Paleoceanographic changes in the late Pliocene promoted rapid diversification in pelagic seabirds

Joan Ferrer Obiol[1], Helen James[2], R. Chesser[2], Vincent Bretagnolle[3], Jacob Gonzales-Solis[1], Julio Rozas[1], Andreanna Welch[4], and Marta Riutort[1]

[1]Universitat de Barcelona Facultat de Biologia
[2]Smithsonian Institution
[3]Centre d'Études Biologiques de Chizé, CNRS & La Rochelle Université
[4]Durham University Department of Biosciences

October 12, 2020

Abstract

In marine environments, paleoceanographic changes can act as drivers of diversification and speciation, even in highly mobile marine organisms. Shearwaters are a group of pelagic seabirds with a well-resolved phylogeny that are globally distributed and show periods of both slow and rapid diversification. Using reduced representation sequencing data, we explored the role of paleoceanographic changes on diversification and speciation in these highly mobile pelagic seabirds. We performed molecular dating, applying a multispecies coalescent approach (MSC) to account for the high levels of incomplete lineage sorting (ILS). We identified a major effect of the Pliocene marine megafauna extinction, followed by a period of high dispersal and rapid speciation. Biogeographic analyses showed that dispersal appears to be favoured by surface ocean currents, and we found that founder and vicariant events are the main processes of diversification. Body mass appears to be a key phenotypic trait potentially under selection during shearwater diversification, and it shows significant associations with life strategies and local conditions. We also found incongruences between the current taxonomy and patterns of genomic divergence, suggesting revisions to alpha taxonomy. Globally, our findings extend our understanding on the drivers of speciation and dispersal of highly mobile pelagic seabirds and shed new light on the important role of paleoceanographic events.

JOAN FERRER-OBIOL[1,2,*], HELEN F. JAMES[3], R. TERRY CHESSER[3,4], VINCENT BRETAGNOLLE[5], JACOB GONZÁLEZ-SOLÍS[6,2], JULIO ROZAS[1,2], ANDREANNA J. WELCH[7], MARTA RIUTORT[1,2,*]

[1]*Departament de Genètica, Microbiologia i Estadística, Facultat de Biologia, Universitat de Barcelona, Barcelona, Catalonia, Spain*

[2]*Institut de Recerca de la Biodiversitat (IRBio), Barcelona, Catalonia, Spain*

[3]*Department of Vertebrate Zoology, National Museum of Natural History, Smithsonian Institution, Washington, DC, USA*

[4]*U.S. Geological Survey, Patuxent Wildlife Research Center, Laurel, MD, USA*

[5]*Centre d'Études Biologiques de Chizé, CNRS & La Rochelle Université, 79360, Villiers en Bois, France*

[6]*Departament de Biologia Evolutiva, Ecologia i Ciències Ambientals, Facultat de Biologia, Universitat de Barcelona, Barcelona, Catalonia, Spain*

[7]*Department of Biosciences, Durham University, Durham, UK*

Correspondence to be sent to: Departament de Genètica, Microbiologia i Estadística, Facultat de Biologia, Universitat de Barcelona, Barcelona, Catalonia, Spain;

**E-mail: joan.ferrer.obiol@gmail.com*and *mriutort@ub.edu*

ABSTRACT

In marine environments, paleoceanographic changes can act as drivers of diversification and speciation, even in highly mobile marine organisms. Shearwaters are a group of pelagic seabirds with a well-resolved phylogeny that are globally distributed and show periods of both slow and rapid diversification. Using reduced representation sequencing data, we explored the role of paleoceanographic changes on diversification and speciation in these highly mobile pelagic seabirds. We performed molecular dating, applying a multispecies coalescent approach (MSC) to account for the high levels of incomplete lineage sorting (ILS). We identified a major effect of the Pliocene marine megafauna extinction, followed by a period of high dispersal and rapid speciation. Biogeographic analyses showed that dispersal appears to be favoured by surface ocean currents, and we found that founder and vicariant events are the main processes of diversification. Body mass appears to be a key phenotypic trait potentially under selection during shearwater diversification, and it shows significant associations with life strategies and local conditions. We also found incongruences between the current taxonomy and patterns of genomic divergence, suggesting revisions to alpha taxonomy. Globally, our findings extend our understanding on the drivers of speciation and dispersal of highly mobile pelagic seabirds and shed new light on the important role of paleoceanographic events.

Keywords: seabirds; diversification; speciation; molecular dating; biogeography; taxonomy

INTRODUCTION

Speciation is a key evolutionary process that results from the independent evolution and adaptation of populations and ultimately acts as a major driver responsible for the generation of species-level biodiversity (Kopp, 2010; Schluter & Pennell, 2017). Species richness is unevenly distributed across the Tree of Life, and its current patterns of distribution result from biotic and abiotic processes that operate over space and time (Benton, 2009; Simpson, 1953; Vargas & Zardoya, 2014). Evidence for the mechanisms that promote population differentiation and speciation are currently better understood in terrestrial than in marine environments (Butlin et al., 2012; Coyne & Orr, 2004; Nosil, 2012), where the lack of obvious physical barriers would suggest that neutral processes of panmixia, or isolation-by-distance, will prevail, especially in highly mobile species (Moura et al., 2013). However, counterintuitive evidence of fine-scale differentiation among populations and species in a number of marine taxa has been described as the "marine species paradox" (Bierne, Bonhomme, & David, 2003; Palumbi, 1994). Thus, there is a need for explicit evaluations of the role of selective processes in driving patterns of differentiation in marine systems.

In species complexes that are geographically widespread, the gradual evolution of reproductive isolation in allopatry can make species delimitation challenging, especially in young radiations (Carstens, Pelletier, Reid, & Satler, 2013; Cutter, 2013). Many allospecies first tend to differ from their close relatives at traits subjected to sexual and other forms of social selection (Price, 2008; Seddon et al., 2013). When this occurs, our ability to delimit species may be further hampered by morphological stasis, especially when changes in ecological niche in allopatry are minimal (Fišer, Robinson, & Malard, 2018). In cases of morphological stasis and limited behavioural information, genomic data can provide informed hypotheses on species limits of allopatric taxa and can be conclusive in parapatric or sympatric taxa. Despite the extent of disagreement about how genomic data should be applied to species delimitation (Leaché, Zhu, Rannala, & Yang, 2018; Sukumaran & Knowles, 2017), agreement exists that genomic data can provide additional perspective on species limits when used together with other data types such as phenotypic and ecological information.

Seabirds of the order Procellariiformes present some of the most extreme examples of the marine speciation paradox. Procellariiformes are highly mobile pelagic seabirds with a high dispersal ability and perform some of the longest animal migrations on Earth (covering more than 120,000 km a year) (Shaffer et al., 2006; González-Solís, Croxall, Oro, & Ruiz, 2007; Weimerskirch, Delord, Guitteaud, Phillips, & Pinet, 2015). However, Procellariiformes also show high philopatry to their breeding grounds (Coulson, 2002), which is expected to limit gene flow and therefore reinforce genetic differentiation (Friesen, Burg, & McCoy, 2007).

Shearwaters are a monophyletic group in the family Procellariidae, and they offer an excellent case study for examining the mechanisms of population differentiation and speciation in marine environments. First, shearwaters are globally distributed and breed mostly in allopatry. Second, the current taxonomy recognises three genera and 30 species with a well-resolved phylogeny showing clear periods of rapid diversification (Ferrer-Obiol et al. under review). Third, the three recognised genera exhibit different ecologies and degrees of species richness. Fourth, their high mobility makes them an ideal model to evaluate the roles of founder events and vicariance using biogeographic analyses. Fifth, abiotic and biotic factors are known to promote speciation in the shearwaters and related Procellariiformes; for instance, paleoceanographic changes such as the Pleistocene climatic oscillations can act as historical drivers of speciation (Gómez-Díaz, González-Solís, Peinado, & Page, 2006; Silva et al., 2015) and intrinsic biotic factors such as different foraging strategies and allochrony can also promote speciation (Friesen, Smith, et al., 2007; Lombal, Wenner, Lavers, & Austin, 2018; Rayner et al., 2011). Sixth, species limits are controversial, mostly due to high morphological stasis (Austin, Bretagnolle, & Pasquet, 2004; Austin, 1996); indeed, only a few phenotypic traits, such as vocalisation characteristics, slight plumage colour differences and in particular, body size, may differ between closely related species. A comprehensive study using genomic data will assist in resolving species delimitation within the context of the factors that promote diversification and speciation.

To accurately relate historical environmental and oceanographic changes to the timing of speciation events, it is necessary to estimate accurate divergence times. Analyses based on concatenation can lead to biases in branch lengths and misleading age estimates, particularly at recent timescales (Angelis & Dos Reis, 2015; McCormack, Heled, Delaney, Peterson, & Knowles, 2011; Mendes & Hahn, 2016). For such events, the multispecies coalescent model (MSC) offers a more accurate solution by incorporating the effects of incomplete lineage sorting (ILS), which is likely the most common source of phylogenetic incongruence in rapid diversification events (Edwards et al., 2016; Maddison, 1997; Suh, Smeds, & Ellegren, 2015).

The reconstruction of ancestral ranges and evaluation of alternative biogeographic models are critical to our understanding of shearwater diversification throughout the world in light of environmental and oceanographic events. Of particular interest is the importance of founder events during the evolution of shearwaters. The foundation of colonies is believed to be a rare event in most seabird species despite their great potential for long-range dispersal (Milot, Weimerskirch, & Bernatchez, 2008). However, in several shearwater species, contemporary colony foundation events have been reported (Munilla, Genovart, Paiva, & Velando, 2016; Storey & Lien, 1985). Understanding the relative importance of founder and vicariant events during the evolution of shearwaters can have important implications for the conservation of these endangered pelagic seabirds.

Here, we use paired-end double-digest restriction site-associated DNA sequencing (PE-ddRAD-Seq) for almost all extant shearwater taxa to explore the drivers of diversification and speciation in this group of pelagic seabirds. Specifically, we produce the first time-calibrated shearwater species tree using the MSC to account for the high levels of ILS affecting the shearwater phylogeny. We then infer the biogeographic history of the group by estimating ancestral ranges and evaluating the roles of founder events, vicariance and surface ocean currents in driving their diversification. Furthermore, we explore the ecological forces responsible for the variability in a key phenotypic trait, body size. Finally, we assess the validity of the current taxonomy of the group by analysing genomic patterns of recent shared ancestry and differentiation among shearwater taxa.

MATERIAL AND METHODS

Sampling and sequence data generation

We collected 68 blood or tissue samples from 25 of the 32 recognised species of shearwaters (Gill, Donsker, & Rasmussen, 2020) (Table S1) representing all the major lineages in the group (Austin et al., 2004). Species that could not be included (*Puffinus heinrothi* , *P. bannermani* ,*P. bryani* , *P. myrtae* , *P. auricularis* , *P. persicus* and *P. subalaris*) breed in remote islands, have very limited distributions and/or are categorised as critically endangered by the IUCN Red List of Threatened Species (http://www.iucnredlist.org/). Sampling was conducted under permits issued by the relevant authorities. Sequence data for 51 of these samples were generated previously in a recent phylogenomic study (Ferrer-Obiol et al. under review).

For the new samples generated here, we extracted genomic DNA using the Qiagen DNeasy Blood and Tissue Kit according to the manufacturer's instructions (Qiagen GmbH, Hilden, Germany). We used a Qubit Fluorometer (Life Technologies) to quantify and standardise DNA concentrations of all samples at 10 ng/ul. Approximately 250 ng of genomic DNA of each sample was sent to the Genomic Sequencing and Analysis Facility, University of Texas at Austin, to perform ddRAD library preparation following the Peterson et al. (2012) protocol. DNA was fragmented using an uncommon cutter *EcoRI* and a common cutter *MspI* in a single reaction. Illumina adaptors containing sample-specific barcodes and Illumina indexes were ligated onto the fragments and four pools were produced differing by their Illumina index. Barcodes differed by at least two base pairs to reduce the chance of inaccurate barcode assignment. Pooled libraries were size selected (between 150 and 300 bp after accounting for adapter length) using a Pippin Prep size fractionator (Sage Science, Beverly, Ma). Libraries were amplified in a final PCR step prior to sequencing on a single lane of an Illumina HiSeq4000 platform using a 150-bp paired-end (PE) metric.

PE-ddRAD-Seq data filtering and assembly

Raw reads were demultiplexed and cleaned using process_radtags in Stacks v2.41 (Rochette, Rivera-Colón, & Catchen, 2019). To maximise the amount of biological information, we built loci using the forward reads with parameters optimised for shearwater data (see Ferrer-Obiol et al. under review*)* using the USTACKS-CSTACKS-SSTACKS core clustering algorithm. We used the TSV2BAM program to incorporate reverse reads by matching the set of forward read IDs in each locus. We then assembled a contig for each locus, called SNPs using the Bayesian genotype caller (BGC; Maruki & Lynch, 2015, 2017) and phased haplotypes using GSTACKS. Subsequently, we mapped the GSTACKS catalog to the Balearic shearwater (*Puffinus mauretanicus*) genome assembly (Cuevas et al., 2019) using BWA-MEM v. 0.7.17 (Li, 2013). We sorted and indexed the mapped reads using SAMtools v.1.4 (Li, 2011; Li et al., 2009) and integrated alignment positions to the catalog using STACKS-INTEGRATE-ALIGNMENTS (Paris, Stevens, & Catchen, 2017). Finally, we used the POPULATIONS program to filter SNP data requiring a minor allele frequency (MAF) above 5% and an observed heterozygosity below 50% to generate datasets for downstream analysis.

Species tree inference

To estimate a time-calibrated species tree for shearwaters, we applied the SNP-based MSC approach of (Stange, Sánchez-Villagra, Salzburger, & Matschiner, 2018) implemented in the SNAPP v.1.3. (Bryant, Bouckaert, Felsenstein, Rosenberg, & RoyChoudhury, 2012) package of the program BEAST 2 v.2.5.0 (R. Bouckaert et al., 2019). To prepare a suitable dataset for this method, we selected a maximum of two individuals per subspecies (51 individuals in total) and we exported called variants to variant call format (VCF). Because SNAPP assumes a single nucleotide substitution rate, we performed the analyses including only transitions to reduce heterogeneity in the evolutionary rate. We further processed the VCF file with VCFtools v.0.1.15 (Danecek et al., 2011) to include only biallelic SNPs without missing data, to mask genotypes if the per-sample read depth was below 5 or above 150, or if the genotype quality was below 30.

Finally, we selected a single SNP per ddRAD locus to remove potentially linked SNPs (minimum distance between SNPs > 500 bp). After filtering, we retained a dataset of 1397 transitions.

We followed recommendations of Stange et al. (2018) by constraining the root of the species tree to follow a normal distribution with a mean of 20.23 Mya and a standard deviation (SD) of 2 as reported by Ferrer-Obiol et al. (*under review*) based on three fossil calibrations (see calibration strategy B there-in) and a relaxed clock. SD was calculated to fit the posterior distribution for the root in Ferrer-Obiol et al. (*under review*). This divergence time estimate for the root was further supported by a global study on birds using relaxed clocks (Jetz, Thomas, Joy, Hartmann, & Mooers, 2012). As we were mainly interested in SNAPP's ability to estimate divergence times rather than the tree topology, we fixed the species tree topology to that inferred by Ferrer-Obiol et al. (*under review*) using UCE and ddRAD data. We also tested the robustness of divergence-time estimates by performing two additional analyses. Firstly, we explored the effects of fixing the topology by also performing the analysis without the topology being fixed. We also evaluated the use of fossil calibrations using three different calibration points based on those described in strategy B of Ferrer-Obiol et al. (*under review*). We used the ruby script snapp_prep.rb (https://github.com/mmatschiner/snapp_prep) to prepare the XML file for SNAPP analyses. For each analysis, we conducted three replicate runs, each with a run length of 500,000 Markov-chain Monte Carlo (MCMC) iterations. Convergence and stationarity were confirmed (effective sample sizes > 300) using Tracer v.1.7.1 (Rambaut, Drummond, Xie, Baele, & Suchard, 2018). The first 10% of each MCMC was discarded as burn-in, and posterior distributions of run replicates were merged to generate maximum-clade-credibility (MCC) trees with node heights set to mean age estimates with TreeAnnotator (Heled & Bouckaert, 2013). SNAPP trees were visualised in Densitree v.2.2.7 (Bouckaert, 2010).

The finite-sites model implemented in SNAPP allows the estimation of both branch lengths (times) and population sizes (ϑ) (Bryant et al., 2012). Because the Stange et al. (2018) approach only estimates a single value of ϑ for all branches, we also constructed a SNAPP phylogeny without any age constraint, in order to estimate ϑ values for each branch.

Ancestral range estimation

Biogeographic analyses were performed to estimate ancestral ranges and to examine patterns of shearwater dispersal across five broad areas. The five areas were chosen based on contemporary shearwater breeding ranges: Southern Australia and New Zealand (A), Southern Ocean (B), North and Tropical Pacific Ocean (C), Tropical Indian Ocean (D), and North Atlantic Ocean and Mediterranean Sea (E). We set the limit between areas A and B at the Subtropical Front (Sutton, 2001). The R package BioGeoBEARS v. 1.1.2 (Matzke, 2013) was used to estimate ancestral ranges using likelihood versions of three models: dispersal-extinction-cladogenesis (DEC; Ree & Smith, 2008), dispersal-vicariance (DIVA; Ronquist, 1997), and BayArea (Landis, Matzke, Moore, & Huelsenbeck, 2013), and the time-calibrated shearwater tree (Figure 1). We compared ancestral range estimates of these models with and without the founder-event speciation parameter (j) under two scenarios: one that allowed unrestricted dispersal between all areas and another that limited dispersal between areas connected by major surface ocean currents from the Pliocene to the present, when most of the shearwater diversification occurred (Figure 1). Corrected Akaike Information Criterion (AICc) and AICc weights were used to select the best-fit scenario for the models with and without the j parameter separately, because the DEC + j model has been criticised for not being statistically comparable to the DEC model (Ree & Sanmartín, 2018). To infer the ancestral range of the shearwaters' most recent common ancestor (MRCA), we used the ranges of the two most closely related outgroup lineages (for which no genetic data are available): genus*Procellaria* , and genera *Pseudobulweria* and *Bulweria*(Estandía, 2019). *Pseudobulweria rostrata* , *Bulweria bulwerii* , *Procellaria westlandica* and *Procellaria cinerea*were chosen because they represent the totality of ranges within their clades. Divergence times between the outgroups and shearwaters and among the outgroups were retrieved from the TimeTree database (Kumar, Stecher, Suleski, & Hedges, 2017). Outgroups were incorporated into the time-calibrated shearwater tree using the bind.tree function from the ape package (Paradis & Schliep, 2019) in R.

Phylogenetic comparative analyses

To evaluate potential predictors of body size variation in shearwaters, we retrieved data for four body size measures: 1) mean body mass; 2) range of body mass (maximum body mass - minimum body mass); 3) wing length; and 4) total body length, and five predictors: 1) minimum, 2) mean and 3) maximum breeding latitudes (in absolute values), 4) latitudinal range occupied by a species year-round (maximum latitude - minimum latitude where the species is present either during the breeding or the non-breeding period) and 5) migratory strategy (long-distance migrant, short distance migrant or dispersive/sedentary). Additionally, we retrieved wingspan measurements to obtain a mean body mass measure corrected by body surface (mean body mass / (body length x wingspan). Data were retrieved for all recognised species of shearwaters (Gill et al. 2020) with the exception of Heinroth's shearwater (*Puffinus heinrothi*), because no information about its phylogenetic placement is available. The majority of morphometric, distributional and behavioural data were retrieved from Birds of the World (Billerman, Keeney, Rodewald, & Schulenberg, 2020) and additional morphometric data were extracted from (Onley & Scofield, 2013). We performed phylogenetic generalised least squares regressions (PGLS) for all body size measures against each of the potential predictors using the R package caper (Orme et al., 2013) and we adjusted p-values by FDR correction for multiple testing. Due to the reduced number of species, we only performed univariate regressions to avoid overfitting (Mundry, 2014). Following recommendations of (Revell, 2010), we simultaneously estimated the λ parameter (Pagel, 1999) to account for deviations from a pure Brownian motion (BM). PGLS analyses were run using the time-calibrated tree with the species for which no sequencing data was available incorporated into the phylogeny using the bind.tip function from the R package phytools (Revell, 2012) according to the phylogenetic position and branch lengths from previous phylogenetic studies (Austin et al., 2004; Pyle, Welch, & Fleischer, 2011); (Martínez-Gómez, Matías-Ferrer, Sehgal, & Escalante, 2015); (Kawakami, Eda, Izumi, Horikoshi, & Suzuki, 2018). We estimated ancestral states for body size measures using the function fastAnc in the R package phytools and visualised the reconstructions with phenograms using the R package ggtree (Yu, Smith, Zhu, Guan, & Lam, 2017). We also reconstructed ancestral states for migratory behaviour using maximum likelihood (ML) with the function rerootingMethod in the R package phytools.

To evaluate the effect of life-history traits (LHT) on the nucleotide substitution rate and the equilibrium of GC content (GC*), we modelled the correlation between these two parameters, and their correlations with mean body mass and the number of breeding pairs as a multivariate Brownian motion in CoEvol (Lartillot & Poujol, 2011). We ran two independent Markov Chain Monte Carlo (MCMC) chains, and stopped the process after reaching convergence (effective sample size $> 1{,}000$ and discrepancy between chains < 0.05 for all statistics; 5,000 generations) using tracecomp from the Coevol package. The number of breeding pairs for each species were retrieved from Birds of the World (Billerman et al. 2020) and BirdLife International (2020).

The time-calibrated tree was also used to calculate evolutionary distinctness (ED) scores and EDGE scores (i.e., evolutionary distinctness and globally endangered status; Isaac, Turvey, Collen, Waterman, & Baillie, 2007), based on IUCN Red List of Threatened Species threat-status (GE, as of June 2020; http://www.iucnredlist.org/), calculated in the R package caper. EDGE scores for each species were calculated as follows: $\mathrm{EDGE} = \ln(1 + \mathrm{ED}) + \mathrm{GE} \times \ln(2)$.

Patterns of recent coancestry and sequence divergence

To explore congruence between the current shearwater taxonomy and the genetic structure among species, we used fineRADstructure v0.3.2 (Malinsky, Trucchi, Lawson, & Falush, 2018) to infer the shared ancestry among all individuals. fineRADstructure uses haplotype linkage information to derive a co-ancestry matrix based on the most recent coalescent events. We exported haplotypes for loci present in at least 75% of the individuals to RADpainter format using POPULATIONS, resulting in a set of haplotypes for 8,049 PE-ddRAD loci containing a total of 63,492 SNPs. RADpainter was used to infer a coancestry matrix and the fineSTRUCTURE MCMC clustering algorithm was used to assign individuals into clusters, with a burn-in period of 100,000 generations and an extra 100,000 MCMC iterations sampled every 1,000 generations.

To arrange the clusters based on their relationships within the coancestry matrix, we built a tree within fineSTRUCTURE using default parameters. To visualise the results, we used the R scripts fineradstructureplot.r and finestructurelibrary.r (available at*http://cichlid.gurdon.cam.ac.uk/fineRADstructure.html*).

As an additional approach to examining congruence between the current shearwater taxonomy and genomic divergence, we examined the distribution of pairwise genetic distances using loci present in at least 95% of the individuals (1,525 loci; 11,055 SNPs). Briefly, we exported variants into a VCF file using POPULATIONS in Stacks, we converted the VCF file into a DNAbin object using the R package vcfR (Knaus & Grunwald, 2017), and we calculated pairwise distances using the dist.dna function from the ape package in R.

RESULTS

We recovered an average of 1,227,032 (SD = 815,798) PE-ddRAD reads per sample (Table S1) that were assembled to an average of 24,621 loci per sample, with a mean coverage per sample of 39x (SD = 19). Locus length ranged from 140 to 239 bp with a median of 198 bp (SD = 25.5).

Bayesian divergence time estimation with SNP data

The SNAPP phylogeny revealed largely the same topology as a previous phylogenetic study based on the same data (Ferrer-Obiol et al. under review), except for the poorly supported relationship between*Ardenna* and *Puffinus* and the relationship between *A. grisea* and *A. tenuirostris* (Table S2; Figure S1). Both incongruences were already identified in the previous study using different methods and datasets, and were caused by high levels of ILS.

Using a constraint for the age of the root, we estimated the time-calibrated tree shown in Figure 1. The time to the most recent common ancestor (TMRCA) of *Puffinus* was the oldest among the three genera, estimated at 9.98 Mya (95% HPD: 12.26-7.65 Mya). The TMRCA of *Ardenna* was inferred to be 4.52 Mya (95% HPD: 5.64-3.53 Mya) and the TMRCA of *Calonectris* 3.54 Mya (95% HPD: 4.57-2.55 Mya). If the divergence times are accurate, then shearwater speciation increased during the Pliocene reaching a peak by the late Pliocene (˜2.58 Mya; Figure 1), when most of the modern biogeographical groups of shearwaters were already present.

Using the same three fossil calibrations (see Material and Methods), shearwater divergence times inferred using the MSC were 28-94% younger than those estimated by Ferrer-Obiol et al. (under review) using concatenation (Table S2). MSC analyses using these fossil calibrations resulted in slightly older estimates (8.9% older on average) compared to the same analyses using a single age constraint on the root (Figure S2, Table S2). Conversely, fixing the phylogeny had very little effect on age estimates (0.4% older on average).

The mean population size across all shearwater species estimated by SNAPP was N=63,555 individuals (95% HPD: 50,390–77,155) when assuming the lowest generation time estimated for a shearwater species (13 years; Genovart et al., 2016), and N=43,485 individuals (95% HPD: 34,477–52,790) when assuming the highest estimated value (19 years; Birdlife International 2020). However, SNAPP analysis without age constraints showed a notable variation in ϑ estimates even between sister species (Figure S3) suggesting frequent changes in population size in the evolutionary history of shearwaters.

Biogeographic Analysis

Under all tested models, ancestral range estimation analyses, including a dispersal matrix restricting dispersal between areas connected by main historical and present surface ocean currents, had lower AICc than models with an unrestricted dispersal matrix (Table 1). DIVALIKE and DEC models had lower AICc than BAYAREALIKE models, especially when the founder event parameter (j) was not included, suggesting that vicariance is an important mode of speciation in shearwaters. The slightly lower AICc for DIVALIKE models compared to DEC models further supported the importance of widespread vicariance. However, in models including founder event speciation, the j parameter ranged from 0.0874 to 0.1733 and the rate of range expansion (d) was an order of magnitude smaller, showing that founder events have a higher probability of explaining most of the data than range expansion. Indeed, the likelihood ratio test (LRT) between the best DIVALIKE and DIVALIKE + j models showed that DIVALIKE + j was strongly favoured ($P = 1.9 \times 10^{-5}$).

Under the best DIVALIKE + j model, the South Australia - New Zealand area showed the highest support as the ancestral region of shearwaters (marginal ML probability = 0.44), followed by the North and Tropical Pacific (0.33) (Figure 1 and Figure S4). The origin of *Ardenna*was also traced to the South Australia - New Zealand area (0.54). On the other hand, *Calonectris* had an unequivocal origin in the Northern Hemisphere (North Atlantic and North and Tropical Pacific = 0.45, North and Tropical Pacific = 0.45), whereas the ancestral area of the MRCA of*Puffinus* was estimated as either the North and Tropical Pacific (0.37), the South Australia - New Zealand area (0.27) or both (0.16).

Phylogenetic generalized least squares of body size

The PGLS analyses recovered several significant correlations (FDR < 0.05) between body size measures and the predictors (Figure 2a; Table S3). Mean body mass showed significant correlations with all predictors, suggesting that this trait is strongly influenced by ecological factors. Overall, migratory strategy and latitudinal range were the best predictors, suggesting that body size in shearwaters is associated with movement capacity. Indeed, migratory strategy explained 75% of the variance in mean body mass (Figure S5a; long-distance migrants were the heaviest and sedentary/dispersive species the lightest) and latitudinal range explained 67% of the variance in body mass range (Figure S5b shows the positive correlation between body mass range and latitudinal range occupied by a species year-round). Breeding latitude was also a good predictor of mean body mass, with the strongest correlation recovered for maximum breeding latitude (Figure S5c; adjusted R^2=0.212). As shown in the phenogram of ancestral state reconstructions for body mass in Figure 2b, striking differences in body mass between sister clades are common in shearwaters, showing that body mass changes may be important during speciation. The ancestral state reconstruction of migratory behaviour showed that the MRCAs of *Calonectris* and *Ardenna* were most likely long-distance migrants (Figure S6; marginal ML probability = 0.94 and 0.86, respectively). Conversely, the MRCA of *Puffinus* was most likely either a short-distance migrant or a sedentary species (marginal ML probability = 0.47 and 0.37, respectively).

Effects of life-history traits on the substitution rate

Results from a previous study found that rates of mitochondrial DNA (mtDNA) evolution were slower for larger taxa in the Procellariiformes (Nunn & Stanley, 1998), yet we did not find any significant correlations between the substitution rate of our PE-ddRAD-Seq dataset and the LHT (Table S4). This may have been influenced by the high variance in the substitution rates of our dataset. However, consistent with GC-biased gene conversion (gBGC): a recombination-associated mechanism that leads to the preferential fixation of G and C in AT/GC heterozygotes, we found that GC* had a positive significant correlation with the number of breeding pairs (correlation coefficient = 0.684, posterior probability = 0.94). These results are in agreement with the hypothesis that the impact of gBGC is strongest in species with high population sizes (Weber, Boussau, Romiguier, Jarvis, & Ellegren, 2014).

Genomic divergence and taxonomy

The fineRADStructure analysis identified three major clusters corresponding to the three shearwater genera (Figure 3). Further subdivisions within each group largely supported the most recent shearwater phylogeny (Ferrer-Obiol et al. under review), and all the species and subspecies included in the study were recovered as unique clusters by the fineStructure clustering algorithm (Lawson, Hellenthal, Myers, & Falush, 2012), except for *P. bailloni nicolae* and *P. bailloni dichrous* , *P. mauretanicus* and *P. yelkouan* , and*A. creatopus* and *A. carneipes* , that were in each case, clustered into a single population.

Overall, the distributions of genetic distances were consistent with the current taxonomy. However, the distributions of distances within and among species showed some overlap (Figure 4). The genetic distances between *A. creatopus* and *A. carneipes* , and between*P. mauretanicus* and *P. yelkouan* , were within the distribution of genetic distances within the same subspecies (first interval delimited in Figure 4b to ease qualitative comparison). In addition, the genetic distances between *P. boydi* and *P. baroli* , and between the different Atlantic *Calonectris* species were within the interval of genetic distances among different subspecies (second interval in Figure 4b).

DISCUSSION

This study presents a fundamental analysis of the drivers of diversification and speciation in a major group of seabirds, by constructing the first MSC time-calibrated species tree and biogeographical analysis for shearwaters based on a fully resolved phylogeny. This allowed us to explicitly explore the paleoceanographic events that may have driven the diversification of the group, as well as to infer their ancestral range using formal biogeographic analyses and evaluate the role of dispersal, vicariance and founder events in shearwater diversification. We also discuss the role of body size in shearwater diversification, and we consider potential ecological and evolutionary forces that may have shaped its evolution. Lastly, we used the evidence uncovered here to explain incongruences between the current taxonomy and the patterns of genomic divergence.

Limitations of divergence time estimation using SNAPP

A limitation of the SNAPP approach for absolute divergence time estimation is the assumption of equal and constant population sizes on all branches of the phylogeny (Stange et al., 2018). This assumption is clearly violated in our analysis, as shown in the SNAPP results without any age constraint (Figure S3). As a result, our divergence times might be slightly overestimated for lineages with larger population sizes than the overall estimation and vice versa. An additional limitation is the heterogeneity in substitution rates among shearwater lineages (Ferrer-Obiol et al. under review), which would likely benefit from the use of a relaxed clock (Drummond, Ho, Phillips, & Rambaut, 2006; Rannala & Yang, 2007) instead of the strict clock model implemented by (Stange et al., 2018). Nonetheless, previous analyses to select a clock model for this PE-ddRAD dataset showed that the strict clock model obtained the best marginal likelihoods (Ferrer Obiol et al. under review). We therefore do not expect this limitation to significantly reduce accuracy. Despite these limitations, given the relatively high overall population size estimated in this study, and given the shallow timescales encompassed by the shearwater phylogeny, we argue that the older divergence times estimated by concatenation analyses (Table S2) are most likely caused by a higher degree of node age error in the latter analyses, potentially caused by failing to fully consider the role of ILS (Angelis & Dos Reis, 2015; Ogilvie, Bouckaert, & Drummond, 2017; Ogilvie, Heled, Xie, & Drummond, 2016).

Biogeographic history of shearwaters

Our biogeographic analyses indicate that vicariance and founder events are probably the main mechanisms of speciation in shearwaters, as expected by their global distribution and high mobility. Unlike other Procellariiformes (Friesen, Smith, et al., 2007), sympatric speciation has not been described in shearwaters. Indeed, very few records of sister species inhabiting the same island exist in the wild and are limited to marginal overlaps between parapatric species (Navarro, Forero, et al., 2009). The biogeographic analyses suggest that shearwater dispersal is favoured by surface ocean currents; nevertheless, we cannot draw firm conclusions given the reduced differences in log-likelihood (< 3 units) between ancestral range estimation models with or without a dispersal matrix that restricted dispersal to areas connected by surface ocean currents (Table 1). Several studies have shown that winds are a major determinant of foraging ranges and migratory routes of seabirds, especially in the Procellariiformes (González-Solís et al., 2009; Weimerskirch, Louzao, de Grissac, & Delord, 2012). Winds are also a primary driver of surface ocean currents; hence, our study suggests that winds could also be an important determinant of species dispersal in the Procellariiformes.

Ancestral range estimation analyses inferred the South Australia - New Zealand area as the ancestral region of shearwaters with the highest support followed by the Northern and Tropical Pacific (Figure S4). The South Australia - New Zealand area is currently a hotspot of global seabird biodiversity (Croxall et al., 2012) and has the greatest number of shearwater species breeding in any single area (Dickinson & Remsen, 2013). On the other hand, the coast of California harbours the highest diversity of shearwater fossils from extinct species and some of the oldest ones (Miller, 1961). These observations suggest that the current biogeographic analyses represent a more probable hypothesis of the ancestral area of shearwaters than previous hypotheses, which suggested that the North Atlantic was the ancestral area based on the relatively rich shearwater fossil record in this area (Austin, 1996; Kuroda, 1954). The phylogenetic position of the oldest North Atlantic shearwater fossil species (*P. raemdonckii* and *P. arvernensis*) is still unclear (Olson, 1985) and the age of *P. micraulax,* which was believed to be the oldest shearwater fossil species (lower Miocene, Hawthorne Formation, Florida) is uncertain (Ferrer-Obiol et al., in review). Thus, earlier suggestions of the North Atlantic as the ancestral area of shearwaters may have been misled by these uncertainties in the fossil record.

The MRCA of *Calonectris* had a North Pacific and North Atlantic distribution. Fossils of at least 5 species have been described from the North Atlantic dating back to ~14 Mya (Olson, 2008; Olson, 2009; Olson, & Rasmussen, 2001), supporting this area as a speciation hotspot for the genus. However, considering the mobility of the genus and given that the MRCA of *Calonectris* was probably a long-distance migrant (Figure S6), we can not eliminate the possibility that the regions where these fossils were found were not the breeding areas for the species. The estimated divergence time (~3.54 Mya) between the North Pacific and the North Atlantic clades is very similar to previous estimates based on mtDNA rates (~3.44 Mya; Gómez-Díaz et al., 2006) and suggests a vicariant event as the result of the gradual closure of the isthmus of Panama, as has been observed in other marine organisms (Lessios, 2008).

Our analyses indicate that *Ardenna* had a South Australia - New Zealand origin and, thereafter, some lineages colonised the Southern Ocean (Figure 1), which disagrees with the North Atlantic origin of*Ardenna* proposed by (Austin, 1996) based on the fossil record. Extant species are long-distance trans-equatorial migrants that can be locally common on North American and European coasts (Carey, Phillips, Silk, & Shaffer, 2014; Shaffer et al., 2006) and based on our ancestral state reconstruction, the MRCA of *Ardenna* was also most likely a long-distance migrant (Figure S6). We suggest that extinct taxa were also long-distance migrants breeding in the Southern Hemisphere, and that the fossils found in the North Atlantic likely represent birds that died during the non-breeding period.

The ancestors of *Puffinus* acquired the strongest diving adaptations of the three genera (Olson, & Rasmussen, 2001); these allow them to routinely dive to depths of 55 m (Shoji et al., 2016), providing advantages for reaching prey in the nutrient poor tropical and subtropical waters of the Pacific (inaccessible to most other tropical seabirds; Burger, 2001), where the MRCA of *Puffinus* most probably originated based on the current ancestral range estimation analyses and the fossil record (Miller, 1961). Although we could not obtain samples for *P. subalaris*from the Galapagos; in a previous study this species formed a clade with*P. nativitatis* (Austin et al., 2004), which further supports a Tropical Pacific origin of*Puffinus* . Most extant *Puffinus* species are short-distance migrants or dispersers that remain close to their breeding sites throughout the year (e.g. Ramos et al., 2020). Their lower dispersal compared to other shearwater genera may have reduced gene flow and promoted higher species richness. The population sizes of*Puffinus* species tend to be small and many had the highest EDGE scores (Table 2), which is a metric that identifies those threatened species that deserve particular attention because of their unique evolutionary history. Predation by invasive alien species is the main current threat for seabirds (Croxall et al., 2012) and is a principal cause of population declines among *Puffinus*species (Rodríguez et al., 2019). Enhanced by predation, intra- and inter-specific competition for nest sites plays an important role in limiting populations of small Procellariiformes, such as *Puffinus*shearwaters (Monteiro, Ramos & Furness 1996; Ramos, Monteiro, Sola, & Moniz, 1997). At sea, fisheries bycatch is also a main threat for *Puffinus* shearwaters (Bugoni, Mancini, Monteiro, Nascimento, & Neves, 2008; Cortés, Arcos, & González-Solís, 2017) and one that could drive some species to extinction unless conservation measures are put in place (Genovart et al., 2016). These are likely some of the main reasons why *Puffinus*shearwaters have the highest number of endangered species among the shearwaters.

Across the three genera, the Pliocene-Pleistocene boundary appeared as a period of high and rapid speciation and dispersal (Figure 1). For instance, *Puffinus* spread from the Pacific to the North Atlantic, the Southern Ocean, and the Indian Ocean during a rapid radiation. During the Cenozoic, the largest global sea-level changes and oscillations occurred in the Pliocene and Pleistocene (Lisiecki & Raymo, 2007; Miller et al., 2005; Zachos, Pagani, Sloan, Thomas, & Billups, 2001). Neritic waters, which represent the main foraging grounds for medium and large shearwaters, especially during the breeding period, suffered a significant sudden reduction at the end of the Pliocene followed by extreme fluctuation and gradual reduction over the Pleistocene (Pimiento et al., 2017). Global oceanographic changes, such as the end of permanent el Niño, the closure of the Isthmus of Panama and the formation of the Arctic ice cap (Fedorov et al., 2006; O'Dea et al., 2016) may have been the cause of such reduction. This reduction has been hypothesised to be the cause of a three-fold increase in the extinction rate of megafauna associated with coastal habitats (O'Dea et al., 2007; Pimiento et al., 2017). In shearwaters, ˜36% of the known extinct fossil species are from the Pliocene (Howard, 1971; Olson, 1985; Olson, & Rasmussen, 2001); together with the long stems in the three shearwater genera (Figure 1), this suggests that the late Pliocene extinction severely affected the group. The subsequent burst of speciation and dispersal was probably promoted by Pleistocene climatic shifts that probably promoted geographic splitting and bottlenecks (Avise & Walker, 1998; Gómez-Díaz et al., 2006). An increase in diversification during this period has also been detected in other seabird groups such as penguins (Cole et al., 2019; Vianna et al., 2020) and even in deep sea species (Eilertsen & Malaquias, 2015).

Body mass as a key phenotypic trait

In the Procellariiformes, body mass is a trait closely related to fitness at the intraspecific level. For instance, body condition (body mass corrected by overall body size) of the progenitors affects breeding success in several species (Barbraud & Chastel, 1999; Chastel, Weimerskirch, & Jouventin, 1995; Tveraa, Sether, Aanes, & Erikstad, 1998). On the other hand, at the interspecific level, the drivers of body mass variation are poorly understood despite the high variation exhibited by the Procellariiformes (Nunn & Stanley, 1998). Our results shed some new light on potential behavioural and distributional drivers that may be affecting body mass variation in the Procellariiformes, although caution must be taken at interpreting our findings that are merely correlational.

Migratory strategy was the best evaluated predictor for mean body mass (Figure 2). This correlation is likely twofold and additive. On the one hand, migratory species tend to be larger (i.e. longer wings in migratory species; Marchetti, Price, & Richman, 1995; Minias et al., 2015) as shown by the significant correlations of all the other body size measures with migratory strategy (Figure 2a). On the other hand, migratory behaviour allows the exploitation of additional resources leading to a higher accumulation of fat deposits. The weaker but significant correlation between migratory strategy and mean body mass when corrected by body surface (Table S3) supports the twofold and additive effect of this correlation.

Within an endothermic species or a group of closely related endothermic species, individuals inhabiting colder habitats and higher latitudes tend to be larger than those inhabiting warmer environments and lower latitudes (Bergmann, 1848). This geographical pattern in body size holds for birds throughout the world at the intraspecific (Ashton, 2002; Meiri & Dayan, 2003) and interspecific level (Bergmann, 1848) although the mechanisms responsible for the generation of this trend are subject to much debate (Ashton, 2002; Meiri, 2011). In the shearwaters, this pattern has also been shown to apply to intraspecific body size variation in the Streaked Shearwater (*Calonectris leucomelas* ; Yamamoto et al., 2016). Among shearwater species, we also found a positive significant correlation between breeding latitude and mean body mass (Figure 2 and Figure S5C), despite previous studies that have shown that conformity to Bergmann's Rule tends to be weaker for migratory and enclosed nesting species (Mainwaring & Street, 2019; Meiri & Dayan, 2003).

The correlation was strongest between maximum breeding latitude and mean body mass corrected by body surface ($R^2 = 0.387$; Table S3), suggesting that heavier bodies, independent of body size, might provide a better adaptation to thrive in higher and colder latitudes. However, these correlations could also be indirectly driven by a higher tendency of species living in higher latitudes to be migratory and/or by differences in diving behaviour, which could not be explored in this study.

The strong association between body mass range and latitudinal range is likely twofold. On the one hand, exploiting larger foraging areas may allow for ecological segregation between sexes and size dimorphism (De Felipe et al., 2019). Indeed, ecological segregation has been shown to be the most likely cause of size dimorphism in other Procellariiformes (González-Solís, 2004). On the other hand, larger body mass differences may arise between individuals that are more efficient at exploiting the available resources compared to those that are less efficient. This might provide the substrate for sexual selection to act on body mass. Higher body condition has been associated with higher breeding success in several species of Procellariiformes (Barbraud & Chastel, 1999; Barbraud & Weimerskirch, 2005; Chastel et al., 1995).

Considerations of shearwater taxonomy

Species delimitation in shearwaters is a challenging and controversial topic, partly due to their remarkably similar morphology (Austin et al., 2004). Conflict has arisen amongst morphological studies, and analyses based on genetic data (i.e., mtDNA and microsatellites), and also between different genetic datasets (Austin, 1996; Genovart et al., 2013; Gómez-Díaz, González-Solís, & Peinado, 2009). In addition, despite being a promising trait for species delimitation, analyses of shearwater vocalizations are limited (Bretagnolle, 1996). Genome-wide datasets have the potential to provide fine-scale population structure and genomic divergence estimates that can inform taxonomy. Despite the high resolution of our PE-ddRAD dataset, fineRADStructure analysis showed no structure between two species pairs, *P. mauretanicus* and *P. yelkouan* , and *A. creatopus* and*A. carneipes* (Figure 3). Furthermore, although we do not consider there to be a genetic cutoff for species-level divergence, the genetic divergence between these recently diverged species were the lowest amongst any pair of species and overlapped with the genetic divergences observed between individuals of the same subspecies (Figure 4). *P. mauretanicus* and *P. yelkouan* were granted species status based on morphological, osteological and reciprocal monophyly using cytochrome b sequences (Heidrich, Amengual, & Wink, 1998; Sangster, Knox, Helbig, & Parkin, 2002). However, more recently, a lack of correspondence at the individual level was found between phenotypic characters, stable isotopes analyses, nuclear and mtDNA, and was attributed to admixture between the two species (Genovart, Juste, Contreras-Diaz, & Oro, 2012; Militao, Gomez-Diaz, Kaliontzopoulou, & Gonzalez-Solis, 2014). *A. creatopus* and*A. carneipes* are widely considered as two different species in taxonomic checklists (Carboneras, & Bonan, 2019; Gill et al., 2020), but some authors have argued that they should be considered conspecific based on the lack of uniform differentiation in colour and size (Bourne, 1962) and on low mtDNA differentiation (Penhallurick & Wink, 2004). These species pairs differ in plumage colouration and body size, which are known to be labile traits even within species of shearwaters. Dark and pale phases can be found within a single species (i.e.,*A. pacifica*) and some species exhibit a continuum from pale to dark (i.e., *P. mauretanicus*). Body size covaries with migratory behaviour (see previous section), can be under selection (Barbraud, 2000; Navarro, Kaliontzopoulou, & Gonzalez-Solis, 2009), and thus could evolve rapidly under strong selection pressures. In addition to the aforementioned species pairs, other shearwater species showed weak patterns of population structure and genetic distances within the interval among different subspecies: *P. boydi* and *P. baroli* , and the three Atlantic *Calonectris* species. These species complexes are the subject of ongoing taxonomic debate (Genovart et al., 2013; Gomez-Diaz et al., 2009; Olson, 2010; Sangster, Collinson, Helbig, Knox, & Parkin, 2005; Ramos et al., 2020). As a final consideration, our genomic data, together with ongoing taxonomic debate, suggest that these taxa should not be granted species status. Future studies should use species delimitation approaches combining genomic data with a thorough morphological revaluation including a detailed evaluation of vocalisations. Further research is also required to include the taxa that could not be sampled during this study, particularly taxa from the tropical Pacific that breed in remote islands and have very limited distributions and low population sizes.

ACKNOWLEDGEMENTS

We thank Gary Nunn, Jeremy J. Austin, Chris Gaskin, Juan E. Martinez-Gomez, and Maite Louzao for collecting tissue and blood samples from the field and/or providing tissue samples from museum skins for this study. We would also like to thank the many institutions that provided tissue loans for this research: Smithsonian National Museum of Natural History, Louisiana State University Museum of Natural Science, University of Washington Burke Museum, American Museum of Natural History, the University of Kansas Biodiversity Institute, and the Museum National d'Histoire Naturelle. We thank the pertinent authorities for issuing the permits needed for this work. Finally, we would like to thank Josephine R. Paris for her comments on an early version of this manuscript and Martí Franch for his wonderful shearwater illustrations. Research funding was provided by the Fundación BBVA (program "Ayudas a Equipos de Investigación Científica 2017", project code 062_17 to M.R.).

REFERENCES

Angelis, K., & Dos Reis, M. (2015). The impact of ancestral population size and incomplete lineage sorting on Bayesian estimation of species divergence times. Current Zoology, 61(5), 874–885.

Ashton, K. G. (2002). Patterns of within-species body size variation of birds: strong evidence for Bergmann's rule. Global Ecology and Biogeography, 11(6), 505–523.

Austin, J. J. (1996). Molecular phylogenetics of Puffinus shearwaters: preliminary evidence from mitochondrial cytochrome b gene sequences. Molecular Phylogenetics and Evolution, 6(1), 77–88.

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), 647–864.

Avise, J. C., & Walker, D. E. (1998). Pleistocene phylogeographic effects on avian populations and the speciation process. Proceedings of the Royal Society of London. Series B: Biological Sciences, 265, 457–463. doi: 10.1098/rspb.1998.0317

Barbraud, C. (2000). Natural selection on body size traits in a long-lived bird, the snow petrel Pagodroma nivea. Journal of Evolutionary Biology, 13(1), 81-88.

Barbraud, C., & Chastel, O. (1999). Early body condition and hatching success in the snow petrel Pagodroma nivea. Polar Biology, 21(1), 1–4.

Barbraud, C., & Weimerskirch, H. (2005). Environmental conditions and breeding experience affect costs of reproduction in blue petrels. Ecology, 86(3), 682–692.

Benton, M. J. (2009). The Red Queen and the Court Jester: species diversity and the role of biotic and abiotic factors through time. Science, 323(5915), 728–732.

Bergmann, C. (1848). Über die Verhältnisse der Wärmeökonomie der Thiere zu ihrer Größe. Gottinger Studien, 3: 384 595–708.

Bierne, N., Bonhomme, F., & David, P. (2003). Habitat preference and the marine-speciation paradox. Proceedings of the Royal Society of London. Series B: Biological Sciences, 270(1522), 1399–1406.

Billerman, S. M., Keeney, B. K., Rodewald, P. G., & Schulenberg, T. S. (2020). Birds of the World. Cornell Laboratory of Ornithology, Ithaca, NY, USA. https://birdsoftheworld.org/bow/home

BirdLife International (2020) IUCN Red List for birds. Downloaded from http://www.birdlife.org on 17/08/2020.

Bouckaert, R. R. (2010). DensiTree: Making sense of sets of phylogenetic trees. Bioinformatics, 26(10), 1372–1373.

Bouckaert, R., Vaughan, T. G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., … Drummond, A. J. (2019). BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. PLoS Computational Biology, 15(4), e1006650.

Bourne, W. R. P. (1962). Handbook of North American Birds. Yale University Press New Haven.

Bretagnolle, V. (1996). Acoustic communication in a group of nonpasserine birds, the petrels. Ecology and Evolution of Acoustic Communication in Birds, 160, 177.

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

Bugoni, L., Mancini, P. L., Monteiro, D. S., Nascimento, L., & Neves, T. S. (2008). Seabird bycatch in the Brazilian pelagic longline fishery and a review of capture rates in the southwestern Atlantic Ocean. Endangered Species Research, 5, 137–147.

Burger, A. E. (2001). Diving depths of shearwaters. The Auk, 118(3), 755-759.

Butlin, R., Debelle, A., Kerth, C., Snook, R. R., Beukeboom, L. W., Castillo, R. F. C., … & van de Zande, L. (2012). What do we need to know about speciation? Trends in Ecology & Evolution, 27(1), 27–39.

Carboneras, C., Bonan, A. (2019). Petrels, Shearwaters (Procellariidae). In: del Hoyo, J., Elliott, A., Sargatal, J., Christie, D.A. & de Juana, E. (eds.). Handbook of the Birds of the World Alive. Lynx Edicions, Barcelona.

Carey, M. J., Phillips, R. A., Silk, J. R. D., & Shaffer, S. A. (2014). Trans-equatorial migration of Short-tailed Shearwaters revealed by geolocators. Emu - Austral Ornithology, 114(4), 352–359.

Carstens, B. C., Pelletier, T. A., Reid, N. M., & Satler, J. D. (2013). How to fail at species delimitation. Molecular Ecology, 22(17), 4369–4383.

Chastel, O., Weimerskirch, H., & Jouventin, P. (1995). Body Condition and Seabird Reproductive Performance: A Study of Three Petrel Species. Ecology, 76(7), 2240–2246.

Cole, T. L., Ksepka, D. T., Mitchell, K. J., Tennyson, A. J. D., Thomas, D. B., Pan, H., … Waters, J. M. (2019). Mitogenomes Uncover Extinct Penguin Taxa and Reveal Island Formation as a Key Driver of Speciation. Molecular Biology and Evolution, 36(4), 784–797.

Cortés, V., Arcos, J. M., & González-Solís, J. (2017). Seabirds and demersal longliners in the northwestern Mediterranean: factors driving their interactions and bycatch rates. Marine Ecology Progress Series, 565, 1–16.

Coulson, J. (2002). Colonial breeding in seabirds. Biology of marine birds, 87-113.

Coyne, J. A., & Orr, H. A. (2004). Speciation (Vol. 37). Sinauer Associates Sunderland, MA.

Croxall, J. P., Butchart, S. H. M., Lascelles, B., Stattersfield, A. J., Sullivan, B., Symes, A., & Taylor, P. (2012). Seabird conservation status, threats and priority actions: a global assessment. Bird Conservation International, 22, 1–34. doi: 10.1017/S0959270912000020

Cuevas-Caballé, C., Ferrer-Obiol, J., Genovart, M., Rozas, J., González-Solís, J., Riutort, M. (2019). Conservation genomics applied to the Balearic shearwater. G10K-VGP/EBP 2019. doi: 10.13140/RG.2.2.15751.21923

Cutter, A. D. (2013). Integrating phylogenetics, phylogeography and population genetics through genomes and evolutionary theory. Molecular Phylogenetics and Evolution, 69(3), 1172–1185.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., . . . 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. Bioinformatics, 27(15), 2156–2158.

De Felipe, F., Reyes-González, J. M., Militão, T., Neves, V. C., Bried, J., Oro, D., . . . & González-Solís, J. (2019). Does sexual segregation occur during the non-breeding period? A comparative analysis in the spatial ecology of three Calonectris shearwaters. Ecology and Evolution, 9(18), 10145-10162.

Dickinson, E. C., & Remsen, J. R. (2013). The Howard and Moore complete checklist of the birds of the World. Vol. 1. Non-passerines. Aves Press, Eastbourne, UK.

Drummond, A. J., Ho, S. Y. W., Phillips, M. J., & Rambaut, A. (2006). Relaxed phylogenetics and dating with confidence. PLoS Biology, 4(5), 699–710.

Edwards, S. V., Xi, Z., Janke, A., Faircloth, B. C., McCormack, J. E., Glenn, T. C., . . . Davis, C. C. (2016). Implementing and testing the multispecies coalescent model: A valuable paradigm for phylogenomics. Molecular Phylogenetics and Evolution, 94, 447–462.

Eilertsen, M. H., & Malaquias, M. A. E. (2015). Speciation in the dark: diversification and biogeography of the deep-sea gastropod genus Scaphander in the Atlantic Ocean. Journal of Biogeography, 42(5), 843–855.

Estandia, A. (2019). Genome-wide phylogenetic reconstruction for Procellariiform seabirds is robust to molecular rate variation. Masters Thesis, Durham University.

Fedorov, A. V., Dekens, P. S., McCarthy, M., Ravelo, A. C., deMenocal, P. B., Barreiro, M., . . . Philander, S. G. (2006). The Pliocene paradox (mechanisms for a permanent El Niño). Science, 312(5779), 1485–1489.

Ferrer-Obiol, J., James, H.F., Chesser, R.T., Bretagnolle, V., González-Solís, J., Rozas, J., Riutort, M., & Welch, A.J. (under review). Integrating Sequence Capture and Restriction-Site Associated DNA Sequencing to resolve Recent Radiations of Pelagic Seabirds.

Fišer, C., Robinson, C. T., & Malard, F. (2018). Cryptic species as a window into the paradigm shift of the species concept. Molecular Ecology, 27(3), 613–635.

Friesen, V. L., Burg, T. M., & McCoy, K. D. (2007). Mechanisms of population differentiation in seabirds. Molecular Ecology, 16(9), 1765–1785.

Friesen, V. L., Smith, A. L., Gomez-Diaz, E., Bolton, M., Furness, R. W., González-Solís, J., & Monteiro, L. R. (2007). Sympatric speciation by allochrony in a seabird. Proceedings of the National Academy of Sciences, 104(47), 18589–18594.

Genovart, M., Arcos, J. M., Álvarez, D., McMinn, M., Meier, R., Wynn, R., . . . Oro, D. (2016). Demography of the critically endangered Balearic shearwater: the impact of fisheries and time to extinction. Journal of Applied Ecology. doi: 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.

Genovart, M., Thibault, J. C., Igual, J. M., Bauzà-Ribot, M. del M., Rabouam, C., & Bretagnolle, V. (2013). Population Structure and Dispersal Patterns within and between Atlantic and Mediterranean Populations of a Large-Range Pelagic Seabird. PloS One, 8(8). doi: 10.1371/journal.pone.0070711

Gill, F., Donsker, D., & Rasmussen, P. (2020). IOC World Bird List (v10. 1). doi : 10.14344/IOC.ML.10.1.

Gómez-Díaz, E., González-Solís, J., & Peinado, M. A. (2009). Population structure in a highly pelagic seabird, the Cory's shearwater Calonectris diomedea: An examination of genetics, morphology and ecology. Marine Ecology Progress Series, 382, 197–209.

Gómez-Díaz, E., González-Solís, J., Peinado, M. A., & Page, R. D. M. (2006). Phylogeography of the Calonectris shearwaters using molecular and morphometric data. Molecular Phylogenetics and Evolution, 41(2), 322–332.

González-Solis, J. (2004). Sexual size dimorphism in northern giant petrels: ecological correlates and scaling. Oikos, 105(2), 247-254.

Gonzalez-Solis, J., Croxall, J. P., Oro, D., & Ruiz, X. (2007). Trans-equatorial migration and mixing in the wintering areas of a pelagic seabird. Frontiers in Ecology and the Environment, 5(6), 297-301.

Gonzalez-Solis, J., Felicisimo, A., Fox, J. W., Afanasyev, V., Kolbeinsson, Y., & Munoz, J. (2009). Influence of sea surface winds on shearwater migration detours. Marine Ecology Progress Series, 391, 221–230.

Heidrich, P., Amengual, J., & Wink, M. (1998). Phylogenetic relationships in mediterranean and North Atlantic shearwaters (Aves: Procellariidae) based on nucleotide sequences of mtDNA. Biochemical Systematics and Ecology, 26(2), 145–170.

Heled, J., & Bouckaert, R. R. (2013). Looking for trees in the forest: summary tree from posterior samples. BMC Evolutionary Biology, 13, 221.

Howard, H. (1971). Pliocene avian remains from Baja California. Los Angeles County Museum of Natural History.

Isaac, N. J. B., Turvey, S. T., Collen, B., Waterman, C., & Baillie, J. E. M. (2007). Mammals on the EDGE: conservation priorities based on threat and phylogeny. PloS One, 2(3), e296.

Jetz, W., Thomas, G. H., Joy, J. B., Hartmann, K., & Mooers, A. O. (2012). The global diversity of birds in space and time. Nature, 491(7424), 444–448.

Kawakami, K., Eda, M., Izumi, H., Horikoshi, K., & Suzuki, H. (2018). Phylogenetic position of endangered Puffinus lherminieri bannermani. Ornithological Science, 17(1), 11–18.

Knaus, B. J., & Grunwald, N. J. (2017). vcfr: a package to manipulate and visualize variant call format data in R. Molecular Ecology Resources, 17(1), 44–53.

Kopp, M. (2010). Speciation and the neutral theory of biodiversity. BioEssays, 32, 564–570. doi: 10.1002/bies.201000023

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.

Kuroda, N. (1954). On the Classification and Phylogeny of the Order Tubinares, Particularly the Shearwaters (Puffinus), with Special Considarations [ie Considerations] on Their Osteology and Habit Differentiation. Kuroda.

Landis, M. J., Matzke, N. J., Moore, B. R., & Huelsenbeck, J. P. (2013). Bayesian analysis of biogeography when the number of areas is large. Systematic Biology, 62(6), 789–804.

Lartillot, N., & Poujol, R. (2011). A phylogenetic model for investigating correlated evolution of substitution rates and continuous phenotypic characters. Molecular Biology and Evolution, 28(1), 729–744.

Lawson, D. J., Hellenthal, G., Myers, S., & Falush, D. (2012). Inference of population structure using dense haplotype data. PLoS Genetics, 8(1), e1002453.

Leache, A. D., Zhu, T., Rannala, B., & Yang, Z. (2018). The Spectre of Too Many Species. Systematic Biology, 68(1), 168–181.

Lessios, H. A. (2008). The great American schism: divergence of marine organisms after the rise of the Central American Isthmus. Annual Review of Ecology, Evolution, and Systematics, 39, 63-91. doi: 10.1146/annurev.ecolsys.38.091206.095815

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics , 27(21), 2987–2993.

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

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., . . . Durbin, R. (2009). 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and samtools. Bioinformatics , 25(16), 2078–2079.

Lisiecki, L. E., & Raymo, M. E. (2007). Plio–Pleistocene climate evolution: trends and transitions in glacial cycle dynamics. Quaternary Science Reviews, 26(1), 56–69.

Lombal, A. J., Wenner, T. J., Lavers, J. L., & Austin, J. J. (2018). Genetic divergence between colonies of Flesh-footed Shearwater Ardenna carneipes exhibiting different foraging strategies. Conservation Genetics, 19(1), 27-41.

Maddison, W. P. (1997). Gene Trees in Species Trees. Systematic Biology, 46(3), 523–536.

Mainwaring, M. C., & Street, S. E. (2019). Conformity to Bergmann's rule in birds depends on nest design and migration. bioRxiv, 686972. doi: 10.1101/686972

Malinsky, M., Trucchi, E., Lawson, D. J., & Falush, D. (2018). RADpainter and fineRADstructure: Population Inference from RADseq Data. Molecular Biology and Evolution, 35(5), 1284–1290.

Marchetti, K., Price, T., & Richman, A. (1995). Correlates of Wing Morphology with Foraging Behaviour and Migration Distance in the Genus Phylloscopus. Journal of Avian Biology, 26(3), 177–181.

Martinez-Gomez, J. E., Matias-Ferrer, N., Sehgal, R. N. M., & Escalante, P. (2015). Phylogenetic placement of the critically endangered Townsend's Shearwater (Puffinus auricularis auricularis): evidence for its conspecific status with Newell's Shearwater (Puffinus a. newelli) and a mismatch between genetic and phenotypic differentiation. Journal of Ornithology, 156(4), 1025–1034.

Maruki, T., & Lynch, M. (2015). Genotype-Frequency Estimation from High-Throughput Sequencing Data. Genetics, 201(2), 473–486.

Maruki, T., & Lynch, M. (2017). Genotype Calling from Population-Genomic Sequencing Data. G3: Genes|Genomes|Genetics, 7(5), 1393–1404.

Matzke, N. J. (2013). BioGeoBEARS: BioGeography with Bayesian (and likelihood) evolutionary analysis in R Scripts. R Package, Version 0. 2, 1, 2013.

McCormack, J. E., Heled, J., Delaney, K. S., Peterson, A. T., & Knowles, L. L. (2011). Calibrating divergence times on species trees versus gene trees: implications for speciation history of Aphelocoma jays. Evolution; International Journal of Organic Evolution, 65(1), 184–202.

Meiri, S. (2011). Bergmann's Rule–what's in a name?. Global Ecology and Biogeography, 20(1), 203-207.

Meiri, S., & Dayan, T. (2003). On the validity of Bergmann's rule. Journal of Biogeography, 30(3), 331–351.

Mendes, F. K., & Hahn, M. W. (2016). Gene Tree Discordance Causes Apparent Substitution Rate Variation. Systematic Biology, 65(4), 711–721. doi: 10.1093/sysbio/syw018

Militao, T., Gomez-Diaz, E., Kaliontzopoulou, A., & Gonzalez-Solis, J. (2014). Comparing multiple criteria for species identification in two recently diverged seabirds. PloS One, 9(12), e115650.

Miller, K. G., Kominz, M. A., Browning, J. V., Wright, J. D., Mountain, G. S., Katz, M. E., . . . Pekar, S. F. (2005). The Phanerozoic record of global sea-level change. Science, 310(5752), 1293–1298.

Miller, L. (1961). Birds from the Miocene of Sharktooth Hill, California. The Condor, 63(5), 399–402.

Milot, E., Weimerskirch, H., & Bernatchez, L. (2008). The seabird paradox: dispersal, genetic structure and population dynamics in a highly mobile, but philopatric albatross species. Molecular Ecology, 17(7), 1658-1673.

Minias, P., Meissner, W., Włodarczyk, R., Ożarowska, A., Piasecka, A., Kaczmarek, K., & Janiszewski, T. (2015). Wing shape and migration in shorebirds: a comparative study. Ibis, 157(3), 528–535.

Monteiro, L. R., Ramos, J. A., & Furness, R. W. (1996). Past and present status and conservation of the seabirds breeding in the Azores Archipelago. (1996). Biological Conservation, 78(3), 319–328.

Moura, A. E., Nielsen, S. C. A., Vilstrup, J. T., Moreno-Mayar, J. V., Gilbert, M. T. P., Gray, H. W. I., . . . Hoelzel, A. R. (2013). Recent diversification of a marine genus (Tursiops spp.) tracks habitat preference and environmental change. Systematic Biology, 62(6), 865–877.

Mundry, R. (2014). Statistical Issues and Assumptions of Phylogenetic Generalized Least Squares. Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology, 131–153. doi: 10.1007/978-3-662-43550-2_6

Munilla, I., Genovart, M., Paiva, V. H., & Velando, A. (2016). Colony foundation in an oceanic seabird. PloS one, 11(2), e0147222.

Navarro, J., Forero, M. G., González-Solís, J., Igual, J. M., Bécares, J., & Hobson, K. A. (2009). Foraging segregation between two closely related shearwaters breeding in sympatry. Biology Letters, 5, 545–548. doi: 10.1098/rsbl.2009.0150

Navarro, J., Kaliontzopoulou, A., & González-Solís, J. (2009). Sexual dimorphism in bill morphology and feeding ecology in Cory's shearwater (Calonectris diomedea). Zoology , 112(2), 128–138.

Nosil, P. (2012). Ecological Speciation. Oxford University Press.

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.

O'Dea, A., Jackson, J. B. C., Fortunato, H., Smith, J. T., D'Croz, L., Johnson, K. G., & Todd, J. A. (2007). Environmental change preceded Caribbean extinction by 2 million years. Proceedings of the National Academy of Sciences of the United States of America, 104(13), 5501–5506.

O'Dea, A., Lessios, H. A., Coates, A. G., Eytan, R. I., Restrepo-Moreno, S. A., Cione, A. L., . . . Jackson, J. B. C. (2016). Formation of the Isthmus of Panama. Science Advances, 2(8), e1600883.

Ogilvie, H. A., Bouckaert, R. R., & Drummond, A. J. (2017). StarBEAST2 Brings Faster Species Tree Inference and Accurate Estimates of Substitution Rates. Molecular Biology and Evolution, 34(8), 2101–2114.

Ogilvie, H. A., Heled, J., Xie, D., & Drummond, A. J. (2016). Computational Performance and Statistical Accuracy of *BEAST and Comparisons with Other Methods. Systematic Biology, 65(3), 381–396.

Olson, S. L. (1985). The fossil record of birds. Avian biology.

Olson, S. L. (2008). A new species of shearwater of the genus Calonectris (Aves: Procellariidae) from a middle Pleistocene deposit on Bermuda. Proceedings of the Biological Society of Washington, 121(3), 398-409.

Olson, S. L. (2009). A new diminutive species of shearwater of the genus Calonectris (Aves: Procellariidae) from the Middle Miocene Calvert Formation of Chesapeake Bay. Proceedings of the Biological Society of Washington, 122(4), 466–470.

Olson, S. L. (2010). Stasis and turnover in small shearwaters on Bermuda over the last 400 000 years (Aves: Procellariidae: Puffinus lherminieri group). Biological Journal of the Linnean Society, 99(4), 699–707.

Olson, S. L., & Rasmussen, P. C. (2001). Miocene and Pliocene birds from the Lee Creek Mine, North Carolina. Smithsonian Contributions to Paleobiology, 90, 233–365.

Onley, D., & Scofield, P. (2013). Albatrosses, Petrels and Shearwaters of the World. Bloomsbury Publishing.

Orme, D., Freckleton, R., Thomas, G., Petzoldt, T., Fritz, S., & Others. (2013). The caper package: comparative analysis of phylogenetics and evolution in R. R Package Version, 5.2, 1–36.

Pagel, M. (1999). The Maximum Likelihood Approach to Reconstructing Ancestral Character States of Discrete Characters on Phylogenies. Systematic Biology, 48(3), 612–622.

Palumbi, S. R. (1994). Genetic divergence, reproductive isolation, and marine speciation. Annual Review of Ecology and Systematics, 25(1), 547–572.

Paradis, E., & Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics , 35(3), 526–528.

Paris, J. R., Stevens, J. R., & Catchen, J. M. (2017). Lost in parameter space: a road map for stacks. Methods in Ecology and Evolution, 8, 1360–1373. doi: 10.1111/2041-210x.12775

Penhallurick, J., & Wink, M. (2004). Analysis of the taxonomy and nomenclature of the Procellariiformes based on complete nucleotide sequences of the mitochondrial cytochrome b gene. Emu, 104, 125–147. doi: 10.1071/MU01060

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

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, 1(8), 1100–1106.

Price, T. (2008). Speciation in birds. Roberts and Co..

Pyle, P., Welch, A. J., & Fleischer, R. C. (2011). A New Species of Shearwater (Puffinus) Recorded from Midway Atoll, Northwestern Hawaiian Islands. The Condor, 113(3), 518–527.

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. Systematic Biology, 67(5), 901–904.

Ramos, J. A., Monteiro, L. R., Sola, E., & Moniz, Z. (1997). Characteristics and Competition for Nest Cavities in Burrowing Procellariiformes. The Condor, 99(3), 634–641.

Ramos, R., Paiva, V. H., Zajková, Z., Precheur, C., Fagundes, A. I., Jodice, P. G., … & González-Solís, J. (2020). Spatial ecology of closely related taxa: the case of the little shearwater complex in the North Atlantic Ocean. Zoological Journal of the Linnean Society, zlaa045. doi: 10.1093/zoolinnean/zlaa045

Rannala, B., & Yang, Z. (2007). Inferring speciation times under an episodic molecular clock. Systematic Biology, 56(3), 453–466.

Rayner, M. J., Hauber, M. E., Steeves, T. E., Lawrence, H. a., Thompson, D. R., Sagar, P. M., … Shaffer, S. a. (2011). Contemporary and historical separation of transequatorial migration between genetically distinct seabird populations. Nature Communications, 2, 332.

Ree, R. H., & Sanmartín, I. (2018). Conceptual and statistical problems with the DEC+ J model of founder-event speciation and its comparison with DEC via model selection. Journal of Biogeography, 45(4), 741–749.

Ree, R. H., & Smith, S. A. (2008). Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. Systematic Biology, 57(1), 4–14.

Revell, L. J. (2010). Phylogenetic signal and linear regression on species data: Phylogenetic regression. Methods in Ecology and Evolution, 1(4), 319–329.

Revell, L. J. (2012). phytools: an R package for phylogenetic comparative biology (and other things). Methods in Ecology and Evolution, 3(2), 217–223.

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

Rodríguez, A., Arcos, J. M., Bretagnolle, V., Dias, M. P., Holmes, N. D., Louzao, M., . . . & Ronconi, R. A. (2019). Future directions in conservation research on petrels and shearwaters. Frontiers in Marine Science, 6, 94.

Ronquist, F. (1997). Dispersal-Vicariance Analysis: A New Approach to the Quantification of Historical Biogeography. Systematic Biology, 46(1), 195–203.

Sangster, G., Collinson, J. M., Helbig, A. J., Knox, A. G., & Parkin, D. T. (2005). Taxonomic recommendations for British birds: third report. Ibis, 147(4), 821–826.

Sangster, G., Knox, A. G., Helbig, A. J., & Parkin, D. T. (2002). Taxonomic recommendations for European birds. Ibis, 144(1), 153–159.

Schluter, D., & Pennell, M. W. (2017). Speciation gradients and the distribution of biodiversity. Nature, 546(7656), 48–55.

Seddon, N., Botero, C. A., Tobias, J. A., Dunn, P. O., Macgregor, H. E. A., Rubenstein, D. R., . . . Safran, R. J. (2013). Sexual selection accelerates signal evolution during speciation in birds. Proceedings of the Royal Society B: Biological Sciences, 280(1766), 20131065.

Shaffer, S. A., Tremblay, Y., Weimerskirch, H., Scott, D., Thompson, D. R., Sagar, P. M., . . . Costa, D. P. (2006). Migratory shearwaters integrate oceanic resources across the Pacific Ocean in an endless summer. Proceedings of the National Academy of Sciences of the United States of America, 103(34), 12799–12802.

Shoji, A., Dean, B., Kirk, H., Freeman, R., Perrins, C. M., & Guilford, T. (2016). The diving behaviour of the Manx Shearwater Puffinus puffinus. Ibis, 158(3), 598–606.

Silva, M. C., Matias, R., Wanless, R. M., Ryan, P. G., Stephenson, B. M., Bolton, M., . . . Coelho, M. M. (2015). Understanding the mechanisms of antitropical divergence in the seabird White-faced Storm-petrel (Procellariiformes: Pelagodroma marina) using a multilocus approach. Molecular Ecology, 24(12), 3122–3137.

Simpson, G. G. (1953). The major features of evolution. (No. 575 S55).

Stange, M., Sánchez-Villagra, M. R., Salzburger, W., & Matschiner, M. (2018). Bayesian Divergence-Time Estimation with Genome-Wide Single-Nucleotide Polymorphism Data of Sea Catfishes (Ariidae) Supports Miocene Closure of the Panamanian Isthmus. Systematic Biology, 67(4), 681–699.

Storey, A. S., & Lien, J. (1985). Development of the first North American colony of Manx Shearwaters. The Auk, 395-401.

Suh, A., Smeds, L., & Ellegren, H. (2015). The dynamics of incomplete lineage sorting across the ancient adaptive radiation of neoavian birds. PLoS Biology, 13(8), 1–18.

Sukumaran, J., & Knowles, L. L. (2017). Multispecies coalescent delimits structure, not species. Proceedings of the National Academy of Sciences of the United States of America, 114(7), 1607–1612.

Sutton, P. (2001). Detailed structure of the Subtropical Front over Chatham Rise, east of New Zealand. Journal of Geophysical Research, 106(C12), 31045–31056.

Tveraa, T., Sether, B., Aanes, R., & Erikstad, K. E. (1998). Regulation of food provisioning in the Antarctic petrel; the importance of parental body condition and chick body mass. Journal of Animal Ecology, 67, 699–704. doi: 10.1046/j.1365-2656.1998.00234.x

Vargas, P., & Zardoya, R. (2014). The tree of life. Sunderland, MA.

Vianna, J. A., Fernandes, F. A. N., Frugone, M. J., Figueiró, H. V., Pertierra, L. R., Noll, D., … Bowie, R. C. K. (2020). Genome-wide analyses reveal drivers of penguin diversification. Proceedings of the National Academy of Sciences of the United States of America. 117(36), 22303-22310.

Weber, C. C., Boussau, B., Romiguier, J., Jarvis, E. D., & Ellegren, H. (2014). Evidence for GC-biased gene conversion as a driver of between-lineage differences in avian base composition. Genome Biology, 15(12), 549.

Weimerskirch, H., Delord, K., Guitteaud, A., Phillips, R. A., & Pinet, P. (2015). Extreme variation in migration strategies between and within wandering albatross populations during their sabbatical year, and their fitness consequences. Scientific Reports, 5, 1–7.

Weimerskirch, H., Louzao, M., de Grissac, S., & Delord, K. (2012). Changes in wind pattern alter albatross distribution and life-history traits. Science, 335(6065), 211–214.

Yamamoto, T., Kohno, H., Mizutani, A., Yoda, K., Matsumoto, S., Kawabe, R., … Takahashi, A. (2016). Geographical variation in body size of a pelagic seabird, the streaked shearwater Calonectris leucomelas. Journal of Biogeography, 43(4), 801–808.

Yu, G., Smith, D. K., Zhu, H., Guan, Y., & Lam, T. T. Y. (2017). ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods in Ecology and Evolution, 8(1), 28–36.

Zachos, J., Pagani, M., Sloan, L., Thomas, E., & Billups, K. (2001). Trends, rhythms, and aberrations in global climate 65 Ma to present. Science, 292(5517), 686–693.

DATA ACCESSIBILITY

Raw PE-ddRAD reads are archived on the European Nucleotide Archive (ENA) under the accession number PRJEB38458. Files containing processed data at several stages including: VCF format and XML files for SNAPP analyses, PHYLIP format file for COEVOL analysis, fineRADStructure input files and MCC SNAPP trees, are deposited in the Dryad Digital Repository (*https://doi.org/1XXXXX*).

AUTHOR CONTRIBUTIONS

R.T.C., H.F.J., J.G., V.B., A.J.W. and J.F. contributed to data collection and all authors contributed to study design. J.F. analysed the data and wrote the manuscript with input from all authors. All authors read and approved the final manuscript.

ORCID

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

Helen F. James http://orcid.org/0000-0002-2495-6133

R. Terry Chesserhttp://orcid.org/0000-0003-4389-7092

Vincent Bretagnollehttp://orcid.org/0000-0002-2320-7755

Jacob González-Solís http://orcid.org/0000-0002-8691-9397

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

Andreanna J. Welchhttp://orcid.org/0000-0002-5947-4695

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

FIGURES AND TABLES

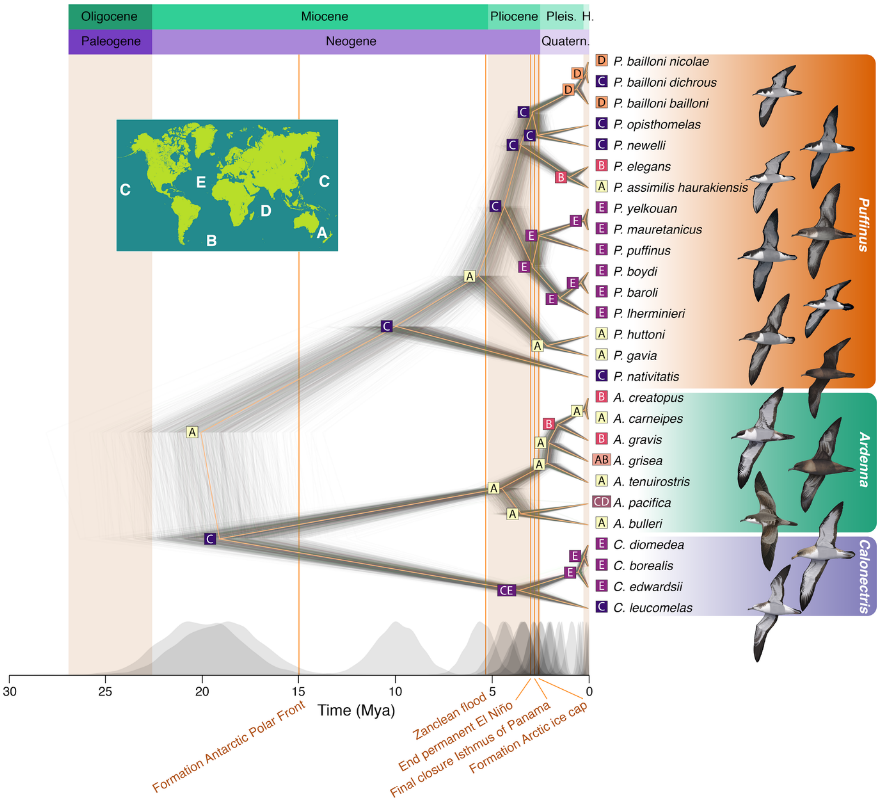

Figure 1 Time-calibrated species tree of the shearwaters using a constraint on the root age and a fixed topology. Geological periods and epochs are labeled above the tree. Posterior densities of divergence times are shown below the species tree. Note the speciation peak during the late Pliocene - early Pleistocene. Ancestral ranges were estimated under the DIVALIKE + j model using a dispersal matrix restricting dispersal between areas connected by main historical and present surface ocean currents in BioGeoBEARS and are shown as boxes at nodes and tips coded according to the map (Inset). Posterior estimates of divergence times are summarized in Table S2. Illustrations by Martí Franch are representative shearwater species depicted by their lineages.

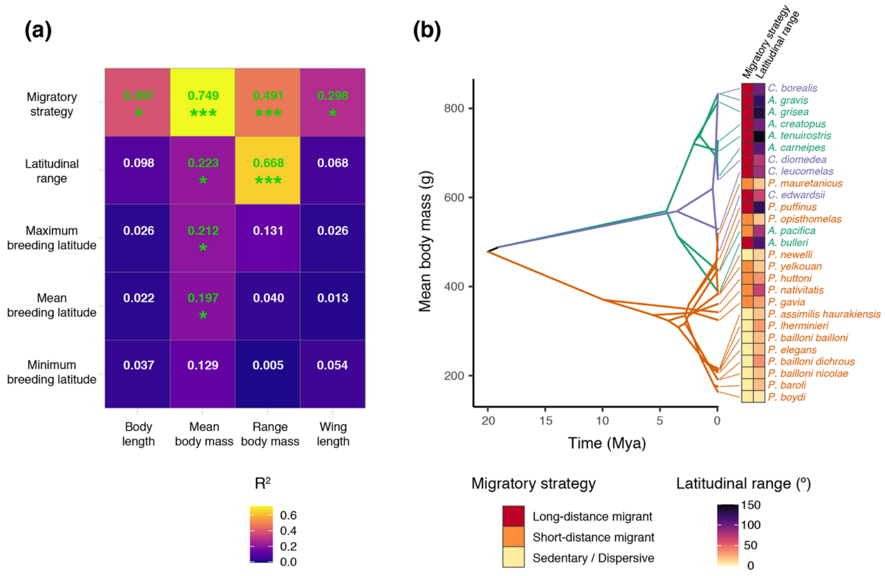

Figure 2 Migratory strategy and latitudinal range are the best predictors of body size. (a) Heatmap showing adjusted R^2 values for PGLS analyses of body size measures against the predictors. Positive correlations coefficients were recovered for each test. Numbers within each tile show the adjusted R^2 values and are coloured in green when significant after adjusting p-values by FDR correction for multiple testing (FDR: *** $< 0.001 >$ ** $< 0.01 >$ * < 0.05). (b) Phenogram of mean body mass constructed in Phytools (Revell, 2012) showing abrupt differences in mean body mass between sister clades. Edge colours indicate the three genera: *Calonectris* (purple),*Ardenna* (green) and *Puffinus* (orange). Heatmaps next to the phenogram show the migratory strategy and the latitudinal range for each species.

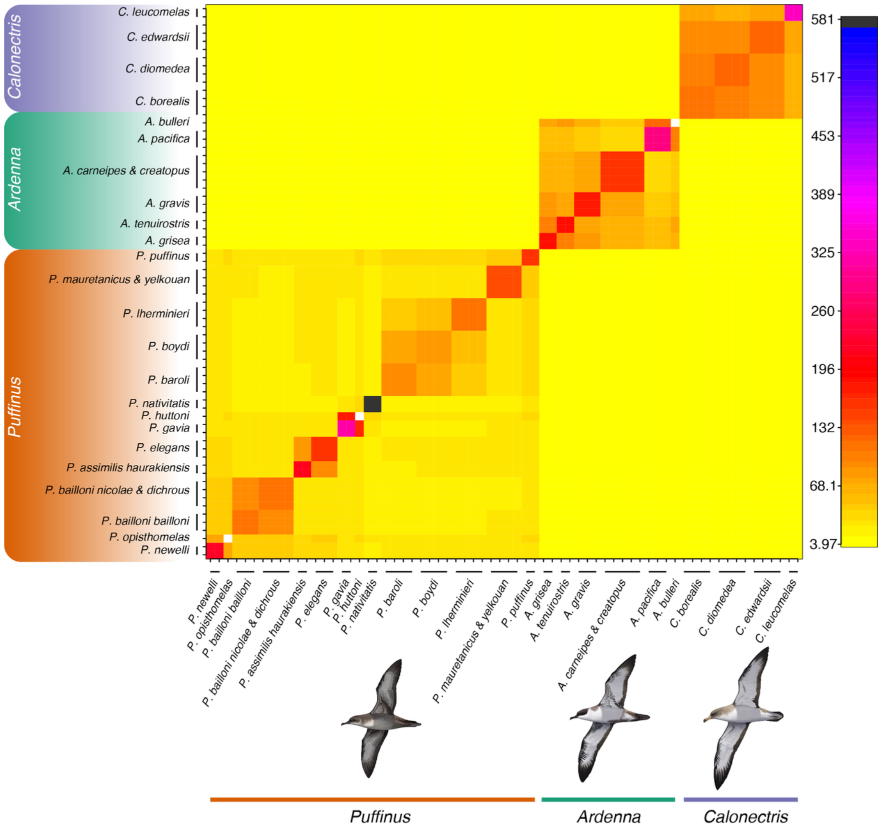

Figure 3 Clustered fineRADStructure coancestry matrix based on 8,049 PE-ddRAD loci. Pairwise coancestry coefficients are colour coded from low (yellow) to high (black). Every name represents a discernible discrete cluster based on the pairwise matrix of coancestry coefficients, defined by a posterior probability > 0.9 in the finestructure tree. Note that the three major clusters represent the three genera and that most species and subspecies included in the study are recovered as unique clusters.

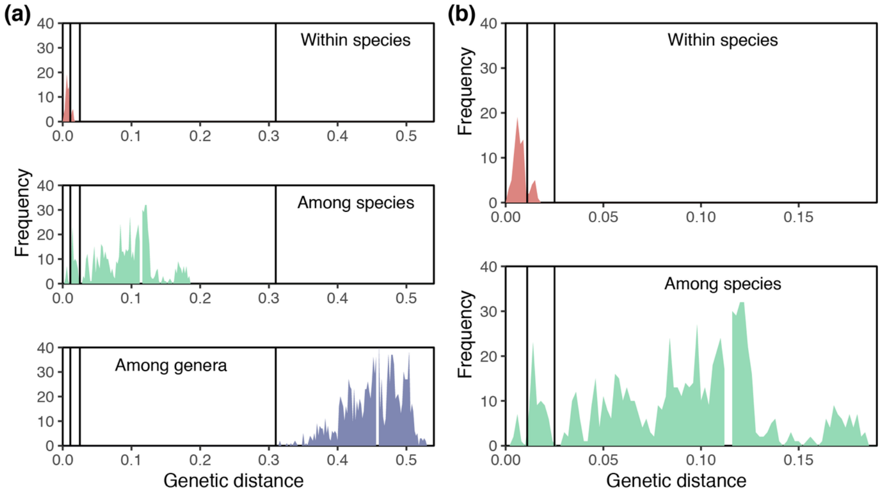

Figure 4 The distributions of genetic distances within and among species overlap. (a) Distribution of

genetic distances at different taxonomic levels in the shearwaters (upper panel: within species, middle panel: among species and lower panel: among genera) according to current taxonomy. Vertical bars show proposed orientative limits to assist visualisation across panels for values within subspecies, among subspecies, among species and among genera from left to right. (b) Zoom-in of the distributions of genetic distances within and among species. Comparisons between *P. mauretanicus* and*P. yelkouan* and between *A. carneipes* and *A. creatopus* fall in the within subspecies range, and comparisons between*P. baroli* and *P. boydi* and between the three Atlantic*Calonectris* species fall in the among subspecies range.

Table 1 Comparison of models of ancestral range estimation for the shearwaters. Models with and without the founder event parameter (j) are shown separately and for each case the model with the highest AICc weight is shown in bold.

Model	Dispersal	LnL	Parameters	*d*	*e*	*j*	AICc	AICc weight (%)
DEC	Unrestricted	-61.35	2	0.0155	0.0027	0	127.14	4.0
DEC	Restricted to areas connected by currents	-60.17	2	0.0193	0.0009	0	124.78	13.0
DIVALIKE	Unrestricted	-60.38	2	0.0230	0.0045	0	125.20	10.5
DIVALIKE	Restricted to areas connected by currents	-58.45	2	0.0295	0.0035	0	121.34	**72.5**
BAYAREALIKE	Unrestricted	-92.31	2	0.0441	0.1288	0	189.06	0.0
BAYAREALIKE	Restricted to areas connected by currents	-90.71	2	0.0648	0.1309	0	185.86	0.0
DEC + J	Unrestricted	-52.77	3	0.0057	$1x10^{-12}$	0.1020	112.46	1.9
DEC + J	Restricted to areas connected by currents	-49.97	3	0.0071	$1x10^{-12}$	0.1644	106.86	31.8
DIVALIKE + J	Unrestricted	-52.16	3	0.0081	$1x10^{-12}$	0.0874	111.24	3.6
DIVALIKE + J	Restricted to areas connected by currents	-49.33	3	0.0100	$1x10^{-12}$	0.1410	105.58	**60.3**
BAYAREALIKE + J	Unrestricted	-54.24	3	0.0049	$1x10^{-7}$	0.1375	115.40	0.4
BAYAREALIKE + J	Unrestricted	-52.76	3	0.0055	$1x10^{-7}$	0.1733	112.44	1.95

Table 2 Number of breeding pairs, conservation status, evolutionary distinctness and EDGE scores for shearwater species and subspecies in the study.

Scientific name	Breeding pairs	IUCN Red List Status (GE score for EDGE calculation; IUCN 2019)	Evolutionary distinctness (ED)	EDGE Score
Ardenna bulleri	350,000	Vulnerable (2)	6.2	3.4
Ardenna carneipes	74,000	Near-threatened (1)	3.6	2.2
Ardenna creatopus	29,573	Vulnerable (2)	3.6	2.9
Ardenna gravis	6,800,000	Least concern (0)	4.4	1.7
Ardenna grisea	4,400,000	Near-threatened (1)	4.7	2.4
Ardenna pacifica	4,966,000	Least concern (0)	6.2	2.0
Ardenna tenuirostris	14,800,000	Least concern (0)	4.7	1.7
Calonectris borealis	252,500	Least concern (0)	5.3	1.8
Calonectris diomedea	182,000	Least concern (0)	5.3	1.8
Calonectris edwardsii	6,312	Near-threatened (1)	5.5	2.6
Calonectris leucomelas	1,000,000	Near-threatened (1)	7.5	2.8
Puffinus assimilis haurakiensis	10,000	Least concern (0)	3.3	1.5
Puffinus bailloni bailloni	4,080	Least concern (0)	2.6	1.3
Puffinus bailloni dichrous	60,500	Least concern (0)	2.3	1.2
Puffinus bailloni nicolae	120,000	Least concern (0)	2.3	1.2
Puffinus baroli	3,360	Vulnerable (2)	2.6	2.7
Puffinus boydi	5,000	Near-threatened (1)	2.6	2.0
Puffinus elegans	16,100	Least concern (0)	3.3	1.5
Puffinus gavia	100,000	Least concern (0)	4.8	1.8
Puffinus huttoni	114,000	Endangered (3)	4.8	3.8
Puffinus lherminieri	15,700	Near-threatened (1)	3.2	2.1
Puffinus mauretanicus	3,142	Critically endangered (4)	2.7	4.1
Puffinus nativitatis	50,000	Least concern (0)	10.6	2.4
Puffinus newelli	5,000	Critically endangered (4)	4.0	4.4
Puffinus opisthomelas	41,000	Near-threatened (1)	4.0	2.3
Puffinus puffinus	399,500	Least concern (0)	3.9	1.6
Puffinus yelkouan	22,928	Vulnerable (2)	2.7	2.7