















Comparative Genomics Supports Ecologically Induced Selection as a Putative Driver of Banded Penguin Diversification

Fabiola León ^{1,2,3,4} Eduardo Pizarro ^{1,2,3,4} Daly Noll ^{1,2,3,4} Luis R. Pertierra ^{3,5}
Patricia Parker ⁶ Marcela P.A. Espinaze ⁷ Guillermo Luna-Jorquera ^{8,9}
Alejandro Simeone ¹⁰ Esteban Frere ¹¹ Gisele P.M. Dantas ¹² Robin Cristofari ¹³
Omar E. Cornejo ¹⁴ Rauri C.K. Bowie ^{15,*} Juliana A. Vianna ^{1,2,3,4,*}

¹Pontificia Universidad Católica de Chile, Facultad de Ciencias Biológicas, Instituto para el Desarrollo Sustentable, Santiago, Chile

²Millennium Institute Center for Genome Regulation (CRG), Santiago, Chile

³Millennium Institute of Biodiversity of Antarctic and Subantarctic Ecosystems (BASE), Santiago, Chile

⁴Millennium Nucleus of Patagonian Limit of Life (LiLi), Santiago, Chile

⁵Department of Biogeography and Global Change, Museo Nacional de Ciencias Naturales (MNCN-CSIC), Madrid, Spain

⁶Department of Biology, University of Missouri St. Louis and Saint Louis Zoo, St. Louis, MO 63121-4400, USA

⁷Department of Conservation Ecology and Entomology, Faculty of AgriScience, Stellenbosch University, Stellenbosch 7602, South Africa

⁸Center for Ecology and Sustainable Management of Oceanic Islands (ESMOI), Departamento de Biología Marina, Universidad Católica del Norte, Coquimbo, Chile

⁹Centro de Estudios Avanzados en Zonas Áridas (CEAZA), Universidad Católica del Norte, Coquimbo, Chile

¹⁰Facultad de Ciencias de la Vida, Universidad Andrés Bello, Departamento de Ecología y Biodiversidad, Santiago, Chile

¹¹Centro de Investigaciones de Puerto Deseado, Universidad Nacional de la Patagonia Austral, Puerto Deseado, Argentina

¹²PPG Biologia de Vertebrados, Pontificia Universidade Católica de Minas Gerais, Belo Horizonte, MG 30535-901, Brazil

¹³Institute of Biotechnology, HiLIFE, University of Helsinki, Helsinki, Finland

¹⁴Department of Ecology and Evolutionary Biology, University of California, Santa Cruz, CA 95060, USA

¹⁵Museum of Vertebrate Zoology and Department of Integrative Biology, University of California, Berkeley, CA 94720-3160, USA

*Corresponding authors: E-mails: jvianna@uc.cl; bowie@berkeley.edu.

Associate editor: Rebekah Rogers

Abstract

The relative importance of genetic drift and local adaptation in facilitating speciation remains unclear. This is particularly true for seabirds, which can disperse over large geographic distances, providing opportunities for intermittent gene flow among distant colonies that span the temperature and salinity gradients of the oceans. Here, we delve into the genomic basis of adaptation and speciation of banded penguins, Galápagos (*Spheniscus mendiculus*), Humboldt (*Spheniscus humboldti*), Magellanic (*Spheniscus magellanicus*), and African penguins (*Spheniscus demersus*), by analyzing 114 genomes from the main 16 breeding colonies. We aim to identify the molecular mechanism and genomic adaptive traits that have facilitated their diversifications. Through positive selection and gene family expansion analyses, we identified candidate genes that may be related to reproductive isolation processes mediated by ecological thermal niche divergence. We recover signals of positive selection on key loci associated with spermatogenesis, especially during the recent peripatric divergence of the Galápagos penguin from the Humboldt penguin. High temperatures in tropical habitats may have favored selection on loci associated with spermatogenesis to maintain sperm viability, leading to reproductive isolation among young species. Our results suggest that genome-wide selection on loci associated with molecular pathways that underpin thermoregulation, osmoregulation, hypoxia, and social behavior appears to have been crucial in local adaptation of banded penguins. Overall, these results contribute to our understanding of how the complexity of biotic, but especially abiotic, factors, along with the high dispersal capabilities of these marine species, may promote both neutral and adaptive lineage divergence even in the presence of gene flow.

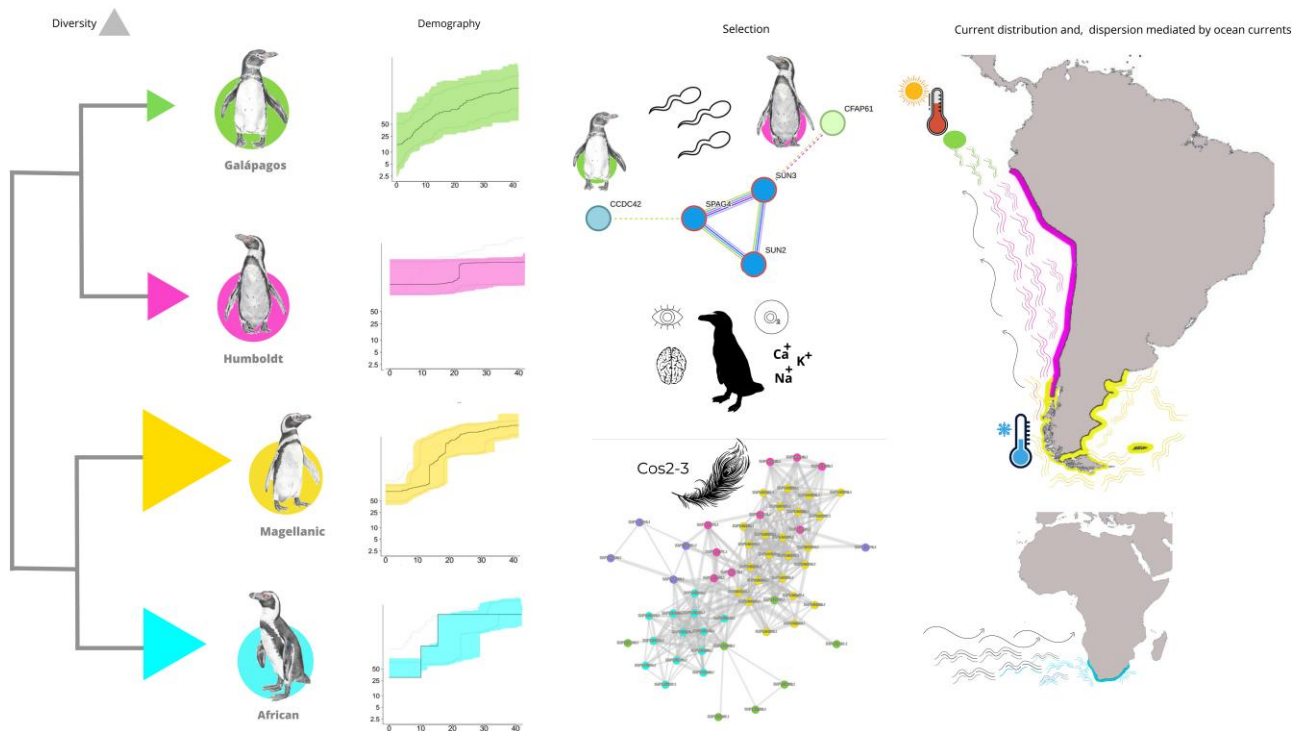
Received: November 21, 2023. Revised: June 27, 2024. Accepted: July 10, 2024

© The Author(s) 2024. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Open Access

Graphical abstract



Key words: penguin, genomics, diversification, adaptation, ecological niche.

Introduction

Strong environmental gradients in temperature and salinity across oceanic basins and the geographic isolation of islands are key factors that underpin the remarkable evolutionary radiation of many phenotypically unique plants and animal lineages (Warren et al. 2015). Islands are also hypothesized to have been fundamental to the diversification of several seabird lineages across the globe (e.g. shearwaters, Procellariidae, Obiol et al. 2023; tropicbirds, Phaethontidae, Varela et al. 2024). Given the high dispersal capacity of seabirds with shallow genetic divergence among species being characteristic of several clades (e.g. gulls, Laridae, Sonsthagen et al. 2016; shags, Phalacrocoracidae, Rawlence et al. 2022; skuas, Stercorariidae, Mikkelsen and Weir 2023), it seems that long periods of isolation in allopatry are unlikely to be the only, or even the most important, driver of lineage diversification in seabirds, thereby fundamentally differing from drivers of lineage formation among landbirds (Cai et al. 2020). This raises the question of the relative importance of accumulating genetic differences through genetic drift versus through natural selection (local adaptation) in shaping lineage formation among seabirds. This question is particularly pertinent given that both genetic drift and local adaptation are modulated by the extent of gene flow among populations and lineages (Via 2009). While the extensive movement of seabirds across large geographical

ranges facilitates genetic exchange among populations, factors such as oceanic barriers, degree of population isolation, and specialized breeding behaviors (prezygotic barriers) can lead to divergence by restricting gene flow (Welch et al. 2012; Danckwerts et al. 2021; Kersten et al. 2021).

The mechanisms driving the accumulation of differences between populations can be prezygotic (e.g. morphology and breeding behavior) or postzygotic (mechanical incompatibility and infertility), often working synergistically to result in reproductive isolation. During the early stages of speciation, genetic divergence between populations is expected to occur at a few key loci (Via and West 2008; Feder et al. 2012). In the late stages of speciation, genome-wide divergence is likely to be observed due to selection acting on multiple adaptive loci to restrict gene flow (Feder and Nosil 2010). The sex chromosomes are among the genomic regions that undergo rapid differentiation, primarily due to their smaller effective population size that makes selective sweeps more likely. Divergence of sex chromosomes is further influenced by the presence of loci associated with reproduction located on these chromosomes, which can ultimately lead to postzygotic isolation through positive selection acting on these loci (Dufresnes and Crochet 2022).

Banded penguins (genus *Spheniscus*) are exceptional for their ability to have successfully colonized both tropical and temperate latitudes and are one of the youngest

lineages among living penguins having originated along the South America coast some 1.8 million years ago (Mya) (Vianna et al. 2020). Four species are recognized in the genus. Three species are distributed along the coast of South America, including the Galápagos (*Spheniscus mendiculus*), Humboldt (*Spheniscus humboldti*), and Magellanic penguins (*Spheniscus magellanicus*), and the African penguin (*Spheniscus demersus*) is distributed along the south and west coasts of South Africa and Namibia. The Galápagos penguin holds the distinction of being the northernmost penguin species, straddling the equator (García-Borboroglu and Boersma 2013). Magellanic penguins range across both the Pacific and Atlantic coasts of South America, while the Humboldt penguin is restricted to the Humboldt Current in the Pacific Ocean.

Here, we evaluate several classic speciation models, often formulated for landbirds, using banded penguins as an exemplar system with which to further our understanding of the speciation process in seabirds. First, we explore support for peripatric speciation between Galápagos and Humboldt penguins, representing an island–continent system. Second, we explore the evidence for parapatric speciation across the range boundary between Humboldt and Magellanic penguins. Third, we explore the extent of speciation in allopatry between the sister species African and Magellanic penguins. We hypothesize that speciation events in banded penguins were driven not only by geographical distance between continents or continent–islands but also by environmental heterogeneity, leading to selection across the genome to facilitate adaptation to local conditions. This is particularly significant given that most extant penguin lineages are adapted to the low water temperatures and high salinity of the sub-Antarctic and Antarctic (Thomas et al. 2020; Vianna et al. 2020; Cole et al. 2022), while banded penguins must navigate a trade-off between inhabiting subtropical and tropical regions and maintaining thermal equilibrium during reproductive seasons, especially in the face of heat-waves and shifts in local salinity levels.

Results

A total of 114 penguin genomes were obtained, covering a range of per-sample depth of coverage values between 3× and 7× (see [supplementary figs. S1 and S2](#) and [table S1, Supplementary Material](#) online). On average, 7 individuals from each of the 16 major breeding colonies were sequenced, thereby enabling us to sample from across the geographic range encompassed by each of the 4 banded penguin species ([Fig. 1](#); [supplementary table S2, Supplementary Material](#) online).

Diversity and Diversification

Nucleotide diversity, heterozygosity, and the number of private alleles were lowest in the Galápagos penguin and greatest in the Magellanic penguin ([Fig. 1](#); [supplementary figs. S3 and S4](#) and [tables S3 and S4, Supplementary](#)

[Material](#) online). The percentage of sequence dissimilarity within species was lower in the Galápagos penguin (<0.004%) relative to the other 3 banded penguin species ([supplementary fig. S4a, Supplementary Material](#) online). Relatedness coefficients were consistently highest among Galápagos penguins and lowest among Magellanic penguins ([supplementary fig. S4b, Supplementary Material](#) online). Tajima's D was positive for all 4 species, indicating a deficit of rare alleles, a result consistent with population contraction ([supplementary fig. S4c, Supplementary Material](#) online).

We performed a principal component analysis (PCA) with data set 1B (refer to [supplementary table S3, Supplementary Material](#) online). The PCA revealed that individuals were clustered by species ([Fig. 2a](#)), where the first principal component explained 27.7% of the variance and the second principal component explained 9.4% of the variance. Sex-linked sites show higher divergence between the Magellanic and African penguins than between Humboldt and Galápagos penguins ([supplementary figs. S5 and S6, Supplementary Material](#) online). Phylogenetic analyses performed on 4,740 ultraconserved elements (UCE) and 16,966 coding sequences (CDS) each placed the Galápagos and Humboldt penguins as sister species and the African and Magellanic penguins as sister species ([Fig. 2b](#); [supplementary fig. S7, Supplementary Material](#) online), supporting previous phylogenomic hypotheses of species relationships (Pan et al. 2019; Vianna et al. 2020; Cole et al. 2022).

Pairwise comparisons of the 2D site frequency spectrum further indicated that the Galápagos and Humboldt penguins are sister species, with the Galápagos and African penguins exhibiting the highest level of sequence divergence compared to the other species pairs examined ([supplementary fig. S8, Supplementary Material](#) online). The results from admixture analyses recovered $K = 4$ as the optimal number of clusters. This result corroborates the PCA and phylogenomic analyses, with all 4 species being delimited with little indication of recent interspecific admixture ([Fig. 2c](#)). In contrast, the results from Treemix recovered 2 vectors whose placement suggests that gene flow may have occurred over the entire history of the diversification of banded penguins. A vector with low migration weight (close to 0) extends from the ancestral taxon of the Humboldt–Galápagos clade to the Magellanic penguin ([Fig. 2d](#)), and there is an intraspecific vector between populations of Humboldt penguin. Results from using the ABBA-BABA test (D-statistic Z-score > 3) are consistent with the Treemix results, suggesting that interspecific gene flow over the evolutionary history of banded penguins has occurred between Humboldt and Magellanic penguins ([supplementary table S5, Supplementary Material](#) online).

The degree of intraspecific population structure observed among sampled colonies spanning each species distributional range was limited, with levels of genomic differentiation between populations within species (F_{st}) varying between 0.001 and 0.006 ([supplementary table S6, Supplementary Material](#) online). PCA, EEMS, and

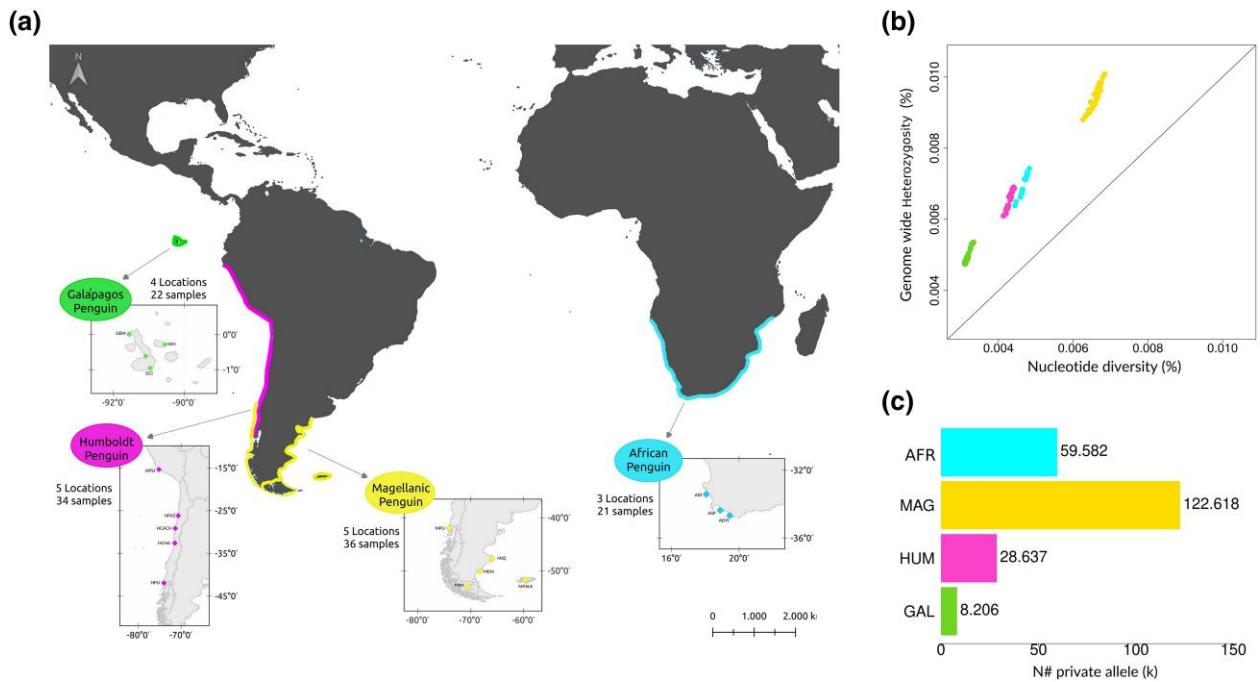


Fig. 1. Geographic distribution and genetic diversity of banded penguin species. a) Geographic distribution of the breeding banded penguin colonies sampled in this study with the number of individuals sampled indicated. Color codes correspond to the 4 banded penguin (*Spheniscus*) species. b) Nucleotide diversity vs. genome-wide heterozygosity. c) Number of private alleles restricted to each species: African (AFR), Magellanic (MAG), Humboldt (HUM), and Galápagos Penguin (GAL).

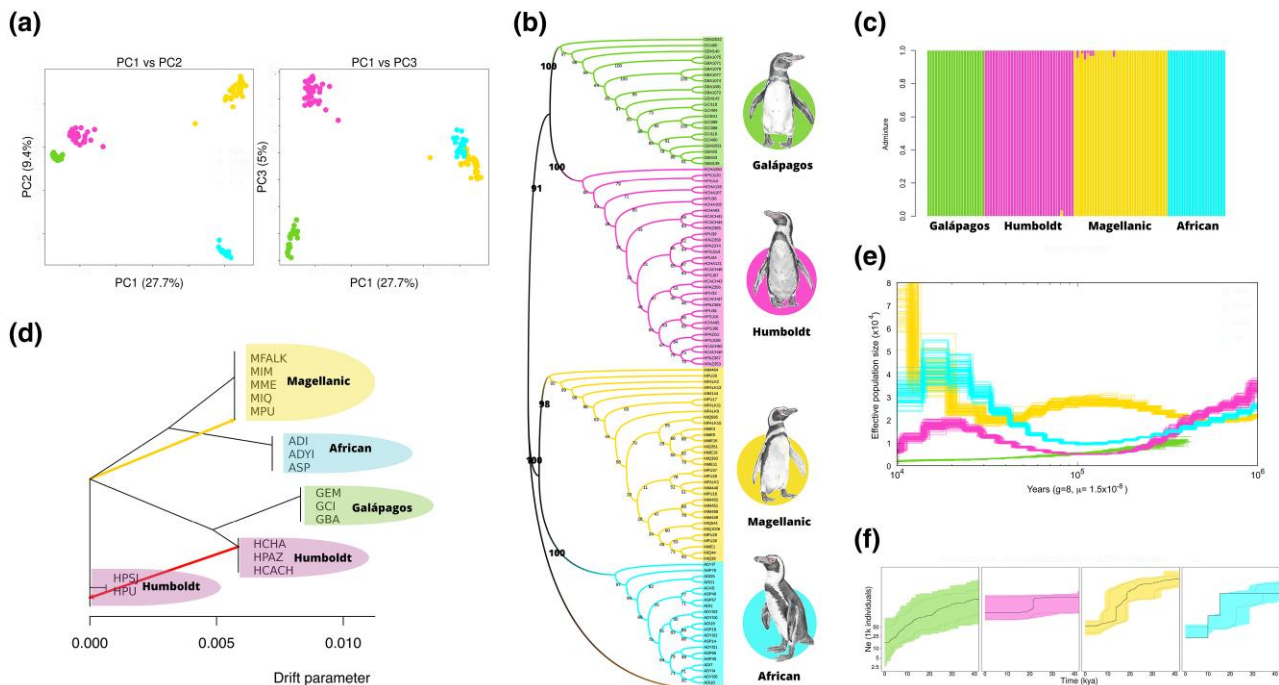


Fig. 2. Population structure and demographic history of banded penguins. a) PCA based on genome-wide SNP data. b) Maximum likelihood topology generated with 4,700 UCES. c) Estimates of admixture ($K = 4$). d) Maximum likelihood tree generated using Treemix, with the most significant ancestral periods of introgression indicated. Note that here we made use of neutral gene regions from whole-genome sequence data from a subset of individuals belonging to each of the breeding colonies. e) Demographic history inferred from the PSMC using a single genome of high coverage ($\sim 30\times$). f) Demographic inference carried out with the set of populations belonging to the same species with stairway plot 2. Breeding colonies of Galápagos penguin: El Muñeco (GEM), Caleta Iguana (GCI), and Bartolomé (GBA); Humboldt penguin: Punta San JUAN (HPSJ), Pan de Azucar (HPAZ), Isla Chañaral (HCHA), Isla Cachagua (HCACH), and Puñihuil (HPU); Magellanic penguin: Puñihuil (MPU), Isla Magdalena (MIM), Malvinas/Falkland Island (MFALK), Monte Entrance (MME), and Isla Quiroga (MIQ); African penguin: Dassen Island (ADI), Stony Point (ASP), and Dyer Island (ADYI).

BayesAss also support limited population structure within species (supplementary table S7 and figs. S9 and S10, Supplementary Material online).

Demographic Inference

Both the pairwise sequentially Markovian coalescent (PSMC) models (Li and Durbin 2011) (Fig. 2e) and the stairway plot 2 models (Liu and Fu 2020) (Fig. 2f) supported past changes in population size for each of the 4 banded penguin species. In the PSMC analysis, the effective population size (N_e) of the Galápagos penguin shows a consistent decline toward the present. Humboldt and African penguins reached a low N_e point around 100,000 years before present (BP), after which both species started to increase. In contrast to the other species, the Magellanic penguin shows an expansion between 500,000 and 100,000 years BP. Stairway plots indicated that all 4 banded penguin species experienced a putative decline in N_e at approximately 20,000 years BP, during the last glacial maximum. However, confidence intervals indicate considerable uncertainty for the Humboldt penguin which may have maintained a stable population size. Magellanic, African, and Galápagos penguins have continued to decline over the past 40,000 years. Demographic inferences of individuals from each sampled population (supplementary fig. S11, Supplementary Material online) are consistent with species-wide inferences of changes in N_e (Fig. 2e).

Detection of Outlier Loci, Gene Family Expansion, and Contraction

A total of 1,532 single nucleotide polymorphisms (SNPs) on autosomal scaffolds (between Galápagos–Humboldt, Humboldt–Magellanic, and Magellanic–African) were identified through species pair comparisons using 3 commonly used methods to detect outlier loci: OUTFLANK (Whitlock and Lotterhos 2015), PCAdapt (Luu et al. 2017), and GWDS (de Jong et al. 2021) (Fig. 3; supplementary table S8, Supplementary Material online). The top 4 enriched biological processes (Fig. 3a) were cellular process (GO:0009987), biological regulation (GO:0065007), metabolic process (GO:0008152), and response to stimulus (GO:0050896). SNPs were distributed across CDSs, genes, messenger RNAs (mRNA), and pseudogenes (supplementary table S8, Supplementary Material online). Depending on the species pair compared, between 364 and 645 SNPs were retrieved by all 3 methods (Fig. 3a1 to a3; supplementary table S8, Supplementary Material online). Signals of selection associated with these biological functions are linked to the ecological habits of the banded penguins, highlighting molecular adaptation to hypoxia, osmoregulation, and visual and olfactory stimuli, as well as muscular development and cognitive capabilities related to social behaviors, memory, and learning (Fig. 3b; supplementary tables S9 and S10, Supplementary Material online). F_{st} of outlier SNPs was lower between Galápagos and Humboldt penguins and higher between Humboldt and Magellanic

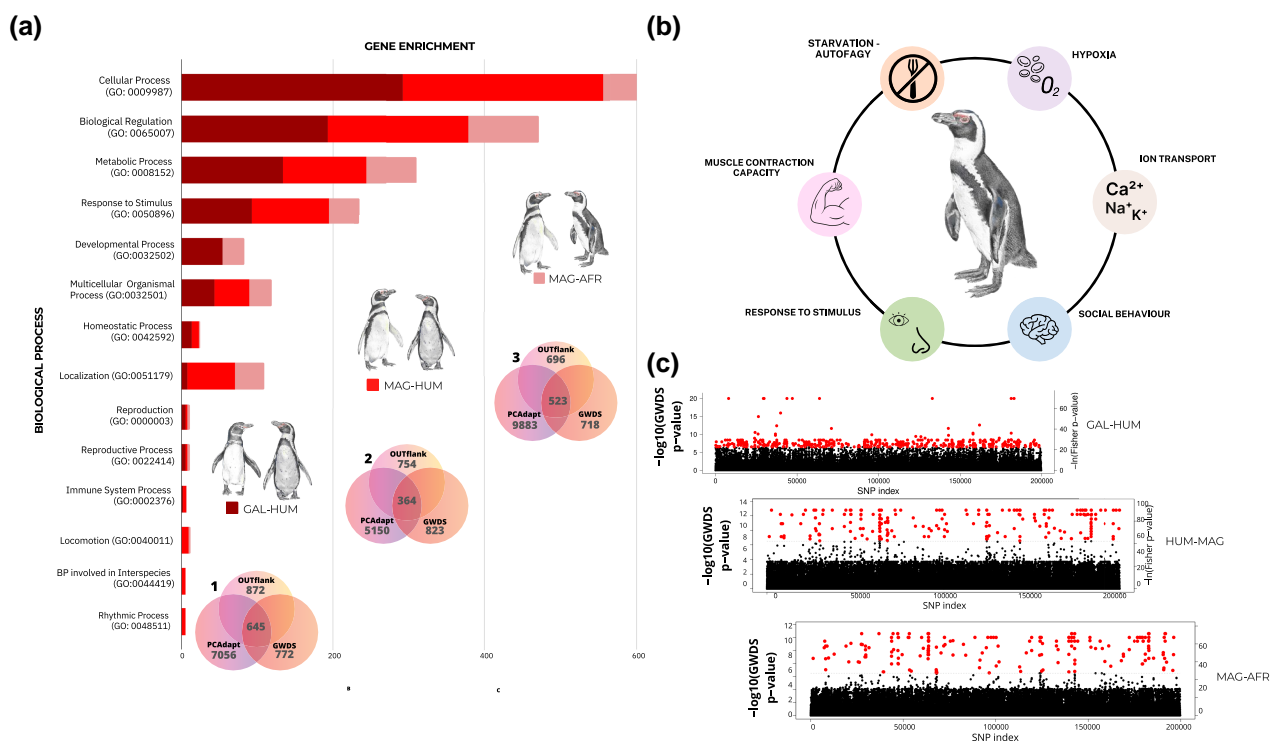


Fig. 3. Whole-genome divergent selection between pairwise *Spheniscus* species. a) The primary GO and enrichment analysis conducted in the selection analysis between species (a1 to a3). Venn diagrams illustrate the number of outlier SNPs recovered using OUTFLANK, PCAdapt, and GWDS and their intersection. b) The biological functions inferred to be under selection in all 4 banded penguin species. c) Manhattan plots illustrate the SNPs between species and their degree of differentiation.

and Magellanic and African penguins (Fig. 3c). The biological processes showing the highest gene enrichment between Galápagos and Humboldt species (Fig. 3a) were spermatogenesis (GO:0007283), response to heat (GO:0031072), and morphogenetic regulation (GO:0009653; [supplementary table S11, Supplementary Material online](#)). Comparison between Galápagos and Humboldt penguin recovered 645 SNPs under selection; 183 (28%) SNPs under divergent selection are involved in spermatogenesis (e.g. *SPAG4*, *ADCY10*, *MROH2B*, *DRC1*, *SUN3*, *ARHGAP24*, *HSD17B11*, *YTHDC1*, *SUN2*, *CFAP61*, and *CCDC42*; Fig. 3b), with several of these loci functionally linked to each other (Fig. 3; [supplementary table S9, Supplementary Material online](#)).

A Human Phenotype Ontology analysis revealed enriched phenotypic characteristics associated with the smaller body sizes of Galápagos penguins compared to Humboldt penguins. Among the most prominent phenotypic features were “growth delay,” “microcephaly,” “short digit,” “hypogonadism,” “short nose,” “hypothyroidism,” “short toe,” “short finger,” “short neck,” and “small nail.”

We used molecular data to identify the sex of each individual revealing a ratio of 52% males to 48% females in our data set. We found 4 sexual scaffolds inferred to be “Z” VULB01013104.1 ($P = 3.96446519251287e-83$), VULB01007854.1 ($P = 3.54081731216743e-102$), VULB01004053.1 ($P = 7.39397430999582e-136$), and VULB01013990.1 ($P = 3.77227071588691e-65$). In males (2 copies of the Z-chromosome), 622 SNPs were recovered across the 4 putative Z-chromosome scaffolds for the Galápagos–Humboldt (60 SNPs), Humboldt–Magellanic (219 SNPs), and Magellanic–African penguin comparisons (343 SNPs) ([supplementary fig. S12, Supplementary Material online](#)). The pairwise comparisons indicate a diverse range of biological processes in common, including molecular transducer activity binding (GO:0005488), structural molecule activity (GO:0005198), and catalytic activity (GO:0003824). Among the uniquely enriched biological processes for the Z-chromosome scaffolds in pairwise species comparisons are GAL-HUM molecular function regulator activity (GO:0098772), HUM-MAG cytoskeletal motor activity (GO:0003774), ATP-dependent activity (GO:0140657), and MAG-AFR translation regulator activity (GO:0045182). We confirmed the presence of a gene block on the Z-chromosome (*DCC*, *MEX3C*, *POLI*, *MAPK4*, *PARP8*, *RAB27B*, and *TCF4*) that has been lost in other bird lineages such as Galliformes and passerines (Fig. 4; [Friocourt et al. 2017](#); [Patthey et al. 2017](#)).

Restricting the analyses to females (single copy of the Z-chromosome), 828 SNPs were detected to be under selection on the Z-chromosome across the Galápagos–Humboldt (159 SNPs), Humboldt–Magellanic (293 SNPs), and Magellanic–African (376 SNPs) comparisons ([supplementary fig. S13, Supplementary Material online](#)). The findings suggest a broad array of shared biological processes, including binding (GO:0005488), structural molecule activity (GO:0005198), molecular function regulator activity (GO:0098772), and catalytic activity (GO:0003824).

OrthoFinder ([Emms and Kelly 2019](#)) identified 299 gene families that are shared among all banded penguin species (Fig. 5; [supplementary tables S12 to S25, Supplementary Material online](#)), whereas 325 gene families were categorized as single-copy orthogroups including all species. Gene family expansion and contraction were investigated using CAF5 ([Mendes et al. 2021](#)), with the gene birth rate estimated to be 0.00100 when accounting for duplications per gene per million years. Employing this approach, a total of 888 gene families across the 5 species (4 banded species and little penguin as outgroup) were found to have experienced noteworthy expansion, and 1719 experienced contraction events (Fig. 5). The primary biological and molecular processes that have accelerated the rate of evolution among species are associated with intermediate filament bundle assembly (P -value $5.05E-27$), visceral muscle development (P -value $2.62E-09$), and nucleosome assembly (P -value 0.00; [supplementary fig. S14, Supplementary Material online](#)). Several of the expanding gene families among banded penguin species could be involved in biological processes related to the response to environmental stimuli such as olfactory receptors (Fig. 5; [supplementary figs. S15a and S16, Supplementary Material online](#)), feather diversification (Fig. 5d; [supplementary fig. S15b, Supplementary Material online](#)), osmoregulation, and visual receptors, among others (e.g. histone gene families, Fig. 5e). The Galápagos penguin exhibits 1,009 rapidly evolving gene families, consisting of 299 gene family expansions and 710 contractions. Notably, the expanded gene families in Galápagos penguins were predominantly associated with biological processes such as spermatogenesis (P -value $2.9E-04$; Fig. 5; [supplementary fig. S17, Supplementary Material online](#)), gamete generation (P -value $2.14E-08$), and reproduction structures (P -value $7.7E-32$). Additionally, the results suggest an exclusive expansion and accelerated evolutionary rate in the heat shock gene families of the Galápagos penguin and genes involved with spermatogenesis and osmoregulation (Fig. 5f to i). In terms of gene families that have contracted, these genes encoded various proteins such as odorant receptors, keratin proteins, and sodium chloride–channel proteins, among others.

Comparisons of the Occupancy of Niche Space among Banded Penguin Species

The Galápagos and Humboldt penguins show uneven dynamic niche occupation (Fig. 6; [supplementary figs. S18 to S24 and tables S26 to S31, Supplementary Material online](#)). The ecological niche of the Humboldt penguin spans a much higher range of environmental tolerances, whereas the niche occupied by the Galápagos penguin is restricted to the environment of the Galápagos’ Islands ([supplementary fig. S18, Supplementary Material online](#)). This results in the Galápagos penguin occupying a novel niche (Schöener’s D overlap of 10%) compared to the Humboldt penguin. In the Galápagos penguin, the species occupancy of islands near the equator represents a unique thermal niche expansion that comprises 99% of its

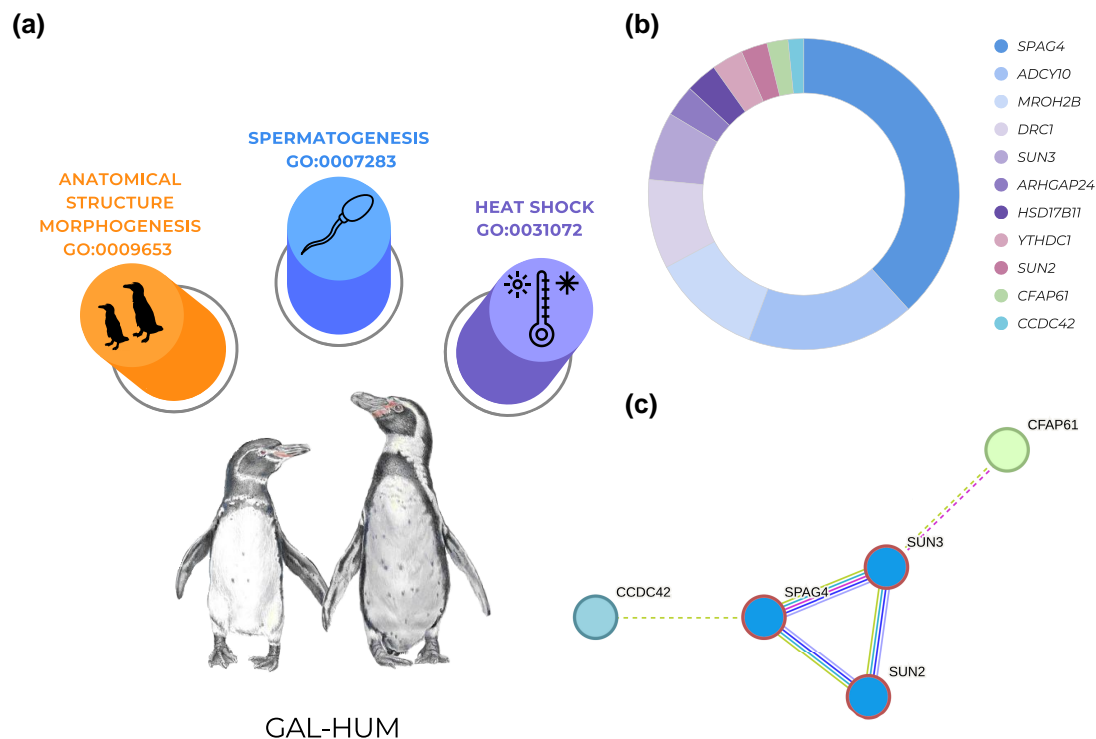


Fig. 4. Enrichment results of the SNPs under divergent selection between Galápagos and Humboldt penguins. a) The GO of the most relevant biological processes, along with SNPs under selection, identified throughout the genomes of Galápagos and Humboldt penguins. b) The relative proportion of SNPs under selection, linked to genes associated with functions pertaining to spermatogenesis and sperm motility. c) Associations between spermatogenesis loci revealing functional relationships between genes. Interaction types include direct interactions, gene expression regulation, or participation in shared biochemical pathways.

niche space (supplementary table S26, Supplementary Material online). This difference is visualized as a thermal gap in Fig. 6 showing that the colonization of the Galápagos Islands represents a steep transition from its sister species, the Humboldt penguin, by being adapted to sea-water that is on average 2 to 4 degrees warmer than the waters occupied by the Humboldt penguin (supplementary fig. S18, Supplementary Material online). Differences observed in the chlorophyll content of the sea are not as pronounced, with 25.6% overlap in trophic levels between Galápagos and Humboldt penguins (supplementary table S27, Supplementary Material online).

Magellanic and Humboldt penguins represent the only pair of banded penguin species that have a partially sympatric range, coexisting along the Southern Pacific coastline of South America. Consequently, their macroecological niche overlap is the highest among banded penguins, reaching 53%, and their thermal niche overlap is 55% comprising nearly 100% of the niche stability (retained niche space) of Humboldt over Magellanic penguins. Thus, Humboldt penguins do not occupy any unique niche thermal space with respect to Magellanic penguins (Fig. 6; supplementary table S5, Supplementary Material online). The unique part of the Magellanic penguin thermal niche (some 37%) can be attributed to the cold waters (down to 5 °C) of the southernmost parts of its range, where Humboldt penguins do not reach. The higher thermal limit of Humboldt and Magellanic penguins is remarkably

similar, showing a high degree of niche conservatism, with neither species occupying waters >20 °C suggesting this isotherm may be an important ecophysiological barrier (table S26, Supplementary Material online). The salinity and chlorophyll levels encompassed by the distributional ranges of Humboldt and Magellanic penguins overlap (50% and 74%; supplementary figs. S19 and S20, Supplementary Material online) with a small expansion of the niche for Magellanic penguins (12% and 20%; supplementary fig. S21 and tables S27 to S29, Supplementary Material online).

Magellanic and African penguins consistently exhibit a high degree of niche overlap with the Humboldt penguin, with percentages of 53% and 32%, respectively. Additionally, they share an 18% overlap with each other. The thermal niche, salinity, and chlorophyll levels of the niche space occupied by these species remain consistent, indicating a strong niche conservatism in environmental conditions (supplementary figs. S22 to S24 and tables S26 to S29, Supplementary Material online).

We also recovered niche differences among populations within species that are consistent with the observation of 2 genetic clusters obtained by EEMS for Galápagos, Humboldt, and Magellanic penguins (supplementary figs. S22 to S24 and tables S30 and S31, Supplementary Material online).

The results of the redundancy analysis (RDA) reveal a statistically significant association of a set of environmental

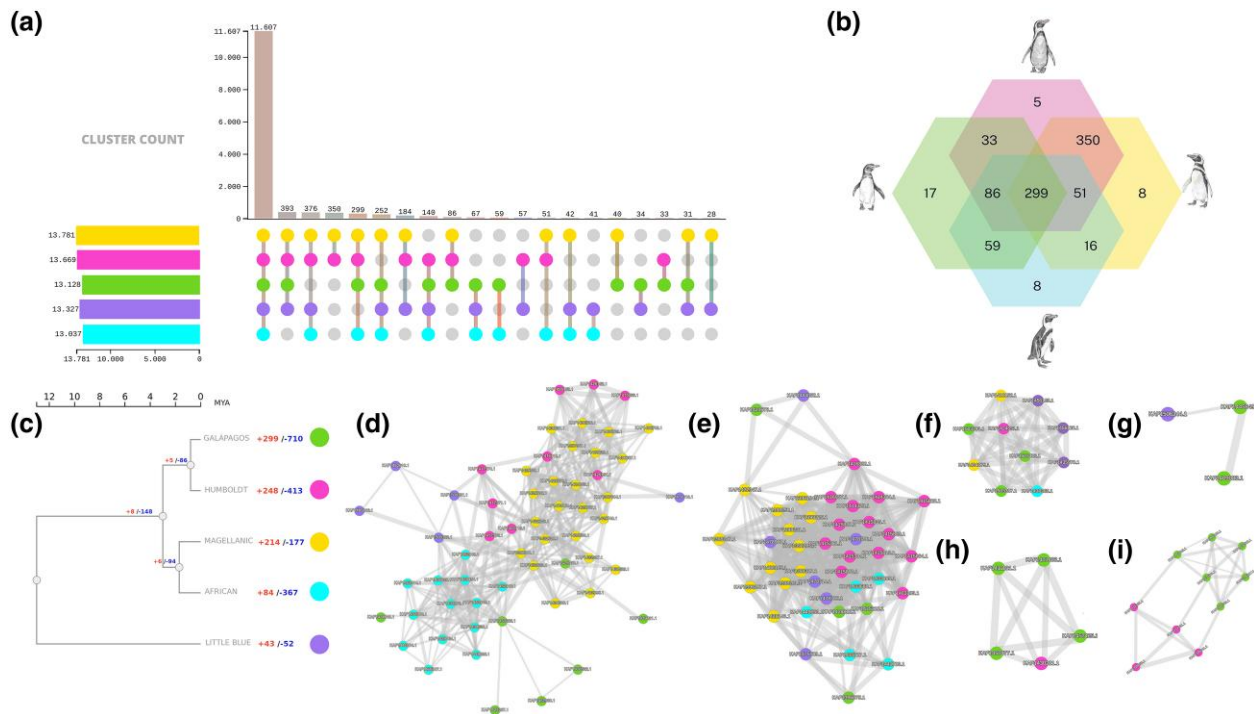


Fig. 5. Comparative gene family analyses among banded and little blue penguins. a) Cluster and protein counts of orthogroups found among banded and little blue penguins. b) Cluster and protein count of orthogroups found among banded penguins. c) Maximum likelihood tree based on single-copy genes showing the gain (red) and loss (blue) of genes across banded penguin lineages. d) Feather keratin Cos2-3 family gene expansion. e) Histone H2B family gene expansion. f) The gene family associated with heat stress response includes the heat shock transcription Y-linked gene. (g and h) The adenylate cyclase type 10 and chromodomain Y-like gene are part of the spermatogenesis expanded gene family. i) The gain in the osmoregulation gene family involved sodium channel genes.

and genetic variation among banded penguin species (supplementary figs. S25 and S26 and tables S26 to S29, Supplementary Material online). These results from an RDA suggest that Galápagos penguins tend to be positively associated with the average speed of present ocean currents (VELO) and the annual mean temperature variable (BIO1), while Magellanic penguin tends to be positively correlated with chlorophyll levels in the ocean (CHLO).

Discussion

Young species assemblages that occur across environmental gradients provide excellent systems to investigate molecular and ecological mechanisms that drive local adaptation. Changes in climate across the southern oceans from the Pleistocene to the present likely led to the current distribution of banded penguins (Vianna et al. 2020). Our results indicated that gene flow following postglacial range expansion appears to have facilitated genetic homogenization of allelic diversity within species, leading to minimal spatial genetic structure within each banded penguin species. Further, our results indicate that both genetic drift among geographically isolated species and local adaptation of populations across ecological gradients have facilitated the diversification of banded penguins and helped to maintain species boundaries in these highly mobile seabirds.

Our analyses suggest that origins of the Galápagos penguin are consistent with the expectations of peripatric speciation. The demographic inference of Galápagos penguins

indicated a low effective population size over several hundred thousand years, which is indicative of a small founding population with genetic drift leading to the loss of low-frequency alleles. We hypothesize that Galápagos penguins colonized the Galápagos Islands long after the origin and establishment of the cold Humboldt Current (between the Neogene and the climatic fluctuations of the Quaternary; Thiel et al. 2007) and likely followed the northerly direction of the Humboldt Current and became isolated in the Galápagos Islands after a rare open ocean dispersal event.

Glacial cycles likely significantly disrupted the Humboldt Current due to the presence of ice sheets along the Southern Pacific coast of South America as suggested by geological evidence (Rabassa et al. 2000) and effective population size reduction in other marine species (Pardo-Gandarillas et al. 2018; Dantas et al. 2019; Weinberger et al. 2021). Effective population size of endemic species reliant on one of the world's largest upwelling systems, such as the Humboldt penguin, appears to have been impacted by the drop in sea surface temperature and disruption of the Humboldt Current during glacial cycles. The changes to oceanic conditions altered nutrient supply and thereby likely influenced the reproductive success and population dynamics of coastal breeding populations of penguins (Kim et al. 2002; Thiel et al. 2007; Montecino and Lange 2009; Dantas et al. 2019).

The bathymetric differences between the Atlantic coast compared to the Pacific coast likely resulted in the exposure of the Patagonian Continental Shelf during glacial conditions (Violante et al. 2014). This geographical feature is

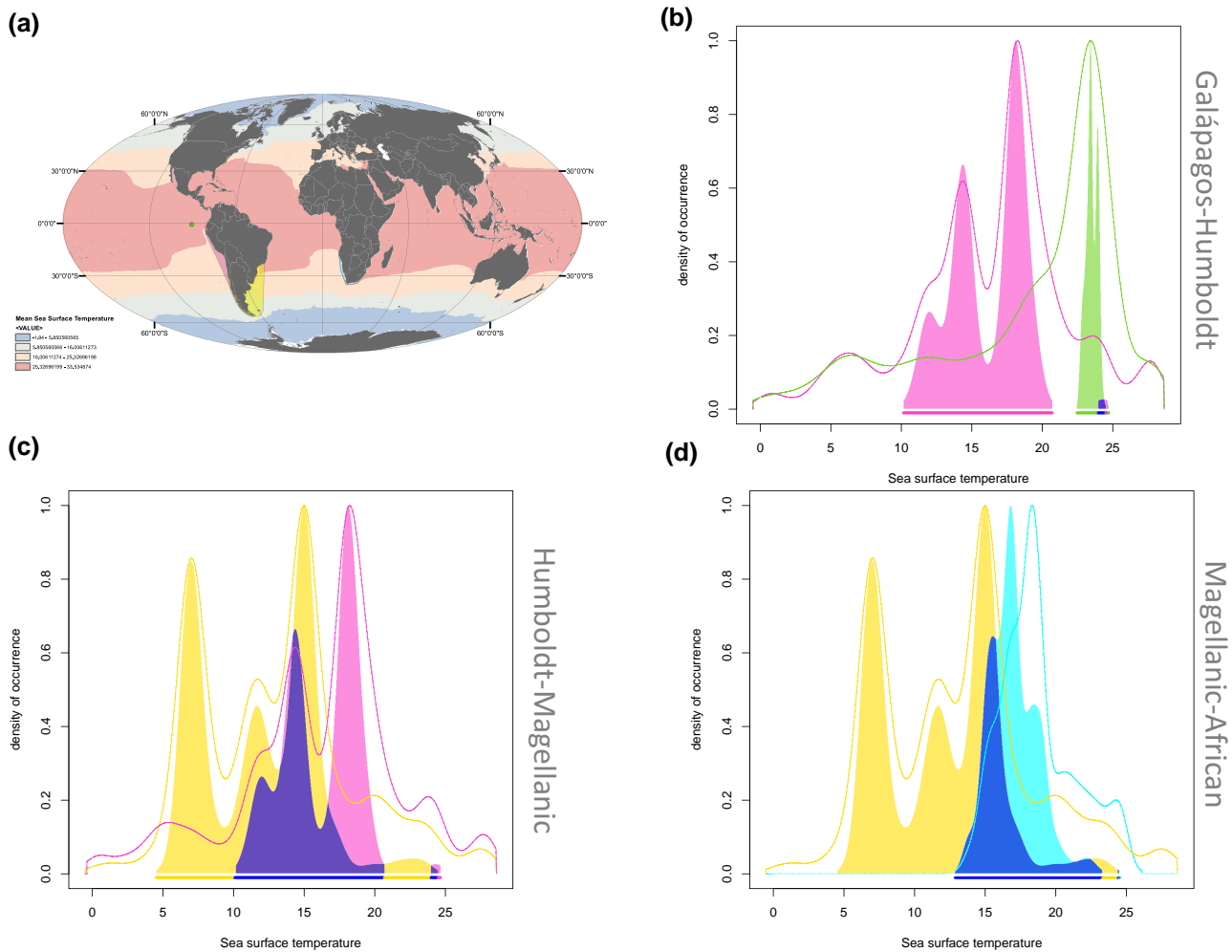


Fig. 6. Banded penguin ecological niche overlap. a) Sea surface temperature across the globe, with breeding colony locations where environmental data were obtained. Graphical representation of the extent of niche overlap (in purple and dark blue) in pairwise comparisons between banded species: b) Galápagos penguins (green) vs. Humboldt (pink); c) Humboldt (pink) vs. Magellanic penguins (yellow); d) Magellanic (yellow) vs. African penguins (light blue).

postulated to have provided emerging nesting sites for intertidal marine vertebrates such as sea lions (Hoffman et al. 2016), rockhopper penguins (Vianna et al. 2020) and Magellanic penguins along the Atlantic coast of South America (Fig. 2) as sea levels fell during the mid-Pleistocene glaciations (Cavallotto et al. 2011).

The limited but ongoing gene flow, the adjacent geographic range with overlapping ecological niches, and ecological niche conservatism support our hypothesis that Humboldt and Magellanic penguins likely diverged in parapatry. Interglacial periods in the South American Austral region would have facilitated secondary contact (Sersic et al. 2011; González-Wevar et al. 2016; Ceballos et al. 2016) and hybridization where species ranges overlap (Simeone et al. 2009; Hibbets et al. 2020; Vianna et al. 2020). Our findings reveal that while such hybridization is infrequent, it has likely persisted for thousands of years, with directional gene flow from Humboldt to Magellanic penguins (Fig. 2d). This result is consistent with the emerging viewpoint that

admixture between different lineages is common and widespread in recently diverged species (Richards et al. 2019), highlighting the relevance of speciation with gene flow as an important mechanism of lineage divergence in seabirds.

Previous studies indicate that South America served as the point of origin for banded penguins (*Spheniscus*; Vianna et al. 2020), with the Benguela current playing a crucial role in facilitating dispersal between South America and Africa, thereby promoting speciation through allopatry during the Pleistocene. Our results suggest that after penguins reached Southern Africa, African penguins experienced decreases in effective population size during glacial periods, which may have impacted reproductive success during adverse climatic periods (Barrable et al. 2002; Quick et al. 2022).

The analyses investigating selection across the genome suggest that outlier SNPs, encompassing diverse gene functions, may be undergoing positive or purifying selection in banded penguins. These include loci associated with

biological processes related to hypoxia, osmoregulation, and sensory perception, which could be advantageous for diving birds, enabling penguins to detect prey within deep and dimly lit waters. These genetic adaptations appear to align with patterns observed for several seabird species (Pincemy et al. 2009; Coffin et al. 2011; Höglund et al. 2019; Mendelson and Safran 2021; Cole et al. 2022; Sin et al. 2022). Additionally, positive selection and gene family expansion for loci associated with environmental stimuli, such as olfactory and visual receptors, feather keratin, and inner ear balance, suggest the potential reliance of penguins on a combination of sensory cues to navigate their environments and locate food sources, although further research is warranted to confirm these hypotheses.

According to our outlier analyses, genetic variants at specific positions within the genes *SPAG4*, *ADCY10*, *SUN3*, *MROH2B*, and *DRC1*—known for their roles in spermatogenesis, sperm motility, and resilience to heat shock (Frohnert et al. 2011; Calvi et al. 2015; Kenigsberg et al. 2017; Yang et al. 2018; Akbari et al. 2019; Pereira et al. 2019; Zhang et al. 2021)—might play a crucial role in upholding species boundaries between Galápagos and Humboldt penguins. Several gametogenesis genes tend to evolve rapidly, are under positive selection (Swanson et al. 2003; Dean et al. 2008; Kopania et al. 2022), and have been described as “speciation genes” for some model species of birds (Irwin 2018), mammals (Mihola et al. 2009), and insects (Orr et al. 2004; Presgraves 2008). Our results suggest that the signals of selection on loci associated with gametogenesis have been stronger and more frequent in island endemic species such as the Galápagos penguin.

Environmental temperature could have shaped how spermatogenesis genes accumulate genomic differences among populations (Tao et al. 2007; Qvarnström et al. 2016). For instance, male infertility has been related to high temperatures that deform and kill sperm cells (Cai et al. 2021; Schou et al. 2021; Gao et al. 2022). Galápagos penguins experienced the highest maximum sea surface temperatures during the Pleistocene glaciations relative to the remaining banded penguin species (Lyle et al. 1992; Liu and Herbert 2004; Vianna et al. 2020). Thus, selection on alleles at loci associated with spermatogenesis may have been necessary to maintain fertility at higher sea surface temperatures for the Galápagos penguin and thereby facilitated divergence from the Humboldt penguin by increasing the degree of postzygotic isolation in this young species pair. Results from our RDA analyses suggest that currently, annual mean temperatures tend to be positively associated with genotype traits in Galápagos penguins. These findings, coupled with those from the ecological niche analyses, suggest that niche divergence related to temperature, along with signals of selection in genes associated with spermatogenesis, may have been key factors in the ecological speciation of Galápagos penguins. Furthermore, our results suggest that species boundaries are being reinforced by the apparent current absence of gene flow between Galápagos and Humboldt penguins. The absence of gene flow is most likely due to the presence

of a thermal isolation barrier (Fig. 6), that Humboldt penguins do not cross despite their ability to travel long distances at sea.

The functional enrichment analysis using the Human Phenotype Ontology database suggests that selection is also acting on loci associated with morphogenesis between Galápagos and Humboldt penguins, particularly in how anatomical structures are generated and organized (supplementary table S11, Supplementary Material online). A reduction in body size may be related to the “island syndrome,” a term that refers to the rapid evolution of reduced dispersal capacity and smaller body sizes associated with high environmental temperatures and dampened temperate extremes on islands (Pörtner and Farrell 2008). This effect is particularly pronounced in amphibious seabirds such as penguins, impacting their ability to achieve thermoneutrality on land and the metabolic costs associated with heat generation while in the sea (Stahel and Nicol 1982; Bevan et al. 2002; Fahlman et al. 2005). Thermal body homeostasis in penguins is influenced by factors such as the high thermal conductivity and specific heat capacity of water during dives, which far exceeds that of air (Stahel and Nicol 1982). Heat stress can also result from extended periods on land during breeding and molting (Wilson and Wilson 1990; Roberts 2016). Notably, body size is crucial for coping with cold stress, governed by the surface area-to-volume ratio (Oswald and Arnold 2012). This factor may restrict certain species, like the little penguin, to temperate regions (Stahel and Nicol 1982) and potentially also be influencing the distribution of Galápagos penguins. Penguins employ diverse mechanisms to counteract heat loss, including dense packing of feathers with no space between feather tracts, layers of blubber, and reduced blood flow to the exposed appendages, especially in response to cold water in upwelling zones (Stonehouse 1967). However, these adaptations present challenges in dissipating heat within burrows, potentially causing thermal stress for banded penguins on land (Boersma and Rebstock 2014; Holt and Boersma 2022; Shaun and Pichegru 2023; Welman and Pichegru 2023). This is particularly pertinent for Galápagos penguins, which may encounter air temperatures exceeding 40 °C (Boersma 1976; Kemper et al. 2007; Boersma and Rebstock 2014).

Based on our findings, we emphasize the importance of identifying the presence of repeated gene blocks on Z sex chromosomes. These blocks, comprised of 12 genes each, have reportedly been lost due to chromosomal rearrangement in other bird lineages and could have significant implications for adaptation and survival in species that retain them (Friocourt et al. 2017; Patthey et al. 2017). Our results suggest that these gene blocks may be under selection in banded penguins, despite the strong purifying selection operating on sex chromosomes. However, caution should be exercised when interpreting signals of selection associated with the sex chromosomes, as faster genetic drift due to the smaller effective population size of sex chromosomes may lead to false signatures of selection (Dean et al. 2015).

Positive selection and gene family expansion for loci associated with the response to environmental stimuli, such as those involved in olfactory and visual receptors, feather keratin, and inner ear balance, suggest that foraging penguins rely on a combination of olfactory, visual, integument, and mechanoreceptors to effectively navigate their environments and locate upwelling waters rich in food sources. While visual perception has long been recognized as crucial for avian foraging, recent research has shed light on the equally significant role of olfactory receptors in Procellariiformes (Sin et al. 2022) and penguins (Coffin et al. 2011) in facilitating foraging. Visual receptors in birds are vital for detecting colors, shapes, and movement of prey when foraging in low-light environments such as underwater, as well as for identifying and avoiding predators such as seals (Höglund et al. 2019, Cole et al. 2022). Differences in visual acuity, feather density and coloration (plumage), and facial patterning could play a role in species-specific mate choice (Pincemy et al. 2009) through sexual selection and thereby serve as important mechanisms in the formation and maintenance of prezygotic reproductive isolation among these young species (Mendelson and Safran 2021). Olfactory receptors enhance avian foraging by allowing birds to detect a variety of odors associated with food availability and also could facilitate the homing ability of seabirds, enabling them to navigate back to their breeding colonies (Silva et al. 2020; Bonadonna and Gagliardo 2021).

Conclusions

Our research revealed a spectrum of genetic divergence among banded penguin species, characterized by differing degrees of shared genetic variation between lineage pairs. These results suggest recent isolation and independent evolutionary trajectories in diverse environments for the 4 species, with support for peripatric, parapatry, and allopatric models of speciation. Isolation and divergence in temperature niches between Galápagos and Humboldt penguins may have expedited the speciation process between these evolutionary young lineages, potentially through selection acting on genes associated with prezygotic reproductive isolation. Signatures of selection acting on loci associated with hypoxia, osmoregulation, social behavior, responses to external environmental stimuli, thermoregulation, autophagy, and starvation appear to be shared traits under selection in banded penguins. Future research should prioritize obtaining highly contiguous genomes (i.e. chromosomes) to determine the importance of structural variation on species adaptation. While we have explored the potential significance of the genes we identified to be under selection and gene family dynamics linked to the putative diversification of banded penguins, a comprehensive understanding of their evolutionary history necessitates incorporation of data from transcriptomic analyses and functional experiments. Such data would significantly improve our understanding of the underlying

molecular adaptations that have enabled banded penguins to occupy waters that span from the subantarctic to the equator.

Materials and Methods

Sampling

A total of 114 blood samples were collected from 16 breeding colonies of the 4 species of *Spheniscus* penguins: Galápagos, Humboldt, Magellanic, and African penguins (Fig. 1; supplementary table S2, Supplementary Material online). Individuals were captured, and approximately 1 to 1.5 mL of blood was extracted by puncturing the brachial vein or the dorsal aspect of the foot, after which the birds were released. The blood was preserved by adding 96% ethanol, or lysis buffer, or by spotting onto Fast Technology for Analysis of nucleic acids cards.

DNA Isolation and Genomic Library Preparation

DNA was isolated from blood samples using a phenol-chloroform or salt extraction protocol (Aljanabi and Martinez 1997). The concentration and integrity of the isolated genomic DNA were determined using a Qubit spectrophotometer (Thermo Fisher Scientific) and by 1% agarose gel electrophoresis. Genomic DNA was fragmented using an ultrasonicator. Once fragmented, the DNA was processed for the purpose of constructing paired-end libraries with the Illumina TruSeqNano kit. The genomic fragments obtained were ligated to polyA tails, index adapters, and barcodes. Six cycles of polymerase chain reaction were carried out, and the amplified libraries were bead purified and then quantified with the Qubit. The resulting libraries were sequenced at ~5× coverage with paired-end 151 base pair reads using an Illumina NovaSeq 6000 S4 platform at the University of California, Berkeley.

Quality Control

The quality of the raw DNA sequences was assessed using FastQC v0.11.4 (Andrews 2010) before and after read and quality trimming. Demultiplexing and the trimming of lower quality flanking regions were carried out (establishing a quality threshold of 25) using a sliding window of 4:15 with Trimmomatic (Bolger et al. 2014).

Resequencing and Variant Discovery

Reads were aligned against the little blue penguin *Eudyptula minor novaehollandiae* genome assembly VULB01 SAMN12384878 (accession: PRJNA556735 ID: 556735; Pan et al. 2019) using BWA-MEM (Li and Durbin 2009). Four of the banded penguin genomes, with approximately 31× coverage each (supplementary table S1, Supplementary Material online), were obtained from Vianna et al. (2020). The aligned SAM and BAM files were processed prior to SNP identification with SAMtools (Daneczek et al. 2021), Picard Tools, RealignerTargetCreator, and IndelRealigner (Van der Auwera et al. 2013) to remove

duplicates, correct relationship pair information, correct unmapped read flags, and obtain overall mapping statistics. We identified variants for each individual using BCFtools mpileup and BCFtools call implemented in BCFtools (Danecek et al. 2021). We used the approach of Nursyifa et al. (2022) to identify scaffolds comprising the sex chromosomes. This method is based on mapping sequence fragments using differences in relative read depth among scaffolds to identify contigs that map to the sex chromosomes. We recovered 4 sex scaffolds that mapped to the Z-chromosome “Z” VULB01013104.1 ($P = 3.96446519251287e-83$), VULB01007854.1 ($P = 3.54081731216743e-102$), VULB01004053.1 ($P = 7.39397430999582e-136$), and VULB01013990.1 ($P = 3.77227071588691e-65$). Variant site calling files (VCF) were filtered by quality, keeping only those with genotype quality equal to or greater than 30 in Phred score, and covered by at least 3 reads and a maximum of 7 reads.

Low coverage genomes ($\sim 5\times$) were used to estimate population structure, genetic diversity, and for selection scans, whereas both high ($4 \sim 30\times$ genomes) and low coverage genomes were used for demographic reconstruction.

The average coverage and average missing data for each individual genome are summarized in the [supplementary figs. S1 and S2, Supplementary Material](#) online, respectively. We identified 60,076,142 SNPs, and specific filters were used to optimize each data set for different analyses ([supplementary table S3, Supplementary Material](#) online). SNPs located on the scaffolds comprising the sex chromosomes represented 7% of the data set. Mitochondrial scaffolds were excluded from all analyses.

Population Structure

To characterize intra- and interspecific genetic variability, we estimated genomic diversity using the following estimators: (i) nucleotide diversity (π) estimated as the average number of nucleotide differences per site between 2 DNA sequences within a population (Nei 1975); (ii) heterozygosity; and (iii) Tajima’s D that compares 2 estimators of genetic diversity π (nucleotide diversity) and θ which is the Watterson’s estimator of the population mutation rate (Tajima 1989) ([supplementary fig. S3, Supplementary Material](#) online). Genome-wide heterozygosity was computed using the formula in the R package sambaR (de Jong et al. 2021).

Here, nH represents the count of heterozygous sites observed for an individual within the SNP data set, n_{ind} signifies the number of nonmissing data points for that individual within the SNP data set, $nsnps$ indicates the total number of SNPs in the data set, and $ntotal$ denotes the total number of sequenced sites, encompassing both monomorphic and polymorphic loci.

We used PCA to explore broad genetic affinities among the banded penguin individuals using Plink 1.9 (Purcell et al. 2007) (Fig. 2; [supplementary fig. S9, Supplementary Material](#) online). We estimated ancestry coefficients of the

different banded penguin individuals at both the species and population level with ADMIXTURE v1.3 (Alexander and Lange 2011). ADMIXTURE uses a cross-validation procedure to select the most probable number of clusters (K) that explain the structure of the data (Fig. 2).

The occurrence of admixture among lineages was further investigated using the interspecific SNP data set with Treemix v1.12 (Fig. 2). Treemix models the relationship among the sample populations with the ancestral population using genome-wide allele frequency data and a Gaussian approximation of genetic drift. The optimal m -value ($m = 2$) was estimated using the OptM R package (Fitak 2021).

To determine the extent of hybridization and introgression between Humboldt and Magellanic penguins, we used Dsuite (Malinsky et al. 2021) to obtain the D-statistic and f_4 -ratios. The D-statistic, also known as the ABBA-BABA test, is commonly used to assess evidence of gene flow between 2 taxa. Under this approach, 4 taxa are analyzed, and the ancestral alleles (“A”) and derivatives (“B”) are considered. The allelic patterns “ABBA” and “BABA” occur with the same frequency in a scenario without introgression, while the excess of either of the allelic patterns is considered to be indicative of introgression between 2 taxa and in the test is reflected by the D-statistic being significantly different from 0 ([supplementary table S5, Supplementary Material](#) online). The contemporary gene flow rates were estimated using BayesAss3-SNP (Musmann et al. 2019; [supplementary table S7, Supplementary Material](#) online). We set the number of iterations to 1,000,000, the burn-in to 100,000, and the delta values to 0.1.

Estimation of Effective Migration Surface

We analyzed patterns of gene flow among banded penguin populations in a spatial context to determine how landscape features impact genetic variation. We used EEMS (Petkova et al. 2016) to estimate gene flow across the landscape. We generated a biallelic matrix of genotypes with Plink 1.9 that was then transformed with bed2diff (available from the EEMS GitHub repository) into a genetic differentiation matrix. Each individual was georeferenced, and a habitat polygon was manually constructed with the help of the Google Maps API v3 Tool (<http://www.birdtheme.org/useful/v3tool.html>). The study area was covered with a dense regular grid composed of triangular demes. EEMS was run independently for each species using the `runeems_SNPS` script and the default setting for 10 million steps and a 1 million step burn-in with 400 demes. We used the R-scripts described at <https://github.com/dipetkov/reemplots2> to visualize the results.

Demographic History

We made use of PSMC version 0.6.5-r67 (Li and Durbin 2011) to reconstruct the demographic history of each species over deep time using both the high coverage (Fig. 2)

and low coverage genomes (supplementary fig. S11, Supplementary Material online). For the PSMC analysis, we first called variants for each individual. To do so, we used SAMtools version 1.3.1 with HTSLib 1.3.1 and the vcfutils.pl script from BCFtools 1.3.1 with the command “samtools mpileup -C50 -uf ref.fa aln.bam | BCFtools view -c - | vcfutils.pl vcf2fq -d 10 -D 100 | gzip > diploid.fq.gz”. Following the recommendation of the PSMC documentation (<https://github.com/lh3/psmc>), we used a third of the average read depth as the minimum read depth (-d) and at least twice the average read depth as the maximum read depth (-D) (-d 3 -D 12). The generated consensus fasta file was made through the fq2psmcfa command, and then, PSMC was run with parameters “-N25 -t15 -r5 -p 4 + 25*2 + 4 + 6”. Once we obtained the psmcfa files, sex-linked scaffolds and CDS regions were removed with the seqtk (Li 2012) and seqkit (Shen et al. 2016) tools. Prior to bootstrapping, we carried out the *splitfa* function, which is a tool for splitting a multi-FASTA file into individual files, with each file containing a single sequence, and then, the inference of population size history of each pseudohaploid fasta sequences was made with the *psmc* command. We assumed a nucleotide substitution rate of $m = 1.91 \times 10^{-9}$ substitutions/site/year based on the chicken lineage (*Gallus gallus*) multiplied by the generation time of banded penguin species ($g = 8$) as described by Vianna et al. (2020).

To estimate demographic history over more recent time periods (last 40,000 years), we made use of stairway plots v2.0 using the total set of neutral SNPs of all genomes (low and high coverage) for each species. This method makes use of a likelihood approach to determine values that best reproduce the observed SFS and then uses this information to estimate changes in N_e through time. We generated the frequency spectra of folded sites through ANGSD realSFS (Korneliussen et al. 2014). Stairway plot 2 was run on the folded SFS with the same mutation rate parameters and generation time estimates as used for the PSMC analyses.

Genome-Wide Locus Phylogeny

We obtained UCEs and CDSs from the pseudohaploid fasta files. We used BCFtools norm to align BAM reads to the left, to perform the normalization of the indels, and to check if alleles match the reference. Then, a fasta consensus sequence was generated for each individual with BCFtools consensus. We identified missing sites with bedtools genomecov and masked them with bedtools maskfasta. Once such sites were masked, UCE loci were extracted with PhyLUCE (Faircloth 2016) using these functions `phyluce_probe_run_multiple_lastzs_sqlite` and `phyluce_probe_slice_sequence_from_genomes`. UCEs present in more than 70% of the taxa were retained for analysis. CDS and exon loci were extracted with Gffreads (Pertea and Pertea 2020). Each of the UCEs, CDS, and exon data sets was separately aligned with MAFFT v7.245 (Katoh and Standley 2013). Phylogenetic trees were estimated

from the concatenated data obtained with catfasta2-phym1.pl <https://github.com/nylander/catfasta2phym1> using IQTREE v1.5.3 (Minh et al. 2020) which makes use of the maximum likelihood optimality criteria. The nucleotide substitution model used was GTR + G with branch support determined using an ultrafast bootstrap (Minh et al. 2013).

Detecting Signatures of Selection across the Genome

Given the recent divergence of the banded penguin species from each other, they constitute an ideal clade with which to identify regions of recent genomic differentiation and candidate loci under selection. We compared sister species Galápagos–Humboldt and Magellanic–African through *Fst*-based selection analyses because genetic divergence due to background drift is minimized. We also compared Humboldt–Magellanic penguins due to their recent introgression events. To screen as many SNPs as possible, we filtered the whole genome raw vcf files with the following parameters using vcfutils `-minQ 30 -minDP 3 -max-missing 1 -min count 2` and no missing data. Outlier analyses were conducted via several R packages, including OUTFLANK, that works to detect unusually high or low levels of genetic differentiation between populations through *Fst* pairwise genetic differentiation metrics using false discovery rate (FDR) to reduce false positive; PCAadapt, which combines PCA with linear regression; and GWDS that conducts a SNP-by-SNP analysis comparing allele frequencies between pairs of populations in a data set of biallelic SNPs with Bonferroni correction. We conducted an independent selection analyses on autosomes and sex scaffolds. The genomic positions of outlier SNPs were mapped using the reference genome *Eudyptula minor novaehollandiae* GFF file, which enabled the identification of genes, CDSs, mRNA, and other genome regions.

For analysis of selection on the sex scaffolds, we filtered the raw whole genome VCF files using VCFtools with the following parameters: `-minQ 30, -minDP 3, -max-missing 1, -min-count 2`, and no missing data. Then, we constructed a VCF file containing only the Z sexual scaffolds. After this, we generated separate VCF files for males (2 copies of the Z-chromosome) and females (a single copy of the Z-chromosome) and subsequently performed selection analyses based on *Fst*, similar to those described in the previous section.

We performed functional enrichment analyses including Gene Ontology (GO and Human Phenotype Ontology (HPO) using Uniprot (The UniProt Consortium 2015), PANTHER (Thomas et al. 2022), and gProfiler (Kolberg et al. 2020). We used Fisher’s exact test with FDR correction to compute significance of associations. GO terms in the categories of biological process, molecular function, and cellular component with a FDR of less than 0.05 were considered significantly enriched. We evaluate the functional interactions of proteins encoded by genes using STRING (Szklarczyk et al. 2021).

Protein Family Evolution Analyses

We used 5 genome protein sequences of penguin species, including Galápagos SAMN12384884, Humboldt SAMN12384883, Magellanic SAMN12384882, and African SAMN12384881 penguins to identify gene families (Pan et al. 2019). Little penguin was chosen as the outgroup. We employed OrthoFinder v3 (Emms and Kelly 2019) to infer orthogroups identifying a total of 11,607 gene families (Fig. 5a), encompassing a vast repertoire of 64,058 genes when compared to little penguin (supplementary tables S12 to S25, Supplementary Material online). The evolution of gene families (gain and loss) was analyzed using CAFE v5 (Mendes et al. 2021), with the lambda parameter used for calculating birth and death rates. We used single-copy genes to infer an ultrametric tree with FastTree2 (Price et al. 2010) and calibrated with the divergence time (13 Mya) of the most recent common ancestor between little penguin and Banded penguins, obtained from Vianna et al. (2020).

Ecological Niche Overlap between Species

Macroecological niche overlap analyses were performed with the R package Ecospat following the work of Broennimann et al. (2012) and Di Cola et al. (2017). Schöener D overlap values indicate the degree of superposition of the 2 paired units compared. In addition to the niche D overlap, which is the same to both paired species, the dynamic niche evolution is presented as an indication of the niche differentiation levels of each species pair. The outcomes of these analyses are presented as niche unfilling, stability, and expansion and respectively indicate unfilling: the percentage of niche of the paired species for comparison that was not occupied by the target species of study; stability: the percentage of the niche of the studied species that overlapped between the species pair; and expansion: the percentage of the niche of the studied species that is exclusive to that species relative to the other species in the pairwise comparison. Niche stability and expansion add to 1. Thus, stability and expansion are given as percentages.

To this end, mean sea surface temperature, mean sea surface salinity, and mean sea surface chlorophyll ocean variables were retrieved from the BioOracle 2.0 repository with a spatial resolution of 10 × 10 km. Spatial occurrences for all 4 banded penguin species were retrieved from their records in the sea reported in GBIF. Spatial records were downloaded and filtered with the R package spocc (<https://github.com/ropensci/spocc>) to match the resolution of the ocean variables considered. Isolated records from distant locations (>5,000 km) to known colonies were discarded since they are considered either errors or vagrant (nonreproductive) individuals. In addition, for 3 of the 4 species, an additional contrast of niche overlap was calculated for the regional subpopulations that showed a high degree of isolation. In the case of the Galápagos penguin, this was done between records of Isabella and Santiago islands. In the case of the Humboldt penguin, breeding populations of this species were broken into north and south subpopulations using as a natural break

reflecting a discontinuity in the species distribution around the Atacama corridor. Lastly, Magellanic penguin subpopulations were subdivided between the Atlantic and the Pacific coasts using the Magellan Strait as a discontinuity.

The multidimensional niche comprising all 3 variables integrated was calculated and mapped in a sPCA for all 6 paired combinations of the 4 banded penguin species as well within the 3 regional intraspecific populations of Magellanic, Galápagos, and Humboldt penguins (supplementary fig. S18, Supplementary Material online). The Schöener D overlap index was estimated (supplementary tables S21 to S23, Supplementary Material online). We then performed a sPCA that depicts the full niche space distribution clouds and the contour of the 95% centroid (discontinued lines, supplementary fig. S18, Supplementary Material online). In addition, the extent of individual variable overlap was estimated and graphically mapped (supplementary figs. S18 to S21, Supplementary Material online). The shared niche space (typically named stability) together with the unique niche of the 2 paired species (typically referred to as niche unfilling and niche expansion) for each individual variable was summarized in supplementary tables S25 to S30, Supplementary Material online. The figures show stability in gray and the unique niche space of each species in its respective guiding color. In the case of intraspecific regional subpopulation comparisons, these are represented with contrasting shades of the same color (dark and light, supplementary figs. S22 to S24, Supplementary Material online).

Genomic Environmental Association

We utilized RDA, a canonical ordination method developed by van den Wollenberg (1977) and Legendre and Legendre (2012), to investigate the variance in response variables. We used only autosomal scaffolds. For interspecific analysis, we use data set 1B, and for intraspecific analysis, we use data set 2A. The resulting vcf file for each set of SNPs files was converted into a lfm file for input into the population RDA approach. RDA was conducted using a systematic workflow in R, employing various packages such as vegan (Dixon 2003), LEA (Frichot and François 2015), permute (<https://github.com/gavinsimpson/permute>), and corplot (Wei et al. 2017).

The climatic characteristics of each of each site were compared across 19 bioclimatic variables extracted with R software packages: terra (Hijmans et al. 2022), raster (<https://github.com/rsatial/raster>), and rgdal (Bivand et al. 2015) at 30 arc-second resolutions from the CHELSA database (Karger et al. 2017) covering the period 1980 to 2010. The median elevation was obtained from the SRTM4.1 global topography data set (Amatulli et al. 2018). To choose the environmental variables for the analysis, the correlation between variables was examined with matrices and plots. Variables with R-square values under 0.77 were considered as variables with low correlation and were kept for further analysis (supplementary table S32, Supplementary Material online). Genetic and environmental data were inspected for data structure

and filtered for missing values. The best subsetting of variables explored in the RDA analysis (i.e. those maximizing the R-square value while being highly significant) comprised present surface chlorophyll mean (Chlo), present surface current velocity mean (Velo), annual mean temperature (BIO1), and annual precipitation (BIO12) (supplementary table S32 and figs. S25 and S26, Supplementary Material online). The RDA was executed with genetic data regressed on the selected environmental variables that were previously standardized.

We examine the relationship between geographical distances, spatial autocorrelation, and environmental variables within a designated study area. By incorporating Moran's I statistic, we identify spatial patterns to assess the environmental drivers influencing spatial distributions. Subsequently, we derived the dbMEM and integrated it as a covariate in the RDA analysis to evaluate the relationship among environmental variables while accounting for geographical distance.

We conducted RDAs among banded penguin species and within species. The results were examined using several analytical approaches including assessments of eigenvalues and adjusted R-squared values, and significance tests were performed with 999 permutations for (i) the RDA model, (ii) the terms of the model (added sequentially), and (iii) the axes of the model (supplementary table S33, Supplementary Material online). Visualization of the RDA outputs was achieved through scatter plots.

Supplementary Material

Supplementary material is available at *Molecular Biology and Evolution* online.

Acknowledgments

Thanks are due to the reviewers for their contribution to improve this manuscript. Special thanks are due to Fabián Rodríguez León for his contribution to the elaboration of the selection analysis plots; Constanza Weinberger for the MAP; and Astolfo Mata for the illustrations of the banded penguins. The Geryon cluster at the Centro de Astro-Ingeniería UC was extensively used for the calculations performed in this paper.

Author Contributions

F.L., R.C.K.B., and J.A.V. conceptualized and designed the study. F.L., E.P., D.N., R.C., O.E.C., R.C.K.B., and J.A.V. analyzed and interpreted the genomic data. L.R.P. performed the ecological niche modeling. P.P., M.P.A.E., G.L.-J., A.S., E.F., and G.D. provided the samples. All authors read and approved the final manuscript.

Funding

This work was supported by the Instituto Antártico Chileno-Inach (RG_48_20, RT-12_14); Agencia Nacional

de Investigación y Desarrollo (ANID)—Programa Iniciativa Milenio—ICN2021_044 (CGR), ICN2021_002 (BASE), and NCN2021-050 (Lili); and FONDECYT 1210568, 1150517, and 11110060. BASAL CATA PFB-06, the Anillo ACT-86, FONDEQUIP AIC-57, and QUIMAL 130008 provided funding for several improvements to the Geryon cluster.

Conflict of Interest

The authors declare that they have no competing interests.

Data Availability

Banded penguin raw fastq reads and reconstructed genomes have been deposited in the GenBank database (PRJNA939132, BioSample accession numbers SAMN33459127 to SAMN33459236). All scripts are available at GitHub (<https://github.com/lafabi>). All other data needed are provided in either the main text or in the supplementary material.

References

- Akbari A, Pipitone GB, Anvar Z, Jaafarinia M, Ferrari M, Carrera P, Totonchi M. ADCY10 frameshift variant leading to severe recessive asthenozoospermia and segregating with absorptive hypercalciuria. *Hum Reprod*. 2019;**34**(6):1155–1164. <https://doi.org/10.1093/humrep/dez048>.
- Alexander DH, Lange K. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics*. 2011;**12**(1):246. <https://doi.org/10.1186/1471-2105-12-246>.
- Aljanabi SM, Martinez I. Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. *Nucleic Acids Res*. 1997;**25**(22):4692–4693. <https://doi.org/10.1093/nar/25.22.4692>.
- Amatulli G, Domisch S, Tuanmu MN, Parmentier B, Ranipeta A, Malczyk J, Jetz W. A suite of global, cross-scale topographic variables for environmental and biodiversity modeling. *Sci Data*. 2018;**5**(1):180040. <https://doi.org/10.1038/sdata.2018.40>.
- Andrews S. 2010. FastQC: A quality control tool for high throughput sequence data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- Barrable A, Meadows ME, Hewitson BC. Environmental reconstruction and climate modelling of the Late Quaternary in the winter rainfall region of the Western Cape, South Africa. *S Afr J Sci*. 2002;**98**(11):611–616. <https://hdl.handle.net/10520/EJC97414>
- Bevan RM, Butler PJ, Woakes AJ, Boyd IL. The energetics of gentoo penguins, *Pygoscelis papua*, during the breeding season. *Funct Ecol*. 2002;**16**(2):175–190. <https://doi.org/10.1046/j.1365-2435.2002.00622.x>.
- Bivand R, Keitt T, Rowlingson B, Pebesma E, Sumner M, Hijmans R, Rouault E, Bivand MR. 2015. Package “rgdal”. Bindings for the geospatial data abstraction library, p. 172. <https://cran.r-project.org/web/packages/rgdal/index>. [accessed 2017 October 15].
- Boersma P. An ecological and behavioral study of the Galápagos penguin. *Living Bird*. 1976;**15**:43–93.
- Boersma P, Rebstock GA. Climate change increases reproductive failure in Magellanic penguins. *PLoS One*. 2014;**9**(1):e85602. <https://doi.org/10.1371/journal.pone.0085602>.
- Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;**30**(15):2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
- Bonadonna F, Gagliardo A. Not only pigeons: avian olfactory navigation studied by satellite telemetry *Evolution*. 2021;**33**(3):273–289. <https://doi.org/10.1080/03949370.2021.1871967>.

- Broennimann O, Fitzpatrick MC, Pearman PB, Petitpierre B, Pellissier L, Yoccoz NG, Thuiller W, Fortin M-J, Randin C, Zimmermann NE, et al. Measuring ecological niche overlap from occurrence and spatial environmental data. *Glob Ecol Biogeogr*. 2012;**21**(4):481–497. <https://doi.org/10.1111/j.1466-8238.2011.00698.x>.
- Cai H, Qin D, Peng S. Responses and coping methods of different testicular cell types to heat stress: overview and perspectives. *Biosci Rep*. 2021;**41**(6):BSR20210443. <https://doi.org/10.1042/BSR20210443>.
- Cai T, Shao S, Kennedy JD, Alström P, Moyle RG, Qu Y, Lei F, Fjeldså J. The role of evolutionary time, diversification rates and dispersal in determining the global diversity of a large radiation of passerine birds. *J Biogeogr*. 2020;**47**(7):1612–1625. <https://doi.org/10.1111/jbi.13823>.
- Calvi A, Wong ASW, Wright G, Wong ESM, Loo TH, Stewart CL, Burke B. SUN4 is essential for nuclear remodeling during mammalian spermiogenesis. *Dev Biol*. 2015;**407**:321–330. <https://doi.org/10.1016/j.ydbio.2015.09.010>.
- Cavallotto JL, Violante RA, Hernández-Molina FJ. Geological aspects and evolution of the Patagonian continental margin. *Biol J Linn Soc*. 2011;**103**(2):346–362. <https://doi.org/10.1111/j.1095-8312.2011.01683.x>.
- Ceballos SG, Lessa EP, Licandeo R, Fernandez DA. Genetic relationships between Atlantic and Pacific populations of the nototheniid fish *Eleginops maclovinus*: the footprints of quaternary glaciations in Patagonia. *Heredity* (Edinb). 2016;**116**(4):372–377. <https://doi.org/10.1038/hdy.2015.106>.
- Coffin HR, Watters JV, Mateo JM. Odor-based recognition of familiar and related conspecifics: a first test conducted on captive Humboldt penguins (*Spheniscus humboldti*). *PLoS One*. 2011;**6**(9):e25002. <https://doi.org/10.1371/journal.pone.0025002>.
- Cole TL, Zhou C, Fang M, Pan H, Ksepka DT, Fiddaman SR, Emmerling CA, Thomas DB, Bi X, Fang Q, et al. Genomic insights into the secondary aquatic transition of penguins. *Nat Commun*. 2022;**13**(1):3912. <https://doi.org/10.1038/s41467-022-31508-9>.
- Dankwerts DK, Humeau L, Pinet P, Mcquaid CD, Le Corre M. 2021. Extreme philopatry and genetic diversification at unprecedented scales in a seabird. *Sci Rep*. **11**(1):6834. <https://doi.org/10.1038/s41598-021-86406-9>.
- Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, et al. 2021. Twelve years of SAMtools and BCFtools. *GigaScience*. **10**(2):giab008. <https://doi.org/10.1093/gigascience/giab008>.
- Dantas G, Oliveira L, Santos A, Flores M, Melo D, Simeone A, González-Acuña D, Luna-Jorquera G, Le Bohec C, Valdés-Velasquez A, et al. 2019. Uncovering population structure in the Humboldt penguin (*Spheniscus humboldti*) along the Pacific coast at South America. *PLoS One*. **14**(5):e0215293. <https://doi.org/10.1371/journal.pone.0215293>.
- Dean MD, Good JM, Nachman MW. Adaptive evolution of proteins secreted during sperm maturation: an analysis of the mouse epididymal transcriptome. *Mol Biol Evol*. 2008;**25**(2):383–392. <https://doi.org/10.1093/molbev/msm265>.
- Dean R, Harrison PW, Wright AE, Zimmer F, Mank JE. Positive selection underlies faster-Z evolution of gene expression in birds. *Mol Biol Evol*. 2015;**32**(10):2646–2656. <https://doi.org/10.1093/molbev/msv138>.
- de Jong MJ, de Jong JF, Hoelzel AR, Janke A. Sambar: an R package for fast, easy and reproducible population-genetic analyses of biallelic SNP data sets. *Mol Ecol Resour*. 2021;**21**(4):1369–1379. <https://doi.org/10.1111/1755-0998.13339>.
- Di Cola V, Broennimann O, Petitpierre B, Breiner FT, D’Amen M, Randin C, Engler R, Pottier J, Pio D, Dubuis A, et al. Ecospat: an R package to support spatial analyses and modeling of species niches and distributions. *Ecography*. 2017;**40**(6):774–787. <https://doi.org/10.1111/ecog.02671>.
- Dixon P. VEGAN, a package of R functions for community ecology. *J Veg Sci*. 2003;**14**(6):927–930. <https://doi.org/10.1111/j.1654-1103.2003.tb02228.x>.
- Dufresnes C, Crochet P-A. Sex chromosomes as supergenes of speciation: why amphibians defy the rules? *Philos Trans R Soc B Biol Sci*. 2022;**377**(1856):20210202. <https://doi.org/10.1098/rstb.2021.0202>.
- Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol*. 2019;**20**(1):238. <https://doi.org/10.1186/s13059-019-1832-y>.
- Fahlman A, Schmidt A, Handrich Y, Woakes AJ, Butler PJ. Metabolism and thermoregulation during fasting in king penguins, *Aptenodytes patagonicus*, in air and water. *Am J Physiol Regul Integr Comp Physiol*. 2005;**289**(3):R670–R679. <https://doi.org/10.1152/ajpregu.00130.2005>.
- Faircloth BC. PHYLUCE is a software package for the analysis of conserved genomic loci. *Bioinformatics*. 2016;**32**(5):786–788. <https://doi.org/10.1093/bioinformatics/btv646>.
- Feder JL, Gejji R, Yeaman S, Nosil P. Establishment of new mutations under divergence and genome hitchhiking. *Philos Trans R Soc B Biol Sci*. 2012;**367**(1587):461–474. <https://doi.org/10.1098/rstb.2011.0256>.
- Feder JL, Nosil P. The efficacy of divergence hitchhiking in generating genomic island during ecological speciation. *Evolution*. 2010;**64**(6):1729–1747. <https://doi.org/10.1111/j.1558-5646.2009.00943.x>.
- Fitak RR. Optm: estimating the optimal number of migration edges on population trees using Treemix. *Biol Methods Protoc*. 2021;**6**(1):bpab017. <https://doi.org/10.1093/biomethods/bpab017>.
- Frichot E, François O. LEA: an R package for landscape and ecological association studies. *Methods Ecol Evol*. 2015;**6**(8):925–929. <https://doi.org/10.1111/2041-210X.12382>.
- Friocourt F, Lafont AG, Kress C, Pain B, Manceau M, Dufour S, Chédotal A. Recurrent DCC gene losses during bird evolution. *Sci Rep*. 2017;**7**(1):37569. <https://doi.org/10.1038/srep37569>.
- Frohner C, Schweizer S, Hoyer-Fender S. SPAG4L/SPAG4L-2 are testis-specific SUN domain proteins restricted to the apical nuclear envelope of round spermatids facing the acrosome. *Mol Hum Reprod*. 2011;**17**(4):207–218. <https://doi.org/10.1093/molehr/gaq099>.
- Gao Y, Wang C, Wang K, He C, Hu K, Liang M. The effects and molecular mechanism of heat stress on spermatogenesis and the mitigation measures. *Syst Biol Reprod Med*. 2022;**68**(5-6):331–347. <https://doi.org/10.1080/19396368.2022.2074325>.
- García-Borboroglu P, Boersma D. Penguins: natural history and conservation. Seattle, WA: University of Washington Press; 2013. 360 pp.
- González-Wevar CA, Rosenfeld S, Segovia NI, Hüne M, Gérard K, Ojeda J, Mansilla A, Brickle P, Díaz A, Poulin E. Genetics, gene flow, and glaciation: the case of the South American limpet *Nacella mytilina*. *PLoS One*. 2016;**11**(9):e0161963. <https://doi.org/10.1371/journal.pone.0161963>.
- Hibbets EM, Schumacher KI, Scheppler HB, Boersma PD, Bouzat JL. Genetic evidence of hybridization between Magellanic (*Spheniscus magellanicus*) and Humboldt (*Spheniscus humboldti*) penguins in the wild. *Genetica*. 2020;**148**(5-6):215–228. <https://doi.org/10.1007/s10709-020-00106-2>.
- Hijmans RJ, Bivand R, Forner K, Ooms J, Pebesma E, Sumner MD. Package “terra”. Vienna, Austria: Maintainer; 2022.
- Hoffman JI, Kowalski GJ, Klimova A, Eberhart-Phillips LJ, Staniland IJ, Baylis AM. Population structure and historical demography of South American sea lions provide insights into the catastrophic decline of a marine mammal population. *R Soc Open Sci*. 2016;**3**(7):160291. <https://doi.org/10.1098/rsos.160291>.
- Höglund J, Mitkus M, Olsson P, Lind O, Drews A, Bloch NI, Kelber A, Strandh M. Owls lack UV-sensitive cone opsin and red oil droplets, but see UV light at night: retinal transcriptomes and ocular media transmittance. *Vision Res*. 2019;**158**:109–119. <https://doi.org/10.1016/j.visres.2019.02.005>.
- Holt KA, Boersma PD. Unprecedented heat mortality of Magellanic penguins. *Condor*. 2022;**124**(1):duab052. <https://doi.org/10.1093/ornithapp/duab052>.

- Irwin DE. Sex chromosomes and speciation in birds and other ZW systems. *Mol Ecol*. 2018;**27**(19):3831–3851. <https://doi.org/10.1111/mec.14537>.
- Karger DN, Conrad O, Böhner J, Kawohl T, Kreft H, Soria-Auza RW, Zimmermann NE, Linder HP, Kessler M. Climatologies at high resolution for the earth's land surface areas. *Sci Data*. 2017;**4**(1):1–20. <https://doi.org/10.1038/sdata.2017.122>.
- Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013;**30**(4):772–780. <https://doi.org/10.1093/molbev/mst010>.
- Kemper J, Underhill LG, Roux J-P, Bartlett PA, Chesselet YJ, James JAC, Jones R, Uhongora N-N, Wepener S. Breeding patterns and factors influencing breeding success of African penguins *Spheniscus demersus* in Namibia. In: Kirkman SP, editor. Final report of the BCLME (Benguela Current Large Marine Ecosystem) project on top predators as biological indicators of ecosystem change in the BCLME. Cape Town: Avian Demography Unit; 2007. pp. 89–99.
- Kenigsberg S, Lima PD, Maghen L, Wyse BA, Lackan C, Cheung AN, Tsang BK, Librach CL. The elusive MAESTRO gene: its human reproductive tissue-specific expression pattern. *PLoS One*. 2017;**12**(4):e0174873. <https://doi.org/10.1371/journal.pone.0174873>.
- Kersten O, Star B, Leigh DM, Anker-Nilssen T, Strøm H, Danielsen J, Descamps S, Erikstad KE, Fitzsimmons MG, Fort J, et al. Complex population structure of the Atlantic puffin revealed by whole genome analyses. *Commun Biol*. 2021;**4**(1):922. <https://doi.org/10.1038/s42003-021-02415-4>.
- Kim JH, Schneider RR, Hebbeln D, Müller PJ, Wefer G. Last deglacial sea-surface temperature evolution in the Southeast Pacific compared to climate changes on the South American continent. *Quat Sci Rev*. 2002;**21**(18-19):2085–2097. [https://doi.org/10.1016/S0277-3791\(02\)00012-4](https://doi.org/10.1016/S0277-3791(02)00012-4).
- Kolberg L, Raudvere U, Kuzmin I, Vilo J, Peterson H. Gprofiler2—an R package for gene list functional enrichment analysis and namespace conversion toolset g: profiler. *F1000Res*. 2020;**9**:ELIXIR-709. <https://doi.org/10.12688/f1000research.24956.2>.
- Kopania EE, Larson EL, Callahan C, Keeble S, Good JM. Molecular evolution across mouse spermatogenesis. *Mol Biol Evol*. 2022;**39**(2):msac023. <https://doi.org/10.1093/molbev/msac023>.
- Korneliusen TS, Albrechtsen A, Nielsen R. ANGSD: analysis of next generation sequencing data. *BMC Bioinformatics*. 2014;**15**(1):356. <https://doi.org/10.1186/s12859-014-0356-4>.
- Legendre P, Legendre L. Canonical analysis. In: *Developments in environmental modelling*. 24. Québec (Canada): Elsevier; 2012. p. 625–710.
- Li H. Seqtk toolkit for processing sequences in FASTA/Q formats. *GitHub*. 2012;**767**:69.
- Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*. 2009;**25**(14):1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>.
- Li H, Durbin R. Inference of human population history from whole genome sequence of a single individual. *Nature*. 2011;**475**(7357):493. <https://doi.org/10.1038/nature10231>.
- Liu X, Fu Y-X. Stairway plot 2: demographic history inference with folded SNP frequency spectra. *Genome Biol*. 2020;**21**(1):280. <https://doi.org/10.1186/s13059-020-02196-9>.
- Liu Z, Herbert TD. High-latitude influence on the eastern equatorial Pacific climate in the early Pleistocene epoch. *Nature*. 2004;**427**(6976):720–723. <https://doi.org/10.1038/nature02338>.
- Luu K, Bazin E, Blum MGB. Pcadapt: an R package to perform genome scans for selection based on principal component analysis. *Mol Ecol Resour*. 2017;**17**(1):67–77. <https://doi.org/10.1111/1755-0998.12592>.
- Lyle MW, Prahl FG, Sparrow MA. Upwelling and productivity changes inferred from a temperature record in the central equatorial Pacific. *Nature*. 1992;**355**(6363):812–815. <https://doi.org/10.1038/355812a0>.
- Malinsky M, Matschiner M, Svardal H. Dsuite - fast D-statistics and related admixture evidence from VCF files. *Mol Ecol Resour*. 2021;**21**(2):584–595. <https://doi.org/10.1111/1755-0998.13265>.
- Mendelson TC, Safran RJ. Speciation by sexual selection: 20 years of progress. *Trends Ecol Evol (Amst)*. 2021;**36**(12):1153–1163. <https://doi.org/10.1016/j.tree.2021.09.004>.
- Mendes FK, Vanderpool D, Fulton B, Hahn MW. CAFE 5 models variation in evolutionary rates among gene families. *Bioinformatics*. 2021;**36**(22-23):5516–5518. <https://doi.org/10.1093/bioinformatics/btaa1022>.
- Mihola O, Trachtulec Z, Vlcek C, Schimenti JC, Forejt J. A mouse speciation gene encodes a meiotic histone H3 methyltransferase. *Science*. 2009;**323**(5912):373–375. <https://doi.org/10.1126/science.1163601>.
- Mikkelsen EK, Weir JT. Phylogenomics reveals that mitochondrial capture and nuclear introgression characterize skua species proposed to be of hybrid origin. *Syst Biol*. 2023;**72**(1):78–91. <https://doi.org/10.1093/sysbio/syac078>.
- Minh BQ, Nguyen MAT, von Haeseler A. Ultrafast approximation for phylogenetic bootstrap. *Mol Biol Evol*. 2013;**30**(5):1188–1195. <https://doi.org/10.1093/molbev/mst024>.
- Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol*. 2020;**37**(5):1530–1534. <https://doi.org/10.1093/molbev/msaa015>.
- Montecino V, Lange CB. The Humboldt current system: ecosystem components and processes, fisheries, and sediment studies. *Prog Oceanogr*. 2009;**83**(1-4):65–79. <https://doi.org/10.1016/j.pocean.2009.07.041>.
- Mussmann SM, Douglas MR, Chafin TK, Douglas ME. BA3-SNPs: contemporary migration reconfigured in BayesAss for next-generation sequence data. *Methods Ecol Evol*. 2019;**10**(10):1808–1813. <https://doi.org/10.1111/2041-210X.13252>.
- Nei M. 1975. Molecular population genetics and evolution. <https://www.cabdirect.org/cabdirect/abstract/19750118335>
- Nursyifa C, Brüniche-Olsen A, Garcia-Erill G, Heller R, Albrechtsen A. Joint identification of sex and sex-linked scaffolds in non-model organisms using low depth sequencing data. *Mol Ecol Resour*. 2022;**22**(2):458–467. <https://doi.org/10.1111/1755-0998.13491>.
- Obiol JF, Herranz J, Paris JR, Whiting JR, Rozas J, Riutort M, González-Solís J. Species delimitation using genomic data to resolve taxonomic uncertainties in a speciation continuum of pelagic seabirds. *Mol Phylogenet Evol*. 2023;**179**:107671. <https://doi.org/10.1016/j.ympev.2022.107671>.
- Orr HA, Masly JP, Presgraves DC. Speciation genes. *Curr Opin Genet Dev*. 2004;**14**(6):675–679. <https://doi.org/10.1016/j.gde.2004.08.009>.
- Oswald SA, Arnold JM. Direct impacts of climatic warming on heat stress in endothermic species: seabirds as bioindicators of changing thermoregulatory constraints. *Integr Zool*. 2012;**7**(2):121–136. <https://doi.org/10.1111/j.1749-4877.2012.00287.x>.
- Pan H, Cole TL, Bi X, Fang M, Zhou C, Yang Z, Ksepka DT, Hart T, Bouzat JL, Argilla LS, et al. High-coverage genomes to elucidate the evolution of penguins. *GigaScience*. 2019;**8**(9):giz117. <https://doi.org/10.1093/gigascience/giz117>.
- Pardo-Gandarillas MC, Ibáñez CM, Yamashiro C, Méndez MA, Poulin E. Demographic inference and genetic diversity of *Octopus mimus* (Cephalopoda: Octopodidae) throughout the Humboldt current system. *Hydrobiologia*. 2018;**808**(1):125–135. <https://doi.org/10.1007/s10750-017-3339-4>.
- Patthey C, Tong YG, Tait CM, Wilson SI. Evolution of the functionally conserved DCC gene in birds. *Sci Rep*. 2017;**7**(1):42029. <https://doi.org/10.1038/srep42029>.
- Pereira CD, Serrano JB, Martins F, da Cruz e Silva OA, Rebelo S. Nuclear envelope dynamics during mammalian spermatogenesis: new insights on male fertility. *Biol Rev*. 2019;**94**(4):1195–1219. <https://doi.org/10.1111/brv.12498>.
- Pertea G, Pertea M. GFF utilities: GffRead and GffCompare. *F1000Res*. 2020;**9**:304. <https://doi.org/10.12688/f1000research.23297.1>.

- Petkova D, Novembre J, Stephens M. Visualizing spatial population structure with estimated effective migration surfaces. *Nat Genet.* 2016;**48**(1):94–100. <https://doi.org/10.1038/ng.3464>.
- Pincemy G, Dobson FS, Jouventin P. Experiments on colour ornaments and mate choice in king penguins. *Anim Behav.* 2009;**78**(5):1247–1253. <https://doi.org/10.1016/j.anbehav.2009.07.041>.
- Pörtner HO, Farrell AP. Physiology and climate change. *Science.* 2008;**322**(5902):690–692. <https://doi.org/10.1126/science.1163156>.
- Presgraves DC. Sex chromosomes and speciation in *Drosophila*. *Trends Genet.* 2008;**24**(7):336–343. <https://doi.org/10.1016/j.tig.2008.04.007>.
- Price MN, Dehal PS, Arkin AP. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS One.* 2010;**5**(3):e9490. <https://doi.org/10.1371/journal.pone.0009490>.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007;**81**(3):559–575. <https://doi.org/10.1086/519795>.
- Quick LJ, Chase BM, Carr AS, Chevalier M, Grobler BA, Meadows ME. A 25,000 year record of climate and vegetation change from the southwestern Cape Coast, South Africa. *Quat Res.* 2022;**105**:82–99. <https://doi.org/10.1017/qua.2021.31>.
- Qvarnström A, Ålund M, McFarlane SE, Sirkiä PM. Climate adaptation and speciation: particular focus on reproductive barriers in *Ficedula* flycatchers. *Evol Appl.* 2016;**9**(1):119–134. <https://doi.org/10.1111/eva.12276>.
- Rabassa J, Coronato A, Bujalesky G, Salemme M, Roig C, Meglioli A, Heusser C, Gordillo S, Roig F, Borrromei A, et al. Quaternary of Tierra del Fuego, southernmost South America: an updated review. *Quat Int.* 2000;**68–71**:217–240. [https://doi.org/10.1016/S1040-6182\(00\)00046-X](https://doi.org/10.1016/S1040-6182(00)00046-X).
- Rawlence NJ, Salis AT, Spencer HG, Waters JM, Scarsbrook L, Mitchell KJ, Phillips RA, Calderón L, Cook TR, Bost C-A, et al. Rapid radiation of Southern Ocean shags in response to receding sea ice. *J Biogeogr.* 2022;**49**(5):942–953. <https://doi.org/10.1111/jbi.14360>.
- Richards EJ, Servedio MR, Martin CH. Searching for sympatric speciation in the genomic era. *BioEssays.* 2019;**41**(7):1900047. <https://doi.org/10.1002/bies.201900047>.
- Roberts JL. 2016. African penguin (*Spheniscus demersus*) distribution during the non-breeding season: preparation for, and recovery from, a moulting fast. [MSc thesis]. Cape Town: University of Cape Town.
- Schou MF, Bonato M, Engelbrecht A, Brand Z, Svensson EI, Melgar J, Muvhali PT, Cloete SWP, Cornwallis CK. Extreme temperatures compromise male and female fertility in a large desert bird. *Nat Commun.* 2021;**12**(1):666. <https://doi.org/10.1038/s41467-021-20937-7>.
- Sersic AN, Cosacov A, Cocucci AA, Johnson LA, Pozner R, Avila LJ, Sites JW Jr, Morando M. Emerging phylogeographical patterns of plants and terrestrial vertebrates from Patagonia. *Biol J Linnean Soc.* 2011;**103**(2):475–494. <https://doi.org/10.1111/j.1095-8312.2011.01656.x>.
- Shaun W, Pichegru L. Nest microclimate and heat stress in African penguins *Spheniscus demersus* breeding on Bird Island, South Africa. *Bird Conserv Int.* 2023;**33**:e34. <https://doi.org/10.1017/S0959270922000351>.
- Shen W, Le S, Li Y, Hu F. SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS One.* 2016;**11**(10):e0163962. <https://doi.org/10.1371/journal.pone.0163962>.
- Silva M, Chibucos M, Munro J, Daugherty S, Coelho M, Silva J. Signature of adaptive evolution in olfactory receptor genes in Cory's Shearwater supports molecular basis for smell in procellariiform seabirds. *Sci Rep.* 2020;**10**(1):543. <https://doi.org/10.1038/s41598-019-56950-6>.
- Simeone A, Hiriart-Bertrand L, Reyes-Arriagada R, Halpern M, Dubach J, Wallace R, Pütz K, Lüthi B. Heterospecific pairing and hybridization between wild Humboldt and Magellanic penguins in southern Chile. *Condor.* 2009;**111**:544–550. <https://doi.org/10.1525/cond.2009.090083>.
- Sin SYW, Cloutier A, Nevitt G, Edwards SV. Olfactory receptor sub-genome and expression in a highly olfactory procellariiform seabird. *Genetics.* 2022;**220**(2):iyab210. <https://doi.org/10.1093/genetics/iyab210>.
- Sonsthagen SA, Wilson RE, Chesser T, Pons J-M, Crochet P-A, Driskell A, Dove C. Recurrent hybridization and recent origin obscure phylogenetic relationships within the “white-headed” gull (*Larus* sp.) complex. *Mol Phylogenet Evol.* 2016;**103**:41–54. <https://doi.org/10.1016/j.ympev.2016.06.008>.
- Stahel CD, Nicol SC. Temperature regulation in the little penguin, *Eudyptula minor*, in air and water. *J Comp Physiol B.* 1982;**148**(1):93–100. <https://doi.org/10.1007/BF00688892>.
- Stonehouse, B. The general biology and thermal balances of penguins. *Adv Ecol Res* 1967;**4**:131–196. .
- Swanson WJ, Nielsen R, Yang Q. Pervasive adaptive evolution in mammalian fertilization proteins. *Mol Biol Evol.* 2003;**20**(1):18–20. <https://doi.org/10.1093/oxfordjournals.molbev.a004233>.
- Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirsch R, Pyysalo S, Doncheva NT, Legeay M, Fang T, Bork P, et al. The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* 2021;**49**(D1):D605–D612. <https://doi.org/10.1093/nar/gkaa1074>.
- Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics.* 1989;**123**(3):585–595. <https://doi.org/10.1093/genetics/123.3.585>.
- Tao Y, Masly JP, Araripe L, Ke Y, Hartl DL. A sex-ratio meiotic drive system in *Drosophila simulans*. I: an autosomal suppressor. *PLoS Biol.* 2007;**5**(11):e292. <https://doi.org/10.1371/journal.pbio.0050292>.
- The UniProt Consortium. 2015. UniProt: a hub for protein information. *Nucleic Acids Res.* **43**(D1):D204–D212. <https://doi.org/10.1093/nar/gku989>.
- Thiel M, Macaya E, Acuna E, Arntz W, Bastias H, Brokordt K, Camus P, Castilla J, Castro L, Cortes M, et al. The Humboldt current system of northern and central Chile: oceanographic processes, ecological interactions and socioeconomic feedback. *Oceanogr Mar Biol.* 2007;**45**:195–344. <https://doi.org/10.1201/9781420050943.ch6>.
- Thomas PD, Ebert D, Muruganujan A, Mushayahama T, Albou L-P, Mi H. PANTHER: making genome-scale phylogenetics accessible to all. *Protein Sci.* 2022;**31**(1):8–22. <https://doi.org/10.1002/pro.4218>.
- Thomas DB, Tennyson AJD, Scofield RP, Heath TA, Pett W, Ksepka DT. Ancient crested penguin constrains timing of recruitment into seabird hotspot penguin recruitment into seabird hotspot. *Proc R Soc B: Biol Sci.* 2020;**287**(1932):20201497. <https://doi.org/10.1098/rspb.2020.1497>.
- Van Den Wollenberg AL. Redundancy analysis an alternative for canonical correlation analysis. *Psychometrika.* 1977;**42**(2):207–219. <https://doi.org/10.1007/BF02294050>.
- Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, et al. From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. *Curr Protoc Bioinformatics.* 2013;**43**(1):11.10.1–11.10.33. <https://doi.org/10.1002/0471250953.bi1110s43>.
- Varela AI, Brokordt K, Vianna JA, Frugone MJ, Ismar-Rebitz SMH, Gaskin CP, Carlile N, O'Dwyer T, Adams J, VanderWerf EA, et al. Are threatened seabird colonies of the Pacific Ocean genetically vulnerable? The case of the red-tailed tropicbird, *Phaethon rubricauda*, as a model species. *Biodivers Conserv.* 2024;**33**(3):1165–1184. <https://doi.org/10.1007/s10531-024-02791-3>.
- Via S. Natural selection in action during speciation. *Proc Natl Acad Sci USA.* 2009;**106**(supplement_1):9939–9946. <https://doi.org/10.1073/pnas.0901397106>.
- Via S, West J. The genetic mosaic suggests a new role for hitchhiking in ecological speciation. *Mol Ecol.* 2008;**17**(19):4334–4345. <https://doi.org/10.1111/j.1365-294X.2008.03921.x>.

- Vianna JA, Fernandes FAN, Frugone MJ, Figueiró HV, Pertierra LR, Noll D, Bi K, Wang-Claypool CY, Lowther A, Parker P, et al. Genome-wide analyses reveal drivers of penguin diversification. *Proc Natl Acad Sci USA*. 2020;**117**(36):22303–22310. <https://doi.org/10.1073/pnas.2006659117>.
- Violante RA, Paterlini CM, Marcolini SI, Costa IP, Cavallotto JL, Laprida C, Dragani W, García Chapori N, Watanabe S, Totah V, et al. Chapter 6 The argentine continental shelf: morphology, sediments, processes and evolution since the last glacial maximum. *Geol Soc Lond Mem*. 2014;**41**(1):55–68. <https://doi.org/10.1144/M41.6>.
- Warren BH, Simberloff D, Ricklefs RE, Aguilée R, Condamine FL, Gravel D, Morlon H, Mouquet N, Rosindell J, Casquet J, et al. Islands as model systems in ecology and evolution: prospects fifty years after MacArthur-Wilson. *Ecol Lett*. 2015;**18**(2):200–217. <https://doi.org/10.1111/ele.12398>.
- Wei T, Simko V, Levy M, Xie Y, Jin Y, Zemla J. Package “corrplot”. *Statistician*. 2017;**56**(316):e24. <https://cran.r-project.org/web/packages/corrplot/corrplot.pdf>
- Weinberger CS, Vianna JA, Faugeron S, Marquet PA. Inferring the impact of past climate changes and hunting on the South American sea lion. *Diversity Distrib*. 2021;**27**(12):2479–2497. <https://doi.org/10.1111/ddi.13421>.
- Welch AJ, Fleischer RC, James HF, Wiley AE, Ostrom PH, Adams J, Duvall F, Holmes N, Hu D, Penniman J, et al. Population divergence and gene flow in an endangered and highly mobile seabird. *Heredity (Edinb)*. 2012;**109**(1):19–28. <https://doi.org/10.1038/hdy.2012.7>.
- Welman S, Pichegru L. Nest microclimate and heat stress in African Penguins *Spheniscus demersus* breeding on Bird Island, South Africa. *Bird Conserv Int*. 2023;**33**:1–9. <https://doi.org/10.1017/S0959270922000351>.
- Whitlock MC, Lotterhos KE. Reliable detection of loci responsible for local adaptation: inference of a null model through trimming the distribution of FST. *Am Nat*. 2015;**186**(S1):S24–S36. <https://doi.org/10.1086/682949>.
- Wilson RP, Wilson MPT. Foraging ecology of breeding *Spheniscus* Penguins. In: Davis LS, Darby JT, editors. *Penguin biology*. San Diego (CA): Academic Press; 1990. p. 181–206.
- Yang K, Adham IM, Meinhardt A, Hoyer-Fender S. Ultra-structure of the sperm head-to-tail linkage complex in the absence of the spermatid-specific LINC component SPAG4. *Histochem Cell Biol*. 2018;**150**(1):49–59. <https://doi.org/10.1007/s00418-018-1668-7>.
- Zhang J, He X, Wu H, Zhang X, Yang S, Liu C, Liu S, Hua R, Zhou S, Zhao S, et al. Loss of DRC1 function leads to multiple morphological abnormalities of the sperm flagella and male infertility in human and mouse. *Hum Mol Genet*. 2021;**30**(21):1996–2011. <https://doi.org/10.1093/hmg/ddab171>.