Molecular Phylogenetics and Evolution 179 (2023) 107671

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Species delimitation using genomic data to resolve taxonomic uncertainties in a speciation continuum of pelagic seabirds

Joan Ferrer Obiol [a,b,c,*], Jose M. Herranz [d,e], Josephine R. Paris [f,g], James R. Whiting [g,h], Julio Rozas [a,b], Marta Riutort [a,b], Jacob González-Solís [b,i]

[a] Departament de Genètica, Microbiologia i Estadística, Facultat de Biologia, Universitat de Barcelona (UB), Barcelona, Catalonia, Spain
[b] Institut de Recerca de la Biodiversitat (IRBio), Barcelona, Catalonia, Spain
[c] Department of Environmental Science and Policy, University of Milan, Milan, Italy
[d] National Institute for the Study of Liver and Gastrointestinal Diseases, CIBERehd, Carlos III Health Institute, Madrid, Spain
[e] Program of Hepatology, Center for Applied Medical Research (CIMA), University of Navarra, Pamplona, Spain
[f] Department of Health, Life and Environmental Sciences, University of l'Aquila, Coppito, Italy
[g] Department of Biosciences, University of Exeter, Exeter, UK
[h] Department of Biological Sciences, Faculty of Sciences, University of Calgary, Calgary, Canada
[i] Departament de Biologia Evolutiva, Ecologia i Ciències Ambientals, Facultat de Biologia, Universitat de Barcelona (UB), Barcelona, Catalonia, Spain

ARTICLE INFO

Keywords:
Population divergence
Taxonomy
Species delimitation
Management
Seabirds
Shearwaters

ABSTRACT

Speciation is a continuous and complex process shaped by the interaction of numerous evolutionary forces. Despite the continuous nature of the speciation process, the implementation of conservation policies relies on the delimitation of species and evolutionary significant units (ESUs). *Puffinus* shearwaters are globally distributed and threatened pelagic seabirds. Due to remarkable morphological status the group has been under intense taxonomic debate for the past three decades. Here, we use double digest Restriction-Site Associated DNA sequencing (ddRAD-Seq) to genotype species and subspecies of North Atlantic and Mediterranean *Puffinus* shearwaters across their entire geographical range. We assess the phylogenetic relationships and population structure among and within the group, evaluate species boundaries, and characterise the genomic landscape of divergence. We find that current taxonomies are not supported by genomic data and propose a more accurate taxonomy by integrating genomic information with other sources of evidence. Our results show that several taxon pairs are at different stages of a speciation continuum. Our study emphasises the potential of genomic data to resolve taxonomic uncertainties, which can help to focus management actions on relevant taxa, even if they do not necessarily coincide with the taxonomic rank of species.

1. Introduction

How populations diverge and become new species is one of the most fundamental questions in evolutionary biology. With the increasing availability of genome-wide data, we can now characterise genome-wide patterns of divergence and investigate the interplay of multiple evolutionary processes, such as gene flow, mutation, recombination, drift, and selection that together shape the genomic landscape. Understanding how these evolutionary processes interact and ultimately result in new species remains challenging (Nosil and Feder, 2012; Ravinet et al., 2017).

Despite the continuous nature of speciation, the implementation of efficient conservation policies relies on the delimitation of species and evolutionary significant units (ESUs, Crandall et al., 2000; Moritz, 2002). The general lineage concept (GLC), which considers species as separately evolving metapopulation lineages (De Queiroz, 2007) provides a good framework for statistical species delimitation. Within the GLC framework, the combination of high-resolution genome-wide data with the development of multispecies coalescent (MSC) delimitation approaches has emerged as a powerful approach to test different hypotheses of lineage divergence (Knowles and Carstens, 2007; Yang and Rannala, 2010) and alternative species delimitation hypotheses (Leaché et al., 2014). Such methods are being used in a growing number of studies to delimit species in a wide range of taxa (Abdelkrim et al., 2018; Ewart et al., 2020; Hosegood et al., 2020; Newton et al., 2020; Tonzo et al., 2019). However, the high resolution of genomic data makes it

* Corresponding author at: Departament de Genètica, Microbiologia i Estadística, Facultat de Biologia, Universitat de Barcelona, Barcelona, Catalonia, Spain.
E-mail address: joan.ferrer.obiol@gmail.com (J. Ferrer Obiol).

https://doi.org/10.1016/j.ympev.2022.107671
Received 5 August 2022; Received in revised form 28 October 2022; Accepted 17 November 2022
Available online 25 November 2022

difficult to distinguish population structure from species boundaries when using MSC methods (Chambers and Hillis, 2020; Sukumaran and Knowles, 2017). Introgression can further hinder species delimitation, especially in cases of limited geographical sampling (Chambers and Hillis, 2020; Chan et al., 2020). Appropriate geographical sampling, including contact zones among putative species, can overcome the issue of over-splitting caused by sampling limitations. Combining appropriate geographical sampling with other sources of evidence such as morphological, ecological or phenological data, provides a robust framework for species delimitation (Carstens et al., 2013; Chambers and Hillis, 2020). Furthermore, more recent implementations using unsupervised machine learning algorithms can avoid the issue of over-splitting attributed to MSC methods (Chambers and Hillis, 2020; Sukumaran and Knowles, 2017) and have demonstrated accurate species delimitation in several organisms (DeRaad et al., 2022; Derkarabetian et al., 2019; Newton et al., 2020).

One animal group with an urgent need of well-defined species and ESUs is the shearwaters, a globally distributed group of medium-sized pelagic seabirds. Over 50 % of shearwater species are listed as threatened by the IUCN (https://www.iucnredlist.org). Shearwaters face several anthropogenic threats, both at their breeding colonies and at sea (Croxall et al., 2012; Dias et al., 2019; Rodríguez et al., 2019). Inland, shearwater populations are severely affected by the introduction of invasive alien species, such as cats and rats, which predate on eggs, chicks and even adult birds (Holmes et al., 2019; Spatz et al., 2017). At sea, the main threat is fisheries bycatch (Bugoni et al., 2008; Cortés et al., 2017; Oppel et al., 2011), which could drive some species to extinction unless conservation measures are promptly implemented (Genovart et al., 2016; Oro et al., 2004).

However, resolving the evolutionary relationships among shearwaters has long posed a challenge (Austin, 1996; Austin et al., 2004). Species limits are controversial, mostly due to high morphological stasis in the group (Austin et al., 2004). A recent phylogenomic study showed that *Puffinus* shearwaters from the North Atlantic and Mediterranean constitute a monophyletic group that is divided into a clade of medium-sized taxa (*P. puffinus*, *P. mauretanicus*, *P. yelkouan*) and a clade of small-sized taxa (*P. lherminieri*, *P. baroli*, *P. boydi*) (Ferrer Obiol et al., 2021). However, the group is still under contentious ongoing taxonomic debate (Sangster et al., 2005; Olson, 2010; Genovart et al., 2012; Ramos et al., 2020; Rodríguez et al., 2020).

Taxa in the medium-sized group were originally considered to be conspecific and were placed together under *P. puffinus* (Mathews, 1934) until the end of the 1980 s. However, analysis of morphological data and vocalisations (Bourne et al., 1988; Bretagnolle, 1992) resulted in a split of the Mediterranean and North Atlantic lineages into two different species (*P. puffinus* and *P. yelkouan*). *P. yelkouan* included two subspecies (*mauretanicus* and *yelkouan*) that more recently, were elevated to species status based on morphological characters and reciprocal monophyly of cytochrome *b* (cyt b) sequences (Heidrich et al., 1998; Sangster et al., 2002). However, this split has not been unanimously integrated in bird taxonomies (i.e. Christidis, 2014). On the other hand, *P. puffinus* was considered a monotypic species. Recently, the Canary Islands populations have been described as a new subspecies (*P. p. canariensis*) based on multiple sources of evidence (Rodríguez et al., 2020). In addition, there is some uncertainty as to the taxonomic affinities of the Madeiran population of *P. puffinus* (Gil-Velasco et al., 2015; Rodríguez et al., 2020).

Small-sized shearwater species are under even more contentious taxonomic debate. Since Austin (1996) identified *lherminieri*, *baroli* and *boydi* to be a monophyletic group, the three taxa have been considered either as one, two or three different species (del Hoyo et al. 2014; Olson, 2010; Sangster et al., 2005) and ancestral introgression has been detected between *lherminieri* and *boydi* (Ferrer Obiol et al., 2021). Within *lherminieri*, populations breeding on islets off Panama and Northern Venezuela are on average smaller and have been named as a separate subspecies (*P. l. loyemilleri*; Wetmore, 1959). However, Austin et al. (2004) showed a lack of differentiation between *P. l. loyemilleri* and *P. l. lherminieri* based on cyt b sequences. Given all these uncertainties, and in order to develop effective conservation measures, there is an urgent need to robustly review the taxonomic status of the currently recognized eight taxa (*P. puffinus puffinus*, *P. p. canariensis*, *P. mauretanicus*, *P. yelkouan*, *P. baroli*, *P. boydi*, *P. lherminieri lherminieri*, *P. l. loyemilleri*) of North Atlantic and Mediterranean *Puffinus* shearwaters.

Shearwaters also provide a unique opportunity to study the process of speciation. Shearwaters are highly mobile seabirds that are highly philopatric (Coulson, 2016). In these island breeding species, philopatry is known to play a role in population differentiation (Friesen et al., 2007). This process is often intensified by differences in non-breeding and foraging distributions, commonly caused by the effects of winds and oceanic fronts (González-Solís et al., 2009; Weimerskirch et al., 2012). Such differences in non-breeding and foraging distributions can also result in differences in the time of arrival to the breeding colonies and ultimately lead to speciation due to allochrony (Rayner et al., 2011). Shearwaters are nocturnal burrowing species in their breeding colonies which rely on acoustics and olfaction to detect potential partners (Warham, 1996). The aforementioned processes often result in genetic differentiation between morphologically similar lineages as morphological changes are not expected to play a significant role in the evolution of reproductive isolation.

Here, we use double digest Restriction-Site Associated DNA sequencing (ddRAD-Seq) to: (a) quantify the genomic levels of variation among and within every species and subspecies of North Atlantic and Mediterranean *Puffinus* shearwaters, and evaluate their phylogenetic relationships; (b) explore the number of independently evolving lineages by applying multiple species delimitation approaches and; c) to use these results to integrate morphological, behavioural and ecological evidence, with the aim of evaluating and updating the taxonomy of the group. We use these results to inform on the conservation and management of the group.

2. Material and methods

2.1. Sampling, DNA extraction and ddRAD sequence data generation

We collected blood or tissue samples from a total of 42 individuals of the eight recognised North Atlantic and Mediterranean *Puffinus* shearwater taxa across their geographical ranges (Fig. 1, Table S1). We also included *Puffinus nativitatis* and *Calonectris borealis* as outgroups (Table S1). Data for 18 individuals of the ingroup taxa and the outgroups were previously generated by Ferrer Obiol et al. (2021). Genomic DNA extraction and ddRAD library construction for the rest of individuals were performed as described in Ferrer Obiol et al. (2021). Briefly, we used the Qiagen DNeasy Blood and Tissue Kit to extract genomic DNA according to the manufacturer's instructions (Qiagen GmbH, Hilden, Germany) and DNA extracts were sent to the Genomic Sequencing and Analysis Facility, University of Texas at Austin. ddRAD library preparation was performed following the Peterson et al. (2012) protocol using an uncommon cutter (*EcoRI*) and a common cutter (*MspI*). ddRAD libraries were sequenced on an Illumina HiSeq4000 platform using 150 bp paired-end (PE) sequencing.

2.2. ddRAD data processing

ddRAD data were processed using Stacks v2.41 (Rochette et al., 2019). Raw reads were quality-filtered and demultiplexed using *process_radtags*. Loci were built *de novo* using the forward reads with the *ustacks-cstacks-sstacks* core clustering algorithm with optimised parameters for shearwater data (as per Ferrer Obiol et al. 2021). Reverse reads were incorporated using *tsv2bam* and *gstacks* was used to assemble a contig for each locus, calling SNPs using the Bayesian genotype caller (BGC; Maruki and Lynch, 2015, 2017) and phasing resultant haplotypes.

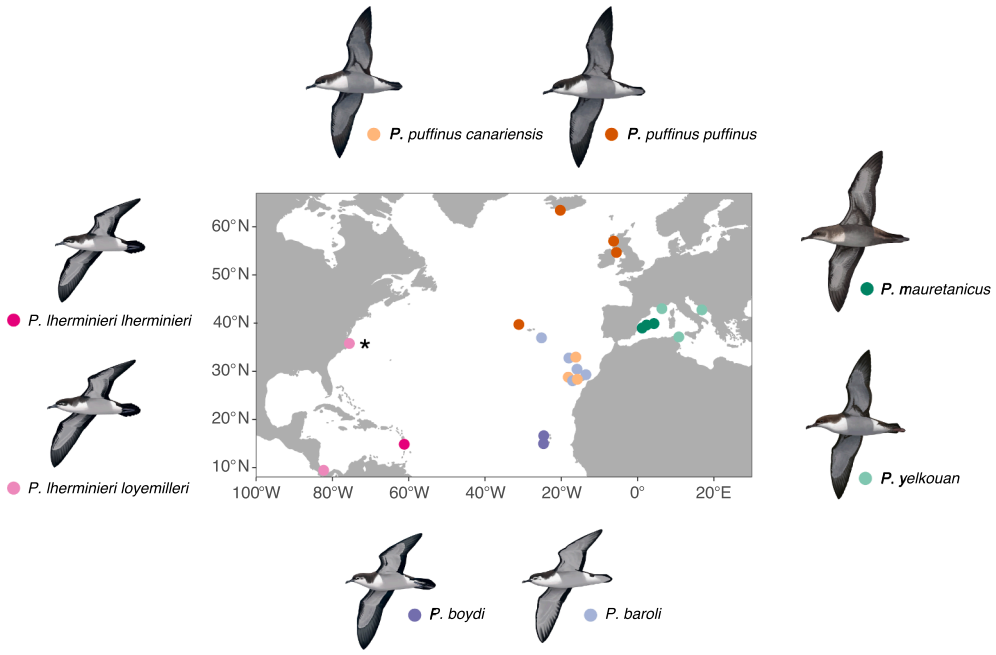

Fig. 1. Map showing sampling localities of the eight recognised taxa of North Atlantic and Mediterranean *Puffinus* shearwaters. Shearwaters were sampled at breeding colonies with the exception of the sampling site with an asterisk where stranded individuals were sampled. Illustrations by Martí Franch © represent *Puffinus* shearwater taxa included in this study. Colours represent different taxa and are consistent across all figures.

We mapped *gstacks* catalog loci to the Balearic shearwater genome assembly (Cuevas-Caballé et al., 2022) using BWA mem 0.7.17 (Li, 2013), we sorted them using SAMtools v.0.1.19 (Li et al., 2009) and we integrated alignment positions to the catalog using *stacks-integrate-alignments* (Paris et al., 2017). The *populations* module was used to export data in various formats for downstream analyses, requiring a minimum allele frequency (MAF) above 5 % and an observed heterozygosity below 50 % to process a SNP. For SNP-based analyses, we further filtered VCF files using VCFtools v.0.1.15 (Danecek et al., 2011) to include only biallelic SNPs and to mask genotypes if the per-sample read depth was $<$ 5 or $>$ 150, or if the genotype quality was $<$ 30. Table S2 includes the subsets of the total dataset that were used for each of the downstream analyses, which included a maximum of 16,339 loci and 141,767 SNPs.

2.3. Analysis of genomic variation among and within taxa

We described genomic diversity within and among taxa using several summary statistics. Per taxon nucleotide diversity (π), inbreeding coefficient (F_{IS}), the ratio of polymorphic SNPs, and pairwise F_{ST} between taxa were calculated using the Stacks2 *populations* program. We calculated the ratio of nonsynonymous to synonymous mutations following Perrier et al. (2017) (see Supplemental Information). Due to the low sample sizes for both subspecies of *P. lherminieri*, all summary statistics were calculated at the species level. To explore if patterns of genome-wide diversity relate to census size, we retrieved the number of breeding pairs from Birds of the World (Billerman et al. 2020) and BirdLife International (2020).

We studied the genomic variation among and within taxa using Principal Component Analysis (PCA) implemented in the R package adegenet (Jombart and Ahmed, 2011). We also performed a maximum-likelihood (ML) model-based clustering analysis to calculate individual ancestries using ADMIXTURE v.1.3.0 (Alexander and Lange, 2011). We tested K = 1–10 and the optimal K was determined using the lowest cross-validation errors estimates across 10 independent runs (Alexander et al., 2009; Evanno et al., 2005). Additional values of K were also assessed. To examine finer levels of genetic structure, hierarchical analyses were performed on individual clusters identified using the optimal K.

We used fineRADstructure v0.3.2 (Malinsky et al., 2018) to infer shared ancestry among all individuals. RADpainter was used to infer a coancestry matrix and the fineSTRUCTURE Monte Carlo Markov Chains (MCMC) clustering algorithm was used to assign individuals into clusters, running 100,000 MCMC iterations (following a burn-in period of 100,000 iterations) sampled every 1,000 generations. A tree of relationships based on the coancestry matrix was built in fineSTRUCTURE using default parameters. We used available R scripts (https://cichlid.gurdon.cam.ac.uk/fineRADstructure.html) to visualise the results. To detect finer-scale genetic structuring, we also performed fineRADstructure analyses for each of the three main groups detected by our phylogenetic and populations structure analyses (*P. puffinus*, *P. mauretanicus* - *P. yelkouan* and *P. lherminieri* - *P. baroli* - *P. boydi*).

To further infer geographic structuring and visualise genealogical patterns, we computed Neighbor-net phylogenetic networks (Bryant and Moulton, 2004), implemented in SplitsTree5 v.5.0.16 (Huson and Bryant, 2006).

2.4. Clustering-based species discovery analyses

We performed clustering-based species discovery analyses as outlined by Derkarabetian et al. (2019), which include both traditional clustering approaches and novel applications of unsupervised machine-learning (UML) algorithms. First, we used the R package adegenet to perform discriminant analysis of principal components (DAPC) (Jombart et al., 2010). We compared the goodness-of-fit for successive k-means clustering schemes based on the Bayesian Information Criterion (BIC).

We then used two UML algorithms to perform dimensionality reduction: Random Forest (RF; Breiman, 2001) and *t*-distributed Stochastic Neighbour Embedding (*t*-SNE; Van der Maaten and Hinton, 2008), followed by two unsupervised clustering methods: partitioning around medoids using the cluster R package (Maechler et al. 2022) and hierarchical agglomerative clustering using the mclust R package (Scrucca et al., 2016). A detailed description of these analyses is included in the Supplemental Information.

2.5. MSC species delimitation

To determine the number of independently evolving lineages, we applied two coalescent-based species delimitation approaches: BPP v.4.0 (Flouri et al., 2018) and BFD* (Leaché et al., 2014). BPP was run using option A11, which performs a joint comparison of species assignment and species tree models (Rannala and Yang, 2017; Yang and Rannala, 2014). To ensure computational tractability, we performed BPP analyses using two subsets of 500 loci, which has been shown to provide sufficient power for species delimitation (e.g. Tonzo et al., 2019). Subset 1 contained loci with at least four variable sites as such loci provide greater power in species delimitation (Huang, 2018). We also selected a random subset of loci to evaluate the effects of including less informative loci on species delimitation. We followed the approach of Huang and Knowles (2016) to test for the impact of different evolutionary and demographic scenarios by using different inverse-gamma distributed diffuse priors ($\alpha = 3$) for the population sizes (θ) and root ages (τ_0) (Table 2). Each analysis was run for 100,000 generations, sampling every 10 generations after a burnin of 100,000 generations.

We used BFD* (Leaché et al., 2014) with a matrix of 500 SNPs with no missing data, to rank ten competing species delimitation hypotheses (SDH) based on the five most popular world bird lists (IOC v.10.2: Gill et al., 2020; Clements v2019: Clements et al., 2019; HBW & Birdlife International: del Hoyo et al., 2014; Howard & Moore v.4.1: Christidis, 2014; Peters: Peters 1931), and also using the results from the genetic clustering and phylogenetic analyses performed here (Table 2). For each SDH, we conducted species tree estimation and calculated marginal likelihood estimates (MLE) using SNAPP v.1.4.2 (Bryant et al., 2012). For MLE calculation, we performed path sampling analyses with 40 steps for 100,000 iterations after a burnin of 12,000 iterations and setting alpha to 0.3. Every analysis was run twice using different seeds to assess consistency. Because the number of SNPs included in the analysis has the potential to impact model ranks when using BFD* (Leaché et al., 2014), we also performed additional analyses using 2000 SNPs with no missing data. To ensure computational tractability using this larger number of SNPs, we performed analyses separately for each of the three main groups detected by our phylogenetic and population structure analyses (see Supplemental Information). For both types of analyses, models were ranked by their MLE, and MLEs were compared using Bayes Factors (Kass and Raftery, 1995).

2.6. Phylogenetic analyses

To infer the phylogenetic relationships of the studied taxa and to evaluate the monophyly of clusters identified in the previous analyses, we estimated phylogenies based on concatenation and coalescent approaches using *C. borealis* and *P. nativitatis* as outgroups. For concatenation analyses, we used the MPI version of ExaBayes v.1.5 (Aberer et al., 2014) and raxml-ng v.0.6.0 (Kozlov et al., 2019) to estimate unpartitioned Bayesian and maximum-likelihood (ML) phylogenies, respectively. For ExaBayes, two independent runs with four coupled chains for 1,000,000 generations were performed and assessed for stationarity (effective sample sizes (ESS) > 300 for all model parameters) in Tracer v.1.7 (Rambaut et al., 2018). For raxml-ng, 50 ML tree searches were conducted with the GTR + G substitution model. Following the best tree search, we generated 500 non-parametric bootstrap replicates. To directly model incomplete lineage sorting (ILS), we also inferred species trees using two MSC methods: the Bayesian SNP-based SNAPP v.1.4.2 (Bryant et al., 2012) in BEAST v.2.5.0 (Bouckaert et al., 2019), and the summary method of ASTRAL-III (Zhang et al., 2018). For SNAPP, we used uninformative priors as we do not assume strong a priori knowledge about the parameters. Two replicates were run for 100,000 burn-in iterations, followed by 1,000,000 MCMC iterations. Tree and parameter estimates were sampled every 1000 MCMC iterations. Convergence and stationarity were confirmed (ESS > 300) using Tracer. For ASTRAL-III, we used RAxML v.8 (Stamatakis, 2014) to estimate gene trees for each ddRAD locus running 100 rapid bootstrap replicates followed by a thorough ML search. We then used ASTRAL-III to estimate a species tree from the best-scoring ML gene trees and bootstrap replicates.

To estimate divergence times, we applied the MSC approach of Stange et al. (2018) implemented in SNAPP. To avoid the inclusion of potentially introgressed individuals, we performed the analysis on two individuals from the most geographically distant populations per taxon. We performed the analysis excluding the outgroups and only used transitions to reduce rate heterogeneity. We followed the recommendations of Stange et al. (2018) in specifying an age constraint on the root of the ingroup as a normal distribution with a mean of 2.87 Mya and a standard deviation (SD) of 0.39. Mean and SD values were calculated to fit the posterior distribution for the ingroup root in Ferrer Obiol et al. (2022), which also used the Stange et al. (2018) approach with a molecular clock model calibrated with fossil constraints to estimate a time-calibrated species tree of all shearwaters. We conducted three replicate runs, each of 1,500,000 MCMC iterations after 100,000 burn-in iterations. More details on the phylogenetic analyses can be found in the Supplemental Information.

2.7. Characterisation of genomic landscapes of divergence

To assess the patterns of genetic diversity and differentiation across the genome of three taxon pairs with low differentiation and representing three different initial stages of differentiation (*P. mauretanicus* and *P. yelkouan*, *P. p. puffinus* and *P. p. canariensis*, and *P. boydi* and *P. baroli*), we calculated per locus π, D_{XY} and Weir and Cockerham F_{ST} using the R package PopGenome (Pfeifer et al., 2014). Because D_{XY} is associated with within-group diversity, we also calculated net divergence, D_a (Nei and Li, 1979), to capture only the differences that have accumulated since the taxa split. Finally, we constructed haplotype networks for the most highly differentiated loci in each taxon pair using the R package pegas (Paradis, 2010) to visualise relationships among taxa.

We used a liftover approach to map ddRAD loci to the Anna's Hummingbird (*Calypte anna*) chromosome-level genome assembly (diverged from shearwaters between 62.7 and 71.1 Ma (Jarvis et al., 2015)), which represents one of the best available genome assemblies within non-passerine Neoaves. Details of the liftover approach can be found in the Supplemental Information.

3. Results

3.1. Patterns of genome-wide diversity

π ranged from 0.00147 (*P. lherminieri*) to 0.00214 (*P. boydi*) and the proportion of polymorphic SNPs varied markedly from 28.4 % in *P. lherminieri* to 50.4 % in *P. puffinus puffinus*. Inbreeding (F_{IS}) was relatively low for most taxa, ranging from 0.0808 (*P. baroli*) to 0.1506 (*P. puffinus puffinus*). Among a mean of 5542 (range = 3611–6397) polymorphic loci per taxon, an average of 115 (85–128) polymorphic loci had significant BLAST hits to *P. mauretanicus* proteins (Table S3). We identified 59–108 synonymous (mean = 91) and 19–31 (mean = 24) non-synonymous mutations per taxa. Per-locus π distributions only varied slightly across taxa (Figure S1), with the exception of *P. lherminieri*, which showed a much higher proportion of low π values. Accordingly, *P. lherminieri* had one of the highest F_{IS} values, the lowest ratio of polymorphic SNPs, and the highest ratio of non-synonymous to synonymous mutations amongst all taxa (Table 1). This suggests a reduction of diversity in this species despite the relatively high number of breeding pairs. Indeed, the number of breeding pairs did not appear to have a strong effect on genome-wide levels of genetic diversity (Table 1). For instance, we found relatively high levels of global π in taxa with low census size (i.e. *P. boydi*). On the other hand, recently diverged sister taxa (*P. mauretanicus* and *P. yelkouan*, *P. p. puffinus* and *P. p. canariensis*,

Table 1
Genetic characteristics and number of breeding pairs for each taxon. Global π, inbreeding coefficient (F_{IS}) and the ratio of polymorphic SNPs are reported for each taxon. The ratio of non-synonymous to synonymous mutations was calculated based on significant hits from a BLAST query of polymorphic loci against the *P. mauretanicus* annotated proteins. The number of breeding pairs was retrieved from Birds of the World (Billerman et al. 2020) and BirdLife International (2020). Summary statistics for *P. lherminieri* were calculated at the species level, due to the low sample sizes for both subspecies of *P. lherminieri*.

Taxon	Number of breeding pairs	Global π	F_{IS}	Ratio of polymorphic SNPs	Ratio of nonsynonymous mutations
P. baroli (n = 9)	3,360	0.00178	0.0808	0.483	0.326
P. boydi (n = 4)	5,000	0.00214	0.1105	0.446	0.255
P. lherminieri (n = 5)	15,700	0.00147	0.1308	0.284	0.441
P. mauretanicus (n = 6)	3,142	0.00186	0.1011	0.431	0.241
P. yelkouan (n = 6)	22,928	0.00184	0.0887	0.428	0.211
P. puffinus puffinus (n = 7)	399,500	0.00212	0.1506	0.504	0.176
P. puffinus canariensis (n = 5)	800	0.00210	0.1411	0.478	0.268

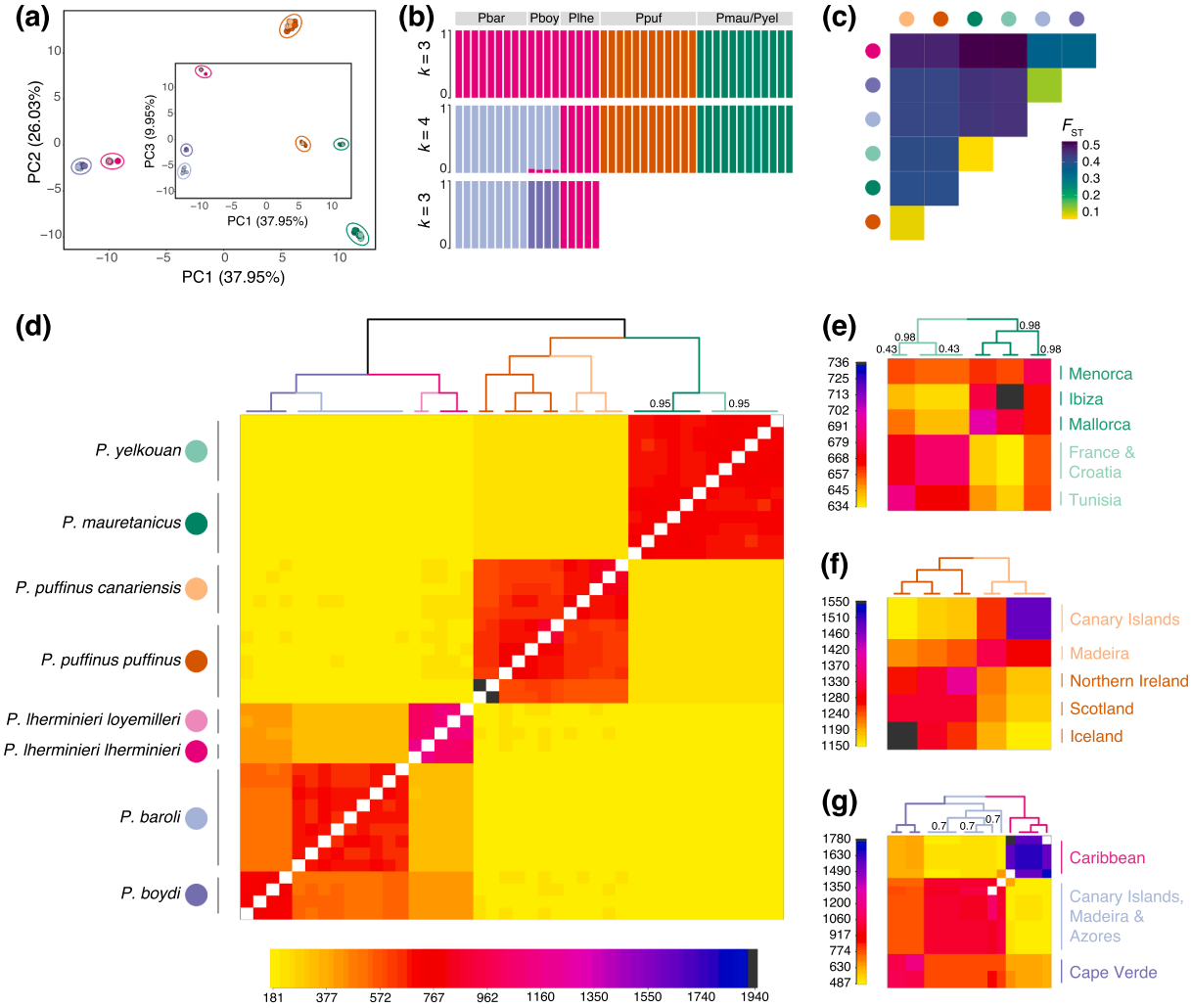

Fig. 2. Population structure of the eight recognised taxa of North Atlantic and Mediterranean *Puffinus* shearwaters. (a) Principal Components Analysis (PCA) showing PC1 (38 % variance) and PC2 (26 % variance). Inset shows PC1 versus PC3 (10 % of variance). (b) ADMIXTURE results for $K = 3$ and $K = 4$ which had the lowest cross-validation error and results for $K = 3$ for the small-sized taxa only. Facet labels above the plots represent: *P. baroli* (Pbar), *P. boydi* (Pboy), *P. lherminieri* (Plhe), *P. puffinus* (Ppuf), *P. mauretanicus* (Pmau) and *P. yelkouan* (Pyel). (c) Heatmap of pairwise F_{ST} estimates between *Puffinus* shearwater taxa. (d-g) Patterns of shared coancestry inferred from fineRADstructure. Each panel represents a heatmap showing coancestry coefficients between shearwater samples. Coancestry coefficients are colour coded from low (yellow) to high (blue-black) corresponding to the values in the legend. Atop each heatmap is a fineRADstructure clustering dendrogram based on the matrix of coancestry coefficients with branches coloured by taxon following the same colour code used next to the taxon labels on the left of the panel. Branch supports are shown for branches with posterior probabilities < 1. (d) Coancestry coefficients among all samples. (e) Average coancestry coefficients among all samples of *P. mauretanicus* and *P. yelkouan*, (f) *P. puffinus*, and (g) *P. lherminieri*, *P. boydi* and *P. baroli*. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

and *P. boydi* and *P. baroli*) had very similar estimates for most of the within-taxon diversity statistics, showing a strong phylogenetic signal and suggesting that despite strong philopatry, gene flow could be tempering a potential loss of genetic diversity. However, for each taxon pair, the ratio of non-synonymous to synonymous mutations was always higher for the taxon with the lowest census sizes (Table 1), suggesting that in these taxa selection has become less effective due to a higher incidence of genetic drift.

3.2. Clustering-based species discovery

PCA showed a clear separation between small-sized taxa and medium-sized taxa along PC1 (37.95 % of the variance) (Fig. 2a) and a further subdivision of each group into two (*P. puffinus* from *P. mauretanicus* and *P. yelkouan*, and *P. lherminieri* from *P. boydi* and *P. baroli*). PC2 (26.03 %) further separated the medium-sized *P. puffinus* from *P. mauretanicus* and *P. yelkouan* highlighting higher differentiation among the medium-sized taxa compared to the small-sized taxa. PC3 (9.95 %) further separated the three small-sized taxa into three different groups. The variance explained by PCs decreased abruptly after PC3 (Figure S2). ADMIXTURE identified $K = 4$ as the optimum number of clusters (Fig. 2b; *P. lherminieri*, *P. boydi-P. baroli*, *P. puffinus*, *P. mauretanicus-P. yelkouan*), although the cross-validation error in ADMIXTURE was lowest for $K = 3$–5. Increasing K to 5 did not provide additional interpretable resolution; however, analysing the small-sized taxa separately resulted in a complete discrimination of the three species (Fig. 2b).

Overall, fineRADstructure analyses, which emphasises recent coancestry, recovered the same three main groups recovered by ADMIXTURE analysis with $K = 3$. Additionally, fineRADstructure detected finer-scale genetic structure within each of the three main groups (Fig. 2e,f,g). fineRADstructure showed an incipient separation between *P. mauretanicus* and *P. yelkouan*, and also showed that *P. mauretanicus* from Menorca share higher levels of recent coancestry with individuals of *P. yelkouan* (Fig. 2e). Within *P. puffinus*, each sampling locality appeared as a distinct cluster, with the first division separating the individuals from the Canary Islands and Madeira from the North Atlantic populations and the Azores. Coancestry values in this species appeared to follow a pattern of isolation-by-distance (Fig. 2f). Within the small-sized species group, fineRADstructure clearly showed a complete discrimination of the three species (Fig. 2d and g; *P. baroli*, *P. boydi* and *P. lherminieri*). Finally, fineRADstructure recovered the two subspecies of *P. lherminieri* either as distinct groups or *P. l. lherminieri* as paraphyletic.

Our clustering-based species discovery analyses resulted in two SDH, supporting $K = 4$ and $K = 5$ species, respectively. Among the six different analyses, three supported the SDH with four species and three supported the SDH with five species (Fig. 3). Sample assignments for the SDH with $K = 4$ species were identical to sample assignments for ADMIXTURE with $K = 4$, and the SDH with $K = 5$ further separated *P. boydi* and *P. baroli* as two different species. Clustering analyses did not distinguish the Mediterranean species *P. mauretanicus* and *P. yelkouan* or the subspecies of *P. puffinus* and *P. lherminieri*. Taxa that were not found to be distinct using these analyses had low pairwise F_{ST} values ($F_{ST} < 0.12$, Fig. 2c).

3.3. MSC species delimitation

MSC species delimitation analyses using BPP consistently supported a SDH with five species with sample assignments identical to the clustering-based SDH with $K = 5$ (Fig. 3). The results of the analysis were largely robust to both the subsets of loci and the prior combinations used (Table 2).

The 500 SNPs BFD* analyses showed the strongest support for a four species model (H6; Table 3), identical to the clustering-based SDH with $K = 4$ (Fig. 3). The SDH based on a current taxonomy that received the highest support was the Howard & Moore World Bird List (v.4.1: Christidis, 2014) (H3, 2lnBF = 8.4). Increasing the number of SNPs and performing the analyses by group had a significant impact on model ranks with a tendency towards inferring more splits (Table S4; Fig. 3). This is probably due to the higher resolution provided by more sequence data and showcases the over-splitting issue attributed to many coalescent-based species delimitation approaches.

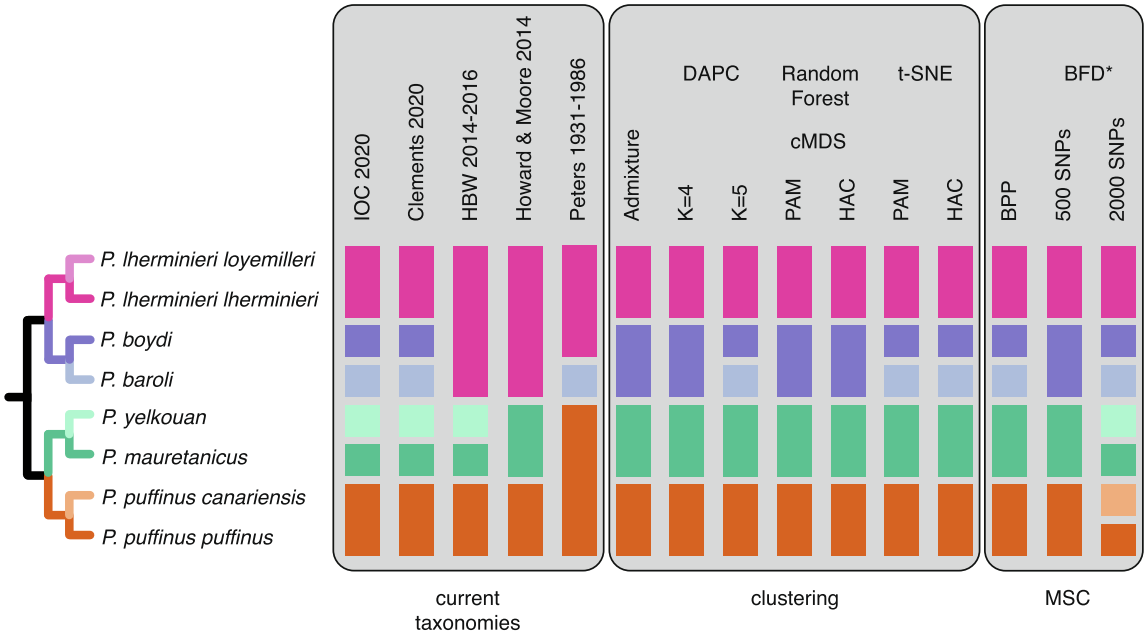

Fig. 3. Proposed taxonomies of North Atlantic and Mediterranean *Puffinus* shearwaters and species delimitation results from clustering-based and multispecies coalescent (MSC) approaches. The species tree depicted on the left is a schematic representation summarising all results from the phylogenetic analyses among the studied taxa.

Table 2
BPP species delimitation analysis results for each subset of loci (random: minimum 1 SNP per locus and informative: minimum 4 SNPs per locus) and different combinations of population size (θ) and root age (τ_0) priors. We report the number of species and the species inferred by each analysis, and the posterior probability of the number of species.

Min. num. of SNPs per locus	Population size prior (θ)	Root age prior (τ_0)	Number of Species	Species	Posterior probability
1	IG(3, 0.002)	IG(3, 0.003)	5	*baroli, boydi, lherminieri, mauretanicus/yelkouan, puffinus*	0.86
1	IG(3, 0.002)	IG(3, 0.03)	5	*baroli, boydi, lherminieri, mauretanicus/yelkouan, puffinus*	0.89
1	IG(3, 0.02)	IG(3, 0.003)	5	*baroli, boydi, lherminieri, mauretanicus/yelkouan, puffinus*	0.96
1	IG(3, 0.02)	IG(3, 0.03)	5	*baroli, boydi, lherminieri, mauretanicus/yelkouan, puffinus*	0.96
4	IG(3, 0.002)	IG(3, 0.003)	6	*baroli, boydi, lherminieri, mauretanicus/yelkouan, puffinus, puffinus* Ireland	0.47
4	IG(3, 0.002)	IG(3, 0.03)	6	*baroli, boydi, lherminieri, mauretanicus/yelkouan, puffinus, puffinus* Ireland	0.65
4	IG(3, 0.02)	IG(3, 0.003)	5	*baroli, boydi, lherminieri, mauretanicus/yelkouan, puffinus*	0.98
4	IG(3, 0.02)	IG(3, 0.03)	5	*baroli, boydi, lherminieri, mauretanicus/yelkouan, puffinus*	0.98

3.4. Phylogenetic analyses

Phylogenetic analyses recovered the five clusters identified by clustering-based and BPP analyses as monophyletic groups (Fig. 4). Phylogenetic trees recovered the same topology as in a previous shearwater phylogenomic study by Ferrer Obiol et al. (2021), and coalescent-based analyses confidently resolved the short internode separating the small-sized and the medium-sized taxa (node 8 in Fig. 4c). Neighbour-net networks and SNAPP analyses showed high levels of reticulation within and among *P. mauretanicus* and *P. yelkouan*, within *P. puffinus* and within *P. lherminieri* suggesting the presence of gene flow (Fig. 4a and b). Accordingly, phylogenetic analyses did not recover *P. mauretanicus*, *P. yelkouan*, and *P. puffinus* and *P. lherminieri* subspecies as monophyletic groups (Fig. 4c). Moreover, divergence time estimates between the two *P. lherminieri* subspecies and between *P. mauretanicus* and *P. yelkouan* included the present in the 95 % HPD intervals (Fig. 4d), suggesting that these taxa have not yet fully diverged. On the other hand, the split between *P. p canariensis* and *P. p. puffinus* was inferred during the Last Glacial Period.

3.5. Genome-wide differentiation in three recently diverged taxon pairs

Per-locus π densities completely overlapped between each of the three taxon pairs showing low differentiation, with the exception of the *P. baroli* and *P. boydi* pair, where *P. baroli* had a higher proportion of loci with low π compared to *P. boydi* (Fig. 5a). The variation in F_{ST} between the three recently diverged taxon pairs highlighted that the pairs represent three different stages of differentiation (Fig. 5b). Differentiation between *P. mauretanicus* and *P. yelkouan* was the lowest (mean F_{ST} = 0.04, 99th percentile = 0.32), followed by *P. p. puffinus* and *P. p. canariensis* (mean F_{ST} = 0.06, 99th percentile = 0.44), and by *P. boydi* and *P. baroli* (mean F_{ST} = 0.1, 99th percentile = 0.65). Across the genome, pairwise F_{ST} showed only a few regions of high differentiation. In two of the three pairwise taxon comparisons, regions of high differentiation were particularly concentrated on the Z chromosome (Figure S3), likely due to higher drift because of the Z chromosome's smaller effective population size. There was little overlap in differentiation peaks among different pairs (Figure S4) and the number of observed overlaps did not significantly differ from random expectations (Table S5). Variation in net divergence (D_a) showed a similar pattern to F_{ST} (Fig. 5c), with mean values of 0.14 % (*P. mauretanicus* versus *P. yelkouan*), 0.26 % (*P. p. puffinus* versus *P. p. canariensis*) and 0.69 % (*P. boydi* versus *P. baroli*), supporting the idea that these three pairs represent different stages of differentiation and not differences in the amount of within population diversity (π_{within}). In agreement with the low levels of differentiation within each of the three taxon pairs, the majority of the genome showed π_{within} only slightly lower than D_{XY}, with most loci clustered along the 1:1 line (the expectation under panmixia), despite marked heterogeneity in both π_{within} and D_{XY} (Fig. 5d). However, the coefficient of determination (R^2) of the regression between π_{within} and D_{XY} also reflected three different levels of differentiation. As expected by the observed levels of differentiation, the comparison between *P. mauretanicus* and *P. yelkouan* had the highest coefficient of determination ($R^2 = 0.94$) and the comparison between *P. boydi* and *P. baroli* the lowest ($R^2 = 0.74$).

Haplotype networks of the four most differentiated loci for the *P. boydi* and *P. baroli* comparison showed species diagnostic haplotypes. On the other hand, haplotype networks of the most differentiated loci for the *P. mauretanicus* and *P. yelkouan,* and the *P. p. puffinus* and *P. p. canariensis* comparisons only showed allele frequency differences (Figure S5) and no fixed differences.

4. Discussion

4.1. North Atlantic and Mediterranean puffinus shearwaters: How many species?

North Atlantic and Mediterranean *Puffinus* shearwaters have recently been identified as a monophyletic group using phylogenomic data (Ferrer Obiol et al. 2021). This recent finding provided us with an opportunity to delve further into the population structure and species delimitation of a group that is under highly contentious ongoing taxonomic debate (Genovart et al., 2012; Olson, 2010; Ramos et al., 2020; Rodríguez et al., 2020; Sangster et al., 2005). Moreover, current world bird lists (Table 3; Fig. 3) disagree about the number of North Atlantic and Mediterranean *Puffinus* shearwater species. Our species delimitation analyses found no support for any of the previously proposed taxonomies for the group. Taking our present results together with a recent phylogenomic study (Ferrer Obiol et al., 2022), and multiple additional lines of evidence (Flood and van der Vliet, 2019; Genovart et al., 2012; Gil-Velasco et al., 2015; Militão et al., 2014; Ramos et al., 2020; Rodríguez et al., 2020) under an integrative taxonomic framework, we recommend a more accurate taxonomy.

We base our taxonomy on defining a species when at least half of the nine species delimitation methods agree, and when additional evidence supports the species status. We define a subspecies when it constitutes an evolutionary significant unit (ESU) below the species level; when we find no agreement between species delimitation methods (less than half of the methods support the species), but when we do find evidence of genetic differentiation, and additional evidence for morphological or ecological distinctiveness. On this basis, we propose the following taxonomy including suggested common names:

Genus *Puffinus* Brisson, 1760.

Species *Puffinus lherminieri* Lesson, 1839, Audubon's Shearwater

Species *Puffinus baroli* Bonaparte, 1856, North Atlantic Little Shearwater

Table 3
BFD* analysis results for competing species delimitation hypothesis (SDH) based on five of the most popular world bird lists (H1 - H4), our genetic clustering and BPP analyses (H6 - H7) and other proposed taxonomic proposals (H5, H8 - *H*10). For each SDH, the number of species, marginal likelihood estimates (MLE), Bayes factors (2 × lnBF) and its rank are shown.

Species delimitation hypothesis (SDH)	Species	Num. species	Rank	MLE	2lnBF
H1: IOC 2020 & Clements 2020	*baroli, boydi, lherminieri, mauretanicus, yelkouan, puffinus*	6	8	−23,729.2	128.8
H2: HBW 2014-2016	*baroli-boydi-lherminieri, mauretanicus, yelkouan, puffinus*	4	5	−23,700.9	72.2
H3: Howard & Moore 2014	*baroli-boydi-lherminieri, mauretanicus-yelkouan, puffinus*	3	2	−23,669.0	8.4
H4: Peters 1931-1986	*baroli (assimilis), boydi-lherminieri, mauretanicus-yelkouan-puffinus*	3	10	−24,107.0	884.4
H5: All taxa	*baroli, boydi, l. lherminieri, l. loyemilleri mauretanicus, yelkouan, p. puffinus, p. canariensis*	8	9	−23,802.9	276.2
H6: ADMIXTURE & DAPC $K = 4$	*baroli-boydi, lherminieri, mauretanicus-yelkouan, puffinus*	4	1	−23,664.8	–
H7: BPP	*baroli, boydi, lherminieri, mauretanicus-yelkouan, puffinus*	5	3	−23,698.8	68.0
H8: Reassign Menorca	*baroli-boydi, lherminieri, mauretanicus, yelkouan* (incl. Menorca), *puffinus*	5	4	−23,700.7	71.8
H9: Split *P. puffinus*	*baroli-boydi, lherminieri, mauretanicus-yelkouan, p. puffinus, p. canariensis*	5	6	−23,701.4	73.2
H10: Split *P. puffinus* & reassign Madeira	*baroli-boydi, lherminieri, mauretanicus-yelkouan, p. puffinus* (incl. Madeira), *p. canariensis*	5	7	−23,703.3	77.0

Species *Puffinus boydi* Mathews, 1912, Cape Verde Little Shearwater
Species *Puffinus puffinus* Brünnich, 1764, Manx Shearwater
Subspecies *puffinus* Brünnich, 1764
Subspecies *canariensis* Rodríguez et al, 2020
Species *Puffinus yelkouan* Acerbi, 1827, Mediterranean Shearwater
Subspecies *yelkouan* Acerbi, 1827, Yelkouan Shearwater
Subspecies *mauretanicus* Lowe, 1921, Balearic Shearwater

Below we discuss the consideration of each taxon on a case-by-case basis.

Firstly, we find no support for the split of *mauretanicus* and *yelkouan* into two different species. Genetic clustering analyses and species delimitation analyses did not recover two distinct groups (Fig. 3), phylogenetic analyses failed to recover reciprocal monophyly between the two taxa (Fig. 4c), and coalescent-based divergence time estimation included the present time in the 95 % HPD (Fig. 4d), suggesting that *mauretanicus* and *yelkouan* have not definitively diverged. Pairwise F_{ST} was extremely low (F_{ST} = 0.04) and we found no fixed differences (species-diagnostic SNPs) between them. Moreover, *mauretanicus* and *yelkouan* were the least differentiated pair amongst the ones examined, showing an F_{ST} density curve that was the most skewed towards low values. These results, together with a gradient of phenotypes(Genovart et al., 2012; Militão et al., 2014), partially overlapping non-breeding distributions (Austin et al., 2019), nearly indistinguishable vocalisations (Yésou et al., 1990), and a lack of correspondence at the individual level between phenotypic characters, stable isotope analyses, microsatellites and mtDNA (Genovart et al., 2012; Militão et al., 2014), lead us to propose that the two Mediterranean taxa should be considered as conspecific. However, fine-scale population structure analysis based on recent coancestry was able to separate *yelkouan* and *mauretanicus* into two distinct groups with finer-scale structure at the population level, especially in *mauretanicus* (Fig. 2e).

Our analyses also suggest that *mauretanicus* and *yelkouan* may be at a very initial stage of speciation, which is in contrast with a previous hypothesis suggesting a scenario of admixture between two well-differentiated species based on deeply divergent mitochondrial haplotypes (Genovart et al., 2012, 2005). Such deep mtDNA divergences are commonly found within species (Bernardo et al., 2019; Morales et al., 2015), and mito-nuclear discordance has been attributed to multiple mechanisms including adaptive introgression of mtDNA, demographic disparities, sex-biased asymmetries, or as a result of differences in effective population size between mitochondrial and nuclear regions (Toews and Brelsford, 2012). In some cases, epistatic interactions between the nuclear genome and mitochondrial haplotypes can form the basis of reproductive incompatibilities (Sloan et al., 2017). Exploring the potential drivers of mito-nuclear discordance in *mauretanicus* and *yelkouan* represents an ongoing area of research. Taken together, these analyses lead us to propose that these two taxa should be defined as two separate ESUs. We hypothesise that differentiation may be occurring due to different migratory strategies and associated changes in breeding phenology (Austin et al., 2019), which appears to be a common mode of population differentiation in Procellariiformes (Rayner et al., 2011). Our analyses also showed that individuals from Menorca, which has been previously described as a hybrid population (Genovart et al., 2012, 2005), seem to be somewhat intermediate between *mauretanicus* from the other Balearic Islands and *yelkouan*. Future ongoing research using whole-genome sequencing should clarify whether this population should be included within the *mauretanicus* ESU or the *yelkouan* ESU.

In agreement with Rodríguez et al. (2020), our species delimitation analyses did not generally support an upgrade of the two *P. puffinus* subspecies into separate species. However, fineRADstructure analyses showed low coancestry between Canary Islands individuals and northern populations. In addition, the F_{ST} density curves showed a slightly higher differentiation between *P. p. puffinus* and *P. p. canariensis* than between *mauretanicus* and *yelkouan*. Individuals from Madeira showed higher levels of coancestry with Canary Islands individuals than with northern populations suggesting that they should belong to *P. p. canariensis*. Our dating analyses showed that the Canary Islands populations diverged from its northern counterparts during the last glacial period (Fig. 4d). These analyses, together with morphological differences, support the need to consider *P. p. canariensis* as an independent ESU from *P. p. puffinus*.

In the small-sized species group, our phylogenetic and clustering analyses were able to recover each of the three taxa as monophyletic/distinct groups, and five of the nine species delimitation approaches supported a scenario with three species. Divergence dating analyses placed the split between the West Atlantic clade (*lherminieri*) and the East Atlantic clade (*boydi* and *baroli*) at ~ 1 Mya (Fig. 4d), and the divergence between *boydi* and *baroli* in the late Pliocene (~120,000 year ago), which is considerably more recent than has been previously proposed (at least 400,000 years ago (Olson, 2010)). Despite their shared ecological plasticity (Ramos et al., 2020) and high overlap in morphological characters (Flood and van der Vliet, 2019), *boydi* and *baroli* can be easily identified using a combination of plumage characteristics and flight behaviour (Flood and van der Vliet, 2019). Thus, among the pairs analysed that represent different stages of divergence in the speciation continuum, *boydi* and *baroli* constitute the only case that justifies separate species status. We also found evidence of fine-scale population structure within *P. lherminieri* suggesting that ESUs should be defined below the species level. However, our sample sizes are too low and too sparse to be able to draw strong conclusions here. Future phylogeographic studies are required in order to properly assess population structure in this species, which has suffered a 95 % reduction in population size since the arrival of humans in the Caribbean (Mackin, 2016).

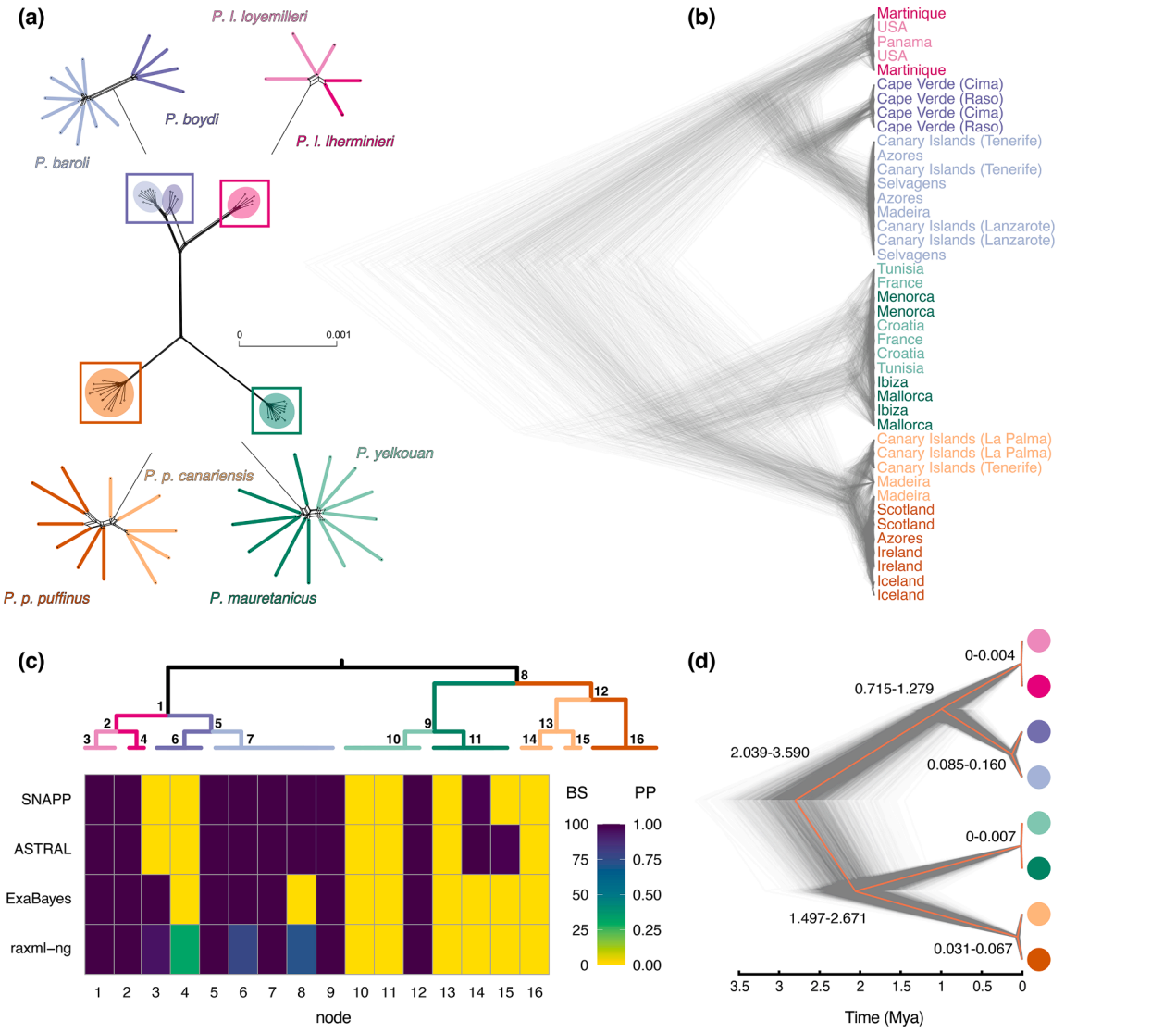

Fig. 4. Phylogenetic analyses of *Puffinus* shearwaters using 15,525 ddRAD loci. (a) Neighbour-net network. Squares represent regions of the network that are shown in more detail adjacently. Note that reticulation denotes non-tree-like areas. (b) Cloudogram of SNAPP trees from the posterior tree distribution showing topological and branch length variation. Tip labels represent sampling localities. (c) Heatmap summarising phylogenetic analyses using different methods. The nodes shown in the heatmap are those from the fineRADtructure dendrogram (shown above). Within *P. p. canariensis*, node 14 represents a monophyletic group containing all Canary Islands individuals and node 15 a monophyletic group containing all Madeiran individuals. Bootstrap support values or posterior probabilities are colour-coded as represented in the legend. (d) Time-calibrated SNAPP species tree (5403 transition sites). Individual trees shown in grey are samples from the posterior tree distribution and a maximum-clade-credibility summary tree is shown in orange. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4.2. Conservation implications

Conservation and management decisions rely on the classification of diversity into species and ESUs. Procellariiform species tend to function as metapopulations, with several populations representing independent ESUs (Friesen et al., 2007; Rexer-Huber et al., 2019; Taylor et al., 2019). Significant population declines within an ESU, or the complete loss of an ESU, can significantly reduce the overall genetic diversity of a species, thus reducing its ability to cope with environmental perturbations (Cristofari et al., 2019; Friesen et al., 2007). In such cases, conservation and management of ESUs should be a priority (Funk et al., 2012; Palsbøll et al., 2007). This should be a major consideration for *P. puffinus*, for which the Canary Islands and Madeira populations could be at threat due to their low census sizes (Table 1). These populations harbour unique genetic diversity and a targeted conservation plan integrating evidence derived from these genetic data together with previous phenological, morphological, acoustic and mtDNA data should be developed to ensure preservation of these populations (Rodríguez et al., 2020).

The proposed lumping of *mauretanicus* and *yelkouan* based on low genetic differentiation does not preclude focussed conservation efforts on these two ESUs. Indeed, the detection of fine-scale population structure should be integrated as new evidence for future conservation plans. Currently, *mauretanicus* and *yelkouan* are catalogued as Critically Endangered and Vulnerable, respectively, and are severely affected by longline fisheries bycatch in the Mediterranean (Cortés et al., 2017; Genovart et al., 2016; Oppel et al., 2011). Identifying the origins of seabirds affected by bycatch is vital to identify the populations most severely affected by fishing practices. A previous integrative approach (Militão et al., 2014) was able to correctly identify 96 % individuals to *mauretanicus* or *yelkouan* but lacked the resolution required to assign individuals to populations. However, our genomic dataset provided resolution at the population level (Fig. 4b) and therefore opens up the possibility to develop a management-relevant genetic assay to

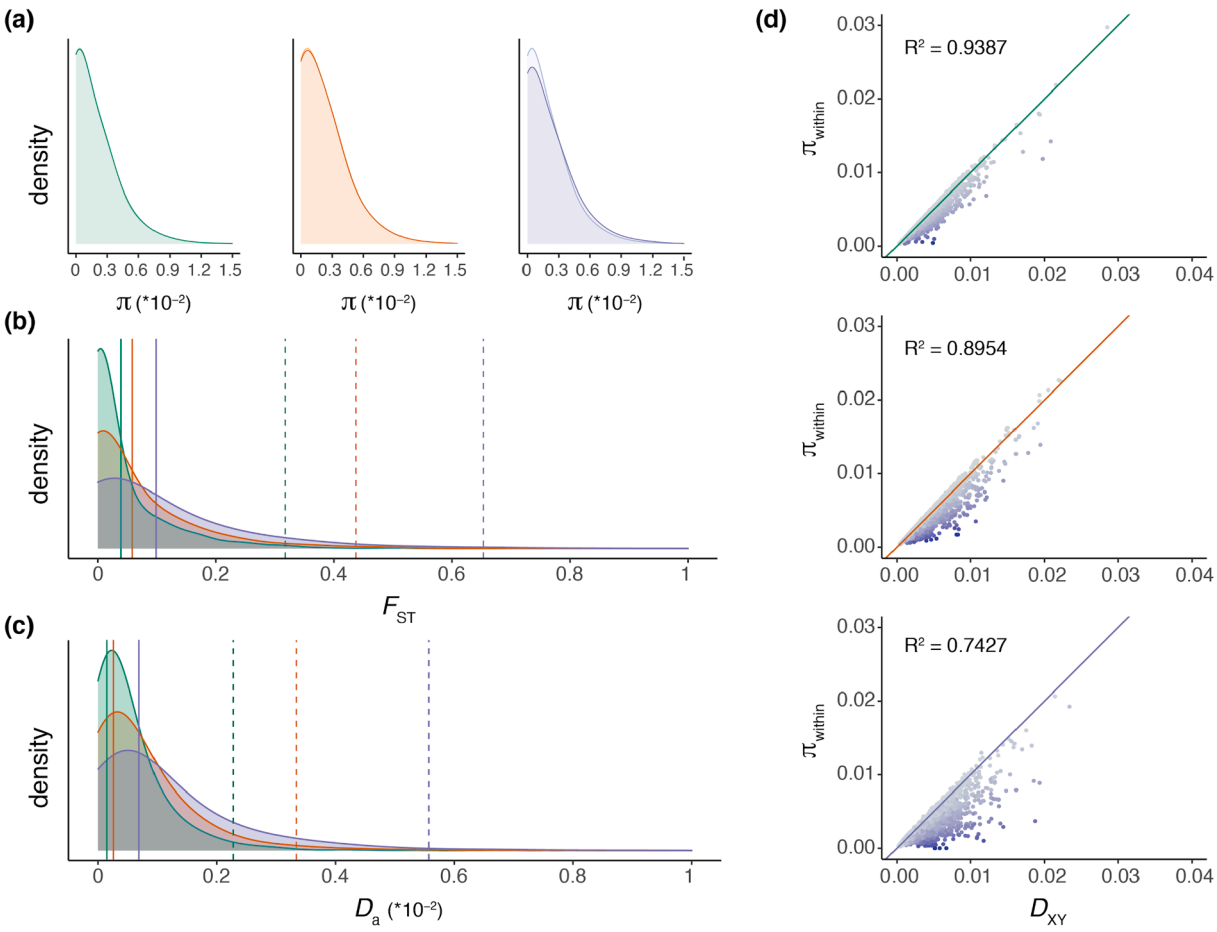

Fig. 5. Genome-wide differentiation in three recently diverged taxon pairs. (a) Smoothed density distributions of per-locus π estimates for both taxa in each taxon pair. From left to right: *P. mauretanicus* (green) and *P. yelkouan* (light green), *P. p. puffinus* (light orange) and *P. p. canariensis* (orange), *P. boydi* (purple) and *P. baroli* (light purple). Smoothed distributions of per-locus F_{ST} (b) and D_a (c) estimates for each taxon pair: *P. mauretanicus* versus *P. yelkouan* (green), *P. p. puffinus* versus *P. p. canariensis* (orange), *P. boydi* versus *P. baroli* (purple). Vertical continuous and dashed lines indicate mean and 99% percentiles, respectively. (d) Relationship between D_{XY} and within-taxon nucleotide diversity (π_{within}). High F_{ST} loci (represented by increasing blue colour) are characterised by reduced π_{within} compared to D_{XY}. Taxon pair colour coding for panels (c) and (d) is identical as in panel (b). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

determine the origin of shearwaters that die from fisheries bycatch, using a selection of the most informative SNPs. Such approaches have proved successful in the genetic assignment of other marine organisms (Jenkins et al., 2019; Meek et al., 2016; Nielsen et al., 2012), including seabirds (Baetscher et al., 2022), and can provide a critical tool for species management.

Despite taxonomy and species conservation being two separate fields of biology with different goals, species are commonly used as the conservation unit in which conservation effort should focus (Mace 2004). Conservation practitioners and decision makers often neglect conservation units under the species level, and governments and international agencies tend to only take conservation action for threatened species. However, there is a growing tendency to fully recognise intraspecific units for conservation, moving away from species-based conservation, towards conservation of intraspecific units that guarantee the maintenance of genetic diversity and evolutionary processes (Coates et al. 2018). *P. mauretanicus* and *P. yelkouan* are good examples of differentiated populations worthy of consideration as intraspecific units that represent unique morphological, ecological and genetic diversity for conservation of biodiversity. We stress that developing action plans for *P. mauretanicus* and *P. yelkouan* should not depend on whether these taxa are classified as separate species or not.

5. Conclusions

Our analysis using high resolution genome-wide data reveals the phylogenetic relationships and population structure of a group of highly mobile pelagic seabirds, the North Atlantic and Mediterranean *Puffinus* shearwaters. By integrating across multiple methods, we provide a robust framework for species delimitation. We highlight that none of the current taxonomies provide an accurate delineation of these shearwater species and propose a more accurate taxonomy. By characterising fine-scale population level genetic structure, we further highlight the need for management of ESUs below the species level. By focusing on the genetic differentiation between three recently diverged taxon pairs, we also provide insight into the process of genomic differentiation in island-breeding marine organisms across the speciation continuum. Our findings have important implications for the conservation of these endangered seabirds as they provide detailed information of species limits and population connectivity, providing sufficient resolution for genetic assignment of shearwater bycatch in the North Atlantic Ocean and the Mediterranean Sea.

CRediT authorship contribution statement

Joan Ferrer Obiol: Conceptualization, Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Visualization,

Supervision. **Jose M. Herranz:** Formal analysis. **Josephine R. Paris:** Conceptualization, Writing – review & editing. **James R. Whiting:** Conceptualization, Writing – review & editing. **Julio Rozas:** Resources, Writing – review & editing, Funding acquisition, Supervision. **Marta Riutort:** Resources, Writing – review & editing, Funding acquisition, Supervision. **Jacob González-Solís:** Resources, Writing – review & editing, Funding acquisition, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank Xavier Gumà Saumell for his help with DNA extractions and preliminary analyses. We thank many volunteers and Jeremy J. Austin, Yann Kolbeinsson, Meritxell Genovart, Maite Louzao, Karen Bourgeois, Vincent Bretagnolle, Veronica Neves, Frank Zino Karen McCoy, Elena Gomez-Diaz and Andreanna J. Welch for collecting and/or providing shearwater blood samples from the field. We would also like to thank the Smithsonian National Museum of Natural History for providing tissue loans. We thank the following institutions for issuing permits and providing logistical support: Cabildo Insular de La Palma and Lanzarote, Gobierno de Canarias, Govern Illes Balears, Secretaria Regional do Ambiente da Região Autónoma dos Açores, Direcção Nacional do Ambiente from Cape Verde, Copeland Bird Observatory and Port Cros National Park. Finally, we thank R. Andrew King for helpful comments on an earlier draft of the manuscript and Martí Franch for his wonderful shearwater illustrations. This work was supported by the Fundación BBVA (program "Ayudas a Equipos de Investigación Científica 2017", project code 062_17 to M.R.).

Data accessibility

Raw ddRAD sequencing data used for this study are archived on the European Nucleotide Archive (ENA) under accession number PRJEB38458. FINESTRUCTURE input files, PHYLIP format alignment files for phylogenetic analyses, XML files for SNAPP analyses, Newick format phylogenetic trees and custom Python scripts are available from the github repository https://github.com/jferrerobiol/sp_delim_shearwaters.

References

Abdelkrim, J., Aznar-Cormano, L., Buge, B., Fedosov, A., Kantor, Y., Zaharias, P., Puillandre, N., 2018. Delimiting species of marine gastropods (Turridae, Conoidea) using RAD sequencing in an integrative taxonomy framework. Mol. Ecol. 27, 4591–4611.

Aberer, A.J., Kobert, K., Stamatakis, A., 2014. ExaBayes: massively parallel bayesian tree inference for the whole-genome era. Mol. Biol. Evol. 31, 2553–2556.

Acerbi, G., 1827. Biblioteca Ital. 47, 297.

Alexander, D.H., Lange, K., 2011. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. BMC Bioinf. 12, 246.

Alexander, D.H., Novembre, J., Lange, K., 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 19, 1655–1664.

Austin, J.J., 1996. Molecular phylogenetics of Puffinus shearwaters: preliminary evidence from mitochondrial cytochrome b gene sequences. Mol. Phylogenet. Evol. 6, 77–88.

Austin, J.J., Bretagnolle, V., Pasquet, E., 2004. A global molecular phylogeny of the small Puffinus shearwaters and implications for systematics of the Little-Audubon's Shearwater complex. Auk 121, 647–864.

Austin, R.E., Wynn, R.B., Votier, S.C., Trueman, C., McMinn, M., Rodríguez, A., Suberg, L., Maurice, L., Newton, J., Genovart, M., Péron, C., Grémillet, D., Guilford, T., 2019. Patterns of at-sea behaviour at a hybrid zone between two threatened seabirds. Sci. Rep. 9, 14720.

Baetscher, D.S., Beck, J., Anderson, E.C., Ruegg, K., Ramey, A.M., Hatch, S., Nevins, H., Fitzgerald, S.M., Carlos Garza, J., 2022. Genetic assignment of fisheries bycatch reveals disproportionate mortality among Alaska Northern Fulmar breeding colonies. Evol. Appl. 15, 447–458.

Bernardo, P.H., Sánchez-Ramírez, S., Sánchez-Pacheco, S.J., Álvarez-Castañeda, S.T., Aguilera-Miller, E.F., Mendez-de la Cruz, F.R., Murphy, R.W., 2019. Extreme mitonuclear discordance in a peninsular lizard: the role of drift, selection, and climate. Heredity 123, 359–370.

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

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

Bonaparte, C.L., 1856. Conspect. Gen. Avium 2, 204.

Bouckaert, R., Vaughan, T.G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., Heled, J., Jones, G., Kühnert, D., De Maio, N., Matschiner, M., Mendes, F.K., Müller, N.F., Ogilvie, H.A., du Plessis, L., Popinga, A., Rambaut, A., Rasmussen, D., Siveroni, I., Suchard, M.A., Wu, C.-H., Xie, D., Zhang, C., Stadler, T., Drummond, A.J., 2019. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. PLoS Comput. Biol. 15, e1006650.

Bourne, W.R.P., Mackrill, E.J., Paterson, A.M., Yésou, P., 1988. The Yelkouan Shearwater Puffinus (puffinus?) yelkouan. Br. Birds 81, 306–319.

Breiman, L., 2001. Random Forests. Mach. Learn. 45, 5–32.

Bretagnolle, V., 1992. Variation géographique des vocalisations de pétrels ouest-paléarctiques et suggestions taxonomiques. Alauda 60, 251–252.

Brisson, M. J., 1760. Ornithologia sive synopsis methodica sistens avium divisionem in ordines, sectiones, genera, species, ipsarumque varietates (Vol. 4). Cl. Jean-Baptiste Bauche.

Brünnich, M.T. 1764. Orn. Borealis.

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. Mol. Biol. Evol. 29, 1917–1932.

Bryant, D., Moulton, V., 2004. Neighbor-net: an agglomerative method for the construction of phylogenetic networks. Mol. Biol. Evol. 21, 255–265.

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

Carstens, B.C., Pelletier, T.A., Reid, N.M., Satler, J.D., 2013. How to fail at species delimitation. Mol. Ecol. 22, 4369–4383.

Chambers, E.A., Hillis, D.M., 2020. The Multispecies Coalescent Over-Splits Species in the Case of Geographically Widespread Taxa. Syst. Biol. 69, 184–193.

Chan, K.O., Hutter, C.R., Wood Jr, P.L., Grismer, L.L., Das, I., Brown, R.M., 2020. Gene flow creates a mirage of cryptic species in a Southeast Asian spotted stream frog complex. Mol. Ecol. 29, 3970–3987.

Christidis, L., 2014. The Howard and Moore Complete Checklist of the Birds of the World, version 4.1.

Clements, J.F., Schulenberg, T.S., Iliff, M.J., Billerman, S.M., Fredericks, T.A., Sullivan, B. L., Wood, C.L., 2019. The eBird/Clements Checklist of Birds of the. WORLD v2019.

Coates, D.J., Byrne, M., Moritz, C., 2018. Genetic diversity and conservation units: dealing with the species-population continuum in the age of genomics. Front. Ecol. Evol. 6, 165.

Cortés, V., Arcos, J.M., González-Solís, J., 2017. Seabirds and demersal longliners in the northwestern Mediterranean: factors driving their interactions and bycatch rates. Mar. Ecol. Prog. Ser. 565, 1–16.

Coulson, J.C., 2016. A Review of Philopatry in Seabirds and Comparisons with Other Waterbird Species. Cowa 39, 229–240.

Crandall, K.A., Bininda-Emonds, O.R., Mace, G.M., Wayne, R.K., 2000. Considering evolutionary processes in conservation biology. Trends Ecol. Evol. 15, 290–295.

Cristofari, R., Plaza, P., Fernández, C.E., Trucchi, E., 2019. Unexpected population fragmentation in an endangered seabird: the case of the Peruvian diving-petrel. Sci. Rep.

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 Conserv. Int. https://doi.org/10.1017/S0959270912000020.

Cuevas-Caballé, C., Ferrer Obiol, J., Vizueta, J., Genovart, M., Gonzalez-Solís, J., Riutort, M., Rozas, J., 2022. The First Genome of the Balearic Shearwater (Puffinus mauretanicus) Provides a Valuable Resource for Conservation Genomics and Sheds Light on Adaptation to a Pelagic lifestyle. Genome Biol. Evol. 14 https://doi.org/10.1093/gbe/evac067.

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., 1000 Genomes Project Analysis Group, 2011. The variant call format and VCFtools. Bioinformatics 27, 2156–2158.

De Queiroz, K., 2007. Species concepts and species delimitation. Syst. Biol. 56, 879–886.

del Hoyo, J., Collar, N.J., Christie, D.A., Elliott, A., Fishpool, L.D.C., Boesman, P., Kirwan, G.M., 2014. HBW and BirdLife International Illustrated Checklist of the Birds of the World, Volume 1. Lynx Edicions in association with BirdLife International, Barcelona, Spain and Cambridge, UK.

DeRaad, D.A., McCormack, J.E., Chen, N., Peterson, A.T., Moyle, R.G., 2022. Combining Species Delimitation, Species Trees, and Tests for Gene Flow Clarifies Complex Speciation in Scrub-Jays. Syst. Biol. https://doi.org/10.1093/sysbio/syac034.

Derkarabetian, S., Castillo, S., Koo, P.K., Ovchinnikov, S., Hedin, M., 2019. A demonstration of unsupervised machine learning in species delimitation. Mol. Phylogenet. Evol. 139, 106562.

Dias, M.P., Martin, R., Pearmain, E.J., Burfield, I.J., Small, C., Phillips, R.A., Yates, O., Lascelles, B., Borboroglu, P.G., Croxall, J.P., 2019. Threats to seabirds: A global assessment. Biol. Conserv. 237, 525–537.

Evanno, G., Regnaut, S., Goudet, J., 2005. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. Mol. Ecol. 14, 2611–2620.

Ewart, K.M., Lo, N., Ogden, R., Joseph, L., Ho, S.Y.W., Frankham, G.J., Eldridge, M.D.B., Schodde, R., Johnson, R.N., 2020. Correction: Phylogeography of the iconic

Australian red-tailed black-cockatoo (Calyptorhynchus banksii) and implications for its conservation. Heredity 125, 167.
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. Syst. Biol. 70, 976–996.
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. J. Biogeogr. 49, 171–188.
Flood, R.L., van der Vliet, R., 2019. Variation and identification of Barolo Shearwater and Boyd's Shearwater. Dutch Birding 41, 215–237.
Flouri, T., Jiao, X., Rannala, B., Yang, Z., 2018. Species tree inference with BPP using genomic sequences and the multispecies coalescent. Mol. Biol. Evol. 35, 2585–2593.
Friesen, V.L., Burg, T.M., McCoy, K.D., 2007. Mechanisms of population differentiation in seabirds: Invited review. Mol. Ecol. 16, 1765–1785.
Funk, W.C., McKay, J.K., Hohenlohe, P.A., Allendorf, F.W., 2012. Harnessing genomics for delineating conservation units. Trends Ecol. Evol. 27, 489–496.
Genovart, M., Juste, J., Oro, D., 2005. Two sibling species sympatrically breeding: A new conservation concern for the critically endangered Balearic shearwater. Conserv. Genet. 6, 601–606.
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. J. Hered. 103, 330–341.
Genovart, M., Arcos, J.M., Álvarez, D., McMinn, M., Meier, R., Wynn, R., Guilford, T., Oro, D., 2016. Demography of the critically endangered Balearic shearwater: the impact of fisheries and time to extinction. J. Appl. Ecol. 53, 1158–1168.
Gill, F., Donsker, D., Rasmussen, P., 2020. IOC World Bird List (v10. 1). doi: 10.14344/IOC.ML.10.1.
Gil-Velasco, M., Rodriguez, G., Menzie, S., Arcos, J.M., 2015. Plumage variability and field identification of Manx, Yelkouan and Balearic Shearwaters. Br. Birds 108, 518–539.
González-Solís, J., Felicísimo, A., Fox, J.W., Afanasyev, V., Kolbeinsson, Y., Muñoz, J., 2009. Influence of sea surface winds on shearwater migration detours. Mar. Ecol. Prog. Ser. 391, 221–230.
Heidrich, P., Amengual, J., Wink, M., 1998. Phylogenetic relationships in mediterranean and North Atlantic shearwaters (Aves: Procellariidae) based on nucleotide sequences of mtDNA. Biochem. Syst. Ecol. 26, 145–170.
Holmes, N.D., Spatz, D.R., Oppel, S., Tershy, B., Croll, D.A., Keitt, B., Genovesi, P., Burfield, I.J., Will, D.J., Bond, A.L., Wegmann, A., Aguirre-Muñoz, A., Raine, A.F., Knapp, C.R., Hung, C.-H., Wingate, D., Hagen, E., Méndez-Sánchez, F., Rocamora, G., Yuan, H.-W., Fric, J., Millett, J., Russell, J., Liske-Clark, J., Vidal, E., Jourdan, H., Campbell, K., Springer, K., Swinnerton, K., Gibbons-Decherong, L., Langrand, O., Brooke, M. de L., McMinn, M., Bunbury, N., Oliveira, N., Sposimo, P., Geraldes, P., McClelland, P., Hodum, P., Ryan, P.G., Borroto-Páez, R., Pierce, R., Griffiths, R., Fisher, R.N., Wanless, R., Pasachnik, S.A., Cranwell, S., Micol, T., Butchart, S.H.M., 2019. Globally important islands where eradicating invasive mammals will benefit highly threatened vertebrates. PLoS One 14, e0212128.
Hosegood, J., Humble, E., Ogden, R., de Bruyn, M., Creer, S., Stevens, G.M.W., Abudaya, M., Bassos-Hull, K., Bonfil, R., Fernando, D., Foote, A.D., Hipperson, H., Jabado, R.W., Kaden, J., Moazzam, M., Peel, L.R., Pollett, S., Ponzo, A., Poortvliet, M., Salah, J., Senn, H., Stewart, J.D., Wintner, S., Carvalho, G., 2020. Phylogenomics and species delimitation for effective conservation of manta and devil rays. Mol. Ecol. 29, 4783–4796.
Huang, J.-P., 2018. What Have Been and What Can Be Delimited as Species Using Molecular Data Under the Multi-Species Coalescent Model? A Case Study Using Hercules beetles (Dynastes; Dynastidae). Insect Systematics and Diversity. 2, 3.
Huang, J.-P., Knowles, L.L., 2016. The Species versus Subspecies Conundrum: Quantitative Delimitation from Integrating Multiple Data Types within a Single Bayesian Approach in Hercules Beetles. Syst. Biol. 65, 685–699.
Huson, D.H., Bryant, D., 2006. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. 23, 254–267.
Jarvis, E.D., Mirarab, S., Aberer, A.J., Li, B., Houde, P., Li, C., Ho, S.Y.W., Faircloth, B.C., Nabholz, B., Howard, J.T., Suh, A., Weber, C.C., da Fonseca, R.R., Alfaro-Núñez, A., Narula, N., Liu, L., Burt, D., Ellegren, H., Edwards, S.V., Stamatakis, A., Mindell, D. P., Cracraft, J., Braun, E.L., Warnow, T., Jun, W., Gilbert, M.T.P., Zhang, G., Avian Phylogenomics Consortium, 2015. Phylogenomic analyses data of the avian phylogenomics project. Gigascience 4, 4.
Jenkins, T.L., Ellis, C.D., Triantafyllidis, A., Stevens, J.R., 2019. Single nucleotide polymorphisms reveal a genetic cline across the north-east Atlantic and enable powerful population assignment in the European lobster. Evol. Appl. 12, 1881–1899.
Jombart, T., Ahmed, I., 2011. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics 27, 3070–3071.
Jombart, T., Devillard, S., Balloux, F., 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet. 11, 94.
Kass, R.E., Raftery, A.E., 1995. Bayes Factors. J. Am. Stat. Assoc. 90, 773–795.
Knowles, L.L., Carstens, B.C., 2007. Delimiting species without monophyletic gene trees. Syst. Biol. 56, 887–895.
Kozlov, A.M., Darriba, D., Flouri, T., Morel, B., Stamatakis, A., 2019. RAxML-NG: A fast, scalable, and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics 1–3.
Leaché, A.D., Fujita, M.K., Minin, V.N., Bouckaert, R.R., 2014. Species delimitation using genome-wide SNP Data. Syst. Biol. 63, 534–542.
Lesson, R.P., 1839. Rev. Zool. Soc. Cuvierienne 2, 102.
Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079.
Li, H., 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv [q-bio.GN].
Lowe, P.R., 1921. Bull. Brit. Orn. Club 41, 141.
Mace, G.M., 2004. The role of taxonomy in species conservation. Philos. Trans. R. Soc. Lond. B Biol. Sci. 359, 711–719.
Mackin, W.A., 2016. Current and former populations of Audubon's Shearwater (Puffinus lherminieri) in the Caribbean region. Condor 118, 655–673.
Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., Hornik, K., 2022. cluster: Cluster Analysis Basics and Extensions. R package version 2 (1), 3.
Malinsky, M., Trucchi, E., Lawson, D.J., Falush, D., 2018. RADpainter and fineRADstructure: Population Inference from RADseq Data. Mol. Biol. Evol. 35, 1284–1290.
Maruki, T., Lynch, M., 2015. Genotype-Frequency Estimation from High-Throughput Sequencing Data. Genetics 201, 473–486.
Maruki, T., Lynch, M., 2017. Genotype Calling from Population-Genomic Sequencing Data. G3: Genes|Genomes|Genetics 7, 1393–1404.
Mathews, G.M., 1912. Birds of Australia, 549–551.
Mathews, G.M., 1934. A check-list of the order Procellariiformes.
Meek, M.H., Baerwald, M.R., Stephens, M.R., Goodbla, A., Miller, M.R., Tomalty, K.M.H., May, B., 2016. Sequencing improves our ability to study threatened migratory species: Genetic population assignment in California's Central Valley Chinook salmon. Ecol. Evol. 6, 7706–7716.
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, e115650.
Morales, H.E., Pavlova, A., Joseph, L., Sunnucks, P., 2015. Positive and purifying selection in mitochondrial genomes of a bird with mitonuclear discordance. Mol. Ecol. 24, 2820–2837.
Moritz, C., 2002. Strategies to protect biological diversity and the evolutionary processes that sustain it. Syst. Biol. 51, 238–254.
Nei, M., Li, W.H., 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc. Natl. Acad. Sci. U. S. A. 76, 5269–5273.
Newton, L.G., Starrett, J., Hendrixson, B.E., Derkarabetian, S., Bond, J.E., 2020. Integrative species delimitation reveals cryptic diversity in the southern Appalachian Antrodiaetus unicolor (Araneae: Antrodiaetidae) species complex. Mol. Ecol. 29, 2269–2287.
Nielsen, E.E., Cariani, A., Mac Aoidh, E., Maes, G.E., Milano, I., Ogden, R., Taylor, M., Hemmer-Hansen, J., Babbucci, M., Bargelloni, L., Bekkevold, D., Diopere, E., Grenfell, L., Helyar, S., Limborg, M.T., Martinsohn, J.T., McEwing, R., Panitz, F., Patarnello, T., Tinti, F., Van Houdt, J.K.J., Volckaert, F.A.M., Waples, R.S., FishPopTrace consortium, Albin, J.E.J., Vieites Baptista, J.M., Barmintsev, V., Bautista, J.M., Bendixen, C., Bergé, J.-P., Blohm, D., Cardazzo, B., Diez, A., Espiñeira, M., Geffen, A.J., Gonzalez, E., González-Lavín, N., Guarniero, I., Jeráme, M., Kochzius, M., Krey, G., Mouchel, O., Negrisolo, E., Piccinetti, C., Puyet, A., Rastorguev, S., Smith, J.P., Trentini, M., Verrez-Bagnis, V., Volkov, A., Zanzi, A., Carvalho, G.R., 2012. Gene-associated markers provide tools for tackling illegal fishing and false eco-certification. Nat. Commun. 3, 851.
Nosil, P., Feder, J.L., 2012. Genomic divergence during speciation: causes and consequences. Philos. Trans. R. Soc. Lond. B Biol. Sci. 367, 332–342.
Olson, S.L., 2010. Stasis and turnover in small shearwaters on Bermuda over the last 400 000 years (Aves: Procellariidae: Puffinus lherminieri group): small Bermuda shearwaters. Biol. J. Linn. Soc. Lond. 99, 699–707.
Oppel, S., Raine, A.F., Borg, J.J., Raine, H., Bonnaud, E., Bourgeois, K., Breton, A.R., 2011. Is the Yelkouan shearwater Puffinus yelkouan threatened by low adult survival probabilities? Biol. Conserv. 144, 2255–2263.
Oro, D., Aguilar, J.S., Igual, J.M., Louzao, M., 2004. Modelling demography and extinction risk in the endangered Balearic shearwater. Biol. Conserv. 116, 93–102.
Palsbøll, P.J., Bérubé, M., Allendorf, F.W., 2007. Identification of management units using population genetic data. Trends Ecol. Evol. 22, 11–16.
Paradis, E., 2010. pegas: an R package for population genetics with an integrated–modular approach. Bioinformatics 26, 419–420.
Paris, J.R., Stevens, J.R., Catchen, J.M., 2017. Lost in parameter space: a road map for stacks. Methods Ecol. Evol. 8, 1360–1373.
Perrier, C., Ferchaud, A.L., Sirois, P., Thibault, I., Bernatchez, L., 2017. Do genetic drift and accumulation of deleterious mutations preclude adaptation? Empirical investigation using RADseq in a northern lacustrine fish. Mol. Ecol. 26, 6317–6335.
Peters, J.L., 1931. Check-list of Birds of the World, 1931–1986. Harvard University Press/Museum of Comparative Zoology.
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, e37135.
Pfeifer, B., Wittelsbürger, U., Ramos-Onsins, S.E., Lercher, M.J., 2014. PopGenome: an efficient Swiss army knife for population genomic analyses in R. Mol. Biol. Evol. 31, 1929–1936.
Rambaut, A., Drummond, A.J., Xie, D., Baele, G., Suchard, M.A., 2018. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. Syst. Biol. 67, 901–904.
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. Zool. J. Linn. Soc. 191, 482–502.
Rannala, B., Yang, Z., 2017. Efficient Bayesian Species Tree Inference under the Multispecies Coalescent. Syst. Biol. 66, 823–842.

Ravinet, M., Faria, R., Butlin, R.K., Galindo, J., Bierne, N., Rafajlović, M., Noor, M.A.F., Mehlig, B., Westram, A.M., 2017. Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow. J. Evol. Biol. 30, 1450–1477.

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. Nat. Commun. 2, 332.

Rexer-Huber, K., Veale, A.J., Catry, P., Cherel, Y., 2019. Genomics detects population structure within and between ocean basins in a circumpolar seabird: The white-chinned petrel. Mol. Ecol. 28, 4552–4572.

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

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., Hodum, P., Humphries, G.R.W., Le Corre, M., Lebarbenchon, C., Martin, R., Melvin, E.F., Montevecchi, W.A., Pinet, P., Pollet, I.L., Ramos, R., Russell, J.C., Ryan, P.G., Sanz-Aguilar, A., Spatz, D.R., Travers, M., Votier, S.C., Wanless, R.M., Woehler, E., Chiaradia, A., 2019. Future Directions in Conservation Research on Petrels and Shearwaters. Front. Mar. Sci. 6, 94.

Rodríguez, A., Rodríguez, B., Montelongo, T., Garcia-Porta, J., Pipa, T., Carty, M., Danielsen, J., Nunes, J., Silva, C., Geraldes, P., Medina, F.M., Illera, J.C., 2020. Cryptic differentiation in the Manx Shearwater hinders the identification of a new endemic subspecies. J. Avian Biol. 51 https://doi.org/10.1111/jav.02633.

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

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

Scrucca, L., Fop, M., Murphy, T.B., Raftery, A.E., 2016. mclust 5: Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models. R J. 8, 289–317.

Sloan, D.B., Havird, J.C., Sharbrough, J., 2017. The on-again, off-again relationship between mitochondrial genomes and species boundaries. Mol. Ecol. 26, 2212–2236.

Spatz, D.R., Holmes, N.D., Reguero, B.G., Butchart, S.H.M., Tershy, B.R., Croll, D.A., 2017. Managing Invasive Mammals to Conserve Globally Threatened Seabirds in a Changing Climate. Conserv. Lett. 10, 736–747.

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313.

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

Sukumaran, J., Knowles, L.L., 2017. Multispecies coalescent delimits structure, not species. Proc. Natl. Acad. Sci. U. S. A. 114, 1607–1612.

Taylor, R.S., Bolton, M., Beard, A., Birt, T., Deane-Coe, P., Raine, A.F., González-Solís, J., Lougheed, S.C., Friesen, V.L., 2019. Cryptic species and independent origins of allochronic populations within a seabird species complex (Hydrobates spp.). Mol. Phylogenet. Evol. 139, 106552.

Toews, D.P.L., Brelsford, A., 2012. The biogeography of mitochondrial and nuclear discordance in animals. Mol. Ecol. 21, 3907–3930.

Tonzo, V., Papadopoulou, A., Ortego, J., 2019. Genomic data reveal deep genetic structure but no support for current taxonomic designation in a grasshopper species complex. Mol. Ecol. 28, 3869–3886.

Van der Maaten, L.V.D., Hinton, G., 2008. Visualizing data using t-SNE. J. Mach. Learn. Res. 9, 2579–2605.

Warham, J., 1996. The behaviour, population biology and physiology of the petrels. Academic Press.

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

Wetmore, A., 1959. The birds of Isla Escudo de Veraguas, Panamá. Smithsonian Miscellaneous Collections.

Yang, Z., Rannala, B., 2010. Bayesian species delimitation using multilocus sequence data. Proc. Natl. Acad. Sci. U. S. A. 107, 9264–9269.

Yang, Z., Rannala, B., 2014. Unguided species delimitation using DNA sequence data from multiple Loci. Mol. Biol. Evol. 31, 3125–3135.

Yésou, P., Paterson, A.M., Mackrill, E.J., Bourne, W.R.P., 1990. Plumage variation and identification of the "Yelkouan Shearwater." Br. Birds 83.

Zhang, C., Rabiee, M., Sayyari, E., Mirarab, S., 2018. ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinf. 19, 15–30.