


SCIENTIFIC REPORTS



OPEN

Unexpected population fragmentation in an endangered seabird: the case of the Peruvian diving-petrel

Robin Cristofari^{1,2,3,4}, Paula Plaza⁵, Claudia E. Fernández⁵, Emiliano Trucchi^{6,7}, Nicolas Gouin^{8,9}, Céline Le Bohec^{2,3,10}, Carlos Zavalaga¹¹, Joanna Alfaro-Shigueto¹¹ & Guillermo Luna-Jorquera ^{5,9}

In less than one century, the once-abundant Peruvian diving petrel has become the first endangered seabird of the Humboldt Current System (HCS). This small endemic petrel of the South American Pacific coast is now an important indicator of ongoing habitat loss and of the success of local conservation policies in the HCS - an ecoregion designated as a priority for the conservation of global biodiversity. Yet so far, poorly understood life history traits such as philopatry or dispersal ability may strongly influence the species' response to ecosystem changes, but also our capacity to assess and interpret this response. To address this question, we explore the range-wide population structure of the Peruvian diving petrel, and show that this small seabird exhibits extreme philopatric behavior at the island level. Mitochondrial DNA sequences and genome-wide SNP data reveal significant isolation and low migration at very short distances, and provide strong evidence for questioning the alleged recovery in the Peruvian and Chilean populations of this species. Importantly, the full demographic independence between colonies makes local population rescue through migration unlikely. As a consequence, the Peruvian diving petrel appears to be particularly vulnerable to ongoing anthropogenic pressure. By excluding immigration as a major factor of demographic recovery, our results highlight the unambiguously positive impact of local conservation measures on breeding populations; yet at the same time they also cast doubt on alleged range-wide positive population trends. Overall, the protection of independent breeding colonies, and not only of the species as a whole, remains a major element in the conservation strategy for endemic seabirds. Finally, we underline the importance of considering the philopatric behavior and demographic independence of breeding populations, even at very fine spatial scales, in spatial planning for marine coastal areas.

The large-scale oceanographic processes that structure the southern Pacific Ocean also fuel the upwelling of nutrient-rich subsurface waters along the West coast of South America¹, resulting in one of the most highly productive oceanic systems on the planet: the Humboldt Current System (HCS)^{2,3}. Local cape and island topology

¹Department of Biology, University of Turku, 20014 Turun Yliopisto, Turku, Finland. ²Université de Strasbourg, Centre National de la Recherche Scientifique (CNRS), Institut Pluridisciplinaire Hubert Curien (IPHC) UMR 7178, F-67000, Strasbourg, France. ³Laboratoire International Associé (LIA-647 BioSensib, CSM-CNRS-Unistra), 8 Quai Antoine 1er, Monaco, 98000, Monaco. ⁴Centre for Ecological and Evolutionary Synthesis (CEES), Department of Biosciences, University of Oslo, Postboks 1066, Blindern, Oslo, Norway. ⁵Millennium Nucleus for Ecology and Sustainable Management of Oceanic Island (ESMOI), Departamento de Biología Marina, Facultad de Ciencias del Mar, Universidad Católica del Norte, Larrondo 1281, Coquimbo, Chile. ⁶Department of Botany and Biodiversity Research, University of Vienna, Rennweg 14 A-1030, Vienna, Austria. ⁷Department of Life Sciences and Biotechnology, University of Ferrara, 44121, Ferrara, Italy. ⁸Instituto de Investigación Multidisciplinar en Ciencia y Tecnología, Universidad de La Serena, Av. Raul Bitran Nachary, La Serena, Chile. ⁹Centro de Estudios Avanzados en Zonas Áridas (CEAZA), Larrondo 1281, Coquimbo, Chile. ¹⁰Centre Scientifique de Monaco - Département de Biologie Polaire, 8, quai Antoine 1er, MC, 98000, Monaco. ¹¹Universidad Científica del Sur, Lima, Antigua Panamericana Sur Km 19, Lima, Peru. Robin Cristofari and Paula Plaza contributed equally. Correspondence and requests for materials should be addressed to G.L.-J. (email: gluna@ucn.cl)

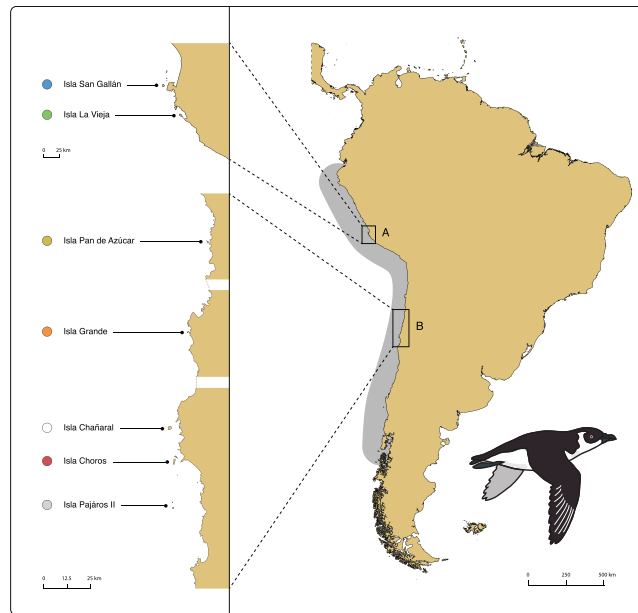


Figure 1. Current breeding range of the Peruvian diving petrel. (A) Peruvian and (B) Chilean range of the species. On the left, close-up of the range of the species (dashed lines indicate discontinuity in our representation of the shoreline). The grey area represents the historical range of the species. Isla Chañaral is currently unoccupied, and Isla Pájaros II, in grey, could not be sampled for this study. Islands sampled for the study (Choros, Grande, Pan de Azúcar, La Vieja and San Gallán) are indicated by colored circles.

increases and stabilizes the upwelling activity, and generates productivity and biodiversity hotspots⁴. Together with the high endemism of its coastal ecosystems^{5–7}, this makes the HCS an ecoregion of global interest for the conservation of marine biodiversity^{8,9}. Concurrently, its exposure to direct threats² and strong sensitivity to climate change^{3,10} make reliable information about ongoing biological processes a vital necessity. The use of seabirds as important integrative bioindicator species has been regularly stressed in this context: as upper predators, they respond both to overall ecosystem productivity and to trophic network restructuring^{11–13}. However, current knowledge of HCS seabird populations is limited: demographic surveys are often patchy and lack a consistent methodological framework, resulting in over- or underestimation of population trends^{14,15}.

One difficulty in studying seabird populations is their dependence on both marine and terrestrial ecosystems^{16–18}, which makes them sensitive to changes on both sides of the shoreline. At sea, threats involve direct impacts of fisheries such as competition for food^{13,19–23} and bycatch^{24,25}, and indirect impacts of human activities such as climate change^{19,26,27}. On land, the disruption of breeding habitats can be caused by the spread of diseases²⁸, sudden habitat disappearance²⁹ and/or growing interaction with terrestrial predators - including humans³⁰. Unlike the changes in the marine environment, disturbances on land may have local effects; failure to clearly assess the scale at which these effects occur can lead us to confuse local effects with general trends^{12,31,32}, which may result in flawed conservation strategies.

The Peruvian diving petrel (*Pelecanoides garnotii*), a small seabird endemic to the HCS, is a dramatic example of the way multiple direct and indirect threats may converge on a single species. This small insular petrel was once one of the most abundant seabirds of the HCS. In the early 20th century, it was documented from 6°S (in Isla Lobos de Tierra, Peru) to 42°S (in Isla Chiloé, Chile)^{33,34}. Yet the Peruvian diving petrel was driven to near-extinction by human-induced threats that included habitat destruction through guano extraction^{34–36}, poaching of adults and egg collection³⁵, bycatch^{35,37}, human-introduced predators^{15,38} and habitat competitors³⁹, food competition with fisheries^{35,40}, light pollution (Sea Shepherd Chile, pers. com.), and the effects of ongoing climate change on the productivity of the HCS³. As a consequence of these multiple threats, only seven nesting sites remain active in the previous 35° latitudinal range of the species^{36,38,41,42}. The species' breeding range has been fragmented into two main areas, separated by 1,300 km (Fig. 1) — one in central Peru, between 8°S and 14°S (Isla Corcovado with a handful breeding pairs, Isla San Gallán and Isla La Vieja with 12,000–13,000 pairs in total^{40,43,44}), and one in northern Chile between 26°S and 29°S (Islas Pan de Azúcar, with an estimated ~220 pairs, Isla Choros, with 9,516 pairs, Isla Grande, with ~200 pairs, and Isla Pájaros II, with ~120 pairs^{15,45–47}). Historical population sizes are unknown, but available information suggests a rapid decline. A striking example can be found in Isla Chañaral, in northern Chile, where ~100,000 pairs bred in 1938³⁸: this population is now extinct¹⁵.

In contrast to this global decline, however, marginal range re-expansion and local population growth have been suggested³⁶. In the past decades, small breeding groups have been discovered on the Peruvian Corcovado Island⁴³, and non-breeding individuals have been sighted off northern Peru^{36,42,48}. Rapid population growth was observed on La Vieja Island, Peru⁴⁴, as well as on some Chilean islands^{15,49}. However, the processes underlying these possible population recovery events are poorly understood, and it is yet unknown whether the remaining populations have effectively gone through a bottleneck, as suggested by their severe decline. Most importantly,

the scarcity of knowledge on the species' life history, phenology, and at-sea distribution casts serious doubts on the few existing surveys. It cannot be excluded that active breeding locations have only been visited by observers outside of the breeding peak, or that visits have been too few to assess population trends. Distinguishing recovery from increasing knowledge about a previously underestimated population is not a trivial problem¹⁴. As an example, recent census estimates for Isla Pan de Azúcar vary between ~220⁴⁵, ~600⁵⁰ and ~3,300 breeding pairs⁵¹ based on a similar methodology, without any decisive argument to explain these discrepancies. It may be that the species did not go through a population bottleneck at all, and that its resilience to habitat disruption has been much higher than previously thought.

The effects of local demographic processes in the Peruvian diving-petrel may be further blurred by the lack of genetic structure generally observed in HCS seabirds. Indeed, while most studies conducted in seabirds globally have emphasized the importance of genetic or phylogeographic structure in species dynamics^{52,53}, the HCS stands out as an exception: to date, surveys have revealed a lack of genetic structure at the continental scale (e.g. Peruvian booby *Sula variegata*⁵⁴, or Peruvian pelican *Pelecanus thagus*⁵⁵), or very high gene flow between colonies (Humboldt penguin *Spheniscus humboldti*⁵⁶). Promotion of high dispersal by foraging concentrations around localized upwellings⁵⁵ has been proposed as a central underlying mechanism. If the Peruvian diving petrel fits into the same panmictic picture, demographic independence of colonies may not be assumed. In that case, observed local population growth should be interpreted as the result of complex processes, and only the range-wide assessment may be informative as to the status of the species. If, on the other hand, high philopatry and low dispersal are the norm, we can expect demographic independence between breeding locations. In that case, population trends observed at single locations may reflect actual local demographic responses (e.g. local conservation measures such as reduced guano extraction in several protected islands, and the creation of marine protected areas), rather than range-wide and species-scale processes. In addition, local gene pools should be informative about the recent demographic events at each location. Here, we use a combination of range-wide mitochondrial data and genome-wide genetic markers to (a) estimate the importance of philopatric behavior and demographic independence in the Peruvian diving-petrel, and (b) understand recent population history and its implications for the future of the species. In particular, we examine whether the current distribution of the species is better explained by local extirpation and recolonization, or rather by unexpected resilience.

Material and Methods

Ethics statement. This study was carried out in accordance with the recommendations of the Corporación Nacional Forestal de Chile (permit N°38/2012), the Servicio Agrícola y Ganadero de Chile (Resolution N° 7238/2013), and the Servicio Nacional de Áreas Naturales Protegidas por el Estado, Ministerio del Ambiente de Perú (permit N° 22-2018-SERNANP - RNP/J). The protocol was approved by the Bioethics Committee of the Universidad Católica del Norte, Coquimbo, Chile.

Sample collection and DNA extraction. Blood sampling of Peruvian diving petrels was performed between April and May of 2012, outside of the reproductive peak season to minimize disturbance. A total of 109 individuals, representing seven colonies of the Peruvian diving petrel, were sampled along its breeding range (Fig. 1): two in Isla Choros, Chile (29°16'S 71°32'W, CHA and CHB, 32 and 11 individuals, respectively), one in Isla Grande de Atacama, Chile (27°15'S 70°59'W, IGA, 7 individuals), two in Isla Pan de Azúcar, Chile (26°09'S 70°41'W, AZA and AZB: 27 and 11 individuals, respectively), one on Isla La Vieja, Peru (14°15'S 76°12'W, ILV, 12 individuals), and one on Isla San Gallán, Peru (13°50'S 76°26'W, ISG, 9 individuals). Adults were captured using a mist net placed at the entrance of the colony at dusk: capturing adults as they returned to their burrows minimized the chance of sampling prospecting individuals (which presumably do not excavate a burrow until they settle for breeding). Blood was sampled from the retia in the interdigital membrane of the foot, using a heparinized microcapillary, dried on a Whatman filter paper and stored at -20 °C. DNA extraction was performed using a standard spin-column protocol (Qiagen DNeasy® Blood and Tissue kit) with minor modifications. All collected samples (N = 109) were sequenced for a fragment of the mitochondrial cytochrome-b gene. A subset of 21 samples from two islands (Isla Choros: CHA = 6, CHB = 5, and Isla Pan de Azúcar: AZA = 5, AZB = 5) was selected for single-digest RAD sequencing. We randomly selected these samples from a subset for which DNA molecular weight was above 10,000 bp (as assessed visually on a 1% agarose gel), and total available DNA quantity was above 500 ng (the starting quantity for our RAD sequencing library preparation protocol⁵⁷).

Mitochondrial marker sequencing. A 900 bp-long fragment of the mitochondrial cytochrome-b gene was amplified for all samples using either two or four specific primers designed from the sequence published by Nunn and Stanley⁵⁸: as 5'-3', yun1f: GCCCAAACCTCCGAAAATCCCA and yun2r: GGTGATGGAGG CTAGTTGGCCG or with internal primers yun1r: GCCTGATTCGTGAAGGAAGGTGAGG and yun2f: CCACC CTAACCCGATTCTTCGCC. Amplification and sequencing initially used the four primer design, but it appeared that using only two primers yielded equally good sequences, and that design was preferred for the latest sequences (Isla Grande de Atacama, Isla San Gallán and Isla La Vieja). PCR were performed using MasterMix® (Qiagen) premixed TAQ-polymerase, dNTPs and MgCl₂ in a total reaction volume of 12.5 µL (6.75 µL MasterMix, 2 µL of each primer pair, 0.5 µL DNA, H₂O to a final volume of 12.5 µL). Amplification used the following conditions: 5' at 94 °C, followed by 35 cycles of 30" at 94 °C, 30" at 57 °C and 1' at 72 °C, and finally 5' at 72 °C. PCR products were then purified using Illustra™ ExoStar™, and Sanger sequencing was performed at the ABI lab of the University of Oslo, on an ABI PRISM® 3733 analyzer. We used Geneious® v6.1.2 to process the reads: we performed visual assessment of read quality, then trimmed both ends of each read to keep only bases with an error probability below 2.5%. Finally, sequences were assembled, and a manual control was performed on the consensus sequence. All sequences were aligned and trimmed to keep only the region sequenced successfully in all samples.

RAD library preparation and SNP typing. SNP discovery and sequencing was performed on a subset of individuals (Isla Choros: CHA = 6, CHB = 5, and Isla Pan de Azúcar: AZA = 5, AZB = 5) following a single-digest RAD-sequencing protocol⁵⁷, modified as detailed by Cristofari *et al.*³² and sequenced on a half Illumina HiSeq2500 lane (V3 chemistry) at the Norwegian Sequencing Center, using paired-end, 2 × 95 bp reads. As no reference genome was available for the Peruvian diving petrel or any closely related species, data processing followed a two-step protocol. First, loci were built *de novo*, using the Stacks v1.28 pipeline^{59,60}, and a synthetic reference RADome was built – an approach suggested by Hoffman *et al.*⁶¹ and that allows the use of well-established SNP calling algorithms in the absence of a closely-related species’ reference genome, while avoiding biases towards slowly-evolving regions due to differential success of alignment. Second, this reference RADome was used to perform alignment-based SNP typing. The workflow was as follows: (a) sequence de-multiplexing and *de novo* locus assembly was done according to in-line barcodes using Stacks’ *process_radtags.pl* script and *denovo_map.pl* pipeline, with a minimum depth of 3X, and a maximum of 5 mismatches allowed between alleles at a single locus (both within and between individuals). This level is, in our experience, reasonable for seabirds: it is the “convergence level” in King penguins, according to Paris *et al.*⁶², and indeed, changes in parameters within the 3–6 range only marginally affected our assembly (both in terms of total number of loci, and number of polymorphic loci). The corresponding paired-end contigs were assembled using Velvet⁶³ and Stacks, setting a minimum contig length of 200 bp to filter out multiple non-overlapping short paired-end reads. (b) The resulting loci were concatenated together with their respective paired-end contigs, separated by a padding of 500 ‘N’s, to form the scaffolds of a reduced reference genome, or RADome⁶¹. At this point, loci comprising multiple paired-end contigs (a potential sign of collapsed paralogous loci) were removed. The raw paired-end fastq files were then mapped onto this RADome using Bowtie2⁶⁴, with paired-end, concordant mapping only, a maximum distance between both reads of 500 bp (corresponding to our fixed-length N-padding), and otherwise default settings. The resulting alignments were filtered using Samtools 0.1.19⁶⁵, PicardTools 1.113 (<http://picard.sourceforge.net>), and custom R and shell scripts: we discarded orphaned reads and low-quality pairs (a minimum Phred-scaled quality score of 30 was used), and further filtered the remaining aligned reads to remove PCR and optical duplicates. (c) SNP and genotype calling was restricted to the first-in-pair reads. We then used GATK’s HaplotypeCaller pipeline⁶⁶ to call SNPs and genotypes simultaneously in all samples, using first reads only, with standard parameters except for population heterozygosity which was set to 0.01, and retaining only di-allelic, non-indel sites sequenced in at least 75% of samples. We retained only 1 SNP per locus (arbitrarily set as the first SNP in read order). We repeated key analyses using ANGSD 0.900^{67,68} to estimate allele frequency and genotype likelihoods, without hard-calling genotypes, an approach that has been shown to be much more robust for low-coverage data^{68,69}.

Analysis of population structure. Summary statistics (nucleotide diversity π and Tajima’s D) were calculated using DnaSP⁷⁰ for mtDNA, and adegenet⁷¹ for SNP data. Analysis of molecular variance was performed for both datasets using Arlequin v3.5⁷², in a locus-by-locus format for SNP data, and haplotypic format for mtDNA. We performed 15,000 permutations to assess significance levels. Additionally, significance level for pairwise Φ_{ST} estimates from mtDNA were corrected using the Benjamini-Yekutieli procedure⁷³ using the R package *multtest*⁷⁴. On the other hand, pairwise SNP-based comparisons were hierarchical (one pairwise comparison within each island, and one pairwise comparison between the two sampled islands), and each was considered to test for an independent null hypothesis, so that no correction was applied. A neighbor net was built in SplitsTree⁷⁵ either directly from the mtDNA sequences, or from a matrix of pairwise Hamming distances calculated in PLINK v1.9⁷⁶ from SNP data. Additionally, a maximum likelihood tree was built from the cytochrome-b sequences in RAxML⁷⁷, using a simple HKY + G substitution model as supported by the Akaike Information Criterion (AIC)-based model selection performed in *JModelTest*⁷⁸, and otherwise default parameters. This tree was used to build a Fitch-distance haplotype network using Fitchi⁷⁹.

Principal components analysis was performed on genome-wide SNP data either using called genotypes in *adegenet*, or raw genotype likelihoods in ngsTools⁸⁰. Admixture between populations was estimated using two different approaches: (1) a classical model-based clustering approach relying on the hypothesis of Hardy-Weinberg equilibrium (HWE)⁸¹, as implemented with an empirical Bayes algorithm either in fastStructure⁸² for called genotypes, or in ngsAdmix⁸³ for raw genotype likelihoods, retaining only sites that did not violate HWE, with a number of components ranging from 1 to 4, and 100 bootstrap replicates. Model complexity was chosen using Evanno’s ΔK method⁸⁴. (2) In order to assess the impact of the HWE assumption, we used a “model-free” approach implemented in the *adegenet* *pca-coordinate* k-means clustering algorithm⁸⁵.

Model-based estimation of population history. We relied on two separate coalescent-based approaches to estimate the amount of gene flow between the northern and southern ends of the species’ Chilean range (i.e. between Isla Choros and Isla Pan de Azúcar) based on genome-wide SNP data. First, population sizes and migration rates were co-estimated based on multilocus short-haplotype data in a Bayesian framework, as implemented in *Migrate-n*⁸⁶. We selected three random sets of 50 95-bp-long nuclear loci containing 4 to 6 polymorphisms each as an unbiased representation of the neutrally evolving part of the genome (a protocol described in Trucchi *et al.*⁸⁷), phasing being provided by the Illumina read identity. In order to correct for potential allele dropout, we randomly sampled one haplotype (i.e. allele) only for each individual, at each locus, as proposed by Cristofari *et al.*³². We ran a cold chain and 3 heated chains under a static heating scheme, raising the cold chain to a power of 1.5, 3 and 1e6 and proposing chain swapping every 100 steps. Chains were run for 50,000,000 generations, recording every 500 generations, with a 5,000,000-generation burn-in, specifying uniform priors both for population sizes (Θ , bounded between 0 and 0.1 with a δ of 0.01), and for the migration rates (M , bounded at 4,000 with a δ of 400). Proper mixing under these conditions was ensured by using the highest parametrization model and checking convergence between the three independent random datasets. Models were ordered by log Bayes factors⁸⁸, using the thermodynamic integration approximation of marginal likelihood ds. Four models were tested: (1) the two

	AZA	AZB	CHA	CHB	IGA	ISG	ILV
AZA	—	0.269 ± 0.004	0.001 ± 0.000	0.000 ± 0.000	0.180 ± 0.004	0.000 ± 0.000	0.000 ± 0.000
AZB	0.015	—	0.250 ± 0.004	0.001 ± 0.000	0.734 ± 0.004	0.001 ± 0.000	0.001 ± 0.000
CHA	0.129**	0.016	—	0.012 ± 0.001	0.660 ± 0.0045	0.005 ± 0.000	0.001 ± 0.000
CHB	0.398**	0.364**	0.165	—	0.0467 ± 0.002	0.256 ± 0.0045	0.185 ± 0.004
IGA	0.046	−0.051	−0.030	0.205	—	0.048 ± 0.002	0.022 ± 0.001
ISG	0.421**	0.417**	0.209*	0.032	0.212	—	0.879 ± 0.003
ILV	0.443**	0.435**	0.227**	0.0412	0.248	−0.086	—

Table 1. Pairwise Φ_{ST} values computed from the cytochrome-b mtDNA sequences between each Peruvian diving petrel breeding colony. Below diagonal: mean Φ_{ST} values. Above diagonal: p -values for significance as assessed against a random distribution of the same sequences, with 10,000 permutations. Significant differences from random expectations based on permutations of the same dataset are signalled by asterisks, after Benjamini-Yekutieli correction (* for a corrected p -value ≤ 0.05 , and ** for a corrected p -value ≤ 0.01). AZA/AZB and CHA/CHB are the two breeding colonies of Isla Pan de Azúcar and Isla Choros, respectively, IGA is Isla Grande de Atacama, ILV is Isla La Vieja and ISG is Isla San Gallán.

islands are fully isolated populations; (2) the two islands are independent populations exchanging migrants with symmetrical gene flow; (3) the two islands are independent population exchanging migrants with asymmetrical gene flow; and (4) the two islands belong to a single, panmictic population.

Co-estimation of population size, gene flow and population history was performed by explicit model testing in *fastsimcoal2*⁸⁹, through AIC comparison of the optimized composite likelihood of the two-dimensional folded allele frequency spectrum under six different models (see input files in Supplementary information). We first tested two simple models: (1) a full isolation model with two stable, independent populations, and (2) a classical isolation-with-migration model, with population sizes, migration rates and divergence time as free parameters. We then tested four more complex scenarios (3–6), in which the two populations diverged after the last glacial maximum, ca. 7,000 generations ago (a conservative biogeographic estimate for most taxa). Under model (3) human activities have no impact on the populations, and constant gene flow is maintained between the islands until the present, with no population size changes. Under model (4) population size does not change, but gene flow is disrupted at some point during the past. Under models (5) and (6), population size changes during the past, and either gene flow remains constant (model 5) or gene flow also changes (model 6). We generated 50 nonparametric spectrum bootstrap replicates and performed 50 parallel runs for each model and each bootstrap replicate, keeping only the one with the highest composite log-likelihood. Each run required a maximum of 80 ECM cycles, simulating 100,000 spectra at each step.

Models were calibrated using the general background genomic mutation rate for water birds, 1.6×10^{-3} substitutions.site⁻¹.Myr⁻¹, as established from a wide panel of genomic data by Zhang *et al.*⁹⁰, or $\mu = 9.6 \times 10^{-9}$ substitutions.site⁻¹.generation⁻¹ considering a generation time of 6 years⁹¹. *Migrate-n* models, however, were based on RAD loci containing 4 to 6 polymorphic sites: these cannot be calibrated directly using the general background substitution rate. Considering that most of the RADome is assumed to evolve neutrally, the number of polymorphic sites per locus is expected to follow a Poisson distribution, and the background substitution rate μ applies for the expected value λ of that distribution. Therefore, after verifying the goodness-of-fit of the Poisson model we extracted the expected substitution rates for loci with $S = 4$, $S = 5$ or $S = 6$ SNPs, as $S/(\lambda/\mu)$, and used their weighted average to calibrate our reconstructions³².

Range-wide process analysis from mtDNA data. In order to better understand range-wide demographic processes, we used sampling locations as the only prior hypothesis on phylogeographic processes, and inferred locations at ancestral nodes in a Bayesian framework as implemented in *BEAST2*^{92,93}. We partitioned data so as to set a separate model for third-codon positions, using a simple HKY + G mutation model as inferred in *jModelTest2*⁷⁸ with three gamma-distributed categories. Since we worked at an intraspecific level without external prior information on population history, we chose to remain within a simple constant-population coalescence model with a strict molecular clock. The sampling chain was run for 100,000,000 generations, recording every 5,000th state. This allowed the generation of sufficient effective sample sizes ($ESS > 500$) for all parameters in the model. Clock calibration used the estimate provided for the cytochrome-b gene by Nunn and Stanley⁵⁸ for medium-sized Procellariiformes (the order of seabirds that includes the *Pelecanoides* diving-petrel genus), of 0.90% per million years.

Results

Range-wide mtDNA analysis. Analysis of cytochrome-b sequences unequivocally indicated a high degree of reproductive isolation between the main breeding locations across the range of the Peruvian diving-petrel. Pairwise Φ_{ST} values were high, and the majority deviated significantly from a random distribution of the same sequences (Table 1). This was particularly the case for longer-distance comparisons. The two neighboring Peruvian colonies of Isla La Vieja and Isla San Gallán, on the other hand, did not differ significantly. Within the islands of Choros and Pan de Azúcar, differences between colonies were insignificant. Analysis of molecular variance, with colonies gathered into four groups (a-both Choros colonies, b-Isla Grande de Atacama, c-both Pan de Azúcar colonies, and d-both Peruvian colonies) also supported significant differentiation between groups. Difference between groups accounted for 14.87% of the total variation, while difference between colonies within

	AZR	CHR	IGA	ILV	ISG	ALL
Haplotype diversity	$H_{AZR} = 0.711$	$H_{CHR} = 0.632$	$H_{IGA} = 0.810$	$H_{ILV} = 0.667$	$H_{ISG} = 0.778$	$H_{ALL} = 0.765$
Nucleotide diversity	$\pi_{AZR} = 0.00125$	$\pi_{CHR} = 0.00220$	$\pi_{IGA} = 0.00204$	$\pi_{ILV} = 0.00185$	$\pi_{ISG} = 0.00203$	$\pi_{ALL} = 0.00213$
Nucleotide diversity (nuclear SNP)	$\pi_{AZR} = 0.00226$	$\pi_{CHR} = 0.00224$	—	—	—	—
Pairwise differences	$K_{AZR} = 0.997$	$K_{CHR} = 1.75$	$K_{IGA} = 1.62$	$K_{ILV} = 1.47$	$K_{ISG} = 1.61$	$K_{ALL} = 1.70$
Tajima's D	$D_{AZR} = -0.649$	$D_{CHR} = -0.430$	$D_{IGA} = -1.02$	$D_{ILV} = 0.384$	$D_{ISG} = -0.673$	$D_{ALL} = -0.672$
Tajima's D (nuclear SNP)	$D_{AZR} = -0.361$	$D_{CHR} = -0.364$	—	—	—	—

Table 2. Summary statistics of the cytochrome-b mtDNA haplotype sequences for the Peruvian diving petrel breeding colonies. AZR: Isla Pan de Azúcar, CHR: Isla Choros, IGA: Isla Grande de Atacama, ILV: Isla La Vieja, ISG: Isla San Gallán, ALL: pooled samples. ns: p -value > 0.05. Values are computed from the cytochrome B sequence, or from nuclear SNP data when specified.

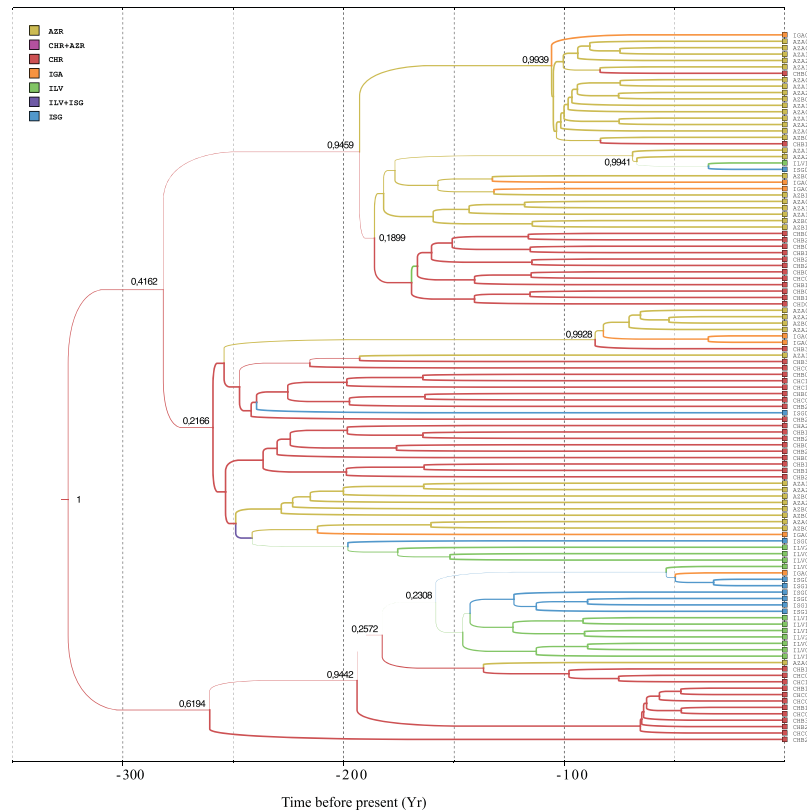


Figure 2. Maximum-likelihood phylogenetic tree of the Peruvian diving petrel based on cytochrome-b sequences and the HKY + G substitution model. Calibrated time tree with inferred ancestral location. Colors indicate the location, either sampled (at the tips), or inferred (along the branches). Posterior probability is indicated at the main nodes.

each group accounted for 12.11%. Differentiation between groups was significant ($\Phi_{ST} = 0.270$, p -value < 0.05), and so was differentiation between colonies within groups ($\Phi_{SC} = 0.142$, p -value < 0.05). All 109 pooled sequences contained 14 segregating sites and 14 different haplotypes. Despite rather asymmetric sample sizes, standard indices of molecular diversity were homogeneous (Table 2).

The coalescent-based joint reconstruction of the gene tree for cytochrome-b and geographic location probability at ancestral nodes performed in *BEAST2* provided more information on the present-day structure of Peruvian diving petrel populations (Fig. 2). The highest posterior probability was for a common ancestor for all samples in Isla Choros, and subsequent colonization of the Chilean and Peruvian range from there - although a common ancestor in Isla Pan de Azúcar cannot be rejected with confidence. Using a common estimate of mitochondrial substitution rate in medium-sized Procellariiformes, this common ancestry is inferred to be rather recent - about the early 17th century. Most of our samples from Peru seem to stem from a single migration event, although a few others are inferred to be more recent migrants from Isla Choros and Isla Pan de Azúcar. Most remarkably, samples from Isla Grande de Atacama appear to result from multiple independent migration events, with sources in Choros, Pan de Azúcar and Peru. These observations are consistent with the maximum-likelihood-tree-based Fitch network (Fig. 3D).

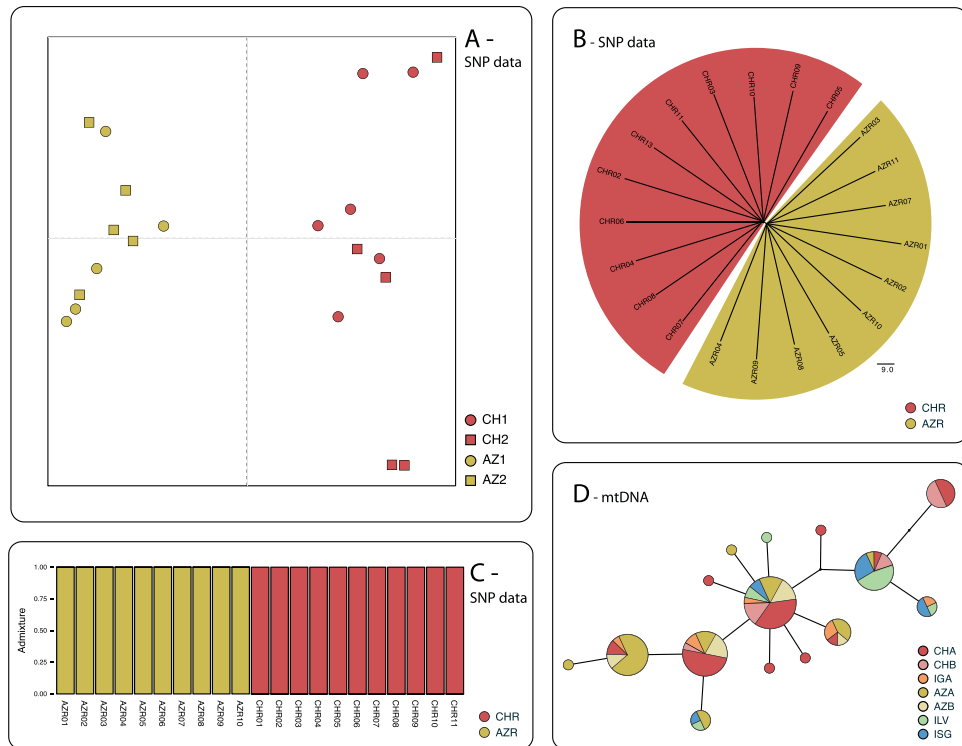


Figure 3. Population genetic structure of the Peruvian diving petrel. **(A)** Principal component analysis performed on genome-wide SNP data generated for the Chilean breeding colonies of Isla Choros (CH1 and CH2) and Isla Pan de Azúcar (AZ1 and AZ2), using raw genotype likelihoods calculated by ngsTools; **(B)** NeighborNet of the Isla Choros (CHR) and Isla Pan de Azúcar (AZR) samples built from nuclear SNP data; **(C)** Individual components of admixture between samples of Isla Choros (CHR) and Isla Pan de Azúcar (AZR), as inferred from genome-wide SNP genotype likelihood data using ngsAdmix (median values from 500 bootstrap replicates); **(D)** Fitch distance haplotype graph based on the mitochondrial cytochrome-b gene sequences of all the breeding colonies sampled for the study (disc area proportional to the number of haplotype copies). AZA/AZB and CHA/CHB refer to the two breeding colonies sampled in Isla Pan de Azúcar and Isla Choros, respectively, IGA to Isla Grande de Atacama, ILV to Isla La Vieja and ISG to Isla San Gallán. Color codes on all figures are as in Fig. 1.

A focus on two islands using genome-wide SNP data. We found genome-wide short-locus RAD-sequencing and mitochondrial data to be in good agreement. The Illumina sequencing lane yielded 182,319,948 paired-end reads with a mean phred-scaled quality score of 37, of which 144,048,478 were retained after barcode and cut-site sequence filtering. The final database included 65,582 distinct RAD loci, of which 37,170 were polymorphic, with an average of 1.3 SNPs per locus. After PCR-duplicate removal, SNP calling from the higher-quality reads in GATK yielded 43,175 SNPs, of which 32,964 were genotyped in at least 75% of individuals, with a mean depth of 3.2X and a median depth of 3.1X. Nucleotide diversity computed from genome-wide SNP data was very similar across islands (see Table 2 - and pooled diversity $\pi = 0.00233$), in perfect agreement with mtDNA-based values (see Table 2), and at a level in accordance with the expectation for long-lived seabirds (Romiguier *et al.* 2014). Here again, Tajima's D supports neutral evolution at the variable sites (see Table 2). Small sample sizes have been shown to bias genome-wide estimates of population size in cases of non-neutral sequence evolution, or exponential population growth^{94,95}. Neither appear to be the case in our data, although strictly quantitative interpretation of the demography is of course impossible.

All analyses support a high level of genetic divergence at such a short geographic distance for seabirds^{52–56}. The fixation index was high for genome-wide SNP data ($F_{ST} = 0.049 \pm 0.008$, averaged over 500 SNP windows). Analysis of molecular variance also supports significant differentiation between Isla Choros and Isla Pan de Azúcar. Based on SNP data, the difference between islands accounts for 2.51% of the total variation, while differences between colonies within each island are virtually nonexistent (-0.099% per AMOVA). Genetic differentiation between island groups is significant ($\Phi_{ST} = 0.025$, p -values < 0.05), but differentiation between colonies within islands is not ($\Phi_{SC} = -0.001$, p -value = 0.795). Principal component analysis of the SNP data clearly supports a two-population structure (Fig. 3A). Although the first component only accounts for 6.62% of the total variance, it discriminates widely between individuals from Isla Choros and Isla Pan de Azúcar - while the second and following components only summarize inter-individual variance. The idea of a low-level but highly consistent separation is also supported by a Hamming-distance-based neighbor net (Fig. 3B); terminal edges are much longer than structural inner edges, yet sorting is complete between the two islands (a structure that is however rather poorly supported by a mtDNA haplotype network, see Fig. 3D). Finally, clustering analyses, whether based on genotype likelihood (Fig. 3C) or on genotype calls, strongly support a two-population model, over

	Model 4		Model 6	
	Median	CI95%	Median	CI95%
T0	(7,000)		(7,000)	
M _{AC} 0	2.756e-4	[2.397e-4; 3.025e-4]	2.853e-4	[2.466e-4; 3.128e-4]
M _{CA} 0	1.146e-5	[1.121e-5; 1.209e-5]	1.135e-5	[1.018e-5; 1.207e-5]
N _A 0	—	—	1,014	[1,010; 1,019]
N _C 0	—	—	95,542	[88,815; 101,124]
T1	207	[82; 375]	75	[6; 231]
M _{AC} 1	4.268e-3	[2.553e-3; 1.041e-2]	1.103e-2	[3.724e-3; 1.116e-1]
M _{CA} 1	1.170e-5	[1.124e-5; 1.240e-5]	1.224e-5	[1.122e-5; 2.226e-5]
N _A 1	1,013	[1,010; 1,018]	1,027	[1,011; 17,702]
N _C 1	91,35	[87,977; 95,509]	23,832	[10,068; 36,104]

Table 3. Parameter estimates from *fastsimcoal2* demographic reconstructions. Model structures are illustrated in Fig. 4. Times are given in generations, population sizes in effective individuals, and migration rates are mutation-scaled. Parameter names are as follows: T0 is the fixed divergence, set at 7,000 generations. T1 is the time of regime change. M_{AC}0, M_{CA}0, M_{AC}1 and M_{CA}1 are the migration rates from Pan de Azúcar to Choros and from Choros to Pan de Azúcar, during period 0 and period 1, respectively. For model 4, N_A1 and N_C1 are the (constant) population sizes in Pan de Azúcar and Choros. For model 6, N_A0, N_C0, N_A1 and N_C1 are the population sizes in Pan de Azúcar and Choros, during period 0 and period 1, respectively.

both panmixia and higher complexity models. It should be noted, however, that our sampling design does not allow us to positively differentiate a two-population model from the discrete sampling of an isolation-by-distance gradient, that may have formerly included the intermediate (but extinct) Isla Chañaral population, as well as the current Isla Grande de Atacama population.

Isolation and migration analysis. *Migrate-n* models of connectivity support a structure with low migration between separated populations. Bayes factor (BF) model comparison across three independent replicates allows unambiguous rejection of both the panmixia and full isolation hypotheses. BF did not allow a clear choice between symmetrical and asymmetrical migrations, although since BF does not take into account the number of estimated parameters, the lower-parameterization symmetrical migration model should be preferred. Under this model, $\Theta_{\text{CHR}} = 0.00697 \pm 8.0829e^{-5}$, $\Theta_{\text{AZR}} = 0.00747 \pm 0.000212$ and $M = 2075 \pm 10.6$. Using the reconstructed substitution rate for highly polymorphic RAD loci $\mu = 3.55 \cdot e^{-8}$ substitutions site⁻¹.generation⁻¹, inferred effective population sizes scale to $N_{\text{CHR}} \approx 58,600$ and $N_{\text{AZR}} \approx 62,800$, with a symmetric migration rate of $M \approx 3\%$ effective migrants per generation.

Joint minor-allele frequency spectrum-based inferences (as performed in *fastsimcoal2*) were globally consistent with this model, both in terms of population sizes and migration rates, which fall within the same orders of magnitude. Parameter values and confidence intervals are given in Table 3. AIC supports a model with a recent change in the migration regime (about 200 generations ago), but no population size change (our fourth model, see Fig. 4A). Population sizes are widely asymmetric despite equal prior specification. Migration from Choros to Pan de Azúcar is inferred stable, but increases at least ten-fold from Pan de Azúcar to Choros. The second-best AIC score supports a model with changes in migration regime and in population size (model 6, see Fig. 4B): inferred migration patterns were similar, although change occurs at a slightly more recent time. Population size in Pan de Azúcar was stable or increased slightly, but decreased in Choros.

Discussion

A highly structured population. The range-wide colonies of the Peruvian diving petrel stand out as remarkably structured, despite short geographic distances and a tumultuous recent history. Our study brings strong support to the idea of distinctive populations in each island group and limited gene flow between them. Isla Grande de Atacama, however, stands out as an exception: the small population seems mostly a sink for migrant birds from all other populations, and may stem from a recent re-colonization event. Genome-wide SNP-based analyses are in full accordance with mtDNA sequence analyses, and support a complete sorting of individuals between Isla Choros and Isla Pan de Azúcar in Chile, albeit with moderate differentiation.

The good agreement between nuclear and mitochondrial signals in the Isla Choros - Isla Pan de Azúcar comparison allows us to rely confidently on mtDNA sequence data alone for range-wide analysis. Some aspects, however, deserve consideration. The clearer sorting of individuals observed with nuclear data compared to mtDNA (see e.g. Fig. 3B vs. Fig. 3D) is most likely a consequence of the much higher resolution of the large neutral nuclear marker dataset compared to the relatively short and non-recombining functional cytochrome-b gene sequence. Sex-specific dispersal has often been proposed as an explanation in cases of discrepancies between nuclear and mitochondrial signals. However, this interpretation has been questioned⁹⁶, and the difference in coalescence rate of haploid and diploid markers has been convincingly proposed instead. In our case, the much larger sample size used for mtDNA analysis (83 individuals for Choros-Azúcar comparison) compared to nuclear analysis (21 individuals) may also account for a large part of this apparent difference. The resolution and the strength of the signal, however, should not be confused: the weaker, but clearer nuclear signal may characterize the true isolation processes more accurately than the larger and more contrasted mitochondrial dataset⁹⁴.

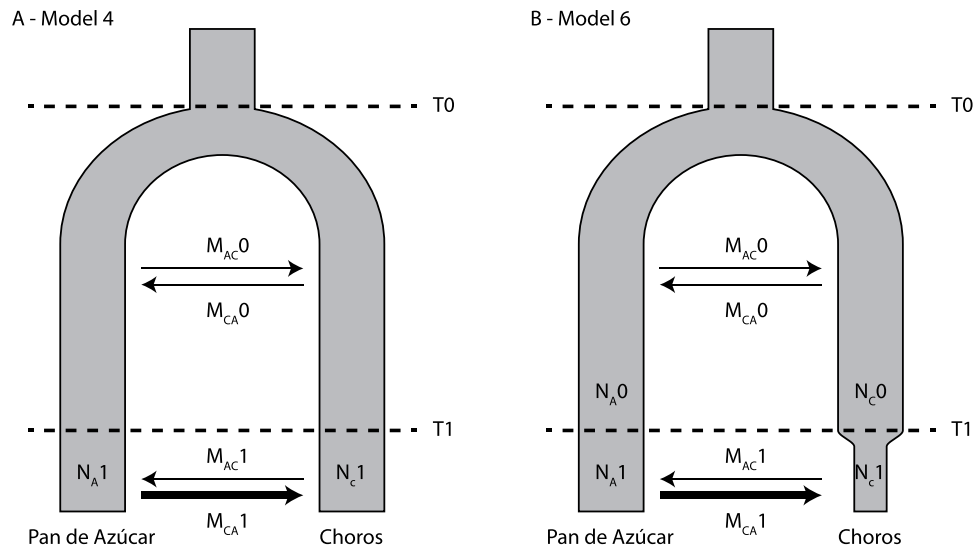


Figure 4. Selected *fastsimcoal2* models of demographic history. (A) Model 4 infers a change in the rate of migration from Isla Pan de Azúcar to Isla Choros in a recent past, and (B) model 6 infers change in both migration rate from Isla Pan de Azúcar to Isla Choros, and in population size at Isla Choros. Parameter names and values are detailed in Table 3.

Reconstructed population sizes and gene flow between Isla Choros and Isla Pan de Azúcar are globally consistent between haplotype-based and spectrum-based approaches, although with considerably higher precision for the spectrum-based inferences. Both *Migrate-n* and *fastsimcoal2* models allow us to reject panmixia as well as full isolation between colonies, and support partial isolation, with ongoing gene flow. However, the models differ in two important points. First, *Migrate-n* suggests equal population sizes, with $N_e \approx 60,000$ for each population, while *fastsimcoal2* infers unequal (yet stable) population sizes, with $N_{CHR} \approx 90,000$ and $N_{AZR} \approx 1,000$. Despite these contrasting results, however, both models propose an effective population size that is considerably higher than the present-day census: a total population of $\sim 90,000$ to $\sim 120,000$ effective breeders is in clear opposition to the currently observed population of $\sim 19,000$ breeders in Isla Choros⁴⁷ and ~ 4400 in Isla Pan de Azúcar⁴⁵, but more in tune with the historical populations (e.g. 100,000 breeders on the major breeding location of Isla Chañaral in 1938, and similarly large colonies elsewhere³⁸)—a direct testimony to the rapidity and extent of the population collapse in the past decades⁹⁷. Second, *Migrate-n* inferred symmetric and relatively high gene flow ($\sim 3\%$ per generation), whereas *fastsimcoal2* supports much lower, asymmetric gene flow: northward migration (from Isla Choros to Isla Pan de Azúcar) is stable and very low ($\sim 0.001\%$ per generation), while southward migration is higher, and increases from $\sim 0.03\%$ to $\sim 0.4\%$ in the recent past. Thus the haplotype-based reconstruction generally suggests a “homogenized” system, with population sizes averaged by higher gene flow. In contrast, the spectrum-based approach offers a more nuanced reconstruction that better reflects the observed state of the populations. The higher resolution of the spectrum-based approach, however, is expected. First, it is able to use the information from the full RADome, and not only from a random subset of loci. Second, haplotype-based methods chiefly capture events occurring at a substitution (i.e. millennial) scales, while spectrum-based approaches also capture drift in allele frequencies that may occur in a few generations, thus making them more suitable for the recovery of recent population history. Therefore, we may more appropriately consider *Migrate-n* reconstructions as a general framework that does not contradict the more precise reconstruction given by our *fastsimcoal2* model.

A northward colonization from Chile. Our results allow us to understand better the recent history of the species in its entire range. First, the large divergent gene pool of Isla Pan de Azúcar excludes the idea of a local extinction or near-extinction followed by gradual re-colonization on that island. Indeed, local divergence on Isla Pan de Azúcar is comparable to that of the “refugial” Isla Choros population, and models proposing more recent colonization or steep population growth in the former are systematically rejected. Thus our data strongly support the idea that the Peruvian diving petrel was, in fact, always remnant in Isla Pan de Azúcar, despite having been overlooked by several successive surveys, probably made outside the peak breeding activity season. Our data also suggest that the minimum population size in Isla Pan de Azúcar never went under a critical threshold, since no loss of diversity appears in our data compared to historically more stable populations (e.g. Isla Choros or the Peruvian islands). If indeed the present-day population is growing (as suggested by the comparison of recent studies^{45,50,51}) and not just stagnating, it is receiving very few migrants from other colonies in the process. This is in contrast to the small colony of Isla Grande de Atacama, which seems to be entirely made up of migrants from nearby colonies, and may well have gone extinct at some point in the century - it was *de facto* first identified in 2000¹⁵. Finally, although fueled by immigration from Chile in the past, the Peruvian colonies clearly exhibit signs consistent with the founder effect, with most samples from Isla La Vieja and Isla San Gallán stemming back to a single common ancestral haplotype.

Conservation actions must be island-based. The observed level of population isolation has important implications for the conservation of the Peruvian diving petrel. While it has been suggested that the panmictic character of seabird populations (such as the HCS populations of Peruvian pelican or Peruvian booby) increases their robustness in the face of environmental change^{32,55}, the high fragmentation of the Peruvian diving petrel population, on the other hand, should add to its already great vulnerability to land and marine habitat destruction. Indeed, high philopatry and low dispersal imply that the loss of a safe breeding island associated with a productive foraging area will lead to a local extinction and irremediable loss of genetic diversity.

An associated prediction is that current populations are demographically independent⁹⁸. The observed local trajectories are unlikely to be an artifact of changing migratory fluxes: they should represent the true local demographic trend. However, this observation is double-edged. In an increasingly well-monitored population such as the one in Isla Choros, the local census increase⁴⁷ can confidently be interpreted as the positive outcome of successful conservation measures - in particular the creation of the Humboldt Penguin National Reserve (that includes Isla Choros and the surrounding marine area). This is a strong encouragement to persevere in such endeavors. However, this demographic independence leaves us with serious doubts when censuses are inconsistent over time, as is the case in Isla Pan de Azúcar, as we cannot account for large and rapid population size changes in such a long-lived species without inferring important migratory flux. A conservative interpretation of census data is necessary in this case, and we should consider that faulty or mistimed surveys better account for the observed census changes¹⁴. Thus the population increase in Pan de Azúcar, and probably also in the Peruvian islands, should be interpreted carefully, and should not lead us to reconsider the species' status until more reliable time-series data are available.

Most importantly, our results make it especially clear that any conservation strategy for the Peruvian diving petrel in the HCS should focus on each island and the surrounding marine areas⁷ as a significant conservation unit. Within islands, colonies do not stand out as genetically separate entities, although philopatric behavior may still apply at that scale for stretches of a few generations. But the distinctness of each island's gene pool, as well as their inferred demographic independence, makes the conservation of each of the seven remaining breeding islands an absolute necessity for the survival of this once extremely abundant and now endangered seabird species. Indeed, despite the positive trend observed in the last decade on Isla Choros, important threats remain both for the nesting sites and for the marine habitats of diving petrels. One of the four islands where petrels nest (Isla Pájaros II, currently with no legal protection) is still under a valid permit of guano extraction, and ongoing extraction of guano was confirmed on the island as recently as 2003¹⁵. Future plans for mining projects include building two large harbors (<http://www.conocedominga.cl>, and <http://www.capmineria.cl/proyecto-puerto-cruz-grande/>) near the central breeding location of the specie, Isla Choros. This could affect significantly both the coastal foraging areas of the specie (through increased marine transit) and nesting areas (through direct effects of increased human intrusion), and could also increase mortality induced by lights. Our results clearly show that the conservation of every breeding location is essential for preserving the full remaining genetic diversity of this already severely reduced specie, and maximizing the chances of recovery after its massive historical decline. Thus they add compelling support to previous recommendations of the need to provide formal protection to several HCS islands that harbor a biodiversity of both regional and global importance, including colonies of endemic seabirds, and that urgently need strict legal protection status⁷.

Data Availability

All data used in present study are available from the NCBI portal. Mitochondrial sequences have reference numbers MK113717 to MK113815, while raw Illumina reads are available from the Sequence Read Archive (www.ncbi.nlm.nih.gov/sra) under the BioProject accession number PRJNA498806. Input files for *fastsimcoal2* are available from www.figshare.com, with DOI 10.6084/m9.figshare.7262138.

References

1. Albert, A., Echevin, V., Lévy, M. & Aumont, O. Impact of nearshore wind stress curl on coastal circulation and primary productivity in the Peru upwelling system. *Journal of Geophysical Research: Oceans* **115** (2010).
2. Thiel, M. *et al.* The Humboldt Current System of northern and central Chile: oceanographic processes, ecological interactions and socioeconomic feedback. *Oceanography and Marine Biology* **45**, 195–344 (2007).
3. Gutiérrez, D., Akester, M. & Naranjo, L. Productivity and sustainable management of the Humboldt Current large marine ecosystem under climate change. *Environmental Development* **17**, 126–144 (2016).
4. Montecino, V. *et al.* Bio-physical interactions off western South-America. *The sea* **14**, 329–390 (2006).
5. Luna-Jorquera, G., Simeone, A. & Aguilar, R. Ecofisiología de animales endotermos en un desierto cálido y un mar frío: el caso de las aves marinas de la corriente de Humboldt. *Fisiología ecológica y evolutiva. Teoría y casos de estudios en animales* 297–316 (2003).
6. Weichler, T., Garthe, S., Luna-Jorquera, G. & Moraga, J. Seabird distribution on the Humboldt Current in northern Chile in relation to hydrography, productivity, and fisheries. *ICES Journal of Marine Science: Journal du Conseil* **61**, 148–154 (2004).
7. Luna-Jorquera, G., Fernández, C. E. & Rivadeneira, M. M. Determinants of the diversity of plants, birds and mammals of coastal islands of the Humboldt current systems: implications for conservation. *Biodiversity and Conservation* **21**, 13–32 (2012).
8. Olson, D. M. & Dinerstein, E. The Global 200: A Representation Approach to Conserving the Earth's Most Biologically Valuable Ecoregions. *Conservation Biology* **12**, 502–515 (1998).
9. Spalding, M. D. *et al.* Marine ecoregions of the world: a bioregionalization of coastal and shelf areas. *BioScience* **57**, 573–583 (2007).
10. Aravena, G., Broitman, B. & Stenseth, N. C. Twelve years of change in coastal upwelling along the central-northern coast of Chile: spatially heterogeneous responses to climatic variability. *PLoS One* **9**, e90276 (2014).
11. Parsons, M. *et al.* Seabirds as indicators of the marine environment. *ICES Journal of Marine Science* **65**, 1520–1526 (2008).
12. Mallory, M. L., Robinson, S. A., Hebert, C. E. & Forbes, M. R. Seabirds as indicators of aquatic ecosystem conditions: a case for gathering multiple proxies of seabird health. *Mar Pollut Bull* **60**, 7–12 (2010).
13. Cury, P. M. *et al.* Global seabird response to forage fish depletion—one-third for the birds. *Science* **334**, 1703–1706 (2011).
14. Mattern, T., Ellenberg, U., Luna-Jorquera, G. & Davis, L. P. Chile: Recent Increase or Past Underestimate of Penguin Numbers? *Waterbirds* **27**, 368–376 (2004).

15. Simeone, A. *et al.* Breeding distribution and abundance of seabirds on islands off north-central Chile. *Revista Chilena de Historia Natural* **76**, 323–333 (2003).
16. Blais, J. M. *et al.* Arctic seabirds transport marine-derived contaminants. *Science* **309**, 445–445 (2005).
17. Ellis, J. C. Marine birds on land: a review of plant biomass, species richness, and community composition in seabird colonies. *Plant Ecology* **181**, 227–241 (2005).
18. McCauley, D. J. *et al.* From wing to wing: the persistence of long ecological interaction chains in less-disturbed ecosystems. *Sci Rep* **2**, 409 (2012).
19. Trathan, P. N. *et al.* Pollution, habitat loss, fishing, and climate change as critical threats to penguins. *Conserv Biol* **29**, 31–41 (2015).
20. Tasker, M. L. *et al.* The impacts of fishing on marine birds. *ICES Journal of Marine Science: Journal du Conseil* **57**, 531–547 (2000).
21. Becker, B. H. & Beissinger, S. R. Centennial decline in the trophic level of an endangered seabird after fisheries decline. *Conservation Biology* **20**, 470–479 (2006).
22. Karpouzi, V. S., Watson, R. & Pauly, D. Modelling and mapping resource overlap between seabirds and fisheries on a global scale: a preliminary assessment. *Marine Ecology Progress Series* **343**, 87–99 (2007).
23. Wagner, E. L. & Boersma, P. D. Effects of fisheries on seabird community ecology. *Reviews in Fisheries Science* **19**, 157–167 (2011).
24. Weimerskirch, H., Catard, A., Prince, P. A., Cherel, Y. & Croxall, J. P. Foraging white-chinned petrels (*Procellaria aequinoctialis*) at risk: from the tropics to Antarctica. *Biological Conservation* **87**, 273–275 (1999).
25. Lewison, R. L., Crowder, L. B., Read, A. J. & Freeman, S. A. Understanding impacts of fisheries bycatch on marine megafauna. *Trends in Ecology & Evolution* **19**, 598–604 (2004).
26. Le Bohec, C. *et al.* King penguin population threatened by Southern Ocean warming. *Proc Natl Acad Sci USA* **105**, 2493–2497 (2008).
27. Sydeman, W. J., Thompson, S. A. & Kitaysky, A. Seabirds and climate change: roadmap for the future. *Marine Ecology Progress Series* **454**, 107–117 (2012).
28. Barbosa, A. & Palacios, M. J. Health of Antarctic birds: a review of their parasites, pathogens and diseases. *Polar Biology* **32**, 1095–1115 (2009).
29. Ancel, A. *et al.* Emperors in hiding: when ice-breakers and satellites complement each other in Antarctic exploration. *PLoS One* **9**, e100404 (2014).
30. Croxall, J. P. *et al.* Seabird conservation status, threats and priority actions: a global assessment. *Bird Conservation International* **22**, 1–34 (2012).
31. Lynch, H. J., Naveen, R., Trathan, P. N. & Fagan, W. F. Spatially integrated assessment reveals widespread changes in penguin populations on the Antarctic Peninsula. *Ecology* **93**, 1367–1377 (2012).
32. Cristofari, R. *et al.* Full circumpolar migration ensures evolutionary unity in the Emperor penguin. *Nature Communications* **7**, 11842 (2016).
33. Coker, R. Habits and economic relations of the guano birds of Peru. *Proceedings of the US National Museum* **56**, 449:511 (1919).
34. Murphy, R. C. *Oceanic birds of South America: a study of species of the related coasts and seas, including the American quadrant of Antarctica, based upon the Brewster-Sanford collection in the American Museum of Natural History* (Macmillan Co., 1936).
35. Hays, C. The Peruvian diving petrel in Peru. *Oryx* **23**, 102–105 (1989).
36. Figueroa, J., Suazo, E. & Santillan, L. Registros actuales e históricos del Potoyunco peruano *Pelecanoides Garnotii* en el Perú. *The Biologist* **9**, 19–37 (2011).
37. García-Godos, I. & Goya, E. Diet of the Peruvian diving petrel *Pelecanoides garnotii* at La Vieja island, Peru, 1997–2000: potential fishery interactions and conservation implications. *Marine Ornithology* **34**, 33–41 (2006).
38. Modinger, B. A. & Duffy, D. C. Animal introduction to Isla Chanaral, Chile; their history and effect on seabirds. *Cormorant* **15**, 3–6 (1987).
39. Fernández Zamora, C. Midiendo la densidad reproductiva del petrel buceador (*Pelecanoides garnotii*) antes de la erradicación de conejos europeos sobre las islas de Coquimbo, Chile. *Post-graduate thesis, Universidad Católica del Norte, Coquimbo, Chile* (2011).
40. Jahncke, J. & Goya, E. The status of the Peruvian diving-petrel population at its main breeding areas along the coast of Peru. *Colonial Waterbirds* 94–97 (1998).
41. Vilina, Y. A. Status of the Peruvian diving petrel, *Pelecanoides garnotii*, on Chañaral Island, Chile. *Colonial Waterbirds* 137–139 (1992).
42. Figueroa, J. & Stucchi, M. Peruvian diving petrels *Pelecanoides garnotii* on Lobos de afuera islands, Peru. *Marine Ornithology* **36**, 189–190 (2008).
43. Valverde Romero, M. B. First record of the endangered Peruvian diving-petrel *Pelecanoides garnotii* breeding on Corcovado island, Peru. *Marine Ornithology* **34**, 75–76 (2006).
44. The International Union for the Conservation of Nature Red List. (2018)
45. Mattern, T., Ellenberg, U. & Luna-Jorquera, G. A South American marine otter *Lontra felina* preys upon chicks of the Peruvian diving petrel *Pelecanoides garnotii*. *Marine Ornithology* **30**, 95–96 (2002).
46. Luna-Jorquera, G. & Cortes, M. Estudio del ensamble de aves y mamíferos marinos al interior del Área Marina y Costera Protegida de Múltiples Usos Isla Grande de Atacama. *Proyecto Conservación de la Biodiversidad de Importancia Mundial a lo largo de la Costa Chilena* (2007).
47. Martínez Palma, P. *Reporte Anual Regional del cumplimiento de las actividades del Plan de Conservación de la especie Pelecanoides garnotii (Yunco)*. Departamento de Areas Silvestres Protegidas, region Coquimbo (2014).
48. Figueroa, J. & Suazo, E. Distribución De Las Aves Marinas Frente A La Costa Norte-Centro Del Perú En El Invierno De 2010 Y Su Interacción Potencial Con La Pesquería. *The Biologist* **10**, 41–71 (2012).
49. Fernández, C., Miranda Urbina, D., Portflitt Toro, M. & Luna-Jorquera, G. Estimación de la densidad reproductiva del yunco *Pelecanoides garnotii* en Isla Choro. *XI Congreso Chileno de Ornitología* **24** (2014).
50. Páez, J., Castillo, A. & Alcota, F. Nuevos antecedentes sobre los sitios de nidificación del Pato Yunco (*Pelecanoides garnotii*, Lesson 1832) en Isla Pan de Azúcar, Parque Nacional Pan de Azúcar, III Region. (2006).
51. Turgis, Y. & Téran, D. Informe de inspección de nidificación del Yunco (*Pelecanoides garnotii*) en Isla Pan de Azúcar, Parque Nacional Pan de Azúcar. *CONAF* (2009).
52. Friesen, V. L. Speciation in seabirds: why are there so many species and why aren't there more. *J Ornithol* **156**, 27–39 (2015).
53. Friesen, V. L., Burg, T. M. & McCoy, K. D. Mechanisms of population differentiation in seabirds. *Molecular Ecology* **16**, 1765–1785 (2007).
54. Taylor, S. A. *et al.* Panmixia and high genetic diversity in a Humboldt Current endemic, the Peruvian Booby (*Sula variegata*). *Journal of Ornithology* 1–8 (2010).
55. Jeyasingham, W. S., Taylor, S. A., Zavalaga, C. B., Simeone, A. & Friesen, V. L. Specialization to cold water upwellings may facilitate gene flow in seabirds: new evidence from the Peruvian pelican *Pelecanus thagus* (Pelecaniformes: Pelecanidae). *Journal of avian biology* **44**, 297–304 (2013).
56. Schlosser, J. A. *et al.* Evidence for gene flow differs from observed dispersal patterns in the Humboldt penguin, *Spheniscus humboldti*. *Conservation Genetics* **10**, 839–849 (2009).
57. Baird, N. A. *et al.* Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* **3**, e3376 (2008).
58. Nunn, G. B. & Stanley, S. E. Body size effects and rates of cytochrome b evolution in tube-nosed seabirds. *Molecular Biology and Evolution* **15**, 1360–1371 (1998).

59. Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W. & Postlethwait, J. H. Stacks: building and genotyping loci de novo from short-read sequences. *G3: Genes, Genomes, Genetics* **1**, 171–182 (2011).
60. Catchen, J. M., Hohenlohe, P. A., Bassham, S., Amores, A. & Cresko, W. A. Stacks: an analysis tool set for population genomics. *Molecular ecology* **22**, 3124–3140 (2013).
61. Hoffman, J. I. *et al.* High-throughput sequencing reveals inbreeding depression in a natural population. *Proc Natl Acad Sci USA* **111**, 3775–3780 (2014).
62. Paris, J. R., Stevens, J. R. & Catchen, J. M. Lost in parameter space: a road map for STACKS. *Methods in Ecology and Evolution* **8**, 1360–1373 (2017).
63. Zerbino, D. R. & Birney, E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome research* **18**, 821–829 (2008).
64. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature methods* **9**, 357–359 (2012).
65. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
66. DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature genetics* **43**, 491–498 (2011).
67. Kim, S. Y. *et al.* Estimation of allele frequency and association mapping using next-generation sequencing data. *BMC bioinformatics* **12**, 231 (2011).
68. Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* **15**, 356 (2014).
69. Nielsen, R., Korneliussen, T., Albrechtsen, A., Li, Y. & Wang, J. SNP calling, genotype calling, and sample allele frequency estimation from new-generation sequencing data. *PLoS one* **7**, e37558 (2012).
70. Rozas, J. & Rozas, R. DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* **15**, 174–175 (1999).
71. Jombart, T. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**, 1403–1405 (2008).
72. Excoffier, L., Laval, G. & Schneider, S. Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol Bioinform Online* **1**, 47–50 (2005).
73. Benjamini, Y. & Yekutieli, D. The control of false discovery rate in multiple testing under dependency. *The Annals of Statistics* **29**(4), 1165–1188 (2001).
74. Pollard, K. S., Dudoit, S., van der Laan, M. J. Multiple Testing Procedures: R multtest Package and Applications to Genomics, in *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*, Gentleman, R., Carey, V., Huber, W., Irizarry, R., Dudoit, S. (Editors). Springer, pp. 251–272 (2005).
75. Huson, D. H. & Bryant, D. Application of phylogenetic networks in evolutionary studies. *Molecular biology and evolution* **23**, 254–267 (2006).
76. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559–575 (2007).
77. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
78. Posada, D. jModelTest: phylogenetic model averaging. *Mol Biol Evol* **25**, 1253–1256 (2008).
79. Matschiner, M. Fitchi: Haplotype genealogies based on Fitch distances. (2015).
80. Fumagalli, M., Vieira, F. G., Linderoth, T. & Nielsen, R. ngsTools: methods for population genetics analyses from Next-Generation Sequencing data. *Bioinformatics* **1–2** (2014).
81. Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959 (2000).
82. Raj, A., Stephens, M. & Pritchard, J. K. fast STRUCTURE: Variational Inference of Population Structure in Large SNP Data Sets. *Genetics* **197**, 573–589 (2014).
83. Skotte, L., Korneliussen, T. S. & Albrechtsen, A. Estimating individual admixture proportions from next generation sequencing data. *Genetics* **195**, 693–702 (2013).
84. Evanno, G., Regnaut, S. & Goudet, J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* **14**, 2611–2620 (2005).
85. Liu, N. & Zhao, H. A non-parametric approach to population structure inference using multilocus genotypes. *Hum Genomics* **2**, 353 (2006).
86. Beerli, P. & Palczewski, M. Unified framework to evaluate panmixia and migration direction among multiple sampling locations. *Genetics* **185**, 313–326 (2010).
87. Trucchi, E. *et al.* King penguin demography since the last glaciation inferred from genome-wide data. *Proceedings of the Royal Society B: Biological Sciences* **281**, 20140528 (2014).
88. Kass, R. E. & Raftery, A. E. Bayes factors. *Journal of the American statistical association* **90**, 773–795 (1995).
89. Excoffier, L., Dupanloup, I., Huerta-Sanchez, E., Sousa, V. C. & Foll, M. Robust demographic inference from genomic and SNP data. *PLoS Genet* **9**, e1003905 (2013).
90. Zhang, G. *et al.* Comparative genomics reveals insights into avian genome evolution and adaptation. *Science* **346**, 1311–1320 (2014).
91. Welch, A. J., Yoshida, A. A. & Fleischer, R. C. Mitochondrial and nuclear DNA sequences reveal recent divergence in morphologically indistinguishable petrels. *Molecular Ecology* **20**, 1364–1377 (2011).
92. Lemey, P., Rambaut, A., Drummond, A. J. & Suchard, M. A. Bayesian phylogeography finds its roots. *PLoS Comput Biol* **5**, e1000520 (2009).
93. Bouckaert, R. *et al.* BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput Biol* **10**, e1003537 (2014).
94. Subramanian, S. The effects of sample size on population genomic analyses – implications for the tests of neutrality, *BMC Genomics* **17**(123) (2016).
95. Hoban, S. M., Gaggiotti, O. E. & Bertorelle, G. The number of markers and samples needed for detecting bottlenecks under realistic scenarios, with and without recovery: a simulation based study. *Molecular ecology* **22**, 3444–3450 (2013).
96. Zink, R. M. & Barrowclough, G. F. Mitochondrial DNA under siege in avian phylogeography. *Molecular Ecology* **17**, 2107–2121 (2008).
97. Roman, J. & Palumbi, S. R. Whales before whaling in the North Atlantic. *Science* **301**, 508–510 (2003).
98. Waples, R. S. & Gaggiotti, O. What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular ecology* **15**, 1419–1439 (2006).

Acknowledgements

We thank Diego Miranda-Urbina, Nicole Licuime Castillo, Gina Mori, Celia Caceres, Francisco Bermedo and Natalia Ortiz for their invaluable help on the field, as well as Rasmie Hereme at the CEAZA, Morten Skage, Marianne Selander Hansen and Ave Tooming-Klunderud of the Norwegian Sequencing Center for their assistance in the laboratory. This study was conducted with the support of the French National Research Agency (ANR) “PICASO” grant (ANR-2010-BLAN-1728-01), from Marie Curie Intra European Fellowships (FP7-PEOPLE-IEF-2008, European Commission; project no. 235962 to CLB and FP7-PEOPLE-IEF-2010, European

Commission; project no. 252252 to ET), from the Centre Scientifique de Monaco through funds allocated to the Laboratoire International Associé 647 'BioSensib' (CSM/CNRS-University of Strasbourg), and the Réseau thématique pluridisciplinaire international "NUTrition et RESistance aux Stress environnementaux" (RTPI NUTRESS) (CSM/CNRS-University of Strasbourg). This study was performed on the Abel Cluster, owned by the University of Oslo and the Norwegian metacenter for High Performance Computing (NOTUR), and operated by the Department for Research Computing at USIT, the University of Oslo. The work in Peru was supported by a grant from American Bird Conservancy #374D to Carlos Zavalaga. The work in Chile was supported by the Millennium Nucleus for Ecology and Sustainable Management of Oceanic Island ESMOI-ICM.

Author Contributions

R.C., C.L.B., and G.L.-J. planned and designed the study; R.C. and P.P. executed field and laboratory research and contributed equally to the study; C.E.F., C.Z. and J.A.S. contributed to field research; E.T. and N.G. contributed to laboratory research. R.C. and G.L.-J. drafted and revised the manuscript. All coauthors reviewed and gave final approval for the manuscript.

Additional Information

Competing Interests: The authors declare that the research was conducted in the absence of any competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019