

# Sea surface temperature, rather than land mass or geographical distance, may drive genetic differentiation in a species complex of highly-dispersive seabirds

Lucas Torres<sup>1</sup>, Eric Pante<sup>2</sup>, Jacob González-Solís<sup>3</sup>, Amelia Viricel<sup>2</sup>, Cecile Ribout<sup>1</sup>, Frank Zino<sup>4</sup>, Will MacKin<sup>4</sup>, Carine Precheur<sup>4</sup>, Julie Tourmetz<sup>5</sup>, Licia Calabrese<sup>6</sup>, Teresa Militão<sup>7</sup>, Laura Zango<sup>3</sup>, Hadoram Shirihaï<sup>4</sup>, and Vincent Bretagnolle<sup>1</sup>

<sup>1</sup>CEBC

<sup>2</sup>LIENSs

<sup>3</sup>University of Barcelona

<sup>4</sup>Affiliation not available

<sup>5</sup>SEOR

<sup>6</sup>Island Conservation Society

<sup>7</sup>Universitat de Barcelona

July 20, 2020

## Abstract

Seabirds, particularly Procellariiformes, are highly mobile organisms with a great capacity for long dispersal, though simultaneously showing high philopatry, two conflicting characteristics that may lead to contrasted patterns of genetic population structure. Landmasses were suggested to explain differentiation patterns observed in seabirds, but philopatry, isolation-by-distance, segregation between breeding and non-breeding zones, and oceanographic conditions (sea surface temperatures) may also contribute to differentiation patterns. No study has simultaneously contrasted the multiple factors contributing to the diversification of seabird species, especially in the grey zone of speciation. We conducted a multi-locus phylogeographic study on a widespread shearwater species complex (*Puffinus lherminieri/bailloni*), showing highly homogeneous morphology. We sequenced three mitochondrial and six nuclear markers on all extant populations (five nominal lineages, 13 populations). We found sharp differentiation among populations separated by the African continent with both mitochondrial and nuclear markers, while only mitochondrial markers allowed characterizing the five nominal lineages. No differentiation could be detected within these five lineages, questioning the strong level of philopatry showed by these shearwaters. Finally, we propose that Atlantic populations likely originated from the Indian Ocean. Within the Atlantic, a stepping-stone process accounts for the current distribution. Based on our divergence times estimates, we suggest that the observed pattern of differentiation mostly resulted from variation in sea surface temperatures.

## Keywords

*Puffinus*, divergence, phylogeography, multi-locus, philopatry, mito-nuclear discordance

## Introduction

Species are cornerstone in evolution and are used as study units in both evolutionary and conservation biology (Butlin et al. 2012, Wake 2006). Though divergence with gene-flow has been theorized and observed (review in Pinho & Hey 2010), the model of allopatric speciation predominates by far in the literature (Stroud & Losos 2016, Butlin et al. 2012). In this model, a physical barrier to gene flow catalyzes genetic

differentiation between populations, through selection and/or genetic drift, eventually followed by other pre- or post-zygotic barriers (Coyne & Orr 2004). In practice however, the mechanisms that impede gene flow and promote differentiation are multifactorial and still poorly understood (Butlin et al. 2012, Price 2008). In particular, geographic barriers alone may not explain the differentiation of populations in highly dispersive species, e.g. marine birds (e.g. Genovart et al. 2007), birds of prey (e.g. Doyle et al. 2016), mammals (Hassanin et al. 2018) or plants (Sanz et al. 2014). Seabirds are a case in point: their wide geographic distribution and dispersal ability should theoretically maintain high levels of gene flow, but many seabirds show surprisingly strong geographic population structure, a pattern attributed to their high degree of philopatry (Abbott & Double 2003, Smith et al. 2007a).

Present or historic landmasses were identified as the most important barriers to gene flow in seabirds (Friesen 2015, Friesen et al. 2007a) though they cannot explain every differentiation patterns, and other factors such as philopatry, isolation-by-distance, or segregation between breeding and non-breeding zones may play a role. So far, however, no study has contrasted all these processes simultaneously. Moreover, most previous studies were based on maternally-inherited mitochondrial markers (Friesen 2015), while gene flow is expected to vary between sexes given stronger male philopatry in seabirds and other birds in general (Greenwood 1980). Sex-specific patterns of divergence are expected, and indeed mito-nuclear discordance has been detected in seabirds, with more genetic structure in mt than in nuDNA (Burg & Croxall 2001, Deane 2013, Gangloff et al. 2013, Silva et al. 2015, Welch et al. 2011, Friesen et al. 2006, Welch Fleicher 2012, Genovart 2012, Sonsthagen et al. 2016; but see Pons et al. 2014). Such discordance was suggested to result from incomplete lineage sorting in nuDNA due to a higher effective sample size than mtDNA (McKay & Zink 2010), but other mechanisms were proposed such as adaptive introgression of mtDNA, demographic disparities and sex-biased asymmetries (review in Toews & Brelsford 2012). Disentangling these latter processes can be difficult (McKay & Zink 2010), thus a combined use of mitochondrial and biparentally-inherited nuclear markers, as well as coalescent theory-based methods using the lineage species concept coupled with population genetics within a phylogeographic context, are advocated (Hudson & Coyne 2002).

Shearwaters (order Procellariiformes) are long-lived birds showing slow demographic rates (Warham 1996). They breed in large colonies on remote oceanic islands, are pelagic (González-Solís et al. 2007) and highly philopatric (Brooke 2004). Geographic structuring of populations can be strong (Genovart et al. 2012, Gómez-Díaz et al. 2009), and the systematics of some taxa is still highly controversial, especially for the little and Audubon's Shearwaters *Puffinus assimilis-lherminieri* complex (Austin et al. 2004). This widespread small-sized, black-and-white shearwater breeds from equatorial to sub-arctic seas (see map in Fig. 1). Previous studies, based only on the mitochondrial gene *cytb*, indicated that population genetic clusters matched with geographic distribution (Austin et al. 2004, Kawakami et al. 2018). Using this single marker, three distinct lineages were recognized in the North Atlantic: *lherminieri* in the Caribbean and off Brazil, *baroli* in the Azores, Canaries and Madeira, and *boydi* in Cape Verde (Fig. 1). These taxa are characterised by non-overlapping breeding and non-breeding distributions at sea (Ramos et al. in review), and are thus geographically and genetically (for at least one marker) separated. Still, they are morphologically and ecologically highly similar (Precheur et al. 2016, Calabrese et al. in review), a pattern typical of the first stages of the speciation process, the so-called grey-zone (De Queiroz 2007). Unsurprisingly, their taxonomic ranking has been hotly debated, e.g. *baroli* being considered as belonging to *assimilis* (Shirihai et al. 1995), *lherminieri* (Austin et al. 2004) or a species of its own (Sangster et al. 2005). Lineages belonging to this complex are also found in the Indian and Pacific Ocean, with populations breeding in the Seychelles (*nicolae*), Réunion (*bailloni*) and many islands in the Pacific Ocean (*dichrous*). Breeding populations of *bailloni* are characterised by different breeding phenologies on the northern and southern parts of Réunion (Bretagnolle & Attie 1996), potentially impacting genetic structuration (Friesen et al. 2007b). Indian Ocean birds were alternatively considered as a *P. lherminieri* subspecies (Warham 1990) or subspecies of a *bailloni* pan-tropical taxon (Austin et al. 2004). The exact taxonomic status of these six lineages is thus largely unresolved.

Here we consider these six lineages, covering Atlantic and the Indo-Pacific branches of the *Puffinus assimilis-lherminieri* complex. We sampled birds from all but one known breeding localities in the North Atlantic (four Caribbean breeding sites for *lherminieri*, two main breeding sites in Cape Verde for *boydi*, one breeding site

in Azores, Madeira, Selvagens Canary Islands for *baroli*) and in the northern Indian Ocean (northern and southern populations of *bailloni* in Réunion and Seychelles for *nicolae*; Fig. 1). Only a single breeding locality was unsampled, the Fernando de Noronha archipelago, off Brazil (Fig. 1). We analysed three mitochondrial and six nuclear markers to delineate genetic units and investigated patterns of genetic differentiation and divergence among populations. As conflicting geographic patterns between mitochondrial and nuclear markers (hereon referred to as mito-nuclear discordance; Toews & Brelsford 2012) was systematically found in petrels so far (references above), we inferred population structure among females and males independently to test for sex-biased dispersal on nuDNA. We then used multispecies coalescent inference and an ABC framework to investigate evolutionary scenarios of breeding site colonisation over the last million years, trying to disentangle contrasted processes shaping evolutionary history and the contemporary population structure of this complex, such as landmass presence, isolation-by-distance, oceanographic conditions (sea surface temperature) and climatic oscillations.

## Material and Methods

### 2.1 Sampling, extraction and PCR amplification of gDNA

A total of 276 birds (all adults, i.e. having already dispersed) spanning the entire known breeding distribution of *Puffinus lherminieri* and four populations of interest of *P. bailloni* were selected (Supplementary Material 1). Four individuals from the Pacific Ocean (taxon *dichrous*, by far the most widespread lineage in the Pacific (Onley & Scofield 2007); Fig. 1) were also used. Individuals were sexed using PCR amplification (with the 2250F and 2781R primers (Fridolfsson & Ellegren 1999)). The overall sex ratio was unbiased (120 females, 107 males; Pearson's  $\chi^2$  with Yates' continuity correction,  $p=0.42$ ; 49 individuals could not be sexed successfully, see Supplementary Material 1).

Total genomic DNA was extracted from blood samples (except for the population of the Bahamas, for which samples were derived from toepads collected on dead birds, *lherminieri* 1 and 2 on Figure 1) using NucleoSpin® Tissue XS Kit (Macherey & Nagel, Düren, Germany). Samples were incubated overnight in 4 mg of Proteinase K. Purified genomic DNA was eluted twice in 50  $\mu$ L of TE buffer pre-heated at 70°C. DNA concentration was measured using Nanodrop spectrophotometry. Three mitochondrial markers (*cox1*, *cytb*, and the mitochondrial Control Region, CR) and six nuclear markers (Beta-fibrinogen exon6 through 8,  $\beta\phi\beta$ ; Cold Shock Domain-containing E1 intron5, *csde*; Interferon Regulatory Factor 2 intron2, *irf2*; PAX interacting protein 1 intron20, *pax*; Recombination activating protein 1, *rag1*; and Tropomyosin 1 alpha exon7, *tpm*) were targeted (primer sequences, PCR profile and conditions: Supplementary Material 2). These markers were previously shown to be polymorphic within and between petrel species (Gangloff et al. 2013, Silva et al. 2011).

### 2.2 Quality control of genetic data

While checking chromatograms we found several cases of double peaks on the sequences of *cox1*, *cytb* and CR (Supplementary Material 3). Bird blood contains relatively few mitochondria, and it is therefore likely to amplify nuclear copies of mitochondrial markers, or “numts” (Sorenson & Quinn 1998). Such nuclear copies may diverge from the original mitochondrial genes since they are non-coding, which result in double-peaks on the chromatograms. To check this scenario and avoid such copies, we digested nuclear DNA with the ExonucleaseV (ExoV; NEB-M0345S) and sequenced again the mitochondrial markers for all individuals showing double-peaks for *cox1*, plus 5 individuals showing no double-peaks randomly chosen, using a protocol modified from (Jayaprakash et al. 2015, see Supplementary Material 3). Before running analyses, we checked that coding sequences contained no stop codons or indels. Some analyses required phased data (e.g. \*BEAST analysis), so the gametic phase of nuclear markers was determined probabilistically using PHASE 2.1 (Stephens et al. 2001) implemented in DNAsp v.5.10.01 (Librado & Rozas 2009). Additional Genbank sequences (Supplementary Material 1) were aligned to our sequences using MAFFT v 7.187 (Katoh et al. 2002).

### 2.3 Population diversity, differentiation, and divergence

Each population was described by calculating haplotype frequencies, inter-haplotype distances, haplotype diversity and nucleotide diversity ( $\pi$ ), as well as expected and observed heterozygosity for nuclear markers using DNAsp v.5.10.01 (Librado & Rozas 2009) and Genetix v4.05.2 (Belkhir et al. 2004). We evaluated signals for departures from neutrality or demographic changes by estimating Tajima's D (Tajima 1989) and Fu's  $F_s$  (Fu 1997) for each locus, with Arlequin v.3.1 (Excoffier et al. 2005). Differentiation among populations was estimated by performing AMOVAs, and calculating pairwise  $F_{ST}$  and  $\Phi_{ST}$  and the population average pairwise differences  $D_{XY}$ , using Arlequin. For AMOVAs, samples were stratified into five groups, corresponding to the five nominal lineages (*lherminieri*, *boydi* and *baroli* in the Atlantic, *nicolae* and *bailloni* in the Indian Ocean), and populations (i.e. sampling localities; Fig. 1) within these groups. The matrix of genetic distances among all pairs of haplotypes was computed using the K2P model of substitution for concatenated mitochondrial markers, and TN93 for concatenated nuclear markers, as determined using jModelTest2. We used a Mantel test to measure the level of correlation among genetic distances and geographic distances (Smouse et al. 1986). Geographic distance was calculated as the shortest distance between two populations without crossing land. Statistical significance (AMOVAs, Pairwise  $F_{ST}$  and Mantel tests) was estimated using 1000 permutations. To visualize relationships among lineages, we inferred NeighborNet networks using SplitsTree v 4.14.2 (Huson & Bryant 2006), with different dataset combinations: all markers independently, concatenated mitochondrial markers, concatenated nuclear markers.

## 2.4 Estimation of sex-biased dispersal using nuclear markers

To detect sex-biased dispersal, we separated sequences from females and males into two separate datasets, excluding three populations (Saint-Barth  l  my, Funchal and Selvagem) represented by fewer than five individuals from each sex. Sex-biased dispersal was tested at both intra- and inter-lineage scales, expecting females to be less structured than males since males are presumably more philopatric (Greenwood 1980). We calculated average pairwise relatedness for each sex, within each population, using the triadic likelihood estimator (Wang 2007) implemented in Coancestry (Wang 2011). To test if the difference of mean relatedness between males and females of each population was significant, we used the test of difference between sex by bootstrapping samples 1000 times and recalculating difference in means between sex for each bootstrap. Observed and simulated differences was then compared, and if the observed difference fell outside of the 95% confidence interval, we considered it to be significant.

If females disperse more than males, females sampled from a single population will be a mixture of residents and immigrants. The female sample will therefore deviate from the Hardy-Weinberg equilibrium and show a deficit of heterozygotes (Wahlund effect).  $F_{IS}$  calculated for the female sample is thus expected to be larger than the male  $F_{IS}$  (Goudet et al. 2002). We estimated  $F_{IS}$  separately for females and males for all tested populations, and evaluated its significance using 1000 permutations with Genetix. Conversely, we expect  $F_{ST}$  (Goudet et al. 2002) to be higher in philopatric males than in females. We calculated  $F_{ST}$  for each pair of populations within the two datasets, with Arlequin.

## 2.5 Phylogeographic scenarios

Reciprocal monophyly was inferred on all markers as partition by gene trees, which were used to evaluate the degree of divergence among lineages using MrBayes v. 3.2.6 (Ronquist et al. 2012; Supplementary Material 4). To reconstruct the scenario of divergence among the different populations, we used species trees, inferred using two different methods. We first ran an analysis with \*BEAST v 2.2.0 (Bouckaert et al. 2014) using the three mitochondrial markers. We choose to link time-trees for the three mitochondrial markers, since they are on the same plasmid, where no recombination is expected at the time scale considered in this study. However, the three markers have different composition and different evolutionary rates, so we did not link the molecular clock and evolution models. We tested the hypothesis of molecular clock with the Clock Test using ML implemented in MEGA v7.0.20 (Kumar et al. 2016). This hypothesis was rejected (p-value < 0.0005). We therefore used, for each marker, an uncorrelated lognormal relaxed clock model. The clock rate was fixed to  $0.00553 \pm 0.00115$  substitution per site per million year for *cox1* and  $0.00631 \pm 0.0035$  for *cytb* (rates inferred for Procellariiformes by (Nabholz et al. 2016)). The rate for the control region was estimated by the model as no rate is published for Procellariiformes. Existing rates for other birds could not be used

either because they belong to groups that are too distant (e.g. Moas (Baker et al. 2005) or Peafowls (Kimball et al. 1997)), and the control region is highly variable among groups (Ruokonen & Kvist 2002). Published rates for the CR were however set in \*BEAST as priors and did not affect the results (data not shown). We used for each marker a model consistent with the result of jModelTest2, and a Yule process species tree prior with a continuous population size model. As for the MrBayes analysis, we ran each MCMC chain with  $50 \times 10^6$  generations, sampled every 1,000 generations, the first 25% of generations were discarded as burn-in and we inspected the stationarity of the chains using Tracer (Rambaut et al. 2018). To test whether the colonization of the Atlantic could result from birds of the Pacific passing through the Panama Isthmus, individuals from the central Pacific (Marquesas archipelago, taxon *dichrous*) were added to the \*BEAST inference. This taxon can be considered as the best representative taxon for the central Pacific, as it is the most widespread and numerous, since *polynesiae* is considered synonym to *dichrous* (Austin et al. 2004), while *bannermani* is best recognised as a species on its own (Kawakami et al. 2018). The taxon *gunax* (from Vanuatu) has never been sequenced, and actually the location of its breeding colonies is unknown. Three *dichrous cytb* sequences retrieved from Genbank (AY219949-AY219951) and three individuals from our own collection were used. We ran a second \*BEAST analysis by adding all the nuclear markers independently to the three mitochondrial markers, using the same MCMC parameters. Clock rates priors were set to 0.019 substitutions per site per million year for *βφβ* and 0.013 substitution per site per million year for *rag1*, since these rates were estimated for birds (Groth & Barrowclough 1999; Pritchitko & Moore 1997). The clock rate for nuclear markers was estimated by the model as no rate has been produced for petrels. In this latter case, we kept only the individuals for which we had sequences for all markers. To investigate the demographic history of lineages, we estimated population size through time by estimating Extended Bayesian Skyline plot as implemented in \*BEAST, for each genetic unit previously delimited, considering a generation time of 15 years (Precheur et al. 2016).

We also used a coalescent-based ABC approach to explore the best demographic scenario describing the dataset of the combined mitochondrial and nuclear markers using the program DIYABC v. 2.1.0 (Cornuet et al. 2014). ABC methods consist in the simulation of datasets similar to the real dataset in terms of population and marker sizes. First, in the Indian Ocean, we tested if the three populations emerged simultaneously in a radiation event or in two disjoint events by comparing the posterior probability of these three scenarios, and in the case of the latter, which population was ancestral to the two others and which one of the two remaining populations was ancestral to the other. This hierarchical strategy was applied to each lineage independently, then to the three Atlantic lineages, and finally considering the five lineages together (see Supplementary Material 5 for a description of all tested scenarios). For each possible scenario,  $10^6$  pseudo-observed datasets were simulated, with the same ploidy and number of loci per population as observed in the real dataset. We fixed uniform priors for population sizes, times of size variation and divergence and mutation and admixture rates priors (see Supplementary Material 4 for details), from which we simulated the datasets. Summary statistics were calculated from the simulated datasets and compared to the same statistics obtained from the real dataset. The Euclidean distance was calculated between the statistics obtained for each normalized simulated dataset and those for the observed dataset (Beaumont et al. 2002). Posterior probability of each scenario was then calculated using a logistic regression on summary statistics produced by the 1% of the simulated datasets closest to the real dataset. To reduce the dimensionality of the data, a linear discriminant analysis was preliminarily applied to the summary statistics (Estoup et al. 2012). The scenario with the highest posterior probability value with a non-overlapping 95% confidence interval (95% CI) was selected.

## Results

### 3.1 Patterns of genetic diversity, numts and the presence of a duplicated region

We obtained an average of 192 sequences per marker (length range 307-1323 bp; Table 1). Mitochondrial data produced 148 polymorphic sites yielding 150 haplotypes, while nuclear data exhibited a total of 111 variable sites and 150 alleles (see Table 1 for summary of polymorphic sites, haplotypes and diversities per marker). Mitochondrial markers were twice as variable as nuclear markers, though nuclear haplotype and

nucleotide diversity in  $\beta\phi\beta$  reached a level similar to mitochondrial haplotype diversity (Table 1).

None of the coding markers (*cox1*, *cytb* and *pax*) presented any insertion, deletion, nonsense or stop-codon following translation (see Supplementary Material 1). Double peaks on Sanger chromatograms were however detected for each of the three mtDNA markers. While all double peaks at *cox1* were removed by the exonuclease treatment, 60 CR sequences (33%) still showed double-peaks at 73 positions, as well as 37 positions for 22 individuals (10%) for *cytb*. Double-peaks were not specific to any population or sex, and were not linked to the position of the individuals in the sequencing plate (see Supplementary Material 3). Only 12 (5%) individuals showed double-peaks both at CR and *cytb*, so the presence of double peaks seemed unlinked between the two markers. Replicating DNA extractions, PCR and sequencing confirmed these results, making laboratory contamination unlikely. Contamination in the field was also unlikely since new sampling supplies were used for every sample. Given that only 10% of the *cytb* sequences presented such ambiguities (which may be due to heteroplasmy, Torres et al 2018), we removed such sequences for further analyses expecting little impact on the analyses. However, for the CR, since a third of the total sequences were involved, we kept all CR data in further analyses, considering two haplotypic phases for MrBayes and \*BEAST analyses (although this is a violation of the assumption that mixed sequences to be phased are under Hardy-Weinberg equilibrium; see discussion below).

### 3.2 Population structure and sex-biased dispersal

Mitochondrial and nuclear results from Fu's  $F_s$  indicated no deviation from neutrality (Supplementary Material 7). However, for Tajima's  $D$  tests, three localities displayed significant negative Tajima's  $D$  for all mitochondrial and nuclear markers: South Reunion (Indian Ocean, see map in Figure 1), Saint-Barthélemy (W Atlantic) and Raso (E Atlantic). In addition, three localities presented significant negative Tajima's  $D$  at mitochondrial loci only (Selvagem and Funchal, E Atlantic; South Reunion) and two at nuclear loci only (Vila, E Atlantic; North Reunion). Patterns of population structure at seven out of 13 localities might therefore be influenced by selection and/or recent demographic changes, in addition to neutral processes.

Gene trees and phylogenetic inference with \*BEAST and MrBayes revealed a hierarchical structure composed of two well-supported (posterior probabilities PP [?] 0.95) reciprocally monophyletic clades corresponding to the two oceans, within which individuals from the five lineages further clustered into monophyletic sub-clades (Fig. 2b,c and S7). All except one of these sub-clades (East Atlantic *boydi*) were supported in \*BEAST (PP [?] 0.95) using all concatenated markers. For both mtDNA markers, and all markers concatenated, the central-Pacific *dichrous* lineage was nested within the *bailloni* / *nicolae* clade, although node supports for *dichrous* position within Indian Ocean clade were weak (between 0.27 and 0.73). Assignment to an ocean basin based on nuclear haplotype networks was however discordant from the mitochondrial data for 33 individuals, for at least one nuclear locus (Fig. 3b and S8): 15 Atlantic individuals fell closely to the Indian phylogroup, and 18 Indian Ocean individuals clustered within the Atlantic group. All of these 33 individuals showed the mitochondrial signature expected based on their geographical sampling location. Interestingly, a *baroli* individual showed one haplotypic phase clustering with the *baroli* phylogroup (mother) while the other haplotypic phase (father) clustered with the *nicolae* phylogroup for four nuclear markers (the two remaining could not be assigned to any particular lineages). This individual might be the result of hybridization, although further analyses based on additional markers would be necessary to detect more robustly hybridization among lineages. The ambiguous assignment of the other individuals might be due to introgression or incomplete lineage sorting (see below).

In parallel, we used an AMOVA framework with the five nominal lineages now defined a priori, to examine how genetic variants partitioned among and within these taxonomic units. Most of the genetic variance was due to inter-lineage differentiation (88.5% and 58.4% for mitochondrial and nuclear markers, respectively). The variance among sampling localities within lineages accounted for 0.5 to 4.1%, while variance within sampling localities represented 11.0 to 37.5%. Pairwise  $F_{ST}$  showed consistently higher values among, than within lineages for both marker types, with mostly non-significant values within each lineage (Table 2a). Moreover, 24 nuclear  $F_{ST}$  values were found non-significant versus 10 mitochondrial  $\Phi_{ST}$  values (Table 2a). Population average pairwise differences led to similar results, with high structuration for the five nominal

lineages (Supplementary Material 7).

Genetic distance increased clearly with geographic distance (Fig. 4a), but Mantel tests were performed only between pairs within the same Ocean (given that each Ocean taxon is likely different species). Tests confirmed that genetic and geographic distances were strongly correlated to each other, both for mtDNA and nuclear markers when analyzing pairs of populations within an Ocean basin ( $r=0.88$  and  $0.70$ ,  $n=45$ ,  $p < 0.005$  for mtDNA and nuDNA, respectively, Fig. 4a,b). Between breeding sites and within lineages, isolation by distance could not be reliably tested as the number of populations was too low, but visually it seemed that there was no relation between geographic and genetic distances (Fig. 4).

Indian Ocean populations (*nicolae* and *bailloni*) showed stronger female dispersion as indicated by significantly stronger deficit of heterozygotes and a significantly lower average relatedness in females (Table 3). Conversely, in *baroli*,  $F_{IS}$  was significantly higher for males and they were less related to each other than females, suggestive of male-biased dispersal. Finally, an ambiguous pattern was found for both *therminieri* (males had stronger deficit of heterozygotes and were significantly more related to each other than females) and *boydi* (female  $F_{IS}$  was significantly higher than male  $F_{IS}$ , though females were more related to each other than for males at one sampling locality). In addition, population structure within lineages, as measured with  $F_{ST}$ , was similar between sexes, but between oceans a larger range of  $F_{ST}$  values was observed for males with higher maximum values, suggesting that males were more structured at least for some pairs of populations (e.g. *therminieri* vs. Indian Ocean lineages; Table 2b). Overall males seemed more structured than females between oceans, suggesting that females disperse farther, but genetic signal for sex-biased dispersal varied geographically: female-biased in the Indian Ocean, male-biased or inconclusive in the Atlantic Ocean.

### 3.3 Reconstructing scenario of breeding site colonization

A split between ancestral Atlantic and Indian populations (Fig. 2a) occurring 1.76 My ago (95% CI range 0.99-2.60) was inferred based on all 9 gene trees (2.71 My ago (1.17-4.72) using only mitochondrial markers). West and east Atlantic ancestral populations split around 1.38 My ago (0.78-2.04), *baroli* and *boydi* split at 0.85 My ago (0.44-1.32), and *nicolae* and *bailloni* at 0.70 My ago (0.33-1.13). The \*BEAST analysis based on all nine markers showed the same topology though with generally lower divergence times and higher confidence intervals (Supplementary Material 8). ABC analyses also supported a similar scenario of ancestral population divergence: best retained topologies using mtDNA markers and all nine markers suggested, starting from oldest to newest splits, *nicolae* being ancestral, leading to the appearance of *boydi* (Fig 2d, Supplementary Material 5). Then *therminieri* diverged from *boydi*, and *baroli* diverged from *boydi*. Finally, *bailloni* diverged from *nicolae* (Fig. 2d Supplementary Material 4). Our phylogenetic trees placed the Central Pacific taxon *dichrous* within the Indian clade, thus supporting the putative scenario of Atlantic lineages diversifying from Indian Ocean rather than from Pacific ancestors (Fig. 2b,c). Within *therminieri*, ABC analyses suggested a northerly stepping stone colonization process, from Martinique to the Bahamas (Supplementary Material 5). Similarly, the most likely scenario of population divergence within *baroli* was colonization from the Canaries to the more northerly Azores.

Finally, Bayesian Skyline analyses inferred current effective population sizes to be around  $10^4$  individuals (see Supplementary Material 10). Mean  $N_e$  seemed to increase slowly over time, but high confidence intervals precluded detection of any sudden change in population sizes. Confidence intervals on current population sizes were also very large, in particular for the individuals breeding on Réunion (see Supplementary Material 10). The demographic parameter estimations were consistent with a constant population size over time for each lineage, as suggested from Fu's  $F_s$  results (which are more suitable than Tajima's  $D$  to estimate population expansion; Ramirez-Soriano et al. 2008; Supplementary Material 7).

## Discussion

### 4.1 Mito-nuclear discordance and sex-biased dispersal

At the inter-lineage scale, we observed more genetic structure at mitochondrial than at nuclear loci. This dissimilarity has been observed for numerous Procellariiformes species (e.g. Gangloff et al. 2013, Silva et al.

2015, Welch et al. 2011) as well as other organisms (see Toews & Brelsford 2012 for a review). One likely explanation is a difference in the pace of molecular evolution between mitochondrial and nuclear markers, with the latter having a slower substitution rate than the former, and the former being more polymorphic at the intra-specific scale (see Brown 1985), at least in birds (Helm-Bychowski 1984; Mindel et al. 1996). However, we did observe high levels of intra-lineage diversity for some nuclear loci such as  $\beta\text{globin}$  (see also Gangloff et al. 2013, Silva et al. 2015), and therefore the difference of structuration cannot be solely attributed to a difference of marker variability. We therefore suspect that incomplete lineage sorting and retention of ancestral polymorphisms at nuclear loci also contribute. Indeed, effective population size of mitochondrial DNA is four times smaller than that of nuclear DNA due to uniparental inheritance. Lineage sorting will therefore be faster in mtDNA than in nuDNA, being inversely proportional to the effective population size (Funk & Omland 2003). Incomplete lineage sorting is actually thought to be the main cause of mito-nuclear discordance when associated to a pattern of loss of geographic differentiation on nuclear markers (McKay & Zink 2010; Toews & Brelsford 2012). We also found patterns suggestive of introgression in nuclear markers. Hybridisation with introgression has been documented in shearwaters (Genovart et al. 2007, Gómez-Díaz et al. 2009), other Procellariiformes (Brown et al. 2010,) and other seabirds (Gay et al. 2009; Morris-Pocock et al. 2011; Pons et al. 2014). The likelihood of Indian Ocean petrels visiting breeding Atlantic Petrels may be supported by recent tracking of *Pterodroma arminjoniana* breeding on Round Island (Mauritius), which showed that some individuals foraged around South Trinidad Is, off Brazil, and even in the northern Atlantic (Booth Jones et al. 2017), although flight capacities of *Pterodroma* are far higher than *Puffinus*. Introgression can also blur phylogeographic signals by mixing alleles from distinct populations, and is considered as the second main cause of mito-nuclear discordance (McKay & Zink 2010). Incomplete lineage sorting and introgression are however difficult to distinguish, and additional unlinked markers would be required to disentangle these phenomena. Finally, as the mitochondrial markers represent only the female evolutionary history, sex-biased dispersal favoring females may alternatively explain why the population structure inferred based on nuclear markers conflicts with female-inherited mtDNA markers (see Petit & Excoffier (2009). These authors suggested that the markers associated with the most dispersing sex should better delimitate species, as they will show stronger intra-specific gene flow from colonizing lineages, reducing the effects of genetic drift and lowering the probability of fixating introgressed alleles. Dispersal was indeed stronger in females in some populations, particularly in the larger and the putatively ancestral lineage, *nicolae*. Sex-biased dispersal was however more uncertain for *therminieri* and *boydi*, while for *baroli* dispersal was inferred to be male-biased. The sample size for *baroli* was theoretically large enough to robustly detect a bias in  $F_{ST}$ ,  $F_{IS}$ , and Relatedness (Goudet et al. 2002). Sex-biased dispersal may therefore have further contributed to the observed mito-nuclear discordance, at least in some lineages.

## 4.2 Sequencing artifacts due to mtDNA duplication and uncertainties about molecular clock rates

The mitochondrial genome of several Procellariiformes presents tandem repeats (Abbott et al. 2005; Lounsbury et al. 2015), including *P. therminieri* (Torres et al. 2018). In this taxon, a duplicated region comprising two copies of the CR was found, yielding the presence of double peaks on chromatograms. By treating gDNA with an exonuclease (which effectively removed all linear DNA), we did not observe triple- or quadruple-peaks on chromatograms, and thus hypothesized that only two copies of the CR sequences were amplified by PCR. Therefore we considered two haplotypic phases for all CR sequences, assuming that removing all CR sequences would have led to a loss of information and of statistical power in the analyses. To check the robustness of this approach however, we replicated all our analyses using two other data subsets: one in which all ambiguous CR sequences were removed, and another one from which all CR sequences were removed. Removing all CR sequences led to a strong loss of information, an increase of the estimated differentiation and a decrease of the estimated divergence times (Supplementary Material 6). Removing only the ambiguous sequences led to estimations close to the estimations of the complete dataset. This suggests that the noise caused by the multiple copies of CR was swamped by the signal contained in that marker and so our analyses using all the individuals are not significantly biased by these sequences.

Another major issue concerns the choice of a molecular clock rate for dating the splitting events. We found



a clear hierarchical structuration of populations in the Atlantic and Indo-Pacific Oceans, which diverged around 1.76 to 2.71 My ago either using all markers or only mtDNA markers, respectively. In contrast, using the same taxa (but fewer specimens and only *cytb*), Austin et al. (2004) dated this split at 3.2-3.8 My, and rather suggested that the closure of the Isthmus of Panama erected a barrier to gene flow between Indo-Pacific and Atlantic populations. This difference in divergence times might be due to taxon sampling, gene sampling, or the calibration of the molecular clock (0.631%/My here, vs. 0.9%/My in Austin et al). This latter parameter is probably a major contributor to the difference observed between our studies, since using the same value as Austin et al, and only sequences from *cytb*, we found a divergence time of 2.2-3.8 My, thus very close to Austin et al. Many substitution rates for both Mt DNA markers were proposed for petrels, ranging from 0.19%/My (Pacheco et al. 2011) to 1.544%/My (Pereira & Baker 2006) for *cox1*, and from 0.18%/My (Pacheco et al. 2011) to 0.88-0.92%/My (Nunn & Stanley 1998; the latter rate was used by Austin et al), 1.022%/My (Pereira & Baker 2006) and 1.89%/My (Weir & Schluter 2008) for *cytb*. Using lowest rates (Pacheco et al. 2011) resulted in a divergence time between Atlantic and Indo-Pacific lineages around 12.8 My ago (Supplementary Material 10), an unlikely value given the presence of an ocean between the Americas before the Isthmus of Panama erected, despite the fact that *Puffinus* is dating back to the Oligocene (Henderson & Gill 2010). Here we used the substitution rates that were estimated in Nabholz et al. 2016, as the most recent review. There are two further arguments against Austin et al. (2004) scenario: first, the closure of the Isthmus of Panama has been actually dated at 2.8 My ago (O'Dea et al. 2016), thus later than Austin et al's scenario. Second, we found that the pacific taxon, *dichrous*, was not ancestral to the Atlantic taxon, but was embedded within Indian ocean taxa. The use of precise fossil calibration would bring a more robust estimation of divergence times and so a supplemental evidence of the origin of the colonization.

### 4.3 Inferring key drivers of diversification in the small *Puffinus*

Based on our inferred date of the Indian –Atlantic lineage split, gene exchange between Indian and Atlantic birds have apparently occurred after the closure of the Isthmus of Panama: we suggest that this happened through individuals going around South Africa, since African continent is an insurmountable barrier for Procellariiformes (Silva et al. 2015). Indeed, these shearwaters are tropical or subtropical species (at least currently). Off South Africa, until 1 My ago, sea surface temperatures (SST) were approximately 2°C higher than today (Bell et al. 2015), suggesting that migration between Atlantic and Indian oceans may have remained possible for such birds. From 2.0, a strong decrease of SST occurred in both oceans (Bell et al. 2015), and gene flow between Indian and Atlantic may thus have ceased, in agreement with our estimated time of divergence (1.76 My ago). Once the Atlantic birds were isolated from Indo-Pacific populations, differentiation started to occur among Atlantic lineages 1.38 to 1.90 My ago (respectively for the 9 markers or only the mtDNA markers). This period also corresponds to a further decrease of the SST in the Atlantic (Bell et al. 2015), a southward shift of the subtropical front and warmer waters in the Southern Ocean (Maiorano et al. 2009), and sea ice development in the North extending southwards from the Arctic to the current Great lakes in the USA (Webb & Bartlein 1992). Cold temperatures, preventing the colonization of potential northern breeding sites such as the Azores, may have forced shearwaters to spread over the two sides of the Atlantic (Fig. 2d). The third step in the colonization of Atlantic breeding grounds concerns the divergence of *baroli* from *boydi*, which would have occurred around 0.85-1.26 My ago, a period that corresponds to a stabilization of the SST at the current level in the North Atlantic. The ice melt may then have allowed northward colonization on both sides of the Atlantic, from Cape Verde to the Canaries, Madeira and the Azores, and from the lesser Antilles to the Bahamas and Bermuda. Similar timing of divergence has been suggested between *Calonectris edwardsii* (Cape Verde) and *C. diomedea* (North Atlantic and Mediterranean) 0.7-0.9 My ago (Gómez-Díaz et al. 2006) and *Puffinus olsoni* (Canaries) and *Puffinus puffinus* (North Atlantic) 0.2-1.0 My (Ramirez et al. 2010). In the Indian Ocean, a reversed pattern of southward colonization occurred, shearwaters colonizing from Seychelles, partly continental in origin, to Réunion 0.70-1.01 My ago, precisely at the time strong volcanic activity ended on Réunion (Gillot & Nativel 1989). Mauritius was probably colonized long before (Mauritius age: 8My; McDougall & Chamalaun 1969), then Rodrigues and Réunion (both about the same age, 2 My; McDougall 1971), but shearwaters are now

extinct on Mauritius and Rodrigues, following human colonization since the 17th century, so no sample is available. The southward movement was however not necessarily related to changes in SST, but rather to availability of volcanic islands that eventually emerged in a southward direction.

Mantel tests showed a strong correlation between geographic and genetic distance between lineages, i.e. at large scale. Indeed Little shearwaters are poor flyers compared to other Procellariiformes. However, isolation-by-distance within each lineage was visually not detected, therefore distance alone could not be a factor of population divergence at this smaller scale; we argue that sea temperature could rather be the most important factor of divergence in our case at this scale. Seabirds indeed depend on both sea and the islands where they breed, thus sea temperature has strong impacts on seabird phenology, breeding, survival and abundance (see Sydeman et al. 2012 for a review). We suggest that foraging ecology, which strongly depends on SST, is an important process shaping divergence among lineages. The segregation of foraging areas among populations is an important factor of differentiation among seabirds (Friesen 2015; Friesen et al. 2007), as shown for other shearwaters (Genovart et al. 2007; Gómez-Díaz et al. 2006), petrels (Gangloff et al. 2013; Welch et al. 2012), storm-petrels (Deane 2013; Smith et al. 2007) and albatrosses (Alderman et al. 2005; Burg & Croxall 2001). Assessment of non-breeding and breeding distributions at sea of the little shearwaters complex revealed that all three Atlantic taxa show rather separated foraging and wintering areas (Ramos et al. in review), and further suggest that *boydi* rather than *lherminieri* was ancestral in the North Atlantic. Indeed, *boydi* is more flexible in its foraging ecological niche, suggesting ancestral behavior, with a far larger potential distribution at sea covering all central North Atlantic (Zajková et al. 2017, see map in Ramos et al. in review). In addition, over the last My, SST oscillations gradually increased in amplitude with up to five degree difference in range (Bell et al. 2015; Herbert et al. 2011). These oscillations were showed to correlate with changes in marine productivity (Martí et al. 2009), prey species diversity (Yasuhara & Cronin 2008), atmospheric circulation (Chang et al. 2000) and sea level (e.g. in Atlantic Nascimento et al. 2011, Zazo et al. 2010). It is likely that these oscillations also contributed to divergence in the North Atlantic, and we suggest that significant Tajima's D tests found in almost half of the populations studied represent traces of past bottlenecks and population expansions as it has already been shown in other taxa (Ramakrishnan et al. 2005; Weber et al. 2004; Zhu et al. 2006).

A last interesting side effect of our study is that it suggests that the small black and white shearwaters have shown a very recent radiation speciation event, with not less than 13 species radiating in just 1.46 million years since *P. puffinus*, *P. assimilis* and *P. newelli* clades are either embedded in *lherminieri* or *orbailoni* clades. All these species are rather coastal shearwaters (compared to more pelagic species such as the larger shearwaters), and such high rate of speciation may be the result of the high climatic oscillation that occurred over the last 2 million years which may have favored high rates of colonization and extinction on coastal islands.

#### 4.4 Sea surface temperature as a major diversification driver in marine organisms?

Conversely to pelagic fishes that show no or low inter-ocean structure (Díaz-Jaimes et al. 2010, Ely et al. 2005) and no intra-ocean structure (Nomura et al. 2014, Taguchi et al. 2015), sea mammal's gene flow is shaped by sea temperature which drives structuration between ocean basins and among breeding areas (Alexander et al. 2016, Jackson et al. 2014, Richard et al. 2018, Fontaine et al. 2014, Viricel & Rosel 2014). SST plays also a role in the diversification of sea turtles, since only cold-adapted species are able to exchange genes among oceans (Dutton et al. 1999). Finally, organisms with a pelagic larval phase show globally low structuration (Kelly & Palumbi 2010) but when detected, structuration is often linked to sea temperature (Benestan et al. 2015; Teske et al. 2005, 2018). Therefore, SST appears as a generic driver of diversification in marine organisms, though their patterns of structuration are generally considerably weaker (see Bowen et al. 2016 for a review) than those found here. We suggest that the reason for this discrepancy lies in the fact that these seabirds are central place foragers, i.e. they still depend on terrestrial habitats for breeding, the latter being impacted for instance by glaciations. They are therefore highly sensitive to any latitudinal change of SST in comparison to island distribution, which acts as a constraint since an optimal area in regard to SST may lack any island for breeding. Indeed, in marine organisms with low dispersal abilities, patterns of

structuration and divergence are more similar to the patterns found here on shearwaters. For instance, timing of population divergence between Atlantic and Indian Ocean lineages and within the Atlantic in the seahorse *Hippocampus* 'kuda complex' (Floeter et al. 2007) fits to our estimates. This species has no planktonic larval duration (Lourie et al. 2005) and long dispersal events are considered rare and implying a few individuals (Teske et al. 2005). Moreover, Cape Agulhas is known to be a phylogeographic break among several costal species, due to the difference of currents and sea temperatures between the two oceans (review in Teske et al. 2011). On the other extreme, seabirds do not compare either to terrestrial organisms living on islands, despite being highly philopatric, simply because they can disperse easily, if an island is available to being colonised.

In this small shearwater complex, geographical barriers and/or isolation by distance may have been a major driver of differentiation at large scale (typically, between Oceans) while SST has been a more important driver at smaller scale (within Oceans), with shearwaters shifting their breeding latitudes with a changing SST. However, since these seabirds depend on the geographical distribution of their breeding islands and because these seabirds are not the best seabird fliers, this distribution becomes a major constraint resulting in the present geographical structure, promoting local adaptation to small scale ecological constraints and reducing gene flow. Therefore, petrels and shearwaters present an interesting case study where diversification processes rely more (or at least equally) on ecological factors, in particular sea surface temperature, rather than distance or continental barriers, in contrast to either "true" marine organisms or terrestrial organisms. Strict marine organisms can disperse far more, or alternatively are unconstrained by island distribution, and thus show much less geographical structure within taxa. The terrestrial organisms tend to disperse far less, and isolation-by-distance tends to be a main driver of population differentiation (Meirmans 2012; Vekemans & Hardy 2004). Indeed terrestrial organisms, such as lizards or birds in Macaronesia (Almalki et al. 2017; Brehm et al. 2003), geckos in Cape Verde (Arnold et al. 2008) or birds in America (Patel et al. 2011), have revealed strong splits between islands with no shared haplotypes for the same mitochondrial markers.

## Acknowledgments

We thank the University of La Rochelle (ULR) for funding LT PHD, and the ULR and CEBC molecular core facilities for lab support. We also thank the Parc Naturel Régional de la Martinique and the DEAL Martinique for help and funding of the research program on shearwaters. Blood samples of Audubon's shearwaters of St Barth were obtained with the agreement of Agence Territoriale de l'Environnement de Saint-Barthélemy. We are also grateful to the Direção Nacional do Ambiente of Cape Verde, the Cabildo de Lanzarote and Gobierno de Canarias of Spain, and the Direcção Regional do Ambiente from Azores of Portugal for permission and logistical support in conducting fieldwork in the respective countries and to the Spanish Government (CGL2009-11278/BOS, CGL2013-42585-P, CGL2016-78530-R) and Fondos FEDER for funding. We warmly thank Pierre-André Crochet, Sarah Samadi and Joan Ferrer-Obiol for great comments on previous draft that very much improved our manuscript.

## Literature cited

- Abbott CL, Double MC, Trueman JWH, Robinson A, Cockburn A. 2005. An unusual source of apparent mitochondrial heteroplasmy: Duplicate mitochondrial control regions in *Thalassarche* albatrosses. *Mol. Ecol.* 14(11):3605–13
- Alderman R, Double MC, Valencia J, Gales RP. 2005. Genetic affinities of newly sampled populations of Wandering and Black-browed Albatross. *Emu* . 105:169–79
- Alexander A, Steel D, Hoekzema K, Mesnick SL, Engelhaupt D, et al. 2016. What influences the worldwide genetic structure of sperm whales (*Physeter macrocephalus*)? *Mol. Ecol.* 25(12):2754–72
- Almalki M, Kupán K, Carmona-Isunza MC, López P, Veiga A, et al. 2017. Morphological and Genetic Differentiation Among Kentish Plover *Charadrius alexandrinus* Populations in Macaronesia. *Ardeola* . 64(1):3–16
- Arnold EN, Vasconcelos R, Harris DJ, Mateo JA, Carranza S. 2008. Systematics, biogeography and evolution

- of the endemic *Hemidactylus* geckos (Reptilia, Squamata, Gekkonidae) of the Cape Verde Islands: Based on morphology and mitochondrial and nuclear DNA sequences. *Zool. Scr.* 37(6):619–36
- Baker AJ, Huynen LJ, Haddrath O, Millar CD, Lambert DM. 2005. Reconstructing the tempo and mode of evolution in an extinct clade of birds with ancient DNA : The giant moas of New Zealand. *PNAS* . 102(23):8257–62
- Beaumont MA, Zhang W, Balding DJ. 2002. Approximate Bayesian computation in population genetics. *Genetics* . 162(4):2025–35
- Belkhir K, Porsa P, Chikhi L, Raufaste N, Bonhomme F. 1996. *GENETIX 4.05, Logiciel Sous Windows TM Pour La Genetique Des Populations.*
- Bell DB, Jung SJA, Kroon D. 2015. The Plio-Pleistocene development of Atlantic deep-water circulation and its influence on climate trends. *Quat. Sci. Rev.* 123:265–82
- Benestan L, Gosselin T, Perrier C, Sainte-Marie B, Rochette R, Bernatchez L. 2015. RAD genotyping reveals fine-scale genetic structuring and provides powerful population assignment in a widely distributed marine species, the American lobster (*Homarus americanus*). *Mol. Ecol.* 24(13):3299–3315
- Bouckaert R, Heled J, Kuhnert D, Vaughan T, Wu C-H, et al. 2014. BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. *PLoS Comput. Biol.* 10(4):e1003537
- Bowen BW, Gaither MR, DiBattista JD, Iacchei M, Andrews KR, et al. 2016. Comparative phylogeography of the ocean planet. *Proc. Natl. Acad. Sci.* 113(29):7962–69
- Brehm A, Jesus J, Spinola H, Alves C, Vicente L, Harris DJ. 2003. Phylogeography of the Madeiran endemic lizard *Lacerta dugesii* inferred from mtDNA sequences. *Mol. Phylogenet. Evol.* 26(2):222–30
- Brown RM, Nichols RA, Faulkes CG, Jones CG, Bugoni L, et al. 2010. Range expansion and hybridization in Round Island petrels (*Pterodroma* spp.): evidence from microsatellite genotypes. *Mol. Ecol.* 19(15):3157–70
- Brown WM. 1985. The mitochondrial genome of animals. In *Molecular Evolutionary Genetics*
- Burg TM, Croxall JP. 2001. Global relationships amongst black-browed and grey-headed albatrosses: analysis of population structure using mitochondrial DNA and microsatellites. *Mol. Ecol.* 10:2647–60
- Chang P, Saravan R, Ji L, Heger GC. 2000. The Effect of Local Sea Surface Temperatures on Atmospheric Circulation over the Tropical Atlantic Sector. *J. Clim.* (1985):2195–2216
- Cornuet JM, Pudlo P, Veyssier J, Dehne-Garcia A, Gautier M, et al. 2014. DIYABC v2.0: A software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. *Bioinformatics* . 30(8):1187–89
- Deane P. 2013. *What traits predispose the Band-rumped Storm-petrel, Oceanodroma castro, to ecological speciation in the absence of physical barriers to gene flow?*
- Diaz-Jaimes P, Uribe-Alcocer M, Rocha-Olivares A, Garcia-de-Leon FJ, Nortmoon P, Durand JD. 2010. Global phylogeography of the dolphinfish (*Coryphaena hippurus*): The influence of large effective population size and recent dispersal on the divergence of a marine pelagic cosmopolitan species. *Mol. Phylogenet. Evol.* 57(3):1209–18
- Dutton PH, Bowen BW, Owens DW, Barragan A, Davis SK. 1999. Global phylogeography of the leatherback turtle (*Dermochelys coriacea*). *J. Zool.* 248(3):397–409
- Ely B, Vinas J, Alvarado Bremer JR, Black D, Lucas L, et al. 2005. Consequences of the historical demography on the global population structure of two highly migratory cosmopolitan marine fishes: The yellowfin tuna (*Thunnus albacares*) and the skipjack tuna (*Katsuwonus pelamis*). *BMC Evol. Biol.* 5:1–9

- Estoup A, Lombaert E, Marin J-M, Guillemaud T, Puldo P, et al. 2012. Estimation of demo-genetic model probabilities with Approximate Bayesian Computation using linear discriminant analysis on summary statistics. *Mol. Ecol. Resour.* 12(5):846–55
- Excoffier L, Laval G, Schneider S. 2005. Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evol. Bioinforma.* 1:47–50
- Floeter SR, Rocha L., Roberston DR, Joyeux JC, Smith-Vaniz WF, et al. 2007. Atlantic reef fish biogeography and evolution. *J. Biogeogr.* 35(1):
- Fontaine MC, Roland K, Calves I, Austerlitz F, Palstra FP, et al. 2014. Postglacial climate changes and rise of three ecotypes of harbour porpoises, *Phocoena phocoena*, in western Palearctic waters. *Mol. Ecol.* 23(13):3306–21
- Fridolfsson A-K, Ellegren H. 1999. A simple and universal method for molecular sexing of non-ratite birds. *J. avian Biol.* 116–21
- Friesen VL. 2015. Speciation in seabirds: why are there so many species...and why aren't there more? *J. Ornithol.*
- Friesen VL, Burg TM, McCOY KD. 2007. Mechanisms of population differentiation in seabirds. *Mol. Ecol.* 16(9):1765–85
- Fu YX. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* . 147(2):915–25
- Funk DJ, Omland KE. 2003. Species-Level Paraphyly and Polyphyly: Frequency, Causes, and Consequences, with Insights from Animal Mitochondrial DNA. *Annu. Rev. Ecol. Evol. Syst.* 34(1):397–423
- Gangloff B, Zino F, Shirihai H, Gonzalez-Solis J, Couloux A, et al. 2013. The evolution of north-east Atlantic gadfly petrels using statistical phylogeography. *Mol. Ecol.* 22(January):495–507
- Gay L, Neubauer G, Zagalska-Neubauer M, Pons JM, Bell DA, Crochet PA. 2009. Speciation with gene flow in the large white-headed gulls: Does selection counterbalance introgression? *Heredity (Edinb)*. 102(2):133–46
- Genovart M, Oro D, Juste J, Bertorelle G. 2007. What genetics tell us about the conservation of the critically endangered Balearic Shearwater? *Biol. Conserv.* 137(2):283–93
- Gillot P-Y, Nativel P. 1989. Eruptive history of the Piton de la Fournaise volcano, Reunion Island, Indian Ocean. *J. Volcanol. Geotherm. Res.* 36(1–3):53–65
- Gomez-Diaz E, Gonzalez-Solis J, Peinado MA. 2009. Population structure in a highly pelagic seabird, the Cory's shearwater *Calonectris diomedea*: an examination of genetics, morphology and ecology. *Mar. Ecol. Prog. Ser.* 382(Palumbi 1994):197–209
- Gomez-Diaz E, Gonzalez-Solis J, Peinado MA, Page RDM. 2006. Phylogeography of the *Calonectris* shearwaters using molecular and morphometric data. *Mol. Phylogenet. Evol.* 41(2):322–32
- Goudet J, Perrin N, Waser P. 2002. Tests for sex-biased dispersal using bi-parentally inherited genetic markers. *Mol. Ecol.* 11(6):1103–14
- Groth JG, Barrowclough GF. 1999. Basal Divergences in Birds and the Phylogenetic Utility of the Nuclear RAG-1 Gene. *Mol. Phylogenet. Evol.* 12(2):115–23
- Helm-Bychowski KM. 1984. *Evolution of Nuclear and Mitochondrial DNA in Gallinaceous Birds*
- Henderson N, Gill BJ. 2010. A mid-Pliocene shearwater skull ( Aves : Procellariidae : Puffinus ) from the Taihape Mudstone , central North Island , New Zealand. *New Zeal. J. Geol. Geophys.* 53(4):
- Herbert TD, Years PM, Herbert TD, Peterson LC, Lawrence KT, Liu Z. 2011. Tropical Ocean Temperatures Over the Past 3.5 Million Years. . 1530(2010):

- Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* 23(2):254–67
- Jackson JA, Steel DJ, Beerli P, Congdon BC, Olavarria C, et al. 2014. Global diversity and oceanic divergence of humpback whales (Megaptera novaeangliae). *Proc. R. Soc. B Biol. Sci.*281(1786):20133222–20133222
- Jayaprakash AD, Benson EK, Gone S, Liang R, Shim J, et al. 2015. Stable heteroplasmy at the single-cell level is facilitated by intercellular exchange of mtDNA. *Nucleic Acids Res.* 43(4):2177–87
- Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30(14):3059–66
- Kelly RP, Palumbi SR. 2010. Genetic structure among 50 species of the northeastern pacific rocky intertidal community. *PLoS One* . 5(1):
- Kimball RT, Braun EL, Ligon JD. 1997. Resolution of the phylogenetic position of the Congo peafowl , *Afropavo congensis* : a biogeographic and evolutionary enigma. *Proc. R. Soc. London B Biol. Sci.*264(1387):1517–23
- Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* 33(7):1870–74
- Librado P, Rozas J. 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* . 25(11):1451–52
- Lounsbury ZT, Brown SK, Collins PW, Henry RW, Newsome SD, Sacks BN. 2015. Next-generation sequencing workflow for assembly of nonmodel mitogenomes exemplified with North Pacific albatrosses (Phoebastria spp.). *Mol. Ecol. Resour.* 15(4):893–902
- Lourie SA, Green DM, Vincent ACJ. 2005. Dispersal, habitat differences, and comparative phylogeography of Southeast Asian seahorses (Syngnathidae: Hippocampus). *Mol. Ecol.* 14(4):1073–94
- Maiorano P, Marino M, Flores JA. 2009. The warm interglacial Marine Isotope Stage 31: Evidences from the calcareous nannofossil assemblages at Site 1090 (Southern Ocean). *Mar. Micropaleontol.*71(3–4):166–75
- Martinez-GA, Rosell-mele A, Geibert W, Gersonde R. 2009. Links between iron supply , marine productivity , sea surface temperature , and CO<sub>2</sub> over the last 1 . 1 Ma. *Paleoceanography* . 24:1–14
- McDougall I. 1971. The geochronology and evolution of the young volcanic island of Réunion, Indian Ocean. *Geochim. Cosmochim. Acta* . 35(3):261–88
- McDougall IAN, Chamalaun FH. 1969. Isotopic dating and geomagnetic polarity studies on volcanic rocks from Mauritius, Indian Ocean. *Geol. Soc. Am. Bull.* 80(8):1419–42
- McKay BD, Zink RM. 2010. The causes of mitochondrial DNA gene tree parphyly in birds. *Mol. Phylogenet. Evol.* 54(2):647–50
- Meirmans PG. 2012. The trouble with isolation by distance. *Mol. Ecol.* 21:2839–46
- Mindel DP, Knight A, Baer C, Huddleston CJ. 1996. Slow Rates of Molecular Evolution in Birds and the Metabolic Rate and Body Temperature Hypotheses. *Mol. Biol. Evol.* (Mol Biol Evol) . 13(2):422
- Morris-Pocock J a., Anderson DJ, Friesen VL. 2011. Mechanisms of global diversification in the brown booby (*Sula leucogaster*) revealed by uniting statistical phylogeographic and multilocus phylogenetic methods. *Mol. Ecol.* 20(13):2835–50
- Nabholz B, Lanfear R, Fuchs J. 2016. Body mass-corrected molecular rate for bird mitochondrial DNA. *Mol. Ecol.* 25(18):4438–49

- Nascimento L De, Delgado JD, Garcı E, Whittaker RJ. 2011. A reconstruction of Palaeo-Macaronesia , with particular reference to the long-term biogeography of the Atlantic island laurel forests. . 226–46
- Nomura S, Kobayashi T, Agawa Y, Margulies D, Scholey V, et al. 2014. Genetic population structure of the Pacific bluefin tuna *Thunnus orientalis* and the yellowfin tuna *Thunnus albacares* in the North Pacific Ocean. *Fish. Sci.* 80(6):1193–1204
- Nunn GB, Stanley SE. 1998. Body size effects and rates of cytochrome b evolution in tube-nose(d seabirds. *Mol. Biol. Evol.* 15(10):1360–71
- O’Dea A, Lessios HA, Coates AG, Eytan RI, Restrepo-Moreno SA, et al. 2016. Formation of the Isthmus of Panama. *Sci. Adv.* 2(8):1–12
- Onley D, Scofield RP. 2007. *Albatrosses, Petrels and Shearwaters of the World*. Princeton field guides.
- Pacheco MA, Battistuzzi FU, Lentino M, Aguilar RF, Kumar S, Escalante AA. 2011. Evolution of modern birds revealed by mitogenomics: Timing the radiation and origin of major orders. *Mol. Biol. Evol.* 28(6):1927–42
- Patel S, Weckstein JD, Patané JSL, Bates JM, Aleixo A. 2011. Temporal and spatial diversification of *Pteroglossus arazari* (AVES: Ramphastidae) in the neotropics: Constant rate of diversification does not support an increase in radiation during the Pleistocene. *Mol. Phylogenet. Evol.* 58(1):105–15
- Pereira SL, Baker AJ. 2006. A mitogenomic timescale for birds detects variable phylogenetic rates of molecular evolution and refutes the standard molecular clock. *Mol. Biol. Evol.* 23(9):1731–40
- Petit RJ, Excoffier L. 2009. Gene flow and species delimitation. *Trends Ecol. Evol.* 24(7):386–93
- Pons JM, Sonsthagen S, Dove C, Crochet PA. 2014. Extensive mitochondrial introgression in North American Great Black-backed Gulls (*Larus marinus*) from the American Herring Gull (*Larus smithsonianus*) with little nuclear DNA impact. *Heredity (Edinb)*. 112(3):226–39
- Precheur C, Barbraud C, Martail F, Mian M, Nicolas J, et al. 2016. Some like it hot : effect of environment on population dynamics of a small tropical seabird in the Caribbean region. *Ecosphere* . 7(10):1–18
- Prychitko TM, Moore WS. 1997. Comparative Evolution of the Mitochondrial Cytochrome b Gene and Nuclear -Fibrinogen Intron 7 in Woodpeckers. *Mol. Biol. Evol.* 17(7):1101–11
- Ramakrishnan U, Hadly EA, L MJ. 2005. Detecting past population bottlenecks using temporal genetic data. *Mol. Ecol.* 14:2915–22
- Ramirez O, Illera JC, Rando JC, Gonzalez-Solis J, Alcover JA, Lalueza-Fox C. 2010. Ancient DNA of the Extinct Lava Shearwater (*Puffinus olsoni*) from the Canary Islands Reveals Incipient Differentiation within the *P. puffinus* Complex. *PLoS One* . 5(12):e16072
- Richard G, V Titova O, D. Fedutin I, Steel D, Meschersky I, et al. 2018. Cultural Transmission of Fine-Scale Fidelity to Feeding Sites May Shape Humpback Whale Genetic Diversity in Russian Pacific Waters. *J. Hered.* 1–11
- Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, et al. 2012. Mrbayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61(3):539–42
- Ruokonen M, Kvist L. 2002. Structure and evolution of the avian mitochondrial control region. *Mol. Phylogenet. Evol.* 23:422–32
- Scherer RP, Bohaty SM, Dunbar RB, Esper O, Flores JA, et al. 2008. Antarctic records of precession-paced insolation-driven warming during early Pleistocene Marine Isotope Stage 31. *Geophys. Res. Lett.* 35(3):1–5
- Silva MC, Duarte M a, Coelho MM. 2011. Anonymous nuclear loci in the white-faced storm-petrel *Pelagodroma marina* and their applicability to other Procellariiform seabirds. *J. Hered.* 102(3):362–65

- Silva MC, Matias R, Wanless RM, Ryan PG, Stephenson BM, et al. 2015. Understanding the mechanisms of antitropical divergence in the seabird White-faced Storm-petrel (Procellariiformes: Pelagodroma marina) using a multilocus approach. *Mol. Ecol.* 24(12):3122–37
- Smith AL, Monteiro L, Hasegawa O, Friesen VL. 2007. Global phylogeography of the band-rumped storm-petrel (*Oceanodroma castro*; Procellariiformes: Hydrobatidae). *Mol. Phylogenet. Evol.* 43(3):755–73
- Smouse PE, Long JC, Sokal RR. 1986. Multiple Regression and Correlation Extensions of the Mantel Test of Matrix Correspondence. *Syst. Zool.* 35(4):627
- Sorenson MD, Quinn TW. 1998. Numts: a challenge for avian systematics and population biology. *Auk* . 115:214–21
- Stephens M, Smith NJ, Donnelly P. 2001. A New Statistical Method for Haplotype Reconstruction from Population Data. *Am. J. Hum. Genet.* 68(4):978–89
- Taguchi M, King JR, Wetklo M, Withler RE, Yokawa K. 2015. Population genetic structure and demographic history of Pacific blue sharks (*Prionace glauca*) inferred from mitochondrial DNA analysis. *Mar. Freshw. Res.* 66(3):267–75
- Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* . 123(3):585–95
- Teske PR, Hamilton H, Palsboll PJ, Choo CK, Gabr H, et al. 2005. Teske et al 2005 Long distance dispersal. . 286:249–60
- Teske PR, Sandoval-Castillo J, Golla TR, Emami-Khoyi A, Tine M, et al. 2018. Thermal selection drives biodiversity origination across the Atlantic/Indian Ocean boundary. *bioRxiv* . 1–10
- Teske PR, Von der Heyden S, McQuaid CD, Barker NP. 2011. A review of marine phylogeography in southern Africa. *S. Afr. J. Sci.* 107(5/6):1–11
- Toews DPL, Brelsford A. 2012. The biogeography of mitochondrial and nuclear discordance in animals. *Mol. Ecol.* 21(16):3907–30
- Torres L, Welch AJ, Zanchetta C, Chessier RT, Manno M, et al. 2018. Evidence for a duplicated mitochondrial region in Audubon’s shearwater based on MinION sequencing. *Mitochondrial DNA Part A* . 1–8
- Van Loenen AL. 2013. *Reconstructing the Genetic Legacy of Cook’s Petrels ( Pterodroma cookii )*
- Vekemans X, Hardy O. 2004. New insights from fine-scale spatial genetic structure analyses in plant populations. *Mol. Ecol.* 13:921–35
- Viricel A, Rosel PE. 2014. Hierarchical population structure and habitat differences in a highly mobile marine species: the Atlantic spotted dolphin. *Mol. Ecol.* 23(20):
- Wang J. 2007. Triadic IBD coefficients and applications to estimating pairwise relatedness. *Genet. Res.* 89(3):135–53
- Wang J. 2011. Coancestry: A program for simulating, estimating and analysing relatedness and inbreeding coefficients. *Mol. Ecol. Resour.* 11(1):141–45
- Webb T, Bartlein PJ. 1992. Global Changes During the Last 3 Million Years: Climatic Controls and Biotic Responses. *Annu. Rev. Ecol. Syst.* 23(1):141–73
- Weber DS, S SB, Ehman NL. 2004. Genetic Consequences of a Severe Population Bottleneck in the Guadalupe Fur Seal ( *Arctocephalus townsendi* ). *J. Hered.* 95(2):144–53
- Weir JT, Schluter D. 2008. Calibrating the avian molecular clock. *Mol. Ecol.* 17(10):2321–28



Welch AJ, Fleischer RC, James HF, Wiley AE, Ostrom PH, et al. 2012. Population divergence and gene flow in an endangered and highly mobile seabird. *Heredity (Edinb)*. 109(1):19–28

Yasuhara M, Cronin TM. 2008. CLIMATIC INFLUENCES ON DEEP-SEA OSTRACODE ( CRUSTACEA ) DIVERSITY FOR THE LAST THREE MILLION YEARS. *Ecology* . 89(11):53–65

Zazo C, Goy JL, Hillaire-marcel C, Dabrio CJ, Gonzalez-delgado JA, et al. 2010. Sea level changes during the last and present interglacials in Sal Island ( Cape Verde archipelago ). *Glob. Planet. Change* . 72(4):302–17

Zhu Q, Zheng X, Luo J, Gaut BS, Ge S. 2006. Multilocus Analysis of Nucleotide Variation of *Oryza sativa* and Its Wild Relatives : Severe Bottleneck during Domestication of Rice. *Mol. Biol. Evol.*24(3):857–88

## Data accessibility statement

DNA sequences: currently submitted to Genbank

## Author’s contributions

VB, LT and EP conceived the study. VB, LC, WM, CP, HS, JS, JT, FZ, JGS, TM and LZ collected samples. CR and LT performed the labwork. LT, EP and AV analyzed the data. LT, EP, VB and AV wrote the manuscript. All authors approved the final manuscript.

## Conflict of Interest

The authors declare no conflict of interest.

Table 1: Summary statistics of polymorphisms for the nine markers used in this study.

N is the number of sequences obtained (with numbers in parentheses referring to the sequences obtained from this study, obtained from previous studies and downloaded from Genbank); L is the length of the sequences in bp; S is the number of polymorphic segregating sites; h / a: h is the number of corresponding haplotypes for mitochondrial markers, and a is the number of alleles for nuclear markers; hd is the haplotype diversity and  $\pi$  is the nucleotide diversity. Due to the presence of ambiguities in the sequence for CR, two haplotype phases were considered here.

Marker	N	L	S	h / a	hd	$\pi$	He	Ho
<i>cox1</i>	225 (212, 10, 3)	577	18	12	0.843	0.01301	-	-
<i>cytb</i>	230 (205, 3, 22)	877	50	48	0.951	0.01394	-	-
CR	181	307	80	98	0.993	0.07497	-	-
<i>tpm</i>	184	427	2	3	-	0.00011	0.0041	0.0008
<i>irf2</i>	182	542	6	6	-	0.00048	0.0133	0.000
<i>csde</i>	223	542	15	13	-	0.0006	0.0007	0.0003
<i>pax</i>	255 (227, 28)	515	9	10	-	0.00159	0.0055	0.0008
<i>rag1</i>	162	1323	29	65	-	0.00337	0.0109	0.0026
$\beta\phi\beta$	92	1067	50	53	-	0.0083	0.0074	0.0014

Table 2: Population differentiation, according to the types of genetic markers and sex.

b.		lherminieri	lherminieri	lherminieri	boydi	boydi	baroli	baroli	bailloni
		<b>Allencay</b>	<b>Longcay</b>	<b>Martinique</b>	<b>Raso</b>	<b>Cima</b>	<b>Mclara</b>	<b>Vila</b>	<b>Nort</b>
lherminieri	<b>Allencay</b>	-	0	0.33*	0.55***	0.49***	0.5***	0.47***	0.81***
lherminieri	<b>Longcay</b>	0	-	0	0.47***	0.4**	0.42***	0.43***	0.81***
lherminieri	<b>Martinique</b>	0.45**	0.38*	-	0.2*	0.35**	0.42**	0.41***	0.71***
boydi	<b>Raso</b>	0.6***	0.48***	0.24*	-	0	0	0.29*	0.81***
boydi	<b>Cima</b>	0.49***	0.34***	0.19*	0	-	0	0	0.77*

b.		lherminieri	lherminieri	lherminieri	boydi	boydi	baroli	baroli	bailloni
baroli	<b>Mclara</b>	0.67***	0.56***	0.52**	0.34**	0	-	0	0.76**
baroli	<b>Vila</b>	0.71***	0.61***	0.69**	0.44**	0	0	-	0.77**
bailloni	<b>North Reunion</b>	0.8***	0.77***	0.86***	0.9***	0.81***	0.92***	0.94***	-
bailloni	<b>South Reunion</b>	0.69***	0.64***	0.65***	0.75***	0.69***	0.73***	0.77***	0.3*
nicolae	<b>Seychelles</b>	0.78***	0.75***	0.8***	0.85***	0.73***	0.86***	0.89***	0.67**

a.	lherminieri	lherminieri	lherminieri	lherminieri	boydi	boydi	baroli	baroli	baroli	baroli	bailloni	bailloni
	Allencay	Longcay	Martinique	St Barthélémey	Raso	Cima	Funchal	Mclara	Selvagem	Vila	North Re-union	South Re-union
lherminieri	Allencay	0	0.18*	-	0.47***	0.25***	0.45*	0.5***	0.46***	0.47***	0.69***	0.68***
lherminieri	Longcay	0.08*	-	0.11	-	0.42***	0.19***	0.45***	0.44***	0.39***	0.41***	0.69***
lherminieri	Martinique	0.22***	-	-	0.16*	-	0.38*	0.39***	0.42***	0.43***	0.78***	0.65***
lherminieri	St Barthélémey	0.19*	0.35***	-0.13	-	-0.28	-0.21	-0.11	-0.07	-0.43	-0.48	0.52***
boydi	Raso	0.81***	0.83***	0.82***	0.77***	-	-	-	0.29***	0.37*	0.35***	0.79***
boydi	Cima	0.76***	0.78***	0.74***	0.71***	0.01	-	-	0.15*	0.15*	0.11*	0.66***
baroli	Funchal	0.81***	0.84***	0.85***	0.72***	0.7***	0.56***	-	-	-	-0.3	0.78***
baroli	Mclara	0.78***	0.81***	0.82***	0.73***	0.67***	0.61***	0.08	-	0.33***	0.06	0.76***
baroli	Selvagem	0.8***	0.83***	0.83***	0.73***	0.64***	0.53***	0.36	0.10	-	0.14*	0.68***
baroli	Vila	0.85***	0.86***	0.84***	0.81***	0.69***	0.65***	0.06	-	0.03	-	0.75***
bailloni	North Re-union	0.93***	0.93***	0.89***	0.91***	0.87***	0.84***	0.89***	0.85***	0.88***	0.89***	-
bailloni	South Re-union	0.93***	0.94***	0.9***	0.92***	0.89***	0.87***	0.91***	0.88***	0.88***	0.91***	-0.16
nicolae	Seychelles	0.93***	0.93***	0.9***	0.91***	0.89***	0.87***	0.91***	0.89***	0.9***	0.92***	0.76***

a. Pairwise  $\Phi_{ST}$  values for mitochondrial markers (below diagonal) and  $F_{ST}$  for nuclear markers (above diagonal).

Border indicates the separation between intra and inter lineage comparisons. Triple band indicates the separation between intra- and inter-ocean comparisons. \*:  $p < 0.05$ ; \*\*\*:  $p < 0.001$

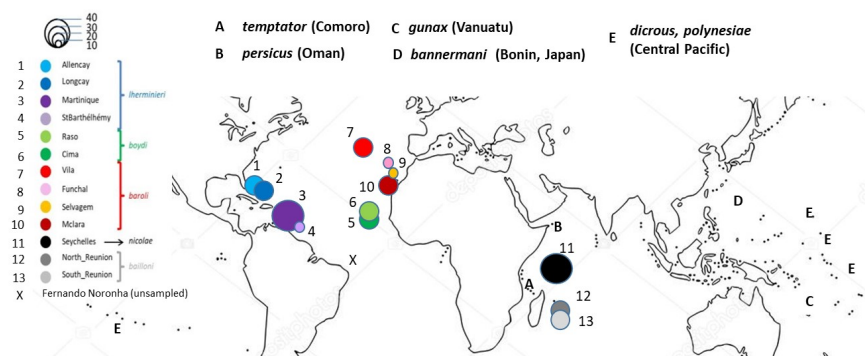
b. Pairwise  $F_{ST}$  for nuclear markers for females (below diagonal) and males (above diagonal).

Table 3: Sex-specific  $F_{IS}$  and Relatedness indices

Mean and confidence interval of  $F_{IS}$  and relatedness are indicated for each population with a sufficient sample size and for each lineage The observed difference of mean is considered as significant when not in the confidence interval (indicated in bold and by an asterisk)

Lineage	Population	Female Sample Size	Male Sample Size	$F_{IS}$ Female	$F_{IS}$ Male	Observed Related- ness Female	Observed Related- ness Male	Observed difference of relatedness	95% inf. dif
<i>lherminieri</i>	Allencay	8	10	<b>0.79</b> [ <b>0.77-</b> <b>0.90</b> ]	<b>0.35</b> [ <b>0.27-</b> <b>0.43</b> ]	0.45	0.55	-0.10	-0.1 0.1
<i>lherminieri</i>	Longcay	8	12	<b>0.19</b> [ <b>0.13-</b> <b>0.20</b> ]	<b>0.60</b> [ <b>0.55-</b> <b>0.69</b> ]	0.62	0.71	-0.09	-0.1 0.0
<i>lherminieri</i>	Bahamas	16	22	<b>0.63</b> [ <b>0.59-</b> <b>0.67</b> ]	<b>0.76</b> [ <b>0.75-</b> <b>0.79</b> ]	0.55	0.62	-0.07	-0.1 0.0
<i>lherminieri</i>	Martinique	19	22	0.71 [0.64- 0.71]	0.62 [0.59- 0.66]	<b>0.70*</b>	<b>0.63*</b>	<b>0.07</b>	-0.1 0.0
<i>boydi</i>	Raso	11	7	0.53 [0.49- 0.58]	0.51 [0.49- 0.55]	<b>0.49*</b>	<b>0.38*</b>	<b>0.11</b>	-0.1 0.1
<i>boydi</i>	Cima	10	9	<b>0.77</b> [ <b>0.76-</b> <b>0.78</b> ]	<b>0.66</b> [ <b>0.64-</b> <b>0.73</b> ]	0.56	0.57	-0.01	-0.1 0.0
<i>baroli</i>	Mclara	6	9	<b>0.36</b> [ <b>0.29-</b> <b>0.39</b> ]	<b>0.58</b> [ <b>0.55-</b> <b>0.60</b> ]	0.59	0.55	-0.04	-0.1 0.0
<i>baroli</i>	Vila	10	8	<b>0.48</b> [ <b>0.44-</b> <b>0.54</b> ]	<b>0.64</b> [ <b>0.63-</b> <b>67</b> ]	0.47	0.45	0.02	-0.1 0.0
<i>bailloni</i>	North Reunion	14	8	<b>0.77</b> [ <b>0.75-</b> <b>0.82</b> ]	<b>0.33</b> [ <b>0.23-</b> <b>0.43</b> ]	<b>0.68*</b>	<b>0.78*</b>	<b>-0.10</b>	-0.1 0.0
<i>bailloni</i>	South Reunion	17	11	<b>0.58</b> [ <b>0.58-</b> <b>0.66</b> ]	<b>0.42</b> [ <b>0.34-</b> <b>0.49</b> ]	<b>0.6*</b>	<b>0.72*</b>	<b>-0.12</b>	-0.1 0.0
<i>nicolae</i>	All populations	14	10	<b>0.61</b> [ <b>0.59-</b> <b>0.68</b> ]	<b>0.42</b> [ <b>0.39-</b> <b>0.46</b> ]	<b>0.6*</b>	<b>0.72*</b>	<b>-0.12</b>	-0.1 0.0

Figure 1: World map distribution of *Puffinus lherminieri* complex (see Austin et al. 2004 for *assimilis* ).



Numbers represent breeding localities that were sampled for this study (colour codes identical across figures). The size of filled circles corresponds to sample size (number of individuals). “X” represents the only breeding colony from North Atlantic that was not sampled here. Other letters (A to E) for other taxa within the complex.

Figure 2: Gene trees and scenario of breeding site colonization

a. Gene trees obtained by Bayesian inference for all markers, node bars correspond to the 95% confidence interval of the estimated divergence times. The scale corresponds to time before present in Million years (My). b. Gene trees obtained using \*BEAST for all mitochondrial markers with dichrous haplotypes in yellow. c. Gene trees obtained using \*BEAST for all mitochondrial and nuclear markers with dichrous sequences in yellow. In b. and c. only individuals sequenced for all mitochondrial markers and all markers respectively are showed. Only the posterior values  $>0.90$  are shown. d. Scenario of colonization inferred based on DIYABC. Branch colours correspond to ancestral populations, dates to mean divergence times of trees in My inferred by \*BEAST analyses. Each bifurcation corresponds to a divergence-colonization event. \* Indicates the split between Atlantic and Indian ancestral populations. The analysis excludes the population of Funchal, due to low sample size (4 individuals).

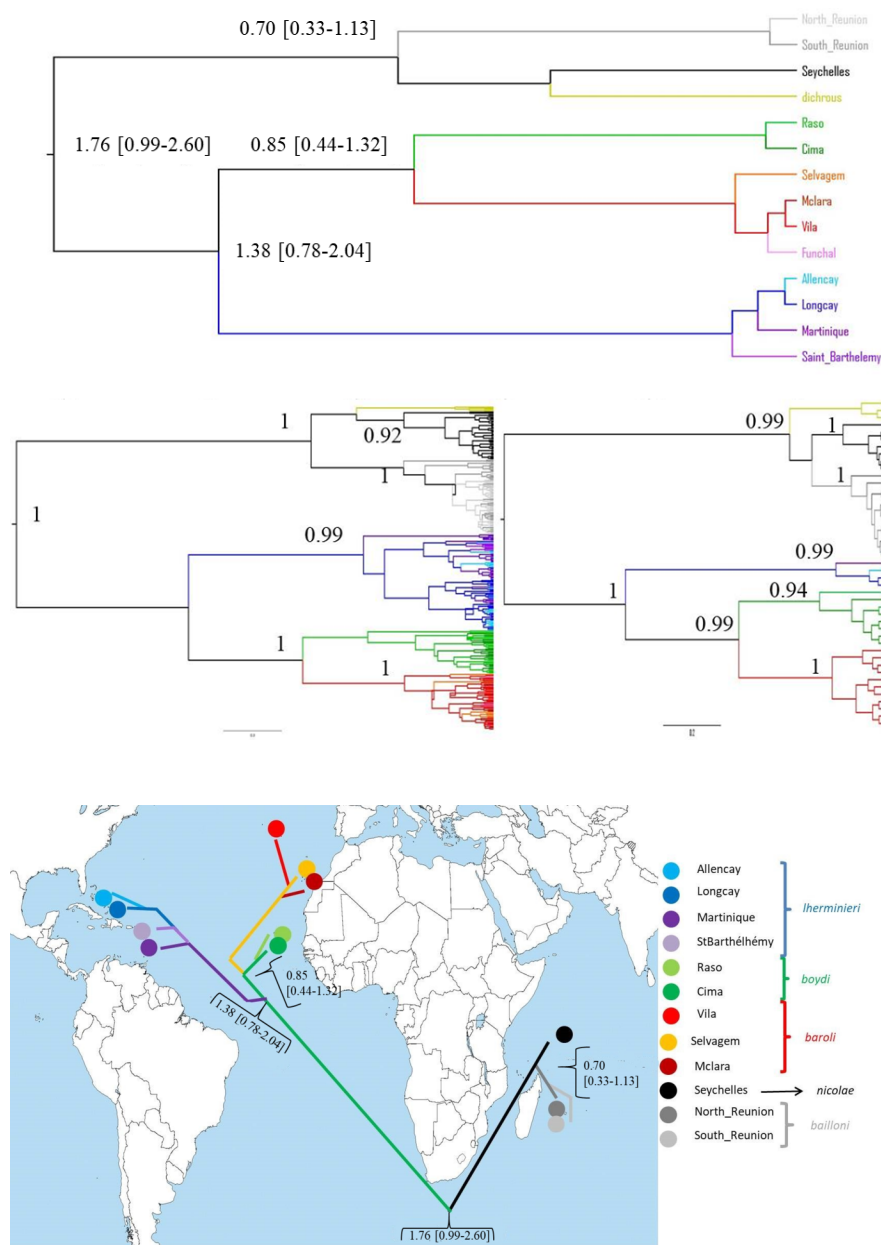


Figure 3: NeighborNet networks obtained using mitochondrial markers (a.) and nuclear markers (b.).

The scale bars indicate the sequence divergence (number of substitutions per site) represented by the length of a branch.



Allencay



Longcay



Martinique



StBarthélemy



Raso



Cima



Vila



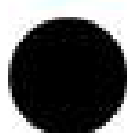
Funchal



Selvagem



Mclara



Seychelles

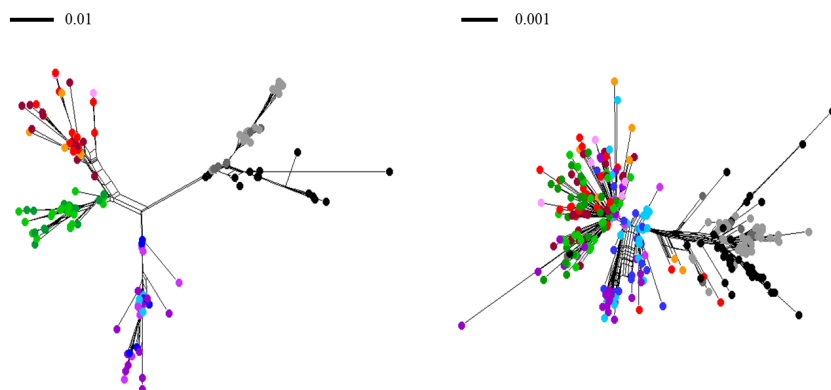
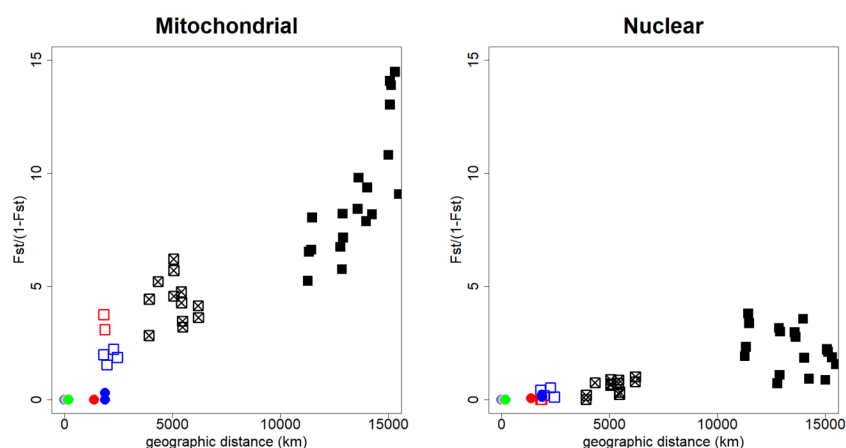


Figure 4: Correlation between genetic and geographic distance



Relationships between genetic ( $F_{ST}/(1-F_{ST})$ ) and geographic distances. Genetic distances were calculated for all concatenated mitochondrial markers (a.) and all concatenated nuclear markers (b.). Geographic distances were calculated as the shortest distances between pairs of populations without crossing land. Black squares: pairs of populations from Atlantic and Indian oceans. Light squares: pairs of population between *therminieri* and *baroli-boydi* (crossed), *boydi* and *baroli* (in blue) and *bailloni* and *nicolae* (in red). Coloured circles include only pairs of populations belonging to the same lineage.