Genome-wide analyses reveal drivers of penguin diversification

Juliana Vianna, Flávia Fernandes, María José Frugone, Henrique Figueiró, Luis Perttierra, Daly Noll, Ke Bi, Cynthia Wang-Claypool, Andrew Lowther, Patricia Parker, et al.

To cite this version:

Juliana Vianna, Flávia Fernandes, María José Frugone, Henrique Figueiró, Luis Perttierra, et al.. Genome-wide analyses reveal drivers of penguin diversification. Proceedings of the National Academy of Sciences of the United States of America , National Academy of Sciences, 2020, 117 (36), pp.22303-22310. 10.1073/pnas.2006659117. hal-02999306

HAL Id: hal-02999306
https://hal.archives-ouvertes.fr/hal-02999306
Submitted on 17 Nov 2020
Title: Genome-wide analyses reveal drivers of penguin diversification

Authors: Juliana A. Vianna¹*, Flávia A. N. Fernandes ², María José Frugone³, Henrique Figueiró⁴, Luis R. Pertierra⁶, Daly Noll¹,³, Ke Bi⁶, Cynthia Y. Wang-Claypool⁶, Andrew Lowther⁷, Patricia Parker⁸, Celine Le Bohec⁹,¹⁰, Francesco Bonadonna¹¹, Barbara Wienecke¹², Pierre Pistorius¹³, Antje Steinfurth¹⁴, Christopher Burridge¹⁵, Gisele P. M. Dantas¹⁶, Elie Poulin³, W. Brian Simison¹⁷, Jim Henderson¹⁷, Eduardo Eizirik⁴, Mariana F. Nery², Rauri C. K. Bowie⁶*

Affiliations:

¹Pontificia Universidad Católica de Chile, Facultad de Agronomía e Ingeniería Forestal, Departamento de Ecosistemas y Medio Ambiente, Vicuña Mackenna 4860, Santiago, Chile.

²Universidade Estadual de Campinas (UNICAMP), Sao Paulo, Brazil.

³Instituto de Ecología y Biodiversidad (IEB), Departamento de Ciencias Ecológicas, Facultad de Ciencias, Universidad de Chile, Santiago, Chile.

⁴Pontificia Universidad Católica do Rio Grande do Sul, Escola de Ciências, Brazil.

⁵Museo Nacional de Ciencias Naturales, Departamento de Ecología Evolutiva, CSIC, C/José Gutiérrez Abascal, 2, 28006 Madrid, Spain.

⁶Museum of Vertebrate Zoology and Department of Integrative Biology, 3101 Valley Life Science Building, University of California, Berkeley, Berkeley, CA 94720, USA.

⁷Norwegian Polar Institute, N-9297 Tromsø, Norway.

⁸University of Missouri St Louis and Saint Louis Zoo, One University Blvd., St. Louis, MO, 63121-4400, USA.

⁹Université de Strasbourg, CNRS, IPHC UMR 7178, 23 rue Becquerel, F-67000 Strasbourg, France.
Author contributions: J.A.V., F.A.N.F., M.N., R.C.K.B. conceived and designed the study and contributed to preparing the manuscript. H.F., K.B., E.E., B.S. and J.H. contributed bioinformatics. L.P. developed the ecological niche models. M.J.F., D.N., and E.P. facilitated analyses and contributed to the interpretation of results. M.A.L., P.P., C.L.B., F.B., C.G., B.W., P.P., A.S., C.B., G.P.M.D, and C.W. undertook field work, provided samples, and contributed to the manuscript. All co-authors revised and commented on the manuscript and approved the final version;

Competing interests: The authors declare no competing interests.
Corresponding authors:
Juliana A. Vianna, Pontificia Universidad Católica de Chile, Facultad de Agronomía e Ingeniería Forestal, Departamento de Ecosistemas y Medio Ambiente, Vicuña Mackenna 4860, Macul, Santiago, Chile, phone 56-988199882, jvianna@uc.cl
Rauri C. K. Bowie, Museum of Vertebrate Zoology and Department of Integrative Biology, 3101 Valley Life Science Building, University of California, Berkeley, CA 94720, USA. Phone 510-643-1617, bowie@berkeley.edu

Keywords: Penguin, Antarctica, Genome, Ancestral niche, Ancestral distribution.
Extant penguins are the only flightless family of diving birds. They comprise at least 18 modern species, distributed from polar to tropical environments in the Southern Hemisphere. The history of their diversification and adaptation to these diverse environments remains controversial. We employed 22 new genomes from 18 penguin species to reconstruct the order, timing, and location of their diversification, track changes in thermal niche through time, and test for associated adaptation across the genome. The penguin crown-group originated during the Miocene in New Zealand and Australia, and not in Antarctica as previously thought. *Aptenodytes* appears as sister taxon of all other penguin species. We show that lineage diversification in penguins was driven by changing climatic conditions and by the development of the Antarctica Circumpolar Current (ACC). Changes in thermal niche were accompanied by adaptations in genes that govern thermoregulation and oxygen metabolism. Our reconstruction of changes in ancestral effective population sizes ($N_e$) recovered three different demographic trajectories in deeper time, the most common (11/18 penguin species) being an increased $N_e$ between 0.04 and 0.07 Mya; $N_e$ drops precipitously after the Last Glacial Maximum (LGM) most likely as a consequence of the overall decline in marine productivity following the last glaciation.

Significance Statement

Penguins have long been of interest to the scientific community. However, their evolutionary history is not well understood. Using genomes we investigate macro- and microevolutionary drivers of penguin diversification. Crown-group penguins diverged in the early Miocene in the Australia/New Zealand region. We found that *Aptenodytes* (emperor and king penguins) form a separated clade, sister of all other penguin lineages, in contrast with previous studies that placed them as sister taxa of *Pygoscelis*. Analyses suggest that penguins first radiated from temperate environments and then occupied cold Antarctic waters. Although the onset of the ACC (35 Mya) was suggested to have a role on penguins diversification, our findings of more recent time estimates of divergence
suggest that the intensification of the ACC (11.6 Mya) may have promoted their
diversification and geographic expansion.

Introduction

Few organisms have been as successful at colonizing the globe as seabirds, a large
ecological assemblage of oceanic and nearshore species that undergo some of the most
remarkable foraging and migratory journeys on Earth (1, 2). Despite their ubiquitous
presence, surprisingly little is known about the mechanisms that spurred their
diversification and allowed for their adaptation to diverse and often dynamic oceanic
habitats.

As the only living clade of flightless diving birds, penguins (order Sphenisciformes)
occupy both terrestrial and marine habitats. They forage across a wide range of ocean
temperatures and depths, from Antarctic to tropical waters (3). Our understanding of
penguin diversification and adaptation is hampered by disagreements about their
phylogenetic relationships (4–10) and the chronology of their radiation. When these
estimations are made using few genetic markers from different sections of the genome,
discordant results are to be expected as genomic regions vary in their mutation rates and
evolutionary history, including unknown patterns of gene introgression when different
species hybridize (11). Accordingly, estimates of ordinal crown-age vary between the
Miocene and Eocene (9.9–47.6 million years ago [Mya]) (4, 5, 8, 9). The earliest crown-
group fossil dates to the late Miocene (12).

In addition, reconstructions of ancestral distributions and climatic niches are critical
to extend our knowledge of penguin diversification, but existing hypotheses are
conflicting: (i) an Antarctic origin, with later expansion towards warmer areas (5) or (ii) a
sub-Antarctic origin, with subsequent colonization of Southern Ocean islands and
Antarctica (7, 13). Testing these alternative hypotheses on a broad scale requires
accurate knowledge of the evolutionary history of penguins, including branching patterns
and timing of diversification.
The extent of ice and changes in the currents of the Southern Ocean during repeated glacial cycles likely played a significant role in structuring penguin populations and lineage diversification (14-16). Biogeographic boundaries in the Southern Ocean, particularly the Antarctic Polar Front (APF) and the sub-tropical Front (STF), serve as barriers to dispersal of some penguin species (14, 15, 17, 18). Differences in abiotic (e.g., temperature and salinity) and biotic (e.g., types of food resources) variables on either side of these two fronts may promote local adaptation and enable niche divergence among penguins (19). Furthermore, associated currents have varied significantly over time in latitude and strength in response to changing global circulation patterns (20). These changes have been implicated in the colonization, isolation, and local extinction of some penguin populations and species (4, 9).

We report here the reconstruction of the history of penguin diversification and adaptation using 22 newly sequenced genomes representing 18 extant species and one outgroup. We determined phylogenetic relationships and biogeographic histories, and consequently the divergence of niche space through time. Finally, we also assessed the extent of introgression between species that may have contributed to previous disagreements on phylogenetic reconstructions, and adaptations of penguins across environmental gradients.

Results

The penguin and petrel outgroup genomes were sequenced to ~30x coverage (Tables S1–S3). All, but one genome sequences contained >90% of the conserved single-copy genes identified across avian genomes (complete BUSCOs >90%; Table S4, Figs. S1–4). From the assembled genomes we extracted 23,108 loci: Ultra Conserved Elements (UCE: 4,057 loci), coding sequences (CDS: 11,011), and introns (8,040). We applied species tree and concatenation methods to construct a phylogenetic hypothesis for extant penguins using all loci together, and for each set of loci independently (Figs. 1, S5–8, Table S5–S6).

Phylogenetic history and patterns of introgression of crown-group penguins
All analyses supported a similar phylogenetic hypothesis, all placing *Aptenodytes* as sister taxon of all other extant penguin species (Fig. 1), distinct from other clades comprising the genera *Pygoscelis*, *Spheniscus+Eudyptula*, and *Megadyptes+Eudyptes*. The phylogeny based on mitogenomes was similar to that retrieved from the genomic datasets, with a few minor differences in *Eudyptes* penguins that are likely explained by genome-wide introgression among closely related species (Figs. 2, S7, S9, Table S7) as we found evidence that some species may have hybridized during the course of their diversification. Genomic introgressions were detected between i) erect-crested and the ancestral rockhopper penguin species (17–23%); ii) erect-crested and macaroni/royal (25%) penguins (Fig. 2), and iii) an ancestor of the Galápagos/humboldt and magellanic penguins (11%; Fig. 2). Our divergence time estimated the divergence of the crown-group penguins in the early Miocene at 21.9 Mya (95% CI 19.06–25.19; Figs. 1, S10, Table S8).

**Biogeographic history and ancestral niche reconstruction**

The reconstruction of ancestral geographical occurrence identified the coastlines of Australia, New Zealand, and nearby islands as the most likely range of the ancestor of extant penguins (Figs. 1, S11; Table S9). The first branching event (20.3–19.7 Mya) led to the establishment of the genus *Aptenodytes* in the Antarctic, and reconstructions of the ancestral *Pygoscelis* species indicate that they colonized the Antarctic Peninsula (area C; Fig. 1, Table S9) soon after *Aptenodytes*, pointing to a long history of occupation in Antarctica. In the mid-Miocene, the lineage leading to the *Spheniscus/Eudyptula* ancestor colonized the South American coast (A), with members of the genera *Eudyptes*, *Eudyptula*, *Megadyptes*, and *Spheniscus* progressively diversifying and colonizing warmer oceanic environments (Fig. 1).

Consistent with the above findings, our reconstruction of sea surface temperatures (SST) as a proxy for ecological niche disparity through time (DTT) suggests that penguins originated from areas with a maximum SST of 9ºC (Figs. 1, S12-S13) which is in alignment with present-day temperatures of sub-Antarctic waters (21). Southern and eastern rockhopper penguins have retained a thermal preference for a maximum SST of 9ºC, while other closely related species (northern Rockhopper, fiordland, erect-crested) have shifted towards warmer at-sea conditions (SST of 11–17ºC; Figs. 1, S13). This
thermal shift was most likely driven by divergence across the STF over the past 4.5 Mya. In contrast, Macaroni penguins have shifted to occupy colder at-sea conditions to feed in the nutrient-rich waters off Antarctica (Fig. S13). Lower latitude geographical locations, such as the South African continental coasts (G) and the Galápagos Islands (J), were colonized in the Pleistocene (0.59 Mya). Galápagos penguins are present in Pacific tropical waters near the equator, where SST reach up to 27°C, and exhibit a significant increase in thermal tolerance compared to their sister-species, humboldt penguins (>6°C higher; DTT results, Fig. 1).

**Genes under positive selection**

We detected 104 genes under positive selection (BEB value >0.95 and FDR q<0.1) using a site model across all branches of the Spheniscidae (Table S10-12, Fig. S14). Using a gene interaction network analysis, we found that many of these positively selected loci are functionally connected and formed two major clusters (Fig. 3): one primarily related to broad cellular functions, and the other containing genes that affect specific phenotypes including immunity, renal function and cardiovascular activities (e.g., blood pressure, oxygen metabolism, coagulation). Our gene ontology analysis revealed a concordant pattern of pathway enrichment, including terms related to angiotensin regulation, blood pressure, and oxygen metabolism (Fig. 3, Table S11–12). All of these adaptations are related to diving and maintenance of high core body temperatures.

**Ancestral effective populations sizes \((N_e)\)**

Our reconstruction of changes in ancestral effective population sizes \((N_e)\) recovered three different demographic trajectories in deeper time. Eleven species of penguin (emperor, king, Adélie, chinstrap, gentoo [Kerguelen Population], humboldt, magellanic, African, eastern rockhopper, little) increased \(N_e\) between 0.04 and 0.07 Mya, with \(N_e\) dropping precipitously during the Last Glacial Maximum. In contrast, three species (northern rockhopper, southern rockhopper, erect-crested) increased \(N_e\) during the Penultimate Glaciation Period (PGP) between 0.130–0.194, but were in decline by 0.07 Mya. Finally,
two species (gentoo [Antarctica, Falkland Is. populations], Galápagos) have been in steady decline since at least the Naynayxungla Glaciation (0.50–0.72 Mya; Fig. 4, S16). Why Galápagos penguins and southern populations of gentoo penguins have been in decline over such a considerable period of time remains uncertain, as there are no large life-history or ecological differences that distinguishes these two species from other penguins (Table S2). One possibility is that both Galápagos penguins and southern populations of gentoo penguins represent recent divergence events (Fig 1. S7, S10), and hence, the demographic reconstructions may be reflective of a founder event with larger deep time $N_e$ reflecting ancestral population sizes.

For erect-crested, southern and northern rockhoppers, and magellanic penguins distributed around the tip of South America and on islands north of the APF, $N_e$ was highest during the PGP with a decline starting in or shortly after the subsequent interglacial. This pattern could reflect changes in ecosystem productivity over deep time, but it could be an artifact of the inability of the PSMC method to account for introgression. Our analyses suggest that for these four species ancestral introgression (Fig. S9) would have elevated estimates of $N_e$ and could have offset the time interval when a peak in $N_e$ occurred (Fig. S16). The largest ancestral $N_e$ we observed was for Adélie penguins ($>10^5$), a primarily cold-water species (Fig. 1) widely distributed south of the APF; decreases in temperatures and increases in ice cover may have promoted the expansion of that population. Smaller historical peaks for $N_e$ ($\sim 2-3*10^4$) were recovered for island endemic species, such as fiordland (southern New Zealand) and northern rockhopper penguins (Tristan da Cunha and Gough Island).

Discussion
Phylogenetic reconstructions recovered by this study show that the large emperor and king penguins (i.e., *Aptenodytes*) are sister to all other extant penguins, refuting the hypothesis that *Aptenodytes* and *Pygoscelis* are sister-taxon (4, 8, 9). While genome-wide data used in our trees allowed for the detection of the deep lineage split between *Aptenodytes* and all other penguins, the short internal branch recovered by our analysis at this split likely indicates a rapid diversification event (<1 Mya) (22). Such rapid speciation events may explain the discrepancies among the previous proposed
phylogenetic hypothesis generated using that few molecular markers, which are known to be insufficient in resolving such deep and short branches (23). As the clades become more divergent, the phylogenetic position of the remaining genera and species are consistent with phylogenetic hypotheses generated in other studies (4–10). Reticulated evolution may also have contributed to some inconstancies in penguin phylogenetic reconstructions. Genome analyses detected deep and shallow introgression events across the phylogeny of extant penguins. Species of crested penguins (Eudyptes) appear to have exchanged genes throughout much of their evolutionary history. The directionality of introgression among these lineages is consistent with the clockwise direction of the Antarctic Circumpolar Current (ACC) connecting sub-Antarctic islands and promoting dispersal. We also detected extensive genomic introgression between an ancestor of the galápagos/humboldt penguin and the magellanic penguin; such introgression is still observed at present between magellanic and humboldt penguins for which hybridization has been reported in the wild (24).

Our divergence time estimates are consistent with the fossil record (12), placing the ancestor of crown-group penguins in the early Miocene. Our biogeographic reconstruction for crown-group penguins pointed to the coastlines of Australia and New Zealand and nearby islands as the most likely range of the ancestor of extant penguins, supporting earlier suggestion (7), rather than Antarctica as place of origin (5). Because of the recency of most diversification events (2–9 Mya), we propose an alternative hypothesis regarding their geographic expansion, namely that the opening of the Drake Passage around the tip of South America led eventually to the full development of a deep, strong ACC by ~11.6 Mya (26) which in turn may have contributed to colonization of new areas and diversification of penguins, rather than the initial formation of the ACC ~35 Mya (26)(Fig. 1). Global climate cooling intensified with the strengthening of the ACC and possibly led to the extinction of several species inhabiting Antarctica (20, 27). The ACC also reinforced the isolation of taxa inhabiting Antarctica from those on continental shelves and islands to the north (20, 27), while promoting eastward colonization of available sub-Antarctic islands.

Acceleration of the decline in surface temperatures since the Pliocene (20) may have facilitated the colonization of new areas such as islands in the Indian Ocean. This
hypothesis is supported by our analysis of the divergence of the four gentoo penguin lineages across thermal and salinity gradients to the north and south of the APF (Figs. S5–S8). The Pliocene-Pleistocene cooling, along with the expansion of ice shelves across the Southern Ocean (20), would have reduced connectivity among penguin populations and facilitated speciation across Pygoscelis, Spheniscus, Eudyptes and Aptenodytes between 2 and 5.5 Mya (Fig. 1).

During the Quaternary glaciations (1.8–0.01 Mya), sea-ice is thought to have reached c. 40ºS at the coast of South America (28). This extent of sea-ice would have promoted the northern expansion of Spheniscus to the subtropics, and subsequently enabled colonization of the Galápagos Islands, home to rich feeding resources for penguins due to the upwelling of cooler, nutrient rich waters. Strong north-flowing currents – the Humboldt Current and the Benguela Current – would have further facilitated penguin colonization of subtropical habitats in the Pacific (Galápagos) and Atlantic (southern Africa), respectively. Neither of these current systems penetrate beyond the equator; this, together with their sub-Antarctic origins and preference for cooler waters, may explain why crown-group penguins never successfully colonized the Northern Hemisphere.

The genes for which signals of positive selection were detected are associated with thermoregulation (e.g., vasoconstriction/vasodilation: ENPEP, MME, BDKRB2), osmoregulation (e.g., balance of fluids and salt: SCL6A19, ACE2, AGT), and diving capacity (e.g., oxygen storage: MB), thereby reflecting adaptations that enabled penguins to colonize both colder and warmer habitats, away from their ancestral thermal maximum of c. 9ºC (Fig. 1, S12-13). In Antarctica, emperor penguins are exposed to temperatures as low as -40ºC and forage in waters of -1.8ºC (29), whereas Galápagos penguins are exposed to SST >27ºC and air temperatures exceeding 40ºC. These high temperatures are mostly associated with El Niño events, which may cause heat stress, high mortality, and low recovery of the penguin colonies (30, 31). Regulating blood pressure selectively through the constriction of blood vessels can further conserve oxygen consumption and facilitate the maintenance of high core body temperatures (32).

Penguins spend most of their lives at sea, often performing prolonged dives while foraging. They store oxygen in their lungs, blood, and muscles (33), and their rates of
oxygen consumption can be very low (34). The two largest penguin species, the emperor and king penguins, can achieve depths of >300 meters and maximum dive durations of 22 and 8 minutes, respectively (33, 35, 36). Smaller penguin species tend to dive in shallow waters (<50 m) with dives of 1–2 minutes in duration (33), although the chinstrap penguins often dive off-shelf (>200 m, (37)). In this sense, nucleotide differences in Myoglobin (MB, overall positive selection, Z=2.645, p=0.005) across species groups could be associated with differences in diving capacity. For example, we found several non-synonymous substitutions that were common within Pygoscelis, Eudyptes, and Aptenodytes penguins, but differed in their amino acid composition (Fig. S15). It is possible that these non-synonymous mutations encode greater oxygen-binding capacity, which would facilitate the deep and prolonged dives performed by Aptenodytes and some species of Pygoscelis penguins compared with Eudyptes (higher dN/dS ratios, Table S13).

Our results suggest that adaptive evolution, implicating genes involved in multiple interconnected genetic pathways, has increased the foraging success and survival of penguin species across diverse temperature and salinity gradients. Foraging success is associated with reproductive success (38, 39) and also with survival during long periods of fasting while caring for eggs and chicks (40). Collectively, such adaptations would have enabled the radiation of penguin species across the Southern Hemisphere. Penguins have a remarkable evolutionary history. Their radiation from the temperate coasts of New Zealand and Australia into other parts of the Southern Hemisphere was facilitated by changes in global circulation patterns over the past 20 million years. Our analysis detected positive selection across several gene networks, suggesting that molecular adaptation enabled the establishment of penguin populations in Antarctic and tropical regions, and enhanced the ability of some species to dive deeply. Demographic reconstructions over the past million years show that most penguin species have declined during the severe ice conditions during the LGM in the Southern Ocean, a result concordant with that recovered for several other bird species (41, 42).

Our results suggest that penguins originated from areas with a maximum SST of 9°C and diversified over millions of years, occupying colder Antarctic and warmer tropical
waters. As such, it seems unlikely that locally adapted species will be able to keep pace with rapid climate change, especially as marine species may be more vulnerable to global warming than terrestrial species (43, 44). This vulnerability is especially pertinent for penguins, as illustrated by the massive mortality of Adélie penguin chicks and by the relocation of emperor penguins in response to suboptimal sea ice conditions (45, 46). As large-scale genomic studies become increasingly feasible and data for more sophisticated global climate models become available for niche modelling, the application of approaches like those in the present study hold significant promise for revealing new insights into the evolutionary history and climatic vulnerability of many of the world’s poorly understood taxonomic groups.

Materials and Methods
Detailed methods for each of the sections below are provided in the supplementary documents.

Genome sequencing and assembly
The genomes of 18 extant penguin species (22 individuals) as well as the southern giant petrel (*Macronectes giganteus*) were sequenced to ~30x coverage with 150 bp paired-end reads using an Illumina HiSeq X platform at MedGenome (USA; Table S1). We used the giant petrel as outgroup for the analyses reported below. Briefly, duplicate sequences were removed using Super Deduper ([https://github.com/dstreett/Super-Deduper](https://github.com/dstreett/Super-Deduper)) and the reads were then filtered. We aligned the resulting cleaned reads of each individual to the emperor penguin reference genome ([http://gigadb.org/dataset/100005](http://gigadb.org/dataset/100005); scaffold-level assembly) using LAST ([http://last.cbrc.jp/](http://last.cbrc.jp/)). The extent to which genome assemblies were complete was assessed using the Benchmarking Universal Single-Copy Orthologs, BUSCO v2 dataset (47)(Fig. S2) and KAT spectra-cn plots (48)(Fig. S3, S4). The CDS, intron and mitogenome sequences were extracted for each genome using the reference genome GFF. We extracted 120 bp Ultra Conserved Element (UCE) loci with 750 bp of length of the flanking sequence padded to each side with scripts from the PHYLUCE pipeline (49). Sequences were aligned using MAFFT (50).
**Estimation of phylogeny, divergence times, and interspecific introgression**

To account for potential genome-wide incompatibilities between taxa and loci, we used Astral III (48) to estimate species tree phylogenies for each of the UCE, intron, and CDS data sets, and for all three data sets combined (Fig S8). We used RAxML-NG (v. 0.5.1b BETA) (51) to generate independent gene trees for each locus of the UCE, intron, and CDS alignments. As input to Astral III, we merged all of the “best” trees generated by RAxML-NG into a single file (Supplement). For the phylogenomic concatenated analysis using all the data from UCEs, CDS, intron, and the mitogenome we carried out maximum likelihood (ML) analyses in IQ-TREE, (52), with 1000 bootstrap replicates (Fig. S5-S7). We estimated divergence times in BEAST v2.5.2 (53) using the computer resources available through the CIPRES Science Gateway (Fig S7, S10) (54) calibrating the topology with five fossils (Table S8). This phylogeny was used for all subsequent analysis that required this information. Analyses to investigate the extent of interspecific introgression across the phylogeny were performed using a partitioned $D$-statistics approach implemented in DFOIL (55) for possible taxa combinations (eight) respecting the priors of a symmetrical tree composed of four taxa and an outgroup, one ingroup clade being younger than the other (Fig. 2, S9, Table S7). To perform the tests, we split the genome-wide alignments into 100 kb, non-overlapping windows with Bedtools and custom scripts.

**Ancestral Distribution and Niche Reconstruction**

For the historical biogeographic analysis, we estimated the ancestral range of the extant penguin species in the R package BioGeoBEARS (56) implementing three models of ancestral area reconstruction with and without long distance dispersal (the parameter j: “jump dispersal”). We subdivided the extant penguin geographic distribution into 10 different areas. Occurrence records for all penguin species and six marine variables were used to created raw models with MaxEnt–Javascript (57). Niche overlap was estimated between all penguin species for the set of variables considered. We used the package ‘Phyloclim’ to create Predicted Niche Occupancy (PNO) profile values and plots. Subsequently we combined this information with the phylogenetic tree to generate the
Divergence Through Time (DTT) plots and climatic tolerance chronograms depicted in Figure 1.

Detection of signatures of positive selection

We performed a dN/dS ratio test using the CodeML algorithm implemented in ETE3 (58, 59) for all species and sub-groups in the phylogeny. Due to the large number of analyzed genes, we performed a multiple comparisons (false-discovery rate [FDR]) test implemented in R. Genes that persisted on the list (i.e. remained significantly different from neutral expectations, supporting a positive selection regime) were then used to perform a gene ontology analysis using WebGestalt (WEB-based GEne SeT AnaLysis Toolkit) (60) and network analysis performed with StringDB (61) (Figure S14-S15).

Demographic history

To address questions about how climate may have influenced effective population size for each penguin species, we performed a demographic analysis using a pairwise sequential Markovian coalescent (PSMC) method (62). PSMC was run with parameters “-N25 -t15 -r5 -p 4 + 25*2 + 4 + 6”, and an estimated generation time (g) for the different penguin species (Figure 4, S16, Table S2) using a substitution rate derived from chicken pedigrees (63).

References

1. Egevang C, et al. (2010) Tracking of Arctic terns <em>Sterna paradisaea</em> reveals longest animal migration. Proceedings of the National Academy of Sciences 107(5):2078-2081.
2. Péron C & Grémillet D (2013) Tracking through Life Stages: Adult, Immature and Juvenile Autumn Migration in a Long-Lived Seabird. PLOS ONE 8(8):e72713.
3. Thomas DB & Fordyce RE (2012) Biological Plasticity in Penguin Heat-Retention Structures. Anatomical Record-Advances in Integrative Anatomy and Evolutionary Biology 295(2):249-256.
4. Gavryushkina A, et al. (2017) Bayesian Total-Evidence Dating Reveals the Recent Crown Radiation of Penguins. Systematic Biology 66(1):57-73.
5. Baker AJ, Pereira SL, Haddrath OP, & Edge KA (2006) Multiple gene evidence for expansion of extant penguins out of Antarctica due to global cooling. Proc Biol Sci 273(1582):11-17.
6. Clarke JA, et al. (2007) Paleogene equatorial penguins challenge the proposed relationship between biogeography, diversity, and Cenozoic climate change. Proc Natl Acad Sci U S A 104(28):11545-11550.

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

8. Subramanian S, Beans-Picon G, Swaminathan SK, Millar CD, & Lambert DM (2013) Evidence for a recent origin of penguins. Biol Lett 9(6):20130748.

9. Cole TL, et al. (2019) Mitogenomes Uncover Extinct Penguin Taxa and Reveal Island Formation as a Key Driver of Speciation. Mol Biol Evol 36(4):784-797.

10. Pan H, et al. (2019) High-coverage genomes to elucidate the evolution of penguins. GigaScience 8(9).

11. Figueiró HV, et al. (2017) Genome-wide signatures of complex introgression and adaptive evolution in the big cats. Science Advances 3(7):e1700299.

12. 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. Journal of Vertebrate Paleontology 38(2):e1445636.

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

14. Vianna JA, et al. (2017) Marked phylogeographic structure of Gentoo penguin reveals an ongoing diversification process along the Southern Ocean. Mol Phylogenet Evol 107:486-498.

15. Frugone MJ, et al. (2018) Contrasting phylogeographic pattern among Eudyptes penguins around the Southern Ocean. Sci Rep 8(1):17481.

16. Munro KJ & Burg TM (2017) A review of historical and contemporary processes affecting population genetic structure of Southern Ocean seabirds. Emu - Austral Ornithology 117(1):4-18.

17. Frugone MJ, et al. (2019) More than the eye can see: Genomic insights into the drivers of genetic differentiation in Royal/Macaroni penguins across the Southern Ocean. Molecular Phylogenetics and Evolution 139:106563.

18. Clucas GV, et al. (2018) Comparative population genomics reveals key barriers to dispersal in Southern Ocean penguins. Molecular Ecology 27(23):4680-4697.

19. Pertierra L, et al. (under review) Integrated phylogenomic and niche analyses reveal cryptic speciation in Gentoo penguins driven by local adaptation. Molecular Ecology.

20. Halanych KM & Mahon AR (2018) Challenging Dogma Concerning Biogeographic Patterns of Antarctica and the Southern Ocean. Annual Review of Ecology, Evolution, and Systematics 49:355-378.

21. Dunstan PK, et al. (2018) Global patterns of change and variation in sea surface temperature and chlorophyll a. Scientific reports 8(1):14624-14624.

22. Weisrock DW, Harmon LJ, & Larson A (2005) Resolving deep phylogenetic relationships in salamanders: analyses of mitochondrial and nuclear genomic data. Syst Biol 54(5):758-777.

23. Oliveros CH, et al. (2019) Earth history and the passerine superradiation. Proceedings of the National Academy of Sciences 116(16):7916-7925.

24. Simeone A, et al. (2009) Heterospecific Pairing and Hybridization between Wild Humboldt and Magellanic Penguins in Southern Chile (BIOONE) pp 544-550, 547.
Goldner A, Herold N, & Huber M (2014) Antarctic glaciation caused ocean circulation
changes at the Eocene-Oligocene transition. *Nature* 511(7511):574-577.

Dalziel IWD, *et al.* (2013) A potential barrier to deep Antarctic circumpolar flow until
the late Miocene? *Geology* 41(9).

Crame JA (2018) Key stages in the evolution of the Antarctic marine fauna. *Journal of
Biogeography* 45(5):986-990.

Glasser NF, Jansson KN, Harrison S, & Kleman J (2008) The glacial geomorphology and
Pleistocene history of South America between 38°S and 56°S. *Quaternary Science
Reviews* 27(3):365-390.

Williams CL, Hagelin JC, & Kooyman GL (2015) Hidden keys to survival: the type,
density, pattern and functional role of emperor penguin body feathers. *Proceedings of the
Royal Society B-Biological Sciences* 282(1817).

Boersma PD (1998) Population trends of the Galapagos penguin: Impacts of El Nino and
La Nina. *Condor* 100(2):245-253.

Boersma D (1975) Adaptations of Galapagos penguins for life in two different
environments. *The biology of penguins*, ed Stonehouse B (Macmillan, London), pp 101-
114.

Butler PJ & Jones DR (1997) Physiology of diving of birds and mammals. *Physiological
Reviews* 77(3):837-899.

Ponganis PJ & Kooyman GL (2000) Diving physiology of birds: a history of studies on
polar species. *Comparative Biochemistry and Physiology a-Molecular and Integrative
Physiology* 126(2):143-151.

Williams CL, Meir JU, & Ponganis PJ (2011) What triggers the aerobic dive limit?
Patterns of muscle oxygen depletion during dives of emperor penguins. *The Journal of
Experimental Biology* 214(11):1802-1812.

Wieneke B, Robertson G, Kirkwood R, & Lawton K (2007) Extreme dives by free-
ranging emperor penguins. *Polar Biology* 30(2):133-142.

Putz K & Cherel Y (2005) The diving behaviour of brooding king penguins (Aptenodytes
patagonicus) from the Falkland Islands: variation in dive profiles and synchronous
underwater swimming provide new insights into their foraging strategies. *Marine Biology
147(2):281-290.

Kokubun N, Takahashi A, Mori Y, Watanabe S, & Shin HC (2010) Comparison of diving
behavior and foraging habitat use between chinstrap and gentoo penguins breeding in the
South Shetland Islands, Antarctica. *Marine Biology* 157(4):811-825.

Kowalczyk ND, Reina RD, Preston TJ, & Chiaradia A (2015) Environmental variability
drives shifts in the foraging behaviour and reproductive success of an inshore seabird.
*Oecologia* 178(4):967-979.

Berlincourt M & Arnould JPY (2015) Influence of environmental conditions on foraging
behaviour and its consequences on reproductive performance in little penguins. *Marine
Biology* 162(7):1485-1501.

Thiebot JB, *et al.* (2014) Adjustment of pre-moult foraging strategies in Macaroni
Penguins Eudyptes chrysophalus according to locality, sex and breeding status. *Ibis
156(3):511-522.

Nadachowska-Brzyska K, Li C, Smeds L, Zhang G, & Ellegren H (2015) Temporal
Dynamics of Avian Populations during Pleistocene Revealed by Whole-Genome
Sequences. *Curr Biol* 25(10):1375-1380.
42. Cole TL, et al. (2019) Receding ice drove parallel expansions in Southern Ocean penguins. Proceedings of the National Academy of Sciences 116(52):26690-26696.

43. Pinsky ML, Eikeset AM, McCauley DJ, Payne JL, & Sunday JM (2019) Greater vulnerability to warming of marine versus terrestrial ectotherms. Nature 569(7754):108-111.

44. Wiens JJ (2016) Climate-Related Local Extinctions Are Already Widespread among Plant and Animal Species. PLoS biology 14(12):e2001104-e2001104.

45. Cimino MA, Lynch HJ, Saba VS, & Oliver MJ (2016) Projected asymmetric response of Adélie penguins to Antarctic climate change. Scientific Reports 6:28785.

46. Fretwell PT & Trathan PN (2019) Emperors on thin ice: three years of breeding failure at Halley Bay. Antarctic Science 1:1-6.

47. Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, & Zdobnov EM (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31(19):3210-3212.

48. Mapleson D, Garcia Accinelli G, Kettleborough G, Wright J, & Clavijo BJ (2017) KAT: a K-mer analysis toolkit to quality control NGS datasets and genome assemblies. Bioinformatics (Oxford, England) 33(4):574-576.

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

50. Katoh K & Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30(4):772-780.

51. Kozlov AM, Darriba D, Flouri T, Morel B, & Stamatakis A (2018) RAxML-NG: A fast, scalable, and user-friendly tool for maximum likelihood phylogenetic inference. bioRxiv:447110.

52. Nguyen L-T, Schmidt HA, von Haeseler A, & Minh BQ (2015) IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Molecular Biology and Evolution 32(1):268-274.

53. Bouckaert R, et al. (2014) BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput Biol 10(4):e1003537.

54. Miller MA, Pfeiffer W, & Schwartz T (2010) Creating the CIPRES Science Gateway for inference of large phylogenetic trees. 2010 Gateway Computing Environments Workshop (GCE), pp 1-8.

55. Pease JB & Hahn MW (2015) Detection and polarization of introgression in a five-taxon phylogeny. Systematic Biology 64(4):651–662

56. Matzke N (2013) BioGeoBEARS: BioGeography with Bayesian (and Likelihood) Evolutionary Analysis in R Scripts. University of California, Berkeley, Berkeley, CA.

57. Phillips SJ, Anderson RP, & Schapire RE (2006) Maximum entropy modeling of species geographic distributions. Ecological Modelling 190(3-4):231-259.

58. Huerta-Cepas J, Serra F, & Bork P (2016) ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. Mol Biol Evol 33(6):1635-1638.

59. Yang Z (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. Computer applications in the biosciences : CABIOS 13(5):555-556.

60. Wang J, Vasaikar S, Shi Z, Greer M, & Zhang B (2017) WebGestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. Nucleic Acids Res 45(W1):W130-W137.
Acknowledgments: We thanks to Andrea Polanowski, Claudia Godoy and Klemens Pütz for assistance with sample acquisition; Funding: Financial support was provided by INACH RT_12–14, Fondecyt Project 1150517, GAB PIA CONICYT ACT172065, NSF DEB-1441652, French Polar Institute Paul-Emile Victor (IPEV, Progs. 137 and 354), and a Lakeside grant from the California Academy of Sciences;

Data and materials availability: Penguins and giant petrel raw fastq reads, reconstructed genomes (BioProject PRJNA530615, BioSample accession SAMN11566608-SAMN11566630) and mitogenomes (MK760983-MK761004, MK761006) were deposited in Genbank. All UCE, CDS, intron and mitogenome alignments, dated phylogenies and codes for all data analyses are available at Dryad (https://doi.org/10.5061/dryad.pk0p2ngj2). All other data needed to evaluate the conclusions in this paper are present either in the main text or the supplementary materials.

Supplementary Materials:
Materials and Methods
Figures S1-S16
Tables S1-S13
References (1-42)
**Fig. 1. Evolutionary history of penguins.** Phylogenetic hypothesis of penguin species and divergence time estimates using UCE dataset. Red arrows represent the four fossil calibration points (fifth point corresponds to the node with the outgroup which not represented). Each node is represented by the ancestral distribution before the cladogenesis event using Ancestral range reconstruction based on the best-fit model (DIVALIKE+J) and is associated with one or more of the ten geographic locations depicted on the map in the lower right (different letters A–J and colors); areas at branch tips represent the current range of species. Average world surface temperature in the past is represented by the white graph behind the phylogeny (41) and onset of the strengthening of the Antarctic Circumpolar Current (ACC) by a dashed red line. At the top right: Ecological niche disparity through time (DTT) for penguins (top right), with the phylogeny projected onto niche parameter space on the y-axis (maximum surface water temperature).
temperature) with predicted niche occupancy (PNO) over time (x-axis) reconstructed for internal nodes.

Fig. 2. Summary of introgression among penguin taxa. Arrows represent the percentage of introgression (>2%) between taxa for all 8 combinations of the five-taxon statements evaluated (see Fig. S9, Table S7).
Fig. 3. Analysis of genes under positive selection in extant penguin lineages. The main panel depicts results from the network analysis of positively selected genes, retrieving two connected main clusters, one associated with general cellular functions (green) and another grouped genes related to functions associated with osmoregulation (renal function), immunity, thermoregulation (e.g. blood pressure) and diving ability (e.g. oxygen metabolism) which are classified at the right pie-chart around the penguin.
**Fig. 4. Demographic history of penguins.** Pairwise Sequentially Markovian Coalescent (PSMC) plot depicting the demographic history of each lineage inferred from genomic data represented by different colored lines (see Fig. S16 for bootstrapped curves). The period of the Naynayxungla glaciation, Penultimate glaciation and the Last glaciation are represented in the plots, as well as the Last Glacial Maximum (LGM).