PNAS
www.pnas.org

Supplementary Information for

Genome-wide analyses reveal drivers of penguin diversification

Juliana A. Vianna, Flávia A. N. Fernandes, Maria José Frugone, Henrique V. Figueiró, Luis R. Pertierra, Daly Noll, Ke Bi, Cynthia Y. Wang-Claypool, Andrew Lowther, Patricia Parker, Celine Le Bohec, Francesco Bonadonna, Barbara Wienecke, Pierre Pistorius, Antje Steinfurth, Christopher Burridge, Gisele P. M. Dantas, Elie Poulin, W. Brian Simison, Jim Henderson, Eduardo Eizirik, Mariana F. Nery, Rauri C. K. Bowie

Juliana A. Vianna
Email: jvianna@uc.cl
Rauri C. K. Bowie
Email: bowie@berkeley.edu

This PDF file includes:

- Supplementary text
- Figures S1 to S16
- Tables S1 to S13
- SI References 124

www.pnas.org/cgi/doi/10.1073/pnas.2006659117

Supplementary Information Text

Materials and Methods

Sampling

The genomes of 18 extant penguin species (22 individuals) were sequenced. We sequenced the genome of macaroni penguins from two populations and gentoo penguins from geographic locations corresponding to four divergent lineages previously Identified (1-3; Table S1, S2). We sequenced the genome of a giant petrel from Antarctica as an outgroup for the analyses reported below. Blood samples were collected in the wild through our fieldwork for most penguin species, with the exception of erect-crested (AMNH 211990) and fiordland (AMNH 17808) penguins where tissues were provided from museums. A blood sample from an African penguin was kindly sampled for us by the California Academy of Sciences aquarium, and DNA was isolated from a preserved museum round-skin specimen of the yellow-eyed penguin (MVZ 149367).

Animal handling and sample collection were approved by the Ethics and Welfare Committee of Pontificia Universidad Católica de Chile to conduct this research. Samples from all penguins were collected under permit: Subsecretaría de pesca y acuicultura, Chile, permits 2086-2014, 110-2012; Pontificia Universidad Católica de Chile INACH 44/2012, 45/2016, 46/2016; Terres Australes et Antarctiques Francaises, 2012-126, 2012-111 and 2012-117, 2007-145; Convenio de cooperación internacional Fundación Charles Darwin, Parque Nacional Galápagos, Missuri U and Sant Louis Zoo; DEC Western Australia permits SF004716, CE000772; UNISINOS; Programa Antártico Brasileiro PPECEUA04.2017, PROANTAR nov. 2010, A14-SCI-ZOO-012; Nelson Mandela University Research Ethics Committee; Department of primary industries, parks, water and environment, Tasmania TFA 09198; 431; Ministry of research; Comite d'Ethique pour l'Expérimentation Animale Languedoc Roussillon APAFIS#9497-2017100417123472 v2. Samples from Antarctica were collected with permits in accordance to Annex II, Article 3 of the Protocol on Environmental Protection for Antarctic Research (SCAR) provided by the countries involved in this study.

Genome sequencing and assembly

DNA was isolated using a salt extraction protocol (4) with slight modifications (2). A total of 100 ng of genomic DNA was fragmented to 350 base pairs (on average) using a Covaris E220 focused ultrasonicator (USA). The sheared DNA was used to construct paired-end libraries using the Illumina TruSeq Nano kit. Fragmented DNA was treated with end repair mix, with A-tailing mix, and a ligation enzyme to attach the Illumina indexed adapter and barcode. Six PCR cycles were used to enrich ligated DNA using Illumina adapters. Amplified libraries were purified with beads and quantified using a Qubit fluorometer and the library size was measured with an Agilent TapeStation. The libraries were then sequenced to ~30x coverage (Table S3) with 150 paired-end reads using an Illumina HiSeq X platform at MedGenome (USA).

To process the raw reads obtained for each species, exact duplicates were removed using Super Deduper (https://github.com/dstreett/Super-Deduper). The reads were then filtered using Cutadapt (5) and Trimmomatic (6) to trim adapters and low-quality reads. The remaining overlapping paired-end reads were merged using FLASH (7). We then aligned the resulting cleaned reads of each individual to an emperor penguin draft genome (http://gigadb.org/dataset/100005; scaffold-level assembly) using LAST (http://last.cbrc.jp/). LAST is more sensitive to finding similarities and is capable of finding similar regions in spite of mismatches and gaps between reads and the reference genome. LAST handles sequence quality data better during mapping than some other algorithms (8). For each penguin species, the resulting alignment in LAST MAF format was converted to sorted BAM format using maf-convert and SAMtools (9). Finally, we used SAMTools/bcftools and vcfutils.pl vcf2fq (9) to call genotypes in order to generate individual consensus sequences (in fastq format). We kept a consensus base only when the sequence depth was >10X and masked sites within a 5 bp window around an indel. We converted fastq to fasta using seqtk (https://github.com/lh3/seqtk) for each individual.

After mapping the data of each individual (Table S1) to the reference genome of the emperor penguin, each reconstructed genome was of equal length. Therefore, we used the reference genome GFF based on the annotations from GeneWise for gene predictions, to extract coding sequences (CDS) and introns (of each gene) for each individual. The CDS and intron sequences for all species were then aligned using MAFFT (10) and trimAl (11) was used to trim ambiguous regions in the alignment (Commands: trimal

-in $in -out $out -block 20 -resoverlap 0.3 -seqoverlap 30 -keepseqs -gappyout). We used a conservative filter and removed loci if more than 30% missing data (Ns) was present across 30% of the individuals. We also removed loci from the alignment if the proportion of shared polymorphic sites at any locus was greater than 20%. Penguins and giant petrel raw fastq reads, reconstructed genomes (BioProject PRJNA530615, BioSample accession SAMN11566608-SAMN11566630) and mitogenomes (MK760983-MK761004, MK761006) were deposited in Genbank. All UCE, CDS, intron and mitogenome alignments, dated phylogenies and scripts for all data analyses are available at Dryad (https://doi.org/10.5061/dryad.pk0p2ngj2). Scripts are also available at the Github indicated below.

Code for the above available at:
https://github.com/CGRL-QB3-UCBerkeley/denovoTargetCapturePhylogenomics/blob/master/1-ScrubReads_e

Assessment of genome assemblies

The statistics of the final assembled genomes were assessed using asmstats.pl (https://github.com/calacademy-research/ccgutils/tree/master/asmstats), a modified version of the assemblathon_stats.pl statistics program from Assemblathon 2 (12) (Table S3, Fig. S1).

Benchmarking Universal Single-Copy Orthologs (BUSCO v2 (13) with the argument --limit 20 in genome mode (Table S4, Fig. S2) were used to show the reference-based assembly performed was sufficient to recover nearly all the single-copy protein sequences predicted in the avian lineage dataset (4915 BUSCO aves_odb9). BUSCO uses the gene prediction program Augustus v3.2.6 (Augustus: Gene Prediction, RRID:SCR_008417; (14)), HMMER 3.1b2 (http://hmmer.org/), and the Basic Local Alignment Search Tool (BLAST+) v 2.6 (National Center for Biotechnology Information [NCBI] BLAST).

Genomes were evaluated using KAT spectra-cn plots (15) to assess motif representation and to detect possible bias in assembly of using the emperor penguin species genome as a reference for all the species. We generated a k-mer spectrum plot (histograms plotting number of distinct k-mers at each frequency) for each of the 22 penguin genomes to obtain information such as the data quality (level of errors, sequencing biases, completeness of sequencing coverage with the KAT sect option, and potential contamination) and genomic complexity (size, karyotype, levels of heterozygosity, and repeat content) (15). We used the function Kat comp and k-mer length of 27 bp to compare K-mers generated from clean reads to those from the assembled genome for each species, and in comparison to the reference emperor penguin genome (NCBI ASM69914, Fig. S3, S4). Spectra-cn plots provide information of k-mer content type and the number of copies from the fastq reads that became part of the assembly (15). These histogram plots are derived from a matrix of the total number of distinct k-mers found in the fastq reads versus the frequency of each distinct k-mer found in the fastq reads. The total number of times a distinct k-mer occurs in the assembly is then used to paint the histogram assigned colors. Black is assigned to k-mers not found in the assembly and red is assigned to k-mers that occur once in the assembly (see figures for color assignments).

Code for the above available at:
https://github.com/calacademy-research/ccgutils/tree/master/asmstats

Gene trees and species tree

We performed the Ultra Conserved Elements (UCE) capture using scripts from the PHYLUCE pipeline available at https://github.com/faircloth-lab/phyluce (16). This pipeline includes an *in silico* pipeline for identifying UCE loci from whole genomes, by aligning the genome's contigs and scaffolds to a probe set of 5060 UCE loci specific for tetrapods (17). The script "probe_run_multiple_lastzs_sqlite" was used to align the 120 bp length UCE probes to the genomes. The UCE loci and their respective 750 bp flanking regions were then sliced from the genomes using the script "probe_slice_sequence_from_genomes". Because our 30x genomes are not complete, some missing data is expected, which means that not all UCEs were recovered for all taxa (18). To mitigate the effect of missing data, we included in the final alignment only UCE loci recovered from a minimum of 18 individuals (we sequenced 23 individuals). Sequences were aligned using MAFFT (10) commands: mafft --inputorder --leavegappyregion --threadit 0 --allowshift --unalignlevel 1 --anysymbol --retree 1 --thread $thread --maxiterate 1000 --globalpair $raw >

$mafft_aln_raw. The sequences were concatenated using the script “catfasta2phyml.pl” (available at https://github.com/nylander/catfasta2phyml). Although some UCE may occur within introns, and/or can overlap with coding regions (19), there are likely to be very few in number and consequently these were not filtered.

The mitochondrial sequence from each taxon was recovered by extracting pairs of reads that matched penguin mitochondrial sequences from NCBI GenBank previously published by (20). We used BLATQ version 1.02 (https://zenodo.org/record/61136#.XQFsc29Kjys) to match reads from the fastq pair of files (forward and reverse strand reads) for each individual; then we used a custom script to extract reads into a pair of fastq files representing the mitochondrial reads. For each individual the reads were imported into Geneious R11 (https://www.geneious.com) and the De Novo Assemble tool was used to generate the consensus assembly for the mitochondrion. Assembled mitogenomes were aligned with other available penguins (20); with our dataset comprising a total of 32 mitochondrial genomes and the giant petrel outgroup.

The most appropriate models of sequence evolution for each alignment were determined using ModelFinder (21) as implemented in IQ-TREE version 1.6.8 (22). For the concatenated phylogenomic analyses (without partitions) we performed analyses with and without the inclusion of the yellow-eyed penguin due to the more fragmented nature of the assembled genome given that we sequenced this individual from a skin sample off an old museum study skin (MVZ 149367). The full dataset with yellow-eyed penguin omitted comprised a total of 23,108 loci (CDS 11,011 loci; intron 8,040 loci; UCE 4,057 loci; Fig. S5, Table S5-S6); and with yellow-eyed penguin included comprised 9,103 loci (CDS 4,668 loci; intron 2,610 loci; UCEs 1,825 loci; Fig. S6, Table S5-S6); both datasets included the mitogenome (Fig. S7). The nucleotide substitution models selected were GTR+F+R2 for CDS, GTR+F+R2 for intron, GTR+F+R3 for UCE, and TVM+F+I+G4 for the mitogenome. The same models were selected for each marker with and without the yellow-eyed penguin included in the dataset. We carried out maximum likelihood (ML) analyses in IQ-TREE, with 1000 bootstrap replicates using the ultrafast bootstrap approximation (UFBoot, Fig. S5-S6) (23).

To account for potential genome-wide incompatibilities among taxa and loci, we used the species tree summary method incorporated by Astral-III (24) to estimate a species-level phylogeny that included yellow-eyed penguin for each of the UCE (1,825 loci), intron (2,610 loci), and CDS (4,668 loci) data sets, and one for all three datasets combined (9,103 loci) (Fig. S8). We used RAxML-NG (v. 0.5.1b BETA) (47) to generate independent gene trees for each locus of the UCE, intron, and CDS alignments. As input for Astral-III, we combined all of the RAxML-NG (v. 0.5.1b BETA) (25) maximum likelihood trees from each locus alignment into a single file. To achieve this, we wrote a python scripts enabling the parallelization of RAxML-NG analyses to generate input trees for Astral-III and a set of scripts to run Astral-III. Branch support values were calculated as local posterior probabilities computed in the context of the multispecies coalescent by a quartet-based support algorithm using quadripartition (the four clusters around a branch) (26). We chose the Astral-III summary method over full species-tree inference under the multispecies coalescent (e.g. *BEAST) because of the large size of the dataset (27).

Code for the above available at:
https://github.com/CGRL-QB3-UCBerkeley/denovoTargetCapturePhylogenomics/blob/master/4-TransExonCapPhyloV3.3.8
https://github.com/freitas-lucas/UCEs
https://github.com/calacademy-research/RAxML-NG-Astral-III
https://github.com/calacademy-research/RAxML_Astral_trees

Estimation of Divergence times

We estimated divergence times in BEAST v2.5.2 (27) for UCEs and mitogenomes including all taxa using the computer resources available through the CIPRES Science Gateway (Fig. S7, S10) (28). We did not use a fixed topology (e.g. the maximum likelihood tree) once we recovered the same topology with BEAST. Dating was performed using the GTRGAMMA model with base frequencies empirically estimated. We used a relaxed-lognormal clock under the calibrated Yule speciation process with fossil calibration distributions following (29). The five fossil calibrations we used are detailed in Table S8 and the position of the four internal calibration points are depicted on Figure 1. The calibration used followed the parameters used by Cole et al. (2019).

The root (i.e. Sphenisciformes / Procellariiformes node) was calibrated using the oldest known Sphenisciformes fossil, *Waimanu manneringi* (30)*,* with a lognormal prior distribution mean of 3.7, standard deviation of 1.0, and offset of 60.5 million years ago (Mya). The mean and standard deviation values of the lognormal prior specify a minimum age prior of 60.5 Mya, which corresponds to the minimum age of the locality of the species' type, and a maximum constraint of 72.1 Mya, which corresponds to the age of the boundary of the uppermost stage of the Late Cretaceous, the Maastrichtian stage, as described by (29).

A second fossil calibration was placed in the crown Spheniscidae where *Aptenodytes* diverges from the other genera. We used the fossil species *Madrynornis mirandus*, and following (31) with a uniform prior distribution and an offset of zero (0.0). The minimum age prior was 9.7 Mya, which corresponds to the minimum age estimate of the only specimen of this fossil from the Puerto Madryn Formation (10 ± 0.3 Mya, (31)), and a maximum age prior of 25.2 Mya, which is considered as the maximum age bound of crown penguins, in accordance with the estimated maximum age of New Zealand's Kokoamu Greensand formation, in which a wide variety of crown and stem fossil penguins have been found (29). *Madrynornis mirandus* has been considered a close relative to *Eudyptes* (31, 32), to *Spheniscus/Eudyptula* (33), and as a sister-taxon to Spheniscidae (34). Even though the exact position of the fossil is uncertain, its position at the base of the crown-group is the most commonly accepted placement (33).

The third calibration prior was placed at the stem node of *Pygoscelis* using the fossil *Pygoscelis calderensis* with a lognormal distribution, mean of 6.0, standard deviation of 1.0, and offset of 6.3 Mya. The mean and standard deviation values of the lognormal prior specify a minimum age of 6.3 Mya given that the fossil was found in the Bahía Inglesa Formation of Chile with an estimated K-Ar age of 7.6 ± 1.3 Mya (35). The maximum age prior was 25.2 Mya, following the same rationale for constraining the crown Spheniscidae as outlined above.

The fourth calibration prior was placed along the node uniting *Spheniscus* and *Eudyptula*, placing the fossil *Spheniscus muizoni* using a lognormal prior distribution, mean of 4.4, standard deviation of 1.0, and an offset of 9.2 Mya. The mean and standard deviation values of the lognormal prior specify a minimum age of 9.2 Mya derived from the stratigraphic position of the fossil in the Pisco Formation (Peru), which recent studies estimate to have a minimum age bound of 9.2 Mya (36). The maximum age prior was set to 23.03 Mya, which corresponds with the onset of the Miocene; to date no *Spheniscus* or *Eudyptula* fossils have been found in Oligocene deposits (37, 38).

Finally, the fifth calibration prior was placed along the node uniting *Eudyptes* and *Megadyptes*, in which an undescribed *Eudyptes sp.* fossil with a lognormal prior distribution, mean of 7.04, standard deviation of 1.0, and offset of 3.06. The mean and standard deviation values of the lognormal prior specify a minimum age of 3.06 Mya, which is the minimum age bound of the Tangahoe Formation (New Zealand) (39) in which the fossil was found, and a maximum age prior of 25.2 Mya, following the same rationale for constraining the crown Spheniscidae as outlined above. The conservative maximum bounds in the Oligocene (i.e. 25.2 Mya) used in most calibrations in our study, as well as other studies of crown-group penguins (e.g. (29)) was chosen in the absence of a better-known fossil interval for most of the fossil the taxa.

The MCMC was run over two independent runs: 500 million for the UCE, and 100 million generations for the mitogenome datasets for each run, sampling parameters every 10,000 generations. The output log files were analyzed in Tracer v.1.7.1 (40) and trees were summarized using TREEANNOTATOR (27) with a 10% burn-in. We ran the sample from the priors to evaluate how divergent our posterior probability distribution is for both UCE and mitogenomes of the calibrated phylogenies with and without the sequence data. The ESS values obtained for UCE were 42 without data and 3550 with data, while mitogenomes were 28 and 2595, respectively. The results suggest that the data have sufficient signal to estimate the parameters without being dominated by the prior selection. These results are expected since the selection of tree prior (under a Yule process) and molecular clock has almost no impact on the diversification rates estimates when sequence data are sufficiently informative and have low to moderate substitution rate heterogeneity (41).

The subsequent analysis to detect signatures of positive selection (see below) used the calibrated phylogeny for all taxa rooted on the giant petrel as an outgroup. For ancestral range and niche reconstruction (see below) we used the calibrated phylogeny pruned to one exemplar per species with redundant taxa removed: (I) macaroni/royal penguins reduced to a single lineage following recent papers (42, 43) and our analyses (Fig. S5-S8; macaroni penguin from Antarctica was retained); (II) we reduced the four gentoo penguin lineages to a single lineage by choosing the most divergent lineage of gentoo penguin

(Crozet Archipelago); and (III) removed the outgroup. Pruning was performed using the package ape (Analyses of Phylogenetics and Evolution) v5.3 (44).

Historical biogeographic analyses

For the historical biogeographic analysis, we estimated the ancestral range of the extant penguin species in the R package BioGeoBEARS (45). BioGeoBEARS implements three models of ancestral area reconstruction: Dispersal-Extinction-Cladogenesis (DEC) (46), Dispersal-Vicariance Analysis (DIVA) (47), and BayArea (48). Each of these methods allows the inclusion of a set of anagenetic (e.g. dispersal; extinction) and cladogenetic (e.g. sympatry; vicariance) processes as parameters (Table S9, Fig. S11). In addition, BioGeoBEARS can also include the parameter j ("jump dispersal"), which accounts for founder-event speciation (49) as an additional variable to the above models. The founder-event speciation parameter from BioGeoBEARS allows a daughter lineage to have a different range from the parental lineage during a cladogenetic event, an important feature to be considered for seabirds that often breed on remote islands (50).

BioGeoBEARS estimates the likelihood of the ancestral states (i.e. ancestral range) at the nodes of the phylogeny, using as input the dated phylogenetic tree and a geographic file containing an adjacent matrix coding the current distribution of the species. The phylogenetic tree used as input in BioGeoBEARS must have only one representative of each species (or monophyletic populations) at the tips of the tree. Thus, we pruned the dated phylogeny (Fig. 1) to include only 18 taxa (one representative of each penguin species, for details see above). The geographic adjacent matrix depicts the areas where each species natively breeds in a presence/absence matrix. Each area is considered as a character state in the tree, and the output tree contains the set of areas (i.e. range) covered by the ancestral lineages of Spheniscidae. We tested all previously mentioned models alone (i.e. DEC, DIVALIKE, BAYAREALIKE), and with the addition of the founder-event speciation parameter j (i.e. DEC+J, DIVALIKE+J, BAYAREALIKE+J) (51).

For determination of each species' occurrence when coding the adjacent matrix, we followed the current distribution of penguin colonies or breeding range along continental margins and islands as described in Bertelli and Giannini (52), Handbook of the Birds of the World and BirdLife International (53). We subdivided the extant penguin geographic distribution into 10 different areas: A) South American coasts and Falkland/Malvinas Islands; B) Scotia Arc Islands; C) Antarctic Peninsula; D) Continental Antarctic; E) Tristan da Cunha and Gough Island; F) Bouvet Island; G) South African coast; H) Indian ocean Islands; I) Australia/New Zealand coasts and nearby islands; and J) Galápagos Islands (Fig. 1).

Our selection of 10 areas is based on the current understand of the biogeography, ecology, and biology of penguins. We took into consideration the location of nesting sites along continental margins and on island coasts, the extent to which biogeographic units are isolated (completely or partially) from each other due to oceanic fronts (e.g. Antarctic Polar Front, and Subtropical Front) or geographic distance, and degree to which the faunal composition of penguin species differs across the Southern Ocean.

Specifically, we separated the South American coast (area A) from the Galápagos Islands (area J), because the Galápagos penguins are endemic to the Galápagos Islands whereas the two South American *Spheniscus* overlap in distribution along the continental margin. The South American coast and the Falkland/Malvinas Islands (area A) are home to the Humboldt and Magellanic penguins, and are separated from the Scotia Arc Islands (area B) by the Antarctic Polar Front, which forms a major biogeographic break between these areas for several marine species including penguins (1). These two areas (A and B) as well as the Antarctic Peninsula (area C), and Continental Antarctica (D), are separated by each having distinct penguin species composition: the Antarctic Peninsula encompasses the range for the three *Pygoscelis* species (Adélie, chinstrap and gentoo penguins), and on Continental Antarctica Adélie and emperor penguins occur, which show reduced population structure (54, 55) suggesting that Antarctica need not be further subdivided. Tristan da Cunha and Gough islands (area E) are situated to the north of the Subtropical Front and the only penguin species to occupy these islands is the northern rockhopper. Bouvet Island (area F) is thousands of kilometers distant from the other penguin breeding colonies distributed to the south of the Antarctic Polar Front. The South African coast (area G) is home to a single endemic penguin species, the African penguin. The western Indian ocean encompasses a group of islands (area H: Marion, Prince Edward, Crozet, Kerguelen, Amsterdam Islands) that are similar in penguin faunal composition. Finally, Australia and New Zealand (I) both share little penguin.

Ancestral Niche Reconstruction

Occurrence records for all penguin species were retrieved from GBIF. Spatial data was filtered by IUCN distribution ranges with a resolution of 10x10 km. Six marine variables (max/min temperature, max/min salinity, and max/min primary productivity) taken from the Bio-Oracle repository (56) were used to create raw models with MaxEnt–Javascript (57). From here we estimated the extent of niche overlap between all penguin species for the set of variables considered (58). We used the package 'Phyloclim' to create Predicted Niche Occupancy (PNO, Fig. S12) profile values and plots (59). Subsequently we combined this information with the phylogenetic tree for penguins generated in this study. This was done by calculating the weighted means of niche dimensions for each species and expressing them through phylogenetic relatedness over time to estimate ancestral climatic tolerances (59). As a result, we obtained Divergence Through Time plots and climatic tolerance chronograms (Fig. 1, S13).

Code for the above available at:
https://github.com/mjfrugone/Predicted-Niche-Ocupancy.git

Interspecies introgression

Introgression analyses were performed using a partitioned *D*-statistics approach implemented in DFOIL (60). All clades were sampled following the software requirement of a symmetrical tree composed of four taxa and an outgroup, one ingroup clade being younger than the other, for eight combinations of taxa (Fig. S9, Table S7). To perform the tests, we split the genome-wide alignments into 100 kb, non-overlapping windows with Bedtools and custom scripts. The 100kb is the recommended window length based on the simulations performed by the authors of DFOIL. Windows smaller than 100kb has an increase proportion of false positive results and windows bigger than 100kb may include two sample regions with differing recombination rates. We considered a window to be introgressed if it had a p-value < 0.05. We parsed DFOIL output to obtain the percentage of windows with each type of introgression signal.

Code for the above available at:
https://github.com/henriquevf/Introgression-penguins

Demographic history

To address questions about climate and effective population size for each penguin, we performed a demographic analysis using a pairwise sequential Markovian coalescent (PSMC) method (61) based on whole-genome sequencing to elucidate effective population size through time. In order to prepare our data for input into an analysis using PSMC version 0.6.5-r67 (61), we used Samtools version 1.3.1 with HTSlib 1.3.1 (9, 62), bcftools version 1.3.1 (63), and the vcfutils.pl script from bcftools to call variants from the alignment derived from LAST with the command "samtools mpileup -C50 -uf ref.fa aln.bam | bcftools view -c - | vcfutils.pl vcf2fq -d 10 -D 100 | gzip > diploid.fq.gz". As per the recommendation of the PSMC documentation (https://github.com/lh3/psmc), we used a third of the average read depth as the minimum read depth (-d) and at least twice the average read depth as the maximum read depth (-D) (-d 10 -D 100). PSMC was run with parameters "-N25 -t15 -r5 -p 4 + 25*2 + 4 + 6". We estimated generation time (g) for the different penguin species based on the average maximum age at sexual maturity multiplied by a factor of two as suggested by (64): *Spheniscus* ssp. and little (g=8), chinstrap, gentoo, yellow-eyed, king (g=6), emperor (g=12), macaroni and royal (g=15) and the remaining *Eudyptes* species and Adélie (g=10) (Fig. 3, S16 Table S2). We assumed a nucleotide substitution rate of m = 1.91 x 10^{-9} substitutions/site/year based on the chicken lineage (*Gallus gallus*) (65) multiplied by the generation time of each species. We performed 100 bootstraps to estimate uncertainty in the estimates of changes in effective population size through time.

We compared the generation time we estimated with that provided by IUCN, that estimate by Forcada et al. (66), and that based on the formula suggested by Lande et al. (67). The differences among these estimates of generation time was approximately three generations. Re-running our analyses with these different generation times does not significantly influence our estimates of the timing of the demographic events, but results in a more recent decrease of N_e for some species during the last glaciation.

Detection of signatures of positive selection

Coding sequences were extracted from all species based on the available annotation for the emperor penguin using gffread (https://github.com/gpertea/gffread). Every taxon was tested for the presence of in-frame start and stop codons and premature stop codons with custom scripts. We performed a dN/dS ratio test using the Codeml algorithm implemented in ETE3 (68, 69). We performed a site test (M7 vs M8) for all species and sub-groups in the phylogeny, focusing on the 4,562 genes that could be reliably compared across all taxa and that contained a start and stop codon at the beginning and end of the CDS, respectively. Table S10 reports the number of loci obtained after filtering for each species and their average length. The calibrated tree obtained from analyzing the UCE dataset was used. Due to the large number of analyzed genes, we performed a multiple comparison (false-discovery rate [FDR]) test implemented in R, with a q-value threshold of 0.1 using the alpha values extracted from the standard output of ETE3 as input (Table S11). Genes that persisted on the list (i.e. remained significantly different from neutral expectations, supporting a positive selection regime) were then used to perform a gene ontology and network analysis. We used WebGestalt (WEB-based GEne SeT AnaLysis Toolkit) (70) for the gene ontology analyses with the following parameters: overrepresentation enrichment analysis (ORA) as method of interest, and ‘geneontology’ as the functional database (Table S11-S13). The network analysis was performed with StringDB (71) with default parameters using human as reference (Fig. 2, S14, S15). For the genes of the network analyses we performed the overall and pairwise genome Z-tests (d_N/d_S) to further evaluate positive and purifying selection (72). We further analyzed the sites under positive selection for the genes with the lowest *p*-value for the branch-site results (Table S12) and belonging to the network cluster related to specific functions (e.g. renal function, circulatory system, immunity, Figure 3). We employed the Mixed Effects Model of Evolution (MEME; 73) and Fast Unconstrained Bayesian Approximation (FUBAR; 74) tools implemented on the Datamonkey server (http://www.datamonkey.org; 75, 76; Figure S15). Default significance thresholds were used for both tools (p-value < 0.1 for MEME and posterior probability > 0.9 for FUBAR).

Code for the above available at:
https://github.com/henriquevf/Selection-penguins/

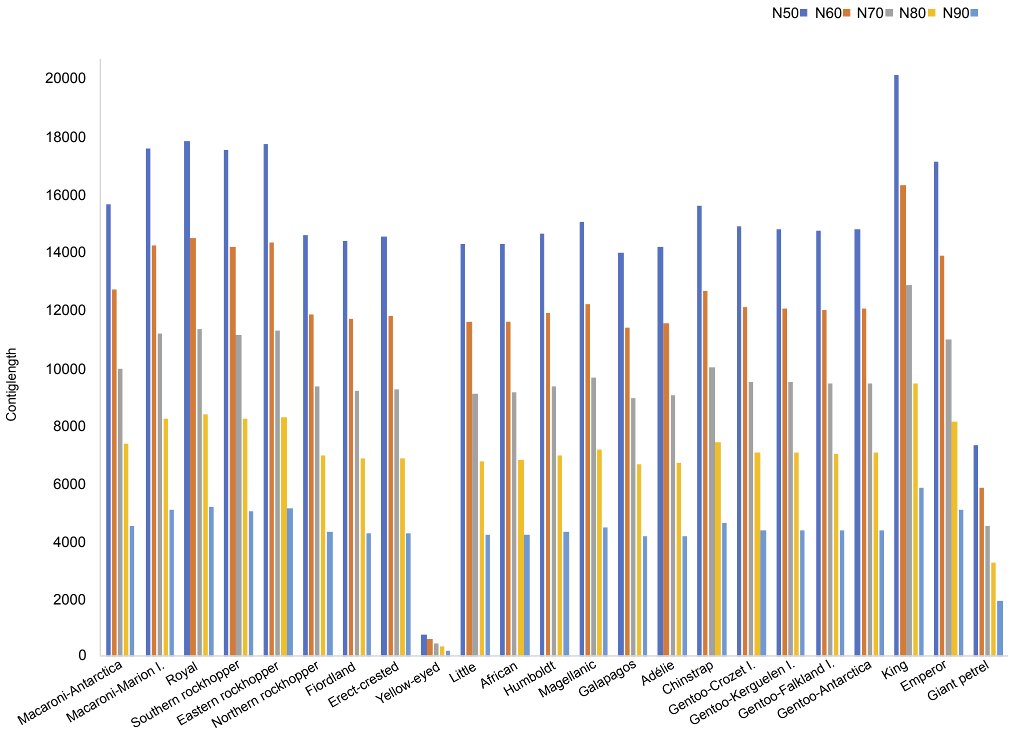


Fig S1. Data quality control for the 22 penguin genomes we sequenced. Summary statistics of scaffold assembly lengths, with the N50 (mean contig length) ranging between 13.9 kb for Galápagos penguin and 20.1 kb for king penguin: the yellow-eyed penguin was an exception, N50 = 0.75 kb. Genomes can currently be assembled using two main methods for assembling high throughput sequencing reads into longer contiguous genomic sequences to recover partial to near-complete genome sequences. De novo approaches assemble overlapped reads to build longer contigs and therefore reconstruct progressively the genome sequence (77). However, this method is generally unable to produce a single contig corresponding to the whole genome as non-contiguous sequences are frequent and require complex gap filling algorithms to infer their respective position (77). Assembly methods based on using a reference genome avoids such problems, as non-overlapping contigs will be placed in relation to each other based on the position of homologous sequences in the reference genome (78). Even if some areas cannot be filled by any reads, the vast majority of the partial genome will be accurately mapped, and this method has been widely used in studies similar to ours (79). Such accuracy will principally depend on the degree of identity between the reference and sequenced genomes, and phylogenetic proximity is generally considered as the best criteria for choosing a reference genome to assemble against (78). In the present study, we preferred the reference assembly method as a high-quality penguin species genome was already available and complied with the criteria of close phylogenetic relationship. Using a high-quality reference genome also has the advantage that with the mapping of each of our species to it, we can use the existing intron-exon boundaries to extract contiguous loci. Our phylogenetic reconstruction and divergence time estimates was based on a UCE dataset, the choice of genome assembly method should not have an effect on the phylogenetic hypothesis presented here and all other analyses based on this phylogeny (e.g. DTT, adaptation, introgression analyses) as our approach is similar to what several other authors have adopted when extracting UCE loci from short-read data. Further, all our nuclear genomic datasets (exon, UCE, introns) recovered very similar topologies with similar branches lengths between the exon and intron trees. This suggests that there is no particular bias in our approach.

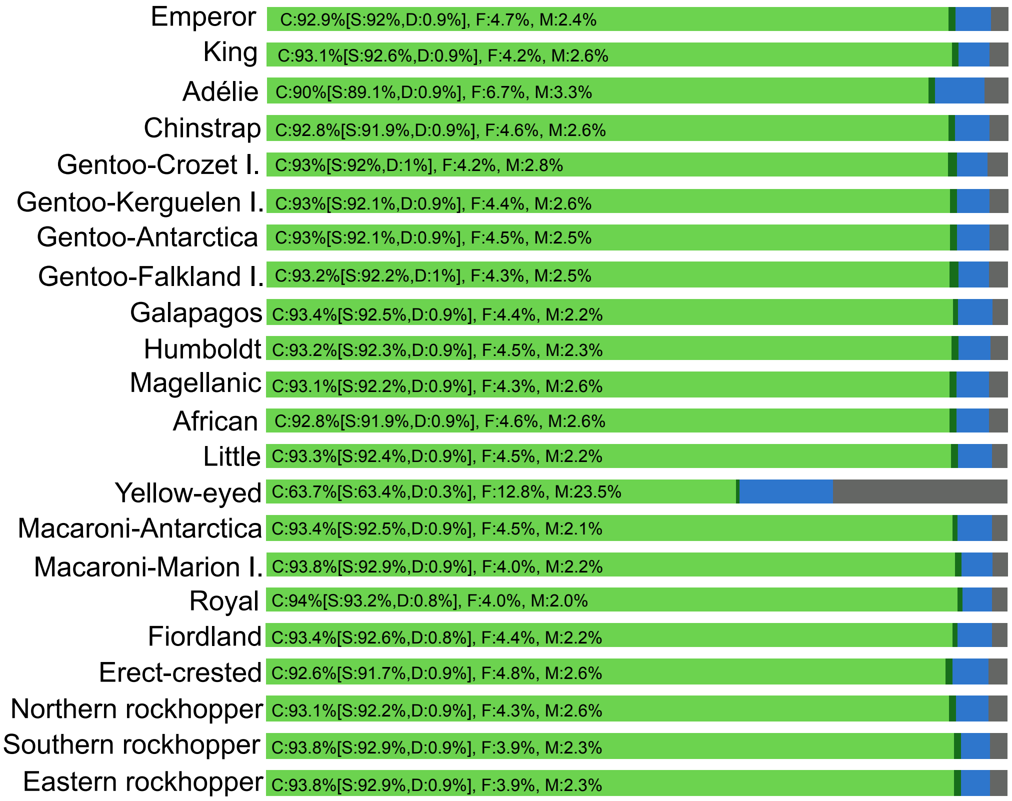


Fig S2. Results of benchmarking genome quality using the Universal Single-Copy Orthologs (BUSCO) locus set (n = 4915). Bar charts produced with the BUSCO plotting tool show proportions of BUSCO loci classified as complete (C, green + dark green), complete single-copy (S, green), complete duplicated (D, dark green), fragmented (F, blue), and missing (M, gray). Genome sequences had high BUSCO completeness scores (over 90%, with the exception of the yellow-eyed penguin ~63.7%, sequenced from a historical skin sample).

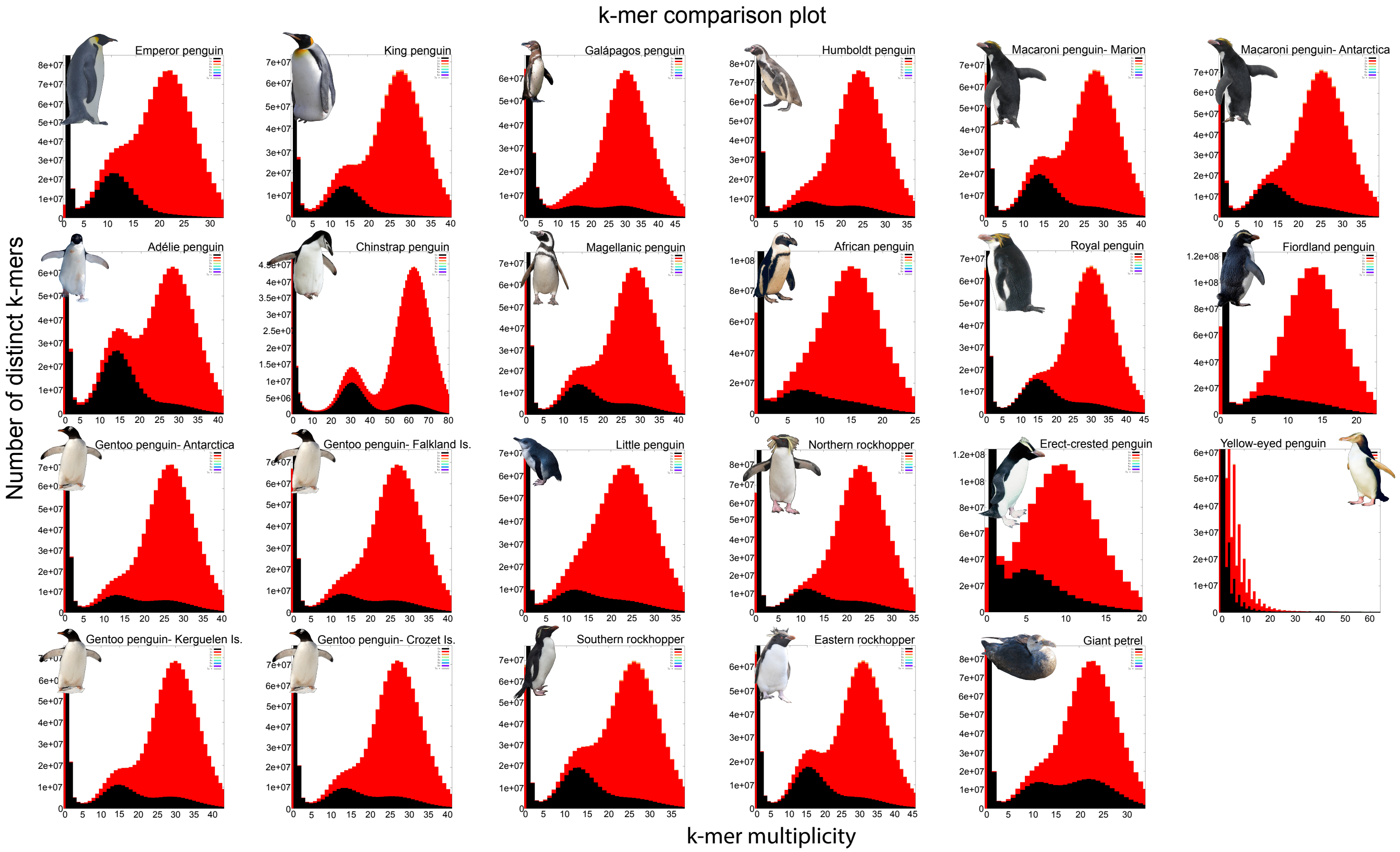


Fig. S3. KAT spectra-cn plots for all 22 penguin genomes and for the giant petrel, generated by comparing the paired-end reads to the reconstructed scaffolds of each penguin taxon. Plots are colored to show the number of times k-mers from the reads appear in the assembly; frequency of occurrence (multiplicity; x-axis) and number of distinct k-mers (y-axis). Black represents k-mers absent from the assembly; red represents k-mers that appear once in the assembly; the other colors represent k-mers that appear multiple times, but these are very few and hence not readily visible on the figures. Most genomes exhibited similar spectra-cn plots, with the majority of k-mers occurring in the assembly once (red distribution) and few duplicates detected. Individual penguin spectra-cn plots showed a black distribution (zero occurrences in the assembly) at frequencies ranging from 5-15 or 5-25. These black distributions represent k-mers spanning heterozygous regions where the De Bruijn graphing method rejects one of

the two SNPs spanning k-mers, reflecting the level of heterozygosity. Heterozygous content was high in chinstrap and Adélie penguins, followed by macaroni, southern and eastern rockhoppers, Magellanic, and king and emperor penguins. The exception was the genome sequence obtained from the historical specimen of the yellow-eyed penguin, where the spectra-cn plot revealed that half of the reads were not represented in the assembly (black), likely a consequence of the more fragmented nature of the genome assembly due to the presence of degraded DNA.

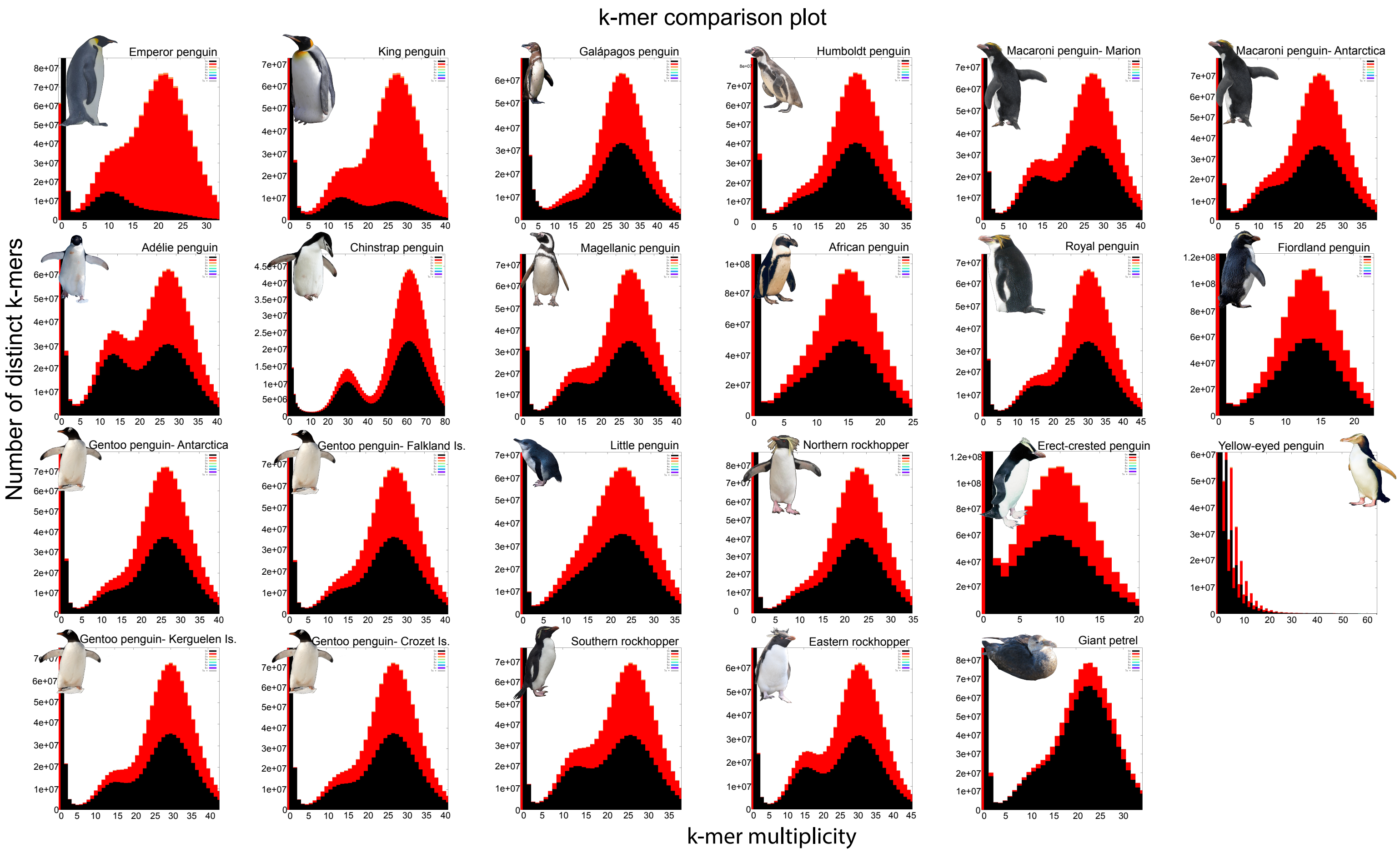


Fig S4. KAT spectra-cn plots for all 22 penguin genomes and for the giant petrel, generated by comparing the paired-end reads to the emperor penguin reference genome **(80)**. Plots follow those in Figure S3. Plots show a similar pattern as Figure S3, however, as expected, they show a higher number of k-mers to be absent from the assembly.

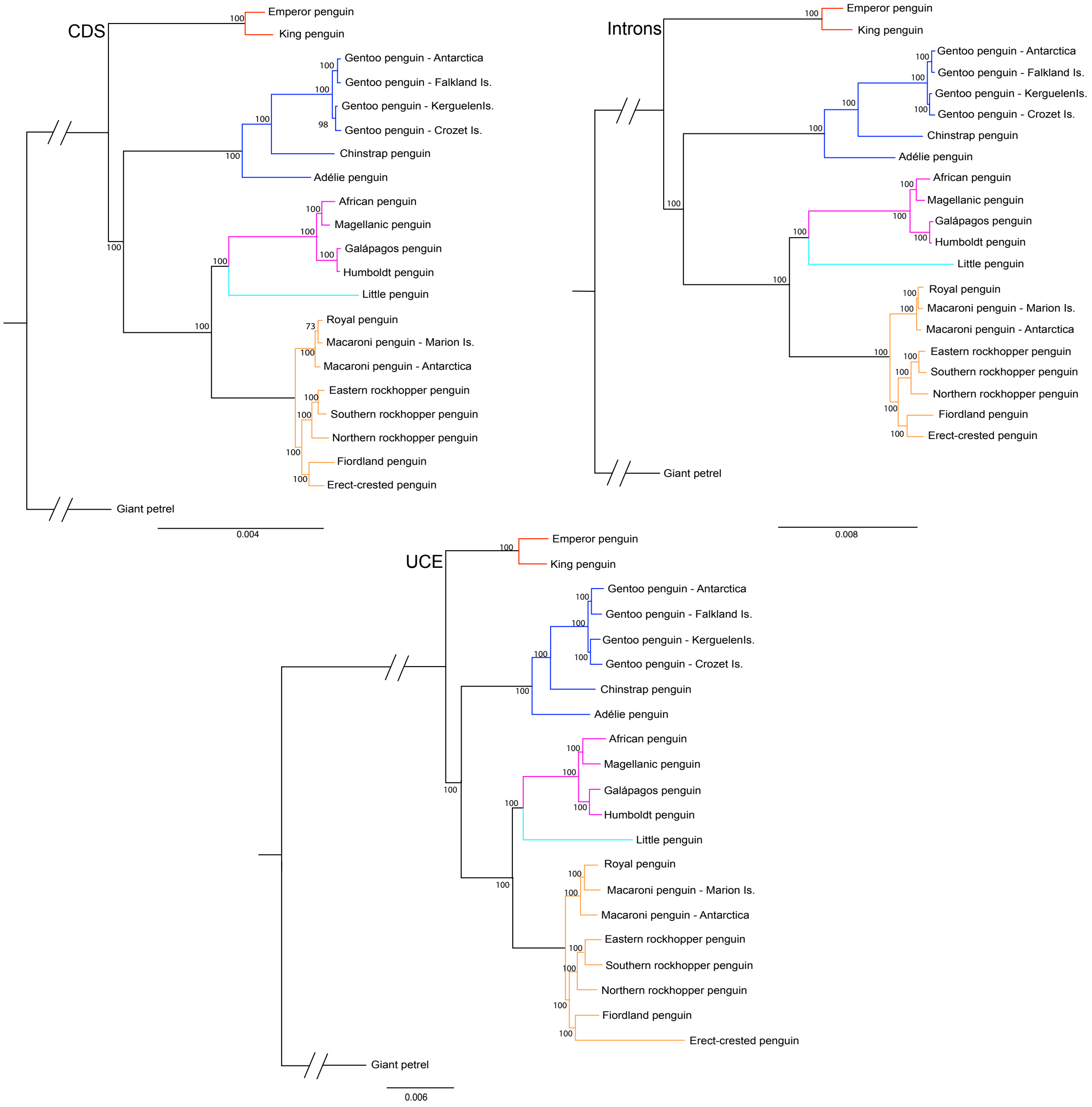


Fig S5. Maximum likelihood (ML) phylogenetic reconstruction of concatenated data for 21 genomes (omitting yellow-eyed penguin) representing 18 penguin species with the giant petrel used as the outgroup. ML topology for CDS, introns and UCEs. Node values indicate bootstrap support values. The total number of loci used for tree reconstruction excluding the yellow-eyed penguin genome was 23,108 loci: CDS 11,011 loci; intron 8,040 loci; UCE 4,057 loci. Absence of monophyly was recovered for royal and macaroni penguins; a result consistent with those of studies investigating population genetics of these taxa, which have suggested that they be considered a single species (42, 43). All phylogenies generated from the different datasets recovered a similar topology. All trees support the genus *Aptenodytes*, emperor and king penguins, as the first diverging clade (as suggested by 32, 52, 81-83) as opposed to *Aptenodytes* being placed closer to the tip of the phylogeny and sister to the smaller *Pygoscelis* penguins (29, 84, 85). Moreover, the short internal branches recovered by our analysis at the divergence of *Aptenodytes* from other penguins likely indicates that a rapid radiation occurred in less than 1 million years, possibly by

multiple cladogenetic events occurring simultaneously in various lineages (86). This may have led to the recovery of the *Aptenodytes/Pygoscelis* clade by previous studies that used a small number of molecular markers (or small fragments of the mitogenome (29)), which may have not separated concomitantly when the lineages diverged. Fewer than 10 loci may be incapable of recovering rapid divergence events (87).

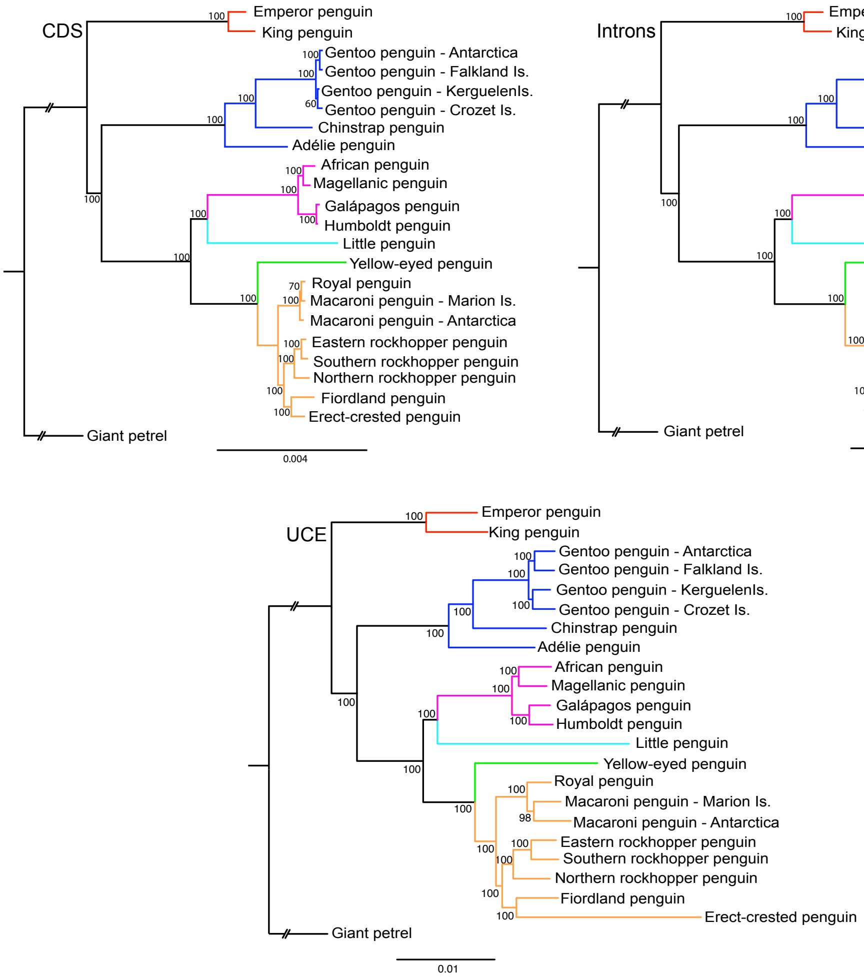

CDS
Emperor penguin
King penguin
Gentoo penguin - Antarctica
Gentoo penguin - Falkland Is.
Gentoo penguin - KerguelenIs.
Gentoo penguin - Crozet Is.
Chinstrap penguin
Adélie penguin
African penguin
Magellanic penguin
Galápagos penguin
Humboldt penguin
Little penguin
Yellow-eyed penguin
Royal penguin
Macaroni penguin - Marion Is.
Macaroni penguin - Antarctica
Eastern rockhopper penguin
Southern rockhopper penguin
Northern rockhopper penguin
Fiordland penguin
Erect-crested penguin
Giant petrel
100
60
70
0.004
Introns
Emperor penguin
King penguin
Gentoo penguin - Antarctica
Gentoo penguin - Falkland Is.
Gentoo penguin - KerguelenIs.
Gentoo penguin - Crozet Is.
Chinstrap penguin
Adélie penguin
African penguin
Magellanic penguin
Galápagos penguin
Humboldt penguin
Little penguin
Yellow-eyed penguin
Royal penguin
Macaroni penguin - Marion Is.
Macaroni penguin - Antarctica
Eastern rockhopper penguin
Southern rockhopper penguin
Northern rockhopper penguin
Fiordland penguin
Erect-crested penguin
Giant petrel
100
0.008
UCE
Emperor penguin
King penguin
Gentoo penguin - Antarctica
Gentoo penguin - Falkland Is.
Gentoo penguin - KerguelenIs.
Gentoo penguin - Crozet Is.
Chinstrap penguin
Adélie penguin
African penguin
Magellanic penguin
Galápagos penguin
Humboldt penguin
Little penguin
Yellow-eyed penguin
Royal penguin
Macaroni penguin - Marion Is.
Macaroni penguin - Antarctica
Eastern rockhopper penguin
Southern rockhopper penguin
Northern rockhopper penguin
Fiordland penguin
Erect-crested penguin
Giant petrel
100
98
0.01

Fig S6. Maximum likelihood (ML) phylogenetic reconstruction of concatenated data for 22 genomes representing 18 penguin species with the giant petrel used as the outgroup. ML topology for CDS, introns and UCEs. Node values indicate bootstrap support values. The total number of loci used for tree reconstruction with the 23 individuals was 9,103 loci: CDS 4,668; intron 2,610; and UCE 1,825 loci. Absence of monophyly was recovered for royal and macaroni penguins; a result consistent with those of studies investigating population genetics of these taxa, which have suggested that they be considered a single species (42, 43).

Node	BPS	Node age (Mya)	HPD (95%)
1	1	54.4	
2	1	23.22	9.78 - 25.20
3	1	6.03	2.64 - 9.63
4	1	19.93	16.17 - 23.16
5	1	12.36	8.68 - 16.03
6	1	8.20	4.99 - 11.66
7	1	2.71	1.51 - 4.15
8	1	0.83	0.43 - 1.29
9	1	0.31	0.13 - 0.53
10	1	15.42	12.17 - 18.53
11	1	12.58	9.71 - 15.38
12	1	3.50	2.21 - 4.98
13	1	1.23	0.74 - 1.76
14	1	2.15	1.22 - 3.28
15	1	10.32	7.37 - 13.22
16	1	5.80	4.00 - 7.78
17	1	0.41	0.19 - 0.65
18	1	4.76	3.09 - 6.58
19	1	3.19	1.87 - 4.67
20	1	1.32	0.67 - 2.04
21	1	3.19	1.79 - 4.83
22	1	4.58	2.38- 6.92

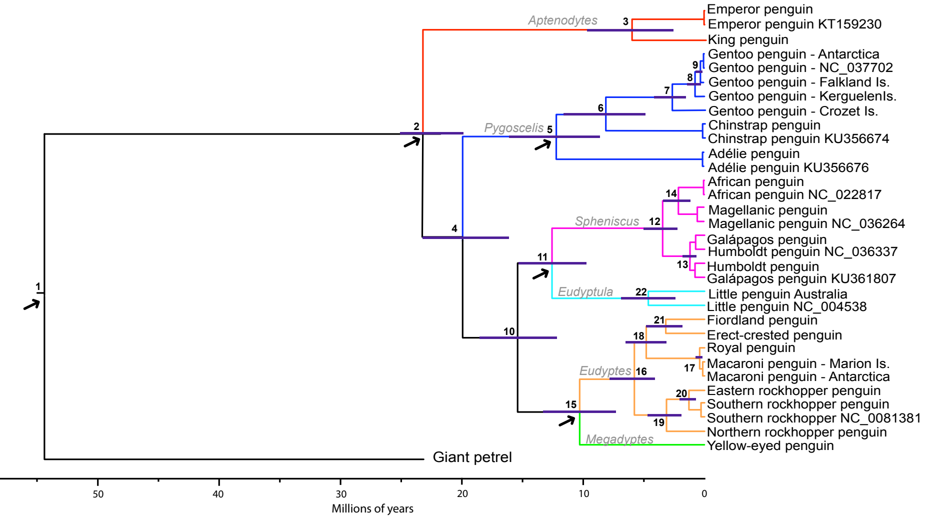

Aptenodytes
3
Emperor penguin
Emperor penguin KT159230
King penguin
Gentoo penguin - Antarctica
9
Gentoo penguin - NC_037702
8
Gentoo penguin - Falkland Is.
7
Gentoo penguin - KerguelenIs.
6
Gentoo penguin - Crozet Is.
2
Pygoscelis
5
Chinstrap penguin
Chinstrap penguin KU356674
Adélie penguin
Adélie penguin KU356676
African penguin
14
African penguin NC_022817
Magellanic penguin
Spheniscus
12
Magellanic penguin NC_036264
4
Galápagos penguin
Humboldt penguin NC_036337
11
13
Humboldt penguin
Galápagos penguin KU361807
1
Eudyptula
22
Little penguin Australia
Little penguin NC_004538
21
Fiordland penguin
10
18
Erect-crested penguin
Royal penguin
Eudyptes
16
17
Macaroni penguin - Marion Is.
Macaroni penguin - Antarctica
20
Eastern rockhopper penguin
15
Southern rockhopper penguin
19
Southern rockhopper NC_0081381
Northern rockhopper penguin
Megadyptes
Yellow-eyed penguin
Giant petrel
50
40
30
20
10
0
Millions of years

Fig S7. Bayesian phylogenetic (BA) inference using complete mitochondrial genomes for 32 individuals representing 18 penguin species. Numbers indicated on nodes corresponds to the accompanying table: Bayesian posterior probabilities (BPS), node age (Mya), HPD (95%). Arrows represent the five fossil calibration points used. In addition to the 22 newly generated mitogenomes, an additional 10 published penguin mitogenomes were include representing seven penguin species (20), plus little penguin from New Zealand (30); African penguin from South Africa (88); and southern rockhopper penguin (89). The giant petrel was used to root the phylogeny. The phylogeny showed a similar topology to that of our nuclear gene topology, with the exception of relationships among the seven *Eudyptes* species, and among the four lineages of gentoo penguins sampled. A deep divergence was observed between little penguin sequences from New Zealand and Australia (see also 90 and these lineages likely each warrant species status: *Eudyptula novaehollandiae* (Australia) and *Eudyptula minor* (New Zealand) (65). The lack of divergence between the four genomes of Humboldt and Galápagos penguins, and the three genomes of macaroni and royal penguins, may be explained by incomplete lineage sorting (ILS) and/or recent introgression. Although the topology of the phylogeny based on mtDNA is not in agreement with Cole et al. (29), their divergence time estimates were similar to our results.

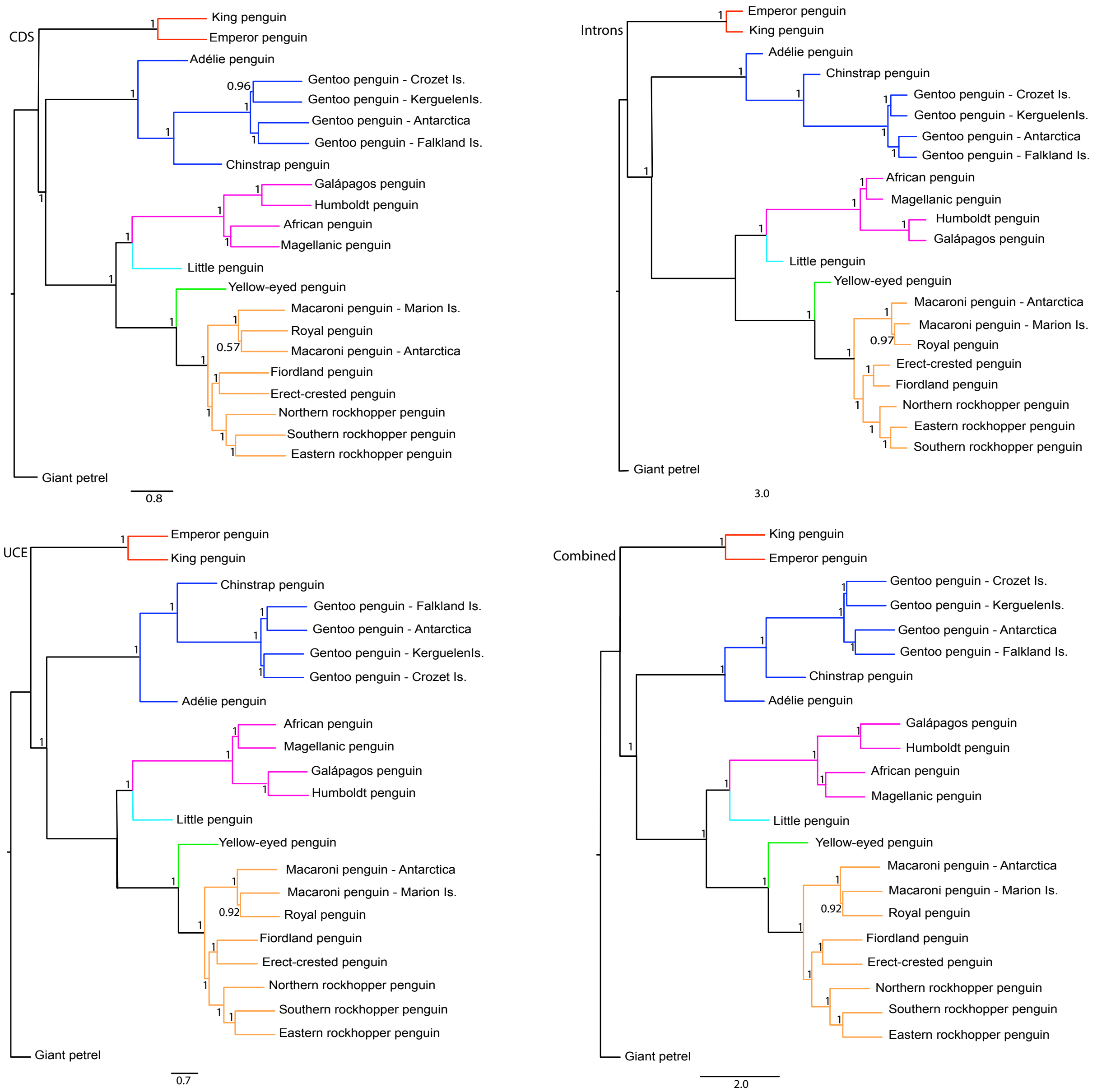

CDS
King penguin
Emperor penguin
Adélie penguin
0.96
Gentoo penguin - Crozet Is.
Gentoo penguin - KerguelenIs.
Gentoo penguin - Antarctica
Gentoo penguin - Falkland Is.
Chinstrap penguin
Galápagos penguin
Humboldt penguin
African penguin
Magellanic penguin
Little penguin
Yellow-eyed penguin
Macaroni penguin - Marion Is.
Royal penguin
0.57
Macaroni penguin - Antarctica
Fiordland penguin
Erect-crested penguin
Northern rockhopper penguin
Southern rockhopper penguin
Eastern rockhopper penguin
Giant petrel
0.8
Introns
Emperor penguin
King penguin
Adélie penguin
Chinstrap penguin
Gentoo penguin - Crozet Is.
Gentoo penguin - KerguelenIs.
Gentoo penguin - Antarctica
Gentoo penguin - Falkland Is.
African penguin
Magellanic penguin
Humboldt penguin
Galápagos penguin
Little penguin
Yellow-eyed penguin
Macaroni penguin - Antarctica
Macaroni penguin - Marion Is.
0.97
Royal penguin
Erect-crested penguin
Fiordland penguin
Northern rockhopper penguin
Eastern rockhopper penguin
Southern rockhopper penguin
Giant petrel
3.0
UCE
Emperor penguin
King penguin
Chinstrap penguin
Gentoo penguin - Falkland Is.
Gentoo penguin - Antarctica
Gentoo penguin - KerguelenIs.
Gentoo penguin - Crozet Is.
Adélie penguin
African penguin
Magellanic penguin
Galápagos penguin
Humboldt penguin
Little penguin
Yellow-eyed penguin
Macaroni penguin - Antarctica
Macaroni penguin - Marion Is.
0.92
Royal penguin
Fiordland penguin
Erect-crested penguin
Northern rockhopper penguin
Southern rockhopper penguin
Eastern rockhopper penguin
Giant petrel
0.7
Combined
King penguin
Emperor penguin
Gentoo penguin - Crozet Is.
Gentoo penguin - KerguelenIs.
Gentoo penguin - Antarctica
Gentoo penguin - Falkland Is.
Chinstrap penguin
Adélie penguin
Galápagos penguin
Humboldt penguin
African penguin
Magellanic penguin
Little penguin
Yellow-eyed penguin
Macaroni penguin - Antarctica
Macaroni penguin - Marion Is.
0.92
Royal penguin
Fiordland penguin
Erect-crested penguin
Northern rockhopper penguin
Southern rockhopper penguin
Eastern rockhopper penguin
Giant petrel
2.0

Fig S8. Species tree retrieved using the program ASTRAL-III with individual gene trees generated for each locus and the combined dataset using RAxML-NG. Our phylogenetic hypotheses comprised 22 genomes representing 18 penguin species with the giant petrel used as the outgroup. The species tree for CDS, introns, UCEs, and the combined dataset are depicted. The total number of loci used for tree reconstruction with yellow-eyed penguin included was 9,103 loci: CDS 4,668; intron 2,610; and UCE 1,825 loci. As some UCEs may occur within introns or can overlap with coding regions, these different datasets may not completely independent. Branch support values are local posterior probabilities computed in the context of the multispecies coalescent by a quartet-based support algorithm using quadripartitioning (the four clusters around a branch; 26). All trees generated from the different datasets recovered a similar topology. Node values are posterior probabilities. All trees support the genus *Aptenodytes*, emperor and king penguins, as the first diverging clade (as suggested by 32, 52, 81-83) as opposed to *Aptenodytes* being placed closer to the tip of the phylogeny and sister to the smaller *Pygoscelis* penguins as recovered in some phylogenetic hypotheses (29, 84, 85).

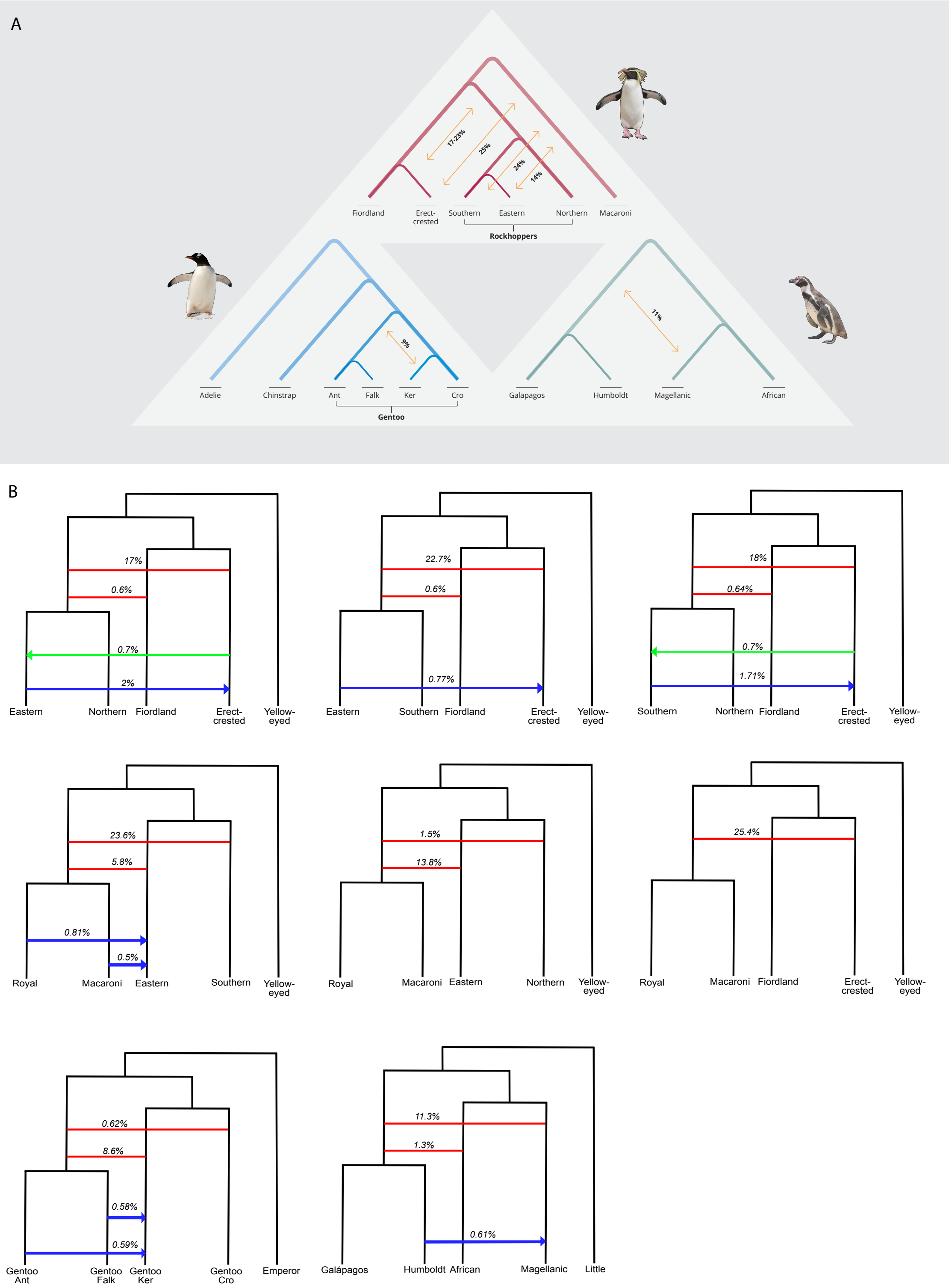


Fig S9. Introgression among species of penguins estimated using DFOIL. (A) Summary of introgression recovered in the eight 5-taxon statement analyses in (B), (B) DFOIL results of introgression among eight combinations of five penguin taxa. The arrows represent gene flow over 0.5% between taxa. Blue and

green arrows represent gene flow in two different directions between extant taxa. *Eudyptes* penguins appear to have exchanged genes through much of their evolutionary history, with extensive genome-wide introgression occurring between erect-crested and the ancestral rockhopper penguin species (17-23%) and between erect-crested and macaroni/royal (25%) penguins. We also recovered extensive genomic introgression between an ancestor of the Galápagos/Humboldt penguin and the Magellanic penguin (11%); Magellanic and Humboldt penguins are known to hybridize in the wild (91). Gentoo penguins exhibited genomic introgression (8.6%) between the Antarctic-South American clade with the lineage from Kerguelen Island. Signatures of more recent introgression episodes included genomic contributions from eastern and southern rockhopper penguins to the erect-crested penguin, as well the gentoo penguins from Antarctica and South America (Falkland/Malvinas) to Kerguelen. These results are consistent with the clockwise direction of the Antarctic Circumpolar Current (ACC) connecting sub-Antarctic islands and promoting dispersal and admixture.

Node	BPS	Node age (Mya)	HPD (95%)
1	1	55.80	
2	1	21.93	19.06 - 25.19
3	1	2.64	1.00 - 4.44
4	1	20.31	17.37 - 22.87
5	1	8.41	6.36 - 11.24
6	1	4.87	2.42 - 7.84
7	1	1.26	0.74 - 2.00
8	1	0.90	0.27 - 1.69
9	1	0.47	0.01 - 0.87
10	1	14.08	11.25 - 16.79
11	1	12.57	9.38 - 16.11
12	1	2.65	1.60 - 3.76
13	1	0.59	0.01 - 1.36
14	1	1.51	0.04 - 2.67
15	1	6.41	4.86 - 8.14
16	1	4.62	2.92 - 6.75
17	1	0.81	0.14 - 1.29
18	1	3.06	0.61 - 4.73
19	1	1.60	0.34 - 2.63
20	1	0.80	0.02 - 1.42
21	1	1.87	0.17 - 2.99

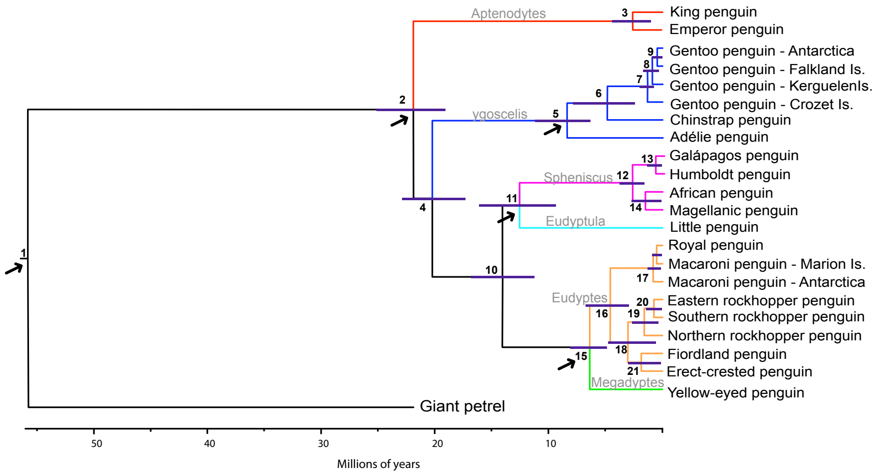
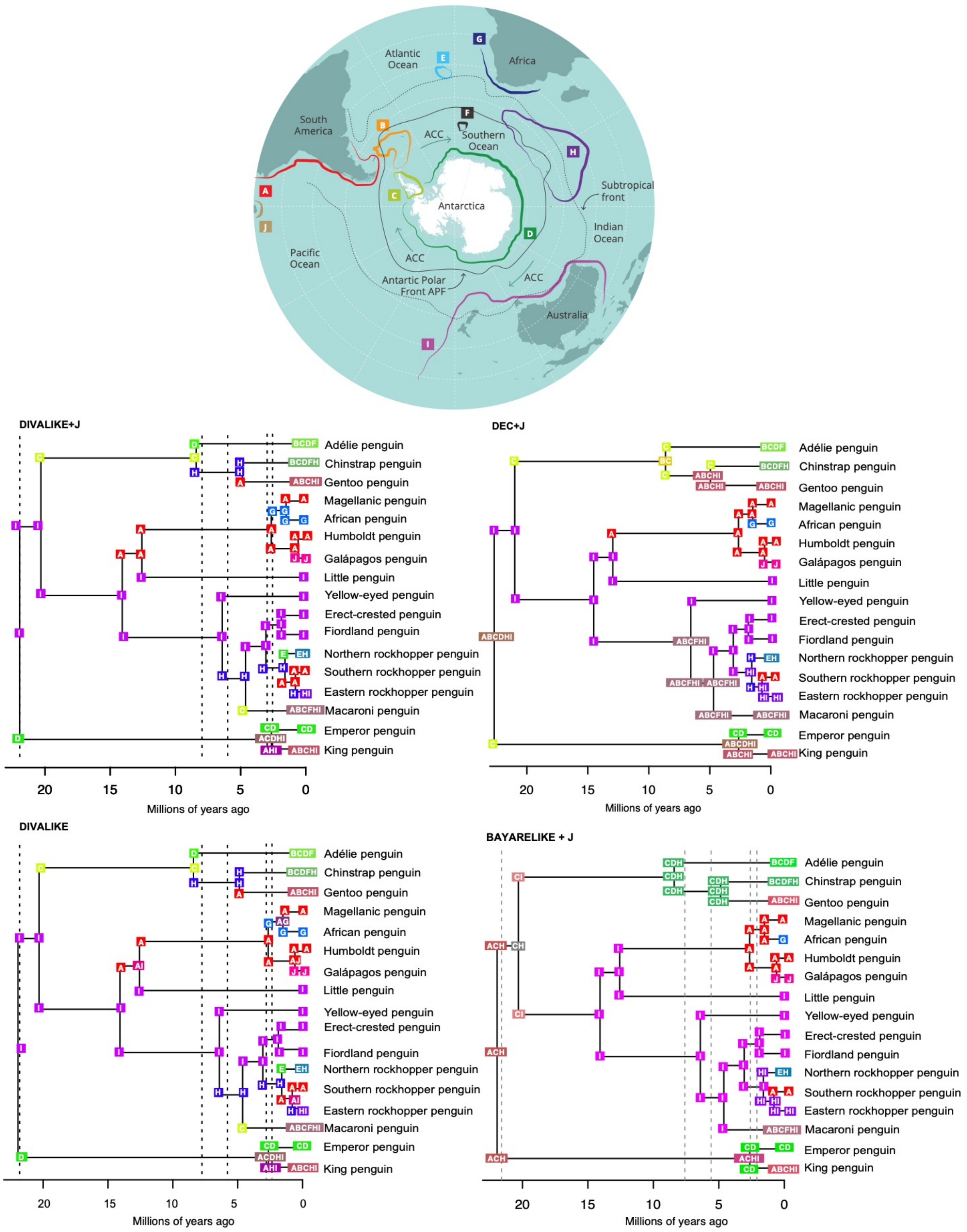


Fig S10. UCE Bayesian Phylogenetic (BA) inference with divergence time estimation generated using the software BEAST for penguins with the giant petrel as outgroup. Arrows represent the five fossil calibration points. Node numbers corresponds to those on the accompanying table: Bayesian posterior probabilities (BPS), node age (Mya), and HPD (95%) are reported. The ancestor of crown-group penguins was estimated at the early Miocene (UCE: 21.93 Mya; Mitogenome: 23.22 Mya). Initial opening of the Drake Passage was estimated between 45 and 24 Mya according to different authors (88, 89), with the most agreed date around 34 Mya (90, 91). Between 34 to 14 Mya, an ancestral volcanic arc east of the Drake Passage gateway was limiting the full ACC onset and the temperature in the Southern was relatively constant (92). During the middle Miocene climate transition, 14.2 to 13.8 Mya, the Southern Ocean cooled 6°C to 7°C due the intensification of atmospheric and oceanic circumpolar circulation increasing the isolation of Antarctic organisms (93). This period encapsulates the divergence between *Spheniscus*+*Eudyptula* and *Megadyptes*+*Eudyptes* (UCE: 14.08 Mya, Mitogenome 15.42 Mya) followed by the *Spheniscus/Eudyptula* split (UCE: 12.57 Mya, Mitogenome: 12.58 Mya). Full onset of the Drake Passage and the ACC is estimated around 11.6 Mya (94, 95), contributing to the diversification of several organisms of Antarctic marine fauna (96-102) and the colonization of new areas and diversification of most penguins (< 9 Mya). Approximately 10 Mya, ice became permanent in eastern Antarctica and around 5 Mya in Western Antarctica (103). Recent speciation of endemic penguins (< 5 Mya) might have occurred with island formation during the Pleistocene (e.g. Galápagos, Snares, Macquarie, Gough, Antipodes Islands), as has been proposed (29).


Atlantic Ocean
Africa
South America
Southern Ocean
ACC
Subtropical front
Antarctica
Indian Ocean
Pacific Ocean
Antartic Polar Front APF
Australia
A
B
C
D
E
F
G
H
I
J
DIVALIKE+J
DEC+J
DIVALIKE
BAYARELIKE + J
Adélie penguin
Chinstrap penguin
Gentoo penguin
Magellanic penguin
African penguin
Humboldt penguin
Galápagos penguin
Little penguin
Yellow-eyed penguin
Erect-crested penguin
Fiordland penguin
Northern rockhopper penguin
Southern rockhopper penguin
Eastern rockhopper penguin
Macaroni penguin
Emperor penguin
King penguin
20
15
10
5
0
Millions of years ago

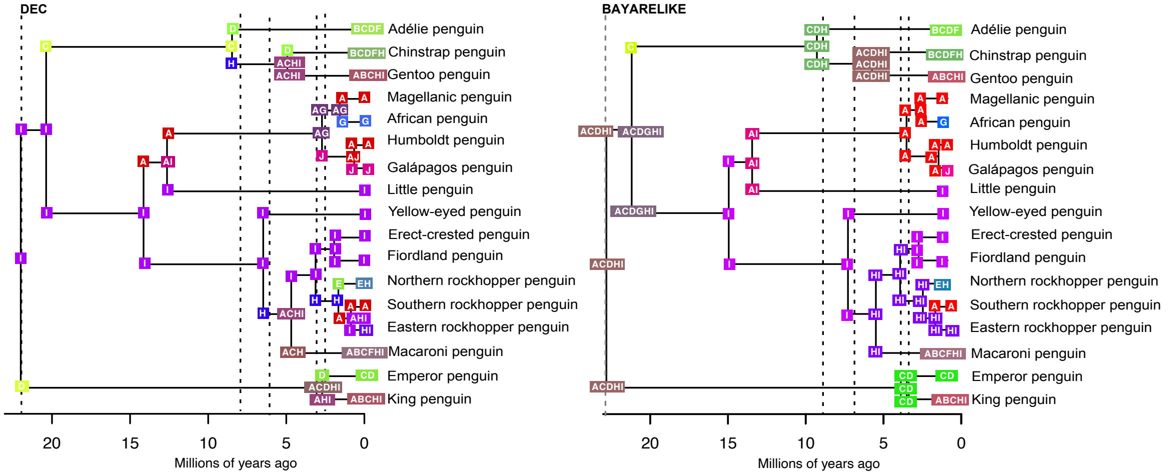


Fig S11. Historical biogeographic analysis using BioGeoBEARS. Results of the six models: DIVALIKE, DEC, BAYAREALIKE with and without the parameter j, which attempts to model speciation through a founder-event. The letters at the nodes indicate the distribution according to the accompanying map: A-South America coast, B-Scotia Arc Islands, C-Antarctic Peninsula, D-Continental Antarctica (except Peninsula), E-Gough Island, F-Bouvet Island, G- Southern African coast, H-islands in the Indian Ocean and Amsterdam Island, I-New Zealand and Australia region, and J-Galápagos Island. The models are shown according to the likelihood, with the best model at the top left (DIVALIKE + j; Fig. 1), followed by DEC+j, DIVALIKE, BAYAREALIKE+j, DEC, BAYAREALIKE. DIVALIKE and DEC suggest an ancestral distribution along the coasts of New Zealand and Australia as suggested by the optimal model (DIVALIKE + j; Fig. 1), whereas the models DEC+J and BAYAREALIKE suggest a broad ancestral area that includes New Zealand and Australia but also South America, Scotia Arc, Antarctic Peninsula, Antarctica, Indian Ocean and Amsterdam Island. The BAYAREALIKE+j model recovers a very different ancestral area prediction that encompasses: South America, the Antarctic Peninsula, and islands in the Indian Ocean and Amsterdam Island.

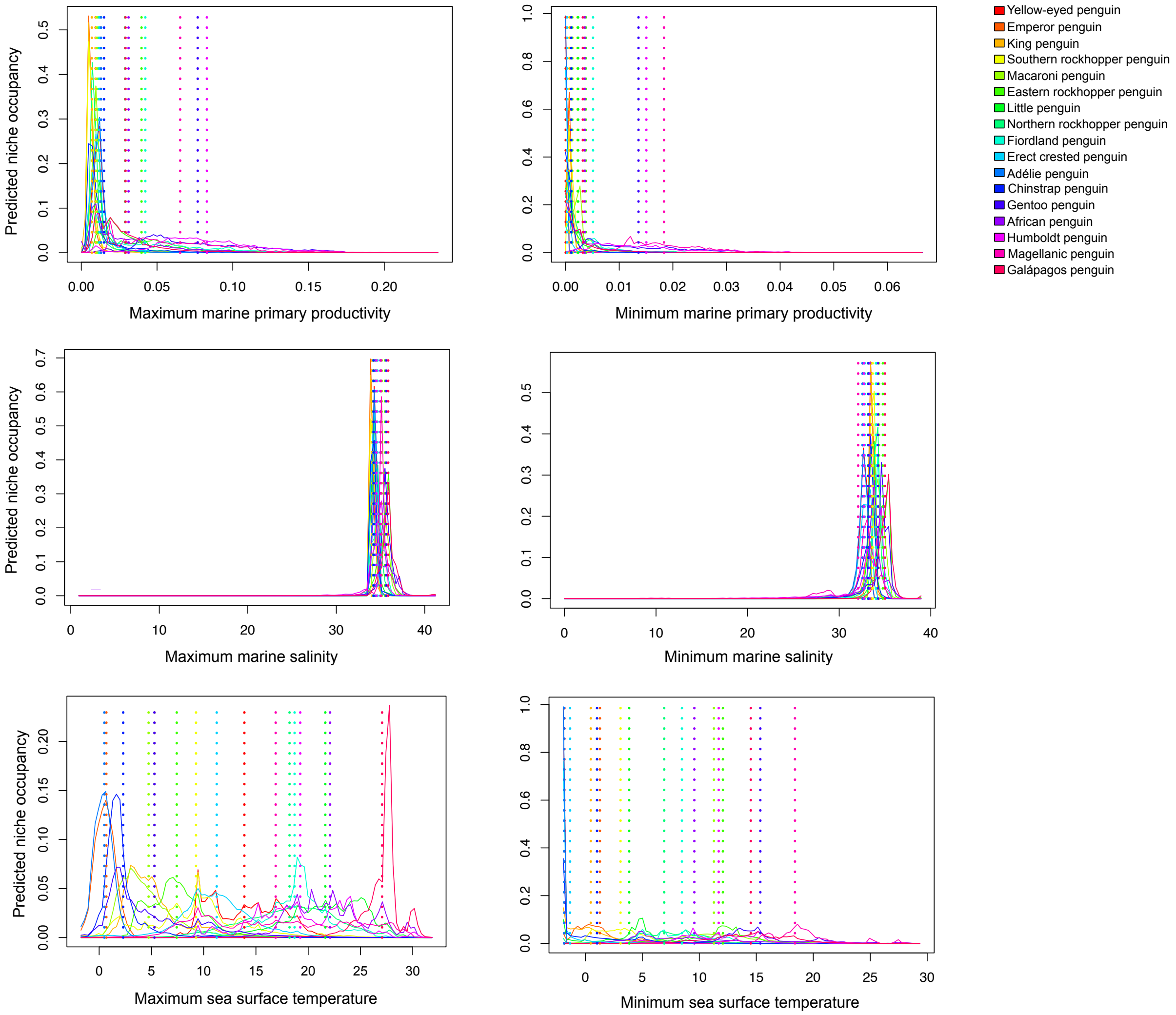


Fig S12. Predicted Niche Occupancy (PNO) profiles for 18 penguin species. The x-axes represent the six evaluated bioclimatic variables (maximum and minimum marine temperature, maximum and minimum salinity, and maximum and minimum primary productivity); the y-axes describe the degree of suitability (niche occupancy) for each penguin species. Overlapping peaks in niche occupancy suggest similar bioclimatic tolerances, and the breadth of the curve indicates the bioclimatic tolerance specificity. Each species is indicated by a different color in the caption. The dashed lines are the average values for each species. The plot for Salinity shows greater overlap among penguin species than primary productivity. Penguins that occupy areas that experience extreme temperatures (Antarctica and Galápagos) behave as stenothermal species, compared to the eurythermal species distributed across an intermediate range of min and max temperatures.

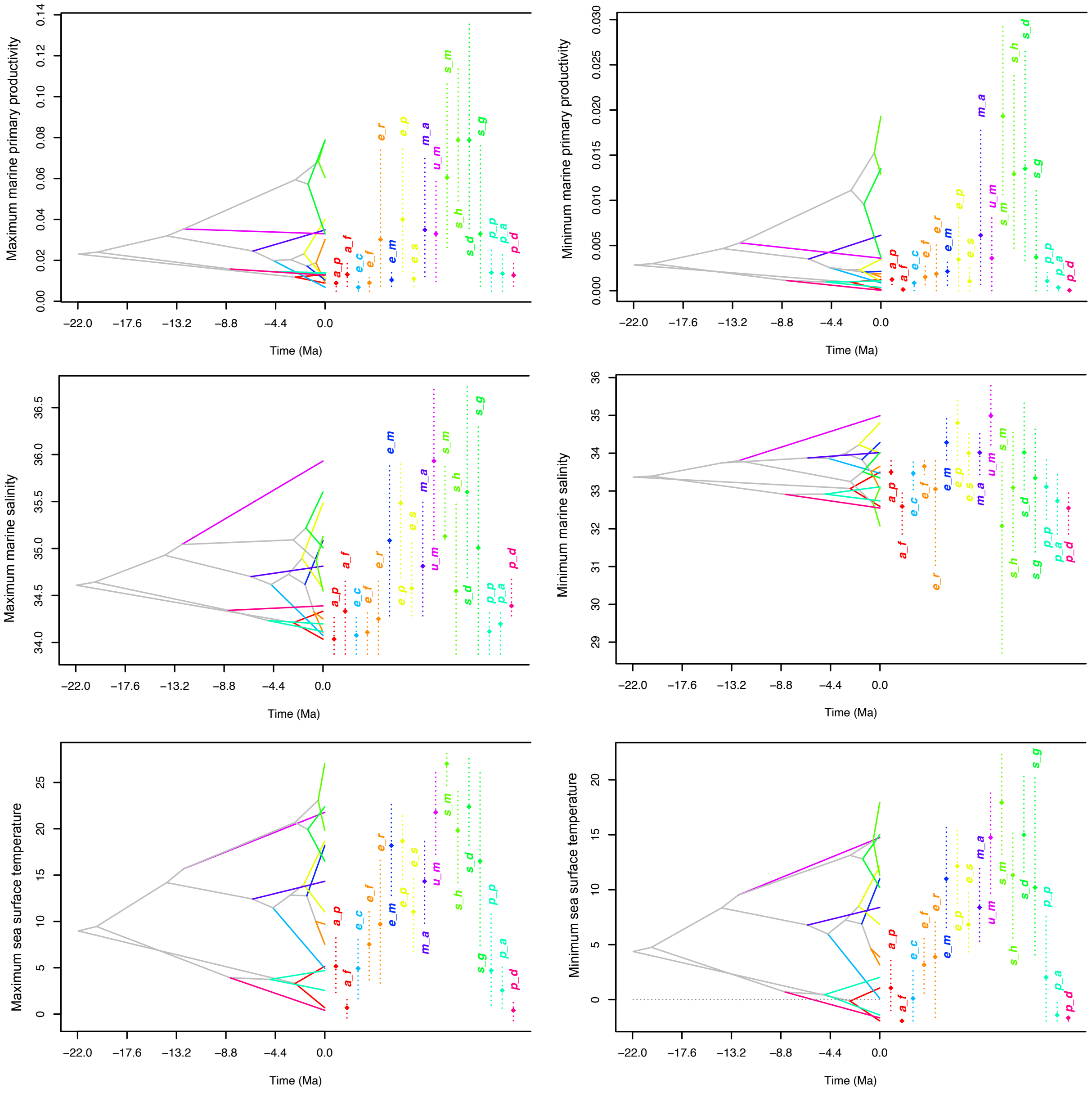


Fig S13. Ecological niche disparity through time (DTT) for 18 penguin species. Our phylogenetic hypothesis of penguin relationships (Fig. 1) is projected into niche parameter space represented on the y-axis (maximum and minimum temperature, maximum and minimum salinity, and maximum and minimum primary productivity) and the mean environmental tolerance of the PNO profiles are reconstructed across the internal nodes of the topology. Vertical dashed line for each species of penguin represent the 80% central density of environmental tolerance (i.e. confidence interval), and the diamond point of the same color represents the mean. Key: king (a_p), emperor (a_f), gentoo (p_p), chinstrap (p_a), Adélie (p_d), macaroni (e_c), eastern rockhopper (e_f), southern rockhopper (e_r), fiordland (e_p), erect-crested (e_s), yellow-eyed (m_a), little (u_m), Galápagos (s_m), Humboldt (s_h), African (s_d), Magellanic (s_g) penguins. Convergent niche evolution among different penguin species from distinct clades are suggested by crossing branches in the topology, whereas similar climatic origins are suggested by internal nodes that overlap.

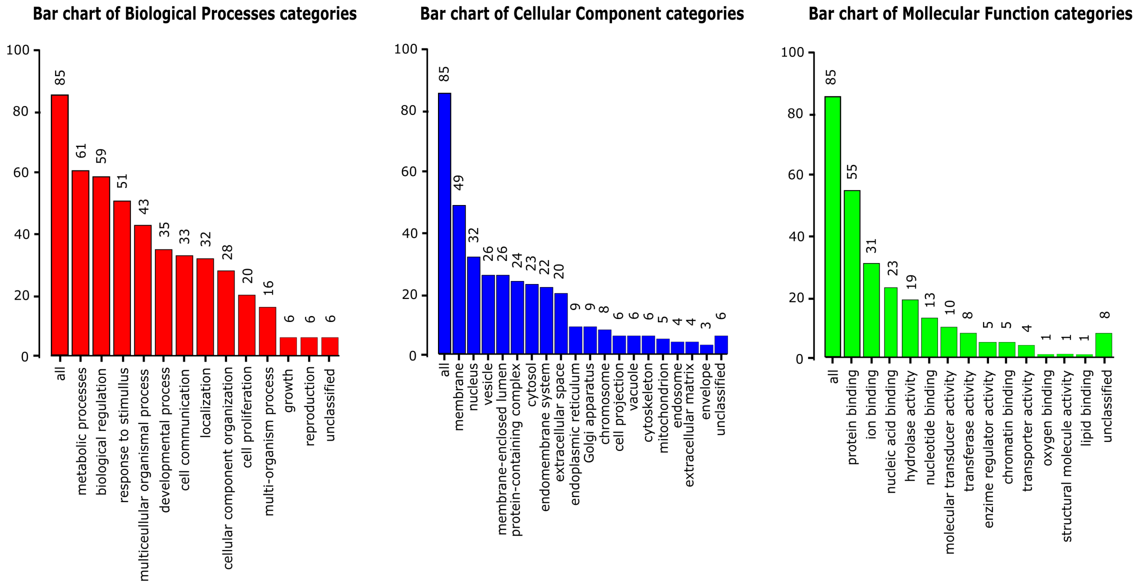


Fig S14. Classification of genes detected to be under positive selection by functional category. From our initial dataset of 104 genes under positive selection, after classification using Gene Ontology a total of 80 genes were determined to represent: 135 biological processes, 85 cellular components, and 48 protein processes. The remaining 24 genes do not show any clear relationship to the categories presently represented in gene ontology databases.

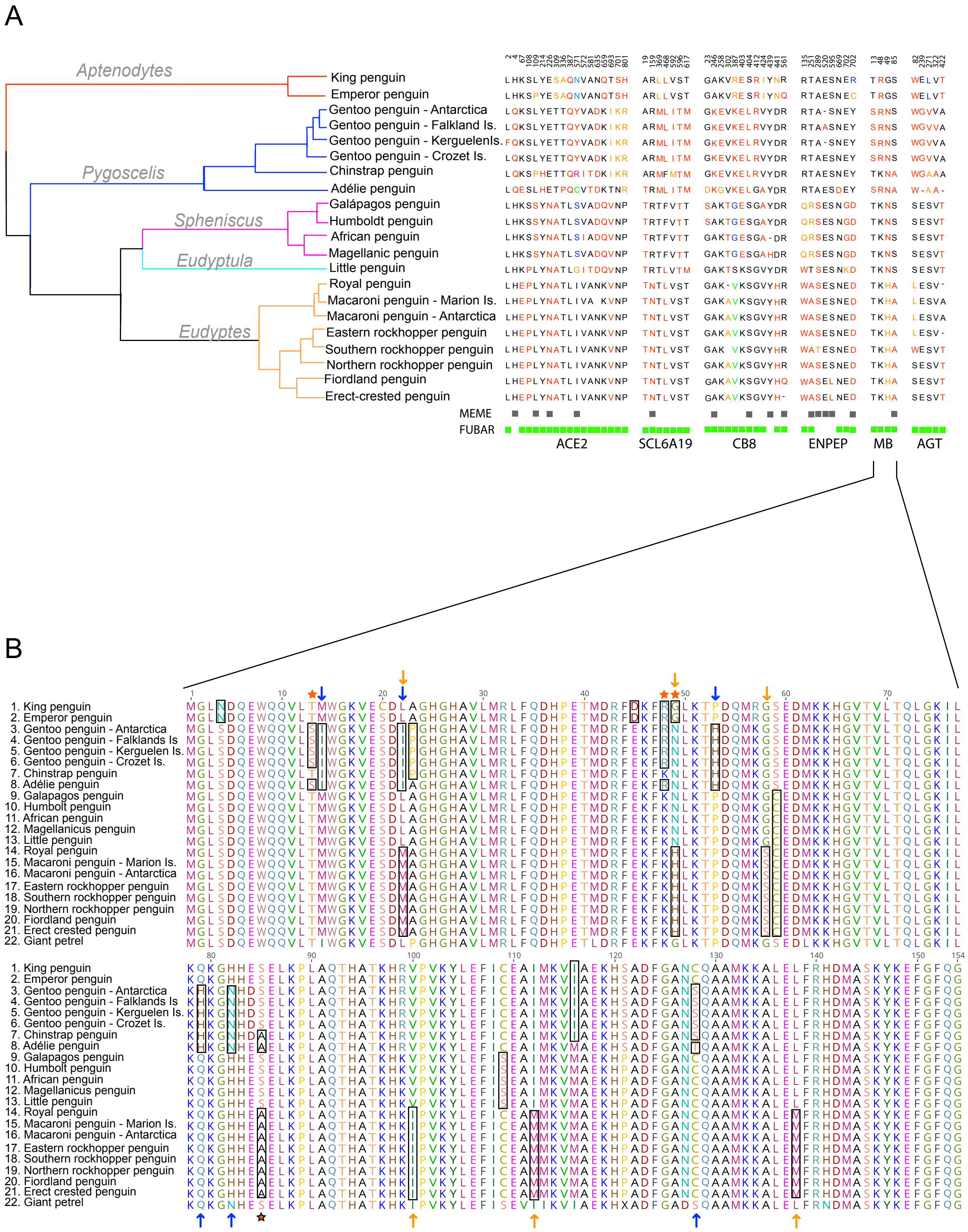
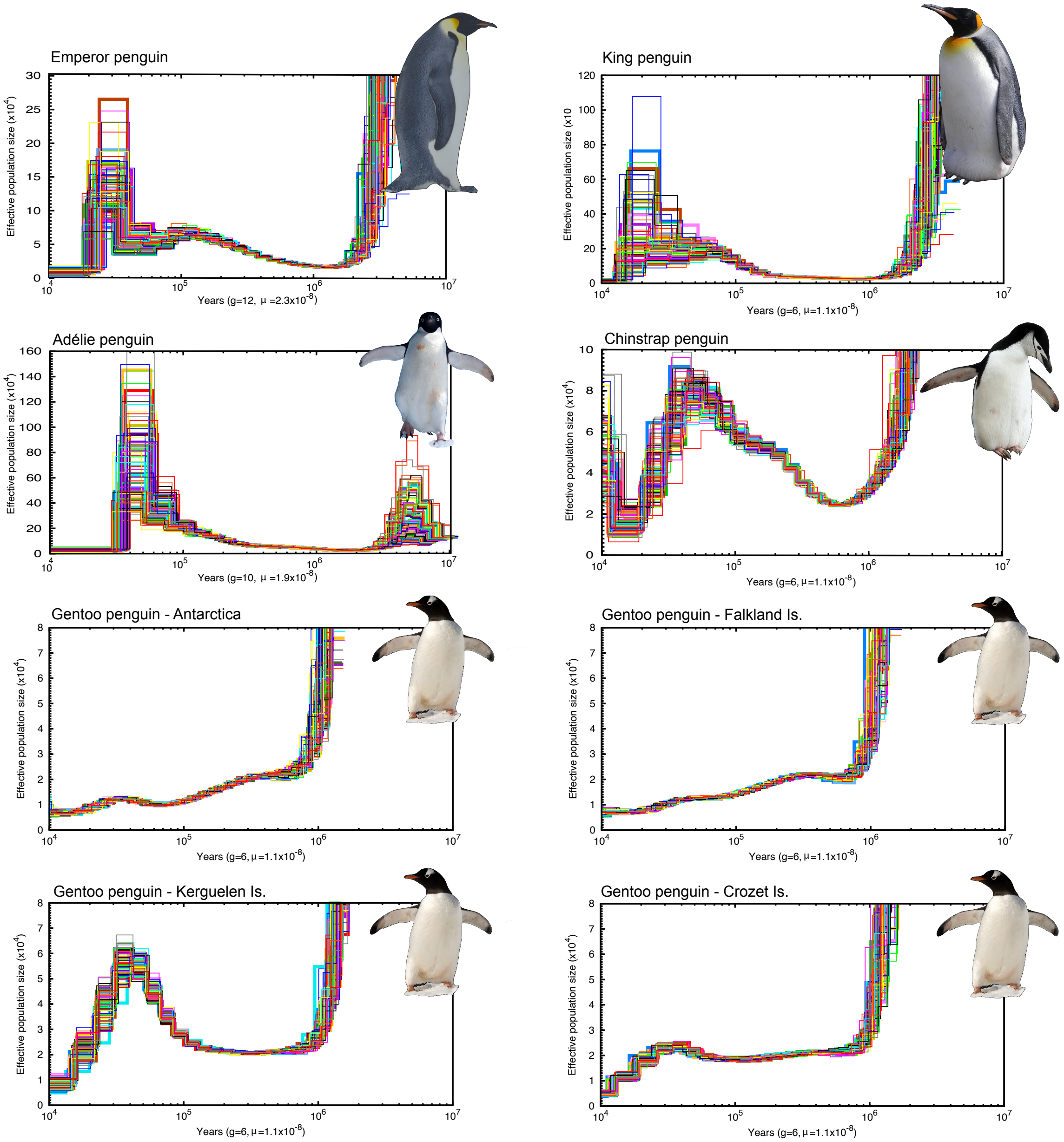
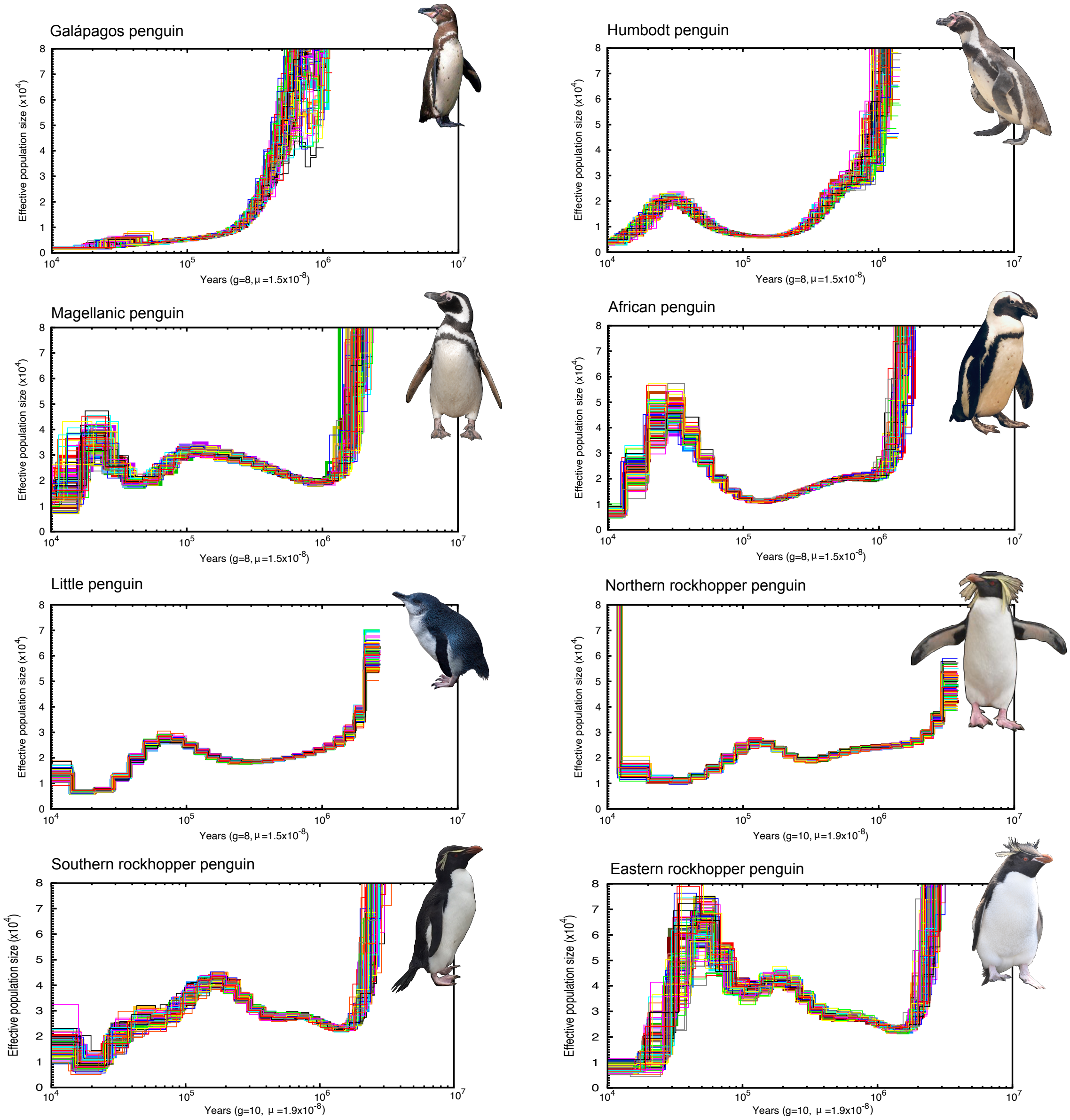

A
Aptenodytes
Pygoscelis
Spheniscus
Eudyptula
Eudyptes
King penguin
Emperor penguin
Gentoo penguin - Antarctica
Gentoo penguin - Falkland Is.
Gentoo penguin - KerguelenIs.
Gentoo penguin - Crozet Is.
Chinstrap penguin
Adélie penguin
Galápagos penguin
Humboldt penguin
African penguin
Magellanic penguin
Little penguin
Royal penguin
Macaroni penguin - Marion Is.
Macaroni penguin - Antarctica
Eastern rockhopper penguin
Southern rockhopper penguin
Northern rockhopper penguin
Fiordland penguin
Erect-crested penguin
MEME
FUBAR
ACE2
SCL6A19
CB8
ENPEP
MB
AGT
B
1. King penguin
2. Emperor penguin
3. Gentoo penguin - Antarctica
4. Gentoo penguin - Falklands Is
5. Gentoo penguin - Kerguelen Is.
6. Gentoo penguin - Crozet Is.
7. Chinstrap penguin
8. Adélie penguin
9. Galapagos penguin
10. Humbolt penguin
11. African penguin
12. Magellanicus penguin
13. Little penguin
14. Royal penguin
15. Macaroni penguin - Marion Is.
16. Macaroni penguin - Antarctica
17. Eastern rockhopper penguin
18. Southern rockhopper penguin
19. Northern rockhopper penguin
20. Fiordland penguin
21. Erect crested penguin
22. Giant petrel
1
10
20
30
40
50
60
70
80
90
100
110
120
130
140
150
154

Fig S15. Alignment of the positively selected sites and amino acids. A-Results of the six genes (*ACE2, SLC6A19, C8B, ENPEP, MB, AGT*) with the lowest p-value for the branch-site results (Table S12) and belonging to the network cluster related to specific functions (e.g. renal function, circulatory system, immunity, Figure 3). Positively selected sites were identified by MEME (posterior probability > 0.90) and FUBAR (p-value ≤ 0.10) for all six genes with the exception of *AGT*, for which sites were identified only by FUBAR. Sites under selection for MEME and FUBAR respectively were: *ACE2*: 4/17; *SLC6A19*: 1/7; *C8B*: 4/13, ENPEP: 5/5, MB: 1/4, AGT: 0/5. The gene under positive selection with the lowest p-value and represented in the network is the angiotensin-converting enzyme 2 (*ACE2*) and its function is related to regulation of cardiovascular and renal function. Such as the genes SLC6A19, *ENPEP*, MB and *AGT*, the function of *ACE2* is related to thermoregulation and/or osmoregulation. Recently ACE2 has been widely studied since it is used by the SARS-COV2 as a cell receptor to invade human cells (108). However, in birds the *ACE2* gene shows very low susceptibility to SARS-COV2 (109). Between the six genes, *ACE2* also showed the higher number of sites under selection (FUBAR: 17 sites and FUBAR and MEME: 3 sites). Another interesting gene is Myoglobin (*MB*) an essential gene for marine mammals and seabirds (110-112), as the increase of oxygen storage is necessary to their diving capacities, but it is also a gene involved in cold adaptation. Shivering and non-shivering thermogenesis promote MB production in king penguin chicks (112-114). For the emperor penguin, MB production occurs in chicks and juveniles, but increases in adults where it appears to be required for exercise/hypoxia of diving activity (115). B- Amino acid alignment of the CDS of Myoglobin (*MB*) across penguins; polymorphic sites of *MB* are indicated by bars around the amino acids. Members of *Pygoscelis* (taxon 3 to 8) share 6 non-synonymous sites (blue arrows) for *MB*, and *Eudyptes* species (taxon 14 to 21) share 6 different non-synonymous sites (orange arrows). These two genera have significantly different dN/dS ratios relative to each other (Table S13). Positive selected sites suggested by FUBAR are indicated with an orange star (Ser-Ala; neutral/polar and large to neutral/nonpolar and small), and the site indicated by both methods is indicated by a black and orange star, which is an amino acid shared between all *Eudyptes* and Adélie and chinstrap penguins, but differing from all remaining penguins.


Emperor penguin
Effective population size (x10^4)
0
5
10
15
20
25
30
10^4
10^5
10^6
10^7
Years (g=12, μ =2.3x10^{-8})
King penguin
Effective population size (x10
0
20
40
60
80
100
120
10^4
10^5
10^6
10^7
Years (g=6, μ=1.1x10^{-8})
Adélie penguin
Effective population size (x10^4)
0
20
40
60
80
100
120
140
160
10^4
10^5
10^6
10^7
Years (g=10, μ =1.9x10^{-8})
Chinstrap penguin
Effective population size (x10^4)
0
2
4
6
8
10
10^4
10^5
10^6
10^7
Years (g=6, μ=1.1x10^{-8})
Gentoo penguin - Antarctica
Effective population size (x10^4)
0
1
2
3
4
5
6
7
8
10^4
10^5
10^6
10^7
Years (g=6, μ=1.1x10^{-8})
Gentoo penguin - Falkland Is.
Effective population size (x10^4)
0
1
2
3
4
5
6
7
8
10^4
10^5
10^6
10^7
Years (g=6, μ=1.1x10^{-8})
Gentoo penguin - Kerguelen Is.
Effective population size (x10^4)
0
1
2
3
4
5
6
7
8
10^4
10^5
10^6
10^7
Years (g=6, μ=1.1x10^{-8})
Gentoo penguin - Crozet Is.
Effective population size (x10^4)
0
1
2
3
4
5
6
7
8
10^4
10^5
10^6
10^7
Years (g=6, μ=1.1x10^{-8})


Galápagos penguin
Humbodt penguin
Magellanic penguin
African penguin
Little penguin
Northern rockhopper penguin
Southern rockhopper penguin
Eastern rockhopper penguin
Effective population size (x10^4)
0
1
2
3
4
5
6
7
8
10^4
10^5
10^6
10^7
Years (g=8, μ =1.5x10^{-8})
Years (g=10, μ =1.9x10^{-8})
Years (g=10, μ =1.9x10^{-8})

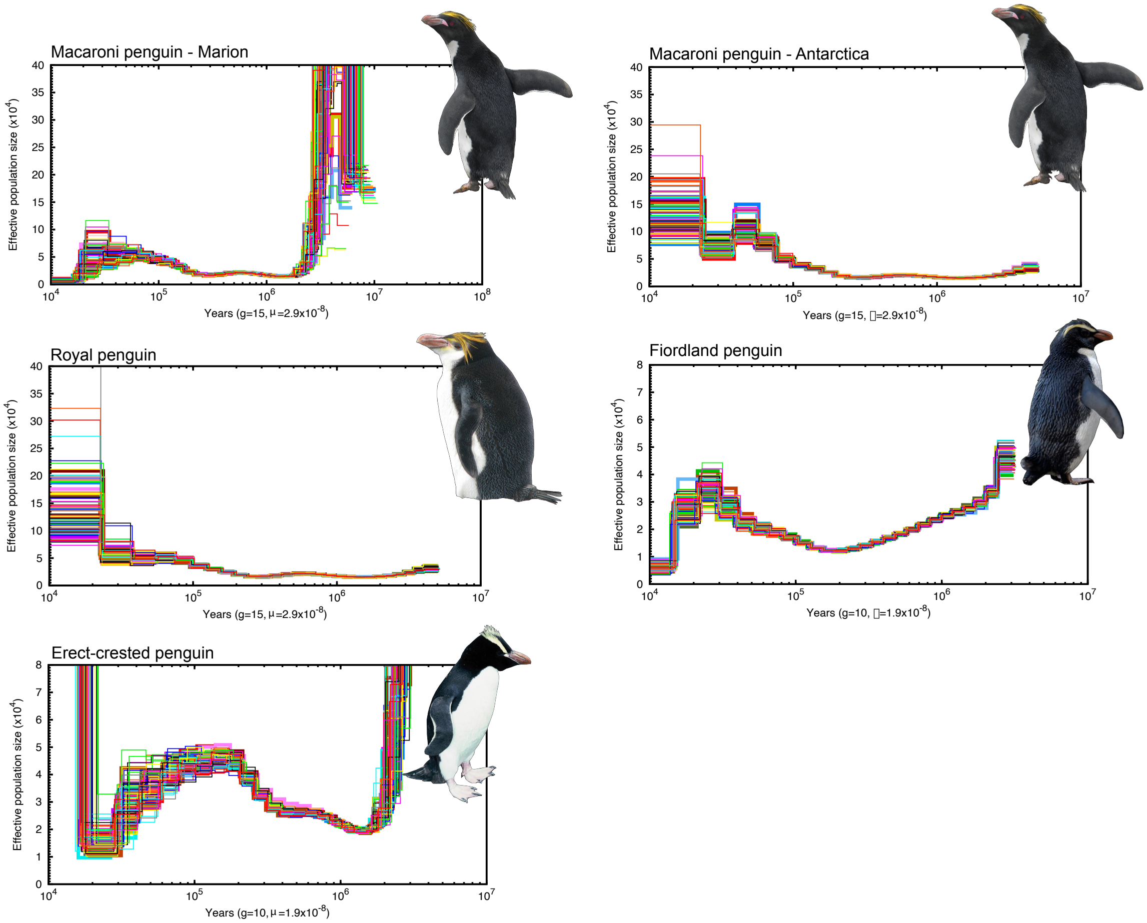


Fig S16. Estimation of Demographic histories using Pairwise Sequentially Markovian Coalescent (PSMC) plots for 21 penguin taxa. The yellow-eyed penguin was excluded from these analyses due to the fragmented nature of its assembled genome. Confidence in the PSMC plots was evaluated using 100 bootstrap replicates. Results were consistent across bootstrap replicates, with some variation around the magnitude of the increase in N_e (y-axis) for Adélie, king, and emperor penguins: penguin species are given above each panel and generation time (g) and mutation rate (μ) used for each penguin taxon are detailed below the x-axis (time). Two different responses are observed across the different penguin species with some species increasing in N_e at ~ 0.04 – 0.07 Mya and some not. High levels of introgression detected among *Eudyptes* species (Fig. S9) could generate an artifact reflected in their demographic history by elevating N_e, therefore the PSMC plot should be interpreted with caution – there are no current PSMC-like methods that account for the influence of introgression. Four divergent lineages of gentoo penguin we sequenced, based on the results of Vianna *et al.* (1), and we recovered different demographic histories. Similarly, results from Frugone et al. (42, 43) suggest that populations of macaroni/royal penguins distributed to the north and south of the Antarctic Polar Front are genetically distinct; our PSMC results indicate that each of these populations have experienced different demographic histories. These results are consistent with knowledge on the extent of ice sheets during glacial periods (116) that suggest that each lineage would have been differentially impacted during glacial periods. Other studies have evaluated demographic history, frequently over more recent history, using mtDNA (117-121) and SNPs from reduced representation sequencing approaches (e.g. RAD-seq; 122-124). Our results recovered a similar shape to the PSMC curve for the emperor and Adélie penguins as estimated by a previous study (80). Our data suggest that the emperor penguin underwent two periods of population expansion, and during the second expansion event reached maximum N_e more recently than the Adélie penguin. Our estimates of maximum N_e for Adélie and emperor penguins are broadly consistent with those estimated by Li et al. (80): Adélie -

48 * 10^4 (bootstrap 23 to 140 * 10^4) vs. 60 * 10^4 (bootstrap 40 to 80 * 10^4); emperor 12 * 10^4 (bootstrap 7 to 27 * 10^4) vs. 25 * 10^4 (20-24 * 10^4). Cristofari et al. (122) estimated N_e for king and emperor penguins over the past 150,000 years using SNP data, while we evaluated N_e from 10,000 to 0.5-1 Mya. We omitted the first 10,000 years as PSMC is known to have high error rates over this period given how the method makes use of blocks of genome-wide data rather than the site-frequency spectrum to estimate N_e. Cristofari et al. (122) recover a decline in N_e during the LGM for the king penguin and only a single demographic expansion for emperor penguins at c. 120,000 years – results that differ from the PSMC curves we report. The difference between Cristofari et al. (122) and our study is likely a consequence of the two studies using different mutation rates and generation times when estimating the PSMC curves.

Table S1. Details for all penguin taxa sequenced: common name, latin name, location, latitude and longitude, sample source and NCBI biosample code. The museum/aquarium vouchers numbers are: CAS10091 for African penguin, MVZ149367 for yellow-eyed penguin, AMNH17808 for fiordland penguin, AMNH211990 for erected-crested penguin.

Species	Common name	Location	Latitude	Longitude	Sample Source	Biosample code
Aptenodytes patagonicus	King penguin	Inutil Bay, Chile	53º25' S	68º19'' W	blood	SAMN11566628
Aptenodytes forsteri	Emperor penguin	Pointe Géologie, Adélie Land, Antarctica	66º40' S	140º01' E	blood	SAMN11566629
Pygoscelis adeliae	Adélie penguin	Lagotellerie, Antarctica	67º53' S	67º23' W	blood	SAMN11566622
Pygoscelis antarcticus	Chinstrap penguin	Narebski, Barton Peninsula, Antarctica	62º14' S	58º46' W	blood	SAMN11566623
Pygoscelis papua	Gentoo penguin	Antarctica	64º49' S	62º51' W	blood	SAMN11566624
Pygoscelis papua	Gentoo penguin	Falkland/Malvinas	52º20' S	59º21' W	blood	SAMN11566625
Pygoscelis papua	Gentoo penguin	Kerguelen island	49º16' S	70º32' E	blood	SAMN11566626
Pygoscelis papua	Gentoo penguin	Crozet Island	46º25' S	50º24' E	blood	SAMN11566627
Spheniscus mendiculus	Galápagos penguin	Galápagos islands	0º40' S	91º16' W	blood	SAMN11566620
Spheniscus humboldti	Humboldt penguin	Pan de Azucar, Chile	26º09' S	70º40' W	blood	SAMN11566619
Spheniscus magellanicus	Magellanic penguin	Puñihuil, Chiloé, Chile	45º55' S	74º02' W	blood	SAMN11566617
Spheniscus demersus	African penguin	Aquarium, California Academy of Science			blood	SAMN11566618
Eudyptula minor	penguin	Cheyne Island, Western Australia	34º35' S	118º45' E	blood	SAMN11566621
Megadyptes antipodes	Yellow-Eyed penguin	Southern Islands, New Zealand	45º33' S	170º02' E	skin snip & toe-pad	SAMN11566616
Eudyptes chrysolophus	Macaroni penguin	Elephant Island, Antarctica	61º05' S	55º00' W	blood	SAMN11566614
Eudyptes chrysolophus	Macaroni penguin	Marion Island	46º50' S	37º48' E	blood	SAMN11566613
Eudyptes schlegeli	Royal penguin	Macquarie Island, Tasmania	54º29' S	158º56' E	blood	SAMN11566615
Eudyptes pachyrhynchus	Fiordland penguin	Ocean off Albany, Western Australia, Australia	35º03' S	117º57' E	blood	SAMN11566612
Eudyptes sclateri	Erect-crested penguin	Bounty Island, New Zealand	47º45' S	179º02' E	blood	SAMN11566611
Eudyptes moseleyi	Northern Rockhopper penguin	Amsterdam Island	37º50' S	77º31' E	blood	SAMN11566608
Eudyptes chrysocome	Southern rockhopper penguin	Terhalten Island, Chile	55º26' S	67º03' W	blood	SAMN11566610
Eudyptes filholi	Eastern rockhopper penguin	Kerguelen island	49º28' S	69º57' E	blood	SAMN11566609
Macronectes giganteus - outgroup	Giant petrel	Avian Island, Antarctica	67º45' S	68º53' W	blood	SAMN11566630

Table S2. Ecological information for extant penguins, including distribution, latitudinal range, body mass, foraging type, mean foraging depth, diet, and generation time used for Pairwise Sequentially Markovian Coalescent (PSMC) analyses based on the average maximum age at sexual maturity according to the authors listed (reference g) multiplied by a factor of two.

Species	Geographic distribution	Latitude range	Body mass (Kg)		Foraging Type	Mean foraging depth	Diet			Generationtime PSMC	
			Males	Felmales			Fish	Cephalopods	Crustaceans	Years	Reference time
Emperor penguin	Circum-Antarctica	66-77º S	21.9-40.0	22.1-32.0	Offshore	20-200 m	✓✓✓	✓	✓	12	Mougin and van Beveren 1979
King penguin	sub-Antarctica	46-54º S	9.3-13.8	8.4-13.2	Offshore	100-200 m	✓	✓	✓✓✓	6	Garcia-Borboroglu and Boersma 2013
Adélie penguin	Circum-Antarctica	54-77º S	4.2-5.5	3.8-4.8	Offshore	3-98 m	✓✓		✓✓✓	10	Ainley et al. 1983
Chinstrap penguin	Antarctica, Scotia Sea region	53-65º S	3.5-5.1	3.2-4.5	Offshore	31-121m	✓		✓✓✓	6	Garcia-Borboroglu and Boersma 2013
Gentoo penguin	Antarctica, sub-Antarctica	54-65º S	5.5-7.7	4.9-7.2	Inshore	30-90m	✓✓✓	✓✓✓	✓✓✓	6	Williams 1995
Galápagos penguin	Galápagos	0º S	1.9-2.6	1.4-2.1	Inshore	6-52 m	✓✓✓	✓	✓	8	Garcia-Borboroglu and Boersma 2013
Humboldt penguin	Peru, Chile	11-41º S	4.7-5.0	4.9-4.1	Inshore	< 30m	✓✓✓	✓✓✓	✓✓✓	8	Araya 2000
Magellanic penguin	Argentina, Chile	41-55º S	4.0-5.2	3.2-4.4	Offshore	30-90m	✓✓✓	✓✓	✓✓	8	Garcia-Borboroglu and Boersma 2013
African penguin	South Africa	32-34º S	3.1-3.6	2.7-3.3	Inshore	30-85m	✓✓✓	✓		8	Crawford et al. 1999
Little penguin	Australia, New Zealand	32-47º S	1.0-1.3	0.8-1.2	Either		✓✓✓	✓✓	✓✓	8	Mougin and van Beveren 1979; Perman and Steen 2000
Yellow-Eyed penguin	New Zealand, Auckland Is.	43-52º S	3.9-8.9	3.6-8.4	Inshore	40-120 m	✓✓✓	✓✓	✓	6	Richdale 1957
Fiordland penguin	New Zealand	43-47º S	2.4-3.9	2.2-3.1	Offshore		✓	✓✓✓	✓✓	10	Garcia-Borboroglu and Boersma 2013
Erect-crested penguin	Antipodes Is., Bountry Is.	47-49º S	4.9-5.5	4.8-5.4	Offshore					15	Garcia-Borboroglu and Boersma 2013
Southern rockhopper penguin	South America, Falkland Is.	50-54º S	2.0-2.8	1.5-2.8	Either	14m	✓✓✓	✓✓✓	✓✓✓	10	Garcia-Borboroglu and Boersma 2013
Northern Rockhopper penguin	sub-Antarctica	37º S	2.6-3.3	2.8-3.5	Offshore		✓✓	✓	✓✓✓	10	Guinard et al. 1998, Simeone et al. Unpublish
Eastern rockhopper penguin	sub-Antarctica (Indian Ocean Is and Pacific Ocean)	46-54º S	2.1-3.2	2.0-3.2	Offshore	40m	✓✓✓	✓✓✓	✓✓✓	10	Garcia-Borboroglu and Boersma 2013
Macaroni/Royal penguins	North Antarctic Peninsula, sub-Antarctica (Indian Ocean Is and Pacific Ocean)	54-61º S	3.1-6.6	2.8-6.3	Offshore	20-35 m	✓✓✓	✓	✓✓	15	Croxall and Davis, 1999

Table S3. Summary of penguin genome sequencing results and assembly quality metrics. The genomes of 22 penguins from 18 species were sequenced (Table S1). We assembled the genomes of all penguins to ~30x coverage using emperor penguin as a reference (66).

Species	Coverage	No. of contigs	Number of contigs in scaffolds	Number of contigs not in scaffolds	Longest contig	No. of contigs > 1000 bp	Total size of contigs	N50 (bp)
Macaroni penguin-Antarctica	29	160,228	159,596	632	124,946	115,272	1,163,517,104	15,603
Macaroni penguin-Marion	31	141,188	140,520	668	130,906	105,255	1,167,467,463	17,518
Royal penguin	34	139,221	138,529	692	154,947	103,813	1,167,662,777	17,753
Southern rockhopper penguin	34	145,468	144,774	694	147,757	105,657	1,168,557,712	17,460
Eastern rockhopper penguin	35	139,802	139,114	688	154,723	104,511	1,167,742,180	17,645
Northern Rockhopper penguin	29	167,064	166,589	475	109,281	119,787	1,148,990,222	14,524
Fiordland penguin	26	168,080	167,602	478	138,629	121,119	1,148,260,405	14,333
Erect-crested penguin	18	167,276	166,817	459	115,472	120,523	1,149,679,527	14,458
Yellow-Eyed penguin	6	2,152,762	2,152,688	74	20,154	240,165	1,011,762,526	749
Little penguin	28	169,972	169,503	469	133,945	121,901	1,147,664,646	14,220
African penguin	26	169,351	168,896	455	133,931	121,828	1,148,419,732	14,220
Humboldt penguin	26	163,183	162,727	456	129,040	119,504	1,149,336,756	14,582
Magellanic penguin	31	157,810	157,326	484	133,949	116,894	1,150,338,177	14,979
Galápagos penguin	31	171,436	170,980	456	110,226	123,722	1,147,203,018	13,916
Adélie penguin	30	173,296	172,858	438	110,227	122,170	1,144,972,554	14,117
Chinstrap penguin	67	155,353	154,862	491	114,170	113,214	1,150,431,292	15,547
Gentoo penguin-Crozet I.	30	158,326	157,848	478	110,226	117,882	1,149,081,850	14,811
Gentoo penguin-Kerguelen I.	32	159,641	159,164	477	110,234	118,344	1,148,959,576	14,743
Gentoo penguin-Falkland I.	30	160,673	160,203	470	136,813	118,594	1,149,024,085	14,696
Gentoo penguin-Antarctica	29	160,551	160,078	473	110,222	118,431	1,149,169,261	14,737
King penguin	32	120,609	119,815	794	162,643	94,268	1,173,234,245	20,065
Emperor penguin	25	134,943	134,351	592	138,667	105,539	1,156,469,197	17,049
Giant Petrel	23	373,028	372,834	194	77,036	201,876	1,115,383,760	7,295

Table S4. Results of benchmarking genome quality using the Universal Single-Copy Orthologs (BUSCO) dataset. Percentage of complete (C), fragmented (F), and missing (M) vertebrate single-copy orthologs are detailed. Genome sequences had high completeness scores (over 90%, with the exception of the yellow-eyed penguin (63.7%) sequenced from a skin-snip taken from a museum specimen).

Species	Location	% C	% S	% D	%F	% M	Complete BUSCOs (C)	Complete and single-copy BUSCOs (S)	Complete and duplicated BUSCOs (D)	Fragmented BUSCOs (F)	Missing BUSCOs (M)	Total BUSCO searched
Emperor penguin	Adélie Land, Antarctica	92.9	92	0.9	4.7	2.4	4564	4520	44	231	120	4915
King penguin	Inutil Bay, Chile	93.1	92.6	0.9	4.2	2.6	4595	4551	44	204	116	4915
Adélie penguin	Lagotellerie, Antarctica	90	89.1	0.9	6.7	3.3	4424	4380	44	328	163	4915
Chinstrap penguin	Narebski, Antarctica	92.8	91.9	0.9	4.6	2.6	4563	4517	46	226	126	4915
Gentoo penguin	Crozet	93	92	1	4.2	2.8	4574	4524	50	208	133	4915
Gentoo penguin	Kerguelen Island	93	92.1	0.9	4.4	2.6	4572	4528	44	217	126	4915
Gentoo penguin	Antarctica	93	92.1	0.9	4.5	2.5	4568	4525	43	222	125	4915
Gentoo penguin	Falkland/Malvinas Island	93.2	92.2	1	4.3	2.5	4582	4533	49	211	122	4915
African penguin	Aquarium Calacademy	93.1	92.2	0.9	4.4	2.5	4577	4534	43	218	120	4915
Humboldt penguin	Pan de Azucar, Chile	93.2	92.3	0.9	4,5	2.3	4585	4539	46	219	111	4915
Magellanic penguin	Puñihuil, Chiloe	93.1	92.2	0.9	4.3	2.6	4578	4533	45	211	126	4915
Galápagos penguin	Galápagos Island	93.4	92,5	0.9	4.4	2.2	4587	4545	42	214	114	4915
Little penguin	Cheyne Island, Australia	93.3	92.4	0.9	4.5	2.2	4583	4541	42	221	111	4915
Yellow-Eyed penguin	Southern Island, New Zealand	63.7	63.4	0.3	12.8	23,5	3132	3117	15	631	1152	4915
Southern rockhopper penguin	Terhalten Island	93.8	92.9	0.9	3.9	2.3	4612	4567	45	192	111	4915
Macaroni penguin	Elephant island, Antarctica	93.4	92,5	0.9	4.5	2.1	4589	4544	45	219	107	4915
Macaroni penguin	Marion Island	93.8	92.9	0.9	4	2.2	4612	4566	46	195	108	4915
Eastern rockhopper penguin	Kerguelen Island	93.8	92.9	0.9	3.9	2.3	4607	4564	43	191	117	4915
Northern Rockhopper penguin	Amsterdam Island	93.1	92.2	0.9	4.3	2.6	4578	4533	45	210	127	4915
Fiordland penguin	Western Australia	93.4	92.6	0.8	4.4	2.2	4589	4550	39	216	110	4915
Royal penguin	Macquarie Island	94	93.2	0.8	4	2	4622	4582	40	196	97	4915
Erect-crested penguin	Bounty Island, New Zealand	92.6	91.7	0.9	4.8	2.6	4553	4509	44	238	124	4915

Table S5. The total number of introns, CDS, and UCE loci identified for each penguin genome and the outgroup, with an average of 13,460 introns, 16,039 CDS and 4,817 UCEs.

Species	Location	Intron	CDS	UCE
Emperor penguin	Adélie Land, Antarctica	13504	16033	4823
King penguin	Inutil Bay, Chile	13513	16054	4806
Adélie penguin	Lagotellerie, Antarctica	13502	16023	4819
Chinstrap penguin	Narebski, Antarctica	13504	16030	4817
Gentoo penguin	Crozet	13503	16030	4828
Gentoo penguin	Kerguelen Island	13504	16034	4823
Gentoo penguin	Antarctica	13503	16031	4822
Gentoo penguin	Falkland/Malvinas Island	13505	16034	4826
African penguin	Aquarium Cal. Acad. Sci.	13501	16027	4818
Humboldt penguin	Pan de Azucar, Chile	13503	16030	4820
Magellanic penguin	Puñihuil, Chiloe	13503	16034	4823
Galápagos penguin	Galápagos Island	13503	16028	4821
Little penguin	Cheyne Island, Australia	13505	16032	4957
Yellow-Eyed penguin	Southern Island, New Zealand	12453	16059	4503
Southern rockhopper penguin	Terhalten Island	13513	16065	4804
Macaroni penguin	Elephant island, Antarctica	13511	16048	4845
Macaroni penguin	Marion Island	13513	16058	4809
Eastern rockhopper penguin	Kerguelen Island	13514	16060	4809
Northern Rockhopper penguin	Amsterdam Island	13505	16034	4863
Fiordland penguin	Western Australia	13504	16033	4861
Royal penguin	Macquarie Island	13514	16060	4849
Erect-crested penguin	Bounty Island, New Zealand	13506	16036	4825
Giant petrel	Avian Island, Antarctica	13504	16033	4836

Table S6. Summary of loci used for phylogenetic analyses: the total length of the alignment (bp), the number of loci, and the average length of each locus for CDS, Intron and UCE loci, respectively (loci with extensive missing data omitted – see Table S5 for unfiltered loci). For the intron loci, each intron locus is comprised of all introns in a gene concatenated together as they are not independent of each other.

All taxa	**CDS**	**Intron**	**UCE**
total length of alignment (bp)	6999894	63106707	2929733
number of loci	4668	2610	1825
average length of each locus (bp)	1499.5	24178.8	1605.3
Omitting yellow-eyed penguin			
total length of alignment (bp)	17461296	202550995	6491478
number of loci	11011	8040	4057
average length of each loci (bp)	1585.8	25192.9	1600.1

Table S7. Results of the D-FOIL test for a total of eight different combinations of 5-taxon statements. The p-values for the four D-FOIL components, DFO, DIL, DFI, DOL (F:First, O:Outer, I:Inner, L:Last) are detailed.

Dataset	Combination	DFO (p-value)	DIL (p-value)	DFI (p-value)	DOL (p-value)
1	Eastern rockhopper, northern rockhopper, fiordland, yellow-eyed penguin	-0.202 (0.140)	-0.199 (0.142)	0.152 (0.211)	0.156 (0.207)
2	Southern rockhopper, northern rockhopper, fiordland, erect-crested, yellow-eyed penguin	-0.203 (0.139)	-0.199 (0.141)	0.119 (0.230)	0.123 (0.227)
3	Eastern rockhopper, southern rockhopper, fiordland, erect-crested, yellow-eyed penguin	-0.206 (0.138)	-0.207 (0.139)	0.064 (0.290)	0.062 (0.288)
4	Royal, macaroni, eastern rockhopper, northern rockhopper, yellow-eyed penguin	0.150 (0.207)	0.150 (0.206)	-0.027 (0.343)	-0.028 (0.345)
5	Royal, macaroni, eastern rockhopper, southern rockhopper, yellow-eyed penguin	0.053 (0.298)	0.053 (0.296)	-0.027 (0.344)	-0.027 80.344)
6	Royal, macaroni, fiordland, erect-crested, yellow-eyed penguin	-0.207 (0.134)	-0.207 (0.135)	-0.030 (0.353)	-0.030 (0.352)
7	Gentoo-Antarctica, gentoo-Falkland, gentoo-Kerguelen, gentoo-Crozet, chinstrap penguin	0.396 (0.139)	0.375 (0.143)	-0.016 (0.245)	-0.040 (0.263)
8	Galápagos, Humboldt, African, Magellanic, little penguin	0.198 (0.235)	-0.199 (0.235)	-0.237 (0.254)	-0.273 (0.253)

Table S8. Details of the five fossil taxa used to calibrate divergence time across crown-group penguins, including the date and age interval (Mya) of the fossils, and the node (Mya) at which the calibration point was applied.

Node	Age interval (Mya)	Fossil calibration taxa
Sphenisciformes/Procellariiformes (root)	60.5 – 72.1	*Waimanu manneringi*
Spheniscidae	9.7 – 25.2	*Madrynornis mirandus*
Pygoscelis	6.3 – 25.2	*Pygoscelis calderensis*
Spheniscus/Eudyptula	9.2 – 23.03	*Spheniscus muizoni*
Eudyptes/Megadyptes	3.06 – 25.2	*Eudyptes sp.*

Table S9. Models tested using BioGeoBEARS with summary statistics: LnL indicates the log-likelihood of the model given our data; numparams indicates the number of parameters in the analyses; columns d, e and j indicate the likelihood of their respective parameters; AICc indicates the corrected Akaike information criterion, and AICc_wt the weighted AICc of the tested models. DIVALIKE+J was the model selected and is indicated by an asterisk.

Model	LnL	numparams	d	e	j	AICc	AICc_wt
DEC	-78.47	2	0.026	0.017	0	161.8	0.072
DEC+J	-75.68	3	0.015	0.0009	0.093	159.2	0.26
DIVALIKE	-77.26	2	0.027	0.010	0	159.4	0.24
DIVALIKE+J*	-75.23	3	0.022	5.5e-08	0.058	158.3	0.41
BAYAREALIKE	-88.1	2	0.031	0.12	0	181.1	4.7e-06
BAYAREALIKE+J	-78.44	3	0.020	0.034	0.12	164.7	0.016

Table S10. Number of coding sequence regions available for each species after filtering for premature stop codons. We used 4,562 CDS loci in our selection analyses (see methods) that overlapped all species in our dataset and for which a start and stop codon were present at the beginning and end of the CDS, respectively.

Species	Location	CDS	Average Sequence Length
Emperor penguin	Adélie Land, Antarctica	5470	1592.64
King penguin	Inutil Bay, Chile	5403	1615.64
Adélie penguin	Lagotellerie, Antarctica	5372	1609.52
Chinstrap penguin	Narebski, Antarctica	5404	1613.58
Gentoo penguin	Crozet	5443	1607.57
Gentoo penguin	Kerguelen Island	5440	1607.58
Gentoo penguin	Antarctica	5446	1607.57
Gentoo penguin	Falkland/Malvinas Island	5442	1607.57
African penguin	Aquarium Calacademy	5400	1597.58
Humboldt penguin	Pan de Azucar, Chile	5404	1601.59
Magellanic penguin	Puñihuil, Chiloe	5424	1614.61
Galápagos penguin	Galápagos Island	5391	1612.60
Little penguin	Cheyne Island, Australia	5410	1606.54
Southern rockhopper penguin	Terhalten Island	5454	1610.64
Macaroni penguin	Elephant island, Antarctica	5362	1613.57
Macaroni penguin	Marion Island	5447	1613.63
Eastern rockhopper penguin	Kerguelen Island	5458	1611.64
Northern Rockhopper penguin	Amsterdam Island	5451	1608.60
Fiordland penguin	Western Australia	5430	1597.58
Royal penguin	Macquarie Island	5446	1610.64
Erect-crested penguin	Bounty Island, New Zealand	5448	1598.39
Giant petrel	Avian Island, Antarctica	5387	1591.37

Table S11. List of enriched gene sets: Gene ontology analyses for the positively selected genes.

Enriched gene set	N	FDR
Angiotensin maturation	3	0.034638
Regulation of angiotensin levels in blood	3	0.036825
Regulation of systemic arterial blood pressure mediated by a chemical signal	4	0.055572
Positive regulation of reactive oxygen species metabolic process	5	0.055572
Double-strand break repair	7	0.055572
Regulation of innate immune response	9	0.055572
Regulation of defense response	14	0.015593
Regulation of response to stress	20	0.015593

Table S12. Branch-site results for selection across the sampled penguin species. Candidate genes under positive selection, p-values for the gene (significance threshold: alpha<0.05) and gene description based on Gene Ontology and GeneCards. Genes are ordered based on the p-values recovered. The 46 genes present in the network (Fig. 3) which show two or more connections in function are indicated by an asterisk.

Contig	p-value	q-value	Gene name	Description of the gene or encoded protein (from geneCards)
Contig001918	0	0	*CD8A*	CD8 antigen mediates efficient cell-cell interactions in the immune system
Contig007559	0	0	*ACE2**	Family of dipeptidyl carboxydipeptidases, regulation of cardiovascular and renal function
Contig007985	0	0	*GUCY2C*	Transmembrane protein that functions as a receptor for endogenous peptides guanylin and uroguanylin
Contig007990	0	0	*EMP1**	Epitelial Membrane Protein
Contig008215	0	0	*RNASEL**	Component in the interferon-regulated 2-5A system that functions in the antiviral and antiproliferative roles of interferons
Contig009143	0	0	*LOC103913744*	
Contig009200	0	0	*SMPD3*	Catalyzes ceramide formation. Ceramide mediates numerous cellular functions, such as apoptosis and growth arrest
Contig010445	0	0	*HELZ2**	Nuclear transcriptional co-activator for peroxisome proliferator activated receptor alpha. The encoded protein contains a zinc finger and is a helicase that appears to be part of the peroxisome proliferator activated receptor alpha interacting complex.
Contig010969	0	0	*TICAM1*	Is involved immunity against pathogens
Contig013409	0	0	*KIAA1211*	B(0) transmembrane protein
Contig014755	0	0	*LOC103919005*	
Contig000718	1,E-06	2,E-04	*SLC6A19**	the protein associated to this gene transports certain amino acids into cells. B^0AT1 is found primarily in the membrane of intestinal cells and are associated to nutrients absorbtion from food. In the kidneys, B^0AT1 reabsorbs neutral amino acids into the bloodstream.
Contig008076	1,E-06	2,E-04	*FOXM1**	The protein encoded by this gene is a transcriptional activator involved in cell proliferation.
Contig015452	1,E-06	2,E-04	*RFTN1*	Involved in protein trafficking via association with clathrin and AP2 complex
Contig001071	2,E-06	3,E-04	*C8B**	Codes one of the three subunits of the complement component 8 (C8) protein. C8 is one component of the membrane attack complex, which mediates cell lysis, and it initiates membrane penetration of the complex.
Contig003547	2,E-06	3,E-04	*PPP1R3A*	Act as a glycogen-targeting subunit for PP1. PP1 is involved in cell division, regulation of glycogen metabolism, muscle contractility and protein synthesis.
Contig006407	2,E-06	3,E-04	*MYO15B*	Encode a sodium-activated potassium channel subunit
Contig013786	2,E-06	3,E-04	*KCNT1*	This gene encodes a sodium-activated potassium channel subunit which is thought to function in ion conductance and developmental signaling pathways.
Contig015700	2,E-06	3,E-04	*ENPEP**	Among its related pathways are Agents Acting on the Renin-Angiotensin System Pathway, Pharmacodynamics and Peptide hormone metabolism.
Contig006608	4,E-06	6,E-04	*CALCOCO2*	Xenophagy-specific receptor required for autophagy-mediated intracellular bacteria degradation.
Contig008677	7,E-06	1,E-03	*ABCC10*	The protein encoded by this gene is a member of the superfamily of ATP-binding cassette (ABC) transporters. ABC proteins transport various molecules across extra- and intra-cellular membranes.
Contig007859	9,E-06	1,E-03	*MB**	The encoded protein is a haemoprotein contributing to intracellular oxygen storage and transcellular facilitated diffusion of oxygen
Contig015042	1,E-05	2,E-03	*CORO2B*	May play a role in the reorganization of neuronal actin structure.

Contig008230	2,E-05	2,E-03	*MARCO*	The protein encoded by this gene is a member of the class A scavenger receptor family and is part of the innate antimicrobial immune system.
Contig006706	2,E-05	2,E-03	*LOC103918151*	
Contig012327	2,E-05	3,E-03	*LOC103913537*	
Contig014887	2,E-05	3,E-03	*LOC103912710*	
Contig000873	3,E-05	3,E-03	*MPHOSPH9*	Angiotensinogen precursor, expressed in the liver and is cleaved by the enzyme renin in response to lowered blood pressure.
Contig010691	3,E-05	3,E-03	*LOC103895549*	
Contig001492	3,E-05	3,E-03	*AGT**	The protein encoded by this gene, is expressed in the liver and is cleaved by the enzyme renin in response to lowered blood pressure. Is involved in maintaining blood pressure and in the pathogenesis of essential hypertension and preeclampsia.
Contig007178	3,E-05	4,E-03	*RHBDF1*	Rhomboid protease-like protein which has no protease activity but regulates the secretion of several ligands of the epidermal growth factor receptor. Indirectly activates the epidermal growth factor receptor signaling pathway and may thereby regulate sleep, cell survival, proliferation and migration
Contig004692	4,E-05	4,E-03	*ZBTB42*	The protein encoded by this gene is a member of the C2H2 zinc finger protein family. Transcriptional repressor. Specifically binds DNA and probably acts by recruiting chromatin remodeling multiprotein complexes.
Contig013099	4,E-05	4,E-03	*LOC103907580*	
Contig010126	4,E-05	4,E-03	*LUZP1*	This gene encodes a protein that contains a leucine zipper motif. The exact function of the encoded protein is not known.
Contig004569	4,E-05	4,E-03	*TDP1**	Member of the phospholipase D family. Involved in repairing stalled topoisomerase I-DNA complexes. Has a role in repair of free-radical mediated DNA double-strand breacks
Contig015308	5,E-05	4,E-03	*UPF2**	This gene encodes a protein that is part of a post-splicing multiprotein complex involved in both mRNA nuclear export and mRNA surveillance.
Contig011340	5,E-05	4,E-03	*FUT7*	The protein encoded by this gene is a Golgi stack membrane protein that is involved in the creation of sialyl-Lewis X antigens.
Contig002045	5,E-05	5,E-03	*LOC101876482*	
Contig003143	6,E-05	5,E-03	*USPL1*	Involved in protein desumoylation
Contig010213	6,E-05	5,E-03	*ATP13A3*	ATPase Family, transport of cations across membranes
Contig004635	7,E-05	5,E-03	*EML1*	Echinoderm microtubule-associated protein-like. Modulates the assembly and organization of the microtubule cytoskeleton, and probably plays a role in regulating the orientation of the mitotic spindle and the orientation of the plane of cell division
Contig004893	7,E-05	5,E-03	*CHAF1B**	Required for the assembly of histone octamers onto newly-replicated DNA
Contig006216	7,E-05	6,E-03	*RAD52**	Related to DNA recombination and repair
Contig012901	1,E-04	8,E-03	*FGB**	The protein encoded by this gene is the beta component of fibrinogen, a blood-borne glycoprotein comprised of three pairs of nonidentical polypeptide chains. In addition, various cleavage products of fibrinogen and fibrin regulate cell adhesion and spreading, display vasoconstrictor and chemotactic activities, and are mitogens for several cell types.
Contig014626	1,E-04	8,E-03	*CCDC73*	Cell-surface proteins with role in regulation of cell development, activation, growth and motility
Contig012690	1,E-04	8,E-03	*LOC103913273*	
Contig009719	1,E-04	8,E-03	*TSPAN8**	The proteins mediate signal transduction events that play a role in the regulation of cell development, activation, growth and motility.
Contig002016	0.0001	0.0085	*SLC26A9*	The product of this gene is a highly selective chloride ion channel regulated by WNK kinases. Alternative splicing results in multiple transcript variants encoding differing isoforms.
Contig015942	1,E-04	9,E-03	*IL18RAP**	The protein encoded by this gene is an accessory subunit of the heterodimeric receptor for interleukin 18 (IL18), a proinflammatory cytokine involved in inducing cell-mediated immunity.

Contig004964	1,E-04	9,E-03	*GPR82*	The protein encoded by this gene is an orphan G protein-coupled receptor of unknown function.
Contig011912	2,E-04	1,E-02	*TENM2**	Involved in neural development, regulating the establishment of proper connectivity within the nervous system. Promotes the formation of filopodia and enlarged growth cone in neuronal cells
Contig013612	2,E-04	1,E-02	*AKAP1*	The A-kinase anchor proteins (AKAPs) are a group of structurally diverse proteins, which have the common function of binding to the regulatory subunit of protein kinase A (PKA) and confining the holoenzyme to discrete locations within the cell.
Contig011182	2,E-04	1,E-02	*CDPF1*	Cysteine Rich DPF Motif Domain Containing 1
Contig010454	2,E-04	1,E-02	*RAD21L1**	Meiosis-specific component of some cohesin complex required during the initial steps of prophase I in male meiosis.
Contig005165	2,E-04	1,E-02	*LOC103916848*	
Contig002300	2,E-04	1,E-02	*F2**	Coagulation factor II is proteolytically cleaved to form thrombin in the first step of the coagulation cascade which ultimately results in the stemming of blood loss. F2 also plays a role in maintaining vascular integrity during development and postnatal life. Peptides derived from the C-terminus of this protein have antimicrobial activity against E. coli and P. aeruginosa.
Contig009500	2,E-04	1,E-02	*STOX1**	DNA binding protein
Contig008168	2,E-04	1,E-02	*NPL**	Gene encodes a member of the N-acetylneuraminate lyase sub-family of (beta/alpha)(8)-barrel enzymes.
Contig005680	3,E-04	1,E-02	*RUFY3*	This gene encodes a RPIP8, UNC-14, and NESCA domain-containing protein that is required for maintenance of neuronal polarity and and axon growth
Contig008788	3,E-04	1,E-02	*OSGEPL1**	Involved in mitochondrial genome maintenance.
Contig001645	3,E-04	2,E-02	*SETX**	Protein encoded contain a DNA/RNA helicase domain, involved in DNA/RNA processing
Contig003495	3,E-04	2,E-02	*PLXNB2*	Transmembrane receptors that participate in axon guidance and cell migration in response to semaphorins
Contig013796	3,E-04	2,E-02	*CARD9**	Adapter protein that plays a key role in innate immune response to a number of intracellular pathogens
Contig014194	3,E-04	2,E-02	*KNG1**	This gene uses alternative splicing to generate two different proteins- high molecular weight kininogen (HMWK) and low molecular weight kininogen (LMWK). HMWK is essential for blood coagulation and assembly of the kallikrein-kinin system. Also, bradykinin, a peptide causing numerous physiological effects, is released from HMWK. Bradykinin also functions as an antimicrobial peptide with antibacterial and antifungal activity.
Contig009432	4,E-04	2,E-02	*AFTPH*	May play a role in membrane trafficking.
Contig012154	4,E-04	2,E-02	*IRF1**	The encoded protein activates the transcription of genes involved in the body's response to viruses and bacteria, playing a role in cell proliferation, apoptosis, the immune response, and DNA damage response.
Contig002792	4,E-04	2,E-02	*ACSL5*	Acyl-CoA Synthetase Long Chain Family Member 5. Activate long-chain fatty acids for both synthesis of cellular lipids, and degradation via beta-oxidation
Contig008467	5,E-04	2,E-02	*MME**	The protein encoded by this gene is a type II transmembrane glycoprotein and a common acute lymphocytic leukemia antigen
Contig000726	5,E-04	2,E-02	*PHF3*	This gene encodes a member of a PHD finger-containing gene family. This gene may function as a transcription factor and may be involved in glioblastomas development.
Contig011577	6,E-04	2,E-02	*FYCO1*	The encoded protein plays a role in microtubule plus end-directed transport of autophagic vesicles
Contig005040	6,E-04	2,E-02	*ZNF292*	This gene encodes a growth hormone-dependent, zinc finger transcription factor that functions as a tumor suppressor. May be involved in transcriptional regulation.
Contig004617	6,E-04	2,E-02	*BDKRB2**	This gene encodes a receptor for bradykinin. The 9 aa bradykinin peptide elicits many responses including vasodilation, edema, smooth muscle spasm and pain fiber stimulation.
Contig005588	6,E-04	2,E-02	*ROS1*	This proto-oncogene, highly-expressed in a variety of tumor cell lines. The protein may function as a growth or differentiation factor receptor.
Contig012421	6,E-04	2,E-02	*TRAF7**	Signal transducers for members of the Tumor necrosis factor (TNF) receptor superfamily

Contig002799	6,E-04	2,E-02	*DCLRE1A**	This gene encodes a conserved protein that is involved in the repair of DNA interstrand cross-links. DNA cross-links suppress transcription, replication, and DNA segregation.
Contig007834	7,E-04	3,E-02	*CUL1**	Core component of multiple cullin-RING-based SCF (SKP1-CUL1-F-box protein) E3 ubiquitin-protein ligase complexes, which mediate the ubiquitination of proteins involved in cell cycle progression, signal transduction and transcription.
Contig012841	7,E-04	3,E-02	*LOC103925145*	
Contig004858	7,E-04	3,E-02	*URB1*	Encodes a member of the KH-domain protein subfamily
Contig004776	8,E-04	3,E-02	*PCBP3**	This gene encodes a member of the KH-domain protein subfamily. Proteins of this subfamily, also referred to as alpha-CPs, bind to RNA with a specificity for C-rich pyrimidine regions. Alpha-CPs play important roles in post-transcriptional activities and have different cellular distributions
Contig002592	8,E-04	3,E-02	*ZFAND4**	ZFAND4 (Zinc Finger AN1-Type Containing 4) is a Protein Coding gene. An important paralog of this gene is RPS27A.
Contig008790	9,E-04	3,E-02	*ASNSD1**	Asparagine Synthetase Domain Containing
Contig015747	9,E-04	3,E-02	*ODC1**	Protein homodimerization activity and ornithine decarboxylase activity
Contig013582	9,E-04	3,E-02	*BRIP1**	Member of ResQDEAH helicase family. Involved in the repair of DNA double-strand breaks by homologous recombination
Contig001025	9,E-04	3,E-02	*ORC1**	Component of the origin recognition complex (ORC) involved in the initiation of DNA replication
Contig000277	1,E-03	3,E-02	*LOC103924632*	
Contig001865	1,E-03	4,E-02	*ACD**	Involved in the regulation of telomere length and protection
Contig014199	1,E-03	4,E-02	*AHSG**	It is involved in several processes, including endocytosis, brain development, and the formation of bone tissue
Contig005311	1,E-03	4,E-02	*GTF2IRD1*	May be a transcription regulator involved in cell-cycle progression and skeletal muscle differentiation. plays a role in craniofacial and cognitive development
Contig007073	1,E-03	4,E-02	*GLIPR1**	This gene encodes a protein with similarity to both the pathogenesis-related protein (PR) superfamily and the cysteine-rich secretory protein (CRISP) family
Contig004531	1,E-03	4,E-02	*LOC113480798*	
Contig006655	1,E-03	4,E-02	*PPFIA1*	Obsolete signal transducer activity
Contig008336	1,E-03	5,E-02	*LOC103915646*	
Contig004632	1,E-03	5,E-02	*LOC103920698*	
Contig007136	1,E-03	5,E-02	*LOC105414297*	
Contig014145	1,E-03	5,E-02	*LOC103919172*	
Contig009830	2,E-03	5,E-02	*ZNF518B*	May be involved in transcriptional regulation
Contig004787	2,E-03	5,E-02	*ITGB2**	This gene encodes an integrin beta chain. Integrins participate in cell adhesion and cell-surface mediated signalling. Plays a role in immune response.
Contig000492	2,E-03	5,E-02	*ARHGAP5*	May mediate cytoskeleton changes .
Contig013992	2,E-03	5,E-02	*PABPC4**	Binds the poly(A) tail of mRNA. May be involved in cytoplasmic regulatory processes of mRNA metabolism. PABPC4 was also identified as an antigen.
Contig007716	2,E-03	5,E-02	*TM4SF4**	Regulates the adhesive and proliferative status of intestinal epithelial cells
Contig009580	2,E-03	5,E-02	*LOC103914698*	
Contig014367	2,E-03	5,E-02	*WDHD1**	Brings together MCM2-7 helicase and DNA polymerase alpha/primase complex to initiate DNA replication
Contig004102	2,E-03	5,E-02	*CLK1*	Involved in gene splicing regulation
Contig014022	2,E-03	5,E-02	*CCDC79**	Involved in meiotic telomere attachment to the nucleus inner membrane

Table S13. Codon-based tests of positive selection for Myoglobin (*MB*) between pairwise penguin genomes. Significant Z-values (d_N/d_S test) of selection for *MB* are indicated in dark gray ($p < 0.05$) and near significant values in light gray ($p = 0.055$). *MB* overall positive selection, Z = 2.645, p = 0.005.

	Southern rockhopper	Eastern rockhopper	Erect-crested	Macaroni-Antarctica	Fiordland	Royal	Macaroni-Marion	Emperor	King	Little	Northern rockhopper	African	Galápagos	Humboldt	Magellanic	Gentoo-Antarctica	Gentoo-Kerguelen	Gentoo-Crozet	Gentoo-Falkland	Chinstrap	Adélie
Southern rockhopper																					
Eastern rockhopper	0.000																				
Erect-crested	-1.006	-1.006																			
Macaroni-Antarctica	0.000	0.000	-1.006																		
Fiordland	0.000	0.000	-1.006	0.000																	
Royal	0.000	0.000	-1.006	0.000	0.000																
Macaroni-Marion	0.000	0.000	-1.006	0.000	0.000	0.000															
Emperor	0.911	0.911	0.339	0.911	0.911	0.911	0.911														
King	1.032	1.032	0.451	1.032	1.032	1.032	1.032	1.001													
Little	**1.609**	**1.609**	0.445	**1.609**	**1.609**	**1.609**	**1.609**	-0.063	0.068												
Northern rockhopper	0.000	0.000	-1.006	0.000	0.000	0.000	0.000	0.911	1.032	1.609											
African	**1.609**	**1.609**	0.445	**1.609**	**1.609**	**1.609**	**1.609**	-0.063	0.068	0.000	**1.609**										
Galápagos	**1.609**	**1.609**	0.445	**1.609**	**1.609**	**1.609**	**1.609**	-0.063	0.068	0.000	**1.609**	0.000									
Humboldt	**1.609**	**1.609**	0.445	**1.609**	**1.609**	**1.609**	**1.609**	-0.063	0.068	0.000	**1.609**	0.000	0.000								
Magellanic	**1.609**	**1.609**	0.445	**1.609**	**1.609**	**1.609**	**1.609**	-0.063	0.068	0.000	**1.609**	0.000	0.000	0.000							
Gentoo-Antarctica	**3.549**	**3.549**	**2.234**	**3.549**	**3.549**	**3.549**	**3.549**	0.354	0.479	1.104	**3.549**	1.104	1.104	1.104	1.104						
Gentoo-Kerguelen	**3.549**	**3.549**	**2.234**	**3.549**	**3.549**	**3.549**	**3.549**	0.354	0.479	1.104	**3.549**	1.104	1.104	1.104	1.104	0.000					
Gentoo-Crozet	**3.549**	**3.549**	**2.234**	**3.549**	**3.549**	**3.549**	**3.549**	0.354	0.479	1.104	**3.549**	1.104	1.104	1.104	1.104	0.000	0.000				
Gentoo-Falkland	**3.549**	**3.549**	**2.234**	**3.549**	**3.549**	**3.549**	**3.549**	0.354	0.479	1.104	**3.549**	1.104	1.104	1.104	1.104	0.000	0.000	0.000			
Chinstrap	**3.134**	**3.134**	**1.835**	**3.134**	**3.134**	**3.134**	**3.134**	0.487	0.609	0.965	**3.134**	0.965	0.965	0.965	0.965	**1.739**	**1.739**	**1.739**	**1.739**		
Adélie	**1.814**	**1.814**	1.018	**1.814**	**1.814**	**1.814**	**1.814**	-0.027	0.086	0.314	**1.814**	0.314	0.314	0.314	0.314	**2.252**	**2.252**	**2.252**	**2.252**	**2.470**	

References

1. Pertierra LR, *et al.* (2020) Cryptic speciation in gentoo penguins is driven by geographic isolation and regional marine conditions: Unforeseen vulnerabilities to global change. *Diversity and Distributions* n/a(n/a).
2. Vianna JA, *et al.* (2017) Marked phylogeographic structure of Gentoo penguin reveals an ongoing diversification process along the Southern Ocean. *Mol Phylogenet Evol* 107:486-498.
3. Levy H, *et al.* (2020) Evidence of Pathogen-Induced Immunogenetic Selection across the Large Geographic Range of a Wild Seabird. *Molecular Biology and Evolution* 37(6):1708-1726.
4. Aljanabi SM & Martinez I (1997) Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. *Nucleic Acids Res* 25(22):4692-4693.
5. Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *2011* 17(1):3.
6. Bolger AM, Lohse M, & Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15):2114-2120.
7. Magoc T & Salzberg SL (2011) FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 27(21):2957-2963.
8. Frith MC, Wan R, & Horton P (2010) Incorporating sequence quality data into alignment improves DNA read mapping. *Nucleic Acids Research* 38(7):e100-e100.
9. Li H, *et al.* (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16):2078-2079.
10. Katoh K & Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30(4):772-780.
11. Capella-Gutierrez S, Silla-Martinez JM, & Gabaldon T (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25(15):1972-1973.
12. Bradnam KR, *et al.* (2013) Assemblathon 2: evaluating de novo methods of genome assembly in three vertebrate species. *GigaScience* 2(1).
13. Kriventseva EV, Zdobnov EM, Simão FA, Ioannidis P, & Waterhouse RM (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31(19):3210-3212.
14. Stanke M & Morgenstern B (2005) AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic acids research* 33(Web Server issue):W465-W467.
15. Mapleson D, Garcia Accinelli G, Kettleborough G, Wright J, & Clavijo BJ (2017) KAT: a K-mer analysis toolkit to quality control NGS datasets and genome assemblies. *Bioinformatics (Oxford, England)* 33(4):574-576.
16. Faircloth BC (2016) PHYLUCE is a software package for the analysis of conserved genomic loci. *Bioinformatics* 32(5):786-788.

17. Faircloth BC*, et al.* (2012) Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Syst Biol* 61(5):717-726.
18. Streicher JW, Schulte JA, II, & Wiens JJ (2015) How Should Genes and Taxa be Sampled for Phylogenomic Analyses with Missing Data? An Empirical Study in Iguanian Lizards. *Systematic Biology* 65(1):128-145.
19. Jarvis ED*, et al.* (2014) Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science (New York, N.Y.)* 346(6215):1320-1331.
20. Ramos B*, et al.* (2018) Landscape genomics: natural selection drives the evolution of mitogenome in penguins. *BMC Genomics* 19(1):53.
21. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, & Jermiin LS (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods* 14:587.
22. Nguyen L-T, Schmidt HA, von Haeseler A, & Minh BQ (2015) IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution* 32(1):268-274.
23. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, & Vinh LS (2018) UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol Biol Evol* 35(2):518-522.
24. Zhang C, Rabiee M, Sayyari E, & Mirarab S (2018) ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC bioinformatics* 19(Suppl 6):153.
25. Kozlov AM, Darriba D, Flouri T, Morel B, & Stamatakis A (2018) RAxML-NG: A fast, scalable, and user-friendly tool for maximum likelihood phylogenetic inference. *bioRxiv*:447110.
26. Sayyari E & Mirarab S (2016) Fast Coalescent-Based Computation of Local Branch Support from Quartet Frequencies. *Molecular Biology and Evolution* 33(7):1654-1668.
27. Bouckaert R*, et al.* (2014) BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput Biol* 10(4):e1003537.
28. Miller MA, Pfeiffer W, & Schwartz T (2010) Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *2010 Gateway Computing Environments Workshop (GCE)*, pp 1-8.
29. Cole TL*, et al.* (2019) Mitogenomes Uncover Extinct Penguin Taxa and Reveal Island Formation as a Key Driver of Speciation. *Mol Biol Evol* 36(4):784-797.
30. Slack KE*, et al.* (2006) Early penguin fossils, plus mitochondrial genomes, calibrate avian evolution. *Mol Biol Evol* 23(6):1144-1155.
31. Hospitalache CAT, C.; Donato, M.; Cozzuol M. (2007) A new Miocene penguin from Patagonia and its phylogenetic relationships. *Acta Palaeontologica Polonica* 52(2):299-314.
32. Ksepka DT & Clarke JA (2010) The Basal Penguin (Aves: Sphenisciformes) Perudyptes devriesi and a Phylogenetic Evaluation of the Penguin Fossil Record. *Bulletin of the American Museum of Natural History* 337:1-77.
33. Degrange FJ, Ksepka DT, & Tambussi CP (2018) Redescription of the oldest crown clade penguin: cranial osteology, jaw myology, neuroanatomy, and phylogenetic affinities of Madrynornis mirandus. *Journal of Vertebrate Paleontology* 38(2):e1445636.

34. Hoffmeister CM, Briceño JDC, & Nielsen SN (2014) The Evolution of Seabirds in the Humboldt Current: New Clues from the Pliocene of Central Chile. *PLOS ONE* 9(3):e90043.
35. Marquardt C*, et al.* (2000) Estratigrafía del Cenozoico Superior en el área de Caldera (26°45'- 28°S), III Región de Atacama. *In Congreso Geológico Chileno, No. 9, Actas: 504-508. Puerto Varas.*
36. Brand L, Urbina M, Chadwick A, DeVries TJ, & Esperante R (2011) A high resolution stratigraphic framework for the remarkable fossil cetacean assemblage of the Miocene/Pliocene Pisco Formation. *Peru. J S Am Earth Sci* 31:414.
37. Ksepka DT & Thomas DB (2012) Multiple cenozoic invasions of Africa by penguins (Aves, Sphenisciformes). *Proc Biol Sci* 279(1730):1027-1032.
38. Thomas DB & Ksepka DT (2013) A history of shifting fortunes for African penguins. *Zoological Journal of the Linnean Society* 168(1):207-219.
39. Naish TR*, et al.* (2005) An integrated sequence stratigraphic, palaeoenvironmental, and chronostratigraphic analysis of the Tangahoe Formation, southern Taranaki coast, with implications for mid-Pliocene (c. 3.4–3.0 Ma) glacio-eustatic sea-level changes. *J Roy Soc New Zeal* 35:151-196.
40. Rambaut A, Drummond AJ, Xie D, Baele G, & Suchard MA (2018) Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Systematic Biology* 67(5):901-904.
41. Sarver BAJ*, et al.* (2018) The choice of tree prior and molecular clock does not substantially affect phylogenetic inferences of diversification rates. *bioRxiv*:358788.
42. Frugone MJ*, et al.* (2018) Contrasting phylogeographic pattern among Eudyptes penguins around the Southern Ocean. *Sci Rep* 8(1):17481.
43. Frugone MJ*, et al.* (2019) More than the eye can see: Genomic insights into the drivers of genetic differentiation in Royal/Macaroni penguins across the Southern Ocean. *Molecular Phylogenetics and Evolution* 139:106563.
44. Paradis E & Schliep K (2019) ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35(3):526-528.
45. Matzke N (2013) BioGeoBEARS: BioGeography with Bayesian (and Likelihood) Evolutionary Analysis in R Scripts. *University of California, Berkeley, Berkeley, CA*. .
46. Ree RH & Smith SA (2008) Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Syst Biol* 57(1):4-14.
47. Ronquist F (1997) Dispersal-Vicariance Analysis: A New Approach to the Quantification of Historical Biogeography. *Systematic Biology* 46(1):195-203.
48. Landis MJ, Matzke NJ, Moore BR, & Huelsenbeck JP (2013) Bayesian Analysis of Biogeography when the Number of Areas is Large. *Systematic Biology* 62(6):789-804.
49. Templeton AR (1980) The theory of speciation via the founder principle. *Genetics* 94(4):1011-1038.
50. Matzke NJ (2014) Model selection in historical biogeography reveals that founder-event speciation is a crucial process in Island Clades. *Systematic Biology* 63(6):951-970.

51. Matzke NJ (2013) Probabilistic historical biogeography: new models for founder-event speciation, imperfect detection, and fossils allow improved accuracy and model-testing. *Frontiers of Biogeography* 5(4).
52. Bertelli S & Giannini NP (2005) A phylogeny of extant penguins (Aves: Sphenisciformes) combining morphology and mitochondrial sequences. *Cladistics* 21(3):209-239.
53. International HotBotWaB (2017) Handbook of the Birds of the World and BirdLife International digital checklist of the birds of the world. Version 9.1.
54. Cristofari R*, et al.* (2016) Full circumpolar migration ensures evolutionary unity in the Emperor penguin. *Nat Commun* 7:11842.
55. Clucas GV*, et al.* (2018) Comparative population genomics reveals key barriers to dispersal in Southern Ocean penguins. *Molecular Ecology* 27(23):4680-4697.
56. Assis J*, et al.* (2018) Bio-ORACLE v2.0: Extending marine data layers for bioclimatic modelling. *Global Ecology and Biogeography* 27(3):277-284.
57. Phillips SJ, Anderson RP, & Schapire RE (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190(3):231-259.
58. Warren DL, Glor RE, & Turelli M (2008) Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. *Evolution* 62(11):2868-2883.
59. Evans ME, Smith SA, Flynn RS, & Donoghue MJ (2009) Climate, niche evolution, and diversification of the "bird-cage" evening primroses (Oenothera, sections Anogra and Kleinia). *The American naturalist* 173(2):225-240.
60. Pease JB & Hahn MW (2015) Detection and polarization of introgression in a five-taxon phylogeny. *Systematic Biology* 64(4):651–662
61. Li H & Durbin R (2011) Inference of human population history from individual whole-genome sequences. *Nature* 475(7357):493-496.
62. Li H, Handsaker B, Marshall J, & Danecek P (2015) Samtools. Version 1.3 with HTSlib 1.3.1 [01 October 2016];http://www.htslib.org.
63. Narasimhan V*, et al.* (2016) BCFtools/RoH: a hidden Markov model approach for detecting autozygosity from next-generation sequencing data. *Bioinformatics* 32(11):1749-1751.
64. Nadachowska-Brzyska K, Li C, Smeds L, Zhang G, & Ellegren H (2015) Temporal Dynamics of Avian Populations during Pleistocene Revealed by Whole-Genome Sequences. *Curr Biol* 25(10):1375-1380.
65. Nam K*, et al.* (2010) Molecular evolution of genes in avian genomes. *Genome Biol* 11(6):R68.
66. FORCADA J & TRATHAN PN (2009) Penguin responses to climate change in the Southern Ocean. *Global Change Biology* 15(7):1618-1630.
67. Lande R, Engen S, & Saether B (2003) Stochastic Population Dynamics in Ecology and Conservation. *Oxford University Press*.
68. Huerta-Cepas J, Serra F, & Bork P (2016) ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. *Mol Biol Evol* 33(6):1635-1638.
69. Yang Z (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. *Computer applications in the biosciences : CABIOS* 13(5):555-556.

70. Wang J, Vasaikar S, Shi Z, Greer M, & Zhang B (2017) WebGestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. *Nucleic Acids Res* 45(W1):W130-W137.
71. Szklarczyk D*, et al.* (2019) STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* 47(D1):D607-d613.
72. Kumar S, Stecher G, & Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol* 33(7):1870-1874.
73. Murrell B*, et al.* (2012) Detecting Individual Sites Subject to Episodic Diversifying Selection. *PLOS Genetics* 8(7):e1002764.
74. Murrell B*, et al.* (2013) FUBAR: a fast, unconstrained bayesian approximation for inferring selection. *Mol Biol Evol* 30(5):1196-1205.
75. Pond SLK & Frost SDW (2005) Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* 21(10):2531-2533.
76. Weaver S*, et al.* (2018) Datamonkey 2.0: A Modern Web Application for Characterizing Selective and Other Evolutionary Processes. *Molecular Biology and Evolution* 35(3):773-777.
77. Li R*, et al.* (2010) De novo assembly of human genomes with massively parallel short read sequencing. *Genome research* 20(2):265-272.
78. Liu W*, et al.* (2018) RGAAT: A Reference-based Genome Assembly and Annotation Tool for New Genomes and Upgrade of Known Genomes. *Genomics, Proteomics & Bioinformatics* 16(5):373-381.
79. Figueiró HV*, et al.* (2017) Genome-wide signatures of complex introgression and adaptive evolution in the big cats. *Science Advances* 3(7):e1700299.
80. Li C*, et al.* (2014) Two Antarctic penguin genomes reveal insights into their evolutionary history and molecular changes related to the Antarctic environment. *GigaScience* 3(1):27-27.
81. Ksepka D.T. BS, Giannini N. (2006) The phylogeny of the living and fossil Sphenisciformes (penguins). *Cladistics* 22:412-441.
82. Clarke JA*, et al.* (2007) Paleogene equatorial penguins challenge the proposed relationship between biogeography, diversity, and Cenozoic climate change. *Proc Natl Acad Sci U S A* 104(28):11545-11550.
83. Baker AJ, Pereira SL, Haddrath OP, & Edge KA (2006) Multiple gene evidence for expansion of extant penguins out of Antarctica due to global cooling. *Proc Biol Sci* 273(1582):11-17.
84. Subramanian S, Beans-Picon G, Swaminathan SK, Millar CD, & Lambert DM (2013) Evidence for a recent origin of penguins. *Biol Lett* 9(6):20130748.
85. Gavryushkina A*, et al.* (2017) Bayesian Total-Evidence Dating Reveals the Recent Crown Radiation of Penguins. *Systematic Biology* 66(1):57-73.
86. Weisrock DW, Harmon LJ, & Larson A (2005) Resolving deep phylogenetic relationships in salamanders: analyses of mitochondrial and nuclear genomic data. *Syst Biol* 54(5):758-777.
87. Oliveros CH*, et al.* (2019) Earth history and the passerine superradiation. *Proceedings of the National Academy of Sciences* 116(16):7916-7925.

88. Labuschagne C, Kotzé A, Grobler JP, & Dalton DL (2014) The complete sequence of the mitochondrial genome of the African Penguin (Spheniscus demersus). *Gene* 534(1):113-118.
89. Watanabe M*, et al.* (2006) New candidate species most closely related to penguins. *Gene* 378:65-73.
90. Grosser S, Burridge CP, Peucker AJ, & Waters JM (2015) Coalescent Modelling Suggests Recent Secondary-Contact of Cryptic Penguin Species. *PLOS ONE* 10(12):e0144966.
91. Simeone A*, et al.* (2009) *Heterospecific Pairing and Hybridization between Wild Humboldt and Magellanic Penguins in Southern Chile* (BIOONE) pp 544-550, 547.
92. Livermore R, Nankivell A, Eagles G, & Morris P (2005) Paleogene opening of Drake Passage. *Earth and Planetary Science Letters* 236:459-470.
93. Lyle M, Gibbs S, Moore TC, & Rea DK (2007) Late Oligocene initiation of the Antarctic Circumpolar Current: Evidence from the South Pacific. *Geology* 35(8):691-694.
94. Barker PF & Thomas E (2004) Origin, signature and palaeoclimatic influence of the Antarctic Circumpolar Current. *Earth-Science Reviews* 66(1):143-162.
95. Goldner A, Herold N, & Huber M (2014) Antarctic glaciation caused ocean circulation changes at the Eocene-Oligocene transition. *Nature* 511(7511):574-577.
96. Hansen J, Sato M, Russell G, & Kharecha P (2013) Climate sensitivity, sea level and atmospheric carbon dioxide. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 371(2001):20120294.
97. Shevenell AE, Kennett JP, & Lea DW (2004) Middle Miocene Southern Ocean Cooling and Antarctic Cryosphere Expansion. *Science* 305(5691):1766-1770.
98. Dalziel IWD*, et al.* (2013) A potential barrier to deep Antarctic circumpolar flow until the late Miocene? *Geology* 41(9).
99. Pearce JA*, et al.* (2014) Composition and evolution of the Ancestral South Sandwich Arc: Implications for the flow of deep ocean water and mantle through the Drake Passage Gateway. *Global and Planetary Change* 123:298-322.
100. González-Wevar CA*, et al.* (2017) Following the Antarctic Circumpolar Current: patterns and processes in the biogeography of the limpet Nacella (Mollusca: Patellogastropoda) across the Southern Ocean. *Journal of Biogeography* 44(4):861-874.
101. Near TJ*, et al.* (2012) Ancient climate change, antifreeze, and the evolutionary diversification of Antarctic fishes. *Proceedings of the National Academy of Sciences* 109(9):3434-3439.
102. Strugnell JM*, et al.* (2011) The Southern Ocean: Source and sink? *Deep Sea Research Part II: Topical Studies in Oceanography* 58(1):196-204.
103. Strugnell JM, Rogers AD, Prodöhl PA, Collins MA, & Allcock AL (2008) The thermohaline expressway: the Southern Ocean as a centre of origin for deep-sea octopuses. *Cladistics* 24(6):853-860.
104. Crame JA (2018) Key stages in the evolution of the Antarctic marine fauna. *Journal of Biogeography* 45(5):986-994.

105. Halanych KM & Mahon AR (2018) Challenging Dogma Concerning Biogeographic Patterns of Antarctica and the Southern Ocean. *Annual Review of Ecology, Evolution, and Systematics* 49:355-378.
106. Dueñas LF*, et al.* (2016) The Antarctic Circumpolar Current as a diversification trigger for deep-sea octocorals. *BMC evolutionary biology* 16:2-2.
107. Zachos JC, Shackleton NJ, Revenaugh JS, Pälike H, & Flower BP (2001) Climate response to orbital forcing across the Oligocene-Miocene boundary. *Science* 292(5515):274-278.
108. Zhou P*, et al.* (2020) A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 579(7798):270-273.
109. Damas J*, et al.* (2020) Broad Host Range of SARS-CoV-2 Predicted by Comparative and Structural Analysis of ACE2 in Vertebrates. *bioRxiv*:2020.2004.2016.045302.
110. Nery MF, Arroyo JI, & Opazo JC (2013) Accelerated Evolutionary Rate of the Myoglobin Gene in Long-Diving Whales. *Journal of Molecular Evolution* 76(6):380-387.
111. Mirceta S*, et al.* (2013) Evolution of mammalian diving capacity traced by myoglobin net surface charge. *Science* 340(6138):1234192.
112. Noren S, Williams T, Pabst DA, McLellan WA, & Dearolf J (2001) The development of diving in marine endotherms: Preparing the skeletal muscles of dolphins, penguins, and seals for activity during submergence. *Journal of comparative physiology. B, Biochemical, systemic, and environmental physiology* 171:127-134.
113. Duchamp C*, et al.* (1991) Nonshivering thermogenesis in king penguin chicks. I. Role of skeletal muscle. *Am J Physiol* 261(6 Pt 2):R1438-1445.
114. Duchamp C, Rouanet JL, & Barré H (2002) Ontogeny of thermoregulatory mechanisms in king penguin chicks (Aptenodytes patagonicus). *Comp Biochem Physiol A Mol Integr Physiol* 131(4):765-773.
115. Ponganis PJ, Welch TJ, Welch LS, & Stockard TK (2010) Myoglobin production in emperor penguins. *J Exp Biol* 213(11):1901-1906.
116. Hodgson DA*, et al.* (2014) Terrestrial and submarine evidence for the extent and timing of the Last Glacial Maximum and the onset of deglaciation on the maritime-Antarctic and sub-Antarctic islands. *Quaternary Science Reviews* 100:137-158.
117. Younger JL*, et al.* (2015) Too much of a good thing: sea ice extent may have forced emperor penguins into refugia during the last glacial maximum. *Glob Chang Biol* 21(6):2215-2226.
118. Peña M F*, et al.* (2014) Have Historical Climate Changes Affected Gentoo Penguin (Pygoscelis papua) Populations in Antarctica? *PLOS ONE* 9(4):e95375.
119. Clucas GV*, et al.* (2014) A reversal of fortunes: climate change 'winners' and 'losers' in Antarctic Peninsula penguins. *Scientific Reports* 4(1):5024.
120. Mura-Jornet I*, et al.* (2018) Chinstrap penguin population genetic structure: one or more populations along the Southern Ocean? *BMC Evol Biol* 18(1):90.
121. Dantas GPM*, et al.* (2018) Demographic history of the Magellanic Penguin (Spheniscus magellanicus) on the Pacific and Atlantic coasts of South America. *Journal of Ornithology* 159(3):643-655.

122. Cristofari R*, et al.* (2018) Climate-driven range shifts of the king penguin in a fragmented ecosystem. *Nature Climate Change* 8(3):245-251.

123. Trucchi E*, et al.* (2014) King penguin demography since the last glaciation inferred from genome-wide data. *Proceedings of the Royal Society B: Biological Sciences* 281(1787):20140528.

124. Cole TL*, et al.* (2019) Receding ice drove parallel expansions in Southern Ocean penguins. *Proceedings of the National Academy of Sciences* 116(52):26690-26696.