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

YOSHAN MOODLEY,* † JUAN F. MASELLO,‡ THERESA L. COLE,‡ † LUCIANO CALDERON,‡ GOPI K. MUNIMANDA,† MARCO R. THALI,¶ RACHAEL ALDERMAN,** RICHARD J. CUTHBERT,†† MANUEL MARIN,‡ †† MELANIE MASSARO,¶¶ JOAN NAVARRO,*** RICHARD A. PHILLIPS,††† PETER G. RYAN,§§§ CRISTIÁN G. SUAZO,‡ YVES CHEREL,§§§ HENRI WEIMERSKIRCH§§§ and PETRA QUILLFELDT‡

*Department of Zoology, University of Venda, Private Bag X5050, Thohoyandou 0950, South Africa, †Department of Integrative Biology and Evolution, Konrad Lorenz Institute for Ethology, University of Veterinary Medicine Vienna, Savoenstr. 1a, A-1160 Vienna, Austria, ‡Justus Liebig University Gießen, Department of Animal Ecology & Systematics, Heinrich-Buff-Ring 38, D-35392 Gießen, Germany, §Trace and Environmental DNA Laboratory, Department of Environment and Agriculture, Curtin University, Perth, WA 6102, Australia, ¶Ecogenics GmbH, Grabenstrasse 11a, 8952 Zurich-Schlieren, Switzerland, **Department of Primary Industries, Parks, Water and Environment, GPO Box 44, Hobart, Tas. 7001, Australia, ††Royal Society for the Protection of Birds (RSPB), The Lodge, Sandy, Bedfordshire SG19 2DL, UK, †††Section of Ornithology, Natural History Museum of Los Angeles County, 900 Exposition Boulevard, Los Angeles, CA 90007, USA, §§§Feather Link Inc., 1013 Westchester Way, Cincinnati, OH 45244, USA, §§School of Environmental Sciences, Charles Sturt University, PO Box 789, Albury, NSW 2640, Australia, ***Department of Conservation Biology, Estación Biológica de Doñana (EBD-CSIC), Avda. Américo Vespucio s/n, Sevilla 41092, Spain, †††British Antarctic Survey, Natural Environment Research Council, High Cross, Madingley Road, Cambridge CB3 0ET, UK, †††Percy FitzPatrick Institute, DST/NRF Centre of Excellence, University of Cape Town, Rondebosch 7701, South Africa, §§§Centre d’Études Biologiques de Chizé, UMR 7372 CNRS-Université de La Rochelle, 79360 Villiers-en-Bois, France

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 nonamplification, or disruptions that may result in decreased polymorphism in nontarget species. Furthermore, high mutation rates and constraints on allele size may also with evolutionary time, promote 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 nonamplifying 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 nonlinear 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.

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

Received 23 July 2014; revision received 12 January 2015; accepted 13 January 2015
Introduction

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

Petrels (Procellariidae) of the genus Pachyptila, the prions, are ideal monitors of ocean productivity as they feed mainly on zooplankton, which responds rapidly to changes in recent years (Ovenden et al. 2004; Cherel et al. 2002; Quillfeldt et al. 2007, 2008). Prions are highly mobile and have a wide distribution in sub-Antarctic waters (Onley & Scofield 2007).

Detailed studies on diet, breeding biology and behavioural ecology have been published for several prion species (Strange 1980; Bretagnolle et al. 1990; Liddle 1994; Ridoux 1994; Reid et al. 1999; Cherel et al. 2002; Quillfeldt et al. 2003, 2007, 2008; Navarro et al. 2013). Much less was known about distributions at sea, particularly during the nonbreeding season. However, recent stable isotope and tracking studies indicate considerable ecological segregation among populations breeding in the Atlantic and Indian sectors of the Southern Ocean (Cherel et al. 2002, 2006; Quillfeldt et al. 2010, 2013). This spatial and temporal segregation could potentially lead to population differentiation, with consequences for taxonomy and conservation status.

Prions are generally clustered into 2 groups, the species with, or without, palatal lamellae, and hence filtering apparatus (Prince & Morgan 1987). The former (the so-called ‘whale birds’) includes the Antarctic prion Pachyptila desolata, Salvin’s prion P. salvini and broad-billed prion P. vittata, and the latter, the thin-billed prion P. belcheri, fairy prion P. turtur and fulmar prion P. crassirostris. As yet, there are no phylogeographic studies of any prion species, and only scattered genetic information exists (e.g. Ovenden et al. 1991). Nor is there agreement regarding the number of prion species or their genetic relationships (see Brooke 2004; Penhallurik & Wink 2004; Rheindt & Austin 2005; Onley & Scofield 2007). Using enzyme electrophoresis, Barowcough et al. (1981) concluded that Antarctic prions were closely related to blue petrels Halobaena caerulea, and Viot et al. (1993) that Antarctic prions, thin-billed prions and Salvin’s prions were very closely related. The low variation at the mitochondrial cytochrome b gene also suggests that Antarctic prions, thin-billed prions and blue petrels are closely related species (Nunn & Stanley 1998). Based on the same locus, Penhallurik & Wink (2004) invoked the multidimensional biological species concept to suggest that all prions represent just two species. However, this last study was heavily criticized (Rheindt & Austin 2005). Consequently, taxonomic authorities and field guides still follow the scheme outlined by Bretagnolle et al. (1990) which concluded from the combination of morphometrics, breeding biology, genetics and calls that Antarctic, Salvin’s, thin-billed and fairy prions were distinct but closely related species. The recent description of distinct thin- and broad-billed morphs, within broad-billed prions that also show strong differences in breeding phenology at Gough Island, raises further questions regarding intra- and interspecific relationships of this group (Ryan et al. 2014). More genetic data from the many populations of prions that breed in the sub-Antarctic could potentially shed important light on these open questions.

Microsatellite loci are powerful tools in population and evolutionary genetics that could provide the resolution for detailed analyses of several aspects of prion biology. Given their high mutation rate, they are ideal for testing hypotheses relating to fine-scale spatiotemporal segregation and for the estimation of demographic parameters such as gene flow, effective population size and genetic variability (Brunford & Wayne 1993; Sunnucks 2000). They are also conserved among closely related species (Moore et al. 1991) making them potentially useful for clarifying interspecific genetic relationships between recently diverged taxa (e.g. Dawson et al. 2010), although their high mutation rate means that loci developed for one species may not always be useful for others. This is because mutations at primer binding sites may lead to increased nonamplification (Moodley et al. 2006), or disruptions within tandemly repeated elements may lead to a reduced level of observed polymorphism in the nontarget species (Garza et al. 1995; Primmer et al. 2005). The taxa of interest must therefore be closely related to maximize utility and offset the costs of microsatellite development. Hence, the testing of newly developed microsatellite loci for cross-species utility has become fairly standard practice (Li et al. 2003; Bried et al. 2008; Dawson et al. 2010; Jan et al. 2012; Huang et al. 2014); however, few surveys have reported statistical trends in their multispecies data sets.

Here, we used next-generation sequencing technology to develop a set of 26 polymorphic microsatellite markers from a shotgun genome sequence of the
thin-billed prion in order to test for genetic structure among the different populations of this species and to provide an accurate estimation of demographic parameters. We also tested these microsatellite loci for cross-species amplification in other prions, the closely related blue petrel (all Procellariiformes, Procellaridae) and the more distantly related Wilson’s storm-petrel Oceanites oceanicus (Procellariiformes, Hydrobatidae). While prion species appear to be very closely related, we predict, nevertheless, that average observed microsatellite genetic diversity will decrease, whereas the number of nonamplifying alleles will increase, with increasing phylogenetic distance from the species of origin (thin-billed prion). As any increase in null allele frequency could bias the estimation of essential population parameters in prion. As any increase in null allele frequency could bias the estimation of essential population parameters in prion. As any increase in null allele frequency could bias the estimation of essential population parameters in prion. As any increase in null allele frequency could bias the estimation of essential population parameters in prion. As any increase in null allele frequency could bias the estimation of essential population parameters in prion. As any increase in null allele frequency could bias the estimation of essential population parameters in prion. As any increase in null allele frequency could bias the estimation of essential population parameters in prion. As any increase in null allele frequency could bias the estimation of essential population parameters in prion.

Lastly, high microsatellite mutation rates coupled with mutational limits on allele sizes (Ostrander et al. 1993; Bowcock et al. 1994) increase the probability of convergent evolution of allele size classes. Given this highly homoplastic scenario, population parameters and evolutionary hypotheses inferred under the assumption that alleles of the same size share a most recent common ancestor (i.e. are identical by descent) could be biased, even in comparisons between sister taxa (Paetkau et al. 1997). We expect, therefore, that microsatellite genetic distance between species will be biased to lower values as the evolutionary time separating species increases.

Materials and methods

Molecular methods

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

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

After initial testing for amplification and polymorphism, microsatellite loci were visualised in the seven species of petrel through florescent labelling of universal M13 primers, as described in Schuelke (2000). PCR was then performed in a final volume of 10 μL including 1 × Qiagen PCR buffer, 2 mM dNTPs, 2 μM M13-tailed forward primer, 2 μM reverse primer, 2 μM of universal M13 primer 5′-end-labelled, 0.5 U Hotstar Taq (Qiagen) and 10 ng template. The PCR program comprised an initial denaturation step of 95°C for 15 min, the cycling parameters were 30 cycles at 95°C for 30 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, 72°C for 45 s and a final extension step of 72°C for 30 min. PCR products were visualized on a 1.5% agarose gel to confirm successful amplification and to examine negative controls. Products were run on a AB 3130xl genetic analyser along with a ROX size-standard. We repeated all PCRs for individuals that failed to amplify at >4 loci and for those loci with >10% missing data.

We used an 880-bp fragment of the mitochondrial cytochrome b gene to estimate the phylogenetic relationships among the seven taxa in our data set, as two of our hypotheses required an independent estimate of interspecific phylogenetic distance. While we do not believe that mtDNA distances are unbiased, we do feel that its slower mutation rate, relative to that of microsatellites, would ensure its linearity among the seven species on our comparison. Furthermore, cytochrome b has been
used previously to successfully infer relatedness among a much broader sample of Procellariiformes (Nunn & Stanley 1998), and previous morphological and behavioural analyses (Bretagnolle et al. 1990) did not specifically quantify interspecific distances among the species in our study. Generic avian cytochrome b primers (e.g. Patterson et al. 2011) were problematic for some samples. Therefore, we designed specific primers (CytB_Pri_F: 5’-CTAGCTATACACTACCCGC-3’ and CytB_Pri_R: 5’-CTAGTTGGCCGATGATGATG-3’) for our study group from an alignment of those samples that we successfully sequenced. PCRs were conducted in 20 µL reaction volumes containing 100 ng DNA template, 10 mM of each primer, 10 mM dNTPs (Roth, Karlsruhe), 2 mM MgCl₂, 5 U Thermus aquaticus polymerase (BioLabs Taq DNA polymerase) in a 1x PCR reaction buffer. Thermocycling included initial denaturation at 94°C for 2 min, 30 cycles of denaturation at 94°C for 30 s, annealing at 60°C for 45 s and extension at 72°C for 1 min, followed by a final extension step 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 ABI 3130xl 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 rescoring by a separate individual, with a resulting error rate of <5%. The probability of deviation from Hardy–Weinberg equilibrium (HWE) and nonrandom 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 (Hₒ) and expected heterozygosity (Hₑ)] were estimated in CERVUS 3.0.3 (Kalinowski et al. 2007) and MSA 4.05 (Dieringer & Schlötterer 2003). The inbreeding coefficient (Fᵢₒ) 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 (Hₒ) and the proportion of missing data (nonamplifying loci after 3x repeat PCRs, with standardized DNA quantity and quality) in each species against phylogenetic (mtDNA, cytochrome b) p-distances, calculated in MEGA (Table S1, Supporting Information). 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 phylogegetic 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 data set 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–maximization (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 pairwise Fₛₜ and Dₑ (genetic distance of Cavalli-Sforza & Edwards 1967) between species using INA and ENA data sets. We reconstructed UPGMA species trees from these triangular Fₛₜ and Dₑ 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 pairwise Mantel regressions performed on the same three phylogenetic data groupings used in Fig. 1. As interspecific microsatellite distance, we used INA and ENA triangular matrices of both Fₛₜ and Dₑ and checked their linearity against the matrix of pairwise maximum-likelihood cytochrome b distances calculated previously (Table S1, Supporting Information). All Mantel regressions were calculated in GenAlEx 6.5 (Peakall & Smouse 2012). To formally test the hypothesis that Fₛₜ and Dₑ microsatellite distances were nonlinearly related to mtDNA
Table 1 Locus characteristics and genetic variation at 25 newly isolated microsatellites in the target species, the thin-billed prion *Pachyptila belcheri*

<table>
<thead>
<tr>
<th>Locus code and name</th>
<th>Motif†</th>
<th>Primer sequences 5′–3′</th>
<th>Size (bp), range</th>
<th>N</th>
<th>A</th>
<th>H_O</th>
<th>H_E</th>
<th>F_IS</th>
<th>F (null)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A Pacbel_00386</td>
<td>(AC)_{14}</td>
<td>F: GCATGTCTCATAACAAACG&lt;br&gt;R: TCACTGAAACCAAGACTAGCC</td>
<td>120–142</td>
<td>72</td>
<td>11</td>
<td>0.764</td>
<td>0.811</td>
<td>0.058***</td>
<td>0.014</td>
</tr>
<tr>
<td>D Pacbel_02653</td>
<td>(AC)_{12}</td>
<td>F: AGCCATAGCTCTAACAGTTTC&lt;br&gt;R: TGGAGCAATTCCTAGGGG</td>
<td>132–170</td>
<td>77</td>
<td>12</td>
<td>0.325</td>
<td>0.639</td>
<td>0.494***</td>
<td>0.203</td>
</tr>
<tr>
<td>E Pacbel_03731</td>
<td>(AC)_{14}</td>
<td>F: TTAGGGACTGTCACACGAC&lt;br&gt;R: TAGCAGCTGGAAGACATCG</td>
<td>122–268</td>
<td>47</td>
<td>19</td>
<td>0.392</td>
<td>0.972</td>
<td>0.599***</td>
<td>0.293</td>
</tr>
<tr>
<td>F Pacbel_04240</td>
<td>(AC)_{14}</td>
<td>F: GCCATGTCCTGGGCAAGCG&lt;br&gt;R: GCATCTGCTGTGGGATGTC</td>
<td>166–254</td>
<td>47</td>
<td>19</td>
<td>0.511</td>
<td>0.815</td>
<td>0.376***</td>
<td>0.164</td>
</tr>
<tr>
<td>G Pacbel_04355</td>
<td>(AC)_{14}</td>
<td>F: TACCAGGGCACAATTCTGGG&lt;br&gt;R: GGGAAAAATACAGGAGATGCTTG</td>
<td>158–212</td>
<td>69</td>
<td>20</td>
<td>0.579</td>
<td>0.931</td>
<td>0.379***</td>
<td>0.181</td>
</tr>
<tr>
<td>H Pacbel_04991</td>
<td>(GT)_{14}</td>
<td>F: TGTCATGAGCTCAGTACAGT&lt;br&gt;R: TGCAGTCATTTCAGGTTTGG</td>
<td>86–106</td>
<td>74</td>
<td>11</td>
<td>0.757</td>
<td>0.877</td>
<td>0.138 ns</td>
<td>0.064</td>
</tr>
<tr>
<td>I Pacbel_05765</td>
<td>(TG)_{12}</td>
<td>F: TACTGGTTCACAATAATCTACTG&lt;br&gt;R: CCTAGCTTTCGACACAAAGGATG</td>
<td>156–172</td>
<td>73</td>
<td>9</td>
<td>0.849</td>
<td>0.808</td>
<td>0.052 ns</td>
<td>0.000</td>
</tr>
<tr>
<td>J Pacbel_08509</td>
<td>(GA)_{13}</td>
<td>F: GCAAGTTTAAAGGCAACAAAAAACCTCT&lt;br&gt;R: ACAAAGAAAACCATACCAATAC</td>
<td>120–144</td>
<td>74</td>
<td>8</td>
<td>0.689</td>
<td>0.690</td>
<td>0.002 ns</td>
<td>0.000</td>
</tr>
<tr>
<td>K Pacbel_08988</td>
<td>(CT)_{12}</td>
<td>F: CCGGAAAGAGAGCTTAAAAGGG&lt;br&gt;R: GCCAGCAAGATGCATCAG</td>
<td>184–202</td>
<td>74</td>
<td>10</td>
<td>0.757</td>
<td>0.692</td>
<td>0.094 ns</td>
<td>0.000</td>
</tr>
<tr>
<td>L Pacbel_09021</td>
<td>(GT)_{12}</td>
<td>F: AACTGTTTTGCTCACAACCCCC&lt;br&gt;R: ATGCCTTTCCCGCAGGATCC</td>
<td>146–170</td>
<td>75</td>
<td>9</td>
<td>0.600</td>
<td>0.809</td>
<td>0.260**</td>
<td>0.118</td>
</tr>
<tr>
<td>M Pacbel_09570</td>
<td>(GC)_{13}</td>
<td>F: GCTTTATTTAAGAGCAAAAACCC&lt;br&gt;R: ACAAGGAAACACCTTACAC</td>
<td>92–110</td>
<td>73</td>
<td>10</td>
<td>0.822</td>
<td>0.830</td>
<td>0.010 ns</td>
<td>0.026</td>
</tr>
<tr>
<td>N Pacbel_10033</td>
<td>(AC)_{12}</td>
<td>F: GCAAGGCGGTTCCTGT&lt;br&gt;R: GCCACCTCACACAACACAG</td>
<td>102–120</td>
<td>77</td>
<td>7</td>
<td>0.688</td>
<td>0.811</td>
<td>0.152 ns</td>
<td>0.058</td>
</tr>
<tr>
<td>O Pacbel_10895</td>
<td>(AC/AT)_{8}</td>
<td>F: AGCTTTCTGTCTTG&lt;br&gt;R: TGCTTTCTACAG</td>
<td>158–196</td>
<td>75</td>
<td>19</td>
<td>0.720</td>
<td>0.890</td>
<td>0.192**</td>
<td>0.092</td>
</tr>
<tr>
<td>P Pacbel_12344</td>
<td>(AC)_{12}</td>
<td>F: CCAACGCCTGCCC&lt;br&gt;R: GCCGTGCAGACGTGAATAG</td>
<td>92–116</td>
<td>74</td>
<td>11</td>
<td>0.419</td>
<td>0.805</td>
<td>0.482***</td>
<td>0.211</td>
</tr>
<tr>
<td>Q Pacbel_15327</td>
<td>(CA)_{13}</td>
<td>F: TTCTTGTAGCAGTAGG&lt;br&gt;R: ACCTCATGTGTAAAACCTGCC</td>
<td>146–162</td>
<td>75</td>
<td>9</td>
<td>0.627</td>
<td>0.674</td>
<td>0.071*</td>
<td>0.038</td>
</tr>
<tr>
<td>R Pacbel_16671</td>
<td>(GT)_{13}</td>
<td>F: TGAAGGTATGC&lt;br&gt;R: CTGAAGCATTAGCACCTGCC</td>
<td>126–134</td>
<td>75</td>
<td>5</td>
<td>0.650</td>
<td>0.598</td>
<td>0.071 ns</td>
<td>0.000</td>
</tr>
<tr>
<td>S Pacbel_16989</td>
<td>(CA)_{12}</td>
<td>F: TGCTTTTGGGACG&lt;br&gt;R: TCGTGACATCTCTTCTGGAG</td>
<td>100–120</td>
<td>75</td>
<td>10</td>
<td>0.653</td>
<td>0.670</td>
<td>0.025**</td>
<td>0.026</td>
</tr>
<tr>
<td>T Pacbel_17529</td>
<td>(AG)_{14}</td>
<td>F: TGCAAGTGCT&lt;br&gt;R: AATGCAATTTGCTTCCGGG</td>
<td>142–164</td>
<td>76</td>
<td>12</td>
<td>0.842</td>
<td>0.821</td>
<td>0.026 ns</td>
<td>0.000</td>
</tr>
<tr>
<td>U Pacbel_17944</td>
<td>(TG)_{13}</td>
<td>F: TACACCTGTTCT&lt;br&gt;R: GGAAGACGAGCAGCAATAC</td>
<td>228–254</td>
<td>75</td>
<td>12</td>
<td>0.880</td>
<td>0.828</td>
<td>0.063 ns</td>
<td>0.000</td>
</tr>
<tr>
<td>V Pacbel_19907</td>
<td>(GT)_{12}</td>
<td>F: TTCTCTTACG&lt;br&gt;R: CCATACCTGGGAGGAGGAC</td>
<td>166–184</td>
<td>74</td>
<td>8</td>
<td>0.622</td>
<td>0.642</td>
<td>0.033 ns</td>
<td>0.000</td>
</tr>
<tr>
<td>W Pacbel_20784</td>
<td>(AG)_{12}</td>
<td>F: GCACGTGCT&lt;br&gt;R: ATGGTAGTACAC&lt;br&gt;TA</td>
<td>122–158</td>
<td>76</td>
<td>50</td>
<td>0.500</td>
<td>0.833</td>
<td>0.401***</td>
<td>0.177</td>
</tr>
</tbody>
</table>

†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. deserta (des), broad-billed prions P. vittata (vit), fairy prions P. turtur (tur), Salvin’s prions P. salvinii (sal), blue petrels Halobaena caerulea (car) and Wilson’s storm-petrels Oceanites oceanicus (oce).
### 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 4 to 16, the observed heterozygosity (H<sub>O</sub>) from 0.325 to 0.880 and the expected heterozygosity (H<sub>E</sub>) 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 the primer pairs worked successfully in 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 Pachyptila belcheri (P<sub>0.001</sub>; Table S1 (Supporting Information), 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, d.f. = 1, P < 0.001, effect of locus: F = 3.046, d.f. = 24, P < 0.001; Group 2 Pachyptila and
Halobaena, effect of distance: \( F = 0.468, \) d.f. = 1, \( P < 0.001, \) effect of locus: \( F = 2.843, \) d.f. = 24, \( P < 0.001; \) Group 3 only Pachyptila, effect of distance: \( F = 1.207, \) d.f. = 1, \( P < 0.001, \) effect of locus: \( F = 2.106, \) d.f. = 24, \( P < 0.001). \) The proportion of missing data increased slightly (Group 1, \( f = 0.1 + 1.0 \times x, \) \( R^2 = 0.04, \) \( P = 0.006; \) Group 2, \( f = 0.1 + 0.4 \times x, \) \( R^2 = 0.004, \) \( P = 0.473; \) Group 3, \( f = 0.01 + 8.6 \times x, \) \( R^2 = 0.31, \) \( P < 0.001) \) with phylogenetic distance from thin-billed prion (Fig. 1), and a GLM confirmed this trend for individual loci (Group 1, effect of distance: \( F = 0.361, \) d.f. = 1, \( P = 0.003, \) effect of locus: \( F = 1.296, \) d.f. = 24, \( P = 0.134; \) Group 2, effect of distance: \( F = 0.017, \) d.f. = 1, \( P = 0.428, \) effect of locus: \( F = 0.805, \) d.f. = 24, \( P = 0.191; \) Group 3, effect of distance: \( F = 1.24, \) d.f. = 1, \( P < 0.001, \) effect of locus: \( F = 0.911, \) d.f. = 24, \( P < 0.001). \)

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

Mantel regressions (Fig. 3) indicated that a large proportion of DNA sequence variance in the cytochrome \( b \) data could be significantly explained by the multilocus
microsatellite distance statistics $F_{ST}$ and $D_{C}$ ($R_{xy} > 0.9$ in all cases, see Fig. 3). However, the relationship between microsatellite and mtDNA distance values changed markedly among the three groups of species tested and depended on which species were included. The slopes of the regressions including all species (Group 1) were lowest and increased incrementally as more phylogenetically similar taxa (Group 2, only Pachyptila and Halobaena species; 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 nonlinear 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**

*Nonamplification 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 nonamplifying (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 nonamplification 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.

*Utility in analysing interspecific relationships*

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

### Linearity of microsatellite genetic distances

The high proportion of explained variation in Mantel regressions of microsatellite and mtDNA genetic distance implies that variation was similarly distributed between the both microsatellite and mitochondrial data sets. However, when regression analyses of the three groups of

### Table 3

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Best fit linear function</th>
<th>AIC linear</th>
<th>Best fit polynomial function</th>
<th>AIC poly.</th>
<th>Delta AIC</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F_{ST}$ INA</td>
<td>$y = 2.325x + 0.024$</td>
<td>−67.2</td>
<td>$y = -12.599x^2 + 4.285x - 0.015$</td>
<td>−72.5</td>
<td>5.3</td>
<td>0.013</td>
</tr>
<tr>
<td>$F_{ST}$ ENA</td>
<td>$y = 2.437x + 0.014$</td>
<td>−72.5</td>
<td>$y = -9.259x^2 + 3.877x - 0.014$</td>
<td>−75.3</td>
<td>2.8</td>
<td>0.045</td>
</tr>
<tr>
<td>$D_{C}$ INA</td>
<td>$y = 2.826x + 0.314$</td>
<td>−53.4</td>
<td>$y = -19.726x^2 + 5.894x + 0.253$</td>
<td>−61.3</td>
<td>7.9</td>
<td>0.004</td>
</tr>
<tr>
<td>$D_{C}$ ENA</td>
<td>$y = 2.802x + 0.312$</td>
<td>−56.3</td>
<td>$y = -17.195x^2 + 5.477x + 0.259$</td>
<td>−62.5</td>
<td>6.3</td>
<td>0.009</td>
</tr>
</tbody>
</table>

© 2015 John Wiley & Sons Ltd
varying species diversity imposed a linear relationship between microsatellite and mtDNA distance, the slope of the regression changed considerably, suggesting that the true relationship was nonlinear. Instead, at lower distance values (e.g. among congeners; Group 3), the relationship appears linear, but microsatellite distance gradually reaches a plateau with increasing mtDNA distance (Fig. 3A,B, Groups 1 and 2), implying that the latter statistic is a better estimator of relationships among distantly related taxa. We tested the hypothesis that microsatellite distances are not linear with evolutionary distance by fitting both linear and polynomial functions to the $F_{ST}$ and $D_C$ data and found that in both cases a quadratic 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 nonlinearity 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 microsatellites in phylogenetic reconstruction beyond the genus level, as even at that level this may be associated with high variance.

Acknowledgements

The work was funded by a grant provided by the German Science Foundation DFG (Qu 148/5). We would like to thank the New Island Conservation Trust with assistance from Ian, Maria and Georgina Strange, and Benno H. Lüthi, Klemens Pütz and Gerhard Meyer from the Antarctic Research Trust, for crucial support during the fieldwork. We extend special thanks to Ruth Brown for her fieldwork support at Bird Island (South Georgia) and to Jaime A. Cursach for his assistance during fieldwork at Diego Ramirez (Chile). We thank Antje Schreiner for help with the laboratory work.

References


© 2015 John Wiley & Sons Ltd
Cross-species utility of microsatellites for seabirds


**Data accessibility**


**Supporting Information**

Additional Supporting Information may be found in the online version of this article:

Table S1 Phylogenetic p-distances and maximum likelihood distances (cytochrome *b*) in *Pachyptila* species, blue petrels and Wilson’s storm-petrels.