The genome of the Balearic shearwater (*Puffinus mauretanicus*), a Critically Endangered seabird: a valuable resource for evolutionary and conservation genomics

Cristian Cuevas-Caballé[1], Joan Ferrer Obiol[1], Joel Vizueta[2], Meritxell Genovart[3], Jacob Gonzales-Solis[4], Marta Riutort[1], and Julio Rozas[1]

[1]Universitat de Barcelona
[2]University of Copenhagen
[3]Mediterranean Institute of Advanced Studies
[4]University of Barcelona

April 16, 2024

Abstract

The Balearic shearwater (*Puffinus mauretanicus*) is the most threatened seabird in Europe. The fossil record suggests that human colonisation of the Balearic Islands resulted in a sharp decrease of the population size. Currently, populations continue to be decimated mainly due to predation by introduced mammals and bycatch in longline fisheries, and some studies predict their extinction by 2070. We present the first high-quality reference genome for the species which was obtained by a combination of short and long-read sequencing. Our hybrid assembly includes 4,169 scaffolds, with a scaffold N50 of 2.1 Mbp, a genome length of 1.2 Gbp, and BUSCO completeness of 96%, which is amongst the highest across sequenced avian species. This reference genome allowed us to study critical aspects relevant to the conservation status of the species, such as an evaluation of overall heterozygosity levels and the reconstruction of its historical demography. Our phylogenetic analysis using whole-genome information resolves current uncertainties in the order Procellariiformes systematics. Comparative genomics analyses uncover a set of candidate genes that may have played an important role into the adaptation to a pelagic lifestyle of Procellariiformes, including those for the enhancement of fishing capabilities, night vision and the development of natriuresis. This reference genome will be the keystone for future developments of genetic tools in conservation efforts for this Critically Endangered species.

Introduction

The genomic sequence of a species accumulates valuable information on the evolutionary history, including demographic and selective events, and on the evolution of genes and traits (Feng et al., 2020; Foote et al., 2015; Jarvis et al., 2014; Nadachowska-Brzyska, Li, Smeds, Zhang, & Ellegren, 2015), information that it is also crucial for the emerging field of conservation genomics (Allendorf, 2017). The genetic diversity within a species represents a reservoir of adaptive variation that can help populations to cope with environmental variability (Dussex et al., 2021). Understanding the processes that shape genetic diversity and its distribution pattern within species is paramount to assess the conservation status or the factors responsible for a species decline (Brüniche-Olsen, Kellner, Belant, & DeWoody, 2021) (Wang et al., 2021). This knowledge can inform the proposal of effective conservation and management plans, as for instance the definition of management units (Funk, McKay, Hohenlohe, & Allendorf, 2012). In this context, next-generation sequencing (NGS) techniques allow the analysis of an increased density of markers across the genome, providing unprecedented accuracy in the estimations of population genetic parameters relevant for scientific-based conservation recommendations (Supple & Shapiro, 2018).

Among the Critically Endangered species listed by the IUCN Red List (IUCN 2021), we find the Balearic shearwater (*Puffinus mauretanicus* Lowe, 1921) (Figure 1a) belonging to the most diverse order of seabirds, the Procellariiformes. This order has a worldwide distribution and comprises more than 140 species (IUCN 2021) in four families: petrels and shearwaters (Procellariidae); northern storm petrels (Hydrobatidae); southern storm petrels (Oceanitidae) and albatrosses (Diomedeidae). All species show many morphological, physiological and life history traits associated with their adaptation to a pelagic lifestyle. They are long-lived with deferred sexual maturity, low fecundity (all lay a single egg), colonial breeders, socially (and mostly sexually) monogamous, highly philopatric and with prominent salt gland at the base of the bill, adaptations for underwater vision to fish as well as a particularly acute sense of smell, among other traits (Brooke, 2004). Within this apparent homogeneity, the group shows a large variation in body mass and lifestyles, ranging from 20g to 15kg, from bodies shaped for diving (e.g. short strong wings used for wing-propelled diving) to those prepared for an extremely vagile lifestyle (thin elongated wings) and from a continuous flapping to dynamic soaring flight modes. Currently, their phylogenetic relationships present some conflicting issues (Estandía et al., 2021; Hackett et al., 2008), such as the position of albatrosses, whether storm petrels (Families Hydrobatidae and Oceanitidae) constitute a monophyletic group, and whether diving-petrels (genus*Pelecanoides*) should be considered an independent family from Procellariidae.

The Balearic shearwater is a medium-sized pelagic seabird endemic to the Balearic Islands. Its population size is undergoing a fast annual decline of 7.4-14% (Genovart et al., 2016; Oro, Aguilar, Igual, & Louzao, 2004) mostly due to bycatch in longline fisheries and predation by invasive mammals in the colonies (Arcos, Louzao, & Oro, 2008; Louzao, Arcos, Hyrenbach, Sola, & Oro, 2004; Martí & Ruiz, 2004). Currently, it has a reduced number of breeding pairs (estimated as*ca.* 3,200, Arcos, 2011, with a total population size up to 30,000 individuals due to the vast contingent of floaters (Arcos et al., 2012; Arroyo et al., 2016). Genetic studies based on mtDNA and microsatellites found that this species has low levels of genetic diversity and high inbreeding coefficients (Genovart, Juste, Contreras-Díaz, & Oro, 2012). Although local inbreeding and natal philopatry can contribute to a reduction in population size, the actual worst menace for the species comes from human activities, and a population viability study based on demographic modeling predicted that the species would become extinct by 2070 (Genovart et al., 2016).

Indeed, a population viability study based on demographic modeling predicted that the species would become extinct by 2070 (Genovart et al., 2016). Moreover, studies based on mitochondrial markers (Genovart, Juste, & Oro, 2005) and also on morphology and migratory behaviour (Austin et al., 2019), suggested a possible ongoing hybridization and introgression process between Balearic and Mediterranean (*P. yelkouan*) shearwaters, which may represent an additional threat for the species.

Here, we (1) provide a high-quality reference genome for the Balearic shearwater along its structural and functional annotations; (2) estimate genome-wide heterozygosity and the historical demography of the species by performing Multiple Sequentially Markovian Coalescent (MSMC) analyses; (3) revisit the phylogeny of the order by using this genome together with seven additional Procellariiformes genomes released by the B10K (Feng et al., 2020), and (4) uncover genes putatively involved in Procellariiformes adaptation to pelagic life. The high-quality genome of the most endangered seabird in Europe presented here will be the base for further population-based conservation genomics studies.

Materials and Methods

Sampling, DNA and RNA extraction and sequencing

We sampled two Balearic shearwater adults and one chick. Adults were sampled on Sa Cella colony, Mallorca (male) and on Sa Conillera, Eivissa (unsexed) in 2004, while the chick was sampled on Conills islet (Mallorca) in July 2019. From here on the animals will be referred to as male-Mll, unsexed-Ei and chick-Mll, respectively. Special permits to obtain the samples were issued by Conselleria de Medi Ambient, Agricultura i Pesca (Govern de les Illes Balears, Spain).

We extracted DNA from blood samples preserved in absolute ethanol for both adults. The DNA extraction for the male-Mll was performed with DNeasy Blood & Tissue Kit (Qiagen) following the manufacturer's

instructions, and with Blood & Cell Culture DNA Mini Kit (Qiagen) for the unsexed-Ei. RNA was extracted from the chick-Mll's blood cells preserved in RNAlater 1:5 using the RNeasy Mini Kit (Qiagen) according to the manufacturer's protocols. We performed the quality control with gel electrophoresis and NanoDrop One (Thermo Fisher Scientific, Waltham, MA, USA), and the quantification with an Invitrogen Qubit Fluorometer 2.0 (Broad Range kit).

We obtained the reference genome combining short-read and long-read sequencing libraries, and using RNA-seq data to assist with the annotation. First, an Illumina TruSeq DNA PCR Free library (insert size = 350 bp) was prepared by Macrogen (South Korea) using DNA from male-Mll, and sequenced using two HiSeq X Ten runs (2x150bp). Second, long-read libraries were prepared, from the DNA of unsexed-Ei, using the Ligation kit SQK-LSK109 1D from ONT (Oxford Nanopore Technologies) (N50 of 9431 bp) at CNAG (Centro Nacional de Análisis Genómico, Spain) and sequenced through five runs of MinION on FLO-MIN106 flow cells. Third, we prepared RNA sequencing (RNA-seq) libraries from the chick-Mll's RNA using the TruSeq RNA Sample Prep Kit v2 with Ribo-Zero, and we sequenced the libraries on a NovaSeq 6000 (2x100bp) (Macrogen, South Korea).

Genome Assembly

We performed a *de novo* hybrid genome assembly with MaSuRCA 3.3.1 (Zimin et al., 2017), using short (Illumina) and long (ONT) reads. Before the assembly step, we filtered the ONT reads with a Phred quality score (Q [?] 5) using the NanoFilt software (included in NanoPack, De Coster, D'Hert, Schultz, Cruts, & Van Broeckhoven, 2018). Paired-end Illumina reads were parsed into MaSuRCA without any preprocessing, as adapters and errors are handled by the QuORUM error corrector (Marcais, Yorke, & Zimin, 2015), which is part of the MaSuRCA pipeline. MaSuRCA was run applying the following parameters: fragment mean (422), fragment stdev (312) and estimated genome size (1.2 Gbp). The resulting assembly was screened for contaminants with BlobTools v1.0 (Laetsch & Blaxter, 2017) -x bestsumorder. Assembly completeness was assessed with BUSCO 4.0.2 (Seppey, Manni, & Zdobnov, 2019) using the 8,338 single-copy conserved genes in aves_odb10 database (Kriventseva et al., 2019).

Transcriptome Assembly

We trimmed RNA-seq raw reads for adapters with BBDuk (https://sourceforge.net/projects/bbmap/) (k = 17, tpe option), and used STAR 2.7.3a (Dobin et al., 2013) to map the filtered reads to the newly assembled reference genome. We obtained the transcriptome assembly with Trinity 2.8.6 (Grabherr et al., 2011) using the genome-guided bam mode (–genome_guided_max_intron 82945). Transcripts were clustered with CD-HIT (Fu, Niu, Zhu, Wu, & Li, 2012; W. Li & Godzik, 2006) 4.8.1 (-c 0.98) and coding regions (CDS) were predicted with TransDecoder 5.5.0 (https://github.com/TransDecoder/).

Mitogenome Assembly

We trimmed adapters from Illumina raw reads with BBDuk (k = 23, tpe option), before using them as input to NOVOPlasty 2.7.2 (Dierckxsens, Mardulyn, & Smits, 2017). The *Puffinus lherminieri* mitogenome (MH206163.1) was used as seed using the following parameters: Genome Range (16000-24000), Insert size (422), Insert range (1.74) and Insert range strict (1.3). The annotation was performed using the MITOS WebServer (Bernt et al., 2013).

Repeat annotation

We generated a *de novo* repeat library of the genome with RepeatModeler – 1.0.11 (Smit, Hubley, & Green,) on scaffolds >100 kbp. This library was combined with all avian and ancestral consensus repeats from Dfam_Consensus-20181026 (Storer, Hubley, Rosen, Wheeler, & Smit, 2021), RepBase-20181026 (Jurka et al., 2005) and the repeat annotation of the Cory's shearwater (*Calonectris borealis*) (Feng et al., 2020), which represents the most closely related sequenced genome. Redundancies among libraries were removed with the script ReannTE_MergeFasta.pl (*https://github.com/4ureliek/ReannTE*). We then ran RepeatMasker 4.0.7 (Smit et al.,) using the combined library as a reference, with the following parameters: -xsmall -e ncbi -s -gccalc -no_is -gff.

Structural and Functional Annotation

We performed the structural annotation with BRAKER 2.1.2 (https://github.com/Gaius-Augustus/BRAKER) (–etpmode) using data from both the Cory's shearwater proteome (Feng et al., 2020), and the RNA-Seq data generated in this work. Since the inclusion of RNA-Seq data appeared detrimental, we excluded this piece of information to perform the final annotation using the soft-masked genome with BRAKER 2.1.2 (–prg=gth –trainFromGth).

We made the functional annotation of the predicted genes using a similarity-based approach. We determined the protein domains with InterProScan 5.31-70.0 (Jones et al., 2014), used BLASTP (Altschul, Gish, Miller, Myers, & Lipman, 1990; Camacho et al., 2009) (-evalue 1e-5; -max_target_seqs 10) against the Swiss-Prot database (Boutet et al., 2016) and the Cory's shearwater and the Zebra finch reference (UP000007754) proteomes. Transcripts were annotated in the same manner. We also annotated the ncRNAs using cmscan from INFERNAL 1.1.2 (Nawrocki, Kolbe, & Eddy, 2009) with the covariance models (CMs) from the Rfam 14.1 database, and tRNA genes using tRNAscan-SE 2.0.5 (Chan & Lowe, 2019).

Demographic History

We used MSMC2 (Schiffels & Wang, 2020) to infer the historical demography of the Balearic shearwater. MSMC2 implements a MSMC model, which allows the estimation of the effective population size (N_e) over time. To generate input files for MSMC2, we mapped Illumina short reads to scaffolds larger than 1 Mbp (343 scaffolds spanning 71.8% of the assembled genome) using BWA-MEM 0.7.17 (H. Li & Durbin, 2009), as recommended in Gower et al., 2018). First, we called the SNPs using samtools mpileup (Samtools mpileup 1.9 -q 20 -Q 20 -C 50) and then bcftools 1.9 -c -V indels. The input files were then generated by converting the SNPs obtained to MSMC input format using the bamCaller.py script accounting for the mean coverage of each scaffold. Multiple sequentially Markovian coalescent (MSMC) for two haplotypes, known as PSMC', was run with MSMC2 with time patterning specified as -p 1*4+30*2+1*4+1*6+1*10.

We ran 100 bootstraps of 29 pseudo-chromosomes (Yamashina & Udagawa, 1954) sampling 20 chunks of 1.508.752 bp with replacement using multihetsep_bootstrap.py. We scaled time and population size using a generation time for the Balearic shearwater of 12.8 years (Genovart et al., 2016) and the Northern fulmar (*Fulmarus glacialis*) mutation rate (2.89×10^{-9} substitutions per nucleotide per generation, Nadachowska-Brzyska et al., 2015).

Genome-wide heterozygosity

We estimated genome-wide heterozygosities using information of a single individual from all eight Procellariiformes species studied. We applied the Robinson et al., 2019 method, with minor modifications to take genome fragmentation into consideration, since we included genome assemblies with varying amounts of contiguity. The DNA sequence data (genome assemblies and whole-genome sequencing data) were downloaded from NCBI (PRJNA261828, PRJNA545868; Feng et al., 2020; Jarvis et al., 2014 *)*. For each species, adapter-trimmed reads were aligned to its genome assembly using bwa mem (Heng Li, 2013), bam files were merged using Picard-Tools (*http://broadinstitute.github.io/picard/*) and variants were called using the GATK 4.1.9 HaplotypeCaller and GenotypeGVCFs (McKenna et al., 2010). Sites with a coverage $<1/3$X or >2X of the average coverage depth (of the particular genome) were filtered out using VCFtools 0.1.15 (Danecek et al., 2011). We computed per-site heterozygosity as the proportion of heterozygous sites per total number of called genotypes within a single individual in nonoverlapping 25Kb windows across each scaffold. Windows with less than 50% of net sites (those excluding missing or filtered sites), were excluded from the analysis.

Orthology inference

We performed the phylogenomic and comparative genomics analyses including information from 12 species with an available genome assembly: eight Procellariiformes (*P. mauretanicus* , *Thalassarche chlororhynchos* , *Hydrobates tethys* , *Oceanites oceanicus* ,*Fregetta grallaria* , *Pelecanoides urinatrix* , *Fulmarus glacialis* and *Calonectris borealis*), and four outgroups,*Aptenodytes forsteri* , *Pygoscelis adeliae*(Sphenisciformes);

Egretta garzetta , *Phalacrocorax carbo*(Pelecaniformes). We inferred orthologous genes across the proteomes of these 12 species using OrthoFinder 2.3.8 (Emms & Kelly, 2019) with default parameters.

Phylogenetic relationships

We built a multiple sequence alignment (MSA) for each 1:1 orthologs with PRANK v.100802 (Loytynoja, 2014), using both coding sequences (CDS) (-codon -noxml -notree -F) and amino acid sequences (-noxml -notree -F). Individual alignments were concatenated with catfasta2phyml v1.1.0 (https://github.com/nylander/catfasta2phyml) to create a CDS supermatrix and an amino acid supermatrix. Only locus with data for all the twelve species have been considered. Fourfold degenerate sites (4D) for the CDS supermatrix were extracted with MEGA X (Kumar, Stecher, Li, Knyaz, & Tamura, 2018). We performed unpartitioned maximum likelihood (ML) phylogenetic analyses using IQ-TREE 1.6.12 (Nguyen, Schmidt, Von Haeseler, & Minh, 2015) (-bb 1000) both for 4D and amino acid supermatrices. Optimal models of sequence evolution were obtained with ModelFinder (Kalyaanamoorthy, Minh, Wong, Von Haeseler, & Jermiin, 2017) according to Bayesian information criterion (BIC), and the resulting best-fit models were GTR+F+R2 for 4D and HIVb+F+R3 for amino acid supermatrix. Node support was obtained with Ultrafast Bootstrap (Hoang, Chernomor, Von Haeseler, Minh, & Vinh, 2018).

To explicitly account for incomplete lineage sorting (ILS) under the Multispecies Coalescent Model (MSC), we inferred the species tree using the summary coalescent approach, as implemented in ASTRAL-III 5.6.3 (C. Zhang, Rabiee, Sayyari, & Mirarab, 2018). We first obtained all gene trees (for each 1:1 orthologous genes) using IQ-TREE 1.6.12, and inferred the species tree and its normalized score (from both CDS and amino acid gene trees) using ASTRAL-III.

We generated an ultrametric tree with r8s v.1.81 (Sanderson, 2003) using the 4D supermatrix ML tree. We used four calibration points (in myr): root (max_age=84 min_age=73, (Braun et al., 2011), most recent common ancestor (MRCA) of Spheniscidae (min_age=12.6, (Subramanian, Beans-Picon, Swaminathan, Millar, & Lambert, 2013)), MRCA Procellariiformes (min_age=49, Claramunt & Cracraft, 2015), MRCA Procellariidae (min_age=14, (Prum et al., 2015)), retrieved from TimeTree (Kumar, Stecher, Suleski, & Hedges, 2017), and references therein). We used the penalized likelihood (PL) method and the Truncated Newton (TN) algorithm, smoothing parameter was set to 100.

Positive Selection Analysis

We evaluated the selective constraints of genes that could be associated with pelagic lifestyle. For this purpose, we performed the analysis with HyPhy 2.5 (Kosakovsky Pond, Frost, & Muse, 2005), using Procellariiformes data (1:1 MSAs). Prior to the analysis, non-reliable positions across all 1:1 orthologs MSAs were filtered with ZORRO (Wu, Chatterji, & Eisen, 2012) (default options; MSA with average quality < 5 were filtered). We used the aBSREL method (Smith et al., 2015) to test for positive selection, and the RELAX method (Wertheim, Murrell, Smith, Pond, & Scheffler, 2015) to test for relaxed/intensified selection. We also performed a Gene ontology (GO) enrichment analysis of the candidate genes using the GOstats (Falcon & Gentleman, 2007) R package against the background GOs of 1:1 orthologs.

Gene family evolution

We estimated gene turnover rates, the number of gene gains and losses across the phylogeny lineages, and inferred gene family contractions and expansions using BadiRate 1.7 (Librado, Vieira, & Rozas, 2012). For the analysis we first inferred the orthogroups with OrthoFinder 2.3.8, and we used the calibrated ultrametric tree estimated with r8s. We tested, under the Birth-Death-Innovation (BDI) model for turnover rates, several biological relevant hypotheses with three different branch models: Free Rates (FR), Global Rates (GR) and Branch-specific Rates (BR), and chose the best model based on the lowest AIC value. To ensure an appropriate convergence we ran multiple times each model.

Results

Sequencing data and Genome assembly and annotation

Illumina paired-end (2x150 bp) sequencing of the male-Mll yielded a throughput of 147.7 Gbp (Table 1), representing a mean coverage of 118x. The five runs of ONT sequencing of the unsexed-Ei resulted in a 10x coverage with a read N50 of 9,431 bp. RNA sequencing of the chick-Mll (2x100 bp) yielded 15 Gbp of data.

We obtained a hybrid assembly with MaSuRCA formed by 4,169 scaffolds, with an N50 of 2.1 Mbp, and an assembly length of 1.21 Gbp (Table 2, Figure 1b). The completeness analysis using BUSCO yields a value of 95.9%, and only 0.3% of the complete genes were duplicated and 1.1% were fragmented (Table 2). Our *de novo* repeat annotation analysis shows that 9.95% of the genome consists of repetitive regions (Table S1), which is within the range of previously sequenced avian genomes (G. Zhang et al., 2014). Among repeat elements, long interspersed nuclear elements (LINEs) were the most abundant (4.45% of the genome). The genome annotation process resulted in a total of 21,959 protein-coding genes, of which 18,769 (85.5%) have at least one GO associated term, and 19,218 (87.5%) have hits across the surveyed curated databases (Table S2).

Blood transcriptome assembly from the chick-Mll resulted in 224,904 transcripts (Table S3). However, BUSCO completeness was only 62.4%, which was far below genome completeness, probably due to the RNA coming from a single not very transcriptionally active tissue.

The assembly of the mitogenome of *P. mauretanicus* resulted in a single contig of 19,885 bp long, with a coverage (Illumina reads) of 371x, which is around three times higher than the coverage of the nuclear genome. This mitogenome has the same gene order as other published Procellariiformes' mitogenomes (Figure S1). The mitogenome has two copies of the *nad6* gene, as predicted in *P. lherminieri* (Torres et al., 2018)*;* the later feature was also confirmed analysing the mean coverage (illumina reads) across genes (Table S4).

Historical demography of the Balearic shearwater

Balearich shearwater PSMC' analysis showed support for a steady growth in population size from an originally low population size followed by a sudden increase ˜200 kya (Figure 1c). High population size did not last long and suffered a sudden decrease to nearly one tenth of the population coinciding with the end of the glacial period before the last interglacial period (119-128 kya) and a prolonged period of low sea level (Figure 1c). Hereafter, N_e remained stable until ˜10 kya ago, as more recent MSMC2 time segments are regarded as being unreliable (Schiffels & Wang, 2020).

Genome-wide heterozygosity in *P. mauretanicus* was 0.0024, which is within the range of genome-wide heterozygosities estimated for other Procellariiformes (ranging from 0.0014 in *T. chlororhynchos* to 0.0037 in *P. urinatrix)* (Figure 2a). Among Procellariiformes, small-bodied species tended to have higher mean heterozygosities but also higher variance than large-bodied species (Figure 2b).

Phylogenetic relationships

OrthoFinder analysis estimated 6,172 single copy (1:1) ortholog genes across the 12 genomes surveyed. With this data we generated three supermatrices: 1) CDS supermatrix of 10,534,506 bp long to extract the 4D sites, 2) 4D supermatrix with 1,512,677 4-fold degenerate sites, and 3) the amino acid supermatrix including 3,466,564 sites. Phylogenetic analyses using the 4D and the amino acid supermatrices recovered the same topology with full support at all nodes (ultrafast bootstrap = 100; Figures S2 and S3).

The analysis performed to explicitly account for incomplete lineage sorting (ILS) with ASTRAL using either the individual gene sequences (CDS gene trees) or the individual amino acid sequences (amino acid gene trees), produced species trees with the same topology as those obtained by ML using 4D or amino acid supermatrices (Figures S4 and Figure S5). The normalized quartet score (proportion of input gene tree quartet trees in agreement with the species tree) was 0.78 for CDS gene trees and 0.64 for amino acid gene trees.

The ultrametric tree (Figure 3) obtained using r8s from the 4D supermatrix ML tree summarizes the recovered topology. In this topology, the Atlantic yellow-nosed albatross (*T. chlororhynchos* , Diomedeidae) is the sister group to all the other Procellariiformes. We also find that storm petrels (Hydrobatidae and

Oceanitidae) do not constitute a monophyletic group. In addition, diving petrels (*Pelecanoides*) are included within Procellariidae.

Comparative Genomics and Positive selection analyses

To identify genes associated with adaptation to a pelagic lifestyle in the Procellariiformes, we performed a positive selection analysis across 12 species including eight Procellariiformes species applying the HyPhy aBSREL model. We identified the hallmark of positive selection in 20 (out of the 6,172 single-copy orthologs genes), after correcting for multiple testing (Table S5). The enriched GO analysis uncovered terms related with striated muscle cell differentiation, nutrient reservoir activity, response to starvation, visual learning, positive regulation of neural retina development, olfactory receptor activity or natriuresis (Table S6). We also performed an analysis to assess the global impact of natural selection in Procellariformes (both positive and negative selection), which uncovered a total of 310 genes (Table S7). The GO terms enriched in these genes include wound healing, response to wounding, inflammatory response, sensory perception of sound, smell and chemical stimulus, neurological system process, defense response, response to stress, camera-type eye development, renal system and chloride transport among others (Table S8).

Using OrthoFinder 2.3.8, we identified 182,487 N:N orthogroups across all genes identified in the 12 analysed genomes. This data, together with the estimated ultrametric tree, was used to estimate gene gains, losses, and number of genes in the ancestral nodes using BadiRate; for the analysis we selected the Free Rates (FR) model, since it was the best fitted branch model. The analysis was conducted including all orthogroups, and the minimum number of gains and losses per branch is represented in Figure 3. Our analysis showed a tendency to gain genes in Procellariiformes (+442/-34), while the branch leading to albatrosses (Diomedeidae) showed an opposite effect, with a noticeable loss of genes (+464/-3258); the branch leading to the rest of the Procellariiformes (+379/-15) is in the line of the general behavior of the tubenoses (Table S9). Within the order, families Oceanitidae and Hydrobatidae present the same trend, with the branch leading to *H. thethys*presenting a stronger gene loss balance (+325/-966) than the branch leading to the ancestor of Oceanitidae (*O. oceanicus* and *F. grallaria* (+182/-414)).

We identified three gene families significantly expanded in the branch leading to the Procellariiformes (Table S9). These families encode zinc finger proteins (OG0000000), olfactory receptors (OG0000084) and avian histones (OG0000224).

Discussion

A high-quality genome assembly for the most endangered seabird in Europe

The assembly length and the GC content of the Balearic shearwater hybrid assembly presented here are similar to those reported in the seven Procellariiformes genomes released by the B10K (Feng et al., 2020). Albeit repetitive content is remarkably higher (+33.4%) in the Balearic shearwater in comparison to the other genomes of the order, but within the range of avian genomes (G. Zhang et al., 2014). This difference of up to a third can be due to the fact that we included a Procellariiform (*C. borealis*) repeat library prior to running RepeatMasker, achieving a more precise library that encloses clade related repeats that are present in the genome but not found by the de novo RepeatModeler library. The genome assembly completeness (BUSCO 95.9%) is slightly higher than the obtained for other recently published bird genomes (Feng et al., 2020; Prost et al., 2019), and even higher than genome assemblies including optical mapping (Penalba et al., 2020). Despite not being a chromosome-scale assembly, contiguity is also quite high (N50 2.1 Mbp), and higher than recent avian MaSuRCA hybrid assemblies (Gan et al., 2019; Leroy et al., 2019).

The retrieved proteome (21,959 protein-coding genes) is similar to previous genomes (Liu et al., 2021; Recuerda et al., 2021), but higher than the B10K 2020 genomes used in the comparative studies in this work (mean of 16K). This is probably due to the B10K annotation pipeline being fully based on homology, whilst we also used *de novo*prediction. The functional annotation quality in terms of genes having at least a GO term (85.9%) is comparable to recent chromosome-scale genomes (Recuerda et al., 2021).

The mitogenome of *P. mauretanicus* spans 19,885 bp, exhibiting the same order and the *nad6* gene duplica-

tion observed in *P. lherminieri* (Torres et al. 2018). We did not find any *cob*duplication as it occurs in the Diomedeidae family (Abbott, Double, Trueman, Robinson, & Cockburn, 2005). Our result supports Torres et al. (2018) conclusion that *nad6* duplication could be widespread in Procellariiformes, and, like *cob* , could have undergone various events of deletion or addition during the diversification of the order. Nevertheless, since some of the reported duplications could be artificial (Formenti et al., 2021; Urantowka, Kroczak, & Mackiewicz, 2020), to fully identify the true number of gene duplications/deletions will require additional and specific experimental analyses.

Heterozygosity levels and historical demography

Current level of intraspecific heterozygosity is a relevant parameter to determine the adaptive capacity of a population (or species) (Orsted, Hoffmann, Sverrisdottir, Nielsen, & Kristensen, 2019). Since the Balearic shearwater is categorised as Critically Endangered by the IUCN, we could naively expect low heterozygosity levels in the species when compared to other Procellariiformes. However, the fossil record suggests that the Balearic shearwater had a very large population (>30.000 pairs) until the arrival of human settlers in the Balearic Islands (Alcover, J.A., Bover, P., Segui, 1991), which hunted shearwaters (Ramis, 2018) and introduced invasive mammals that also predated on them (Pinya & Carretero, 2011). In line with Genovart, Oro, Juste, & Bertorelle, (2007) results using mtDNA markers, we observed relatively high genome-wide heterozygosity levels, suggesting that the very recent demographic decline in the species is not yet visible in its genetic diversity.

Regarding the historical demography, our PSMC' analysis shows an increase in N_e in the Balearic shearwater from around 1 Mya to later expand to reach high population sizes, until around 150,000 ya, when it suddenly suffered a sharp decline, resulting in lower N_e values maintained until 10,000 years ago. Since current PSMC' analysis is based on the analysis of a single genome, we could not reliably infer more recent events (Schiffels & Wang, 2020). Pimiento et al., 2017 had shown that the Plio-Pleistocene eustatic variations resulted in a loss of neritic zones as sea level regressed, this may represent a loss of coastal habitat availability, which added to other oceanographic alterations (changes in ocean circulation or productivity) may have been the drivers of great population losses in marine megafauna, including seabirds. In the case of the Balearic shearwater, the PSMC' analysis shows an abrupt decay ofN_e associated with a long period of low sea level during the Penultimate Glacial Period (˜194-135 kya) which may have resulted in an important loss of neritic zones.

Here, we have observed a negative correlation between heterozygosity and body size within the Procellariiformes, where small-bodied species (*O. oceanicus* , *F. grallaria* , *H. tethys* and *P. urinatrix*) have higher heterozygosities than large-bodied species (*F. glacialis* , *C. borealis* , *P. mauretanicus* and*T. chlororhynchos*). Although controversial, contrasting heterozygosity levels between species with different body-sizes has also been reported in different species including Procellariiformes (Estandia et al. 2021; and references therein). Since body-size also correlates with population size and with other life-history traits, current data does not allow to determine the biological meaning of such correlation effect (Estandia et al., 2021; Mackintosh et al., 2019).

In view of the critical population declines affecting the Balearic shearwater populations, understanding its impacts on current genetic diversity of the species and among colonies will be crucial to assess the conservation status of the Balearic shearwater. Future ongoing research, using a more powerful population genomics approach, will allow to reconstruct the most recent demographic history of the species and to test the fossil-based hypotheses of a recent loss of population due to human colonization of the island, as well as why heterozygosity values have not decayed.

Phylogeny of Procellariiformes

The study of the evolutionary relationships within the order Procellariiformes had until recently been based mainly in the phylogenetic analyses of a single gene, the mitochondrial cytochrome b or on supertree approaches combining life history, morphological and sequence data (Kennedy & Page, 2002; Nunn & Stanley, 1998; Penhallurick & Wink, 2016). However, these approaches did not show enough resolution for this group, leaving several open questions. The main points that remain contentious are: 1) which family is the sister to

the rest of the Procellariiformes (Diomedeidae or Hydrobatidae), 2) which is the phylogenetic position of the diving petrels (*Pelecanoides sp.*) and whether they should be placed on their own family, 3) the monophyly of the storm petrels as well as the phylogenetic relationships among the speciose Procellariidae (J. J. Austin, Bretagnolle, & Pasquet, 2004; Brown et al., 2011; Obiol et al., 2020; Welch, Olson, & Fleischer, 2014). More recently, the first study to use genomic data to resolve the backbone Procellariiformes phylogeny (Estandia et al., 2021) reported a well-resolved phylogeny of 51 species using 4,365 ultraconserved elements (UCEs). This phylogeny recovered the albatrosses (Diomedeidae) as the sister group to the rest of Procellariiformes, the diving petrels included within Procellariidae, and the storm petrels constituting a paraphyletic group with Oceanitidae and Hydrobatidae being two separate monophyletic groups, and Hydrobatidae as sister group of Procellariidae. Our phylogenomic results using a smaller taxon sampling but a more extensive phylogenomic dataset (of up to 6,172 genes), agrees with those of (Estandia et al., 2021), supporting that these phylogenetic relationships are definitive.

Adaptation to a pelagic lifestyle in Procellariiformes

Our selection inference uncovered 20 genes evolving under positive selection in Procellariiformes, being therefore candidates to be actively involved in the adaptation of the order to a pelagic lifestyle. Indeed, the GO's Enrichment analysis of these genes reveals biological processes related to striated muscle cell differentiation, response to starvation, and nutrient reservoir activity, that may be related to the high energy expenditure during the vast distances they cover in the open ocean, while visual related genes could be related with underwater vision to fish and night vision (Hayes, Martin, & Brooke, 1991; Martin & De, 1991; Mitkus, Nevitt, Danielsen, & Kelber, 2016). Positive selection of genes related to natriuresis also makes sense for Procellariiformes since this biological process plays a key role to maintain the osmotic equilibrium in a sodium-rich environment like the ocean (Gutierrez, 2014; "Water and Salt Balance in Seabirds," 2001), which Procellariiformes perform thanks to the development of salt glands (modified nasal glands engaged in secretion of salts). Olfactory receptors, also found here among enriched GOs of positively selected genes, showed signature of adaptive evolution in shearwaters (C. Silva et al., 2020), and are crucial to Procellariiformes for navigation (Gagliardo et al., 2013; Padget, Dell'Ariccia, Gagliardo, Gonzalez-Solis, & Guilford, 2017; Pollonara et al., 2015), partner recognition and mating (Bonadonna & Nevitt, 2004; Hoover et al., 2018; Strandh et al., 2012), finding their own burrows (Bonadonna & Bretagnolle, 2002) or foraging (Bastos et al., 2020; Nevitt, 2008; Nevitt, Veit, & Kareiva, 1995; Yung, Sin, Cloutier, Nevitt, & Edwards, 2019).

We also inferred genes with intensified natural selection in Procellariiformes, for which the GOs annotations (Table S8) are similar and coherent to those in the candidate set of genes with positive selection in all tubenoses, or, in other words, related to the adaptation of the order to a pelagic lifestyle. For example, molecular functions such as sensory perception of sound, smell and chemical stimulus, neurological system process, camera-type eye development are related with oceanic navigation. On the other hand, functions such as homeostatic process, renal system, renal response, chloride transport and regulation of ion transport point to the need of maintaining osmotic equilibrium. We also found intensified natural selection in genes participating in functions related to immune response (like inflammatory response, defense response, response to wounding, wound healing, positive regulation of phagocytosis, etc.), accompanied by a relaxation of natural selection in regulators of blood constituents, induction of bacterial agglutination, regulation of antigen processing and presentation, viral budding via host ESCRT complex or macrophage antigen processing and presentation. As Procellariiformes exposure to parasites is high (Khan et al., 2019) and their life-history traits favour parasite maintenance within populations (McCoy et al., 2016), the tuning between the intensification and relaxation of natural selection in multiple biological processes and molecular functions related with immune response would have emerged following an arms race-like model. For example, as many parasites of tubenoses are blood-feeding, the intensified natural selection on the thrombin-activated receptor signaling pathway (GO:0070493), may be an evolutionary response to counter the anticoagulant activity that most blood-feeding parasites present (Bensaoud et al., 2018).

Among the gene families expanded in the branch of Procellariiformes, the one encoding olfactory receptors

is remarkable as it is coherent not only with the finding of a gene with positive selection in all Procellariiformes with the same functional annotation (g16276.t1 in the*P. mauretanicus* reference annotation, Table 4) but also with 3 olfactory receptors genes with intensification selection in the same branch (g14377.t1, g16276.t1 and g17936.t1 in *P. mauretanicus*reference annotation, Table S8). This triple evidence highlights the importance of how the adaptation to a pelagic life resulted in the enhancement of the olfactory function in Procellariiformes, as we discussed already in the paragraphs above. Moreover, Silva et al. 2020 obtained similar results of positive selection in olfactory genes in*C. borealis* . Physiologically, tubenoses have one of the largest olfactory bulb to brain size (OB) ratio of all birds (Cobb, 1968)

Conclusions

Our study highlights the utility of the hybrid assembly strategy using Illumina and ONT at recovering high quality genome assemblies, especially regarding contiguity and completeness. Comparative genomics analyses identified candidate genes under selection to have played a major role in the adaptation of the Procellariiformes to a pelagic lifestyle such as changes in sensory perception, navigation, natriuresis and physiological adaptations. Regarding the phylogeny of Procellariiformes, our results gave full support to recent genomic based hypotheses in which albatrosses (Diomedeidae) are sister to the rest of Procellariiformes, storm petrels are paraphyletic and diving petrels are included within Procellariidae. The high-quality genome presented in this work will be a great tool for future population genomic analyses, that will reveal with more precision the genetic variability of the species, its recent demographic history and the potential introgression with its sister species, the Mediterranean shearwater (*P. yelkouan*). The data obtained will be of great help in future proposals of conservation and management plans for the species.

Acknowledgements

We are grateful to David Garcia and Maite Louzao for kindly providing samples and the Govern Illes Balears for research permits (CEP19/2019). This research was supported by Fundacion Banco Bilbao Vizcaya (Spain), Project 062-17; by the Ministerio de Economia y Competitividad of Spain, projects CGL2016-78530-R, PGC2018-093924-B-100 and PID2019-103947GB.

Author contributions

JFO, JGS, MR and JR, conceived the study. CCC, JFO, MG and JGS, performed samplings. CCC, JFO performed wet lab work. CCC, JFO and JV performed the bioinformatic analyses. CCC, JFO, MR and JR, interpreted the genomic data. MG, JGS discussed on biological data. CCC, JFO, MR and JR drafted the first version of the manuscript. All authors revised and approved the final version of the manuscript.

ORCID

Cristian Cuevas-Caballe https://orcid.org/0000-0003-4292-9777

Joan Ferrer Obiol https://orcid.org/0000-0002-1184-5434

Joel Vizueta https://orcid.org/0000-0003-0139-3013

Meritxell Genovart https://orcid.org/0000-0003-2919-1288

Jacob Gonzalez-Solis https://orcid.org/0000-0002-8691-9397

Marta Riutort https://orcid.org/0000-0002-2134-7674

Julio Rozas https://orcid.org/0000-0002-6839-9148

DATA AVAILABILITY STATEMENT

The whole-genome shotgun project has been deposited at DDBJ/ENA/GenBank under the Bioproject ID PRJNA780920, and BioSample ID SUB10672806. The raw reads are also included in the Bioproject repository. Other relevant datasets, such as those including the structural and functional annotations, are available

in https://github.com/molevol-ub/Puffinus_mauretanicus_genome, and in the Supplementary Material online.

References

Abbott, C. L., Double, M. C., Trueman, J. W. H., Robinson, A., & Cockburn, A. (2005). An unusual source of apparent mitochondrial heteroplasmy: duplicate mitochondrial control regions in Thalassarche albatrosses. *Molecular Ecology* , *14* (11), 3605–3613. https://doi.org/10.1111/J.1365-294X.2005.02672.X

Alcover, J.A., Bover, P., Segui, B. (1991). No Title. In*Paleoecologia de les illes. In:Alcover, J.A. (Ed.), Ecologia de les Illes, Monografies de la Societatd'Histo'ria Natural de les Balears*(pp. 169–204).

Allendorf, F. W. (2017). Genetics and the conservation of natural populations: allozymes to genomes. *Molecular Ecology* ,*26* (2), 420–430. https://doi.org/10.1111/mec.13948

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology* , *215* (3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Arcos. (2011). Arcos, J.M. (compiler) 2011. International species action plan for the Balearic shearwater, Puffinus mauretanicus. SEO/BirdLife & BirdLife International.

Arcos, Arroyo, Becares, Mateos-Rodriguez, Rodriguez, Munoz, ... Oro. (2012). Arcos, J.M., Arroyo, G.M., Becares, J., Mateos-Rodriguez, M., Rodriguez, B., Munoz, A.R. et al. (2012) New estimates at sea suggest a larger global population of the Balearic Shearwater Puffinus mauretanicus. In *Proceedings of the 13th Medmaravis Pan-Mediterranean Symposium* .

Arcos, Louzao, & Oro. (2008). Fisheries ecosystem impacts and management in the Mediterranean: seabirds point of view. *American Fisheries Society Symposium* , 587–596.

Arroyo, G. M., Mateos-Rodriguez, M., Munoz, A. R., De La Cruz, A., Cuenca, D., & Onrubia, A. (2016). New population estimates of a critically endangered species, the Balearic Shearwater Puffinus mauretanicus, based on coastal migration counts. *Bird Conservation International* , *26* (1), 87–99. https://doi.org/10.1017/S095927091400032X

Austin, J. J., Bretagnolle, V., & Pasquet, E. (2004). A Global Molecular Phylogeny of the Small Puffinus Shearwaters and Implications for Systematics of the Little-Audubon's Shearwater Complex. *The Auk* , *121* (3), 847–864. https://doi.org/10.1093/AUK/121.3.847

Austin, R. E., Wynn, R. B., Votier, S. C., Trueman, C., McMinn, M., Rodriguez, A., ... Guilford, T. (2019). Patterns of at-sea behaviour at a hybrid zone between two threatened seabirds.*Scientific Reports* , *9* (1). https://doi.org/10.1038/s41598-019-51188-8

Bastos, R., Martins, B., Cabral, J. A., Ceia, F. R., Ramos, J. A., Paiva, V. H., ... Santos, M. (2020). Oceans of stimuli: an individual-based model to assess the role of olfactory cues and local enhancement in seabirds' foraging behaviour. *Animal Cognition 2020 23:4* , *23* (4), 629–642. https://doi.org/10.1007/S10071-020-01368-1

Bensaoud, C., Nishiyama, M. Y., Ben Hamda, C., Lichtenstein, F., Castro De Oliveira, U., Faria, F., ... Chudzinski-Tavassi, A. M. (2018). De novo assembly and annotation of Hyalomma dromedarii tick (Acari: Ixodidae) sialotranscriptome with regard to gender differences in gene expression. *Parasites & Vectors 2018 11:1* , *11* (1), 1–16. https://doi.org/10.1186/S13071-018-2874-9

Bernt, M., Donath, A., Juhling, F., Externbrink, F., Florentz, C., Fritzsch, G., ... Stadler, P. F. (2013). MITOS: Improved de novo metazoan mitochondrial genome annotation. *Molecular Phylogenetics and Evolution* , *69* (2), 313–319. https://doi.org/10.1016/j.ympev.2012.08.023

Bonadonna, F., & Bretagnolle, V. (2002). Smelling home: a good solution for burrow-finding in nocturnal petrels? *Journal of Experimental Biology* , *205* (16), 2519–2523. https://doi.org/10.1242/JEB.205.16.2519

Bonadonna, F., & Nevitt, G. A. (2004). Partner-specific odor recognition in an Antarctic seabird. *Science (New York, N.Y.)* ,*306* (5697), 835. https://doi.org/10.1126/SCIENCE.1103001

Boutet, E., Lieberherr, D., Tognolli, M., Schneider, M., Bansal, P., Bridge, A. J., ... Xenarios, I. (2016). Uniprotkb/swiss-prot, the manually annotated section of the uniprot knowledgebase: How to use the entry view. In *Methods in Molecular Biology* (Vol. 1374, pp. 23–54). Humana Press Inc. https://doi.org/10.1007/978-1-4939-3167-5_2

Braun, E. L., Kimball, R. T., Han, K. L., Iuhasz-Velez, N. R., Bonilla, A. J., Chojnowski, J. L., ... Yuri, T. (2011). Homoplastic microinversions and the avian tree of life. *BMC Evolutionary Biology* , *11* (1). https://doi.org/10.1186/1471-2148-11-141

Brooke, M. (2004). Albatrosses and petrels across the world, 499.

Brown, R. M., Jordan, W. C., Faulkes, C. G., Jones, C. G., Bugoni, L., Tatayah, V., ... Nichols, R. A. (2011). Phylogenetic Relationships in Pterodroma Petrels Are Obscured by Recent Secondary Contact and Hybridization. *PLOS ONE* , *6* (5), e20350. https://doi.org/10.1371/JOURNAL.PONE.0020350

Bruniche-Olsen, A., Kellner, K. F., Belant, J. L., & DeWoody, J. A. (2021). Life-history traits and habitat availability shape genomic diversity in birds: implications for conservation. *Proceedings of the Royal Society B* , *288* (1961). https://doi.org/10.1098/RSPB.2021.1441

C. Silva, M., Chibucos, M., Munro, J. B., Daugherty, S., Coelho, M. M., & C. Silva, J. (2020). Signature of adaptive evolution in olfactory receptor genes in Cory's Shearwater supports molecular basis for smell in procellariiform seabirds. *Scientific Reports 2020 10:1* ,*10* (1), 1–11. https://doi.org/10.1038/s41598-019-56950-6

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics* , *10* (1), 1–9. https://doi.org/10.1186/1471-2105-10-421

Chan, P. P., & Lowe, T. M. (2019). tRNAscan-SE: Searching for tRNA genes in genomic sequences. In *Methods in Molecular Biology* (Vol. 1962, pp. 1–14). Humana Press Inc. https://doi.org/10.1007/978-1-4939-9173-0_1

Claramunt, S., & Cracraft, J. (2015). Evolutionary Ecology: A new time tree reveals Earth history's imprint on the evolution of modern birds.*Science Advances* , *1* (11). https://doi.org/10.1126/SCIADV.1501005/SUPPL_FILE/1501005_SM.PDF

Cobb, S. (1968). The Size of the Olfactory Bulb in 108 Species of Birds.*The Auk* , *85* (1), 55–61. https://doi.org/10.2307/4083624

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., ... Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics* , *27* (15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

De Coster, W., D'Hert, S., Schultz, D. T., Cruts, M., & Van Broeckhoven, C. (2018). NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* , *34* (15), 2666–2669. https://doi.org/10.1093/bioinformatics/bty149

Dierckxsens, N., Mardulyn, P., & Smits, G. (2017). NOVOPlasty: De novo assembly of organelle genomes from whole genome data. *Nucleic Acids Research* , *45* (4), 18. https://doi.org/10.1093/nar/gkw955

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., ... Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* , *29* (1), 15–21. https://doi.org/10.1093/bioinformatics/bts635

Dussex, N., van der Valk, T., Morales, H. E., Wheat, C. W., Diez-del-Molino, D., von Seth, J., ... Dalen, L. (2021). Population genomics of the critically endangered kākāpō. *Cell Genomics* ,*1* (1), 100002.

https://doi.org/10.1016/J.XGEN.2021.100002

Emms, D. M., & Kelly, S. (2019). OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biology* , *20* (1), 1–14. https://doi.org/10.1186/s13059-019-1832-y

Estandía, A., Chesser, R. T., James, H. F., Levy, M. A., Obiol, J. F., Bretagnolle, V., ... Welch, A. J. (2021). Substitution Rate Variation in a Robust Procellariiform Seabird Phylogeny is not Solely Explained by Body Mass, Flight Efficiency, Population Size or Life History Traits. *bioRxiv* , 2021.07.27.453752. https://doi.org/10.1101/2021.07.27.453752

Falcon, S., & Gentleman, R. (2007). Using GOstats to test gene lists for GO term association. *Bioinformatics* , *23* (2), 257–258. https://doi.org/10.1093/bioinformatics/btl567

Feng, S., Stiller, J., Deng, Y., Armstrong, J., Fang, Q., Reeve, A. H., ... Zhang, G. (2020). Dense sampling of bird diversity increases power of comparative genomics. *Nature* , *587* (7833), 252–257. https://doi.org/10.1038/s41586-020-2873-9

Foote, A. D., Liu, Y., Thomas, G. W. C., Vinař, T., Alföldi, J., Deng, J., ... Gibbs, R. A. (2015). Convergent evolution of the genomes of marine mammals. *Nature Genetics 2015 47:3* , *47* (3), 272–275. https://doi.org/10.1038/ng.3198

Formenti, G., Rhie, A., Balacco, J., Haase, B., Mountcastle, J., Fedrigo, O., ... Bukhman, Y. (2021). Complete vertebrate mitogenomes reveal widespread repeats and gene duplications. *Genome Biology* , *22* (1), 1–22. https://doi.org/10.1186/S13059-021-02336-9/FIGURES/5

Fu, L., Niu, B., Zhu, Z., Wu, S., & Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* , *28* (23), 3150–3152. https://doi.org/10.1093/bioinformatics/bts565

Funk, W. C., McKay, J. K., Hohenlohe, P. A., & Allendorf, F. W. (2012, September 1). Harnessing genomics for delineating conservation units. *Trends in Ecology and Evolution* . Elsevier Current Trends. https://doi.org/10.1016/j.tree.2012.05.012

Gagliardo, A., Bried, J., Lambardi, P., Luschi, P., Wikelski, M., & Bonadonna, F. (2013). Oceanic navigation in Cory's shearwaters: Evidence for a crucial role of olfactory cues for homing after displacement. *Journal of Experimental Biology* , *216* (15), 2798–2805. https://doi.org/10.1242/JEB.085738

Gan, H. M., Falk, S., Morales, H. E., Austin, C. M., Sunnucks, P., & Pavlova, A. (2019). Genomic evidence of neo-sex chromosomes in the eastern yellow robin. *GigaScience* , *8* (9), 1–10. https://doi.org/10.1093/gigascience/giz111

Genovart, M., Arcos, J. M., Álvarez, D., McMinn, M., Meier, R., B. Wynn, R., ... Oro, D. (2016). Demography of the critically endangered Balearic shearwater: the impact of fisheries and time to extinction. *Journal of Applied Ecology* , *53* (4), 1158–1168. https://doi.org/10.1111/1365-2664.12622

Genovart, M., Juste, J., Contreras-Díaz, H., & Oro, D. (2012). Genetic and phenotypic differentiation between the critically endangered balearic shearwater and neighboring colonies of its sibling species. *Journal of Heredity* , *103* (3), 330–341. https://doi.org/10.1093/jhered/ess010

Genovart, M., Juste, J., & Oro, D. (2005). Two sibling species sympatrically breeding: A new conservation concern for the critically endangered Balearic shearwater. *Conservation Genetics* ,*6* (4), 601–606. https://doi.org/10.1007/s10592-005-9010-z

Genovart, M., Oro, D., Juste, J., & Bertorelle, G. (2007). What genetics tell us about the conservation of the critically endangered Balearic Shearwater? *Biological Conservation* , *137* (2), 283–293. https://doi.org/10.1016/J.BIOCON.2007.02.016

Gower, G., Tuke, S., Rohrlach, A. B., Soubrier, J., Llamas, B., Bean, N., & Cooper, A. (2018). Population size history from short genomic scaffolds: how short is too short? *bioRxiv* , 382036. https://doi.org/10.1101/382036

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., ... Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* , *29* (7), 644–652. https://doi.org/10.1038/nbt.1883

Gutiérrez, J. S. (2014). Living in Environments with Contrasting Salinities: A Review of Physiological and Behavioural Responses in Waterbirds. *https://doi.org/10.13157/arla.61.2.2014.233* ,*61* (2), 233–256. https://doi.org/10.13157/ARLA.61.2.2014.233

Hackett, S. J., Kimball, R. T., Reddy, S., Bowie, R. C. K., Braun, E. L., Braun, M. J., ... Yuri, T. (2008). A phylogenomic study of birds reveals their evolutionary history. *Science* ,*320* (5884), 1763–1768. https://doi.org/10.1126/science.1157704

Hayes, B., Martin, G. R., & Brooke, M. de L. (1991). Novel Area Serving Binocular Vision in the Retinae of Procellariiform Seabirds.*Brain, Behavior and Evolution* , *37* (2), 79–84. https://doi.org/10.1159/000114348

Hoang, D. T., Chernomor, O., Von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: Improving the ultrafast bootstrap approximation.*Molecular Biology and Evolution* , *35* (2), 518–522. https://doi.org/10.1093/molbev/msx281

Hoover, B., Alcaide, M., Jennings, S., Sin, S. Y. W., Edwards, S. V., & Nevitt, G. A. (2018). Ecology can inform genetics: Disassortative mating contributes to MHC polymorphism in Leach's storm-petrels (Oceanodroma leucorhoa). *Molecular Ecology* , *27* (16), 3371–3385. https://doi.org/10.1111/MEC.14801

IUCN 2021. The IUCN Red List of Threatened Species. Version 2021-1. https://www.iucnredlist.org. (n.d.).

Jarvis, E. D., Mirarab, S., Aberer, A. J., Li, B., Houde, P., Li, C., ... Zhang, G. (2014). Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* , *346* (6215), 1320–1331. https://doi.org/10.1126/science.1253451

Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., ... Hunter, S. (2014). InterProScan 5: Genome-scale protein function classification. *Bioinformatics* , *30* (9), 1236–1240. https://doi.org/10.1093/bioinformatics/btu031

Jurka, J., Kapitonov, V. V., Pavlicek, A., Klonowski, P., Kohany, O., & Walichiewicz, J. (2005). Repbase Update, a database of eukaryotic repetitive elements. *Cytogenetic and Genome Research* ,*110* (1–4), 462–467. https://doi.org/10.1159/000084979

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., Von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods* , *14* (6), 587–589. https://doi.org/10.1038/nmeth.4285

Kennedy, M., & Page, R. D. M. (2002). Seabird Supertrees: Combining Partial Estimates of Procellariiform Phylogeny. *The Auk* ,*119* (1), 88–108. https://doi.org/10.1093/AUK/119.1.88

Khan, J. S., Provencher, J. F., Forbes, M. R., Mallory, M. L., Lebarbenchon, C., & McCoy, K. D. (2019). Parasites of seabirds: A survey of effects and ecological implications. *Advances in Marine Biology* , *82* , 1–50. https://doi.org/10.1016/BS.AMB.2019.02.001

Kosakovsky Pond, S. L., Frost, S. D. W., & Muse, S. V. (2005). HyPhy: Hypothesis testing using phylogenies. *Bioinformatics* ,*21* (5), 676–679. https://doi.org/10.1093/bioinformatics/bti079

Kriventseva, E. V., Kuznetsov, D., Tegenfeldt, F., Manni, M., Dias, R., Simão, F. A., & Zdobnov, E. M. (2019). OrthoDB v10: Sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. *Nucleic Acids Research* , *47* (D1), D807–D811. https://doi.org/10.1093/nar/gky1053

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms.*Molecular Biology and Evolution* , *35* (6), 1547–1549. https://doi.org/10.1093/molbev/msy096

Kumar, S., Stecher, G., Suleski, M., & Hedges, S. B. (2017). TimeTree: A Resource for Timelines, Timetrees, and Divergence Times.*Molecular Biology and Evolution* , *34* (7), 1812–1819. https://doi.org/10.1093/molbev/msx116

Laetsch, D. R., & Blaxter, M. L. (2017). BlobTools: Interrogation of genome assemblies. *F1000Research* , *6* , 1287. https://doi.org/10.12688/f1000research.12232.1

Leroy, T., Anselmetti, Y., Tilak, M.-K., Berard, S., Csukonyi, L., Gabrielli, M., ... Nabholz, B. (2019). A bird's white-eye view on avian sex chromosome evolution. *bioRxiv* , 505610. https://doi.org/10.1101/505610

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Retrieved from http://arxiv.org/abs/1303.3997

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* , *25* (14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, W., & Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences.*Bioinformatics* , *22* (13), 1658–1659. https://doi.org/10.1093/bioinformatics/btl158

Librado, P., Vieira, F. G., & Rozas, J. (2012). BadiRate: Estimating family turnover rates by likelihood-based methods.*Bioinformatics* , *28* (2), 279–281. https://doi.org/10.1093/bioinformatics/btr623

Liu, J., Wang, Z., Li, J., Xu, L., Liu, J., Feng, S., ... Zhou, Q. (2021). A new emu genome illuminates the evolution of genome configuration and nuclear architecture of avian chromosomes.*Genome Research* , *31* (3), gr.271569.120. https://doi.org/10.1101/GR.271569.120

Louzao, Arcos, Hyrenbach, Sola, D., & Oro. (2004). Resultados preliminares sobre el hábitat de alimentación de la Pardela Balear en el Levante Ibérico Peninsular. *Anuari Ornitològic de Les Balears: Revista D'observació Estudi I Conservació Dels Aucells* , 61–67.

Löytynoja, A. (2014). Phylogeny-aware alignment with PRANK.*Methods in Molecular Biology* , *1079* , 155–170. https://doi.org/10.1007/978-1-62703-646-7_10

Mackintosh, A., Laetsch, D. R., Hayward, A., Charlesworth, B., Waterfall, M., Vila, R., & Lohse, K. (2019). The determinants of genetic diversity in butterflies. *Nature Communications 2019 10:1* , *10* (1), 1–9. https://doi.org/10.1038/s41467-019-11308-4

Marçais, G., Yorke, J. A., & Zimin, A. (2015). QuorUM: An error corrector for Illumina reads. *PLoS ONE* , *10* (6). https://doi.org/10.1371/journal.pone.0130821

Martí, & Ruiz. (2004). Ruiz, A., Martí, R. (Eds.) (2004). La pardela balear. SEO/BirdLife-Conselleria de Medi Ambient del Govern de les Illes Balears. Madrid.

Martin, G. R., & De, M. (1991). The eye of a procellariiform seabird, the Manx shearwater, Puffinus puffinus: visual fields and optical structure. *Brain, Behavior and Evolution* , *37* (2), 65–78. https://doi.org/10.1159/000114347

McCoy, K. D., Dietrich, M., Jaeger, A., Wilkinson, D. A., Bastien, M., Lagadec, E., ... Lebarbenchon, C. (2016). The role of seabirds of the Iles Eparses as reservoirs and disseminators of parasites and pathogens. *Acta Oecologica (Montrouge, France)* , *72* , 98. https://doi.org/10.1016/J.ACTAO.2015.12.013

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., ... DePristo, M. A. (2010). The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* , *20* (9), 1297–1303. https://doi.org/10.1101/gr.107524.110

Mitkus, M., Nevitt, G. A., Danielsen, J., & Kelber, A. (2016). Vision on the high seas: spatial resolution and optical sensitivity in two procellariiform seabirds with different foraging strategies. *The Journal of Experimental Biology* , *219* (Pt 21), 3329–3338. https://doi.org/10.1242/JEB.140905

Nadachowska-Brzyska, K., Li, C., Smeds, L., Zhang, G., & Ellegren, H. (2015). Temporal Dynamics of Avian Populations during Pleistocene Revealed by Whole-Genome Sequences. *Current Biology* ,*25* (10), 1375. https://doi.org/10.1016/J.CUB.2015.03.047

Nawrocki, E. P., Kolbe, D. L., & Eddy, S. R. (2009). Infernal 1.0: Inference of RNA alignments. *Bioinformatics* , *25* (10), 1335–1337. https://doi.org/10.1093/bioinformatics/btp157

Nevitt, G. A. (2008). Sensory ecology on the high seas: the odor world of the procellariiform seabirds. *Journal of Experimental Biology* ,*211* (11), 1706–1713. https://doi.org/10.1242/JEB.015412

Nevitt, G. A., Veit, R. R., & Kareiva, P. (1995). Dimethyl sulphide as a foraging cue for Antarctic Procellariiform seabirds. *Nature 1995 376:6542* , *376* (6542), 680–682. https://doi.org/10.1038/376680ao

Nguyen, L. T., Schmidt, H. A., Von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* ,*32* (1), 268–274. https://doi.org/10.1093/molbev/msu300

Nunn, G. B., & Stanley, S. E. (1998). Body size effects and rates of cytochrome b evolution in tube-nosed seabirds. *Molecular Biology and Evolution* , *15* (10), 1360–1371. https://doi.org/10.1093/OXFORDJOURNALS.MOLBEV.A025864

Obiol, J. F., James, H., Chesser, R., Bretagnolle, V., Gonzales-Solis, J., Rozas, J., ... Riutort, M. (2020). Paleoceanographic changes in the late Pliocene promoted rapid diversification in pelagic seabirds. *Authorea Preprints* . https://doi.org/10.22541/AU.160253833.37656353/V1

Oro, D., Aguilar, J. S., Igual, J. M., & Louzao, M. (2004). Modelling demography and extinction risk in the endangered Balearic shearwater. *Biological Conservation* , *116* (1), 93–102. https://doi.org/10.1016/S0006-3207(03)00180-0

Ørsted, M., Hoffmann, A. A., Sverrisdóttir, E., Nielsen, K. L., & Kristensen, T. N. (2019). Genomic variation predicts adaptive evolutionary responses better than population bottleneck history. *PLoS Genetics* , *15* (6). https://doi.org/10.1371/JOURNAL.PGEN.1008205

Padget, O., Dell'Ariccia, G., Gagliardo, A., González-Solís, J., & Guilford, T. (2017). Anosmia impairs homing orientation but not foraging behaviour in free-ranging shearwaters. *Scientific Reports 2017 7:1* , *7* (1), 1–12. https://doi.org/10.1038/s41598-017-09738-5

Penhallurick, J., & Wink, M. (2016). Analysis of the taxonomy and nomenclature of the Procellariiformes based on complete nucleotide sequences of the mitochondrial cytochrome b gene. *http://dx.doi.org/10.1071/MU01060* , *104* (2), 125–147. https://doi.org/10.1071/MU01060

Peñalba, J. V., Deng, Y., Fang, Q., Joseph, L., Moritz, C., & Cockburn, A. (2020). Genome of an iconic Australian bird: High-quality assembly and linkage map of the superb fairy-wren (Malurus cyaneus). *Molecular Ecology Resources* , *20* (2), 560–578. https://doi.org/10.1111/1755-0998.13124

Pimiento, C., Griffin, J. N., Clements, C. F., Silvestro, D., Varela, S., Uhen, M. D., & Jaramillo, C. (2017). The Pliocene marine megafauna extinction and its impact on functional diversity. *Nature Ecology & Evolution 2017 1:8* , *1* (8), 1100–1106. https://doi.org/10.1038/s41559-017-0223-6

Pinya, S., & Carretero, M. A. (2011). The Balearic herpetofauna: a species update and a review on the evidence. *Acta Herpetologica* ,*6* (1), 59–80. https://doi.org/10.13128/ACTA_HERPETOL-9579

Pollonara, E., Luschi, P., Guilford, T., Wikelski, M., Bonadonna, F., & Gagliardo, A. (2015). Olfaction and topography, but not magnetic cues, control navigation in a pelagic seabird: displacements with shearwaters in the Mediterranean Sea. *Scientific Reports 2015 5:1* ,*5* (1), 1–10. https://doi.org/10.1038/srep16486

Prost, S., Armstrong, E. E., Nylander, J., Thomas, G. W. C., Suh, A., Petersen, B., ... Irestedt, M. (2019). Comparative analyses identify genomic features potentially involved in the evolution of birds-of-paradise. *GigaScience* , *8* (5), 1–12. https://doi.org/10.1093/gigascience/giz003

Prum, R. O., Berv, J. S., Dornburg, A., Field, D. J., Townsend, J. P., Lemmon, E. M., & Lemmon, A. R. (2015). A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature 2015 526:7574* , *526* (7574), 569–573. https://doi.org/10.1038/nature15697

Ramis, D. (2018). Animal Exploitation in the Early Prehistory of the Balearic Islands. *Journal of Island and Coastal Archaeology* ,*13* (2), 265–278. https://doi.org/10.1080/15564894.2017.1334721

Recuerda, M., Vizueta, J., Cuevas-Caballé, C., Blanco, G., Rozas, J., & Milá, B. (2021). Chromosome-Level Genome Assembly of the Common Chaffinch (Aves: Fringilla coelebs): A Valuable Resource for Evolutionary Biology. *Genome Biology and Evolution* , *13* (4). https://doi.org/10.1093/GBE/EVAB034

Robinson, J. A., Räikkönen, J., Vucetich, L. M., Vucetich, J. A., Peterson, R. O., Lohmueller, K. E., & Wayne1, R. K. (2019). Genomic signatures of extensive inbreeding in Isle Royale wolves, a population on the threshold of extinction. *Science Advances* , *5* (5), 757–786. https://doi.org/10.1126/sciadv.aau0757

Sanderson, M. J. (2003). r8s: Inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock.*Bioinformatics* , *19* (2), 301–302. https://doi.org/10.1093/bioinformatics/19.2.301

Schiffels, S., & Wang, K. (2020). MSMC and MSMC2: The Multiple Sequentially Markovian Coalescent. *Methods in Molecular Biology* ,*2090* , 147–166. https://doi.org/10.1007/978-1-0716-0199-0_7

Seppey, M., Manni, M., & Zdobnov, E. M. (2019). BUSCO: Assessing genome assembly and annotation completeness. In *Methods in Molecular Biology* (Vol. 1962, pp. 227–245). Humana Press Inc. https://doi.org/10.1007/978-1-4939-9173-0_14

Smit, A. F. ., Hubley, R., & Green, P. (n.d.). RepeatMasker.

Smith, M. D., Wertheim, J. O., Weaver, S., Murrell, B., Scheffler, K., & Kosakovsky Pond, S. L. (2015). Less is more: An adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Molecular Biology and Evolution* , *32* (5), 1342–1353. https://doi.org/10.1093/molbev/msv022

Storer, J., Hubley, R., Rosen, J., Wheeler, T. J., & Smit, A. F. (2021). The Dfam community resource of transposable element families, sequence models, and genome annotations. *Mobile DNA* ,*12* (1), 1–14. https://doi.org/10.1186/s13100-020-00230-y

Strandh, M., Westerdahl, H., Pontarp, M., Canbäck, B., Dubois, M. P., Miquel, C., … Bonadonna, F. (2012). Major histocompatibility complex class II compatibility, but not class I, predicts mate choice in a bird with highly developed olfaction. *Proceedings of the Royal Society B: Biological Sciences* , *279* (1746), 4457–4463. https://doi.org/10.1098/RSPB.2012.1562

Subramanian, S., Beans-Picón, G., Swaminathan, S. K., Millar, C. D., & Lambert, D. M. (2013). Evidence for a recent origin of penguins.*Biology Letters* , *9* (6). https://doi.org/10.1098/RSBL.2013.0748

Supple, M. A., & Shapiro, B. (2018). Conservation of biodiversity in the genomics era. *Genome Biology* , *19* (1), 1–12. https://doi.org/10.1186/s13059-018-1520-3

Torres, L., Welch, A. J., Zanchetta, C., Chesser, R. T., Manno, M., Donnadieu, C., … Pante, E. (2018). Evidence for a duplicated mitochondrial region in Audubon's shearwater based on MinION sequencing.*https://doi.org/10.1080/24701394.2018.1484116* , *30* (2), 256–263. https://doi.org/10.1080/24701394.2018.1484116

Urantówka, A. D., Kroczak, A., & Mackiewicz, P. (2020). New view on the organization and evolution of Palaeognathae mitogenomes poses the question on the ancestral gene rearrangement in Aves. *BMC Genomics 2020 21:1* , *21* (1), 1–25. https://doi.org/10.1186/S12864-020-07284-5

Wang, P., Burley, J. T., Liu, Y., Chang, J., Chen, D., Lu, Q., … Zhang, Z. (2021). Genomic Consequences of Long-Term Population Decline in Brown Eared Pheasant. *Molecular Biology and Evolution* ,*38* (1), 263–273. https://doi.org/10.1093/MOLBEV/MSAA213

Water and Salt Balance in Seabirds. (2001). *Biology of Marine Birds* , 485–502. https://doi.org/10.1201/9781420036305-17

Welch, A. J., Olson, S. L., & Fleischer, R. C. (2014). Phylogenetic relationships of the extinct St Helena petrel, Pterodroma rupinarum Olson, 1975 (Procellariiformes: Procellariidae), based on ancient DNA.*Zoological Journal of the Linnean Society* , *170* (3), 494–505. https://doi.org/10.1111/ZOJ.12078

Wertheim, J. O., Murrell, B., Smith, M. D., Pond, S. L. K., & Scheffler, K. (2015). RELAX: Detecting relaxed selection in a phylogenetic framework. *Molecular Biology and Evolution* ,*32* (3), 820–832. https://doi.org/10.1093/molbev/msu400

Wu, M., Chatterji, S., & Eisen, J. A. (2012). Accounting for alignment uncertainty in phylogenomics. *PLoS ONE* , *7* (1), 30288. https://doi.org/10.1371/journal.pone.0030288

Yamashina, Y., & Udagawa, T. (1954). The chromosomes of the streaked shearwater Puffinus leucomelas (Vieillot). *Journal of the Yamashina Institute for Ornithology* , *1* (5), 220–221.

Yung, S., Sin, W., Cloutier, A., Nevitt, G., & Edwards, S. V. (2019). Olfactory receptor subgenome and expression in a highly olfactory procellariiform seabird. *bioRxiv* , 723924. https://doi.org/10.1101/723924

Zhang, C., Rabiee, M., Sayyari, E., & Mirarab, S. (2018). ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* , *19* (6), 15–30. https://doi.org/10.1186/s12859-018-2129-y

Zhang, G., Li, C., Li, Q., Li, B., Larkin, D. M., Lee, C., ... Froman, D. P. (2014). Comparative genomics reveals insights into avian genome evolution and adaptation. *Science* , *346* (6215), 1311–1320. https://doi.org/10.1126/science.1251385

Zimin, A. V., Puiu, D., Luo, M. C., Zhu, T., Koren, S., Marçais, G., ... Salzberg, S. L. (2017). Hybrid assembly of the large and highly repetitive genome of Aegilops tauschii, a progenitor of bread wheat, with the MaSuRCA mega-reads algorithm. *Genome Research* ,*27* (5), 787–792. https://doi.org/10.1101/gr.213405.116

Figure legends

Figure 1. (a) Map depicting the known Balearic shearwater breeding colonies in the Balearic Islands. Circle size is proportional to population size as shown in the legend. Modified from Arcos (2011) (b) Snail plot (Challis et al. 2020) summarizing genome assembly statistics. From inside to outside, the light-grey spiral shows the cumulative scaffold count on a log scale with white scale lines depicting changes of order of magnitude. Dark-grey segments show the distribution of scaffold lengths, and the plot radius is scaled to the longest scaffold (shown in red). Orange and light-orange rings represent the N50 and N90 scaffold lengths, respectively. Blue and light-blue rings show GC, AT and N percentages along the genome assembly. (c) MSMC2 reconstruction of effective population size estimates (N_e) over time, estimated using generation time of 12.8 years and mutation rate of 2.89 x 10^{-9}substitutions per nucleotide per generation. Light-brown vertical bars represent interglacial periods. Upper panel represents global temperature changes as inferred from the EPICA (European Project for Ice Coring in Antarctica) Dome C ice core (Augustin et al. 2004). Lower panel represents sea level changes inferred from a stack of 57 globally distributed benthic δ18O records (Lisiecki & Raymo 2005).

Figure 2. Comparison of genome-wide heterozygosity among Procellariiformes. (a) Density plots showing the distribution of individual nucleotide diversity (π) values in nonoverlapping 25Kb windows for each of the eight Procellariiformes species with an available reference genome. Scientific names of large-bodied and small-bodied species are shown in green and orange, respectively. Color-scale represents π values tail probabilities as shown in the legend. The white line depicts median values and black lines depict 25th and 75th percentiles. (b) Density plots showing the distribution ofπ values in large-bodied and small-bodied species groups.

Figure 3. Ultrametric tree based on the 4D CDS ML tree calibrated with r8s. Minimum number of gains

(green) and losses (red) per branch are represented according to BadiRate analysis. Numbers in ancestral nodes and in the tips (in parenthesis) indicate the inferred number of genes. Illustrations of seabird species were reproduced with permission from Lynx Edicions and Martí Franch.

Hosted file

`Genome Assembly Form MER_20211125.docx` available at https://authorea.com/users/738972/articles/712898-the-genome-of-the-balearic-shearwater-puffinus-mauretanicus-a-critically-endangered-seabird-a-valuable-resource-for-evolutionary-and-conservation-genomics

Hosted file

`Table1.docx` available at https://authorea.com/users/738972/articles/712898-the-genome-of-the-balearic-shearwater-puffinus-mauretanicus-a-critically-endangered-seabird-a-valuable-resource-for-evolutionary-and-conservation-genomics

Hosted file

`Table2.docx` available at https://authorea.com/users/738972/articles/712898-the-genome-of-the-balearic-shearwater-puffinus-mauretanicus-a-critically-endangered-seabird-a-valuable-resource-for-evolutionary-and-conservation-genomics