PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
FACULTAD DE AGRONOMÍA E INGENIERÍA FORESTAL
DIRECCIÓN DE INVESTIGACIÓN Y POSTGRADO
MAGISTER EN RECURSOS NATURALES

Is the environment shaping the genetic structure of the Humboldt penguin population?

Thesis presented as a requisite to obtain the degree of
Master in Natural Resources

by:

Valentina Muñoz Farías

Thesis committee:
Advisor: Juliana A. Vianna
Co-advisor: Nicolás I. Segovia
Examiners: Elie Poulin, Sylvain Faugeron

September 2019
Santiago, Chile

Acknowledgements

I would like to thank FONDECYT 1150517, the Laboratorio de Genética Molecular de la Facultad de Agronomía e Ingeniería Forestal de la Pontificia Universidad Católica de Chile where the laboratory step of this study was done, and the Museum of Vertebrate Zoology University of California Berkeley where the samples were sequenced.

Many thanks to CONAF and Subpesca institutions who were fundamental to this project. I would also like to thank Nicole Sallaberry-Pincheira, Daly Noll, Cynthia Wang, Natalia Kandalaft, Ke Bi, María Eugenia Lopez and Dana Lin for their support and help in field sampling, laboratory data processing and the bioinformatics analyses.

This study was done with the Subpesca (110), CONAF and DGFFS-minag permits.

I would also like to thank the reviewers: Sylvain Faugeron and Elie Poulin, and my co-advisor and mentor Nicolás Segovia, who had to deal with me and helped me with the data analysis, even at odd hours, weekends and holidays.

Personally, I would like to thank everyone who had to bear with me during this process: my family for their support and understanding. Special thanks to Paula Farias for the funding and backing and Felipe Farias for holding and taking care of me. Also to Braulio Muñoz, Juan Pablo Larenas, María Elena Farías, Alejandra Cáceres, Javiera Larenas, José Pablo Larenas and Diego Muñoz, I truly love you all. I would also like to thank the best lab partners, with whom I shared suffering, pizzas, and joy: Isidora Mura, Mariola Tobar, and Felipe Córdova. To the family I chose: Sebastian Kraft, Rodrigo Muñoz, Sebastian Diaz, Karina Robles, and Daniela Baeza I love you all very much and I am deeply grateful for all the support, love, laughs, confidence, consolation and emotional containment. Finally, I would like to thank Juliana Vianna for being an encourager and an empowered inspiration; thanks for the confidence and for pushing me to be a better scientist who cares about conservation.

I dedicate this work to my niece Agustina,

so she will growth with hope in the future.

Index

Is the environment shaping the genetic structure of the Humboldt penguin population?

Valentina Muñoz

Laboratorio de Biodiversidad Molecular, Facultad de Agronomía e Ingeniería Forestal, Pontificia Universidad Católica de Chile, Santiago, Chile.

Abstract

Specialization to highly productive systems (as the Humboldt Current) might lead to genetic structure due to favorable local conditions, or it can promote movements, during water warming episodes, that lead to dispersion and gene flow. Although previous studies have already stated the genetic structure of the Humboldt penguin (*Spheniscus humboldti*), the role of the environment over this pattern has never been tested before. We used genome-wide neutral SNPs and redundancy analysis (RDA) to account for the genetic structure of the Humboldt penguin and its association with the geographical and environmental setting of its distribution range. We found a slight although significant genetic structure, with three genetic groups: first, the Peruvian colony; the main Chilean colony; and a cluster comprising the remaining Chilean colonies. The statistical significance of the latitudinal vector on the model, might be an indicator of isolation by distance. Furthermore, chlorophyll-a, as an indicator of primary productivity, was also significant, suggesting that the genetic structure might be product of preference to local environmental conditions. We highlight the importance of the geographic and environmental configuration of the Humboldt Current System over the population genetic structure of *S. humboldti*. Therefore, in the scenario of accelerated climate change, the genetic structure, which is determined by local environmental conditions and high distance in a latitudinal gradient, must be taken into account when considering management and conservation strategies for this vulnerable species.

Key words: Humboldt Current, RDA model, isolation by distance, specialization to cold waters.

Introduction

One of the most important high productivity environments of the world is the Humboldt Current System (HCS), which flows towards north at the Pacific east coast from southern Chile (45°S) up to Ecuador (near 0º S) (Thiel *et al.*, 2007; Silva *et al.*, 2009). This system presents important latitudinal gradients of environmental conditions and oceanographic factors (Rojas de Mendiola, 1981; Pennington *et al.*, 2006; Escribano *et al.*, 2007; Thiel *et al.*, 2007). It is a nutrient-rich complex of water masses of sub-Antarctic and equatorial origin (Rojas de Mendiola, 1981; Daneri *et al.*, 2000; Thiel *et al.*, 2007; Silva *et al.*, 2009). Consequently, it supports plenty of marine life that serves as a food resource for top predators, including marine birds that inhabits and breed in the area (Daneri *et al.*, 2000). The HCS is an area of conservation concern because of its biodiversity hotspots (Friedlander *et al.*, 2016) and high endemism (Schlatter and Simeone, 1999). Also, as its productivity depends on the upwelling of cold, nutrient-rich waters, the HCS is susceptible climate change, consequent water warming and large-scale climatic fluctuations. During episodes termed El Niño events (EN), warm water masses from the Pacific west coast move towards the east coast elevating the mean temperature of the HCS, reducing its primary production (Wieters *et al.*, 2003; Pennington *et al.*, 2006). Therefore, this exerts a bottom up effect on the ecology, survival and dispersal of the components of the food chain (Stenseth *et al.*, 2002; Thiel *et al.*, 2007; Saba *et al.*, 2008; De Oliveira *et al.*, 2012).

The susceptibility of the HCS can lead seabird populations to migrate to cold nutrient-rich waters during water warming episodes, accordingly to the *upwelling hypothesis*. This hypothesis has been proposed to describe why seabirds that are specialized to cold waters lose their philopatric tendencies, increases gene flow and reach panmixia over large spatial scales (Menge and Menge, 2013). Some HCS species have shown this pattern of high specialization to cold waters leading to weak genetic structure: the Galápagos penguin (*Spheniscus mendiculus*), the Peruvian (*Sula variegata*) and the blue-footed boobies (*S. nebouxii*), and the Peruvian pelican (*Pelecanus thagus*) (Nims *et al.*, 2008; Taylor, Maclagan, *et al.*, 2011; Taylor, Zavalaga, *et al.*, 2011; Jeyasingham *et al.*, 2013). However, and contrary to the upwelling hypothesis, the Peruvian diving-petrel (*Pelecanoides garnotii*) was recently reported as extremely philopatric and, thus, with a marked genetic structure (Cristofari *et al.*, 2019).

Understanding the role of the environment in shaping the genetic structure of populations associated to highly productive systems is particularly challenging. This is due to the

contrasting patterns this specialization can bring: on one side, it might lead to philopatry and genetic structure due to favorable local conditions. Or, on the other hand, it can promote movements during water warming and consequent low productivity episodes, that lead to dispersion and gene flow. Penguins are commonly associated to these systems of cold-to temperate waters with high productivity that concentrate high amounts of biomass. Thus, they might prefer to feed in nearby areas since the cost of traveling long distances to forage could be high for flightless seabirds (Schmidt-nielsen, 1972). This can result on to sedentarism, philopatry and, consequently, the appearance of genetic structure in a population. Secondly and contrary to philopatry and genetic structure, individuals might move away of the natal colony due to increased dispersal for foraging during periods of warm waters and low productivity (the upwelling hypothesis). In addition, natal dispersal might also contribute to gene flow, and thus, to the homogenization of a population (Greenwood, 1980).

The contrasting pattern between philopatry and gene flow has been previously reported in some penguin species. For example, penguins of the genus *Pygoscelis* have shown to be philopatric at different scales (Ainley *et al.*, 1995). The Gentoo penguin (*P. papua*) is the only species in this genus that has shown to be strictly philopatric, exhibiting a high degree of genetic differentiation due to an oceanographic barrier impeding dispersion (Vianna *et al.*, 2017). In contrast, other studies have stated that the Chinstrap penguin (*P. antarcticus*) and the Adelie penguin (*P. adeliae*) exhibit a very subtle sign of population genetic structure (Roeder *et al.*, 2001; Mura-Jornet *et al.*, 2018). Unlike in *P. papua*, this probably occurs due to facilitated dispersion in the absence of barriers restricting gene flow. Furthermore, The Emperor penguin (*Aptenodytes forsteri*) was also reported to have a strong -although not strictly- philopatric behavior (Larue *et al.*, 2015), showing to be panmictic throughout the whole Antarctic shelf (Cristofari *et al.*, 2016).

The Humboldt penguin (*S. humboldti*) is endemic to the HCS inhabiting most of the Peruvian and Chilean coasts (Schlosser *et al.*, 2009; Garcia Borboroglu and Boersma, 2013). It has been documented that the latitudinal gradient of environmental conditions of this system may affect the distribution and behavior of the Humboldt penguins: during EN, intensified rainfall and flooding of nests were followed by major reproductive failures of colonies at central Chile (Meza *et al.*, 1998; Simeone *et al.*, 2002). Moreover, climatic fluctuations can also have indirect effects on the Humboldt penguin as sea-surface temperature anomalies (SSTA) are associated to a reduction in prey numbers, consequently increasing foraging

efforts (Culik and Luna-Jorquera, 1997a; Culik *et al.*, 2000), and a likely related nest abandonment (Simeone *et al.*, 2002). In addition, their main prey items, the South American pilchard (*Sardinops sagax*) and the Peruvian anchoveta (*Engraulis ringens*) aggregate and migrate latitudinally depending on hydrological conditions (e.g. Gutiérrez *et al.*, 2007) which would promote foraging migrations and, in extreme scenarios, death of adults by starvation during food shortage episodes (Jaksic, 2004). This is in agreement with the work of Hays (1986), who reported a dramatic decline of the Humboldt penguin population in Peru during the strongest EN registered, in the years 1982 to 1983. Later, Culik *et al*. (2000) described long migration trips of this species as the productivity decreased during one of the strongest EN. Finally, displacement patterns for this species associated with climatic fluctuations are also debated by Vianna *et al.* (2014). Therefore, the environment appears to be playing an undiscovered role in Humboldt penguin populations: it might be acting as a homogenizing factor, by promoting dispersion and gene flow as a result of the specialization to cold waters, or, in contrast, giving specific local conditions that may favor philopatry and genetic structure.

Previous studies have described the Humboldt penguin as a sedentary, inshore-feeder and likely a year-round resident species, and thus as highly philopatric, presenting both colony and nest fidelity (Culik and Luna-Jorquera, 1997b; Culik *et al.*, 1998; Teare *et al.*, 1998; Croxall and Davis, 1999; Wallace *et al.*, 1999; Araya *et al.*, 2000; Simeone and Wallace, 2013). In contrast, the species also shows a natal dispersal trend (Simeone and Wallace, 2013). Consequently, there is debate about the genetic structure of the Humboldt penguin populations. The genetic structure of the Humboldt penguin was first explored by Schlosser *et al.* (2009), who detected a slight, albeit significant, signal of genetic structure, likely due to long-term gene flow, apparently affected by a pattern of isolation by distance (IBD). In opposition, Dantas *et al.*. (2019) found a very marked genetic structure for the Humboldt penguin population without an IBD pattern. Both studies used microsatellite data to test for the genetic structure of this penguin. However, the role of the environment over the genetic structure of the Humboldt penguin has never been tested before.

Genome wide neutral markers associated to environmental data (sea surface temperature and chlorophyll-a) might give us a robust sight of how a population is potentially adapted to their habitat's local conditions. As well, the association of neutral single nucleotide polymorphisms (SNPs) to the spatial distribution may provide an insight of a pattern of IBD,

which structures the population by effect of the distance they have to travel to make gene flow effective.

The aim of this study is to assess population genetic structure of the Humboldt penguin associated to the geographical and environmental setting of its distribution range considering colonies in Chile and Peru. Thus, to account for the genetic structure, we evaluated SNPs scattered throughout its genome (ddRAD-seq). In addition, we used statistical models to test for local environmental features shaping its population genetic structure, regarding an indicator of productivity (chlorophyll a), and an indicator of climatic conditions (sea surface temperature), besides the association to the spatial distribution of the colonies regarding vectors for latitude and longitude to test for isolation by distance. The obtained results will support and link up what has previously been done for this species with a new whole-genome approach, reinforcing the knowledge of its population status and thus, helping to take the best decisions for its management and conservation.

Hypothesis

Environmental conditions are playing a role in shaping the population structure of the Humboldt penguin.

Objective

General objective

Assess the population structure of the Humboldt penguin (*Spheniscus humboldti*), related to the geographical and environmental setting of its distribution range, considering five colonies in Chile and one in Peru.

Specific objectives

- To determine the population structure from an individual scope, in order to assign individuals to clusters.
- To determine the population structure from a clustering scope to resolve its spatial genetic structure.
- To test for environmental factors shaping the genetic structure of the population.

Methods

Library preparation and ddRAD-seq

Genomic DNA was extracted from blood samples from 57 Humboldt penguins obtained from the brachial or foot vein, at the localities shown in Figure 1. The samples were preserved in 96% ethanol and the DNA was isolated using the salt protocol described in Vianna *et al.*, (2017). Total genomic DNA was quantified using Qubit dsDNA Broad Range Assay Kit (Thermo Fisher Scientific). Double-digest libraries were prepared using 500 ng of genomic DNA following the protocol described by Peterson *et al.*, (2012). Digestion was performed at 37º C for 3 hours with 0.5 U of EcoRI and SphI-HF (New England Biolabs) simultaneously, followed by a ligation step, by which each individual was assigned to one of 24 unique barcodes. Digested DNA was quantified using Qubit dsDNA High Sensitive Assay Kit. Pools of 24 individuals were grouped together after adapter ligation (total of 20 pools). The size selection of fragment was performed by Pippin Prep (Sage Science), where fragments of 300-400 bp were isolated. Each pool was amplified using 12 cycles in 25 µl indexing PCR ligation, at a final concentration of 0.8 uM of P5 and P7 Illumina adapters, using 0.4 U of Phusion HF DNA polymerase (Thermo Scientific) and 20 ng of template DNA. DNA libraries were quantified using the DNA 1000 Kit in a 2100 Bioanalyzer (Agilent Technologies). Pools were combined in equimolar concentration to form a single genomic library to perform the multiplex sequencing in Illumina HiSeq 4000 equipment.

Construction of species-specific reference genome

In order to improve the efficiency and accuracy of short read mapping, and to reduce alignment bias to a divergent genome, we reconstructed a species-specific reference genome for the Humboldt penguin. We prepared libraries for the genome sequencing of Humboldt penguins using an Illumina TruSeq Nano kit following the manufacturer's instructions. In brief, 100 ng of genomic DNA was fragmented to 350 bp segments using an ultrasonicator. After cleaning with beads, fragmented DNA was treated with end repair mix and then with A-tailing to add an adenine to the 3'-end, to which indexing adapters were ligated. Ligated DNA fragments were amplified and purified with beads prior to quantification using a Qubit fluorometer. Library size measured with an Agilent TapeStation (Agilent Technologies Inc). The library was sequenced to ~40x coverage with 150 paired reads using an Illumina HiSeq X Ten platform at MedGenome (USA). To process raw reads, exact duplicates were removed by using Super Deduper (https://github.com/dstreett/Super-

Deduper). The reads were then filtered using Cutadapt (Martin, 2011) and Trimmomatic (Bolger *et al.*, 2014) to trim adapter contaminations and low-quality reads. Overlapping paired-end (PE) reads were merged using FLASH (Magoč and Salzberg, 2011). We then aligned the resulting cleaned reads to an Emperor Penguin (*Aptenodytes forsteri*) draft genome (http://gigadb.org/dataset/100005) using LAST (http://last.cbrc.jp/). The resulting alignment was converted to sorted BAM format using SAMtools (Li et al., 2009). We then used *samtools mpileup, bcftools, vcfutils.pl vcf2fq* and *seqtk* (https://github.com/lh3/seqtk) to convert alignments into FASTA format reference genome. The species-specific reference fasta sequence was evaluated for completeness by comparing it against the Emperor Penguin draft genome.

RAD data processing

We used a custom PERL pipeline evoking various external programs for processing ddRAD-seq data. The pipelines are available in https://github.com/CGRL-QB3-UCBerkeley/RAD. Raw fastq reads were first de-multiplexed based on the sequences of internal barcodes with a tolerance of one mismatch. De-multiplexed reads were removed if the expected cutting site (also one mismatch allowed) is not found at the beginning of the sequences. The reads were then filtered using Cutadapt and Trimmomatic to trim adapter contaminations and low-quality reads. The resulting cleaned reads of each individual were aligned to the reference genome, using BWA-MEM (Li, 2013). For the resulting alignment, we then used Picard (http://picard.sourceforge.net) to add read groups and GATK (McKenna *et al.*, 2010) to perform realignment around indels. After realignment, we filtered raw variants using SNPcleaner (github.com/tplinderoth/ngsQC/tree/master/snpCleaner). We only considered sites in which at least 70% of the individuals in our dataset had coverage of at least 3x. We also filtered sites near 5bp around an indel. The minimum RMS mapping quality for a variable site to keep is 10. Min p-value for base quality bias is 1e-100. We also removed sites showing excessive heterozygosity using the one-tailed Fisher's exact test for HWE ($p<0.0001$). For the following variant calling and population genetic analysis, we only focused on sites that passed the above filters.

Variant calling

Demultiplexed fastq files were aligned against the reference genome consensus sequence (in fasta format) of the Humboldt penguin using Bowtie2 v. 2.2.3 (Langmead and Salzberg, 2012). The obtained SAM files were sorted and converted to BAM files using SAMTools

0.1.19 (Li *et al.*, 2009). These files were used as input data for the STACKS pipeline using the *ref_map.pl* program. At this step we used no filters, in order to obtain every SNP present in the dataset. The output file in vcf format was filtered and cleaned using Tassel 5 (Bradbury *et al.*, 2007) where we only kept sites with a minimum allele frequency (maf) of 0.05, a maximum heterozygosity of 0.8, and we only considered in the analysis sites that were present in at least 70% of the individuals.

Selecting neutral SNPs

We performed three independent runs in Bayescan (Foll and Gaggiotti, 2008) to eliminate loci under selection. The three datasets were overlapped in a Venn diagram to select the SNPs under selection that were consistent in the three independent runs. This final SNPs dataset was saved as vcf and plink format, which were transformed, as necessary, to other format files using PGDSpider v. 2.1.1.5 (Lischer and Excoffier, 2012).

Genetic diversity and population genetic structure

We used Arlequin v 3.5, PLINK v. 1.90 (Purcell *et al.*, 2007), and custom R scripts to calculate genetic diversity indices, which are summarized in Table 1.

In order to account for genetic groups presents in the population, we used the *find.clusters* function of the Adegenet (Brian *et al.*, 2018) package for R. With this same package, we performed a principal component analysis (PCA), and a discriminant analysis of principal components (DAPC). We also run StrAuto v. 1.0 (Chhatre and Emerson, 2017) which is based on Python (v. 2.7.14) and Structure v. 2.3.4 (Pritchard *et al.*, 2000), and uses SructureHarvester web v. 0.6.94 (Earl and vonHoldt, 2012) to collate the obtained results. We used CLUMPP v. 1.1.2 (Jakobsson and Rosenberg, 2007) and *distruct* v. 1.1 (Rosenberg, 2004) to plot these results. We calculated pairwise F_{ST} values using PLINK v. 1.90 (Purcell *et al.*, 2007).

To account for migration patterns and geographic barriers that may be restraining gene flow between the sampled colonies, we used EEMS software to estimate effective migration surfaces (Petkova *et al.*, 2015), which visualizes spatial population structure by calculating genetic differences (as F_{ST}) values and contrasts them with geographic data.

We performed an isolation by distance (IBD) analysis (Wright, 1943) by a Mantel Test (Mantel, 1967) using Arlequin. To complement the IBD analysis, and to test for environmental factors shaping the genetic structure found in the 963 neutral SNPs present

in the six sampled colonies, we performed redundancy analysis (RDA, summarized in Table 3) for which we used the packages *adespatial*, *SoDA* and *Vegan* in R. For this, we first formulated dbMEM (distance-based Moran's eigenvector maps) models, in order to account for spatial arrangements, as latitude and longitude (dbMEMs), in the genetic structure of the populations. The dbMEMs are independent variables (vectors) representing the spatial configuration (latitude and longitude) in a cartesian plane (X-Y) associated to a distance matrix. We also used oceanographic traits, as sea surface temperature (SST) and chlorophyll-a (Chl-a) as local environmental conditions that might be shaping the genetic structure of the population. We took SST and Chl-a satellite data from 2005 to 2015 from the MODIS-aqua online database (https://giovanni.gsfc.nasa.gov/giovanni/) and obtained the historical mean value for each location using custom Python scripts.

We first performed two partial models (partial-environmental and partial-spatial) to explain the genotypic variation by effects of: i) the environment (SST and Chl-a) while conditioning the geographic configuration (dbMEMs); and ii) the spatial configuration while conditioning the environmental variables. Second, we performed a total model regarding the four environmental features (dbMEM1, dbMEM2, SST and Chl-a). Finally, we used the *ordistep* function over all variables to select the best candidates for explaining the genetic structure. This function, which uses the Akaike's information criterion (AIC), selected dbMEM1 (latitude vector) and Chl-a as the best candidate variables, so we performed a fourth model with this suggestion (partial environmental-spatial model).

Results

Summary statistics

We obtained a total number of 11,000 raw SNPs using the *ref_map.pl* function from the STACKS pipeline. Using Tassel 5, we filtered 8,994 loci, retaining 2,006 SNPs that passed the filtering stage. Finally, using Bayescan we separated and kept a final dataset of 963 high-confidence neutral SNPs for the 57 individuals from the 6 Humboldt penguin colonies sampled (Figure 1, Table 1). For the calculation of the genetic diversity indices, Arlequin only considered loci with less than 5% of missing values, thus, of the 963 loci found for the entire population we only used from a minimum of 792 SNPs (Pan de Azúcar) to a maximum of 917 SNPs (Puñihuil). The allelic richness (A_R) ranged from 1.28 to 1.29, being Isla Choros the locality with the greatest number of alleles (Table 1). The expected heterozygosity ranged from 0.31 at Pan de Azucar to 0.55 at Puñihuil with an average of 0.36. The observed heterozygosity, similarly, ranged from 0.30 to 0.54 with an average of 0.35. No differences between observed and expected heterozygosity were found. The nucleotide diversity (π) was low and similar within all colonies (0.15), being the Peruvian locality the one with the lower value (0.14).

Genetic structure

To account for the genetic structure of the Humboldt penguin population, we used four methods with two different approaches: we analyzed the population using a clustering method (assigning individuals to groups) with PCA (Figure 2a), DAPC (Figure 4b) and STRUCTURE (Figure 3); and frequency-based inferences to account for the spatial structure of the clusters, if present, with F_{ST} values (Figure 4a, Table 2) and EEMS (Figure 2b).

The results obtained here for the Humboldt penguin suggest a subtle signal of population genetic structure. The *find.clusters* function of the *Adegenet* package and the STRUCTURE analysis suggested K=2 as the most likely number of clusters within the population. As well, the first principal component (PC1 at the X axis) of the PCA (Figure 2a) shows the separation of the individuals in two groups. The EEMS analysis (Figure 2b) shows the posterior mean migration rates (m) in $\log_{10}(m)$, with positive migration rates representing gene flow, and negative migration rates representing genetic barriers. The darker the color, the higher the migration signal. With this analysis, we encountered clear genetic segregation of the Peruvian colony Punta San Juan (SJ) due to the lack of migration obtained for this locality

(orange halos). Further, the Chilean colonies appeared to have a positive migration rate among all localities (although low, due to the light-blue halo). This result agrees with the separation shown by the second principal component (PC2) of the PCA (Figure 2a).

Also, according to the AMOVA analysis between the groups conformed by Punta San Juan and the other Chilean colonies, the proportion of variation attributable to within-population differences was high (98.37%), whereas only 1.33% and 0.3% occurred among groups and among groups within populations, respectively. This result indicates that major diversity is encountered inside each colony and, thus, the genetic structure signaling among Chilean and Peruvian groups is weak.

In addition, the results of *find.clusters*, STRUCTURE and PCA analyses suggested that the two genetic groups were composed by random individuals and are equally distributed through the entire distribution range, without a spatial genetic structure consistent with the geographic configuration of the sampled colonies (PC1, Figure 2a; K=2, Figure 3). Thus, the suggestion of two genetic groups as the most likely number of clusters within the population appears to be either a sequencing or variant calling bias.

The consistence between the genetic structure and the geography is recovered when K=3 is plotted (Figure 3) and forced in the *find.clusters* function for the DAPC construction (Figure 4b), in which Punta San Juan is separated from the Chilean colonies. This three groups were also showed in the F_{ST} analysis (Figure 4a and Table 2). The groups would be composed by the northernmost colony at Peru, Punta San Juan (SJ); Chañaral (CH), the colony with the largest aggregation of penguins (Wallace & Araya, 2015); and a cluster composed by Pan de Azúcar (PA), Isla Choros (IC), Cachagua (CA) and Puñihuil (PU). Although the DAPC shows this separation, the genetic differences among groups were still subtle since the sum of the two first principal components (PC1 and PC2) does not reach to explain 20% of the variance. Similarly, the F_{ST} values reported here are very low, although significant, ranging from 0.007 to 0.023 (Table 2). Nevertheless, according to F_{ST} values, there are three statistically significant groups -the same obtained with the DAPC- since Punta San Juan is statistically different from all other Chilean locations, as well as Chañaral (Figure 4).

Redundancy analysis: environmental association to genetic structure

For the analysis of the models (Table 3), we first discarded the two partial models for their significance: p = 0.35 for the partial-environmental model; p= 0.037 for the partial-spatial

model. Thus, we kept the partial environmental-spatial model chosen by AIC (Figure 5), because it is statistically significant (p = 0.001). We can evince a slight signal of population genetic structure by looking at the pink dot cluster in the middle of the plot (Figure 5). In this partial geo-oceanographic model, the two axes rise to account for 57.3% and 42.7% of the genetic variation, respectively, and the ANOVA suggested both variables (Chl-a and dbMEM1) as statistically significant predictors of the variation (see Table 3).

The mantel test performed to test for isolation by distance (IBD) was not significant (p = 0.099, r = 0.53, Figure 6). Although, the statistical significance of the latitudinal vector (dbMEM1) in both the total model and the partial environmental-spatial model (see Table 3), might be an indicator of IBD. As well, local conditions of productivity (chlorophyll a) appears to be playing a role in limiting gene flow among the Humboldt penguin population, as it is statistically significant within the partial environmental-spatial model (Table 3). Furthermore, Chañaral seems to be separated to the other Chilean colonies by means of chlorophyll-a differences, while Punta San Juan differentiates from the Chilean cluster because of the latitudinal dbMEM1 (Figure 5). Finally, although the mantel test was not statistically significant ($p > 0.05$), there seems to be a tendency of greater genetic similarities at shorter distances (Figure 6).

Discussion

Our study reports the population genetic structure of the Humboldt penguin breeding colonies using a genome-wide approach to detect neutral SNPs, and its association with environmental features tested using statistical models. We revealed a slight although significant genetic structure among the six studied colonies which is largely explained by local environmental conditions (as chlorophyll-a) and the geographic configuration (latitudinal distances) of its distribution range.

Summary statistics

The allelic richness result seems low for biallelic markers such as SNPs, although, there is to consider that A_R was standardized to a sample size of n=2 (Puñihuil). Also, the study of Ryynänen *et al.*, (2007) reported 1.26 as the minimum value of allelic richness for a salmon species.

The genetic diversity found here is higher than the reported by Clucas *et al.*, (2016) for the King penguin for both expected heterozygosity (H_e) and nucleotide diversity (Cluca's H_e~0.11 and π~0.12). As well, H_e reported here was higher than the reported by Frugone *et al.*, (2019) for Macaroni (*Eudyptes chrysolophus*) and Royal (*E. schlegeli*) penguins (Frugone's H_e~0.26). Although, the nucleotide diversity was lower in our value (Frugone's π~0.27). There is to notice that both studies that we are comparing with our study used a different number of SNPs and sampling sizes.

Genetic structure

Punta San Juan is shown as a different genetic group in all the performed analyses. As the northernmost location and most isolated one, there seems more likely that all the analyses show this colony as genetically different to the other Chilean colonies. The studies of Schlosser *et al.*, (2009) and Dantas *et al.*, (2019) also found significant genetic structure of Punta San Juan with respect to Chilean colonies when using microsatellite data. Furthermore, the study of Sallaberry-Pincheira *et al.*, (2016) also reported a high genetic structure of this locality compared to other Chilean colonies for both MHC I and II (an adaptive marker with multiple genes involved in the immune system). As a marker used to evaluate selection driven by diseases, the genetic difference found for MHC I and II at this locality were attributed to the high immunological pressure exerted by a higher diversity of pathogens associated to lower latitudes (Sallaberry-Pincheira *et al.*, 2016). Also, Punta San

Juan is the largest aggregation of the Humboldt penguin at Peru, and there is evidence of strong colony fidelity at this location (Araya *et al.*, 2000; Culik, Hennicke and Martin, 2000).. Thus, this study confirms what was already stated for the genetic differentiation of Punta San Juan with respect to Chilean colonies.

The genetic divergence of Chañaral seems unlikely -yet observed- due to the closeness of the colonies and the already reported gene flow and high mobility rates of this species (Simeone & Wallace 2013; Schlosser, et al., 2009). The F_{ST} analysis suggests Chañaral as genetically different to the other Chilean colonies. Although the values are very low (≤ 0.01), they are statistically significant. As well, the RDA model chosen here separates Chañaral considering differences regarding Chl-a. Seabirds in general tend to be philopatric and to aggregate in large groups due to the benefits of coloniality, which includes habitat and nesting stability, predator avoidance, food availability and enhanced reproductive performance (Danchin and Wagner, 1997; Dubois, Cézilly and Pagel, 1998). Chañaral is the main colony of the Humboldt penguin at Chile (Mattern *et al.*, 2006) bearing around 80% of the total Chilean population (Wallace and Araya, 2015). Therefore, as the largest and premier colony, Chañaral might be an advantageous location to breed, promoting philopatry and ultimately the genetic structure found. Thus, the genetic structure found for this colony might be attributed to philopatry, although a sex-biased analysis should be performed in order to clear up this suggestion. Nevertheless, sedentarism was previously reported for the Humboldt penguin at this location and other Chilean colonies (Schlosser *et al.*, 2009; Simeone and Wallace, 2013) and Dantas *et al.*, (2019) reported a philopatry rate of 98% at Chañaral, conditions that may be explaining the genetic divergence encountered at this location.

Dantas *et al.*, (2019) also revealed a significant genetic separation of Pan de Azúcar with a strong philopatric signature, which was not found in our study. Foraging efforts for the Humboldt penguin are reported to be in nearby areas, within a radius of 20 to 35 km around Pan de Azúcar island (Culik and Luna-Jorquera, 1997b; Culik *et al.*, 1998), reinforcing the sedentary nature of the Humboldt penguin at this location. The findings of Dantas *et al.*, (2019) are also contrasting the GeneClass findings of our work (Table 4), which revealed that Pan de Azúcar would be a source of migrants mainly for Cachagua and Chañaral. The small sampling sizes used here might be underestimating the genetic structure found for Pan de Azúcar. There is also to consider the different methodological approaches used to detect the genetic structure, which might also be contributing to the contradiction found.

Also, as Slatkin (1985) stated, 9 migrants per generation are sufficient to mask genetic differentiation. Although, further analyses and a comparative study for this site are needed in order to clarify the genetic difference of Pan de Azúcar.

Geographic and environmental factors shaping the genetic structure

The geographic configuration of the Humboldt penguin colonies evaluated here seems to be playing a role in determining its genetic structure. Furthermore, geographic distance and local conditions of productivity appear to be shaping the slight, although significant, genetic structure of the Humboldt penguin population. There are several studies questioning the interpretations of Mantel test and suggesting the use of redundancy analysis (RDA) instead, which also allows to include environmental data in order to explain genetic structure (Meirmans, 2012, 2015; Diniz-Filho *et al.*, 2013; Guillot and Rousset, 2013; Legendre, Fortin and Borcard, 2015; Szulkin *et al.*, 2016). Thus, as dbMEM1 is significantly explaining the genetic structure found, an IBD pattern would be responsible for restricting the gene flow between Peru and Chile. This agrees with Schlosser *et al.*, (2009) and Sallaberry-Pincheira *et al.*, (2016) who also found an IBD pattern for *S. humboldti*.

As well, productivity (as chlorophyll-a), would be also playing a role over the genetic structure found here, especially at differentiating Chañaral from the other Chilean colonies. The study of Hennicke and Culik, (2005), encountered differences in the foraging times of penguins at Pan de Azúcar and Puñihuil, due to higher productivity and consequent greater abundance and prey availability in the southern location. This behavior might be explaining the separation of Puñihuil and Chañaral by means of Chl-a at the partial environmental-spatial model (albeit a more thorough analysis could clarify the chlorophyll association to Chañaral). Thus, a strict acclimation to local environmental conditions might be strengthening the genetic structure of the Humboldt penguin. Furthermore, a local adaptation pattern might be responsible for this result, which was already reported for this species (Sallaberry-Pincheira *et al.*, 2016). Although, further analysis using adaptive markers, such us SNPs under selection, would give us a sight of evolutive adaptation to local environmental conditions.

There is a recent need to disentangle the importance of geography and environment over the genetic structure on vertebrates, invertebrates and even in plant species (Sexton, Hangartner and Hoffmann, 2014; Wang and Bradburd, 2014; Nadeau *et al.*, 2016). In this case, as well as in the study of Nadeau *et al.*, (2016), Benestan *et al.*, (2016) and Frugone

et al., (2019) both IBD and isolation by environment (IBE) act together to structure a population. IBE is the pattern where genetic differentiation increases independent of IBD (Wang and Bradburd, 2014). Although, as the HCS is a latitudinal gradient of heterogeneous environmental factors, with nutrient-rich zones concentrated in upwelling areas, some oceanographic traits (IBE) are related to latitude (IBD). Here, chlorophyll-a and latitude are independently shaping the genetic structure of the Humboldt penguin population, as they are not significantly correlated (Figure 7). Thus, IBD and IBE can be disentangled here, but further analyses are needed in order to clear up the independent roles and hierarchy of geography and environment.

Finally, the results of this study do not support the upwelling hypothesis for explaining the genetic structure of the Humboldt penguin, as we did not find a panmictic population. By the contrary, this results support the hypothesis of this study, as the slight but significant genetic structure found here might be determined by high geographic distances limiting gene flow, and, independently, by the effects of local productivity conditions.

Conclusion

This is the first study to link a genome-wide technique with statistical approaches to account for the geographic and environmental relevance in the genetic structure of the Humboldt penguin population. Here, we confirm what has already been found by other authors regarding the genetic structure of Peru against Chile, in addition to the importance of the geographic and environmental configuration of the HCS over this structure. Local environmental conditions and isolation by distance found here corroborate that the genetic structure reported for the Humboldt penguin population is determined by environmental local conditions in a latitudinal gradient.

The Humboldt penguin is a vulnerable species and their major threats are guano harvesting, fisheries, and human disturbances according to the IUCN red list. Additionally, in the climate change scenario, adaptation to accelerated environmental changes might become challenging. Thus, these results should be considered when making decisions regarding the management and conservation strategies of this vulnerable species.

Finally, a higher number of SNPs and expanding the sampling efforts in order to cover the whole species distribution, would clarify the patterns shaping the genetic structure found here.

References

Ainley, D. G., Nur, N. and Woehler, E. J. (1995) 'Factors Affecting the Distribution and Size of Pygoscelid Penguin Colonies in the Antarctic', *The Auk*, 112(1), pp. 171–182. doi: 10.2307/4088776.

Araya, B. *et al.* (2000) *Population and habitat viability assessment for the Humboldt penguin (Spheniscus humboldti), Final Report*. IUCN/SSC C. Apple Valley, Minnesota.

Benestan, L. *et al.* (2016) 'Seascape genomics provides evidence for thermal adaptation and current-mediated population structure in American lobster (Homarus americanus)', *Molecular ecology*, 25(20), pp. 5073–5092. doi: 10.1111/mec.13811.

Bolger, A. M., Lohse, M. and Usadel, B. (2014) 'Trimmomatic: a flexible trimmer for Illumina sequence data', *Bioinformatics*, 30(15), pp. 2114–2120. doi: 10.1093/bioinformatics/btu170.

Bradbury, P. J. *et al.* (2007) 'TASSEL: Software for association mapping of complex traits in diverse samples', *Bioinformatics*, 23(19), pp. 2633–2635. doi: 10.1093/bioinformatics/btm308.

Brian, A. *et al.* (2018) 'Adegenet'.

Chhatre, V. E. and Emerson, K. J. (2017) 'StrAuto: Automation and parallelization of STRUCTURE analysis', *BMC Bioinformatics*. BMC Bioinformatics, 18(1), pp. 1–5. doi: 10.1186/s12859-017-1593-0.

Clucas, G. V. *et al.* (2016) 'Dispersal in the sub-Antarctic: king penguins show remarkably little population genetic differentiation across their range', *BMC Evolutionary Biology*. BMC Evolutionary Biology, 16(1), pp. 1–14. doi: 10.1186/s12862-016-0784-z.

Cristofari, R. *et al.* (2016) 'Full circumpolar migration ensures evolutionary unity in the Emperor penguin', *Nature Communications*, 7(11842), pp. 1–9. doi: 10.1038/ncomms11842.

Cristofari, R. *et al.* (2019) 'Unexpected population fragmentation in an endangered seabird: the case of the Peruvian diving-petrel', *Scientific Reports*, 9(1), pp. 1–13. doi: 10.1038/s41598-019-38682-9.

Croxall, J. P. and Davis, L. S. (1999) 'Penguins: paradoxes and patterns', *Marine Ornithology*, 27, pp. 1–12.

Culik, B. M. *et al.* (1998) 'Humboldt penguins monitored via VHF telemetry', *Marine Ecology Progress Series*, 162, pp. 279–286. doi: 10.3354/meps162279.

Culik, B. M., Hennicke, J. and Martin, T. (2000) 'Humboldt penguins outmanoeuvring El Niño.', *The Journal of experimental biology*, 203(Pt 15), pp. 2311–22. doi: 10.1038/28303.

Culik, B. M. and Luna-Jorquera, G. (1997a) 'Satellite tracking of Humboldt penguins (Spheniscus humboldti) in northern Chile', *Marine Biology*, 128(4), pp. 547–556. doi: 10.1007/s002270050120.

Culik, B. M. and Luna-Jorquera, G. (1997b) ‘The humboldt penguin Spheniscus humboldti: A migratory bird?’, *Journal fur Ornithologie*, 138(3), pp. 325–330. doi: 10.1007/BF01651558.

Danchin, E. and Wagner, R. H. (1997) ‘the Emergence of New Perspectives’, *Tree*, 12(9), pp. 342–347.

Daneri, G. *et al.* (2000) ‘Primary production and community respiration in the Humboldt Current System off Chile and associated oceanic areas’, *Marine Ecology Progress Series*, 197, pp. 41–49. doi: 10.3354/meps197041.

Dantas, G. P. M. *et al.* (2019) ‘Uncovering population structure in the Humboldt penguin (Spheniscus humboldti) along the Pacific coast at South America’, *PLoS ONE*, 14(5), pp. 1–19. doi: 10.1371/journal.pone.0215293.

Diniz-Filho, J. A. F. *et al.* (2013) ‘Mantel test in population genetics’, *Genetics and Molecular Biology*, 36(4), pp. 475–485. doi: 10.1590/S1415-47572013000400002.

Dubois, F., Cézilly, F. and Pagel, M. (1998) ‘Mate fidelity and coloniality in waterbirds: A comparative analysis’, *Oecologia*, 116(3), pp. 433–440. doi: 10.1007/s004420050607.

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

Escribano, R. *et al.* (2007) ‘Seasonal and inter-annual variation of mesozooplankton in the coastal upwelling zone off central-southern Chile’, *Progress in Oceanography*, 75(3), pp. 470–485. doi: 10.1016/j.pocean.2007.08.027.

Foll, M. and Gaggiotti, O. (2008) ‘A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective’, *Genetics*, 180(2), pp. 977–993. doi: 10.1534/genetics.108.092221.

Friedlander, A. M. *et al.* (2016) ‘Marine biodiversity in Juan Fernández and Desventuradas islands, Chile: Global endemism hotspots’, *PLoS ONE*, 11(1). doi: 10.1371/journal.pone.0145059.

Frugone, M. J. *et al.* (2019) ‘More than the eye can see: genomic insights into the drivers of genetic differentiation in Royal/Macaroni penguins across the Southern Ocean’, *Molecular Phylogenetics and Evolution*, p. 47. doi: 10.1016/j.ympev.2019.106563.

Garcia Borboroglu, P. and Boersma, P. D. (eds) (2013) *Penguins: Natural History and Conservation*. Seattle: University of Washington Press.

Greenwood, P. J. (1980) ‘Mating systems , philopatry and dispersal in birds and mammals’, *Animal Behavior*, 28(1960), pp. 1140–1162.

Guillot, G. and Rousset, F. (2013) ‘Dismantling the Mantel tests’, *Methods in Ecology and Evolution*, 4(4), pp. 336–344. doi: 10.1111/2041-210x.12018.

Gutiérrez, M. *et al.* (2007) ‘Anchovy (Engraulis ringens) and sardine (Sardinops sagax) spatial dynamics and aggregation patterns in the Humboldt Current

ecosystem, Peru, from 1983-2003', *Fisheries Oceanography*, 16(2), pp. 155–168. doi: 10.1111/j.1365-2419.2006.00422.x.

Hays, C. (1986) 'Effects of the 1982-1983 El Nino on Humboldt penguin colonies in Peru', *Biological Conservation*, 36(2), pp. 169–180. doi: 10.1016/0006-3207(86)90005-4.

Hennicke, J. C. and Culik, B. M. (2005) 'Foraging performance and reproductive success of Humboldt penguins in relation to prey availability', *Marine Ecology Progress Series*, 296, pp. 173–181. doi: 10.3354/meps296173.

Jakobsson, M. and 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), pp. 1801–1806. doi: 10.1093/bioinformatics/btm233.

Jaksic, F. M. (2004) 'El niño effects on avian ecology: lessons learned from the southeastern pacific', *Ornitologia Neotropical*, 15, pp. 61–72.

Jeyasingham, W. S. *et al.* (2013) 'Specialization to cold-water upwellings may facilitate gene flow in seabirds: New evidence from the Peruvian pelican Pelecanus thagus (Pelecaniformes: Pelecanidae)', *Journal of Avian Biology*, 44(3), pp. 297–304. doi: 10.1111/j.1600-048X.2012.00004.x.

Langmead, B. and Salzberg, S. L. (2012) 'Fast gapped-read alignment with Bowtie 2', *Nature Methods*, 9(4), pp. 357–359. doi: 10.1038/nmeth.1923.

Larue, M. A. *et al.* (2015) 'Emigration in emperor penguins: Implications for interpretation of long-term studies', *Ecography*, 38(2), pp. 114–120. doi: 10.1111/ecog.00990.

Legendre, P., Fortin, M. J. and Borcard, D. (2015) 'Should the Mantel test be used in spatial analysis?', *Methods in Ecology and Evolution*, 6(11), pp. 1239–1247. doi: 10.1111/2041-210X.12425.

Li, H. *et al.* (2009) 'The Sequence Alignment/Map format and SAMtools', *Bioinformatics*, 25(16), pp. 2078–2079. doi: 10.1093/bioinformatics/btp352.

Li, H. (2013) 'Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM', *arXiv preprint*, 00(00), pp. 1–3. doi: arXiv:1303.3997 [q-bio.GN].

Lischer, H. E. L. and Excoffier, L. (2012) 'PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs', *Bioinformatics*, 28(2), pp. 298–299. doi: 10.1093/bioinformatics/btr642.

Magoč, T. and Salzberg, S. L. (2011) 'FLASH: fast length adjustment of short reads to improve genome assemblies', *Bioinformatics*, 27(21), pp. 2957–2963. doi: 10.1093/bioinformatics/btr507.

Mantel, N. (1967) 'Cancer Research', *American Association for Cancer Research*, 27(2 part I), pp. 209–220. doi: 10.1038/214637b0.

Martin, M. (2011) 'Cutadapt removes adapter sequences from high-throughput sequencing reads', *EMBnet.journal*, 17(1), pp. 10–12. doi: doi:https://doi.org/10.14806/ej.17.1.200.

Mattern, T. *et al.* (2006) ‘Humboldt Penguin Census on Isla Chañaral, Chile: Recent Increase or Past Underestimate of Penguin Numbers?’, *Waterbirds*, 27(3), pp. 368–376. doi: 10.1675/1524-4695(2004)027[0368:hpcoic]2.0.co;2.

McKenna, A. *et al.* (2010) ‘The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data’, *Genome research*, 20(9), pp. 1297–1303.

Meirmans, P. G. (2012) ‘The trouble with isolation by distance Why not look’, pp. 2839–2846.

Meirmans, P. G. (2015) ‘Seven common mistakes in population genetics and how to avoid them over confidence need diff analyses diff scales needed may not be able to do patterns errors’, *Molecular Ecology*, 24, pp. 3223–3231.

Menge, B. A. and Menge, D. N. L. (2013) ‘Dynamics of coastal meta-ecosystems: The intermittent upwelling hypothesis and a test in rocky intertidal regions’, *Ecological Society of America*, 83(3), pp. 283–310. doi: 10.1890/12-1706.1.

Meza, J. *et al.* (1998) *Censos de pingüino de Humboldt (Spheniscus humboldti) en el Monumento Natural Isla Cachagua y Santuario de la Naturaleza Islote Pájaro Niño, 1990-1997*.

Mura-Jornet, I. *et al.* (2018) ‘Chinstrap penguin population genetic structure: one or more populations along the Southern Ocean?’, *BMC Evolutionary Biology*. BMC Evolutionary Biology, 18(1), p. 90. doi: 10.1186/s12862-018-1207-0.

Nadeau, S. *et al.* (2016) ‘The challenge of separating signatures of local adaptation from those of isolation by distance and colonization history: The case of two white pines’, *Ecology and Evolution*, 6(24), pp. 8649–8664. doi: 10.1002/ece3.2550.

Nims, B. D. *et al.* (2008) ‘Low genetic diversity and lack of population structure in the endangered Galápagos penguin (Spheniscus mendiculus)’, *Conservation Genetics*, 9(6), pp. 1413–1420. doi: 10.1007/s10592-007-9465-1.

De Oliveira, L. R., Fraga, L. D. and Majluf, P. (2012) ‘Effective population size for South American sea lions along the Peruvian coast: The survivors of the strongest El Niño event in history’, *Journal of the Marine Biological Association of the United Kingdom*, 92(8), pp. 1835–1841. doi: 10.1017/S0025315411001871.

Pennington, J. T. *et al.* (2006) ‘Primary production in the eastern tropical Pacific: A review’, *Progress in Oceanography*, 69(2–4), pp. 285–317.

Peterson, B. K. *et al.* (2012) ‘Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species’, *PLoS ONE*, 7(5), pp. 1–11. doi: 10.1371/journal.pone.0037135.

Petkova, D., Novembre, J. and Stephens, M. (2015) ‘Visualizing spatial population structure with estimated effective migration surfaces’, *Nature Genetics*. Nature Publishing Group, 48(1), pp. 94–100. doi: 10.1038/ng.3464.

Pritchard, J. K., Stephens, M. and Donnelly, P. (2000) ‘Inference of Population Structure Using Multilocus Genotype Data’, *Genetics Society of America*, 155(2), pp. 945–959. doi: 10.1111/j.1471-8286.2007.01758.x.

Purcell, S. *et al.* (2007) 'PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses', *The American Journal of Human Genetics*, 81(3), pp. 559–575. doi: 10.1086/519795.

Roeder, A. D. *et al.* (2001) 'Gene flow on the ice: Genetic differentiation among Adélie penguin colonies around Antarctica', *Molecular Ecology*, 10(7), pp. 1645–1656. doi: 10.1046/j.0962-1083.2001.01312.x.

Rojas de Mendiola, B. (1981) 'Seasonal phytoplankton distribution along the Peruvian coast', 1, pp. 348–356. doi: 10.1029/CO001p0348.

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

Ryynänen, H. J. *et al.* (2007) 'A comparison of biallelic markers and microsatellites for the estimation of population and conservation genetic parameters in Atlantic salmon (Salmo salar)', *Journal of Heredity*, 98(7), pp. 692–704. doi: 10.1093/jhered/esm093.

Saba, V. S. *et al.* (2008) 'An oceanographic context for the foraging ecology of eastern Pacific leatherback turtles: Consequences of ENSO', *Deep-Sea Research Part I: Oceanographic Research Papers*, 55(5), pp. 646–660. doi: 10.1016/j.dsr.2008.02.006.

Sallaberry-Pincheira, N. *et al.* (2016) 'Contrasting patterns of selection between MHC I and II across populations of Humboldt and Magellanic penguins', *Ecology and Evolution*, 6(20), pp. 7498–7510. doi: 10.1002/ece3.2502.

Schlatter, R. and Simeone, A. (1999) 'Estado del conocimiento y conservación de las aves en mares chilenos.', *Estudios oceanológicos*, 18, pp. 25–33.

Schlosser, J. A. *et al.* (2009) 'Evidence for gene flow differs from observed dispersal patterns in the Humboldt penguin, Spheniscus humboldti', *Conservation Genetics*, 10(4), pp. 839–849. doi: 10.1007/s10592-008-9644-8.

Schmidt-nielsen, K. (1972) 'Locomotion : Energy Cost of Swimming , Flying , and Running', *Science*, 177(4045), pp. 222–228.

Sexton, J. P., Hangartner, S. B. and Hoffmann, A. A. (2014) 'Genetic isolation by environment or distance: Which pattern of gene flow is most common?', *Evolution*, 68(1), pp. 1–15. doi: 10.1111/evo.12258.

Silva, N., Rojas, N. and Fedele, A. (2009) 'Water masses in the Humboldt Current System: Properties, distribution, and the nitrate deficit as a chemical water mass tracer for Equatorial Subsurface Water off Chile', *Deep-Sea Research Part II: Topical Studies in Oceanography*, 56(16), pp. 1004–1020. doi: 10.1016/j.dsr2.2008.11.001.

Simeone, A. *et al.* (2002) 'Oceanographic and climatic factors influencing breeding and colony attendance patterns of Humboldt penguins Spheniscus humboldti in central Chile', *Marine Ecology Progress Series*, 227(Lack 1954), pp. 43–50. doi: 10.3354/meps227043.

Simeone, A. and Wallace, R. S. (2013) 'Evidence of philopatry and natal dispersal in Humboldt Penguins', *Emu*, 114(1), pp. 69–73. doi: 10.1071/MU13021.

Slatkin, M. (1985) ‘Rare Alleles as Indicators of Gene Flow’, *Evolution*, 39(1), pp. 53–65. doi: 10.2307/2408516.

Stenseth, N. C. *et al.* (2002) ‘Ecological effects of climate change’, *Science*, 297(August), pp. 1292–1296.

Szulkin, M. *et al.* (2016) ‘Population genomic footprints of fine-scale differentiation between habitats in Mediterranean blue tits’, *Molecular Ecology*, 25(2), pp. 542–558. doi: 10.1111/mec.13486.

Taylor, S. A., Maclagan, L., *et al.* (2011) ‘Could specialization to cold-water upwelling systems influence gene flow and population differentiation in marine organisms? A case study using the blue-footed booby, Sula nebouxii’, *Journal of Biogeography*, 38(5), pp. 883–893. doi: 10.1111/j.1365-2699.2010.02445.x.

Taylor, S. A., Zavalaga, C. B., *et al.* (2011) ‘Panmixia and high genetic diversity in a Humboldt Current endemic, the Peruvian Booby (Sula variegata)’, *Journal of Ornithology*, 152(3), pp. 623–630. doi: 10.1007/s10336-010-0628-3.

Teare, J. A. *et al.* (1998) ‘Nest-site fidelity in Humboldt penguins (Spheniscus humboldti) at Algarrobo, Chile.’, *Penguin Conservation*, 11, pp. 22–23.

Thiel, M. *et al.* (2007) ‘the Humboldt Current System of Northern-Central Chile Oceanographic Processes , Ecological Interactions’, *Oceanography and Marine Biology: An Annual Review*, 45(3), pp. 195–344. doi: Book_Doi 10.1201/9781420050943.

Vianna, J. A. *et al.* (2014) ‘Changes in Abundance and Distribution of Humboldt Penguin Spheniscus Humboldti’, *Marine Ornithology*, 42, pp. 153–159.

Vianna, J. A., Noll, D., Mura-jornet, I., *et al.* (2017) ‘Comparative genome-wide polymorphic microsatellite markers in Antarctic penguins through next generation sequencing’, *Genetics and Molecular Biology*, 40(3), pp. 676–687.

Vianna, J. A., Noll, D., Dantas, G. P. M., *et al.* (2017) ‘Marked phylogeographic structure of Gentoo penguin reveals an ongoing diversification process along the Southern Ocean’, *Molecular Phylogenetics and Evolution*, 27, pp. 486–498. doi: 10.1016/j.ympev.2016.12.003.

Wallace, R. S. *et al.* (1999) ‘Movements of Humboldt Penguins from a breeding colony in Chile’, *Waterbirds*, 22(3), pp. 441–444. doi: 10.2307/1522121.

Wallace, R. S. and Araya, B. (2015) ‘Humboldt penguin Spheniscus humboldti population in Chile: Counts of moulting birds, February 1999–2008’, *Marine Ornithology*, 43(1), pp. 107–112.

Wang, I. J. and Bradburd, G. S. (2014) ‘Isolation by environment’, *Molecular Ecology*, 23(23), pp. 5649–5662. doi: 10.1111/mec.12938.

Wieters, E. A. *et al.* (2003) ‘Alongshore and temporal variability in chlorophyll a concentration in Chilean nearshore waters’, *Marine Ecology Progress Series*, 249, pp. 93–105. doi: 10.3354/meps249093.

Wright, S. (1943) ‘Isolation by Distance’, *Genetics*, 28(2), pp. 114–138. doi: Article.

Resumen

La especialización a sistemas altamente productivos y heterogéneos (como la corriente de Humboldt) podría llevar a filopatría y a una estructura genética, debido a condiciones locales favorables. O bien, puede promover movimientos durante episodios de calentamiento del agua, que impulsan la dispersión y el flujo génico. Aunque estudios previos ya han establecido la estructura genética del pingüino de Humboldt (*Spheniscus humboldti*), el papel del medio ambiente sobre este patrón no se ha estudiado antes. Utilizamos una aproximación de genoma completo con SNPs neutrales análisis de redundancia (RDA) para dar cuenta de la estructura genética del pingüino de Humboldt y su asociación con el entorno geográfico y ambiental de su rango de distribución. Encontramos una estructura genética leve, aunque significativa, con tres grupos genéticos: primero, la colonia peruana Punta San Juan; segundo Chañaral, la principal colonia chilena; y un grupo que comprende las colonias chilenas restantes (Pan de Azúcar, Isla Choros, Cachagua y Puñihuil). Aunque la prueba de Mantel para el aislamiento por distancia (IBD) no fue significativa, la significación estadística del vector latitudinal (dbMEM1) en dos de nuestros modelos RDA, podría ser un indicador de IBD. Lo anterior ya que la latitud se muestra como el factor principal que separa al extremo norte, la localidad Punta San Juan en Perú, de las colonias chilenas. Además, la clorofila-a, como indicador de productividad primaria, también fue significativa en nuestro modelo de RDA, lo que sugiere que la estructura genética encontrada podría ser producto de las condiciones ambientales locales. Destacamos la importancia de la configuración geográfica y ambiental del Sistema de Corriente de Humboldt sobre la estructura genética de la población de *S. humboldti*. Por lo tanto, en el escenario de un acelerado cambio climático, la estructura genética, que está determinada por las condiciones locales en un gradiente latitudinal, debe tenerse en cuenta al considerar las estrategias de manejo y conservación para esta especie vulnerable.

Annex 1: Tables

Table 1. Diversity indices obtained for the 57 individual's 963 neutral SNPs. N: Number of sampled individuals per colony; S: number of polymorphic sites; Usable loci: number of loci with less than 5% missing values; H_o: observed heterozygosity; H_e: expected heterozygosity; π: nucleotide diversity; N_A: mean number of alleles per locus; A_R: average allelic richness.

Locality	ID	N	S	Usable loci	H_o	H_e	π	N_A	A_R
Punta San Juan	SJ	12	364	816	0.31	0.32	0.143 ± 0.071	1.88	1.275
Pan de Azúcar	PA	14	380	792	0.30	0.31	0.149 ± 0.073	1.95	1.289
Chañaral	CH	11	411	879	0.31	0.32	0.149 ± 0.074	1.92	1.289
Isla Choros	IC	8	382	847	0.31	0.33	0.149 ± 0.075	1.89	1.293
Cachagua	CA	10	380	827	0.32	0.33	0.151 ± 0.076	1.89	1.280
Puñihuil	PU	2	247	917	0.54	0.55	0.147 ± 0.097	1.51	1.280

Table 2. Genetic and geographic distances. F_{ST} values above, geographic distance in kilometers below. Bold values are significant: ***<0.001, **<0.01, *<0.05.

Locality	SJ	PA	CH	IC	CA	PU
SJ	-	**0.018*** **	**0.023*** **	**0.019*** **	**0.021*** **	**0.019***
PA	1288	-	**0.007** **	-0.003	-0.004	0.003
CH	1620	331	-	**0.010*** **	**0.009*** **	0.004
IC	1643	356	26	-	-0.001	-0.006
CA	1951	715	394	369	-	0.007
PU	2956	1780	1450	1425	1063	-

Table 3. Redundancy Analysis (RDA) and models employed using dbMEMs and oceanographic variables. Geno: Genotypes; SST: sea surface temperature; Chl-a: chlorophyll a concentration; dbMEM1: latitude vector; dbMEM2: longitude vector. Partial models

watch for the relative influence of environmental variables controlling (Condition) the effect of variables inside the parentheses. SSV: statistically significant variable within the model and respective p-value.

SNPs	RDA	Analysis	Model	Model *p*-value	Adjusted R^2	SSV (*p*-value)
963 neutral	Partial	Environmental	Geno ~ Chl-a + SST + Condition (dbMEM1 + dbMEM2)	0.348	0.036	-
		Spatial	Geno ~ dbMEM1 + dbMEM2 + Condition (Chl-a + SST)	**0.037**	0.039	dbMEM1 (0.009)
		Environmental-Spatial	Geno ~ dbMEM1 + Chl-a	**0.001**	0.046	dbMEM1 (0.001) Chl-a (0.016)
	Total	All variables	Geno ~ Chl-a + SST + dbMEM1 + dbMEM2	**0.001**	0.081	SST (0.001) dbMEM1 (0.009)

Table 4. Detection of first-generation migrants (GeneClass2). The simulation algorithm used was Rannala & Mountain (1997). 31 individuals were found to come from a different colony than the origin (probability of assignation < 0.01). Values in bold correspond to self-recruitment. -* Puñihuil had not an assigned probability for the two individuals sampled.

Number of migrants		From SJ	PA	CH	IC	CA	PU
To	**SJ** (12)	**10**	2	-	-	-	-
	PA (14)	-	**8**	1	1	4	-
	CH (11)	-	4	**5**	-	2	-
	IC (8)	-	7	1	**0**	-	-
	CA (10)	-	8	1	-	**1**	-
	PU (2)	-	-	-	-	-	**-***

Annex 2: Figures

Figure 1. Sampling locations of the Humboldt penguin at Peru (Punta San Juan) and Chile (Pan de Azúcar, Chañaral, Isla Choros, Cachagua and Puñihuil).

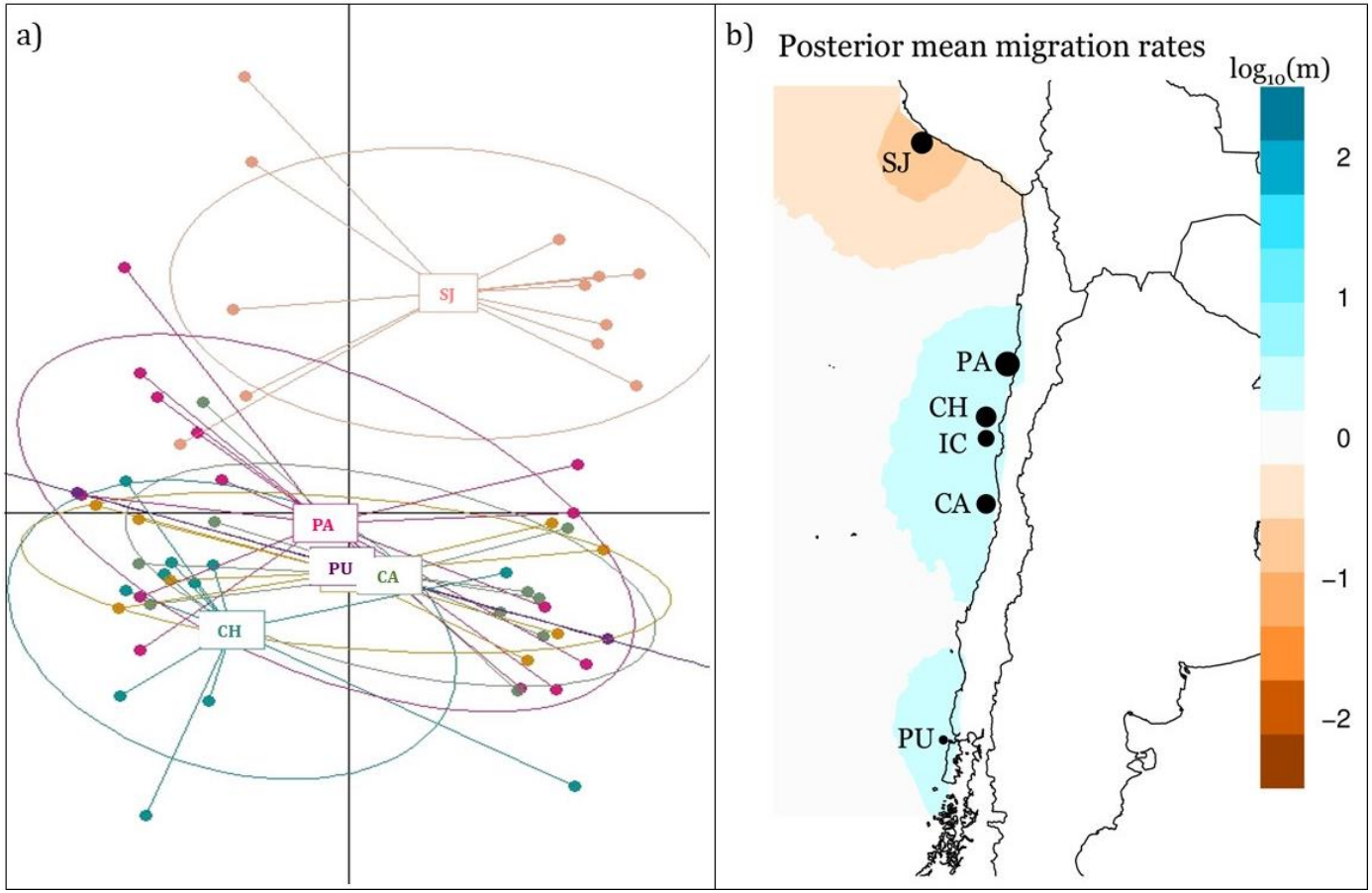

Figure 2. Genetic structure of the Humboldt penguin determined by a PCA (a) and estimating effective migration surfaces or EEMS (b). Both figures show the separation of Punta San Juan (SJ) as a different genetic group. The PCA analysis separates SJ by the PC2 and the EEMS shows a gene flow barrier between Chilean and Peruvian colonies.

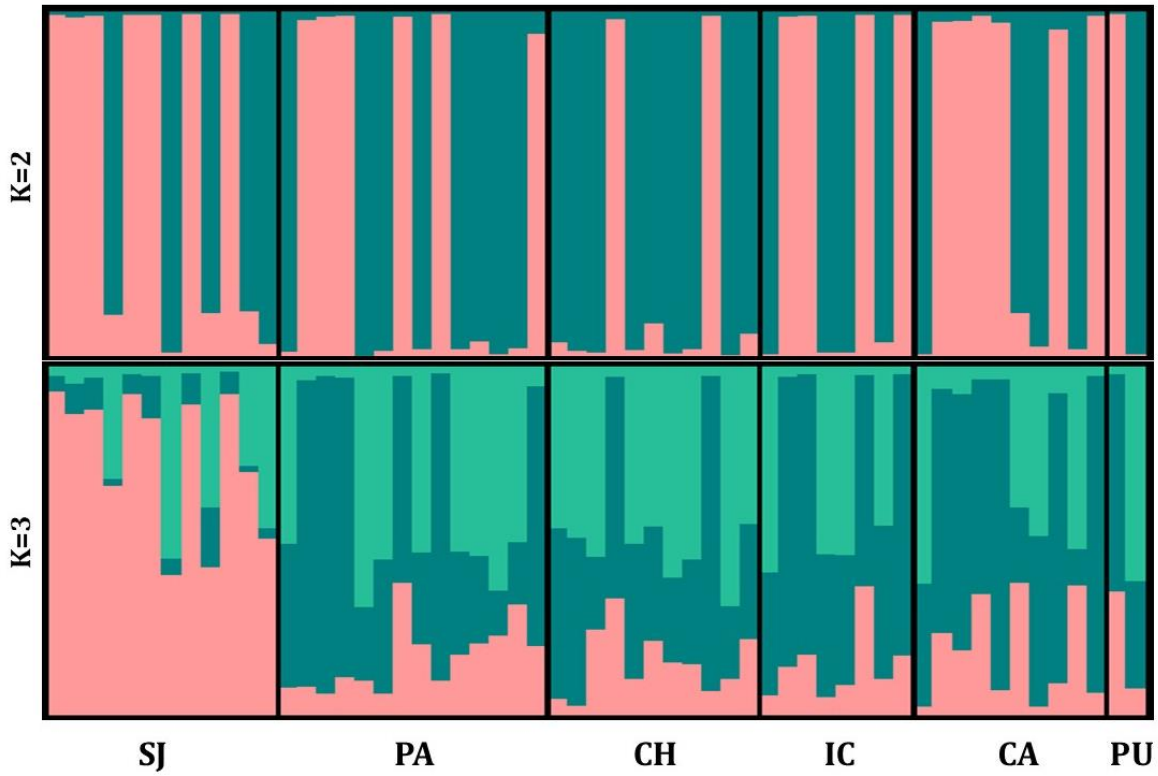

Figure 3. STRUCTURE analysis showing K=2 (above) and K=3 (below). The Evanno method used by structure showed K=2 as the most likely number of clusters present in the population, however, the genetic structure is lost here. Curiously, the population structure consistent with geography seen in the other analyses is recovered when plotting K=3, although there is no third cluster.

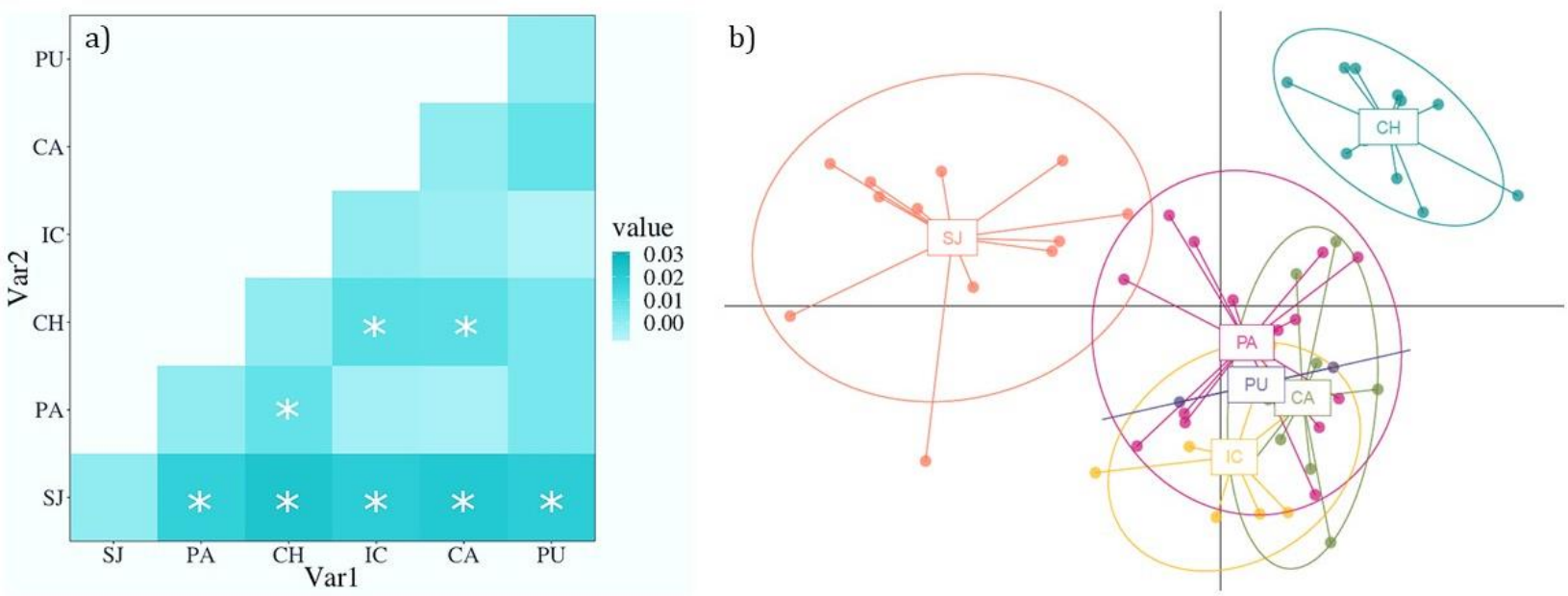

Figure 4. Pairwise F_{ST} values (a) and DAPC (b) showing a slight signal of genetic structure of the Humboldt penguin population. Punta San Juan (SJ) colony at Peru differentiates from all Chilean colonies, as well as Chañaral (CH).

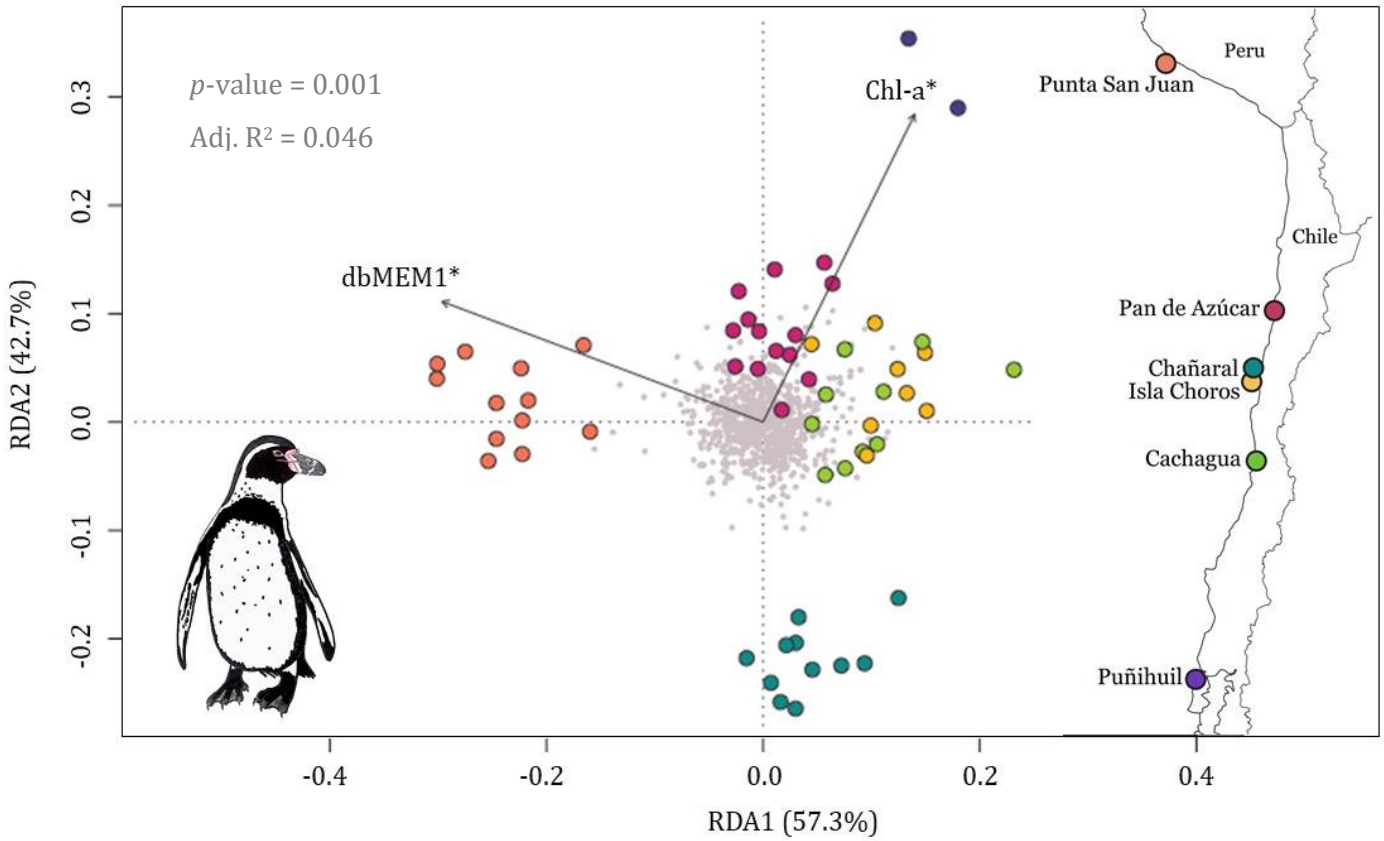

Figure 5. Redundancy analysis (RDA) performed on the 963 SNPs under neutral selection (rose dots in the middle). This model corresponds to the environmental-spatial partial model explaining genotypic variation using variables chosen by *ordistep* (Chl-a and dbMEM1). * for both variables, $p < 0.05$ according to ANOVA.

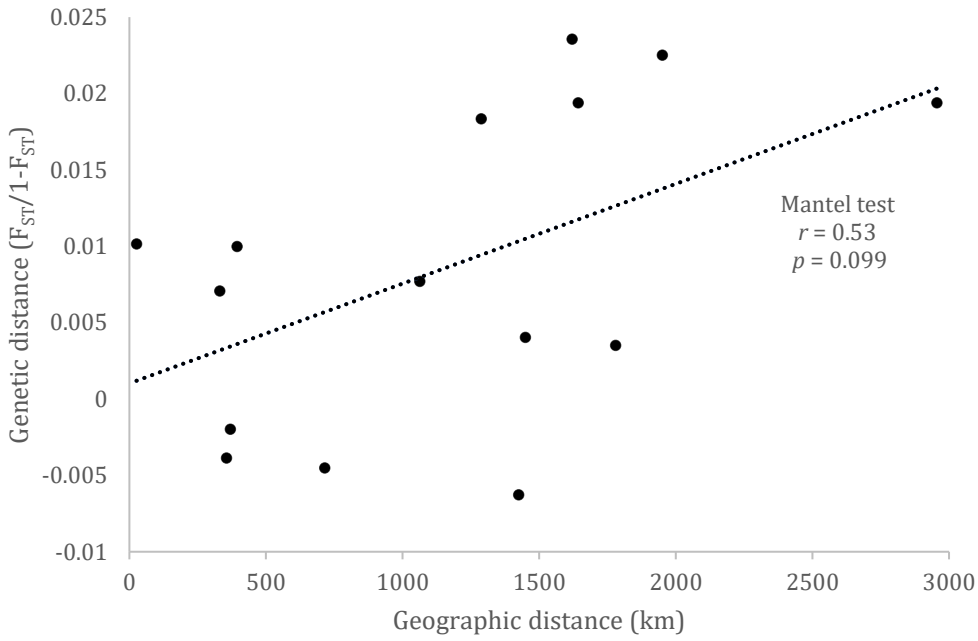

Figure 6. Correlation between geographic and genetic distances and Mantel test coefficient and *p*-value. There is a trend of correlation between geography and genetic structure, although the Mantel test was not significant.

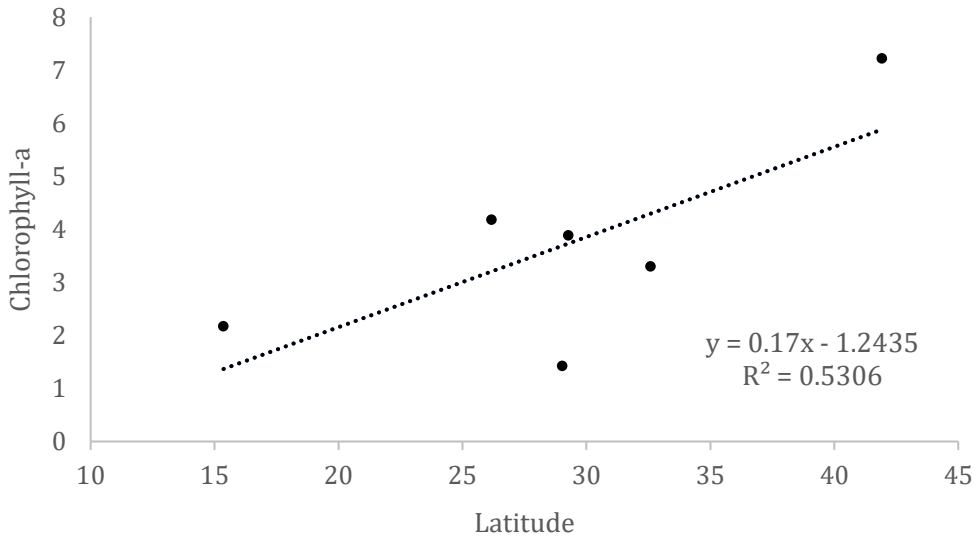

Figure 7. Latitude vs. Chlorophyll-a concentration. Correlation test (method pearson, kendall, and spearman) was not statistically significant (p = 0.1).