communications biology

A Nature Portfolio journal

Article

https://doi.org/10.1038/s42003-026-10081-7

Integrative evidence reveals adaptive divergence and speciation in gentoo penguins

Daly Noll[1,2,3,4] ✉, Jane Younger[5], Luis R. Pertierra[3,6,7], Michelle Greve[6], Eduardo J. Pizarro[3,4,8], Fabiola León[3,4,9], Debora Y. C. Brandt[10], Joshua Tyler[11], Gemma Clucas[12], Hila Levy[13], W. Brian Simison[14], Julie McInnes[5,15], Pierre Pistorius[16], Céline Le Bohec[17,18,19], Francesco Bonadonna[19], Phil N. Trathan[20], Andrés Barbosa[7,27], Andrea Raya Rey[21,22,23], Gisele P. M. Dantas[24], Rauri C. K. Bowie[25], Elie Poulin[2,3] & Juliana A. Vianna[3,4,26] ✉

Understanding lineage divergence is crucial for uncovering cryptic biodiversity. Adaptive divergence, geographic isolation and life-history traits drive speciation in heterogeneous environments. The gentoo penguin complex (*Pygoscelis* spp.), historically treated as a single species, provides an ideal system to examine divergence across its full distribution. Here, we show the existence of four divergent evolutionary lineages (northern, southern, southeastern, and eastern), supported by phylogenomic and lineage-specific selective pressures, despite ancestral gene flow. South Georgia and Macquarie individuals whose status has been debated, were included. Genomic scans reveal lineage-specific signals of positive selection in genes related to thermoregulation, oxygen transport, metabolism, and skeletal development, consistent with ecological and morphological differentiation across the Antarctic Polar Front. Future niche projections indicate severe habitat losses for three lineages, whereas the southern gentoo may expand its range. We propose a taxonomic revision recognizing four distinct gentoo penguin species, including *Pygoscelis kerguelensis* sp. nov., with important conservation implications.

Marine environments are generally perceived as interconnected ecosystems, with reduced physical barriers to species dispersal[1], facilitating gene flow between distant regions. However, despite this connectivity, physico-chemical differences in the oceans, such as temperature, salinity and nutrient availability, can create ecological barriers that promote adaptive divergence of lineages through natural selection[2]. Over time, populations diverging from a common ancestor accumulate genetic and phenotypic differences that influence their physiology and shape ecological interactions[3,4], ultimately leading to adaptive divergence and speciation[5,6].

Although seabirds are highly mobile and generally show weak population genetic structure, some species exhibit strong genetic differentiation[7,8]. Such cases are often associated with ecological specialization and environmental heterogeneity[9] and can also be influenced by life-history traits[10,11]. This pattern of population divergence is well documented[9,12,13], and several mechanisms have been proposed to explain how restrictions in gene flow can promote speciation. These include geographic isolation, philopatry, differences in ocean regimes, as well as variation in non-breeding and foraging distribution, and breeding phenology[7]. From an evolutionary perspective, such processes may drive population differentiation along a continuum of divergence, ultimately favoring the emergence of new species. In seabirds, this differentiation often unfolds under a mosaic of genetic, ecological and behavioral pressures, resulting in lineages that can be recognized as distinct evolutionary entities even in the absence of strict reproductive isolation[14]. The distinction between what a species is and how it should be delimited has been the focus of extensive debate in evolutionary biology. The general lineage concept (GLC) provides a unifying framework by defining species as separately evolving metapopulation lineages and interprets the different traditional criteria as alternative lines of evidence for lineage independence rather than fixed boundaries of speciation[15]. Under this framework, species are defined not by a single property but by their status as independently evolving lineage, reflecting both the origin and maintenance of lineage independence supported by concordant sources of evidence. In practice, the integration of genomic data with ecological, morphological, phylogenetic approaches and biogeographic information now provides the resolution needed to detect patterns of divergence and to assess the degree of lineage independence across complex evolutionary landscapes. Investigating how

A full list of affiliations appears at the end of the paper. ✉e-mail: dalynoll@gmail.com; juliana.vianna@unab.cl

lineage differentiation unfolds in these highly mobile species provides valuable insights into the role of ecological barriers in the speciation. A deeper understanding of these processes in a microevolutionary context is essential for accurately detecting cryptic species, especially among vulnerable taxa, to avoid underestimating biodiversity and overestimating their distribution ranges[16]. Moreover, recognizing how evolutionary processes shape genetic diversity and ecological variability is crucial for assessing the resilience of these lineages under future climate change scenarios, because populations with limited adaptive potential may face an increased risk of extinction[17], emphasizing the need to integrate evolutionary history into conservation planning.

Among seabirds, the gentoo penguin complex (*Pygoscelis papua*) represents a potential case of adaptive divergence, with morphological, ecological, and genetic differences observed among lineages across the Southern Ocean. Recent studies have identified four to six divergent evolutionary lineages[18–21] inhabiting environments with limited ecological niche overlap[18], suggesting that local conditions can play a crucial role in their adaptive differentiation[9,18]. Genetic divergence in gentoo penguins has been linked to geographic and environmental isolation[18], with some studies suggesting that selective abiotic[18] or biotic[22] pressures have driven segregation of mitochondrial haplotypes and rapid local adaptation[23]. However, intrinsic traits also contribute to lineage differentiation in this group. Gentoo penguins exhibit strong year-round philopatry, showing high fidelity to nesting sites and limited dispersal throughout their life cycle[24,25]. These life-history traits can severely restrict the gene flow among distant colonies, promoting divergence even in the absence of obvious physical barriers. This pattern contrasts with their sister species, Adélie (*Pygoscelis adeliae*) and chinstrap (*Pygoscelis antarcticus*) penguins, which exhibit higher connectivity among breeding colonies[19,26]. Based on these patterns of divergence, Pertierra et al.[18] proposed recognizing four ecologically separated gentoo penguin lineages as subspecies: northern gentoo (*Pygoscelis papua papua*) from South America and Falkland/Malvinas Islands, southern gentoo (*P. papua ellsworthi*) from Antarctica and Maritime Antarctica, the southeastern gentoo from Kerguelen Islands (awaiting formal description), and the eastern gentoo (*P. papua taeniata*) from Crozet and Marion Island. Additionally, a potential fifth evolutionary unit has been identified at Macquarie Island, although its taxonomic status remains unresolved[18]. Clucas et al.[19] reported a sixth lineage from South Georgia, which was later formally described as the South Georgia gentoo (*P. poncetii*)[20]. Tyler et al.[20] even recommended elevating the lineages from Falkland/Malvinas Islands, Antarctic Peninsula, South Georgia Island and Kerguelen Islands to species level based on their high genetic and morphological differentiation. However, studies that include colonies from the Indian Ocean, specifically Crozet and Marion Islands located north of the Antarctic Polar Front, strongly support the delimitation of four primary taxa[9,18,21] (Falkland/Malvinas Islands, Antarctic Peninsula, Crozet/Marion Islands, and Kerguelen Islands). In this sense, based on the high genetic and phylogenetic divergence of the four main lineages described by Pertierra et al.[18], the taxonomic status of the colonies from Macquarie and South Georgia Islands remains uncertain. In particular, the South Georgia gentoo penguin may represent an intermediate position in the continuum of speciation, as it corresponds to a population that has diverged morphologically and genomically (SNP-based divergence), but retains ancestral mitochondrial haplotypes with Antarctica[27], unlike the other lineages[21,23].

Our central hypothesis is that gentoo penguins diversified primarily through geographic isolation, reinforced by adaptive responses to distinct ecological pressures across this biogeographic barrier. In this study, we integrate genomic, ecological and morphological analyses across their near-complete geographic distribution to assess the relative roles of natural selection, genetic drift, and life-history traits in shaping lineage divergence. This integrative framework allows us to explicitly link functional genomic divergence with ecological differentiation and potential future persistence. Our approach combines genome-wide data with species distribution modeling and morphological analyses to evaluate whether the observed divergence corresponds to independent evolutionary lineages, in accordance with the General Lineage Concept. We provide a unified view of how environmental and intrinsic factors interact to drive diversification and adaptation in this group. Finally, we discuss the implications of these findings for systematic classification and offer suggestions for the conservation of these evolutionary lineages in the Southern Ocean. This study represents the most comprehensive evaluation of divergence and taxonomic status in gentoo penguins to date.

Results and Discussion

Patterns of genomic divergence and diversification history

We re-sequenced the genomes of 64 gentoo penguins from ten breeding colonies distributed across the Southern Ocean to investigate their evolutionary history and adaptive divergence (Fig. 1a, Supplementary Table 1). The average genome coverage was approximately 11.7x (Supplementary Table 2), yielding 9,158,384 SNPs (Supplementary Table 3). For population structure analysis, we removed sex-related scaffolds (Supplementary Table 4), filtered sites with a minor allele frequency (MAF) below 0.05, performed linked disequilibrium (LD) pruning, and excluded outlier loci (see Adaptive Divergence section below). For phylogenomic analyses, we

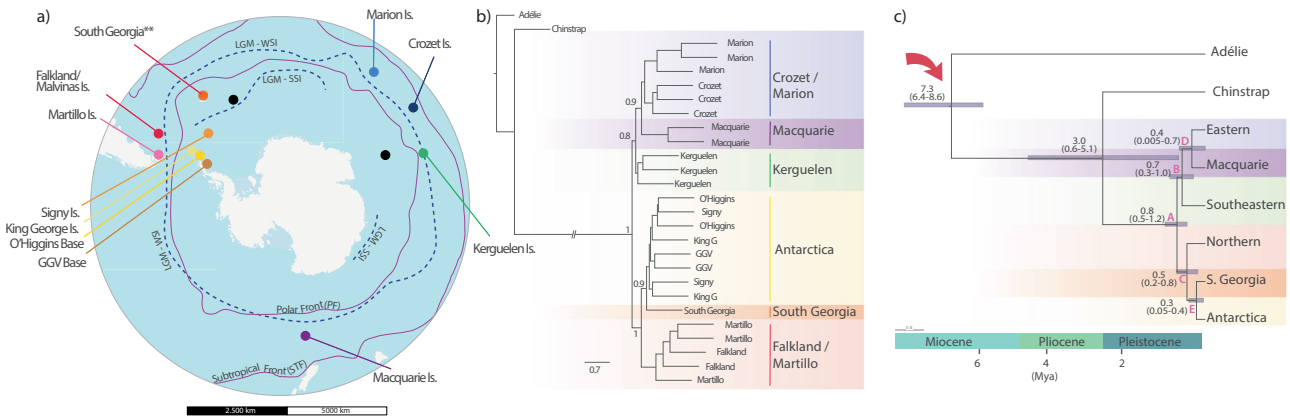

Fig. 1 | Geographic and evolutionary context of gentoo penguin divergence. **a** Geographic distribution of gentoo penguins and sampled breeding colonies (colored circles). Black dots represent unsampled colonies. The extent of sea ice during the Last Glacial Maximum (LGM) is shown for winter (LGM-WSI, solid line) and summer (LGM-SSI, dashed line), based on reconstructions by Gersonde et al.[104]. The South Georgia individual (**) was included only in phylogenetic analyses. **b** Species tree inferred from UCE data using ASTRAL-III, showing well-supported relationships among colonies. **c** Time-calibrated species tree inferred with StarBEAST3 using 100 UCE loci. Horizontal bars represent 95% highest posterior density (HPD) intervals for node ages, and the red arrow marks the fossil-calibrated node used for time estimation (*Pygoscelis grandis*; 6.3–20Mya).

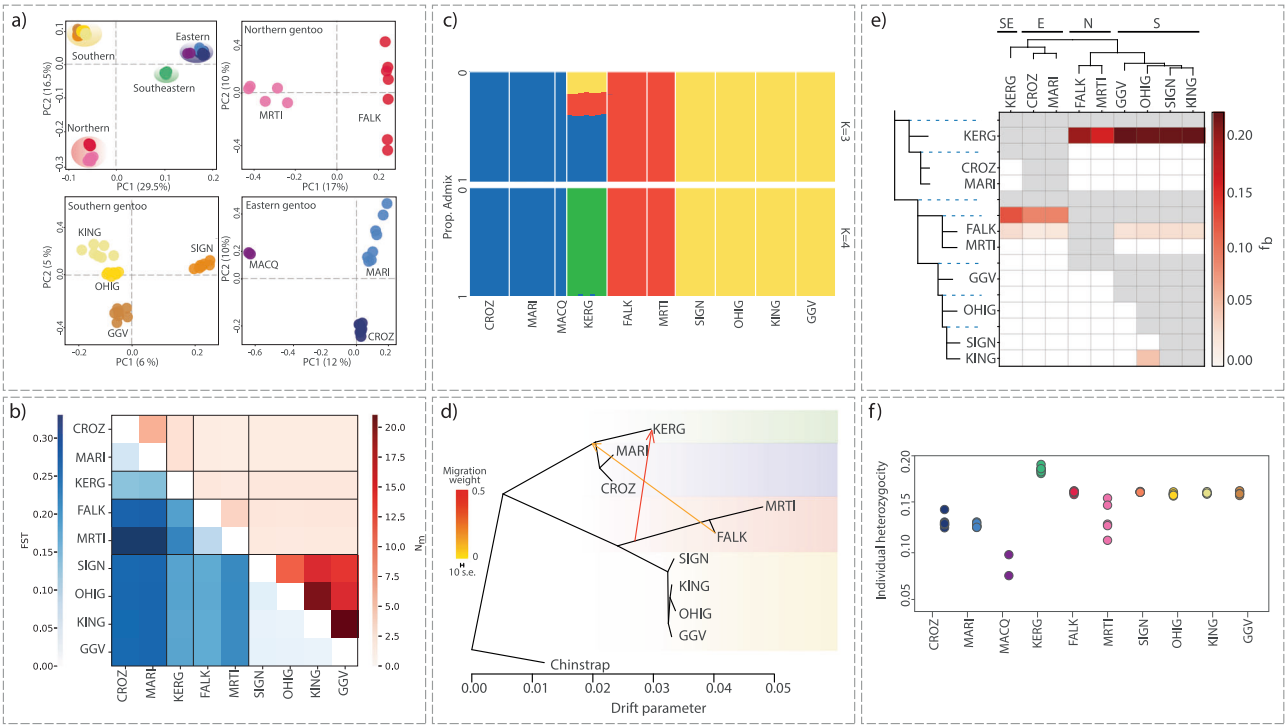

Fig. 2 | Phylogeographic structure and signatures of historical gene flow among gentoo penguin lineages. a Principal Component Analysis (PCA) showing clustering by geographic lineage and for each lineage individually: Northern, Southern, Eastern, and Southeastern, and separate within-lineage PCAs for Northern, Southern, and Eastern lineages (Southeastern not shown due to a single sampling locality). **b** Pairwise genetic differentiation (F_{ST}, below diagonal) and gene flow (Nm, above diagonal) among colonies. **c** ADMIXTURE results for $K = 3$ and $K = 4$. While $K = 4$ reflects the four main genetic lineages, the $K = 3$ model (with lower cross-validation error) indicates admixture in the Kerguelen population. **d** TreeMix graph with three inferred migration edges. Arrow colors represent migration weights (scale on left). **e** fb statistics from Dsuite reveal evidence of excess allele sharing between certain lineages. Heatmap shows levels of shared ancestry; values closer to red indicate stronger signals. **f** Individual heterozygosity estimates for each colony, with variation across lineages reflecting differences in genetic diversity.

incorporated the previously published genome assembly of a South Georgia gentoo penguin from Cole et al.[28], extracted 4,035 Ultraconserved Elements (UCEs) from 28 individuals, and inferred a Maximum Likelihood species tree using IQ-TREE – ASTRAL. Subsequently, we subsampled 100 loci using the method described by Koch[29] to ensure a robust and phylogenetically informative dataset suitable for downstream species tree reconstruction and divergence time estimation.

The phylogenomic and divergence time estimation analyses (Fig. 1b, c) provide a consistent evolutionary framework for the diversification of this group, recovering the four main lineages previously described[9,18,21]: northern gentoo (South America), southern gentoo (Antarctic Peninsula and maritime Antarctica), southeastern gentoo (from Kerguelen Islands), and the eastern gentoo (subantarctic colonies north of the Antarctic Polar Front in the Indian and Pacific sectors of the Southern Ocean). Our UCE-based phylogeny strongly supports the grouping of South Georgia and southern gentoo lineages into a monophyletic group, revealing lower levels of divergence between them compared to that observed among the four main lineages (Fig. 1b, c). Similarly, our analysis supports the clustering of the Macquarie Island individuals (included only in the phylogenetic and assignment analyses) as a sister group of the eastern gentoo penguin lineage (Fig. 1b, c). Divergence time estimation for the ancestral node of gentoo penguins' dates back to the mid-Pleistocene, approximately 0.8 million years ago (Fig. 1c).

In a phylogeographic context, principal component analysis (PCA) also assigns individuals from Macquarie Island to the eastern lineage (Fig. 2a). High F_{ST} values (0.1–0.3) and low gene flow (N_m) between geographically distant breeding colonies (Fig. 2b, Supplementary data Fig. 2b) further support the existence of genetically isolated groups. This pattern was reinforced by a hierarchical AMOVA, which shows ~19% of the variance among lineages ($\Phi_{Lineage} = 0.199$, $P < 0.001$). Comparable levels of genomic differentiation have been reported in other penguins and seabirds—for instance, Frugone et al.[13] document substantial genetic divergence among three rockhopper penguin taxa using SNP data supporting their recognition as distinct species. The higher genomic differentiation observed among the four lineages of gentoo penguins therefore indicates a more advanced stage of divergence, consistent with potential long-term reproductive and ecological isolation. Although our study lacks population-level data for Macquarie and South Georgia colonies, their well-supported phylogenetic position as sister clades to the eastern and southern lineages, respectively, indicates coherent genomic differentiation at a lower hierarchical level than that observed among the main four species-level lineages. Comparable levels of genomic differentiation (F_{ST} ~ 0.04–0.1) in *Puffinus* shearwaters have been interpreted as representing early to intermediate stages of divergence[30], consistent with differentiation at the subspecies level.

Traces of past connectivity among lineages are revealed by ADMIXTURE (Fig. 2c, Supplementary data Fig. 2c), TreeMix (Fig. 2d), and Dsuite (Fig. 2e, Supplementary data Fig. 2e) analyses, particularly between South America, Kerguelen, and the Crozet/Marion Islands. While ADMIXTURE with $K = 4$ clearly separates the four lineages, the model with $K = 3$ (lower cross-validation error, Supplementary Fig. 1) indicates genetic admixture in the Kerguelen lineage, involving alleles from other lineages. This evidence of admixture may reflect either the retention of ancestral alleles or historical gene flow. Moreover, TreeMix analyses detected migration edges from South America to the Crozet/Marion Islands and, even more strongly, to Kerguelen (Fig. 2d). Although three migration events were inferred in the best-fitting model, one of them, from the Martillo Island colony to the outgroup, is likely an artifact caused by low sample size, as the dataset included only one chinstrap penguin (outgroup sample) and five Martillo Island individuals (Supplementary Fig. 2). These results support the hypothesis that the Falkland/Malvinas and Kerguelen Islands served as glacial refugia during the Pleistocene[31,32], promoting population persistence and isolation, while islands such as Crozet, Marion, and Macquarie likely functioned as secondary dispersal areas. During the Pleistocene, the

expansion of sea ice forced endothermic species like penguins to abandon their terrestrial ice-free nesting sites on the continent[33], driving them to disperse to peripheral regions, such as the Subantarctic region[34]. Subsequent sea ice retreat likely facilitated isolation of the newly established colonies. In this sense, past connectivity between lineages could reflect sporadic dispersal events, especially considering that genomic divergence observed among lineages have been shaped by life-history traits, such as year-round fidelity to nesting sites[24,25].

Individual genome-wide heterozygosity analyses reveal the genetic consequences of glacial and postglacial periods. Gentoo penguins from Kerguelen exhibit higher levels of genetic diversity (Fig. 2f, Supplementary data Fig. 2f) compared to other colonies, a pattern characteristic of historical contact zones[35], reinforcing the potential historical role of the Kerguelen Islands as a glacial refugium. This suggests that the region facilitated the persistence of diverse ancestral populations and served as a connection between South American and Indian Ocean colonies. Conversely, lower levels of genetic diversity observed in individuals from Crozet, Marion and Macquarie Islands (Fig. 2f, Supplementary data Fig. 2f) may have resulted from bottlenecks or founder effects during glacial periods, a common scenario in postglacially colonized regions[36]. This reduction in genome-wide genetic diversity is consistent with the expected effects of genetic drift, which increases in strength as effective population size decreases[37].

Evidence of adaptive divergence

Birds exhibit a faster rate of adaptation compared to other vertebrates[38], enabling them to occupy a broad range of environments. This accelerated adaptation frequently occurs when populations diverge and follow independent evolutionary trajectories in isolation, often facilitated by standing genetic variation, thereby enabling rapid adaptive divergence and ecological specialization[39]. In this context, gentoo penguins likely spread from subantarctic glacial refugia into new habitats, expanding their range to polar waters in Antarctica and to more temperate regions north of the Antarctic Polar Front. Our analysis of relative (F_{ST}) and absolute (D_{XY}) genomic divergence across sliding windows revealed a consistent pattern of high F_{ST} and low D_{XY} (Supplementary Fig. 3). While these patterns may indicate divergent natural selection between pairs of lineages[40], we must also consider the contribution of genetic drift, potentially amplified by geographic isolation. The low genome-wide D_{XY} values suggest a relatively recent divergence, possibly reflecting insufficient time for the accumulation of new lineage-specific mutations[41]. In contrast, the elevated F_{ST} may result from allele frequencies changes driven either by drift or by selection[42].

To further explore the relative contribution of these processes, we next examined genome-wide signals of selection among lineages. Specifically, we focused on identifying genomic regions potentially shaped by natural selection. We used two complementary selective sweep methods: RAiSD[43] (which integrates signals of reduced genetic diversity, shifts in the allele frequency spectrum and linkage disequilibrium), and nS_L[44] (which detects of positive selection based on extended haplotype homozygosity patterns). After intersecting the outlier windows identified by both methods, thereby minimizing potential background noise from genetic drift, we found 105, 132, 195, and 165 genes within shared outlier windows in the northern, southern, southeastern, and eastern gentoo penguin lineages, respectively (Supplementary Fig. 4, Supplementary Table 5). Although demographic history—including bottlenecks, genetic drift and ancestral gene flow—can distort the SFS and generate long haplotypes that mimic sweep-like patterns, we performed lineage-specific analysis and combined RAiSD and nS_L to obtain a more robust set of candidate loci for positive selection.

RDA models were built separately for terrestrial and marine environments to assess association between allele frequencies and environmental gradients. We used 31,234 outlier SNPs located exclusively within genes to assess the correlation between genetic differentiation and environmental variables (Supplementary Table 6). The marine model included Maximum Sea Ice Cover, Maximum and Minimum Primary Productivity, Maximum and Minimum Salinity, and Maximum Ocean Temperature (Supplementary Table 7). The terrestrial model included Mean Diurnal Range (Bio2), Isothermality (Bio3), Temperature Annual Range (Bio7), Mean Temperature of the Warmest Quarter (Bio10), Precipitation Seasonality (Bio15), Precipitation of the Wettest Quarter (Bio16), and Precipitation of the Warmest Quarter (Bio18). Both models revealed significant association between adaptive genetic variation and environmental conditions ($p = 0.0001$; Fig. 3, Supplementary data Fig. 3). The terrestrial and marine RDA models explained 60.7% ($F = 3.03$) and 64,95% ($F = 2.72$) of the total variance, respectively (Fig. 3, Supplementary data Fig. 3, Supplementary Table 7), consistent with previous analysis based on neutral SNPs[18]and reinforcing the role of local environmental pressures in shaping adaptive divergence among lineages.

Eastern gentoo penguins were associated with terrestrial variables such as high precipitation and stable thermal conditions, while the southern lineage showed an association with ice cover and high salinity, suggesting potential adaptive responses linked to extreme polar environments. In the case of the southeastern lineage, the analysis showed a strong association with ocean temperature, consistent with seasonal variation in the Antarctic Polar Front[45], indicating potential adaptations to thermal shifts. In contrast, northern gentoo penguins were strongly associated with primary productivity. In line with the RDA results, the high primary productivity associated with the northern lineage coincides with GO term enrichment (Supplementary Figs. 5, 6, 7) in digestion-related processes (GO:0007586), including genes such as *KCNQ1*, *AKR1D1* and *STRAP*, as well as pathways involved in cardiac conduction and muscle excitation (*CACNA1C*, *SCN2A*, *KCNQ1* genes), suggesting possible metabolic and physiological optimizations to sustain high foraging activity (Fig. 4, Supplementary Tables 5, 8). In southern gentoo penguins, selection signals highlight thermoregulation-related processes (GO:0060048, *cardiac muscle contraction*; GO:0014896, *muscle hypertrophy*), with genes *RYR2*, *TTN*, *KCNQ1* and *IGFI* and the AKT-mTOR focal adhesion pathway (*VEGFD*, *IGF1*, *LAMA2*, *LAMA4*, *KITLG*, *MYB*, *PIK3R1*, *PPP2R2A*, *CREB3L1*) supporting heat generation[46] in response to cold environments. Additionally, adipogenesis-related genes (*LRP5*, *ASXL2*, *METRNL*, GO:0045600) could be associated with lipid accumulation and localization (*MYB*, *ATP8A1*, *KCNK9*, *ATP8A2*, GO:1905952) as a key physiological strategy. Notably, *METRNL* was also identified as overexpressed in emperor penguins relative to king penguins in a comparative gene expression study of cold-adapted species[47], suggesting potential convergence in metabolic pathways associated with extreme cold environment. Furthermore, genes involved in light perception (*GRM8*, *IMPG1*, *PDE6C*, *RP1*, *POU6F2*, *BBS9*, *KCNK9*, *ATP8A2*, *CCDC66*, GO:0007601, GO:0050953) suggest possible adaptations to seasonal variation in daylight and ice reflectivity (Fig. 4, Supplementary Tables 5, 8). In southeastern gentoo penguins, positive selection signals were found in genes related to neuronal plasticity (GO:0099536, GO:0060078), with genes such as *CHRNB3*, *GRIN2A*, *MEF2C*, and *CHRNA6* associated with synaptic signaling and excitatory postsynaptic potential regulation, which may suggest behavioral and physiological modifications likely related with seasonal environment variation. In addition, pathways related to muscular adaptation (GO:0051146, GO:0030048, GO:0006937), with the genes *BMP4*, *MEF2C*, *MAML1*, *CACNB2*, *PARVA*, *CALCA*, *PDE4B*, and *GSTO1* suggesting body composition adaptations likely for efficient foraging activities (Fig. 4, Supplementary Tables 5, 8). Finally, the enrichment analysis of the eastern lineage (Fig. 4, Supplementary Tables 5, 8) suggests an emphasis on energy efficiency in carbohydrate metabolism (GO:1901137, GO:0032868, GO:0045912) including genes such as *GUCY1A2*, *GUCY2C*, *ST8SIA1*, *B3GALT2*, *NME7*, *FGFR2*, *FLT4*, *KDR*, *ROR2*, *MAPKAP1*, *SYAP1*, *LEPR*, *PRKN* and *PRKG1*. Signals of efficient diving, reflected in a differentiated capability of transport and use of oxygen, are evident in genes related to angiogenesis (*KDR*, *VEGFD*, *FLT4* genes), mitochondrial fission (*KDR*, *PRKN*, *RALA* genes) and respiratory processes (lung alveolus development GO:0048286, *KDR*, *PRKN*, *RALA* genes), which could favor prolonged underwater activity in low-productivity habitats, maximizing foraging while minimizing energetic expenditure.

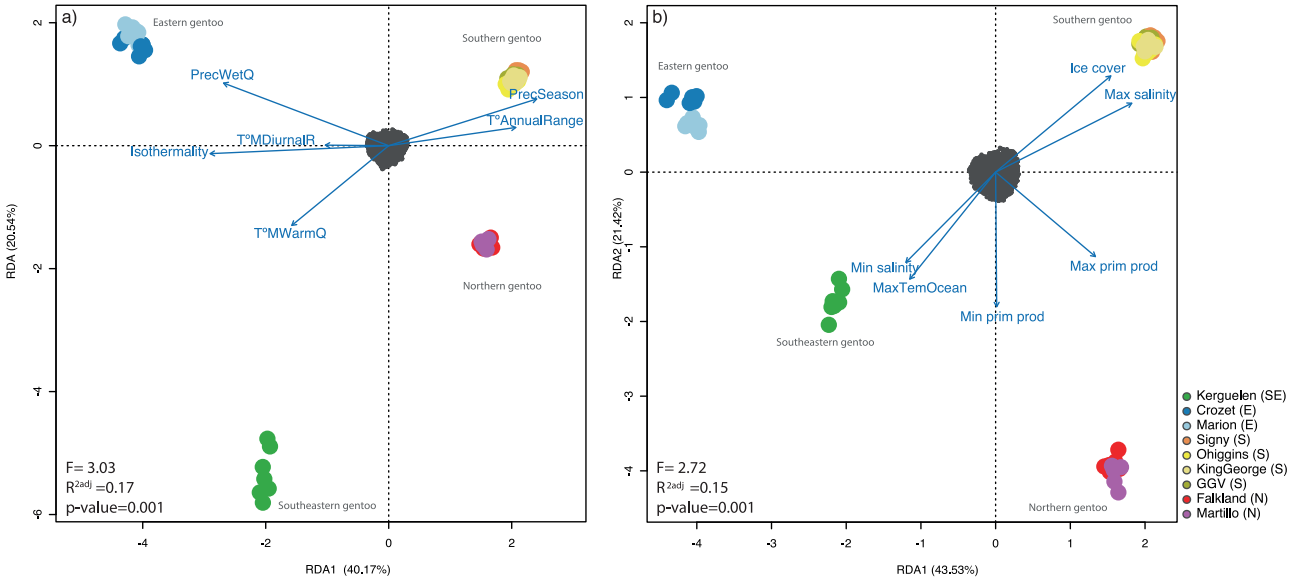

Fig. 3 | Redundancy analysis (RDA) showing associations between environmental variables and putatively adaptive genetic variation. RDA was performed using SNPs located within genes identified as outliers in selection scans. **a** Associations between terrestrial bioclimatic variables and genetic variation. Variables include: Mean Diurnal Range (T°MDiurnalR), Isothermality, Temperature Annual Range (T°AnnualRange), Mean Temperature of Warmest Quarter (T°MWarmQ), Precipitation Seasonality (PrecSeason), and Precipitation of Wettest Quarter (PrecWetQ). **b** Associations between marine environmental variables and genetic variation. Variables include: Minimum and Maximum Primary Productivity (prim prod), Minimum and Maximum Salinity, Maximum Ocean Temperature (TemOcean), and Sea Ice Cover.

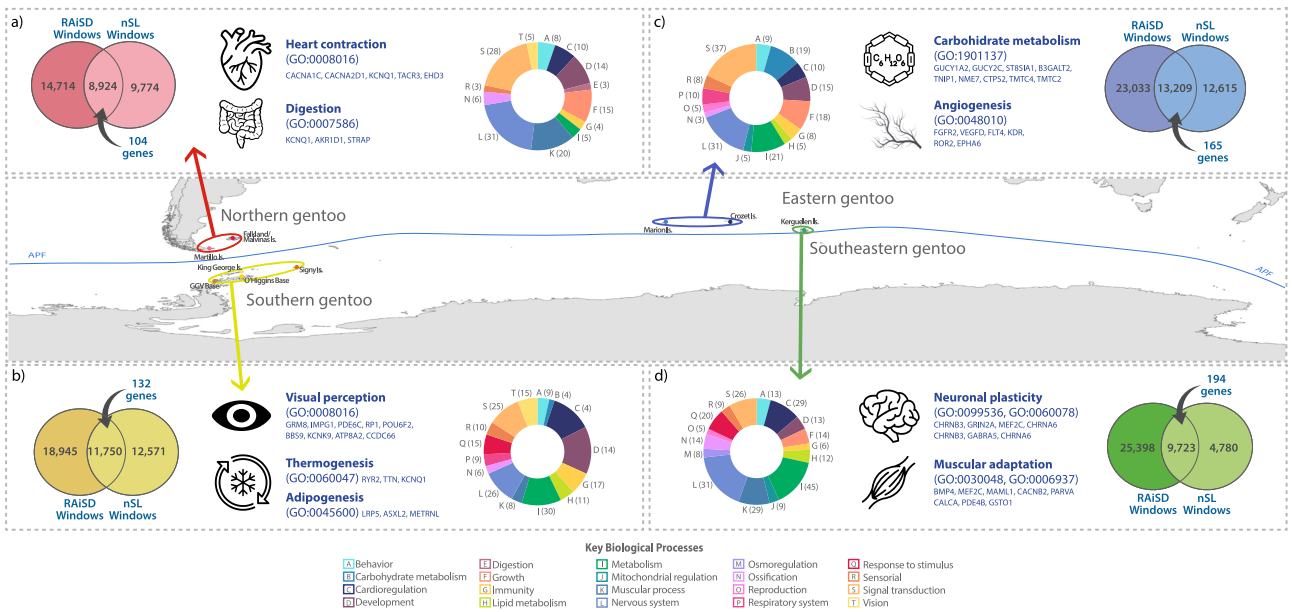

Fig. 4 | Lineage-specific functional signatures of selection in gentoo penguins. Each panel shows enriched biological processes identified in genes under positive selection for each of the four gentoo penguin lineages. Selection signals were detected using two complementary methods (RAiSD and nS_L), and the number of overlapping genomic windows and resulting candidate genes is shown in the Venn diagrams. **a** Northern gentoo penguins exhibit selection in processes related to digestion and heart contraction, likely reflecting adaptations to high foraging demands in productive waters. **b** In the southern lineage, enriched functions include thermogenesis, adipogenesis, and visual perception, consistent with adaptation to extreme cold and variable light conditions. **c** Eastern gentoo penguins show selection in genes involved in carbohydrate metabolism and angiogenesis, suggesting energy efficiency and oxygen transport optimization during extended diving. **d** The southeastern lineage presents signals related to neuronal plasticity and muscular adaptation, possibly enabling behavioral flexibility and efficient locomotion in seasonal environments. Circular diagrams summarize the number of enriched GO terms grouped into 20 major biological process categories (see legend below). The Antarctic Polar Front (APF) is shown in blue. Icons extracted from Canva Pro; map base layer generated with ArcMap (Esri).

Initial studies in gentoo penguins have revealed significant morphological differences between colonies located north and south of the Antarctic Polar Front (APF)[48,49]. In line with these findings, our study demonstrates that colonies north of the APF (northern, eastern, and southeastern gentoo penguins) exhibit signals of positive selection in genes related to growth regulation, skeletal development, and metabolic adaptation (Fig. 5a, Supplementary Table 8) including *BMP4*, *MATN1*, *MEF2C*, and *DDR2* in the southeastern lineage; *CACNA1C*, *IGF1* and *PLCB4* in the northern lineage; and, *FGFR2*, *LEPR*, and *STC1* in the eastern lineage (Fig. 5a, Supplementary Table 8). In contrast, southern gentoo penguin show selection in anabolic

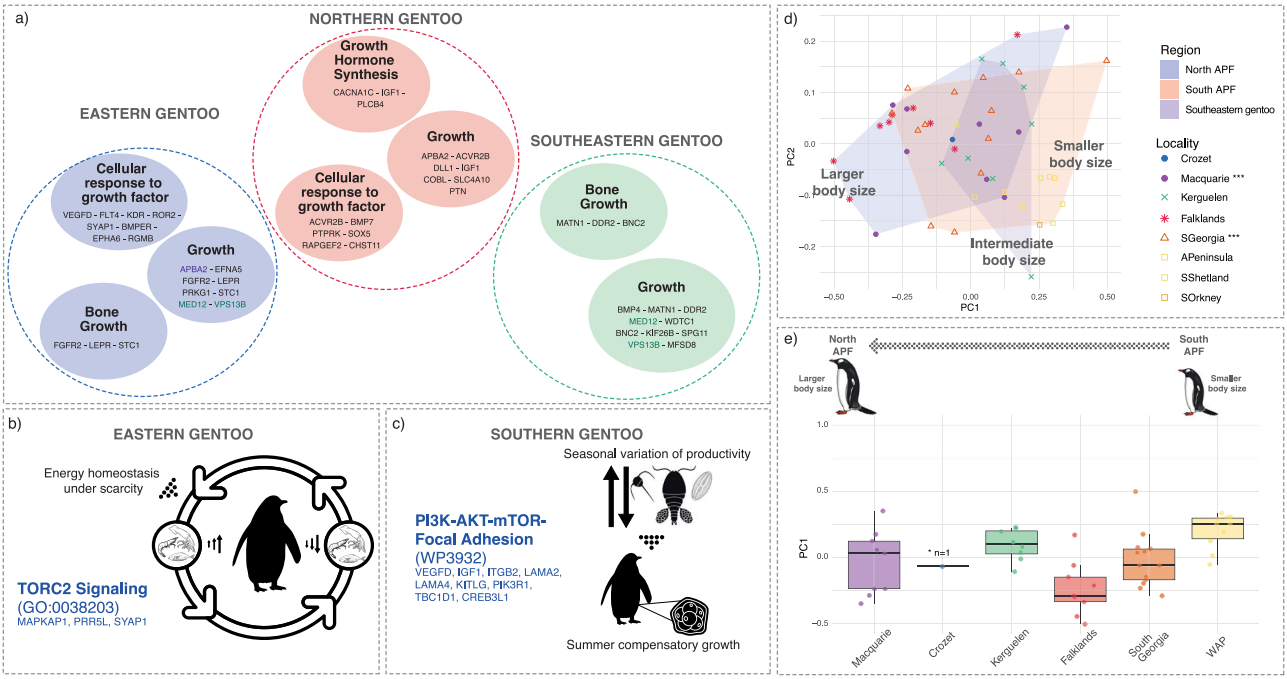

Fig. 5 | Genomic, functional and morphological differentiation among gentoo penguin lineages in relation to the Antarctic Polar Front (APF). a Genes under positive selection related to growth regulation, skeletal development, and response to growth factors identified in northern (red), southeastern (green), and eastern (blue) lineages. Genes were detected within intersected outlier windows and grouped by functional category. **b**, **c** Lineage-specific enrichment of key signaling pathways associated with energy regulation and growth dynamics. The TORC2 signaling pathway in the eastern lineage mediates energy homeostasis and cell survival under nutrient limitation, consistent with adaptations to low-productivity environments. In contrast, the PI3K-AKT-mTOR and focal adhesion signaling pathways in the southern lineage support compensatory growth and metabolic activation during short periods of summer resource abundance. **d** Principal Component Analysis (PCA) of body size measurements from museum specimens showing differentiation according to position relative to the APF: north (blue), south (red), and at intermediate position southeastern gentoo (purple). **e** Geographic variation in PC1 scores, reflecting body size gradients across sampling localities. Individuals from colder southern environments tend to be smaller (Western Antarctic Peninsula, WAP), whereas those north of the APF are generally larger, consistent with an inverse Bergmann's rule pattern. Penguin illustrations and icons in (**b**, **c**, **e**) extracted from Canva Pro.

and thermogenic pathways, consistent with their extreme polar environment (Supplementary Table 8). Functionally, this contrast between lineages is captured by two major signaling axes. The TORC2 signaling pathway in the eastern lineage (Fig. 5b), which regulates energy homeostasis, stress resistance, and cell survival under nutrient limitation, could be involved in maintaining cellular function during chronic resource scarcity. Notably, in the Crozet population, Bost & Jouventin[50] documented a winter onset of breeding that coincides with the seasonal peak in prey availability under foraging constraints. This finding suggests that gentoo penguins have historically modulated their reproductive phenology in response to environmental energy dynamics rather than temperature per se. Conversely, the PI3K–AKT–mTOR/focal adhesion axis (Fig. 5c), which controls protein synthesis, cell growth, and metabolic activation, appears to underlie compensatory growth and enhanced metabolism during periods of resource abundance in the Antarctic summer[51].

Morphometric analyses of museum specimens (Supplementary Table 9) revealed significant differences both at the lineage level and according to each geographic position relative to the Antarctic Polar Front (APF). This study expands significantly upon previous work by Tyler et al.[20], incorporating 11 additional museum specimens from locations that were not included in their study, such as Macquarie, Crozet and South Orkney Islands. We followed the same morphological measurement protocols described by Tyler et al.[20] to ensure comparability. Multivariate comparisons among all pairs of lineages showed that northern, southern (Antarctic Peninsula and maritime Antarctica), southeastern and South Georgia gentoo penguins each differ significantly in morphology (Supplementary Tables 10, 11, $p < 0.05$). By contrast, the eastern gentoo penguin (including samples from Macquarie and Crozet Islands) exhibits an intermediate size and cannot be distinguished from most other lineages based on these morphometric traits, except when compared to the southern gentoo penguin (Fig. 5d, e, Supplementary data Fig. 5d, Supplementary Table 11, $p = 0.003$). This overlap in morphospace is evident in the PCA (Supplementary Fig. 8). Similarly, we observed significant differences among colonies north and south of the APF ($p = 0.01518$, Supplementary Table 12). When colonies were grouped into three groups, north APF (Falkland/Malvinas, Crozet, Macquarie Islands), south of APF (Antarctic Peninsula, maritime Antarctica and South Georgia), and Kerguelen (treated separately due to its intermediate position at the Subantarctic region), MANOVA results ($p = 0.0066$, Fig. 5d, e, Supplementary data Fig. 5d, Supplementary Table 12) indicated that geographic position relative to the APF exerts a significant influence on morphological differentiation[48]. Notably, in the PCA (Fig. 5d, Supplementary data Fig. 5d) Kerguelen colonies showed intermediate morphological traits between the north and south of APF groups, likely reflecting the mixed oceanographic conditions surrounding the islands. The convergence of Antarctic and Subantarctic waters generates transitional thermal and productivity regimes, which may influence growth rates and body size through their effects on energy availability and foraging ecology. Within southern gentoo penguins, the morphological analysis also distinguishes South Georgia individuals from those in the Antarctic Peninsula and maritime Antarctica, underscoring the complex interplay between local environmental pressures. Morphological comparisons between Crozet/Marion and Macquarie colonies were limited by the low number of specimens measured from Crozet/Marion which constrains the strength of inference for fine-scale morphological differentiation in these regions. Although previous studies have reported biometric differences among some of these populations—for example Macquarie gentoo exhibit relatively minor biometric differences compared to those from Falkland/Malvinas, Kerguelen and South Georgia

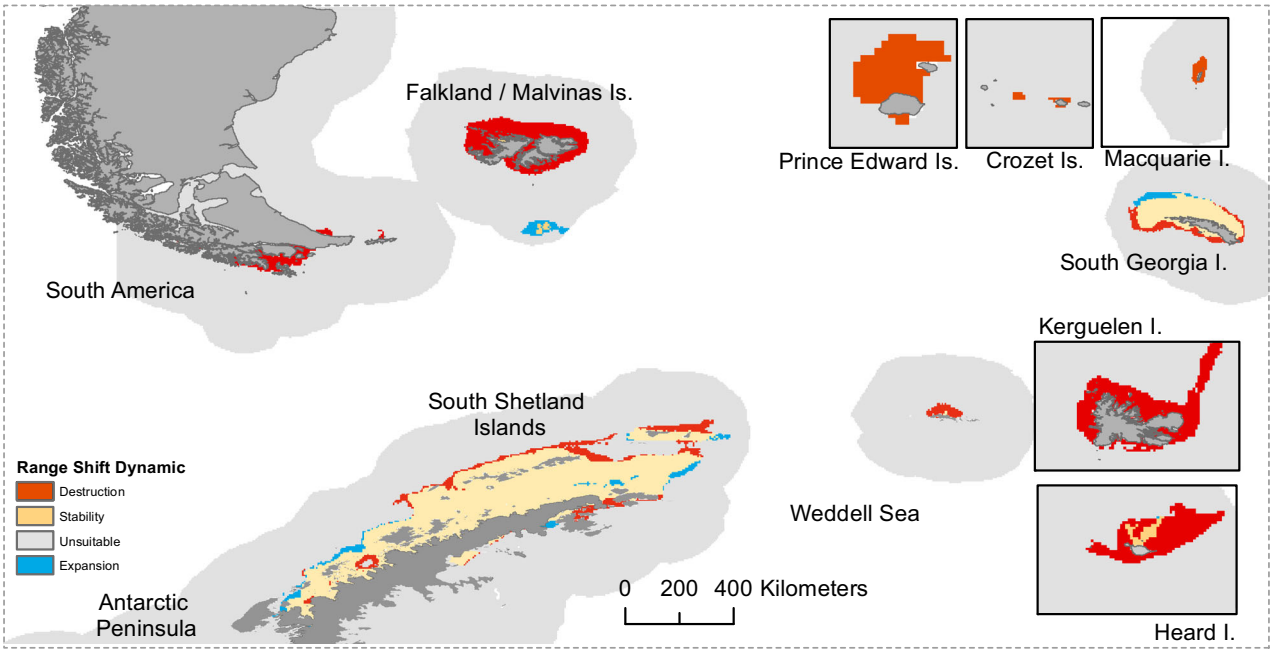

Fig. 6 | Future niche suitability changes in gentoo penguins. Range shift projection for the four main gentoo penguin lineages under the RCP4.5 climate scenario for the year 2050 based on species distribution models. Colored areas indicate predicted range dynamics relative to the current conditions.

populations, whereas Crozet colonies have the largest body size of all populations[25]—additional sampling will be necessary to fully resolve morphological variation within the eastern and southeastern lineage. According to Stonehouse[48] and Valenzuela-Guerra (REF), gentoo[52] penguins show an inverse trend to Bergmann's rule, with individuals from colder regions tending to be smaller. Although the genomic and morphological datasets came from distinct but overlapping lineages and localities, our analyses focused on general patterns rather than direct genotype-phenotype associations.

Beyond morphological and metabolic adaptations, behavioral divergence also plays a role in the adaptive divergence of gentoo penguins. Our results suggest signals of positive selection in genomic regions containing several genes associated with vocal learning, primarily through different subunits of glutamate receptors. Within outlier windows, we identified four genes (*GRM1*, *GRM8*, *GRIN2A*, *CADPS2*) in the southeastern lineage, three (*GRM1*, *GRM8*, *GRIK2*) in the southern lineage, and two (*GRM8*, *GRIK2*) in the eastern lineage (Supplementary Table 8). These genes are involved in vocal learning and adult vocal performance, playing a key role in the regulation of the song system in birds[53,54]. Recent studies have shown that vocal divergence can contribute to assortative mating and reproductive isolation in birds, particularly in taxa with strong philopatry and restricted gene flow[55,56]. In this context, our findings reinforce the hypothesis that vocal communication in gentoo penguins also may have been subject to adaptive pressures, potentially reflecting ecological and social drivers of divergence among lineages[57].

Macroenvironmental niche potential, future range shift projections, population trends and conservation considerations

Using the biomod2 R package[58,59], we explored potential changes in marine habitat suitability for gentoo penguins under current conditions and moderate future climate change scenario projected for the year 2050, based on Representative Concentration Pathway 4.5 (RCP 4.5). This pathway represents an intermediate stabilization trajectory of greenhouse gas emissions and provides a realistic framework to evaluate potential habitat changes under plausible warming conditions, avoiding the more extreme impacts predicted under high-emission scenarios. This analysis was conducted using four ecologically relevant predictors: temperature, salinity, pH and primary productivity. Our results suggest that the northern, eastern and southeastern lineages may experience a severe contraction of predicted suitable habitat by 2050, with near-complete loss of suitable environmental space across large portions of their current distribution (Fig. 6). Notably, the southeastern gentoo penguin may retain some suitable habitat around Heard Island. Conversely, the southern gentoo penguin is projected to moderately expand its suitable habitat by approximately 9%, potentially benefiting from improved climatic conditions in the Western Antarctic Peninsula, particularly in the central and southern regions, as previously reported[60–62]. Incorporating bathymetry as an additional predictor into our models suggests a more optimistic projection, with increased habitat suitability in deeper marine areas that may become favorable for foraging (Supplementary Fig. 10). However, these projections should be interpreted with caution. Our species distribution models (SDMs) rely on four environmental predictors and assume relative niche stability under future climate scenarios. As correlative models, they do not account for behavioral flexibility, microhabitat heterogeneity, prey switching, dispersal capacity or other forms of ecological plasticity. While they indicate substantial reduction in predicted habitat suitability, they should not be interpreted as definitive forecasts of extinction. Gentoo penguins may possess adaptive capacities that allow them to persist beyond their currently modeled ecological niches.

Although genotype-environment associations revealed by RDA and SDM analyses and projections were conducted independently, both tell a consistent story of adaptive divergence and future vulnerability they converge in highlighting the same set of environmental variables—particularly temperature and primary productivity—as key ecological drivers across lineages. While RDA reveals genomic signals associated with key climatic variables, niche projections allow us to visualize how those variables could affect the distribution of each lineage under future scenarios—highlighting environmental factors that have shaped ecological differences and that could influence their persistence.

From an integrative perspective, our results indicate that each gentoo penguin lineage constitutes an independently evolving unit with distinct genomic, ecological and morphological profiles. Previous studies have also reported consistent behavioral differentiation (e.g., vocal communication, philopatry, breeding phenology)[50,57,63,64] and lineage-specific demographic histories inferred with PSMC[9] for the same lineages analyzed here. These trajectories revealed asynchronous demographic

responses to late Pleistocene glacial cycles, with repeated expansions and contractions consistent with fluctuating habitat availability. The projected loss of suitable habitat for northern, eastern and southeastern lineages is particularly concerning given that the IUCN currently classifies gentoo penguins as a single, widely distributed species in the Southern Ocean, listed as 'Least Concern'. Our study provides genomic evidence supporting the recognition of four highly differentiated evolutionary lineages (northern, southern, southeastern and eastern gentoo), each with distinct population trends, ecology, behavior, genetic diversity, and habitat projections. The global "Least Concern" status is largely influenced by the southern gentoo penguin, which comprise approximately 65% of the global abundance, compared to 24% for northern, 9% for southeastern, and 2% for eastern gentoo penguin[62]. Although the total global number of gentoo penguins has increased ~11% since 2013—driven primarily by the southern gentoo colonies particularly in the Antarctic Peninsula[60,62]—reliable estimates for other regions remain scarce[62]. Many population censuses are outdated, including those from Crozet (1981), Kerguelen (1985), Heard and South Georgia (1987), Macquarie (2006) and Falkland/Malvinas (2010), underscoring the urgent need for updated surveys to accurately assess the status of each distinct evolutionary lineage.

The Western Antarctic Peninsula has undergone significant environmental changes in recent decades—warmer air temperatures, increased precipitations, and reduced sea ice extent, thickness, and seasonal duration[65–67], likely favoring the southward expansion of gentoo penguins. In contrast, the eastern gentoo penguin is declining. On Macquarie Island there has been a 50% decrease in breeding pairs over the last three generations[68], and on the Crozet archipelago, the monitored colony on Possession Island declined by 32.6% between 1992 and 2018[69]. These contrasting demographic trajectories demonstrate that extinction risk is not uniform across lineages. We therefore emphasize the need to reassess the conservation status of each lineage independently. Establishing lineage-specific monitoring programs, updating colony censuses, and considering separate IUCN assessments would better align management priorities with the true evolutionary and ecological boundaries of these taxa. Such population-level evaluations would also help identify populations at higher risk under ongoing climate-driven habitat shifts, providing a more accurate foundation for regional conservation strategies.

Systematic analysis: taxonomic resurrection and new species description

The taxonomic status of diagnostic gentoo penguin lineages has been widely debated since the 1920s, with ongoing disagreement on the number of unique lineages and their taxonomic rank. Before the genomic era, four taxonomic revisions based on morphological data were conducted[48,49,70,71], with the last pre-genomics taxonomy classification recognizing two subspecies, *Pygoscelis papua ellsworthi*, south of the Antarctic Polar Front and *P. p. papua* more broadly distributed across the subantarctic region[49]. However, our study reinforces that gentoo penguin lineages have evolved independently under distinct selective and ecological pressures, as evidenced by genomic, adaptive, morphological and ecological aspects.

This study is the first to cover nearly the entire geographic distribution of gentoo penguins, including all evolutionary units previously described in the literature. Our taxonomic assessment supports the existence of four highly differentiated lineages consistent with species-level divergence (Fig. 7). We applied an integrative framework grounded in the General Lineage Concept (GLC)[72] to assign taxonomic ranks, combining genomic, ecological and morphological information. Under the GLC, species are viewed as separately evolving segments of metapopulation lineages and the various operational criteria—genomic, ecological, morphological, or behavioral—represent alternative lines of evidence for lineage independence rather than discrete definitions. Within this framework we recognize a species when lineages show strong and consistent genomic differentiation, supported by phylogenomic and mitochondrial analyses from other studies[9,18,19,21–23,27,57], and independent ecological or morphological distinctiveness. We recognize a subspecies when a population represents an evolutionarily significant unit (ESU) below the species level, showing phylogenomic differentiation and coherent morphological or ecological divergence.

This work also provides the first genomic characterization of the Macquarie Island colonies, which reveals a clear phylogenomic affinity with the Crozet and Marion colonies despite their distant geographic location. Species trees analysis suggests that Macquarie and South Georgia colonies occupy an intermediate stage in the speciation continuum. Consistent with previous studies[19,28] and morphological differentiation highlighted by Tyler et al.[20], our results show substantial divergence between gentoo penguins from Antarctica and South Georgia, placing the latter as a sister clade to the southern gentoo (Fig. 1b, c). However, mitochondrial markers reveal

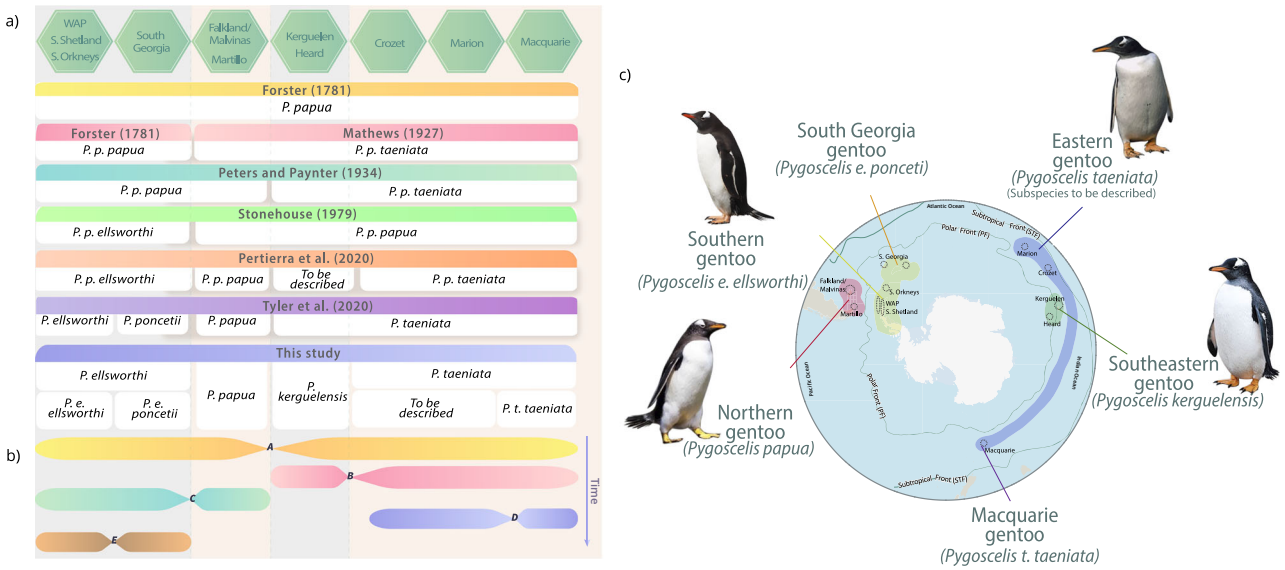

Fig. 7 | Synonymy and divergence of gentoo penguins. a Synonymy of gentoo penguins from their description by Forster (1781) as *Pygoscelis papua* to the taxonomic revision suggested by this study. **b** Graphical representation of lineage divergences based on splitting times across the calibrated molecular phylogeny (Fig. 1c). **c** Geographic distribution of the four species of gentoo penguins around the Southern Ocean. Penguin photographs by Daly Noll, Andrea Raya-Rey and Francesco Bonadonna. Map in panel c created with Quantarctica (Norwegian Polar Institute, https://www.npolar.no/quantarctica/).

haplotype overlap among colonies from South Georgia, the Antarctic Peninsula, and the South Shetland Islands[27], and the phylogenomic divergence observed (Fig. 1b, c) is significantly lower than among the four main lineages. Therefore, we consider the South Georgia population as a morphologically divergent subspecies within the southern lineage, rather than a distinct species. A similar pattern is evident for Macquarie population, which clusters closely with the eastern gentoo in PCA, admixture (Fig. 2a, b) and phylogenetic reconstructions, confirming its placement within the eastern gentoo lineage.

Based on our results, we recognized the existence of four species of gentoo penguins:

- The northern gentoo, *Pygoscelis papua*, is restricted to the Falkland/Malvinas and Martillo Islands in South America.
- The southern gentoo, *P. ellsworthi*, is the only polar species distributed across the Antarctic Peninsula, the South Orkneys, the South Shetlands, and South Georgia Islands, the latter showing signs of incipient genomic divergence and important morphological differentiation. In this context, we propose designating the South Georgia gentoo penguin as the subspecies *P. e. poncetii*, while those inhabiting the remainder of the Antarctic distribution correspond to *P. e. ellsworthi*.
- The eastern gentoo, *P. taeniata*[71], includes colonies on the Crozet, Marion, and Macquarie islands, represented by the subspecies *P. t. taeniata* from Macquarie Island and an as-yet undescribed subspecies to be described from Crozet and Marion islands.
- Finally, while molecular data are currently available only for gentoo penguins from Kerguelen Island, satellite tracking has recorded juveniles moving to Heard Island, likely foraging[73]. This suggests a close ecological connection between both regions, which can be classified as a single evolutionary unit. Therefore, we recognize these populations as part of the southeastern gentoo, *P. kerguelensis*, which we formally describe below:

***Pygoscelis kerguelensis* sp. nov.**

Common Name. Southeastern gentoo penguin.

Holotype. American Museum of Natural History (AMNH) SKIN 525824. Adult male collected by Robert Hall at Kerguelen Island on 31 December 1897. The specimen was prepared as a museum flat skin and used in the morphological analysis.

Paratypes. Specimens used in the morphological analyses. AMNH SKIN 442462, AMNH SKIN 442459: Adults collected by George Comer at Kerguelen Island in 1888 (Frances Allyn expedition). NHM 80.11.18.771, NHM 80.11.18.772, NHM 80.11.18.773, NHM 90.5.5.4: Adults collected during the Challenger Expedition (and determined by Philip Sclater and Osbert Salvin) at Christmas Harbour, Kerguelen Island in January 1874. NHM 41.7.8.2: Adult collected during the Admiralty Antarctic Expedition at Kerguelen Island.

Etymology. The species name *kerguelensis* is derived from the Kerguelen Islands, where this species originates.

Diagnosis. Morphologically, *P. kerguelensis* is characterized by the following mean values for morphological traits: Culmen length 53.96 mm, bill width at the base of 15.56 mm, bill height at gonys angle 15.05 mm, bill width at gonys angle 8.22 mm, flipper width 50.98 mm, radius length 53.39 mm, manus length 125.27 mm, tarsus length 31.87 mm and middle toe length 63.64 mm. This species is of intermediate size, being smaller than *P. papua*, larger than *P. e. ellsworthi*, and similar in size to *P. taeniata* and *P. e. poncetii*. *P. kerguelensis* can be distinguished from *P. papua*, *P. e. ellsworthi*, and *P. e. poncetii* based on morphological traits summarized in Supplementary Table 9 and Fig. 5d, e, Supplementary data Fig. 5d.

Comparisons. Principal component analysis (PCA, Fig. 5d. e, Supplementary data Fig. 5d) shows that size is the key delimiter among gentoo species, with *P. e. ellsworthi* representing the smallest gentoo followed by *P. kerguelensis, P. e. poncetii*, and significantly different of *P. papua* (Supplementary Table 13). Morphological separation of *P. kerguelensis* from all other taxa except *P. taeniata* as supported by MANOVA analysis (southeastern/southern $p = 0.02$; southeastern/northern $p = 0.04$; southeastern/South Georgia $p = 0.0009$, Supplementary Table 11). Phylogenetic reconstruction using genomic data found significant differences among all four species of gentoo penguin, with the maximum likelihood phylogeny resolving each species with high support values (Fig. 1b).

Description of holotype. *Plumage*: The holotype present black head with white band over the crown from eyebrow to eyebrow. Dark blue-gray back with a white ventral side between breasts and vent. Dark blue-gray flippers edged with white. *Bill*: Black-tipped orange bill. *Feet*: Orange/pink feet.

Measurements of the holotype. Culmen length: 56.25 mm, bill width at base: 17.71 mm, bill height at gonys angle: 15.10 mm, bill width at gonys angle: 8.74 mm, flipper width: 47.61 mm, radius length: 54.49 mm, manus length: 119.66 mm, tarsus length: 31.14 mm, middle toe length: 68.43 mm.

Description of paratypes. No discernible variation in coloration was found within *P. kerguelensis*, with all paratypes matching the description given for the holotype: black head with white band over the crown from eyebrow to eyebrow. Back dark blue gray with white on the ventral side between breasts and vent. Flippers dark blue gray edged with white. Black tipped orange bill. Orange/pink feet.

Our taxonomic analysis confirms that gentoo penguins represent four distinct species (Fig. 7), each following its own evolutionary trajectories, facing different population dynamics, and exposed to unique environmental threats. We urge that future conservation assessments, including IUCN Red List evaluations, recognize each gentoo penguin species separately. A comprehensive understanding of cryptic biodiversity will be critical in guiding effective conservation strategies.

Methods

Sampling, DNA extraction, sequencing, and read mapping

This study covers nearly the entire geographic distribution of gentoo penguins in the Southern Ocean (Supplementary Table 1; Crozet, Marion, Kerguelen, Falkland/Malvinas, Martillo, South Georgia Islands, and Antarctica; Fig. 1a). Only the South Sandwich and Heard/McDonald Islands were not included. A total of 58 DNA samples were reused from previous studies, representing individuals analyzed using mitochondrial[21,22] and ddRAD[18] data. In addition, two individuals from Macquarie Island were newly sampled to complement the geographical representation of gentoo penguin lineages. Macquarie Island samples were collected under Department of Primary Industries, Parks, Water and Environment Tasmania staff permits, and we have complied with all relevant ethical regulations for animal use. Tissue samples were obtained from naturally deceased individuals in the context of population monitoring from archived collections. Genomic DNA was originally extracted using a salt protocol[74] with modifications described by Vianna et al.[21]. Briefly, after digestion with proteinase K and lysis buffer, samples were incubated overnight. Ammonium acetate (10 M) was added instead of NaCl to precipitate proteins, and samples were centrifuged for 20 min at 14,000 rpm prior to ethanol precipitation. These samples were selected based on DNA integrity and geographic coverage.

Whole genome resequencing was then constructed from 100 ng of genomic DNA, sheared to ~350 bp to construct pair-end libraries using the Illumina TruSeq Nano kit. After six PCR cycles for enrichment, sequencing was performed to ~15x coverage using an Illumina HiSeq X platform at MedGenome Inc. To increase sample the size at locations, we incorporated raw reads from four gentoo penguins individuals sequenced by Vianna

et al.[9] which belongs to the same breeding colonies analyzed in this study (SRA/GenBank access SAMN11566624 from GGV Base in Antarctic Peninsula, SAMN11566625 from Falkland/Malvinas Islands, SAMN11566626 from Kerguelen Islands, and SAMN11566627 from Crozet Islands). Additionally, we included Adélie (*P. adeliae*; GenBank access SAMN11566622) and chinstrap (*P. antarcticus*; GenBank access SAMN11566623) penguin genomes sequenced by Vianna et al.[9], which were mapped to the reference genome to be used as outgroup for phylogenomic reconstructions. Also, we include the South Georgia gentoo penguin assembly (GenBank GCA_030674165.1) for further phylogenomic reconstructions.

Raw NGS reads were filtered using readCleaner (https://github.com/tplinderoth/ngsQC/) to remove PCR duplicates, adapters, low-quality bases, low-complexity reads, and potential human contaminants. Reads shorter than 36 bp or containing ≥50% of ambiguous bases ('N's) were discarded. High-quality reads were then mapped to gentoo penguin reference genome[75] (BGI_Ppap.V1, GenBank access GCA_010090195.1) using the BWA-MEM2 algorithm[76]. This reference genome has a total size of 1.3 Gb, with a scaffold N50 of 2.8 Mb, contig N50 of 101.3 kb, and includes 18,750 scaffolds and is classified at the scaffold level. Reads in BAM format with a mapping quality greater than 10 and base quality above 20 were retained, properly paired and sorted using Samtools v.1.3.1[77]. Duplicated reads were marked with the MarkDuplicatesSpark option in GATK v. 3.3.0[78] and local realignment around small indels was performed against the reference to reduce false-positive SNP calls using RealignerTargetCreator and IndelRealigner implemented in GATK.

Variant calling, filtering, and fasta consensus generation

Variant calling was performed on 66 individuals (64 gentoo penguins, Adélie, and chinstrap penguins as outgroups) using bcftools mpileup/call[79]. Variants were then normalized with bcftools norm, and SNPs were filtered according to quality metrics using snpCleaner.pl (https://github.com/tplinderoth/ngsQC/tree/master/snpCleaner), retaining only variants with a minimum Phred score of 30 and ≤5% of missing data. These thresholds reduce the influence of depth heterogeneity and the Wahlund effect on heterozygosity and population structure estimates[80]. From this filtered VCF, a FASTA consensus was generated for each sample using bcftools consensus. Low coverage regions identified by Bedtools genomecov were masked with maskfasta[81]. Additionally, scaffolds associated with sex chromosomes were identified based on sequencing-depth approaches described by Nursyifa, Brüniche-Olsen[82], which detect sex scaffolds through coverage depth patterns using the SATC function in R (Sex Assignment Through Coverage, Supplementary Table 4). For all analysis, SNPs located on sex-linked scaffolds were excluded.

Genomic diversity, population structure and ancestral migration

To obtain a neutral dataset for population structure, genetic diversity and ancestral migration analyses, we applied additional filtering steps. Specifically, sites with a minor allele frequency (MAF) < 0.05 were removed using VCFtools[83]and SNPs in linkage disequilibrium (LD) were pruned using PLINK v1.9[84] (--indep-pairwise 50 10 0.1) across the entire dataset to retain a common set of informative SNPs for comparative analyses across lineages. Furthermore, SNPs located within genomic regions under positive selection were excluded from this dataset (see *Detection of signals of selection within lineages* section for details on outlier identification).

We performed principal component analysis (PCA) using PLINK v1.9, plotting eigenvectors with Matplotlib in Python3. Individual ancestries were then inferred using ADMIXTURE v.1.3.0, testing values of K from 1 to 6 (Supplementary Fig. 1). The optimal number of genetic clusters was determined by the lowest cross-validation (CV) error, and admixture proportions were visualized using Pophelper (http://pophelper.com/). To quantify genetic differentiation, pairwise F_{ST} were computed among breeding colonies using VCFtools (--weir-fst-pop), and the effective number of migrants per generation (Nm) was estimated following Wright's equation:

$$Nm = \frac{1 - Fst}{4 * Fst}$$

where high F_{ST} values indicate strong genetic differentiation and low Nm values suggest limited gene flow. Finally, we conducted a hierarchical AMOVA (Lineage/Locality) in R using the '*poppr*' package, based on 999 permutations, to partition genomic variance among lineages and among localities within lineages.

To infer historical signals of gene flow among lineages, we ran TreeMix v.1.13[85], a maximum likelihood method that models population splits, and migration edges based on allele frequency covariance. First, we built a maximum likelihood tree without migration edges and then we ran models incorporating between one to five migrations events (m = 1 to 5), with 10 independent runs per migration value, totaling 50 bootstrap replicates. We used a block size of 500 SNPs (-k 500) to account for linkage disequilibrium. The optimal number of migration edges was determinate by comparing likelihood scores (Supplementary Fig. 2).

To investigate potential introgression among gentoo penguin lineages, we used Dsuite v0.5[86], a software suite designed to detect gene flow using D-statistic and F-branch statistics. We first ran *Dtrios* module to evaluate deviations from the expected allele frequency patterns under a strict bifurcating tree using the KS-test for homoplasy to assess statistical significance. We then computed F-branch statistics to estimate the proportion of shared alleles across branches, allowing the identification of specific introgression events.

Finally, we estimated genome-wide heterozygosity for each individual using VCFtools with the --het option. This measure was used to compare genetic diversity across samples and to identify potential signals of inbreeding or demographic processes. For TreeMix and Dsuite, we used the same dataset employed in the population structure analyses, with the addition of chinstrap penguin samples merged to allow for phylogenetic reconstruction admixture detection required by these analyses.

Estimation of species trees and divergence time

To reconstruct the phylogenetic relationships among gentoo penguin lineages, we analyzed 28 samples: 25 randomly selected gentoo penguins from our dataset 2–3 samples per locality), one South Georgia gentoo penguin from the previously published genome assembly by Cole et al.[28] (GenBank access GCA_030674165.1), and two additional outgroup species (one chinstrap and one Adélie penguin). We extracted 4,035 Ultraconserved elements (UCEs) from genomic sequences following the standard pipeline implemented in the PHYLUCE package[87]. Briefly, we download the 5K-UCE probe set (uce-5k-probes.fasta; available at https://github.com/faircloth-lab/uce-probe-sets). Genomic sequences of each sample were masked and indexed, and UCE probes were identified using phyluce_probe_run_multiple_lastzs_sqlite. Matching sequences were then extracted using phyluce_probe_slice_sequence_from_genomes, including flanking regions of 750 bp, which have been described as phylogenetically informative[88]. UCEs were further filtered following criteria from Vianna et al.[9]: we retained loci containing a maximum of one ambiguous nucleotide ("N") per sample. From these filtered loci, we selected the subset of UCEs that were present across all 28 penguin samples, which were individually aligned with MAFFT[89]. To obtain a robust and phylogenetically informative dataset, we subsampled 100 loci (Supplementary Table 14) from 3478 UCE using the method proposed by Koch[29], which prioritizes markers based on their evolutionary signal and topological informativeness for downstream phylogenomic inference. The output of the R script provided by Koch (https://github.com/mongiardino/genesortR) was ranked according to key properties (i.e., average bootstrap support, treeness, and Robinson-Foulds similarity) to the reference species tree. These metrics, corresponding to phylogenetic properties such as bootstrap support, treeness, and Robinson-Foulds (RF) similarity, were used to sort the dataset in spreadsheet software. Loci consistently scoring higher across these criteria were retained. The final

subset of 100 loci exhibited markedly higher values in bootstrap support (over 64.2), treeness (greater than 0.687), and RF similarity compared with the rest of the dataset (Supplementary Table 15). The selected set of UCE alignments used in phylogenomic analyses is available in the repository associated with this study (see Data Availability statement). Individual gene trees for each of the 100 subsampled UCE loci were inferred using IQ-TREE v 2.2.0.3[90], with substitution models automatically selected using ModelFinder Plus. Branch support was assessed by 1,000 ultrafast bootstrap replicates and *Pygoscelis adeliae* was specified as the outgroup for each locus. These resulting gene trees were combined as input to infer a coalescent-based species tree with ASTRAL v 5.7.8[91]. We performed divergence time estimation using StarBEAST3[92],based on the same subset of 100 UCE loci. Alignments were concatenated and imported into BEAUTi v2.7.7, using the StarBEAST template, specifying a multispecies coalescent model with an uncorrelated strict molecular clock. Substitution models we selected using bModelTest v.1.3.3. We applied lognormal prior on the root node of the *Pygoscelis* clade (offset = 6.3 Mya, mean = 7.0, SD = 1.0) to incorporate fossil calibrations based on *Pygoscelis grandis* (6.3–20Mya)[93]. MCMC chains were run for 100 million generations with parameters sampled every 5000 steps. Two independent runs were performed and combined. Convergence and effective sample sizes (ESS > 200) were assessed using Tracer v1.7.2. The maximum clade credibility (MCC) tree was summarized in TreeAnnotator v.2.7.3, discarding the initial 10% of sampled trees as burn-in.

Patterns of genome divergence

To determine whether the observed genomic patterns of divergence might reflect adaptive processes rather than purely neutral differentiation, we estimated genome-wide F_{ST} and D_{XY}. These two metrics are complementary: F_{ST} describes the relative difference in allele frequencies between populations, while D_{XY} represent the mean number of substitutions per site, providing an absolute measure of genetic distance. We first generated a new VCF file using bcftools mpileup and – bcftools call (with the -m option) to include both variant and invariant sites. This approach allowed us to retain complete site information across the genome, which is essential for accurate estimates of genetic diversity and divergence. By including invariant positions, we avoided overestimating diversity metrics and ensured that missing data were not mistakenly treated as invariant. We then used pixy[94] to calculate pairwise values of F_{ST} and D_{XY} between lineages in 200 kb genomic windows, allowing us to explore patterns of divergence and identify regions potentially associated with adaptive processes[40].

Detection of signals of selection within lineages

To evaluate the relationship between bioclimatic variables and putatively adaptive genetic differentiation, we performed a Redundancy Analysis (RDA) using the R package 'vegan'[95]. The analysis was conducted separately for marine and terrestrial variables. We used a matrix of allele frequencies derived from 31,234 SNPs located within genes, specifically within intersected outlier windows, identified in previous genome-wide selective sweep analyses (see details in the next paragraph). Climatic information was obtained from Pertierra et al.[18] who conducted their analysis using neutral loci (Supplementary Table 6). Environmental variables were standardized and filtered to remove collinear predictors ($|r| > 0.95$; Supplementary Tables 16, 17, Supplementary Fig. 9). The selected variables included Maximum Sea Ice Cover, Maximum Primary Productivity, Minimum Primary Productivity, Maximum Salinity, Minimum Salinity, Maximum Ocean Temperature for marine environment; Minimum Ocean Temperature (o14) was excluded due to high collinearity ($|r| > 0.95$). For the terrestrial environment, the selected variables were Mean Diurnal Range (T°MDiurnalR, Bio2), Isothermality (Bio3), Temperature Annual Range (T°AnnualRange, Bio7), Mean Temperature of Warmest Quarter (T°MWarmQ, Bio10), Precipitation Seasonality (PrecSeason, Bio15), Precipitation of Wettest Quarter (PrecWetQ, Bio16) and Precipitation of Warmest Quarter (Bio18). The RDA was implemented with the function rda() using scaled predictors, and the overall significance of the constrained axes was assessed through permutation tests (999 iterations) using anova.cca(). Multicollinearity was verified with variance inflation factors (vif.cca()), and eigenvalues were used to summarize the variance explained by each axis. The analysis was performed independently for marine and terrestrial environments to minimize redundancy among variables and facilitate ecological interpretation. Because the goal was to explore adaptive-environmental associations within lineages based on outlier loci, we did not include neutral structure as a covariate (i.e., no partial RDA was performed).

We implemented two complementary selective sweep methods to identify potential genomic regions under selection. Among the available tools, we selected RAiSD[43] specifically because it was designed to minimize the confounding effects of demography. RAiSD estimates the *μ statistic* by analyzing the spatial patterns of individual SNPs across the genome integrating three signatures of selection—reduction in genetic diversity, distortions in the allele frequency spectrum (SFS), and increased linkage disequilibrium (LD). By relying on SNP-level spatial patterns rather than only aggregated summary statistics, RAiSD reduces the risk of false positives associated with demographic processes such as bottlenecks or genetic drift. Additionally, we ran nS_L[44] (number of segregating sites by length) using Selscan[96], which quantifies extended haplotype homozygosity (EHH) decay to detect signals of positive selection. Because nS_L counts the number of segregating sites within extended haplotypes, it is less sensitive to variation in recombination rates and demographic effects. Subsequently, we independently analyzed the RAiSD and nS_L outputs for each gentoo penguin lineage. Outlier windows were defined using a 99th-percentile threshold (RAiSD and upper nS_L) and a 1st-percentile lower threshold for nS_L. Windows identified by both methods were intersected using the Python package PyRanges, producing shared genomic regions with robust evidence of selective sweep. Next, SNPs located within these intersected outlier windows were identified by intersection with the SNP dataset (in BED format) using PyRanges. Finally, gene-associated SNPs were extracted by intersecting these positions with annotated genes from the gentoo penguin genome (NCBI GCA_010090195.1, GTF format), using Python libraries pandas and PyRanges. The resulting dataset, containing SNPs exclusively within genes from outlier regions, was then used for downstream analyses.

To explore the functional relevance of genes located within the candidate selective sweep regions, we conducted a functional enrichment analysis using METASCAPE[97]. For each lineage, we submitted the list of genes located within intersected outlier windows. Based on the "Enrichment" summary output, we integrate overrepresented terms from multiple curated databases, including GO Biological Process, KEGG, Reactome and WikiPathways. All terms listed in the enrichment output for each lineage were compiled into a single comparative table (Supplementary Table 8), allowing qualitative examination of shared and lineage-specific functional processes potentially involved in adaptive divergence.

Morphological variation

We followed the morphometric approach described in Tyler et al.[20] to evaluate the potential taxonomic groupings. Our dataset incorporates the original individuals measured in Tyler et al.[20], complemented with additional specimens from Macquarie, Crozet and South Orkneys, providing a broader geographic representation of gentoo penguin lineages. Details of the specimens re-used from Tyler et al.[20] and newly measured for this study are provided in Supplementary Table 9. The same person (JY) measured all adult gentoo penguin skins, available at the Natural History Museum in Tring (UK) and the American Museum of Natural History (New York, USA). These collections comprised a total of 50 individuals from: South Georgia ($n = 13$); Kerguelen ($n = 8$); Macquarie Island ($n = 9$); Crozet ($n = 1$); the Falkland/Malvinas Islands ($n = 9$) and the Southern lineage (South Orkneys, South Shetlands, Western Antarctic Peninsula; $n = 10$), totaling 50 individuals (Supplementary Table 9). Nine linear measurements were taken to represent key morphological traits[98]: culmen length (CL; taken along the medial line), bill width at the base (BWB), bill height at gonys angle (BH), bill width at gonys angle (BWG), flipper width (FW; shortest distance from anterior surface of flipper above the radiale to the posterior side of the flipper), radius length (RL), manus length (ML; indent at radiale/radius/ulna

to distal flipper tip), tarsus length (TML; anterior surface) and middle toe length (MTL; digit I11 excluding nail)[20]. Measurements were taken with Mitutoyo digital calipers with an accuracy of 0.01 mm. All measurements were repeated three times, outliers were checked by confirming that all measurements were within one standard deviation, and then averaged[20]. Measures were log-transformed prior to statistical analyses. To identify traits that differed significantly between sexes, we carried out an analysis of variance (ANOVA) of sex within lineage for each trait and found that only Flipper width (FW) had a statistically significant difference between sexes ($p = 0.0333$). This trait was therefore excluded from subsequent analyses to remove any potential bias introduced by uneven sampling of sexes. Both univariate and multivariate traits analyses of traits were used to investigate morphological differentiation between lineages, and also between colonies located north and south of the Polar Front. We carried out pairwise ANOVAs to determinate whether any individual traits differed among lineages, and pairwise multivariate analysis of variance (MANOVA) on the multi-trait dataset to assess overall morphological differentiation, using the F statistic for significance testing[20]. We also executed principal component analysis (PCA) on the retained variables to visualize morphological variation. Boxplots of the first principal component (PC1) were used to illustrate morphological trends across different localities, and convex hulls in the PCA plots highlighted the extent of morphological overlap. All analyses were performed using R v.4.4 (R Core Team, 2013), primarily using the base statistics package, dplyr for data handling, and ggplot2 for visualization.

Future niche projections

Species distribution models (SDMs) for the four gentoo penguins were created using an ensemble forecasting approach[99] implemented in the biomod2 R package[58,59]. Occurrence data were obtained from GBIF, filtered by regional context and complemented with information used in Pertierra et al.[18]. Records falling on land cells were relocated to the nearest marine cell (see R script in the GitHub repository, Supplementary Table 18). Environment layers characterizing marine conditions for the present (average 2000–2020) and future conditions from the RCP4.5 emissions pathway scenario for 2050 were retrieved from the BioOracle 2.2 repository[100,101]. To select a mechanistically meaningful set of predictors and reduce the collinearity, we applied both Pearson correlation and variable inflation (VIF) tests. Ultimately, four ecologically relevant marine variables (temperature, salinity, pH and primary productivity) were selected to maximize parsimony and comparability among lineage-specific models. An additional SDM including bathymetry was also constructed, providing insights into realized distribution ranges, however, because bathymetry represent a static environmental constraint, potentially limiting future range projections, we excluded this layer from the final models to better characterize the dynamic ecological niche independent of geography. Ensemble modeling was performed using five common algorithms: Random Forest, Generalized Linear Models (GLM), Maxnet, Generalized Additive Models (GAM) and Generalized Boosting Models (GBM). For each gentoo penguin lineage, models were calibrated using three random background datasets (each of 500 records) and tested using 5-fold cross validation. Model performance was consistently high based on the True Skill Statistic[102] (Supplementary Fig. 11). The importance of the variables was provided in Supplementary Fig. 12. Ensemble predictions were derived by applying the committee average pooling method, equally weighted individual model above a threshold of confidence (TSS > 0.6). Final habitat suitability models for present and future were binarizing using the optimal threshold maximizing sensitivity plus specificity. Models were then compared using range size function, distinguish areas of predicted range destruction (present but not future suitability), range stability (present and future suitability), and range expansion (future but not present suitability). All analyses were performed in R software, and the complete script is available on the GitHub repository[103].

Statistics and reproducibility

Statistical analyses were conducted using R v4.4 and Python3. Population genomic analyses including PCA, F_{ST} estimation, heterozygosity calculations, and linkage disequilibrium pruning were performed using PLINK v1.9 and VCFtools. Hierarchical AMOVA analyses were conducted using the R package *poppr* with statistical significance assessed through 999 permutations. Redundancy analyses (RDA) were performed using the R package *vegan*, with the significance of constrained axes evaluated using permutation tests (999 iterations).

Sample sizes for each analysis are reported in the relevant sections of the manuscript and Supplementary Tables. For genomic analyses, each sequenced genome represents a biologically independent sample. Morphometric analyses were conducted on museum specimens ($n = 50$ individuals), with each specimen measured independently. Measurements were repeated three times per specimen and averaged prior to statistical analyses. Species Distribution models were evaluated using 5-fold cross-validation and ensemble modeling procedures implemented in the *biomod2* R package.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Whole-genome resequencing data generated and analyzed in this study are available in the NCBI Sequence Read Archive under BioProject PRJNA970270. Previously published sequencing data used in this study are available in the SRA under accession number SAMN11566624, SAMN11566625, SAMN11566626 and SAMN11566627. The South Georgia gentoo penguin genome assembly used for phylogenomic analyzes is available in GenBank under accession GCA_030674165.1. GenBank accession numbers and BioSample identifiers for all genome assemblies are listed in the file "genbank_biosample_access.tsv" within the same repository. Source data underlying the graphs and charts presented in the main figures are provided as Supplementary Data in the file Supplementary_source_figure_data.xlsx. Environmental layers used for species distribution models were obtained from the Bio-ORACLE v2.2 database. ZooBank registration of *P. kerguelensis*: urn:lsid:zoobank.org:pub:49BC3CED-73F5-41F7-A952-4792C79EF2A0.

Code availability

Custom scripts and analysis pipelines used in this study are publicly available at https://github.com/dalynoll/Gentoo-penguin-adaptive-divergence and have been permanently archived in Zenodo (https://doi.org/10.5281/zenodo.17399581). The repository includes scripts for variant filtering, population genomic analyses, selective sweep detection, species distribution models, and figure generation. All analyses were performed using publicly available software as described in the Methods section, including BWA-MEM2, bcftools, PLINK v1.9, TreeMix v1.13, Dsuite v0.5, IQ-TREE v2.2.0.3, ASTRAL v5.7.8, StarBEAST3, RAiSD, Selscan, and the R packages vegan, biomod2, poppr, dplyr and ggplot2.

Received: 29 August 2024; Accepted: 8 April 2026;
Published online: 23 April 2026

References

1. Norris, R. D. Pelagic species diversity, biogeography, and evolution. *Paleobiology* **26**, 236–258 (2000).
2. Nosil, P., Funk, D. J. & Ortiz-Barrientos, D. Divergent selection and heterogeneous genomic divergence. *Mol. Ecol.* **18**, 375–402 (2009).
3. Ravinet, M. et al. Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow. *J. Evol. Biol.* **30**, 1450–1477 (2017).
4. Feder, J. L., Egan, S. P. & Forbes, A. A. Ecological adaptation and speciation: the evolutionary significance of habitat avoidance as a postzygotic reproductive barrier to gene flow. *In. J. Ecol.* **2012**, 1–15 (2012).

5. Nosil, P., Egan, S. P. & Funk, D. J. Heterogeneous genomic differentiation between walking-stick ecotypes: "isolation by adaptation" and multiple roles for divergent selection. *Evolution* **62**, 316–336 (2008).
6. Savolainen, O., Lascoux, M. & Merilä, J. Ecological genomics of local adaptation. *Nat. Rev. Genet.* **14**, 807–820 (2013).
7. Friesen, V. L. Speciation in seabirds: why are there so many species…and why aren't there more? *J. Ornithol.* **156**, 27–39 (2015).
8. Munro, K. J. & Burg, T. M. A review of historical and contemporary processes affecting population genetic structure of Southern Ocean seabirds. *Emu - Austral Ornithol.* **117**, 4–18 (2017).
9. Vianna, J. A. et al. Genome-wide analyses reveal drivers of penguin diversification. *Proc. Natl. Acad. Sci. USA* **117**, 22303–22310 (2020).
10. Masaki, E. et al. Genetic characteristics of the Black-footed Albatross Diomedea nigripes</ span> on the Bonin Islands and their implications for the species' demographic history and population structure. *Ornithol. Sci.* **7**, 109–116 (2008).
11. Cagnon, C., Lauga, B., HéMery, G. & MouchèS, C. Phylogeographic differentiation of storm petrels (Hydrobates pelagicus) based on cytochrome b mitochondrial DNA variation. *Marine Biol.* **145**, 1257–1264 (2004).
12. Steeman, M. E. et al. Radiation of extant cetaceans driven by restructuring of the oceans. *Syst. Biol.* **58**, 573–585 (2009).
13. Frugone, M. J. et al. Taxonomy based on limited genomic markers may underestimate species diversity of rockhopper penguins and threaten their conservation. *Divers. Distrib.* **27**, 2277–2296 (2021).
14. Friesen, V. L., Burg, T. M. & McCoy, K. D. Mechanisms of population differentiation in seabirds. *Mol. Ecol.* **16**, 1765–1785 (2007).
15. De Queiroz, K. Species concepts and species delimitation. *Syst. Biol.* **56**, 879–886 (2007).
16. Lourenço, W. R. Scorpion incidents, misidentification cases and possible implications for the final interpretation of results. *J. Venomous Animals Toxins Trop. Dis.* **22**, 21 (2016).
17. Stanton, D. W. G. et al. More grist for the mill? Species delimitation in the genomic era and its implications for conservation. *Conserv. Genet.* **20**, 101–113 (2019).
18. Pertierra, L. R. et al. Cryptic speciation in gentoo penguins is driven by geographic isolation and regional marine conditions: unforeseen vulnerabilities to global change. *Divers. Distrib.* **26**, 958–975 (2020).
19. Clucas, G. V. et al. Comparative population genomics reveals key barriers to dispersal in Southern Ocean penguins. *Mol. Ecol.* **27**, 4680–4697 (2018).
20. Tyler, J., Bonfitto, M. T., Clucas, G. V., Reddy, S. & Younger, J. L. Morphometric and genetic evidence for four species of gentoo penguin. *Ecol. Evol.* **10**, 13836–13846 (2020).
21. Vianna, J. A. et al. Marked phylogeographic structure of Gentoo penguin reveals an ongoing diversification process along the Southern Ocean. *Mol. Phylogenet Evol.* **107**, 486–498 (2017).
22. Levy, H. et al. Evidence of pathogen-induced immunogenetic selection across the large geographic range of a wild seabird. *Mol. Biol. Evol.* **37**, 1708–1726 (2020).
23. Noll, D. et al. Positive selection over the mitochondrial genome and its role in the diversification of gentoo penguins in response to adaptation in isolation. *Sci. Rep.* **12**, 3767 (2022).
24. Tanton, J. L., Reid, K., Croxall, J. P. & Trathan, P. N. Winter distribution and behaviour of gentoo penguins Pygoscelis papua at South Georgia. *Polar Biol.* **27**, 299–303 (2004).
25. Bost C.-A. & Jouventin P. Evolutionary ecology of gentoo penguins (Pygoscelis papua). In: *Penguin Biology* (eds Davis LS, Darby JT) (Academic Press, 1990).
26. Mura-Jornet, I. et al. Chinstrap penguin population genetic structure: one or more populations along the Southern Ocean? *BMC Evol. Biol.* **18**, 90 (2018).
27. Levy, H. et al. Population structure and phylogeography of the Gentoo Penguin (Pygoscelis papua) across the Scotia Arc. *Ecol. Evol.* **6**, 1834–1853 (2016).
28. Cole, T. L. et al. Genomic insights into the secondary aquatic transition of penguins. *Nat. Commun.* **13**, 3912 (2022).
29. Koch, N. M. Phylogenomic subsampling and the search for phylogenetically reliable loci. *Mol. Biol. Evol.***38**, 4025–4038 (2021).
30. Ferrer Obiol, J. et al. Species delimitation using genomic data to resolve taxonomic uncertainties in a speciation continuum of pelagic seabirds. *Mol. Phylogenet. Evol.* **179**, 107671 (2023).
31. Hall, K. J. Review of present and quaternary periglacial processes and landforms of the maritime and sub-Antarctic region. *South Afr. J. Sci.* **98**, 71–81 (2002).
32. Lau, S. C. Y., Wilson, N. G., Silva, C. N. S. & Strugnell, J. M. Detecting glacial refugia in the Southern Ocean. *Ecography* **43**, 1639–1656 (2020).
33. Hall, K. Quaternary glaciation of the sub-Antarctic Islands. in *Developments in Quaternary Sciences* (eds Ehlers J, Gibbard PL) (Elsevier, 2004).
34. Thatje, S., Hillenbrand, C.-D., Mackensen, A. & Larter, R. Life hung by a thread: endurance of Antarctic fauna in glacial periods. *Ecology* **89**, 682–692 (2008).
35. Hewitt, G. The genetic legacy of the quaternary ice ages. *Nature* **405**, 907–913 (2000).
36. Hewitt, G. M. Some genetic consequences of ice ages, and their role in divergence and speciation. *Biol. J. Linnean Soc.* **58**, 247–276 (1996).
37. Delmore, K. E. et al. Comparative analysis examining patterns of genomic differentiation across multiple episodes of population divergence in birds. *Evol. Lett.* **2**, 76–87 (2018).
38. Zhang, G. et al. Comparative genomics reveals insights into avian genome evolution and adaptation. *Science* **346**, 1311–1320 (2014).
39. Tusso, S., Nieuwenhuis, B. P. S., Weissensteiner, B., Immler, S. & Wolf, J. B. W. Experimental evolution of adaptive divergence under varying degrees of gene flow. *Nat. Ecol. Evol.* **5**, 338–349 (2021).
40. Irwin, D. E., Alcaide, M., Delmore, K. E., Irwin, J. H. & Owens, G. L. Recurrent selection explains parallel evolution of genomic regions of high relative but low absolute differentiation in a ring species. *Mol. Ecol.* **25**, 4488–4507 (2016).
41. Sharma, M. et al. Genetic diversity and population genetic structure analysis of echinococcus granulosus sensu stricto complex based on mitochondrial DNA signature. *PLoS ONE* **8**, e82904 (2013).
42. Liu, Y., Yu, W., Wu, B. & Li, J. Patterns of genomic divergence in sympatric and allopatric speciation of three Mihoutao (Actinidia) species. *Horticult. Res.* **9**, uhac054 (2022).
43. Alachiotis, N. & Pavlidis, P. RAiSD detects positive selection based on multiple signatures of a selective sweep and SNP vectors. *Commun. Biol.* **1**, (2018).
44. Ferrer-Admetlla, A., Liang, M., Korneliussen, T. & Nielsen, R. On detecting incomplete soft or hard selective sweeps using haplotype structure. *Mol. Biol. Evol.* **31**, 1275–1291 (2014).
45. Pauthenet, E. et al. Seasonal meandering of the polar front upstream of the Kerguelen Plateau. *Geophys. Res. Lett.* **45**, 9774–9781 (2018).
46. Fennel, Z. J., Amorim, F. T., Deyhle, M. R., Hafen, P. S. & Mermier, C. M. The heat shock connection: skeletal muscle hypertrophy and atrophy. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **323**, R133–R148 (2022).
47. Paris, J. R. et al. Gene expression shifts in emperor penguin adaptation to the extreme Antarctic environment. *Mol. Ecol.* **34**, e17552 (2024).
48. Stonehouse, B. Geographic variation in Gentoo penguins Pygoscelis papua. *Ibis* **112**, 52–57 (1970).
49. Murphy, R. C. A new zonal race of the Gentoo Penguin. *Auk* **64**, 454–455 (1947).

50. Bost, C. A. & Jouventin, P. The breeding performance of the Gentoo Penguin Pygoscelis papua at the northern edge of its range. *Ibis* **133**, 14–25 (1991).
51. Heilmayer, O. et al. Temperature effects on summer growth rates in the Antarctic scallop, Adamussium colbecki. *Polar Biol.* **28**, 523–527 (2005).
52. Valenzuela-Guerra, P. et al. Geographic morphological variation of Gentoo penguin(Pygoscelis papua) and sex identification: using morphometric characters and molecular markers. *Polar Biol.* **36**, 1723–1734(2013).
53. Shi, Z. et al. miR-9 regulates basal ganglia-dependent developmental vocal learning and adult vocal performance in songbirds. *eLife* **7**, e29087 (2018).
54. Lovell, P. V., Huizinga, N. A., Friedrich, S. R., Wirthlin, M. & Mello, C. V. The constitutive differential transcriptome of a brain circuit for vocal learning. *BMC Genom.* **19**, 231 (2018).
55. Lima, R. D. & Vaz, R. V. Divergence in vocalizations indicates cryptic speciation in Camptostoma tyrannulets. *Ornithology* **142**, ukae058 (2025).
56. Brambilla, M., Janni, O., Guidali, F. & Sorace, A. Song perception among incipient species as a mechanism for reproductive isolation. *J. Evol. Biol.* **21**, 651–657 (2008).
57. de Dinechin, M. et al. The biogeography of Gentoo Penguins (Pygoscelis papua). *Can. J. Zool.* **90**, 352–360 (2012).
58. Guéguen, M., Blancheteau, H. & Thuiller, W. biomod2: ensemble platform for species distribution modeling. R package version 4.3-2-1. (2025).
59. Thuiller W., Georges D., Engler R., Breiner F. *biomod2: Ensemble Platform for Species Distribution Modeling* (2016).
60. Herman, R. W. et al. Whole genome sequencing reveals stepping-stone dispersal buffered against founder effects in a range expanding seabird. *Mol. Ecol.* **33**, e17282 (2024).
61. Herman, R. W. & Lynch H. J. Age-structured model reveals prolonged immigration is key for colony establishment in Gentoo Penguins. *Ornithol. Appl.* **124**, duac014 (2022).
62. Herman, R. et al. Update on the global abundance and distribution of breeding Gentoo Penguins (Pygoscelis papua). *Polar Biol.* **43**, 1947–1956 (2020).
63. Lynch, M. A. & Lynch, H. J. Variation in the ecstatic display call of the Gentoo Penguin (Pygoscelis papua) across regional geographic scales. *Auk* **134**, 894–902 (2017).
64. Rößler, H., Lynch, M., Torres Ortiz, S., Næsbye Larsen, O. & Beaulieu, M. Neighbors matter: vocal variation in Gentoo Penguins depends on the species composition of their colony. *Ornithology* **139**, ukac031 (2022).
65. Turner, J. et al. Antarctic climate change during the last 50 years. *Int. J. Climatol.* **25**, 279–294 (2005).
66. Turner, J. et al. Unprecedented springtime retreat of Antarctic sea ice in 2016. *Geophys. Res. Lett.* **44**, 6868–6875 (2017).
67. Meredith, M. P. & King, J. C. Rapid climate change in the ocean west of the Antarctic Peninsula during the second half of the 20th century. *Geophys. Res. Lett.* **32**, n/a-n/a (2005).
68. Pascoe, P. et al. Trends in gentoo penguin (Pygoscelis papua) breeding population size at Macquarie Island. *Polar Biol.* **43**, 877–886 (2020).
69. Barbraud, C. et al. Population trends of penguins in the French Southern Territories. *Polar Biol.* **43**, 835–850 (2020).
70. ForsterJ. R. *Indische Zoologie oder systematische Beschreibungen seltener und unbekannter Thiere aus Indien* (Johann Jacob Gebauer, 1781).
71. Mathews, G. M. Systema avium australasianarum: A systematic list of the birds of the Australasian region. British Ornithologists' Union (1927)
72. de Queiroz, K. The general lineage concept of species and the defining properties of the species category. In Species: New Interdisciplinary Essays (eds. Wilson, R. A.) 49–89 (MIT Press, 1999).
73. Thiebot, J.-B., Lescroël, A., Pinaud, D., Trathan, P. N. & Bost, C.-A. Larger foraging range but similar habitat selection in non-breeding versus breeding sub-Antarctic penguins. *Antarct. Sci.* **23**, 117–126 (2011).
74. Aljanabi, S. & Martinez, I. Universal and rapid salt-extraction of high quality genomic DNA for PCR- based techniques. *Nucleic Acids Res.* **25**, 4692–4693 (1997).
75. Pan, H. et al. High-coverage genomes to elucidate the evolution of penguins. *Gigascience* **8**, giz117 (2019).
76. Vasimuddin, M., Misra, S., Li, H. & Aluru, S. *Efficient Architecture-Aware Acceleration of BWA-MEM for Multicore Systems* (IEEE, 2019).
77. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
78. Van der Auwera, G. A. & O'Connor, B. D. Genomics in the Cloud: Using Docker, GATK, and WDL in Terra. O'Reilly Media (2020).
79. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993 (2011).
80. Kardos, M. & Waples, R. S. Low-coverage sequencing and Wahlund effect severely bias estimates of inbreeding, heterozygosity and effective population size in North American wolves. *Mol. Ecol.* **34**, e17415 (2024).
81. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
82. 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.* **22**, 458–467 (2022).
83. Danecek, P. et al. The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
84. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4**, s13742–015 (2015).
85. Pickrell, J. K. & Pritchard, J. K. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* **8**, e1002967 (2012).
86. Malinsky, M., Matschiner, M. & Svardal, H. Dsuite - Fast D-statistics and related admixture evidence from VCF files. *Mol. Ecol. Resour.* **21**, 584–595 (2021).
87. Faircloth, B. C. PHYLUCE is a software package for the analysis of conserved genomic loci. *Bioinformatics* **32**, 786–788 (2016).
88. Smith, B. T., Harvey, M. G., Faircloth, B. C., Glenn, T. C. & Brumfield, R. T. Target capture and massively parallel sequencing of ultraconserved elements for comparative studies at shallow evolutionary time scales. *Syst. Biol.* **63**, 83–95 (2014).
89. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment Software Version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
90. Minh, B. Q. et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
91. Zhang, C., Rabiee, M., Sayyari, E. & Mirarab, S. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinform.* **19**, 153 (2018).
92. Douglas, J., Jiménez-Silva, C. L. & Bouckaert, R. StarBeast3: adaptive parallelized Bayesian inference under the multispecies coalescent. *Syst. Biol.* **71**, 901–916 (2022).
93. Walsh, S. A. & Suárez, M. E. New penguin remains from the Pliocene of Northern Chile. *Hist. Biol.* **18**, 119–130 (2006).
94. Korunes, K. L. & Samuk, K. pixy : unbiased estimation of nucleotide diversity and divergence in the presence of missing data. *Mol. Ecol. Resour.* **21**, 1359–1368 (2021).

95. Oksanen, J. et al. vegan: community ecology package 2.5-6. *Computer software]* https://CRAN.R-proje.ct.org/packa.ge=vegan, (2019).
96. Szpiech, Z. A. selscan 2.0: scanning for sweeps in unphased data. *Bioinformatics*. **40**, btae006 (2024).
97. Zhou, Y. et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* **10**, 1523 (2019).
98. Baldwin, S. P., Oberholser, H. C., Worley, L. G. & Valentine, J. M. Measurements of birds. Cleveland Museum of Natural History, Cleveland (1931).
99. Thuiller, W., Lafourcade, B., Engler, R. & Araújo, M. B. BIOMOD – a platform for ensemble forecasting of species distributions. *Ecography* **32**, 369–373 (2009).
100. Assis, J. et al. Bio-ORACLE v3.0. Pushing marine data layers to the CMIP6 Earth System Models of climate change research. *Glob. Ecol. Biogeogr.* **33**, e13813 (2024).
101. Assis, J. et al. Bio-ORACLE v2.0: extending marine data layers for bioclimatic modelling. *Glob. Ecol. Biogeogr.* **27**, 277–284 (2018).
102. Allouche, O., Tsoar, A. & Kadmon, R. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *J. Appl. Ecol.* **43**, 1223–1232 (2006).
103. Noll, D. et al. Custom scripts and analysis pipelines for: Integrative evidence reveals adaptive divergence driving speciation in gentoo penguins. Zenodo https://doi.org/10.5281/zenodo.17399581 (2025).
104. Gersonde, R., Crosta, X., Abelmann, A. & Armand, L. Sea-surface temperature and sea ice distribution of the Southern Ocean at the EPILOG Last Glacial Maximum: a circum-Antarctic view based on siliceous microfossil records. *Quaternary Science Reviews* **24**, 869–896 (2005).

Acknowledgements

Financial support for this study was provided by ANID Fondecyt postdoctorado 3240563, ANID Subvención a la Instalación en la Academia (SIA) 85250093, INACH RT_12-14, Fondecyt 1150517, PIA ACT172065 GAB, Millennium Institute Biodiversity of Antarctic and Subantarctic Ecosystems (BASE), Millennium Institute Center of Genome Regulation (CRG), ANID - MILENIO ICN2021_002 and ICN2021_044, the CNPq (431463/2016-6) and the PROANTAR, IPEV programs 137 ANTAVIA and 354 ETHOTAAF and by the Spanish Research Agency (CGL2004-01348, CGL2007-60369, POL2006-06635 and CTM2015-64720-R). The Geryon cluster at the Centro de Astro-Ingeniería UC was extensively used for the calculations performed in this paper. The ANID BASAL project FB21000, BASAL CATA PFB-06, Anillo ACT-86, FONDEQUIP AIC-57, and QUIMAL 130008 provided funding for several improvements to the Geryon cluster. L.R.P. and M.G. were funded by the South African Department of Science and Innovation through the EU-Biodiversa ASICS project. The work conducted by HL was carried out in a nongovernmental capacity; the findings and conclusions in this article are those of the author(s) and do not necessarily represent the views of the U.S. Fish and Wildlife Service. We thank Sarah Crofts, Micky Reeves, and Jonathan Handley from Falklands Conservation for their contribution to the sampling. We thank Mark Adams (Natural History Museum, UK), Paul Sweet (American Museum of Natural History) and Judy Clarke (Tasmanian Museum and Art Gallery, Aust) for facilitating access to their gentoo penguin collections.

Author contributions

D.N. conceived and designed the study, coordinated the project, performed analyses, interpreted the data, and led the writing of the manuscript with support from J.A.V.; R.C.K.B., E.P., G.C., and H.L. contributed substantially to data interpretation and manuscript development; J.Y., L.R.P., M.G., E.J.P., F.L., D.Y.C.B., J.T., and W.B.S. contributed to data analyses and manuscript editing; J.M., P.P., C.L.B., F.B., P.N.T., A.B., A.R.R., and G.P.M.D. provided samples and contributed to manuscript review. All authors approved the final version of the manuscript.

Competing interests

The authors declare no competing interests

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42003-026-10081-7.

Correspondence and requests for materials should be addressed to Daly Noll or Juliana A. Vianna.

Peer review information *Communications Biology* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editors: Luciano Matzkin and Johannes Stortz. A peer review file is available.

[1]Departamento de Ciencias Biológicas y Biodiversidad, Universidad de Los Lagos, Osorno, Chile. [2]Facultad de Ciencias, Universidad de Chile, Santiago, Chile. [3]Millennium Institute Biodiversity of Antarctic and Subantarctic Ecosystems (BASE), Santiago, Chile. [4]Millennium Institute Center for Genome Regulation (CRG), Santiago, Chile. [5]Institute for Marine and Antarctic Studies, University of Tasmania, Battery Point, TAS, Australia. [6]Department of Plant and Soil Sciences, University of Pretoria, Pretoria, South Africa. [7]Museo Nacional de Ciencias Naturales, CSIC, Madrid, Spain. [8]Pontificia Universidad Católica de Chile, Facultad de Ciencias Biológicas, Santiago, Chile. [9]Center for Oceanology and Antarctic Studies (COEA), Venezuelan Institute for Scientific Research (IVIC), Caracas, Venezuela. [10]Department of Genetics, Evolution and Environment, University College London, London, UK. [11]Jean Golding Institute, University of Bristol, Bristol, UK. [12]Cornell Lab of Ornithology, Cornell University, Ithaca, NY, USA. [13]U.S. Fish and Wildlife Service, Washington, DC, USA. [14]Center for Comparative Genomics, California Academy of Sciences, San Francisco, CA, USA. [15]Australian Antarctic Division, Kingston, TAS, Australia. [16]Department of Zoology, Centre of Excellence at the Percy FitzPatrick Institute for African

Ornithology (11DST/NRF), Nelson Mandela University, Port Elizabeth, South Africa. [17]Université de Strasbourg, CNRS, IPHC UMR 7178, F-67000, Strasbourg, France. [18]Département de Biologie Polaire, Centre Scientifique de Monaco, Monaco City, Monaco. [19]CEFE, University of Montpellier, CNRS, EPHE, IRD, Montpellier, France. [20]British Antarctic Survey, Cambridge, UK. [21]Centro Austral de Investigaciones Científicas – Consejo Nacional de Investigaciones Científicas y Técnicas (CADIC-CONICET), Ushuaia, Argentina. [22]Instituto de Ciencias Polares, Ambiente y Recursos Naturales, Universidad Nacional de Tierra del Fuego, Ushuaia, Argentina. [23]Wildlife Conservation Society, Buenos Aires, Argentina. [24]PPG in Vertebrate Biology, Pontificia Universidade Católica de Minas Gerais, Belo Horizonte, Brazil. [25]Museum of Vertebrate Zoology and Department of Integrative Biology, 3101 Valley Life Science Building, University of California, Berkeley, CA, USA. [26]One Health Institute, Faculty of Life Sciences, Universidad Andres Bello, Santiago, Chile. [27]Deceased: Andrés Barbosa ✉e-mail: dalynoll@gmail.com; juliana.vianna@unab.cl