

# Mitogenomes uncover extinct penguin taxa and reveal island formation as a key driver of speciation

Theresa L. Cole<sup>a,b</sup>, Daniel T. Ksepka<sup>c,1</sup>, Kieren J. Mitchell<sup>d,1</sup>, Alan J. D. Tennyson<sup>e,1</sup>, Daniel B. Thomas<sup>f,1</sup>, Hailin Pan<sup>g,h,i</sup>, Guojie Zhang<sup>g,h,i</sup>, Nicolas J. Rawlence<sup>a</sup>, Jamie R. Wood<sup>b</sup>, Pere Bover<sup>d,j</sup>, Juan L. Bouzat<sup>k</sup>, Alan Cooper<sup>d</sup>, Steven Fiddaman<sup>l</sup>, Tom Hart<sup>l</sup>, Gary Miller<sup>m,n</sup>, Peter G. Ryan<sup>o</sup>, Lara D. Shepherd<sup>e</sup>, Janet M. Wilmshurst<sup>b,p</sup>, Jonathan M. Waters<sup>a</sup>

<sup>a</sup> Department of Zoology, University of Otago, PO Box 56, Dunedin 9054, New Zealand.

<sup>b</sup> Manaaki Whenua Landcare Research, PO Box 69040, Lincoln, Canterbury 7640, New Zealand.

<sup>c</sup> Bruce Museum, Greenwich, Connecticut, 06830, USA.

<sup>d</sup> Australian Centre for Ancient DNA, School of Biological Sciences, University of Adelaide, Adelaide, SA, 5005, Australia.

<sup>e</sup> Museum of New Zealand Te Papa Tongarewa, PO Box 467, Wellington 6011, New Zealand.

<sup>f</sup> Institute of Natural and Mathematical Sciences, Massey University, Auckland, New Zealand.

<sup>g</sup> State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, 650223, Kunming, China.

<sup>h</sup> China National Genebank, BGI-Shenzhen, 518083, Shenzhen, Guangdong, China.

<sup>i</sup> Centre for Social Evolution, Department of Biology, Universitetsparken 15, University of Copenhagen, DK-2100, Copenhagen, Denmark.

<sup>j</sup> ARAID Foundation. IUCA-Grupo Aragosaurus. Universidad de Zaragoza. C/ Pedro Cerbuna, 12. 50009 Zaragoza, Spain.

<sup>k</sup> Bowling Green State University, Department of Biological Sciences, Bowling Green, Ohio, 43402, USA.

<sup>l</sup> Department of Zoology, University of Oxford, South Parks Road, Oxford, OX13PS, UK.

<sup>m</sup> Division of Pathology and Laboratory Medicine, University of Western Australia, Crawley, Western Australia, 6009, Australia.

<sup>n</sup> Institute for Marine and Antarctic Studies, University of Tasmania, Hobart, Tasmania, 7001, Australia.

<sup>o</sup> FitzPatrick Institute of African Ornithology, DST-NRF Centre of Excellence, University of Cape Town, Rondebosch 7701, South Africa.

<sup>p</sup> School of Environment, University of Auckland, Auckland, New Zealand.

<sup>1</sup> DTK, KJM, AJDT and DBT contributed equally to this work

Corresponding author: Theresa L. Cole, [tesscole1990@gmail.com](mailto:tesscole1990@gmail.com)

**Abstract**

The emergence of islands has been linked to spectacular radiations of diverse organisms. Although penguins spend much of their lives at sea, they rely on land for nesting, and a high proportion of extant species are endemic to geologically-young islands. Islands may thus have been crucial to the evolutionary diversification of penguins. We test this hypothesis using a fossil-calibrated phylogeny of mitochondrial genomes (mitogenomes) from all extant and recently extinct penguin taxa. Our temporal analysis demonstrates that numerous recent island-endemic penguin taxa diverged following the formation of their islands during the Plio-Pleistocene, including the Galápagos (Galápagos Islands), northern rockhopper (Gough Island), erect-crested (Antipodes Islands), Snares crested (Snares) and royal (Macquarie Island) penguins. Our analysis also reveals two new recently extinct island-endemic penguin taxa from New Zealand's Chatham Islands: *Eudyptes warhami* sp. nov. and a dwarf subspecies of the yellow-eyed penguin, *Megadyptes antipodes richdalei* ssp. nov. *Eudyptes warhami* diverged from the Antipodes Islands erect-crested penguin between 1.1 million years ago (Ma) and 2.5 Ma, shortly after the emergence of the Chatham Islands (~3 Ma). This new finding of recently-evolved taxa on this young archipelago provides further evidence that the radiation of penguins over the last 5 Ma has been linked to island emergence. Mitogenomic analyses of all penguin species, and the discovery of two new extinct penguin taxa, highlight the importance of island formation in the diversification of penguins, as well as the extent to which anthropogenic extinctions have affected island-endemic taxa across the Southern Hemisphere's isolated archipelagos.

**Keywords:** Sphenisciformes, ancient DNA, fossil calibrations, *Eudyptes warhami*, *Megadyptes antipodes richdalei*

## Introduction

Biologists have long considered oceanic islands as natural ‘laboratories’ for evolutionary studies (Darwin 1859), with archipelago formation underpinning dramatic biological radiations in many remote regions of the globe (Shaw and Gillespie 2016). In particular, numerous studies have highlighted island isolation as a crucial prerequisite for species formation and adaptive radiation (Darwin 1859; Cowie and Holland 2006; Losos and Ricklefs 2009; Bacon et al. 2012). Soon after emergence, islands (whether volcanic or tectonic in origin [Paulay 1994]) can be rapidly colonised by diverse arrays of dispersing taxa (Fleisher et al. 1998; Gathorne-Hardy and Jones 2000; Mendelson and Shaw 2005; Gillespie et al. 2012), presenting unique opportunities for local adaptation and diversification of species (Waters et al. 2013). The resultant island-endemic taxa can also be particularly prone to extinction (Shaw and Gillespie 2006; Cowie and Holland 2006; Wood et al. 2017).

Penguins (Sphenisciformes) are iconic flightless marine birds that inhabit all major landmasses and many islands in the Southern Hemisphere (Fig. 1 [a]). Approximately 20 extant species are recognized, with some debate over species boundaries between recently-diverged populations. The group has a rich fossil record extending back >60Ma (Slack et al. 2006), with over 50 extinct species documented (Ksepka et al. 2012). Several phylogenetic studies have attempted to pinpoint the timing, and thereby the drivers, of penguin diversification (Baker et al. 2006; Ksepka et al. 2006; Subramanian et al. 2013; Gavryushkina et al. 2017; Frugone et al. 2018). Previous studies have invoked circumpolar ocean currents and/or Antarctic cooling as key drivers of penguin evolution and biogeography (Baker et al. 2006; Frugone et al. 2018). However, one third of all extant penguin species are endemic to geologically young islands (<5 Ma; Maund et al. 1988; Gamble and Morris 1989; Adamson et al. 1996; Sinton et al. 2018) suggesting that founder speciation may also have played an important role in recent penguin cladogenesis. An alternative explanation is that island endemic penguins represent relictual populations of formerly more widespread species.

Here, we test these competing hypotheses using 41 near-complete mitochondrial genomes (mitogenomes), representing all extant and recently extinct penguin taxa. By using well-justified fossil calibrations, we resolve the timing and mechanisms of modern penguin diversification. We demonstrate that many penguin divergences correlate with the formation of islands, providing a new model for understanding penguin evolution. Furthermore, we describe two new recently extinct penguin taxa from the Chatham Islands, demonstrating that

while islands have been key in many recent penguin speciation events, the resulting restricted distributions have also made such lineages particularly susceptible to anthropogenic extinction.

## Results

### *Genetic evidence for two extinct penguin lineages from the Chatham Islands*

We analysed 65 bones (Table S1) from the Chatham Islands to test for the existence of an extinct endemic crested penguin (*Eudyptes*) species, as proposed by Tennyson and Millener (1994) based on morphological evidence. Most were poorly preserved, but we obtained partial cytochrome oxidase subunit I (COI) sequences from 22, and partial control region (CR) sequences from eight. Phylogenetic analyses of these data identified erect-crested penguin (*Eudyptes sclateri*) and two other distinct genetic lineages: one corresponding to *Eudyptes* clade X (Cole et al. 2019), and another within the yellow-eyed penguin genus *Megadyptes* (Fig. S1–S4). The latter discovery was unexpected, as the bones had appeared too small to belong to *Megadyptes* and thus had originally been identified as *Eudyptes*.

### *Mitochondrial phylogenetic analysis*

Phylogenetic analysis of near-complete mitogenomes recovered a clade grouping the two largest-bodied and most polar-adapted penguin genera, *Aptenodytes* and *Pygoscelis*, as sister to all other extant penguins (Fig. S5–S9). This basal split was also found in two recent studies (Subramanian et al. 2013; Gavryushkina et al. 2017), whereas others have recovered *Aptenodytes* as an independent lineage, sister to all other extant penguins (Bertelli and Gianni 2005; Baker et al. 2006; Ksepka et al. 2006). Our phylogeny agrees with previous studies in recovering *Spheniscus* + *Eudyptula* and *Eudyptes* + *Megadyptes* clades (Bertelli and Gianni 2005; Baker et al. 2006; Ksepka et al. 2006). The Chathams *Eudyptes* taxon is distinct, sister to *E. sclateri* (posterior probability: 1.0). The individuals recognised as *Eudyptes* clade X from mainland New Zealand belong to this newly-recognised species. We also found that individuals belonging to the previously unrecognised dwarf Chathams *Megadyptes* taxon form a monophyletic clade (posterior probability: 0.88), though their precise phylogenetic relationship to the other *Megadyptes* lineages could not be confidently resolved (Fig. S5).

### *Pairwise distances and taxonomic status*

To explore genetic divergences and potential taxonomic status of penguin lineages, we compared mitogenomes in a pairwise matrix, excluding positions with missing data (Fig. 1

[b]; Tables S2–S3). The Chathams *Eudyptes* and the erect-crested penguin (*E. sclateri*) are 1.9% divergent, with 268 differential SNPs between the two species. In contrast, we observed just 0.2% divergence between the closely-related (Christidis and Boles 2008) royal penguin (*E. chrysolophus schlegeli*) and macaroni penguin (*E. c. chrysolophus*). Despite clear phenotypic differences (Warham 1975; Shaughnessy 1975; Woehler 1995), it appears that these genetically similar lineages may still be in the earliest stages of diversification, supporting the conclusions of Frugone et al. (2018). Our data reveal more substantial divergences among three rockhopper penguin species (De Dinechin 2009): The southern rockhopper (*E. chrysocome*) and eastern rockhopper (*E. filholi*) penguins are 0.7% divergent, and both show 1.8% divergence relative to the northern rockhopper penguin (*E. moseleyi*). The recently proposed recognition of two little penguin taxa (Grosser et al. 2017) is supported by 2.9% divergence detected between *Eudyptula minor* and *E. novaehollandiae*.

*Megadyptes antipodes* and the new Chathams *Megadyptes* taxon had only 0.1% sequence divergence, and both were 0.3% divergent from *M. waitaha*. These values are substantially smaller than the divergences observed between some other sister species pairs of extant penguins (mean 2.2%, range 0.8–5.2%; Table S3). In contrast, the Chathams *Eudyptes* taxon was 1.9% divergent from its sister species *E. sclateri*, a value exceeding those between sister-pairs of several widely accepted penguin species (e.g. *E. robustus* and the Fiordland crested penguin [*E. pachyrhynchus*] are 0.8% divergent).

#### *Calibrated phylogenetic analysis*

Our divergence estimates of major phylogenetic clades are consistent with the proposed Miocene origin of crown penguins (Subramanian et al. 2013; Gavryushkina et al. 2017), and show that a large proportion of penguin species (16 out of 23 studied taxa) have diverged over the past 2 Ma (Fig. 1 [c]; Table S4). While some divergence time estimates substantially postdate island formation, we observed no instances where an island-endemic penguin diverged from its sister taxon prior to the corresponding island group emergence (see Fig. 1 [c]; Table S4; and Fig. S7–S9 for alternative calibration approaches).

#### **Systematic palaeontology**

Morphological observations (Table S5–S10), combined with our molecular results, support recognition of the Chathams *Eudyptes* (originally proposed by Tennyson and Millener 1994)

as a distinct species and the dwarf Chathams *Megadyptes* as a subspecies of the extant *Megadyptes antipodes* (Fig. 2).

Aves Linnaeus, 1758

Sphenisciformes Sharpe, 1891

*Eudyptes* Vieillot, 1816

*Eudyptes warhami*, n. sp. Cole, Tennyson, Ksepka & Thomas

Holotype. NMNZ S.33007. Skull (Fig. 2 [c1 and c2]).

Etymology. The specific epithet honours John Warham (1919–2010), who carried out pioneering studies on *Eudyptes* penguins.

Type locality. Foredune c.200 m west of Tahatika Creek, Chatham Island. Collected by P.R. Millener, 20 Jan 1993. Age: <7000 years BP; the maximum age of the dunes. An *Anas chathamica* bone from the site dates to  $1529 \pm 57$   $^{14}\text{C}$  years BP (1405–1185 cal BP) (Millener 1999).

Paratype and Stratigraphic context. All bones in the type series are from Chatham Island unless otherwise stated: NMNZ S.24277, left carpometacarpus; NMNZ S.25157, right humerus; NMNZ S.26908, skull; NMNZ S.27259, right coracoid, Chatham Islands; NMNZ S.47917, left coracoid; NMNZ S.30440, mandible; NMNZ S.47921, left tibiotarsus, Mangere Island; CM Av.6816, largely complete skull, Chatham Islands; CM Av.27407, left humerus; CM Av.27867, left humerus (Fig. S10). Most bones in the type series were isolated elements collected from eroded dune surfaces. However, the paratype from Mangere Island was from a soil deposit that contained European-era remains (but had possible rabbit disturbance). No *E. warhami* bones were found in articulated association.

Diagnosis. Characterized by elongate ovoid premaxilla in dorsal view and relatively shallow mandible. Distinguished from *E. chrysocome*, *E. filholi* and *E. moseleyi* by larger size (Table S5). The largest specimens, including the holotype, rival the largest extant *Eudyptes* taxon (*E. chrysolophus schlegeli*). Distinguished from *E. pachyrhynchus* and *E. robustus* by relatively elongate premaxilla. Distinguished from *E. chrysolophus chrysolophus* and *E. chrysolophus schlegeli* by proportionally shallower mandible. Distinguished from *E. sclateri* by more bowed premaxilla (dorsally) and notably shallower mandible. Distinguished from the Pliocene *Eudyptes calauina* by smaller and more slender humerus (max. length 70 mm in *E. warhami* vs. ~81 mm in *E. calauina*).

**Distribution.** Presumably once widespread along coastlines of the Chatham archipelago. The type series includes specimens from northern Chatham Is. (43.71°S) to Mangere Is. (44.27°S). The referred specimen series (see Supplementary Information) indicates that the species ranged westward to the east coast of mainland New Zealand.

**Description.** As typical of larger *Eudyptes* species, the beak of *E. warhami* is more elongate than in smaller congeners. The rostral portion of the upper beak is markedly swollen as in *Eudyptes* and unlike *Megadyptes*. The jugal bar is strongly curved (Fig. 2 [c2]), Fig. S10) as in *Eudyptes*, and more so than in *Megadyptes*. The deep salt gland fossae are bounded by a shelf of bone. The mandibular ramus deepens strongly near the midpoint; a feature observed within *Eudyptes* and *Pygoscelis* penguins that is associated with a preference for planktonic prey (Zusi 1975). Similar to extant *Eudyptes* species, the coracoidal fenestra is completely enclosed by a bridge of bone extending from the procoracoid process. The humerus is relatively wide, lacks a pronounced notch between the head and the dorsal tubercle, and has a posterior trochlear process which projects beyond the ventral border of the shaft.

***Megadyptes* Milne-Edwards, 1880**

***Megadyptes antipodes* (Hombron & Jacquinot, 1841)**

***Megadyptes antipodes richdalei* n. ssp., Tennyson & Cole**

**Holotype.** NMNZ S.26921. Partial skeleton (Fig. 2 [d]; Fig. S11).

**Etymology.** The subspecies epithet honours Lance Richdale (1900–1983), who carried out pioneering studies on *Megadyptes* ecology.

**Type locality.** Foredune east of Maunganui, Chatham Is. Collected by P.R. Millener, 21 Feb 1989. Age: <7000 years BP; the maximum age of the dunes. A *Diaphorapteryx hawkinsi* bone from the site dates to  $1860 \pm 150$  <sup>14</sup>C years BP (1996–1314 cal BP) (Millener 1999).

**Paratype and Stratigraphic context.** All bones in the type series were collected from eroded dune surfaces and were from the Chatham archipelago unless otherwise stated: NMNZ S.47918 right coracoid; NMNZ S.30968, mandible; NMNZ S.47765, premaxilla; NMNZ S.45876, skull Chatham Islands (Fig. 2 [b]); CM Av11287, left humerus; CM (ACAD12997; unregistered) (Fig. S12), left humerus; AM LB12063, proximal left tibiotarsus, Pitt Island.

**Diagnosis.** A *Megadyptes* penguin, smaller than both described *Megadyptes* taxa (*M. antipodes* and *M. waitaha*) (Tables S6–S10). *M. a. richdalei* n. ssp. represents a genetic lineage comprising distinct haplotypes ( $\geq 15$  private mitochondrial SNPs) not detected in other living or extinct *Megadyptes* populations (Fig. S4 and Tables S2–S3).

**Distribution.** Chatham and Pitt Islands (from 43.73°S to 44.23°S). Presumably occurred around all coasts of the Chatham archipelago.

**Description.** *M. a. richdalei* is the smallest *Megadyptes* penguin, being on average smaller than *M. waitaha* (Fig. 2 [b]; Tables S6–S7) (though size distributions overlap; see Table S6). Both recently-extinct taxa are smaller than *M. antipodes*, with almost no overlap between the extant and extinct taxa in bone lengths. The skull closely resembles that of *M. antipodes* and differs from the contemporaneous *E. warhami* in its more slender upper beak and shallower mandible (without pronounced deepening at midpoint). We observed no osteological differentiation between the three *Megadyptes* taxa that could not be accounted for either by size or individual variation (as reflected in *M. antipodes* specimens), suggesting the proposed postcranial differences between *M. antipodes* and *M. waitaha* (Boessenkool et al. 2009a) cannot consistently differentiate these taxa. Thus, we consider the Chatham taxon to represent an instance of isometric dwarfing and recommend the two recently extinct forms be recognized as subspecies of *Megadyptes antipodes*. We follow the New Zealand Bird Checklist Committee (Gill et al. 2010) in defining subspecies using the Diagnostic Species Concept, where it is expected that a subspecies will meet a 75% diagnosable criterion (Amadon 1949; Patten and Unit 2002).

## Discussion

### *Timing of penguin evolution linked to island emergence*

While some studies have suggested that crown penguins began radiating in the Eocene (Baker et al. 2006), our divergence estimates (Fig. 1 [c]; Table S4) indicate a Miocene origin of crown penguins (Subramanian et al. 2013; Gavryushkina et al. 2017). Moreover, our results support the hypothesis that a large proportion of penguins diverged within the past 2 Ma (Gavryushkina et al. 2017). We propose that this diversification pulse was tied to the emergence of islands, which created new opportunities for isolation and speciation.

While island emergence can spawn diverse biological radiations, several studies have detected island endemic lineages substantially pre-dating island formation (Lewin 1985; Fraser et al. 2010). However, our estimated divergence dates for all island-endemic penguin taxa (i.e., those restricted to one island/archipelago) are consistently younger than the islands they inhabit (Table S4). While the ancestral distributions of clades are not always clear, the finding of numerous recently-evolved taxa endemic to geologically-young islands (Fleischer et al. 1998; Mendelson and Shaw 2005) strongly suggests that these lineages evolved *in situ*,



rather than being relicts of formerly widespread species. Our study thus provides clear temporal genetic evidence linking penguin speciation to island formation (Fig. 1 [c]).

Our analysis found that *Eudyptes sclateri*/*E. warhami* diverged from *E. robustus*/*E. pachyrhynchus* 3.5–1.7 Ma, following the emergence of the Antipodes Islands (5 Ma) (Gamble and Morris 1989). Divergence estimates (95% Highest Posterior Densities) for both *Megadyptes antipodes richdalei* (0.4–0.1 Ma) and *Eudyptes warhami* (2.5–1.1 Ma) similarly post-date the emergence of the Chatham Islands 3 Ma (Campbell et al. 2008), and are concordant with divergence estimates for many endemic Chatham lineages, including plants (Heenan et al. 2010), insects (Trewick 2000) and birds (Mitchell et al. 2014a; Wood et al. 2014) (see Fig. S13). While the age of the Snares is unclear, our analysis of *E. robustus* suggests that they have been emergent for at least 1.4–0.5 Ma. The divergence between *E. moseleyi* and *E. chrysocome*/*E. filholi* 2.7–1.2 Ma corresponds with the emergence of Gough Island around 2.5 Ma [Maund et al. 1988]), and populations presumably dispersed to the younger islands of the Tristan da Cunha archipelago, Amsterdam Island, and St Paul Islands (McDougall and Ollier 1982) via the Antarctic Circumpolar Current. The divergence between *E. chrysolophus chrysolophus* and the Macquarie Island endemic *E. chrysolophus schlegeli* (0.2–0.0 Ma) is concordant with the geological uplift of Macquarie Island 0.7 Ma (Adamson et al. 1996). The Galápagos endemic *Spheniscus mendiculus* diverged from its sister taxon *S. humboldti* 1.6–0.6 Ma shortly after the formation of several islands within this young archipelago (3 Ma). Similar founder speciation has previously been inferred for numerous Galápagos endemic taxa (Parent et al. 2008), including invertebrates (Parent and Crespi 2006; Sequeira et al. 2008), reptiles (Caccone et al. 1999) and birds (Bollmer et al. 2006).

Our finding that many recent speciation events among penguins are temporally linked to island formation may provide important clues for understanding evolutionary patterns in other island-endemic taxa. Islands are clearly speciation hotspots for terrestrial taxa, but the role of island emergence as a driver of speciation in marine taxa remains less clear (however see Mitchell et al. 2014a). The shag genus *Leucocarbo* similarly has endemic taxa associated with almost every sub-Antarctic island (Marchant and Higgins 1990), providing a possible parallel example of recent founder speciation in the Southern Ocean. Time-calibrated genomic analysis provides an exceptional new tool for understanding the origins of such iconic southern biodiversity.

### Vulnerability of island taxa to human-induced extinctions

Our study uncovered two new island-endemic penguin taxa: *Eudyptes warhami* and *Megadyptes antipodes richdalei*. The presence of their bones in middens, and lack of reliable historical sightings, suggests that these taxa were extirpated shortly after human settlement on the Chatham Islands (post 13th century AD; Maxwell and Smith 2015). These findings thus potentially represent important new examples of human-driven, Holocene extinction in the Pacific. *Eudyptes warhami* bones (cf. *E.* clade X) excavated from coastal middens demonstrate that the species was also hunted on mainland New Zealand (Cole et al. 2019). However, this does not prove the presence of a local breeding colony. In fact, many extant island endemic *Eudyptes* disperse widely during the non-breeding period: *E. pachyrhynchus* (breeds only on mainland New Zealand) and *E. robustus* (breeds only on the Snares) are commonly observed in southern Australia during winter (Woehler 1992; Cole et al. 2017; Mattern et al. 2018); and *E. sclateri* (breeds only on the Antipodes Islands) is commonly observed in New Zealand (Robertson et al. 2017). As *E. warhami* is relatively rare in mainland subfossil penguin assemblages (represented by only seven specimens among hundreds of genetically-identified penguin bones [Boessenkool et al. 2009a; Rawlence et al. 2015a; Grosser et al. 2016; Cole et al. 2019]), these mainland records fit the pattern that would be expected for non-breeding individuals. In contrast, no *Megadyptes antipodes richdalei* bones have been detected in mainland subfossil assemblages (Boessenkool et al. 2009a; Rawlence et al. 2015a; Cole et al. 2019). This pattern fits with the limited dispersal exhibited by extant *Megadyptes* populations (Boessenkool et al. 2009b).

### Conclusions

We find strong evidence for a Neogene radiation of crown penguins, and provide the first compelling evidence that island emergence drove Plio-Pleistocene penguin diversification. Such processes may also have driven diversification in the deeper past, as fossil data show much higher penguin diversities than present once existed in New Zealand (Ksepka and Ando 2011), Antarctica (Jadwiszczak 2006), Australia (Park and Fitzgerald 2012), and Africa (Thomas and Ksepka 2013). However, as most fossils from these regions are restricted to continental localities, and many islands have scant fossil records, the role of island formation in penguin diversification in the deep past remains obscured. Accordingly, if rates of island-mediated speciation were as high throughout the Cenozoic as in the Plio-Pleistocene, it is conceivable that fossils for a major proportion of extinct penguin taxa will never be found.

Previous studies based on traditional species concepts have struggled to account for recently-evolved biological diversity. Particularly relevant are scenarios of species ‘divergence with gene flow’, where introgression may occur among closely-related lineages (Rheinhardt and Edwards; 2011). Although hybridisation between closely related penguin species within *Spheniscus* and *Eudyptes* have occasionally been reported, solid confirmatory genetic evidence is lacking (White and Clausen 2002; Simeone et al. 2009; Morrison and Sagar 2014). While our study does not address the possibility of introgression among penguin taxa, future genome-wide analyses will provide insights into this question for penguins.

While our results reinforce the importance of islands in generating biodiversity, they also underscore the role of humans as agents of biodiversity loss, especially via the extinction of island-endemic taxa (Duncan et al. 2013). Today only *Eudyptula minor* breeds on the Chatham Islands, yet 500 years ago the archipelago held substantial penguin diversity, with two endemic taxa (*Eudyptes warhami* and *Megadyptes antipodes richdalei*) alongside *Eudyptula minor* and possibly *Eudyptes sclateri*. As many of the bones were from middens, our results provide direct evidence that *E. warhami* was hunted by humans. Although no *Megadyptes antipodes richdalei* remains examined in this study were directly associated with human activity, the near-simultaneous disappearance of both this subspecies and *Eudyptes warhami* suggests that both extirpations were linked to the arrival of humans to the Chatham Islands. Our results further emphasise the value of ancient DNA for elucidating biodiversity shifts, including the dramatic rise and fall of island avifauna (Waters and Groszer 2016).

## Materials and Methods

### *DNA extraction, amplification and sequencing from historical samples*

Historical skin samples from *Eudyptes filholi*, *E. robustus*, *E. sclateri*, *E. chrysolophus schlegeli* and *E. c. chrysolophus* were obtained from the Museum of New Zealand Te Papa Tongarewa (Te Papa). Holocene fossil and archaeological bones from the Chatham Islands identified as *Eudyptes* based on morphology ( $n=37$ ), and two each of *Megadyptes antipodes antipodes*, *M. a. waitaha*, *M. a. richdalei* and *Eudyptes* clade X (Cole et al. 2019) were sourced from Te Papa, Canterbury Museum, and Auckland War Memorial Museum (Table S1). To avoid duplicate sampling of individuals, either left or right elements were sampled from any one site, or bones were sampled from different stratigraphic units within a site. DNA extractions were performed following rigorous ancient DNA protocols (Cooper and Poinar 2000) at four purpose-built ancient DNA laboratories: Department of Zoology

(University of Otago, Dunedin) following Rohland et al. (2010) (bone) or Rawlence et al. (2015b) (museum skins); Landcare Research (Lincoln) following Thomson et al. (2014) (bone); Te Papa (Wellington) using the manufacturer's protocol from the Qiagen DNeasy Tissue Kit (Qiagen) (museum skins); and the Australian Centre for Ancient DNA (ACAD) following Brotherton et al. (2013).

For species identification we followed Boessenkool et al. (2009a), Cole et al. (2017) and Cole et al. (2019), amplifying  $\leq 499$  bp of cytochrome oxidase 1 (COI) (four overlapping 140–164 bp regions), 131 bp of control region (CR) in *Eudyptes*, and  $\leq 402$  bp of CR in *Megadyptes* (two overlapping 229–255 bp regions). PCRs (total volume = 12.5  $\mu$ L) were performed using 2mg/mL BSA (Sigma), 1 x PCR buffer, 2 mM MgSO<sub>4</sub>, 80  $\mu$ M dNTP, 0.4  $\mu$ M each primer, 0.625 U HiFi Platinum Taq (Invitrogen) and 1  $\mu$ L DNA extract on a BIO-RAD MyCycler thermal cycler as follows: 94°C for 3 min; 55 cycles of 94°C for 30 s, 55°C for 30 s and 68°C for 45 s; 68°C for 10 min. PCR products were purified using SPRIselect (Beckman Coulter, Inc., Indianapolis, IN, USA) and sequenced at the Manaaki Whenua Landcare Research sequencing facility (Auckland) on an Applied Biosystems 3500xL Genetic Analyzer. Contiguous sequences of COI and CR were assembled using Geneious v8.1.8 (Biomatters; Kearse et al. 2012) from high quality bi-directional reads and checked manually. Due to post-mortem DNA damage, when inconsistency between sequences from a given individual was observed (e.g. G-A and C-T transitions), additional PCRs and bidirectional sequencing were conducted, and a majority rule consensus was applied (Brotherton et al. 2007).

Phylogenetic trees were created using BEAST v2.4.7 (Bouckaert et al. 2014), with a relaxed log-normal clock and yule speciation model, 100 million MCMC generations sampling tree parameters every 1000 generations, and a burn-in of 10%. Analyses were run in triplicate and combined using Log Combiner v2.4.7. We implemented the Akaike Information Criterion in JmodelTest2 (Darriba et al. 2012) to determine the most appropriate model of sequence evolution (Jukes Cantor for all genetic markers). We created two maximum clade credibility phylogenies using COI (Fig. S1-S2), and one using the *Megadyptes* CR (Fig. S4). As is typical with ancient DNA, our data contained some missing sequence data. We constructed the first phylogeny which contained all COI data (Fig. S2), and a second phylogeny that contained only samples with 3–4 of the 4 overlapping fragments (75–100% of the four fragments) (Fig. S1). In addition, we included one sequence representative of all extant spheonisciform species, and wandering albatross (*Diomedea exulans*) as outgroup. For the

*Megadyptes* phylogeny, we obtained 20 and 25 *M. antipodes antipodes* and *M. a. waitaha* sequences from GenBank, and one *Eudyptes chrysocome* sequence as outgroup (Fig. S4; Table S11). We used PopArt (Leigh and Bryant 2015) to create the minimum spanning haplotype network (Bandelt et al. 1999) that included all *Eudyptes* CR sequences (Fig. S3).

#### *Enriched mitogenomes from sub-fossil bones and museum skins*

One to two Illumina libraries were created for *Eudyptes filholi* (AD253), *E. robustus* (AD270), *E. pachyrhynchus* (AD266) *E. sclateri* (AD302), *E. chrysolophus schlegeli* (AD415, AD416, AD417, AD419), *E. warhami* (cf *E.* clade X; AD156, AD157, AD161, AD162, AD309, AD342), *Megadyptes antipodes antipodes* (AD93, AD94), *M. a. waitaha* (AD91, AD289), and *M. a. richdalei* (AD88, AD95, ACAD12997) following Meyer and Kircher (2010), but using truncated adapters with unique 7-mer barcode sequences and a partial uracil-DNA-glycosylase treatment (Rohland et al. 2015). We used real-time PCR (rtPCR) to determine the appropriate number of cycles to amplify each library (see Carøe et al. 2018): two 10 µL reactions were run per library containing 1 µL of a 1:5 dilution of post-Bst product, 1x High Fidelity PCR Buffer, 2 mM MgSO<sub>4</sub>, 0.25 mM dNTPs, 0.2 µM IS7 and IS8 primers (see Meyer and Kircher 2010), 0.004x ROX (Life Tech), 0.2x SYBR (Life Tech), 0.56 M DMSO (Sigma-Aldrich), and 0.2 U Platinum Taq DNA Polymerase High Fidelity (ThermoFisher). The rtPCRs were run on a LightCycler 96 (Roche) as follows: 94°C for 6 min; 40 cycles of 94°C for 30 sec, 60°C for 30 sec, and 68°C for 40 sec; and a high resolution melt (95°C for 1 min, 40°C for 1 min, then a ramp from 65°C to 97°C). Each library was divided into 8 x 25 µL PCRs containing: 3 µL of post-Bst library product, 1x High Fidelity PCR Buffer, 2 mM MgSO<sub>4</sub>, 0.25 mM dNTPs, 0.2 µM IS7 and IS8 primers, and 0.5 U Platinum Taq DNA Polymerase High Fidelity (ThermoFisher). PCRs were run in a heated-lid thermal cycler as follows: 94°C for 6 min; 10 to 22 cycles (determined by rtPCR) of 94°C for 30 sec, 60°C for 30 sec, and 68°C for 40 sec; and 68°C for 10 min.

Libraries were enriched for avian mitochondrial DNA (except the CR) with commercially synthesised biotinylated 80-mer RNA baits (Arbor Biosciences, MI, USA), designed from published mitogenome sequences for 27 modern birds (Neornithes), including two penguins (*Eudyptes* and *Megadyptes*) (see Mitchell et al. 2014b). DNA-hybridisation enrichment was performed according to the manufacturer's recommendations (myBaits protocol v1) except the incubation step, which we extended to 44 hr (2 hr at 60°C, 12 hr at 55°C, 12 hr at 50°C, 17 hr at 55°C). After washing the bound DNA, the baits and DNA library were eluted in PCR

master mix, which was then divided into 5 x 25 µL reactions comprising: 1x AmpliTaq Gold Buffer, 2.5 mM MgCl<sub>2</sub>, 0.25 mM dNTPs, 0.4 µM indexed full-length adapter primers (IS4 and indexing primer; see Meyer and Kircher 2010), 1.25 U AmpliTaq Gold DNA Polymerase (Thermo Fisher). PCRs were run in a heated-lid thermal cycler as follows: 94°C for 6 min; 15 cycles of 94°C for 30 sec, 60°C for 30 sec, and 72°C for 45 sec; and 72°C for 10 min. Each amplified library was diluted to 2 nM and run on an Illumina MiSeq using 2 x 150 bp (paired-end) sequencing chemistry.

Reads were demultiplexed using ‘sabre’ (<http://github.com/najoshi/sabre>) (default parameters: no mismatches allowed). Adapter sequences were removed and paired-end reads were merged using AdapterRemoval v2.1.2 (Schubert et al. 2016). Low quality bases were trimmed (<Phred20—minquality 4) and merged reads <25 bp were discarded (--minlength 25). Read quality was visualised using fastQC v0.10.1 (<http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc>) before and after trimming to make sure the trimming was efficient. Collapsed reads from *Eudiptes sclateri* (AD302), *E. warhami* (AD342) and *Megadyptes antipodes richdalei* (ACAD12997) were mapped against the *Eudiptes chrysocome* mitogenome (GenBank accession AP009189) using BWA v0.7.8 (Li and Durbin 2009) (aln -1 1024, -n 0.01, -o 2). Reads with a mapping quality Phred score >30 were selected using the SAMtools v.1.4 (Li et al. 2009) view command (-q 30), and duplicate reads were discarded using ‘FilterUniqueSAMCons.py’ (Kircher 2012). We created 50% (majority-rule) consensus sequences in Geneious. Collapsed reads from all other samples were mapped to these three new reference sequences (as described above), preliminary species assignments were made based on the reference to which the most reads mapped, and 50% consensus sequences were created based on the closest matching reference: *E. sclateri* (AD302) for all non-*warhami* *Eudiptes* samples, *E. warhami* (AD342) for all *E. warhami* samples and *Megadyptes antipodes richdalei* (ACAD12997) for all *Megadyptes* samples. All reads for each sample were then re-mapped to their respective consensus (as described above) and final higher-stringency consensus sequences were created: 85% majority for all samples except *Eudiptes warhami*, which were called at 100% majority. A nucleotide was only called for sites with ≥3x coverage depth (sites with insufficient depth were called as 'N'). For one *Megadyptes antipodes richdalei* sample (AD161 and AD88), we were able to combine the two sets of reads from the two libraries to create one sequence. For these samples, there were no ambiguities between reads that overlapped. We used

MapDamage v.2.0.6 (Jónsson et al. 2013) to ensure that damage patterns in our data were consistent with authentic ancient DNA (Fig. S14–S17; Table S12).

#### *Mitogenomes from contemporary blood*

We created mitogenomes from whole blood of *Aptenodytes patagonicus* (Fortuna Bay, South Georgia), *Eudyptes moseleyi* (Amsterdam Island), *E. chrysolophus chrysolophus* (Marion Island) and *E. c. schlegeli* (Green Gorge, Macquarie Island) at the Beijing Genomics Institute, Hong Kong. DNA was extracted at BGI Hong Kong or the University of Oxford using a HiPure Blood DNA Midi Kit II or the Qiagen DNeasy Tissue Kit, respectively. We constructed 250 bp insert size libraries, and performed whole-genome paired-end sequencing on a BGISEq 500 platform with 150 bp read length. Mitogenomes were *de novo* assembled using SOAPdenovo-Trans (Xie et al. 2014) with approximately 6 Gb data for each species. We raised the linkage support during the scaffolding step to improve accuracy and to avoid connections between mitochondrial reads and nuclear mitochondrial DNA segments (NUMTs). We used a hidden Markov model method (Krogh et al. 1994; Durbin et al. 1998; Wheeler and Eddy 2013) to filter candidate mitochondrial sequences and remove potential false positive mitochondrial scaffolds, NUMTs, or sequences that did not map to avian mitochondrial genes, which we considered contamination.

#### *Phylogenetic analysis*

We aligned all mitogenomes (including 15 obtained from GenBank; Table S11) using the MUSCLE algorithm implemented in Geneious. As most of the mitogenomes were not enriched for the CR, we excluded this from genomes sequenced on the BGISEq 500 platform or those from GenBank. We provided PartitionFinder v2.1.1 (Lanfear et al. 2016) with an input of 30 regions: for each protein coding gene, a region corresponding to every third position counting from the first position (i.e. position 1, 4, 7, etc.) and a region corresponding to every third position counting from the second position (i.e. position 2, 5, 8, etc.); 12S rRNA; 16S rRNA; concatenated tRNAs; and concatenated non-coding regions (108 bp) (Table S13). The two regions we defined for each protein coding gene correspond to the first and second codon positions, except for the last 178 bp of ND3, which are frameshifted by one position due to a single nucleotide insertion. Regions were defined by aligning our sequences with the *E. chrysocome* mitogenome (GenBank accession AP009189) using Geneious, and extracting and concatenating the gene regions for downstream analyses. Optimal partitioning schemes were chosen based on the Bayesian Information Criterion

without a maximum likelihood starting tree (Lanfear et al. 2012; Guindon et al. 2010). We used Geneious to calculate genetic divergences among species using both the entire alignment (Table S2) and excluding any positions with missing data in the alignment (Table S3). We analysed all partitioned alignments in BEAST to infer the topology of Sphenisciformes under a phylogenetic framework (Fig. S5–S6). We explored two rooting strategies, using the outgroup *Phoebastria albatrus* (a member of Procellariiformes, commonly recognised as the sister order to penguins; Ksepka et al. 2006) and rooting the tree between *Aptenodytes/Pygoscelis* and all other penguins. All initial phylogenetic analyses were implemented with a Yule process speciation prior, under a single lognormal relaxed clock model. All phylogenetic analyses were conducted via the CIPRES Science Gateway v3.3 (Miller et al. 2010).

#### *Phylogenetic analysis using fossil calibrations*

To calibrate mitogenomic evolution, we implemented a single lognormal relaxed clock model and constrained the ages of five nodes (four when *Phoebastria albatrus* was excluded) based on fossils (Fig. 1 [c] and Fig. S7–S9; (see *Fossil calibrations* section for details). To test the relative contribution of including the calibration stem penguin *Waimanu manneringi* we ran multiple phylogenetic tests using the alignment that sampled one individual per species, with *Phoebastria albatrus* as an outgroup (Fig. S7). The tests used the following calibration points: 1) only *Waimanu manneringi* (Fig. S7 [c]); 2) all initial calibrations (including *W. manneringi*) (Fig. S7 [b]); and 3) all initial calibrations (without *W. manneringi*) (Fig. S7 [a]). The performance of each test was assessed by running the analysis without the data (i.e. priors only, see Warnock et al. 2015) to determine the relative contribution of the data and the priors to the posterior node age estimates. With *W. manneringi* the effective priors for the crown nodes were much older than the initial specifications, yet without *W. manneringi* the split between Sphenisciformes and Procellariiformes became unrealistically young (based on fossil evidence) (see Fig. S7). This conflict could be due to a rate slowdown in the large-bodied penguin clade (Ksepka and Phillips 2015) and may be exacerbated by the long branch separating crown penguins from Procellariiformes (see Fig. S7). Therefore, we removed the outgroup *Phoebastria albatrus*, and ran the analysis on just Sphenisciformes without the *Waimanu manneringi* calibration point. Each test was run at a minimum of 50 million MCMC, and for each new alignment we re-ran Partition Finder and a single uncalibrated phylogenetic analysis was performed (as above) to determine the topology (especially for *Aptenodytes/Pygoscelis* placements [which remained as sister taxa in all analyses]). To assess



the impact of tree prior choice we ran all analyses (with and without *Phoebeastria albatrus*) using either a birth-death or calibrated Yule process (Fig. S8–S9). Results did not differ substantially, so we consider only those that used the birth-death speciation prior (Fig. 1 [c] and Fig. S8–S9 [a]). After comparing the performance of each test, the final analyses were run using only the penguin taxa, with four internal calibration points (including a uniform prior for crown penguins) (Fig. 1 [c]). This final analysis was run for 100 million generations, and we sampled trees and parameter values every 1000 generations. Parameter values were monitored and compared between chains in Tracer v1.6 to ensure convergence and ESSs >200. We combined sampled trees and parameter values from each chain using Log Combiner. The first 10% of each chain was discarded as burn-in using TreeAnnotator v1.8.3. We visualised the maximum clade credibility tree using FigTree v1.4.2.

#### *Fossil calibrations*

Calibrated Node 1: Sphenisciformes-Procellariiformes; Taxon: *Waimanu manneringi*; Specimen: CM zfa35, part skeleton (holotype); Justification: Phylogenetic analyses (all based on CM zfa35 as the only published specimen of *W. manneringi*) universally support *Waimanu* being along the stem penguin lineage (e.g. Slack et al. 2006; Ksepka et al. 2006; Gavryushkina et al. 2017; Chávez Hoffmeister 2014a). Minimum Age Constraint: 60.5 Ma; Maximum Age Constraint: 72.1 Ma; Prior Distribution: lognormal; Mean: 3.7 (in real space); Standard Deviation: 1.0; Offset: 60.5; Justification: Biostratigraphy (*Hornibrookina teuriensis* and *Chaismolithus bidens*) indicates a minimum age for the type locality of 60.5 Ma (Cooper 2004; Slack et al. 2006; Ogg et al. 2008). The maximum age constraint is the lower bound of the Maastrichtian Stage. Maastrichtian sites have yielded fossil diving birds such as the Northern Hemisphere hesperornithids and the Southern Hemisphere *Neogaeornis*, *Vegavis*, and *Polarornis*, indicating preservation potential for marine diving birds globally, and specifically within the geographic range of modern penguins.

Calibrated Node 2: Crown Spheniscidae; Taxon: *Madrynornis mirandus*; Specimen: MEF-PV 100, part skeleton (holotype); Justification: *M. mirandus* was originally considered a close relative of *Eudyptes* (Acosta Hospitaleche et al. 2007; Ksepka and Clarke 2010) and later to possibly represent the sister taxon to crown Spheniscidae (Chávez Hoffmeister 2014a; Chávez Hoffmeister 2014b). The most recent phylogenetic analysis using new characters from the holotype suggests that *Madrynornis* is more closely related to *Spheniscus* and *Eudyptula*, though this was weakly supported (trees placing the fossil with *Eudyptes* were

only one step longer). Nevertheless, seven synapomorphies support crown status for *Madrynornis*, most compellingly the widely separated fossa temporalis, elongate processus retroarticularis, and small foramen ilioischadicum (Degrange et al. 2018). Given the strong evidence that *Madrynornis* is a crown penguin but uncertainty over the precise relationships of this taxon, we use *Madrynornis* to calibrate the penguin crown; Minimum Age Constraint: 9.7 Ma; Maximum Age Constraint: 25.2 Ma; Prior Distribution: uniform; Justification: The single specimen of *M. mirandus* was collected from the “Entrerriense” sequence of the Puerto Madryn Formation (Acosta Hospitaleche et al. 2007); deposited at  $10.0 \pm 0.3$  Ma (Scasso et al. 2001). The maximum age is the upper boundary of New Zealand’s Kokoamu Greensand, which has yielded many penguin specimens of a wide range of body sizes and at least five different species (all stem taxa). Because the boundary between the Kokoamu Greensand and the overlying Otekaike Limestone likely occurs near the upper Whaingaroan/Duntroonian boundary, we use the age of this boundary (25.2 Ma) in lieu of a more refined date. This maximum age for the crown is consistent with observations that Oligocene units in Australia and South America have also yielded exclusively stem penguins (no Oligocene penguins have yet been reported from Antarctica or Africa).

Calibrated Node 3: *Spheniscus-Eudyptula*; Taxon: *Spheniscus muizoni*; Specimen: MNHN PPI 147, part skeleton (holotype); Justification: Göhlich (2007) listed characters supporting placement in the genus *Spheniscus*. This has been supported by subsequent phylogenetic analyses (e.g. Ksepka and Clarke 2010; Chávez Hoffmeister 2014b). Unambiguous synapomorphies of *Spheniscus* in the calibrating specimen include the straight proximal border of the fossa tricipitalis in ventral view (only seen in *Spheniscus* and some *Palaeospheniscus* specimens) and extremely deep sulcus longitudinalis dorsalis medialis (only seen in *Spheniscus* and the otherwise very dissimilar *Aptenodytes*); Minimum Age Constraint: 9.2 Ma; Maximum Age Constraint: 23.03 Ma; Prior Distribution: lognormal; Mean: 4.4 (in real space); Standard Deviation: 1.0; Offset: 9.2; Justification: PPI 147 was collected from the Cerro la Bruja locality of the Pisco Formation in Peru. The original age estimate of 11–13 Ma for *Spheniscus muizoni* was based on general faunal divisions (Göhlich 2007). However, subsequent work (Brand et al. 2011) provides a revised age of 9.2 Ma for Cerro la Bruja. The maximum age is the base of the Miocene (23.03 Ma [Gradstein 2012]). This encompasses: 1) The well-studied South American early Miocene Gaiman Formation, which has yielded abundant stem penguin fossils (e.g. *Palaeospheniscus*, *Paraptenodytes*, *Eretiscus*) but no crown penguins. 2) Miocene record of Australia, which has yielded

specimens interpreted as being either stem penguins or too incomplete to be assigned to either the stem or crown (Park et al. 2016). 3) Miocene-Pliocene record of New Zealand, which has yielded crown species, none of which fall within the *Spheniscus-Eudyptula* clade. 4) The African record, which is limited to middle/late Miocene penguins of indeterminate status (Thomas and Ksepka 2013) and several early Pliocene crown species, none of which fall within the *Spheniscus-Eudyptula* clade (Ksepka and Thomas 2012).

Calibrated Node 4: *Eudyptes-Megadyptes*; Taxon: *Eudyptes* sp.; Specimen: NMNZ S.046318, part skeleton; Justification: The strongly arched jugal bar in this specimen is a derived feature of *Eudyptes* (Thomas et al. unpublished data). Although it also occurs in some *Pygoscelis* species, tarsometatarsi referred to this species have the derived condition of the foramen vasculare proximale medialis perforating the crista medialis hypotarsi (rather than exiting distal to the crest) supporting a position close to *Eudyptes* and ruling out a relationship with *Pygoscelis*; Minimum Age Constraint: 3.06 Ma; Maximum Age Constraint: 25.2 Ma; Prior Distribution: lognormal; Mean: 7.04 (in real space); Standard Deviation: 1.0; Offset: 3.06; Justification: S.046318 is from the Late Pliocene Tangahoe Formation, Taranaki, New Zealand (Naish et al. 2005). The Tangahoe Formation has been tightly constrained between 3.36 and 3.06 Ma and is within the local Waipipian stage (3.7–3.0 Ma) and the international Piacenzian stage (3.6–2.58 Ma) (Naish et al. 2005; Raine et al. 2015). The formation was dated using magnetostratigraphic correlation to the  $\delta^{18}\text{O}$  timescale from Ocean Drilling Program Site 846, and the presence of Waipipian stage macro- and microfossils (Naish et al. 2005). *Eudyptes calauina* from the Horcón Formation of Chile is of similar age but is known from less complete material. Chávez Hoffmeister (2014b) recovered *E. calauina* within a polytomy including all extant species of *Eudyptes*. The Horcón Formation is considered Late Pliocene but no tighter dates are available for the horizon from which *E. calauina* is known. Thus, that species may be slightly older or younger than the Taranaki *Eudyptes*. Because many species of the *Eudyptes* + *Megadyptes* clade occur on islands and have no pre-Holocene fossil records, we used a conservative Oligocene maximum that follows the same justification as that for crown penguins as a whole.

Calibrated Node 5: *Aptenodytes-Pygoscelis*; Taxon: *Pygoscelis calderensis*; Specimen: SGO-PV 790, part skull; Justification: *P. calderensis* was described on the basis of three partial skulls. The holotype preserves a very shallow temporal fossa; a derived feature which is present only in *Aptenodytes* and *Pygoscelis*. It also preserves a shelf of bone bordering the

supraorbital salt gland fossa, a derived feature which occurs in *Pygoscelis* (as well as *Megadyptes* and *Eudyptes*), but is absent in *Aptenodytes*. Together, these features support placement at least to the stem of *Pygoscelis*; Minimum Age Constraint: 6.3 Ma; Maximum Age Constraint: 25.2 Ma; Prior Distribution: lognormal; Mean: 6.0 (in real space); Standard Deviation: 1.0; Offset: 6.3; Justification: SGO-PV 790 was collected from a phosphatic horizon in the Bahía Inglesa Formation, several meters beneath an ash layer in the uppermost Lechero Member which gave a K-Ar age of  $7.6 \pm 1.3$  Ma (Marquardt et al. 2000; Godoy et al. 2003). The ash layer thus provides a minimum age for the fossil. Because both *Pygoscelis* and *Aptenodytes* occur predominantly in Antarctica and sub-Antarctic islands today, and the fossil record from Antarctica is relatively poor, we used a conservative Oligocene maximum that follows the same justification as that for crown penguins as a whole.

#### *Systematic Palaeontology*

We measured 12 elements from eight *Eudyptes* taxa ( $n=87$ ) (Table S5), and up to 23 elements from each *Megadyptes* subspecies ( $n=57$ ) (Tables S6–S10). Radiocarbon dates of terrestrial birds from the same localities as described material (Millener 1999) were re-calibrated using the SHCal13 atmospheric curve (Hogg et al. 2013) via OxCal v4.3.2 (Bronk Ramsey 2017).

#### **Acknowledgements and Funding Information**

We thank M. Rayner (Auckland War Memorial Museum) for bone measurements; V. De Pietri, P. Scofield, E. Burns and T. Webster for access to Canterbury and Otago Museum collections; J-C. Stahl for photographs; E. Fordyce, J. Scott, W. Lee, N. Mortimer, M. Knapp and P. Scofield for discussions; and S. Petersen and A. Black for field assistance. Ethics approval came from the University of Otago, University of Oxford, University of Cape Town, Zoological Society of London and BGI Hong Kong. Collection permits were provided by the Government of South Georgia and the South Sandwich Islands and the South African National Antarctic Programme. This work was supported by the Marsden Fund (UOO1112, LCR1102), BGI Hong Kong, Australian Research Council, a National Science Fund Award (DEB 1556615), a Rutherford Discovery Fellowship (RDF-MNZ1201), Landcare Research and the University of Otago. TLC was funded by an Otago University Postgraduate Scholarship. Sequences and calibrated tree files are available on GenBank (MK290241–MK290284) and/or Figshare (DOI: 10.6084/m9.figshare.c.4329029), and nomenclatural acts have been registered in ZooBank (urn:lsid:zoobank.org:act:640EE978-13B5-4913-BB4E-6E16395325AC; urn:lsid:zoobank.org:act:169A6F00-1564-4977-B56B-09005591A88E).

## Figure Legends

Fig. 1. (a) Maps of penguin breeding ranges, adapted from Ramos et al. (2018). Only the pre-human breeding range of *Megadyptes antipodes antipodes* is shown. GAL: Galápagos Islands; FAL: Falkland Islands; SG: South Georgia; SS: South Sandwich Islands; SO: South Orkney Islands; BOU: Bouvet; GOU: Gough Island; TDC: Tristan da Cunha; PEI: Prince Edward Islands; CRZ: Crozet Islands; KER: Kerguelen Islands; HEA: Heard Island; AMS: Amsterdam and St Paul Islands; MAC: Macquarie Island; CAM: Campbell Island; SNA: Snares; AUC: Auckland Islands; ANT: Antipodes Islands; CHA: Chatham Islands; NZ: New Zealand. (b) Heatmap showing the percentage of pairwise genetic similarity calculated from 14,117 bp of 23 crown penguin mitogenomes (excluding all alignment columns containing missing data). (c) Dated phylogeny of penguins inferred from mitogenomes. Fossil calibrations are marked with large yellow circles. Posterior probabilities for all clades were >0.99, except those marked with orange squares (0.88 for *Megadyptes* and 0.74 for *Eudyptes*). 95% Highest Posterior Densities are shown as bars associated with each node. The divergence dates of the emergence of the five island archipelagos (coloured in concordance to their respective taxa) are shown under the phylogeny. Taxon-specific symbols are consistent between (a), (b) and (c).

Fig. 2. Skulls of (a) *Megadyptes antipodes antipodes*: NMNZ OR.24638; (b) *Megadyptes antipodes richdalei* paratype: NMNZ S.45876 (AD88); (c) *Eudyptes warhami* holotype: NMNZ S.33007 (AD161) in (c1) dorsal view and (c2) left lateral view (jugal bar is disarticulated); (d) *Megadyptes antipodes richdalei* holotype: NMNZ S.26921 (AD95); (d1) mandible in ventral view, (d2) right and (d3) left coracoid in ventral view, (d4) sternum in ventral view, (d5) left and (d6) right side of pelvis in medial view, (d7) right and (d8) left femur in cranial view, (d9) right and (d10) tibiotarsus in cranial view, (d11) right tarsometatarsus in cranial view, (d12) left and (d13) right humerus in caudal view, (d14) left and (d15) right radius in dorsal view, (d16) left and (d17) right ulna in dorsal view, (d18) left and (d19) right carpometacarpus in dorsal view. Photos: Jean-Claude Stahl, Te Papa.

## References

Acosta Hospitaleche C, Tambussi C, Donato M, Cozzuol M. 2007. A new Miocene penguin from Patagonia and its phylogenetic relationships. *Acta Palaeontol Pol.* 52:299–314.

- Adamson DA, Selkirk PM, Price DM, Selkirk JM. 1996. Pleistocene uplift and palaeoenvironments of Macquarie Island: evidence from paleobeaches and sedimentary deposits. *Pap Proc Roy Soc Tasmania*. 130:25–32.
- Amadon D. 1949. The seventy-five percent rule for subspecies. *Condor*. 51:250–258.
- Bacon CD, Baker WJ, Simmons MP. 2012. Miocene dispersal drives island radiations in the palm tribe Trachycarpeae (Arecaceae). *Syst Biol*. 61:426–442.
- Baker AJ, Pereira SL, Haddrath OP, Edge KA. 2006. Multiple gene evidence for expansion of extant penguins out of Antarctica due to global cooling. *P Roy Soc Lond B Bio*. 273:11–17.
- Bandelt H, Forster P, Röhl A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol*. 16:37–48.
- Bertelli S, Giannini NP. 2005. A phylogeny of extant penguins (Aves: Sphenisciformes) combining morphology and mitochondrial sequences. *Cladistics*. 21:209–239.
- Boessenkool S, Austin JJ, Worthy TH, Scofield P, Cooper A, Seddon PJ, Waters JM. 2009a. Relict or colonizer? Extinction and range expansion of penguins in southern New Zealand. *P Roy Soc B-Biol Sci*. 276:815–821.
- Boessenkool S, Star B, Waters JM, Seddon PJ. (2009b). Multilocus assignment analyses reveal multiple units and rare migration events in the recently expanded yellow-eyed penguin (*Megadyptes antipodes*). *Mol Ecol* 18:2390–2400.
- Bollmer JL, Kimball RT, Whiteman NK, Sarasola JH, Parker PG. 2006. Phylogeography of the Galápagos hawk (*Buteo galapagoensis*): a recent arrival to the Galápagos Islands. *Mol Phylogenet Evol*. 39:237–247.
- Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu C-H, Xie D, Suchard MA, Rambaut A, Drummond AJ. 2014. BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. *Plos Comput Biol*. 10: e1003537.

- Brand L, Urbina M, Chadwick A, DeVries TJ, Esperante R. 2011. A high resolution stratigraphic framework for the remarkable fossil cetacean assemblage of the Miocene/Pliocene Pisco Formation, Peru. *J S Am Earth Sci.* 31:414–425.
- Bronk Ramsey C. 2017. Methods for Summarizing Radiocarbon Datasets. *Radiocarbon.* 59:1809–1833.
- Brotherton P, Endicott P, Sanchez JJ, Beaumont M, Barnett R, Austin J, Cooper A. 2007. Novel high-resolution characterisation of ancient DNA reveals C > U-type base modification events as the sole cause of *post mortem* miscoding lesions. *Nucleic Acids Res.* 35:5717–5728.
- Brotherton P, Haak W, Templeton J, Brandt G, Soubrier J, Adler CJ, Richards SM, Der Sarkissian C, Ganslmeier R, Friederich S. et al. 2013. Neolithic mitochondrial haplogroup H genomes and the genetic origins of Europeans. *Nat Commun* 4:1764.
- Caccone A, Gibbs JP, Ketmaier V, Suatoni E, Powell JR. 1999. Origin and evolutionary relationships of giant Galápagos tortoises. *Proc Natl Acad Sci USA* 96:13223–13228
- Campbell HJ, Begg J, Beu A, Carter B, Curtis N, Davies G, Emberson R, Given D, Goldberg J, Holt K, et al. 2008. Geological consideration relating to the Chatham Islands, mainland New Zealand, and the history of New Zealand terrestrial life. *Geol Soc New Zeal Misc Publ* 126.
- Carøe C, Gopalakrishnan S, Vinner L, Mak SST, Sinding MHS, Samaniego JA, Wales N, Sicheritz-Pontén T, Gilbert MTP. 2017. Single-tube library preparation for degraded DNA. *Methods Ecol Evol.* 9:410–419.
- Chávez Hoffmeister MC. 2014a. Phylogenetic characters in the humerus and tarsometatarsus of penguins. *Pol Polar Res.* 35:469–496.
- Chávez Hoffmeister MC, Briceño JDC, Nielsen SN. 2014b. The evolution of seabirds in the Humboldt Current: new clues from the Pliocene of central Chile. *PLoS ONE.* 9:e90043.

- Christidis L, Boles WE. 2008. Systematics and taxonomy of Australian birds. Collingwood: CSIRO Publishing.
- Cole TL, Waters JM, Shepherd LD, Rawlence NJ, Joseph L, Wood JR. 2017. Ancient DNA reveals that the ‘extinct’ Hunter Island penguin (*Tasidyptes hunteri*) is not a distinct taxon. *Zool J Linn Soc-Lond.* 182:459–464.
- Cole TL, Rawlence NJ, Dussex N, Ellenberg U, Houston DM, Mattern T, Miskelly CM, Morrison KW, Scofield RP, Tennyson AJ, et al. 2019. Ancient DNA of crested penguins: Testing for temporal genetic shifts in the world’s most diverse penguin clade. *Mol Phylogenet Evol* 131:72–79.
- Cooper A, Poinar HN. 2000. Ancient DNA: do it right or not at all. *Science.* 289:1139.
- Cooper RA. 2004. The New Zealand geological timescale. *Inst Geol Nucl Sci Mono.* 22.
- Cowie RH, Holland BS. 2006. Dispersal is fundamental to biogeography and the evolution of biodiversity on oceanic islands. *J Biogeogr.* 33:193–198.
- Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods.* 9:772.
- Darwin C. 1859. On the origins of species by means of natural selection. London: Murray.
- De Dinechin M, Ottval R, Quillfeldt P, Jouventin P. 2009. Speciation chronology of rockhopper penguins inferred from molecular, geological and palaeoceanographic data. *J Biogeogr.* 36:693–702.
- Degrange FJ, Ksepka DT, Tambussi CP. 2018. Redescription of the oldest crown clade penguin: cranial osteology, jaw myology, neuroanatomy, and phylogenetic affinities of *Madrynornis mirandus*. *J Vertebr Paleontol.* 38:e1445636.
- Duncan RP, Boyer AG, Blackburn TM. 2013. Magnitude and variation of prehistoric bird extinctions in the Pacific. *Proc Natl Acad Sci USA.* 201216511.



- Durbin R, Eddy SR, Krogh A, Mitchison GX. 1998. Biological sequence analysis: probabilistic models of proteins and nucleic acids. Cambridge: Cambridge University Press.
- Fleischer RC, McIntosh CE, Tarr CL. 1998. Evolution on a volcanic conveyor belt: using phylogeographic reconstructions and K-Ar-based ages of the Hawaiian Islands to estimate molecular evolutionary rates. *Mol Ecol.* 7:533–545.
- Fraser CI, Nikula R, Spencer HG, Waters JM. 2010. Kelp genes reveal effects of subantarctic sea ice during the Last Glacial Maximum. *P Natl Acad Sci USA.* 106:3249–3253.
- Frugone MJ, Lowther A, Noll D, Ramos B, Pistorius P, Dantas GPM, Petry MV, Bonadonna F, Steinfurth A, Polanowski A, Raya Rey A, Lois NA, Pütz K, Trathan P, Wienecke B, Poulin E, Vianna JA. 2018. Contrasting phylogeographic pattern among *Eudyptes* penguins around the Southern Ocean. *Nat Sci Rep.* 8:17481.
- Gamble JA, Morris PA. 1989. Sub-Antarctic and Chatham Islands. In: Johnson RJ, editor. Intraplate volcanism in eastern Australia and New Zealand. Cambridge (MA): Cambridge University Press. p. 169–171.
- Gathorne-Hardy FJ, Jones DT. 2000. The recolonization of the Krakatau islands by termites (Isoptera), and their biogeographical origins. *Biol J Linn Soc.* 71:251–267.
- Gavryushkina A, Heath TA, Ksepka DT, Stadler T, Welch D, Drummond AJ. 2017. Bayesian total-evidence dating reveals the recent crown radiation of penguins. *Syst Biol.* 66:57–73.
- Gill B, Bell BD, Chambers GK, Medway DG, Palma RL, Scofield RP, Tennyson AJD, Worthy TH. 2010. Checklist of the birds of New Zealand. *New Zealand: Ornithological Society of New Zealand.*
- Gillespie RG, Baldwin BG, Waters JM, Fraser CI, Nikula R, Roderick GK. 2012. Long-distance dispersal: a framework for hypothesis testing. *Trends Ecol Evol.* 27:47–56.

- Godoy E, Marquardt C, Blanco N. 2003. Carta Caldera: Región de Atacama. Servicio Nacional de Geología y Minería, Carta Geológica de Chile, Serie Geología Básica, escala 1:100.000, Santiago.
- Göhlich UB. 2007. The oldest fossil record of the extant penguin genus *Spheniscus*—a new species from the Miocene of Peru. *Acta Palaeontol. Pol* 52:285–298.
- Gradstein FM. 2012. The Geologic Time Scale 2012. Amsterdam: Elsevier.
- Grosser S, Rawlence NJ, Anderson CNK, Smith IWG, Scofield RP, Waters JM. 2016. Invader or resident? Ancient-DNA reveals rapid species turnover in New Zealand little penguins. *Proc R Soc B*. 283:20152879.
- Grosser S, Scofield RP, Waters JM. 2017. Multivariate skeletal analyses support a taxonomic distinction between New Zealand and Australian *Eudyptula* penguins (Sphenisciformes: Spheniscidae). *Emu*. 117:276–283.
- Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol*. 59:307–321.
- Heenan PB, Mitchell AD, De Lange PJ, Keeling J, Paterson AM. 2010. Late-Cenozoic origin and diversification of Chatham Islands endemic plant species revealed by analyses of DNA sequence data. *New Zeal J Bot*. 48:83–136.
- Hogg AG, Hua Q, Blackwell PG, Niu M, Buck CE, Guilderson TP, Heaton TJ, Palmer JG, Reimer PJ, Turney CS. 2013. SHCal13 Southern Hemisphere Calibration, 0-50,000 Years cal BP. *Radiocarbon*. 55.
- Jadwiszczak P. 2006. Eocene penguins of Seymour Island, Antarctica: the earliest record, taxonomic problems and some evolutionary considerations. *Pol Polar Res*. 27:287–302.

- 870 Jónsson H, Ginolhac A, Schubert M, Johnson P, Orlando L. 2013. mapDamage 2.0: fast  
 871 approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics*.  
 872 29:1682–1684.
- 873
- 874 Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A,  
 875 Markowitz S, Duran C, et al. 2012. Geneious Basic: An integrated and extendable desktop  
 876 software platform for the organization and analysis of sequence data. *Bioinformatics*  
 877 28:1647–1649.
- 878
- 879 Kircher M. 2012. Analysis of high-throughput ancient DNA sequencing Data. In *Ancient*  
 880 *DNA: Methods and Protocols*. Humana Press. p.197–228.
- 881
- 882 Krogh A, Brown M, Mian IS, Sjölander K, Haussler D. 1994. Hidden Markov models in  
 883 computational biology: Applications to protein modeling. *J Mol Biol*. 235:1501–1531.
- 884
- 885 Ksepka DT, Bertelli S, Giannini NP. 2006. The phylogeny of the living and fossil  
 886 Sphenisciformes (penguins). *Cladistics*. 22:412–441.
- 887
- 888 Ksepka DT, Clarke JA. 2010. The basal penguin (Aves: Sphenisciformes) *Perudyptes*  
 889 *devriesi* and a phylogenetic evaluation of the penguin fossil record. *B Am Mus Nat Hist*.  
 890 337:1–77.
- 891
- 892 Ksepka DT, Ando T. 2011. Penguins past, present, and future: trends in the evolution of the  
 893 Sphenisciformes. In: Dyke G and Kaiser G, editors. *Living Dinosaurs: The evolutionary*  
 894 *history of modern birds*. Chichester (UK): John Wiley & Sons Ltd. p. 155–186.
- 895
- 896 Ksepka DT, Fordyce RE, Ando T, Jones CM. 2012. New fossil penguins (Aves:  
 897 Sphenisciformes) from the Oligocene of New Zealand reveal the skeletal plan of stem  
 898 penguins. *J Vertebr Paleontol*. 32:235–254.
- 899
- 900 Ksepka DT, Thomas DB. 2012. Multiple cenozoic invasions of Africa by penguins (Aves,  
 901 Sphenisciformes). *Proc R Soc. B*. 279:1027–1032.
- 902

- Ksepka DT, Phillips MJ. 2015. Avian diversification patterns across the K-Pg boundary: Influence of calibrations, datasets, and model misspecification. *Ann Mo Bot Gard.* 100: 300–328.
- Lanfear R, Calcott B, Ho SY, Guindon S. 2012. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Mol Biol Evol.* 29:1695–1701.
- Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T., Calcott, B. 2016. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Mol Biol Evol.* 34:772–773.
- Leigh JW, Bryant D. 2015. PopART: full-feature software for haplotype network construction. *Methods Ecol Evol.* 6:1110–1116.
- Lewin R. 1985. Hawaiian *Drosophila*: Young islands, old flies. *Science.* 229:1072–1074.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 25:1754–1760.
- Li H, Handsaker B, Wysoker A, Fennel T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map (SAM) format and SAMtools. *Bioinformatics.* 25:2078–2079.
- Losos JB, Ricklefs RE. 2009. Adaptation and diversification on islands. *Nature.* 457:830
- Marchant S, Higgins PJ. 1990. Handbook of Australian, New Zealand and Antarctic birds. Vol. 1, Pt. B. Melbourne: Oxford University Press.
- Marquardt C, Blanco N, Godoy E, Lavenu A, Ortlieb L, Marchant M, Guzmán N. 2000. Estratigrafía del Cenozoico Superior en el área de Caldera (26°45' - 28°S), III Región de Atacama. In Congreso Geológico Chileno, No. 9, Actas: 504-508. Puerto Varas.

- Mattern T, Pütz K, Garcia-Borboroglu P, Ellenberg U, Houston DM, Long R, Lüthi B, Seddon PJ. 2018. Marathon penguins – Reasons and consequences of long-range dispersal in Fiordland penguins / Tawaki during the pre-moult period. *PLoS One*. 13:e0198688.
- Maund JG, Rex DC, Le Roex AP, Reid DL. 1988. Volcanism on Gough Island: a revised stratigraphy. *Geol Mag*. 125:175–181.
- Maxwell JJ, Smith IWG. 2015. A reassessment of settlement patterns and subsistence at Point Durham, Chatham Island. *Archaeol Ocean*. 50:162–174.
- McDougall I, Ollier CD. 1982. Potassium-argon ages from Tristan da Cunha, south Atlantic. *Geol Mag*. 119:87–93.
- Mendelson TC, Shaw KL. 2005. Rapid speciation in an arthropod. *Nature*. 433:375–376
- Meyer M, Kircher M. 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols*. 2010(6):pdb-prot5448.
- Millener PR. 1999. The history of the Chatham Islands' bird fauna of the last 7000 years-a chronicle of change and extinction. *Sm C Paleob*. 89.
- Miller MA, Pfeiffer W, Schwartz T. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: Proceedings of the Gateway Computing Environments Workshop (GCE), 14 Nov. 2010, New Orleans, LA. p 1–8.
- Mitchell KJ, Wood JR, Scofield RP, Llamas B, Cooper A. 2014a. Ancient mitochondrial genome reveals unsuspected taxonomic affinity of the extinct Chatham duck (*Pachyanas chathamica*) and resolves divergence times for New Zealand and sub-Antarctic brown teals. *Mol Phylogenet Evol*. 70:420–428.
- Mitchell KJ, Llamas B, Soubrier J, Rawlence NJ, Worthy TH, Wood J, Lee MSY, Cooper A. 2014b. Ancient DNA reveals elephant birds and kiwi are sister taxa and clarifies ratite bird evolution. *Science*. 6186:898–900.

- 970 Morrison KW, Sagar PM. 2014. First record of interbreeding between a Snares crested  
 971 (*Eudyptes robustus*) and erect-crested penguin (*E. sclateri*). *Notornis*. 61:109–112.  
 972
- 973 Naish TR, Wehland F, Wilson GS, Browne GH, Cook RA, Morgans HE, Rosenberg M, King  
 974 PR, Smale D, et al. 2005. An integrated sequence stratigraphic, palaeoenvironmental, and  
 975 chronostratigraphic analysis of the Tangahoe Formation, southern Taranaki coast, with  
 976 implications for mid-Pliocene (c. 3.4–3.0 Ma) glacio-eustatic sea-level changes. *J Roy Soc*  
 977 *New Zeal* 35:151–196.  
 978
- 979 Ogg J, Ogg GG, Gradstein FM. 2008. The Concise Geologic Time Scale. Cambridge:  
 980 Cambridge University Press.  
 981
- 982 Parent CE, Crespi BJ. 2006. Sequential colonization and diversification of Galápagos  
 983 endemic land snail genus *Bulimulus* (Gastropoda, Stylommatophora). *Evolution*. 60:2311–  
 984 2328.  
 985
- 986 Parent CE, Cannone A, Petren K. 2008. Colonization and diversification of Galápagos  
 987 terrestrial fauna: a phylogenetic and biogeographical synthesis. *Philos T R Soc B B*. 368.  
 988
- 989 Park TR, Fitzgerald EM. 2012. A review of Australian fossil penguins (Aves:  
 990 Sphenisciformes). *Mem Mus Victoria*. 69:309–325.  
 991
- 992 Park T, Fitzgerald EMG, Gallagher SJ, Tomkins E, Allan T. 2016. New Miocene Fossils and  
 993 the History of Penguins in Australia. *Plos One*. 11:e0153915.  
 994
- 995 Patten MA, Unit P. 2002. Diagnosability versus mean differences of sage sparrow  
 996 subspecies. *The Auk*. 119:26–35.  
 997
- 998 Paulay G. 1994. Biodiversity on Oceanic Islands: Its Origin and Extinction. *Am Zool*.  
 999 34:134–144.  
 1000

- 1001 Raine JJ, Beu AG, Boyes AF, Campbell H, Cooper RA, Crampton JS, Crundwell MP, Hollis  
 1002 CJ, Morgans HE. 2015. Revised calibration of the New Zealand Geological Timescale:  
 1003 NZGT2015/1. GNS Science report 2012/39 (GNS Science, Lower Hutt NZ) p. 53  
 1004
- 1005 Ramos B, González-Acuña D, Loyola DE, Johnson WE, Parker PG, Massaro M, Dantas GP,  
 1006 Miranda MD, Vianna JA. 2018. Landscape genomics: natural selection drives the evolution  
 1007 of mitogenome in penguins. *BMC Genomics*. 19:53.  
 1008
- 1009 Rawlence NJ, Perry GLW, Smith IWG, Scofield RP, Tennyson AD, Matisoo-Smith EA,  
 1010 Boessenkool S, Austin JJ, Waters JM. 2015a. Radiocarbon-dating and ancient DNA reveal  
 1011 rapid replacement of extinct prehistoric penguins. *Quat Sci Rev*. 112:59–65.  
 1012
- 1013 Rawlence NJ, Kennedy M, Anderson CN, Prost S, Till CE, Smith IW, Scofield RP, Tennyson  
 1014 AJ, Hamel J, Lallas C, et al. 2015b. Geographically contrasting biodiversity reductions in a  
 1015 widespread New Zealand seabird. *Mol Ecol* 24:4605–4616.  
 1016
- 1017 Rheindt FE, Edwards SV. 2011. Genetic introgression: an integral but neglected component  
 1018 of speciation in birds. *The Auk*. 128:620–632.  
 1019
- 1020 Robertson HA, Baird K, Dowding JE, Elliott GP, Hitchmough RA, Miskelly CM, McArthur  
 1021 N, O'Donnell CFJ, Sagar CFJ, Scofield RP, et al. 2017. Conservation status of New Zealand  
 1022 birds, 2016. New Zealand Threat Classification Series 19. New Zealand Department of  
 1023 Conservation.  
 1024
- 1025 Rohland N, Siedel H, Hofreiter M. 2010. A rapid column-based ancient DNA extraction  
 1026 method for increased sample throughput. *Mol Ecol Resour*. 10:677–683.  
 1027
- 1028 Rohland N, Harney E, Mallick S, Nordenfelt S, Reich D. 2015. Partial uracil-DNA-  
 1029 glycosylase treatment for screening of ancient DNA. *Philos Trans R Soc Lond B Biol Sci*.  
 1030 370:20130624.  
 1031
- 1032 Scasso RA, McArthur JM, Del Rio CJ, Martinez S, Thirlwall MF. 2001.  $^{87}\text{Sr}/^{86}\text{Sr}$  Late  
 1033 Miocene age of fossil molluscs in the 'Entrerriense' of the Valdés Peninsula (Chubut,  
 1034 Argentina). *J S Am Earth Sci*. 14:319–329.

- 1035
- 1036 Schubert M, Lindgreen S, Orlando L. 2016. AdapterRemoval v2: rapid adapter trimming,  
1037 identification, and read merging. *BMC Res Notes*. 12:88.
- 1038
- 1039 Sequeira AS, Sijapati M, Lanteri AA, Roque Albelo L. 2008. Nuclear and mitochondrial  
1040 sequences confirm complex colonization patterns and clear species boundaries for flightless  
1041 weevils in the Galápagos archipelago. *Phil Trans R Soc. B*. 363:3439–3451.
- 1042
- 1043 Shaughnessy PD. 1975. Variation in facial colour of the Royal Penguin. *Emu*. 75:147–152.
- 1044
- 1045 Shaw KL, Gillespie RG. 2016. Comparative phylogeography of oceanic archipelagos:  
1046 Hotspots for inferences of evolutionary process. *P Natl Acad Sci USA*. 113:7986–7993.
- 1047
- 1048 Simeone A, Hiriart-Bertrand L, Reyes-Arriagada R, Halpern M, Dubach J, Wallace R, Pütz  
1049 K, Luthi B. 2009. Heterospecific pairing and hybridization between wild Humboldt and  
1050 Magellanic penguins in southern Chile. *The Condor*. 111:544–550.
- 1051
- 1052 Sinton WC, Hauff F, Hoernle K, Werner R. 2018. Age progressive volcanism opposite Nazca  
1053 plate motion: Insights from seamounts on the northeastern margin of the Galapagos Platform.  
1054 *Lithos*. 310:342–354.
- 1055
- 1056 Slack KE, Jones CM, Ando T, Harrison GL, Fordyce RE, Arnason U, Penny D. 2006. Early  
1057 penguin fossils, plus mitochondrial genomes, calibrate avian evolution. *Mol Biology Evol*.  
1058 23:1144–1155.
- 1059
- 1060 Subramanian S, Beans-Picón G, Swaminathan SK, Millar CD, Lambert DM. 2013. Evidence  
1061 for a recent origin of penguins. *Biol Letters*. 9:20130748.
- 1062
- 1063 Tennyson AJD, Millener PR. 1994. Bird extinctions and fossil bones from Mangere Island,  
1064 Chatham Islands. *Notornis*. 41 supplement:165–178.
- 1065
- 1066 Thomas DB, Ksepka DT. 2013. A history of shifting fortunes for African penguins. *Zool J*  
1067 *Linn Soc-Lond*. 168:207–219.
- 1068



- Thomson VA, Lebrasseur O, Austin JJ, Hunt TL, Burney DA, Denham T, Rawlence NJ, Wood JR, Gongora J, Flink LG, et al. 2014. Using ancient DNA to study the origins and dispersal of ancestral Polynesian chickens across the Pacific. *P Natl Acad Sci USA* 111:4826–4831.
- Trewick SA. 2000. Molecular evidence for dispersal rather than vicariance as the origin of flightless insect species on the Chatham Islands, New Zealand. *J Biogeogr.* 27:1189–1200.
- Warham J. 1975. The crested penguins. In: *The biology of penguins*. p. 189–269.
- Warnock RCM, Parham JF, Joyce WG, Lyson TR, Donoghue PCJ. 2015. Calibration uncertainty in molecular dating analyses: there is no substitute for the prior evaluation of time priors. *Proc R Soc B.* 282:20141013.
- Waters JM, Fraser CI, Hewitt GM. 2013. Founder takes all: density-dependent processes structure biodiversity. *Trends Ecol Evol.* 28:78–85.
- Waters JM, Grosser S. 2016. Managing shifting species: ancient DNA reveals conservation conundrums in a dynamic world. *Bioessays.* 38:1177–1184.
- Wheeler TJ, Eddy SR. 2013. nhmmer: DNA homology search with profile HMMs. *Bioinformatics.* 29:2487–2489.
- White RW, Clausen AP. 2002. Rockhopper *Eudyptes chrysocome chrysocome* x Macquarie *E. chrysolophus* penguin hybrids apparently breeding in the Falkland Islands. *Mar Ornithol.* 30:40–42.
- Woehler EJ. 1992. Records of vagrant penguins from Tasmania. *Mar Ornithol.* 20:61–73.
- Woehler EJ. 1995. Bill morphology of Royal and Macaroni Penguins, and geographic variation within eudyptid penguins. *The penguins*. Chipping Norton: Surrey Beatty & Sons Pty Ltd, 319–330.

- 1102 Wood JR, Mitchell KJ, Scofield RP, Tennyson AJ, Fidler AE, Wilmshurst JM, Llamas B,  
1103 Cooper A. 2014. An extinct nestorid parrot (Aves, Psittaciformes, Nestoridae) from the  
1104 Chatham Islands, New Zealand. *Zool J Linn Soc-Lond.* 172:185–199.  
1105
- 1106 Wood JR, Alcover JA, Blackburn TM, Bover P, Duncan RP, Hume JP, Luoyes J, Meijer HJM,  
1107 Rando JC, Wilmshurst JM. 2017. Island extinctions: patterns, processes, and potential for  
1108 ecosystem restoration. *Environ Conserv* 44:348–358.  
1109
- 1110 Xie Y, Wu G, Tang J, Luo R, Patterson J, Liu S, Huang W, He G, Gu S, Li S, et al. 2014.  
1111 SOAPdenovo-Trans: de novo transcriptome assembly with short RNA-Seq reads.  
1112 *Bioinformatics* 30:1660–1666.  
1113
- 1114 Zusi RL. 1975. An interpretation of skull structure in penguins. In: Stonehouse B, editor. *The*  
1115 *Biology of Penguins*. Baltimore: University Park Press. p. 55–84.

Figures



