (5' to 3': 12S rRNA, 16S rRNA), 22 tRNAs (5' to 3': tRNA-Phe, tRNA-Val, tRNA-Leu-UUR, tRNA-Ile, tRNA-Gln, tRNA-Met, tRNA-Trp, tRNA-Ala, tRNA-Asn, tRNA-Cys, tRNA-Tyr, tRNA-Ser-UCN, tRNA-Asp, tRNA-Lys, tRNA-Gly, tRNA-Arg, tRNA-His, tRNA-Ser-AGY, tRNA-Leu-CUN, tRNA-Thr, tRNA-Pro, tRNA-Glu), and the non-coding control region for mapping reads to.

BWA-MEM (Li 2013) was used to map the trimmed paired-end reads to the reference sequences to produce BAM files. SAMtools (Danecek et al. 2021) was used to sort and fix mates of BAM files. SAMtools flagstat was used to compute summary statistics of mapped reads. SAMtools depth was used to compute the read coverage depth at each position, and this was used to calculate minimum, mean, maximum, and standard deviation of read coverage depth for each region of each sample. Samples with fewer than 200 mapped reads and without coverage throughout the mitogenome were discarded. The minimum, mean, maximum, and standard deviation of coverage depth per mitochondrial region were examined for each retained

sample to determine regions of elevated coverage indicative of duplications. Coverage depth of each base pair position was also used to determine sudden increases in coverage.

Table 4.1 Species and corresponding GenBank accessions of complete and annotated mitogenome sequences used as references for mapping reads to, and the subject taxa each species was used as a reference for.

Reference species	GenBank ID	Subject taxa
Phoebastria albatrus	KJ735514	*Phoebastria*
		Diomedea
Thalassarche melanophris	AY158677	*Phoebetria*
		Thalassarche
Oceanites oceanicus	MN356163	*Oceanites*
Fregetta grallaria	MN356296	*Fregetta*
Pelagodroma marina	KC875856	*Nesofregetta*
		Garrodia
		Pelagodroma
Oceanodroma castro	MH433599	*Oceanodroma castro* and related species[1]
Oceanodroma tristrami	LC541440	*Oceanodroma tristrami* and related species[2]
Pagodroma nivea	MT726204	*Pagodroma*
Daption capense	MH924023	*Thalassoica*
		Daption
Fulmarus glacialis	MN356131	*Macronectes*
		Fulmarus
Ardenna pacifica	MT948201	*Bulweria*
		Pseudobulweria
		Procellaria
		Calonectris
		Ardenna
		Puffinus
Pelecanoides urinatrix	MN356319	*Pelecanoides*
		Halobaena
		Pachyptila
Aphrodroma brevirostris	AY158678	*Aphrodroma*
		Pterodroma
Spheniscus demersus	KC914350	*Spheniscus demersus*

[1]*Hydrobates pelagicus, Oceanodroma furcatus, O. jabejabe, O. castro, O. monteiroi*.

[2]*O. melania, O. microsoma, O. tethys, O. tristrami, O. matsudairae, O. monorhis, O. homochroa, O. hornbyi, O. markhami, O. leucorhous, O. cheimomnestes, O. socorroensis*.

4.2.3 Examining the evolutionary patterns of mitogenome duplications in Procellariiformes

To examine the evolutionary history of mitogenome duplications, the results of mapping reads to references and inference of duplicated regions was mapped onto the consensus Procellariiformes phylogeny produced in Chapter 2, pruned to contain only taxa for which samples with sufficient mapped reads and coverage had been retained for analyses.

Ancestral state reconstructions of duplication of the mitogenome were conducted using Mesquite version 3.61 (Maddison & Maddison 2019). To obtain accurate branch lengths for the consensus reference tree used for ancestral state reconstructions, a tree was constructed using the aligned 75% matrix of 4,335 UCE loci (see Chapter 2) trimmed to contain only samples with sufficient mapped reads and coverage under maximum likelihood (ML) using RAxML-NG version 1.1.0 (Kozlov et al. 2019) with the GTR+G substitution model and 20 ML searches with 10 random starting trees and 10 parsimony starting trees. This tree was used as the reference for ancestral state reconstruction using ML with the Markov k-state 1-parameter (Mk1) model. Ancestral state reconstruction was also conducted using an unordered maximum parsimony model. As tandem duplication followed by a varying degree of deletion or degeneration is the likely explanation of varying duplication states observed in animals (Boore 2000), the presence/absence of any detected duplication was considered for ancestral state reconstructions.

4.3 Results

4.3.1 The presence of off-target mitochondrial sequences in a sequence capture dataset of Procellariiformes

Sequenced reads were searched against a reference database of 2,158 Procellariiformes CYTB sequences from GenBank. Of the 335 sequenced Procellariiformes samples, four were from museum toepad specimens and not included. Of the 331 retained samples, 213 (64.4%) contained reads matching to CYTB sequences. However, this figure was strongly influenced by the input sample type used for DNA extraction and sequencing. Of the 179 tissue samples, 163 (91.9%) contained reads matching to CYTB sequences, whereas of the 152 blood samples, 50 (32.9%) contained reads matching to CYTB sequences.

Reads were assembled into contigs to test for the presence of putative complete assembled mitochondrial contigs and partially complete assembled mitochondrial contigs. Of the contig assemblies for 335 Procellariiformes samples, 60 contained assembled contigs of 15-20kb in

length that could represent complete assembled mitogenomes. A further 59 samples contained assembled contigs of 10kb-14,999bp in length that could represent partial assembled mitogenomes. Searching 1kb sections of contigs of samples from different genera spanning the diversity of Procellariiformes in the 15-20kb size range using BLASTn found that 9/10 were mitochondrial in origin, as assessed by matching to previously uploaded whole mitogenome sequences on GenBank. For the sample containing the one searched contig in the 15-20kb category that was not mitochondrial in origin the second largest assembled contig was 15,645bp in length and was mitochondrial in origin. Searching 1kb sections of contigs of samples from different Procellariiformes genera in the 10kb-14,999kb size range found that 7/10 were mitochondrial in origin. Assembled contigs in the size range 15-20kb that were mitochondrial in origin had a mean coverage 175.8 per base pair, and assembled contigs in the 10kb-14,999kb size range that were mitochondrial in origin had a mean coverage of 165.9 per base pair.

4.3.2 Mapping sequenced reads to reference mitogenomes

Duplications in the mitogenomes of Procellariiformes species were investigated by mapping reads to 13 complete and annotated reference mitogenomes downloaded from GenBank. A further complete and annotated mitogenome of *Spheniscus demersus* was used as a reference for mapping reads of this species as an outgroup. Reads were mapped to all protein-coding genes, rRNAs, tRNAs, and the control region. Of the 335 Procellariiformes samples, 154 contained sufficient mapped mitochondrial reads and coverage to assess potential duplication of mitochondrial sequences. These 154 samples had a mean of 36,325.7 (95% CI: ±10,343.4) reads mapped to their reference, with a mean of 0.97% (95% CI: ±0.3) mapped reads of the total sequenced reads per sample. For these 154 samples, mean coverage of mitochondrial reference sequences was 147.7 (95% CI: ±42.4).

4.3.3 The evolutionary patterns of mitogenome duplications in Procellariiformes

The 154 retained samples comprised 115 taxa of 101 species and an additional 14 subspecies, including the outgroup species *Spheniscus demersus*. All four families of Procellariiformes were represented within the dataset with multiple species per family (Diomedeidae: 13 species, Oceanitidae: 7 species, Hydrobatidae: 10 species, Procellariidae: 70 species). All genera except for *Hydrobates* were present, and this genus is paraphyletic with respect to the genus *Oceanodroma* (see Chapter 2), that was present and represented by 10 species.

This dataset, spanning the diversity of Procellariiformes was used to examine the evolutionary history of mitogenome duplications using the consensus tree of the order produced in Chapter 2, which was pruned to contain the same taxon set (Fig. 4.1). Different arrangements of mitogenome duplications were found. In most species duplication of ND6 to the control region was present. Duplication of a large 3' fragment of CYTB starting approximately 600-700bp into the 1,143bp length of the gene was present in some species, with the highest frequency in the families Diomedeidae and Oceanitidae, in the fulmarine clade of Procellariidae, and seldom in the remaining Procellariidae radiation. These duplications of CYTB were evidenced by an instant and significant increase in coverage that was replicated in each species for which this region was duplicated. In few species sporadically throughout the tree, ND6 was not duplicated and only the control region was duplicated. The absence of any duplicated region was rare throughout the tree, with only two monophyletic clades of three taxa each with this arrangement found in Hydrobatidae and Procellariidae.

In the family Diomedeidae, evidence of duplication of the mitogenome was present throughout. The genera *Phoebastria* and *Diomedea* showed evidence of duplication of the 3' end of CYTB. The genera *Phoebetria* and *Thalassarche* did not replicate this pattern, and exhibited the more prevalent pattern of duplication of ND6 to the control region.

Oceanitidae showed a similar pattern of duplications in each genus and species. Duplication of a 3' fragment of CYTB was apparent in the genera *Nesofregetta*, *Garrodia*, *Pelagodroma*, and some taxa of *Oceanites*, and absent in *Fregetta*.

Within Hydrobatidae, the two subspecies of *Oceanodroma furcatus* showed duplication only of the control region. The monophyletic clade of *O. melania*, *O. microsoma*, and *O. tethys* provided a rare example of a lack of any detected mitogenome duplication. Evidence of CYTB duplication was not found in Hydrobatidae species.

The large family Procellariidae presented a dominant pattern of mitogenome duplication from ND6 to the control region in most species. In the fulmarine clade, the genera *Pagodroma*, *Thalassoica*, and *Macronectes* also showed evidence of duplication of a 3' fragment of CYTB, similar to those found in in some Diomedeidae and Oceanitidae species. This was absent in the genera *Daption* and *Fulmarus*.

The divergent diving-petrel genus *Pelecanoides* showed duplication of only the control region in the species *P. georgicus*, that is sister to all other congeners. The remaining *Pelecanoides* taxa, of the species *P. magellani* and *P. urinatrix*, provided the second example of a monophyletic clade with no detected signal of any duplication in the mitogenome.

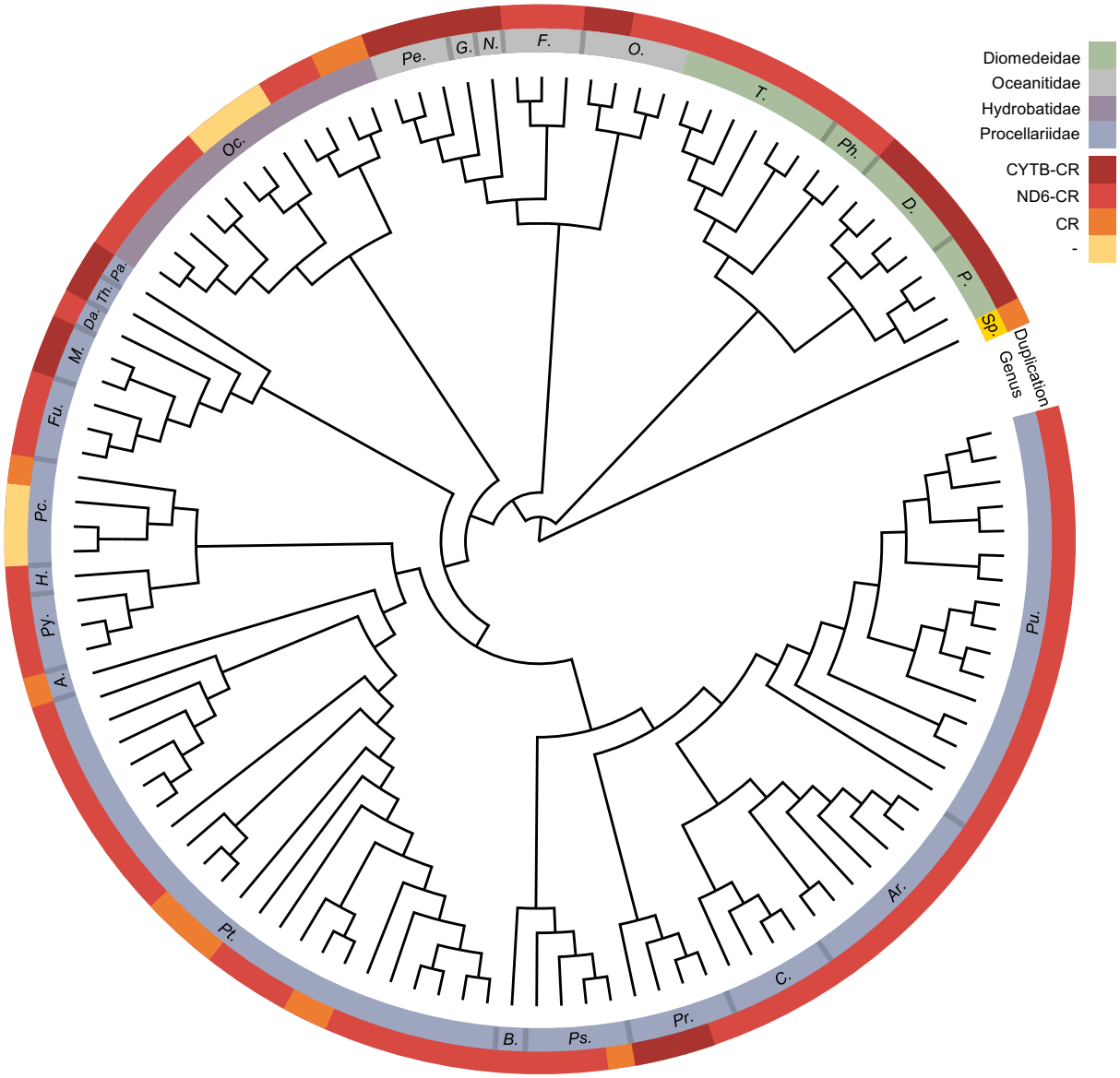


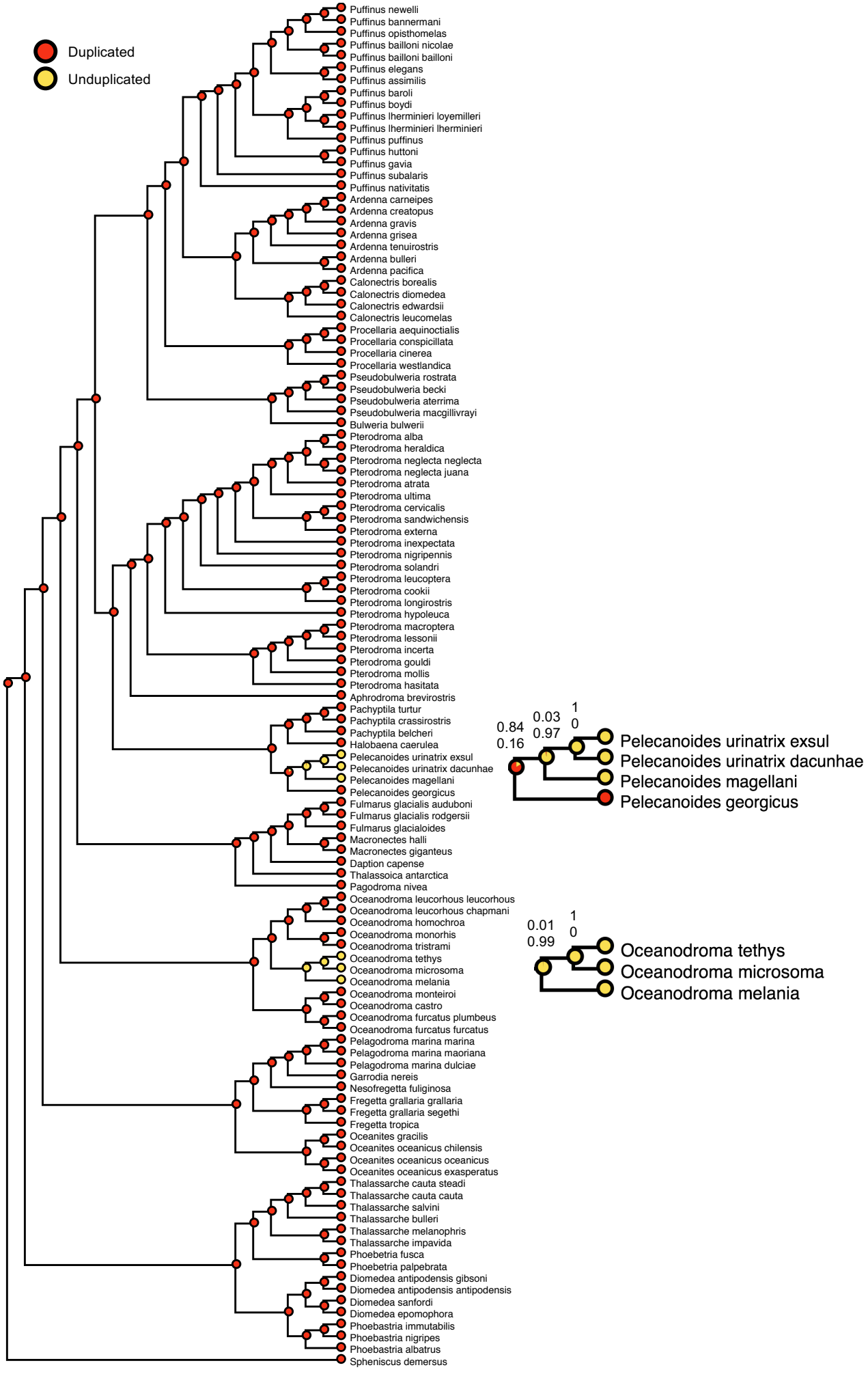
Fig. 4.1 Phylogenetic tree of Procellariiformes species with mapped duplications of the mitogenome. The inner track shows families by colour as indicated by the legend, with genera denoted within each family. Evidence of the duplication of the mitogenome is indicated in the outer track according to the legend. Evidence of duplication of CYTB to the control region (CR) is shown in dark red, ND6 to the control region in light red, the control region only in orange, and no duplication in yellow. Genus names are abbreviated as follows: *Sp. Spheniscus* (Sphenisciformes; outgroup), *P. Phoebastria*, *D. Diomedea*, *Ph. Phoebetria*, *T. Thalassarche*, *O. Oceanites*, *F. Fregetta*, *N. Nesofregetta*, *G. Garrodia*, *Pe. Pelagodroma*, *Oc. Oceanodroma*, *Pa. Pagodroma*, *Th. Thalassoica*, *Da. Daption*, *M. Macronectes*, *Fu. Fulmarus*, *Pc. Pelecanoides*, *H. Halobaena*, *Py. Pachyptila*, *A. Aphrodroma*, *Pt. Pterodroma*, *B. Bulweria*, *Ps. Pseudobulweria*, *Pr. Procellaria*, *C. Calonectris*, *Ar. Ardenna*, *Pu. Puffinus*.

Halobaena and *Pachytpila* displayed the typical pattern of duplication of ND6 to the control region. In *Aphrodroma*, only the control region showed evidence of duplication. In the speciose genus *Pterodroma* duplication was widespread, with the only variation presented by duplication solely of the control region in the sister species *P. cookii* and *P. leucoptera*, *P. solandri*, and the monophyletic species *P. sandwichensis* and *P. cervicalis*.

The large clade of Procellariidae encompassing the genera *Bulweria*, *Pseudobulweria*, *Procellaria*, *Calonectris*, *Ardenna*, and *Puffinus* showed minimal deviation from the prevalent pattern of duplication of ND6 to the control region. Within the genus *Pseudobulweria*, one species, *P. rostrata*, showed signals of duplication of the control region only. In the genus *Procellaria*, three of four species presented the additional duplication of a 3' fragment of CYTB from approximately 600-700bp into the gene, similar to that found in some fulmarine, Oceanitidae, and Diomedeidae species. The shearwater genera *Calonectris*, *Ardenna*, and *Puffinus* showed no variation from the duplicated region of ND6 to the control region in any species.

Ancestral state reconstructions using both ML and maximum parsimony methods indicated that duplication of the mitogenome was the ancestral, plesiomorphic state at the root node of Procellariiformes (Figs. 4.2 and 4.3). Loss of duplication in the mitogenome of three monophyletic *Oceanodroma* species and two monophyletic *Pelecanoides* species was a derived, apomorphic state. In *Pelecanoides*, the deepest diverging species, *P. georgicus*, showed evidence of duplication, and the remaining species *P. magellani* and *P. urinatrix* did not. The ancestral node of the genus *Pelecanoides* was likely to have had a duplicated mitogenome (proportional likelihood values: 0.84/0.16), although this was the only node for which the decision threshold did not reject either state. The ancestral node of *Oceanodroma melania*, *O. microsoma*, and *O. tethys* was not likely to have had a duplicated mitogenome (proportional likelihood values: 0.01/0.99). The most parsimonious state at these nodes was concordant with those inferred by ML.

For these analyses, the single outgroup species *Spheniscus demersus* of Sphenisciformes, the sister order to Procellariiformes, was used and recorded as showing evidence of duplication of the control region. Sphenisciformes contains six genera and 18 species (HBW and BirdLife International 2022), and variation in duplication between species is possible. Recording the outgroup species as polymorphic did not alter the results of maximum parsimony reconstructions that mitogenome duplication was present at the root node of Procellariiformes.


Duplicated
Unduplicated
Puffinus newelli
Puffinus bannermani
Puffinus opisthomelas
Puffinus bailloni nicolae
Puffinus bailloni bailloni
Puffinus elegans
Puffinus assimilis
Puffinus baroli
Puffinus boydi
Puffinus lherminieri loyemilleri
Puffinus lherminieri lherminieri
Puffinus puffinus
Puffinus huttoni
Puffinus gavia
Puffinus subalaris
Puffinus nativitatis
Ardenna carneipes
Ardenna creatopus
Ardenna gravis
Ardenna grisea
Ardenna tenuirostris
Ardenna bulleri
Ardenna pacifica
Calonectris borealis
Calonectris diomedea
Calonectris edwardsii
Calonectris leucomelas
Procellaria aequinoctialis
Procellaria conspicillata
Procellaria cinerea
Procellaria westlandica
Pseudobulweria rostrata
Pseudobulweria becki
Pseudobulweria aterrima
Pseudobulweria macgillivrayi
Bulweria bulwerii
Pterodroma alba
Pterodroma heraldica
Pterodroma neglecta neglecta
Pterodroma neglecta juana
Pterodroma atrata
Pterodroma ultima
Pterodroma cervicalis
Pterodroma sandwichensis
Pterodroma externa
Pterodroma inexpectata
Pterodroma nigripennis
Pterodroma solandri
Pterodroma leucoptera
Pterodroma cookii
Pterodroma longirostris
Pterodroma hypoleuca
Pterodroma macroptera
Pterodroma lessonii
Pterodroma incerta
Pterodroma gouldi
Pterodroma mollis
Pterodroma hasitata
Aphrodroma brevirostris
Pachyptila turtur
Pachyptila crassirostris
Pachyptila belcheri
Halobaena caerulea
Pelecanoides urinatrix exsul
Pelecanoides urinatrix dacunhae
Pelecanoides magellani
Pelecanoides georgicus
Fulmarus glacialis auduboni
Fulmarus glacialis rodgersii
Fulmarus glacialoides
Macronectes halli
Macronectes giganteus
Daption capense
Thalassoica antarctica
Pagodroma nivea
Oceanodroma leucorhous leucorhous
Oceanodroma leucorhous chapmani
Oceanodroma homochroa
Oceanodroma monorhis
Oceanodroma tristrami
Oceanodroma tethys
Oceanodroma microsoma
Oceanodroma melania
Oceanodroma monteiroi
Oceanodroma castro
Oceanodroma furcatus plumbeus
Oceanodroma furcatus furcatus
Pelagodroma marina marina
Pelagodroma marina maoriana
Pelagodroma marina dulciae
Garrodia nereis
Nesofregetta fuliginosa
Fregetta grallaria grallaria
Fregetta grallaria segethi
Fregetta tropica
Oceanites gracilis
Oceanites oceanicus chilensis
Oceanites oceanicus oceanicus
Oceanites oceanicus exasperatus
Thalassarche cauta steadi
Thalassarche cauta cauta
Thalassarche salvini
Thalassarche bulleri
Thalassarche melanophris
Thalassarche impavida
Phoebetria fusca
Phoebetria palpebrata
Diomedea antipodensis gibsoni
Diomedea antipodensis antipodensis
Diomedea sanfordi
Diomedea epomophora
Phoebastria immutabilis
Phoebastria nigripes
Phoebastria albatrus
Spheniscus demersus
0.84
0.16
0.03
0.97
1
0
Pelecanoides urinatrix exsul
Pelecanoides urinatrix dacunhae
Pelecanoides magellani
Pelecanoides georgicus
0.01
0.99
1
0
Oceanodroma tethys
Oceanodroma microsoma
Oceanodroma melania


Duplicated
Unduplicated
Puffinus newelli
Puffinus bannermani
Puffinus opisthomelas
Puffinus bailloni nicolae
Puffinus bailloni bailloni
Puffinus elegans
Puffinus assimilis
Puffinus baroli
Puffinus boydi
Puffinus lherminieri loyemilleri
Puffinus lherminieri lherminieri
Puffinus puffinus
Puffinus huttoni
Puffinus gavia
Puffinus subalaris
Puffinus nativitatis
Ardenna carneipes
Ardenna creatopus
Ardenna gravis
Ardenna grisea
Ardenna tenuirostris
Ardenna bulleri
Ardenna pacifica
Calonectris borealis
Calonectris diomedea
Calonectris edwardsii
Calonectris leucomelas
Procellaria aequinoctialis
Procellaria conspicillata
Procellaria cinerea
Procellaria westlandica
Pseudobulweria rostrata
Pseudobulweria becki
Pseudobulweria aterrima
Pseudobulweria macgillivrayi
Bulweria bulwerii
Pterodroma alba
Pterodroma heraldica
Pterodroma neglecta neglecta
Pterodroma neglecta juana
Pterodroma atrata
Pterodroma ultima
Pterodroma cervicalis
Pterodroma sandwichensis
Pterodroma externa
Pterodroma inexpectata
Pterodroma nigripennis
Pterodroma solandri
Pterodroma leucoptera
Pterodroma cookii
Pterodroma longirostris
Pterodroma hypoleuca
Pterodroma macroptera
Pterodroma lessonii
Pterodroma incerta
Pterodroma gouldi
Pterodroma mollis
Pterodroma hasitata
Aphrodroma brevirostris
Pachyptila turtur
Pachyptila crassirostris
Pachyptila belcheri
Halobaena caerulea
Pelecanoides urinatrix exsul
Pelecanoides urinatrix dacunhae
Pelecanoides magellani
Pelecanoides georgicus
Fulmarus glacialis auduboni
Fulmarus glacialis rodgersii
Fulmarus glacialoides
Macronectes halli
Macronectes giganteus
Daption capense
Thalassoica antarctica
Pagodroma nivea
Oceanodroma leucorhous leucorhous
Oceanodroma leucorhous chapmani
Oceanodroma homochroa
Oceanodroma monorhis
Oceanodroma tristrami
Oceanodroma tethys
Oceanodroma microsoma
Oceanodroma melania
Oceanodroma monteiroi
Oceanodroma castro
Oceanodroma furcatus plumbeus
Oceanodroma furcatus furcatus
Pelagodroma marina marina
Pelagodroma marina maoriana
Pelagodroma marina dulciae
Garrodia nereis
Nesofregetta fuliginosa
Fregetta grallaria grallaria
Fregetta grallaria segethi
Fregetta tropica
Oceanites gracilis
Oceanites oceanicus chilensis
Oceanites oceanicus oceanicus
Oceanites oceanicus exasperatus
Thalassarche cauta steadi
Thalassarche cauta cauta
Thalassarche salvini
Thalassarche bulleri
Thalassarche melanophris
Thalassarche impavida
Phoebetria fusca
Phoebetria palpebrata
Diomedea antipodensis gibsoni
Diomedea antipodensis antipodensis
Diomedea sanfordi
Diomedea epomophora
Phoebastria immutabilis
Phoebastria nigripes
Phoebastria albatrus
Spheniscus demersus

Fig. 4.2 (page 103) Ancestral state reconstruction of mitogenome duplications in Procellariiformes species under ML. The two monophyletic clades where duplications were not found are shown inset, and proportional likelihood values are annotated above nodes.

Fig. 4.3 (page 104) Ancestral state reconstruction of mitogenome duplications in Procellariiformes species under maximum parsimony.

4.4 Discussion

In this study, a large-scale sequencing dataset and well resolved phylogenetic tree were used to infer the evolutionary history of mitogenomic duplications among Procellariiformes species. Sequencing libraries were found to frequently contain off-target mitochondrial reads as bycatch and long assembled mitogenomic sequences in many cases. The approach of mapping short reads to reference mitogenome sequences was successfully used to infer duplications through elevated coverage depth, confirming some previously evidenced duplications and providing novel results for a large number of species. Mapping the duplications onto the phylogenetic tree found extensive duplications throughout Procellariiformes, with repeated degeneration of duplicated regions and, rarely, complete loss of such regions. Several different rearrangements were found, clarifying the nature of these duplications in this clade. A common origin of mitogenome duplications preceding the evolutionary divergence of the clade was inferred, providing further evidence of the widespread nature of this trait in birds.

4.4.1 Mitochondrial sequences in sequence capture datasets and their applications

Searching reads in UCE enriched sequencing datasets against mitochondrial references found the frequent occurrence of mitochondrial sequences. These sequences were unintentionally generated as off-target bycatch during a capture sequencing protocol that is not entirely efficient. The high frequency of off-target reads and relative inefficiency of UCE sequencing libraries has been previously found, where a low proportion of reads may be assembled into contigs, and a low proportion of assembled contigs map to UCE loci (Faircloth et al. 2015, Smith et al. 2014). In this study, of the 154 samples that contained sufficient mitochondrial reads to assess potential mitochondrial duplications, a mean of 0.97% of reads mapped to the mitogenome. This is similar to a previous study using bird UCE sequencing data that identified 0.86% of reads as mapping to the mitogenome (Tamashiro et al. 2019). Although this

proportion is low, given that millions of reads are routinely sequenced the total number can be relatively large, as in this study where 0.97% of total reads mapping to the mitogenome corresponded to a mean of 36,325.7 reads per sample.

The high number of mitochondrial reads found may be a consequence of the high ratio of mitogenomic DNA to nuclear DNA in some tissue types, particularly muscle and heart tissues (D'Erchia et al. 2015). It has been suggested that blood samples produce lower amounts of mitochondrial bycatch sequences (Allio et al. 2020, Barker et al. 2015, Raposo do Amaral et al. 2015). This tissue type effect was clearly illustrated here, in that 91.9% of sequencing libraries using DNA extracted from muscle tissues contained mitochondrial reads, whereas only 32.9% of sequencing libraries using DNA extracted from blood contained mitochondrial reads. This is likely to be a consequence of the low levels of mitochondria in the nucleated red blood cells of birds relative to other tissues (Reverter et al. 2017). Thus, the presence and amount of mitochondrial sequences in capture sequencing datasets is strongly dependent on the input tissue type, and studies aiming to use bycatch mitochondrial sequences should aim to use tissue rather than blood samples.

This presence of off-target mitochondrial sequences can strictly be considered "contamination" (Smith et al. 2014). However, it is generally considered beneficial instead, with sufficient amounts of mitochondrial sequences often allowing the assembly of complete, or assumed to be complete, mitogenomic contigs with high coverage depth without the need for additional labwork or costs (Allio et al. 2020, Raposo do Amaral et al. 2015). In this study, 60 samples from a total of 335 were found to contain assembled contigs of 15-20kb in size, most of which were mitogenomic in nature and likely represent complete mitogenomes. A further 59 samples contained assembled contigs of 10kb-14,999bp in size, for which most were also mitogenomic, and represent partial mitogenomic sequences.

The simultaneous generation of nuclear and mitogenomic markers can provide complementary information for increased power in phylogenetic studies (Allio et al. 2020), and allows inference into patterns of evolutionary discordance between these genomes (Kimball et al. 2021, Tamashiro et al. 2019). However, in the context of mitogenome duplications being widespread and common throughout birds, the mitogenome assemblies being generated using these approaches are potentially questionable in terms of size, gene content, and gene order. The copies of duplicated regions are often identical or very similar (e.g. Gibb et al. 2013, Morris-Pocock et al. 2010), and as illustrated here and previously (Torres et al. 2019), sequenced reads will map to either copy of duplicated regions. In these cases, the assembly process is likely to collapse similar or identical reads to produce single copy mitogenomes,

which may not reflect the true gene content or order of duplicated mitogenomes. Thus, the 'complete' mitogenomes being produced using these approaches are likely to be incomplete in many cases in birds. The control region has proved difficult to assemble completely and unambiguously in studies producing mitogenomes from bycatch in UCE sequencing libraries (Kimball et al. 2021, Tamashiro et al. 2019). Tamashiro et al. (2019) found a long insertion between tRNA-Pro and tRNA-Thr in some species, where a duplicated control region is often found (Gibb et al. 2007) that could not be assembled, and they could not determine whether it was a duplicated control region or resulted from poor assembly (Tamashiro et al. 2019). Reanalysis of mitogenomes previously annotated without duplications, and thought to be complete, using thorough PCR methods to detect duplications has brought up several instances where duplications have been found (Akiyama et al. 2017, Gibb et al. 2007, Gibb et al. 2015, Singh et al. 2008, Urantówka et al. 2018, Zhou et al. 2014b). In these cases, difficulties in amplification, sequencing, and assembly of repeated regions precluded duplications from being identified (Gibb et al. 2007, Gibb et al. 2015, Urantówka et al. 2020). Although studies using next-generation sequencing libraries have the capacity to produce large numbers of seemingly complete mitogenomes, these assemblies may provide similar examples where future, more laborious PCR methods can identify unfound duplications. It should also be noted that this is likely to be a particular problem for birds, where duplications are now known to be commonplace, and less so for studies on taxonomic groups where duplications are rare.

4.4.2 The evolution of mitogenome duplications in Procellariiformes and possible functions

Mitogenomic duplications were found to be commonplace across Procellariiformes families, genera, and species in this study. Mapping these duplications onto the Procellariiformes tree revealed taxonomic patterns. Overall loss of duplicated regions occurred in only two monophyletic clades of three taxa each, one in Hydrobatidae of *Oceanodroma melania*, *O. microsoma*, and *O. tethys*, and one in Procellariidae of *Pelecanoides magellani* and *P. urinatrix*. The *Pelecanoides* are divergent among Procellariiformes in morphology (Brooke 2004), and presented long branches relative to other Procellariidae genera in phylogenetic trees (see Chapter 2).

As suggested by the rare loss of duplications in only two monophyletic clades, ancestral state reconstructions unambiguously inferred the presence of duplications in the mitogenome as the ancestral state at the root of Procellariiformes. There was no evidence of independent duplications arising following the loss of duplicated regions, although there were only two

clades of three taxa each and few internal nodes for which this could have been inferred. This suggests that the mitogenomic duplications that are found in extant taxa are the result of an ancient duplication event preceding the emergence and radiation of Procellariiformes. Duplications are frequent in the wider Aequornithes clade (Cho et al. 2009, Gibb et al. 2013, Morris-Pocock et al. 2010, Rodrigues et al. 2017, Urantówka et al. 2020, Zhou et al. 2014b), suggesting that duplicated mitogenomes may be the ancestral state for this large clade in its entirety. There is evidence that duplications are the ancestral state for other bird orders, including those distantly related to Procellariiformes and the Aequornithes (e.g. Mackiewicz et al. 2019, Urantówka et al. 2018, Urantówka et al. 2021). Compared with these other large-scale orders across birds, Procellariiformes are less typical in that duplications are present in almost all species and absent in only a very small minority of species. The suggestion that mitogenome duplications could represent the default state in birds (Gibb et al. 2013) is certainly borne out by this evidence from Procellariiformes.

Several different duplicated regions were found in Procellariiformes, including the 3' end of CYTB to the control region, ND6 to the control region, and the control region only. These match with previously described duplicated states (Cho et al. 2009, Eberhard et al. 2001, Gibb et al. 2007, Mindell et al. 1998). As has been previously shown in *Phoebastria* albatross species, a large 3' region of CYTB was copied (Eda et al. 2010, Lounsberry et al. 2015). This duplicated region comprised the most expansive rearrangement detected and was relatively rare, found in *Diomedea* and *Phoebastria* but not *Phoebetria* and *Thalassarche* in Diomedeidae, in several but not all Oceanitidae genera and species and, in Procellariidae, only in some fulmarine genera and some species of *Procellaria*. Most species showed a duplication of a shorter region not including CYTB, spanning from ND6 to the control region. This pattern was particularly common in Procellariidae, where deviation was rare and less frequent than in other families. Duplication of only the control region was infrequent, without obvious taxonomic patterns. Notably, *Pelecanoides georgicus* showed this state, representing an apparent intermediate state between the duplication of ND6 to the control region typical of Procellariidae, and the absence of any duplications in sister *Pelecanoides* species.

This pattern, and those observed overall in Procellariiformes, are consistent with the tandem duplication and random loss model of mitogenome rearrangement, whereby an ancestral duplication precedes the emergence of Procellariiformes lineages and subsequent loss of function and gradual degradation or deletion has occurred in some regions in some taxa (Bensch & Härlid 2000, Boore 2000, Macey et al. 1997, Mindell et al. 1998, Moritz & Brown 1986, San Mauro et al. 2006). The most expansive duplicated region of CYTB to the control

region was notably present mostly in Diomedeidae and Oceanitidae, the deepest diverging families of Procellariiformes, and is present in several closely related orders (Cho et al. 2009, Gibb et al. 2013, Morris-Pocock et al. 2010). This suggests that this large duplicated region was ancestral in all Procellariiformes, followed by non-uniform degeneration reaching the loss of all duplications in some few species.

The approach used here of applying coverage depth of large-scale sequencing datasets to determine mitogenome duplications was successful in clarifying the rearrangements found in Procellariiformes. This approach, necessitating reads mapping to similar reference sequences, was more adept at identifying intact copies of duplicated regions rather than degenerated copies which are present in some mitogenomic rearrangements (e.g. Eberhard et al. 2001, Gibb et al. 2007, Mindell et al. 1998). Abbott et al. (2005) documented a partial duplication of the 3' end of CYTB containing a very short partially degraded and partially identical copy in *Thalassarche* species (Diomedeidae). Here, signals of duplication of this short fragment were not detected, although larger fragments of CYTB that were duplicated in other species were successfully detected. These shorter fragments are likely to require more intensive PCR and sequencing methods to successfully detect (Gibb et al. 2007). Nevertheless, the possible presence of undetected degenerated copies of some regions in some taxa would not alter the patterns observed across Procellariiformes, where extensive duplications were found in almost all species.

Procellariiformes diverged from their sister order Sphenisciformes (penguins) at least 60.5mya, based on fossil evidence of basal penguin taxa (Ksepka & Clarke 2015, Slack et al. 2006). The persistence of mitogenome duplications in almost all Procellariiformes species with a common origin suggests stasis of this trait over a large timescale. Furthermore, duplicated regions in the mitogenome may represent the ancestral state for the Paleognathae and all birds (Urantówka et al. 2020), representing stasis over a significant evolutionary period of rapid diversification (Suh 2016).

Adherence to the tandem duplication and random loss model would suggest that copies of duplicated regions should eventually degenerate, lose function, and may be deleted (Boore 2000). However, this is often not the case, for example here across Procellariiformes, where both identical copies are often preserved over large timescales. The persistence of duplicated regions, and observations that both copies of identical regions appear to have retained function (e.g. Abbott et al. 2005, Eberhard et al. 2001, Morris-Pocock et al. 2010, Zhou et al. 2014b), suggest the presence of selective pressures to maintain these duplications.

Due to its role in replication and transcription (Boore 1999), the preservation of two control regions may allow more efficient initiation of mitogenomic replication (Jiang et al. 2007, Kumazawa et al. 1996, Umeda et al. 2001). In the chicken, the majority of replication initiation events were found to occur instead in ND6 (Reyes et al. 2005), suggesting a possible benefit to retaining duplicate functional copies of this gene. Increased initiation efficiency could create more efficient energy production through increased mitogenome copy number per mitochondrion, and increased metabolic flexibility to environmental stress (Mackiewicz et al. 2019, Skujina et al. 2016). Mitogenomic duplications are associated with greater longevity in birds (Skujina et al. 2016). As well as possible benefits from increased energy efficiency, this could be due to the redundancy of copied mitochondrial genes providing protection against age-related deterioration of mitochondrial function (Schirtzinger et al. 2012, Skujina et al. 2016). In additional to longevity, mitogenome duplications have been associated with increased flight ability in parrots, that could also be explained by increased energy production and efficiency (Urantówka et al. 2018). Further biochemical study will be needed to investigate these possible relationships between the retention of mitogenome duplications and energy production.

Associations with longevity are particularly interesting in relation to Procellariiformes, as this order is amongst the highest of all birds in terms of lifespan (Wasser & Sherman 2010). Albatross species are especially long-lived, and a *Phoebastria immutabilis* individual is currently the oldest known wild bird with an age of at least 71 years (U.S. Fish & Wildlife Service 2022). Similarly, associations with flight ability are highly relevant to Procellariiformes, many species of which cover vast distances at sea (Shaffer et al. 2006, Warham 1996, Weimerskirch et al. 2014). The duplicated regions were retained in almost all Procellariiformes species, and at a considerably higher proportion than found in other comparable large-scale studies of birds (Caparroz et al. 2018, Mackiewicz et al. 2019, Schirtzinger et al. 2012, Urantówka et al. 2018). Notably, one of two small clades where the loss of mitogenomic duplications was found was in *Pelecanoides*, an atypical genus among Procellariiformes that is not as well adapted to long distance flight and covers lower distances at sea compared to other members of the order (Navarro et al. 2015, Warham 1977). The maintenance of duplicated mitogenomic regions may partly explain the high longevity and strong flight abilities found in many Procellariiformes species.

4.4.3 Conclusions

Procellariiformes offer insights into the evolution and possible functions of mitogenome duplications in birds. The prevalence of duplications in Procellariiformes species, and evidence from related clades, suggests remarkable stasis of this trait across large timescales. As employed here, methods using the high depth coverage of off-target mitochondrial reads generated from an inefficient capture process in next-generation sequencing datasets offer great promise for combining the study of the evolution of nuclear and mitochondrial genomes in non-model organisms. UCEs have been widely used in birds and other vertebrates and invertebrates, and a similar approach to studying rearrangements of the mitogenome could be routinely replicated. There is growing evidence of significant variation in the animal mitogenome (Lavrov & Pett 2016), and further applications of these methods may provide further useful information in this context.

Chapter 5: Review and future directions

The chapters of this thesis have investigated the evolution of the nuclear and mitochondrial genomes in the context of the seabird order Procellariiformes, providing new knowledge and insights both specific to this study group and in general. This work has also allowed further specificity in defining where remaining knowledge gaps lie, and the relevant future questions to be asked. The following sections detail where future work could be directed to further complement the insights gained here and address these gaps. Sections 5.1.1-5.1.3 detail future directions specific to each chapter, and section 5.2 details those more generally.

5.1 Future directions

5.1.1 Future directions in Procellariiformes phylogenetics and systematics

The novel phylogeny of Procellariiformes presented in Chapter 2 provided new insights into the evolution and systematics at shallow and deep scales across the order. The dataset of 4,335 genome-wide loci represented a strong improvement in resolution of the phylogenetics of Procellariiformes compared with previous studies using only the single mitochondrial gene cytochrome *b*. Data was presented for 89% of all species and 85% of all taxa, providing a comprehensive framework of species and subspecies relationships within the clade. Nevertheless, samples could not be obtained for all taxa of the order. The extremely remote and difficult to reach breeding sites of these birds, restricted to single islands or small groups of islands in some taxa, has presented challenges in obtaining samples, and is a reason that a phylogeny on this scale was previously elusive.

For some taxa it is reasonable to predict where they would fall in the Procellariiformes tree. For example, the poorly known *Bulweria fallax* is the only extant member of the genus *Bulweria* other than the widespread and morphologically similar *B. bulweria* (Olson 1975), and it is reasonable to expect a sister relationship between these species. However, there are several taxa that it is impossible to predict phylogenetic positions of, for example of the very poorly known *Puffinus heinrothi* (Kirwan et al. 2020). Further work to obtain samples and place these taxa in the phylogenetic framework provided in this thesis should be a strong priority for future work in Procellariiformes systematics. Museum specimens have been increasingly used in sequence capture studies in recent years (Derkarabetian et al. 2019, McCormack et al. 2016, Roycroft et al. 2022), and a small number were used in this study.

Expanding on this sampling of museum specimens could be an option for taxa for which obtaining modern samples remains difficult or impossible.

The phylogeny of Procellariiformes also highlighted several taxonomic questions, and challenging to resolve nodes in the evolution of the clade, for which further work could provide important contributions. For taxonomic uncertainty in recent divergences, applying RADseq may be useful (Leaché et al. 2015, Rubin et al. 2012). Of three species for which subspecies were not inferred as monophyletic, two were Diomedeidae species where branch lengths were short in all analyses, and genetic diversity is low (Chambers et al. 2009). Applying fine-scale population genetic work to the subspecies of these species, *Diomedea antipodensis* and *Thalassarche cauta*, could elucidate whether these subspecies should be considered synonymous or recently diverged taxa.

The prion genus *Pachyptila* provided a rare example within the Procellariiformes tree with poorly resolved relationships. These species are very morphologically similar (Harper 1980), and show high gene flow, admixture, migration rates and a lack of genetic structure between populations in several species (Masello et al. 2019, Masello et al. 2022, Quillfeldt et al. 2017). RADseq data of *P. turtur* and *P. crassirostris* found that phylogenetic relationships did not match taxonomic names (Shepherd et al. 2022). Conducting an expanded study of *Pachyptila* populations using RADseq data would be useful in determining the number of *Pachyptila* species and their relationships. Of all clades throughout Procellariiformes, and at all scales, the genus *Pachyptila* represents the largest area of taxonomic uncertainty and should be the highest priority in further study of Procellariiformes phylogenetics.

Addressing these, and other uncertainties uncovered, to complement the largely resolved phylogeny presented in this thesis would move towards a full understanding of Procellariiformes systematics and taxonomy. More broadly, the generation of datasets of homologous genomic markers from further clades will eventually allow the integration of these datasets into larger scale phylogenies. For example, Vianna et al. (2020) mined UCEs from whole genome sequences of penguins (Sphenisciformes) for phylogenetic analyses, and Hruska et al. (2023) used UCEs to build a phylogeny of herons (Pelecaniformes). These datasets, combined with that provided for Procellariiformes here, could eventually contribute to a large and high resolution phylogenomic dataset of the waterbird clade Aequornithes.

5.1.2 Future directions in studying the impacts of sex-linked genomic markers on phylogenetic reconstruction

The resolution of the phylogeny of Procellariiformes improved as a result of filtering out sex-linked genomic loci, which produced sex-biased results as shown in Chapter 3. This suggests that sex-linked sites and loci could be considered a relevant subsample for filtering in phylogenetic studies making use of genome-wide markers. This result was clearly illustrated in Procellariiformes, having an effect largely at shallow evolutionary scales. The hypothesised mechanisms accounting for this, of incorporation of W chromosome reads into Z chromosome loci assemblies that introduces alternative alleles and systematically higher missing data in females could be anticipated to affect other taxonomic groups when using similar sequence capture or other methods targeting both autosomal and sex chromosome sequences. Although the heterogametic chromosome is expected to largely degenerate over time (Charlesworth & Charlesworth 2000), there is considerable variation in degeneration of the bird W chromosome (Zhou et al. 2014a). Less degeneration of the heterogametic chromosome could allow more sequences to contaminate assemblies of the homomorphic chromosome in the heterogametic sex only, leading to sex-biased results. Testing this in a close relative of Procellariiformes, such as the sister order Sphenisciformes, and in more taxonomically distant bird, and other, groups should be a priority to determine if this is a local or more general effect.

This study further illustrated the importance in filtering of phylogenomic datasets in interpreting results, as has been demonstrated by previous work on other metrics that can affect inference (Blom et al. 2017, Bossert et al. 2017, Gilbert et al. 2018, Hosner et al. 2016, Hutter & Duellman 2023, Mclean et al. 2019, Molloy & Warnow 2018, Portik & Wiens 2021, Simmons et al. 2016, Streicher et al. 2016, Van Dam et al. 2021, White & Braun 2019, Wiens & Morrill 2011), although with little concordance (e.g. Chan et al. 2020). Sequence capture methods and their applications have received recent reviews (Andermann et al. 2020, Jones & Good 2016, Zhang et al. 2019). Given the growing popularity of studies employing sequence capture methods for phylogenetic inference, the multitude of different filtering parameters implicated in altering results, and the very recent years in which many of these studies have been published indicates that the topic is particularly relevant for a wide-scale review of the best practices in bioinformatic filtering of these datasets.

5.1.3 Future directions in studying the evolution of mitogenome duplications in birds and other organisms

In Chapter 4 mapping sequenced reads to mitogenomes found expansive duplications in almost all Procellariiformes species, with underlying patterns found through applying this to the well resolved tree. Degeneration of duplicated regions of variable extents was widespread, and lost entirely on only two occasions in small clades. The study on Procellariiformes here represents one of the largest in scale of this trait in birds, and further large scale studies combining molecular sequencing datasets with well resolved phylogenies will be needed to refine our understanding of the prevalence and evolution of mitogenome duplications in bird species.

The unduplicated mitochondrial gene order found in the chicken that was originally thought to be ancestral in birds (Gibb et al. 2007) is not shared with other vertebrates (Desjardins & Morais 1990). Duplications are now known to be frequent throughout the bird tree, and have been inferred as the ancestral state in all birds (Urantówka et al. 2020). Rather than ancestral, the unduplicated mitogenome of the chicken, and all other mitogenome rearrangements found in birds, may instead be derived from a single ancestral mitogenome duplication. It is possible that the evolution of a duplicated mitochondrial region may represent an important shift prior to the rapid diversification of bird lineages, with potential functional implications. A duplicated region was inferred to be present in the ancestor of Procellariiformes, providing evidence of remarkable stasis of this trait over more than 60 million years and indicating the presence of selection pressures to maintain these duplications. Duplicated regions often retain multiple functional copies of genes (e.g. Eberhard et al. 2001, Morris-Pocock et al. 2010, Zhou et al. 2014b), further suggesting a benefit to preserving redundant gene copies. Indeed, mitogenome duplications have been associated with longevity and flight ability in birds (Skujina et al. 2016, Urantówka et al. 2018), that are especially intriguing in relevance to the Procellariiformes that occupy the extremes of these traits. Determining the presence or absence of mitogenomic duplications in further clades should allow more expansive comparative work to determine further associations with ecological and life history traits. Furthermore, understanding the possible relationships between duplications and mitogenomic replication and energy production that have been suggested (Jiang et al. 2007, Kumazawa et al. 1996, Mackiewicz et al. 2019, Skujina et al. 2016, Umeda et al. 2001), and that could explain potential functional benefits of duplications, through further biochemical study should be a priority.

5.2 Future directions in Procellariiformes evolution and ecology

Broad-scale phylogenies with dense sampling allow thorough comparative investigations into the evolution of morphological, ecological, and life history traits, and the relationships between form and function at the level of individual populations, species, and entire ecosystems (Feng et al. 2020, Pigot et al. 2020). The phylogeny of Procellariiformes was used in a comparative context to investigate the prevalence and evolution of mitogenome duplications. Beyond this genomic variation, the new tree of Procellariiformes raises many possibilities for further study. The generation of large-scale datasets of functionally associated morphological and ecological traits are a useful resource in this context (e.g the AVONET dataset of Tobias et al. (2022), and the hand-wing index dataset of Sheard et al. (2020)). In particular, the large variation in size and structure among Procellariiformes species is an interesting trait in a phylogenetic comparative context. Although albatrosses (Diomedeidae) are larger than all other species of Procellariiformes, except for overlap with *Macronectes* giant petrels (Procellariidae) in some species, fossil remains indicate that some stem group albatrosses were significantly smaller-bodied in the Oligocene and Miocene and as recently as the late Pliocene (Mayr & Goedert 2017, Mayr & Pavia 2014, Mayr & Tennyson 2020), indicating a potentially complex evolutionary history of this trait in Procellariiformes.

Testing the association of structure of the wing with flight and feeding specialisations in Procellariiformes (Kuroda 1954, Pennycuick 1982, Warham 1977) in a phylogenetic context is also an interesting possibility. Diving using various techniques, including plunge diving, surface diving and wing-propelled swimming (Ashmole 1971), is distributed throughout the newly generated tree of Procellariiformes. Members of Diomedeidae and Procellariidae of multiple genera dive in search of prey, suggesting either a common origin or multiple shifts towards this feeding specialisation. Additionally, the surface pattering behaviour of the storm-petrels is an interesting feeding specialisation given the confirmed paraphyly of the families Oceanitidae and Hydrobatidae. The relationship between form and function in terms of wing morphology has been investigated in waterbirds, with evidence of convergent evolution and strong phylogenetic signal (Baumgart et al. 2021, Wang & Clarke 2014). With more robust phylogenies and increasingly dense taxon sampling these relationships can be investigated at an increasingly fine scale.

The evolutionary relationships of Procellariiformes inferred also raise questions of the biogeography and speciation of this clade. Many seabirds show high philopatry despite high flight and dispersal abilities, which has been called the ‘seabird paradox’ (Milot et al. 2008).

Procellariiformes exemplify this pattern, with the ability to cover extreme oceanic distances yet also extreme philopatry in some species (Danckwerts et al. 2021, Weimerskirch et al. 2014). In seabirds, speciation has arisen in the presence of geographical barriers, and without such barriers, where differences in oceanic regimes, breeding and nonbreeding distributions, and breeding phenology can disrupt gene flow (Friesen 2015, Munro & Burg 2017). Across Procellariiformes, species and populations with high connectivity such as *Pachyptila*, and those with distinct relationships that do not reflect geographic proximity, such as in the populations of *Pelagodroma marina* were found. Within Hydrobatidae, two main clades that largely conformed to Atlantic Ocean and Pacific Ocean breeding species were found, and a similar, but also imperfect pattern, was apparent in *Pterodroma* species. Gaining a further understanding of these patterns could offer insights into the biogeographical and speciation processes that have formed the Procellariiformes populations observed today, and into other similarly oceanic groups that can cover the globe unlike any other organisms.

Appendices

Appendix 2.1: A comparison of Procellariiformes classifications in major avian checklists: IOC World Bird List v13.1 (Gill et al. 2023), The Clements Checklist of Birds of the World v2022 (Clements et al. 2022), Handbook of the Birds of the World and BirdLife International Taxonomic Checklist Version 7 (HBW and BirdLife International 2022), and The Howard and Moore Complete Checklist of the Birds of the World version 4.1 (Christidis et al. 2018). Families are denoted by bold roman text and preceded by 'Family', genera are denoted by bold italic text, species are denoted by non-bold italic text, subspecies are denoted by indented non-bold italic text. Species and subspecies marked with a cross symbol are extinct or likely to be extinct.

IOC 13.1	**Clements v2022**	**HBW-BirdLife Version 7**	**Howard & Moore 4.1**
Family: Diomedeidae	**Family: Diomedeidae**	**Family: Diomedeidae**	**Family: Diomedeidae**
Phoebastria	***Phoebastria***	***Phoebastria***	***Phoebastria***
Phoebastria immutabilis	*Phoebastria immutabilis*	*Phoebastria immutabilis*	*Phoebastria immutabilis*
Phoebastria nigripes	*Phoebastria nigripes*	*Phoebastria nigripes*	*Phoebastria nigripes*
Phoebastria irrorata	*Phoebastria irrorata*	*Phoebastria irrorata*	*Phoebastria irrorata*
Phoebastria albatrus	*Phoebastria albatrus*	*Phoebastria albatrus*	*Phoebastria albatrus*
Diomedea	***Diomedea***	***Diomedea***	***Diomedea***
Diomedea exulans	*Diomedea exulans*	*Diomedea exulans*	*Diomedea exulans*
Diomedea antipodensis	*D. e. antipodensis*	*Diomedea antipodensis*	*D. e. antipodensis*
D. a. gibsoni	*D. e. gibsoni*	*D. a. gibsoni*	*D. e. gibsoni*
Diomedea amsterdamensis	*D. e. amsterdamensis*	*Diomedea amsterdamensis*	*D. e. amsterdamensis*
Diomedea dabbenena	*D. e. dabbenena*	*Diomedea dabbenena*	*D. e. dabbenena*
Diomedea epomophora	*Diomedea epomophora*	*Diomedea epomophora*	*Diomedea epomophora*
Diomedea sanfordi	*D. e. sanfordi*	*Diomedea sanfordi*	*D. e. sanfordi*
Phoebetria	***Phoebetria***	***Phoebetria***	***Phoebetria***
Phoebetria fusca	*Phoebetria fusca*	*Phoebetria fusca*	*Phoebetria fusca*
Phoebetria palpebrata	*Phoebetria palpebrata*	*Phoebetria palpebrata*	*Phoebetria palpebrata*
Thalassarche	***Thalassarche***	***Thalassarche***	***Thalassarche***
Thalassarche melanophris	*Thalassarche melanophris*	*Thalassarche melanophris*	*Thalassarche melanophris*
Thalassarche impavida	*T. m. impavida*	*Thalassarche impavida*	*T. m. impavida*
Thalassarche cauta	*Thalassarche cauta*	*Thalassarche cauta*	*Thalassarche cauta*
T. c. steadi	*T. c. steadi*	*Thalassarche steadi*	*T. c. steadi*
Thalassarche eremita	*Thalassarche eremita*	*Thalassarche eremita*	*T. c. eremita*
Thalassarche salvini	*Thalassarche salvini*	*Thalassarche salvini*	*T. c. salvini*
Thalassarche chrysostoma	*Thalassarche chrysostoma*	*Thalassarche chrysostoma*	*Thalassarche chrysostoma*
Thalassarche chlororhynchos	*Thalassarche chlororhynchos*	*Thalassarche chlororhynchos*	*Thalassarche chlororhynchos*

Thalassarche carteri	*T. c. carteri*	*Thalassarche carteri*	*T. c. carteri*
Thalassarche bulleri	*Thalassarche bulleri*	*Thalassarche bulleri*	*Thalassarche bulleri*
T. b. platei	*T. b. platei*	*T. b. platei*	*T. b. platei*
Family: Oceanitidae	**Family: Oceanitidae**	**Family: Oceanitidae**	**Family: Oceanitidae**
Oceanites	***Oceanites***	***Oceanites***	***Oceanites***
Oceanites oceanicus	*Oceanites oceanicus*	*Oceanites oceanicus*	*Oceanites oceanicus*
O. o. exasperatus	*O. o. exasperatus*	*O. o. exasperatus*	*O. o. exasperatus*
O. o. chilensis	*O. o. chilensis*	*O. o. chilensis*	Not recognised
Oceanites gracilis	*Oceanites gracilis*	*Oceanites gracilis*	*Oceanites gracilis*
O. g. galapagoensis	*O. g. galapagoensis*	*O. g. galapagoensis*	*O. g. galapagoensis*
Oceanites pincoyae	*Oceanites pincoyae*	*Oceanites pincoyae*	Not recognised
Garrodia	***Garrodia***	***Garrodia***	***Garrodia***
Garrodia nereis	*Garrodia nereis*	*Garrodia nereis*	*Garrodia nereis*
Pelagodroma	***Pelagodroma***	***Pelagodroma***	***Pelagodroma***
Pelagodroma marina	*Pelagodroma marina*	*Pelagodroma marina*	*Pelagodroma marina*
P. m. hypoleuca	*P. m. hypoleuca*	*P. m. hypoleuca*	*P. m. hypoleuca*
P. m. eadesorum	*P. m. eadesorum*	*P. m. eadesorum*	*P. m. eadesorum*
P. m. dulciae	*P. m. dulciae*	*P. m. dulciae*	*P. m. dulciae*
P. m. maoriana	*P. m. maoriana*	*P. m. maoriana*	*P. m. maoriana*
P. m. albiclunis	*P. m. albiclunis*	*P. m. albiclunis*	*P. m. albiclunis*
Fregetta	***Fregetta***	***Fregetta***	***Fregetta***
Fregetta grallaria	*Fregetta grallaria*	*Fregetta grallaria*	*Fregetta grallaria*
F. g. leucogaster	*F. g. leucogaster*	*F. g. leucogaster*	*F. g. leucogaster*
F. g. segethi	*F. g. segethi*	*F. g. segethi*	*F. g. segethi*
F. g. titan	*F. g. titan*	*F. g. titan*	*F. g. titan*
Not recognised	Not recognised	*F. g. guttata*	Not recognised
Fregetta tropica	*Fregetta tropica*	*Fregetta tropica*	*Fregetta tropica*
F. t. melanoleuca	*F. t. melanoleuca*	*F. t. melanoleuca*	*F. t. melanoleuca*
Fregetta lineata	*Fregetta lineata*	Not recognised	Not recognised
Fregetta maoriana	*Fregetta maoriana*	*Fregetta maoriana*	*Fregetta maoriana*
Nesofregetta	***Nesofregetta***	***Nesofregetta***	***Nesofregetta***
Nesofregetta fuliginosa	*Nesofregetta fuliginosa*	*Nesofregetta fuliginosa*	*Nesofregetta fuliginosa*
Family: Hydrobatidae	**Family: Hydrobatidae**	**Family: Hydrobatidae**	**Family: Hydrobatidae**
Hydrobates	***Hydrobates***	***Hydrobates***	***Hydrobates***
Hydrobates pelagicus	*Hydrobates pelagicus*	*Hydrobates pelagicus*	*Hydrobates pelagicus*
H. p. melitensis	*H. p. melitensis*	*H. p. melitensis*	Not recognised
Hydrobates furcatus	*Hydrobates furcatus*	*Hydrobates furcatus*	*Hydrobates furcatus*
H. f. plumbeus	*H. f. plumbeus*	*H. f. plumbeus*	*H. f. plumbeus*
Hydrobates hornbyi	*Hydrobates hornbyi*	*Hydrobates hornbyi*	*Hydrobates hornbyi*
Hydrobates monorhis	*Hydrobates monorhis*	*Hydrobates monorhis*	*Hydrobates monorhis*
Hydrobates matsudairae	*Hydrobates matsudairae*	*Hydrobates matsudairae*	*Hydrobates matsudairae*
Hydrobates leucorhous	*Hydrobates leucorhous*	*Hydrobates leucorhous*	*Hydrobates leucorhous*

H. l. chapmani	*H. l. chapmani*	*H. l. chapmani*	*H. l. chapmani*
Hydrobates socorroensis	*Hydrobates socorroensis*	*Hydrobates socorroensis*	*H. l. socorroensis*
Hydrobates cheimomnestes	*Hydrobates cheimomnestes*	*Hydrobates cheimomnestes*	*H. l. cheimomnestes*
Hydrobates homochroa	*Hydrobates homochroa*	*Hydrobates homochroa*	*Hydrobates homochroa*
Hydrobates castro	*Hydrobates castro*	*Hydrobates castro*	*Hydrobates castro*
Hydrobates monteiroi	*Hydrobates monteiroi*	*Hydrobates monteiroi*	*Hydrobates monteiroi*
Hydrobates jabejabe	*Hydrobates jabejabe*	*Hydrobates jabejabe*	Not recognised
Hydrobates tethys	*Hydrobates tethys*	*Hydrobates tethys*	*Hydrobates tethys*
H. t. kelsalli	*H. t. kelsalli*	*H. t. kelsalli*	*H. t. kelsalli*
Hydrobates melania	*Hydrobates melania*	*Hydrobates melania*	*Hydrobates melania*
Hydrobates macrodactylus†	*Hydrobates macrodactylus*†	*Hydrobates macrodactylus*†	*Hydrobates macrodactylus*†
Hydrobates markhami	*Hydrobates markhami*	*Hydrobates markhami*	*Hydrobates markhami*
Hydrobates tristrami	*Hydrobates tristrami*	*Hydrobates tristrami*	*Hydrobates tristrami*
Hydrobates microsoma	*Hydrobates microsoma*	*Hydrobates microsoma*	*Hydrobates microsoma*
Family: Procellariidae	**Family: Procellariidae**	**Family: Procellariidae**	**Family: Procellariidae**
Macronectes	***Macronectes***	***Macronectes***	***Macronectes***
Macronectes giganteus	*Macronectes giganteus*	*Macronectes giganteus*	*Macronectes giganteus*
Macronectes halli	*Macronectes halli*	*Macronectes halli*	*Macronectes halli*
Fulmarus	***Fulmarus***	***Fulmarus***	***Fulmarus***
Fulmarus glacialis	*Fulmarus glacialis*	*Fulmarus glacialis*	*Fulmarus glacialis*
F. g. auduboni	*F. g. auduboni*	Not recognised	*F. g. auduboni*
F. g. rodgersii	*F. g. rodgersii*	*F. g. rodgersii*	*F. g. rodgersii*
Fulmarus glacialoides	*Fulmarus glacialoides*	*Fulmarus glacialoides*	*Fulmarus glacialoides*
Thalassoica	***Thalassoica***	***Thalassoica***	***Thalassoica***
Thalassoica antarctica	*Thalassoica antarctica*	*Thalassoica antarctica*	*Thalassoica antarctica*
Daption	***Daption***	***Daption***	***Daption***
Daption capense	*Daption capense*	*Daption capense*	*Daption capense*
D. c. australe	*D. c. australe*	*D. c. australe*	*D. c. australe*
Pagodroma	***Pagodroma***	***Pagodroma***	***Pagodroma***
Pagodroma nivea	*Pagodroma nivea*	*Pagodroma nivea*	*Pagodroma nivea*
P. n. major	*P. n. major*	*P. n. major*	*P. n. major*
Halobaena	***Halobaena***	***Halobaena***	***Halobaena***
Halobaena caerulea	*Halobaena caerulea*	*Halobaena caerulea*	*Halobaena caerulea*
Pachyptila	***Pachyptila***	***Pachyptila***	***Pachyptila***
Pachyptila vittata	*Pachyptila vittata*	*Pachyptila vittata*	*Pachyptila vittata*
Pachyptila salvini	*Pachyptila salvini*	*Pachyptila salvini*	*Pachyptila salvini*
Pachyptila macgillivrayi	*Pachyptila macgillivrayi*	*Pachyptila macgillivrayi*	*P. s. macgillivrayi*
Pachyptila desolata	*Pachyptila desolata*	*Pachyptila desolata*	*Pachyptila desolata*
Not recognised	*P. d. altera*	Not recognised	Not recognised
Not recognised	*P. d. banksi*	Not recognised	Not recognised
Pachyptila belcheri	*Pachyptila belcheri*	*Pachyptila belcheri*	*Pachyptila belcheri*
Pachyptila turtur	*Pachyptila turtur*	*Pachyptila turtur*	*Pachyptila turtur*

P. t. subantarctica	Not recognised	Not recognised	Not recognised
Not recognised	Not recognised	Not recognised	*P. t. eatoni*
Pachyptila crassirostris	*Pachyptila crassirostris*	*Pachyptila crassirostris*	*Pachyptila crassirostris*
P. c. pyramidalis	*P. c. pyramidalis*	*P. c. pyramidalis*	*P. c. pyramidalis*
P. c. flemingi	*P. c. flemingi*	*P. c. flemingi*	*P. c. flemingi*
Aphrodroma	***Aphrodroma***	***Aphrodroma***	***Aphrodroma***
Aphrodroma brevirostris	*Aphrodroma brevirostris*	*Aphrodroma brevirostris*	*Aphrodroma brevirostris*
Pterodroma	***Pterodroma***	***Pterodroma***	***Pterodroma***
Pterodroma macroptera	*Pterodroma macroptera*	*Pterodroma macroptera*	*Pterodroma macroptera*
Pterodroma gouldi	*Pterodroma gouldi*	*Pterodroma gouldi*	*P. m. gouldi*
Pterodroma lessonii	*Pterodroma lessonii*	*Pterodroma lessonii*	*Pterodroma lessonii*
Pterodroma incerta	*Pterodroma incerta*	*Pterodroma incerta*	*Pterodroma incerta*
Pterodroma solandri	*Pterodroma solandri*	*Pterodroma solandri*	*Pterodroma solandri*
Pterodroma magentae	*Pterodroma magentae*	*Pterodroma magentae*	*Pterodroma magentae*
Pterodroma ultima	*Pterodroma ultima*	*Pterodroma ultima*	*Pterodroma ultima*
Pterodroma mollis	*Pterodroma mollis*	*Pterodroma mollis*	*Pterodroma mollis*
Not recognised	*P. m. dubia*	Not recognised	Not recognised
Pterodroma madeira	*Pterodroma madeira*	*Pterodroma madeira*	*Pterodroma madeira*
Pterodroma feae	*Pterodroma feae*	*Pterodroma feae*	*Pterodroma feae*
Pterodroma deserta	*P. f. deserta*	*Pterodroma deserta*	*P. f. deserta*
Pterodroma cahow	*Pterodroma cahow*	*Pterodroma cahow*	*Pterodroma cahow*
Pterodroma hasitata	*Pterodroma hasitata*	*Pterodroma hasitata*	*Pterodroma hasitata*
Pterodroma caribbaea†	*P. h. caribbaea*†	*Pterodroma caribbaea*†	*P. h. caribbaea*†
Pterodroma externa	*Pterodroma externa*	*Pterodroma externa*	*Pterodroma externa*
Pterodroma neglecta	*Pterodroma neglecta*	*Pterodroma neglecta*	*Pterodroma neglecta*
P. n. juana	*P. n. juana*	*P. n. juana*	*P. n. juana*
Pterodroma arminjoniana	*Pterodroma arminjoniana*	*Pterodroma arminjoniana*	*Pterodroma arminjoniana*
Pterodroma heraldica	*Pterodroma heraldica*	*Pterodroma heraldica*	*P. n. heraldica*
Pterodroma atrata	*Pterodroma atrata*	*Pterodroma atrata*	*Pterodroma atrata*
Pterodroma alba	*Pterodroma alba*	*Pterodroma alba*	*Pterodroma alba*
Pterodroma baraui	*Pterodroma baraui*	*Pterodroma baraui*	*Pterodroma baraui*
Pterodroma sandwichensis	*Pterodroma sandwichensis*	*Pterodroma sandwichensis*	*Pterodroma sandwichensis*
Pterodroma phaeopygia	*Pterodroma phaeopygia*	*Pterodroma phaeopygia*	*Pterodroma phaeopygia*
Pterodroma inexpectata	*Pterodroma inexpectata*	*Pterodroma inexpectata*	*Pterodroma inexpectata*
Pterodroma cervicalis	*Pterodroma cervicalis*	*Pterodroma cervicalis*	*Pterodroma cervicalis*
Pterodroma occulta	*Pterodroma occulta*	*P. c. occulta*	*P. c. occulta*
Pterodroma nigripennis	*Pterodroma nigripennis*	*Pterodroma nigripennis*	*Pterodroma nigripennis*
Pterodroma axillaris	*Pterodroma axillaris*	*Pterodroma axillaris*	*Pterodroma axillaris*
Pterodroma hypoleuca	*Pterodroma hypoleuca*	*Pterodroma hypoleuca*	*Pterodroma hypoleuca*
Pterodroma leucoptera	*Pterodroma leucoptera*	*Pterodroma leucoptera*	*Pterodroma leucoptera*
Not recognised	*P. l. caledonica*	Not recognised	*P. l. caledonica*
Pterodroma brevipes	*Pterodroma brevipes*	*Pterodroma brevipes*	*Pterodroma brevipes*

P. b. magnificens	*P. b. magnificens*	Not recognised	Not recognised
Pterodroma cookii	*Pterodroma cookii*	*Pterodroma cookii*	*Pterodroma cookii*
Pterodroma defilippiana	*Pterodroma defilippiana*	*Pterodroma defilippiana*	*Pterodroma defilippiana*
Pterodroma longirostris	*Pterodroma longirostris*	*Pterodroma longirostris*	*Pterodroma longirostris*
Pterodroma pycrofti	*Pterodroma pycrofti*	*Pterodroma pycrofti*	*Pterodroma pycrofti*
Bulweria	***Bulweria***	***Bulweria***	***Bulweria***
Bulweria bulwerii	*Bulweria bulwerii*	*Bulweria bulwerii*	*Bulweria bulwerii*
Bulweria bifax†	*Bulweria bifax*†	*Bulweria bifax*†	*Bulweria bifax*†
Bulweria fallax	*Bulweria fallax*	*Bulweria fallax*	*Bulweria fallax*
Pseudobulweria	***Pseudobulweria***	***Pseudobulweria***	***Pseudobulweria***
Pseudobulweria aterrima	*Pseudobulweria aterrima*	*Pseudobulweria aterrima*	*Pseudobulweria aterrima*
Pseudobulweria rupinarum†	*Pseudobulweria rupinarum*†	*Pseudobulweria rupinarum*†	*Pseudobulweria rupinarum*†
Pseudobulweria rostrata	*Pseudobulweria rostrata*	*Pseudobulweria rostrata*	*Pseudobulweria rostrata*
P. r. trouessarti	*P. r. trouessarti*	*P. r. trouessarti*	*P. r. trouessarti*
Pseudobulweria becki	*Pseudobulweria becki*	*Pseudobulweria becki*	*Pseudobulweria becki*
Pseudobulweria macgillivrayi	*Pseudobulweria macgillivrayi*	*Pseudobulweria macgillivrayi*	*Pseudobulweria macgillivrayi*
Procellaria	***Procellaria***	***Procellaria***	***Procellaria***
Procellaria cinerea	*Procellaria cinerea*	*Procellaria cinerea*	*Procellaria cinerea*
Procellaria aequinoctialis	*Procellaria aequinoctialis*	*Procellaria aequinoctialis*	*Procellaria aequinoctialis*
P. a. steadi	Not recognised	Not recognised	Not recognised
Procellaria conspicillata	*Procellaria conspicillata*	*Procellaria conspicillata*	*Procellaria conspicillata*
Procellaria parkinsoni	*Procellaria parkinsoni*	*Procellaria parkinsoni*	*Procellaria parkinsoni*
Procellaria westlandica	*Procellaria westlandica*	*Procellaria westlandica*	*Procellaria westlandica*
Calonectris	***Calonectris***	***Calonectris***	***Calonectris***
Calonectris leucomelas	*Calonectris leucomelas*	*Calonectris leucomelas*	*Calonectris leucomelas*
Calonectris diomedea	*Calonectris diomedea*	*Calonectris diomedea*	*Calonectris diomedea*
Calonectris borealis	*C. d. borealis*	*Calonectris borealis*	*Calonectris borealis*
Calonectris edwardsii	*Calonectris edwardsii*	*Calonectris edwardsii*	*Calonectris edwardsii*
Ardenna	***Ardenna***	***Ardenna***	***Ardenna***
Ardenna pacifica	*Ardenna pacifica*	*Ardenna pacifica*	*Ardenna pacifica*
Not recognised	*A. p. chlororhyncha*	*A. p. chlororhyncha*	*A. p. chlororhyncha*
Ardenna bulleri	*Ardenna bulleri*	*Ardenna bulleri*	*Ardenna bulleri*
Ardenna grisea	*Ardenna grisea*	*Ardenna grisea*	*Ardenna grisea*
Ardenna tenuirostris	*Ardenna tenuirostris*	*Ardenna tenuirostris*	*Ardenna tenuirostris*
Ardenna creatopus	*Ardenna creatopus*	*Ardenna creatopus*	*Ardenna creatopus*
Ardenna carneipes	*Ardenna carneipes*	*Ardenna carneipes*	*Ardenna carneipes*
Ardenna gravis	*Ardenna gravis*	*Ardenna gravis*	*Ardenna gravis*
Puffinus	***Puffinus***	***Puffinus***	***Puffinus***
Puffinus nativitatis	*Puffinus nativitatis*	*Puffinus nativitatis*	*Puffinus nativitatis*
Puffinus puffinus	*Puffinus puffinus*	*Puffinus puffinus*	*Puffinus puffinus*
P. p. canariensis	Not recognised	Not recognised	Not recognised
Puffinus yelkouan	*Puffinus yelkouan*	*Puffinus yelkouan*	*Puffinus yelkouan*

Puffinus mauretanicus	*Puffinus mauretanicus*	*Puffinus mauretanicus*	*P. y. mauretanicus*
Puffinus bryani	*Puffinus bryani*	*Puffinus bryani*	*Puffinus bryani*
Puffinus opisthomelas	*Puffinus opisthomelas*	*Puffinus opisthomelas*	*Puffinus opisthomelas*
Puffinus auricularis	*Puffinus auricularis*	*Puffinus auricularis*	*Puffinus auricularis*
Puffinus newelli	*Puffinus newelli*	*Puffinus newelli*	*P. a. newelli*
Puffinus myrtae	*Puffinus myrtae*	*Puffinus myrtae*	*P. a. myrtae*
Puffinus gavia	*Puffinus gavia*	*Puffinus gavia*	*Puffinus gavia*
Puffinus huttoni	*Puffinus huttoni*	*Puffinus huttoni*	*Puffinus huttoni*
Puffinus lherminieri	*Puffinus lherminieri*	*Puffinus lherminieri*	*Puffinus lherminieri*
P. l. loyemilleri	*P. l. loyemilleri*	Not recognised	*P. l. loyemilleri*
Puffinus baroli	*Puffinus baroli*	*P. l. baroli*	*P. l. baroli*
Puffinus boydi	*Puffinus boydi*	*P. l. boydi*	*P. l. boydi*
Puffinus bailloni	*Puffinus bailloni*	*Puffinus bailloni*	*Puffinus bailloni*
P. b. nicolae	*P. b. nicolae*	*P. b. nicolae*	*P. b. nicolae*
P. b. colstoni	Not recognised	Not recognised	Not recognised
P. b. dichrous	*P. b. dichrous*	*P. b. dichrous*	*P. b. dichrous*
P. b. gunax	*P. b. gunax*	*P. b. gunax*	*P. b. gunax*
Not recognised	Not recognised	Not recognised	*P. b. polynesiae*
Puffinus bannermani	*Puffinus bannermani*	*Puffinus bannermani*	*P. b. bannermani*
Puffinus persicus	*Puffinus persicus*	*Puffinus persicus*	*P. b. persicus*
P. p. temptator	*P. p. temptator*	*P. p. temptator*	*P. b. temptator*
Puffinus subalaris	*Puffinus subalaris*	*Puffinus subalaris*	*Puffinus subalaris*
Puffinus heinrothi	*Puffinus heinrothi*	*Puffinus heinrothi*	*Puffinus heinrothi*
Puffinus assimilis	*Puffinus assimilis*	*Puffinus assimilis*	*Puffinus assimilis*
P. a. haurakiensis	*P. a. haurakiensis*	*P. a. haurakiensis*	*P. a. haurakiensis*
P. a. kermadecensis	*P. a. kermadecensis*	*P. a. kermadecensis*	*P. a. kermadecensis*
P. a. tunneyi	*P. a. tunneyi*	*P. a. tunneyi*	*P. a. tunneyi*
Puffinus elegans	*Puffinus elegans*	*Puffinus elegans*	*P. a. elegans*
Pelecanoides	***Pelecanoides***	***Pelecanoides***	***Pelecanoides***
Pelecanoides garnotii	*Pelecanoides garnotii*	*Pelecanoides garnotii*	*Pelecanoides garnotii*
Pelecanoides magellani	*Pelecanoides magellani*	*Pelecanoides magellani*	*Pelecanoides magellani*
Pelecanoides georgicus	*Pelecanoides georgicus*	*Pelecanoides georgicus*	*Pelecanoides georgicus*
P. g. whenuahouensis	*P. g. whenuahouensis*	*Pelecanoides whenuahouensis*	Not recognised
Pelecanoides urinatrix	*Pelecanoides urinatrix*	*Pelecanoides urinatrix*	*Pelecanoides urinatrix*
P. u. dacunhae	*P. u. dacunhae*	*P. u. dacunhae*	*P. u. dacunhae*
P. u. berard	*P. u. berard*	*P. u. berard*	*P. u. berard*
P. u. chathamensis	*P. u. chathamensis*	*P. u. chathamensis*	*P. u. chathamensis*
P. u. exsul	*P. u. exsul*	*P. u. exsul*	*P. u. exsul*
P. u. coppingeri	*P. u. coppingeri*	*P. u. coppingeri*	*P. u. coppingeri*

Appendix 2.2: Taxonomic list of Procellariiformes used for sampling.

Family: Diomedeidae
Phoebastria irrorata
Phoebastria nigripes
Phoebastria immutabilis
Phoebastria albatrus
Diomedea exulans
Diomedea antipodensis antipodensis
Diomedea antipodensis gibsoni
Diomedea amsterdamensis
Diomedea dabbenena
Diomedea epomophora
Diomedea sanfordi
Phoebetria fusca
Phoebetria palpebrata
Thalassarche melanophris
Thalassarche impavida
Thalassarche cauta cauta
Thalassarche cauta steadi
Thalassarche eremita
Thalassarche salvini
Thalassarche chrysostoma
Thalassarche chlororhynchos
Thalassarche carteri
Thalassarche bulleri bulleri
Thalassarche bulleri platei

Family: Oceanitidae
Oceanites oceanicus oceanicus
Oceanites oceanicus exasperatus
Oceanites oceanicus chilensis
Oceanites gracilis gracilis
Oceanites gracilis galapagoensis
Oceanites pincoyae
Garrodia nereis
Pelagodroma marina marina
Pelagodroma marina hypoleuca
Pelagodroma marina eadesorum
Pelagodroma marina dulciae
Pelagodroma marina maoriana
Pelagodroma marina albiclunis
Fregetta grallaria grallaria
Fregetta grallaria leucogaster
Fregetta grallaria segethi
Fregetta grallaria titan
Fregetta tropica tropica
Fregetta tropica melanoleuca
Fregetta lineata
Fregetta maoriana

Nesofregetta fuliginosa

Family: Hydrobatidae
Hydrobates pelagicus pelagicus
Hydrobates pelagicus melitensis
Oceanodroma furcatus furcatus
Oceanodroma furcatus plumbeus
Oceanodroma hornbyi
Oceanodroma monorhis
Oceanodroma matsudairae
Oceanodroma leucorhous leucorhous
Oceanodroma leucorhous chapmani
Oceanodroma socorroensis
Oceanodroma cheimomnestes
Oceanodroma homochroa
Oceanodroma castro
Oceanodroma monteiroi
Oceanodroma jabejabe
Oceanodroma tethys tethys
Oceanodroma tethys kelsalli
Oceanodroma melania
Oceanodroma markhami
Oceanodroma tristrami
Oceanodroma microsoma

Family: Procellariidae
Macronectes giganteus
Macronectes halli
Fulmarus glacialis glacialis
Fulmarus glacialis auduboni
Fulmarus glacialis rodgersii
Fulmarus glacialoides
Thalassoica antarctica
Daption capense capense
Daption capense australe
Pagodroma nivea nivea
Pagodroma nivea major
Halobaena caerulea
Pachyptila vittata
Pachyptila salvini
Pachyptila macgillivrayi
Pachyptila desolata desolata
Pachyptila desolata banksi
Pachyptila belcheri
Pachyptila turtur turtur
Pachyptila turtur eatoni
Pachyptila turtur subantarctica
Pachyptila crassirostris crassirostris
Pachyptila crassirostris pyramidalis
Pachyptila crassirostris flemingi

Aphrodroma brevirostris
Pterodroma macroptera
Pterodroma gouldi
Pterodroma lessonii
Pterodroma incerta
Pterodroma solandri
Pterodroma magentae
Pterodroma ultima
Pterodroma mollis
Pterodroma madeira
Pterodroma feae
Pterodroma deserta
Pterodroma cahow
Pterodroma hasitata
Pterodroma externa
Pterodroma neglecta neglecta
Pterodroma neglecta juana
Pterodroma arminjoniana
Pterodroma heraldica
Pterodroma atrata
Pterodroma alba
Pterodroma baraui
Pterodroma sandwichensis
Pterodroma phaeopygia
Pterodroma inexpectata
Pterodroma cervicalis
Pterodroma occulta
Pterodroma nigripennis
Pterodroma axillaris
Pterodroma hypoleuca
Pterodroma leucoptera leucoptera
Pterodroma leucoptera caledonica
Pterodroma brevipes brevipes
Pterodroma brevipes magnificens
Pterodroma cookii
Pterodroma defilippiana
Pterodroma longirostris
Pterodroma pycrofti
Bulweria bulwerii
Bulweria fallax
Pseudobulweria aterrima
Pseudobulweria rostrata rostrata
Pseudobulweria rostrata trouessarti
Pseudobulweria becki
Pseudobulweria macgillivrayi
Procellaria cinerea
Procellaria aequinoctialis aequinoctialis
Procellaria aequinoctialis steadi
Procellaria conspicillata
Procellaria parkinsoni

Procellaria westlandica
Calonectris leucomelas
Calonectris diomedea
Calonectris borealis
Calonectris edwardsii
Ardenna pacifica pacifica
Ardenna pacifica chlororhyncha
Ardenna bulleri
Ardenna grisea
Ardenna tenuirostris
Ardenna creatopus
Ardenna carneipes
Ardenna gravis
Puffinus nativitatis
Puffinus puffinus puffinus
Puffinus puffinus canariensis
Puffinus yelkouan
Puffinus mauretanicus
Puffinus bryani
Puffinus opisthomelas
Puffinus auricularis
Puffinus newelli
Puffinus myrtae
Puffinus gavia
Puffinus huttoni
Puffinus lherminieri lherminieri
Puffinus lherminieri loyemilleri
Puffinus baroli
Puffinus boydi
Puffinus bailloni bailloni
Puffinus bailloni nicolae
Puffinus bailloni colstoni
Puffinus bailloni dichrous
Puffinus bailloni gunax
Puffinus bannermani
Puffinus persicus persicus
Puffinus persicus temptator
Puffinus subalaris
Puffinus heinrothi
Puffinus assimilis assimilis
Puffinus assimilis haurakiensis
Puffinus assimilis kermadecensis
Puffinus assimilis tunneyi
Puffinus elegans
Pelecanoides garnotii
Pelecanoides magellani
Pelecanoides georgicus georgicus
Pelecanoides georgicus whenuahouensis
Pelecanoides urinatrix urinatrix
Pelecanoides urinatrix dacunhae

Pelecanoides urinatrix berard
Pelecanoides urinatrix chathamensis
Pelecanoides urinatrix exsul
Pelecanoides urinatrix coppingeri

Appendix 2.3: List of sequenced samples. Taxon reference numbers match those used in figures. Museum abbreviations are given in Chapter 2 section 2.2.2.

Taxon and reference number	Museum/collector and sample number	Collecting locality
Aphrodroma brevirostris 1	Gary Nunn KGP-1	Gough Island
Aphrodroma brevirostris 2	UWBM 61673	South Georgia
Ardenna bulleri 1	USNM 613921	North Pacific Ocean
Ardenna bulleri 2	UWBM 82792	Northland, New Zealand
Ardenna carneipes 1	AMNH DOT-17805	Ocean off Albany, Western Australia
Ardenna carneipes 2	UWBM 82812	Off Great Barrier Island, New Zealand
Ardenna carneipes 3	UWBM 85403	North Pacific Ocean
Ardenna carneipes 4	Yves Cherel FFSH01	St. Paul Island, French Southern and Antarctic Lands
Ardenna creatopus 1	AMNH DOT-3131	Santa Clara, Juan Fernandez Islands, Chile
Ardenna creatopus 2	UWBM 61948	North Pacific Ocean
Ardenna gravis 1	Gary Nunn GS-6	Gough Island
Ardenna gravis 2	Jeremy Austin/Spanish AGraJ1	Gough Island
Ardenna grisea 1	Jeremy Austin/Spanish Agri2	Kidney island, Falkland Islands
Ardenna grisea 2	USNM 621351	Johnston Atoll
Ardenna pacifica chlororhyncha 1	Jeremy Austin/Spanish Apac76	Sand Island, Johnston Atoll
Ardenna pacifica chlororhyncha 2	USNM 643456	Kauai, Hawaii, USA
Ardenna tenuirostris 1	Jeremy Austin/Spanish ATenJ2	Woolamai, Victoria, Australia
Ardenna tenuirostris 2	UWBM 51646	Magadan, Russia
Bulweria bulwerii 1	Vincent Bretagnolle E006980	Madeira
Bulweria bulwerii 2	USNM 631387	Johnston Atoll
Calonectris borealis 1	Jeremy Austin/Spanish CBor533	Montaña Clara, Lanzarote, Canary Islands, Spain
Calonectris borealis 2	Jeremy Austin/Spanish CBor542	Montaña Clara, Lanzarote, Canary Islands, Spain
Calonectris diomedea 1	Jeremy Austin/Spanish CDio669	Menorca, Balearic Islands, Spain
Calonectris diomedea 2	USNM 620710	Off Cape Hatteras National Seashore, North Carolina, USA
Calonectris edwardsii 1	Jeremy Austin/Spanish CEdw105	Curral Velho, Cape Verde
Calonectris edwardsii 2	Jeremy Austin/Spanish CEdw660	Curral Velho, Cape Verde
Calonectris leucomelas 1	LSUMZ 16967	Mikura Island, Seven Islands of Izu, Japan
Calonectris leucomelas 2	Jeremy Austin/Spanish CLeu13	Mikura Islands, Japan
Ciconia maguari 1	USNM 614527	Estancia El Tala, Near Puerto Constanza, Entre Rios, Argentina
Daption capense capense 1	BMNH 109516443	South Georgia
Daption capense capense 2	KU 21827	Muriwai Beach, North Island, New Zealand
Diomedea amsterdamensis 1	Yves Cherel AA10	Amsterdam Island, French Southern and Antarctic Lands
Diomedea amsterdamensis 2	Yves Cherel AA15	Amsterdam Island, French Southern and Antarctic Lands
Diomedea amsterdamensis 3	Yves Cherel AA3	Amsterdam Island, French Southern and Antarctic Lands
Diomedea antipodensis antipodensis 1	Theresa Burg DeAn9917	Antipodes Islands, New Zealand
Diomedea antipodensis antipodensis 2	Theresa Burg DeAn9918	Antipodes Islands, New Zealand

Diomedea antipodensis antipodensis 3	UWBM 81028	Southeast of Cape Kidnappers, New Zealand
Diomedea antipodensis gibsoni 1	Theresa Burg DeAd9932	Adams Island, Auckland Islands, New Zealand
Diomedea antipodensis gibsoni 2	Theresa Burg DeAd9933	Adams Island, Auckland Islands, New Zealand
Diomedea antipodensis gibsoni 3	UWBM 81027	100 km north-northeast of East Cape; New Zealand
Diomedea dabbenena 1	Gary Nunn WA-4	Gough Island
Diomedea dabbenena 2	Gary Nunn WA-3	Gough Island
Diomedea epomophora 1	NMV 50490	Unknown
Diomedea exulans 1	Richard Phillips A1	Bird Island, South Georgia
Diomedea exulans 2	Richard Phillips A2	Bird Island, South Georgia
Diomedea exulans 3	Theresa Burg DIEXPM012	Pointe Morne, Kerguelen Islands, French Southern and Antarctic Lands
Diomedea exulans 4	Theresa Burg DIEXPM013	Pointe Morne, Kerguelen Islands, French Southern and Antarctic Lands
Diomedea sanfordi 1	Paul Scofield AV19580	South Bay, Kaikoura, New Zealand
Fregetta grallaria grallaria 1	AM 76074	Lord Howe Island, New South Wales, Australia
Fregetta grallaria grallaria 2	AM 76583	Lord Howe Island, New South Wales, Australia
Fregetta grallaria leucogaster 1	Gary Nunn WBSP-2	Gough Island
Fregetta grallaria leucogaster 2	Yves Cherel WBSP04	St. Paul Island, French Southern and Antarctic Lands
Fregetta grallaria leucogaster 3	Yves Cherel WBSP05	St. Paul Island, French Southern and Antarctic Lands
Fregetta grallaria segethi 1	AMNH DOT-3126	1 km N Vaqueria, Isla Robinson Crusoe, Chile
Fregetta grallaria titan 1	Alice Cibois & Jean-Claude Thibault 2019-23	South Pacific Ocean
Fregetta grallaria titan 2	Alice Cibois & Jean-Claude Thibault 2019-24	South Pacific Ocean
Fregetta tropica tropica 1	NHMO BI-42935	Bouvet Island
Fregetta tropica tropica 2	Petra Quillfeldt BBSP-Ker-AS22	Kerguelen Islands, French Southern and Antarctic Lands
Fregetta tropica tropica 3	Petra Quillfeldt BBSP-KGI-18	King George Island, South Shetland Islands
Fregetta tropica tropica 4	Richard Phillips H5	Bird Island, South Georgia
Fulmarus glacialis auduboni 1	USNM 623297	North of Keflavik, Iceland
Fulmarus glacialis glacialis 1	Hallvard Strom 4214435	Svalbard, Norway
Fulmarus glacialis glacialis 2	Hallvard Strom 4214658	Svalbard, Norway
Fulmarus glacialis rodgersii 1	Theresa Burg SB113	North Saanich, Vancouver Island, Canada
Fulmarus glacialis rodgersii 2	USNM 638827	Aleutians West Census Area, Alaska, USA
Fulmarus glacialoides 1	AMNH DOT-3210	Puerto Williams, Navarino Island, Chile
Fulmarus glacialoides 2	LSUMZ 13541	Scotia Confluence, South Atlantic Ocean
Garrodia nereis 1	BMNH 109516428	Falkland Islands
Garrodia nereis 2	Gary Nunn GBSP-3	Gough Island
Garrodia nereis 3	Richard Phillips RP20	Bird Island, South Georgia
Halobaena caerulea 1	KU 98793	Victoria, Australia
Halobaena caerulea 2	Richard Phillips C7	Bird Island, South Georgia
Halobaena caerulea 3	UWBM 61675	South Georgia
Hydrobates pelagicus melitensis 1	Vincent Bretagnolle MSP-1	Malta
Hydrobates pelagicus pelagicus 1	ZMUC P740	Nolsoy, Faroe Islands
Macronectes giganteus 1	Gary Nunn SPG-G-1	Gough Island
Macronectes giganteus 2	Gary Nunn SPG-G-2	Gough Island

Macronectes giganteus 3	Richard Phillips C1	Bird Island, South Georgia
Macronectes halli 1	Richard Phillips A9	Bird Island, South Georgia
Macronectes halli 2	UWBM 80999	Bounty Platform, New Zealand
Nesofregetta fuliginosa 1	Alice Cibois & Jean-Claude Thibault 2019-115	South Pacific Ocean
Nesofregetta fuliginosa 2	Alice Cibois & Jean-Claude Thibault 2019-117	South Pacific Ocean
Nesofregetta fuliginosa 3	USNM 614206	South Pacific
Oceanites gracilis gracilis 1	Heraldo Norambuena 0G04	Isla Chungungo, Coquimbo, Chile
Oceanites gracilis gracilis 2	Heraldo Norambuena 0G06	Isla Chungungo, Coquimbo, Chile
Oceanites gracilis gracilis 3	ZMUC FS2-20.6.92	Punta Santa Elena, Ecuador
Oceanites oceanicus chilensis 1	AMNH DOT-3175	San Antonia, V Region de Valparaiso, Chile
Oceanites oceanicus chilensis 2	Heraldo Norambuena OOC06	Cordillera Santiago, Metropolitana, Chile
Oceanites oceanicus chilensis 3	Heraldo Norambuena OOC07	Cordillera Santiago, Metropolitana, Chile
Oceanites oceanicus exasperatus 1	Petra Quillfeldt WSP-KGI-02	King George Island, South Shetland Islands
Oceanites oceanicus exasperatus 2	UAM 35626	Vicinity of Palmer Station, Antarctica
Oceanites oceanicus oceanicus 1	BMNH 109516427	Falkland Islands
Oceanites oceanicus oceanicus 2	Petra Quillfeldt WSP-Mayes-42	Kerguelen Islands, French Southern and Antarctic Lands
Oceanodroma castro 1	LSUMZ 41542	Louisiana, USA
Oceanodroma castro 2	Max Levy 8J	Ilheu de Praia, Azores, Portugal
Oceanodroma castro 3	USNM 602013	Nelson, Virginia, USA
Oceanodroma castro 4	ZMUC P1042	Isla Grande, Selvagem, Madeira
Oceanodroma cheimomnestes 1	Yuliana Bedolla AISP60	Guadalupe Island, Mexico
Oceanodroma cheimomnestes 2	Yuliana Bedolla AISP65	Guadalupe Island, Mexico
Oceanodroma furcatus furcatus 1	USNM 638711	Alaska, USA
Oceanodroma furcatus plumbeus 1	SDNHM 51719	Torrey Pines State Reserve, San Diego, California, USA
Oceanodroma homochroa 1	SDNHM 50480	San Diego, California, USA
Oceanodroma homochroa 2	SDNHM 51572	Spring Valley, San Diego County, California, USA
Oceanodroma homochroa 3	Yuliana Bedolla ASSP36	Todos Santos Sur Island, Mexico
Oceanodroma hornbyi 1	Heraldo Norambuena HH03	Pampa Indio Muerto, Atacama, Chile
Oceanodroma hornbyi 2	Heraldo Norambuena HH04	Pampa Indio Muerto, Atacama, Chile
Oceanodroma jabejabe 1	Jacob González-Solís 2801380	Icima, Cape Verde
Oceanodroma jabejabe 2	Jacob González-Solís 2801399	Icima, Cape Verde
Oceanodroma jabejabe 3	Jacob González-Solís 2801894	Icima, Cape Verde
Oceanodroma leucorhous chapmani 1	SDNHM 50807	North Pacific Ocean, off Baja California Sur, Mexico
Oceanodroma leucorhous leucorhous 1	LSUMZ 49240	Florida, USA
Oceanodroma leucorhous leucorhous 2	USNM 639040	Unalga Pass, East of Dutch Harbor, Alaska, USA
Oceanodroma markhami 1	Heraldo Norambuena HM31	Salar Grande, Tarapacá, Chile
Oceanodroma markhami 2	Heraldo Norambuena HM33	Salar Grande, Tarapacá, Chile
Oceanodroma matsudairae 1	YIO 1999-0221	Saitama, Honshu, Japan
Oceanodroma matsudairae 2	YIO 2005-0015	Hahajima, Ogasawara islands, Japan
Oceanodroma melania 1	KU 9119	South of Puerto Triunfo, Usulutan, El Salvador
Oceanodroma melania 2	SDNHM 50481	Balboa Park, San Diego, California, USA

Oceanodroma microsoma 1	LSUMZ 26377	Bay of Panama, Panama
Oceanodroma microsoma 2	LSUMZ 6363	California, USA
Oceanodroma monorhis 1	BMNH 109516461	Verhovsky Island, Russia
Oceanodroma monorhis 2	Jacob González-Solís 2801913	Icima, Cape Verde
Oceanodroma monorhis 3	UWBM 76000	Vladivostok, Russia
Oceanodroma monteiroi 1	Max Levy 2J	Ilheu de Praia, Azores, Portugal
Oceanodroma monteiroi 2	Max Levy 7J	Ilheu de Praia, Azores, Portugal
Oceanodroma monteiroi 3	Vincent Bretagnolle D52648	Ilheu de Praia, Azores, Portugal
Oceanodroma monteiroi 4	Vincent Bretagnolle D53066	Ilheu de Praia, Azores, Portugal
Oceanodroma socorroensis 1	Yuliana Bedolla TOSP140	Guadalupe Island, Mexico
Oceanodroma tethys kelsalli 1	Heraldo Norambuena OT04	Aduana Loa, Antofagasta, Chile
Oceanodroma tethys kelsalli 2	Heraldo Norambuena OT05	Aduana Loa, Antofagasta, Chile
Oceanodroma tethys tethys 1	LSUMZ 15454	Galapagos Islands, Ecuador
Oceanodroma tristrami 1	UWBM 89754	French Frigate Shoals, Northwestern Islands, Hawaii, USA
Oceanodroma tristrami 2	UWBM 91496	French Frigate Shoals, Northwestern Islands, Hawaii, USA
Pachyptila belcheri 1	KU 14900	Aukland airport, Aukland, North Island, New Zealand
Pachyptila belcheri 2	UWBM 120281	New Island, Falkland Islands
Pachyptila crassirostris flemingi 1	NMV 42149	25 miles NE of Heard Island
Pachyptila crassirostris pyramidalis 1	AMNH DOT-2581	Fourty-Fours Island, Chatham Islands, New Zealand
Pachyptila desolata banksi 1	AMNH DOT-14758	South Atlantic, between South Georgia and South Sandwich Island
Pachyptila desolata banksi 2	NMV 43153	Atlas Cove, Heard Island
Pachyptila desolata banksi 3	Richard Phillips E1	Bird Island, South Georgia
Pachyptila desolata desolata 1	Yves Cherel AP01	Kerguelen Islands, French Southern and Antarctic Lands
Pachyptila desolata desolata 2	Yves Cherel AP02	Kerguelen Islands, French Southern and Antarctic Lands
Pachyptila desolata ssp. 1	NHMO BI-35405	Bouvet Island
Pachyptila macgillivrayi 1	Yves Cherel MGPR2	St. Paul Island, French Southern and Antarctic Lands
Pachyptila macgillivrayi 2	Yves Cherel MGPR3	St. Paul Island, French Southern and Antarctic Lands
Pachyptila salvini 1	NMV 34760	Marion Island
Pachyptila salvini 2	SAM 2773	Jack Point, Coorong, South Australia
Pachyptila turtur eatoni 1	Yves Cherel FPR02	St. Paul Island, French Southern and Antarctic Lands
Pachyptila turtur eatoni 2	Yves Cherel FPR03	St. Paul Island, French Southern and Antarctic Lands
Pachyptila turtur subantarctica 1	Paul Scofield AV1208b	Antipodes Islands, New Zealand
Pachyptila turtur turtur 1	AMNH DOT-2580	Little Sister Island, Chatham Islands, New Zealand
Pachyptila turtur turtur 2	UWBM 81011	Southwest of Cape Foulwind, New Zealand
Pachyptila vittata 1	AMNH DOT-2586	Toru Inlet, Snares Island, New Zealand
Pagodroma nivea nivea 1	LSUMZ 13547	Atlantic Ocean
Pagodroma nivea nivea 2	Richard Phillips RP19	South Georgia waters
Pagodroma nivea nivea 3	UWBM 61674	South Georgia
Pagodroma nivea ssp. 1	Yves Cherel SPP04	Adélie Land, Antarctica
Pelagodroma marina dulciae 1	SAM 2962	Wedge Island, South Australia

Pelagodroma marina eadesorum 1	Jacob González-Solís 4202232	Ipassaros, Cape Verde
Pelagodroma marina eadesorum 2	Jacob González-Solís 4202233	Ipassaros, Cape Verde
Pelagodroma marina hypoleuca 1	Vincent Bretagnolle E006973	Madeira
Pelagodroma marina maoriana 1	USNM 614205	North Pacific Ocean
Pelagodroma marina marina 1	Gary Nunn WSP-5	Gough Island
Pelecanoides georgicus 1	NMV 44270	Atlas Cove, Heard Island
Pelecanoides georgicus 2	Yves Cherel SGDP24	Kerguelen Islands, French Southern and Antarctic Lands
Pelecanoides georgicus 3	Yves Cherel SGDP26	Kerguelen Islands, French Southern and Antarctic Lands
Pelecanoides magellani 1	AMNH DOT-3211	Puerto Williams, Navarino Island, Chile
Pelecanoides magellani 2	AMNH DOT-3214	Puerto Williams, Navarino Island, Chile
Pelecanoides urinatrix berard 1	UWBM 70887	South Jason Island, Falkland Islands
Pelecanoides urinatrix chathamensis 1	AMNH DOT-2583	Snares Island, New Zealand
Pelecanoides urinatrix dacunhae 1	Gary Nunn CDP-6	Gough Island
Pelecanoides urinatrix exsul 1	NMV 44247	Atlas Cove, Heard Island
Pelecanoides urinatrix exsul 2	UWBM 60517	Annenkov Island, South Georgia
Pelecanoides urinatrix exsul 3	Yves Cherel CDP10	Kerguelen Islands, French Southern and Antarctic Lands
Pelecanoides urinatrix exsul 4	Yves Cherel CDP12	Kerguelen Islands, French Southern and Antarctic Lands
Pelecanoides urinatrix urinatrix 1	AMNH DOT-11052	North Brother island, New Zealand
Phoebastria albatrus 1	UWBM 55909	Aleutian Islands, Alaska, USA
Phoebastria albatrus 2	UWBM 82972	Oregon, USA
Phoebastria immutabilis 1	USNM 643358	Hawaii, USA
Phoebastria immutabilis 2	UWBM 65680	North Pacific
Phoebastria irrorata 1	Kate Huyvaert 5017	Galapagos Islands, Ecuador
Phoebastria irrorata 2	Kate Huyvaert 703	Galapagos Islands, Ecuador
Phoebastria irrorata 3	Kate Huyvaert 704	Galapagos Islands, Ecuador
Phoebastria nigripes 1	USNM 630958	North Pacific Ocean
Phoebastria nigripes 2	UWBM 84471	Chichagof Island, Alaska, USA
Phoebetria fusca 1	Gary Nunn SA-1	Gough Island
Phoebetria fusca 2	Gary Nunn SA-3	Gough Island
Phoebetria palpebrata 1	BMNH 109516416	South Georgia
Procellaria aequinoctialis aequinoctialis 1	AMNH DOT-14762	South Atlantic, between South Georgia & South Sandwich Islands
Procellaria aequinoctialis aequinoctialis 2	Richard Phillips A7	Bird Island, South Georgia
Procellaria aequinoctialis steadi 1	UWBM 82806	North of Auckland Islands, New Zealand
Procellaria cinerea 1	Gary Nunn GYP-2	Gough Island
Procellaria cinerea 2	UWBM 82802	150 km east of Tolaga Bay, New Zealand
Procellaria conspicillata 1	MCNB 1490-T	Namibia
Procellaria conspicillata 2	MCNB 1506-T	Namibia
Procellaria conspicillata 3	MCNB 1527-T	Namibia
Procellaria westlandica 1	UWBM 82803	120 km off East Cape, New Zealand
Procellaria westlandica 2	ZMUC P770	Punakaiki, South Island, New Zealand
Pseudobulweria aterrima 1	Vincent Bretagnolle 10.36/1995.1	Reunion, France

Pseudobulweria becki 1	Vincent Bretagnolle beckiS1	Unknown
Pseudobulweria macgillivrayi 1	KU 22549	Fiji
Pseudobulweria macgillivrayi 2	Vincent Bretagnolle 10-53/BI26	Gau Island, Fiji
Pseudobulweria rostrata rostrata 1	NZP 7491	American Samoa
Pseudobulweria rostrata trouessarti 1	Vincent Bretagnolle FS79049	New Caledonia
Pterodroma alba 1	USNM 613206	North Pacific Ocean
Pterodroma alba 2	USNM 613853	Pacific Ocean
Pterodroma atrata 1	Vincent Bretagnolle 22/07/11	Henderson Island, Pitcairn Islands
Pterodroma atrata 2	Vincent Bretagnolle 31/07/11	Henderson Island, Pitcairn Islands
Pterodroma atrata 3	USNM 542141	Pitcairn, Pitcairn Islands
Pterodroma baraui 1	Vincent Bretagnolle 2599/26-2E	Mer St Gilles, Reunion, France
Pterodroma baraui 2	Vincent Bretagnolle 2604/26-2G	R. St Etienne, Reunion, France
Pterodroma brevipes brevipes 1	Vincent Bretagnolle ERO11B	Erromango, Vanuatu
Pterodroma brevipes magnificens 1	Vincent Bretagnolle M5	Vanua Lava, Vanuatu
Pterodroma cahow 1	BMNH 109516437	Bermuda
Pterodroma cervicalis 1	AM 72269	Yamba, New South Wales, Australia
Pterodroma cervicalis 2	Paul Scofield AV1304	Kermadec Islands, New Zealand
Pterodroma cookii 1	UWBM 70582	Grayland, Washington, USA
Pterodroma cookii 2	UWBM 82808	Northland, New Zealand
Pterodroma deserta 1	Vincent Bretagnolle 1005245-1	Bugio, Desertas Islands, Madeira
Pterodroma deserta 2	Vincent Bretagnolle 1009930-5	Bugio, Desertas Islands, Madeira
Pterodroma externa 1	AMNH DOT-3106	Mas Afuera, Juan Fernandez Islands, Chile
Pterodroma externa 2	USNM 613868	North Pacific Ocean
Pterodroma feae 1	Vincent Bretagnolle 5500301-1	Fogo, Cape Verde
Pterodroma feae 2	Vincent Bretagnolle 5500302-2	Fogo, Cape Verde
Pterodroma gouldi 1	UWBM 80997	North of Raoul Island, New Zealand
Pterodroma gouldi 2	UWBM 82768	Off Coromandel Peninsula, New Zealand
Pterodroma hasitata 1	Jacob González-Solís 5501064	Sant Antao, Cape Verde
Pterodroma hasitata 2	USNM 596792	Verona, Virginia, USA
Pterodroma hasitata 3	USNM 621363	St Michaels, Miles River, Maryland, USA
Pterodroma heraldica 1	USNM 562780	Pitcairn, Pitcairn Islands
Pterodroma hypoleuca 1	UMMZ 234724	Iwo Jima, Japan
Pterodroma hypoleuca 2	UWBM 55680	North Pacific Ocean
Pterodroma incerta 1	Gary Nunn ATP-1	Gough Island
Pterodroma incerta 2	Gary Nunn ATP-3	Gough Island
Pterodroma inexpectata 1	UWBM 61906	North Pacific Ocean
Pterodroma inexpectata 2	UWBM 61907	North Pacific Ocean
Pterodroma lessonii 1	KU 10639	Toukley, New South Wales, Australia
Pterodroma lessonii 2	UWBM 82791	Northland, New Zealand
Pterodroma leucoptera caledonica 1	USNM 613225	North Pacific Ocean
Pterodroma leucoptera leucoptera 1	Vincent Bretagnolle 42540	Cabbage Tree Island, Australia
Pterodroma longirostris 1	USNM 613899	North Pacific Ocean

Pterodroma longirostris 2	UWBM 55679	North Pacific Ocean
Pterodroma longirostris 3	YIO 2010-0352	Okayama, Honshu, Japan
Pterodroma macroptera 1	AMNH DOT-17854	Cheynes Beach, Albany, Western Australia
Pterodroma macroptera 2	AMNH DOT-17705	Goode Beach, Albany, Western Australia
Pterodroma madeira 1	Vincent Bretagnolle PM0106	Manga Grande, Madeira
Pterodroma madeira 2	Vincent Bretagnolle PM0274-1	Madeira
Pterodroma mollis 1	Gary Nunn SPP-1	Gough Island
Pterodroma mollis 2	Gary Nunn SPP-4	Gough Island
Pterodroma neglecta juana 1	AMNH DOT-3124	Morro Juanango, Isla Robinson Crusoe, Chile
Pterodroma neglecta neglecta 1	USNM 562779	Pitcairn, Pitcairn Islands
Pterodroma nigripennis 1	Vincent Bretagnolle PEAN12	New Caledonia
Pterodroma nigripennis 2	USNM 607570	North Pacific Ocean
Pterodroma occulta 1	Vincent Bretagnolle O3	Vanua Lava, Vanuata
Pterodroma occulta 2	Vincent Bretagnolle O4	Vanua Lava, Vanuata
Pterodroma sandwichensis 1	NZP SOSHAPE1	Kauai, Hawaii, USA
Pterodroma sandwichensis 2	NZP HAPE21298	Maui, Hawaii, USA
Pterodroma solandri 1	UWBM 55582	North Pacific Ocean
Pterodroma ultima 1	USNM 562778	Pitcairn, Pitcairn Islands
Puffinus assimilis haurakiensis 1	Jacob González-Solís 131	Hauraki Gulf, New Zealand
Puffinus assimilis haurakiensis 2	Jacob González-Solís 98	Hauraki Gulf, New Zealand
Puffinus assimilis kermadecensis 1	Paul Scofield AV20400	Raoul Island, Kermadec Islands, New Zealand
Puffinus assimilis tunneyi 1	AMNH DOT-17846	Western Australia, Australia
Puffinus auricularis 1	Yuliana Bedolla TOSH3	Socorro Island, Mexico
Puffinus auricularis 2	Yuliana Bedolla TOSH4	Socorro Island, Mexico
Puffinus bailloni bailloni 1	Vincent Bretagnolle Pufflhe14/2182	Reunion, France
Puffinus bailloni bailloni 2	Vincent Bretagnolle Pufflhe14/2259	Reunion, France
Puffinus bailloni colstoni 1	Vincent Bretagnolle ALDABRASHEAR	Aldabra, Seychelles
Puffinus bailloni dichrous 1	Alice Cibois & Jean-Claude Thibault 2019-93	South Pacific Ocean
Puffinus bailloni dichrous 2	Alice Cibois & Jean-Claude Thibault 2019-96	South Pacific Ocean
Puffinus bailloni dichrous 3	Alice Cibois & Jean-Claude Thibault 2019-97	South Pacific Ocean
Puffinus bailloni dichrous 4	Vincent Bretagnolle PlherUapouVII-98-2	Marquesas Islands, French Polynesia
Puffinus bailloni nicolae 1	Vincent Bretagnolle GE5074	Seychelles
Puffinus bailloni nicolae 2	Vincent Bretagnolle GE50901	Seychelles
Puffinus bannermani 1	Kazuto Kawakami 11275	Ogasawara Islands, Japan
Puffinus bannermani 2	Kazuto Kawakami 11491	Ogasawara Islands, Japan
Puffinus bannermani 3	YIO 2001-0128	Hahajima, Ogasawara Islands, Japan
Puffinus bannermani 4	YIO 2003-0104	Chichijima, Ogasawara islands, Japan
Puffinus baroli 1	Jeremy Austin/Spanish PAAs91	Tenerife, Canary Islands, Spain
Puffinus baroli 2	Jeremy Austin/Spanish PLBaJ2	Vila, Azores, Portugal
Puffinus baroli 3	Jeremy Austin/Spanish PLBaJ4	Vila, Azores, Portugal
Puffinus boydi 1	Jeremy Austin/Spanish PLBoJ2	Raso, Cape Verde
Puffinus boydi 2	Jeremy Austin/Spanish PLBoJX	Raso, Cape Verde

Puffinus elegans 1	Gary Nunn LS-2	Gough Island
Puffinus elegans 2	Yves Cherel LISH01	St. Paul Island, French Southern and Antarctic Lands
Puffinus gavia 1	KU 14876	Near Raglon, Ruapuke Beach, North Island, New Zealand
Puffinus gavia 2	UWBM 82796	Off Coromandel Peninsula, New Zealand
Puffinus huttoni 1	LSUMZ 23388	Marlborough, Kaikoura Ranges, New Zealand
Puffinus huttoni 2	LSUMZ 23389	Marlborough, Kaikoura Ranges, New Zealand
Puffinus lherminieri lherminieri 1	Jeremy Austin/Spanish PLLh20918	Near Oregon Inlet, USA
Puffinus lherminieri lherminieri 2	USNM 620721	North Carolina, USA
Puffinus lherminieri loyemilleri 1	USNM 607632	Tiger Rocks, Off Peninsula Valiente, Panama
Puffinus mauretanicus 1	Jeremy Austin/Spanish PMau15	Sa Conillera, Ibiza, Balearic Islands, Spain
Puffinus mauretanicus 2	Jeremy Austin/Spanish PMau18	Sa Conillera, Ibiza, Balearic Islands, Spain
Puffinus myrtae 1	Alice Cibois & Jean-Claude Thibault 2019-27	South Pacific Ocean
Puffinus myrtae 2	Alice Cibois & Jean-Claude Thibault 2019-28	South Pacific Ocean
Puffinus nativitatis 1	Jeremy Austin/Spanish PNat85	Sand Island, Johnston Atoll
Puffinus nativitatis 2	USNM 613922	North Pacific Ocean
Puffinus newelli 1	NZP NESHMI67	Kauai, Hawaii, USA
Puffinus newelli 2	NZP NESH10250	Hawaii
Puffinus opisthomelas 1	LSUMZ 16802	Ventura County, California, USA
Puffinus opisthomelas 2	LSUMZ 19402	Los Angeles County, California, USA
Puffinus puffinus 1	Jeremy Austin/Spanish PPufJ4	Heimaey island, Iceland
Puffinus puffinus 2	UWBM 80168	Bay Center, Washington, USA
Puffinus subalaris 1	LACM 114638	Off Panama, North Pacific Ocean
Puffinus yelkouan 1	Jeremy Austin/Spanish PYelJ1	Port-Cros, France
Puffinus yelkouan 2	Jeremy Austin/Spanish PYelJ2	Port-Cros, France
Spheniscus demersus 1	USNM 631252	Captive
Sula leucogaster 1	USNM 622596	Johnston Atoll
Thalassarche bulleri bulleri 1	AMNH DOT-2582	Big Solanders Island, New Zealand
Thalassarche bulleri platei 1	UWBM 80961	Chatham Islands, New Zealand
Thalassarche carteri 1	Theresa Burg 3696	Amsterdam Island, French Southern and Antarctic Lands
Thalassarche carteri 2	Theresa Burg 3734	Amsterdam Island, French Southern and Antarctic Lands
Thalassarche cauta cauta 1	AAD AI100	Albatross Island, Tasmania, Australia
Thalassarche cauta cauta 2	Theresa Burg 3668	South Africa bycatch
Thalassarche cauta steadi 1	AAD F36	Disappointment Island, Aukland Islands, New Zealand
Thalassarche cauta steadi 2	Theresa Burg 662	South Africa bycatch
Thalassarche cauta steadi 3	UWBM 82770	Between Snares and Aukland Islands, New Zealand
Thalassarche chlororhynchos 1	Gary Nunn YNA6	Gough Island
Thalassarche chlororhynchos 2	Gary Nunn YNA7	Gough Island
Thalassarche chrysostoma 1	AMNH DOT-2584	Diego Ramirez, Chile
Thalassarche chrysostoma 2	Richard Phillips A3	Bird Island, South Georgia
Thalassarche impavida 1	Theresa Burg A100001	Campbell Island bycatch, New Zealand
Thalassarche impavida 2	UWBM 80981	Campbell Island, New Zealand

Thalassarche impavida 3	UWBM 82777	100 km east of Tolaga Bay, New Zealand
Thalassarche melanophris 1	BMNH 109516419	Falkland Islands
Thalassarche melanophris 2	Richard Phillips A5	Bird Island, South Georgia
Thalassarche melanophris 3	Theresa Burg 1149	South Africa bycatch
Thalassarche melanophris 4	Theresa Burg 21	Bird Island, South Georgia
Thalassarche melanophris 5	UWBM 80962	Chatham Islands, New Zealand
Thalassarche melanophris 6	Yves Cherel BBA-SN-A13-P16	Kerguelen Islands, French Southern and Antarctic Lands
Thalassarche salvini 1	UWBM 81006	Bounty Platform, New Zealand
Thalassarche salvini 2	UWBM 81018	Chatham Islands, New Zealand
Thalassoica antarctica 1	LSUMZ 13536	Scotia Confluence, South Atlantic Ocean
Thalassoica antarctica 2	UWBM 81012	East of Balleny Islands (north of Cape Adare), New Zealand

References

Abadi, S., Azouri, D., Pupko, T., Mayrose, I. (2019) Model selection may not be a mandatory step for phylogeny reconstruction. *Nature Communications*. **10**(1): 934.

Abbott, C.L., Double, M.C. (2003a) Genetic structure, conservation genetics, and evidence of speciation by range expansion in shy and white-capped albatrosses. *Molecular Ecology*. **12**: 2953-2962.

Abbott, C.L., Double, M.C. (2003b) Phylogeography of shy and white-capped albatrosses inferred from mitochondrial DNA sequences: implications for population history and taxonomy. *Molecular Ecology*. **12**: 2747-2758.

Abbott, C.L., Double, M.C., Trueman, J.W., Robinson, A., Cockburn, A. (2005) An unusual source of apparent mitochondrial heteroplasmy: duplicate mitochondrial control regions in *Thalassarche* albatrosses. *Molecular Ecology*. **14**(11): 3605-3613.

Abbott, R.J., Barton, N.H., Good, J.M. (2016) Genomics of hybridization and its evolutionary consequences. *Molecular Ecology*. **25**: 2325-2332.

Aberer, A.J., Kobert, K., Stamatakis, A. (2014) ExaBayes: massively parallel bayesian tree inference for the whole-genome era. *Molecular Biology and Evolution*. **31**(10): 2553-2556.

Abeyrama, D.K., Dempsey, Z.W., Ryan, P.G., Burg, T.M. (2021) Cryptic speciation and population differentiation in the yellow-nosed albatross species complex. *Conservation Genetics*. **22**: 757-766.

Agapow, P.M., Bininda-Emonds, O.R., Crandall, K.A., Gittleman, J.L., Mace, G.M., Marshall, J.C., Purvis, A. (2004) The impact of species concept on biodiversity studies. *The Quarterly Review of Biology*. **79**(2): 161-179.

Akiyama, T., Nishida, C., Momose, K., Onuma, M., Takami, K., Masuda, R. (2017) Gene duplication and concerted evolution of mitochondrial DNA in crane species. *Molecular Phylogenetics and Evolution*. **106**: 158-163.

Alda, F., Ludt, W.B., Elías, D.J., McMahan, C.D., Chakrabarty, P. (2021) Comparing Ultraconserved Elements and Exons for Phylogenomic Analyses of Middle American Cichlids: When Data Agree to Disagree. *Genome Biology and Evolution*. **13**(8): evab161.

Alda, F., Tagliacollo, V.A., Bernt, M.J., Waltz, B.T., Ludt, W.B., Faircloth, B.C., Alfaro, M.E., Albert, J.S., Chakrabarty, P. (2019) Resolving Deep Nodes in an Ancient Radiation of Neotropical Fishes in the Presence of Conflicting Signals from Incomplete Lineage Sorting. *Systematic Biology*. **68**(4): 573-593.

Alexander, W., Falla, R.A., Jouanin, C., Murphy, R.C., Salmonsen, F., Voous, K.H., Watson, G.E., Bourne, W.R.P., Kuroda, N.H., Rowan, M.K., Serventy, D.L., Tickell, W.L.N., Warham, J., Winterbottom, J.M. (1965) The families and genera of petrels and their names. *Ibis*. **107**: 401-405.

Alexander, W.B. (1928) Birds of the Ocean. Putnam, New York.

Alfaro, M.E., Faircloth, B.C., Harrington, R.C., Sorenson, L., Friedman, M., Thacker, C.E., Oliveros, C.H., Černý, D., Near, T.J. (2018) Explosive diversification of marine fishes at the Cretaceous-Palaeogene boundary. *Nature Ecology & Evolution*. **2**(4): 688-696.

Allio, R., Schomaker-Bastos, A., Romiguier, J., Prosdocimi, F., Nabholz, B., Delsuc, F. (2020) MitoFinder: Efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics. *Molecular Ecology Resources*. **20**(4): 892-905.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*. **25**: 3389-3402.

Andermann, T., Fernandes, A.M., Olsson, U., Töpel, M., Pfeil, B., Oxelman, B., Aleixo, A., Faircloth, B.C., Antonelli, A. (2019) Allele Phasing Greatly Improves the Phylogenetic Utility of Ultraconserved Elements. *Systematic Biology*. **68**(1): 32-46.

Andermann, T., Torres Jiménez, M.F., Matos-Maraví, P., Batista, R., Blanco-Pastor, J.L., Gustafsson, A.L.S., Kistler, L., Liberal, I.M., Oxelman, B., Bacon, C.D., Antonelli, A. (2020) A Guide to Carrying Out a Phylogenomic Target Sequence Capture Project. *Frontiers in Genetics*. **10**: 1407.

Anderson, O., Small, C., Croxall, J., Dunn, E., Sullivan, B., Yates, O., Black, A. (2011) Global seabird bycatch in longline fisheries. *Endangered Species Research*. **14**: 91-106.

Andrews, S. (2010) FastQC: a quality control tool for high throughput sequence data. Available at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Arenas, M. (2015) Trends in substitution models of molecular evolution. *Frontiers in Genetics*. **6**: 319.

Arndt, A., Smith, M.J. (1998) Mitochondrial gene rearrangement in the sea cucumber genus *Cucumaria*. *Molecular Biology and Evolution*. **15**(8): 1009-1016.

Ashmole, N.P. (1971) "Seabird ecology and the marine environment" in Farner, D.S., King, J.R., Parkes, K.C. (eds.) Avian Biology vol. 1, Academic Press, New York.

Austin, J.J. (1996) Molecular phylogenetics of *Puffinus* shearwaters: Preliminary evidence from mitochondrial cytochrome *b* gene sequences. *Molecular Phylogenetics and Evolution*. **6**: 77-88.

Austin, J.J., Bretagnolle, V., Pasquet, E. (2004) A global molecular phylogeny of the small *Puffinus* shearwaters and implications for systematics of the Little-Audubon's shearwater complex. *The Auk*. **121**: 847-864.

Bachtrog, D., Mank, J.E., Peichel, C.L. et al. (2014) Sex determination: why so many ways of doing it? *PLoS Biology*. **12**(7): e1001899.

Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., Selker, E.U., Cresko, W.A., Johnson, E.A. (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One*. **3**(10): e3376.

Ballard, J.W., Whitlock, M.C. (2004) The incomplete natural history of mitochondria. *Molecular Ecology*. **13**(4): 729-744.

Bang, B.G. (1966) The olfactory apparatus of tubenosed birds (Procellariiformes). *Acta Anatomica*. **65**(1): 391-415.

Bankevich, A., Nurk, S., Antipov, D. et al. (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*. **19**(5): 455-477.

Bapteste, E., van Iersel, L., Janke, A., Kelchner, S., Kelk, S., McInerney, J.O., Morrison, D.A., Nakhleh, L., Steel, M., Stougie, L., Whitfield, J. (2013) Networks: expanding evolutionary thinking. *Trends in Genetics*. **29**(8): 439-441.

Barker F.K., Oyler-McCance, S., Tomback, D.F. (2015) Blood from a turnip: tissue origin of low-coverage shotgun sequencing libraries affects recovery of mitogenome sequences. *Mitochondrial DNA*. **26**(3): 384-388.

Barnosky, A.D., Matzke, N., Tomiya, S., Wogan, G.O., Swartz, B., Quental, T.B., Marshall, C., McGuire, J.L., Lindsey, E.L., Maguire, K.C., Mersey, B., Ferrer, E.A. (2011) Has the Earth's sixth mass extinction already arrived? *Nature*. **471**(7336): 51-57.

Baumgart, S.L., Sereno, P.C., Westneat, M.W. (2021) Wing Shape in Waterbirds: Morphometric Patterns Associated with Behavior, Habitat, Migration, and Phylogenetic Convergence. *Integrative Organismal Biology*. **3**(1): obab011.

Bayard De Volo, S., Reynolds, R.T., Douglas, M.R., Antolin, M.F. (2008) An improved extraction method to increase DNA yield from molted feathers. *Condor*. **110**: 762-766.

Bedolla-Guzmán, Y., Méndez-Sánchez, F., Aguirre-Muñoz, A. et al. (2019) "Recovery and current status of seabirds on the Baja California Pacific Islands, Mexico, following restoration actions" in Veitch, C.R., Clout, M.N., Martin, A.R., Russell, J.C., West, C.J. (eds.) Island invasives: scaling up to meet the challenge, Occasional Paper SSC no. 62, Gland, Switzerland: IUCN, 531-538.

Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W.J., Mattick, J.S., Haussler, D. (2004) Ultraconserved elements in the human genome. *Science*. **304**: 1321-1325.

Bellott, D.W., Skaletsky, H., Cho, T.-J. et al. (2017) Avian W and mammalian Y chromosomes convergently retained dosage-sensitive regulators. *Nature Genetics*. **49**: 387-394.

Bellott, D.W., Skaletsky, H., Pyntikova, T., Mardis, E.R., Graves, T., Kremitzki, C., Brown, L.G., Rozen, S., Warren, W.C., Wilson, R.K., and Page, D.C. (2010) Convergent evolution of chicken Z and human X chromosomes by expansion and gene acquisition. *Nature*. **466**: 612-616.

Benestan, L., Gosselin, T., Perrier, C., Sainte-Marie, B., Rochette, R., Bernatchez, L. (2015) RAD genotyping reveals fine-scale genetic structuring and provides powerful population assignment in a widely distributed marine species, the American lobster (*Homarus americanus*). *Molecular Ecology*. **24**(13): 3299-3315.

Bensch, S., Härlid, A. (2000) Mitochondrial genomic rearrangements in songbirds. *Molecular Biology and Evolution*. **17**(1): 107-113.

Berlin, S., Ellegren, H. (2006) Fast accumulation of nonsynonymous mutations on the female-specific W chromosome in birds. *Journal of Molecular Evolution*. **62**(1): 66-72.

Betancur-R, R., Naylor, G.J., Ortí, G. (2014) Conserved genes, sampling error, and phylogenomic inference. *Systematic Biology*. **63**(2): 257-262.

BirdLife International (2023) IUCN Red List for Birds. Available at: http://datazone.birdlife.org. Accessed 01/07/2023.

Blanchette, M., Green, E.D., Miller, W., Haussler, D. (2004) Reconstructing large regions of an ancestral mammalian genome in silico. *Genome Research*. **14**(12): 2412-2423.

Blom, M.P.K., Bragg, J.G., Potter, S., Moritz, C. (2017) Accounting for Uncertainty in Gene Tree Estimation: Summary-Coalescent Species Tree Inference in a Challenging Radiation of Australian Lizards. *Systematic Biology*. **66**(3): 352-366.

Bolger, A.M., Lohse, M., Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. **30**(15): 2114-2120.

Bolton, M., Smith, A.L., Gómez-Díaz, E., Friesen, V.L., Medeiros, R., Bried, J., Roscales, J.L., Furness, R.W. (2008) Monteiro's Storm-petrel *Oceanodroma monteiroi*: a new species from the Azores. *Ibis*. **150**: 717-727.

Bond, A.L., Taylor, C., Kinchin-Smith, D., Fox, D., Witcutt, E., Ryan, P.G., Loader, S.P., Weimerskirch, H. (2021) A juvenile Tristan albatross (*Diomedea dabbenena*) on land at the Crozet Islands. *Polar Biology*. **44**: 229-233.

Boore, J.L. (1999) Animal mitochondrial genomes. *Nucleic Acids Research*. **27**(8): 1767-1780.

Boore, J.L. (2000) "The duplication/random loss model for gene rearrangement exemplified by mitochondrial genomes of deuterostome animals" in Sankoff, D., Nadeau, J.H. (eds.) Comparative genomes: empirical and analytical approaches to gene order dynamics, map alignment and the evolution of gene families, Dordrecht: Kluwer Academic Publishers, 133-148.

Booth Jones, K.A., Nicoll, M.A., Raisin, C., Dawson, D.A., Hipperson, H., Horsburgh, G.J., Groombridge, J.J., Ismar, S.M., Sweet, P., Jones, C.G., Tatayah, V., Ruhomaun, K., Norris, K. (2017) Widespread gene flow between oceans in a pelagic seabird species complex. *Molecular Ecology*. **26**: 5716-5728.

Bossert, S., Danforth, B.N. (2018) On the universality of target-enrichment baits for phylogenomic research. *Methods in Ecology and Evolution*. **9**(6): 1453-1460.

Bossert, S., Murray, E.A., Blaimer, B.B., Danforth, B.N. (2017) The impact of GC bias on phylogenetic accuracy using targeted enrichment phylogenomic data. *Molecular Phylogenetics and Evolution*. **111**: 149-157.

Bourne, W.R.P. (1959) A new Little Shearwater from the Tubuai Islands: *Puffinus assimilis myrtae* subsp. nov. *Emu*. **59**: 212-214.

Branstetter, M.G., Longino, J.T., Ward, P.S., Faircloth, B.C. (2017) Enriching the ant tree of life: enhanced UCE bait set for genome-scale phylogenetics of ants and other Hymenoptera. *Methods in Ecology and Evolution*. **8**: 768-776.

Braun, E.L., Cracraft, J., Houde, P. (2019) "Resolving the avian tree of life from top to bottom: The promise and potential boundaries of the phylogenomic era" in Kraus, R.H.S. (ed.) Avian Genomics in Ecology and Evolution: From the Lab into the Wild, Springer International Publishing: Cham, Switzerland, 151-210.

Braun, E.L., Kimball, R.T. (2021), Data types and the phylogeny of Neoaves. *Birds*. **2**: 1-22.

Bravo, G.A., Antonelli, A., Bacon, C.D. et al. (2019) Embracing heterogeneity: coalescing the Tree of Life and the future of phylogenomics. *PeerJ*. **7**: e6399.

Bretagnolle, V. (1993) Adaptive significance of seabird coloration: the case of procellariiforms. *The American Naturalist*. **142**(1): 141-173.

Bretagnolle, V., Attie, C., E. Pasquet, E. (1998) Cytochrome *b* evidence for validity and phylogenetic relationships of *Pseudobulweria* and *Bulweria* (Procellariidae). *The Auk*. **115**: 188-95.

Bretagnolle, V., Attié, C., Mougeot, F. (2000) Audubon's Shearwaters *Puffinus lherminieri* on Reunion Island, Indian Ocean: behaviour, census, distribution, biometrics and breeding biology. *Ibis*. **142**(3): 399-412.

Bretagnolle, V., Flood, R.L., Gaba, S., Shirihai, H. (2022) *Fregetta lineata* (Peale, 1848) is a valid extant species endemic to New Caledonia. *Bulletin of the British Ornithologists' Club*. **142**(1): 111-130

Bretagnolle, V., Shirihai, H. (2010) A new taxon of Collared Petrel *Pterodroma brevipes* from the Banks Islands, Vanuatu. *Bulletin of the British Ornithologists' Club*. **130**: 286-301.

Bretagnolle, V., Zotier, R., Jouventin, P. (1990) Comparative Population Biology of Four Prions (Genus *Pachyptila*) from the Indian Ocean and Consequences for Their Taxonomic Status. *The Auk*. **107**: 305-316.

Bried, J., Pontier, D., Jouventin, P. (2003) Mate fidelity in monogamous birds: A re-examination of the Procellariiformes. *Animal Behaviour*. **65**(1): 235-246.

Brooke, M., Rowe, G. (1996) Behavioural and molecular evidence for specific status of light and dark morphs of the Herald Petrel *Pterodroma heraldica*. *Ibis*. **138**: 420-32.

Brooke, M.d.L. (2004) Albatross and Petrels Across the World. Oxford University Press, Oxford.

Brown, R.M., Jordan, W.C., Faulkes, C.G., Jones, C.G., Bugoni, L., Tatayah, V., Palma, R.L., Nichols, R.A. (2011) Phylogenetic relationships in *Pterodroma* petrels are obscured by recent secondary contact and hybridization. *PLoS ONE*. **6**: e20350.

Brown, W.M., George, M., Wilson, A.C. (1979) Rapid evolution of animal mitochondrial DNA. *Proceedings of the National Academy of Sciences of the United States of America*. **76**(4): 1967-1971.

Bryant, D., Moulton, V. (2004) Neighbor-net: an agglomerative method for the construction of phylogenetic networks. *Molecular Biology and Evolution*. **21**(2): 255-265.

Buermans, H.P., den Dunnen, J.T. (2014) Next generation sequencing technology: Advances and applications. *Biochimica et Biophysica Acta*. **1842**(10): 1932-1941.

Burg, T.M, Catry, P., Ryan, P.G., Phillips, R.A. (2017) Genetic population structure of black-browed and Campbell albatrosses, and implications for assigning provenance of birds killed in fisheries. *Aquatic Conservation: Marine and Freshwater Ecosystems*. **27**: 1156-1163.

Burg, T.M., Bird, H., Lait, L., Brooke, M.d.L. (2014) Colonization pathways of the northeast Atlantic by northern fulmars: a test of James Fisher's 'out of Iceland' hypothesis using museum collections. *Journal of Avian Biology*. **45**: 209-218.

Burg, T.M., Croxall, J.P. (2001) Global relationships amongst black- browed and grey-headed albatrosses: analysis of population structure using mitochondrial DNA and microsatellites. *Molecular Ecology*. **10**: 2647-2660.

Burg, T.M., Croxall, J.P. (2004) Global population structure and taxonomy of the wandering albatross species complex. *Molecular Ecology*. **13**: 2345–2355.

Burleigh, J.G., Kimball, R.T., Braun, E.L. (2015) Building the avian tree of life using a large-scale, sparse supermatrix. *Molecular Phylogenetics and Evolution*. **84**: 53-63.

Burress, E.D., Alda, F., Duarte, A., Loureiro, M., Armbruster, J.W., Chakrabarty, P. (2018) Phylogenomics of pike cichlids (Cichlidae: Crenicichla): the rapid ecological speciation of an incipient species flock. *Journal of Evolutionary Biology*. **31**(1): 14-30.

Cadahía, L., Pinsker, W., Negro, J.J., Pavlicev, M., Urios, V., Haring, E. (2009) Repeated sequence homogenization between the control and pseudo-control regions in the mitochondrial genomes of the subfamily Aquilinae. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*. **312**(3): 171-185.

Caparroz, R., Rocha, A.V., Cabanne, G.S., Tubaro, P., Aleixo, A., Lemmon, E.M., Lemmon, A.R. (2018) Mitogenomes of two neotropical bird species and the multiple independent origin of mitochondrial gene orders in Passeriformes. *Molecular Biology Reports*. **45**(3): 279-285.

Carboneras, C., Bonan, A. (2014) "Northern Storm-petrels (Hydrobatidae)" in del Hoyo, J., Elliott, A., Sargatal, J., Christie, D.A., de Juana, E. (eds.) Handbook of the Birds of the World Alive, Lynx Edicions, Barcelona.

Carlos, C.J., Voisin, J.F. (2008) Identifying giant petrels *Macronectes giganteus* and *M. halli*, in the field and in the hand. *Seabird*. **21**: 1-15.

Castoe, T.A., de Koning, A.P., Kim, H.M., Gu, W., Noonan, B.P., Naylor, G., Jiang, Z.J., Parkinson, C.L., Pollock, D.D. (2009) Evidence for an ancient adaptive episode of

convergent molecular evolution. *Proceedings of the National Academy of Sciences of the United States of Americ*a. **106**(22): 8986-8991.

Castresana, J. (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution*. **17**: 540-552.

Chambers, G.K., Moeke, C., Steel, R., Trueman, J.W.H. (2009) Phylogenetic analysis of the 24 named albatross taxa based on full mitochondrial cytochrome *b* DNA sequences. *Notornis*. **56**: 82-94.

Chan, K.O., Hutter, C.R., Wood, P.L. Jr, Grismer, L.L., Brown, R.M. (2020) Larger, unfiltered datasets are more effective at resolving phylogenetic conflict: Introns, exons, and UCEs resolve ambiguities in Golden-backed frogs (Anura: Ranidae; genus *Hylarana*). *Molecular Phylogenetics and Evolution*. **151**: 106899.

Charlesworth, B., Charlesworth, D. (2000) The degeneration of Y chromosomes. *Philosophical Transactions of the Royal Society B*. **355**: 1563-1572.

Charlton-Howard, H.S., Bond, A.L., Rivers-Auty, J., Lavers, J.L. (2023) 'Plasticosis': Characterising macro- and microplastic-associated fibrosis in seabird tissues. *Journal of Hazardous Materials*. **450**: 131090.

Chen, M.Y., Liang, D., Zhang, P. (2017) Phylogenomic resolution of the phylogeny of laurasiatherian mammals: exploring phylogenetic signals within coding and noncoding sequences. *Genome Biology and Evolution*. **9**(8): 1998-2012.

Chifman, J., Kubatko, L. (2014) Quartet inference from SNP data under the coalescent model. *Bioinformatics*. **30**(23): 3317-3324.

Cho, H.J., Eda, M., Nishida, S., Yasukochi, Y., Chong, J.R., Koike, H. (2009) Tandem duplication of mitochondrial DNA in the black-faced spoonbill, *Platalea minor*. *Genes & Genetic Systems*. **84**(4): 297-305.

Chou, J., Gupta, A., Yaduvanshi, S., Davidson, R., Nute, M., Mirarab, S., Warnow, T. (2015) A comparative study of SVDquartets and other coalescent-based species tree estimation methods. *BMC Genomics*. **16**: S2.

Chown, S.L., Gaston, K.J., Williams, P.H. (1998) Global patterns in species richness of pelagic seabirds: the Procellariiformes. *Ecography*. **21**: 342-350.

Christidis et al. (2018) The Howard and Moore Complete Checklist of the Birds of the World, version 4.1. Available at: https://www.howardandmoore.org

Cirulli, E.T., Goldstein, D.B. (2010) Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nature Reviews Genetics*. **11**(6): 415-425.

Clark, B.L., Carneiro, A.P.B., Pearmain, E.J. et al. (2023) Global assessment of marine plastic exposure risk for oceanic birds. *Nature Communications*. **14**(1): 3665.

Clarke, A., Croxall, J.P., Poncet, S., Martin, A.R., Burton, R. (2012) Important Bird Areas: South Georgia. *British Birds*. **105**: 118-144.

Clayton, D.A. (1991) Replication and transcription of vertebrate mitochondrial-DNA. *Annual Review of Cell and Developmental Biology*. **7**: 453-478.

Clements, J.F., Schulenberg, T.S., Iliff, M.J., T.A. Fredericks, T.A., Gerbracht, J.A., Lepage, D., Billerman, S.M., Sullivan, B.L., C.L. Wood, C.L. (2022) The eBird/Clements checklist of Birds of the World: v2022. Available at: https://www.birds.cornell.edu/clementschecklist/download/

Cloutier, A., Sackton, T.B., Grayson, P., Clamp, M., Baker, A.J., Edwards, S.V. (2019) Whole-Genome Analyses Resolve the Phylogeny of Flightless Birds (Palaeognathae) in the Presence of an Empirical Anomaly Zone. *Systematic Biology*. **68**(6): 937-955.

Collar, N.J. (2018) Taxonomy as tyranny. *Ibis*. **160**: 481-484.

Cooper, J., Weimerskirch, H. (2003) Exchange of the Wandering Albatross *Diomedea Exulans* Between the Prince Edward and Crozet Islands: Implications for Conservation. *African Journal of Marine Science*. **25**: 519-523.

Corl, A., Ellegren, H. (2012) The genomic signature of sexual selection in the genetic diversity of the sex chromosomes and autosomes. *Evolution*. **66**(7): 2138-2149.

Cornwell, W., Nakagawa, S. (2017) Phylogenetic comparative methods. *Current Biology*. **27**(9): R333-R336.

Coues, E. (1864a) A Critical Review of the Family Procellaridae:-Part I.; Embracing the Procellarieae, or Stormy Petrels. *Proceedings of the Academy of Natural Sciences of Philadelphia* **16**: 72-91.

Coues, E. (1864b) A Critical Review of the Family Procellaridae:-Part II.; Embracing the Puffineae. *Proceedings of the Academy of Natural Sciences of Philadelphia*. **16**: 116-144

Coues, E. (1866a) A Critical Review of the Family Procellariidae:-Part III.; Embracing the Fulmareae. *Proceedings of the Academy of Natural Sciences of Philadelphia*. **18**: 25-33.

Coues, E. (1866b) A Critical Review of the Family Procellariidae:-Part IV.; Embracing the Aestrelateae and the Prioneae. *Proceedings of the Academy of Natural Sciences of Philadelphia*. **18**: 134-172.

Coues, E. (1866c) A Critical Review of the Family Procellariidae:-Part V.; Embracing the Diomedeinae and the Halodrominae, With a General Supplement. *Proceedings of the Academy of Natural Sciences of Philadelphia*. **18**: 172-197.

Coulson, J.C. (2002) "Colonial breeding in seabirds" in Schrieber, E.A., Burger, J. (eds.) Biology of Marine Birds, CRC Press, Boca Raton, 87-113.

Cox, J.B. (1980) Some remarks on the breeding distribution and taxonomy of the prions (Procellariidae: *Pachyptila*). *Records of the South Australian Museum*. **18**: 91-121.

Croxall, J.P, Butchart, S.H.M, Lascelles, B., Stattersfield, A.J., Sullivan, B., Symes, A., Taylor, P. (2012) Seabird conservation status, threats and priority actions: a global assessment. *Bird Conservation International*. **22**: 1-34.

Cuthbert, R. (2004) Breeding biology of the Atlantic Petrel, *Pterodroma incerta*, and a population estimate of this and other burrowing petrels on Gough Island, South Atlantic Ocean. *Emu*. **104**: 221-228.

D'Erchia, A.M., Atlante, A., Gadaleta, G. et al. (2015) Tissue-specific mtDNA abundance from exome data and its correlation with mitochondrial transcription, mass and respiratory activity. *Mitochondrion*. **20**: 13-21.

Damas J., Kim J., Farré, M., Griffin, D.K., Larkin, D.M. (2018) Reconstruction of avian ancestral karyotypes reveals differences in the evolutionary history of macro- and microchromosomes. *Genome Biology*. **19**(1): 155.

Danckwerts, D.K., Humeau, L., Pinet, P., McQuaid, C.D., Le Corre, M. (2021) Extreme philopatry and genetic diversification at unprecedented scales in a seabird. *Scientific Reports*. **11**(1): 6834.

Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., Li, H. (2021) Twelve years of SAMtools and BCFtools. *Gigascience*. **10**(2): giab008.

Darriba, D., Weiß, M., Stamatakis, A. (2016) Prediction of missing sequences and branch lengths in phylogenomic data. *Bioinformatics*. **32**(9): 1331-1337.

Davey, J.W., Hohenlohe, P.A., Etter, P.D., Boone, J.Q., Catchen, J.M., Blaxter, M.L. (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*. **12**(7): 499-510.

Davies, T.J., Buckley, L.B. (2011) Phylogenetic diversity as a window into the evolutionary and biogeographic histories of present-day richness gradients for mammals. *Philosophical Transactions of the Royal Society B*. **366**(1576): 2414-2425.

Deagle, B.E., Jarman, S.N., Coissac, E., Pompanon, F., Taberlet, P. (2014) DNA metabarcoding and the cytochrome c oxidase subunit I marker: not a perfect match. *Biology Letters*. **10**(9): 20140562.

Degnan, J.H., Rosenberg, N.A. (2006) Discordance of species trees with their most likely gene trees. *PLoS Genetics*. **2**(5): e68.

Degnan, J.H., Rosenberg, N.A. (2009) Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology & Evolution*. **24**(6): 332-340.

del Hoyo, J., Carboneras, C., Jutglar, F., Collar, N., Kirwan, G.M., Garcia, E.F.J. (2020) "Wandering Albatross (*Diomedea exulans*) version 1.0." in Billerman, S.M., Keeney, B.K., Rodewald, P.G., Schulenberg, T.S. (eds.) Birds of the World, Cornell Lab of Ornithology, Ithaca, NY, USA.

del Priore, L., Pigozzi, M.I. (2017) Broad-scale recombination pattern in the primitive bird *Rhea americana* (Ratites, Palaeognathae). *PLoS One*. **12**(11): e0187549.

Delsuc, F., Brinkmann, H., Philippe, H. (2005) Phylogenomics and the reconstruction of the tree of life. *Nature Reviews Genetics*. **6**(5): 361-75.

Derkarabetian, S., Benavides, L.R., Giribet, G. (2019) Sequence capture phylogenomics of historical ethanol-preserved museum specimens: Unlocking the rest of the vault. *Molecular Ecology Resources*. **19**(6): 1531-1544.

Desjardins, P., Morais, R. (1990) Sequence and gene organization of the chicken mitochondrial genome. A novel gene order in higher vertebrates. *Journal of Molecular Biology*. **212**(4): 599-634.

Dias, M.P., Martin, R., Pearmain, E.J., Burfield, I.J., Small, C., Phillips, R.A., Yates, O., Lascelles, B., Borboroglu, P.G., Croxall, J.P. (2019) Threats to seabirds: a global assessment. *Biological Conservation*. **237**: 525-537.

Dickel, D.E., Ypsilanti, A.R., Pla, R. et al. (2018) Ultraconserved Enhancers Are Required for Normal Development. *Cell*. **172**(3): 491-499.e15.

Diniz-Filho, J.A.F, Loyola, R.D., Raia, P., Mooers, A.O., Bini, L.M. (2013) Darwinian shortfalls in biodiversity conservation. *Trends in Ecology & Evolution*. **28**: 689-695.

dos Reis, M., Donoghue, P.C., Yang, Z. (2016) Bayesian molecular clock dating of species divergences in the genomics era. *Nature Reviews Genetics*. **17**(2): 71-80.

Double, M.C., Gales, R., Reid., T., Brothers, N., Abbott, C.L. (2003) Morphometric comparison of Australian Shy and New Zealand White-capped Albatross. *Emu*. **10**: 287-294.

Drummond, A.J., Ho, S.Y., Phillips, M.J., Rambaut, A. (2006) Relaxed phylogenetics and dating with confidence. *PLoS Biology*. **4**(5): e88.

Dubois, A. (2003) The relationships between taxonomy and conservation biology in the century of extinctions. *Comptes Rendus Biologies*. **326**: 9-21.

Dufour, P., Descamps, S., Chantepie, S., Renaud, J., Guéguen, M., Schiffers, K., Thuiller, W., Lavergne, S. (2020) Reconstructing the geographic and climatic origins of long-distance bird migrations. *Journal of Biogeography*. **47**(1): 155-166.

Eberhard, J.R., Wright, T.F. (2016) Rearrangement and evolution of mitochondrial genomes in parrots. *Molecular Phylogenetics and Evolution*. **94**: 34-46.

Eberhard, J.R., Wright, T.F., Bermingham, E. (2001) Duplication and concerted evolution of the mitochondrial control region in the parrot genus Amazona. *Molecular Biology and Evolution*. **18**(7): 1330-1342.

Eda, M., Kuro-o, M., Higuchi, H., Hasegawa, H., Koike, H. (2010) Mosaic gene conversion after a tandem duplication of mtDNA sequence in Diomedeidae (albatrosses). *Genes & Genetic Systems*. **85**(2): 129-139.

Edwards, S., Bensch, S. (2009) Looking forwards or looking backwards in avian phylogeography? A comment on Zink and Barrowclough 2008. *Molecular Ecology*. **18**: 2930-2933.

Edwards, S.V. (2009) Is a new and general theory of molecular systematics emerging? *Evolution*. **63**(1): 1-19.

Edwards, S.V., Cloutier, A., Baker, A.J. (2017) Conserved Nonexonic Elements: A Novel Class of Marker for Phylogenomics. *Systematic Biology*. **66**(6): 1028-1044.

Edwards, S.V., Liu, L., Pearl, D.K. (2007) High-resolution species trees without concatenation. *Proceedings of the National Academy of Sciences of the United States of America*. **104**: 5936-5941.

Edwards, S.V., Xi, Z., Janke, A., Faircloth, B.C., McCormack, J.E., Glenn, T.C., Zhong, B., Wu, S., Lemmon, E.M., Lemmon, A.R., Leaché, A.D., Liu, L., Davis, C.C. (2016) Implementing and testing the multispecies coalescent model: A valuable paradigm for phylogenomics. *Molecular Phylogenetics and Evolution*. **94**: 447-462.

Ekblom, R., Galindo, J. (2010) Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity*. **107**(1): 1-15.

Ellegren, H. (2014) Genome sequencing and population genomics in non-model organisms. *Trends in Ecology & Evolution*. **29**(1): 51-63.

Engel, M.S., Ceríaco, L.M.P., Daniel G.M. et al. (2021) The taxonomic impediment: A shortage of taxonomists, not the lack of technical approaches. *Zoological Journal of the Linnean Society*. **193**: 381-387.

Esselstyn, J.A., Oliveros, C.H., Swanson, M.T., Faircloth, B.C. (2017) Investigating Difficult Nodes in the Placental Mammal Tree with Expanded Taxon Sampling and Thousands of Ultraconserved Elements. *Genome Biology and Evolution*. **9**(9): 2308-2321.

Faircloth, B.C. (2016) PHYLUCE is a software package for the analysis of conserved genomic loci. *Bioinformatics*. **32**(5): 786-788.

Faircloth, B.C. (2017) Identifying conserved genomic elements and designing universal bait sets to enrich them. *Methods in Ecology and Evolution*. **8**: 1103-1112.

Faircloth, B.C., Branstetter, M.G., White, N.D., Brady, S.G. (2015) Target enrichment of ultraconserved elements from arthropods provides a genomic perspective on relationships among Hymenoptera. *Molecular Ecology Resources*. **15**(3): 489-501.

Faircloth, B.C., McCormack, J.E., Crawford, N.G., Harvey, M.G., Brumfield, R.T., Glenn, T.C. (2012) Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Systematic Biology*. **61**(5): 717-726.

Faircloth, B.C., Sorenson, L., Santini, F., Alfaro, M.E. (2013) A phylogenomic perspective on the radiation of Ray-Finned fishes based upon targeted sequencing of ultraconserved elements (UCEs). *PLoS One*. **8**: e65923.

Falla, R.A. (1940) The genus *Pachyptila* Illiger. *Emu*. **40**: 218-236.

Felsenstein, J. (1978) Cases in which parsimony and compatibility methods will be positively misleading. *Systematic Zoology*. **27**: 401-410.

Feng, S., Stiller, J., Deng, Y., Armstrong, J. et al. (2020) Dense sampling of bird diversity increases power of comparative genomics. *Nature*. **587**(7833): 252-257.

Ferrer Obiol, J., Herranz, J.M., Paris, J.R., Whiting, J.R., Rozas, J., Riutort, M., González-Solís, J. (2023) Species delimitation using genomic data to resolve taxonomic uncertainties in a speciation continuum of pelagic seabirds. *Molecular Phylogenetics and Evolution*. **179**: 107671.

Ferrer Obiol, J., James, H.F., Chesser, R.T., Bretagnolle, V., González-Solís, J. (2022) Palaeoceanographic changes in the late Pliocene promoted rapid diversification in pelagic seabirds. *Journal of Biogeography*. **49**: 171-188.

Ferrer Obiol, J., James, H.F., Chesser, R.T., Bretagnolle, V., González-Solís, J., Rozas, J., Riutort, M., Welch, A.J. (2021) Integrating Sequence Capture and Restriction Site-Associated DNA Sequencing to Resolve Recent Radiations of Pelagic Seabirds. *Systematic Biology*. **70**(5): 976-996.

Fisher, J. (1952) The Fulmar. Collins, London.

Fleming, C.A. (1941) The phylogeny of the prions. *Emu*. **41**: 134-155.

Fridolfsson, A.K., Cheng, H., Copeland, N.G., Jenkins, N.A., Liu, H.C., Raudsepp, T., Woodage, T., Chowdhary, B., Halverson, J., Ellegren, H. (1998) Evolution of the avian sex chromosomes from an ancestral pair of autosomes. *Proceedings of the National Academy of Sciences of the United States of America*. **95**(14): 8147-8152.

Friesen, V.L. (2015) Speciation in seabirds: Why are there so many species…and why aren't there more? *Journal of Ornithology*. **156**: 27-39.

Friesen, V.L., Smith, A.L., Gomez-Diaz, E., Bolton, M., Furness, R.W., González-Solís, J., Monteiro, L.R. (2007) Sympatric speciation by allochrony in a seabird. *Proceedings of the National Academy of Sciences of the United States of America*. **104**: 18589-18594.

Fujita, M.K., Boore, J.L., Moritz, C. (2007) Multiple origins and rapid evolution of duplicated mitochondrial genes in parthenogenetic geckos (*Heteronotia binoei*; Squamata, Gekkonidae). *Molecular Biology and Evolution*. **24**(12): 2775-2786.

Funk, D.J., Omland, K.E. (2003) Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annual Review of Ecology, Evolution, and Systematics*. **34**: 397-423.

Gadagkar, S.R., Rosenberg, M.S., Kumar, S. (2005) Inferring species phylogenies from multiple genes: concatenated sequence tree versus consensus gene tree. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*. **304**(1): 64-74.

Galeotti, P., Rubolini, D., Dunn, P.O., Fasola, M. (2003) Colour polymorphism in birds: causes and functions. *Journal of Evolutionary Biology*. **16**: 635-646.

Gangloff, B., Shirihai, H., Watling, D., Cruaud, C., Couloux, A., Tillier, A., Pasquet, E., Bretagnolle, V. (2012) The complete phylogeny of *Pseudobulweria*, the most endangered seabird genus: systematics, species status and conservation implications. *Conservation Genetics*. **13:** 39-52.

Gangloff, B., Zino, F., Shirihai, H., Gonzalez-Solis, J., Couloux, A., Pasquet, E., Bretagnolle, V. (2013) The evolution of northeast Atlantic gadfly petrels using statistical phylogeography. *Molecular Ecology*. **22**: 495-507.

Garcia-Moreno, J., Mindell, D.P. (2000) Rooting a phylogeny with homologous genes on opposite sex chromosomes (gametologs): a case study using avian CHD. *Molecular Biology and Evolution*. **17**(12): 1826-1832.

Garnett, S.T., Christidis, L. (2017) Taxonomy anarchy hampers conservation. *Nature*. **546**(7656): 25-27.

Gatesy, J., Springer, M.S. (2014) Phylogenetic analysis at deep timescales: unreliable gene trees, bypassed hidden support, and the coalescence/concatalescence conundrum. *Molecular Phylogenetics and Evolution*. **80**: 231-266.

Gibb, G.C., England, R., Hartig, G., McLenachan, P.A., Smith, B.L.T., McComish, B.J., Cooper, A., Penny, D. (2015) New Zealand passerines help clarify the diversification of major songbird lineages during the Oligocene. *Genome Biology and Evolution*. **7**(11): 2983-2995.

Gibb, G.C., Kardailsky, O., Kimball, R.T., Braun, E.L., Penny, D. (2007) Mitochondrial genomes and avian phylogeny: complex characters and resolvability without explosive radiations. *Molecular Biology and Evolution*. **24**: 269-280.

Gibb, G.C., Kennedy, M., Penny, D. (2013) Beyond phylogeny: pelecaniform and ciconiiform birds, and long-term niche stability. *Molecular Phylogenetics and Evolution*. **68**(2): 229-238.

Gilbert, P.S., Chang, J., Pan, C., Sobel, E.M., Sinsheimer, J.S., Faircloth, B.C., Alfaro, M.E. (2015) Genome-wide ultraconserved elements exhibit higher phylogenetic informativeness than traditional gene markers in percomorph fishes. *Molecular Phylogenetics and Evolution*. **92**: 140-146.

Gilbert, P.S., Wu, J., Simon, M.W., Sinsheimer, J.S., Alfaro, M.E. (2018) Filtering nucleotide sites by phylogenetic signal to noise ratio increases confidence in the Neoaves phylogeny generated from ultraconserved elements. *Molecular Phylogenetics and Evolution*. **126**: 116-128.

Gill, F., Donsker, D., Rasmussen, P. (eds.) (2023) IOC World Bird List (v13.1). Available at: https://www.worldbirdnames.org/new/

Glenn, T.C., Faircloth, B.C. (2016) Capturing Darwin's dream. *Molecular Ecology Resources*. **16**(5): 1051-1058.

Godman, F.D.C. (1907-1910) A Monograph of the Petrels (Order Tubinares). Witherby, London.

Goodwin, S., McPherson, J., McCombie, W. (2016) Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics*. **17**: 333–351.

Greenwood, P.J. (1980) Mating systems, philopatry and dispersal in birds and mammals. *Animal Behaviour*. **28**: 1140-1162.

Grosser, S., Robertson, F., Shepherd, L.D., Tennyson, A.J.D., Miskelly, C.M., Taylor, G.A., Robertson, B.C. (2021) Phylogenetic affinities of the Whenua Hou Diving Petrel: implications for conservation. *Emu*. **121**: 102-112.

Hackett, S.J., Kimball, R.T., Reddy, S. et al. (2008) A phylogenetic study of birds reveals their evolutionary history. *Science.* **320**: 1763-1768.

Hahn, M.W., Nakhleh, L. (2016) Irrational exuberance for resolved species trees. *Evolution*. **70**(1): 7-17.

Hallström, B.M., Janke, A. (2010) Mammalian evolution may not be strictly bifurcating. *Molecular Biology and Evolution*. **27**(12): 2804-2816.

Handbook of the Birds of the World and BirdLife International (2022) Handbook of the Birds of the World and BirdLife International digital checklist of the birds of the world Version 7. Available at: http://datazone.birdlife.org/species/taxonomy

Hanna, Z.R., Henderson, J.B., Sellas, A.B., Fuchs, J., Bowie, R.C.K., Dumbacher, J.P. (2017) Complete mitochondrial genome sequences of the northern spotted owl (*Strix occidentalis caurina*) and the barred owl (*Strix varia*; Aves: Strigiformes: Strigidae) confirm the presence of a duplicated control region. *PeerJ*. **5**: e3901.

Harmston, N., Baresic, A., Lenhard, B. (2013) The mystery of extreme non-coding conservation. *Philosophical Transactions of the Royal Society B*. **368**(1632): 20130021.

Harper, P.C. (1980) The field identification and distribution of the prions (genus *Pachyptila*), with particular reference to the identification of stormcast material. *Notornis*. **27**: 235-286.

Harrison, C.S. (1990) Seabirds of Hawaii: Natural History and Conservation. Cornell University Press.

Harrison, P., Perrow, M., Larsson, H. (2021) Seabirds: The New Identification Guide. Lynx Edicions, Barcelona, Spain.

Harrison, P., Sallaberry, M., Gaskin, C.P., Baird, K.A., Jaramillo, A., Metz, S.M., Pearman, M., O'keeffe, M., Dowdall, J., Enright, S., Fahy, K., Gilligan, J., Lillie, J. (2013) A new storm-petrel species from Chile. *The Auk*. **130**: 180-191.

Harvey, M.G., Bravo, G.A., Claramunt, S. et al. (2020) The evolution of a tropical biodiversity hotspot. *Science*. **370**(6522): 1343-1348.

Harvey, M.G., Smith, B.T., Glenn, T.C., Faircloth, B.C., Brumfield, R.T. (2016) Sequence capture versus restriction site associated DNA sequencing for shallow systematics. *Systematic Biology*. **65**(5): 910-924.

Heidrich, P., Amengual, J., Wink, M. (1998) Phylogenetic relationships in Mediterranean and North Atlantic shearwaters (Aves: Procellariidae) based on nucleotide sequences of mtDNA. *Biochemical Systematics and Ecology*. **26:** 145-170.

Heled, J., Drummond, A.J. (2010) Bayesian inference of species trees from multilocus data. *Molecular Biology and Evolution*. **27**(3): 570-80.

Herman, R.W., Winger, B.M., Dittmann, D.L, Harvey, M.G. (2022) Fine-scale population genetic structure and barriers to gene flow in a widespread seabird (*Ardenna pacifica*). *Biological Journal of the Linnean Society*. **137**: 125-136.

Hinch, A.G., Altemose, N., Noor, N., Donnelly, P., Myers, S.R. (2014) Recombination in the human Pseudoautosomal region PAR1. *PLoS Genetics*. **10**(7): e1004503.

Hindwood, K.A. (1945) The Fleshy-footed Shearwater (*Puffinus carneipes*). *Emu*. **44**: 241-248.

Hobolth, A., Dutheil, J.Y., Hawks, J., Schierup, M.H., Mailund, T. (2011) Incomplete lineage sorting patterns among human, chimpanzee, and orangutan suggest recent orangutan speciation and widespread selection. *Genome Research*. **21**: 349-356.

Hodcroft, E. (2021) TreeCollapserCL4. Available at: http://emmahodcroft.com/TreeCollapseCL.html

Hodges, E., Xuan, Z., Balija, V., Kramer, M., Molla, M.N., Smith, S.W., Middle, C.M., Rodesch, M.J., Albert, T.J., Hannon, G.J., McCombie, W.R. (2007) Genome-wide in situ exon capture for selective resequencing. *Nature Genetics*. **39**(12): 1522-1527.

Hortal, J., de Bello, F., Diniz-Filho, J.A.F., Lewinsohn, T.M., Lobo, J.M., Ladle, R.J. (2015) Seven shortfalls that beset large-scale knowledge of biodiversity. *Annual Review of Ecology, Evolution, and Systematics*. **46**: 523-549.

Hosner, P.A., Faircloth, B.C., Glenn, T.C., Braun, E.L., Kimball, R.T. (2016) Avoiding missing data biases in phylogenomic inference: an empirical study in the landfowl (Aves: Galliformes). *Molecular Biology and Evolution*. **33**: 1110-1125.

Hotaling, S., Kelley, J.L., Frandsen, P.B. (2021) Toward a genome sequence for every animal: Where are we now? *Proceedings of the National Academy of Sciences of the United States of America*. **118**(52): e2109019118.

Hruska, J.P., Holmes, J., Oliveros, C., Shakya, S., Lavretsky, P., McCracken, K.G., Sheldon, F.H., Moyle, R.G. (2023) Ultraconserved elements resolve the phylogeny and corroborate patterns of molecular rate variation in herons (Aves: Ardeidae). *Ornithology*. **140**: ukac005.

Huson, D.H., Bryant, D. (2006) Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*. **23**(2): 254-267.

Hutter, C.R., Duellman, W. (2023) Filtration of Gene Trees From 9,000 Exons, Introns, and UCEs Disentangles Conflicting Phylogenomic Relationships in Tree Frogs (Hylidae). *Genome Biology and Evolution*. **15**(5): evad070.

Imber, M.J. (1973) The food of grey-faced petrels (*Pterodroma macroptera gouldi* (Hutton)), with special reference to diurnal vertical migration of their prey. *Journal of Animal Ecology*. **42**(3): 645-662.

Iqbal, Z., Caccamo, M., Turner, I., Flicek, P., McVean, G. (2012) De novo assembly and genotyping of variants using colored de Bruijn graphs. *Nature Genetics*. **44**: 226-232.

Irwin, D.E. (2018) Sex chromosomes and speciation in birds and other ZW systems. *Molecular Ecology*. **27**(19): 3831-3851.

Itoh, Y., Mizuno, S. (2002) Molecular and cytological characterization of SspI-family repetitive sequence on the chicken W chromosome. *Chromosome Research*. **10**(6): 499-511.

Jarvis, E.D., Mirarab, S., Aberer, A.J. et al. (2014) Whole-genome analysis resolve early branches in the tree of life of modern birds. *Science*. **346**: 1320-1331.

Jeffroy, O., Brinkmann, H., Delsuc, F., Philippe, H. (2006) Phylogenomics: the beginning of incongruence? *Trends in Genetics*. **22**(4): 225-231.

Jesus, J., Menezes, D., Gomes, S., Oliveira, P., Nogales, M., Brehm, M. (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 Conservation International*. **19**: 199-214.

Jiang, Z.J., Castoe, T.A., Austin, C.C., Burbrink, F.T., Herron, M.D., McGuire, J.A., Parkinson, C.L., Pollock, D.D. (2007) Comparative mitochondrial genomics of snakes: extraordinary substitution rate dynamics and functionality of the duplicate control region. *BMC Ecology and Evolution*. **7**: 123.

Jones, M.R., Good, J.M. (2016) Targeted capture in evolutionary and ecological genomics. *Molecular Ecology*. **25**(1): 185-202.

Jouventin, P., Mougin, J.L., Stahl, J.C., Weimerskirch, H. (1985) Comparative biology of the burrowing petrels of the Crozet Islands. *Notornis*. **32**: 157-220.

Jouventin, P., Viot, C.R. (1985) Morphological and genetic variability of Snow Petrels *Pagodroma nivea*. *Ibis* **127**: 430-441

Joy, J.B., Liang, R.H., McCloskey, R.M., Nguyen, T., Poon, A.F. (2016) Ancestral Reconstruction. *PLoS Computational Biology*. **12**(7): e1004763.

Kahn, N.W., Quinn, T.W. (1999) Male-driven evolution among Eoaves? A test of the replicative division hypothesis in a heterogametic female (ZW) system. *Journal of Molecular Evolution*. **49**(6): 750-759.

Karin, B.R., Gamble, T., Jackman, T.R. (2020) Optimizing Phylogenomics with Rapidly Evolving Long Exons: Comparison with Anchored Hybrid Enrichment and Ultraconserved Elements. *Molecular Biology and Evolution*. **37**(3): 904-922.

Katoh, K., Standley, D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*. **30**(4): 772-780.

Katzman, S., Kern, A.D., Bejerano, G., Fewell, G., Fulton, L., Wilson, R.K., Salama, S.R., Haussler, D. (2007) Human genome ultraconserved elements are ultraselected. *Science*. **317**(5840): 915.

Kawakami, K., Eda, M., Izumi, H., Horikoshi, K., Suzuki, H. (2018) Phylogenetic position of endangered *Puffinus lherminieri bannermani*. *Ornithology Science*. **17**: 11-18.

Kerr, K.C., Dove, C.J. (2013) Delimiting shades of gray: phylogeography of the Northern Fulmar, *Fulmarus glacialis*. *Ecology and Evolution*. **3**(7): 1915-1930.

Kimball, R.T., Guido, M., Hosner, P.A., Braun, E.L. (2021) When good mitochondria go bad: Cyto-nuclear discordance in landfowl (Aves: Galliformes). *Gene*. **801**: 145841.

Kimball, R.T., Wang, N., Heimer-McGinn, V., Ferguson, C., Braun, E.L. (2013) Identifying localized biases in large datasets: a case study using the avian tree of life. *Molecular Phylogenetics and Evolution*. **69**: 1021-32.

King, N., Rokas, A. (2017) Embracing Uncertainty in Reconstructing Early Animal Evolution. *Current Biology*. **27**(19): R1081-R1088.

Kirwan, G.M., Carboneras, C., Jutglar, F. (2020) "Heinroth's Shearwater (*Puffinus heinrothi*), version 1.0" in Billerman, S.M., Keeney, B.K,. Rodewald, P.G., Schulenberg, T.S. (eds.) Birds of the World, Cornell Lab of Ornithology, Ithaca, NY, USA.

Knowles, L.L. (2009) Estimating species trees: methods of phylogenetic analysis when there is incongruence across genes. *Systematic Biology*. **58**(5): 463-467.

Knowles, L.L., Huang, H., Sukumaran, J., Smith, S.A. (2018) A matter of phylogenetic scale: distinguishing incomplete lineage sorting from lateral gene transfer as the cause of gene tree discord in recent versus deep diversification histories. *American Journal of Botany*. **105**: 376-384.

Kozlov, A.M., Darriba, D., Flouri, T., Morel, B., Stamatakis, A. (2019) RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics*. **35**(21): 4453-4455.

Ksepka, D., Clarke, J. (2015) Phylogenetically vetted and stratigraphically constrained fossil calibrations within Aves. *Palaeontologia Electronica*. **18.1.3FC**: 1-25.

Ksepka, D.T., Stidham, T.A., Williamson, T.E. (2017) Early Paleocene landbird supports rapid phylogenetic and morphological diversification of crown birds after the K-Pg mass extinction. *Proceedings of the National Academy of Sciences of the United States of America*. **114**(30): 8047-8052.

Kubatko, L.S., Degnan, J.H. (2007) Inconsistency of phylogenetic estimates from concatenated data under coalescence. *Systematic Biology*. **56**(1): 17-24.

Kuhl, H., Frankl-Vilches, C., Bakker, A., Mayr, G., Nikolaus, G., Boerno, S.T., Klages, S., Timmermann, B., Gahr, M. (2021) An Unbiased Molecular Approach Using 3'-UTRs Resolves the Avian Family-Level Tree of Life. *Molecular Biology and Evolution*. **38**(1): 108-127.

Kumazawa, Y., Ota, H., Nishida, M., Ozawa, T. (1996) Gene rearrangements in snake mitochondrial genomes: highly concerted evolution of control-region-like sequences duplicated and inserted into a tRNA gene cluster. *Molecular Biology and Evolution*. **13**(9): 1242-1254.

Kumazawa, Y., Ota, H., Nishida, M., Ozawa, T. (1998) The complete nucleotide sequence of a snake (*Dinodon semicarinatus*) mitochondrial genome with two identical control regions. *Genetics*. **150**(1): 313-329.

Kuroda, N. (1954) On the classification and phylogeny of the order Tubinares, particularly the shearwaters (*Puffinus*), with special consideration on their osteology and habit differentiation (Aves). Tokyo, Japan.

Kuroda, N. (1967) Morpho-anatomical analysis of parallel evolution between diving petrel and ancient auk, with comparative osteological data of other species. *Journal of the Yamashina Institute for Ornithology*. **5**: 111-137.

Lanfear, R., Calcott, B., Ho, S.Y., Guindon, S. (2012) Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution*. **29**(6): 1695-1701.

Lanfear, R., Calcott, B., Kainer, D., Mayer, C., Stamatakis, A. (2014) Selecting optimal partitioning schemes for phylogenomic datasets. *BMC Ecology and Evolution*. **14**: 82.

Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T., Calcott, B. (2017) PartitionFinder 2: New Methods for Selecting Partitioned Models of Evolution for Molecular and Morphological Phylogenetic Analyses. *Molecular Biology and Evolution*. **34**(3): 772-773.

Lavrov, D.V., Pett, W. (2016) Animal Mitochondrial DNA as We Do Not Know It: mt-Genome Organization and Evolution in Nonbilaterian Lineages. *Genome Biology and Evolution*. **8**(9): 2896-2913.

Lawrence, H.A., Taylor, G.A., Millar, C.D., Lambert, D.M. (2008) High mitochondrial and nuclear genetic diversity in one of the world's most endangered seabirds, the Chatham Island Taiko (*Pterodroma magentae*). *Conservation Genetics*. **9**: 1293-1301.

Leaché, A.D., Chavez, A.S., Jones, L.N., Grummer, J.A., Gottscho, A.D., Linkem, C.W. (2015) Phylogenomics of phrynosomatid lizards: conflicting signals from sequence capture versus restriction site associated DNA sequencing. *Genome Biology and Evolution*. **7**(3): 706-719.

Lee, J.S., Miya, M., Lee, Y.S., Kim, C.G., Park, E.H., Aoki, Y., Nishida, M. (2001) The complete DNA sequence of the mitochondrial genome of the self-fertilizing fish *Rivulus marmoratus* (Cyprinodontiformes, Rivulidae) and the first description of duplication of a control region in fish. *Gene*. **280**: 1-7.

Legendre, L.J., Guénard, G., Botha-Brink, J., Cubo, J. (2016) Palaeohistological Evidence for Ancestral High Metabolic Rate in Archosaurs. *Systematic Biology*. **65**(6): 989-996.

Lemmon, A.R., Emme, S.A., Lemmon, E.M. (2012) Anchored hybrid enrichment for massively high-throughput phylogenomics. *Systematic Biology*. **61**(5): 727-744.

Lewin, H.A., Robinson, G.E., Kress, W.J. et al. (2018) Earth BioGenome Project: Sequencing life for the future of life. *Proceedings of the National Academy of Sciences of the United States of America*. **115**(17): 4325-4333.

Li, H. (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*. 1303.3997v2 [q-bio.GN].

Liò, P., Goldman, N. (1998) Models of molecular evolution and phylogeny. *Genome Research*. **8**: 1233-1244.

Liu, L., Xi, Z., Wu, S., Davis, C.C., Edwards, S.V. (2015) Estimating phylogenetic trees from genome-scale data. *Annals of the New York Academy of Sciences*. **1360**: 36-53.

Lombal, A.J., Wenner, T.J., Lavers, J.L., Austin, J.J., Woehler, E.J., Hutton, I., Burridge, C.P. (2018) Genetic divergence between colonies of Flesh-footed Shearwater *Ardenna carneipes* exhibiting different foraging strategies. *Conservation Genetics*. **19**(1): 27-41.

Lomolino, M.V. (2004) "Conservation biogeography" in Lomolino, M.V., Heaney, L.R. (eds.) Frontiers of Biogeography: New Directions in the Geography of Nature, Sunderland, Massachusetts: Sinauer Associates, 293-296.

Loomis, L. M. (1918) A review of the albatrosses, petrels and diving petrels. *Proceedings of the California Academy of Sciences*. **4**: 1-187.

Lounsberry, Z.T., Brown, S.K., Collins, P.W., Henry, R.W., Newsome, S.D., Sacks, B.N. (2015) Next-generation sequencing workflow for assembly of nonmodel mitogenomes exemplified with North Pacific albatrosses (*Phoebastria* spp.). *Molecular Ecology Resources*. **15**(4): 893-902.

Mace, G.M. (2004) The role of taxonomy in species conservation. *Philosophical Transactions of the Royal Society B*. **359**(1444): 711-719.

Macey, J.R., Larson, A., Ananjeva, N.B., Fang, Z., Papenfuss, T.J. (1997) Two novel gene orders and the role of light-strand replication in rearrangement of the vertebrate mitochondrial genome. *Molecular Biology and Evolution*. **14**(1): 91-104.

Mackiewicz, P., Urantówka, A.D., Kroczak, A., Mackiewicz, D. (2019) Resolving phylogenetic relationships within Passeriformes based on mitochondrial genes and inferring the evolution of their mitogenomes in terms of duplications. *Genome Biology and Evolution*. **11**: 2824-2849.

Maddison, W.P. (1997) Gene trees in species trees. *Systematic Biology*. **46**(3): 523-536.

Maddison, W.P., Knowles, L.L. (2006) Inferring phylogeny despite incomplete lineage sorting. *Systematic Biology*. **55**(1): 21-30.

Maddison, W.P., Maddison, D.R. (2019) Mesquite: A Modular System for Evolutionary Analysis. Version 3.61. Available at: http://www.mesquiteproject.org

Malinsky, M., Svardal, H., Tyers, A.M., Miska, E.A., Genner, M.J., Turner, G.F., Durbin, R. (2018) Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. *Nature Ecology & Evolution*. **2**(12): 1940-1955.

Mallet, J., Besansky, N., Hahn, M.W. (2016) How reticulated are species? *Bioessays*. **38**(2): 140-149.

Mamanova, L., Coffey, A.J., Scott, C.E., Kozarewa, I., Turner, E.H., Kumar, A., Howard, E., Shendure, J., Turner, D.J. (2010) Target-enrichment strategies for next-generation sequencing. *Nature Methods*. **7**(2): 111-118.

Martínez-Gómez, J.E., Matías-Ferrer, N., Sehgal, R.N.M., Escalante, P. (2015) Phylogenetic placement of the critically endangered Townsend's Shearwater (*Puffinus auricularis auricularis*): evidence for its conspecifc status with Newell's Shearwater (*Puffunus a. newelli*) and a mismatch between genetic and phenotypic differentiation. *Journal of Ornithology*. **156**: 1025-1034.

Masello, J.F., Quillfeldt, P., Sandoval-Castellanos, E. et al. (2019) Additive Traits Lead to Feeding Advantage and Reproductive Isolation, Promoting Homoploid Hybrid Speciation. *Molecular Biology and Evolution*. **36**(8): 1671-1685.

Masello, J.F., Ryan, P.G., Shepherd, L.D. et al. (2022) Independent evolution of intermediate bill widths in a seabird clade. *Molecular Genetics and Genomics*. **297**(1): 183-198.

Mathews, G.M. (1912) Birds of Australia, vol. 2. Witherby, London.

Mathews, G.M. (1934) A check-list of the order Procellariiformes. *Novitates Zoologicae*. **39**: 151-206.

Mathews, G.M. (1948) Systematic notes on petrels. *Bulletin of the British Ornithologists' Club*. **68**: 155-170.

Mathews, G.M., Hallstrom, E.J.L. (1943) Notes on the Order Procellariiformes. Verity Hewitt Bookshop, Canberra.

Mayr, G. (2011) The phylogeny of charadriiform birds (shorebirds and allies) - reassessing the conflict between morphology and molecules. *Zoological Journal of the Linnean Society*. **161**(4): 916-934.

Mayr, G., Goedert, J.L. (2017) Oligocene and Miocene albatross fossils from Washington State (USA) and the evolutionary history of North Pacific Diomedeidae. *The Auk*. **134**: 659-671.

Mayr, G., Pavia, M. (2014) On the true affinities of *Chenornis graculoides* Portis, 1884 and *Anas lignitifila* Portis, 1884 – an albatross and an unusual duck from the Miocene of Italy. *Journal of Vertebrate Paleontology*. **34**: 914-923.

Mayr, G., Tennyson, A.J.D. (2020) A small, narrow-beaked albatross from the Pliocene of New Zealand demonstrates a higher past diversity in the feeding ecology of the Diomedeidae. *Ibis*. **162**(3): 723-734.

McCormack, J.E., Faircloth, B.C., Crawford, N.G., Gowaty, P.A., Brumfield, R.T., Glenn, T.C. (2012) Ultraconserved elements are novel phylogenomic markers that resolve placental mammal phylogeny when combined with species-tree analysis. *Genome Research*. **22**(4): 746-754.

McCormack, J.E., Harvey, M.G., Faircloth, B.C., Crawford, N.G., Glenn, T.C., Brumfield, R.T. (2013a) A phylogeny of birds based on over 1,500 loci collected by target enrichment and high-throughput sequencing. *PLoS One*. **8**(1): e54848.

McCormack, J.E., Hird, S.M., Zellmer, A.J., Carstens, B.C., Brumfield, R.T. (2013b) Applications of next-generation sequencing to phylogeography and phylogenetics. *Molecular Phylogenetics and Evolution*. **66**(2): 526-38.

McCormack, J.E., Tsai, W.L., Faircloth, B.C. (2016) Sequence capture of ultraconserved elements from bird museum specimens. *Molecular Ecology Resources*. **16**(5): 1189-1203.

Mclean, B.S., Bell, K.C., Allen, J.M., Helgen, K.M., Cook, J.A. (2019) Impacts of inference method and data set filtering on phylogenomic resolution in a rapid radiation of ground squirrels (Xerinae: Marmotini). *Systematic Biology*. **68**: 298-316.

Meiklejohn, K.A., Danielson, M.J., Faircloth, B.C., Glenn, T.C., Braun, E.L., Kimball, R.T. (2014) Incongruence among different mitochondrial regions: a case study using complete mitogenomes. *Molecular Phylogenetics and Evolution*. **78**: 314-323.

Meiklejohn, K.A., Faircloth, B.C., Glenn, T.C., Kimball, R.T., Braun, E.L. (2016) Analysis of a Rapid Evolutionary Radiation Using Ultraconserved Elements: Evidence for a Bias in Some Multispecies Coalescent Methods. *Systematic Biology*. **65**(4): 612-627.

Milot, E., Weimerskirch, H., Bernatchez, L. (2008) The seabird paradox: dispersal, genetic structure and population dynamics in a highly mobile, but philopatric albatross species. *Molecular Ecology*. **17**: 1658-1673.

Mindell, D.P., Sorenson, M.D., Dimcheff, D.E. (1998) Multiple independent origins of mitochondrial gene order in birds. *Proceedings of the National Academy of Sciences of the United States of America*. **95**(18): 10693-10697.

Minh, B.Q., Hahn, M.W., Lanfear, R. (2020a) New methods to calculate concordance factors for phylogenomic datasets. *Molecular Biology and Evolution*. **37**: 2727-2733.

Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., Lanfear, R. (2020b) IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution*. **37**: 1530-1534.

Mirarab, S., Bayzid, M.S., Warnow, T. (2016) Evaluating summary methods for multilocus species tree estimation in the presence of incomplete lineage sorting. *Systematic Biology*. **65**: 366-380.

Mo, Y.K., Lanfear, R., Hahn, M.W., Minh, B.Q. (2023) Updated site concordance factors minimize effects of homoplasy and taxon sampling. *Bioinformatics*. **39**(1): btac741.

Molloy, E.K., Warnow, T. (2018) To include or not to include: the impact of gene filtering on species tree estimation methods. *Systematic Biology*. **67**: 285-303.

Mongiardino Koch, N. (2021) Phylogenomic Subsampling and the Search for Phylogenetically Reliable Loci. *Molecular Biology and Evolution*. **38**(9): 4025-4038.

Montevecchi, W.A. (2023) "Invasive species threats to seabirds" in Young, L., VanderWerf, E. (eds.) Conservation of Marine Birds, Academic Press, 57-95.

Moodley, Y., Masello, J.F., Cole, T.L. et al (2015) Evolutionary factors affecting the cross-species utility of newly developed microsatellite markers in seabirds. *Molecular Ecology Resources*. **15**: 1046-1058.

Moritz, C. (1994) Defining 'Evolutionarily Significant Units' for conservation. *Trends in Ecology & Evolution*. **9**(10): 373-375.

Moritz, C., Brown, W.M. (1986) Tandem duplications of D-loop and ribosomal RNA sequences in lizard mitochondrial DNA. *Science*. **233**: 1425-1427.

Moritz, C., Brown, W.M. (1987) Tandem duplications in animal mitochondrial DNAs: variation in incidence and gene content among lizards. *Proceedings of the National Academy of Sciences of the United States of America*. **84**(20): 7183-7187.

Morris-Pocock, J.A., Taylor, S.A., Birt, T.P., Friesen, V.L. (2010) Concerted evolution of duplicated mitochondrial control regions in three related seabird species. *BMC Ecology and Evolution*. **10**: 14.

Mueller, R.L., Boore, J.L. (2005) Molecular mechanisms of extensive mitochondrial gene rearrangement in plethodontid salamanders. *Molecular Biology and Evolution*. **22**(10): 2104-2112.

Mukherjee, S., Seshadri, R., Varghese, N.J. et al. (2017) 1,003 reference genomes of bacterial and archaeal isolates expand coverage of the tree of life. *Nature Biotechnology*. **35**(7): 676-683.

Munro, K.J., Burg, T.M. (2017) A review of historical and contemporary processes affecting population genetic structure of Southern Ocean seabirds. *Emu*. **117**: 4-18.

Murphy, R.C. (1936) Oceanic birds of South America. The American Museum of Natural History, New York.

Nam, K., Ellegren, H. (2008) The chicken (*Gallus gallus*) Z chromosome contains at least three nonlinear evolutionary strata. *Genetics*. **180**(2): 1131-1136.

Navarro, J., Cardador, L., Brown, R., Phillips, R.A. (2015) Spatial distribution and ecological niches of non-breeding planktivorous petrels. *Scientific Reports*. **5**: 12164.

Navarro, J., Votier, S.C., Aguzzi, J., Chiesa, J.J., Forero, M.G., Phillips, R.A. (2013) Ecological segregation in space, time and trophic niche of sympatric planktivorous petrels. *PLoS One*. **8**(4): e62897.

Neves, V.C., Carroll, G., W.C. Schäfer, W.C., Hereward, H.F.R., Quillfeldt, P. (2023) Consistent foraging habitat use by a vulnerable breeding seabird highlights potential areas for protection in the mid-Atlantic Ocean. *Marine Ecology Progress Series*. **716**: 107-121.

Nevitt, G.A. (2000) Olfactory foraging by Antarctic procellariiform seabirds: life at high Reynolds numbers. *The Biological Bulletin*. **198**(2): 245-253.

Nevitt, G.A. (2008) Sensory ecology on the high seas: the odor world of the procellariiform seabirds. *The Journal of Experimental Biology*. **211**: 1706-1713.

Nichols, R. (2001) Gene trees and species trees are not the same. *Trends in Ecology & Evolution*. **16**(7): 358-364.

Nunn, G.B., Cooper, J., Jouventin, P., Robertson, C.J.R., Robertson, G.G. (1996) Evolutionary relationships among extant albatrosses (Procellariiformes, Diomedeidae) established from complete cytochrome-b gene sequences. *The Auk*. **113**: 784-801.

Nunn, G.B., Stanley, S.E. (1998) Body size effects and rates of cytochrome *b* evolution in tube-nosed seabirds. *Molecular Biology and Evolution*. **15**: 1360-1371.

O'Connor, R.E., Romanov, M.N., Kiazim, L.G., Barrett, P.M., Farré, M., Damas, J., Ferguson-Smith, M., Valenzuela, N., Larkin, D.M., Griffin, D.K. (2018) Reconstruction of the diapsid ancestral genome permits chromosome evolution tracing in avian and non-avian dinosaurs. *Nature Communications*. **9**(1): 1883.

Odom, K.J., Hall, M.L., Riebel, K., Omland, K.E., Langmore, N.E. (2014) Female song is widespread and ancestral in songbirds. *Nature Communications*. **5**: 3379.

Ogoh, K., Ohmiya, Y. (2007) Concerted evolution of duplicated control regions within an ostracod mitochondrial genome. *Molecular Biology and Evolution*. **24**(1): 74-78.

Olson, S.L. (1975) Remarks on the generic characters of *Bulweria*. **Ibis**. **117**(1): 111-113.

Onley, D., Scofield, P. (2007) Albatrosses, Petrels and Shearwaters of the World. Princeton University Press, London.

Otto, S.P., Pannell, J.R., Peichel, C.L., Ashman, T.-L., Charlesworth, D., Chippindale, A.K., Delph, L.F., Guerrero, R.F., Scarpino, S.V., McAllister, B.F. (2011) About PAR: the distinct evolutionary dynamics of the pseudoautosomal region. *Trends in Genetics*. **27**: 358-367.

Ouzounis, C.A. (2005) Ancestral state reconstructions for genomes. *Current Opinion in Genetics and Development*. **15**(6): 595-600.

Palma, R.L., Tennyson, A.J.D., Gaskin, C.P., Jaramillo, A. (2012) The scientific name, author, and date for the "Fuegian Storm-Petrel", a subspecies of *Oceanites oceanicus* from southern South America. *Notornis*. **59**: 74-78.

Pamilo, P., Nei, M. (1988) Relationships between gene trees and species trees. *Molecular Biology and Evolution*. **5**(5): 568-583.

Parham, J.F., Donoghue, P.C., Bell, C.J. (2012) Best practices for justifying fossil calibrations. *Systematic Biology*. **61**(2): 346-359.

Pennycuick, C.J. (1982) The flight of petrels and albatrosses (procellariiformes), observed in South Georgia and its vicinity. *Philosophical Transactions of the Royal Society B*. **300**(1098): 75-106.

Philippe, H., Brinkmann, H., Lavrov, D.V., Littlewood, D.T., Manuel, M., Wörheide, G., Baurain, D. (2011) Resolving difficult phylogenetic questions: why more sequences are not enough. *PLoS Biology*. **9**(3): e1000602.

Phillimore, A.B., Owens, I.P. (2006) Are subspecies useful in evolutionary and conservation biology? *Proceedings of the Royal Society B: Biological Sciences*. **273**(1590): 1049-1053.

Phillips, R.A., Cooper, J., Burg, T.M (2018) Breeding-site vagrancy and hybridization in albatrosses. *Ibis*. **160**: 907-913.

Phillips, R.A., Fort, J., Dias, M.P. (2023) "Conservation status and overview of threats to seabirds" in Young, L., VanderWerf, E. (eds.) Conservation of Marine Birds, Academic Press, 33-56.

Phillips, R.A., Waluda, C.M. (2020) Albatrosses and petrels at South Georgia as sentinels of marine debris input from vessels in the southwest Atlantic Ocean. *Environment International*. **136**: 105443.

Picardi, E., Pesole, G. (2012) Mitochondrial genomes gleaned from human whole-exome sequencing. *Nature Methods*. **9**(6): 523-524.

Pickering, S.P.C. (1989) Attendance patterns and behaviour in relation to experience and pair-bond formation in the Wandering Albatross *Diomedea exulans* at South Georgia. *Ibis*. **131**(2): 183-195.

Pigot, A.L., Sheard, C., Miller, E.T., Bregman, T.P., Freeman, B.G., Roll, U., Seddon, N., Trisos, C.H., Weeks, B.C., Tobias, J.A. (2020) Macroevolutionary convergence connects morphological form to ecological function in birds. *Nature Ecology and Evolution*. **4**: 230-239.

Pistorius, P.A., William J. Sydeman, W.J., Watanuki, Y., Thompson, S.A., Orgeret, F. (2023) "Climate change: The ecological backdrop of seabird conservation" in Young, L., VanderWerf, E. (eds.) Conservation of Marine Birds, Academic Press, 245-276.

Portik, D.M., Wiens, J.J. (2021) Do Alignment and Trimming Methods Matter for Phylogenomic (UCE) Analyses? *Systematic Biology*. **70**(3): 440-462.

Presgraves, D.C. (2018) Evaluating signatures of "the large X-effect" during complex speciation. *Molecular Ecology*. **27**(19): 3822-3830.

Prince, P.A., Morgan, R.A. (1987) "Diet and feeding ecology of Procellariiformes" in Croxall, J.P. (ed.) Seabirds: feeding ecology and role in marine ecosystems, Cambridge, Cambridge University Press, 135-171.

Prum, R.O., Berv, J.S., Dornburg, A., Field, D.J., Townsend, J.P., Lemmon, E.M., Lemmon, A.R. (2015) A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature*. **526**: 569-573.

Purvis, A., Gittleman, J., Brooks, T. (2005) "Phylogeny and conservation" in Purvis, A., Gittleman, J., Brooks, T. (eds.) Phylogeny and Conservation, Cambridge University Press, 1-16.

Pyle, P., Welch, A.J., Fleischer, R.C. (2011) A new species of shearwater (*Puffinus*) recorded from Midway Atoll, Northwestern Hawaiian Islands. *Condor*. **113**: 518-527.

Quattrini, A.M., Faircloth, B.C., Dueñas, L.F. et al. (2018) Universal target-enrichment baits for anthozoan (Cnidaria) phylogenomics: New approaches to long-standing problems. *Molecular Ecology Resources*. **18**(2): 281-295.

Quillfeldt, P., Cherel, Y., Masello, J.F., Delord, K., McGill, R.A., Furness, R.W., Moodley, Y., Weimerskirch, H. (2015) Half a world apart? Overlap in nonbreeding distributions of Atlantic and Indian Ocean thin-billed prions. *PLoS One*. **10**: e0125007.

Quillfeldt, P., Moodley, Y., Weimerskirch, H., Cherel, Y., Delord, K., Phillips, R.A., Navarro, J., Calderón, L., Masello, J.F. (2017) Does genetic structure reflect differences in non-breeding movements? A case study in small, highly mobile seabirds. *BMC Ecology and Evolution*. **17**(1): 160.

Quinn, T.W., Wilson, A.C. (1993) Sequence evolution in and around the mitochondrial control region in birds. *Journal of Molecular Evolution*. **37**(4): 417-425.

Rackete, C., Poncet, S., Good, S.D., Phillips, R.A., Passfield, K., Trathan, P. (2021) Variation among colonies in breeding success and population trajectories of wandering albatrosses *Diomedea exulans* at South Georgia. *Polar Biology*. **44**: 221-227.

Rains, D., Weimerskirch, H., Burg, T.M. (2011) Piecing together the global population puzzle of wandering albatrosses: genetic analysis of the Amsterdam albatross *Diomedea amsterdamensis*. *Journal of Avian Biology*. **42**: 69-79.

Rambaut, A., Drummond, A.J., Xie, D., Baele, G., Suchard, M.A. (2018) Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Systematic Biology*. **67**(5): 901-904.

Rand, D.M. (1993) Endotherms, ectotherms, and mitochondrial genome-size variation. *Journal of Molecular Evolution*. **37**(3): 281-295.

Raposo do Amaral, F., Neves, L.G., Resende, M.F., Mobili, F., Miyaki, C.Y., Pellegrino, K.C., Biondo, C. (2015) Ultraconserved Elements Sequencing as a Low-Cost Source of Complete Mitochondrial Genomes and Microsatellite Markers in Non-Model Amniotes. *PLoS One*. **10**(9): e0138446.

Receveur, A., Allain, V., Menard, F., Dhaussy, A.L., Laran, S., Ravache, A., Bourgeois, K., Vidal, E., Hare, S.R., Weimerskirch, H., Borsa, P., Menkès, C. (2022) Modelling marine predator habitat using the abundance of its pelagic prey in the tropical southwestern Pacific. *Ecosystems*. **25**: 757-779.

Reddy, S., Kimball, R.T., Pandey, A. et al. (2017) Why Do Phylogenomic Data Sets Yield Conflicting Trees? Data Type Influences the Avian Tree of Life more than Taxon Sampling. *Systematic Biology*. **66**: 857-879.

Reid, K., Croxall, J.P., Edwards, T.M., Hill, H.J. & Prince, P.A. (1997) Diet and feeding ecology of the diving petrels *Pelecanoides georgicus* and *P. urinatrix* at South Georgia. *Polar biology*. **17**: 17-24.

Reneker, J., Lyons, E., Conant, G.C., Pires, J.C., Freeling, M., Shyu, C.R., Korkin, D. (2012) Long identical multispecies elements in plant and animal genomes. *Proceedings of the National Academy of Sciences of the United States of America*. **109**(19): E1183-E1191.

Reverter, A., Okimoto, R., Sapp, R., Bottje, W.G., Hawken, R., Hudson, N.J. (2017) Chicken muscle mitochondrial content appears co-ordinately regulated and is associated with performance phenotypes. *Biology Open*. **6**(1): 50-58.

Reyes, A., Yang, M.Y., Bowmaker, M., Holt, I.J. (2005) Bidirectional replication initiates at sites throughout the mitochondrial genome of birds. *Journal of Biological Chemistry*. **280**(5): 3242-3250.

Rhie, A., McCarthy, S.A., Fedrigo, O. et al. (2021) Towards complete and error-free genome assemblies of all vertebrate species. *Nature*. **592**: 737-746.

Richardson, M.E. (1984) Aspects of the ornithology of the Tristan da Cunha group and Gough Island, 1972-1974. *Marine Ornithology*. **12**: 123-201.

Ridoux, V. (1994) The diets and dietary segregation of seabirds at the subantarctic Crozet Islands. *Marine Ornithology*. **22**: 1-192.

Robertson, B.C., Stephenson, B.M., Ronconi, R.A., Goldstien, S.J., Shepherd, L., Tennyson, A., Carlile, N., Ryan, P.G. (2016) Phylogenetic affinities of the *Fregetta* storm-petrels are not black and white. *Molecular Phylogenetics and Evolution*. **97**: 170-176.

Robertson, H.A., Baird, K.A., Elliott, G.P., Hitchmough, R.A., McArthur, N.J., Makan, T.D., Miskelly, C.M., O'Donnell, C.F.J., Sagar, P.M., Scofield, R.P., Taylor, G.A., Michel, P. (2021) Conservation status of birds in Aotearoa New Zealand, 2021. *New Zealand Threat Classification Series 36*. Department of Conservation, Wellington.

Robinson, D.F., Foulds, L.R. (1981) Comparison of phylogenetic trees. *Mathematical Biosciences*. **53**: 131-147.

Roch, S., Steel, M. (2015) Likelihood-based tree reconstruction on a concatenation of alignments can be statistically inconsistent. *Theoretical Population Biology*. **100**: 56-62.

Roch, S., Warnow, T. (2015) On the robustness to gene tree estimation error (or lack thereof) of coalescent-based species tree methods. *Systematic Biology*. **64**: 663-676.

Rodrigues, P., Álvarez, P., Verdugo, C. (2017) Complete mitochondrial genome of the Neotropic cormorant (*Phalacrocorax brasilianus*). *Mitochondrial DNA Part A: DNA Mapping, Sequencing, and Analysis*. **28**(3): 401-402.

Rodríguez, A., Arcos, J.M., Bretagnolle, V. et al. (2019) Future directions in conservation research on petrels and shearwaters. *Frontiers in Marine Science*. **6**: 94.

Rodríguez, A., Rodríguez, B., Montelongo, T., Garcia-Porta, J., Pipa, T., Carty, M., Danielsen, J., Nunes, J., Silva, C., Geraldes, P., Medina, F.M., Illera, J.C. (2020) Cryptic differentiation in the Manx shearwater hinders the identification of a new endemic subspecies. *Journal of Avian Biology*. **51**(11): e02633

Rojas-Soto, O., Navarro-Sigüenza, A., Espinosa De Los Monteros, A. (2010) Systematics and bird conservation policies: The importance of species limits. *Bird Conservation International*. **20**(2): 176-185.

Rokas, A., Williams, B.L., King, N., Carroll, S.B. (2003) Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature*. **425**(6960): 798-804.

Roman, L., Hardesty, B.D., Hindell, M.A., Wilcox, C. (2019) A quantitative analysis linking seabird mortality and marine debris ingestion. *Scientific Reports*. **9**: 3202.

Ronquist, F., Sanmartín, I. (2011) Phylogenetic Methods in Biogeography. *Annual Review of Ecology, Evolution, and Systematics*. **42**: 441-464.

Roycroft, E., Moritz, C., Rowe, K.C., Moussalli, A., Eldridge, M.D.B., Portela, R.P., Piggott, M.P., Potter, S. (2022) Sequence Capture From Historical Museum Specimens: Maximizing Value for Population and Phylogenomic Studies. *Frontiers in Ecology and Evolution*. **10**: 931644.

Rubin, B.E., Ree, R.H., Moreau, C.S. (2012) Inferring phylogenies from RAD sequence data. *PLoS One*. **7**(4): e33394.

Rutkowska, J., Lagisz, M., Nakagawa, S. (2012) The long and the short of avian W chromosomes: no evidence for gradual W shortening. *Biology Letters*. **8**(4): 636-638.

Ryan, P.G., Bourgeois, K., Dromzée, S., Dilley, B.J. (2014) The occurrence of two bill morphs of prions *Pachyptila vittata* on Gough Island. *Polar Biology*. **37**: 727-735.

Ryan, P.G., Dean, W.R.J., Moloney, C.L., Watkins, B.P., Milton, S.J. (1990) New information on seabirds at Inaccessible Island and other islands in the Tristan da Cunha group. *Marine Ornithology*. **18**: 43-54.

Saitoh, Y., Saitoh, H., Ohtomo, K., Mizuno, S. (1991) Occupancy of the majority of DNA in the chicken W chromosome by bent-repetitive sequences. *Chromosoma*. **101**(1): 32-40.

Salichos, L., Rokas, A. (2013) Inferring ancient divergences requires genes with strong phylogenetic signals. *Nature*. **497**(7449): 327-331.

Salomon, M., Voisin, J.F. (2010) Ecogeographical variation in the Southern Giant Petrel (*Macronectes giganteus*). *Canadian Journal of Zoology*. **88**(2): 195-203.

Sammler, S., Bleidorn, C., Tiedemann, R. (2011) Full mitochondrial genome sequences of two endemic Philippine hornbill species (Aves: Bucerotidae) provide evidence for pervasive mitochondrial DNA recombination. *BMC Genomics*. **12**: 35.

San Mauro, D., Gower, D.J., Zardoya, R., Wilkinson, M. (2006) A hotspot of gene order rearrangement by tandem duplication and random loss in the vertebrate mitochondrial genome. *Molecular Biology and Evolution*. **23**(1): 227-234.

Sandelin, A., Bailey, P., Bruce, S., Engström, P.G., Klos, J.M., Wasserman, W.W., Ericson, J., Lenhard, B. (2004) Arrays of ultraconserved non-coding regions span the loci of key developmental genes in vertebrate genomes. *BMC Genomics*. **5**: 99.

Sangster, G., Braun, E.L., Johansson, U.S., Kimball, R.T., Mayr, G., Suh. A. (2022) Phylogenetic definitions for 25 higher-level clade names of birds. *Avian Research*. **13**: 100027.

Sausner, J., Torres-Mura, J.C., Robertson, J., Hertel, F. (2016) Ecomorphological differences in foraging and pattering behavior among storm-petrels in the eastern Pacific Ocean. *The Auk*. **133**: 397-414.

Schirtzinger, E.E., Tavares, E.S., Gonzales, L.A., Eberhard, J.R., Miyaki, C.Y., Sanchez, J.J., Hernandez, A., Müeller, H., Graves, G.R., Fleischer, R.C., Wright, T.F. (2012) Multiple independent origins of mitochondrial control region duplications in the order Psittaciformes. *Molecular Phylogenetics and Evolution*. **64**(2): 342-356.

Schluter, D., Price, T., Mooers, A.Ø., Ludwig, D. (1997) Likelihood of ancestor states in adaptive radiation. *Evolution*. **51**(6): 1699-1711.

Schramm, M. (1986) Burrow densities and nest site preferences of petrels (Procellariidae) at the Prince Edwards Islands. *Polar Biology*. **6**: 63-70.

Serventy, D.L. (1950) Taxonomic trends in Australian ornithology with special reference to the work of Gregory Mathews. *Emu*. **49**: 259-267.

Shaffer, S.A., Tremblay, Y., Weimerskirch, H., Scott, D., Thompson, D.R., Sagar, P.M., Moller, H., Taylor, G.A., Foley, D.G., Block, B.A., Costa, D.P. (2006) Migratory shearwaters integrate oceanic resources across the Pacific Ocean in an endless summer. *Proceedings of the National Academy of Sciences of the United States of America*. **103**(34): 12799-12802.

Sheard, C., Neate-Clegg, M.H.C., Alioravainen, N., Jones, S.E.I., Vincent, C., MacGregor, H.E.A., Bregman, T.P., Claramunt, S., Tobias, J.A. (2020) Ecological drivers of global gradients in avian dispersal inferred from wing morphology. *Nature Communications*. **11**(1): 2463.

Shen, X.X., Hittinger, C.T., Rokas, A. (2017) Contentious relationships in phylogenomic studies can be driven by a handful of genes. *Nature Ecology & Evolution*. **1**(5): 126-110.

Shen, X.X., Steenwyk, J.L., Rokas, A. (2021) Dissecting Incongruence between Concatenation- and Quartet-Based Approaches in Phylogenomic Data. *Systematic Biology*. **70**: 997-1014.

Shepherd, L.D., Miskelly, C.M., Bulgarella, M., Tennyson, A.J.D. (2022) Genomic analyses of fairy and fulmar prions (Procellariidae: *Pachyptila* spp.) reveals parallel evolution of bill morphology, and multiple species. *PLoS One*. **17**(9): e0275102.

Shoemaker, J.S., Fitch, W.M. (1989) Evidence from nuclear sequences that invariable sites should be considered when sequence divergence is calculated. *Molecular Biology and Evolution*. **6**: 270-289.

Silva, M.C., Matias, R., Wanless, R.M., Ryan, P.G., Stephenson, B.M., Bolton, M., Ferrand, N., Coelho, M.M. (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.

Simmons, M.P. (2012) Misleading results of likelihood-based phylogenetic analyses in the presence of missing data. *Cladistics*. **28**(2): 208-222.

Simmons, M.P., Gatesy, J. (2015) Coalescence vs. concatenation: Sophisticated analyses vs. first principles applied to rooting the angiosperms. *Molecular Phylogenetics and Evolution*. **91**: 98-122.

Simmons, M.P., Sloan, D.B., Gatesy, J. (2016) The effects of subsampling gene trees on coalescent methods applied to ancient divergences. *Molecular Phylogenetics and Evolution*. **97**: 76-89.

Singh, T.R., Shneor, O., Huchon, D. (2008) Bird mitochondrial gene order: insight from 3 warbler mitochondrial genomes. *Molecular Biology and Evolution*. **25**(3): 475-477.

Singhal, S., Grundler, M., Colli, G., Rabosky, D.L. (2017) Squamate Conserved Loci (SqCL): A Unified Set of Conserved Loci for Phylogenomics and Population Genetics of Squamate Reptiles. *Molecular Ecology Resources*. **17**: e12-e24.

Skaletsky, H., Kuroda-Kawaguchi, T., Minx, P.J. (2003) The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature*. **423**(6942): 825-837.

Skujina, I., McMahon, R., Lenis, V.P.E., Gkoutos, G.V., Hegarty, M. (2016) Duplication of the mitochondrial control region is associated with increased longevity in birds. *Aging*. **8**(8): 1781-1789.

Slack, K.E., Jones, C.M., Ando, T., Harrison, G.L., Fordyce, R.E., Arnason, U., Penny, D. (2006) Early penguin fossils, plus mitochondrial genomes, calibrate avian evolution. *Molecular Biology and Evolution*. **23**(6): 1144-1155.

Smeds, L., Kawakami, T., Burri, R., Bolivar, P., Husby, A., Qvarnström, A., Uebbing, S., Ellegren, H. (2014) Genomic identification and characterization of the pseudoautosomal region in highly differentiated avian sex chromosomes. *Nature Communications*. **5**: 5448.

Smeds, L., Warmuth, V., Bolivar, P. et al. (2015) Evolutionary analysis of the female-specific avian W chromosome. *Nature Communications*. **6**: 7330.

Smith, A.L., Monteiro, L., Hasegawa, O., Friesen, V.L. (2007) Global phylogeography of the band-rumped storm-petrel (*Oceanodroma castro*; Procellariiformes: Hydrobatidae). *Molecular Phylogenetics and Evolution*. **43**(3): 755-773.

Smith, B.T., Harvey, M.G., Faircloth, B.C., Glenn, T.C., Brumfield, R.T. (2014) Target capture and massively parallel sequencing of ultraconserved elements (UCEs) for comparative studies at shallow evolutionary time scales. *Systematic Biology*. **63**(1): 83-95.

Soh, Y.Q., Alföldi, J., Pyntikova T. et al. (2014) Sequencing the mouse Y chromosome reveals convergent gene acquisition and amplification on both sex chromosomes. *Cell*. **159**(4): 800-813.

Song, S., Liu, L., Edwards, S.V., Wu, S. (2012) Resolving conflict in eutherian mammal phylogeny using phylogenomics and the multispecies coalescent model. *Proceedings of the National Academy of Sciences of the United States of America*. **109**(37): 14942-14947.

Spatz, D.R., Jones, H.P., Bonnaud, E., Kappes, P., Holmes, N.D., Bedolla-Guzmán, Y. (2023) "Invasive species threats to seabirds" in Young, L., VanderWerf, E. (eds.) Conservation of Marine Birds, Academic Press, 97-130.

Springer, M.S., Gatesy, J. (2016) The gene tree delusion. *Molecular Phylogenetics and Evolution*. **94**: 1-33.

Steenwyk, J.L., Buida, T.J. Li, Y., Shen, X.X., Rokas, A. (2020) ClipKIT: A multiple sequence alignment trimming software for accurate phylogenomic inference. *PLoS Biology*. **18**(12): e3001007.

Stiglec, R., Ezaz, T., Graves, J.A. (2007) A new look at the evolution of avian sex chromosomes. *Cytogenetic and Genome Research*. **117**: 103-109.

Streicher, J.W., Schulte, J.A. II, Wiens, J.J. (2016) How Should Genes and Taxa be Sampled for Phylogenomic Analyses with Missing Data? An Empirical Study in Iguanian Lizards. *Systematic Biology*. **65**(1): 128-145.

Sudmant, P., Rausch, T., Gardner, E. et al. (2015) An integrated map of structural variation in 2,504 human genomes. *Nature*. **526**: 75-81.

Suh, A. (2016) The phylogenomic forest of bird trees contains a hard polytomy at the root of Neoaves. *Zoologica Scripta*. **45**: 50-62.

Suh, A., Kriegs, J.O., Brosius, J., Schmitz, J. (2011) Retroposon insertions and the chronology of avian sex chromosome evolution. *Molecular Biology and Evolution*. **28**(11): 2993-2997.

Sumner, J.G., Jarvis, P.D., Fernández-Sánchez, J., Kaine, B.T., Woodhams, M.D., Holland, B.R. (2012) Is the general time-reversible model bad for molecular phylogenetics? *Systematic Biology*. **61**(6): 1069-1074.

Swofford, D.L. (2002) PAUP*: Phylogenetic Analysis Using Parsimony (* and other methods), Version 4, Sinauer Associates: Sunderland, MA.

Tamashiro, R.A., White, N.D., Braun, M.J., Faircloth, B.C., Braun, E.L., Kimball, R.T. (2019) What are the roles of taxon sampling and model fit in tests of cyto-nuclear discordance using avian mitogenomic data? *Molecular Phylogenetics and Evolution*. **130**: 132-142.

Tavaré, S. (1986) "Some probabilistic and statistical problems in the analysis of DNA sequences," in Miura, R.M. (ed.) Some Mathematical Questions in Biology - DNA Sequence Analysis, Providence, RI: American Mathematical Society, 57-86.

Taylor, R.S., Bailie, A., Gulavita, P., Birt, T., Aarvak, T., Anker-Nilssen, T., Barton, D.C., Lindquist, K., Bedolla-Guzmán, Y., Quillfeldt, P., Friesen, V.L. (2017) Sympatric population divergence within a highly pelagic seabird species complex (*Hydrobates* spp.). *Journal of Avian Biology*. **49**: e01515.

Taylor, R.S., Bolton, M., Beard, A., Birt, T., Deane-Coe, P., Raine, A.F., González-Solís, J., Lougheed, S.C., Friesen, V.L. (2019) Cryptic species and independent origins of allochronic populations within a seabird species complex (*Hydrobates* spp.). *Molecular Phylogenetics and Evolution*. **139**: 106552.

Techow, N.M.S.M, Ryan, P.G., O'Ryan, C. (2009) Phylogeography and taxonomy of White-chinned and Spectacled Petrels. *Molecular Phylogenetics and Evolution*. **52**(1): 25-33.

Techow, N.M.S.M., O'Ryan, C., Phillips, R.A., Gales, R., Marin, M., Patterson-Fraser, D., Quintana, F., Ritz, M.S., Thompson, D.R., Wanless, R.M., Weimerskirch, H., Ryan, P.G. (2010) Speciation and phylogeography of giant petrels *Macronectes*. *Molecular Phylogenetics and Evolution*. **54**: 472-487.

Tennyson, A., Lawrence, H., Taylor, G., Imber, M. (2013) A hybrid gadfly petrel suggests that soft-plumaged petrels (*Pterodroma mollis*) had colonised the Antipodes Islands by the 1920s. *Notornis*. **60**: 290-295.

Tennyson, A.J.D., Bartle, J.A. (2005) A scientific name for fulmar prions nesting at Auckland and Heard Islands. *Notornis*. **52**: 47-55.

The 1000 Genomes Project Consortium (2015) A global reference for human genetic variation. *Nature*. **526**: 68-74.

Tickell, W.L.N. (1968) "The biology of the great albatrosses, *Diomedea exulans* and *Diomedea epomophora*" in Austin Jr, O.L. (ed.) Antarctic Bird Studies, Washington, DC: American Geophysical Union (AGU), 1-56.

Tobe, S.S., Kitchener, A.C., Linacre, A.M. (2010) Reconstructing mammalian phylogenies: a detailed comparison of the cytochrome B and cytochrome oxidase subunit I mitochondrial genes. *PLoS One*. **5**(11): e14156.

Tobias, J.A., Sheard, C., Pigot, A.L. et al. (2022) AVONET: morphological, ecological and geographical data for all birds. *Ecology Letters*. **25**(3): 581-597.

Toews, D.P., Brelsford, A. (2012) The biogeography of mitochondrial and nuclear discordance in animals. *Molecular Ecology*. **21**(16): 3907-3930.

Tollu, B. (1984) La Quille (Île Saint Paul, Océan Indien), sanctuaire de populations relictes. *L'Oiseau et la Revue francaise d'ornithologie*. **54**: 79-85.

Torres, L., Pante, E., González-Solís, J. et al. (2021) Sea surface temperature, rather than land mass or geographic distance, may drive genetic differentiation in a species complex of highly dispersive seabirds. *Ecology and Evolution*. **11**(21): 14960-14976.

Torres, L., Welch, A.J., Zanchetta, C., Chesser, R.T., Manno, M., Donnadieu, C., Bretagnolle, V., Pante, E. (2019) Evidence for a duplicated mitochondrial region in Audubon's shearwater based on MinION sequencing. *Mitochondrial DNA Part A: DNA Mapping, Sequencing, and Analysis*. **30**(2): 256-263.

Tsuda, Y., Nishida-Umehara, C., Ishijima, J., Yamada, K., Matsuda, Y. (2007) Comparison of the z and w sex chromosomal architectures in elegant crested tinamou (*Eudromia elegans*) and ostrich (*Struthio camelus*) and the process of sex chromosome differentiation in palaeognathous birds. *Chromosoma*. **116**(2): 159-173.

U.S. Fish & Wildlife Service (2022) Wisdom, the World's Oldest Known Wild Bird, Returns to Midway Atoll. Available at: https://www.fws.gov/story/2022-12/wisdom-worlds-oldest-known-wild-bird-returns-midway-atoll. Accessed 01/09/2023.

Umeda, S., Tang, Y., Okamoto, M., Hamasaki, N., Schon, E.A., Kang, D. (2001) Both heavy strand replication origins are active in partially duplicated human mitochondrial DNAs. *Biochemical and Biophysical Research Communications*. **286**(4): 681-687.

Urantówka, A.D., Kroczak, A., Mackiewicz, P. (2020) New view on the organization and evolution of Palaeognathae mitogenomes poses the question on the ancestral rearrangement in Aves. *BMC Genomics*. **21**: 874.

Urantówka, A.D., Kroczak, A., Silva, T., Padrón, R.Z., Gallardo, N.F., Blanch, J., Blanch, B., Mackiewicz, P. (2018) New Insight into Parrots' Mitogenomes Indicates That Their Ancestor Contained a Duplicated Region. *Molecular Biology and Evolution*. **35**(12): 2989-3009.

Urantówka, A.D., Kroczak, A., Strzala, T., Zaniewicz, G., Kurkowski, M., Mackiewicz, P. (2021) Mitogenomes of Accipitriformes and Cathartiformes were subject to ancestral and recent duplications followed by gradual degradation. *Genome Biology and Evolution*. **13**: evab193.

Van Dam, M.H., Henderson, J.B., Esposito, L., Trautwein, M. (2021) Genomic Characterization and Curation of UCEs Improves Species Tree Reconstruction. *Systematic Biology*. **70**(2): 307-321.

Vendrami, D.L., Telesca, L., Weigand, H., Weiss, M., Fawcett, K., Lehman, K., Clark, M.S., Leese, F., McMinn, C., Moore, H., Hoffman, J.I. (2017) RAD sequencing resolves fine-scale population structure in a benthic invertebrate: implications for understanding phenotypic plasticity. *Royal Society Open Science*. **4**(2): 160548.

Verkuil, Y.I., Piersma, T., Baker, A.J. (2010) A novel mitochondrial gene order in shorebirds (Scolopacidae, Charadriiformes). *Molecular Phylogenetics and Evolution*. **57**(1): 411-416.

Vianna, J.A., Fernandes, F.A.N., Frugone, M.J. et al. (2020) Genome-wide analyses reveal drivers of penguin diversification. *Proceedings of the National Academy of Sciences of the United States of America*. **117**(36): 22303-22310.

Voisin, J.F., Bester, M.N. (1981) "The specific status of giant petrels at Gough Island" in Cooper, J. (ed.) Proceedings of the Symposium on Birds of the Sea and Shore, African Seabird Group, Cape Town, 215-222.

Walker, J.F., Brown, J.W., Smith, S.A. (2018) Analyzing contentious relationships and outlier genes in phylogenomics. *Systematic Biology*. **67**(5): 916-924.

Wallace, S.J., Morris-Pocock, J.A., González-Solís, J., Quillfeldt, P., Friesen, V.L. (2017) A phylogenetic test of sympatric speciation in the Hydrobatinae (Aves: Procellariiformes). *Molecular Phylogenetics and Evolution*. **107**: 39-47.

Walsh, H.E., Kidd, M.G., Moum, T., Friesen, V.L. (1999) Polytomies and the power of phylogenetic inference. *Evolution*. **53**: 932-937.

Wang, N., Hosner, P.A., Liang, B., Braun, E.L., Kimball, R.T. (2017) Historical relationships of three enigmatic phasianid genera (Aves: Galliformes) inferred using phylogenomic and mitogenomic data. *Molecular Phylogenetics and Evolution*. **109**: 217-225.

Wang, X., Clarke, J.A. (2014) Phylogeny and forelimb disparity in waterbirds. *Evolution*. **68**: 2847-2860.

Wang, X.Y., Huang, Y., Liu, N., Yang, J., Lei, F.M. (2015) Seven complete mitochondrial genome sequences of bushtits (Passeriformes, Aegithalidae, *Aegithalos*): the evolution pattern in duplicated control regions. *Mitochondrial DNA*. **26**(3): 350-356.

Warham, J. (1977) Wing loadings, wing shapes, and flight capabilities of Procellariiformes. *New Zealand Journal of Zoology*. **4**(1): 73-83.

Warham, J. (1990) The petrels: their ecology and breeding systems. Academic Press, London.

Warham, J. (1996) The Behaviour, Population Biology and Physiology of the Petrels. Academic Press, London.

Warham, J., Watts, R., Dainty, R.J. (1976) The composition, energy content and function of the stomach oils of petrels (order, Procellariiformes). *Journal of Experimental Marine Biology and Ecology*. **23**(1): 1-13

Warnefors, M., Hartmann, B., Thomsen, S., Alonso, C.R. (2016) Combinatorial Gene Regulatory Functions Underlie Ultraconserved Elements in Drosophila. *Molecular Biology and Evolution*. **33**(9): 2294-2306.

Warren, W.C., Clayton, D.F., Ellegren, H. et al. (2010) The genome of a songbird. *Nature*. **464**: 757-762.

Warren, W.C., Hillier, L.W., Tomlinson, C., Minx, P., Kremitzki, M., Graves, T., Markovic, C., Bouk, N., Pruitt, K.D., Thibaud-Nissen, F., Schneider, V. (2017) A new chicken genome assembly provides insight into avian genome structure. *G3: Genes, Genomes, Genetics*. **7**: 109-117.

Wasser, D.E., Sherman, P.W. (2010) Avian longevities and their interpretation under evolutionary theories of senescence. *Journal of Zoology*. **280**: 103-155.

Weimerskirch, H., Cherel, Y., Delord, K., Jaeger, A., Patrick, S.C., Riotte-Lambert, L. (2014) Lifetime foraging patterns of the wandering albatross: Life on the move! *Journal of Experimental Marine Biology and Ecology*. **450**: 68-78.

Weimerskirch, H., Clobert, J., Jouventin, P. (1987) Survival in five southern albatrosses and its relationship with their life history. *Journal of Animal Ecology*. **56**: 1043-1055

Weimerskirch, H., Zotier, R., Jouventin, P. (1989) The avifauna of the Kerguelen Islands. *Emu*. **89**: 15-29.

Welch, A.J., Olson, S.L., Fleischer, R.C. (2014) Phylogenetic relationships of the extinct St. Helena petrel, *Pterodroma rupinarum* Olson (Procellariiformes: Procellariidae), based on ancient DNA. *Zoological Journal of the Linnean Society*. **170**: 494-505.

Wetterstrand, K.A. (2022) DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP). Available at: https://www.genome.gov/sequencingcostsdata. Accessed 01/09/2023.

Wheeler, Q.D., Raven, P.H., Wilson, E.O. (2004) Taxonomy: impediment or expedient? *Science*. **303**(5656): 285.

White, N.D., Braun, M.J. (2019) Extracting phylogenetic signal from phylogenomic data: Higher-level relationships of the nightbirds (Strisores). *Molecular Phylogenetics and Evolution*. **141**: 106611.

Whitfield, J.B., Lockhart, P.J. (2007) Deciphering ancient rapid radiations. *Trends in Ecology & Evolution*. **22**(5): 258-265.

Wiens, J.J., Morrill, M.C. (2011) Missing data in phylogenetic analysis: reconciling results from simulations and empirical data. *Systematic Biology*. **60**(5): 719-731.

Wilson Sayres, M.A. (2018) Genetic Diversity on the Sex Chromosomes. *Genome Biology and Evolution*. **10**(4): 1064-1078.

Wink, M., Heidrich, P., Kahl, U., Swatschek, I., Witt, H.-H., Ristow, D. (1993a) Inter- and intraspecific variation of the nucleotide sequence of the cytochrome *b* gene in Cory's (*Calonectris diomedea*), Manx Shearwater (*Puffinus puffinus*) and the Fulmar (*Fulmarus glacialis*). *Zeitschrift für Naturforschung*. **48**: 504-509.

Wink, M., Heidrich, P., Ristow, D. (1993b) Genetic evidence for speciation of the Manx Shearwater *Puffinus puffinus* and Mediterranean Shearwater *Puffinus yelkouan*. *Die Vogelwelt*. **114**: 226-232.

Winker, K., Glenn, T.C., Faircloth, B.C. (2018) Ultraconserved elements (UCEs) illuminate the population genomics of a recent, high-latitude avian speciation event. *PeerJ*. **6**: e5735.

Witherby, H.F., Jourdain, F.C.R., Ticehurst, N.F., Tucker, B.W. (1940) Handbook of British Birds. Vol. 4. Witherby, London.

Wood, J.R., Lawrence, H.A., Scofield, R., Taylor, G.A., Lyver, P.O., Gleeson, D.M. (2017) Morphological, behavioural, and genetic evidence supports reinstatement of full species status for the Grey-faced Petrel, *Pterodroma macroptera gouldi* (Procellariiformes, Procellariidae). *Zoological Journal of the Linnean Society*. **179**: 201-216.

Woolfe, A., Goodson, M., Goode, D.K. et al. (2004) Highly conserved non-coding sequences are associated with vertebrate development. *PLoS Biology*. **3**: e7.

Wu, D., Hugenholtz, P., Mavromatis, K. et al. (2009) A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea. *Nature*. **462**(7276): 1056-1060.

Xi, Z., Liu, L., Davis, C.C. (2015) Genes with minimal phylogenetic information are problematic for coalescent analyses when gene tree estimation is biased. *Molecular Phylogenetics and Evolution*. **92**: 63-71.

Xu, L., Wa Sin, S.Y., Grayson, P., Edwards, S.V., Sackton, T.B. (2019) Evolutionary Dynamics of Sex Chromosomes of Paleognathous Birds. *Genome Biology and Evolution*. **11**(8): 2376-2390.

Yang, Z. (1993) Maximum-likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites. *Molecular Biology and Evolution*. **10**(6): 1396-1401.

Yang, Z. (1998) On the best evolutionary rate for phylogenetic analysis. *Systematic Biology*. **47**(1): 125-133.

Yang, Z. (2006) Computational Molecular Evolution. Oxford University Press, Oxford.

Yang, Z., Rannala, B. (2012) Molecular phylogenetics: principles and practice. *Nature Reviews Genetics*. **13**(5): 303-314.

Yazdi, H.P., Silva, W.T.A.F., Suh, A. (2020) Why Do Some Sex Chromosomes Degenerate More Slowly Than Others? The Odd Case of Ratite Sex Chromosomes. *Genes*. **11**(10): 1153.

Young, A.D., Gillung, J.P. (2019) Phylogenomics - principles, opportunities and pitfalls of big-data phylogenetics. *Systematic Entomology*. **45**(2): 225-247.

Zhang, C., Rabiee, M., Sayyari, E., Mirarab, S. (2018) ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics*. **19**: 153.

Zhang, Y.M., Williams, J.L., Lucky, A. (2019) Understanding UCEs: A Comprehensive Primer on Using Ultraconserved Elements for Arthropod Phylogenomics. *Insect Systematics and Diversity*. **3**(5): 1-12.

Zhou, Q., Zhang, J., Bachtrog, D., An, N., Huang, Q., Jarvis, E.D., Gilbert, M.T., Zhang, G. (2014a) Complex evolutionary trajectories of sex chromosomes across bird taxa. *Science*. **346**(6215): 1246338.

Zhou, X., Lin, Q., Fang, W., Chen, X. (2014b) The complete mitochondrial genomes of sixteen ardeid birds revealing the evolutionary process of the gene rearrangements. *BMC Genomics*. **15**: 573.

Zuckerkandl, E., Pauling, L. (1965) "Evolutionary divergence and convergence in proteins" in Bryson, V., Vogel, H.J (eds.) Evolving genes and proteins, New York: Academic Press, 97-166.

Žydelis, R., Small, C., French, G. (2013) The incidental catch of seabirds in gillnet fisheries: A global review. *Biological Conservation*. **162**: 76-88.