UNIVERSIDADE ESTADUAL DE CAMPINAS

INSTITUTO DE BIOLOGIA

FLÁVIA AKEMI NITTA FERNANDES

EVOLUTIONARY AND BIOGEOGRAPHIC HISTORY OF THE EXTANT PENGUINS (AVES: SPHENISCIDAE) USING GENOMIC DATA

HISTÓRIA EVOLUTIVA E BIOGEOGRÁFICA DOS PINGUINS VIVENTES (AVES: SPHENISCIDAE) USANDO DADOS GENÔMICOS

CAMPINAS

2019

FLÁVIA AKEMI NITTA FERNANDES

EVOLUTIONARY AND BIOGEOGRAPHIC HISTORY OF THE EXTANT PENGUINS (AVES: SPHENISCIDAE) USING GENOMIC DATA

HISTÓRIA EVOLUTIVA E BIOGEOGRÁFICA DOS PINGUINS VIVENTES (AVES: SPHENISCIDAE) USANDO DADOS GENÔMICOS

Dissertation presented to the Institute of Biology of the University of Campinas in partial fulfillment of the requirements for the degree of Master in Genetics and Molecular Biology, in the area of Animal Genetics and Evolution.

Dissertação apresentada ao Instituto de Biologia da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do Título de Mestra em Genética e Biologia Molecular, na área de Genética Animal e Evolução.

ESTE ARQUIVO DIGITAL CORRESPONDE À VERSÃO FINAL DISSERTAÇÃO DEFENDIDA PELA ALUNA FLÁVIA AKEMI NITTA FERNANDES, E ORIENTADA PELA PROF(A) DR(A) MARIANA FREITAS NERY.

Orientadora: MARIANA FREITAS NERY

CAMPINAS

2019

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Biologia
Mara Janaina de Oliveira - CRB 8/6972

F391e

Fernandes, Flávia Akemi Nitta, 1995-
Evolutionary and biogeographic history of the extant penguins (Aves: Spheniscidae) using genomic data / Flávia Akemi Nitta Fernandes. – Campinas, SP : [s.n.], 2019.

Orientador: Mariana Freitas Nery.
Dissertação (mestrado) – Universidade Estadual de Campinas, Instituto de Biologia.

1. Filogenia. 2. Genomas. 3. Biogeografia. 4. Marcadores genéticos. 5. Pinguins. I. Nery, Mariana Freitas, 1983-. II. Universidade Estadual de Campinas. Instituto de Biologia. III. Título.

Informações para Biblioteca Digital

Título em outro idioma: História evolutiva e biogeográfica dos pinguins viventes (Aves:Spheniscidae) usando dados genômicos
Palavras-chave em inglês:
Phylogeny
Genomes
Biogeography
Genetic markers
Penguins
Área de concentração: Genética Animal e Evolução
Titulação: Mestra em Genética e Biologia Molecular
Banca examinadora:
Mariana Freitas Nery [Orientador]
Juliana José
Gisele Pires de Mendonça Dantas
Data de defesa: 29-04-2019
Programa de Pós-Graduação: Genética e Biologia Molecular

Identificação e informações acadêmicas do(a) aluno(a)
- ORCID do autor: https://orcid.org/0000-0001-7897-8976
- Currículo Lattes do autor: http://lattes.cnpq.br/4326910987159123

Campinas, 29 de abril de 2019

COMISSÃO EXAMINADORA

Prof.(a) Dr.(a) Mariana Freitas Nery

Dr.(a) Juliana José

Prof.(a) Dr.(a) Gisele Pires de Mendonça Dantas

Os membros da Comissão Examinadora acima assinaram a Ata de Defesa, que se encontra no processo de vida acadêmica do aluno.

AGRADECIMENTOS

Este trabalho foi realizado com apoio do Fundo de Apoio ao Ensino à Pesquisa e Extensão (FAEPEX) e Pró-reitoria de Pesquisa da Unicamp (PRP), através da concessão da minha Bolsa Mestrado (vinculada ao auxílio início de carreira) relativa à Sol. FAEPEX 202/17 (PAPDIC_1420/15: Bolsa de Mestrado da aluna Flávia Akemi Nitta Fernandes - Vigência: 01/03/2017 à 28/02/2019).

Agradeço aos meus pais e namorado por sempre me incentivarem, acompanharem e apoiarem durante todos os momentos da minha vida. Por sempre acreditarem em mim e estarem ao meu lado, independente da distância física em que nos encontramos.

Aos meus amigos por me ajudarem e apoiarem, especialmente durante estes dois anos de mestrado.

Aos amigos e colegas de laboratório e departamento pelo auxílio nas análises e discussões.

À Prof.ª Juliana Vianna da Pontificia Universidad Católica de Chile e demais colaboradores do projeto *Penguin Phylogenome*: Rauri Bowie, Ke Bi, Daly Noll, Céline Le Bohec, Christopher Burridge, Claudia Godoy, Patricia Parker, Gisele Dantas, Barbara Wienecke, Andrea Polanowski, Antje Steinfurth, Francesco Bonadonna, Pierre Pistorius, Cynthia Y. Wang-Claypool, Elie Poulin, pelo fornecimento de amostras de animais selvagens e também pertencentes a acervos de museus e liberação de dados moleculares gerados por tais amostras.

À minha orientadora por me proporcionar a primeira oportunidade de trabalhar com o grupo biológico com o qual sempre sonhei em trabalhar.

RESUMO

A história evolutiva e biogeográfica de espécies marinhas é altamente dependente de seus aspectos biológicos e dos oceanos que habitam, como é o caso dos pinguins viventes (Aves: Spheniscidae). Apesar de estudos anteriores terem focado na resolução do tempo e das relações evolutivas entre as espécies de pinguim viventes, assim como em sua correlação com a distribuição geográfica do grupo, as hipóteses na literatura ainda são controversas. Neste estudo, utilizamos elementos ultraconservados (UCEs) como marcadores genômicos para elucidar as principais questões a respeito da história evolutiva dos pinguins: (i) quais as relações evolutivas entre as espécies e gêneros de pinguins viventes, (ii) quando tais eventos de radiação aconteceram, (iii) qual a distribuição geográfica ancestral de Spheniscidae, e (iv) como esses eventos de radiação estão correlacionados com as dinâmicas e geologia dos oceanos do Hemisfério Sul. Para responder tais questões, inferimos uma filogenia robusta e estimamos os tempos de divergência e distribuição geográfica ancestral dos pinguins utilizando dados moleculares em escala genômica. Nossas análises recuperaram a divergência de Spheniscidae no início do Mioceno (21.9 milhões de anos atrás) e uma filogenia na qual *Aptenodytes* é a primeira linhagem a divergir, seguido de *Pygoscelis* e, finalmente, pelo clado que contém *Eudyptes*/*Megadyptes* e *Spheniscus*/*Eudyptula*. A maioria dos eventos de especiação ocorreu durante o Plioceno e Pleistoceno. Nossas análises da história biogeográfica recuperaram uma provável distribuição ancestral na região da Austrália/Nova Zelândia durante o aquecimento global do início do Mioceno, o que teria permitido a seguinte colonização do continente Antártico pelo ancestral de *Aptenodytes.* O surgimento da corrente circumpolar Antártica (ACC) provavelmente seria um fator chave na diversificação e expansão das linhagens de pinguins para suas atuais distribuições sub-Antárticas e sub-Tropicais. Resumidamente, demonstramos como dados genômicos são capazes de auxiliar na resolução de incongruências na história evolutiva dos pinguins. Nossos resultados podem amparar futuras comparações intra e interespecíficas relacionadas à importância de condições climáticas e oceânicas na radiação das diferentes espécies de pinguins, que podem servir de modelo para as possíveis respostas dessas aves no cenário de mudanças climáticas atuais e futuras.

ABSTRACT

The evolutionary and biogeographic history of marine species is highly dependent on their biology and the oceans they inhabit, which is the case of extant penguins (Aves: Spheniscidae). Although previous studies have focused on unraveling the evolutionary relationships of extant penguin species and the connection with their distributional range, the hypotheses in literature remain controversial. In this study, we used UCEs as genomic markers in order to unravel the main questions about the evolutionary history of extant penguins: (i) what are the evolutionary relationships among penguin species and genera, (ii) when radiation events occurred, (iii) what is the most likely geographic range of the ancestor of the crown penguins, and (iv) how the diversification events correlated with the Southern oceans' dynamics and geology. To answer these questions, we inferred a robust phylogenetic tree and estimated the divergence times and ancestral range of penguins using molecular data in genomic scale. Our analyses recovered an early Miocene divergence of Spheniscidae, 21.9 million years ago (Mya), and a phylogenetic relationship in which *Aptenodytes* is the first lineage to diverge, followed by *Pygoscelis* and finally by the clades formed by *Eudyptes/Megadyptes* and *Spheniscus/Eudyptula.* Most speciation events occurred during the Pliocene and Pleistocene, accompanying the global cooling events of the Pleistocene glaciations. Our historical biogeographical analyses recovered the most likely ancestral range for the Spheniscidae family on the Australia/New Zealand region during the early Miocene warming, which may have allowed the following colonization of the Antarctic continent by the *Aptenodytes* ancestor. The onset of the Antarctic circumpolar current (ACC) likely played a key role in the diversification and expansion of penguin lineages to the further sub-Antarctic and sub-tropical locations they currently inhabit. In summary, we demonstrate how genomic data can support the resolution of incongruences in the evolutionary history of penguins. Our results can aid further interspecific and intraspecific comparisons on the importance of ocean and climate conditions in the radiation of different penguin species, which can be useful to model these birds' putative responses to the ongoing climate change conditions.

LIST OF FIGURES

Figure 6. Ancestral range estimation results under the DIVALIKE+J model. a) Geographic locations of the areas inputted in BioGeoBEARS: A) South American coasts and Falkland islands; B) Scotia Arc islands; C) Antarctic Peninsula; D) Antarctic continent; E) Tristan da Cunha and Gough islands; F) Bouvet islands; G) South African coasts; H) Indian Ocean islands; I) Australia/New Zealand coasts and nearby islands; and J) Galapagos islands; b) Ancestral range reconstruction tree. Ranges on the tips of the tree represent the current distribution of the species, ranges on the nodes represent the most likely distribution of the clade ancestral, and ranges on the vertices represent the range distribution right after cladogenesis events.

(APPENDIX) Figure 7. Ancestral range estimation results under the a) DIVALIKE; b) DEC; c) DEC+J; d) BAYAREALIKE; e) BAYAREALIKE+J models. Geographic locations of the areas inputted in BioGeoBEARS: A) South American coasts and Falkland islands; B) Scotia Arc islands; C) Antarctic Peninsula; D) Antarctic continent; E) Tristan da Cunha and Gough islands; F) Bouvet islands; G) South African coasts; H) Indian Ocean islands; I) Australia/New Zealand coasts and nearby islands; and J) Galapagos islands. Ranges on the tips of the tree represent the current distribution of the species, ranges on the nodes represent the most likely distribution of the clade ancestral, and ranges on the vertices represent the range distribution right after cladogenesis events.

LIST OF TABLES

in literature which recovered Miocene estimates. *S: Spheniscus; Ea: Eudyptula; Es: Eudyptes; M: Megadyptes.*

LIST OF ABBREVIATIONS

ACC: Antarctic circumpolar current

APF: Antarctic Polar Front

AU test: Approximately unbiased test

BAYAREALIKE: Bayesian inference of historical biogeography for discrete areas (with likelihood)

COI: Cytochrome oxidase 1 gene

DEC: Dispersal-Extinction-Cladogenesis

DIVALIKE: Dispersal-Vicariance Analysis with likelihood

ELW: Expected likelihood weights

HPD: Highest posterior density

ILS: Incomplete lineage sorting

KH test: Kishino-Hasegawa test

Mya: Million years ago

ML: Maximum likelihood

MMCT: Middle Miocene climate transition

RAG-1: Recombination-activating gene 1

RELL: Resampling estimated log-likelihood

SH test: Shimodaira-Hasegawa test

STF: Subtropical Front

UCEs: Ultraconserved elements

UFBoot: Ultrafast bootstrap

SUMMARY

1. Introduction

1.1. Introducing penguins

The life histories of top marine predators are intrinsically connected with the dynamics of the oceans they inhabit. Their evolutionary and biogeographic history is highly associated with biogeographic barriers and the species' biological features, which are fundamental for determining the distribution of species in this heterogeneous environment (Munro and Berg 2017). Southern Ocean marine birds' distribution and evolution are especially associated with the main Southern Hemisphere ocean fronts (*e.g.* Subtropical Front - STF, Antarctic Polar Front - APF) and currents (*e.g.* Antarctic circumpolar current - ACC) (Bost et al. 2009) (Figure 1). This dependency on ocean fronts and currents is even more prominent in the evolution and biogeographic history of flightless marine birds, such as penguins.

Penguins (Sphenisciformes) are flightless marine birds with a widespread distribution in the Southern oceans. Ranging from the higher latitudes of the Antarctic continent to the lower latitudes of the equatorial region, penguin species may vary in size and annual cycles' length, but share various morphological, physiological and behavioral specializations for the marine life (Borboroglu and Boersma 2015). Among the adaptations that enabled life underwater, we can cite penguin's packed scale-like feathers, which provide thermal insulation for the waterproof effect (Taylor 1986), visual accommodation of the cornea for both air and water vision (Sivak 1976; Sivak and Millodot 1977; Bowmaker and Martin 1985), stiff wing joints (Raikow et al. 1988), and dense bones that counterbalance buoyancy while diving (Meister 1962).

The order Sphenisciformes includes both the stem group of extinct penguins, with more than 50 species, and the crown group of extant species, all included in the Spheniscidae family. The 18 extant penguin species are divided into six genera: *Aptenodytes* (comprising the two larger penguin species), *Pygoscelis* (three widespread Antarctic and sub-Antarctic penguins), *Spheniscus* (banded penguins), *Eudyptes* (crested penguins), *Eudyptula* (with the little penguin as its only species), and *Megadyptes* (with the yellow-eyed penguin as its only species) (Ksepka and Ando 2011). The exact number of extant species is still debatable due to putative deeper

intraspecific divergences (*e.g.* gentoo penguins, Vianna et al. 2017) and recent diversification events (*e.g.* little penguin, Banks et al. 2002; rockhopper penguins, Banks et al. 2006; royal and macaroni penguins, Frugone et al. 2018; and Humboldt and Galapagos penguins, Ramos et al. 2018).

As mentioned before, the ancient and ongoing divergences in penguin's evolutionary history are deeply correlated with the dynamics of the Southern Oceans, as well as with geological and climate change events (Kooyman 2002; Clarke et al. 2007; Clucas et al. 2018). In the last decades, several studies have focused on unraveling the evolutionary relationships and timing of diversification of extant penguins and its relationship with tectonic dynamics and the features of the oceans they inhabit (Jouventin 1982; Schreiweis 1982; O'Hara 1989; McKitrick 1991; Giannini and Bertelli 2004; Bertelli and Giannini 2005; Baker et al. 2006; Bertelli et al. 2006; Ksepka et al. 2006; Acosta Hospitaleche et al. 2007; Ksepka and Clarke 2010; Ksepka et al. 2012; Ksepka and Thomas 2012; Subramanian et al. 2013; Gavryushkina et al. 2017; Cole et al. 2019). Nevertheless, these studies have reached different results and hypotheses regarding penguins' evolutionary and biogeographic histories. Hence, the phylogenetic hypotheses, the estimated timing of diversification and the putative ancestral range of Spheniscidae described in the literature are still controversial.

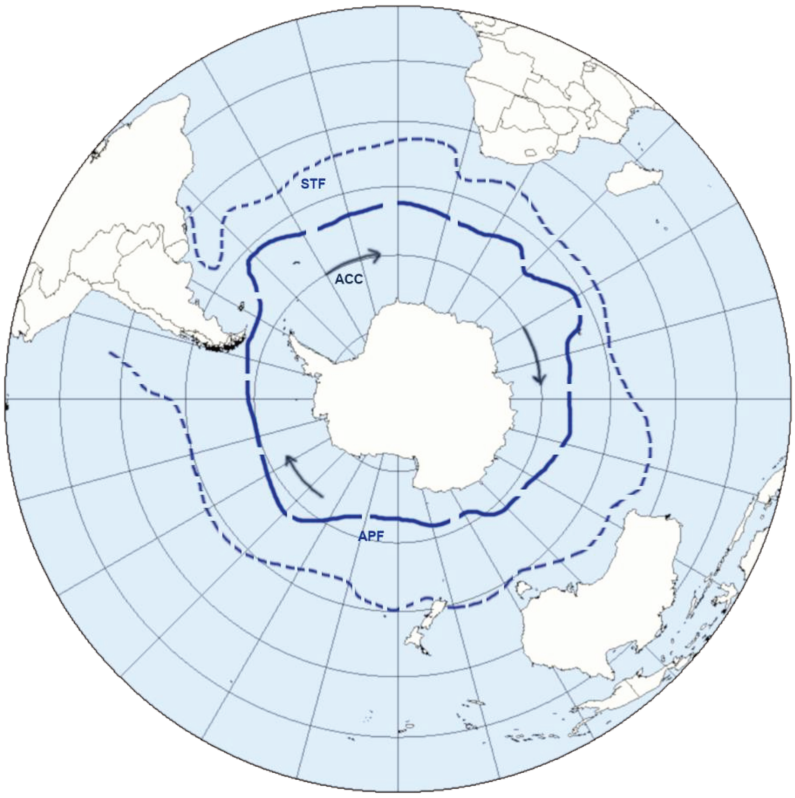


Figure 1. Southern Ocean main currents and fronts. External dotted lines represent the Sub-Tropical Front (STF), internal ones represent the Antarctic Polar Front (APF). Arrows represent the direction in which the Antarctic Circumpolar Current (ACC) flows, from west to east.

1.2. Literature on the evolutionary history of extant penguins

There are two main competing phylogenetic hypotheses at the genus level of extant penguins in literature. The clades formed by *Spheniscus/Eudyptula* and *Eudyptes/Megadyptes* are already well established, but the deeper position of *Aptenodytes* and *Pygoscelis* is still unclear. While some authors recover *Aptenodytes* as the first genus to diverge, followed by *Pygoscelis* and the two remaining extant clades (*i.e. Aptenodytes* sister to all other genera) (Bertelli and Gianni 2005; Baker et al. 2006; Bertelli et al. 2006; Ksepka et al. 2006; Clarke et al. 2007; Ksepka and Clarke 2010), others propose that *Aptenodytes* and *Pygoscelis* constitute a deep clade, sister to all other extant penguins (*i.e. Aptenodytes/Pygoscelis* as sister to all other genera)

(Subramanian et al. 2013; Gavryushkina et al. 2017; Cole et al. 2019) (Figure 2). Species-level relationships are less controversial, although the genus *Eudyptes* holds most of the problematic issues, due to putative introgression events among its taxa (Frugone et al. 2018; Vianna et al. unpublished).

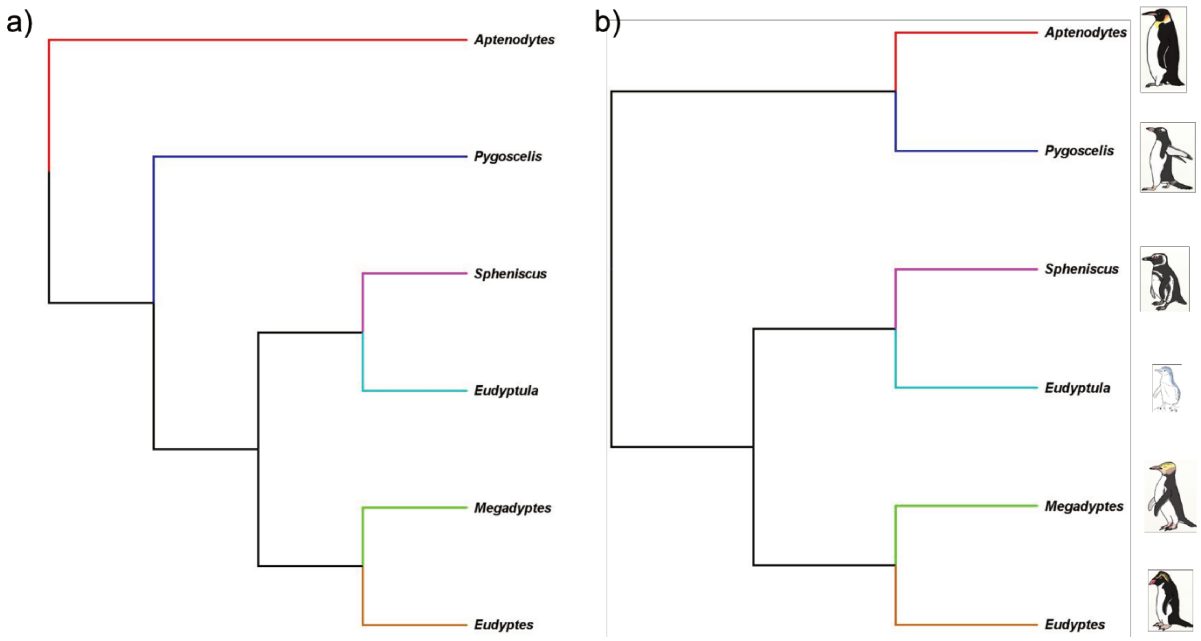


Figure 2. Representation of the two main contrasting phylogenetic hypotheses of Spheniscidae genera in literature (non-scaled cladograms). a) *Aptenodytes* as the sister genus to all other penguin groups, recovered by Bertelli and Gianni (2005), Baker et al. (2006), Bertelli et al. (2006), Ksepka et al. (2006), Clarke et al. (2007) and Ksepka and Clarke (2010); b) *Aptenodytes* and *Pygoscelis* as a clade, sister to the remaining extant groups, recovered by Subramanian et al. (2013), Gavryushkina et al. (2017) and Cole et al. (2019). Branch colors are the same as the ones used in Figures 4 and 5 of the 'Results' section and were used to facilitate the genera identification.

In addition to the unresolved phylogenetic relationships, the timing of the deeper lineage divergences of the crown group are also contrasting in literature. Earlier estimates propose an Eocene divergence of Spheniscidae from the stem group at approximately 40 million years ago (Mya) (Baker et al. 2006). However, the discovery and identification of the fossil *Madrynornis mirandus* (Acosta Hospitaleche et al. 2007),

which is the oldest Spheniscidae fossil known so far (10.0 ± 0.3 Mya), and the use of different tree topologies and estimation methodologies (*e.g.* fossilized birth-death model Stadler 2010), led the latest studies to recover a more recent divergence timing of extant penguins, in the Miocene, from 23 to 12.7 Mya (Subramanian et al. 2013; Gavryushkina et al. 2017; Cole et al. 2019).

Less explored than the phylogenetic and divergence time estimates, the biogeographic history of crown penguins is also an unresolved matter. Dated and not dated phylogenetic studies in the last decade attempted to estimate the ancestral range distribution of the group (Bertelli and Giannini 2005; Baker et al. 2006; Ksepka et al. 2006). Bertelli and Giannini (2005) and Ksepka et al. (2006) proposed an ancestral range for Spheniscidae in the Australia/New Zealand regions as well as in the Antarctic Peninsula, without dating the cladogenesis events. On the other hand, Baker et al. (2006) proposes an Antarctic origin for the extant penguins during the Eocene (~40 Mya) and, according to these authors, crown penguins may have expanded to lower latitudes as a response to global cooling during the second half of the Paleogene (Baker et al. 2006).

Although previous studies that tried to uncover the evolutionary and biogeographic history of extant penguins made use of technologies and tools available at the time they were published (*i.e.* mostly before 2013), there are some methodological considerations, which we focused on resolving in our work, that should be brought up. First, not all studies used a complete taxon dataset (*e.g.* Subramanian et al. 2013 used only 11 living species), therefore, they do not represent a complete evolutionary history of the group. In addition, many previous phylogenetic inferences were performed using parsimony methodologies, which are less time consuming and were likely the most compatible optimization methods for the available computational capacity at the time. However, for phylogenetic analyses using characters with more than two states (*e.g.* nucleotide data have four states: A, T, C, G), parsimony methods are not the most appropriate, once highly inconsistent trees can be generated (Felseinstein 1978).

Finally, most phylogenetic and divergence time estimation studies used a set of five genes alone (*i.e.* 12S, 16S, cytochrome oxidase 1 – COI, cytochrome b, and recombination-activating gene 1 – RAG-1) (Bertelli and Giannini 2005; Baker et al. 2006), together with five introns (Subramanian et al. 2013), or in combination with

morphological characters (Bertelli and Giannini 2005; Bertelli et al. 2006; Ksepka et al. 2006; Clarke et al. 2007; Ksepka and Clarke 2010; Ksepka et al. 2012; Ksepka and Thomas 2012; Gavryushkina et al. 2017). Data from a larger number of molecular loci was only possible to be implemented in penguin macroevolutionary research recently, by Cole et al. (2019), who inferred the dated phylogeny of penguins from the mitogenomes of extinct and extant species.

Many factors can influence molecular phylogenetic inferences, such as alignment and sequencing quality, presence of phylogenetically informative sequences that reflect the evolution of the species (*e.g.* excluding putative paralogs), sequence model selection, and efficiency of tree search algorithm (Talavera and Vila 2011). Thus, the different conclusions reached by phylogenetic studies using molecular data were likely due to the type and small number of genetic markers (*e.g.* most were mitochondrial markers). The incongruences between gene trees and species trees get even more sensitive when involving taxa with large effective population sizes and rapid radiation events (Maddison 1997), which is the case for penguins, since the most recent penguin phylogenetic studies propose that most internal penguin lineages diverged in less than 5 Mya (Gavryushkina et al. 2017; Cole et al. 2019). Moreover, other molecular events, such as introgression or incomplete lineage sorting could have made the resolution of these relationships more difficult (*e.g.* Scally et al. 2012; Wen et al. 2016; Wang et al. 2018).

Aside from the above-mentioned factors, fossil calibration also plays a key role on the timing of extant penguin diversification (Ho and Phillips 2009). Although Sphenisciformes hold a wide fossil record, most of fossil taxa is composed by incomplete skeletons and few bones, which hampers the accurate placement of the extinct specimens on the penguin phylogeny (Ksepka et al. 2012). For example, one of the most controversial, yet fundamental, fossil specimen for the crown penguin dating is *Madrynornis mirandus*. *M. mirandus* has already been considered a close relative to *Eudyptes* (Acosta Hospitaleche et al. 2007; Ksepka and Clarke 2010), to *Spheniscus/Eudyptula* (Degrange et al. 2018), and as a sister taxon to Spheniscidae (Hoffmeister 2014; Hoffmeister et al. 2014). Even though the exact position of the fossil is still uncertain, its crown status is the most certain supposition proposed so far (Degrange et al. 2018). Additionally, different tree topologies can also have an impact on divergence time estimation, once a node calibration applied to a clade that was

recovered by one study may not apply for the same clade in another study because of distinct tree topology (*e.g.* the differences regarding *Aptenodytes* position in Baker et al. 2006 and Subramanian 2013).

Finally, the discrepancies in the penguin biogeographic history in literature can be mostly explained by the different methodologies and area subdivisions used for the ancestral range reconstructions. Ancestral range is estimated using the dated phylogenetic tree of the species, data about its current geographical distribution, and a set of biogeographical parameters, such as dispersal, extinction, sympatry and vicariance (Matzke 2013b). The previous biogeographic inferences were performed based on character mapping on the phylogenies, without the application of specific biogeographic estimation tools (such as SIMMAP, Bollback 2006). Also, the distinct prior geographic subdivisions used by the studies was likely the main reason for the different ancestral ranges recovered. For instance, Baker et al. (2006) divided the current penguin range into three total areas following latitude sectors (60°S, 60° - 45°S, 45°S - 0°), whereas Bertelli and Giannini (2005) considered 10 different areas according to current penguin colony distribution.

We strongly believe that the incongruences regarding penguin evolution in literature are due to the poor choice and small quantity of molecular markers in combination with different analytic tools and priors used by the previous studies, which were not able to recover with much confidence the evolutionary history of Spheniscidae. In this scenario, here we aimed to use nuclear markers captured from high-throughput sequences of penguin species' genomes in order to provide more robust estimates of evolutionary relationships, divergence times and the ancestral range of extant penguins.

1.3. Ultraconserved elements as large-scale molecular markers

Recovering a robust dated phylogeny of extant penguins is the first and fundamental step towards unravelling their evolutionary history. The increased access to low cost high-throughput sequence data of non-model species has allowed the use of genomic data among several evolutionary distant and close taxa (Margulies et al. 2006). Along with that, the emergence of next-generation sequencing technologies has

led to an unprecedented amount of sequencing data in the last decade. As a consequence, phylogenetic analyses in genomic scale, also known as phylogenomics, of non-model organisms have provided the opportunity to refine former controversial evolutionary relationships, especially in cases where taxa diverged in a relatively rapid period of time, such as mammals and many bird groups (*e.g.* Hackett et al. 2008; McCormack et al. 2012; 2013; Nery et al. 2012; Jarvis et al. 2014; Zhang et al. 2014).

The ultraconserved elements (UCEs) are genomic scale molecular markers that have been increasingly used in phylogenomic studies throughout the last decade. UCEs consist on highly conserved regions of the genomes shared among evolutionary-distant taxa, firstly described for human, mouse and rat by Bejerano et al. (2004). Further studies revealed UCEs in more distant related taxa from human, such as fish, insects and even yeast, although the degree of sequence identity and the percentage of compatible alignments were lower in these cases than in closer species (Sandelin et al. 2004; Siepel et al. 2005).

UCEs have been widely and increasingly applied to infer phylogenetic relationships in both deep (Crawford et al. 2012; McCormack et al. 2012; 2013; Faircloth et al. 2013; 2015; Sun et al. 2014; Blaimer et al. 2015; Starrett et al. 2017; Van Dam et al. 2017; 2019; Smith et al. 2018) and shallow scales (Giarla and Esselstyn 2015; Lima et al. 2018; Winker et al. 2018). The use of UCEs for the detection of divergence at different timescales is possible due to the UCE structure. They have a highly conserved region at the center (*i.e.* core region) and two more variable flanking regions at each border. The flanking DNA contains most phylogenetically informative sites that are useful in the reconstruction of the evolutionary history at different taxonomic levels (Faircloth et al. 2012) (Figure 3), which was one of the reasons why we chose this type of marker, once we performed evolutionary analyses within and among the species scale.

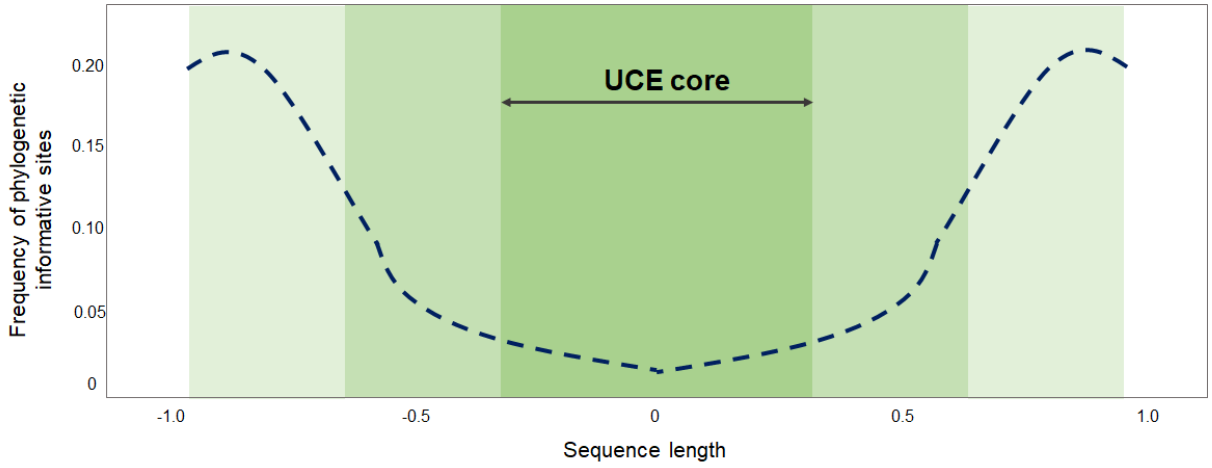


Figure 3. Illustration of the general UCE structure. Horizontal axis represents sequence length, and vertical axis represents the frequency of phylogenetic informative sites, which increase from the UCE core (darker green region) towards the flanking regions (gradually lighter green regions). Adapted from Van Dam et al. (2017).

The functional roles of UCEs are still unknown, although it is well established that most UCEs comprehend non-protein-coding regions (intronic or intergenic) that are involved in the regulation of transcription and in enhancer activities of developmental genes (Bejerano et al. 2004; Stephen et al. 2008). Regardless of the uncharted nature of UCEs, several controversial phylogenies have been successfully inferred using them as a molecular marker (*e.g.* Faircloth et al. 2013; 2015; Crawford et al. 2015; Bryson et al. 2016; Van Dam et al. 2019).

In this study, we used UCEs as genomic markers, in combination with proper analyze tools, in order to unravel the main questions regarding the macroevolutionary history of extant penguins: what are the evolutionary relationships among extant penguin species and genera, and in which timing have these diversification events occurred? What is the most likely geographic range of the ancestor of the crown penguins? How are these diversification events correlated with the Southern oceans' dynamics and geology? To answer these questions, we inferred a robust phylogenetic tree and estimated the divergence times and ancestral range of Spheniscidae family using molecular data in genomic scale, in combination with appropriate inference parameters and models. With this work, we demonstrate how large-scale target

sequence techniques can aid the resolution of incongruences regarding the evolutionary history of close taxa, as is the case for extant penguins.

2. Objectives

2.1. General objectives

Our study aimed to provide accurate phylogenetic estimates on the evolutionary and biogeographic history of extant penguins, unravelling the fundamental questions about this group evolution using high-throughput genomic data. Also, we intended to elucidate how Cenozoic oceanographic and climatological events may have influenced the radiation of this widespread Southern bird family.

2.2. Specific objectives

2.2.1. Biological objectives

Our specific biological objectives with this study were:

a) To estimate a robust and well supported phylogeny of extant penguin species and genera;
b) To estimate the divergence times among extant penguin lineages;
c) To estimate the ancestral range of Spheniscidae;
d) To associate the estimated biogeographic and cladogenetic events with ocean dynamics, continental shifts and global climate changes.

2.2.2. Methodological objectives

Our methodological objectives which make this study innovative in the area:

a) To use biological information derived from genomic data from roughly all extant penguin species in order to recover an evolutionary history more compatible with the species history;
b) To implement appropriate available methodologies and statistic tools for phylogenomics, divergence time, and ancestral range estimation.

3. Material and Methods

3.1. Taxa and sampling locations

Our dataset included seventeen penguin species and the outgroup *Macronectes giganteus* (Giant Petrel) from the order Procellariiformes, considered the evolutionary closest group to Sphenisciformes (Jarvis et al. 2014; Zhang et al. 2014). All extant penguin genera were represented in our study and almost all species, except for the endemic Snares islands penguin (*Eudyptes robustus*) due to the absence of samples. For the species *Eudyptes chrysolophus/E. c. schlegeli* (macaroni and royal penguins, considered as the same species in our study) and *Pygoscelis papua* (gentoo penguins), three and four specimens from distinct locations were sequenced, respectively. For the remaining species, only one representative was sequenced, totalizing 22 penguin individuals and the outgroup species (Table 1).

All sequences used in our study were captured from the reconstructed genomes of penguins generated by the project *Penguin Phylogenome.* The *Penguin Phylogenome* project is an international scientific collaboration among penguin specialists in which 22 genomes belonging to 17 living species were sequenced in order to unravel a complete evolutionary history of penguins, focusing on patterns of macroevolution, demography, biogeography, molecular evolution and ecology based on genomic data. Sample collection, DNA extraction, genome sequencing and assembly were performed by international collaborators from the *Penguin Phylogenome* project.

DNA was extracted from blood samples of wild animals for all penguins and the outgroup, except for *Spheniscus demersus* (African penguin), which came from an aquarium specimen, and *Eudyptes pachyrhynchus* (Fiordland), *E. sclateri* (erect-crested), and *Megadyptes antipodes* (yellow-eyed), which came from preserved museum specimens.DNA extraction was performed using a modified salt extraction protocol (Aljanabi and Martinez 1997; Vianna et al. 2017). Paired-end libraries were constructed using the Illumina TruSeq Nano kit. The 22 penguin genomes and the Giant Petrel (*Macronectes giganteus* – outgroup) genomes were sequenced to ~30x coverage with 150 paired reads using Illumina HiSeq X platform at Medgenome (USA).

Raw reads were trimmed and treated to remove the low-quality ones. Final clean reads were aligned to the Emperor penguin (*Aptenodytes forsteri*) draft genome (http://gigadb.org/dataset/100005; scaffold-level assembly) using LAST (http://last.cbrc.jp/). We had access to the assembled genomes in contigs and scaffolds. More detailed extraction and sequencing methodology are depicted in the supplementary material of Vianna et al. (unpublished).

Table 1. Species included in the study and the locations from where the samples were taken, including the outgroup species. The first column refers to the species scientific name, the second to the common name and the last to the sampling location.

Species		**Sample locations**
Common name	**Scientific name**	
Northern rockhopper penguin	*Eudyptes moseleyi*	Amsterdam Island
Eastern rockhopper penguin	*Eudyptes filholi*	Kerguelen Island
Southern rockhopper penguin	*Eudyptes chrysocome*	Terhalten Island, Chile
Erect-crested penguin	*Eudyptes sclateri*	Bounty Island, New Zealand (Museum)
Fiordland penguin	*Eudyptes pachyrhynchus*	Ocean off Albany, Western Australia, Australia (Museum)
Macaroni penguin	*Eudyptes chrysolophus*	Marion Island
Macaroni penguin	*Eudyptes chrysolophus*	Elephant Island, Antarctica
Royal penguin	*Eudyptes chrysolophus schlegeli*	Macquarie Island, Tasmania
Yellow-eyed penguin	*Megadyptes antipodes*	Southern Islands, New Zealand (Museum)
Magellanic penguin	*Spheniscus magellanicus*	Puñihuil, Chiloé, Chile
African penguin	*Spheniscus demersus*	Aquarium, California Academy of Science
Humboldt penguin	*Spheniscus humboldti*	Pan de Azucar, Chile

Galapagos penguin	*Spheniscus mendiculus*	Galapagos Islands
Little penguin	*Eudyptula minor*	Cheyne Island, Western Australia
Adelie penguin	*Pygoscelis adeliae*	Lagotellerie, Antarctica
Chinstrap penguin	*Pygoscelis antarcticus*	Narebski, Antarctica
Gentoo penguin	*Pygoscelis papua*	Antarctica
Gentoo penguin	*Pygoscelis papua*	Falkland/Malvinas
Gentoo penguin	*Pygoscelis papua*	Kerguelen Islands
Gentoo penguin	*Pygoscelis papua*	Crozet Islands
King penguin	*Aptenodytes patagonicus*	Inutil Bay, Chile
Emperor penguin	*Aptenodytes forsteri*	Pointe Géologie, Adélie Land, Antarctica
Giant petrel - outgroup	*Macronectes giganteus*	Antarctica

3.2. Sequence data treatment, UCE capture and alignment

We first filtered the scaffolds and contigs from the genomes using python scripts available in https://github.com/freitas-lucas/UCEs, in order to withdraw ambiguous nucleotides from the extended IUPAC code (Johnson 2010) different from “N”, once the aligner used by the UCE capture pipeline (LASTZ, Harris 2007), does not recognize sequence characters other than “A”, “T”, “C”, “G” or “N”.

We performed the UCE capture following the scripts from the PHYLUCE pipeline, which is a specific pipeline for UCE data (available in: https://github.com/faircloth-lab/phyluce) (Faircloth 2015). This pipeline is used for UCE loci identification in whole genomes. The identification is done by the alignment of a probe set of 5060 UCE loci specific for tetrapods (Faircloth et al. 2012 - available at https://www.ultraconserved.org/) to the species genome contigs and scaffolds. The script “probe_run_multiple_lastzs_sqlite” was used to align the 120 base pairs (bp) length UCE probes to the genomes. The UCE loci and their respective 500 bp flanking

regions were then sliced from the genomes using the script "probe_slice_sequence_from_genomes".

Once the UCEs are obtained from genomes using a target-sequence capture approach, high levels of missing data are expected, which means that not all UCEs will be found for all taxa in the analysis (Streicher et al. 2015). However, the existence of missing data itself is not a problem for phylogenetic analyses, so that excluding taxa or loci due to their missing data may be prejudicial for the accuracy of the estimates (Wiens and Tiu 2012; Wagner et al. 2013; Huang and Knowles 2014; Jiang et al. 2014; Hosner et al. 2015). Thus, our final dataset was composed by the UCEs that were present in a minimum of 18 of the 23 individuals, in order to maximize the amount of data in our matrix in a feasible level for divergence time analyses, and to deal with the missing data problem. Finally, sequences were aligned using MAFFT v7.407 (Nakamura et al. 2018) with a python loop script.

3.3. Phylogenetic analyses, divergence time estimation and tree topology tests

Phylogenomic analyses were done under concatenated and species-tree approaches. Model selection of sequence evolution was performed with ModelFinder (Kalyaanamoorthy et al. 2017), implemented in IQ-TREE version 1.6.8. (Nguyen et al. 2015). For the concatenated analyses, we carried out both maximum likelihood (ML) and Bayesian phylogenetic analyses. We ran the maximum likelihood analyses in IQ-TREE (Nguyen et al. 2015), with 1,000 bootstrap replicates using the ultrafast bootstrap approximation (UFBoot) (Hoang et al. 2017). The Bayesian phylogenetic analysis and the divergence time estimation were performed in BEAST v2.5.2 (Bouckaert et al. 2014) at CIPRES Science Gateway (Miller et al. 2010), using the GTRGAMMA model with base frequencies empirically estimated. We used a relaxed lognormal distribution clock under the calibrated Yule speciation process. The MCMC was run in two independent runs for 500 million generations, sampling and getting log parameters every 10,000 generations. The output log files were analyzed in Tracer v.1.7.1 (Rambaut et al. 2018) and trees were summarized using TREEANNOTATOR (Bouckaert et al. 2014) with a 10% burn-in.

We used the calibration priors suggested and explained by Cole et al (2019) using four crown fossils for internal calibrations and the oldest stem penguin fossil, *Waimanu manneringi* (Slack et al. 2006s), to calibrate the root of Sphenisciformes/Procellariiformes (Table 2). The maximum age bounds were determined according to the maximum estimated age of the geological layer where they were discovered (Cole et al. 2019).

We also inferred phylogenies using the multispecies coalescent method, in ASTRAL-III (Zhang et al. 2018). In multispecies coalescent approaches, such as the used by ASTRAL-III, tree estimation is made in a two-step approach. First, unrooted "gene trees" (*i.e.* a gene tree for each UCE loci) are estimated by maximum likelihood in IQ-TREE. Then, gene trees are summarized by ASTRAL-III to generate a single species tree, with branch support measured by the local posterior probability (Sayyari and Mirarab 2016).

Additionally, tree topology tests were carried out in IQ-TREE v1.6.8. to assess the log-likelihoods, posterior probabilities and weights of our alignment to recover the two main phylogenetic hypotheses in literature (*i.e.* H1: *Aptenodytes* as sister to all extant penguins; H1': *Aptenodytes* and *Pygoscelis* as sister groups). We performed the topology tests available in the software: the RELL approximation (Kishino et al. 1990) with 1000 bootstrap replicates, the one-sided Kishino-Hasegawa (KH) test (Kishino and Hasegawa 1989), the Shimodaira-Hasegawa (SH) test (Shimodaira and Hasegawa 1999), the expected likelihood weights (ELW) (Strimmer and Rambaut 2002), and the approximately unbiased (AU) test (Shimodaira, 2002).

Table 2. Fossil calibrations used in the divergence time estimation. Age intervals are represented in millions of years ago (Mya). Node numbers correspond to those in Figure 5.

Node number	Description	Age interval (Mya)	Fossil calibration taxa
-	Sphenisciformes /Procellariiformes (root)	60.5 – 72.1	*Waimanu manneringi*
1	Spheniscidae	9.7 – 25.2	*Madrynornis mirandus*

2	*Pygoscelis*	6.3 – 25.2	*Pygoscelis calderensis*
3	*Spheniscus/Eudyptula*	9.2 – 23.03	*Spheniscus muizoni*
4	*Eudyptes/Megadyptes*	3.06 – 25.2	*Eudyptes sp.*

3.4. Ancestral range estimation

For the historical biogeographic analysis, we estimated the ancestral range of the living penguin species in the R package BioGeoBEARS (Matzke 2013a). BioGeoBEARS implements the biogeographical models adopted by the most used methods for inference of ancestral ranges in a likelihood framework: Dispersal-Extinction-Cladogenesis (DEC) (Ree and Smith 2008), Dispersal-Vicariance Analysis (DIVA) (Ronquist 1997) (DIVALIKE), and Bayesian inference of historical biogeography for discrete areas (BAYAREA) (Landis et al. 2013) (BAYAREALIKE). Each of these methods allows the inclusion of a set of anagenetic free parameters (*e.g.* dispersal; extinction) and have fixed cladogenetic parameters (*e.g.* sympatry; vicariance).

The three standard models of the program (*i.e.* DEC, DIVA and BAYAREA) can be distinguished mostly based on the parameters considered by the model and the weights given to each parameter. Briefly, DIVA penalizes events other than vicariance, while DEC gives equal weights to all parameters, and BAYAREA does not consider vicariance (Matzke 2013a). In addition, BioGeoBEARS allows the inclusion of the cladogenetic parameter *j* ("jump dispersal"), which accounts for founder-event speciation (Templeton 2008), in any of the previous models. In the founder-event speciation, a daughter lineage can have a different range from the parental lineage during the cladogenesis event, an important feature to be considered, especially in oceanic island systems (Matzke 2014).

BioGeoBEARS estimates the likelihood of the ancestral ranges at the nodes of the phylogeny, using as input the dated phylogenetic tree and a geographic file containing the current distribution of the species. The phylogenetic tree used as input in BioGeoBEARS must have only one representant of each species (or monophyletic

populations) at the tips of the tree. Thus, we pruned our original 23 taxa tree in order to include only 17 taxa (one from each penguin species) and maintain original branch lengths, choosing the sampled oldest population (according to our phylogenetic analyses) for those species with more than one individual (*i.e. Pygoscelis papua* from Crozet archipelago and *Eudyptes chrysolophus* from Elephant island).

The geographic input file depicts the areas where each species natively breeds in a presence/absence matrix. The determination of each species' areas of occurrence followed the current distribution of penguin colonies in continental coasts and islands, available in the Handbook of the Birds of the World and BirdLife International (2017). We subdivided the extant penguin geographic distribution into 10 different areas according to the penguin species nesting sites in continental and island coasts and based on Bertelli and Giannini's (2005) subdivision: A) South American coasts and Falkland islands; B) Scotia Arc islands; C) Antarctic Peninsula; D) Antarctic continent; E) Tristan da Cunha and Gough islands; F) Bouvet islands; G) South African coasts; H) Indian Ocean islands; I) Australia/New Zealand coasts and nearby islands; and J) Galapagos islands (Figure 6a). Each area is considered as a character state in the tree, and the output tree contains the set of areas (range) covered by the ancestral lineage of Spheniscidae. We tested all previously mentioned models alone (DEC, DIVALIKE, BAYAREALIKE), and with the addition of the founder-event speciation parameter *j* (DEC+J, DIVALIKE+J, BAYAREALIKE+J) (Matzke 2013b).

We performed a constrained analysis, in which we took into consideration that not all 10 areas that are currently inhabited by penguins existed throughout their radiation, such as islands that emerged or were formed after a certain moment in geological time. This was the case of the islands of the Scotia Arc (B), Bouvet (F), Galapagos (J), and Tristan da Cunha and Gough (E) and the time in which each area emerged and became established is depicted in Table 3.

Table 3. Age of islands considered on the biogeographic analyses. First column corresponds to the name of the island / archipelago, second column indicates the minimum age of the area and third column indicates the paper from which the age information comes from.

Islands	Minimum age (Mya)	Paper
Scotia Arc	9.0	Dalziel et al. 2013b
Galapagos	3.0	Parent et al. 2008
Tristan da Cunha and Gough	2.5	Maund et al. 1988
Bouvet	1.0	Prestvik and Winsnes 1981

4. Results

4.1. UCE data information

We initially captured 4,096-4,940 UCE loci from each of the 22 penguins and the outgroup genomes. After quality filtering and sampling for subsets allowing missing data, we recovered a final dataset of 1,971 UCEs present in at least 18 lineages, 78% of total specimens. The final alignment length was of ~2,2 million base pairs (bp) per individual (incomplete matrix), with 73.4% of pairwise identity, with 0.7% identical sites and 39,457 parsimony informative sites. As expected for UCE data (Blair et a. 2018), the sequences showed an AT bias of 62.5% *versus* 37.5% of GC content. The best fit sequence evolution model for our data is TVM+F+R2, a base substitution transversion model, with equal transversion rates and unequal base frequencies, empirically estimated. Additionally, it considers the FreeRate model for rate heterogeneity across sites with two categories (Yang 1995; Soubrier et al. 2012).

4.2. A robust phylogeny of extant penguins

All phylogenetic inference approaches recovered the same genus-level topology, with 100% bootstrap (ML), 1.0 posterior probability (Bayesian inference), and 1.0 local posterior probability (multispecies coalescent) (Figure 4). *Aptenodytes* was recovered as the sister group to all other extant penguins, followed by the divergence of *Pygoscelis*. We also recovered the two main internal clades, *Spheniscus/Eudyptula* and *Eudyptes/Megadyptes*.

In the genus *Pygoscelis*, *Pygoscelis adeliae* is the first pygoscelid species to diverge, followed by *P. antarcticus* and *P. papua*. Within *P. papua*, the Crozet lineage was the first to diverge, followed by Kerguelen and the Atlantic lineages (Falkland and Antarctic). The genus *Spheniscus* had two internal clades, one with the lower latitude *S. mendiculus* and *S. humboldti* species and another with *S. magellanicus* and *S. demersus* species.

Finally, two main *Eudyptes* lineages were found. One including the paraphyletic *E. chrysolophus* (macaroni penguin) and *E. c. schlegeli* (royal penguin), in which the

macaroni sample from Marion Island (Indian Ocean) is more closely related to the New Zealander royal penguin than to the Atlantic macaroni penguin from Elephant Island (located in the West Antarctic Peninsula). The other *Eudyptes* lineage recovered was represented by the New Zealander penguins *E. pachyrhynchus* (Fiordland penguin) and *E. sclateri* (erect-crested penguin) as sister group to the rockhopper penguins. Inside the rockhoppers, *E. chrysocome* (Southern rockhopper) and *E. filholi* (Eastern rockhopper) are evolutionarily closer between each other than to *E. moseleyi* (Northern rockhopper).

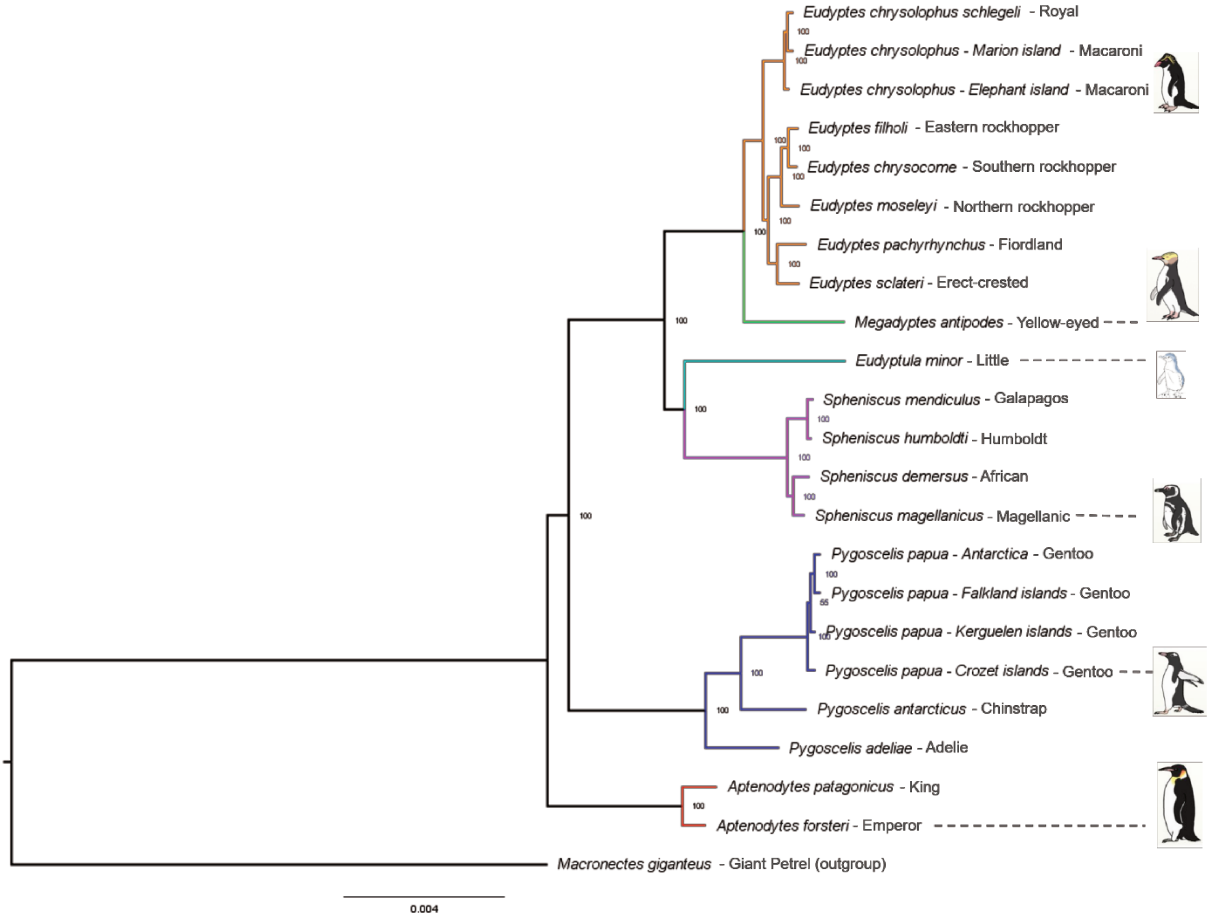


Figure 4. Maximum likelihood phylogenetic tree generated by IQ-TREE of the 23 individuals sampled in our study. Node labels indicate bootstrap values recovered by the ultrafast bootstrap approximation (UFBoot) in IQ-TREE. Branches are colored according to penguin genera: *Aptenodytes* (red), *Pygoscelis* (dark blue), *Spheniscus* (purple), *Eudyptula* (light blue), *Megadyptes* (green) and *Eudyptes* (yellow).

The tree topology tests revealed that the genus-level topology recovered by our study (*i.e.* H1- *Aptenodytes* sister to all extant genera) has a statistically significant superior log-likelihood than the contrasting phylogeny in literature (*i.e.* H1'- *Aptenodytes*+*Pygoscelis* as sister to the other genera). The KH, SH and AU tests rejected the alternative literature, and RELL and ELW returned much higher posterior weights for our topology compared to the alternative one. KH, SH and AU's p-values and RELL and ELW posterior weights are depicted in Table 4.

Table 4. Tree topology tests performed in IQ-TREE. First column corresponds to the hypotheses tested: (H1) corresponds to the *Aptenodytes* sister to all extant genera, and (H1') corresponds to *Aptenodytes* and *Pygoscelis* as a clade sister to the other extant genera. LogL is the log-likelihoods of each model. Bp-RELL represents the bootstrap proportion using RELL method, p-KH is the p-value of one sided Kishino-Hasegawa test, p-SH is the p-value of Shimodaira-Hasegawa test, c-ELW is the model selection probability of the expected likelihood weights (ELW), and p-AU is the p-value of the approximately unbiased (AU) test. Positive signs (+) indicate the best fit topology and negative signs (-) indicate the rejected topology under each test.

H	logL	bp-RELL	p-KH	p-SH	c-ELW	p-AU
1	-3782435	1 (+)	1 (+)	1 (+)	1 (+)	1 (+)
1'	-3782999	0 (-)	0 (-)	0 (-)	7.18^{-200} (-)	1.21^{-05} (-)

4.3. A Miocene origin of Spheniscidae

Divergence time estimation recovered a Miocene divergence of Spheniscidae approximately 21.9 Mya (95% highest posterior density - HPD - interval 19.1 - 25.2 Mya), when *Aptenodytes* diverged from the other penguin taxa (Figure 5). *Pygoscelis* and the clade containing the remaining genera diverged 20.3 Mya (95% HPD 17.4 - 22.9 Mya), and *Eudyptes/Megadyptes* and *Spheniscus/Eudyptula* 14.1 Mya (95% HPD 11.3 - 16.7 Mya). *Eudyptula* diverged from *Spheniscus* 12.6 Mya (95% HPD 9.4 - 16.1

Mya). *Megadyptes* and *Eudyptes* diverged 6.4 Mya (95% HPD 4.9 - 8.1 Mya) (APPENDIX - Table 6).

All species-level divergence events occurred in the past 10 million years. The *Aptenodytes* species diverged 2.6 Mya (95% HPD 1.0 - 4.4 Mya). In the *Pygoscelis* lineage, *P. adeliae* diverged from their congeners 8.4 Mya (95% HPD 6.4 - 11.2 Mya), followed by *P. antarcticus* and *P. papua* at 4.9 Mya (95% HPD 2.4 - 7.8 Mya). The banded penguins have a more recent divergence: *Spheniscus magellanicus/S. demersus* (Magellanic/African penguins) and *S. mendiculus/S. humboldti* (Galapagos/Humboldt penguins) divergence was estimated in 2.7 Mya (95% HPD 1.6 - 3.8 Mya). *S. magellanicus/S. demersus* speciation likely occurred in 1.5 Mya (95% HPD 0.04 - 2.7 Mya) and *S. mendiculus/S. humboldti* in around 591.3 thousand years ago (95% HPD 0.007 - 1.4 Mya). Finally, the two main *Eudyptes* lineages diverged at approximately 4.6 Mya (95% HPD 0.03 - 3.7 Mya). *E. moseleyi, E. chrysocome* and *E. filholi* diverged from *E. pachyrhynchus/E. sclateri* 3.06 Mya (95% HPD 0.1 - 5.3 Mya).

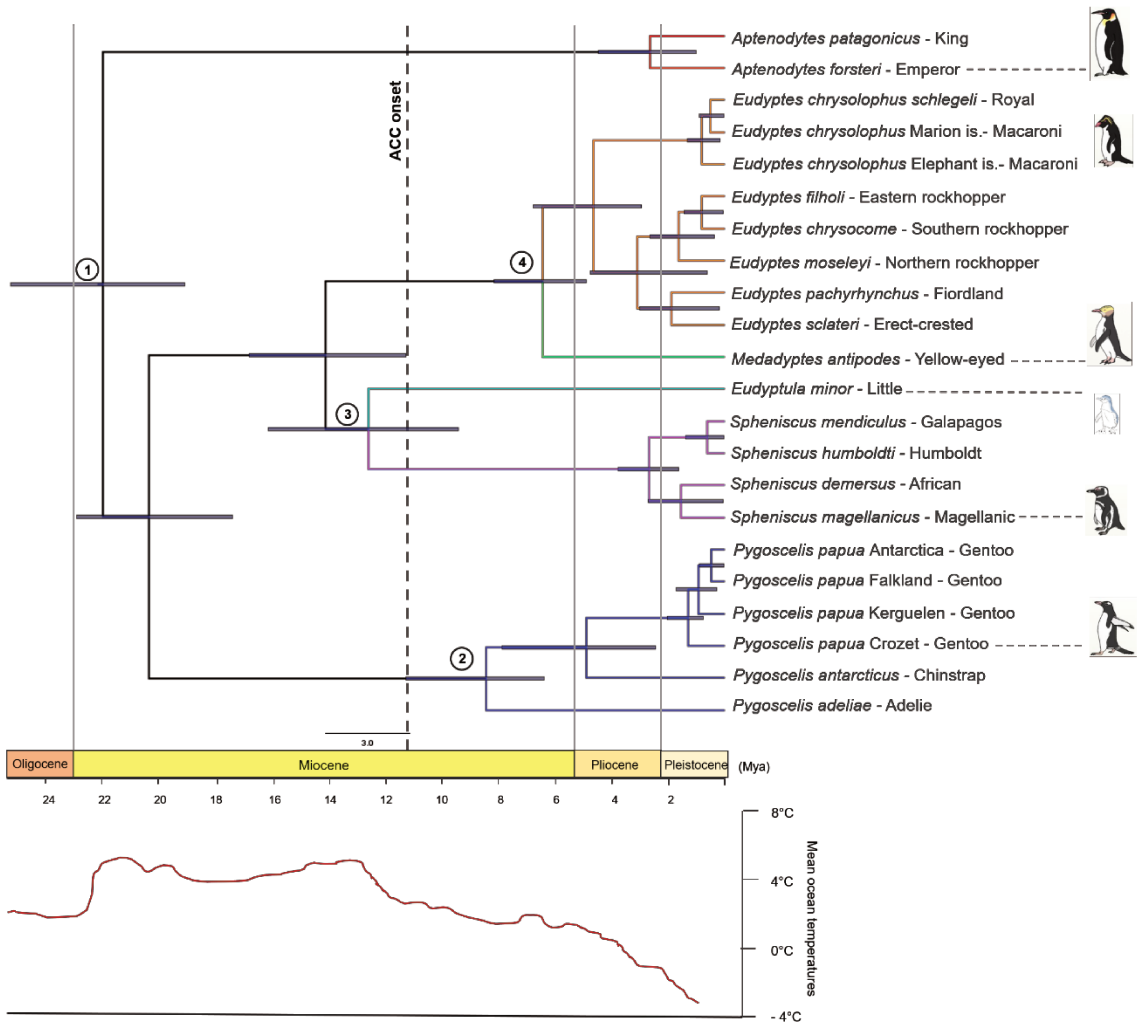


Figure 5. Dated phylogenetic tree of the extant penguin taxa recovered by BEAST v.2.5.2. Numbered nodes indicate calibration points depicted in Table 2: 1) Spheniscidae is calibrated with the fossil *Madrynornis mirandus,* 2) *Pygoscelis* with *Pygoscelis calderensis,* 3) *Spheniscus/Eudyptula* with *Spheniscus muizoni* and 4) *Eudyptes/Megadyptes* with *Eudyptes sp.* Blue bars correspond to the 95% highest posterior densities (HPD) probabilities. The Antarctic circumpolar current (ACC) complete onset is represented by the gray dashed line at 11.6 Mya. Mean ocean temperatures are depicted bellow, following information from Zachos et al. (2014).

4.4. Australia/New Zealand as the ancestral range of crown penguins

Biogeographic history estimation recovered a most likely ancestral range of "I" (*i.e.* Australia/New Zealand coasts and nearby islands) under most models (DEC, DIVALIKE, and DIVALIKE+J) (Figure 6b). Among the six biogeographic tested models, DIVALIKE+J was the best fit to our data (AICc = 158.3) (Table 5), followed by DEC+J (AICc = 159.2). BAYAREALIKE (AICc = 181.1) and BAYAREALIKE+J (AICc = 164.7) were the models the worst fit to our data. Both BAYAREALIKE and BAYARELIKE+J performed worst in our data, as expected, once these models are based on the unrealistic premise that vicariance does not occur (*i.e.* cladogenesis events cannot occur concomitantly with area separation events) (Moyle et al. 2016).

Table 5. BioGeoBEARS tested models' statistics. First column indicates the models (DEC, DIVALIKE, BAYAREALIKE with and without the inclusion of the *j* parameter); "LnL" column indicates the log-likelihood of our data given each model; "d*"* (dispersal)*, "*e" (extinction) and "j" (jump dispersal) columns indicate the likelihood of their respective parameters; "AICc" indicates the corrected Akaike information criterion and "Weights" the weighted AICc of the models. Bold line (DIVALIKE+J statistics) indicates the best fit of the six tested models, with the lowest AICc and highest weight.

Model	LnL	d	e	j	AICc	Weights
DEC	-78.47	0.026	0.017	0	161.8	0.07
DEC+J	-75.68	0.015	0.0009	0.093	159.2	0.26
DIVALIKE	-77.26	0.027	0.010	0	159.4	0.24
DIVALIKE+J	**-75.23**	**0.022**	**5.5e-08**	**0.058**	**158.3**	**0.41**
BAYAREALIKE	-88.10	0.031	0.12	0	181.1	4.7e-06
BAYAREALIKE+J	-78.44	0.020	0.034	0.120	164.7	0.01

When comparing the nested models' (*e.g.* DEC is nested in DEC+J) AICc and log-likelihoods, the models which included the jump dispersal ("*j"*) parameter were the best fit to our data. This implies that models which allow long distance dispersal and cladogenesis events to occur simultaneously are more suitable to analyze the extant penguin radiation. The importance of incorporating the founder-event speciation parameter has been previously shown in oceanic island species, such as wild coffee plants in the Hawaiian island system (Matzke 2014), in the diversification of songbirds

(Moyle et al. 2016), and anole lizards (Poe et al. 2017). This seems to be the case of extant penguins as well, which have a wide distribution in Southern Ocean islands and coasts.

According to our ancestral range estimates, Spheniscidae penguins likely originated in the Australia/New Zealand region, where most of the extant taxa initially diversified. Despite that, an early colonization of the Antarctic continent was performed by *Aptenodytes* clade ancestor, but further radiation of the extant *A. patagonicus* (king) and *A. forsteri* (emperor) species and expansion to other areas only occurred later in the Pliocene. *Pygoscelis* ancestors were the next to step out of the original Australian/New Zealander distribution, colonizing the Antarctic Peninsula by early Miocene, and only later settling in the Antarctic continent, in the case of the *P. adeliae* (Adelie penguin) ancestor, and Indian Ocean islands.

Spheniscus ancestor's expansion to South America likely occurred around 12.7 Mya and a much younger expansion to South Africa occurred only during late Pliocene (~2.65 Mya). According to the DIVALIKE+J model, the ancestor of *Spheniscus/Eudyptula* colonized South America, implying a posterior re-colonization of the Australian / New Zealander quadrant by *Eudyptula*'s ancestor. However, the second-best model, DEC+J, recovered an exclusive *Spheniscus* colonization of South America (APPENDIX Figure 7c), indicating that the *Eudyptula* lineage did not disperse further its original distribution, which is more biologically plausible. The same case is true for *S. magellanicus* (Magellanic) and *S. demersus* (African penguin), which are depicted as an originally South African clade in the DIVALIKE+J model, but in the DEC+J only *S. demersus* colonizes the region at early Pleistocene (~1.5 Mya). Again, the later hypothesis is more plausible than a *S. magellanicus*' South America – South Africa – South America dispersal supported by the first hypothesis. The re-colonization of South America (in the latter case) and Australia and New Zealand (in the *Eudyptula*'s case) by close related clades are both hardly likely due to the east wise flow of Southern Ocean cold currents (*e.g.* the Antarctic Circumpolar Current), as will be discussed in the following sections.

Finally, the *Eudyptes/Megadyptes* ancestor likely inhabited Australia and New Zealand. *Megadyptes antipodes* is the only species whose current distribution remained entirely within the ancestral range, while *Eudyptes* ancestor likely inhabited

the Indian Ocean islands. Within the *Eudyptes* clade, *E. chrysolophus* (Macaroni and Royal penguins) ancestor dispersed to the Antarctic Peninsula at 4.62 Mya, later colonizing the other locations of its current distribution, while the ancestor of the remaining *Eudyptes* re-colonized Australia and New Zealand. The New Zealander penguins *E. sclateri* (Erected-crested) and *E. pachyrhynchus* (Fiordland) diversified within their ancestor's range, while the rockhoppers (*E. chrysocome, E. filholi* and *E. moseleyi*) later dispersed to other locations, such as the newly formed Tristan da Cunha archipelago in the case of *E. moseleyi* (Northern rockhopper) and the Falkland islands in the case of *E. chrysocome* (Southern rockhopper).

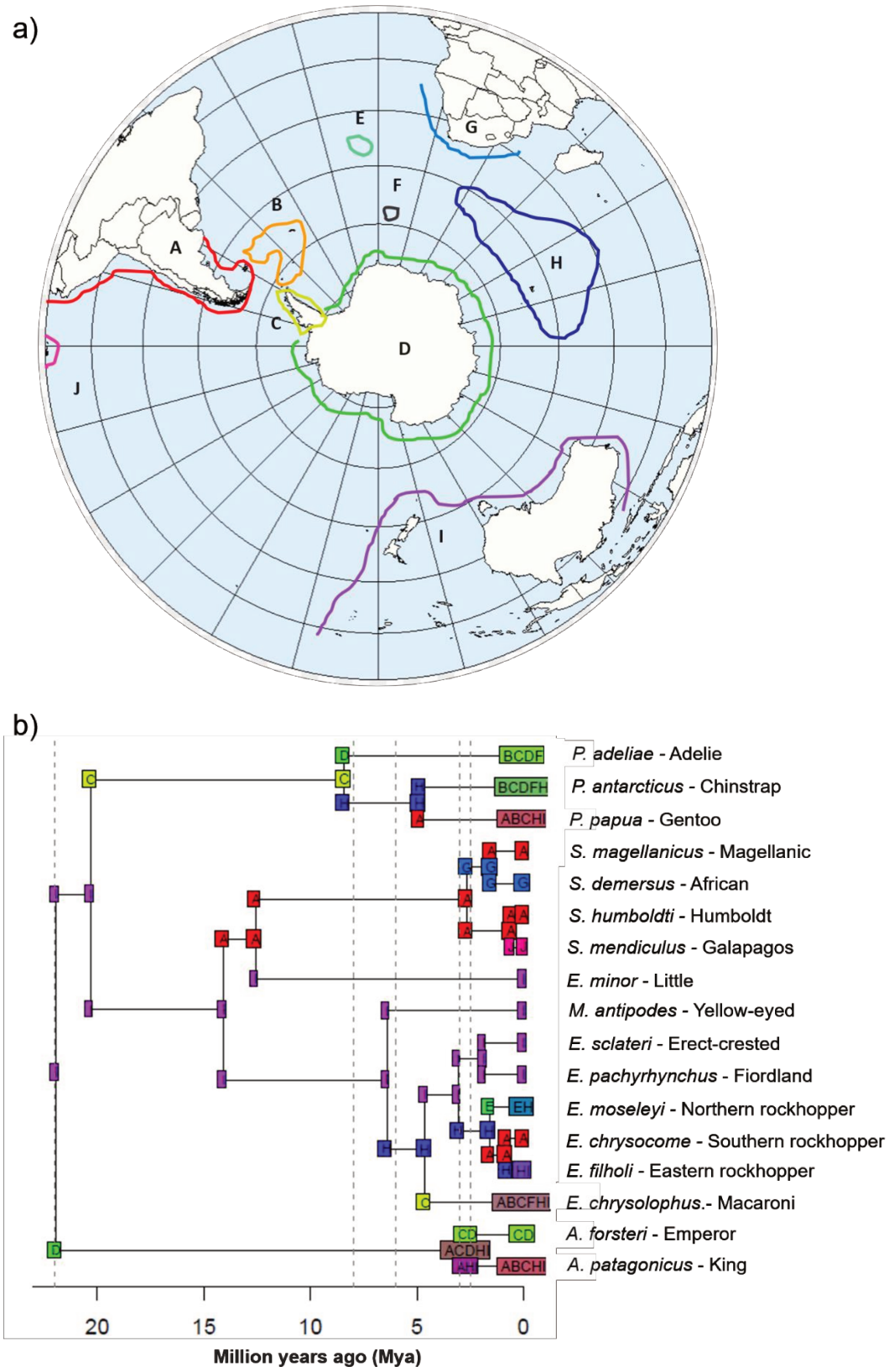


Figure 6. Ancestral range estimation results under the DIVALIKE+J model. a) Geographic locations of the areas inputted in BioGeoBEARS: A) South American coasts and Falkland islands; B) Scotia Arc islands; C) Antarctic Peninsula; D) Antarctic

continent; E) Tristan da Cunha and Gough islands; F) Bouvet islands; G) South African coasts; H) Indian Ocean islands; I) Australia/New Zealand coasts and nearby islands; and J) Galapagos islands; b) Ancestral range reconstruction tree. Ranges on the tips of the tree represent the current distribution of the species, ranges on the nodes represent the most likely distribution of the clade ancestral, and ranges on the vertices represent the range distribution right after cladogenesis events.

5. Discussion

5.1. UCEs unravel the evolutionary history of extant penguins

Genome-wide ultraconserved elements allowed us to recover a robust and highly supported phylogeny of extant penguins, in which *Aptenodytes* is sister group to the remaining penguin genera. Our genus-level phylogeny and tree topology tests found a higher likelihood for our hypothesis, which was also the relationship recovered by previous studies (*e.g.* Bertelli and Gianni 2005, Baker et al. 2006; Bertelli et al. 2006; Ksepka et al. 2006; Clarke et al. 2007; Ksepka and Clarke 2010) and reject the *Aptenodytes/Pygoscelis* clade hypothesis (from Subramanian et al. 2013; Gavryushkina et al. 2017; Cole et al. 2019). These last studies, which recovered a deep *Aptenodytes/Pygoscelis* relationship, used different datasets: Subramanian et al. (2013) used five genes (mitochondrial genes: 12S, 16S, cytochrome oxidase 1 - COI, and cytochrome b; nuclear gene: recombination-activating gene 1 - RAG-1) and five introns, Cole et al. (2019) used mitogenomes, while Gavryushkina et al. (2017) made a total-evidence analysis with morphological and molecular data with the same five genes as Subramanian et al. (2013).

The recovery of a different topology from our nuclear genomic-scale study was likely due to the poor choice of molecular markers by previous works. Although mitochondrial DNA (mtDNA) is free from recombination and, therefore, commonly used for phylogenetic inference, various studies have found inconsistencies in trees inferred with mtDNA and nuclear markers due to introgression and/or incomplete lineage sorting (ILS) (*e.g.* Carr et al. 1986; Funk and Omland 2003; Weisrock et al. 2005; Leaché and McGuire 2006; Wang et al. 2018). Thus, the mitochondrial and nuclear genes (*e.g.* 12S, 16S, RAG-1) used by the three studies that recovered *Aptenodytes/Pygoscelis* may have undergone ILS in the *Aptenodytes* and *Pygoscelis* lineages, recovering a sister lineage relationship. Moreover, the short internal branch recovered by our analysis at the divergence of *Aptenodytes* from other penguins likely indicates that a rapid radiation event occurred in less than 1 million years, possibly by multiple cladogenesis events occurring simultaneously in various lineages (Weisrock et al. 2005). This may have led to the recovery of the *Aptenodytes/Pygoscelis* clade by previous studies that used a small quantity of genes, which may have not separated

concomitantly with the lineages divergence, once fewer than 10 loci may be incapable of recovering rapid divergence events (Oliveros et al. 2019) Not only our coalescent analysis, but also the amount of genome-wide data used in our concatenated tree allowed the detection of the deep "*Aptenodytes*/other penguins" lineage splitting event. In addition, we performed further ML tree reconstructions with other high-throughput molecular markers (*e.g.* introns, coding-sequences, mitogenomes) (data not shown, publication in progress), which also recovered the same genus and species-level tree topology, in agreement with our UCEs tree results.

At the species level, we recovered the already well-established relationships within *Spheniscus* (African/Magellanic and Galapagos/Humboldt penguins) and *Pygoscelis* (Adelie sister to chinstrap/gentoo penguins) (Baker et al. 2006; Gavryushkina et al. 2017; Cole et al. 2019). Regarding the relationships among *Eudyptes*' species, literature is more controversial. Gavryushkina et al. (2017) found the same topology as our study, recovering the clade including *E. moseleyi, E. chrysocome, E. filholi* sister to the *E. pachyrhynchus* and *E. sclateri* penguins (and the Snares penguin, not included in our study) and a second clade including *E. chrysolophus* (macaroni) and *E. chrysolophus schlegeli* (royal) penguins.

As our phylogeny indicates, despite their morphological differences, *E. c. schlegeli* (royal) and *E. chrysolophus* (macaroni) are not separated monophyletic clades from the genetic point of view, as it had already been stated by Frugone et al (2018). Royal penguins are endemic to the sub-Antarctic Macquarie island and have a white face phenotype, while macaroni penguins are widespread in the sub-Antarctic islands of the three main oceans and have a black-face phenotype. A putative secondary contact between the two former species may have led to introgression events that makes it impossible to dissociate between the two of them using identification markers such as COI (Frugone et al. 2018). In this scenario, our results, in combination with Frugone et al. (2018), indicate a major need in the revaluation of the clade's taxonomic status.

Other previous studies which used birds UCEs to explore recent radiation events at the family level also recovered robust and well supported phylogenies. White et al. (2017) recovered a highly supported Nyctibiidae family (potoos) phylogeny using both concatenated and multispecies coalescent methods. They also allowed different levels of missing data (ranging from 75% to 100%), which showed no influence on the

robustness of the tree. Hosner et al. (2015) recovered highly supported landfowl (Aves: Galliformes) phylogenies under both concatenated and coalescent approaches but recovered more robust phylogenies when including only the most informative UCE loci.

All these studies, in the absence of reference genomes, sequenced the UCEs following proper probe design protocols, different from our study, in which we captured the UCEs from whole genomes. Generally, they captured a bigger quantity of UCE loci (2 to 4,8 k UCEs) with smaller flanking regions (UCE total length = ~100 bp in White at al. 2017, and 226 to 386 bp in Hosner et al. 2015) compared to our genome captured UCE dataset, due to quality trimming process. However, their final alignment lengths used for phylogenetic analyses were similar to the 2 billion bp of our study. On the other hand, Meiklejohn et al. (2016) estimated the phylogenetic relationships of the Phasianidae (pheasants and allies) using 1,479 UCEs of 193 to 774 bp (median = 400 bp) length, with a much smaller alignment length compared to our study (~599k bp). They observed low bootstrap support under some multispecies coalescent methods, which they resolved selecting more parsimony informative sites.

Using much smaller alignment lengths, such as the ~599k bp Meiklejohn et al. (2016), may not allow recovering full bootstrap support under multispecies coalescent methods. We therefore evidence, in agreement with Streicher et al. (2015), that the UCE power of phylogenetic resolution lies mostly on the high quantity of data this marker is able to provide, and that using smaller quantities of UCEs (< 1,500 loci) in combination with small UCE sizes (*e.g.* < 500 bp) may impact on the robustness of estimates. Nevertheless, these previous works in combination with our present study highlight the potential of UCEs in recovering robust phylogenetic results inside Aves. Besides their easiness of alignment among evolutionary distant taxa, low levels of saturation (McCormack et al. 2012) and abundance throughout vertebrate genomes (Bejerano et al. 2004), UCEs have shown to be useful in recovering conflicting phylogenetic relationships at the taxonomic level we have studied.

5.2. The importance of fossil calibration on penguin divergence time estimation

Our study recovered an early Miocene (~21.9 Mya) estimation for Spheniscidae origin, which is closer to the latest divergence time estimates of extant penguins

(Subramanian et al. 2013; Gavryushkina et al. 2017; Cole et al. 2019) than to the older Eocene estimate of Baker et al. (2006). The main factors that must have influenced in a more recent estimation are the calibration priors. Baker et al. (2006) only used external calibration priors, in contrast with more recent works. Additionally, the *Madrynornis mirandus* crown fossil, which Cole et al. (2019) and our study used to calibrate the Spheniscidae root, was discovered and identified by Acosta Hospitaleche et al. in 2007, right after Baker et al. (2006) publication.

The absence of a maximum calibration bound in a phylogeny can cause low substitution rates and arbitrary old divergence time estimates (Ho and Phillips 2009). Therefore, the use of *M. mirandus* as a Spheniscidae root calibration constraint pulled the divergence of Spheniscidae estimation to a more recent time, in the Miocene epoch. The inclusion of more recent calibration maximum and its impact on divergence time estimates have already been observed in other studies with different organisms, such as land plants (Morris et al. 2018), arthropods (Blair and Hedges 2004) and more recently in passerine birds (Oliveros et al. 2019).

Previous studies have shown that calibration uncertainty have a major influence over time estimation on phylogenies (Lee et al. 2009; Warnock et al. 2015; Morris et al. 2018). Not only the inclusion of the calibration itself, but the minimum and maximum bounds determined by prior densities have great impact on the estimates (Ho and Phillips 2009; Heled and Drummond 2011; Lee and Skinner 2011). Due to the relative abundance of crown penguin fossil record and the well-known stratigraphic layers in which they have been found (Ksepka and Ando 2011), the determination of calibration distribution and intervals made by Cole et al. (2019) and used by our study are among the most accurate so far under the node dating method we used.

Although Gavryushkina et al. (2017) estimated a much younger age for crown penguins of 12.7 Mya (95% HPD: 9.9 - 15.7), using a total-evidence approach for the dating and the fossilized birth-death model (Stadler 2010), all 95% HPD node age intervals of their study overlap with our estimates. The same is true for the Cole et al. (2019) and Subramanian et al. (2013) estimates, although the later used a smaller dataset of extant species and comparison among various nodes was not possible (APPENDIX, Table 6). However, these studies recovered the contrasting phylogenetic hypothesis (*Aptenodytes* and *Pygoscelis* as sister taxa) and it is recognizable that the

different tree topology has influenced the calibration used Subramanian et al. (2013). This study did not include the *M. mirandus* calibration prior but calibrated the misguided *Aptenodytes/Pygoscelis* clade with *Pygoscelis grandis* (~7.6 Mya) (Walsh and Suárez 2006), which pulled the Spheniscidae to a much younger age than Baker et al. (2006), closer to our estimate. Thus, even though the latest time divergence studies on extant penguin recovered similar age estimates to our results, the divergence nodes recovered by them are presumably wrong representations of the evolutionary history of penguins.

5.3. The influence of ocean dynamics and tectonics on penguin radiation

We estimated an early Miocene origin for Spheniscidae in the Australia/New Zealand region. The patterns of species radiation and range expansion suggests the fundamental role of the Antarctic circumpolar current (ACC) and global climate cooling events in the evolutionary and biogeographic history of extant penguins. The ACC is the largest ocean current in the world and encircles the Southern oceans eastward completely, allowing the admixture of the three main Southern oceans - Atlantic, Indian and Pacific (Barker et al. 2007). The timing of the ACC onset has been a debate (for a review, see Barker et al. 2007), but Dalziel et al. (2013a) has shown evidence that the onset of the deep ACC likely occurred 11.6 Mya (upper age limit), after the opening of the Drake Passage (between South America's Cape Horn and Antarctic Peninsula's South Shetland islands) and complete opening of the East Scotia Sea.

The influence of the ACC on extant penguin evolution and distribution has already been addressed by previous studies (Baker et al. 2006; Vianna et al. 2017; Cole et al. 2019) as well as with other marine species (Macaya and Zuccarello 2010; Nikula et al. 2010; Fraser et al. 2012). While most of the initial penguin radiation events took place in the Australia/New Zealand sector during the Miocene, the *Aptenodytes* and *Pygoscelis* penguins' ancestor colonized the Antarctic region (areas "D" – Antarctic Continent and "C" – Antarctic Peninsula respectively) by early Miocene, while Antarctic ice sheets were reduced compared to their current and previous Oligocene extent (Zachos et al. 2001). The *Aptenodytes* extant species divergence and expansion out of Antarctica, on the other hand, occurred much later during the Pleistocene glaciation, with a range expansion allowed by the ACC onset and possibly

correlated with permanent Antarctic ice sheet establishment (Zachos et al. 2001), especially for the *Aptenodytes patagonicus* (king penguin) lineage, which holds a sub-Antarctic distribution (Borboroglu and Boersma 2015).

Although the *Pygoscelis* ancestor inhabited the Antarctic Peninsula around 20.31 Mya, lineage-specific dispersal events, such as the colonization of Antarctic continent by the *P. adeliae* (Adelie) and the Indian Ocean islands by *P. antarcticus* (Chinstrap) and *P. papua* (Gentoo), occurred only after the ACC opening. Although not explored by our biogeographic analyses, our phylogeny in combination with previous more detailed population studies have shown that the relationship among Crozet, Kerguelen and Atlantic gentoo penguins follows the ACC direction. Vianna et al. (2017) included the same gentoo lineages of our study in addition to populations from Heard (Indian Ocean), Macquarie (Pacific Ocean) and Mirtillo islands (Patagonia) and also identified the pattern of eastern wise migration, following the ACC flow.

This would also be the case of the *Spheniscus/Eudyptula* ancestors, which colonized South America and Falkland Islands (area “A”) by the end of the middle Miocene climate transition (MMCT). This would have been possible due to a putative proto-ACC at approximately 12.57 Mya, which would enable the dispersal from the Australia/New Zealand region to South America in a clockwise direction (Barker et al. 2007), pattern which was already observed for South American and New Zealander mollusks (Beu et al. 1997). The posterior colonization of South African lower latitudes and the Galapagos islands in the Pleistocene occurred over the Pleistocene glaciation phase and may have been aided by the ACC and the north flow of the Benguela and Humboldt currents respectively.

While *Eudyptula*, *Megadyptes* and the *Eudyptes* clade including *E. sclateri* and *E. pachyrhynchus* divergences occurred entirely within the Australia/New Zealand quadrant, the *E. moseleyi, E. chrysocome, E. filholi* and *E. chrysolophus* showed a major range expansion pattern. *E. chrysolophus* likely dispersed to their currently broad distribution on the sub-Antarctic islands in multiple colonization events. For *Eudyptes moseleyi, E. chrysocome* and *E. filholi* (rockhoppers), whose ancestors likely inhabited sub-Antarctic and temperate regions, we can identify that the Subtropical front (STF) must have played an important role as a barrier separating the *Eudyptes moseleyi* species from *E. chrysocome* and *E. filholi*. The STF is a non-circular ocean

front between sub-Antarctic and temperate waters in the Southern oceans and is an effective barrier for gene flow for *Eudyptes* penguins, delimiting the occurrence of the *E. moseleyi* at the north and *E. chrysocome* and *E. filholi* south of the STF (Frugone et al. 2018).

Another oceanic front which played a key role in the diversification of the *Eudyptes* clades was the Antarctic Polar Front (APF). As showed in section 5.1., the macaroni and royal penguins (*E. chrysolophus*) genomic data indicates that the two species are likely experiencing a secondary contact. The relatedness between the royal penguin from the Macquarie islands and the macaroni from Marion island can be explained by the fact that both lineages are located north of the APF, while the more evolutionary distant macaroni penguin from Elephant island is located south of this front. Therefore, the APF likely played a key role in the genetic differentiation among the two macaroni penguins and, in combination with the ACC, may have aided the secondary contact between royal and macaroni penguins north of the APF.

Ocean fronts have been previously shown to be significant barriers for gene flow for organisms with both benthic (Griffiths et al. 2009; Poulin et al. 2014) and pelagic lifestyles, such as fish and cetaceans (Shaw et al. 2004; Rogers et al. 2006; Fontaine et al. 2007). Clucas et al. 2018 explored the effect of Southern Ocean fronts on *Pygoscelis papua* and *Aptenodytes patagonicus*, whose populations are distributed to the north and to the south of the Antarctic Polar Front (APF). Even for the more vagile *A. patagonicus*, the APF showed to be an important barrier for gene flow among colonies, which was more intensely observed for the coastal *P. papua* (Gentoo penguin). In this scenario, we the APF and STF likely played key roles in the macroevolution of penguins within their wide Southern Ocean distribution.

In combination with the ocean circulation dynamics and fronts, global climate changes have played key roles in penguin evolution. Global warmth events in early Miocene may have allowed the Antarctic continent colonization by the *Aptenodytes* ancestor, which was not covered by the present day eastern high-magnitude and thick ice-shelf (Flower and Kennett 1994). In contrast, temperature drops on global climate during the Miocene and Pleistocene may have orientated their expansion and diversification. The middle Miocene climate transition (MMCT) (Shevenell et al. 2004) is likely related with the expansion of the *Spheniscus/Eudyptula* ancestor to South

America. Indeed, the more recent late Pliocene-Pleistocene glaciations (De Schepper et al. 2014) are concurrent with most penguin speciation events and with lower latitude expansion events, such as *S. demersus* and *S. mendiculus* expansions to South Africa and Galápagos islands respectively, and *Aptenodytes patagonicus* to a sub-Antarctic distribution.

Our observation of cladogenesis events accompanied the onset of lower global temperatures periods in the Pliocene-Pleistocene boundary, which may have forced penguins to colonize warmer and ice-free grounds. In this case, we can agree with Baker et al. (2006) regarding that a second wave of cladogenesis events accompanied falls in mean global temperatures and expansion of sea ice, which may have led to the colonization of lower latitude areas and, consequently, radiation of various penguin genera. These events would have also been allowed by north-flowing ocean currents (such as the Humboldt and Benguela currents) and available habitats and niches at the new regions, such as Galapagos islands and South African coasts.

Studying past diversification patterns is essential for speculating the future trends and putative resilience ability of species facing climate changes scenarios. Clucas et al. (2014) modeled possible “winners” (populations which will expand with climate warming) and “losers” (populations which will retract with climate warming) in the global warming scenario through the demographic history of *Pygoscelis* penguins under the Last Glacial Maximum (LGM). They showed that *P. adeliae* and *P. antarcticus*, which likely had population expansions during the LGM, would be “losers” under future climate change, in contrast with *P. papua*, who would be “winners”. Additionally, further modeling studies alert for possible loss and restrictions on *Aptenodytes forsteri* and *A. patagonicus* current distribution range due to possible oceanographic disturbances mainly caused by climate change (Jenouvrier et al. 2017; Cristofari et al. 2018).

Although species response to specific environmental pressures can only be modeled and investigated in detail under microevolutionary studies (Gienapp et al. 2008), our findings point out for the radiation and adaptation of extant penguins in low global temperatures. Our study shows a potential correlation between global cooling and diversification events in penguins, putatively accompanied by ancestral niche expansion and diversification during expansion to new habitats. Thus, we alert for the

importance of investigating more penguin taxa's resilience modeling under the future climate warming scenario in order to detect future population trends that may even affect whole species.

6. Conclusions

Our analyses recovered an early Miocene divergence of Spheniscidae and a phylogenetic relationship in which *Aptenodytes* is the first lineage to diverge, followed by *Pygoscelis* and finally by the clade containing *Eudyptes/Megadyptes* and *Spheniscus/Eudyptula.* Most speciation events occurred during the Pliocene and Pleistocene, according to our estimates, during mean global temperature drops. The Antarctic circumpolar current is likely the fundamental factor that describes the extant penguin distribution patterns. We demonstrate how high-throughput genomic data can support the resolution of incongruences in the evolutionary history of penguins and how tree topology and fossil calibration can highly influence the divergence time estimation of the group. Our results can aid further interspecific and intraspecific comparisons on the importance of ocean and climate conditions in the radiation of different penguin species, which can be useful to model these birds' putative responses to the ongoing climate change conditions.

7. References

Acosta Hospitaleche, C., Tambussi, C., Donato, M. and Cozzuol, M. 2007. A new Miocene penguin from Patagonia and its phylogenetic relationships. Acta Palaeontologica Polonica, 52(2):299-314.

Aljanabi, S.M. and Martinez, I. 1997. Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. Nucleic Acids Research, 25(22):4692-4693.

Baker A.J., Pereira S.L., Haddrath O.P. and Edge K.A. 2006. Multiple gene evidence for expansion of extant penguins out of Antarctica due to global cooling. Proceedings of Royal Society B, 273:11-17.

Banks, J.C., Mitchell, A.D., Waas, J.R. and Paterson, A.M. 2002. An unexpected pattern of molecular divergence within the blue penguin (Eudyptula minor) complex. Notornis, 49(1):29-38.

Banks, J., Van Buren, A., Cherel, Y. and Whitfield, J.B. 2006. Genetic evidence for three species of rockhopper penguins, Eudyptes chrysocome. Polar Biology, 30(1):61-67.

Barker, P.F., Filippelli, G.M., Florindo, F., Martin, E.E. and Scher, H.D. 2007. Onset and role of the Antarctic Circumpolar Current. Deep Sea Research Part II: Topical Studies in Oceanography, 54(21-22):2388-2398.

Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W.J., Mattick, J.S. and Haussler, D. 2004. Ultraconserved elements in the human genome. Science, 304(5675):1321-1325.

Bertelli, S. and Giannini, N.P. 2005. A phylogeny of extant penguins (Aves: Sphenisciformes) combining morphology and mitochondrial sequences. Cladistics, 21(3):209-239.

Bertelli, S., Giannini, N.P. and Ksepka, D.T. 2006. Redescription and phylogenetic position of the early Miocene penguin Paraptenodytes antarcticus from Patagonia. American Museum Novitates, 3525:1-36.

Beu, A.G., Griffin, M. and Maxwell, P.A. 1997. Opening of Drake Passage gateway and Late Miocene to Pleistocene cooling reflected in Southern Ocean molluscan dispersal: evidence from New Zealand and Argentina. Tectonophysics, 281(1-2):83-97.

Blaimer, B.B., Brady, S.G., Schultz, T.R., Lloyd, M.W., Fisher, B.L. and Ward, P.S. 2015. Phylogenomic methods outperform traditional multi-locus approaches in resolving deep evolutionary history: a case study of formicine ants. BMC Evolutionary Biology, 15(1):271.

Blair, J.E. and Hedges, S.B. 2004. Molecular clocks do not support the Cambrian explosion. Molecular Biology and Evolution, 22(3):387-390.

Blair, C., Bryson Jr, R.W., Linkem, C.W., Lazcano, D., Klicka, J. and McCormack, J.E. 2018. Cryptic diversity in the Mexican highlands: thousands of UCE loci help illuminate phylogenetic relationships, species limits and divergence times of montane rattlesnakes (Viperidae: Crotalus). Molecular Ecology Resources, 19(2):349-365.

Bollback, J.P. 2006. SIMMAP: stochastic character mapping of discrete traits on phylogenies. BMC Bioinformatics, 7(1):88.

Borboroglu, P.G. and Boersma, P.D. 2015. Penguins: natural history and conservation. Seattle & London: University of Washington Press.

Bost, C.A., Cotté, C., Bailleul, F., Cherel, Y., Charrassin, J.B., Guinet, C., Ainley, D.G. and Weimerskirch, H. 2009. The importance of oceanographic fronts to marine birds and mammals of the southern oceans. Journal of Marine Systems, 78(3):363-376.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.H., Xie, D., Suchard, M.A., Rambaut, A. and Drummond, A.J. 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Computational Biology, 10(4):e1003537.

Bowmaker, J.K. and Martin, G.R. 1985. Visual pigments and oil droplets in the penguin, Spheniscus humboldti. Journal of Comparative Physiology A, 156(1):71-77.

Bryson Jr, R.W., Faircloth, B.C., Tsai, W.L., McCormack, J.E. and Klicka, J. 2016. Target enrichment of thousands of ultraconserved elements sheds new light on early relationships within New World sparrows (Aves: Passerellidae). The Auk, 133(3):451-458.

Carr, S.M., Ballinger, S.W., Derr, J.N., Blankenship, L.H. and Bickham, J.W. 1986. Mitochondrial DNA analysis of hybridization between sympatric white-tailed deer and mule deer in west Texas. Proceedings of the National Academy of Sciences, 83(24):9576-9580.

Clarke, J.A., Ksepka, D.T., Stucchi, M., Urbina, M., Giannini, N., Bertelli, S., Narváez, Y. and Boyd, C.A. 2007. Paleogene equatorial penguins challenge the proposed relationship between biogeography, diversity, and Cenozoic climate change. Proceedings of the National Academy of Sciences, 104(28):11545-11550.

Clucas, G.V., Dunn, M.J., Dyke, G., Emslie, S.D., Levy, H., Naveen, R., Polito, M.J., Pybus, O.G., Rogers, A.D. and Hart, T. 2014. A reversal of fortunes: climate change 'winners' and 'losers' in Antarctic Peninsula penguins. Scientific Reports, 4:5024.

Clucas, G.V., Younger, J.L., Kao, D., Emmerson, L., Southwell, C., Wienecke, B., Rogers, A.D., Bost, C.A., Miller, G.D., Polito, M.J. and Lelliott, P. 2018. Comparative population genomics reveals key barriers to dispersal in Southern Ocean penguins. Molecular Ecology, 27(23):4680-4697.

Cole, T.L., Ksepka, D.T., Mitchell, K.J., Tennyson, A.J., Thomas, D.B., Pan, H., Zhang, G., Rawlence, N.J., Wood, J.R., Bover, P. and Bouzat, J.L. 2019. Mitogenomes uncover extinct penguin taxa and reveal island formation as a key driver of speciation. Molecular Biology and Evolution, 36(4): 784-797.

Crawford, N.G., Faircloth, B.C., McCormack, J.E., Brumfield, R.T., Winker, K. and Glenn, T.C. 2012. More than 1000 ultraconserved elements provide evidence that turtles are the sister group of archosaurs. Biology Letters, 8(5):783-786.

Crawford, N.G., Parham, J.F., Sellas, A.B., Faircloth, B.C., Glenn, T.C., Papenfuss, T.J., Henderson, J.B., Hansen, M.H. and Simison, W.B. 2015. A phylogenomic analysis of turtles. Molecular Phylogenetics and Evolution, 83:250-257.

Cristofari, R., Liu, X., Bonadonna, F., Cherel, Y., Pistorius, P., Le Maho, Y., Raybaud, V., Stenseth, N.C., Le Bohec, C. and Trucchi, E. 2018. Climate-driven range shifts of the king penguin in a fragmented ecosystem. Nature Climate Change, 8(3):245.

Dalziel, I.W.D., Lawver, L.A., Pearce, J.A., Barker, P.F., Hastie, A.R., Barfod, D.N., Schenke, H.W. and Davis, M.B. 2013a. A potential barrier to deep Antarctic circumpolar flow until the late Miocene?. Geology, 41(9):947-950.

Dalziel, I.W., Lawver, L.A., Norton, I.O. and Gahagan, L.M. 2013b. The Scotia Arc: genesis, evolution, global significance. Annual Review of Earth and Planetary Sciences, 41:767-793.

Degrange, F.J., Ksepka, D.T. and Tambussi, C.P. 2018. Redescription of the oldest crown clade penguin: cranial osteology, jaw myology, neuroanatomy, and phylogenetic affinities of Madrynornis mirandus. Journal of Vertebrate Paleontology, 38(2):e1445636.

De Schepper, S., Gibbard, P.L., Salzmann, U. and Ehlers, J. 2014. A global synthesis of the marine and terrestrial evidence for glaciation during the Pliocene Epoch. Earth-Science Reviews, 135:83-102.

Faircloth, B.C., McCormack, J.E., Crawford, N.G., Harvey, M.G., Brumfield, R.T. and 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. and Alfaro, M.E. 2013. A phylogenomic perspective on the radiation of rayfinned fishes based upon targeted sequencing of ultraconserved elements (UCEs). PLoS One, 8(6):e65923.

Faircloth, B.C., Branstetter, M.G., White, N.D. and 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 BC. 2015. PHYLUCE is a software package for the analysis of conserved genomic loci. Bioinformatics, 32(5):786-788.

Felsenstein, J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. Systematic zoology, 27(4):401-410.

Flower, B.P. and Kennett, J.P. 1994. The middle Miocene climatic transition: East Antarctic ice sheet development, deep ocean circulation and global carbon cycling. Palaeogeography, Palaeoclimatology, Palaeoecology, 108(3-4):537-555.

Fontaine, M.C., Baird, S.J., Piry, S., Ray, N., Tolley, K.A., Duke, S., Birkun, A., Ferreira, M., Jauniaux, T., Llavona, A. and Öztürk, B. 2007. Rise of oceanographic barriers in continuous populations of a cetacean: the genetic structure of harbour porpoises in Old World waters. BMC Biology, 5(1):30.

Fraser, C.I., Nikula, R., Ruzzante, D.E. and Waters, J.M. 2012. Poleward bound: biological impacts of Southern Hemisphere glaciation. Trends in Ecology & Evolution, 27(8):462-471.

Frugone, M.J., Lowther, A., Noll, D., Ramos, B., Pistorius, P., Dantas, G.P.M., Petry, M.V., Bonadonna, F., Steinfurth, A., Polanowski, A. and Rey, A.R. 2018. Contrasting phylogeographic pattern among Eudyptes penguins around the Southern Ocean. Scientific Reports, 8(1):17481.

Funk, D.J. and 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(1):397-423.

Gavryushkina, A., Heath, T. A., Ksepka, D. T., Stadler, T., Welch, D. and Drummond, A. J. 2017. Bayesian total-evidence dating reveals the recent crown radiation of penguins. Systematic Biology, 66(1):57-73.

Giannini, N.P. and Bertelli, S. 2004. Phylogeny of extant penguins based on integumentary and breeding characters. The Auk, 121(2):422-434.

Giarla, T.C. and Esselstyn, J.A. 2015. The challenges of resolving a rapid, recent radiation: empirical and simulated phylogenomics of Philippine shrews. Systematic Biology, 64(5):727-740.

Gienapp, P., Teplitsky, C., Alho, J.S., Mills, J.A. and Merilä, J. 2008. Climate change and evolution: disentangling environmental and genetic responses. Molecular Ecology, 17(1):167-178.

Griffiths, H.J., Barnes, D.K. and Linse, K. 2009. Towards a generalized biogeography of the Southern Ocean benthos. Journal of Biogeography, 36(1):162-177.

Hackett, S.J., Kimball, R.T., Reddy, S., Bowie, R.C., Braun, E.L., Braun, M.J., Chojnowski, J.L., Cox, W.A., Han, K.L., Harshman, J. and Huddleston, C.J. 2008. A

phylogenomic study of birds reveals their evolutionary history. Science, 320(5884):1763-1768.

Handbook of the Birds of the World and BirdLife International. 2017. Handbook of the Birds of the World and BirdLife International digital checklist of the birds of the world. Version 9.1. Available at: http://datazone.birdlife.org/species/taxonomy

Harris, R.S. 2007. Improved Pairwise Alignment of Genomic DNA. Ph.D. Thesis, The Pennsylvania State University.

Heled, J. and Drummond, A.J. 2011. Calibrated tree priors for relaxed phylogenetics and divergence time estimation. Systematic Biology, 61(1):138-149.

Ho, S.Y. and Phillips, M.J. 2009. Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. Systematic Biology, 58(3):367-380.

Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q. and Vinh, L.S. 2017. UFBoot2: improving the ultrafast bootstrap approximation. Molecular Biology and Evolution, 35(2):518-522.

Hoffmeister, M.C. 2014. Phylogenetic characters in the humerus and tarsometatarsus of penguins. Polish Polar Research, 35(3):469-496.

Hoffmeister, M.C., Briceño, J.D.C. and Nielsen, S.N. 2014. The evolution of seabirds in the Humboldt Current: New clues from the Pliocene of central Chile. PloS One, 9(3):e90043.

Hosner, P.A., Faircloth, B.C., Glenn, T.C., Braun, E.L. and Kimball, R.T. 2015. Avoiding missing data biases in phylogenomic inference: an empirical study in the landfowl (Aves: Galliformes). Molecular Biology and Evolution, 33(4):1110-1125.

Huang, H. and Knowles, L.L. 2014. Unforeseen consequences of excluding missing data from next-generation sequences: simulation study of RAD sequences. Systematic Biology, 65(3):357-365.

Jarvis, E.D., Mirarab, S., Aberer, A.J., Li, B., Houde, P., Li, C., Ho, S.Y., Faircloth, B.C., Nabholz, B., Howard, J.T. and Suh, A. 2014. Whole-genome analyses resolve early branches in the tree of life of modern birds. Science, 346(6215):1320-1331.

Jenouvrier, S., Garnier, J., Patout, F. and Desvillettes, L. 2017. Influence of dispersal processes on the global dynamics of Emperor penguin, a species threatened by climate change. Biological Conservation, 212:63-73.

Jiang, W., Chen, S.Y., Wang, H., Li, D.Z. and Wiens, J.J. 2014. Should genes with missing data be excluded from phylogenetic analyses?. Molecular Phylogenetics and Evolution, 80:308-318.

Johnson, A.D. 2010. An extended IUPAC nomenclature code for polymorphic nucleic acids. Bioinformatics, 26(10):1386-1389.

Jouventin, P. 1982. Visual and vocal signals in penguins, their evolution and adaptive characters. Berlin: Paul Parey.

Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K., von Haeseler, A. and Jermiin, L.S. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nature Methods, 14(6):587.

Kishino, H. and Hasegawa, M. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. Journal of Molecular Evolution, 29(2):170-179.

Kishino, H., Miyata, T. and Hasegawa, M. 1990. Maximum likelihood inference of protein phylogeny and the origin of chloroplasts. Journal of Molecular Evolution, 31(2):151-160.

Kooyman, G.L. 2002. Evolutionary and ecological aspects of some Antarctic and sub-Antarctic penguin distributions. Oecologia, 130(4):485-495.

Ksepka, D.T., Bertelli, S. and Giannini, N.P. 2006. The phylogeny of the living and fossil Sphenisciformes (penguins). Cladistics, 22(5):412-441.

Ksepka, D.T. and Clarke, J.A. 2010. The basal penguin (Aves: Sphenisciformes) Perudyptes devriesi and a phylogenetic evaluation of the penguin fossil record. Bulletin of the American Museum of Natural History, 337:1-77.

Ksepka, D.T. and Ando, T. 2011. Penguins past, present, and future: trends in the evolution of the Sphenisciformes. Living Dinosaurs. Oxford: Wiley, 155-186.

Ksepka, D.T., Fordyce, R.E., Ando, T. and Jones, C.M. 2012. New fossil penguins (Aves, Sphenisciformes) from the Oligocene of New Zealand reveal the skeletal plan of stem penguins. Journal of Vertebrate Paleontology, 32(2):235-254.

Ksepka, D.T. and Thomas, D.B. 2012. Multiple Cenozoic invasions of Africa by penguins (Aves, Sphenisciformes). Proceedings of the Royal Society B, 1730(279):1027-1032.

Landis, M. J., Matzke, N. J., Moore, B. R. and Huelsenbeck, J. P. 2013. Bayesian analysis of biogeography when the number of areas is large. Systematic Biology, 62(6):789-804.

Leaché, A.D. and McGuire, J.A. 2006. Phylogenetic relationships of horned lizards (Phrynosoma) based on nuclear and mitochondrial data: evidence for a misleading mitochondrial gene tree. Molecular Phylogenetics and Evolution, 39(3):628-644.

Lee, M.S.Y., Oliver, P.M. and Hutchinson, M.N. 2009. Phylogenetic uncertainty and molecular clock calibrations: a case study of legless lizards (Pygopodidae, Gekkota). Molecular Phylogenetics and Evolution, 50(3):661-666.

Lee, M.S. and Skinner, A. 2011. Testing fossil calibrations for vertebrate molecular trees. Zoologica Scripta, 40(5):538-543.

Lima, M.G., and Silva-Júnior, J.D.S., Černý, D., Buckner, J.C., Aleixo, A., Chang, J., Zheng, J., Alfaro, M.E., Martins, A., Di Fiore, A. and Boubli, J.P. 2018. A phylogenomic perspective on the robust capuchin monkey (Sapajus) radiation: First evidence for extensive population admixture across South America. Molecular Phylogenetics and Evolution, 124:137-150.

Macaya, E.C. and Zuccarello, G.C. 2010. Genetic structure of the giant kelp Macrocystis pyrifera along the southeastern Pacific. Marine Ecology Progress Series, 420:103-112.

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

Margulies, E.H., Chen, C.W. and Green, E.D. 2006. Differences between pair-wise and multi-sequence alignment methods affect vertebrate genome comparisons. Trends in Genetics, 22(4):187-193.

Matzke, N.J. 2013a. BioGeoBEARS: BioGeography with Bayesian (and Likelihood) Evolutionary Analysis in R Scripts. Thesis, University of California.

Matzke, N.J. 2013b. Probabilistic historical biogeography: new models for founder-event speciation, imperfect detection, and fossils allow improved accuracy and model-testing. Frontiers of Biogeography, 5(4):242-248.

Matzke, N.J. 2014. Model selection in historical biogeography reveals that founder-event speciation is a crucial process in Island Clades. Systematic Biology 63:951-970.

Maund, J.G., Rex, D.C., Le Roex, A.P. and Reid, D.L. 1988. Volcanism on Gough Island: a revised stratigraphy. Geological Magazine, 125(2):175-181.

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

McKitrick, M.C. 1991. Phylogenetic analysis of avian hindlimb musculature. Miscellaneous Publications Museum of Zoology University of Michigan, 179:1-89.

Meister, W. 1962. Histological structure of the long bones of penguins. The Anatomical Record, 143(4):377-387.

Miller, M.A., Pfeiffer, W. and Schwartz, T. 2010, November. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In 2010 gateway computing environments workshop (GCE), 1-8.

Morris, J.L., Puttick, M.N., Clark, J.W., Edwards, D., Kenrick, P., Pressel, S., Wellman, C.H., Yang, Z., Schneider, H. and Donoghue, P.C. 2018. The timescale of early land plant evolution. Proceedings of the National Academy of Sciences, 115(10):E2274-E2283.

Moyle, R.G., Oliveros, C.H., Andersen, M.J., Hosner, P.A., Benz, B.W., Manthey, J.D., Travers, S.L., Brown, R.M. and Faircloth, B.C. 2016. Tectonic collision and uplift of Wallacea triggered the global songbird radiation. Nature Communications, 7:12709.

Munro, K.J. and Burg, T.M. 2017. A review of historical and contemporary processes affecting population genetic structure of Southern Ocean seabirds. Emu-Austral Ornithology, 117(1):4-18.

Nakamura, T., Yamada, K.D., Tomii, K. and Katoh, K. 2018. Parallelization of MAFFT for large-scale multiple sequence alignments. Bioinformatics, 34(14):2490-2492.

Nery, M.F., González, D.J., Hoffmann, F.G. and Opazo, J.C. 2012. Resolution of the laurasiatherian phylogeny: evidence from genomic data. Molecular Phylogenetics and Evolution, 64(3):685-689.

Nikula, R., Fraser, C.I., Spencer, H.G. and Waters, J.M. 2010. Circumpolar dispersal by rafting in two subantarctic kelp-dwelling crustaceans. Marine Ecology Progress Series, 405:221-230.

Nguyen, L.T., Schmidt, H.A., von Haeseler, A. and Minh, B.Q. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Molecular Biology and Evolution, 32(1):268-274.

O'Hara, R.J. 1989. An estimate of the phylogeny of the living penguins (Aves: Spheniscidae). American Zoologist, 29(4):11A.

Oliveros, C.H., Field, D.J., Ksepka, D.T., Barker, F.K., Aleixo, A., Andersen, M.J., Alström, P., Benz, B.W., Braun, E.L., Braun, M.J. and Bravo, G.A. 2019. Earth history and the passerine superradiation. Proceedings of the National Academy of Sciences, 201813206.

Parent, C.E., Caccone, A. and Petren, K. 2008. Colonization and diversification of Galápagos terrestrial fauna: a phylogenetic and biogeographical synthesis. Philosophical Transactions of the Royal Society B: Biological Sciences, 363(1508):3347-3361.

Poe, S., Nieto-Montes de Oca, A., Torres-Carvajal, O., De Queiroz, K., Velasco, J.A., Truett, B., Gray, L.N., Ryan, M.J., Köhler, G., Ayala-Varela, F. and Latella, I. 2017. A phylogenetic, biogeographic, and taxonomic study of all extant species of Anolis (Squamata; Iguanidae). Systematic Biology, 66(5):663-697.

Poulin, E., González-Wevar, C., Díaz, A., Gérard, K. and Hüne, M. 2014. Divergence between Antarctic and South American marine invertebrates: What molecular biology

tells us about Scotia Arc geodynamics and the intensification of the Antarctic Circumpolar Current. Global and Planetary Change, 123:392-399.

Prestvik, T. and Winsnes, T.S. 1981. Geology of Bouvetøya, South Atlantic. Skr Norsk Polarinst, 175:41-68.

Raikow, R.J., Bicanovsky, L. and Bledsoe, A.H. 1988. Forelimb joint mobility and the evolution of wing-propelled diving in birds. The Auk, 446-451.

Rambaut, A., Drummond, A.J., Xie, D., Baele, G. and Suchard, M.A. 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Systematic Biology, 67(5):901-904.

Ramos, B., González-Acuña, D., Loyola, D.E., Johnson, W.E., Parker, P.G., Massaro, M., Dantas, G.P., Miranda, M.D. and Vianna, J.A. 2018. Landscape genomics: natural selection drives the evolution of mitogenome in penguins. BMC Genomics, 19(1):53.

Ree, R. H., and Smith, S. A. 2008. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. Systematic Biology, 57(1):4-14.

Rogers, A.D., Morley, S., Fitzcharles, E., Jarvis, K. and Belchier, M. 2006. Genetic structure of Patagonian toothfish (*Dissostichus eleginoides*) populations on the Patagonian Shelf and Atlantic and western Indian Ocean Sectors of the Southern Ocean. Marine Biology, 149(4):915-924.

Ronquist, F. 1997. Dispersal-vicariance analysis: a new approach to the quantification of historical biogeography. Systematic Biology, 46(1):195-203.

Sandelin, A., Bailey, P., Bruce, S., Engström, P.G., Klos, J.M., Wasserman, W.W., Ericson, J. and Lenhard, B. 2004. Arrays of ultraconserved non-coding regions span the loci of key developmental genes in vertebrate genomes. BMC Genomics, 5(1):99.

Sayyari, E. and Mirarab, S. 2016. Fast coalescent-based computation of local branch support from quartet frequencies. Molecular Biology and Evolution, 33(7):1654-1668.

Scally, A., Dutheil, J.Y., Hillier, L.W., Jordan, G.E., Goodhead, I., Herrero, J., Hobolth, A., Lappalainen, T., Mailund, T., Marques-Bonet, T. and McCarthy, S. 2012. Insights into hominid evolution from the gorilla genome sequence. Nature, 483(7388):169.

Schreiweis, D.O. 1982. A comparative study of the appendicular musculature of penguins (Aves, Sphenisciformes). Smithsonian Contributions to Zoology, 341:1–46.

Shaw, P.W., Arkhipkin, A.I. and Al-Khairulla, H. 2004. Genetic structuring of Patagonian toothfish populations in the Southwest Atlantic Ocean: the effect of the Antarctic Polar Front and deep-water troughs as barriers to genetic exchange. Molecular Ecology, 13(11):3293-3303.

Shevenell, A.E., Kennett, J.P. and Lea, D.W. 2004. Middle Miocene southern ocean cooling and Antarctic cryosphere expansion. Science, 305(5691):1766-1770.

Shimodaira, H. and Hasegawa, M. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Molecular Biology and Evolution, 16(8):1114-1114.

Shimodaira, H. 2002. An approximately unbiased test of phylogenetic tree selection. Systematic biology, 51(3):492-508.

Siepel, A., Bejerano, G., Pedersen, J.S., Hinrichs, A.S., Hou, M., Rosenbloom, K., Clawson, H., Spieth, J., Hillier, L.W., Richards, S. and Weinstock, G.M. 2005. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Research, 15(8):1034-1050.

Sivak, J.G. 1976. The role of a flat cornea in the amphibious behavior of the blackfoot penguin (*Spheniscus demersus*). Canadian Journal of Zoology, 54(8):1341-1345.

Sivak, J.G. and Millodot, M. 1977. Optical performance of the penguin eye in air and water. Journal of Comparative Physiology, 119(3):241-247.

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

Smith, B.T., Mauck, W.M., Benz, B. and Andersen, M.J. 2018. Uneven missing data skews phylogenomic relationships within the lories and lorikeets. BioRxiv, 398297.

Soubrier, J., Steel, M., Lee, M.S., Der Sarkissian, C., Guindon, S., Ho, S.Y. and Cooper, A. 2012. The influence of rate heterogeneity among sites on the time dependence of molecular rates. Molecular Biology and Evolution, 29(11):3345-3358.

Stadler T. 2010. Sampling-through-time in birth-death trees. Journal of Theoretical Biology, 267(3):396-404.

Starrett, J., Derkarabetian, S., Hedin, M., Bryson Jr, R.W., McCormack, J.E. and Faircloth, B.C. 2017. High phylogenetic utility of an ultraconserved element probe set designed for Arachnida. Molecular Ecology Resources, 17(4):812-823.

Stephen, S., Pheasant, M., Makunin, I.V. and Mattick, J.S. 2008. Large-scale appearance of ultraconserved elements in tetrapod genomes and slowdown of the molecular clock. Molecular Biology and Evolution, 25(2):402-408.

Streicher, J.W., Schulte, J.A. and Wiens, J.J. 2015. 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.

Strimmer, K. and Rambaut, A. 2002. Inferring confidence sets of possibly misspecified gene trees. Proceedings of the Royal Society of London. Series B: Biological Sciences, 269(1487):137-142.

Subramanian, S., Beans-Picón, G., Swaminathan, S.K., Millar, C.D. and Lambert, D.M. 2013. Evidence for a recent origin of penguins. Biology Letters, 9(6):20130748.

Sun, K., Meiklejohn, K.A., Faircloth, B.C., Glenn, T.C., Braun, E.L. and Kimball, R.T. 2014. The evolution of peafowl and other taxa with ocelli (eyespots): a phylogenomic approach. Proceedings of the Royal Society of London B: Biological Sciences, 281(1790):20140823.

Talavera, G. and Vila, R. 2011. What is the phylogenetic signal limit from mitogenomes? The reconciliation between mitochondrial and nuclear data in the Insecta class phylogeny. BMC Evolutionary Biology, 11(1):315.

Taylor, J.R. 1986. Thermal insulation of the down and feathers of pygoscelid penguin chicks and the unique properties of penguin feathers. The Auk, 103(1):160-168.

Templeton, A.R. 2008. The reality and importance of founder speciation in evolution. Bioessays, 30(5):470-479.

Van Dam, M.H., Lam, A.W., Sagata, K., Gewa, B., Laufa, R., Balke, M., Faircloth, B.C. and Riedel, A. 2017. Ultraconserved elements (UCEs) resolve the phylogeny of Australasian smurf-weevils. PloS One, 12(11):e0188044.

Van Dam, M.H., Trautwein, M., Spicer, G.S. and Esposito, L. 2019. Advancing mite phylogenomics: Designing ultraconserved elements for Acari phylogeny. Molecular Ecology Resources, 19(2):465-475.

Vianna, J.A., Noll, D., Dantas, G.P., Petry, M.V., Barbosa, A., González-Acuña, D., Le Bohec, C., Bonadonna, F. and Poulin, E. 2017. Marked phylogeographic structure of Gentoo penguin reveals an ongoing diversification process along the Southern Ocean. Molecular Phylogenetics and Evolution, 107:486-498.

Wagner, C.E., Keller, I., Wittwer, S., Selz, O.M., Mwaiko, S., Greuter, L., Sivasundar, A. and Seehausen, O. 2013. Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. Molecular Ecology, 22(3):787-798.

Walsh, S.A. and Suárez, M.E. 2006. New penguin remains from the Pliocene of northern Chile. Historical Biology, 18(2):119-130.

Wang, K., Lenstra, J.A., Liu, L., Hu, Q., Ma, T., Qiu, Q. and Liu, J. 2018. Incomplete lineage sorting rather than hybridization explains the inconsistent phylogeny of the wisent. Communications Biology, 1(1):169.

Warnock, R.C., Parham, J.F., Joyce, W.G., Lyson, T.R. and Donoghue, P.C. 2015. Calibration uncertainty in molecular dating analyses: there is no substitute for the prior evaluation of time priors. Proceedings of the Royal Society B: Biological Sciences, 282(1798):20141013.

Weisrock, D.W., Harmon, L.J. and Larson, A. 2005. Resolving deep phylogenetic relationships in salamanders: analyses of mitochondrial and nuclear genomic data. Systematic Biology, 54(5):758-777.

Wen, D., Yu, Y., Hahn, M.W. and Nakhleh, L. 2016. Reticulate evolutionary history and extensive introgression in mosquito species revealed by phylogenetic network analysis. Molecular Ecology, 25(11):2361-2372.

White, N.D., Mitter, C. and Braun, M.J. 2017. Ultraconserved elements resolve the phylogeny of potoos (Aves: Nyctibiidae). Journal of Avian Biology, 48(6):872-880.

Wiens, J.J. and Tiu, J. 2012. Highly incomplete taxa can rescue phylogenetic analyses from the negative impacts of limited taxon sampling. PloS One, 7(8):e42925.

Winker, K., Glenn, T.C. and Faircloth, B.C. 2018. Ultraconserved elements (UCEs) illuminate the population genomics of a recent, high-latitude avian speciation event. PeerJ, 6:e5735.

Yang, Z. 1995. A space-time process model for the evolution of DNA sequences. Genetics, 139(2):993-1005.

Zachos, J., Pagani, M., Sloan, L., Thomas, E. and Billups, K. 2001. Trends, rhythms, and aberrations in global climate 65 Ma to present. Science, 292(5517):686-693.

Zhang, G., Li, C., Li, Q., Li, B., Larkin, D.M., Lee, C., Storz, J.F., Antunes, A., Greenwold, M.J., Meredith, R.W. and Ödeen, A. 2014. Comparative genomics reveals insights into avian genome evolution and adaptation. Science, 346(6215):1311-1320.

Zhang, C., Rabiee, M., Sayyari, E. and Mirarab, S. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinformatics, 19(6):153.

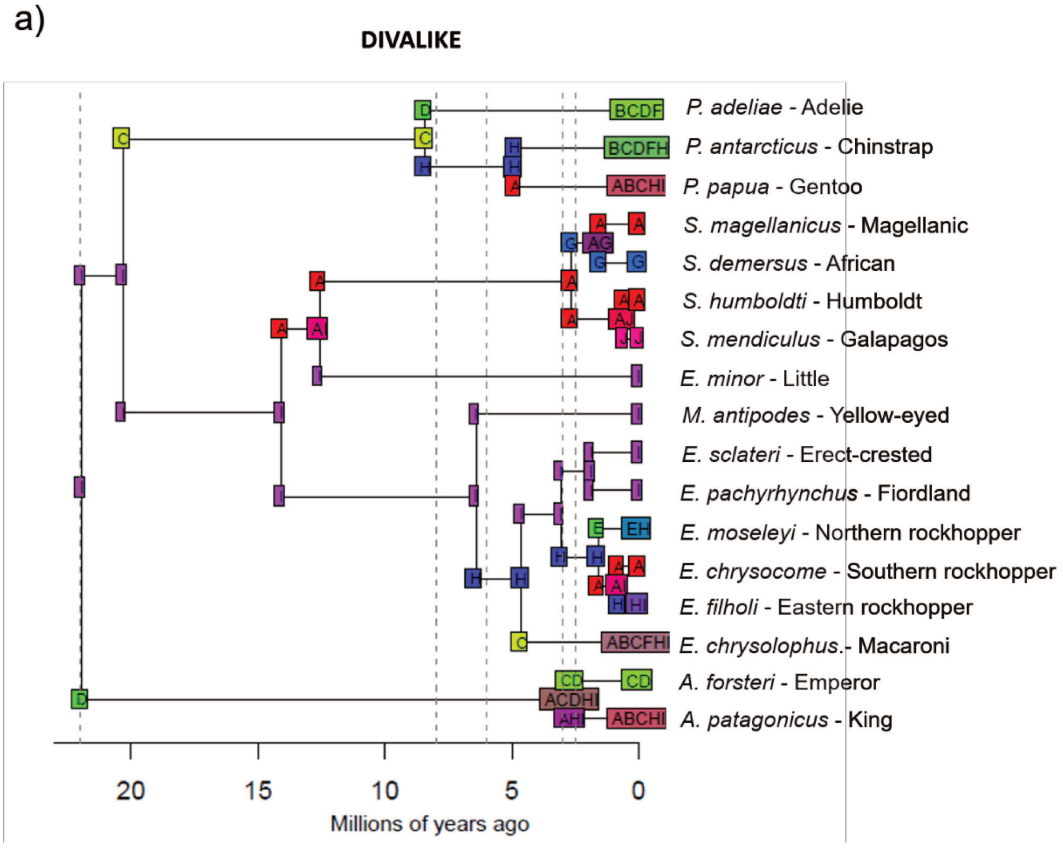
APPENDIX

Table 6. Divergence time estimates in 95% high posterior density (HPD) intervals. First column represents the description of the node, second column the 95% HPDs of our study, and remaining columns represent estimates from previous studies in literature which recovered Miocene estimates. *S: Spheniscus; Ea: Eudyptula; Es: Eudyptes; M: Megadyptes.*

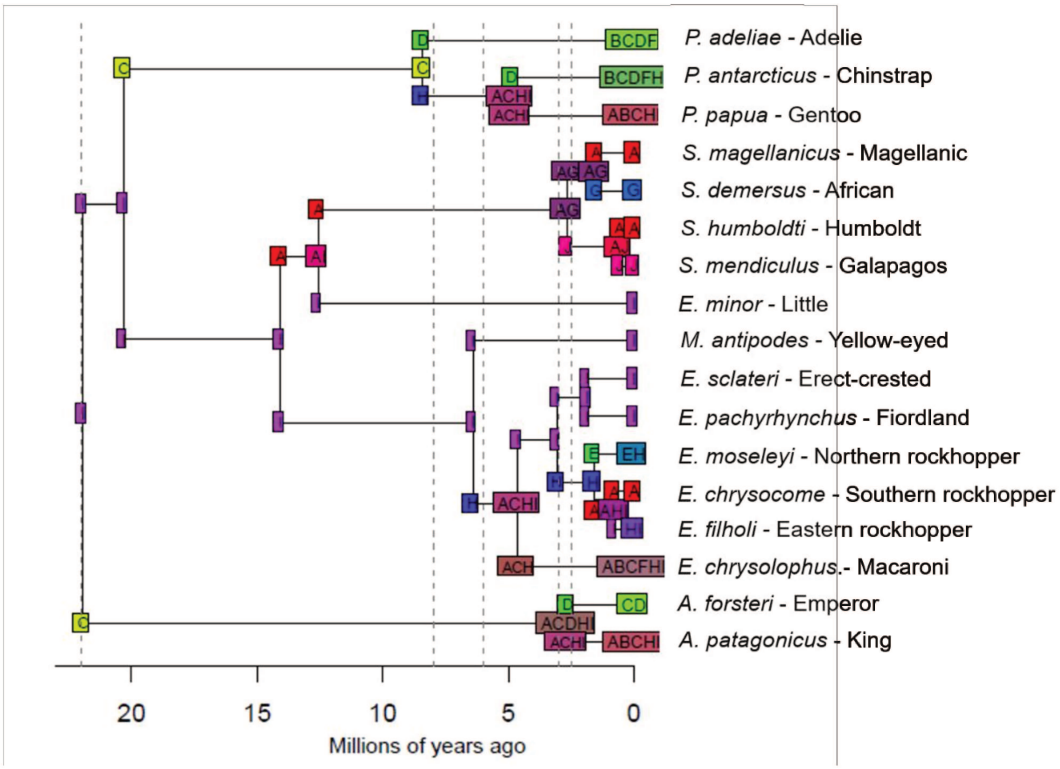
		Literature		
Nodes	**Our study**	**Cole et al. (2019)**	**Gavryushkina et al. (2017)**	**Subramanian et al. (2013)**
Spheniscidae	19.0 - 25.1	22 - 12.5	9.9 - 15.7	17.0 - 23.8
Pygoscelis/ S+Ea+Es+M	17.3 - 22.8	-	-	-
S+Ea/ Es+M	11.2 - 16.7	17.1 - 10.6	7.9 - 12.9	13.6 - 18.2
Spheniscus/ Eudyptula	9.3 - 16.1	13.9 - 9.3	7.1 - 11.9	11.0-15.0
Eudyptes/ Megadyptes	4.8 - 8.1	9.3 - 4.7	3.2 - 6.8	10.0 - 12.2
A. forsteri/ A. patagonicus	1.0 - 4.4	5.1 - 2.0	0.7 - 2.4	-
P. adeliae/ P. antarcticus + P. papua	6.3 - 11.2	10.4 - 4.9	3.3 - 9.0	5.0 - 9.1
P. antarcticus/ P. papua	2.4 - 7.8	8.2 - 3.5	1.6 - 5.2	-
P. papua Crozet/ others	0.7 - 2.0	-	-	-
P. papua Kerguelen/ Atlantic	0.2 - 1.6	-	-	-
P. papua Falkland/ Antarctica	0.008 - 0.8	-	-	-
S. demersus + S. mendiculus/ S. magellanic+S. demersus	1.6 - 3.7	3.0 - 1.4	0.9 - 2.3	-
S. mendiculus/ S. humboldti	0.007 - 1.3	1.6 - 0.6	0.6 - 1.8	-
S. magellanicus/ S. demersus	0.04 - 2.6	2.2 - 0.9	0.5 - 1.6	0.9 - 3.1
Eudyptes/ Eudyptes	2.9 - 6.7	5.3 - 2.7	1.4 - 3.3	5.3 - 2.7

E. chrysolophus Elephant/	0.1 - 1.2	-	-	-
E. schlegeli/ E. chrysolophus Marion	0.009 - 0.8	-	0.1 - 0.7	-
Rockhoppers/ *E. pachyrhynchus + E. sclateri*	0.6 - 4.7	4.9 - 2.4	-	4.9 - 2.4
E. pachyrhynchus/ E. sclateri	0.1 - 2.9	3.5 - 1.7	1.0 -. 2.6	3.5 - 1.7
E. moseleyi/ E. filholi + E. chrysocome	0.3 - 2.6	2.7 - 1.2	0.3 - 1.2	2.7 - 1.2
E. filholi/ E. chrysocome	0.02 - 1.4	1.3 - 0.5	0.09 - 0.5	1.3 - 0.5

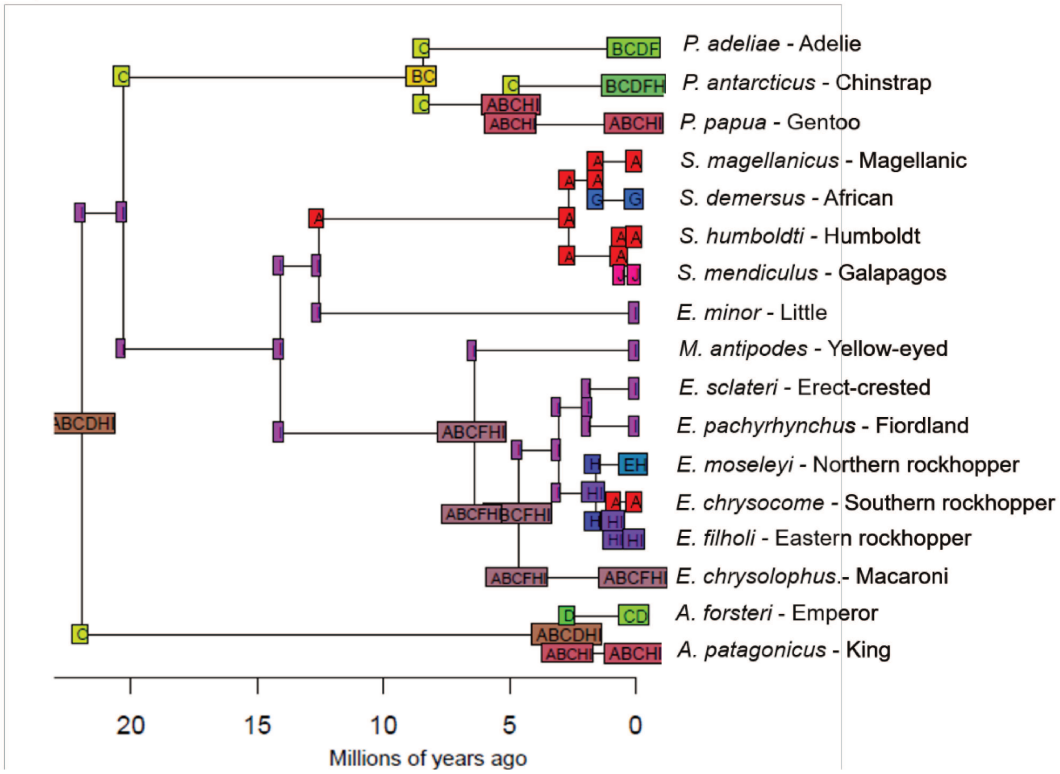
a)



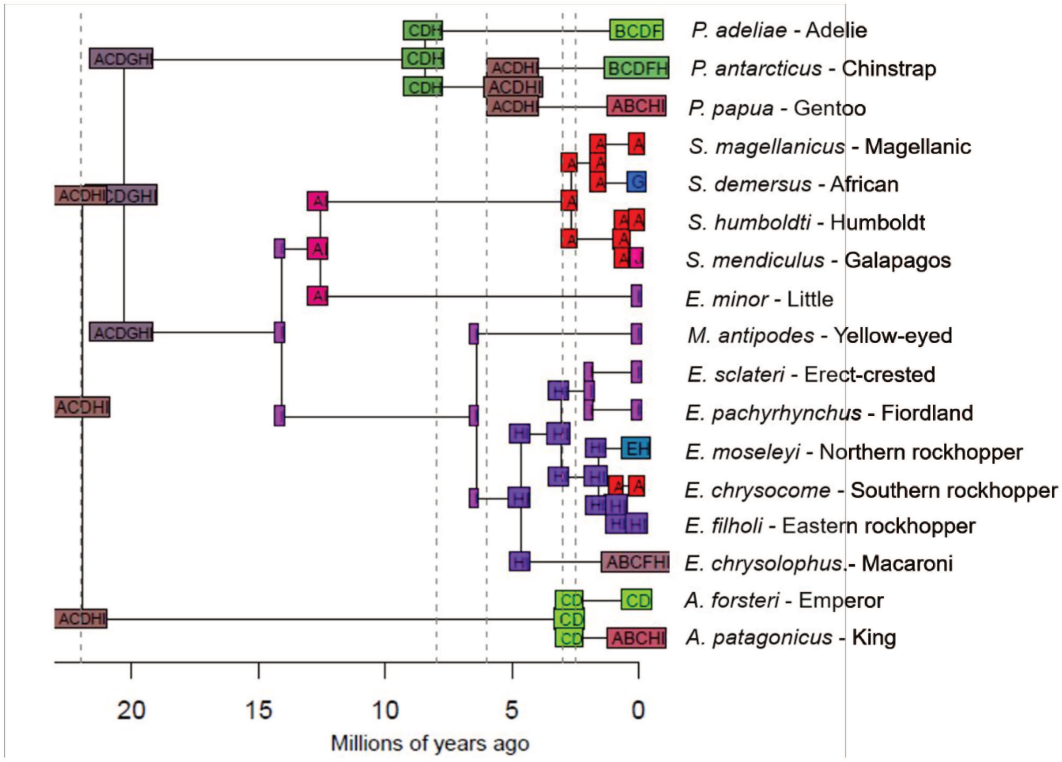
b)


DEC
P. adeliae - Adelie
P. antarcticus - Chinstrap
P. papua - Gentoo
S. magellanicus - Magellanic
S. demersus - African
S. humboldti - Humboldt
S. mendiculus - Galapagos
E. minor - Little
M. antipodes - Yellow-eyed
E. sclateri - Erect-crested
E. pachyrhynchus - Fiordland
E. moseleyi - Northern rockhopper
E. chrysocome - Southern rockhopper
E. filholi - Eastern rockhopper
E. chrysolophus.- Macaroni
A. forsteri - Emperor
A. patagonicus - King
20
15
10
5
0
Millions of years ago

c)


DEC+J
P. adeliae - Adelie
P. antarcticus - Chinstrap
P. papua - Gentoo
S. magellanicus - Magellanic
S. demersus - African
S. humboldti - Humboldt
S. mendiculus - Galapagos
E. minor - Little
M. antipodes - Yellow-eyed
E. sclateri - Erect-crested
E. pachyrhynchus - Fiordland
E. moseleyi - Northern rockhopper
E. chrysocome - Southern rockhopper
E. filholi - Eastern rockhopper
E. chrysolophus.- Macaroni
A. forsteri - Emperor
A. patagonicus - King
20
15
10
5
0
Millions of years ago

d)


BAYAREALIKE
P. adeliae - Adelie
P. antarcticus - Chinstrap
P. papua - Gentoo
S. magellanicus - Magellanic
S. demersus - African
S. humboldti - Humboldt
S. mendiculus - Galapagos
E. minor - Little
M. antipodes - Yellow-eyed
E. sclateri - Erect-crested
E. pachyrhynchus - Fiordland
E. moseleyi - Northern rockhopper
E. chrysocome - Southern rockhopper
E. filholi - Eastern rockhopper
E. chrysolophus.- Macaroni
A. forsteri - Emperor
A. patagonicus - King
20
15
10
5
0
Millions of years ago

e)

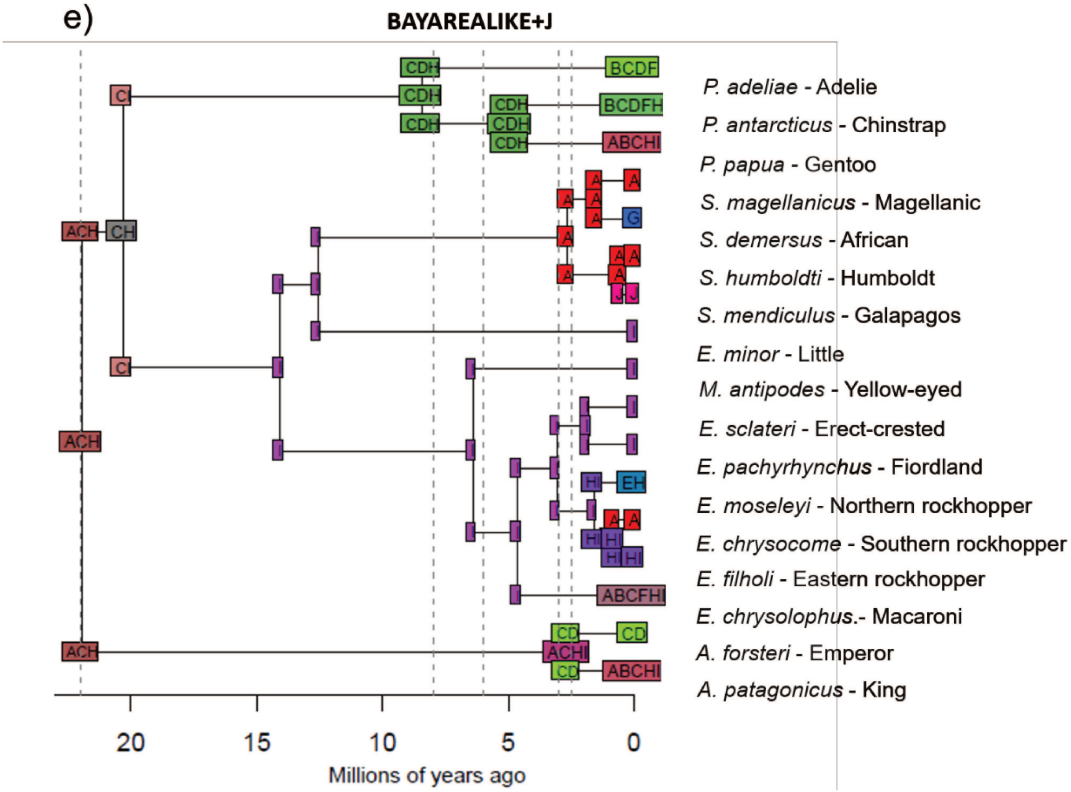

BAYAREALIKE+J
P. adeliae - Adelie
P. antarcticus - Chinstrap
P. papua - Gentoo
S. magellanicus - Magellanic
S. demersus - African
S. humboldti - Humboldt
S. mendiculus - Galapagos
E. minor - Little
M. antipodes - Yellow-eyed
E. sclateri - Erect-crested
E. pachyrhynchus - Fiordland
E. moseleyi - Northern rockhopper
E. chrysocome - Southern rockhopper
E. filholi - Eastern rockhopper
E. chrysolophus.- Macaroni
A. forsteri - Emperor
A. patagonicus - King
20
15
10
5
0
Millions of years ago

Figure 7. Ancestral range estimation results under the a) DIVALIKE; b) DEC; c) DEC+J; d) BAYAREALIKE; e) BAYAREALIKE+J models. Geographic locations of the areas inputted in BioGeoBEARS: A) South American coasts and Falkland islands; B) Scotia Arc islands; C) Antarctic Peninsula; D) Antarctic continent; E) Tristan da Cunha and Gough islands; F) Bouvet islands; G) South African coasts; H) Indian Ocean islands; I) Australia/New Zealand coasts and nearby islands; and J) Galapagos islands. Ranges on the tips of the tree represent the current distribution of the species, ranges on the nodes represent the most likely distribution of the clade ancestral, and ranges on the vertices represent the range distribution right after cladogenesis events.

ATTACHMENT I

COORDENADORIA DE PÓS-GRADUAÇÃO
INSTITUTO DE BIOLOGIA
Universidade Estadual de Campinas
Caixa Postal 6109. 13083-970, Campinas, SP, Brasil
Fone (19) 3521-6378. email: cpgib@unicamp.br

DECLARAÇÃO

Em observância ao **§5° do Artigo 1° da Informação CCPG-UNICAMP/001/15**, referente a Bioética e Biossegurança, declaro que o conteúdo de minha Dissertação de Mestrado, intitulada "HISTÓRIA EVOLUTIVA E BIOGEOGRÁFICA DOS PINGUINS VIVENTES (AVES: SPHENISCIDAE) USANDO DADOS GENÔMICOS", desenvolvida no Programa de Pós-Graduação em Genética e Biologia Molecular do Instituto de Biologia da Unicamp, não versa sobre pesquisa envolvendo seres humanos, animais ou temas afetos a Biossegurança.

Assinatura: Flávia A N Fernandes
Nome do(a) aluno(a): Flávia Akemi Nitta Fernandes

Assinatura: _____
Nome do(a) orientador(a): Mariana Freitas Nery

ATTACHMENT II

Declaração

As cópias de artigos de minha autoria ou de minha co-autoria, já publicados ou submetidos para publicação em revistas científicas ou anais de congressos sujeitos a arbitragem, que constam da minha Dissertação/Tese de Mestrado/Doutorado, intitulada **HISTÓRIA EVOLUTIVA E BIOGEOGRÁFICA DOS PINGUINS VIVENTES (AVES: SPHENISCIDAE) USANDO DADOS GENÔMICOS**, não infringem os dispositivos da Lei n.° 9.610/98, nem o direito autoral de qualquer editora.

Campinas, 25 de junho de 2019

Assinatura : Flávia A N Fernandes
Nome do(a) autor(a): **Flávia Akemi Nitta Fernandes**
RG n.° 38796240-2

Assinatura : _____
Nome do(a) orientador(a): **Mariana Freitas Nery**
RG n.° 34544823-6