Received: 16 August 2019 | Revised: 4 April 2020 | Accepted: 8 April 2020

DOI: 10.1111/ddi.13072

BIODIVERSITY RESEARCH

Diversity and Distributions WILEY

Cryptic speciation in gentoo penguins is driven by geographic isolation and regional marine conditions: Unforeseen vulnerabilities to global change

Luis R. Pertierra[1] | Nicolás I. Segovia[2,3] | Daly Noll[2,4] | Pablo A. Martinez[5] | Patricio Pliscoff[6] | Andrés Barbosa[1] | Pedro Aragón[1,7] | Andrea Raya Rey[8,9,10] | Pierre Pistorius[11] | Phil Trathan[12] | Andrea Polanowski[13] | Francesco Bonadonna[14] | Céline Le Bohec[15,16] | Ke Bi[17] | Cynthia Y. Wang-Claypool[17] | Daniel González-Acuña[18] | Gisele P. M. Dantas[19] | Rauri C. K. Bowie[17] | Elie Poulin[2] | Juliana A. Vianna[4]

[1]Departamento de Ecología Evolutiva, Museo Nacional de Ciencias Naturales, CSIC, Madrid, Spain

[2]Departamento de Ciencias Ecológicas, Instituto de Ecología y Biodiversidad, Universidad de Chile, Santiago, Chile

[3]Departamento de Biología Marina, Facultad de Ciencias del Mar, Universidad Católica del Norte, Coquimbo, Chile

[4]Departamento de Ecosistemas y Medio Ambiente, Facultad de Agronomía e Ingeniería Forestal, Pontificia Universidad Católica de Chile, Santiago, Chile

[5]PIBi-Lab, Departamento de Biologia, Universidade Federal de Sergipe, São Cristóvão, Brazil

[6]Departamento de Ecología, Instituto de Geografía, Pontificia Universidad Católica de Chile, Santiago, Chile

[7]Facultad de Ciecias Biológicas, Departamento de Biodiversidad, Ecología y Evolución, Universidad Complutense de Madrid, Madrid, Spain

[8]Centro Austral de Investigaciones Científicas – Consejo Nacional de Investigaciones Científicas y Técnicas (CADIC-CONICET), Ushuaia, Argentina

[9]Instituto de Ciencias Polares, Ambiente y Recursos Naturales, Universidad Nacional de Tierra del Fuego, Ushuaia, Argentina

[10]Wildlife Conservation Society, Buenos Aires, Argentina

[11]DST/NRF Centre of Excellence at the Percy FitzPatrick Institute for African Ornithology, Department of Zoology, Nelson Mandela University, Port Elizabeth, South

Abstract

Aim: The conservation of biodiversity is hampered by data deficiencies, with many new species and subspecies awaiting description or reclassification. Population genomics and ecological niche modelling offer complementary new tools for uncovering functional units of phylogenetic diversity. We hypothesize that phylogenetically delineated lineages of gentoo penguins (*Pygoscelis papua*) distributed across Antarctica and sub-Antarctic Islands are subject to spatially explicit ecological conditions that have limited gene flow, facilitating genetic differentiation, and thereby speciation processes.

Location: Antarctica and sub-Antarctic area.

Methods: We identify divergent lineages for gentoo penguins using ddRAD-seq and mtDNA, and generated species distribution models (SDMs) based on terrestrial and marine parameters.

Results: Analyses of our genomic data supports the existence of four major lineages of gentoo penguin: (i) spanning the sub-Antarctic archipelagos north of the Antarctic Polar Front (APF); (ii) Kerguelen Island; (iii) South America; and (iv) across maritime Antarctic and the Scotia Arc archipelagos. The APF, a major current system around Antarctica, acts as the most important barrier separating regional sister lineages. Our ecological analyses spanning both the terrestrial (breeding sites) and marine (feeding sites) realms recover limited niche overlap among the major lineages of gentoo penguin. We observe this pattern to correspond more closely with regional differentiation of marine conditions than to terrestrial macroenvironmental features.

Main conclusions: Recognition of regional genetic lineages as discrete evolutionary entities that occupy distinct ecological niches and also differ morphologically should be considered a priority for conservation. Gentoo penguins provide a good example of

Africa

[12]British Antarctic Survey, Cambridge, UK

[13]Australian Antarctic Division, Kingston, Tas., Australia

[14]CEFE UMR 5175, CNRS, Université de Montpellier, Université Paul-Valéry Montpellier, EPHE, Montpellier Cedex 5, France

[15]Universit de Strasbourg, CNRS, IPHC UMR 7178, F-67000, Strasbourg, France

[16]Centre Scientifique de Monaco, Département de Biologie Polaire, Monaco City, Monaco

[17]Department of Integrative Biology, Museum of Vertebrate Zoology, University of California, Berkeley, CA, USA

[18]Laboratorio de Parásitos y Enfermedades de Fauna Silvestre, Facultad de Ciencias Veterinarias, Universidad de Concepción Chillán, Santiago, Chile

[19]PPG in Vertebrate Zoology, Pontificia Universidade Católica de Minas Gerais, Belo Horizonte, Brazil

Correspondence
Juliana A. Vianna, Departamento de Ecosistemas y Medio Ambiente, Pontificia Universidad Católica de Chile, Vicuña Mackenna 4860, Macul, Santiago, Chile.
Email: jvianna@uc.cl

Funding information
Spanish Research Agency, Grant/Award Number: CGL2004-01348, CGL2007-60369 and POL2006-06635; 'Ramón y Cajal' contract, Grant/Award Number: RYC-2011-07670 and MINECO-MNCN; NICHEAPPS project, Grant/Award Number: CGL2014-56416-P; Instituto Antártico Chileno, Grant/Award Number: INACH12-14; ETHOTAAF, Grant/Award Number: IPEV Prog 354; Conselho Nacional de Desenvolvimento Científico e Tecnológico, Grant/Award Number: 482501/2013-8 and 431463/2016-6; Comisión Nacional de Investigación Científica y Tecnológica, Grant/Award Number: GAB PIA CONICYT ACT172065

Editor: April Blakeslee

how conservation policy can be directly impacted by new insights obtained through the integration of larger genomic datasets with novel approaches to ecological modelling. This is particularly pertinent to polar environments that are among the most rapidly changing environments on earth.

KEYWORDS
diversification, ecological niche overlap, gentoo penguin, subspecies

1 | INTRODUCTION

Evolutionary ecology aims to elucidate the spatial pattern of intraspecific genetic diversity and the evolutionary and ecological processes that underpin such patterns. These data enable policymakers to make informed decisions regarding biodiversity conservation and management. However, our understanding of the spatial patterns of biodiversity is often based on incomplete information (Hortal et al., 2015), with many new species and subspecies awaiting description or reclassification and an immense pool of intraspecific diversity having gone largely undocumented. As a result, biodiversity conservation is hampered by such data deficiencies that limit our understanding of the evolutionary patterns and processes that give rise to biodiversity, a situation referred to as the "Darwinian shortfall" (Diniz-Filho, Loyola, Raia, Mooers, & Bini, 2013). In this context, new techniques for studying population-level genomics and spatial variation of the ecological niche offer complementary tools for uncovering functional units of phylogenetic diversity that have heretofore been obscured (Chen et al., 2019; Pahad, Montgelard, & van Vuuren, 2019).

Over the past decade, an increasing number of studies have revealed that the macrofauna of the Southern Ocean shows contrasting patterns of intraspecific diversity, from the existence of single evolutionary units distributed all the way around Antarctica (Cristofari et al., 2016; Díaz, Féral, David, Saucède, & Poulin, 2011) and/or throughout sub-Antarctic, to a multitude of geographic clades, each restricted to a specific area (González-Wevar et al., 2019). Such endemism suggests that the isolation of populations has led to diversification through vicariance after colonization (Chenuil et al., 2018; Halanych & Mahon, 2018; Price, 2007). Oceanic fronts and the great geographic distance that separates Antarctica from the sub-Antarctic archipelagos and islands can limit dispersal and promote the divergence of evolutionary units within species (Clucas et al., 2018; Vianna et al., 2017). Moreover, this divergence may be greater when regional populations become exposed to dissimilar environments. Ultimately, the inter-regional differentiation of their ecological niches can result in shifts in allele frequency among populations that may lead to local adaptation, and given sufficient time, to speciation (De Queiroz, 2007; Graham, Ron, Santos, Schneider, & Moritz, 2004).

Within the Southern Ocean marine ecosystem, seabirds represent key trophic components that are dependent on terrestrial environments for breeding and on marine habitats for feeding. For penguins, little or no population genetic structure has been reported for most species, including species distributed across the Antarctic, chinstrap *Pygoscelis antarcticus* (Freer et al., 2015; Korczak-Abshire, Chwedorzewska, Wąsowicz, & Bednarek, 2012; Mura-Jornet et al., 2018) and emperor *Aptenodytes forsteri* (Cristofari et al., 2016; Younger et al., 2017); and the sub-Antarctic, king *A. patagonicus* (Clucas et al., 2016; Cristofari et al., 2018), macaroni *Eudyptes chrysolophus* and royal penguins *E. schlegeli* (Frugone et al., 2018, 2019). In contrast, rockhopper penguins exhibit considerable population-level philopatry, leading to significant phylogeographic structure across both the sub-Antarctic and sub-tropical oceans, and recently to the designation of three distinct species *E. moseleyi*, *E. filholi* and *E. chrysocome* (Frugone et al., 2018).

Molecular studies of the gentoo penguin *P. papua* have revealed old and cryptic lineage diversification across the Antarctic and sub-Antarctic. This deep genetic structure among populations can be explained by the Antarctic Polar Front (APF) separating colonies, by the large geographic distance among breeding colonies, and life history traits such as high natal philopatry and the coastal lifestyle of gentoo penguins limiting dispersal (Clucas et al., 2018; de Dinechin et al., 2012; Levy et al., 2016; Vianna et al., 2017). Diversification between gentoo penguin colonies from Crozet, Kerguelen, the Falkland/Malvinas Islands and Antarctica took place between 3.6 and 1.3 million years ago (Mya; Vianna et al., 2017). The geographic distribution of the gentoo lineages as recovered using molecular DNA data (Vianna et al., 2017) is partially inconsistent with the present classification of subspecies using morphology: northern gentoo (*P. papua papua*) are distributed north of 60°S across the sub-Antarctic region, and southern gentoo (*P. papua ellsworthii*) are distributed between 60° and 65°S around Antarctica (Stonehouse, 1970). Diversification of gentoo penguin clades (lineages) could be explained in terms of vicariance processes induced and/or reinforced by geographic barriers, followed by selective forces in response to local environmental variables. Penguin species require both terrestrial breeding areas with suitable conditions for thermoregulation that favour their reproduction and nearshore marine habitats that supply sufficient food resources. Understanding the association of each cryptic lineage with its local environment shows affinities or local adaptation which may in turn be used to investigate the drivers and limitations of how lineages may respond to future environmental change. This is particularly relevant in species with deep intraspecific genetic structure to enable accurate designation of the conservation status of member of a species complex by, for example, IUCN or Birdlife International. The gentoo penguin is currently listed as a single species widely distributed across the sub-Antarctic region and part of the Antarctic Peninsula, whose category by IUCN is "Least Concern" due to its stable population trends. However, consideration of the spatial structure of *P. papua* lineages might necessitate revising their conservation status.

In the present study, we first explore the biogeographic extent and drivers of the global *P. papua* distributional range by modelling the respective marine and terrestrial ecological niches of this penguin species as a whole. Using mtDNA sequences and genome-wide SNP, we then establish the genetic relationships among a comprehensive sampling of gentoo penguin populations distributed across the range of the species. Our sampling includes previously unstudied colonies, such as from Marion, Martillo and Macquarie Islands. In addition, we determine and compare the terrestrial and marine ecological niche and macroenvironment envelopes that each lineage occupies. Finally, we investigate the role of ecology as a driver of lineage differentiation among gentoo penguin colonies.

2 | METHODS

2.1 | Species distribution modelling

We evaluated the *Pygoscelis papua* biogeographic range and associated spatial ecological niche drivers with a species distribution model (SDM; Figure 1, Figures S1–S4). Available georeferenced data points for all known gentoo breeding colonies were compiled and supplemented with information on the spatial presence of penguins taken directly from XY-coordinated Maritime Antarctica monitoring sites (Woehler, 1993), from additional literature (Lescröel & Bost, 2006) and from governmental reports on South Atlantic and sub-Antarctic territory dependencies. All point data were filtered by the spatial resolution of environmental data to a 5 arc-min resolution. Climatic and macroecological variables used as environmental predictors for SDMs were extracted from the databases WorldClim2 (terrestrial; Fick & Hijmans, 2017) and Bio-ORACLE 2.0 (marine; Assis et al., 2017). We used a Pearson correlation test ($r > .90$) to examine collinearity between all pairs of variables offered by these repositories, and when these pairs had $r > .90$, we kept those with higher potential biological

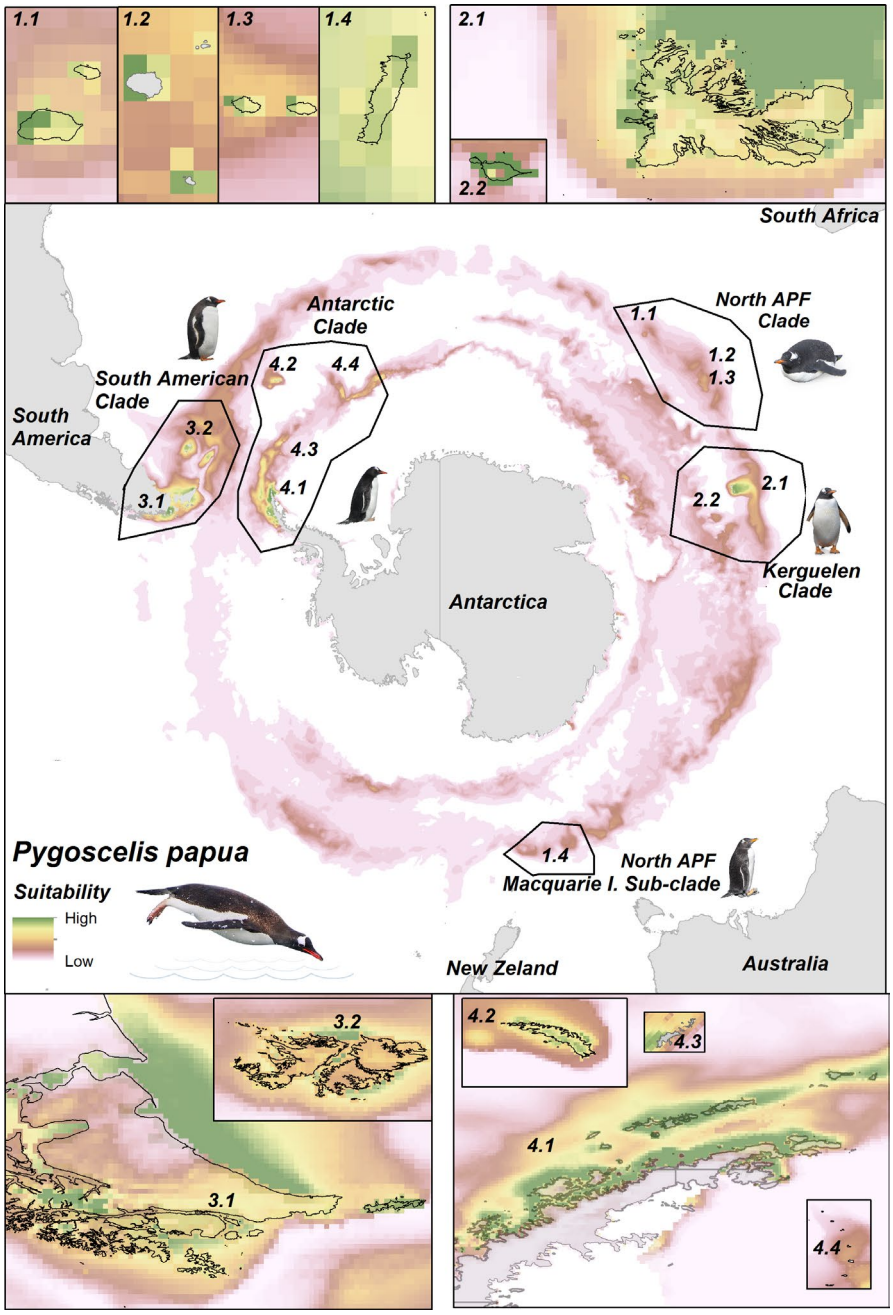

FIGURE 1 Study area with MaxEnt's climatic favourability scores for the global *Pygoscelis papua* breeding areas, including both marine and terrestrial environments. Cells containing both marine and terrestrial scores indicate the terrestrial suitability results. The green to brown colour gradient indicates higher to lower macroclimatic suitability. Inset 1.1, Marion and Prince Edward Islands; 1.1, 1.2, Crozet Island; and 1.4, Macquarie Island. 2.1, Kerguelen Island and 2.2, Heard Island. 3.1, Patagonia and 3.2, Falkland/Malvinas Island. 4.1 Ant. Peninsula; 4.2, S. Georgia Island; 4.3, S. Orkney Island; and 4.4, South Sandwich Islands

relevance. The final selection was composed of 7 variables for modelling the marine environment: sea ice cover maximum (o1), max/min primary productivity (o9/o10), max/min salinity (o11/o12) and max/min surface water temperature (o13/o14); and another set of 7 variables for modelling the terrestrial environment: mean diurnal temperature range (bio2), temperature isothermally (bio3), temperature annual range (bio7), mean temperature of the warmest calendar quarter (bio10), precipitation seasonality (bio15), mean precipitation of the wettest quarter (bio16) and mean precipitation of the warmest quarter (bio18). All SDMs were built with the MaxEnt algorithm (Phillips, Anderson, & Schapire, 2006). Logistic MaxEnt outputs provided suitability gradients that helped us to visualize terrestrial and marine macroenvironmental preferences for the species as a whole (Figure 1). Where the two habitats overlapped, we displayed the terrestrial output in "hybrid" cells and inferred marine suitability based on adjacent cells. We tested the models with a 30% random subset and calculated the true statistic skill (TSS) on the minimum training presence threshold as an indication of the robustness of the models.

2.2 | Samples collection for genetic data

We evaluated genome-wide SNP using ddRAD data and the mtDNA control region for gentoo penguin across the Southern Ocean (Figure 2). For ddRAD data, we analysed a total of 110 individuals, up to 13 individuals per population (Table 1). For mtDNA, we analysed a total of 303 individuals from several locations in Antarctica and the Scotia Arc, from the islands of Kerguelen, Crozet, Marion

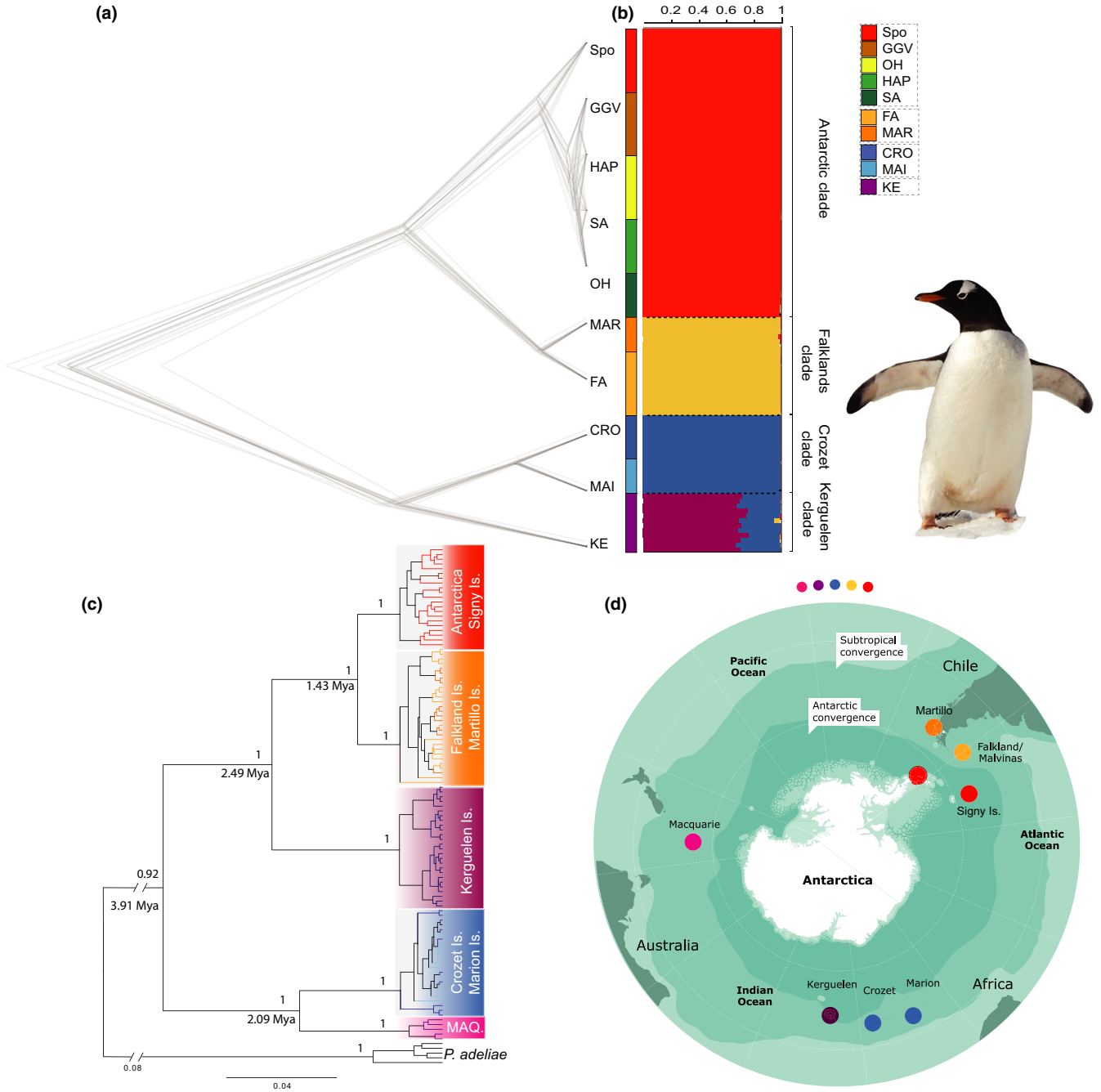

FIGURE 2 Genomic SNP data analyses supporting the existence of four main genetic groups and several sub-clusters or lineages (a–d). (a) SNAPP tree supporting the four main clusters and seven divergent lineages; (b) results of STRUCTURE for all samples (K = 4) which differentiate all four main clades: (1) Antarctica and Scotia Arc; (2) Falkland and Martillo Island; (3) from Crozet and Marion; and (4) Kerguelen Island. Each individual is represented by horizontal lines. (c) mtDNA Bayesian phylogenetic reconstruction of all gentoo penguins for mtDNA HVRI. Bayesian posterior probabilities (BPS) are located above the node and divergence time in Mya below the node. (d) Map with all sampled locations; different colours indicate different clusters obtained in the data analysis: Antarctica and Scotia Arc clade in red, South America clade in orange (lighter colour for Falkland), Crozet and Marion clade in blue (lighter colour for Marion), Kerguelen in purple and Macquarie in pink

and Falkland/Malvinas, Martillo and Macquarie Islands distributed across the sub-Antarctic. The protocol used to capture penguins, sampling procedures and permit details are provided in Appendix S1.

DNA was isolated from blood samples using the salt protocol from Aljanabi and Martinez (1997) with modifications described in Vianna et al. (2017), and from faecal samples using the QIAamp DNA Stool kit (Qiagen). We evaluated degradation of genomic DNA through electrophoresis on a 1% agarose gel. Extractions were quantified using a Qubit fluorometer (Thermo Fisher Scientific).

TABLE 1 Sample locality, sample size and diversity for mtDNA and genome SNP data

				mtDNA					SNP			
Gentoo penguin locations	**Location code**	**Latitude**	**Longitude**	***N***	***H***	***S***	H_d	π	***N***	H_o	H_e	**Allele Richness**
Gabriel Gonzalez Videla base	GGV	64°49′26.12″S	62°51′29.04″W	33	24	31	0.9754	0.0144	13	0.324 ± 0.19	0.316 ± 0.16	1.40 ± 0.49
George Point	GP	64°42′54.96″S	62°41′42.38″W	11	9	15	0.9455	0.0112	—	—	—	—
Bernardo O'Higgins	OH	63°19′14.64″S	57°53′55.25″W	39	16	24	0.9204	0.0118	13	0.302 ± 0.16	0.305 ± 0.14	1.40 ± 0.49
Byers Peninsula	BP	62°39′19.15″S	61°7′6.49″W	13	12	20	0.9872	0.0178	—	—	—	—
Hannah Point	HAP	62°38′2.50″S	60°34′59.27″W	20	16	24	0.9737	0.0143	11	0.314 ± 0.18	0.313 ± 0.14	1.40 ± 0.49
Ardley island	ARD	62°13′2.35″S	58°56′2.28″W	34	17	23	0.9554	0.0152	—	—	—	—
Stranger Point	Spo	62°14′14.86″S	58°35′38.26″W	12	10	20	0.9697	0.0161	13	0.309 ± 0.19	0.309 ± 0.15	1.41 ± 0.49
Admiralty Bay	ADM	62°10′2.55″S	58°25′0.10″W	26	17	24	0.9662	0.0151	—	—	—	—
Elephant Is.	EI	61°7′3.98″S	55°8′25.54″W	16	12	23	0.9583	0.0176	—	—	—	—
Signy Is.	SA	60°43′54.14"S	45°35′58.70"W	15	14	26	0.9905	0.0157	9	0.323 ± 0.18	0.319 ± 0.14	1.38 ± 0.49
Martillo Is.	MAR	54°54′1.79"S	67°23′45.35"W	10	3	8	0.6444	0.0093	7	0.391 ± 0.20	0.371 ± 0.13	1.32 ± 0.47
Falkland Is.	FA	52°20′16.7″S	59°21′48.02″W	19	11	15	0.9357	0.0098	13	0.301 ± 0.18	0.306 ± 0.15	1.43 ± 0.49
Crozet Is.	CRO	46°25′47.30″S	50°24′16.83″E	10	5	9	0.8034	0.0130	9	0.346 ± 0.19	0.345 ± 0.14	1.53 ± 0.50
Marion Is.	MAI	46°58′25.81"S	37°41′54.43"E	13	4	5	0.6538	0.0048	10	0.306 ± 0.17	0.316 ± 0.14	1.53 ± 0.50
Macquarie Is.	MQ	54°38′1.37"S	158°52′32.96"E	5	5	11	1	0.0166	—	—	—	—
Kerguelen Is.	KE	49°16′52.28″S	70°32′27.90″E	27	7	41	0.8667	0.0214	12	0.311 ± 0.17	0.317 ± 0.14	1.67 ± 0.47

Note: N indicates sample size, mtDNA H the number of haplotypes, *S* the number of polymorphic sites, H_d the haplotype diversity, π the nucleotide diversity, SNP H_o the mean observed heterozygosity, H_e the mean expected heterozygosity and F_{IS} the deviation of Wright's F_{IS} index, Allele Richness.

2.3 | ddRAD library preparation

We prepared ddRAD libraries for gentoo penguins, following the protocol described in Peterson, Weber, Kay, Fisher, and Hoekstra (2012). Genomic DNA (500 ng) of each individual was digested using 0.5 μl of EcoRI (0.1 U/μl) and 0.5 μl of SphI-HF (0.1 U/μl) at 37°C for three hours. Each sample was then ligated to one of 24 unique barcodes (P1 and P2). Pools of 24 samples were size-selected for fragments between 300 and 400 base pairs (bp) using Pippin Prep (Sage Science). After size selection, integrity and quantification of samples were assessed using the Agilent 2100 Bioanalyzer system (Agilent). Each library was amplified using 8–10 PCR amplification cycles and dual-indexed using Illumina adapters (P5 and P7; Peterson et al., 2012). A final quantification was performed using the Qubit 2.0 fluorometer (Thermo Fisher Scientific). Libraries were sequenced across three lanes of the Illumina HiSeq 4000 platform at the Vincent J. Coates Genomics Sequencing Laboratory (Q3B, University of California, Berkeley).

2.4 | ddRAD data processing

SNP sets were produced from raw reads assembled to the gentoo penguin reference genome (Appendix S1) using STACKS version 2.2 (Catchen, Hohenlohe, Bassham, Amores, & Cresko, 2013; Rochette, Rivera-Colón, & Catchen, 2019). Quality filtering and demultiplexing was performed using process_radtags truncating all reads to 90 nucleotides to avoid the lower quality bases at the end of the read. For SNP calling, we used a minor allele frequency of 5% and a site minimum count of 80% to restrict the presence of missing data (Ns) in the final dataset using samples with a minimum of 9× of average depth (Table S1). Before data analysis, we estimated Hardy–Weinberg equilibrium (HWE) per locus and per population with ARLEQUIN 3.5.2.2 (Excoffier & Lischer, 2010) using 10,000 permutations. After an FDR correction (*q*-value = 0.05), SNP that appeared in HW disequilibrium in at least 30% of the populations was filtered out. The ddRAD data are available at DataDryad (https://doi.org/10.5061/dryad.s7h44j140).

2.5 | mtDNA sequencing

The mitochondrial control region (hypervariable region 1: HVR1) was PCR-amplified using primers tRNAGlu and AH530 from Roeder, Ritchie, and Lambert (2002). All gentoo penguin mtDNA was Sanger sequenced with an ABI 3730xl at Macrogen (Korea), edited using SEQUENCHER v. 5.1 (Gene Codes) and aligned using CLUSTALX v. 2.1 (Larkin et al., 2007). Polymorphic sites and haplotypes were identified using the program DNASP v. 5.0 (Librado & Rozas, 2009).

2.6 | Diversity indices

We evaluated the differences in genetic diversity for whole genome SNP coverage across the breeding colonies. This was done by calculating the expected heterozygosity (H_e), observed heterozygosity (H_o) and allelic richness (A_r) with rarefied allele counts, using the HIERFSTAT package version 0.04–22 (Goudet, 2005) in R v 3.5.1 (R Core Team, 2018).

For mtDNA HVRI sequences, we characterized the genetic diversity of each population for all species (Table 1). We used ARLEQUIN v. 3.5.2.2 (Excoffier & Lischer, 2010) to calculate the following summary statistics: number of polymorphic sites (*S*), haplotype number (*H*), haplotype diversity (H_d), nucleotide diversity (π) and pairwise difference (Π, average number of nucleotide differences between sequences).

2.7 | Population genetic structure

To assess the influence of varying numbers of loci on determining population genetic structure, we generated 6 random subsets each of 50, 500, 1,000, 2,000, 3,000 and 4,000 loci (Figure S5). For each set of the above number of loci, we performed a DAPC analysis (Jombart, Devillard, & Balloux, 2010) in Adegenet (Jombart, 2008; Jombart & Ahmed, 2011) to estimate both, the number of genetic groups using the Bayesian information criterion (BIC), and to determine the genetic structure of each subset using a number of principal components equal to *N*/3 where *N* is the total number of individuals. We also calculated the pairwise F_{ST} for the total number of SNP, and F_{ST} and φ_{ST} for the mtDNA HVRI data, among locations using ARLEQUIN v. 3.5.2.2. We summarized the results from Arlequin graphically using the R functions for Arlequin XML files (Figure S6). Statistical significance of the estimates was determined with 10,000 permutations. The p-value for pairwise F_{ST} and φ_{ST} between populations was corrected using a FDR.

To determine the number of genetic groups using the total number of SNP after filtering procedures, a Bayesian clustering approach was implemented using STRUCTURE v 2.3.4 (Pritchard, Stephens, & Donnelly, 2000). Clusters (K) varied from one to eleven, corresponding to the number of breeding colonies sampled plus one (Figure 2b). Ten replicate runs were performed in parallel using StrAuto (Chhatre & Emerson, 2017). For each run, the genetic ancestry of each individual was estimated based on the admixture model without any prior population assignment under a correlated frequency model, with 500,000 Markov chain Monte Carlo (MCMC) replicates and with a 10% burn-in period. The 10 replicates obtained for each value of *K* were summarized with CLUMPP (Jakobsson & Rosenberg, 2007) and plotted using DISTRUCT (Rosenberg, 2004). The optimal value of *K* was identified according to the Evanno's method (Evanno, Regnaut, & Goudet, 2005) as implemented in Structure Harvester (Earl & vonHoldt, 2012).

2.8 | Species Tree, phylogenetic reconstruction and divergence time

The species tree SNP data were generated in SNAPP version1.3.0 (Bryant, Bouckaert, Felsenstein, Rosenberg, & RoyChoudhury,

2012) in BEAST2 version 2.4.7 (Suchard et al., 2018) using the full SNP dataset of a subset of five random individuals per sampled site (Figure 2a). Gamma prior distributions (2, 2,000) were used for the ancestral population size parameter (*h*). We used a log-likelihood correction and sampled the coalescent rate and the remaining parameters at default values. We ran two independent runs for each prior using different starting seeds for ≥1 million Markov chain Monte Carlo (MCMC) generations, sampling every 1,000 steps with 10% of trees as the burn-in period. We used TRACER 1.6 to check for convergence of the chains, and the effective sampling size (ESS) for all parameters was >500 (Rambaut, Drummond, Xie, Baele, & Suchard, 2018). Replicated runs were combined using LOGCOMBINER version 2.4.1, and TREE ANNOTATOR v. 2.4.7 was used to generate a consensus tree. The likely species trees were displayed using DENSITREE version 2.2.1 (Bouckaert, 2010).

Bayesian phylogenetic reconstruction and divergence time estimation were implemented in the program BEAST v. 2.4.7. for mtDNA HVRI (Figure 2c, Figure S7). Five Adélie penguin samples were sequenced and then incorporated into the phylogeny (KX925508–KX925512), and a sequence from the emperor penguin (*Aptenodytes forsteri*) was used as the outgroup (Li et al., 2014). The model of nucleotide substitution implemented was determined using JMODELTEST2 v. 2.1.10 (Darriba, Tab oada, Doallo, & Posada, 2012) and the Akaike's information criterion (AIC). The best-fitting model of nucleotide substitution was HKY + I+G. Divergence time among gentoo penguin lineages was based on the placement of the fossil *Pygoscelis grandis* (7.6 ± 1.3 Mya, Walsh & Suárez, 2006), which was used to calibrate the node leading to *Pygoscelis* under a normal distribution. A strict molecular clock model was applied under a Yule process tree prior. Four independent runs were performed using 30 million generations with parameters logged every 1,000 generations; a burn-in of 10% trees was used. The four independent runs were combined using LOGCOMBINER v.2.4.1. The parameter analyses were assessed for convergence and effective sample size (ESS) using TRACER v. 1.6. Finally, TREE ANNOTATOR v. 2.4.7 was used to generate a consensus tree, and FIGTREE v1.4.4 (Rambaut et al., 2018) was used to visualize the tree.

2.9 | Genomic-based species delimitation

Species delimitation hypotheses were tested using the SNP data with a species delimitation method that use Bayes factors (BFD*; Leaché, Fujita, Minin, & Bouckaert, 2014) implemented in SNAPP. Alternative species delimitation scenarios were allowed to be compared with this method in an explicit MSC framework by calculating and comparing marginal likelihood estimates (MLE) for each evaluated model.

We conducted several independent runs in Path Sampler (Lartillot & Philippe, 2006) in BEAST with 12 steps each consisting of 100,000 MCMC generations. We used a burn-in of 10,000 generations, after which we sampled every 100 steps using an alpha value of 0.3. These settings were sufficient to ensure convergence and obtain ESS > 500. The Bayes factor (BF) test statistics were calculated, where BF is the difference in MLE (Marginal L-Estimate) between all competing models. Three competing species delimitation hypotheses were defined based on current taxonomy following Stonehouse, (1970), geographic distribution of putative species and our phylogenomic analyses. To avoid over-parametrization, we ran each model using a gamma distribution (2, 2,000) as prior distribution for the ancestral population size parameter (*h*), that is the "intermediate population size" scenario used for SNAPP analyses.

2.10 | mtDNA species delimitation

Two different species delimitation methods were employed to evaluate the importance of mtDNA lineage structure across the geographic range of gentoo penguins, the Automatic Barcoding Gap Discovery (ABGD) method (a non-tree-based method) and Generalized Mixed Yule Coalescent (GMYC) method (a single locus, tree-based method). The ABGD method uses genetic distance to detect a "barcoding gap" between candidate species based on genetic distance values that are not overlapping among intra- and interspecific comparisons and are independent of tree topology. The ABGD method was performed on the online web server (http://wwwabi.snv.jussieu.fr/public/abgd/) and was run with the default settings (Pmin = 0.001, Pmax = 0.1, Steps = 10, X (relative gap width) = 1.5, Nb bins = 20). The mtDNA HVRI sequence alignment (without outgroup) was used to compute a matrix of pairwise distances using simple uncorrected distance. The GMYC method was implemented in R package SPLITS (Ezard, Fujisawa, & Barraclough, 2009). This method is based on an ultrametric phylogenetic tree such as one calibrated using a molecular clock with dissimilarities of branching rates used to infer species boundaries following a Yule process and neutral coalescent events.

2.11 | Quantification of ecological niche overlap

We examined the ecological niche overlap between the main genetic clusters delineated by SNP by applying the ordination techniques proposed by Broennimann et al. (2012) using the ECOSPAT R package (Di Cola et al., 2017) where we tested Schoener's *D* index as a measure of niche overlap ranging from 0 (no overlap) to 1 (complete overlap). We built comparative SDMs through raw individual MaxEnt models on each of these clusters in both the marine and terrestrial environments of the breeding areas. We ran equivalence tests to evaluate whether the genetic clades occupy non-identical ecological niches, and subsequently ran Schoener's *D* index similarity tests to evaluate niche similarity, that is whether a clade resembled others at more than a random level (i.e., differed from null expectations). For both tests, we conducted 100 null model simulations to compare observed and simulated *D* distributions between each pair of clades (6 permutations) on the overlap of their niches at both at the terrestrial and marine levels. Niche occupancy results for each clade are displayed as density clouds in spatial principal component analyses (sPCA, Figure 3).

2.12 | Bioclimatic variables and population genetic structure

To determine the relative contribution of geographic position and each bioclimatic or ecological variable to the genetic structure of the neutral genotypes, we followed genotypic association analyses with a partial redundancy analysis (RDA; Figure 4) in the VEGAN package (Oksanen et al., 2019). For this, first the spatial genetic structure was estimated using the geographic coordinates of the sampling sites based on distance-based Moran's eigenvector maps, dbMEMs (Dray, Legendre, & Peres-Neto, 2006; Legendre & Legendre, 2012). dbMEMs were determined by converting latitude and longitude into cartesian coordinates using the SODA package in R 3.22, and with these, a matrix of Euclidean distances was calculated using the *dist* function in VEGAN in R (Oksanen et al., 2019). Using this matrix, a rectangular matrix was created with the dbMEMs associated with latitude (dbMEM2) and longitude (dbMEM1) using the *create.dbMEM.model* function in the ADESPATIAL package. Prior to the analysis, genotype data were standardized by removing the broadscale trend using the *decostand* function with the Hellinger's method in VEGAN. A partial RDA was used to evaluate the environmental variables as fixed factors and dbMEM vectors as co-variables to control for the effect of the spatial distribution in the genetic structure. We determined the optimal model with respect to the environmental factors that best explained the genetic variability using the *ordistep* function in VEGAN according to their significance, *F*-ratio and AIC. We used a marginal ANOVA with 10,000 permutations to evaluate the significance of each fixed factor considered.

To estimate the relationship between genetic, geographic and environmental (terrestrial and marine) distances, all values of

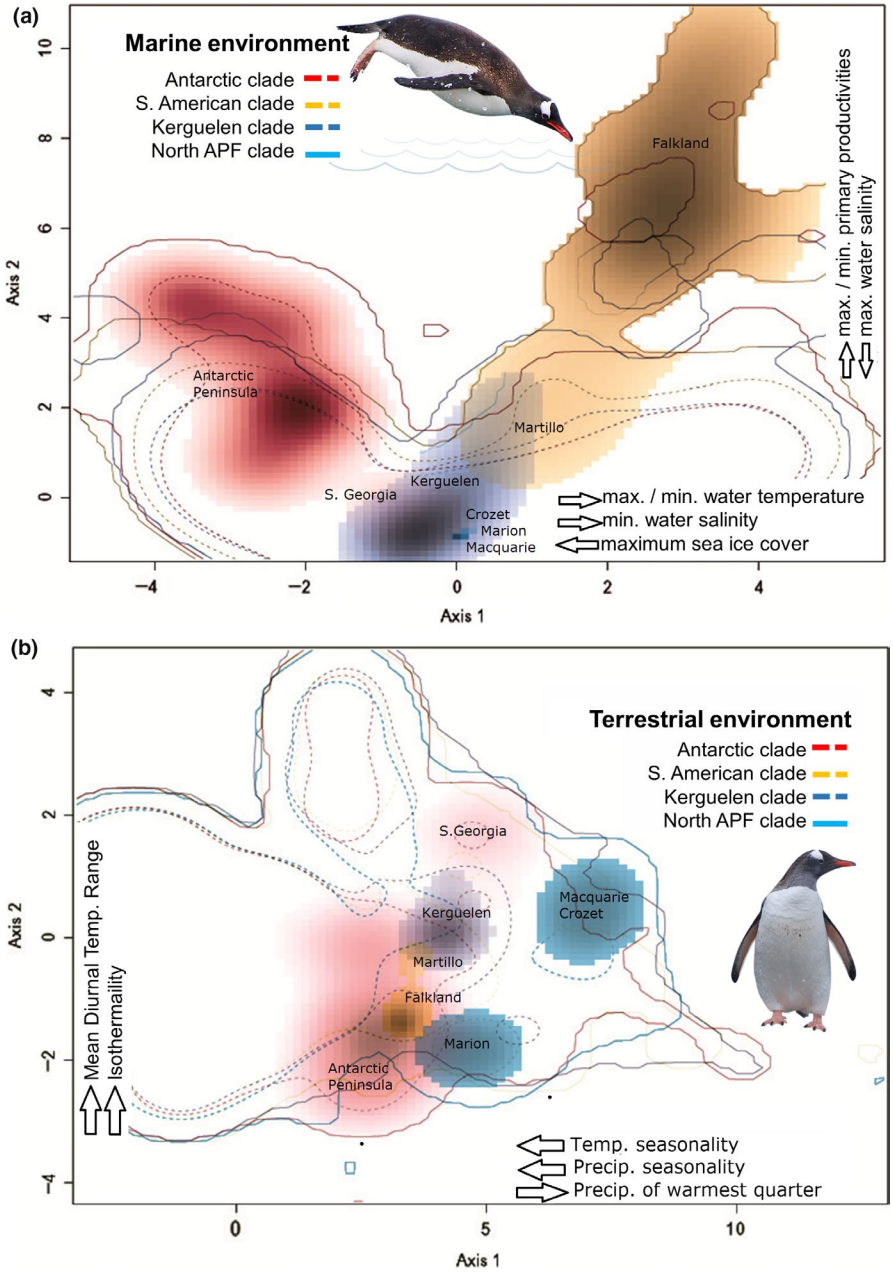

FIGURE 3 Spatialized principal component analysis of the marine (a) and terrestrial (b) niche for the four main gentoo penguin clades, visualized as density clouds. Border densities are highlighted to indicate areas of overlap. Main variable effects per axis are labelled with their sign. Solid and dashed lines, respectively, indicate 50% and 100% of the available background climate estimates. We also delineated the occurrence of density edges in the clades' areas of overlap in order to generate clearer visualization. Each colour represents the different clades, as indicated in Figure 1

matrices were first standardized (x−mean(x)/SD(x)). For each set of environmental variables, we performed a principal component analysis (PCA) using the *prcomp* function in R. Using the first component of each analysis (PC1), we calculated the multi-variable environmental distances between localities on the PC axes. Resulting values were summarized as multi-variable environmental distance matrices for terrestrial, marine and total environment occupied by gentoo penguins, respectively.

We then performed (a) a partial Mantel test using genetic distance (i.e. F_{ST}) and geographic distance between colonies, using environmental distances (marine, terrestrial and combined) as co-variates, and, conversely, (b) we performed a partial Mantel test using genetic distance and environmental distances with geographic distances as a covariate (Figure 5, Figure S8). Partial Mantel test was carried out using the ECODIST package in R (Goslee & Urban, 2007). Finally, to estimate the joint effect of geographic and environmental distances we performed a Multiple Matrix Regression with Randomization analysis (MMRR; Wang, 2013) using the package POPGENREPORT (Adamack & Gruber, 2014; Gruber & Adamack, 2015). For each set of variables, we previously

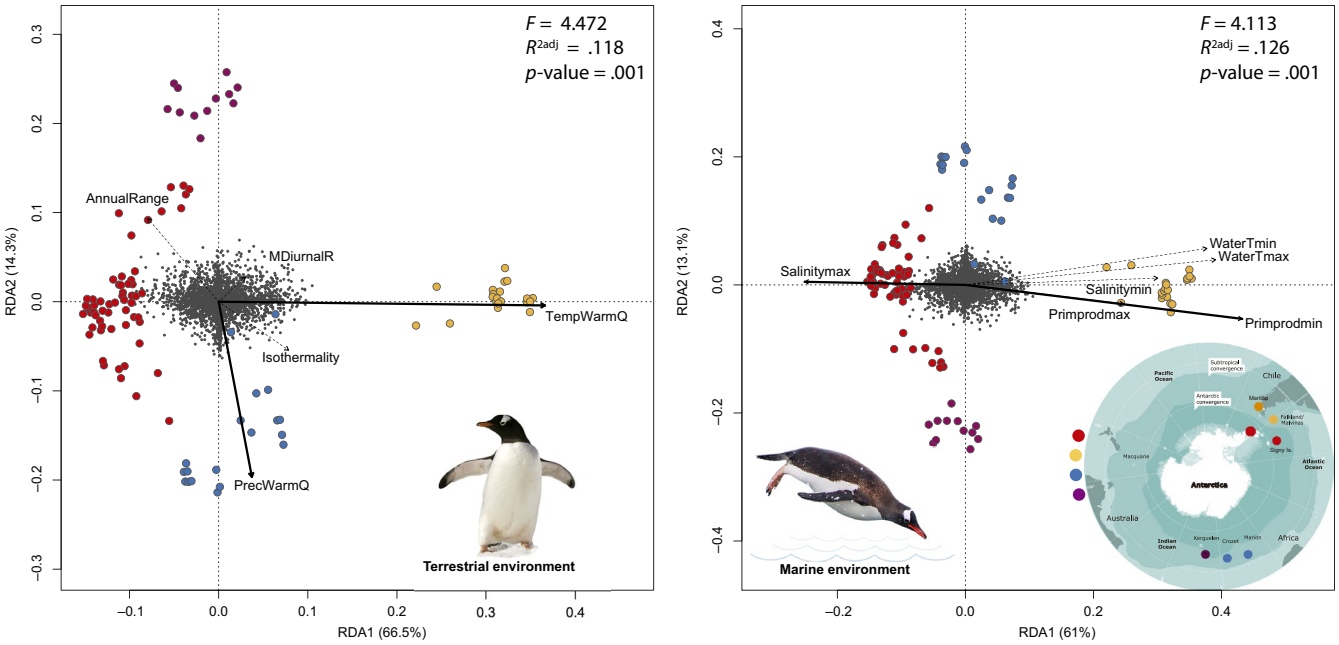

FIGURE 4 Partial redundancy analysis (RDA) showing the relative contribution of terrestrial (left) and marine (right) bioclimatic/environmental variables to the genetic structure of the gentoo penguin controlling for the effect of space (using dbMEM spatial vectors). Plot shows the optimal model obtained with *ordistep* in VEGAN. SNP genotypes are in grey (in the centre of each plot), and individuals are represented by different colours according to location, as indicated on the map (Figure 1). The most relevant variables are represented with a thick black line

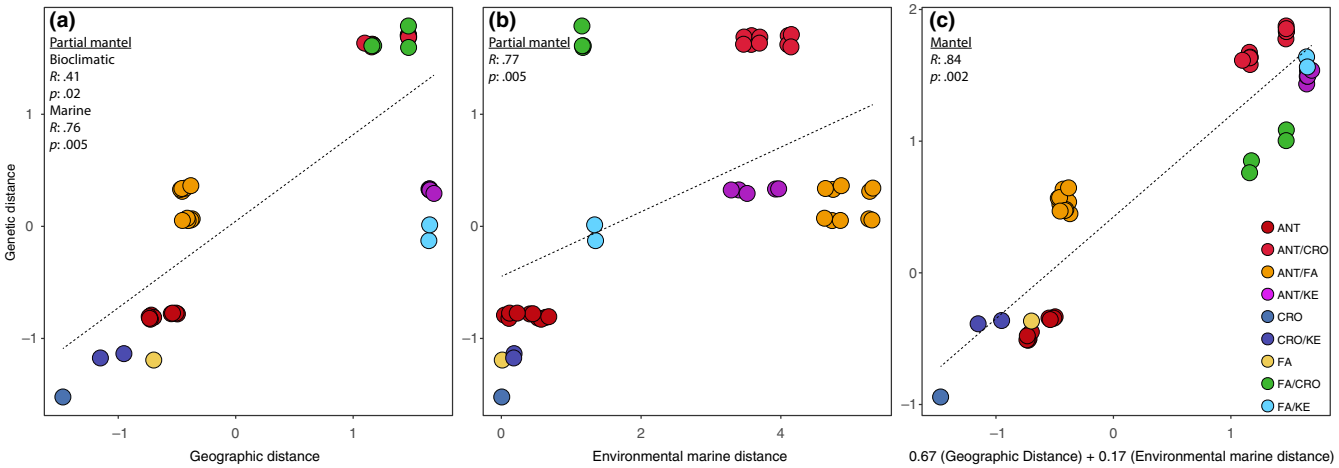

FIGURE 5 Scatter plots showing the relationship of (a) isolation by distance, (b) isolation by marine environment and (c) the joint effect of geographic distance and marine environment based on the results of a Multiple Matrix Regression with Randomization analysis (MMRR). Colour of the points represents pairwise comparisons of each considered local population

performed a Mantel test to evaluate the correlation with the geographic distance in order to fulfil the no-correlation assumption of the MMRR analysis (Wang, 2013). With the weight of their relative contribution on genetic differentiation as measured by MMRR analysis, we constructed a new distance matrix and conducted a Mantel test between the genetic distance and the combined effects of geographic distances and environmental distance.

3 | RESULTS

3.1 | Species distribution modelling

Gentoo penguins are widely distributed across the Southern Ocean and inhabit many of its islands and coasts. Our SDM approach allowed us to identify preferred conditions in both the terrestrial and marine environments inhabited by gentoo penguins during the breeding season. Predictive performance (TSS) was high for both marine (0.85) and terrestrial (0.96) models. Gentoo penguins have a strong preference for breeding around waters of high primary productivity across a range of temperature and salinity levels, with summer temperatures at breeding colonies oscillating around a few degrees above zero, a limited diurnal range of temperatures, and moderate summer precipitation levels (Figure 1, Figures S1–S4).

3.2 | Genetic structure and lineages

Using a reduced genome approach (ddRAD) after HWE filtering, we obtained 4,429 SNP for 110 individuals across the Southern Ocean, with a median coverage of 97.33× (Table S1) and a mean quality score of 35. Using mitochondrial DNA, we identified a total of 145 haplotypes from 303 gentoo penguins. Genetic diversity was similar across populations (SNP H_o = 0.30–0.39; mtDNA H_d = 0.64–1.0; Table 1).

The SNP data showed agreement across the six random subsets of data comprising different numbers of loci for DAPC analyses (Figure S5) and our coalescent-based trees generated using SNAPP with the total number of SNPs revealing four main clusters of gentoo penguins (Figure 2a–d): (i) a clade comprising individuals from Crozet and Marion Island; (ii) a Kerguelen clade (Kerguelen Island); (iii) a South American clade (Falkland/Malvinas and Martillo Islands); and (iv) an Antarctic clade (Antarctica and Scotia Arc). Both methods were also able to distinguish the Crozet Island population from the Marion Island population as well as the Martillo Island population from penguins on the Falkland/Malvinas Islands. Bayesian phylogenetic reconstruction (BA) based on mtDNA supported the existence of the same four divergent clades mentioned above, with lineage divergence dated to about 3.91 Mya (2.06–5.43 Mya). In the mtDNA data, individuals sampled from Macquarie Island were identified as a distinct monophyletic clade sister to individuals sampled from Marion and Crozet Islands, diverging about 2.09 Mya (1.00–2–99 Mya; Figure 2c, Figure S7); these lineages were sister to the Kerguelen lineage (2.49 Mya, 1.26–3.52 Mya), and in turn, the above two clades were sister to two clades comprising individuals sampled from Antarctica and the Scotia Arc and those from the Falkland/Malvinas Islands and Martillo Islands (1.43 Mya, 0.66–1.98 Mya; Figure 2c and Figure S7). However, the position of the Kerguelen clade differs in the genomic and mtDNA phylogenetic trees (Figure 2a,c).

STRUCTURE analyses revealed the same four genetic clusters: (1) Antarctica and the Scotia Arc; (2) the Falkland/Malvinas and Martillo Islands; (3) Crozet and Marion Islands; and (4) Kerguelen Island (Figure 2b).

All F_{ST} comparisons for SNP were significantly different from zero ranging from 0.012 to 0.042 between Antarctic locations, 0.10 between Martillo and the Falkland/Malvinas Islands and 0.052 between Crozet and Marion Islands, with between-clade values varying from 0.21 to 0.66. Significant Φ_{ST}-values for mtDNA were found between clades and for a few pairwise comparisons within clades (Figure S6).

Species delimitation of gentoo penguins was evaluated using SNP data for three different models: (1) current taxonomy reflected by subspecies designations, (2) results from the phylogenomic analyses and (3) geographic distribution. Path Sampler found that the best-fitting model according to SNAPP is the one defined by the four mtDNA clades (MLE = −19,760.218), which is superior to the model which considered each population as a different group (MLE = −19,850.589); the most poorly fitted model reflects the current taxonomy of one species and two subspecies (MLE = −22,515.567).

The methods for species delimitation pointed to the existence of five groups consistent with the genetic clades with the inclusion of the clade from Macquarie Island. The ABGD analysis showed five groups for the first four partitions, with prior maximal intraspecific distances (*P*) ranging from 0.001 to 0.005. The GMYC analysis suggested six clusters and seven entities (including the two species incorporated as outgroups with high probability [confidence interval (CI) = 5–24, lnL of null model = 1,633.632, ML of GMYC model = 1,657.17, p = 5.99e−11]) and the threshold time of 0.95 Mya. These results support the existence of four main clades, along with the colony from Macquarie Island as distinct lineages.

3.3 | Environmental niche overlap

Once we observed the genetic structure among regional gentoo populations, we examined the question of what could be the underlying macroecological driver of lineage divergence. Niche overlap techniques, indicate the degree of ecological characteristics shared by two or more functional groups (Broennimann et al., 2012). Our environmental niche overlap analysis performed independently for the terrestrial and marine environments shows that all pairwise combinations of the ecological niches belonging to the four clades differ significantly in terms of equivalence (i.e., all have non-identical macroecological envelopes) for both the terrestrial and marine

environments (Table 2). The observed overlap in *D*-values was very low, ranging from 0% to 5% in the marine environment where only the Kerguelen clade had a small overlap with the Antarctic (1%) and South American (5%) clades. In the terrestrial environment, the overlap was also relatively small but always present, reaching a high of 15% between the Antarctic and north APF clades. The north APF clade evaluated here comprises gentoo penguins from Crozet, Marion and Macquarie Islands, not including Kerguelen Island which lies on the APF. Notably, the Kerguelen, Antarctic and South American clade overlapped with each other by 6% and 7%, respectively. These findings indicate that the ecological segregation between clades is consistently stronger in the marine than the terrestrial environment.

As all marine and terrestrial niches of the four clades are non-equivalent, we explored the niches in relation to one another (Warren, Glor, & Turelli, 2008). We found no evidence of niche evolution (dissimilarity) for any pairwise comparison. However, our niche similarity tests revealed significant results for the analyses conducted in the context of niche conservation (i.e. the niche for one clade showing greater relatedness to that of another clade than to a random simulation; Table S2). Here, both the Antarctic and South American clades had significant similarities to each other in the terrestrial environment (p = .04 and p = .03, respectively, Table S2), suggesting that these two recently diverged lineages are retaining some common ecological features from the shared ancestral macroclimatic niche; that is, they have not fully differentiated into their respective environments. In contrast, we did not find evidence of niche conservation (similarity) between the north AFP and Kerguelen clades despite their genetic relationship as sister lineages (p = .39 and p = .28, respectively, on terrestrial environment similarity, and p = .19 and p = .23, respectively, on marine environment similarity; Table S2). Therefore, clade niches are more highly differentiated among the oldest sister populations (Crozet vs. Kerguelen) and reduced on the recent diverged ones (Antarctica and South America).

Consistent with the niche overlap scores, the results of our spatial PCA show a stronger segregation of the multivariate ellipsoids in the marine environment. As shown on the first axis in Figure 3a, the increasing temperature and salinity of the marine environment for northern populations such as the Falkland/Malvinas Islands contrast with the marine environment of the southernmost Antarctic populations, related to surrounding sea ice at the latter site. The second axis (Figure 3a) shows a strong positive effect for primary productivity, with the latitudinal extremes (Antarctica and South America) sharing similar tendencies to occupy higher productivity areas. The north APF clade and Kerguelen breeding sites have lower productivity values, possibly related to the observed mismatch of the locally preferred marine and terrestrial environments (seen in Figure 1 insets), where the most favourable (greener) feeding areas are distant from the breeding coastlines. Terrestrial PCA effects are less evident (Figure 3b). Axis 1 shows a positive effect for higher summer temperatures, effectively segregating non-Antarctic populations; the locations from north of APF clade and South American sites sustain higher precipitation in the same warmer period. Variation in seasonal distribution of temperature is higher in the Antarctic and South America than to the north of the APF clade, suggesting a small oceanic effect. Axis 2 indicates higher isothermality and diurnal temperature ranges, but this effect cannot be attributed to any clade in particular.

3.4 | Environmental and genetic redundancy analyses

Optimal models of RDAs for the marine and bioclimatic variables were in general consistent with the niche overlap results described above. The first two axes of the RDA explain 80.8% and 74.1% of the total variance for the terrestrial bioclimatic (F = 4.472) and marine model (F = 4.113), respectively, making both the general models highly significant (p = .001, Figure 4). The best-fit model for bioclimatic variables included temperature and precipitations (Table S3, Figure 4). Both variables were mainly associated with the South American group (Falkland/Malvinas and Martillo Islands) and Crozet/Marion group, respectively. Our modelling using *ordistep* for marine variables revealed the best-fit models included sea water temperatures, salinity and primary productivity, which were strongly associated with populations from the Falkland/Malvinas Islands and Antarctic populations, whereas primary productivity was associated with localities north of APF such as Crozet Island (Table S3; Figure 4). The lower latitude locations such as Crozet and the Falkland/Malvinas Islands were segregated by the significant

TABLE 2 Results of the (lower than random) equivalence tests for PCA between the four clades of gentoo penguin

		Terrestrial		Marine	
Clade a	Clade b	Schoener *D*	*p*-Value	Schoener *D*	*p*-Value
North of the APF	Kerguelen	0.01	.01**	0.00	.01**
North of the APF	South American	0.01	.01**	0.00	.01**
North of the APF	Antarctic	0.15	.01**	0.00	.01**
Kerguelen	Antarctic	0.06	.01**	0.01	.01**
Kerguelen	South American	0.07	.01**	0.05	.01**
Antarctic	South American	0.06	.01**	0.00	.01**

**Significant (<.05).

and positive effects of higher summer temperatures, whereas Antarctica, Signy Island (S. Orkney) and Martillo Island (S. America) were in turn associated mainly with reduced precipitation regimes (Figure 4).

The partial Mantel test was highly significant ($p < .05$) for distance as a response variable (Figure 5a) only differing in the *r*-values when all variables were used ($R = 0.71$), terrestrial bioclimatic ($R = 0.41$) and marine ($R = 0.76$) as co-variates. The marine model was the best-fit based on the results of a Multiple Matrix Regression with Randomization analysis (MMRR) and Mantel *r*-values. A partial Mantel test between genetic distance and environmental marine distance, controlling for geographic distance, was also highly significant ($R = 0.77$, $p = .005$; Figure 5b). Geographic and marine (environmental) distances were not correlated ($R = 0.304$, $p = .064$). In turn, positive and significant correlations were found for both, geographic and bioclimatic distance ($R = 0.74$, $p = .004$) as well as using total environmental distance (marine and bioclimatic [$R = 0.48$, $p = .002$; Figure S8]). For that reason, we only used environmental marine distance from the MMRR analysis to test the influence of the joint effect of geographic distance and environmental distance. In this context, a Mantel test with the jointed effect weighed through a MMRR was the best-fit model ($R = 0.84$, $p = .002$). However, observing the distributions of each point, correlations were mainly explained by the influence of the distinct genetic groups (Figure 5c).

4 | DISCUSSION

Conservation focuses on protecting species and their habitats while inherently assuming a strong degree of niche conservation (Wiens, Stralberg, Jongsomjit, Howell, & Snyder, 2009). The wide geographic distribution of gentoo penguins around the Southern Ocean and part of the South Atlantic spans diverse marine and terrestrial abiotic conditions and suggests that the species, seen as a whole, has a wide tolerance of climatic regimes. Nonetheless, our data indicate that the species is divided into several distinct regional lineages that have adapted to exploit local environmental conditions. These lineages are spread over large distances in the Southern Ocean and subjected to spatially dynamic changes in environmental conditions associated with climate change (Swart, Gille, Fyfe, & Gillett, 2018). Changing conditions in the Southern Ocean involve, among others, coastal water becoming less saline and ocean acidification, which is expected to produce major impacts on the Antarctic biota (Convey & Peck, 2019). The growth trend of gentoo penguins as a whole is not to be taken as a representative fate for each of the genetic and ecologically different clades we identify. Changes in the Southern Ocean will likely affect more intensively peripheral colonies situated at the edges of the distributional range where the species are at the limit of their tolerances (Forcada & Trathan, 2009). However, this issue remains to be explored in depth. Such rapid changes in environmental conditions mean that at least some, if not all, breeding habitat will be at risk of becoming suboptimal over time. This raises important questions about the ecological resilience of previously overlooked cryptic lineages, which lack broad dispersal capabilities and occupy specialized niches. Thus, some of these lineages may not be equally able to adapt to the currently changing macroecological conditions and could be under local risk of extinction (Thomas et al., 2004).

Unveiling cryptic diversification events is essential to implementing informed conservation management strategies. Here, we employed multiple methods centred on using a combination of molecular (genome-wide SNP and mtDNA) and ecological data (niche models and overlap analyses) to detect pronounced diversification among gentoo penguin colonies across the Southern Ocean and to explore the underlying processes that may have led to the observed extent of lineage differentiation. High ecological variability has been described for gentoo penguins across their biogeographic range, with resulting impacts on feeding and breeding biology including laying time, chick growth (Williams, 1995), expression of colour ornaments (Barbosa, Palacios, Valera, & Martinez, 2012), the duration of foraging trips and the availability of prey among colonies (Lescröel, Bajzak, & Bost, 2009). Some of these traits, such as the timing of laying, have a genetic basis, as genotypes may be selected to match resource availability and chick rearing requirements (Charnov & Krebs, 1974), or by photoperiod-, climatic- or resource-related plasticity (Lambrechts, Blondel, Maistre, & Perret, 1997).

The behaviour of gentoo penguins may provide support for the existence of genetic differences because limited gene flow among colonies promotes differentiation and diversification. Gentoo penguins have a greater propensity for being sedentary during the non-breeding period than do other pygoscelid penguin species (Dodino, Hart, Harris, & Rey, 2018; Friesen, Burg, & Mccoy, 2007; Williams, 1995), which partly explains the degree of isolation among colonies. The gentoo penguin is a resident inshore forager (Dimitrijevic et al., 2018; Lescröel & Bost, 2005; Lescröel, Ridoux, & Bost, 2004), an attribute that may limit its dispersal, in contrast with the pelagic behaviour of other penguin species which facilitates inter-colony gene flow (Clucas et al., 2018). Moreover, natal philopatry may explain the population genetic structure detected from genomic data among breeding colonies within each clade we studied. Hence, natural selection may operate across different environments at sea, enabling local adaptation, isolation and over time speciation.

At a regional scale, using SNP data, Clucas et al. (2018) identified three main gentoo penguin clusters (i.e., Kerguelen, Falklands Is., Antarctica and South Georgia). However, the study by Clucas et al. (2018) lacked samples from Crozet and Marion, and Macquarie Islands, which are inhabited by additional lineages as described in this study. Our results reinforce the idea that the APF acts as an important barrier between sister clades (South America vs. Antarctica; and Crozet, Marion and Macquarie Islands vs. Kerguelen Island) which historically (e.g., Kerguelen; Gersonde, Crosta, Abelmann, & Armand, 2005) or currently lie on either side of the APF and show both ecological and genetic differentiation. Climatic and trophic

features can induce morphological changes as seen in the relationship between water temperature and body size. Indeed, morphological differences have been historically reported for at least two subspecies of gentoo penguin: *P. p. ellsworthii* distributed across the South Orkneys, the South Shetland Islands and the Antarctic Peninsula, which has smaller body sizes and bill proportions than *P. p. papua* from the northern parts of the species' distribution across the sub-Antarctica (Stonehouse, 1970). However, spatial variation in morphometrics is also evident in other populations, with a tendency of decreasing size towards the south of the gentoo penguin distribution (Stonehouse, 1970) and within Antarctica (Valenzuela-Guerra, Morales-Moraga, González-Acuña, & Vianna, 2013). Gentoo penguins from Macquarie Island were first described as a distinct subspecies (*P. papua taeniata*; Mathews, 1927) from those distributed across the rest of the sub-Antarctic. The subspecies, *P. papua taeniata*, was later grouped with individuals from Heard, Kerguelen and Marion Island (Peters, 1934), but the penguin population from Crozet Island was not evaluated. Gentoo penguins from Crozet have been reported in the literature as being larger than their counterparts from other locations (Falla, 1937; Stonehouse, 1970), and to resemble those from Marion Island (Crawford, 1952), which are consistently identified in this paper as part of the same genetic clade.

The divergence time estimated between the gentoo penguin clades (3.91–1.43 Mya) was similar to those estimated by Vianna et al. (2017), and the dates are similar to those estimated among species within other penguin genera (Cole et al., 2019). Our results suggest taxonomic recognition for the following four clades based on prior descriptions of morphology, type location, genomic and trophic data: (1) the Southern gentoo penguin, *P. p. ellsworthii*, distributed across Antarctica, South Orkneys, the South Shetlands Islands and South Georgia; (2) the Northern gentoo penguin, *P. p. papua*, restricted to the Falkland/Malvinas Islands and Martillo Islands; (3) the Eastern gentoo penguin, *P. p. taeniata*, first described for Macquarie but which should also include the populations on Crozet and Marion Island within the same mtDNA clade (although mtDNA shows historical divergence between Macquarie and the combined Crozet and Marion Islands, further evaluation using genomic data is necessary to support the possibility of two different taxa); and (4) the Southeastern gentoo penguin, a subspecies from Kerguelen Island, which requires formal description.

Mitochondrial DNA and genomic data support the existence of highly divergent/differentiated clades; however, the Kerguelen clade occupies a different phylogenetic position in each dataset. In the phylogenetic hypothesis constructed using genomic data, the Kerguelen clade is sister to the clade comprising individuals from Crozet and Marion Island, and in the mtDNA to individuals sampled from Antarctica and the Falkland/Malvinas and Martillo Islands. Different tree topologies between biparentally inherited SNP and maternal mtDNA may be explained by the distinct coalescence times of the markers, sex-biased dispersal, and/or introgression between lineages (Funk & Omland, 2003; Maddison, 1997).

In terms of macroecology, gentoo penguin terrestrial niches are less differentiated than those in the marine environment, a distinction we attribute to the high intra-clade homogeneity of sea conditions within feeding areas, in particular locally stable water temperatures and salinity, with larger inter-clade differences across regions appearing to be caused by latitudinal gradients of ocean stratification. Terrestrial features of penguin rookeries are locally more variable within breeding areas due to changing weather conditions, and they sustain a more homogeneous inter-regional optimum driven by the general oceanic climate present across latitudes. This pattern, which reduces the degree of climatic differentiation across terrestrial regions while promoting a rich variety of marine ecosystems, is typical of the Southern Ocean territories. In the case of gentoo penguins, we attribute genetic differentiation primarily to conditions at sea, whereas land conditions are subjected primarily to more local characteristics related to topographic features that drive nesting habitat availability.

Equivalence tests indicate that the niches of all four clades of gentoo penguins differ in terms of both the marine and terrestrial macroenvironment. Our results also suggest one instance of niche conservatism, but only in one of the terrestrial pairwise comparisons: Antarctica and South America. This could be explained by the recent age of these lineages, thereby retaining some common ecological features from the shared ancestral macroclimatic niche; that is, each lineage has not yet diverged to occupy distinct terrestrial environments. This degree of niche conservationism could also explain why gentoo penguin populations in the Antarctic Peninsula are responding positively to a changing clime, increasing their population numbers and expanding southwards as the macroecological conditions become more favourable for them (Trivelpiece et al., 2011). By extrapolation, this similarity in niches would have eroded in clades that have experienced a longer time period of climatic variation enabling local selection to occur for optimal rookery selection, such as on Crozet and Kerguelen Islands. Interestingly, in the Antarctica versus South America comparison, the marine sPCA suggests that these sister clades occurring at environmental extremes on the Scotia Arc have sought to acquire greater marine feeding resources (seen from their position in areas of higher primary production) and diverged from the ancestral state (north of APF clade) which lies between cold and warm water adapted lineages. Thus, we postulate that the marine evolutionary trade-off between thermal stress and gain in primary production where the species expanded its niche towards broader temperature ranges thanks to a higher availability of resources at both thermal extremes. In contrast, our analyses suggest that the terrestrial macroenvironment poses less of a challenge for gentoo penguins than for other penguin species, perhaps due to the species' ability to withstand a wide variation in summer temperatures. Moreover, oscillations in temperature and precipitation are less apparent across breeding sites of sister clades north and south of the APF than changes in the marine environment. This leads us to propose that an important mechanism driving diversification of gentoo penguin lineages, on top of isolation by distance, comes from a trade-off between dispersal (gene flow) and local adaptation to the spatially changing conditions in the marine environment (isolation by environment).

Overall, in the MMRR analyses we find that the genetic distances among gentoo penguin colonies are best explained by the combined effects of geographic distance and marine environmental distance. In the case of geographic distance, this is primarily expressed through the vast longitudinal distribution of oceanic islands and continental land masses across the Southern Ocean. In the case of environmental distances, this is attributed to the rapid change of water conditions due to circulation patterns of oceanic currents that occur across a short latitudinal gradient. The profound environmental gradients and long distances between genetically distinctive regional clades of *P. papua* suggest that the unique functional units will be faced with varying challenges in the face of climate change and as such should be evaluated separately, and not lumped together for the species as a whole.

5 | CONCLUSIONS

Given ongoing processes associated with global change, gentoo penguins face more significant challenges than other penguin species in maintaining healthy population numbers. This is because gentoo penguins are resident and do not migrate to more favourable habitats after breeding, instead relying upon habitats that must supply both summer and winter needs. Gentoo penguins have comparatively limited mobility and rely on the availability of suitable coastal areas for breeding and feeding during the reproductive season (Kowalczyk, Reina, Preston, & Chiaradia, 2015). In the case of the Antarctic populations, an intra-regional expansion southward may be feasible, than is currently present, but other regional populations in the sub-Antarctic islands have narrower opportunities to shift ranges and maintain their present niche. For example, gentoo penguins north of the APF (Crozet and Marion Is.) might find niche refugia only on Kerguelen Island but need to rely on their migratory capacities for this purpose. We found limited genetic migrants between populations north of the APF and Kerguelen Island, suggesting that gentoo penguins are likely to encounter severe difficulties in colonizing new areas given the pace of global change. Colonies of small size, such as gentoo penguin populations on the islands of Crozet and Marion Islands or Macquarie Island (Figure S9), might be under greater threat than populations in Antarctica given their degree of historical isolation. Overall, gentoo penguins comprise separate lineages distributed across Antarctica and the sub-Antarctic, and the local extinction of populations (lineages) would lead to a significantly loss of biodiversity. Exploring and documenting such cryptic diversity is of critical importance before such evolutionary unique lineages are irrevocably lost. This is particularly pertinent to polar environments that are among the most rapidly changing environments on earth.

ACKNOWLEDGMENTS

Financial support was provided by INACH RT_12–14, Fondecyt Project 1150517, GAB (PIA CONICYT ACT172065), the CNPq (482501/2013-8 and 431463/2016-6) and the PROANTAR, IPEV prog 354 ETHOTAAF and by the Spanish Research Agency (CGL2004-01348, CGL2007-60369, POL2006-06635 and CTM2015-64720-R). LRP was contracted by the NICHEAPPS project (CGL2014-56416-P) and recipient of an Antarctic Science Bursary (2017-18 scheme). PA was supported by "Ramón y Cajal" contract (RYC-2011-07670, MINECO-MNCN). Olga Lopez Hilzinger is thanked for graphic designs.

DATA AVAILABILITY STATEMENT

The raw VCF of ddRAD and bioclimatic/oceanographic data are available at DataDryad (https://doi.org/10.5061/dryad.s7h44j140). All gentoo penguin mtDNA sequences are available in GenBank (MK804771–MK804796 and KU514439–KU514493, KF717669–KF717743).

ORCID

Andrés Barbosa https://orcid.org/0000-0001-8434-3649
Elie Poulin https://orcid.org/0000-0001-7736-0969
Juliana A. Vianna https://orcid.org/0000-0003-2330-7825

REFERENCES

Adamack, A. T., & Gruber, B. (2014). PopGenReport: Simplifying basic population genetic analyses in R. *Methods in Ecology and Evolution*, *5*, 384–387. https://doi.org/10.1111/2041-210X.12158

Aljanabi, S. M., & Martinez, I. (1997). Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. *Nucleic Acids Research*, *25*(22), 4692–4693. https://doi.org/10.1093/nar/25.22.4692

Assis, J., Tyberghein, L., Bosch, S., Verbruggen, H., Serrão, E. A., & De Clerck, O. (2017). Bio-ORACLE v2. 0: Extending marine data layers for bioclimatic modelling. *Global Ecology and Biogeography*, *27*(3), 277–284. https://doi.org/10.1111/geb.12693

Barbosa, A., Palacios, M. J., Valera, F., & Martinez, A. (2012). Geographic variation in beak colouration in gentoo penguins *Pygoscelis papua*. *Polar Biology*, *35*, 725–731. https://doi.org/10.1007/s00300-011-1117-8

Bouckaert, R. R. (2010). DensiTree: Making sense of sets of phylogenetic trees. *Bioinformatics*, *26*(10), 1372–1373. https://doi.org/10.1093/bioinformatics/btq110

Broennimann, O., Fitzpatrick, M. C., Pearman, P. B., Petitpierre, B., Pellissier, L., Yoccoz, N. G., … Guisan, A. (2012). Measuring ecological niche overlap from occurrence and spatial environmental data. *Global Ecology and Biogeography*, *21*(4), 481–497. https://doi.org/10.1111/j.1466-8238.2011.00698.x

Bryant, D., Bouckaert, K. R., Felsenstein, J., Rosenberg, N. A., & RoyChoudhury, A. (2012). Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. *Molecular Biology and Evolution*, *29*(8), 1917–1932. https://doi.org/10.1093/molbev/mss086

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: An analysis tool set for population genomics. *Molecular Ecology*, *22*(11), 3124–3140. https://doi.org/10.1111/mec.12354

Charnov, E. L., & Krebs, J. R. (1974). On clutch-size and fitness. *IBIS*, *116*, 217–219. https://doi.org/10.1007/BF00216598

Chen, Y.-C., Nazarizadeh, M., Lei, F.-M., Yang, X.-J., Yao, C.-T., Dong, F., … Hung, C.-M. (2019). The niches of nuthatches affect their lineage evolution differently across latitude. *Molecular Ecology*, *28*(4), 803–817. https://doi.org/10.1111/mec.14980

Chenuil, A., Saucède, T., Hemery, L. G., Eléaume, M., Féral, J.-P., Améziane, N., … Havermans, C. (2018). Understanding processes at the origin of species flocks with a focus on the marine Antarctic

fauna. *Biological Reviews of the Cambridge Philosophical Society*, *93*(1), 481–504. https://doi.org/10.1111/brv.12354

Chhatre, V. E., & Emerson, K. J. (2017). StrAuto: Automation and parallelization of STRUCTURE analysis. *BMC Bioinformatics*, *18*(1), 192. https://doi.org/10.1186/s12859-017-1593-0

Clucas, G. V., Younger, J. L., Kao, D., Emmerson, L., Southwell, C., Wienecke, B., … Hart, T. (2018). Comparative population genomics reveals key barriers to dispersal in Southern Ocean penguins. *Molecular Ecology*, *27*(23), 4680–4697. https://doi.org/10.1111/mec.14896

Clucas, G. V., Younger, J. L., Kao, D., Rogers, A. D., Handley, J., Miller, G. D., … Hart, T. (2016). Dispersal in the sub-Antarctic: King penguins show remarkably little population genetic differentiation across their range. *BMC Evolutionary Biology*, *16*(1), 211. https://doi.org/10.1186/s12862-016-0784-z

Cole, T. L., Ksepka, D. T., Mitchell, K. J., Tennyson, A. J. D., Thomas, D. B., Pan, H., … Waters, J. M. (2019). Mitogenomes uncover extinct penguin taxa and reveal island formation as a key driver of speciation. *Molecular Biology and Evolution*, *36*(4), 784–797. https://doi.org/10.1093/molbev/msz017

Convey, P., & Peck, L. S. (2019). Antarctic environmental change and biological responses. *Science Advances*, *5*(11), eaaz0888. https://doi.org/10.1126/sciadv.aaz0888

Crawford, A. B., & Serventy, D. L. (1952). The birds of Marion Island, South Indian Ocean. *Emu*, *52*, 73–85. https://doi.org/10.1071/MU952073

Cristofari, R., Bertorelle, G., Ancel, A., Benazzo, A., Le Maho, Y., Ponganis, P. J., … Trucchi, E. (2016). Full circumpolar migration ensures evolutionary unity in the Emperor penguin. *Nature Communications*, *7*, 11842. https://doi.org/10.1038/ncomms11842

Cristofari, R., Liu, X., Bonadonna, F., Cherel, Y., Pistorius, P., Le Maho, Y., … Trucchi, E. (2018). Climate-driven range shifts of the king penguin in a fragmented ecosystem. *Nature Climate Change*, *8*(3), 245–251. https://doi.org/10.1038/s41558-018-0084-2

Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2012). jModelTest 2: More models, new heuristics and parallel computing. *Nature Methods*, *9*(8), 772. https://doi.org/10.1038/nmeth.2109

de Dinechin, M., Dobson, F. S., Zehtindjiev, P., Metcheva, R., Couchoux, C., Martin, A., … Jouventin, P. (2012). The biogeography of Gentoo Penguins (*Pygoscelis papua*). *Canadian Journal of Zoology-Revue Canadienne De Zoologie*, *90*(3), 352–360. https://doi.org/10.1139/z2012-016

De Queiroz, K. (2007). Species concepts and species delimitation. *Systematic Biology*, *56*(6), 879–886. https://doi.org/10.1080/10635150701701083

Di Cola, V., Broennimann, O., Petitpierre, B., Breiner, F. T., D'Amen, M., Randin, C., … Guisan, A. (2017). ecospat: An R package to support spatial analyses and modeling of species niches and distributions. *Ecography*, *40*, 774–787. https://doi.org/10.1111/ecog.02671

Díaz, A., Féral, J.-P., David, B., Saucède, T., & Poulin, E. (2011). Evolutionary pathways among shallow and deep sea echinoids of the genus Sterechinus in the Southern Ocean. *Deep Sea Research II*, *587*, 205–211. https://doi.org/10.1016/j.dsr2.2010.10.012

Dimitrijevic, D., Paiva, V. H., Ramos, J. A., Seco, J., Ceia, F. R., Chipev, N., … Xavier, J. C. (2018). Isotopic niches of sympatric Gentoo and Chinstrap Penguins: Evidence of competition for Antarctic krill? *Polar Biology*, *41*, 1665–1669. https://doi.org/10.1007/s00300-018-2306-5

Diniz-Filho, J. A., Loyola, R. D., Raia, P., Mooers, A. O., & Bini, L. M. (2013). Darwinian shortfalls in biodiversity conservation. *Trends in Ecology & Evolution*, *28*(12), 689–695. https://doi.org/10.1016/j.tree.2013.09.003

Dodino, S., Hart, T., Harris, S., & Rey, A. R. (2018). Year-round colony attendance patterns for the Gentoo Penguin (Pygoscelis papua) at Martillo Island, Tierra del Fuego, Argentina. *The Wilson Journal of Ornithology*, *130*(2), 493–501.

Dray, S., Legendre, P., & Peres-Neto, P. R. (2006). Spatial modelling: A comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecological Modelling*, *196*(3), 483–493. https://doi.org/10.1016/j.ecolmodel.2006.02.015

Earl, D. A., & vonHoldt, B. M. (2012). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, *4*(2), 359–361. https://doi.org/10.1007/s12686-011-9548-7

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology*, *14*(8), 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Excoffier, L., & Lischer, H. E. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, *10*(3), 564–567. https://doi.org/10.1111/j.1755-0998.2010.02847.x

Ezard, T., Fujisawa, T., & Barraclough, T. (2009). *splits: SPecies' LImits by Threshold Statistics. R package version 1.0-11/r29 15*. Retrieved from http://R-Forge

Falla, R. A. (1937). *Birds (vol. 1): Rep. Brit., Aust., N.Z. Antarct. Res. Exped. (series B)*.

Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, *37*(12), 4302–4315. https://doi.org/10.1002/joc.5086

Forcada, J., & Trathan, P. N. (2009). Penguin responses to climate change in the Southern Ocean. *Global Change Biology*, *15*(7), 1618–1630. https://doi.org/10.1111/j.1365-2486.2009.01909.x

Freer, J. J., Mable, B. K., Clucas, G., Rogers, A. D., Polito, M. J., Dunn, M., … Hart, T. (2015). Limited genetic differentiation among chinstrap penguin (*Pygoscelis antarctica*) colonies in the scotia arc and western Antarctic peninsula. *Polar Biology*, *38*, 1493–1502. https://doi.org/10.1007/s00300-015-1711-2

Friesen, V. L., Burg, T. M., & Mccoy, K. D. (2007). Mechanisms of population differentiation in seabirds. *Molecular Ecology*, *16*(9), 1765–1785. https://doi.org/10.1111/j.1365-294X.2006.03197.x

Frugone, M. J., López, M. E., Segovia, N. I., Cole, T. L., Lowther, A., Pistorius, P., … Vianna, J. A. (2019). More than the eye can see: Genomic insights into the drivers of genetic differentiation in Royal/Macaroni penguins across the Southern Ocean. *Molecular Phylogenetics and Evolution*, *139*, 106563. https://doi.org/10.1016/j.ympev.2019.106563

Frugone, M. J., Lowther, A., Noll, D., Ramos, B., Pistorius, P., Dantas, G. P. M., … Vianna, J. A. (2018). Contrasting phylogeographic pattern among Eudyptes penguins around the Southern Ocean. *Scientific Reports*, *8*(1), 17481. https://doi.org/10.1038/s41598-018-35975-3

Funk, D. J., & Omland, K. E. (2003). Species-level paraphyly and polyphyly: Frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annual Review of Ecology, Evolution, and Systematics*, *34*, 397–423. https://doi.org/10.1146/annurev.ecolsys.34.011802.132421

Gersonde, R., Crosta, X., Abelmann, A., & Armand, L. K. (2005). *Compilation of diatom LGM SST data from the Southern Ocean*. Retrieved from https://doi.org/10.1594/PANGAEA.227399

González-Wevar, C., Gérard, K., Rosenfeld, S., Saucède, T., Naretto, J., Díaz, A., … Poulin, E. (2019). Cryptic speciation in Southern Ocean Aequiyoldia eightsii (Jay, 1839): Mio-Pliocene trans-Drake separation and diversification. *Progress in Oceanography*, *174*, 44–54. https://doi.org/10.1016/j.pocean.2018.09.004

Goslee, S. C., & Urban, D. L. (2007). The ecodist package for dissimilarity-based analysis of ecological data. *Journal of Statistical Software*, *22*, 1–19. https://doi.org/10.18637/jss.v022.i07

Goudet, J. (2005). hierfstat, a package for r to compute and test hierarchical F-statistics. *Molecular Ecology Notes*, *5*(1), 184–186. https://doi.org/10.1111/j.1471-8286.2004.00828.x

Graham, C. H., Ron, S. R., Santos, J. C., Schneider, C. J., & Moritz, C. (2004). Integrating phylogenetics and environmental niche models to explore speciation mechanisms in dendrobatid frogs. *Evolution*, *58*(8), 1781–1793. https://doi.org/10.1111/j.0014-3820.2004.tb00461

Gruber, B., & Adamack, A. T. (2015). Landgenreport: A new r function to simplify landscape genetic analysis using resistance surface layers. *Molecular Ecology Resources*, *15*, 1172–1178. https://doi.org/10.1111/1755-0998.12381

Halanych, K. M., & Mahon, A. R. (2018). Challenging dogma concerning biogeographic patterns of Antarctica and the Southern Ocean. *Annual Review of Ecology, Evolution, and Systematics*, *49*, 355–378. https://doi.org/10.1146/annurev-ecolsys-121415-032139

Hortal, J., de Bello, F., Diniz-Filho, J. A. F., Lewinsohn, T. M., Lobo, J. M., & Ladle, R. J. (2015). Seven shortfalls that beset large-scale knowledge of biodiversity. *Annual Review of Ecology, Evolution, and Systematics*, *46*(1), 523–549. https://doi.org/10.1146/annurev-ecolsys-112414-054400

Jakobsson, M., & Rosenberg, N. A. (2007). CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, *23*(14), 1801–1806. https://doi.org/10.1093/bioinformatics/btm233

Jombart, T. (2008). Adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*, 1403–1405. https://doi.org/10.1093/bioinformatics/btn129

Jombart, T., & Ahmed, I. (2011). Adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics*, *27*, 3070–3071. https://doi.org/10.1093/bioinformatics/btr521

Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. *BMC Genetics*, *11*, 94. https://doi.org/10.1186/1471-2156-11-94

Korczak-Abshire, M., Chwedorzewska, K. J., Wąsowicz, P., & Bednarek, P. T. (2012). Genetic structure of declining chinstrap penguin (*Pygoscelis antarcticus*) populations from south Shetland Islands (Antarctica). *Polar Biology*, *35*, 1681–1689. https://doi.org/10.1007/s00300-012-1210-7

Kowalczyk, N. D., Reina, R. D., Preston, T. J., & Chiaradia, A. (2015). Environmental variability drives shifts in the foraging behaviour and reproductive success of an inshore seabird. *Oecologia*, *178*(4), 967–979. https://doi.org/10.1007/s00442-015-3294-6

Lambrechts, M. M., Blondel, J., Maistre, M., & Perret, P. (1997). A single response mechanism is responsible for evolutionary adaptive variation in a bird's laying date. *Proceedings of the National Academy of Sciences of the United States of America*, *94*(10), 5153–5155. https://doi.org/10.1073/pnas.94.10.5153

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., … Higgins, D. G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics*, *23*(21), 2947–2948. https://doi.org/10.1093/bioinformatics/btm404

Lartillot, N., & Philippe, H. (2006). Computing Bayes factors using thermodynamic integration. *Systematic Biology*, *55*(2), 195–207. https://doi.org/10.1080/10635150500433722

Leaché, A. D., Fujita, M. K., Minin, V. N., & Bouckaert, R. R. (2014). Species delimitation using genome-wide SNP data. *Systematic Biology*, *63*(4), 534–542. https://doi.org/10.1093/sysbio/syu018

Legendre, P., & Legendre, L. (2012). *Numerical ecology* (3rd ed.). Amsterdam, The Netherlands: Elsevier Science.

Lescröel, A., Bajzak, C., & Bost, C.-A. (2009). Breeding ecology of the gentoo penguin *Pygoscelis papua* at Kerguelen Archipelago. *Polar Biology*, *32*, 1495–1505. https://doi.org/10.1007/s00300-009-0647-9

Lescröel, A., & Bost, C.-A. (2005). Foraging under contrasting oceanographic conditions: The gentoo penguin at Kerguelen Archipelago. *Marine Ecology Progress Series*, *302*, 245–261. https://doi.org/10.3354/meps302245

Lescröel, A., & Bost, C.-A. (2006). Recent decrease in gentoo penguin populations at Iles Kerguelen. *Antarctic Science*, *18*(2), 171–174. https://doi.org/10.1017/S0954102006000198

Lescröel, A., Ridoux, V., & Bost, C.-A. (2004). Spatial and temporal variation in the diet of the gentoo penguin (*Pygoscelis papua*) at Kerguelen Islands. *Polar Biology*, *27*, 206–216. https://doi.org/10.1007/s00300-003-0571-3

Levy, H., Clucas, G. V., Rogers, A. D., Leache, A. D., Ciborowski, K. L., Polito, M. J., … Hart, T. (2016). Population structure and phylogeography of the Gentoo Penguin (Pygoscelis papua) across the Scotia Arc. *Ecology and Evolution*, *6*(6), 1834–1853. https://doi.org/10.1002/ece3.1929

Li, C., Zhang, Y., Li, J., Kong, L., Hu, H., Pan, H., … Zhang, G. (2014). Two Antarctic penguin genomes reveal insights into their evolutionary history and molecular changes related to the Antarctic environment. *GigaScience*, *3*(1), 27–27. https://doi.org/10.1186/2047-217X-3-27

Librado, P., & Rozas, J. (2009). DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, *25*(11), 1451–1452. https://doi.org/10.1093/bioinformatics/btp187

Maddison, W. P. (1997). Gene Trees in Species Trees. *Systematic Biology*, *46*(3), 523–536. https://doi.org/10.1093/sysbio/46.3.523

Mathews, G. M. (1927). *Systema avium Australasianarum. A systematic list of the birds of the Australasian region*. Vol. *2*. London, UK: British Ornithologists' Union.

Mura-Jornet, I., Pimentel, C., Dantas, G. P. M., Petry, M. V., González-Acuña, D., Barbosa, A., … Vianna, J. A. (2018). Chinstrap penguin population genetic structure: One or more populations along the Southern Ocean? *BMC Evolutionary Biology*, *18*(1), 90. https://doi.org/10.1186/s12862-018-1207-0

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., & Wagner, H. (2019). *vegan: Community Ecology Package. R package version 2.3-2*. Retrieved from http://cran.r-project.org/package=vegan

Pahad, G., Montgelard, C., & van Vuuren, B. J. (2019). Phylogeography and niche modelling: Reciprocal enlightenment. *Mammalia*, *84*(1), 10–25. https://doi.org/10.1515/mammalia-2018-0191

Peters, J. L. (1934). *Check-list of birds of the world*. Cambridge, MA: Cambridge.

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double Digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE*, *7*(5), e37135. https://doi.org/10.1371/journal.pone.0037135

Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, *190*(3), 231–259. https://doi.org/10.1016/j.ecolmodel.2005.03.026

Price, T. (2007). *Speciation in Birds*. Greenwood Village, CO: Roberts & Company Publisher.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945–959.

R Core Team (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology*, *67*(5), 901–904. https://doi.org/10.1093/sysbio/syy032

Rochette, N. C., Rivera-Colón, A. G., & Catchen, J. M. (2019). Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. *Molecular Ecology*, *28*(21), 4737–4754. https://doi.org/10.1111/mec.15253

Roeder, A. D., Ritchie, P. A., & Lambert, D. M. (2002). New DNA markers for penguins. *Conservation Genetics*, *3*, 341–344.

Rosenberg, N. A. (2004). distruct: A program for the graphical display of population structure. *Molecular Ecology Notes*, *4*(1), 137–138. https://doi.org/10.1046/j.1471-8286.2003.00566.x

Stonehouse, B. (1970). Geographic variation in gentoo penguins *Pygoscelis papua*. *IBIS*, *112*(1), 52–57. https://doi.org/10.1111/j.1474-919X.1970.tb00075.x

Suchard, M. A., Lemey, P., Baele, G., Ayres, D. L., Drummond, A. J., & Rambaut, A. (2018). Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evolution*, *4*(1), vey016. https://doi.org/10.1093/ve/vey016

Swart, N. C., Gille, S. T., Fyfe, J. C., & Gillett, N. P. (2018). Recent Southern Ocean warming and freshening driven by greenhouse gas emissions and ozone depletion. *Nature Geoscience*, *11*(11), 836–841. https://doi.org/10.1038/s41561-018-0226-1

Thomas, C. D., Cameron, A., Green, R. E., Bakkenes, M., Beaumont, L. J., Collingham, Y. C., … Williams, S. E. (2004). Extinction risk from climate change. *Nature*, *427*(6970), 145–148. https://doi.org/10.1038/nature02121

Trivelpiece, W. Z., Hinke, J. T., Miller, A. K., Reiss, C. S., Trivelpiece, S. G., & Watters, G. M. (2011). Variability in krill biomass links harvesting and climate warming to penguin population changes in Antarctica. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, 7625–7628. https://doi.org/10.1073/pnas.1016560108

Valenzuela-Guerra, P., Morales-Moraga, D., González-Acuña, D., & Vianna, J. A. (2013). Geographic morphological variation of Gentoo penguin (*Pygoscelis papua*) and sex identification: Using morphometric characters and molecular markers. *Polar Biology*, *36*, 1723–1734. https://doi.org/10.1007/s00300-013-1389-2

Vianna, J. A., Noll, D., Dantas, G. P. M., Petry, M. V., Barbosa, A., González-Acuña, D., … Poulin, E. (2017). Marked phylogeographic structure of Gentoo penguin reveals an ongoing diversification process along the Southern Ocean. *Molecular Phylogenetics and Evolution*, *107*, 486–498. https://doi.org/10.1016/j.ympev.2016.12.003

Walsh, S. A., & Suárez, M. E. (2006). New penguin remains from the Pliocene of Northern Chile. *Historical Biology*, *18*(2), 119–130. https://doi.org/10.1080/08912960600640796

Wang, I. J. (2013). Examining the full effects of landscape heterogeneity on spatial genetic variation: A multiple matrix regression approach for quantifying geographic and ecological isolation. *Evolution*, *67*, 3403–3411. https://doi.org/10.1111/evo.12134

Warren, D. L., Glor, R. E., & Turelli, M. (2008). Environmental niche equivalency versus conservatism: Quantitative approaches to niche evolution. *Evolution*, *62*(11), 2868–2883. https://doi.org/10.1111/j.1558-5646.2008.00482.x

Wiens, J. A., Stralberg, D., Jongsomjit, D., Howell, C. A., & Snyder, M. A. (2009). Niches, models, and climate change: Assessing the assumptions and uncertainties. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(Suppl. 2), 19729–19736. https://doi.org/10.1073/pnas.0901639106

Williams, T. D. (1995). *The penguins*. Oxford, UK: Oxford University Press.

Woehler, E. J. (1993). *The distribution and abundance of Antarctic and Subantarctic penguins*. Retrieved from Cambridge, U.K.

Younger, J. L., Clucas, G. V., Kao, D., Rogers, A. D., Gharbi, K., Hart, T., & Miller, K. J. (2017). The challenges of detecting subtle population structure and its importance for the conservation of emperor penguins. *Molecular Ecology*, *26*(15), 3883–3897. https://doi.org/10.1111/mec.14172

BIOSKETCHES

Luis R. Pertierra is a macroecologist with a strong interest in Antarctic biogeography and conservation. He has focused his career in disentangling the various ecological impacts of global environmental change and their effects on Antarctic biodiversity. In particular, he studies the processes of biological invasions and/or range shift redistribution among plants, invertebrates and vertebrate species. **Nicolás I. Segovia** is a marine biologists and molecular ecologist broadly interested in biogeography, phylogeography, population genomics and seascape genomics in marine organisms. This study is part of his collaborations as a postdoctoral researcher at the Instituto de Ecología y Biodiversidad (IEB) and Genomic Antarctic Biodiversity project (www.antarcticgenomics.cl). **Daly Noll** has worked on phylogeography and population genetics of marine vertebrates. She is a PhD student in Evolutionary Biology and studies phylogenomics, adaptation and conservation genomics of gentoo penguin. Her study is part of the Genomic Antarctic Biodiversity project (www.antarcticgenomics.cl) and the Molecular Biodiversity Laboratory (www.biodversidadmolecular.cl).

Author contributions: L.R.P., P.A.M., P.P. and P.A. performed the ecological niche modelling; N.S., K.B., C.Y.W., G.P.M.D., D.N., R.C.K.B., E.P. and J.A.V. performed the genetic data analysis. A.B., A.R.R., P.P., P.T., A.P., F.B., C.L.B. and D.G.A. undertook field work, provided samples and contributed to the manuscript. L.R.P, N.S. and J.A.V. contributed to every analytical step regarding the interpretation of results and in preparing the manuscript. R.C.K.B. helped design the study and advised on analyses to be performed. All authors discussed the results and contributed to the final manuscript.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Pertierra LR, Segovia NI, Noll D, et al. Cryptic speciation in gentoo penguins is driven by geographic isolation and regional marine conditions: Unforeseen vulnerabilities to global change. *Divers Distrib*. 2020;00:1–18. https://doi.org/10.1111/ddi.13072