Evolutionary factors affecting the cross-species utility of newly developed microsatellite markers in seabirds

YOSHAN MOODLEY[1,*], JUAN F. MASELLO[2], THERESA L. COLE[2,3], LUCIANO CALDERON[2], GOPI K. MUNIMANDA[1], MARCO R. THALI[4], RACHAEL ALDERMAN[5], RICHARD J. CUTHBERT[6], MANUEL MARIN[7,8], MELANIE MASSARO[9], JOAN NAVARRO[10], RICHARD A. PHILLIPS[11], PETER G. RYAN[12], CRISTIÁN G. SUAZO[2], YVES CHEREL[13], HENRI WEIMERSKIRCH[13] & PETRA QUILLFELDT[2]

[1] Konrad Lorenz Institute for Ethology, Department of Integrative Biology and Evolution, University of Veterinary Medicine Vienna, Savoyenstr. 1a, A-1160, Vienna, Austria

[2] Justus Liebig University Giessen, Department of Animal Ecology & Systematics, Heinrich-Buff-Ring 38, D-35392, Giessen, Germany

[3] Trace and Environmental DNA Laboratory, Department of Environment and Agriculture, Curtin University, Perth, WA 6102, Australia

[4] Ecogenics GmbH, Grabenstrasse 11a, 8952, Zurich-Schlieren, Swiss

[5] Department of Primary Industries, Parks, Water and Environment, GPO Box 44, Hobart, Tasmania 7001, Australia

[6] Royal Society for the Protection of Birds (RSPB), The Lodge, Sandy, Bedfordshire SG19 2DL, UK

[7] Section of Ornithology, Natural History Museum of Los Angeles County, 900 Exposition Boulevard, Los Angeles, CA 90007, USA

[8] Feather Link Inc., 1013 Westchester Way, Cincinnati, OH 45244, USA

[9] School of Environmental Sciences, Charles Sturt University, PO Box 789, Albury, NSW, 2640, Australia

[10] Department of Conservation Biology, Estación Biológica de Doñana (EBD-CSIC), Avda. Américo Vespucio s/n, Seville 41092, Spain

[11] British Antarctic Survey, Natural Environment Research Council, High Cross, Madingley Road, Cambridge CB3 0ET, UK

[12] Percy FitzPatrick Institute, DST/NRF Centre of Excellence, University of Cape Town, Rondebosch 7701, South Africa

[13] Centre d'Etudes Biologiques de Chizé, UMR 7372 CNRS-Université de La Rochelle, 79360 Villiers-en-Bois, France

Keywords: cross-species transferability, genetic diversity, microsatellite, null alleles, *Pachyptila*, Procellariiformes

*Correspondence: Yoshan Moodley, fax +43-1-489-09-15-801; E-mail: yoshan.moodley@vetmeduni.ac.at

Running title: Cross-species utility of microsatellites for seabirds

Abstract

Microsatellite loci are ideal for testing hypotheses relating to genetic segregation at fine spatio-temporal scales. They are also conserved among closely related species, making them potentially useful for clarifying interspecific relationships between recently diverged taxa. However, mutations at primer binding sites may lead to increased non-amplification, or disruptions that may lead to decreased polymorphism in non-target species. Furthermore, high mutation rates and constraints on allele size may also lead, with evolutionary time, to an increase in convergently evolved allele size classes, biasing measures of interspecific genetic differentiation. Here, we used next-generation sequencing to develop microsatellite markers from a shotgun genome sequence of the sub-Antarctic seabird, the thin-billed prion (*Pachyptila belcheri*), that we tested for cross-species amplification in other *Pachyptila* and related sub-Antarctic species. We found that heterozygosity decreased and the proportion of non-amplifying loci increased with phylogenetic distance from the target species. Surprisingly, we found that species trees estimated from interspecific F_{ST} provided better approximations of mtDNA relationships among the studied species than those estimated using D_C, even though F_{ST} was more affected by null alleles. We observed a significantly non-linear second order polynomial relationship between microsatellite and mtDNA distances. We propose that the loss of linearity with increasing mtDNA distance stems from an increasing proportion of homoplastic allele size classes that are identical in state, but not identical by descent. Therefore, despite high cross-species amplification success and high polymorphism among the closely related *Pachyptila* species, we caution against the use of microsatellites in phylogenetic inference among distantly related taxa.

67 **Introduction**

68

69 Two-thirds of our planet is covered by sea, and albatrosses, petrels and storm-petrels (Aves,
70 Procellariiformes) are, *par excellence*, the seabirds of the open ocean, only coming ashore to breed,
71 usually on remote islands (Brooke 2004). This highly mobile group of seabirds could theoretically
72 maintain high levels of gene flow, but strong philopatry to breeding islands observed in some species
73 (Ovenden *et al*. 1991, Steeves *et al*. 2005, Bicknell *et al*. 2012) may lead to pronounced genetic
74 differentiation between populations. Our understanding of gene flow and genetic structure in petrels
75 has improved considerably in recent years (Smith *et al*. 2007, Lawrence *et al*. 2008, Gangloff *et al*.
76 2012, Wiley *et al*. 2012, Kerr & Dove 2013) but remains poor for the diverse and often widespread
77 species that breed on sub-Antarctic islands.

78 Petrels (Procellariidae) of the genus *Pachyptila*, the prions, are ideal monitors of ocean
79 productivity as they feed mainly on zooplankton, which responds rapidly to changing environmental
80 conditions (Hunt *et al.* 1992, Bocher *et al*. 2001, Cherel *et al.* 2002, Quillfeldt *et al*. 2007, 2008).
81 Prions are highly mobile and have a wide distribution in sub-Antarctic waters (Onley & Scofield
82 2007). Detailed studies on diet, breeding biology and behavioural ecology have been published for
83 several prion species (Strange 1980, Bretagnolle *et al*. 1990, Liddle 1994, Ridoux 1994, Reid *et al*.
84 1999, Cherel *et al.* 2002, Quillfeldt *et al*. 2003, 2007, 2008, Navarro *et al*. 2013). Much less was
85 known about distributions at-sea, particularly during the nonbreeding season. However, recent stable
86 isotope and tracking studies indicate considerable ecological segregation among populations breeding
87 in the Atlantic and Indian sectors of the Southern Ocean (Cherel *et al*. 2002, 2006, Quillfeldt *et al*.
88 2010, 2013). This spatial and temporal segregation could potentially lead to population differentiation,
89 with consequences for taxonomy and conservation status.

90 Prions are generally clustered into 2 groups, the species with, or without, palatal lamellae, and
91 hence filtering apparatus (Prince & Morgan 1987). The former (the so-called "whale birds") includes
92 the Antarctic prion *Pachyptila desolata*, Salvin's prion *P. salvini* and broad-billed prion *P. vittata*, and
93 the latter, the thin-billed prion *P. belcheri*, fairy prion *P. turtur* and fulmar prion *P. crassirostris*. As

94 yet, there are no phylogeographic studies of any prion species, and only scattered genetic information
95 exists (e.g. Ovenden *et al.* 1991). Nor is there agreement regarding the number of prion species or
96 their genetic relationships (see Brooke 2004, Penhallurick & Wink 2004, Rheindt & Austin 2005,
97 Onley & Scofield 2007). Using enzyme electrophoresis, Barrowclough *et al*. (1981) concluded that
98 Antarctic prions were closely related to blue petrels *Halobaena caerulea*, and Viot *et al*. (1993) that
99 Antarctic prions, thin-billed prions and Salvin's prions were very closely related. The low variation at
100 the mitochondrial cytochrome *b* gene also suggests that Antarctic prions, thin-billed prions and blue
101 petrels are closely related species (Nunn & Stanley 1998). Based on the same locus, Penhallurick &
102 Wink (2004) invoked the multidimensional biological species concept to suggest that all prions
103 represent just two species. However, this last study was heavily criticised (Rheindt & Austin 2005).
104 Consequently, taxonomic authorities and field guides still follow the scheme outlined by Bretagnolle
105 *et al*. (1990) which concluded from the combination of morphometrics, breeding biology, genetics and
106 calls, that Antarctic, Salvin's, thin-billed, and fairy prions were distinct but closely related species.
107 The recent description of distinct thin- and broad-billed morphs, within broad-billed prions that also
108 show strong differences in breeding phenology at Gough Island raises further questions regarding
109 intra- and interspecific relationships of this group (Ryan *et al*. 2014). More genetic data from the many
110 populations of prions that breed in the sub-Antarctic could potentially shed important light on these
111 open questions.
112 Microsatellite loci are powerful tools in population and evolutionary genetics that could
113 provide the resolution for detailed analyses of several aspects of prion biology. Given their high
114 mutation rate, they are ideal for testing hypotheses relating to fine-scale spatio-temporal segregation
115 and for the estimation of demographic parameters such as gene flow, effective population size and
116 genetic variability (Bruford & Wayne 1993, Sunnocks 2000). They are also conserved among closely-
117 related species (Moore *et al.* 1991) making them potentially useful for clarifying interspecific genetic
118 relationships between recently diverged taxa (e.g. Dawson *et al.* 2010), although their high mutation
119 rate means that loci developed for one species may not always be useful for others. This is because
120 mutations at primer binding sites may lead to increased non-amplification (Moodley *et al.* 2006) or

121 disruptions within tandemly repeated elements may lead to a reduced level of observed polymorphism
122 in the non-target species (Garza *et al.* 1995, Primmer *et al.* 2005). The taxa of interest must therefore
123 be closely related in order to maximise utility and offset the costs of microsatellite development.
124 Hence, the testing of newly developed microsatellite loci for cross-species utility has become fairly
125 standard practice (Li *et al.* 2003, Bried *et al.* 2008, Dawson *et al.* 2010, Jan *et al.* 2012, Huang *et al.*
126 2014); however, few surveys have reported statistical trends in their multispecies data sets.
127 Here, we used next-generation sequencing technology to develop a set of 26 polymorphic
128 microsatellite markers from a shotgun genome sequence of the thin-billed prion in order to test for
129 genetic structure among the different populations of this species and to provide an accurate estimation
130 of demographic parameters. We also tested these microsatellite loci for cross-species amplification in
131 other prions, the closely related blue petrel (all Procellariiformes, Procellariidae), and the more
132 distantly related Wilson's storm petrel *Oceanites oceanicus* (Procellariiformes, Hydrobatidae). While
133 prion species appear to be very closely related, we predict, nevertheless, that average observed
134 microsatellite genetic diversity will decrease, whereas the number of non-amplifying alleles will
135 increase, with increasing phylogenetic distance from the species of origin (thin-billed prion). As any
136 increase in null allele frequency could bias the estimation of essential population parameters in
137 phylogeographic studies (e.g. Astanei et al. 2005, Wulff et al. 2012, McCormack et al. 2013) and
138 perhaps even alter phylogenetic relationships, we corrected our raw data for null alleles using methods
139 developed by Chapuis & Estoup (2007). This allowed for a comparison of the effect of null alleles on
140 levels of interspecific differentiation.
141 Lastly, high microsatellite mutation rates coupled with mutational limits on allele sizes
142 (Ostrander *et al.* 1993, Bowcock *et al.* 1994) increase the probability of convergent evolution of allele
143 size classes. Given this highly homoplastic scenario, population parameters and evolutionary
144 hypotheses inferred under the assumption that alleles of the same size share a most recent common
145 ancestor (i.e. are identical by descent) could be biased, even in comparisons between sister taxa
146 (Paetkau *et al.* 1997). We expect, therefore, that microsatellite genetic distance between species will
147 be biased to lower values as the evolutionary time separating species increases.

148 **Materials and Methods**

149

150 *Molecular methods*

151 Between 2010–2012, samples (all from adults) from 77 thin-billed prions, 79 Antarctic prions, 118
152 broad-billed prions, 18 Salvin's prions, 35 fairy prions, 99 blue petrels and 6 Wilson's storm-petrels
153 were obtained in breeding colonies located on sub-Antarctic island groups (Noir, Diego Ramirez,
154 Falkland/Malvinas, South Georgia, Tristan da Cunha, Gough, Marion, Kerguelen, Macquarie, and
155 Chatham). Genomic DNA was obtained from different sample types: blood in ethanol (Gough and
156 Diego Ramirez), blood in Queens's lysis buffer (Kerguelen and Falkland/Malvinas), blood on FTA
157 classic cards (Whatman International Ltd., Maidstone, UK; South Georgia and Chatham), muscle in
158 ethanol (Macquarie, Gough, Tristan da Cunha and Noir) and feather quills (Marion, Tristan da Cunha
159 and Falkland/Malvinas). DNA was extracted from blood (n = 313), feather quills (n = 101) or muscle
160 tissue (n = 51) using the Qiagen DNeasy® Tissue kit (Qiagen, Germany). DNA quantity and quality
161 was determined by UV spectrophotometry using a NanoDrop 1000 Spectrophotometer, and all
162 sampled were standardized to a final concentration of 10 ng/μl.

163 Microsatellite-containing genomic sequences were isolated by ecogenics GmbH (Switzerland)
164 from a 1:1 pool of two thin-billed prion individuals from Mayes Island in the Kerguelen Archipelago
165 using a modified high-throughput genomic sequencing approach (Abdelkrim *et al*. 2009). Genomic
166 DNA was nebulised to 300-800bp and ligated into an ssDNA library. These size-selected fragments
167 were then enriched for tandemly repeated element content by using magnetic streptavidin beads and
168 biotin-labelled CT and GT repeat oligonucleotides. This enriched shotgun library was then sequenced
169 on a Roche 454 next generation platform using the GS-FLX titanium reagents. Resulting sequence
170 reads were passed through quality filters and scanned for microsatellite repeats, from the conserved
171 flanking regions of which primer pairs were designed using Primer 3 (Untergasser *et al.* 2012).

172 After initial testing for amplification and polymorphism, microsatellite loci were visualised in
173 the seven species of petrel through florescent labelling of universal M13 primers, as described in

174 Schuelke (2000). PCR was then performed in a final volume of 10 µl including: 1 × Qiagen PCR
175 buffer, 2 mM dNTPs, 2 µM M13-tailed forward primer, 2 µM reverse primer, 2 µM of universal M13
176 primer 5'-end labelled, 0.5 U Hotstar Taq (Qiagen) and 10 ng template. The PCR program comprised
177 an initial denaturation step of 95°C for 15 min, the cycling parameters were: 30 cycles at 95°C for 30
178 s, an annealing temperature of 56°C for 45 s, 72°C for 45 s, 8 cycles of 95°C for 30 s, 53°C for 45 s,
179 72°C for 45 s, and a final extension step of 72°C for 30 min. PCR products were visualised on a 1.5%
180 agarose gel to confirm successful amplification and to examine negative controls. Products were run
181 on a AB 3130*xl* genetic analyser along with a ROX size-standard. We repeated all PCR reactions for
182 individuals that failed to amplify at >4 loci and for those loci with >10% missing data.
183 We used an 880 bp fragment of the mitochondrial cytochrome *b* gene to estimate the
184 phylogenetic relationships among the seven taxa in our data set, since two of our hypotheses required
185 an independent estimate of interspecific phylogenetic distance. While we do not believe that mtDNA
186 distances are unbiased, we do feel that its slower mutation rate, relative to that of microsatellites,
187 would ensure its linearity among the seven species on our comparison. Furthermore, cytochrome *b* has
188 been used previously to successfully infer relatedness among a much broader sample of
189 Procellariiformes (Nunn & Stanley 1998), and previous morphological and behavioural analyses
190 (Bretagnolle *et al*. 1990) did not specifically quantify interspecific distances among the species in our
191 study. Generic avian cytochrome *b* primers (eg. Patterson *et al.* 2011) were problematic for some
192 samples. Therefore, we designed specific primers (CytB_Pri_F: 5'-CTAGCTATACACTACACCGC-
193 3' and
194 CytB_Pri_R: 5'-CTAGTTGGCCGATGATGATG-3') for our study group from an alignment of those
195 samples that we successfully sequenced. PCRs were conducted in 20µl reaction volumes containing
196 100 ng DNA template, 10 mM of each primer, 10 mM dNTPs (Roth, Karlsruhe), 2 mM MgCl, 5 U
197 *Thermus aquaticus* polymerase (BioLabs Taq DNA polymerase) in a 1x PCR reaction buffer.
198 Thermocycling included initial denaturation at 94°C for 2 minutes, 30 cycles of denaturation at 94°C
199 for 30 s, annealing at 60°C for 45 s and extension at 72°C for 1 min, followed by a final extension step
200 of 5 min at 72°C. 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 3130*xl* genetic analyser (Applied Biosystems). Resulting sequences were assembled and aligned in CLC Main Workbench® 6.9.2.

Data analyses

Genotypes were assigned with GeneMarker 1.85 (SoftGenetics LLC, State College, PA, USA). 20% of the samples were re-scored by a separate individual, with a resulting error rate of <5%. The probability of deviation from Hardy–Weinberg equilibrium (HWE) and non-random association of loci was calculated for each locus/species combination using GENEPOP (Raymond & Rousset 1995; Tables 1 and 2). Measures of genetic diversity (number of alleles per locus (A), observed heterozygosity (H_o) and expected heterozygosity (H_e)) were estimated in Cervus 3.0.3 (Kalinowski *et al*. 2007) and MSA 4.05 (Dieringer & Schlötterer 2003). The inbreeding coefficient (F_{IS}) and its significance were estimated with GENEPOP (Raymond & Rousset 1995). Null allele frequencies (F null) per locus and species were obtained using FreeNA (Chapuis & Estoup 2007).

Phylogenetic distances between species at the cytochrome *b* gene were calculated using the maximum likelihood in MEGA (Tamura *et al.* 2013). We reconstructed the mitochondrial species phylogeny by firstly determining the most suitable substitution model for the cytochrome *b* sequence data using the Akaike information criteria (AIC) in JModeltest 2 (Darriba *et al.* 2012), then set the model parameters to the general time reversible (GTR) model with gamma substitution rate heterogeneity estimated from the data using four rate categories.

We examined the cross-species utility of our isolated microsatellite loci by plotting genetic diversity (H_o) and the proportion of missing data (non-amplifying loci after 3x repeat PCRs, with standardised DNA quantity and quality) in each species against phylogenetic (mtDNA, cytochrome *b*) p-distances, calculated in MEGA (Table S1). We also performed this regression separately for three different phylogenetic groupings: Group 1: all seven species; Group 2: *Halobaena* and *Pachyptila* only; Group 3: *Pachyptila* species only. For a more detailed analysis of these relationships, we used a

generalized linear model (GLM, implemented in R, R Development Core Team 2014), to test the effect of phylogenetic distance (as a covariate) and locus (as a factor) on both observed heterozygosity and the proportion of missing data in each of the three species groupings above. A GLM could not be used to test for the effect of phylogenetic distance and locus on null allele frequencies due to a large amount of missing data in Wilson's storm-petrels.

The frequency of null alleles was calculated in our dataset using FreeNA (Chapuis & Estoup, 2007). This method estimates the frequency of null alleles from data sets simulated to contain and not contain null alleles. Then it uses the expectation-maximisation (EM) algorithm of Dempster *et al.* (1977) to adjust homozygote allele frequencies based on true and false homozygote counts, resulting in the estimation of the null allele frequency. Population differentiation indices can then be calculated including null alleles (INA) and also only on the visible allele sizes, thus, excluding null alleles (ENA). To determine the effect of null alleles on cross-species comparisons, we calculated pair-wise F_{ST} (Wright 1943) and D_C (genetic distance of Cavalli-Sforza & Edwards, 1967) between species using INA and ENA data sets. We re-constructed UPGMA species trees from these triangular F_{ST} and D_C matrices, using MEGA 6.06 (Tamura *et al.* 2013).

We investigated the effect of high mutation rates and constraints on allele size on microsatellite genetic distance by observing the change in the slope of pair-wise Mantel regressions performed on the same three phylogenetic data groupings used in Figure 1. As interspecific microsatellite distance, we used INA and ENA triangular matrices of both F_{ST} and D_C and checked their linearity against the matrix of pair-wise maximum likelihood cytochrome *b* distances calculated previously (Table S1). All Mantel regressions were calculated in GenAlEx 6.5 (Peakall & Smouse 2012). To formally test the hypothesis that F_{ST} and D_C microsatellite distances were non-linearly related to mtDNA distance, we fitted the data (INA and ENA for all species) with both linear and polynomial functions, and performed model testing within a GLM framework in R using the Akaike information criterion (AIC).

Results

Shotgun 454 sequencing of two pooled thin-billed prion genomes resulted in 22,220 reads after quality filtering, with an average read-length of 177 bp (total of 3.9 Mb). Of these, 517 (2.3%) contained microsatellite repeat elements with tetra- or trinucleotides of at least six repeat units, or dinucleotides of at least 10 repeat units. Suitable primer design was possible in 166 reads. We tested 36 of these primer pairs for cross-species amplification and polymorphism among four other unrelated Mayes Island *P. belcheri* and three individuals from the closely related *P. desolata* from Verte Island, also in the Kerguelen Archipelago. Twenty-six loci were identified as polymorphic in the target species, showing clear amplification profiles and reliable amplification in both species tested. We further tested the reliability of amplification and genotypic disequilibrium in a larger set of 77 thin-billed prions from across the breeding range (Kerguelen, Falkland/Malvinas, and Isla Noir in southern Chile). Among populations of the target species, the number of alleles (A) per locus ranged from 5 to 48, the observed heterozygosity (H_o) from 0.325 to 0.880, and the expected heterozygosity (H_e) from 0.493 to 0.972 (Table 1). Cross-species amplification was successful for most primer pairs in Antarctic prions, broad-billed prions, fairy prions, Salvin's prions, and blue petrels, whereas only a third of pairs worked successfully that included the distantly related Wilson's storm petrel (Table 2). One of the 26 loci screened (Pacbel_00829) was found to be in significant linkage equilibrium with locus Pacbel_03731 and locus Pacbel_08509, but the latter two loci appeared statistically unlinked. We therefore removed locus Pacbel_00829 from further analyses.

The mean observed heterozygosity decreased in other prion species, blue petrel and Wilson's storm petrel with increasing mtDNA phylogenetic p-distance (Table S1) from the thin-billed prion (Group 1 including all species, f= 0.6 - 2.7*x, $R^2 = 0.21$, $P < 0.001$; Group 2 *Pachyptila* and *Halobaena*, f= 0.6 - 2.0*x, $R^2 = 0.07$, $P = 0.001$; Group 3 only *Pachyptila*, f= 0.7 - 8.4*x, $R^2 = 0.25$, $P < 0.001$; Table S1, Fig. 1). This tendency was also consistent for most loci in a generalized linear model (GLM), using phylogenetic distance as covariate and locus as factor (Group 1 including all species, effect of distance: $F = 2.189$, $df = 1$, $P < 0.001$, effect of locus: $F = 3.046$, $df = 24$, $P < 0.001$; Group 2 *Pachyptila* and *Halobaena*, effect of distance: $F = 0.468$, $df = 1$, $P < 0.001$, effect of locus: F

281 = 2.843, *df* = 24, *P* < 0.001; Group 3 only *Pachyptila*, effect of distance: *F* = 1.207, *df* = 1, *P* < 0.001,
282 effect of locus: *F* = 2.106, *df* = 24, *P* < 0.001). The proportion of missing data increased slightly
283 (Group 1, f= 0.1 + 1.0*x, R^2 = 0.04, *P* = 0.006; Group 2, f= 0.1 + 0.4*x, R^2 = 0.004, *P* = 0.473; Group
284 3, f= 0.01 + 8.6*x, R^2 = 0.31, *P* < 0.001) with phylogenetic distance from thin-billed prion (Fig. 1),
285 and a GLM confirmed this trend for individual loci (Group 1, effect of distance: *F* = 0.361, *df* = 1, *P* =
286 0.003, effect of locus: *F* = 1.296, *df* = 24, *P* = 0.134; Group 2, effect of distance: *F* = 0.017, *df* = 1, *P* =
287 0.428, effect of locus: *F* = 0.805, *df* = 24, *P* = 0.191; Group 3, effect of distance: *F* = 1.24, *df* = 1, *P* <
288 0.001, effect of locus: *F* = 0.911, *df* = 24, *P* < 0.001).

289 The average frequency of null alleles among the loci and species in our total data set was low
290 (0.076 ± 0.085), although values for some loci/species combinations were quite high (0.364). INA and
291 ENA species trees constructed from pair-wise species F_{ST} (Fig. 2A) and D_C (Fig. 2B) values (Table
292 S2) were superimposed onto each other for comparison. F_{ST} values tended to decrease more than D_C
293 when corrected for the presence of null alleles (black relative to grey branches, Fig. 2A/B). This
294 correction did not alter the relationships between taxa for either measure of genetic differentiation.
295 Both trees were compared for topological congruence with the mtDNA phylogeny of the cytochrome *b*
296 gene (Fig. 2C). As with the mtDNA phylogeny, F_{ST} and D_C trees separated the genus *Pachyptila* from
297 outgroup genera *Halobaena* and *Oceanites*, and positioned the fairy prions basally within the
298 *Pachyptila* clade. All trees differed regarding the placement of the most derived *Pachyptila* taxa. F_{ST}
299 more closely approximated the mtDNA phylogeny in that *P. desolata* and *P. salvini* were sister taxa,
300 but they differed with respect to the placement of *P. belcheri* and *P. vittata* (Figure 2).

301 Mantel regressions (Fig. 3) indicated that a large proportion of DNA sequence variance in the
302 cytochrome *b* data could be significantly explained by the multilocus microsatellite distance statistics
303 F_{ST} and D_C (Rxy > 0.9 in all cases, see Fig. 3). However, the relationship between microsatellite and
304 mtDNA distance values changed markedly among the three groups of species tested, and depended on
305 which species were included. The slopes of the regressions including all species (Group 1) were
306 lowest and increased incrementally as more phylogenetically similar taxa (Group 2, only *Pachyptila*
307 and *Halobaena*; Group 3, only *Pachyptila* species) were grouped together. This effect was more

pronounced for D_C, with shallower gradients differentiating F_{ST} regressions. Microsatellite distance statistics calculated including null alleles almost always resulted in a steeper gradient than ENA values, but this difference was smallest in the group that contained all seven species.

We tested the hypothesis that microsatellite distances were non-linear with evolutionary time by model fitting. We found that linear functions provided a closer fit to F_{ST} distances than to D_C distances, but that second order (quadratic) polynomials provided a significantly better fit than linear functions for both distance statistics (Table 3).

Discussion

Non-amplification and null alleles

We amplified microsatellite loci in 432 individual samples in seven species of petrels. Concordant with expectation, we found that genetic diversity decreased, and the proportion of non-amplifying (missing) data, increased with phylogenetic distance from the target species. Although global regressions (Group 1) as well as groups containing *Halobaena* and *Pachyptila* species (Group 2) and *Pachyptila* species (Group 3) were highly significant in most cases, the trend was not observed in all loci, resulting in shallow regression gradients. Nevertheless, this confirmed our expectation that genetic diversity decreases and missing data increases with evolutionary distance from the target species and is compatible with other studies that show increases in non-amplification and decline in polymorphism (e.g. Li *et al.* 2003, Primmer *et al.* 2005, Bried *et al.* 2008, Dawson *et al.* 2010, Jan et al. 2012).

The average frequency of null alleles in our data set was low, and therefore correcting allele frequencies for the presence of null alleles resulted in no change to overall species tree topologies but decreased F_{ST} branch lengths (Fig. 2). D_C branch lengths, on the other hand, differed much less between corrected and uncorrected data sets, implying that this statistic is more robust to the presence of null alleles.

334 *Utility in analysing interspecific relationships*

335 We found that species trees estimated from interspecific microsatellite data were largely congruent
336 with mtDNA relationships among the studied species, with F_{ST} providing a slightly better
337 approximation than D_C distances. This is a surprising result, since F_{ST} is a fixation index, and as such
338 does not satisfy the triangle inequality as would true distance measures like D_C. Our data also showed
339 the inherent problem posed by null alleles in the resolution of interspecific branch lengths, which
340 affected F_{ST} more than D_C (Figs. 2A, B). Furthermore, F_{ST} has often been criticised for inaccurately
341 estimating population differentiation when genetic variation is high (Charlesworth 1998, Balloux &
342 Lugon-Moulin 2002, Carreras-Carbonell *et al.* 2006, Jost 2008). Nevertheless, F_{ST} is perhaps the most
343 reported statistic in population and evolutionary genetics. These comparisons with the cytochrome *b*
344 phylogeny, however, do not account for potential biases in mtDNA itself, nor for differences in tree-
345 building algorithms used for microsatellite (UPGMA) and mtDNA data (maximum likelihood).
346 Therefore, we stress the need for a more thorough reappraisal of the phylogenetic relationships among
347 the prions, using multiple but more slowly evolving nuclear intronic gene sequences.
348
349 *Linearity of microsatellite genetic distances*

350 The high proportion of explained variation in Mantel regressions of microsatellite and mtDNA genetic
351 distance implies that variation was similarly distributed between the both microsatellite and
352 mitochondrial data sets. However, when regression analyses of the three groups of varying species
353 diversity imposed a linear relationship between microsatellite and mtDNA distance, the slope of the
354 regression changed considerably, suggesting that the true relationship was non-linear. Instead, at lower
355 distance values (e.g. among congeners; Group 3), the relationship appears linear, but microsatellite
356 distance gradually reaches a plateau with increasing mtDNA distance (Fig. 3A, B, Groups 1 and 2),
357 implying that the latter statistic is a better estimator of relationships among distantly related taxa. We
358 tested the hypothesis that microsatellite distances are not linear with evolutionary distance by fitting
359 both linear and polynomial functions to the F_{ST} and D_C data and found that in both cases a quadratic
360 function best fitted the data (Table 3).

Interestingly, F_{ST} appeared to remain linear for longer than D_C, especially when corrected for the presence of null alleles (see higher *P* values, Table 3). While this suggests that F_{ST} might be more useful at higher phylogenetic levels, its usefulness is compromised by its higher variance compared to D_C. Because neither microsatellite distance measure maintained linearity in pairwise intergeneric comparisons, we recommend that analyses of genetic differentiation restrict F_{ST} and D_C to studies in which the target species is closely related to the species from which the markers were developed. In either case, ENA correction for null alleles is essential.

Despite high mutation rates of microsatellites, simulations indicate that measures of genetic differentiation will remain linear much longer without constraints in allele size (Nauta & Weissing 1996). Therefore, we propose that the non-linearity we observed at the intergeneric level is a natural consequence of constrained microsatellite allele size that leads to an increase in the number of convergently evolved allele size classes that, while identical in state, are no longer identical by descent in intergeneric pairwise comparisons. Estoup *et al.* (2002) suggested that at the intraspecific level, the high mutation rates of microsatellites will compensate for the inevitable convergent evolution of some allele classes, while Paetkau *et al.* (1997) detected a loss of linearity among closely related sister taxa (brown bear *Ursus arctos* and polar bear *U. maritimus*). Given that variation in our set of microsatellites remains linear within the genus *Pachyptila*, perhaps because of a slightly lower mutation rate, or a slightly larger maximum repeat size, we are confident that population genetic and demographic analyses at this level will not be compromised by constraints in allele size.

Conclusions

We show here that a panel of 25 microsatellite loci developed using next generation sequencing of a thin-billed prion shotgun library may be applied in studies of molecular ecology among congeners; however, this approach may result in a greater proportion of null alleles and lower amounts of genetic diversity in the non-target species. Genetic diversities therefore may not be directly comparable between species, despite the use of the same conserved microsatellite markers. In addition, the contrasting results from the two measures of differentiation lead us to discourage the use of these

388 microsatellites in phylogenetic reconstruction beyond the genus level, as even at that level this may be
389 associated with high variance.
390
391
392 **Acknowledgements**
393
394 The work was funded by a grant provided by the German Science Foundation DFG (Qu 148/5). We
395 would like to thank the New Island Conservation Trust with assistance from Ian, Maria and Georgina
396 Strange, and Benno H. Lüthi, Klemens Pütz and Gerhard Meyer from the Antarctic Research Trust,
397 for crucial support during the fieldwork. We extend special thanks to Ruth Brown for her fieldwork
398 support at Bird Island (South Georgia) and to Jaime A. Cursach for his assistance during fieldwork at
399 Diego Ramirez (Chile). We thank Antje Schreiner for help with the lab work.
400
401
402 **References**
403 Abdelkrim J, Robertson B, Stanton J, Gemmell N (2009) Fast, cost-effective development of species-
404 specific microsatellite markers by genomic sequencing. *BioTechniques* **46**, 185-192.
405 Astanei I, Gosling E, Wilson JIM, Powell E (2005) Genetic variability and phylogeography of the
406 invasive zebra mussel, *Dreissena polymorpha* (Pallas). *Molecular Ecology* **14**, 1655-1666.
407 Balloux F, Lugon-Moulin N (2002) The estimation of population differentiation with microsatellite
408 markers. *Molecular Ecology* **11**, 155-165.
409 Barrowclough GF, Corbin KW, Zink RM (1981) Genetic differentiation in the Procellariiformes.
410 *Comparative Biochemistry and Physiology Part B: Comparative Biochemistry* **69**, 629-632.
411 Bicknell A, Knight M, Bilton D*, et al.* (2012) Population genetic structure and long-distance dispersal
412 among seabird populations: Implications for colony persistence. *Molecular Ecology* **21**, 2863-2876.
413 Bocher P, Cherel Y, Labat JP*, et al.* (2001) Amphipod-based food web: *Themisto gaudichaudii* caught
414 in nets and by seabirds in Kerguelen waters, southern Indian Ocean. *Marine Ecology Progress*

415 *Series* **223**, 261-276.
416 Bowcock A, Ruiz-Linares A, Tomfohrde J*, et al.* (1994) High resolution of human evolutionary trees
417 with polymorphic microsatellites. *Nature* **368**, 455-457.
418 Bretagnolle V, Zotier R, Jouventin P (1990) Comparative population biology of four prions (Genus
419 *Pachyptila*) from the Indian Ocean and consequences for their taxonomic status. *Auk* **107**, 305-316.
420 Bried J, Dubois MP, Jouventin P, Santos RS (2008) Eleven polymorphic microsatellite markers in
421 Cory's shearwater, *Calonectris diomedea*, and cross-species amplification on threatened
422 Procellariiformes. *Molecular Ecology Resources* **8**, 602-604.
423 Brooke M (2004) *Albatrosses and petrels across the world.* Oxford University Press, Oxford.
424 Bruford MW, Wayne RK (1993) Microsatellites and their application to population genetic studies.
425 *Current Opinion in Genetics & Development* **3**, 939-943.
426 Carreras-Carbonell J, Macpherson E, Pascual M (2006) Population structure within and between
427 subspecies of the Mediterranean triplefin fish *Tripterygion delaisi* revealed by highly polymorphic
428 microsatellite loci. *Molecular Ecology* **15**, 3527-3539.
429 Cavalli-Sforza LL, Edwards AW (1967) Phylogenetic analysis. Models and estimation procedures.
430 *American Journal of Human Genetics* **19**, 233-257.
431 Chapuis M-P, Estoup A (2007) Microsatellite null alleles and estimation of population differentiation.
432 *Molecular Biology and Evolution* **24**, 621-631.
433 Charlesworth B (1998) Measures of divergence between populations and the effect of forces that
434 reduce variability. *Molecular Biology and Evolution* **15**, 538-543.
435 Cherel Y, Bocher P, de Broyer C, Hobson KA (2002) Food and feeding ecology of the sympatric thin-
436 billed *Pachyptila belcheri* and Antarctic *P. desolata* prions at Iles Kerguelen, Southern Indian
437 Ocean. *Marine Ecology Progress Series* **228**, 263-281.
438 Cherel Y, Phillips RA, Hobson KA, McGill R (2006) Stable isotope evidence of diverse species-
439 specific and individual wintering strategies in seabirds. *Biology Letters* **2**, 301-303.
440 Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics and
441 parallel computing. *Nature Methods* **9**, 772-772.

442 Dawson DA, Horsburgh GJ, Küpper C*, et al.* (2010) New methods to identify conserved microsatellite
443 loci and develop primer sets of high cross-species utility – as demonstrated for birds. *Molecular*
444 *Ecology Resources* **10**, 475-494.
445 Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM
446 algorithm. *Journal of the Royal statistical Society* **39**, 1-38.
447 Dieringer D, Schlötterer C (2003) Microsatellite analyser (MSA): a platform independent analysis tool
448 for large microsatellite data sets. *Molecular Ecology Notes* **3**, 167-169.
449 Estoup A, Jarne P, Cornuet J-M (2002) Homoplasy and mutation model at microsatellite loci and their
450 consequences for population genetics analysis. *Molecular Ecology* **11**, 1591-1604.
451 Gangloff B, Shirihai H, Watling D*, et al.* (2012) The complete phylogeny of *Pseudobulweria*, the most
452 endangered seabird genus: systematics, species status and conservation implications. *Conservation*
453 *Genetics* **13**, 39-52.
454 Garza JC, Slatkin M, Freimer NB (1995) Microsatellite allele frequencies in humans and chimpanzees,
455 with implications for constraints on allele size. *Molecular Biology and Evolution* **12**, 594-603.
456 Huang D, Zhang Y, Jin M*, et al.* (2014) Characterization and high cross-species transferability of
457 microsatellite markers from the floral transcriptome of *Aspidistra saxicola* (Asparagaceae).
458 *Molecular Ecology Resources* **14**, 569–577.
459 Hunt GL, Priddle J, Whitehouse MJ, Veit RR, Heywood RB (1992) Changes in Seabird Species
460 Abundance Near South Georgia During A Period of Rapid Change in Sea-Surface Temperature.
461 *Antarctic Science* **4**, 15-22.
462 Jan C, Dawson DA, Altringham JD, Burke T, Butlin RK (2012) Development of conserved
463 microsatellite markers of high cross-species utility in bat species (Vespertilionidae, Chiroptera,
464 Mammalia). *Molecular Ecology Resources* **12**, 532-548.
465 Jost LOU (2008) G_{ST} and its relatives do not measure differentiation. *Molecular Ecology* **17**, 4015-
466 4026.
467 Kalinowski ST, Taper ML, Marshall TC (2007) Revising how the computer program cervus
468 accommodates genotyping error increases success in paternity assignment. *Molecular Ecology* **16**,

1099-1106.

Kerr KC, Dove CJ (2013) Delimiting shades of gray: phylogeography of the Northern Fulmar, *Fulmarus glacialis*. *Ecology and Evolution* **3**, 1915-1930.

Lawrence HA, Taylor GA, Millar CD, Lambert DM (2008) High mitochondrial and nuclear genetic diversity in one of the world's most endangered seabirds, the Chatham Island Taiko (*Pterodroma magentae*). *Conservation Genetics* **9**, 1293-1301.

Li G, Hubert S, Bucklin K, Ribes V, Hedgecock D (2003) Characterization of 79 microsatellite DNA markers in the Pacific oyster *Crassostrea gigas*. *Molecular Ecology Notes* **3**, 228-232.

Liddle GM (1994) Interannual variation in the breeding biology of the Antarctic prion *Pachyptila desolata* at Bird Island, South Georgia. *Journal of Zoology* **234**, 12-139.

McCormack JE, Hird SM, Zellmer AJ, Carstens BC, Brumfield RT (2013) Applications of next-generation sequencing to phylogeography and phylogenetics. *Molecular Phylogenetics and Evolution* **66**, 526-538.

Moodley Y, Baumgarten I, Harley E (2006) Horse microsatellites and their amenability to comparative equid genetics. *Animal Genetics* **37**, 258-261.

Moore S, Sargeant L, King T*, et al.* (1991) The conservation of dinucleotide microsatellites among mammalian genomes allows the use of heterologous PCR primer pairs in closely related species. *Genomics* **10**, 654-660.

Nauta MJ, Weissing FJ (1996) Constraints on allele size at microsatellite loci: implications for genetic differentiation. *Genetics* **143**, 1021-1032.

Navarro J, Votier SC, Aguzzi J*, et al.* (2013) Ecological segregation in space, time and trophic niche of sympatric planktivorous petrels. *PLoS ONE* **8**, e62897.

Nunn GB, Stanley SE (1998) Body size effects and rates of cytochrome b evolution in tube-nosed seabirds. *Molecular Biology and Evolution* **15**, 1360-1371.

Onley D, Scofield P (2007) *Albatrosses, petrels and shearwaters of the world.* Christopher Helm, London.

Ostrander EA, Sprague Jr GF, Rine J (1993) Identification and characterization of dinucleotide repeat

(CA)n markers for genetic mapping in dog. *Genomics* **16**, 207-213.

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.

Paetkau D, Waits LP, Clarkson PL, Craighead L, Strobeck C (1997) An empirical evaluation of genetic distance statistics using microsatellite data from bear (Ursidae) populations. *Genetics* **147**, 1943-1957.

Patterson SA, Morris-Pocock JA, Friesen VL (2011) A multilocus phylogeny of the Sulidae (Aves: Pelecaniformes). *Molecular Phylogenetics and Evolution* **58**, 181-191.

Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics* **28**, 2537-2539.

Penhallurick J, Wink M (2004) Analysis of the taxonomy and nomenclature of the Procellariiformes based on complete nucleotide sequences of the mitochondrial cytochrome b gene. *Emu* **104**, 125-147.

Primmer CR, Painter JN, Koskinen MT, Palo JU, Merilä J (2005) Factors affecting avian cross-species microsatellite amplification. *Journal of Avian Biology* **36**, 348-360.

Quillfeldt P, Masello JF, McGill RAR, Adams M, Furness RW (2010) Moving polewards in winter: a recent change in migratory strategy. *Frontiers in Zoology* **7**, 15.11-15.11.

Quillfeldt P, Masello JF, Navarro J, Phillips RA (2013) Year-round distribution suggests spatial segregation of two small petrel species in the South Atlantic. *Journal of Biogeography* **40**, 430-441.

Quillfeldt P, Masello JF, Strange I (2003) Breeding biology of the Thin-billed prion P*achyptila belcheri* at New Island, Falkland Islands, in the poor season 2002/2003: Egg desertion, breeding success and chick provisioning. *Polar Biology* **26**, 746-752.

Quillfeldt P, McGill RAR, Strange IJ*, et al.* (2008) Stable isotope analysis reveals sexual and environmental variability and individual consistency in foraging of Thin-billed prions. *Marine Ecology Progress Series* **373**, 137-148.

Quillfeldt P, Strange I, Masello JF (2007) Sea surface temperatures, variable food supply and

behavioural buffering capacity in Thin-billed prions *Pachyptila belcheri*: breeding success, provisioning and chick begging. *Journal of Avian Biology* **38**, 298-308.

R Development Core Team (2014) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna.

Raymond M, Rousset F (1995) GENEPOP Version 3.1d: population genetics software for exact tests and ecumenism. *Journal of Heredity* **86**, 248-249.

Reid K, Liddle GM, Prince PA, Croxall JP (1999) Measurement of chicks provisioning in Antarctic prions *Pachyptila desolata* using an automated weighing system. *Journal of Avian Biology* **30**, 127-134.

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**, 181-186.

Ridoux V (1994) The diets and dietary segregation of seabirds at the subantarctic Crozet Islands. *Marine Ornithology* **22**, 1-192.

Ryan P, Bourgeois K, Dromzée S, Dilley B (2014) The occurrence of two bill morphs of prions *Pachyptila vittata* on Gough Island. *Polar Biology* **37**, 727-735.

Schuelke M (2000) An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology* **18**, 233-234.

Smith AL, Monteiro L, Hasegawa O, Friesen VL (2007) Global phylogeography of the band-rumped storm-petrel (*Oceanodroma castro*; Procellariiformes: Hydrobatidae). *Molecular Phylogenetics and Evolution* **43**, 755-773.

Steeves TE, Anderson DJ, Friesen VL (2005) A role for nonphysical barriers to gene flow in the diversification of a highly vagile seabird, the masked booby (*Sula dactylatra*). *Molecular Ecology* **14**, 3877-3887.

Strange I (1980) The thin-billed prion, *Pachyptila belcheri,* at New Island, Falkland Islands. *Gerfaut* **70**, 411-445.

Sunnucks P (2000) Efficient genetic markers for population biology. *Trends in Ecology & Evolution* **15**, 199-203.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution* **30**, 2725-2729.

Untergasser A, Cutcutache I, Koressaar T*, et al.* (2012) Primer3–new capabilities and interfaces. *Nucleic Acids Research* **40**, e115.

Viot C, Jouventin P, Bried J (1993) Population genetics of southern seabirds. *Marine Ornithology* **21**, 1-25.

Wiley AE, Welch AJ, Ostrom PH*, et al.* (2012) Foraging segregation and genetic divergence between geographically proximate colonies of a highly mobile seabird. *Oecologia* **168**, 119-130.

Wright S (1943) Isolation by distance. *Genetics* **28**, 114-138.

Wulff A, Hollingsworth PM, Haugstetter J*, et al.* (2012) Ten nuclear microsatellites markers cross-amplifying in *Scaevola montana* and *S. coccinea* (Goodeniaceae), a locally common and a narrow endemic plant species of ultramafic scrublands in New Caledonia. *Conservation Genetics Resources* **4**, 725-728.

Author Contributions

PQ and YM conceived and designed the study. RA, YC, RJC, M Marin, JFM, M Massaro, JN, RAP, PQ, PGR, CGS and HW carried out the extensive fieldwork. MRT isolated the microsatellite sequences. TLC, JFM and GKM screened the samples. JFM, YM, TLC and LC carried out the bioinformatic analyses. YM, JFM and PQ drafted the manuscript. All authors reviewed the final draft of the manuscript.

572 **Data Accessibility**

573

574 DNA sequences: Genbank accessions KP122163-KP122196, KM050769 and KM050770. Shotgun

575 DNA sequence reads, 432 microsatellite genotypes at 25 loci, cytochrome *b* alignment, distance

576 matrices and tree files: Provisional DRYAD entry doi:10.5061/dryad.rc917.

577

578

579 **Supporting Information**

580 **Table S1** Phylogenetic p-distances and maximum likelihood distances (cytochrome *b*) in Pachyptila

581 species, blue petrels and Wilson's storm petrels

582

583 Figure legends

584 **Fig. 1** Observed heterozygosity (A) and proportion of missing data (B) per microsatellite locus
585 (black circles) in seven procellariiform species, against mitochondrial cytochrome *b* phylogenetic
586 distance from the species of origin, the thin-billed prion *Pachyptila belcheri.* Regressions were carried
587 out separately for three different groups of species: Group one (red) includes all species (thin-billed
588 prion *Pachyptila belcheri* (b), Antarctic prion *P. desolata* (d), broad-billed prion *P. vittata* (v), fairy
589 prion *P. turtur* (t), and Salvin's prion *P. salvini* (s), blue petrel *Halobaena caerulea* (h) and Wilson's
590 storm-petrel *Oceanites oceanicus*(o)), Group two (green) includes *Pachyptila* species and *Halobaena*
591 *caerulea* only, and Group three (blue) includes only the closely related *Pachyptila* species. The size of
592 the circles indicates the frequency of particular proportions among microsatellite loci, as shown in the
593 figure panel. The line between both graphs represents the phylogenetic distance between each species.
594 Regression lines of mean values are shown for each group.
595

596 **Fig. 2** Nuclear and mitochondrial phylogenetic relationships among thin-billed prions *Pachyptila*
597 *belcheri*, Antarctic prions *P. desolata*, broad-billed prions *P. vittata*, fairy prions *P. turtur*, Salvin's
598 prions *P. salvini*, blue petrels *Halobaena caerulea*, and Wilson's storm-petrels *Oceanites oceanicus*.
599 **A/B**. UPGMA species trees reconstructed using interspecies F_{ST} (**A**) and D_C distances (**B**) from
600 microsatellite data sets that include null alleles (INA, in grey) and exclude null alleles (ENA, black).
601 **C**. Phylogeny of the mitochondrial cytochrome *b* gene reconstructed via maximum likelihood using
602 the GTR substitution model with gamma-distributed rate heterogeneity.
603

604 **Fig. 3** Mantel regressions of pairwise microsatellite distances F_{ST} (**A**) and D_C (**B**) against
605 mitochondrial cytochrome *b* phylogenetic distance calculated for three groups of Procellariiform
606 species. Microsatellite genetic distances tend to lose linearity for groups that include species
607 increasingly distant to the species of origin, *Pachyptila belcheri*. Microsatellite distances are also
608 given including null alleles (INA, squares and solid lines) and excluding null alleles (ENA, triangles
609 and dashed lines) after Chapuis & Estoup (2007). Species are grouped and colour-coded as in Fig. 1.

610 Black lines describe the polynomial function that best fits the INA (solid lines) and ENA (dashed
611 lines) microsatellite distance data.

Table 1 Locus characteristics of genetic variation at 25 newly-isolated microsatellites in the target species, the thin-billed prion *Pachyptila belcheri.*

Locus code & name	Repeat type[§]	Primer sequences 5'–3'	Size range	*N*	*A*	H_o	H_e	F_{IS}	F (null)
A Pacbel_00386	$(AC)_{14}$	F: GCATGTCTACAAACAAGCACG R: TCACTGGAAACCAGAGTAGGC	120–142	72	11	0.764	0.811	0.058^{ns}	0.014
D Pacbel_02653	$(AC)_{12}$	F: AGCCATAGCTCAGTACAAGTTC R: TGCAGGCATTTCAGGTTTGG	132–170	77	12	0.325	0.639	0.494^{***}	0.203
E Pacbel_03731	$(AC)_{14}$	F: TAGTGGACTGGTCACAGCAC R: TAGCAGCTGGAGAGCATCAG	122–268	74	48	0.392	0.972	0.599^{ns}	0.293
F Pacbel_04240	$(AC)_{14}$	F: CCCATTGTCTGGGCAAAGC R: GCATTCTTTGTGGGGATGGG	166–254	47	19	0.511	0.815	0.376^{***}	0.164
G Pacbel_04355	$(AC)_{17}$	F: TACCAGGGACAATCTGGGTG R: GGAAAAATACAGGAGATGCTTGAG	158–212	69	20	0.579	0.931	0.379^{***}	0.181
H Pacbel_04991	$(GT)_{14}$	F: TGTCCATGAGGTCTGGAAGC R: GGTGGAATACAGGGATGCAC	86–106	74	11	0.757	0.877	0.138^{ns}	0.064
Z Pacbel_07265	$(GT)_{15}$	F: CGTCACTTTAATAGCGCTGGC R: ACCCTGATTTTCCCAGTCCG	148–182	74	15	0.730	0.817	0.108^{ns}	0.040

I	$(TG)_{12}$	F: TCTGGTTTCACAAATACCTACTGC	156–172	73	9	0.849	0.808	-0.052^{ns}	0.000
Pacbel_08509		R: CCTAGTTTCGACACAAAGGATGG							
Ñ	$(GA)_{13}$	F: TTTGGTCAATTTTCCCTCGC	138–154	74	8	0.689	0.690	0.002^{ns}	0.000
Pacbel_08867		R: ACAGAAAACCAATGTTGTTAATAGG							
J	$(CT)_{12}$	F: CTGATCGGTTGTGCTCTGTG	184–202	74	10	0.757	0.692	-0.094^{ns}	0.000
Pacbel_08988		R: GCGGAAAGATCCTAACAAGCC							
K	$(GT)_{12}$	F: ATCTGCGCATGCAGTGATAG	208–254	76	17	0.829	0.901	0.080^{ns}	0.034
Pacbel_09021		R: CACAGCTAGCAGCATTGACC							
L	$(AC)_{12}$	F: AACTGTTTGCTCCACACCAC	146–170	75	9	0.600	0.809	0.260^{**}	0.118
Pacbel_09528		R: ATGGCTTGGAAGTCTCCCTG							
M	$(GT)_{13}$	F: GCTTTATTTAAGAGCAACAAAAACTTC	92–110	73	10	0.822	0.830	0.010^{ns}	0.026
Pacbel_09957		R: ACAAAGCAAACCTAATCATTCCC							
N	$(TG)_{12}$	F: CAACGCGCTTTTGGTTTTGC	102–120	77	9	0.688	0.811	0.152^{ns}	0.058
Pacbel_10033		R: GGCCACTCACCACAATACAAG							
O	$(AC/AT)_{8}$	F: AGCTTTCTGTCTGGTAGCAC	158–196	75	19	0.720	0.890	0.192^{**}	0.092
Pacbel_10895		R: TGCTCCTGCCTAAGCTACG							

S	$(AC)_{12}$	F: CCAAACCCTGCCCCGATG	92–116	74	11	0.419	0.805	0.482^{***}	0.211
Pacbel_12344		R: GCCGTGCAGACGTGAATAG							
T	$(TG)_{13}$	F: CAAGCTGGTTTTCAATGTGCC	254–266	76	7	0.697	0.728	0.042^{ns}	0.020
Pacbel_15293		R: CTGAAGCATTAGCACCTGCC							
Q	$(CA)_{13}$	F: TTCTTGTAGCAGTAGGAGACC	146–162	75	8	0.627	0.674	0.071^{*}	0.038
Pacbel_15327		R: ACCTCATGTGTAAAACCTGCC							
R	$(GT)_{13}$	F: TGAAGGTATGCCTGTCCTCC	126–134	75	5	0.640	0.598	-0.071^{ns}	0.000
Pacbel_16671		R: TCGCTCCCACACACATGC							
V	$(CA)_{12}$	F: TGCTTTTGGACAATGTGGAGG	100–120	75	10	0.653	0.670	0.025^{ns}	0.026
Pacbel_16989		R: TCTGGTACACTTCTCATTTGGAC							
W	$(AG)_{14}$	F: TGCAAGGTCTTGTGATGAAGC	142–164	76	12	0.842	0.821	-0.026^{ns}	0.000
Pacbel_17529		R: AATGCAATTTGTCTGCGGGG							
X	$(TG)_{13}$	F: TACAACCGTTCTCCCTGTGG	228–254	75	12	0.880	0.828	-0.063^{ns}	0.000
Pacbel_17944		R: GGAGAAGCAGGCAGCAATAC							
U	$(GT)_{12}$	F: ATAACCCAGTGTGATGGTGC	204–212	75	5	0.507	0.493	-0.028^{ns}	0.030
Pacbel_17986		R: CACAGCTGCTTAGTGCACAG							

Y Pacbel_19907	$(AG)_{12}$	F: TTTCTCCTTAGCTCGGCAGG R: CCATACTTGGTGGCAGTGTG	166–184	74	8	0.622	0.642	0.033^{ns}	0.000
P Pacbel_20784	$(GT)_{12}$	F: GCAAACGCAAGGCGTACAAG R: ATGGTAGCAAACCTCCTGCC	122–158	76	12	0.500	0.833	0.401^{***}	0.177

[§] Number of repeats indicated in the subscript. Primer annealing temperature, *Ta* = 56°C. *N*: number of individuals with reliable amplification. *A*: number of alleles. H_o: observed heterozygosity. H_e: expected heterozygosity. F_{IS}: inbreeding coefficient. The probabilities of deviation from Hardy–Weinberg equilibrium (HWE) are indicated by asterisks (* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, ns: not significant). F (null): null allele frequency estimate.

Table 2 Cross-species genetic variation of microsatellites isolated from 25 thin-billed prions *Pachyptila belcheri* in Antarctic prions *P. desolata* (des), broad-billed prions *P. vittata* (vit), fairy prions *P. turtur* (tur), Salvin's prions *P. salvini* (sal), blue petrels *Halobaena caerulea* (car) and Wilson's storm-petrels *Oceanites oceanicus* (oce).

Locus code	Sp	*N*	*A*	H_o	H_e	F_{IS}	F (null)	Sp	*N*	*A*	H_o	H_e	F_{IS}	F (null)	Sp	*N*	*A*	H_o	H_e	F_{IS}	F (null)
A	car	97	13	0.835	0.897	0.070^{*}	0.017	des	79	13	0.810	0.862	0.061^{ns}	0.011	tur	33	2	0.424	0.403	-0.054^{ns}	0.000
	oce	1	2	1	1	-1.000^{nc}	0.000	vit	117	11	0.675	0.828	0.185^{***}	0.074	sal	18	10	0.778	0.856	0.093^{ns}	0.033
D	car	95	4	0.042	0.082	0.489^{***}	0.088	des	77	11	0.468	0.754	0.381^{***}	0.167	tur	33	5	0.272	0.525	0.484^{***}	0.167
	oce	6	2	0.167	0.167	0.000^{nc}	0.000	vit	115	9	0.313	0.642	0.513^{***}	0.200	sal	17	8	0.471	0.663	0.297^{*}	0.132
E	car	95	13	0.779	0.795	0.020^{ns}	0.022	des	75	51	0.653	0.971	0.329^{**}	0.157	tur	1	1	0.000	0.000	–	0.001
	oce	6	1	0.000	0.000	–	0.001	vit	111	39	0.649	0.951	0.319^{***}	0.157	sal	17	22	0.588	0.973	0.403^{***}	0.184
F	car	83	13	0.313	0.782	0.601^{***}	0.265	des	48	17	0.438	0.927	0.531^{***}	0.252	tur	17	9	0.235	0.818	0.719^{***}	0.311
	oce	2	3	0.500	0.833	0.500^{ns}	0.001	vit	65	20	0.477	0.872	0.455^{***}	0.204	sal	13	9	0.615	0.840	0.276^{*}	0.104
G	car	93	10	0.731	0.829	0.118^{*}	0.056	des	76	20	0.684	0.930	0.265^{***}	0.127	tur	14	9	0.429	0.820	0.487^{***}	0.200
	oce	5	6	0.400	0.844	0.556^{**}	0.189	vit	116	23	0.750	0.934	0.198^{***}	0.093	sal	13	9	0.385	0.806	0.533^{***}	0.203
H	car	92	8	0.152	0.259	0.413^{***}	0.116	des	78	10	0.795	0.893	0.111^{*}	0.050	tur	33	4	0.333	0.526	0.370^{*}	0.122

	oce	0	0	–	–	–	–	vit	113	11	0.655	0.826	0.208^{***}	0.094	sal	15	9	0.800	0.880	0.094^{ns}	0.043
Z	car	85	28	0.635	0.947	0.331^{***}	0.160	des	72	15	0.694	0.785	0.116^{ns}	0.057	tur	21	6	0.619	0.560	-0.109^{ns}	0.000
	oce	5	5	0.400	0.844	0.556^{*}	0.203	vit	103	14	0.709	0.777	0.088^{**}	0.033	sal	16	7	0.625	0.756	0.178^{ns}	0.048
I	car	95	13	0.832	0.846	0.017^{ns}	0.015	des	75	11	0.720	0.785	0.083^{*}	0.027	tur	16	5	0.438	0.688	0.371^{*}	0.156
	oce	6	2	0.167	0.167	0.000^{nc}	0.000	vit	111	12	0.730	0.756	0.034^{**}	0.032	sal	11	6	0.636	0.805	0.218^{ns}	0.073
Ñ	car	82	10	0.841	0.795	-0.059^{ns}	0.000	des	78	9	0.590	0.664	0.113^{ns}	0.041	tur	26	6	0.577	0.728	0.211^{**}	0.060
	oce	6	2	0.500	0.409	-0.250^{ns}	0.000	vit	106	5	0.623	0.630	0.012^{ns}	0.014	sal	17	6	0.824	0.745	-0.109^{ns}	0.000
J	car	89	12	0.348	0.859	0.596^{***}	0.272	des	77	9	0.792	0.697	-0.138^{ns}	0.000	tur	33	2	0.030	0.088	0.660^{*}	0.105
	oce	6	2	0.500	0.409	-0.250^{ns}	0.000	vit	113	6	0.575	0.635	0.094^{**}	0.040	sal	11	4	0.364	0.688	0.484^{ns}	0.190
K	car	95	22	0.821	0.913	0.101^{*}	0.049	des	75	22	0.907	0.893	-0.015^{ns}	0.000	tur	10	6	0.600	0.832	0.290^{ns}	0.133
	oce	6	2	0.167	0.167	0.000^{nc}	0.000	vit	112	18	0.848	0.911	0.069^{**}	0.019	sal	17	12	0.882	0.895	0.014^{ns}	0.000
L	car	95	9	0.589	0.675	0.127^{*}	0.061	des	72	14	0.653	0.814	0.199^{**}	0.079	tur	32	7	0.375	0.609	0.388^{***}	0.151
	oce	1	1	0.000	0.000	–	0.001	vit	114	13	0.500	0.640	0.219^{***}	0.079	sal	11	5	0.636	0.775	0.186^{ns}	0.050
M	car	85	2	0.059	0.057	-0.024^{ns}	0.000	des	78	9	0.679	0.826	0.179^{***}	0.062	tur	29	5	0.655	0.662	0.011^{ns}	0.031
	oce	4	4	0.250	0.821	0.727^{*}	0.278	vit	116	7	0.431	0.674	0.362^{***}	0.152	sal	15	7	0.533	0.766	0.311^{ns}	0.098
N	car	95	9	0.568	0.842	0.326^{***}	0.151	des	78	9	0.679	0.825	0.178^{*}	0.075	tur	35	5	0.429	0.737	0.422^{***}	0.170
	oce	5	4	0.600	0.644	0.077^{ns}	0.000	vit	117	10	0.752	0.802	0.063^{ns}	0.012	sal	14	6	0.571	0.815	0.307^{ns}	0.119

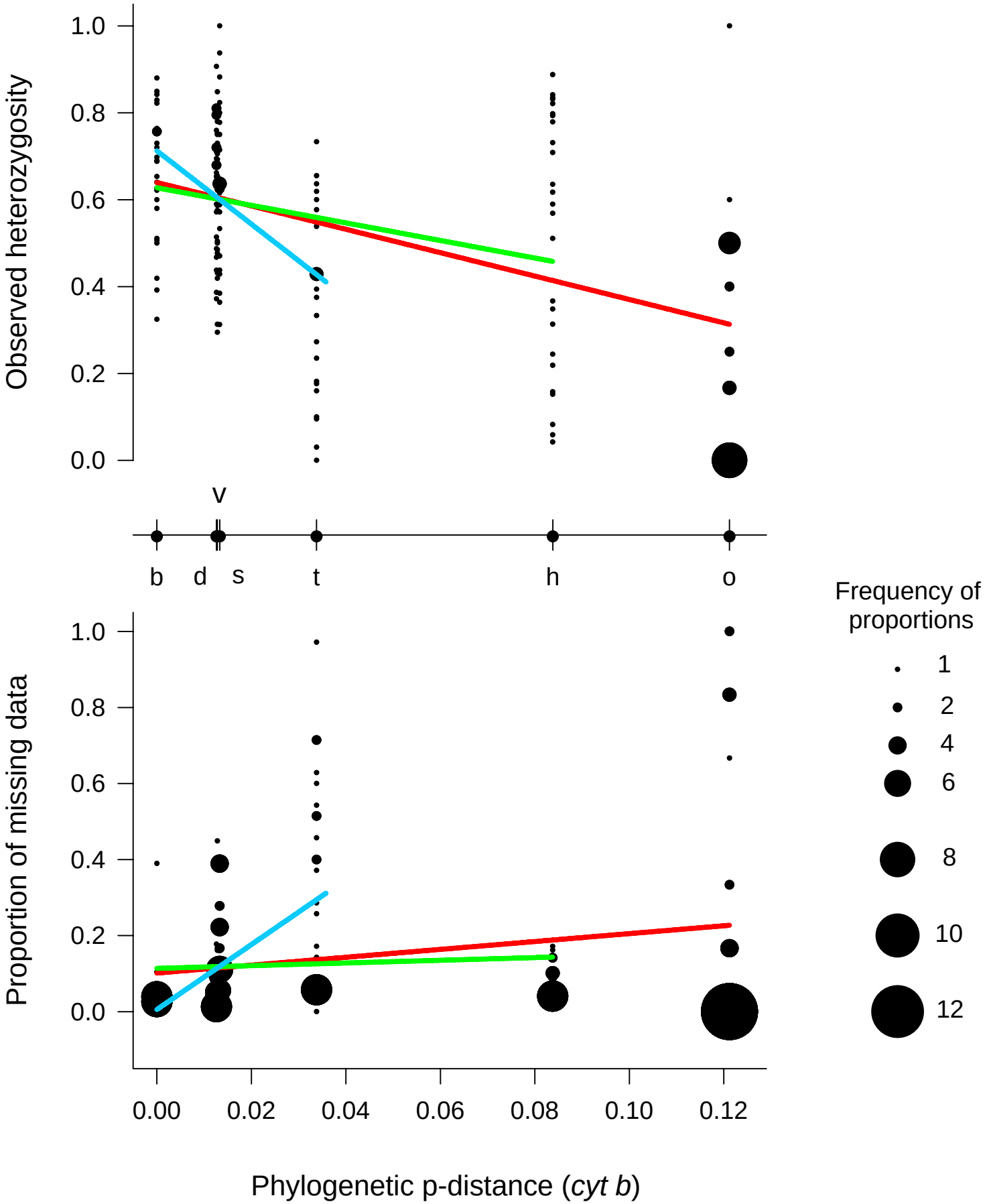
O	car	92	36	0.793	0.946	0.162**	0.068	des	78	17	0.487	0.866	0.439***	0.202	tur	10	4	0.100	0.363	0.735**	0.192
	oce	4	5	0.250	0.893	0.750*	0.300	vit	112	21	0.295	0.868	0.662***	0.306	sal	16	11	0.438	0.855	0.496***	0.209
S	car	90	7	0.367	0.729	0.499***	0.207	des	65	11	0.662	0.838	0.212*	0.088	tur	19	5	0.421	0.616	0.322*	0.149
	oce	5	1	0.000	0.000	–	0.001	vit	104	12	0.587	0.839	0.302***	0.133	sal	14	8	0.714	0.796	0.107ns	0.005
T	car	96	7	0.708	0.730	0.030ns	0.020	des	79	9	0.759	0.774	0.019ns	0.005	tur	13	3	0.538	0.495	-0.091ns	0.000
	oce	6	1	0.000	0.000	–	0.001	vit	114	6	0.640	0.676	0.053ns	0.012	sal	16	6	0.750	0.776	0.035ns	0.000
Q	car	89	14	0.798	0.877	0.091ns	0.040	des	76	12	0.711	0.756	0.061ns	0.000	tur	22	5	0.545	0.636	0.146ns	0.053
	oce	1	1	0.000	0.000	–	0.001	vit	113	10	0.726	0.737	0.015ns	0.005	sal	17	8	0.647	0.775	0.170ns	0.042
R	car	96	5	0.219	0.240	0.091ns	0.018	des	78	9	0.372	0.394	0.056ns	0.004	tur	21	2	0.095	0.093	-0.026ns	0.000
	oce	6	1	0.000	0.000	–	0.001	vit	99	6	0.485	0.514	0.057**	0.046	sal	16	6	0.313	0.512	0.398**	0.126
V	car	95	3	0.158	0.148	-0.070ns	0.000	des	75	12	0.720	0.720	0.000ns	0.000	tur	33	10	0.636	0.734	0.135ns	0.053
	oce	6	1	0.000	0.000	–	0.001	vit	109	10	0.780	0.747	-0.045ns	0.000	sal	14	7	0.643	0.794	0.196ns	0.087
W	car	89	14	0.888	0.852	-0.042ns	0.000	des	78	12	0.795	0.827	0.040ns	0.026	tur	30	10	0.733	0.879	0.168***	0.082
	oce	6	3	0.500	0.439	-0.154ns	0.000	vit	117	11	0.795	0.850	0.065*	0.026	sal	18	8	1.000	0.835	-0.205ns	0.000
X	car	94	10	0.617	0.746	0.173***	0.082	des	79	13	0.810	0.866	0.064ns	0.027	tur	33	8	0.394	0.505	0.222***	0.103
	oce	6	4	0.500	0.455	-0.111ns	0.000	vit	112	12	0.705	0.804	0.123*	0.050	sal	16	8	0.938	0.857	-0.098ns	0.000

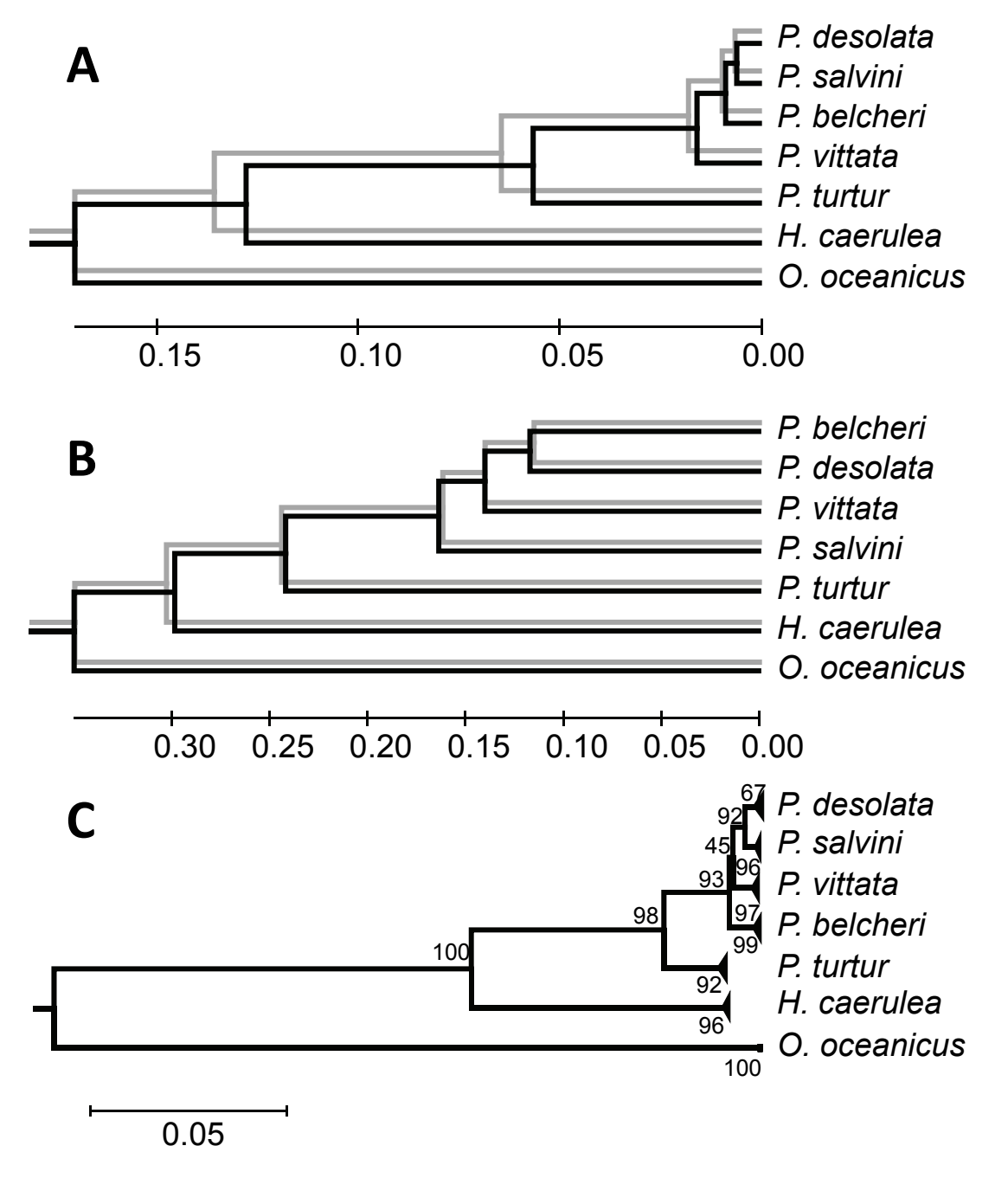
U	car	97	4	0.082	0.080	-0.025^{ns}	0.000	des	77	6	0.571	0.500	-0.143^{ns}	0.000	tur	33	3	0.182	0.224	0.190^{ns}	0.063
	oce	6	1	0.000	0.000	–	0.001	vit	117	4	0.419	0.469	0.107^{ns}	0.021	sal	16	5	0.625	0.558	-0.124^{ns}	0.000
Y	car	94	5	0.511	0.528	0.033^{ns}	0.006	des	72	5	0.514	0.520	0.013^{ns}	0.000	tur	17	3	0.176	0.266	0.343^{ns}	0.093
	oce	0	0	–	–	–	–	vit	104	7	0.692	0.650	-0.066^{ns}	0.000	sal	11	3	0.636	0.636	0.000^{ns}	0.040
P	car	90	10	0.244	0.311	0.214**	0.081	des	75	15	0.387	0.854	0.549***	0.250	tur	25	10	0.160	0.839	0.813***	0.364
	oce	6	3	0.500	0.439	-0.154^{ns}	0.000	vit	117	14	0.504	0.765	0.341***	0.152	sal	14	8	0.429	0.865	0.514**	0.228

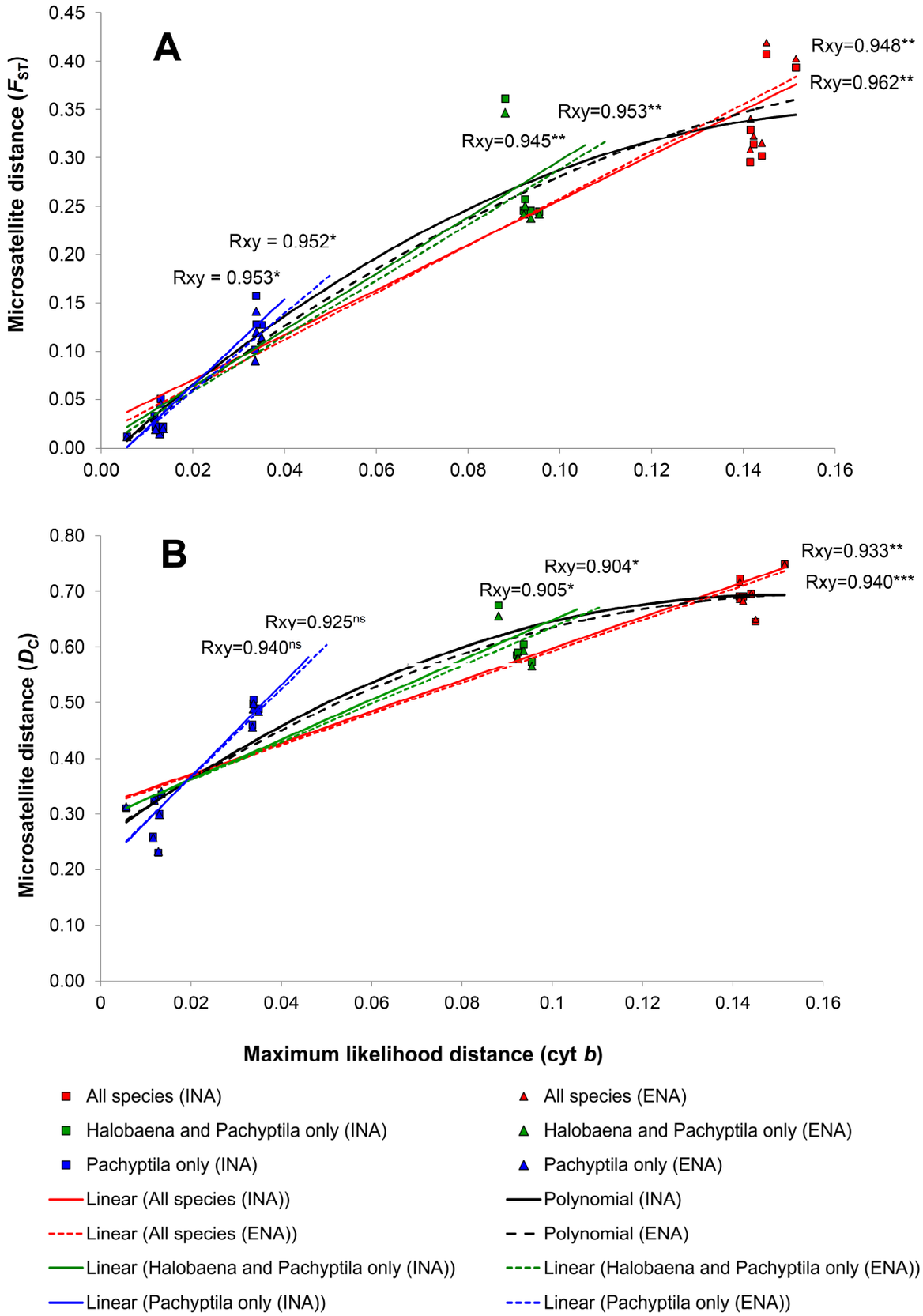
N: number of individuals with reliable amplification. *A*: number of alleles. H_o: observed heterozygosity. H_e: expected heterozygosity. F_{IS}: inbreeding coefficient. The probabilities of deviation from Hardy–Weinberg equilibrium (HWE) are indicated by asterisks (* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, ns: not significant, nc: not calculated). F (null): null allele frequency estimate.

Table 3 Fitting linear and polynomial functions to variance in microsatellite distance statistics across seven sub-Antarctic seabird species.

Statistic	Best fit linear function	AIC linear	Best fit polynomial function	AIC poly.	delta AIC	*P*
F_{ST} INA	y=2.325x+0.024	-67.2	y=-12.599x^2+4.285x-0.015	-72.5	5.3	0.013
F_{ST} ENA	y=2.437x+0.014	-72.5	y=-9.259x^2+3.877x-0.014	-75.3	2.8	0.045
D_C INA	y=2.826x+0.314	-53.4	y=-19.726x^2+5.894x+0.253	-61.3	7.9	0.004
D_C ENA	y=2.802x+0.312	-56.3	y =-17.195x^2+5.477x+0.259	-62.5	6.3	0.009


Observed heterozygosity
1.0
0.8
0.6
0.4
0.2
0.0
v
b
d
s
t
h
o
Proportion of missing data
1.0
0.8
0.6
0.4
0.2
0.0
0.00
0.02
0.04
0.06
0.08
0.10
0.12
Phylogenetic p-distance (cyt b)
Frequency of proportions
1
2
4
6
8
10
12


A
P. desolata
P. salvini
P. belcheri
P. vittata
P. turtur
H. caerulea
O. oceanicus
0.15
0.10
0.05
0.00
B
P. belcheri
P. desolata
P. vittata
P. salvini
P. turtur
H. caerulea
O. oceanicus
0.30
0.25
0.20
0.15
0.10
0.05
0.00
C
67
92
45
96
93
P. desolata
P. salvini
P. vittata
98
97
P. belcheri
100
99
P. turtur
92
H. caerulea
96
O. oceanicus
100
0.05


A
Microsatellite distance (F_{ST})
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.14
0.16
Rxy=0.948**
Rxy=0.962**
Rxy=0.953**
Rxy=0.945**
Rxy = 0.952*
Rxy = 0.953*
B
Microsatellite distance (D_C)
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
0
0.02
0.04
0.06
0.08
0.1
0.12
0.14
0.16
Rxy=0.933**
Rxy=0.940***
Rxy=0.904*
Rxy=0.905*
Rxy=0.925ns
Rxy=0.940ns
Maximum likelihood distance (cyt b)
All species (INA)
All species (ENA)
Halobaena and Pachyptila only (INA)
Halobaena and Pachyptila only (ENA)
Pachyptila only (INA)
Pachyptila only (ENA)
Linear (All species (INA))
Polynomial (INA)
Linear (All species (ENA))
Polynomial (ENA)
Linear (Halobaena and Pachyptila only (INA))
Linear (Halobaena and Pachyptila only (ENA))
Linear (Pachyptila only (INA))
Linear (Pachyptila only (ENA))