



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

Joan Ferrer Obiol^{1,2} | Helen F. James³ |
R. Terry Chesser^{3,4} | Vincent Bretagnolle⁵ | Jacob González-Solís^{2,6} |
Julio Rozas^{1,2} | Andreanna J. Welch⁷ | Marta Riutort^{1,2}

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

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

³Department of Vertebrate Zoology, National Museum of Natural History, Smithsonian Institution, Washington, District of Columbia, USA

⁴U.S. Geological Survey, Eastern Ecological Science Center, Laurel, Maryland, USA

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

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

⁷Department of Biosciences, Durham University, Durham, UK

Correspondence

Joan Ferrer Obiol, Departament de Genètica, Microbiologia i Estadística, Facultat de Biologia, Universitat de Barcelona, Barcelona, Catalonia, Spain.
Email: joan.ferrer.obiol@gmail.com

Funding information

Fundación BBVA; Spanish Government

Handling Editor: Erin Saupe

Abstract

Aim: Palaeoceanographic changes can act as drivers of diversification and speciation, even in highly mobile marine organisms. Shearwaters are a group of globally distributed and highly mobile pelagic seabirds. Despite a recent well-resolved phylogeny, shearwaters have controversial species limits, and show periods of both slow and rapid diversification. Here, we explore the role of palaeoceanographic changes on shearwaters' diversification and speciation. We investigate shearwater biogeography and the evolution of a key phenotypic trait, body size, and we assess the validity of their current taxonomy.

Location: Worldwide.

Taxa: Shearwaters (Order Procellariiformes, Family Procellariidae, Genera *Ardenna*, *Calonectris* and *Puffinus*).

Methods: We generated genomic (ddRAD) data to infer a time-calibrated species tree for the shearwaters. We estimated ancestral ranges and evaluated the roles of founder events, vicariance and surface ocean currents in driving diversification. We performed phylogenetic generalised least squares to identify potential predictors of variability in body size along the phylogeny. To assess the validity of the current taxonomy, we analysed genomic patterns of recent shared ancestry and differentiation among shearwater taxa.

Results: We identified a period of high dispersal and rapid speciation during the Late Pliocene–early Pleistocene. Species dispersal appears to be favoured by surface ocean currents, and founder events are supported as the main mode of speciation in these highly mobile pelagic seabirds. Body mass shows significant associations with life strategies and local conditions. The current taxonomy shows some incongruences with the patterns of genomic divergence.

Main Conclusions: A reduction of neritic areas during the Pliocene seems to have driven global extinctions of shearwater species, followed by a subsequent burst of speciation and dispersal probably promoted by Plio-Pleistocene climatic shifts. Our findings extend our understanding on the drivers of speciation and dispersal of highly

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2021 The Authors. *Journal of Biogeography* published by John Wiley & Sons Ltd.

mobile pelagic seabirds and shed new light on the important role of palaeoceanographic events.

KEYWORDS

biogeography, diversification, molecular dating, seabirds, speciation, taxonomy

1 | 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). In highly mobile species that live in environments with a lack of obvious physical barriers, such as the marine environment, neutral processes of panmixia or isolation-by-distance, are expected to prevail (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 et al., 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 et al., 2013; Cutter, 2013). Many allospecies first tend to differ from their close relatives in 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 et al., 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 et al., 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 under an integrative taxonomic framework.

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) (González-Solís et al., 2007; Shaffer et al., 2006; Weimerskirch et al., 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 et al., 2007).

Shearwaters are a monophyletic group in the family Procellariidae (Nunn & Stanley, 1998), 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 taxonomic classification recognises three genera and 30 species with a recently well-resolved phylogeny showing clear periods of rapid diversification (Ferrer Obiol et al., 2021). 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 biogeographical analyses. Fifth, abiotic and biotic factors are known to promote speciation in the shearwaters and related Procellariiformes; for instance, palaeoceanographic changes such as the Pleistocene climatic oscillations can act as historical drivers of speciation (Gómez-Díaz et al., 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 et al., 2018; Rayner et al., 2011). Sixth, species limits are controversial, mostly due to high morphological stasis (Austin, 1996; Austin et al., 2004); 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.

The end of the Pliocene marked the beginning of a period of great climatic and sea-level oscillations. From the late Pliocene onwards, the global area of the neritic zone abruptly diminished and started experiencing amplified fluctuations (De Boer et al., 2010). This reduction of the neritic zone has been recently hypothesised as a main driver of a marine megafauna extinction event during this period (Pimiento et al., 2017). The reliance of shearwaters on the neritic zone to forage suggests that this extinction event might have severely affected shearwaters. However, this has never been tested. Subsequently, Plio-Pleistocene climatic and sea-level oscillations have been shown to promote population differentiation and speciation in pelagic seabirds, including shearwaters (Gómez-Díaz et al., 2006; Silva et al., 2015).

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 when incomplete lineage sorting (ILS) is prevalent, particularly at recent time-scales (Angelis & Dos Reis, 2015; McCormack et al., 2011). For such events, the multispecies coalescent model (MSC) offers a more accurate solution by incorporating



the effects of ILS, which is likely the most common source of phylogenetic incongruence in rapid diversification events (Edwards et al., 2016; Maddison, 1997).

The reconstruction of ancestral ranges and evaluation of alternative biogeographical 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 et al., 2008). However, in several shearwater species, contemporary colony foundation events have been reported (Munilla et al., 2016; Storey & Lien, 1985). Environmental differences experienced by founder populations have the potential to precipitate speciation (Clegg et al., 2002). On the other hand, the emergence of physical barriers to gene flow can also initiate genetic divergence of populations and ultimately result in allopatric speciation (Coyne, & Orr, 2004). Concordant with this process, seabird populations or species complexes whose breeding distributions are fragmented by land masses tend to show significant genetic differentiation and phylogeographical structure (Friesen, Burg, et al., 2007), and shearwaters are no exception (Austin et al., 2004). Differences in ocean regimes can also act as a barrier to gene flow driving allopatric speciation (Friesen, 2015; Gómez-Díaz et al., 2009). Understanding how these processes have shaped the biogeographical history of shearwaters can provide important information 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 perform divergence dating analyses to evaluate the impact of the Pliocene marine megafauna extinction and the Plio-Pleistocene climatic oscillations. We employ a multispecies coalescent approach (MSC) to account for the high levels of ILS affecting the shearwater phylogeny. We then infer the biogeographical 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 and geographical forces responsible for the variability in a key phenotypic trait, body size. Finally, we assess the validity of the current taxonomic classification of the group by analysing genomic patterns of recent shared ancestry and differentiation among shearwater taxa.

2 | MATERIALS AND METHODS

2.1 | Sampling and sequence data generation

We collected 68 blood or tissue samples from 25 of the 32 recognised species of shearwaters (Gill et al., 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 (see Acknowledgments). Sequence data for 51 of these samples were generated previously in a recent phylogenomic study (Ferrer Obiol et al., 2021).

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). We used a Qubit Fluorometer (Life Technologies) to quantify and standardise DNA concentrations of all samples at 10 ng/μl. 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 indices 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 Blue Pippin Prep size fractionator (Sage Science, Beverly, Ma) with 2% agarose cassettes. Libraries were amplified in a final PCR step for 10 PCR cycles prior to sequencing in a single lane on an Illumina HiSeq4000 platform with 150-bp paired-end (PE) reads.

2.2 | PE-ddRAD-Seq data filtering and assembly

Raw reads were demultiplexed and cleaned using `process_radtags` in `STACKS` v2.41 (Rochette et al., 2019). To maximise the amount of biological information, we built loci using the forward reads with parameters optimised for this shearwater dataset (see Ferrer Obiol et al., 2021) 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 (Maruki & Lynch, 2015, 2017) and phased haplotypes using `GSTACKS`. Subsequently, we mapped the `GSTACKS` catalogue to the Balearic shearwater (*Puffinus mauretanicus*) genome assembly (Cuevas-Caballé 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 catalogue using `STACKS-INTEGRATE-ALIGNMENTS` (Paris et al., 2017). Finally, we used the `POPULATIONS` program to filter single nucleotide polymorphism (SNP) data requiring a minimum minor allele count (MAC) of two to maximise taxon-specific variation and an observed heterozygosity below 50% to generate datasets for downstream analysis.

2.3 | Species tree inference

To estimate a time-calibrated species tree for shearwaters, we applied the SNP-based MSC approach of Stange et al. (2018) implemented in the SNAPP v.1.3. (Bryant et al., 2012) package of the program BEAST2 v.2.5.0 (Bouckaert et al., 2019). To prepare a suitable dataset for this method, we selected one or two individuals per taxon (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 after checking that analysis using all the SNPs yielded comparable results. 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, to remove potentially linked SNPs, we only retained SNPs separated by a minimum distance of 5000 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. (2021) based on three fossil calibrations (see calibration strategy B therein) and a relaxed clock. SD was calculated to fit the posterior distribution for the root in Ferrer Obiol et al. (2021). This divergence time estimate for the root was further supported by a global study on birds using relaxed clocks (Jetz et al., 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. (2021) using UCE and ddRAD data. We also tested the robustness of divergence-time estimates by performing three additional analyses. First, we explored the effects of fixing the topology by also performing the analysis without the topology being fixed. Second, we evaluated the use of fossil calibrations using two additional calibration points based on those described in strategy B of Ferrer Obiol et al. (2021). Briefly, we used minimum age constraints based on fossils for the most recent common ancestor (MRCA) of Crown *Calonectris* and the MRCA of *A. bulleri* and *A. pacifica*. Third, we evaluated the effects of using a parameterisation that should prevent the inclusion of sequencing errors at the expense of removing taxon-specific variation by setting a minimum MAC of three in the POPULATIONS program. 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 for each run (effective sample sizes >300) using TRACER v.1.7.1 (Rambaut et al., 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).

Because the Stange et al. (2018) approach only estimates a single value of θ for all branches, we also constructed a SNAPP phylogeny using the finite-sites model implemented in SNAPP to estimate θ values for each branch. This approach allows the estimation of both branch lengths (times) and population sizes (θ) for each branch (Bryant et al., 2012).

2.4 | Ancestral range estimation

We performed biogeographical analyses 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 et al., 2013), and the time-calibrated shearwater tree. We compared the 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' MRCA, we used the ranges of the two most closely related outgroup lineages (Estandía et al., 2021) (for which no ddRAD data are available): genus *Procellaria*, and genera *Pseudobulweria* and *Bulweria*. *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 Jetz et al. (2012) using the TimeTree database (Kumar et al., 2017). Outgroups were incorporated into the time-calibrated shearwater tree using the `bind.tree` function from the 'ape' package (Paradis & Schliep, 2019) in R.

2.5 | Phylogenetic comparative analyses

To evaluate the potential predictors of body size in shearwaters, we performed univariate and multivariate phylogenetic generalised least squares regressions (PGLS) using the R package 'caper' (Orme et al., 2013). Specifically, we considered four body size measures: mean body mass, range of body mass (maximum body

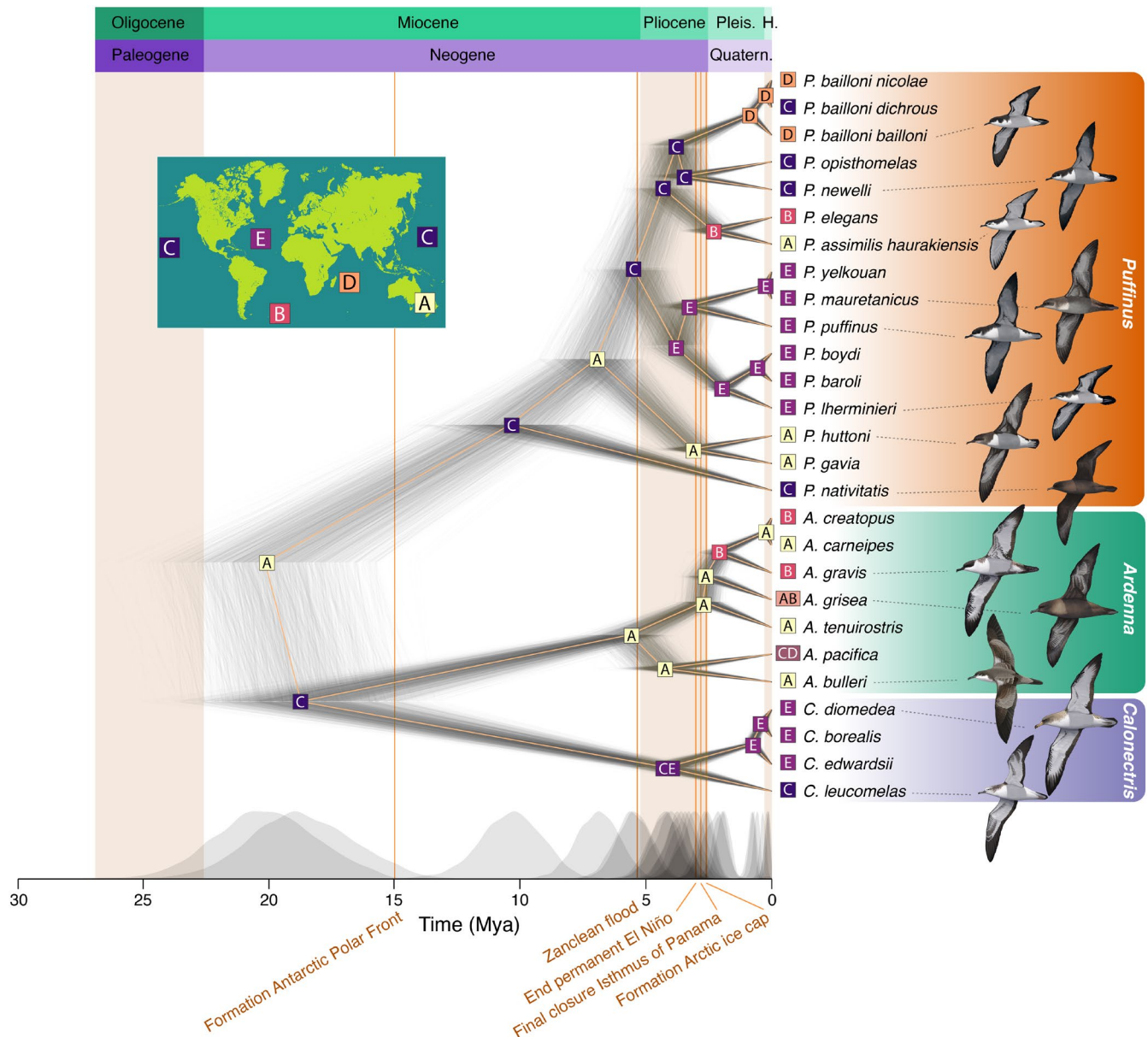


FIGURE 1 Time-calibrated species tree of the shearwaters using a constraint on the root age and a fixed topology. Geological periods and epochs are labelled above the tree. Posterior densities of divergence times are shown below the species tree. Note the diversification peak during the late Pliocene–early 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; 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)). Posterior estimates of divergence times are summarised in Table S2. Illustrations by Martí Franch are representative shearwater species depicted by their lineages

mass – minimum body mass), wing length and total body length, and five predictors: minimum, mean and maximum breeding latitudes (in absolute values), latitudinal range occupied by a species year-round (maximum latitude – minimum latitude of the overall distribution of the species) and 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 \times wingspan)). Data were retrieved from Birds of

the World (Billerman et al., 2020) and Onley and Scofield (2013). Following recommendations of Revell (2010), we simultaneously estimated the λ parameter (Pagel, 1999) to account for deviations from a pure Brownian motion (BM). To ensure the absence of collinearity among predictors in multivariate PGLS, we only incorporated a single breeding latitude predictor (maximum breeding latitude, which showed the strongest correlations in the univariate analyses). For all predictors retained, variance inflation factors were all under six, which is below the suggested threshold of 10 (Hair et al.,

1998). PGLS analyses were run using the time-calibrated tree. Unsourced species were 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; Kawakami et al., 2018; Martínez-Gómez et al., 2015; Pyle et al., 2011). 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 et al., 2017). We also reconstructed ancestral states for migratory strategy using maximum likelihood (ML) with the function `rerootingMethod` in the R package 'phytools'. We also modelled the correlation between the nucleotide substitution rate and the equilibrium GC content (GC*) with mean body mass and the number of breeding pairs as a multivariate Brownian motion in COEVOL (Lartillot & Poujol, 2011; see Supplementary Information Text).

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 et al., 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: $EDGE = \ln(1 + ED) + GE \times \ln(2)$.

2.6 | Patterns of recent coancestry and sequence divergence

To explore congruence between the current shearwater taxonomic classification and the genetic structure among species, we used `FINERADSTRUCTURE` v0.3.2 (Malinsky et al., 2018) to infer the shared ancestry among all individuals. `FINERADSTRUCTURE` uses haplotype linkage information to derive a coancestry 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 8049 PE-ddRAD loci containing a total of 94,811 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 1000 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 taxonomic classification and genomic divergence, we examined the distribution of pairwise genetic distances using loci present in at least 95% of the individuals (1561 loci; 17,675 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 & Grünwald, 2017), and we

calculated pairwise distances using the `dist.dna` function from the 'ape' package in R.

3 | 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 39× (SD = 19). Locus length ranged from 140 to 239 bp with a median of 198 bp (SD = 25.5).

3.1 | 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., 2021), except for the relationship between *A. grisea* and *A. tenuirostris* (Table S2; Figure S1). This incongruence was already identified in the previous study using different methods and datasets, and was caused by high levels of ILS, rate heterogeneity and GC-biased gene conversion.

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 10.39 Mya (95% HPD: 12.67–7.97 Mya). The TMRCA of *Ardenna* was inferred to be 5.58 Mya (95% HPD: 6.84–4.29 Mya) and the TMRCA of *Calonectris* 4.30 Mya (95% HPD: 5.43–3.10 Mya). If the divergence times are accurate, then shearwater speciation increased during the Pliocene reaching a peak by the late Pliocene (Figure 1), when most of the modern biogeographical groups of shearwaters were already present.

Using the same three fossil calibrations (see Materials and Methods), shearwater divergence times inferred using the MSC were on average 1.27 My younger than those estimated by Ferrer Obiol et al. (2021) using concatenation (Table S2). MSC analyses using these fossil calibrations resulted in slightly older estimates (0.13 My older on average) compared to the same analyses using a single age constraint on the root (Figure S2, Table S2). Fixing the topology had a negligible effect on age estimates (0.03 My older on average). Conversely, analysis using a minimum MAC of three resulted in considerably younger age estimates (0.57 My on average), highlighting the importance of maximising taxon-specific variation in divergence dating analyses (see Linck & Battey, 2019 for a full discussion on the impacts of the minor allele frequency threshold on large DNA sequence datasets).

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,

TABLE 1 Comparison of models of ancestral range estimation for the shearwaters

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	1×10^{-12}	0.1020	112.46	1.9
DEC + J	Restricted to areas connected by currents	-49.97	3	0.0071	1×10^{-12}	0.1644	106.86	31.8
DIVALIKE + J	Unrestricted	-52.16	3	0.0081	1×10^{-12}	0.0874	111.24	3.6
DIVALIKE + J	Restricted to areas connected by currents	-49.33	3	0.0100	1×10^{-12}	0.1410	105.58	60.3
BAYAREALIKE + J	Unrestricted	-54.24	3	0.0049	1×10^{-7}	0.1375	115.40	0.4
BAYAREALIKE + J	Unrestricted	-52.76	3	0.0055	1×10^{-7}	0.1733	112.44	1.95

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.

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.

3.2 | Biogeographical 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 some speciation events in shearwaters might be the result of 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).

3.3 | Phylogenetic generalised least squares of body size

Univariate PGLS analyses recovered several significant correlations (false discovery rate (FDR) < 0.05) between body size measures and the predictors (Figure 2a; Table S3). Mean body mass showed significant correlations with nearly 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

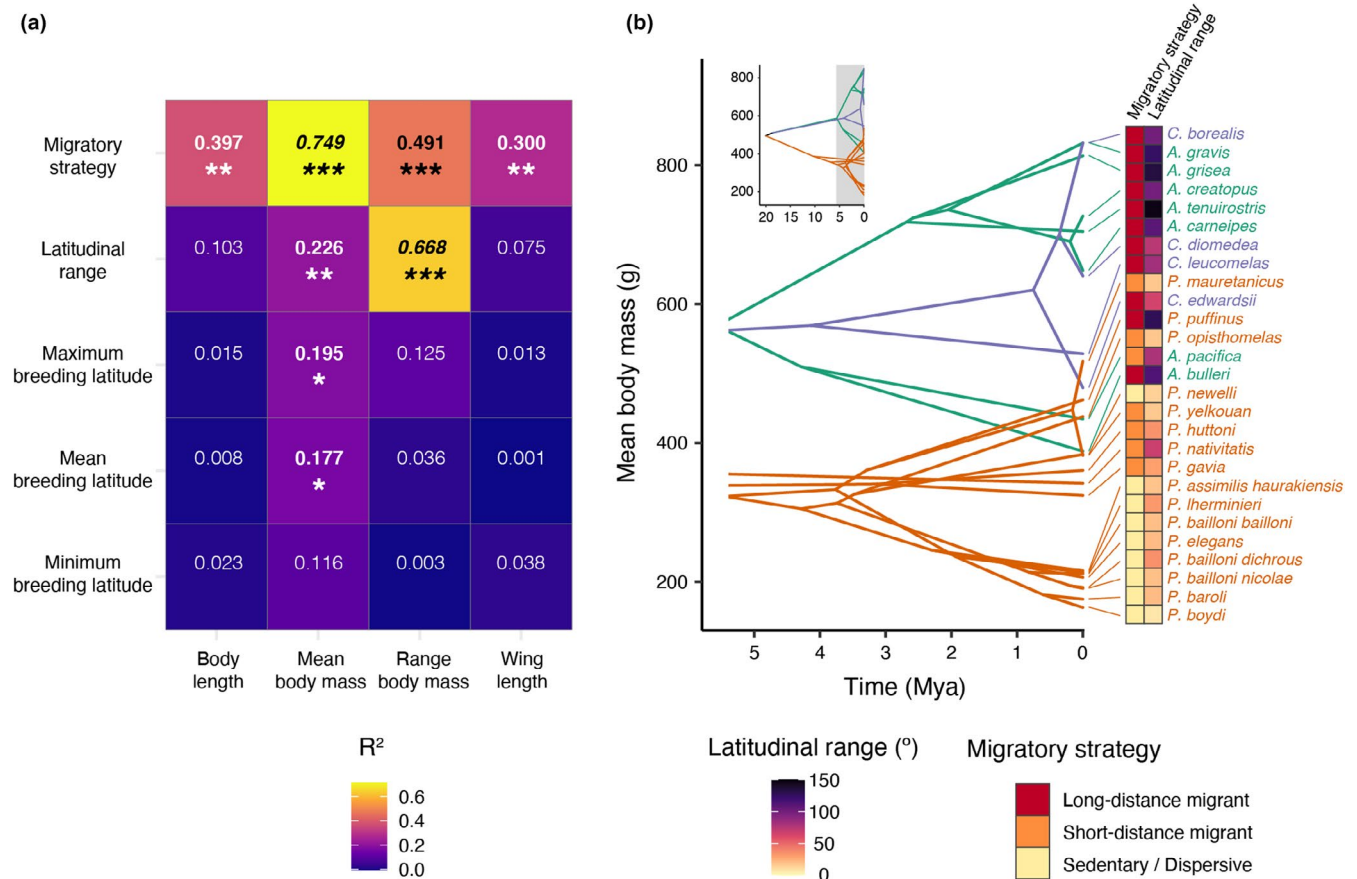


FIGURE 2 Migratory strategy and latitudinal range are the best predictors of body size. (a) Heatmap showing adjusted R^2 values for phylogenetic generalised least squares (PGLS) analyses of body size measures against the predictors. Positive correlation coefficients were recovered for each test. Numbers within each tile show the adjusted R^2 values and are shown in bold when significant after adjusting p -values by false discovery rate (FDR) correction for multiple testing (FDR: $*** < 0.001 >$, $** < 0.01 >$, $* < 0.05 >$) and in bold italics when also significant in multivariate PGLS analyses. (b) Inset: phenogram of mean body mass constructed in 'phytools' showing abrupt differences in mean body mass between sister clades. The values of the reconstructions for mean body mass should be interpreted cautiously, due to the lack of fossil calibrations to calibrate the ancestral state reconstructions. The main plot is an amplification of the same phenogram from the Pliocene onwards, as depicted by the grey shaded area in the inset. 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

mean body mass, with the strongest correlation recovered for maximum breeding latitude (Figure S5c; adjusted $R^2 = 0.195$). Multivariate PGLS showed that migratory strategy was the only significant predictor of mean body mass and latitudinal range was the only significant predictor of body mass range when using migratory strategy, latitudinal range and maximum breeding latitude as predictors (Figure 2a). 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 suggested that the MRCA 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).

3.4 | 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., 2021), and all the species and subspecies included in the study were recovered as unique clusters by the *FINESTRUCTURE* clustering algorithm (Lawson et al., 2012), except for *P. mauretanicus* and *P. yelkouan*, which were aggregated in a single cluster.

Overall, the distributions of genetic distances were consistent with the current taxonomic classification. 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 column in Figure 4). In addition, the genetic distances between *P.*

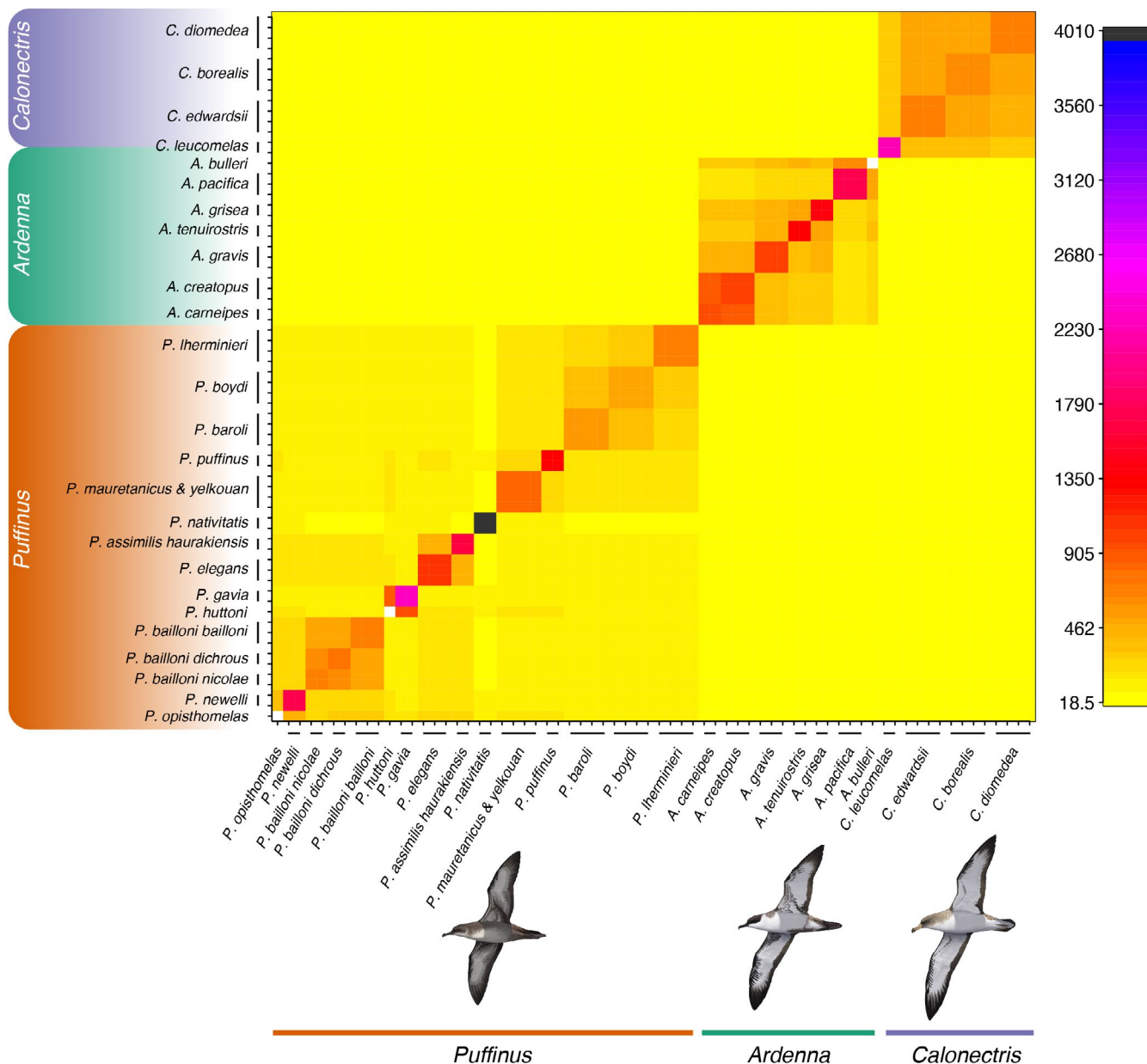


FIGURE 3 Clustered FINERADSTRUCTURE coancestry matrix based on 8049 PE-ddRAD loci. Pairwise coancestry coefficients are colour coded from low (yellow) to high (black). Every name represents a discernible discrete cluster based on the pairwise matrix of coancestry coefficients, defined by a posterior 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. The high coancestry coefficients observed within *P. nativitatis* are likely due to a high level of relatedness between the two individuals

boydi and *P. baroli*, and between the different Atlantic *Calonectris* species were within the interval of genetic distances among different subspecies (second column in Figure 4).

4 | DISCUSSION

This study presents a fundamental analysis of the potential drivers of diversification and speciation in a major group of seabirds, by constructing a MSC time-calibrated species tree and performing

biogeographical analysis for shearwaters based on a fully resolved phylogeny. This allowed us to explore the drivers of shearwater diversification and to reconstruct their biogeographical history. Specifically, we discuss the importance of palaeoceanographic events, founder-event speciation and oceanic currents in shearwater diversification. We also discuss the role of body size in shearwater evolution, and we consider potential ecological and evolutionary forces that may have shaped its evolution. Lastly, we use the evidence uncovered here to explain incongruences between the current taxonomic classification and the patterns of genomic divergence.

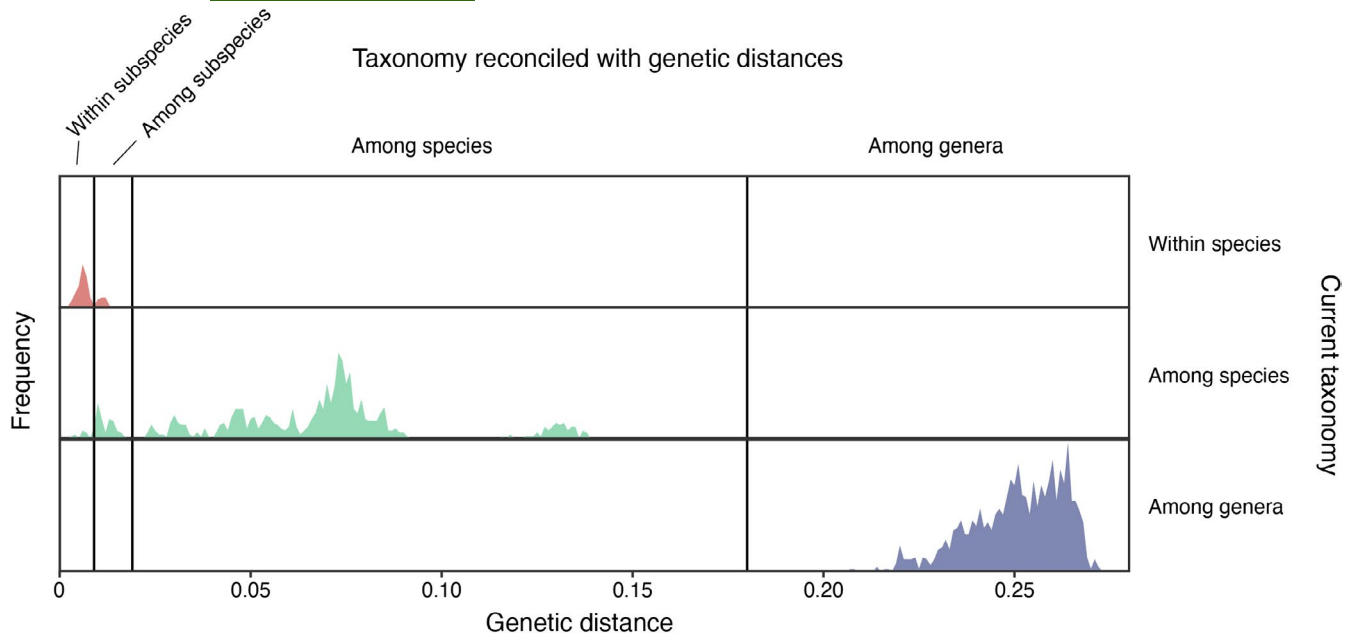


FIGURE 4 The distributions of genetic distances within and among species overlap. Distribution of genetic distances at different taxonomic levels in the shearwaters according to current classification (upper row: within species, middle row: among species and lower row: among genera). Columns show proposed orientative limits for values within subspecies, among subspecies, among species and among genera that would reconcile taxonomic ranks with genetic distances. 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

4.1 | 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., 2021), which would likely benefit from the use of a relaxed clock (Drummond et al., 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., 2021). 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 time-scales 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).

4.2 | Biogeographical history of shearwaters

Our biogeographical analyses suggest that founder events are the main mechanism of speciation in shearwaters, as expected for highly mobile species breeding on islands (Matzke, 2014). 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 biogeographical 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). Ocean currents play a huge role in marine productivity and represent important foraging areas for seabirds throughout the world (Block et al., 2011; Péron et al., 2012). In addition, 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 et al., 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 biogeographical 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). 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) is uncertain (Ferrer Obiol et al., 2021). Thus, we argue that 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 distribution in the North Pacific and North Atlantic or in the North Pacific. Fossils of at least five species have been described from the North Atlantic dating back to ~14 Mya (Olson, 2008, 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 cannot eliminate the possibility that the regions where these fossils were found were not the breeding areas for the species. The estimated divergence time (~4.30 Mya) between the North Pacific and the North Atlantic clades is slightly older than 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; Stange et al., 2018).

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 or even abundant on North American and European coasts (Carey et al., 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). It is thus reasonable to 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. Mass mortality events have been reported for *Ardenna* species in the Northern Hemisphere, for instance, during 1997 hundreds of thousands of *A. tenuirostris* starved to death in the Bering Sea (Hyrenbach et al., 2001) and up to 400,000 individuals were deposited on beaches (Mendenhall, 1997). Such mortality events represent excellent opportunities for fossilisation.

The ancestors of *Puffinus* acquired the strongest diving adaptations of the three genera (Olson, & Rasmussen, 2001). These

adaptations allowed 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). The divergence between the Atlantic and Pacific *Puffinus* clades (Node 4 in Figure S2) occurred ~5.46 Mya, which is more than one My earlier than the split between the Atlantic and Pacific *Calonectris* lineages. This difference could be due to the lower mobility of *Puffinus* shearwaters compared to *Calonectris* and *Ardenna* and suggests another vicariant event as the result of the gradual closure of the Isthmus of Panama. Indeed, 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). The lower dispersal of *Puffinus* 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 conservation 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, intraspecific and interspecific competition for nest sites plays an important role in limiting populations of small Procellariiformes, such as *Puffinus* shearwaters (Ramos et al., 1997). At sea, fisheries bycatch is also a main threat for *Puffinus* shearwaters (Cortés et al., 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.

Divergence time estimation analyses suggested that across the three genera, the late Pliocene was likely a period of high and rapid speciation and dispersal, as shown by the peak in diversification and dispersal during this period (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 (Miller et al., 2005). Neritic waters, which represent the main foraging grounds for medium and large shearwaters, especially during the breeding period, suffered a significant sudden reduction during the late 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 threefold increase in the extinction rate of megafauna associated with coastal habitats (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

TABLE 2 Number of breeding pairs, conservation status, evolutionary distinctness (ED) and Evolutionary Distinct and Globally Endangered (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
<i>Ardenna bulleri</i>	350,000	Vulnerable (2)	6.9	3.5
<i>Ardenna carneipes</i>	74,000	Near Threatened (1)	3.9	2.3
<i>Ardenna creatopus</i>	29,573	Vulnerable (2)	3.9	3.0
<i>Ardenna gravis</i>	6,800,000	Least Concern (0)	4.8	1.8
<i>Ardenna grisea</i>	4,400,000	Near Threatened (1)	5.2	2.5
<i>Ardenna pacifica</i>	4,966,000	Least Concern (0)	6.9	2.1
<i>Ardenna tenuirostris</i>	14,800,000	Least Concern (0)	5.3	1.8
<i>Calonectris borealis</i>	252,500	Least Concern (0)	5.5	1.9
<i>Calonectris diomedea</i>	182,000	Least Concern (0)	5.5	1.9
<i>Calonectris edwardsii</i>	6312	Near Threatened (1)	5.7	2.6
<i>Calonectris leucomelas</i>	1,000,000	Near Threatened (1)	7.9	2.9
<i>Puffinus assimilis haurakiensis</i>	10,000	Least Concern (0)	4.4	1.7
<i>Puffinus bailloni bailloni</i>	4080	Least Concern (0)	3.0	1.4
<i>Puffinus bailloni dichrous</i>	60,500	Least Concern (0)	2.7	1.3
<i>Puffinus bailloni nicolae</i>	120,000	Least Concern (0)	2.7	1.3
<i>Puffinus baroli</i>	3360	Vulnerable (2)	3.1	2.8
<i>Puffinus boydi</i>	5000	Near Threatened (1)	3.1	2.1
<i>Puffinus elegans</i>	16,100	Least Concern (0)	4.4	1.7
<i>Puffinus gavia</i>	100,000	Least Concern (0)	5.9	1.9
<i>Puffinus huttoni</i>	114,000	Endangered (3)	5.9	4.0
<i>Puffinus lherminieri</i>	15,700	Near Threatened (1)	3.8	2.3
<i>Puffinus mauretanicus</i>	3142	Critically Endangered (4)	3.1	4.2
<i>Puffinus nativitatis</i>	50,000	Least Concern (0)	11.0	2.5
<i>Puffinus newelli</i>	5000	Critically Endangered (4)	4.8	4.5
<i>Puffinus opisthomelas</i>	41,000	Near Threatened (1)	4.8	2.5
<i>Puffinus puffinus</i>	399,500	Least Concern (0)	4.7	1.7
<i>Puffinus yelkouan</i>	22,928	Vulnerable (2)	3.1	2.8

suggests that the Pliocene extinction may have severely affected the group. The subsequent burst of speciation and dispersal could have been associated with late Pliocene and Pleistocene climatic shifts that may have promoted geographical splitting and bottlenecks (Avisé & 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 (Spheniscidae; Vianna et al., 2020) and even in deep sea species (Eilertsen & Malaquias, 2015).

4.3 | 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; Tveraa et al., 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). Migratory species tend to have longer wings (Marchetti et al., 1995; Minias et al., 2015) and longer and larger bones to increase the surface available for attachment of the main muscles involved in active flight (Calmaestra & Moreno, 2000). In shearwaters, the significant correlations of all the other body size measures with migratory strategy seem to support these general rules for birds. The stronger correlation between migratory strategy and mean body mass may be the product of body mass being a measure that integrates the effect of different adaptations to a migratory lifestyle. Alternatively, the correlations of other body size

measures with migratory strategy may have higher error associated with the measurements as these may depend on the sampler.

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 levels (Bergmann, 1848) although the mechanisms responsible for the generation of this trend are subject to much debate (Ashton, 2002; Meiri, 2011). In 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, 2021; 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, the lack of significance of breeding latitude as a predictor of mean body size when modelled in a multivariate framework suggests that 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. Indeed, higher body condition has been associated with higher breeding success in several species of Procellariiformes (Barbraud & Chastel, 1999; Barbraud & Weimerskirch, 2005).

4.4 | Considerations of shearwater taxonomic classification

Species delimitation in shearwaters is a challenging and controversial topic, partly due to their remarkably similar morphology (Austin et al., 2004). Conflict has arisen among 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 et al., 2009). In addition, despite being a promising trait for species delimitation, analyses of

shearwater vocalisations 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 *P. mauretanicus* and *P. yelkouan* (Figure 3). Furthermore, although we do not consider there to be a genetic cut-off for species-level divergence, the genetic distances between these recently diverged species and those between *A. creatopus* and *A. carneipes* were the lowest amongst any pair of species and overlapped with the genetic distances observed between individuals of the same subspecies (Figure 4). *P. mauretanicus* and *P. yelkouan* were granted species status based on morphological and osteological characters and reciprocal monophyly using cytochrome b sequences (Heidrich et al., 1998; Sangster et al., 2002). However, more recently, a lack of correspondence at the individual level was found between phenotypic characters, stable isotope analyses, nuclear and mtDNA, and was attributed to admixture between the two species (Genovart et al., 2012; Militão et al., 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 et al., 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. boydii* and *P. baroli*, and the three Atlantic *Calonectris* species. These species complexes are the subject of ongoing taxonomic debate (Genovart et al., 2013; Gómez-Díaz et al., 2009; Olson, 2010; Ramos et al., 2020; Sangster et al., 2005). As a final consideration, our analyses based on genomic data and the ongoing taxonomic debate suggest that the species status of these taxa should be re-evaluated. Future studies should use species delimitation approaches under an integrative taxonomic framework, combining genomic data with a thorough morphological re-evaluation including a detailed evaluation of vocalisations. Further research should also 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. Martínez-Gómez 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 Muséum National d'Histoire Naturelle. We thank the following authorities for issuing the permits needed for this work: Mikurashima Village Office (Japan), Direcção Nacional do Ambiente (Cabo Verde), Gobierno de Canarias (Spain), Junta de Andalucía (Spain), Govern de les Illes Balears (Spain), Organismo Autónomo de Parques Nacionales, Miteco (Spain), Port-Cros National Park (France), Secretaría de Gobernación and Secretaría de Medio Ambiente y Recursos Naturales del Gobierno de México, South Iceland Nature Centre, Northern Ireland Environment Agency (UK), Falkland Islands Government, Tristan da Cunha Government, Região Autónoma dos Açores (Portugal). 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. We also thank Erin Saupe, Rosemary Gillespie and three anonymous reviewers for providing constructive criticism and extensive feedback to improve the article. Research funding was provided by the Fundación BBVA (program 'Ayudas a Equipos de Investigación Científica 2017', project code 062_17 to M.R.), by Fondos FEDER and by Ministerio de Ciencia e Innovación (CGL2009-11278/BOS), Ministerio de Economía y Competitividad (CGL2013-42585-P), and Ministerio de Economía, Industria y Competitividad (CGL2016-78530-R) from the Spanish Government. Any use of trade, product or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

CONFLICT OF INTEREST

Authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

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, PHYMLIP format file for CoEvoL analysis, FINESTRUCTURE input files and MCC SNAPP trees are available from the Dryad Digital Repository: <https://doi.org/10.5061/dryad.qv9s4mwg2>. Scripts used in this paper are available at https://github.com/jferrerobiol/shearwater_biogeography.

ORCID

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

Helen F. James  <https://orcid.org/0000-0002-2495-6133>

R. Terry Chesser  <https://orcid.org/0000-0003-4389-7092>

Vincent Bretagnolle  <https://orcid.org/0000-0002-2320-7755>

Jacob González-Solis  <https://orcid.org/0000-0002-8691-9397>

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

Andreanna J. Welch  <https://orcid.org/0000-0002-5947-4695>

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

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. <https://doi.org/10.1093/czoolo/61.5.874>
- 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. <https://doi.org/10.1046/j.1466-822X.2002.00313.x>
- 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. <https://doi.org/10.1006/mpev.1996.0060>
- 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. [https://doi.org/10.1642/0004-8038\(2004\)121\[0847:AGMPOT\]2.0.CO;2](https://doi.org/10.1642/0004-8038(2004)121[0847:AGMPOT]2.0.CO;2)
- 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. <https://doi.org/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. <https://doi.org/10.1046/j.1420-9101.2000.00151.x>
- Barbraud, C., & Chastel, O. (1999). Early body condition and hatching success in the snow petrel *Pagodroma nivea*. *Polar Biology*, *21*(1), 1–4. <https://doi.org/10.1007/s003000050326>
- Barbraud, C., & Weimerskirch, H. (2005). Environmental conditions and breeding experience affect costs of reproduction in blue petrels. *Ecology*, *86*(3), 682–692. <https://doi.org/10.1890/04-0075>
- 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. *Göttinger 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. <https://birdsoftheworld.org/bow/home>
- BirdLife International. (2020). *IUCN Red List for birds*. Downloaded from <http://www.birdlife.org> on 17/08/2020
- Block, B. A., Jonsen, I. D., Jorgensen, S. J., Winship, A. J., Shaffer, S. A., Bograd, S. J., Hazen, E. L., Foley, D. G., Breed, G. A., Harrison, A. L., & Ganong, J. E. (2011). Tracking apex marine predator movements in a dynamic ocean. *Nature*, *475*(7354), 86–90.
- Bouckaert, R. R. (2010). DensiTree: Making sense of sets of phylogenetic trees. *Bioinformatics*, *26*(10), 1372–1373. <https://doi.org/10.1093/bioinformatics/btq110>
- 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 non-passerine 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. <https://doi.org/10.1093/molbev/mss086>

- Burger, A. E. (2001). Diving depths of shearwaters. *The Auk*, 118(3), 755–759. <https://doi.org/10.1093/auk/118.3.755>
- Calmaestra, R. G., & Moreno, E. (2000). Ecomorphological patterns related to migration: A comparative osteological study with passerines. *Journal of Zoology*, 252, 495–501. <https://doi.org/10.1111/j.1469-7998.2000.tb01232.x>
- Carboneras, C., & Bonan, A. (2019). Petrels, Shearwaters (Procellariidae). In J. del Hoyo, A. Elliott, J. Sargatal, D. A. Christie, & E. de Juana (Eds.), *Handbook of the birds of the world alive*. Lynx Edicions.
- 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. <https://doi.org/10.1071/MU13115>
- 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. <https://doi.org/10.1111/mec.12413>
- Clegg, S. M., Degnan, S. M., Kikkawa, J., Moritz, C., Estoup, A., & Owens, I. P. (2002). Genetic consequences of sequential founder events by an island-colonizing bird. *Proceedings of the National Academy of Sciences*, 99(12), 8127–8132. <https://doi.org/10.1073/pnas.102583399>
- 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. <https://doi.org/10.3354/meps12026>
- Coulson, J. (2002). Colonial breeding in seabirds. In E. A. Schreiber, & J. Burger (Eds.), *Biology of Marine Birds* (pp. 87–113). CRC press.
- Coyne, J. A., & Orr, H. A. (2004). *Speciation* (Vol. 37). Sinauer Associates Sunderland.
- 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. <https://doi.org/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. <https://doi.org/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. <https://doi.org/10.1016/j.ympev.2013.06.006>
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., & Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, 27(15), 2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>
- De Boer, B., van de Wal, R. S. W., Bintanja, R., Lourens, L. J., & Tuentler, E. (2010). Cenozoic global ice-volume and temperature simulations with 1-D ice-sheet models forced by benthic delta O-18 records. *Annals of Glaciology*, 51, 23–33.
- 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.
- 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. <https://doi.org/10.1371/journal.pbio.0040088>
- Edwards, S. V., Xi, Z., Janke, A., Faircloth, B. C., McCormack, J. E., Glenn, T. C., Zhong, B., Wu, S., Lemmon, E. M., Lemmon, A. R., Leaché, A. D., Liu, L., & Davis, C. C. (2016). Implementing and testing the multi-species coalescent model: A valuable paradigm for phylogenomics. *Molecular Phylogenetics and Evolution*, 94, 447–462. <https://doi.org/10.1016/j.ympev.2015.10.027>
- 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.
- Estandía, A., Chesser, R. T., James, H. F., Levy, M. A., Ferrer Obiol, J., Bretagnolle, V., González-Solís, J., & 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*. <https://doi.org/10.1101/2021.07.27.453752>
- 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. (2021). Integrating sequence capture and restriction-site associated DNA sequencing to resolve recent radiations of pelagic seabirds. *Systematic Biology*, 70(5), 976–996. <https://doi.org/10.1093/sysbio/syaa101>
- 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. <https://doi.org/10.1111/mec.14486>
- Friesen, V. L., Burg, T. M., & McCoy, K. D. (2007). Mechanisms of population differentiation in seabirds. *Molecular Ecology*, 16(9), 1765–1785. <https://doi.org/10.1111/j.1365-294X.2006.03197.x>
- 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. <https://doi.org/10.1073/pnas.0700446104>
- Genovart, M., Arcos, J. M., Álvarez, D., McMin, M., Meier, R., B. Wynn, R., Guilford, T., & 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., Thibault, J.-C., Igual, J. M., Bauzá-Ribot, M. D. 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), e70711. <https://doi.org/10.1371/journal.pone.0070711>
- Gill, F., Donsker, D., & Rasmussen, P. (2020). *IOC World Bird List (v10.1)*. <https://doi.org/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. <https://doi.org/10.3354/meps07974>
- 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. <https://doi.org/10.1016/j.ympev.2006.05.006>
- González-Solís, J. (2004). Sexual size dimorphism in northern giant petrels: Ecological correlates and scaling. *Oikos*, 105(2), 247–254. <https://doi.org/10.1111/j.0030-1299.2004.12997.x>
- González-Solís, 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. [https://doi.org/10.1890/1540-9295\(2007\)5\[297:TMAMIT\]2.0.CO;2](https://doi.org/10.1890/1540-9295(2007)5[297:TMAMIT]2.0.CO;2)

- González-Solís, J., Felicísimo, A., Fox, J. W., Afanashev, V., Kolbeinsson, Y., & Muñoz, J. (2009). Influence of sea surface winds on shearwater migration detours. *Marine Ecology Progress Series*, 391, 221–230. <https://doi.org/10.3354/meps08128>
- Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (1998). *Multivariate data analysis*. Pearson Prentice Hall.
- 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. [https://doi.org/10.1016/S0305-1978\(97\)00085-9](https://doi.org/10.1016/S0305-1978(97)00085-9)
- Heled, J., & Bouckaert, R. R. (2013). Looking for trees in the forest: Summary tree from posterior samples. *BMC Evolutionary Biology*, 13, 221. <https://doi.org/10.1186/1471-2148-13-221>
- Howard, H. (1971). Pliocene avian remains from Baja California. *Contributions in Science*, 217, 1–17. <https://doi.org/10.5962/p.241203>
- Hyrenbach, K. D., Baduini, C. L., & Hunt, G. L. Jr. (2001). Line transect estimates of Short-tailed shearwater *Puffinus tenuirostris* mortality in the south-eastern Bering Sea, 1997–1999. *Marine Ornithology*, 29, 11–18.
- 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. <https://doi.org/10.1371/journal.pone.0000296>
- 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., & Grünwald, 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. <https://doi.org/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. <https://doi.org/10.1093/molbev/msx116>
- 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. <https://doi.org/10.1093/sysbio/syt040>
- 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. <https://doi.org/10.1093/molbev/msq244>
- Lawson, D. J., Hellenenthal, G., Myers, S., & Falush, D. (2012). Inference of population structure using dense haplotype data. *PLoS Genetics*, 8(1), e1002453. <https://doi.org/10.1371/journal.pgen.1002453>
- Leaché, A. D., Zhu, T., Rannala, B., & Yang, Z. (2018). The spectre of too many species. *Systematic Biology*, 68(1), 168–181. <https://doi.org/10.1093/sysbio/syy051>
- 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. <https://doi.org/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. <https://doi.org/10.1093/bioinformatics/btr509>
- 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., Marth, G., Abecasis, G., & Durbin, R. (2009). 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and samtools. *Bioinformatics*, 25(16), 2078–2079.
- Linck, E., & Battey, C. J. (2019). Minor allele frequency thresholds strongly affect population structure inference with genomic data sets. *Molecular Ecology Resources*, 19(3), 639–647. <https://doi.org/10.1111/1755-0998.12995>
- Lombal, A. J., Wenner, T. J., Lavers, J. L., Austin, J. J., Woehler, E. J., Hutton, I., & Burridge, C. P. (2018). Genetic divergence between colonies of Flesh-footed Shearwater *Ardenna carneipes* exhibiting different foraging strategies. *Conservation Genetics*, 19(1), 27–41. <https://doi.org/10.1007/s10592-017-0994-y>
- Maddison, W. P. (1997). Gene trees in species trees. *Systematic Biology*, 46(3), 523–536. <https://doi.org/10.1093/sysbio/46.3.523>
- Mainwaring, M. C., & Street, S. E. (2021). Conformity to Bergmann's rule in birds depends on nest design and migration. *Ecology and Evolution*, 11(19), 13118–13127.
- 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. <https://doi.org/10.1093/molbev/msy023>
- 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. <https://doi.org/10.2307/3677316>
- Martínez-Gómez, J. E., Matías-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. <https://doi.org/10.1534/genetics.115.179077>
- Maruki, T., & Lynch, M. (2017). Genotype calling from population-genomic sequencing data. *G3: Genes|genomes|genetics*, 7(5), 1393–1404. <https://doi.org/10.1534/g3.117.039008>
- Matzke, N. J. (2013). *BioGeoBEARS: BioGeography with Bayesian (and likelihood) evolutionary analysis in R Scripts*. R Package, Version 0. 2, 1, 2013.
- Matzke, N. J. (2014). Model selection in historical biogeography reveals that founder-event speciation is a crucial process in Island clades. *Systematic Biology*, 63(6), 951–970. <https://doi.org/10.1093/sysbio/syu056>
- 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. <https://doi.org/10.1111/j.1558-5646.2010.01097.x>
- Meiri, S. (2011). Bergmann's Rule—what's in a name? *Global Ecology and Biogeography*, 20(1), 203–207. <https://doi.org/10.1111/j.1466-8238.2010.00577.x>
- Meiri, S., & Dayan, T. (2003). On the validity of Bergmann's rule. *Journal of Biogeography*, 30(3), 331–351. <https://doi.org/10.1046/j.1365-2699.2003.00837.x>
- Mendenhall, V. M. (1997). *Preliminary report on the 1997 Alaska seabird die-off* (p. 3). U.S. Fish and Wildlife Service.
- Militão, T., Gómez-Díaz, E., Kaliontzopoulou, A., & González-Solís, J. (2014). Comparing multiple criteria for species identification in two recently diverged seabirds. *PLoS One*, 9(12), e115650. <https://doi.org/10.1371/journal.pone.0115650>
- 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. <https://doi.org/10.2307/1365299>

- 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. <https://doi.org/10.1111/j.1365-294X.2008.03700.x>
- 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. <https://doi.org/10.1111/ibi.12262>
- 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.
- Munilla, I., Genovart, M., Paiva, V. H., & Velando, A. (2016). Colony foundation in an oceanic seabird. *PLoS One*, 11(2), e0147222. <https://doi.org/10.1371/journal.pone.0147222>
- Navarro, J., Forero, M. G., González-Solís, J., Igual, J. M., Bécas, J., & Hobson, K. A. (2009). Foraging segregation between two closely related shearwaters breeding in sympatry. *Biology Letters*, 5, 545–548. <https://doi.org/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. <https://doi.org/10.1016/j.zool.2008.05.001>
- 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>
- O'Dea, A., Lessios, H. A., Coates, A. G., Eytan, R. I., Restrepo-Moreno, S. A., Cione, A. L., Collins, L. S., de Queiroz, A., Farris, D. W., Norris, R. D., Stallard, R. F., Woodburne, M. O., Aguilera, O., Aubry, M.-P., Berggren, W. A., Budd, A. F., Cozzuol, M. A., Coppard, S. E., Duque-Caro, H., ... Jackson, J. B. C. (2016). Formation of the isthmus of panama. *Science Advances*, 2(8), e1600883. <https://doi.org/10.1126/sciadv.1600883>
- Olson, S. L. (1985). The fossil record of birds. *Avian Biology*, 8, 79–252.
- 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. <https://doi.org/10.2988/08-11.1>
- 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. <https://doi.org/10.2988/09-19.1>
- 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., Isaac, N., & Pearse, W. (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. <https://doi.org/10.1080/106351599260184>
- Palumbi, S. R. (1994). Genetic divergence, reproductive isolation, and marine speciation. *Annual Review of Ecology and Systematics*, 25(1), 547–572. <https://doi.org/10.1146/annurev.es.25.110194.002555>
- 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. <https://doi.org/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. <https://doi.org/10.1071/MU01060>
- Péron, C., Weimerskirch, H., & Bost, C. A. (2012). Projected poleward shift of king penguins (*Aptenodytes patagonicus*) foraging range at the Crozet Islands, southern Indian Ocean. *Proceedings of the Royal Society B: Biological Sciences*, 279(1738), 2515–2523.
- Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One*, 7(5), e37135. <https://doi.org/10.1371/journal.pone.0037135>
- 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. <https://doi.org/10.1038/s41559-017-0223-6>
- 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. <https://doi.org/10.1093/sysbio/syy032>
- 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. <https://doi.org/10.2307/1370475>
- Ramos, R., Paiva, V. H., Zajková, Z., Precheur, C., Fagundes, A. I., Jodice, P. G. R., Mackin, W., Zino, F., Bretagnolle, V., & 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*, 191(2), 482–502. <https://doi.org/10.1093/zoolinnean/zlaa045>
- Rannala, B., & Yang, Z. (2007). Inferring speciation times under an episodic molecular clock. *Systematic Biology*, 56(3), 453–466. <https://doi.org/10.1080/10635150701420643>
- Rayner, M. J., Hauber, M. E., Steeves, T. E., Lawrence, H. A., Thompson, D. R., Sagar, P. M., Bury, S. J., Landers, T. J., Phillips, R. A., Ranjard, L., & Shaffer, S. A. (2011). Contemporary and historical separation of transequatorial migration between genetically distinct seabird populations. *Nature Communications*, 2, 332. <https://doi.org/10.1038/ncomms1330>
- 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. <https://doi.org/10.1080/10635150701883881>
- Revell, L. J. (2010). Phylogenetic signal and linear regression on species data: Phylogenetic regression. *Methods in Ecology and Evolution*, 1(4), 319–329. <https://doi.org/10.1111/j.2041-210X.2010.00044.x>
- Revell, L. J. (2012). phytools: An R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, 3(2), 217–223. <https://doi.org/10.1111/j.2041-210X.2011.00169.x>
- Rochette, N. C., Rivera-Colón, A. G., & Catchen, J. M. (2019). Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. *Molecular Ecology*, 28(21), 4737–4754. <https://doi.org/10.1111/mec.15253>

- Rodríguez, A., Arcos, J. M., Bretagnolle, V., Dias, M. P., Holmes, N. D., Louzao, M., Provencher, J., Raine, A. F., Ramírez, F., Rodríguez, B., Ronconi, R. A., Taylor, R. S., Bonnaud, E., Borrelle, S. B., Cortés, V., Descamps, S., Friesen, V. L., Genovart, M., Hedd, A., ... Chiaradia, A. (2019). Future directions in conservation research on petrels and shearwaters. *Frontiers in Marine Science*, 6, 94. <https://doi.org/10.3389/fmars.2019.00094>
- Ronquist, F. (1997). Dispersal-vicariance analysis: A new approach to the quantification of historical biogeography. *Systematic Biology*, 46(1), 195–203. <https://doi.org/10.1093/sysbio/46.1.195>
- 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. <https://doi.org/10.1111/j.1474-919X.2005.00483.x>
- Sangster, G., Knox, A. G., Helbig, A. J., & Parkin, D. T. (2002). Taxonomic recommendations for European birds. *Ibis*, 144(1), 153–159. <https://doi.org/10.1046/j.0019-1019.2001.00026.x>
- 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., Uy, J. A. C., Weir, J. T., Whittingham, L. A., & Safran, R. J. (2013). Sexual selection accelerates signal evolution during speciation in birds. *Proceedings of the Royal Society B: Biological Sciences*, 280(1766), 20131065. <https://doi.org/10.1098/rspb.2013.1065>
- Shaffer, S. A., Tremblay, Y., Weimerskirch, H., Scott, D., Thompson, D. R., Sagar, P. M., Moller, H., Taylor, G. A., Foley, D. G., Block, B. A., & 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. <https://doi.org/10.1073/pnas.0603715103>
- 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., & Matschner, 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. <https://doi.org/10.1093/sysbio/syy006>
- Storey, A. S., & Lien, J. (1985). Development of the first North American colony of Manx Shearwaters. *The Auk*, 395–401.
- 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. <https://doi.org/10.1073/pnas.1607921114>
- Sutton, P. (2001). Detailed structure of the Subtropical Front over Chatham Rise, east of New Zealand. *Journal of Geophysical Research*, 106(C12), 31045–31056. <https://doi.org/10.1029/2000JC000562>
- 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. <https://doi.org/10.1046/j.1365-2656.1998.00234.x>
- Vargas, P., & Zardoya, R. (2014). *The tree of life*. Sunderland.
- Vianna, J. A., Fernandes, F. A. N., Frugone, M. J., Figueiró, H. V., Pertierra, L. R., Noll, D., Bi, K. E., Wang-Claypool, C. Y., Lowther, A., Parker, P., Le Bohec, C., Bonadonna, F., Wienecke, B., Pistorius, P., Steinfurth, A., Burridge, C. P., Dantas, G. P. M., Poulin, E., Simison, W. B., ... 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. <https://doi.org/10.1073/pnas.2006659117>
- 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. <https://doi.org/10.1038/srep08853>
- 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.

BIOSKETCH

Joan Ferrer Obiol is a PhD candidate at the University of Barcelona. This work is part of his PhD thesis, which focuses on the use of genomic analyses to study the evolutionary history of shearwaters, an endangered group of pelagic seabirds. Joan is an evolutionary biologist using phylogenomics, population and comparative genomics to answer evolutionary questions using birds as model species. **Marta Riutort** and **Julio Rozas** are the supervisors of this PhD and are professors of genetics at the University of Barcelona. The main goal of the Riutort's laboratory research is to characterise the major events and processes creating and shaping the diversity and distribution of living organisms through the study of their genetic diversity (<http://www.ub.edu/geisan/>).

Author Contributions: R. Terry Chesser, Helen F. James, Jacob González-Solís, Vincent Bretagnolle, Andreanna J. Welch and Joan Ferrer Obiol contributed to data collection and all authors contributed to study design. Joan Ferrer Obiol performed the DNA extractions, processed and analysed the data, and wrote the manuscript with input from all authors. All authors read and approved the final manuscript.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Ferrer Obiol, J., James, H. F., Chesser, R. T., Bretagnolle, V., González-Solís, J., Rozas, J., Welch, A. J., & Riutort, M. (2022). Palaeoceanographic changes in the late Pliocene promoted rapid diversification in pelagic seabirds. *Journal of Biogeography*, 49, 171–188. <https://doi.org/10.1111/jbi.14291>