

# Additive Traits Lead to Feeding Advantage and Reproductive Isolation, Promoting Homoploid Hybrid Speciation

Juan F. Masello,<sup>1</sup> Petra Quillfeldt,<sup>1</sup> Edson Sandoval-Castellanos,<sup>2</sup> Rachael Alderman,<sup>3</sup> Luciano Calderón,<sup>1</sup> Yves Chereil,<sup>4</sup> Theresa L. Cole,<sup>1,5</sup> Richard J. Cuthbert,<sup>6</sup> Manuel Marin,<sup>7,8</sup> Melanie Massaro,<sup>9</sup> Joan Navarro,<sup>10</sup> Richard A. Phillips,<sup>11</sup> Peter G. Ryan,<sup>12</sup> Lara D. Shepherd,<sup>13</sup> Cristián G. Suazo,<sup>1</sup> Henri Weimerskirch,<sup>4</sup> and Yoshan Moodley<sup>\*,14</sup>

<sup>1</sup>Department of Animal Ecology & Systematics, Justus Liebig University Giessen, Germany

<sup>2</sup>Institute of Animal Breeding and Husbandry, University of Kiel, Germany

<sup>3</sup>Department of Primary Industries, Parks, Water and Environment, Hobart, TAS, Australia

<sup>4</sup>Centre d'Etudes Biologiques de Chizé, UMR 7372 CNRS-Université de La Rochelle, Villiers-en-Bois, France

<sup>5</sup>Manaaki Whenua Landcare Research, Canterbury, New Zealand

<sup>6</sup>Royal Society for the Protection of Birds (RSPB), The Lodge, Sandy, Bedfordshire, United Kingdom

<sup>7</sup>Section of Ornithology, Natural History Museum of Los Angeles County, Los Angeles, CA

<sup>8</sup>Feather Link Inc., Cincinnati, OH

<sup>9</sup>Institute for Land, Water and Society, School of Environmental Sciences, Charles Sturt University, Albury, NSW, Australia

<sup>10</sup>Institut de Ciències del Mar CSIC, Barcelona, Spain

<sup>11</sup>British Antarctic Survey, Natural Environment Research Council, High Cross, Cambridge, United Kingdom

<sup>12</sup>FitzPatrick Institute of African Ornithology, DST-NRF Centre of Excellence, University of Cape Town, Rondebosch, Republic of South Africa

<sup>13</sup>Museum of New Zealand Te Papa Tongarewa, Wellington, New Zealand

<sup>14</sup>Department of Zoology, University of Venda, Private Bag X5050, Thohoyandou 0950, Republic of South Africa

\*Corresponding author: E-mail: yoshan.moodley@univen.ac.za.

Associate editor: Claudia Russo

## Abstract

Speciation through homoploid hybridization (HHS) is considered extremely rare in animals. This is mainly because the establishment of reproductive isolation as a product of hybridization is uncommon. Additionally, many traits are underpinned by polygeny and/or incomplete dominance, where the hybrid phenotype is an additive blend of parental characteristics. Phenotypically intermediate hybrids are usually at a fitness disadvantage compared with parental species and tend to vanish through backcrossing with parental population(s). It is therefore unknown whether the additive nature of hybrid traits in itself could lead successfully to HHS. Using a multi-marker genetic data set and a meta-analysis of diet and morphology, we investigated a potential case of HHS in the prions (*Pachyptila* spp.), seabirds distinguished by their bills, prey choice, and timing of breeding. Using approximate Bayesian computation, we show that the medium-billed Salvin's prion (*Pachyptila salvini*) could be a hybrid between the narrow-billed Antarctic prion (*Pachyptila desolata*) and broad-billed prion (*Pachyptila vittata*). Remarkably, *P. salvini*'s intermediate bill width has given it a feeding advantage with respect to the other *Pachyptila* species, allowing it to consume a broader range of prey, potentially increasing its fitness. Available metadata showed that *P. salvini* is also intermediate in breeding phenology and, with no overlap in breeding times, it is effectively reproductively isolated from either parental species through allochrony. These results provide evidence for a case of HHS in nature, and show for the first time that additivity of divergent parental traits alone can lead directly to increased hybrid fitness and reproductive isolation.

**Key words:** homoploid hybrid speciation, additive traits, increased hybrid fitness, Procellariiformes, reproductive isolation, seabird.

## Introduction

Homoploid hybrid speciation (HHS), where distinct species hybridize to form a new reproductively isolated species without a change in chromosome number, is rare in nature

(Rieseberg 1997; Gross and Rieseberg 2005). Zoologists of the 20th century generally believed that gene flow between species would counteract divergence (and speciation; Mayr 1963; Coyne and Orr 2004). The 21st century has thus far seen an increase in the number of putative cases of HHS,

suggesting that HHS could be more common than previously thought (Mavárez et al. 2006). However, despite this trend, compelling evidence for HHS is lacking in all but a few cases (Gross and Rieseberg 2005; Schwarz et al. 2005; Mavárez et al. 2006; Hermansen et al. 2011; Lamichhaney et al. 2018), as several stringent conditions must first be met (Abbott et al. 2013; Schumer et al. 2014).

Three main criteria should be satisfied for successful HHS: 1) reproductive isolation between hybrids and parental species, 2) genetic evidence of hybridization, 3) evidence of reproductive isolation as consequence of hybridization (Schumer et al. 2014). Additionally, it is important that parental populations are species in their own right, possessing distinguishable species-specific characteristics that are sufficiently divergent to yield novel phenotypic characteristics in the resulting hybrid upon secondary contact (Abbott et al. 2013). Yet, parental populations should also not be so divergent that they are reproductively isolated from each other and cannot produce viable hybrids. Moreover, the hybrid phenotype should also confer hybrids with a fitness at least as high as its parent species in order to proliferate and become established (Abbott et al. 2013). If the phenotypic integrity of the hybrids is to be maintained, that is, that the hybrids remain a distinct species in the future, a reduction of continued gene flow with parental populations is required to prevent the formation of a hybrid swarm, where the hybrid phenotype is continually eroded through parental backcrossing (Mallet 2007; Abbott et al. 2013; Schumer et al. 2014; Seehausen et al. 2014). Gene flow may be attenuated when hybridization itself results in the reproductive isolation of the hybrid phenotype. Simulation and empirical studies have shown that reproductive isolation can be achieved through hybridization in a number of ways, including recombinational speciation, where parental incompatibilities are sorted such that a subset act as barriers to gene flow against either parental species (Grant 1958; McCarthy et al. 1995; Schumer et al. 2015); by ecological speciation (Rieseberg 1997; Buerkle et al. 2000; Schwarz et al. 2005) or geographic isolation (Rieseberg 1997; Marques et al. 2016).

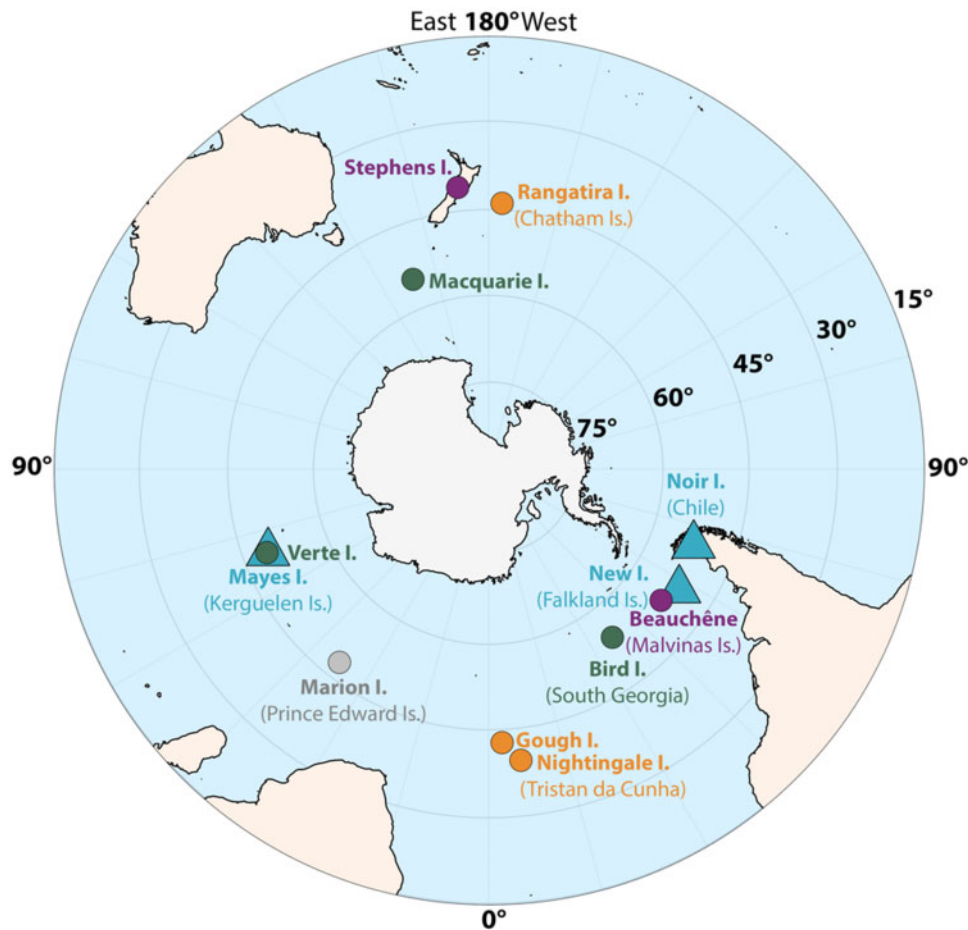
Successful HHS therefore depends on hybrid fitness and reproductive isolation of hybrids from parental populations. In nature, hybrids can show an additive blend of parental characteristics (Brelsford et al. 2011; Elgvin et al. 2011; Hermansen et al. 2011; Lavretsky et al. 2015), since many traits have a polygenic or incomplete dominant basis of inheritance (Anderson 1949). The hybrid phenotype in such cases is approximately intermediate to parents. It is logically rare that such additive hybrid traits could lead to an increase in fitness since intermediate hybrids may not be suited to either parental environment (Fisher 1930; Mayr 1942). It is more common that hybridization events facilitate speciation by producing hybrid phenotypes that are more prominent than those of the parental populations (transgressive hybrid) and have higher adaptive value (deVicente and Tanksley 1993; Rieseberg et al. 1999; Seehausen et al. 2014). Indeed, among some of the proposed cases of HHS, hybrid fitness increased relative to parental species through other genetic means, namely transgressive phenotypes (Nolte et al. 2006;

Lamichhaney et al. 2018); through mosaic phenotypes where hybrids resemble a codominant patchwork of both parental traits (Kunte et al. 2011); through the dominance of one or other parental phenotype and/or through the inheritance of entire linkage blocks (or speciation/divergence islands) that contain the molecular blueprints of each parental phenotype (Nadeau et al. 2012).

It is rarer still that additive traits lead to reproductive isolation (Gavrilets and Losos 2009), as intermediate hybrids will be less divergent and capable of backcrossing with either parental population, thus more likely to form a hybrid swarm (Mavárez and Linares 2008; Hermansen et al. 2011; Schumer et al. 2014). In two suggested cases of HHS where additive traits came into secondary contact, reproductive isolation was facilitated by geographic separation (Marques et al. 2016), and by sexual selection for the novel hybrid trait (Barrera-Guzmán et al. 2018), although the latter case has been recently contested (Rosenthal et al. 2018). Therefore, it has yet to be demonstrated that the additive nature of polygenic or incompletely dominant traits could in itself lead to reproductive isolation.

Adaptive radiations could promote HHS (Seehausen 2004), and play a role in several of the proposed cases (Howarth et al. 2005; Mavárez and Linares 2008; Lamichhaney et al. 2018). This is because phenotypic divergence between speciating lineages occurs faster than reproductive isolation, thereby allowing species to hybridize and produce viable offspring upon secondary contact. We focus on the prions (*Pachyptila* spp.), a monophyletic group of pelagic seabirds, whose distribution is exclusive to the Southern Ocean, where they breed on remote oceanic islands during the austral summer (Quillfeldt et al. 2014) (fig. 1; supplementary table S1, Supplementary Material online). Prions are masters in the art of pelagic flight covering vast oceanic distances in relatively short periods of time (Quillfeldt et al. 2014; Navarro et al. 2015).

Despite the potential for gene flow over vast distances (Quillfeldt et al. 2017), the *Pachyptila* radiation is characterized by the evolution of two disparate feeding strategies, reflecting pronounced differences in bill morphology. All prions possess palatal lamellae of varying size on their upper mandibles to help them filter feed (Murphy 1936). As lamellae are more effective when birds are able to filter larger quantities of water, bill width and bill length are highly correlated (fig. 2; supplementary table S2, Supplementary Material online), and lamellae, therefore, are more pronounced in those species with broader and longer bills (Klages and Cooper 1992). At the larger end of this morpho-ecological axis, the broad-billed prion *Pachyptila vittata* possesses highly pronounced lamellae and feeds almost exclusively on copepods. At the opposite end of the axis, the narrow-billed species such as the thin-billed prion *Pachyptila belcheri* and the fairy prion *Pachyptila turtur* possess only vestigial lamellae (Morgan and Ritz 1982), and are unable to filter feed (Ridoux 1994; Cherel et al. 2002). The diet of narrow-billed species consists mainly of euphausiids, decapods, hyperiid amphipods, cirripeds, small cephalopods and fish (Ridoux 1994; Cherel et al. 2002). The bill of the Antarctic Prion (*Pachyptila desolata*) is slightly broader than the narrowest-billed species (fig. 2;

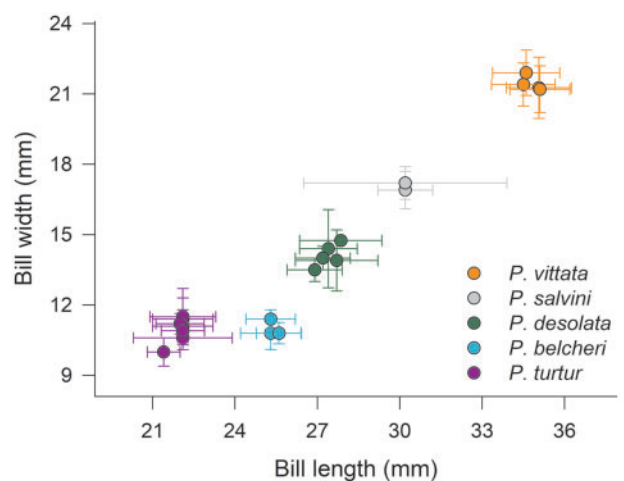


**FIG. 1.** *Pachyptila* (Aves: Procellariiformes) breeding colonies sampled in this study. The different species are color-coded. Thin-billed prions (*Pachyptila belcheri*): light blue triangles. Antarctic prions (*Pachyptila desolata*): light green circles. Broad-billed prions (*Pachyptila vittata*): orange circles. Salvin's prions (*Pachyptila salvini*): grey circle. Fairy prions (*Pachyptila turtur*): purple circles. Colonies are named after the island where they are located: Rangatira I. (Chatham Is.), Noir I. (Chile), New I. (Falkland/Malvinas Is.), Beauchêne I. (Falkland/Malvinas Is.), Bird I. (South Georgia/Georgias del Sur), Gough I., Nightingale I. (Tristan da Cunha), Marion I. (Prince Edward Is.), Verte I. (Kerguelen Is.), Mayes I. (Kerguelen Is.), Macquarie I., Stephens I. (New Zealand).

supplementary table S2, Supplementary Material online), allowing a degree of filter feeding (Cherel et al. 2002).

The Salvin's Prion (*Pachyptila salvini*) possesses a medium-width bill with pronounced palatal lamellae (fig. 2; Murphy 1936) potentially allowing it to feed on prey species available to both broad and narrow-billed species. Genetically, mitochondrial DNA (mtDNA) places *P. salvini* sister to *P. desolata*, yet nuclear microsatellites pair it with either *P. desolata* or *P. vittata*, depending on the method used to estimate genetic distance (Moodley et al. 2015). A hybrid origin for *P. salvini* could potentially explain these results.

Although the prions have attracted much phylogenetic and taxonomic attention (Murphy 1936; Falla 1940; Fleming 1941; Fullagar 1972; Cox 1980; Harper 1980; Jacob and Hoerschelmann 1982; Bretagnolle et al. 1990; Viot et al. 1993; Nunn and Stanley 1998; Penhallurick and Wink 2004; Rheindt and Austin 2005) a reliable species tree is still lacking. Consequently, most studies currently follow the taxonomy by Bretagnolle et al. (1990) which, based on morphometrics, breeding biology, vocalizations, and genetics, classify



**FIG. 2.** Bill width versus bill length for all *Pachyptila* (Aves, Procellariiformes) species sampled. Species are color-coded, circles denote means per colony, and error bars correspond to standard deviations. The means for each colony were obtained from the references in supplementary table S2, Supplementary Material online. Only data from live individuals or fresh corpses were included.

Antarctic, Salvin's, thin-billed, and fairy prions as distinct, yet closely related taxa.

In this study, we perform a morphological and genetic analysis of the *Pachyptila* adaptive radiation and present a remarkable case of HHS, where the additive effects of hybridizing two highly divergent parental species result in both a potentially fitter and reproductively isolated hybrid species.

## Results

### Morphology and Diet

Our meta-analysis of available morphological data confirmed that *Pachyptila* species are distinguishable (fig. 2) not only by bill width ( $F = 364.2$ ,  $n_{\text{individuals}} = 1,524$ ,  $n_{\text{populations}} = 24$ ,  $P < 0.001$ , multiway ANOVA) and bill length ( $F = 931.7$ ,  $n_{\text{individuals}} = 1,375$ ,  $n_{\text{populations}} = 25$ ,  $P < 0.001$ ), but also by wing length ( $F = 81.1$ ,  $n_{\text{individuals}} = 1,673$ ,  $n_{\text{populations}} = 25$ ,  $P < 0.001$ ), tarsus length ( $F = 26.4$ ,  $n_{\text{individuals}} = 1,010$ ,  $n_{\text{populations}} = 21$ ,  $P < 0.001$ ), and tail length ( $F = 6.2$ ,  $n_{\text{individuals}} = 914$ ,  $n_{\text{populations}} = 13$ ,  $P < 0.05$ ) (see details in [supplementary table S2, Supplementary Material online](#)).

All available data on the diet of the *Pachyptila* species studied here were considered in relation to bill morphology. We found that Salvin's prion, possessing a medium-width bill with pronounced palatal lamellae (fig. 3; [supplementary table S2, Supplementary Material online](#)), was able to feed on a larger range of prey species than any other narrow or broad-billed species (fig. 3; [supplementary table S3, Supplementary Material online](#)). Its diet included copepods, hyperiid and gammarid amphipods, euphausiids, cirripeds, isopods, chaetognaths, and small cephalopods and fish ([supplementary table S3, Supplementary Material online](#)). Consequently, the diet of *P. salvini* is 29–49% more diverse than either the

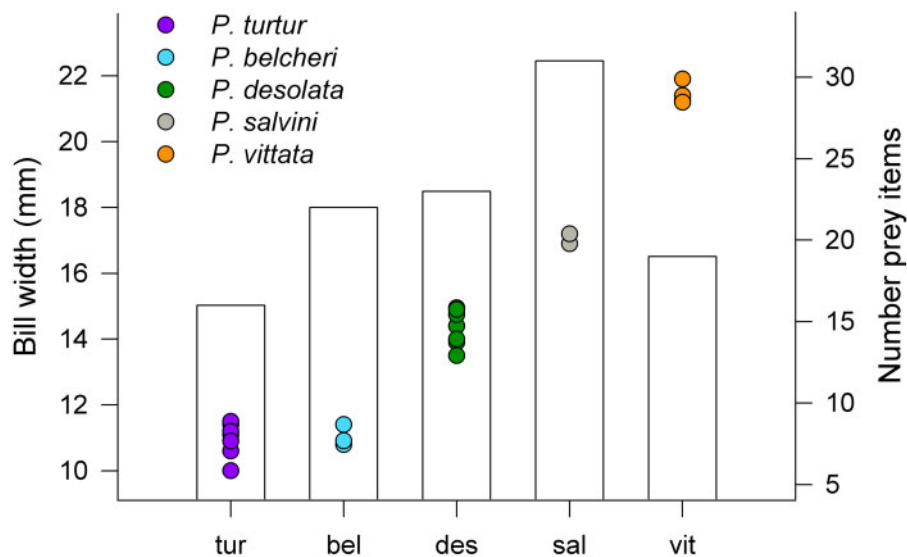
broader- or narrower-billed *Pachyptila* species (fig. 3; [supplementary table S3, Supplementary Material online](#)).

### A Genetic Analysis of the *Pachyptila* Radiation

We undertook a comprehensive genetic survey of the genus *Pachyptila* ([supplementary table S1, Supplementary Material online](#)) from across its distribution in the Southern Ocean (fig. 1). Our data set of five species within the genus, and the closely related blue petrel (*Halobaena caerulea*), consisted of 425 individuals genotyped at 25 microsatellite loci (for details see [Moodley et al. 2015](#)), and 2,842 bp of DNA sequence data (889 bp mtDNA, 1,953 bp nuclear introns; see [supplementary table S4; Supplementary file S2, Supplementary Material online](#)) from 32 individuals representing the diversity within each species.

### Genetic Diversity and Neutrality

Average microsatellite diversity indices showed similarly high values for most *Pachyptila* species, with the exception of fairy prions (*P. turtur*), where heterozygosity and allelic richness (AR) was lower ([supplementary table S5, Supplementary Material online](#)). For more detailed locus-specific values see [Moodley et al. \(2015\)](#). Private AR was generally low among *Pachyptila* species, but higher among the more distant *H. caerulea*. DNA sequence diversity differed greatly depending on species and locus ([supplementary table S5, Supplementary Material online](#)) with highest values observed in *P. turtur* (cytochrome *b*, *cyt b*, and Triosephosphate isomerase, *Tim4*), *P. salvini* (d-crystallin, *d-cryst*, and Ornithine decarboxylase, *OD67*), *P. vittata* (Lipoprotein lipase, *Lipo2*), and *P. belcheri* (Adenylate kinase, *Aden5*). Lowest values were consistently observed across all loci in the outgroup taxon (*H. caerulea*), despite sampling across this species' wide



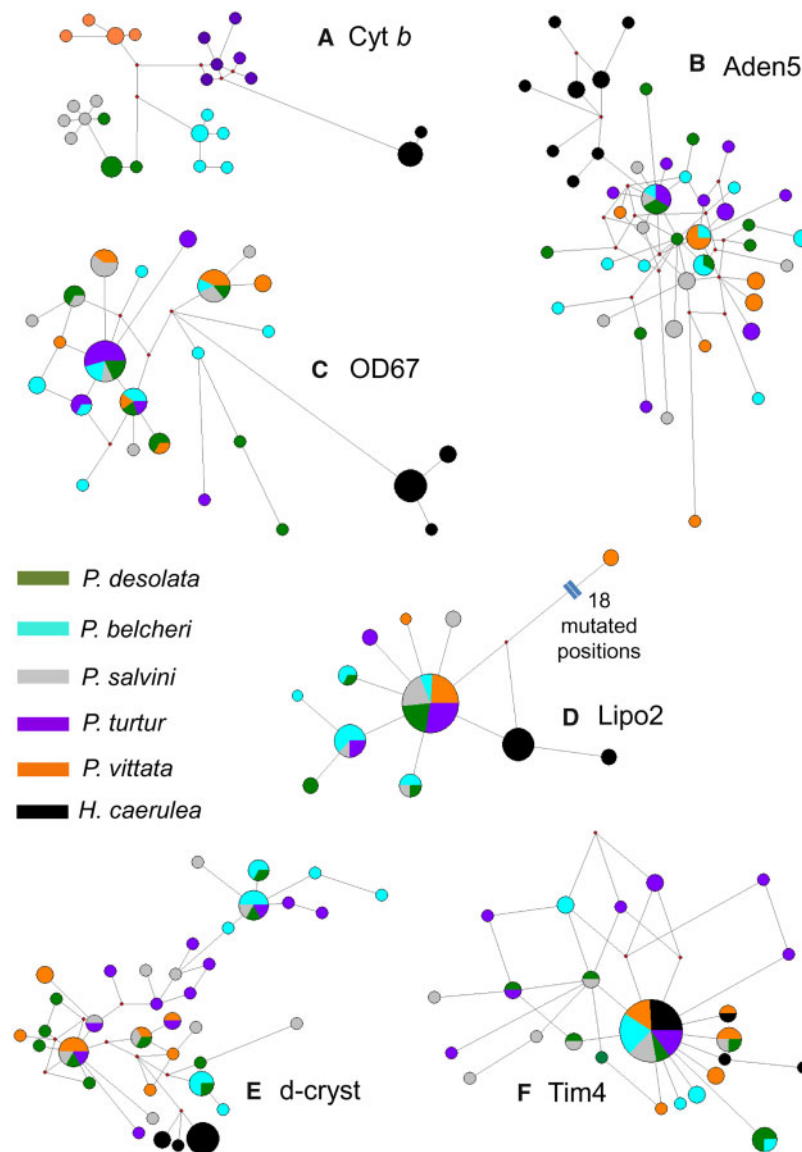
**Fig. 3.** Morphometry and diet. Mean bill width of all *Pachyptila* (Aves, Procellariiformes) species sampled at various colonies (circles color-coded per species), and the number of prey items included in the diet of each species (bars). Sources and sample sizes are provided in [supplementary tables S1 and S2, Supplementary Material online](#). tur: fairy prion *Pachyptila turtur*; bel: thin-billed prion *Pachyptila belcheri*; des: Antarctic prions *Pachyptila desolata*; sal: Salvin's prions *Pachyptila salvini*; vit: broad-billed prions *Pachyptila vittata*. Only data from live individuals or fresh corpses were included. Only prey items reported with an occurrence >3% were included in order to avoid occasional prey and items resulting from secondary ingestion (prey gut contents).

circumpolar distribution (Bird I. in South Georgia/Georgias del Sur, Mayes I. in the Kerguelen Is., and Marion I. in the Prince Edward Is.). Among loci, the mitochondrial *cyt b* locus was most polymorphic and diverse when averaged across all *Pachyptila* species, followed by the nuclear intron Aden5. Yet, when each species was considered separately, variation at *cyt b* was generally lower than for introns Aden5, d-cryst, and OD67 (supplementary table S5, Supplementary Material online), indicating higher levels of allele sharing at the nuclear, relative to the mitochondrial locus. Similarly, low sequence diversity was observed at Lipo2 and Tim4, and this was generally true for species- and genus-level analyses. Although negative values for both Fu's  $F_s$  and Tajima's  $D$  statistics were nonsignificant at the species level (supplementary table S4, Supplementary Material online), one or other of these

values was consistently negative for most species at all loci, potentially indicating demographic histories of population expansion.

### Genetic Structure

Levels of incomplete lineage sorting and/or gene flow among species were first assessed using mtDNA and nuclear intron gene networks in NETWORK (fig. 4). Species-specific structure was more apparent at mtDNA than among nuclear intron genes (fig. 4). The *cyt b* gene structured the genus into its recognized species, with little evidence of haplotype sharing among species (fig. 4). An exception to this was one *P. desolata* haplotype that grouped with *P. salvini*, with the mtDNA diversity of the latter species appearing to have derived from the gene pool of the former. On the other hand, for nuclear



**FIG. 4.** Median-joining networks constructed in NETWORK of all *Pachyptila* (Aves: Procellariiformes) breeding colonies sampled in this study. The networks are based on cytochrome *b* (*cyt b*; A), Adenylate kinase (Aden5; B), Ornithine decarboxylase (OD67; C), Lipoprotein lipase (Lipo2; D), d-crystallin (d-cryst; E), and Triosephosphate isomerase (Tim4; F) sequences. The size of the circles is proportional to haplotype frequency. The length of the line connections is proportional to the number of mutational steps except for one line in D where 18 mutated positions were represented by a double hash-mark for clarity.

intron sequences, allele sharing among *Pachyptila* species was the rule with only the outgroup (*H. caerulea*) partitioning separately for all but one intron (fig. 4B–F). Despite very different levels of allelic diversity among nuclear intron sequences, there was little to no structuring into species-specific lineages within the in-group (fig. 4B–F).

We combined all the information in these six mtDNA and nuclear intron gene networks into a species tree for *Pachyptila* under a Bayesian multispecies coalescent model in BEAST 2 (table 1). A soft-bounded lognormal prior ( $\alpha = 1.2$ ,  $\beta = 4.5$ ) of no later than 5 Ma for the emergence of the genus was based on earliest fossil evidence (see Materials and Methods). This species tree analysis also allowed for a refinement of locus-specific mutation rates, which varied almost 10-fold between highest ( $1.52 \times 10^{-3}$ /site/Ma, *cyt b*) and lowest ( $2.00 \times 10^{-4}$ , Lipoprotein lipase) estimates (supplementary fig. S1, Supplementary Material online). Even the most highly mutating nuclear intron (Aden5) was just under half that of the mtDNA locus (supplementary fig. S1, Supplementary Material online). To resolve more recent divergence and admixture within the radiation, we also carried out analyses of population structure using the model-free discriminant analysis of principal components (DAPC, Jombart et al. 2010) and Bayesian clustering in STRUCTURE (Pritchard et al. 2000) using 326 ingroup multilocus microsatellite genotypes. Both analyses, using Bayesian information criterion (BIC) and the delta K method (Evanno et al. 2005), respectively, estimated an optimal  $K = 3$ . The consensus species tree (fig. 5A) explained 30% of the posterior distribution (PD) of species trees, with the six most common topologies (supplementary fig. S2, Supplementary Material online) accounting for 83%. This tree differed only slightly from the nuclear DNA tree (fig. 5B), in which the position of *P. salvini* was sister to *P. turtur* and not the *P. belcheri*/*P. desolata* clade. The initial split within *Pachyptila* was between broad-billed *P. vittata* and the narrower-billed species (68% PD) occurring 4.8–5.8 Ma (fig. 5A; table 1). Similarly, under a two-population model, genome-wide microsatellite diversity was structured into the broad-billed and the narrow-billed populations (K2; fig. 5D). Within the narrow billed group, the species with the narrowest and shortest bill, *P. turtur*, was first to diverge between 1.1 and 4.0 Ma (fig. 5A; table 1) and in the microsatellites analyses at K3 (fig. 5D).

### The Origin of *P. salvini*

Although the phenotypically intermediate *P. salvini* could be separated from *P. desolata*–*P. belcheri* (51% PD; table 1), it was sister to *P. desolata* and *P. vittata* in 29% and 5% of cases, respectively. STRUCTURE analysis of the microsatellite data provided further resolution of recent population structuring. Instead of partitioning *P. salvini*, microsatellite variation was structured into a *P. desolata* cluster at K4 and the first split among *P. vittata* populations at K5 (fig. 5D). Despite further model fitting up to K10 (supplementary fig. S3, Supplementary Material online), *P. salvini* failed to form a species-specific cluster. Instead, the multilocus profiles of *P. salvini* comprised admixed alleles from *P. desolata* and the *P. vittata* population on Gough Island. This was further

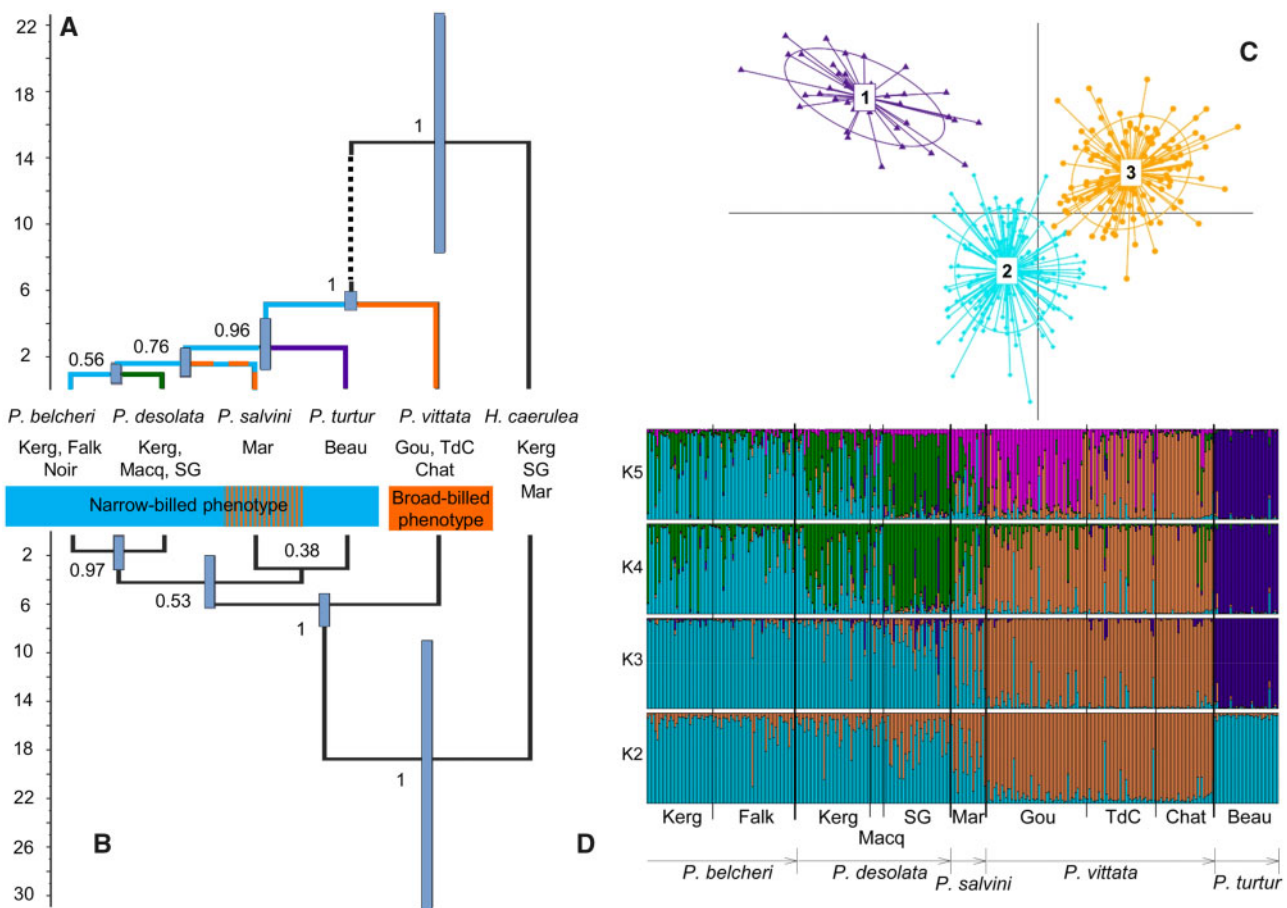
supported by a Bayesian estimation of bidirectional interspecific migration rates inferred in BAYESASS 3.0, which were generally low (table 2), except in the case of *P. salvini*, where up to 96% of its gene pool could have been derived through unidirectional gene flow with *P. desolata*, *P. belcheri*, and *P. vittata*.

In order to confirm a hybrid origin for *P. salvini* and rule out other factors such as common ancestry, lineage sorting, and homoplasy in explaining the observed results, we used the microsatellite data to perform a model choice analysis (hypothesis contrast) based on coalescent simulations within an approximate Bayesian computation (ABC) framework (Beaumont 2010). In a first step, we compared all 105 possible dichotomous species tree topologies as models, with the best model corresponding to a topology that was only one branch-swap different from the nuclear intron-mtDNA consensus tree (fig. 5A). In the second step, we used these two trees (the best supported and intron-mtDNA consensus tree), as well as two randomly chosen trees to construct a balanced, symmetrical set of 16 evolutionary models (hypotheses). The 16 models included four without admixture (models 1–4), four with admixture from *P. desolata* only (models 5–8), four with admixture from *P. vittata* only (models 9–12), and four full hybrid models with parental species *P. desolata* and *P. vittata* (models 13–16; fig. 6). The best model with both a direct acceptance method and a random forests method was model 13, which was similar to the consensus species tree but with *P. salvini* originating from the admixture of *P. desolata* and *P. vittata* (Bayes factor 1.21–138.0, posterior probability 0.166 with the direct method, and 0.4630 with random forests; supplementary figs. S4 and S5, tables S6 and S7, Supplementary Material online). A hybrid origin for *P. salvini* was also supported by the next best model (model 14), in which *P. vittata* remains ancestral to the *Pachyptila*, but where *P. turtur* is most closely related to the *P. belcheri*–*P. desolata* sister group. When pooled together, all hybrid models (models 13–16) received much more support than other models with a direct approach (Bayes factor 4.1–445, posterior probability of 0.529). The random forests approach could only provide a posterior probability for the best supported model (0.4630 for model 13 alone), therefore its support for the combined hybrid models should be even higher.

## Discussion

### Incomplete Lineage Sorting versus Gene Flow

Despite high admixture among species, we were able to infer species-level populations from microsatellite data for all but *P. salvini*, which appeared to comprise a mix of alleles from other species populations. These populations may have arisen during a time when gene flow was limited and/or effective population sizes small, in order for alleles to sort. These conditions would also have promoted the almost complete lineage sorting observed at the *cyt b* locus, perhaps at an elevated level because of the smaller effective population size of mitochondrial DNA. Cytochrome *b*, is widely known for its high phylogenetic resolution among animals and seabird examples include gadfly petrels (genus *Pterodroma*),



**FIG. 5.** Inferred phylogeny and genetic structure for the five *Pachyptila* (Aves, Procellariiformes) species in this study. The species trees were inferred using a Bayesian multispecies coalescent framework in BEAST. Microsatellite genetic structure was inferred in *adegenet* 2.0 and in STRUCTURE 2.0. Reconstruction of species trees using the mtDNA and nuclear intron sequences (A), and using only the nuclear intron markers (B). Blue boxed regions show the 95% Bayesian credible intervals (BCIs) for relative divergence times. Narrow (light blue) and broad-billed (orange) taxa are color-coded. For timing and reliability of species divergence events on the *Pachyptila* species tree see table 1. The blue petrel *Halobaena caerulea* was used as outgroup. (C) Scatterplot showing the distribution of the microsatellite data into three populations in multivariate space after DAPC. Groups are circled by inertia ellipses and colored as in A, figures 1 and 4, with *Pachyptila turtur* represented by triangles, *Pachyptila vittata* by circles and other narrow-billed prions by squares. (D) Microsatellite genetic structure (models K2–K5), using an admixture model with correlated allele frequencies in STRUCTURE. Mayes: Mayes I., Kerguelen Is. New: New I., Falkland/Malvinas Is. Verte: Verte I., Kerguelen Is. Macq: Macquarie I. Bird: Bird I., South Georgia/Georgias del Sur. Mar: Marion I., Prince Edward Is. Gou: Gough I. Night: Nightingale I., Tristan da Cunha. Rang: Rangitara I., Chatham Is. Beau: Beauchêne I., Falkland/Malvinas Is.

**Table 1.** Timing and Reliability of Species Divergence Events on the Most Likely *Pachyptila* (Aves, Procellariiformes) Species Tree.

Node	Taxa Separated on Consensus Tree		Posterior Probability	% of the Posterior Distribution of Trees	Divergence Time (Ma)		
	Clade I	Clade II			Mean	Low <sup>a</sup>	High <sup>a</sup>
1	<i>Halobaena caerulea</i>	All <i>Pachyptila</i>	1	100	14.81	8.31	22.16
2	<i>P. vittata</i>	<i>P. turtur</i> , <i>P. salvini</i> , <i>P. belcheri</i> , <i>P. desolata</i>	1	68	5.16	4.80	5.82
3	<i>P. turtur</i>	<i>P. salvini</i> , <i>P. belcheri</i> , <i>P. desolata</i>	0.96	75	2.50	1.13	4.05
4	<i>P. salvini</i>	<i>P. belcheri</i> , <i>P. desolata</i>	0.76	68	1.48	0.65	2.34
5	<i>P. belcheri</i>	<i>P. desolata</i>	0.56	—	0.90	0.26	1.65

NOTE.—Data correspond to mtDNA (cytochrome *b*) and nuclear intron (Triosephosphate isomerase, d-crystallin, Ornithine decarboxylase, Lipoprotein lipase, Adenylate kinase) genes combined into a species tree for *Pachyptila* under a Bayesian multispecies coalescent model in BEAST 2.

<sup>a</sup>95% highest posterior density (HPD) of node height.

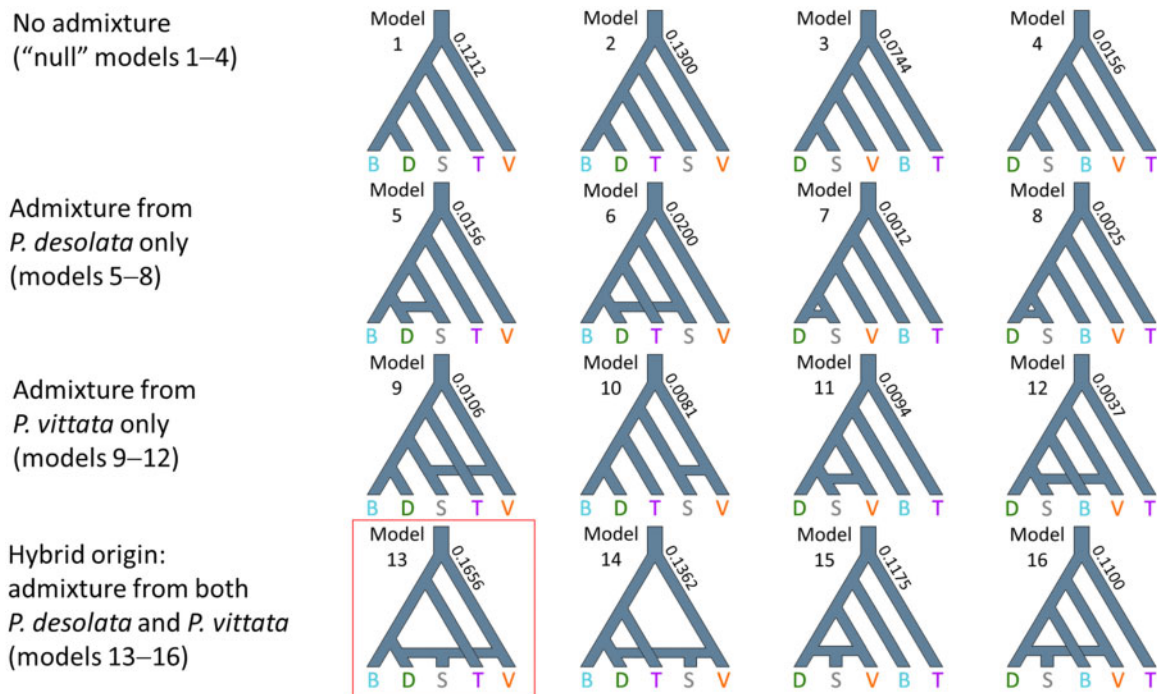
subspecies of White-faced Storm-petrel (*Pelagodroma marina*), and two storm-petrels (genus *Oceanodroma*; Gangloff et al. 2013; Silva et al. 2015, 2016). Yet, the lack of *cyt b* haplotype sharing suggests that mitochondrial gene flow

may be limited in this group, perhaps resulting from female philopatry, as suggested for *P. turtur* (Ovenden et al. 1991), and for other birds (Tiedemann et al. 1999; Caparroz et al. 2009). In general, seabirds are well-known for their philopatric

**Table 2.** Posterior Mean Migration Rates and Standard Deviation of the Marginal Posterior Distribution for Each Estimate.

sp. 1\sp. 2	<i>P. belcheri</i>	<i>P. desolata</i>	<i>P. vittata</i>	<i>P. salvini</i>	<i>P. turtur</i>
<i>P. belcheri</i>	—	0.0075 (0.0069)	0.0061 (0.0057)	0.0044 (0.0043)	0.0047 (0.0046)
<i>P. desolata</i>	0.1094 (0.0264)	—	0.0093 (0.0078)	0.0040 (0.0040)	0.0108 (0.0071)
<i>P. vittata</i>	0.0034 (0.0033)	0.0032 (0.0032)	—	0.0028 (0.0028)	0.0028 (0.0027)
<i>P. salvini</i>	0.3171 (0.0152)	0.3165 (0.0158)	0.3171 (0.0154)	—	0.0216 (0.0198)
<i>P. turtur</i>	0.0106 (0.0100)	0.0107 (0.0101)	0.0095 (0.0092)	0.0140 (0.0121)	—

NOTE.—Mean migration rates ( $m$ ) as a proportion from 0 to 1 and standard deviation (SD) were calculated in BAYESASS 3.0 between populations of the thin-billed prion *Pachyptila belcheri*, and Antarctic prion *P. desolata*, broad-billed prion *P. vittata*, Salvin's prion *P. salvini*, and fairy prion *P. turtur*. Values in italics highlight exceptionally high migration rates into the *P. salvini* gene pool. Values below the diagonal correspond to  $m[1][2]$  ( $\pm$  SD) which is the fraction of individuals in species 1 (column far left) that are migrants derived from species 2 (top line) per generation. Values above the diagonal correspond to  $m[2][1]$  ( $\pm$  SD).



**FIG. 6.** Simulation of coalescent demographic scenarios of all *Pachyptila* (Aves: Procellariiformes) species in this study within an approximate Bayesian computation (ABC) framework. Evolutionary scenarios were grouped as follows: no admixture (“null” models 1–4), admixture from *Pachyptila desolata* only (models 5–8), admixture from *Pachyptila vittata* only (models 9–12), and hybrid origin (admixture from both *P. desolata* and *P. vittata*; models 13–16). The box marks the best model. Species codes are as follow: *Pachyptila salvini* (S), *Pachyptila turtur* (T), *P. desolata* (D), *P. vittata* (V), and *Pachyptila belcheri* (B). The colors in species letters correspond to those of figures 1–4. Numbers upon the schemes represent the estimated model likelihoods. See also [supplementary tables S6–S8](#), [figs. S4–S7](#), [Supplementary Material](#) online.

behavior, which is probably one of the most evident potential barriers to gene flow in this group (Friesen 2015).

Gene flow could also explain high levels of incomplete lineage sorting at nuclear intron loci, but this view is highly unlikely because nuclear microsatellite data could be resolved into species populations. Both introns and microsatellites are nuclear and thus similarly affected by population size. Therefore, gene flow would be expected to decrease microsatellite resolution similarly to that of nuclear introns. Gene flow therefore cannot account for the lack of ancestral structure among intron gene trees, and most of the genetic diversity must be polymorphisms with a much deeper ancestry that have not yet undergone complete lineage sorting. Nuclear introns may be particularly affected by incomplete lineage sorting because of their lower mutation rates ([supplementary material S1: fig. S1](#), [Supplementary Material](#) online),

potentially resulting from stabilizing selection (Castillo-Davis et al. 2002; Hong et al. 2006). Thus, incomplete lineage sorting of nuclear introns is commonly observed in vertebrates (McCracken and Sorenson 2005; Welch et al. 2011; Kutschera et al. 2014), and arises either because species share a recent common ancestor and/or if ancestral effective population sizes were large (Edwards et al. 2005). On the other hand, nuclear intron lineages sampled from outside the radiation among *H. caerulea*, which we estimate shared a distant common ancestor with *Pachyptila*  $\sim 15$  Ma (but could be as recently as 8 Ma) had sorted almost completely.

Despite differences in diversity, mutation rates and levels of incomplete lineage sorting and gene flow, both slower evolving nuclear and mitochondrial DNA sequences and rapidly mutating microsatellite loci revealed very similar hierarchical levels of genetic structure (fig. 5), and this was consistent with



an early divergence between broad and narrow-billed species. The basal position of *P. vittata* (fig. 5) and the occurrence of vestigial palatal lamellae in the narrow-billed species suggest that the broad-billed filter-feeding morpho-ecotype was ancestral among *Pachyptila*, and that the narrow-billed phenotype evolved within the last four million years. At more derived parts of the radiation, both nuclear intron networks and species trees show greater uncertainty among narrow-billed species. This results from inconsistencies between the species tree and the nuclear intron tree topologies (fig. 5B). In both analyses, alternative, though less probable, topologies for relationships among narrow-billed species also occurred in the PD of species trees (fig. 6; supplementary material S1: fig. 4, Supplementary Material online). This evidence implies high levels of gene flow among narrow-billed species, and is supported by high admixture at microsatellite loci. The *Pachyptila* radiation is therefore at a phase where the most ancient population divergence can be reconstructed because of gradual lineage sorting, through on going genetic drift. At the same time, the placement of narrow-billed species divergences remains less certain because drift is overshadowed by gene flow in the more derived part of the tree, from at least 2.5 Ma and possibly longer.

#### A Hybrid Origin for *P. salvini*

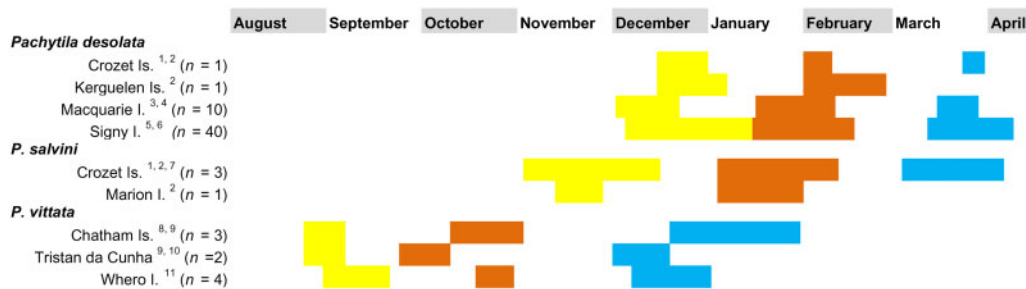
*Pachyptila salvini* comprises a mix of derived alleles obtained through gene flow. This effectively rules out the possibility that *P. salvini* was the common ancestor to, or the genetic intermediate between, narrow and broad-billed phenotypes. Our results suggest that *P. salvini* is a hybrid between *P. desolata* and *P. vittata*, whose intermediate phenotype may confer it with a feeding advantage over parental and other *Pachyptila* species by allowing it to feed on a significantly broader range of prey (fig. 3; supplementary table S3, Supplementary Material online).

Apart from *P. salvini*, all other species in the *Pachyptila* radiation were assigned to species or population clusters. The difficulty in assigning *P. salvini*, even when forcing the data into up to ten clusters, suggests the species possesses a highly admixed genome. This was further supported by a Bayesian estimation of bidirectional interspecific migration rates, where the majority of the variation in *P. salvini* could be attributed to gene flow from other species. A formal testing of the full hybrid model against no-migration or simple introgression models using simulations with powerful ABC-machine learning methods showed that, despite moderate statistical power, the most probable model was that *P. salvini* originated through hybridization between narrow-billed *P. desolata* and broad-billed *P. vittata*. The parental species show some degree of overlap in distribution in areas around New Zealand and in parts of the South Atlantic Ocean (supplementary table S1, Supplementary Material online; Marchant and Higgins 1990). Interestingly, these areas of overlap between *P. desolata* and *P. vittata* are located far away from *P. salvini* breeding colonies (all located in the Indian Ocean). Late Pleistocene climatic oscillations could potentially have brought parental species into greater secondary contact, although we feel this is not a necessary

requirement since prions are highly mobile seabirds, and able to cover the large distances among oceanic islands that offer suitable breeding habitat (Quillfeldt et al. 2013, 2014, 2015, 2017; Navarro et al. 2015). Furthermore, distances between colonies within a species are usually of similar or greater magnitude than between species, with genetic differentiation low and inferred migration rates (table 2) high between island populations within prion species (Quillfeldt et al. 2017). Geography therefore is highly unlikely to restrict gene flow among *Pachyptila* species.

We present strong evidence for the evolution of *P. salvini* through additive morphometric and genetic effects. However, its intermediate position between, and potential feeding advantage over, either parental species do not prove that *P. salvini* is a hybrid species. Hybrid swarms also display similar additive traits, including higher hybrid fitness (Arnold and Hodges 1995; Seehausen 2004). To maintain its integrity, and fitness advantage, any newly evolved hybrid *P. salvini* must also be effectively reproductively isolated from parental species. Since geography may be ruled out as a reproductively isolating mechanism in prions, we looked to behavioral mechanisms instead. Changes in the timing of reproduction (allochryony) were found to attenuate gene flow in other seabirds (Friesen, et al. 2007; Friesen 2015). A summary of the available breeding data for hybrid and parental species suggests that *P. salvini* has an intermediate breeding phenology to either of its parental species (fig. 7). The onset of egg-laying is earliest in *P. vittata* (last third of August to early September), intermediate in *P. salvini* (early to mid-November) and latest in *P. desolata* (early December). Although some overlap in egg-laying might occur between *P. salvini* from Crozet I. and *P. desolata* from Signy I. and Macquarie I., the *P. salvini* population on Marion Island (studied here) ends egg-laying earlier (late-November) than *P. desolata* (early-December). Since egg-laying time is highly correlated with mating time in birds (Arvidsson and Neergaard 1991), these data suggest that the reproduction of *P. salvini* is isolated from both parental species, *P. desolata* and *P. vittata*. This is remarkable, since it would imply that the time of breeding is also additive in nature.

The data and analyses presented here suggest that Salvin's Prion could be a hybrid species in which the additive effects of divergent parental morphological traits have led to an intermediate bill morphology, potentially increasing its fitness, whereas additive parental breeding traits has led directly to reproductive isolation through the evolution of intermediate, and hence allochronous, breeding times. This would be all the more remarkable because intermediate phenotypes may arise through the additive influence of several genes (Anderson 1949), so the chances of independently intermediate phenotypes driving the hybrid speciation process must be extremely low. Future research could confirm our interpretation of the results and unravel the mechanism underlying phenotypically additive pleiotropic effects in this system by means of whole-genome analyses, and investigate potential linkage between genes influencing bill morphology and those controlling breeding phenology.



**Fig. 7.** Phenology. Comparison of egg-laying periods (in yellow), hatching dates (in orange), and chick fledging dates (in light blue) for the Antarctic prion *Pachyptila desolata*, Salvin's prion *Pachyptila salvini* and broad-billed prion *Pachyptila vittata*. Data are given for different colonies of each species (fig. 1; variation between populations). Horizontal bars represent the time span (variation among individuals). *n* indicates the number of breeding seasons investigated (variation between years). Superscript numbers indicate the source references for the data used to create the figure. <sup>1</sup>: Jouventin et al. (1985), <sup>2</sup>: Bretagnolle et al. (1990), <sup>3</sup>: Warham (1969), <sup>4</sup>: Brothers (1984), <sup>5</sup>: Tickell (1962), <sup>6</sup>: Rootes (1988), <sup>7</sup>: Harper (1980), <sup>8</sup>: West and Nilsson (1994), <sup>9</sup>: Marchant and Higgins (1990), <sup>10</sup>: Ryan et al. (2014), <sup>11</sup>: Richdale (1965).

## Materials and Methods

### Species, Samples, and Morphometric Data

Between 2010 and 2012, samples (all from breeding adults) from 77 thin-billed prions (*P. belcheri*), 80 Antarctic prions (*P. desolata*), 118 broad-billed prions (*P. vittata*), 18 Salvin's prions (*P. salvini*), and 34 fairy prions (*P. turtur*) were obtained in breeding colonies located on temperate to subAntarctic island groups (fig. 1; supplementary table S1, Supplementary Material online).

A summary of the morphometric data used in this study, as well as the source references are provided in the supplementary table S2, Supplementary Material online. To obtain missing morphometric data for *P. belcheri* corresponding to New Island, Falkland/Malvinas Islands (fig. 1; supplementary tables S1 and S2, Supplementary Material online), we measured wing length ( $\pm 1$  mm;  $n = 157$ ) with a stopped wing rule, tail length ( $\pm 1$  mm;  $n = 157$ ) with a feather rule, bill length ( $n = 86$ ) and width ( $n = 98$ ), and tarsus length ( $\pm 0.1$  mm;  $n = 155$ ) with callipers.

### Molecular Methods

Genomic DNA extraction for all samples was outlined in a previous study (Moodley et al. 2015). Twenty-five microsatellite loci, originally isolated by shotgun genome sequencing of a thin-billed prion were amplified in all samples following a previous study (Moodley et al. 2015). Microsatellite profiles were checked for null alleles using MICROCHECKER (van Oosterhout et al. 2004) and for deviation from genotypic equilibrium (Hardy–Weinberg equilibrium, HWE) using FSTAT (Goudet 1995). Multiple tests were corrected using a Bonferroni correction. Characteristics and variation of the 25 loci are detailed in Moodley et al. (2015). A subset of 32 samples, five of each of three species and six for *P. belcheri* and *P. turtur*, plus five from the closely related outgroup taxon Blue Petrel *H. caerulea* (from Bird I., South Georgia, Marion I., Prince Edward Is., and Mayes I., Kerguelen Is.), were used to amplify an informative 880 bp fragment of the mitochondrial *cyt b* gene, also following the methods and Polymerase Chain Reaction (PCR) conditions described in Moodley et al. (2015). Within each species, individuals were chosen from as many breeding colonies as possible (supplementary table S1,

Supplementary Material online), thereby capturing as much of the genetic variation of each species as possible.

Partial intron sequences of five nuclear housekeeping loci known to be informative in avian molecular systematics were amplified in the same 32 individuals (six species) used to sequence the *cyt b* gene (supplementary table S4, Supplementary Material online). These nuclear introns included d-crystallin (d-cryst, 330 bp; Morris-Pocock et al. 2008), Lipoprotein lipase (Lipo2, 277 bp; Patterson et al. 2011), Triosephosphate isomerase (Tim4, 317 bp; Patterson et al. 2011), Ornithine decarboxylase (OD67, 581 bp; Patterson et al. 2011), and Adenylate kinase (Aden5, 448 bp; Shapiro and Dumbacher 2001). PCR was conducted in 20  $\mu$ l reaction volumes containing 100 ng DNA template, 10 mM of each primer, 10 mM dNTPs (Roth, Karlsruhe), 2 mM MgCl<sub>2</sub>, 5 U *Thermus aquaticus* polymerase (BioLabs Taq DNA polymerase) in a 1 $\times$  PCR reaction buffer. Thermocycling included an initial denaturation at 94 °C for 2 min, 30 cycles of denaturation at 94 °C for 30 s, annealing at 60 °C (d-Cryst) or 58 °C (for the remaining markers) for 45 s and extension at 72 °C for 1 min, followed by a final extension step of 72 °C for 5 min. Products were purified of excess primers and dNTPs using exonuclease-shrimp alkaline phosphatase (Fermentas Life Sciences following the manufacturer's specifications). PCR products were then sequenced in both directions using Big Dye chemistry (Applied Biosystems) and run on an AB 3130xl genetic analyser (Applied Biosystems). Resulting sequences were assembled and aligned in CLC Main Workbench 6.

### Genetic Diversity and Tests of Neutrality

Genetic diversity was estimated for each population and each species for all marker sets. For microsatellites, the allelic diversity (number of alleles per locus) was estimated using GENETIX 4.05 (Belkhir et al. 2004) and rarefied for differences in sample size using ADZE 1.0 (Szpiech et al. 2008). Unbiased expected heterozygosity (*HE*) and observed heterozygosity (*HO*) was also estimated in GENETIX. Nuclear intron sequences were phased using the Phase algorithm (Stephens et al. 2001), consisting of 100,000 iterations with a burnin proportion of 10%. Nucleotide and haplotype diversities for nuclear

intron and mitochondrial *cyt b* sequences were calculated in ARLEQUIN 3.5 (Excoffier and Lischer 2010). Under neutrality, values for Tajima's *D* (Tajima 1989) and Fu's *F<sub>s</sub>* (Fu 1997) are expected to be approximately zero. A significantly negative *D* value could indicate purifying selection or a population expansion, and a negative *F<sub>s</sub>* value indicates an excess of alleles due to genetic hitchhiking or a population expansion. Regardless, a positive *D* may also result from balancing selection or a population bottleneck and a positive *F<sub>s</sub>* from overdominant selection or a population bottleneck. To infer the nature of selection and population demography acting upon our data set of mtDNA and nuclear intron sequences, we ran both Tajima's and Fu's tests of neutrality using ARLEQUIN 3.5.

### Microsatellite Genetic Structure

We investigated the existence of population genetic structure using two methods: the model-free DAPC (Jombart et al. 2010) implemented in *adegenet* 2.0 (Jombart 2008), and an individual Bayesian clustering algorithm in STRUCTURE 2.0 (Pritchard et al. 2000). All individuals from all species were run at the same time. The optimum number of clusters for DAPC was determined by the lowest BIC. In STRUCTURE, we assumed an admixture model since prions are specialist long distance pelagic birds, it is reasonable that most species have come repeatedly into secondary contact over the course of their radiation. The analysis was run ten times for *K* = 1–10 with each run randomly started, consisting of 500,000 Markov Chain Monte Carlo (MCMC) iterations, assuming correlated allele frequencies and removing the first 100,000 runs as burn-in. The mean likelihood values across multiple values of *K* were determined using STRUCTURE HARVESTER (Earl and vonHoldt 2012) in accordance with Evanno's method ( $\Delta K$ , i.e., the rate of change in the log probability of data between successive *K* values; Evanno et al. 2005). Assignment plots were constructed for all values of *K* that were biologically interpretable. We also performed the same analyses on a reduced data set containing only the 18 loci not deviating from HWE to determine the general consistency of the data (for identity, characteristics and variation of the used loci see Moodley et al. 2015). We used CLUMPAK (Kopelman et al. 2015) for the creation of [supplementary fig. S3, Supplementary Material](#) online. We did not include the sample from Isla Noir in any of these analyses due to missing data in the majority of loci.

### Haplotype Networks

Since we were interested in the potential allele sharing among species, we also represented each gene tree as a reticulating network, rather than a bifurcating tree (fig. 4). The median joining method was implemented in NETWORK 5.0 (Bandelt et al. 1999) with an equal weighting on all nodes and using a correction cost algorithm.

### Parameterizing the *Pachyptila* Species Tree Coalescent

To determine how *Pachyptila* species were related to each other, we reconstructed the phylogeny of the genus using all the available DNA sequence data (mtDNA and nuclear intron markers). The species tree was inferred using a Bayesian

multispecies coalescent framework in BEAST 2 (Bouckaert et al. 2014). The best nucleotide substitution model was determined for each of the six alignments using JMODELTEST (Posada 2008). Site models were thus unlinked for each partition. Each gene tree was unlinked, and a species tree using a birth-death model prior was used to account for gene tree incongruence. The population prior allowed for changes in effective population size, but assumed a constant ancestral population size. Five *H. caerulea* individuals, sequenced for the same marker set, were used as the outgroup of *Pachyptila*. All clock models were unlinked, but to determine the correct clock prior, alternative models assuming lognormal and exponential priors were tested against a strict molecular clock for each partition. After a single run of one billion MCMC iterations, sampling every 10,000 steps and discarding 20% as burn-in, the standard deviations of the posterior marginal distributions of both these parameters included zero in all cases, making them no more likely than a null strict clock for all gene partitions. All subsequent analyses were therefore carried out assuming a strict clock.

### Mutation Rates among Procellariiformes and Coalescence Times

Mutation rates among tube-nosed seabirds are known to be lower than that of other bird orders (Nunn and Stanley 1998; Quillfeldt 2017). Since mutation rates can affect estimates of time to most recent common ancestor (TMRCA) of a set of DNA sequences, we included priors on mutation (clock) rates that were informed by the latest molecular data using BEAST 2 (Bouckaert et al. 2014) and in addition, we informed the minimum height of the *Pachyptila* clade from the fossil record. For *cyt b* we set a strong normally distributed clock prior using a procellariiform-specific mean body-mass corrected mtDNA mutation rate of 0.00189 per site per million years as its mean, with standard deviation set at 0.00025 (Nabholz et al. 2016). Clock rates were also estimated for each of the nuclear intron partition, but with a starting value equal to the recently available germline mutation rate in birds (0.0023 per site per million years). However, it is likely that germline mutation rates are far lower among the tube nosed seabirds (Quillfeldt 2017), so we set wide uniform priors to account for this possibility. We then placed a prior on the height of the monophyletic *Pachyptila* clade, whose height we set at no younger than oldest fossil attributed to the genus, 5 Ma old (Olson 1983, 1985a, 1985b). We set a hard bound of 5 Ma, and accounted for the potential of a much earlier origin of the genus by setting a lognormal distribution ( $\alpha = 1.2$ ,  $\beta = 4.5$ ) where the 95% quartile reached 20 Ma. This fully parameterized model was run five times, each time for one billion simulations, logging parameters every 100,000 steps, and discarding the first 20% as burn-in. MCMC convergence was assessed by viewing MCMC traces directly and by ESS values in TRACER 1.6 (Rambaut and Drummond 2007). To determine the bias brought into the species tree estimation by the higher mutation rate of the *cyt b* gene, we reran the full model, but without the mtDNA partition and compared the resulting species trees.

### Quantifying Interspecific Gene Flow

We inferred gene flow in BAYESASS 3.0 using bidirectional migration rates ( $m$ ) (Wilson and Rannala 2003). This method was successful in the study of immigration and emigration in populations that are not in equilibrium, for example, Friesen et al. (2007). BAYESASS estimates the posterior probability of an individual's history and allows an estimation of the rate and direction of recent dispersal (Genovart et al. 2013). The acceptance rates for the main parameters (i.e., "migration" rate, inbreeding coefficient and allele frequencies) were adjusted during several preliminary runs. Convergence was assessed by checking the trace files in TRACER 1.6 (Rambaut and Drummond 2007). Final parameter estimates were obtained after performing three independent runs using different starting random seed numbers. The MCMC was run for 50,000,000 iterations with a burn-in period of 10,000,000 and a sampling frequency of 5,000 iterations.

### Inferring Historical Structure by ABC

To test the hypotheses of a hybrid origin versus a nonhybrid origin, we employed the microsatellite data to perform an inference by ABC, which is a technique that uses a large number of simulations to perform statistical inference by selecting a sample of simulations with the highest similarity to the empirical data set (Beaumont 2010). Since the number of possible hybridization models is large (from thousands to millions due to the multiple combinations of topologies and admixture events), we performed a model choice analysis (hypotheses contrast) in two stages. In the first step, we tested all the possible 105 rooted dichotomic topologies that could be made with the five *Pachyptila* taxa, with each topology being a competing model. We then took the best-supported model (topology) from this analysis as well as the topology of the mitochondrial consensus tree (which were highly similar) and two random topologies as a basis for the construction of a symmetric set of 16 models. Thus, for each of the four base-topologies we defined four models according to the type of admixture they presented with *P. salvini*: no admixture (models 1–4), admixture with *P. desolata* (models 5–8), admixture with *P. vittata* (models 9–12), and Salvin's prion evolving as a hybrid species from parental species *P. desolata* and *P. vittata* (models 9–12). This design allowed assessing not only individually, but also in groups, the effect of having the correct phylogeny or the correct type of admixture.

We ran 105 million simulations for the first step (one million per model) and analyzed the 100 to 1,000 (~0.001–0.0001%) closest simulations to the empirical data. This 100–1,000 range allowed us to assess the consistency of the acceptance ratios to different thresholds. Summary statistics and priors were the same for this analysis and for the second stage (classic analysis, see below). Because the first step was only intended for guidance in the definition of the admixture scenarios, only Bayes factors were used to select the best model. For this and subsequent analysis we avoided the use of logistic regression because it yielded highly inconsistent results, possibly caused by known pitfalls of the logistic regression approach (Pudlo et al. 2016).

For the second step, we carried out two types of ABC-based model choice analysis: a classic analysis and a machine learning technique known as random forests. The classic approach chooses the best model by inferring model likelihoods out of the model ratios in the accepted sample. In the random forests approach, the inference is seen as a classification problem and the reference table (a table with summary statistics, parameters, and model indicators of all simulations) is used as a training data set that is used by a set of decision trees (the forest) to predict the model corresponding to the empirical data set.

For the classic analysis, 16 million simulations (one million per model) were run, from which the 160–1,600 (0.001–0.01%) closest simulations to the empirical data set were retained for inference. Prior probabilities of each model were equal (0.0625 each), while parameters were sampled from the next priors (see supplementary fig. S6, Supplementary Material online): effective population sizes of all modern and ancestral populations were sampled from log-uniform priors in [10–1000000]; the relative contributions of *P. vittata* and *P. desolata* to *P. salvini* (in hybrid models) were sampled from a uniform prior in [0.001, 0.999]; the time to the hybridization was sampled from a uniform prior in [10, 166666]; the time to the youngest population split was sampled from a uniform prior in [10, 166666]; whereas the times from the intermediate population splits were sampled from uniform priors in [22222, 222222] and [27777, 277777]. Times were set in generations with a generation time set to 18 years. Following recommendations of the software (DIYABC, see below), we employed a Stepwise Mutation Model (SMM) with a mean mutation rate given by a log-uniform prior ( $10^{-6}$ – $10^{-2}$ ) and individual locus rates given by a gamma prior with shape parameter set at 2.0 and constrained to the interval  $10^{-6}$ – $10^{-2}$ . We employed an overall of 50 summary statistics: for each of the five species we used the mean number of alleles, mean genic diversity, mean size variance, and mean Garza–Williamson's  $M$ , whereas for each species pair we used  $F_{st}$ , shared allele distance and  $(\delta\mu)^2$  distance. We visually assessed the fit of the best supported model to the empirical data by means of Principal Components Analysis (PCA) performed on the summary statistics (supplementary fig. S5, Supplementary Material online) and to the whole set of models by means of linear discriminant analysis (supplementary fig. S7, Supplementary Material online). We computed Bayes factors, approached by the acceptance rates (direct estimates), for each pair of models and reported the model that received the highest support.

For the random forest analysis, we ran 160 thousand simulations (10 thousand per model) and loaded the reference table from DIYABC format to R by means of the function "readRefTable" of the "abcrf" package. Using the function "err.abcrf" we estimated the out-of-bag error for different number of trees in order to define an optimal number of trees for the analysis ( $n = 1,000$ ; supplementary fig. S8, Supplementary Material online). We created the random forest with the function "abcrf" and obtained charts of the linear discriminant axes. We estimated the best model with the

function “predict.abcrf” trained on the reference table and using as a predicting point the set of summary statistics (introduced by hand). This function also estimates the posterior probability of the chosen model by means of a regression random forest (Pudlo et al. 2016). Following recommendations (Pudlo et al. 2016; Estoup et al. 2018), the inference was performed with a large set of summary statistics, for which we included all those used in the classic ABC analysis plus classification index for each pair of populations (two indices per pair) and maximum likelihood of admixture, resulting in an overall 71 summary statistics.

Simulations, model choice by the direct method and for logistic regression, as well as PCA analyses were carried out in the software DIYABC 2.1 (Cornuet et al. 2014). For the random forests analysis, the simulations were carried out in DIYABC whereas the rest of the analyses were performed in R (R Core Team 2018) with the package “abcrf” (Pudlo et al. 2016).

## Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

## Acknowledgments

This work was supported by the Deutsche Forschungsgemeinschaft (Germany), Heisenberg program (grant number DFG, Qu 148-5 to P.Q.), a CGS award from the Natural Environment Research Council (NERC), the Institut Polaire Français Paul Emile Victor (Programme N°109, to H.W.), the Terres Australes et Antarctiques Françaises (Kerguelen), Mohamed bin Zayed Species Conservation Fund and the Brian Mason Scientific and Technical Trust (Chatham Islands). The study involved wild individuals and was carried out under permits from the Falkland Islands Government (Environmental Planning: R21.2012), Government of South Georgia and the South Sandwich Islands, Animal Ethic Committee of the Institut Polaire Français Paul Emile Victor, Préfet des Terres Australes et Antarctiques Françaises, the Animal Ethics Committee of Charles Sturt University (15/092), and the New Zealand Department of Conservation. We thank the New Island Conservation Trust with assistance from Ian, Maria and Georgina Strange, Benno H. Lüthi, Klemens Pütz, and Gerhard Meyer from the Antarctic Research Trust, Boglárka Bálint, Ruth Brown, A. Corbeau, J. Ferrer-Obiol, T. Lacombe, H. Mihailou, Gopi K. Munimanda, Colin Miskelly, D. Paris, and M. Passeraut.

## Author Contributions

P.Q., J.F.M., and Y.M. conceived and designed the study. P.Q., R.A., R.C., Y.C., K.D., J.F.M., M.a.M., M.e.M., J.N., R.P., P.R., C.S., and H.W. carried out the extensive fieldwork. J.F.M., T.L.C., L.D.S., and L.C. carried out the lab work. Y.M., J.F.M., and E.S.-C. carried out the bioinformatics analyses. Y.M. and J.F.M. drafted the manuscript. All authors reviewed the final draft of the manuscript.

## References

- Abbott R, Albach D, Ansell S, Arntzen JW, Baird SJE, Bierne N, Boughman J, Brelsford A, Buerkle CA, Buggs R, et al. 2013. Hybridization and speciation. *J Evol Biol.* 26(2):229–246.
- Anderson E. 1949. Introgressive hybridization. New York: John Wiley and Sons.
- Arnold ML, Hodges SA. 1995. Are natural hybrids fit or unfit relative to their parents? *Trends Ecol Evol (Amst).* 10(2):67–71.
- Arvidsson BL, Neergaard R. 1991. Mate choice in the willow warbler—a field experiment. *Behav Ecol Sociobiol.* 29(3):225–229.
- Bandelt HJ, Forster P, Röhl A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol.* 16(1):37–48.
- Barrera-Guzmán AO, Aleixo A, Shawkey MD, Weir JT. 2018. Hybrid speciation leads to novel male secondary sexual ornamentation of an Amazonian bird. *Proc Natl Acad Sci U S A* 115(2):E218–E225.
- Beaumont MA. 2010. Approximate Bayesian computation in evolution and ecology. *Annu Rev Ecol Syst.* 41(1):379–406.
- Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F. 2004. GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. Version 4.05. Montpellier (France): Laboratoire Génome, Populations, Interactions, CNRS UMR 5171, Université de Montpellier II.
- Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu C-H, Xie D, Suchard MA, Rambaut A, Drummond AJ. 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput Biol.* 10(4):e1003537.
- Brelsford A, Mila B, Irwin DE. 2011. Hybrid origin of Audubon’s warbler. *Mol Ecol.* 20(11):2380–2389.
- Bretagnolle V, Zotier R, Jouventin P. 1990. Comparative population biology of four prions (genus *Pachyptila*) from the Indian ocean and consequences for their taxonomic status. *Auk* 107(2):305–316.
- Brothers N. 1984. Breeding, distribution and status of burrow-nesting Petrels at Macquarie Island. *Wildl Res.* 11(1):113–131.
- Buerkle C.A., Morris R.J., Asmussen, M.A. and Rieseberg L.H., 2000. The likelihood of homoploid hybrid speciation. *Heredity*, 84(4):441–451.
- Caparroz R, Miyaki CY, Baker AJ. 2009. Contrasting phylogeographic patterns in mitochondrial DNA and microsatellites: evidence of female philopatry and male-biased gene flow among regional populations of the blue-and-yellow Macaw (*Psittaciformes: Ara Ararauna*) in Brazil. *Auk* 126(2):359–370.
- Castillo-Davis CI, Mekhedov SL, Hartl DL, Koonin EV, Kondrashov FA. 2002. Selection for short introns in highly expressed genes. *Nat Genet.* 31(4):415–418.
- Cherel Y, Bocher P, De Broyer C, Hobson KA. 2002. Food and feeding ecology of the sympatric thin-billed *Pachyptila belcheri* and Antarctic *P. desolata* prions at Iles Kerguelen, Southern Indian Ocean. *Mar Ecol Prog Ser.* 228:263–281.
- Cornuet J-M, Pudlo P, Veysier J, Dehne-Garcia A, Gautier M, Leblois R, Marin J-M, Estoup A. 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):1187–1189.
- Cox J. 1980. Some remarks on the breeding distribution and taxonomy of the prions (Procellariidae: *Pachyptila*). *Rec S Aust Mus.* 18:91–121.
- Coyne J, Orr H. 2004. Speciation. Sunderland (MA): Sinauer.
- deVicente MC, Tanksley SD. 1993. QTL analysis of transgressive segregation in an interspecific tomato cross. *Genetics* 134(2):585–596.
- Earl DA, vonHoldt BM. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour.* 4(2):359–361.
- Edwards SV, Kingan SB, Calkins JD, Balakrishnan CN, Jennings WB, Swanson WJ, Sorenson MD. 2005. Speciation in birds: genes, geography, and sexual selection. *Proc Natl Acad Sci U S A* 102(Suppl 1):6550–6557.
- Elgvin TO, Hermansen JS, Fijarczyk A, Bonnet T, Borge T, Saether SA, Voje KL, Saetre G-P. 2011. Hybrid speciation in sparrows II: a role for sex chromosomes? *Mol Ecol.* 20(18):3823–3837.
- Estoup A, Raynal L, Verdu P, Marin J-M. 2018. Model choice using Approximate Bayesian Computation and Random Forests: analyses

- based on model grouping to make inferences about the genetic history of Pygmy human populations. *J Soc Franç Stat.* 159:167–190.
- Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software structure: a simulation study. *Mol Ecol.* 14(8):2611–2620.
- Excoffier L, Lischer HEL. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour.* 10(3):564–567.
- Falla RA. 1940. The genus *Pachyptila* Illiger. *Emu* 40(3):218–236.
- Fisher R. 1930. The genetical theory of natural selection. Oxford: Clarendon Press.
- Fleming CA. 1941. The phylogeny of the prions. *Emu* 41(2):134–155.
- Friesen V, Smith A, Gomez-Diaz E, Bolton M, Furness R, González-Solís J, Monteiro L. 2007. Sympatric speciation by allochrony in a seabird. *Proc Natl Acad Sci U S A* 104(47):18589–18594.
- Friesen V, Burg T, McCoy K. 2007. Mechanisms of population differentiation in seabirds. *Mol Ecol.* 16(9):1765–1785.
- Friesen VL. 2015. Speciation in seabirds: why are there so many species. . . and why aren't there more? *J Ornithol.* 156 (Suppl 1):S27–S39.
- Fu Y-X. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* 147(2):915–925.
- Fullagar PJ. 1972. Identification of prions – *Pachyptila* spp. *Aust Bird Bander* 10:36–39
- Gangloff B, Zino F, Shirihai H, González-Solís J, Couloux A, Pasquet E, Bretagnolle V. 2013. The evolution of north-east Atlantic gadfly petrels using statistical phylogeography. *Mol Ecol.* 22(2):495–507.
- Gavrilets S, Losos JB. 2009. Adaptive radiation: contrasting theory with data. *Science* 323(5915):732–737.
- Genovart M, Thibault J-C, Igual JM, Bauzá-Ribot MM, Rabouam C, Bretagnolle V. 2013. Population structure and dispersal patterns within and between Atlantic and Mediterranean populations of a large-range pelagic seabird. *PLoS One* 8(8):e70711.
- Goudet J. 1995. FSTAT (version 1.2): a computer program to calculate F-statistics. *J Heredity.* 86(6):485–486.
- Grant V. 1958. The regulation of recombination in plants. *Cold Spring Harb Symp Quant Biol.* 23:337–363.
- Gross B, Rieseberg L. 2005. The ecological genetics of homoploid hybrid speciation. *J Hered.* 96(3):241–252.
- Harper PC. 1980. The field identification and distribution of the prions (genus *Pachyptila*), with particular reference to the identification of storm-cast material. *Notornis* 27:235–286.
- Hermansen JS, Saether SA, Elgvin TO, Borge T, Hjelle E, Saetre G-P. 2011. Hybrid speciation in sparrows I: phenotypic intermediacy, genetic admixture and barriers to gene flow. *Mol Ecol.* 20(18):3812–3822.
- Hong X, Scofield DG, Lynch M. 2006. Intron size, abundance, and distribution within untranslated regions of genes. *Mol Biol Evol.* 23(12):2392–2404.
- Howarth DG, Baum DA, Soltis P. 2005. Genealogical evidence of homoploid hybrid speciation in an adaptive radiation of *Scaevola* (Goodeniaceae) in the Hawaiian Islands. *Evolution* 59(5):948–961.
- Jacob J, Hoerschelmann H. 1982. Chemotaxonomische untersuchungen zur systematik der röhrennasen (Procellariiformes). *J Ornithol.* 123(1):63–84.
- Jombart T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24(11):1403–1405.
- Jombart T, Devillard S, Balloux F. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet.* 11:94.
- Jouventin P, Mougín J-L, Stahl J-C, Weimerskirch H. 1985. Comparative biology of the burrowing petrels of the Crozet Islands. *Notornis* 32:157–220.
- Klages NTW, Cooper J. 1992. Bill morphology and diet of a filter-feeding seabird: the broad-billed prion *Pachyptila vittata* at South Atlantic Gough Island. *J Zool.* 227(3):385–396.
- Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I. 2015. Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Mol Ecol Resour.* 15(5):1179–1191.
- Kunte K, Shea C, Aardema ML, Scriber JM, Juenger TE, Gilbert LE, Kronforst MR. 2011. Sex chromosome mosaicism and hybrid speciation among tiger swallowtail butterflies. *PLoS Genet.* 7(9):e1002274.
- Kutschera VE, Bidon T, Hailer F, Rodi JL, Fain SR, Janke A. 2014. Bears in a forest of gene trees: phylogenetic inference is complicated by incomplete lineage sorting and gene flow. *Mol Biol Evol.* 31(8):2004–2017.
- Lamichhane S, Han F, Webster MT, Andersson L, Grant BR, Grant PR. 2018. Rapid hybrid speciation in Darwin's finches. *Science* 359(6372):224–228.
- Lavretsky P, Engilis A, Eadie JM, Peters JL. 2015. Genetic admixture supports an ancient hybrid origin of the endangered Hawaiian duck. *J Evol Biol.* 28(5):1005–1015.
- Mallet J. 2007. Hybrid speciation. *Nature* 446(7133):279–283.
- Marchant S, Higgins PJ. 1990. Volume 1, Ratites to ducks; Part A, Ratites to petrels. In: Marchant S, Higgins PJ, editors. Handbook of Australian, New Zealand & Antarctic Birds. Melbourne (Australia): Oxford University Press.
- Marques I, Draper D, López-Herranz ML, Garnatje T, Segarra-Moragues JG, Catalán P. 2016. Past climate changes facilitated homoploid speciation in three mountain spiny fescues (*Festuca*, Poaceae). *Sci Rep.* 6:36283.
- Mavárez J, Linares M. 2008. Homoploid hybrid speciation in animals. *Mol Ecol.* 17(19):4181–4185.
- Mavárez J, Salazar CA, Bermingham E, Salcedo C, Jiggins CD, Linares M. 2006. Speciation by hybridization in *Heliconius* butterflies. *Nature* 441(7095):868–871.
- Mayr E. 1942. Systematics and the origin of species, from the viewpoint of a zoologist. New York: Columbia University Press.
- Mayr E. 1963. Animal species and evolution. Cambridge: Harvard University Press.
- McCarthy EM, Asmussen MA, Anderson WW. 1995. A theoretical assessment of recombinational speciation. *Heredity* 74(5):502–509.
- McCracken KG, Sorenson MD. 2005. Is homoplasy or lineage sorting the source of incongruent mtDNA and nuclear gene trees in the stiff-tailed ducks (*Nomonyx-Oxyura*)? *Syst Biol.* 54(1):35–55.
- Moodley Y, Masello JF, Cole TL, Calderon L, Munimanda GK, Thali MR, Alderman R, Cuthbert RJ, Marin M, Massaro M, et al. 2015. Evolutionary factors affecting the cross-species utility of newly developed microsatellite markers in seabirds. *Mol Ecol Resour.* 15(5):1046–1058.
- Morgan WL, Ritz DA. 1982. Comparison of the feeding apparatus in the puttonbird, *Puffinus tenuirostris* (Temminck) and the fairy prion, *Pachyptila turtur* (Kuhl) in relation to the capture of the krill, *Nyctiphanes australis* Sars. *J Exp Mar Biol Ecol.* 59(1):61–75.
- Morris-Pocock J, Taylor S, Birt T, Damus M, Piatt J, Warheit K, Friesen V. 2008. Population genetic structure in Atlantic and Pacific Ocean common murres (*Uria aalge*): natural replicate tests of post-Pleistocene evolution. *Mol Ecol.* 17(22):4859–4873.
- Murphy RC. 1936. Oceanic birds of South America. New York: Macmillan.
- Nabholz B, Lanfear R, Fuchs J. 2016. Body mass-corrected molecular rate for bird mitochondrial DNA. *Mol Ecol.* 25(18):4438–4449.
- Nadeau NJ, Whibley A, Jones RT, Davey JW, Dasmahapatra KK, Baxter SW, Quail MA, Joron M, French-Constant RH, Blaxter ML, et al. 2012. Genomic islands of divergence in hybridizing *Heliconius* butterflies identified by large-scale targeted sequencing. *Philos Trans R Soc B.* 367(1587):343–353.
- Navarro J, Cardador L, Brown R, Phillips RA. 2015. Spatial distribution and ecological niches of non-breeding planktivorous petrels. *Sci Rep.* 5:12164.
- Nolte AW, Freyhof J, Tautz D. 2006. When invaders meet locally adapted types: rapid moulding of hybrid zones between sculpins (*Cottus*, Pisces) in the Rhine system. *Mol Ecol.* 15(7):1983–1993.
- Nunn GB, Stanley SE. 1998. Body size effects and rates of cytochrome b evolution in tube-nosed seabirds. *Mol Biol Evol.* 15(10):1360–1371.

- Olson SL. 1983. Fossil seabirds and changing marine environments in the Late Tertiary of South Africa. *S Afr J Sci*. 79:399–402.
- Olson SL. 1985a. An early Pliocene marine avifauna from Duinefontein, Cape Province, South Africa. *Ann S Afr Mus*. 95:147–164.
- Olson SL. 1985b. Early Pliocene Procellariiformes (Aves) from Langebaanweg, South-western Cape Province, South Africa. *Ann S Afr Mus*. 95:123–145.
- Ovenden J, Wust-Saucy A, Bywater R, Brothers N, White R. 1991. Genetic evidence for philopatry in a colonially nesting seabird, the Fairy Prion (*Pachyptila turtur*). *Auk* 108:688–694.
- Patterson SA, Morris-Pocock JA, Friesen VL. 2011. A multilocus phylogeny of the Sulidae (Aves: Pelecaniformes). *Mol Phylogenet Evol*. 58(2):181–191.
- Penhallurick J, Wink M. 2004. Analysis of the taxonomy and nomenclature of the Procellariiformes based on complete nucleotide sequences of the mitochondrial cytochrome b gene. *Emu* 104(2):125–147.
- Posada D. 2008. jModelTest: phylogenetic model averaging. *Mol Biol Evol*. 25(7):1253–1256.
- Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155(2):945–959.
- Pudlo P, Marin J-M, Estoup A, Cornuet J-M, Gautier M, Robert CP. 2016. Reliable ABC model choice via random forests. *Bioinformatics* 32(6):859–866.
- Quillfeldt P. 2017. Body mass is less important than bird order in determining the molecular rate for bird mitochondrial DNA. *Mol Ecol*. 26(9):2426–2429.
- Quillfeldt P, Cherel Y, Delord K, Weimerkirch H. 2015. Cool, cold or colder? Spatial segregation of prions and blue petrels is explained by differences in preferred sea surface temperatures. *Biol Lett*. 11(4):20141090.
- Quillfeldt P, Masello JF, Navarro J, Phillips RA. 2013. Year-round distribution suggests spatial segregation of two small petrel species in the South Atlantic. *J Biogeogr*. 40(3):430–441.
- Quillfeldt P, Moodley Y, Weimerskirch H, Cherel Y, Delord K, Phillips R, Navarro J, Calderón L, Masello JF. 2017. Does the genetic structure reflect differences in non-breeding movement? A case study in small, highly mobile seabirds. *BMC Evol Biol*. 17:160.161–160.111.
- Quillfeldt P, Phillips RA, Marx M, Masello JF. 2014. Colony attendance and at-sea distribution of thin-billed prions during the early breeding season. *J Avian Biol*. 45(4):315–324.
- R Core Team. 2018. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>. Last accessed on July 30, 2017.
- Rambaut A, Drummond AJ. 2007. Tracer v1.4.
- Rheindt FE, Austin JJ. 2005. Major analytical and conceptual shortcomings in a recent taxonomic revision of the Procellariiformes – a reply to Penhallurick and Wink (2004). *Emu* 105(2):181–186.
- Richdale LE. 1965. Breeding behaviour of the narrow-billed prion and broad-billed prion on Whero Island, New Zealand. *Trans Zool Soc Lond*. 31:87–151.
- Ridoux V. 1994. The diets and dietary segregation of seabirds at the subantarctic Crozet Islands. *Mar Ornithol*. 22:1–192.
- Rieseberg LH. 1997. Hybrid origins of plant species. *Annu Rev Ecol Syst*. 28(1):359–389.
- Rieseberg LH, Archer MA, Wayne RK. 1999. Transgressive segregation, adaptation and speciation. *Heredity* 83(4):363–372.
- Rootes DM. 1988. The status of birds at Signy Island, South Orkney Islands. *Br Antarc Survey Bull*. 80:87–119.
- Rosenthal GG, Schumer M, Andolfatto P. 2018. How the manakin got its crown: a novel trait that is unlikely to cause speciation. *Proc Natl Acad Sci U S A*. 115(18):E4144–E4145.
- Ryan P, Bourgeois K, Dromzée S, Dilley B. 2014. The occurrence of two bill morphs of prions *Pachyptila vittata* on Gough Island. *Polar Biol*. 37(5):727–735.
- Schumer M, Rosenthal GG, Andolfatto P. 2014. How common is homoploid hybrid speciation? *Evolution* 68(6):1553–1560.
- Schumer M, Cui R, Rosenthal GG, Andolfatto P. 2015. Reproductive isolation of hybrid populations driven by genetic incompatibilities. *PLoS Genet*. 11(3):e1005041.
- Schwarz D, Matta BM, Shakir-Botteri NL, McPherson BA. 2005. Host shift to an invasive plant triggers rapid animal hybrid speciation. *Nature* 436(7050):546–549.
- Seehausen O. 2004. Hybridization and adaptive radiation. *Trends Ecol Evol (Amst)*. 19(4):198–207.
- Seehausen O, Butlin RK, Keller I, Wagner CE, Boughman JW, Hohenlohe PA, Peichel CL, Saetre G-P, Bank C, Brännström Å, et al. 2014. Genomics and the origin of species. *Nat Rev Genet*. 15(3):176–192.
- Shapiro LH, Dumbacher JP. 2001. Adenylate kinase intron 5: a new nuclear locus for avian systematics. *Auk* 118(1):248–255.
- Silva MC, Matias R, Wanless RM, Ryan PG, Stephenson BM, Bolton M, Ferrand N, Coelho MM. 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):3122–3137.
- Silva MF, Smith AL, Friesen VL, Bried J, Hasegawa O, Coelho MM, Silva MC. 2016. Mechanisms of global diversification in the marine species Madeiran Storm-petrel *Oceanodroma castro* and Monteiro's Storm-petrel *O. monteiroi*: insights from a multi-locus approach. *Mol Phylogenet Evol*. 98:314–323.
- Stephens M, Smith NJ, Donnelly P. 2001. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet*. 68(4):978–989.
- Szpiech ZA, Jakobsson M, Rosenberg NA. 2008. ADZE: a rarefaction approach for counting alleles private to combinations of populations. *Bioinformatics* 24(21):2498–2504.
- Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123(3):585–595.
- Tickell WLN. 1962. The Dove Prion, *Pachyptila desolata* Gmelin. *Falkland Islands Depend Surv Sci Rep*. 33:1–72.
- Tiedemann R, Von Kistowski KG, Noer H. 1999. On sex-specific dispersal and mating tactics in the common eider *Somateria mollissima* as inferred from the genetic structure of breeding colonies. *Behaviour* 136:1145–1155.
- van Oosterhout C, Hutchinson WF, Wills DP, Shipley P. 2004. Microchecker: software for identifying and correcting genotyping errors in microsatellite data. *Mol Ecol Notes*. 4(3):535–538.
- Viot C, Jouventin P, Bried J. 1993. Population genetics of southern seabirds. *Mar Ornithol*. 21:1–25.
- Warham J. 1969. Notes on some Macquarie Island birds. *Notornis* 16:190–197.
- Welch AJ, Yoshida AA, Fleischer RC. 2011. Mitochondrial and nuclear DNA sequences reveal recent divergence in morphologically indistinguishable petrels. *Mol Ecol*. 20(7):1364–1377.
- West JA, Nilsson R. 1994. Habitat use and burrow densities of burrow-nesting seabirds on South East Island, Chatham Islands, New Zealand. *Notornis* 41:27–37.
- Wilson GA, Rannala B. 2003. Bayesian inference of recent migration rates using multilocus genotypes. *Genetics* 163(3):1177–1191.