Ancient DNA Enables Timing of the Pleistocene Origin and Holocene Expansion of Two Adélie Penguin Lineages in Antarctica

Peter A. Ritchie,*† Craig D. Millar,*‡ Gillian C. Gibb,*† Carlo Baroni,§ and David M. Lambert*†

*Allan Wilson Centre for Molecular Ecology and Evolution, Auckland; †Institute of Molecular BioSciences, Massey University, Auckland; ‡School of Biological Sciences, University of Auckland, Auckland, New Zealand; and §Dipartimento Scienze della Terra, Università di Pisa, and Consiglio Nazionale Ricerche, Centro Studio Geologia Strutturale, Pisa, Italy

The timing of divergent events in history is one of the central goals of contemporary evolutionary biology. Such studies are however dependent on accurate evolutionary rates. Recent developments in ancient DNA analysis enable the estimation of more accurate evolutionary rates and therefore more accurate timing of divergence events. Consequently, this leads to a better understanding of changes in populations through time. We use an evolutionary rate calculated from ancient DNA of Adélie penguins (Pygoscelis adeliae) to time divergent events in their history. We report the presence of two distinct and highly variable mitochondrial DNA lineages and track changes in these lineages through space and time. When the ancient DNA and the phylogenetic rates are used to estimate the time of origin of the lineages, two very different estimates resulted. In addition, these same rates provide very different estimates of the time of expansion of these lineages. We suggest that the rate calculated from ancient DNA is more consistent with the glacial history of Antarctica and requires fewer assumptions than does a narrative based on the phylogenetic rate. Finally, we suggest that our study indicates an important new role for ancient DNA studies in the timing of divergent events in history.

Introduction

Evolutionary biology, over the past 100 years, has focused on first determining phylogenetic relationships among groups of organisms and, second, representing these relationships against time scales. The latter requires, however, accurate estimates of rates of evolution. These can then be used to better understand changing patterns of genetic variation. More precise and higher resolution descriptions of patterns of genetic variation in populations through space and time are central activities in contemporary evolutionary biology. Accurate descriptions of such patterns can be informative of past vicariant events that have had an impact on the evolution of diverse biota. Such vicariant events are typically associated with periods of global environmental change. The Pleistocene epoch (2 Myr BP to 10 kyr BP) was, for example, characterized by a series of global ice ages that were remarkably synchronous (Williams 1989). Over the past 900 kyr, each glacial cycle lasted about 100 kyr and was associated with eccentricity changes in the Earth’s orbit (Imbrie et al. 1992). During these ice ages, many populations and species are thought to have split into refugia (Roberts 1998). Each ice age was followed by relatively short deglacial periods, the most recent of which occurred during the Holocene (<10 kyr BP). The latter warm period was characterized by the expansion of many populations from ice age refugia (Roberts 1998).

In Antarctica, the Late Pleistocene was distinguished by the repeated expansion and collapse of huge marine-based ice shelves, as well as by fluctuations in ice volume on the Antarctic landmass. These changes would have certainly caused large-scale disruptions to animal populations, as well as to their habitats. For example, Adélie penguins (Pygoscelis adeliae) breed in large colonies on ice-free areas close to the sea (Ainley, LaReseche, and Sladen 1983). Ross Island is presently the home to several hundred thousand breeding pairs of birds and is situated at the edge of the Ross Ice Shelf. In Antarctica, even small fluctuations in the volume of ice can have a significant impact on the number and extent of ice-free areas (Colhoun et al. 1992). At the last glacial maximum, the entire Ross Sea coastline (fig. 1) was uninhabitable to Adélie penguins because of the extent of the Ross Ice Shelf, and Ross Island itself was almost 900 km from the sea (Denton et al. 1989). Hence, these areas were uninhabitable to Adélie penguins, and such glacial events have undoubtedly influenced penguin abundance, distribution, and genetic diversity.

Because of the presence of extensive ancient DNA deposits in the undisturbed Antarctic environment, Adélie penguins have enabled the direct estimation of a rate of evolution of the control region of the mitochondrial genome (Lambert et al. 2002). Consequently, this species provides an ideal model to detect any vicariant events precipitated by major global environmental changes; for example, the Pleistocene cooling and later Holocene warming of Antarctica (Roberts 1998). The timing of vicariant events that have structured populations or resulted in speciation has generally been inferred using a “molecular clock” (Stoneking et al. 1992; Rand 1994; Klicka and Zink 1997; Held 2001). To estimate the divergence times between lineages, virtually all authors use phylogenetic rates adopted from other species or populations. Molecular clocks are typically calibrated in absolute time by comparison of sequences from species whose time of divergence is established from either known-age fossils (e.g., Rand 1994), vicariant events (e.g., Held 2001), or archaeological evidence (e.g., Stoneking et al. 1992). However, in some cases, rates estimated in this way result in divergence times that are grossly inconsistent with the fossil record when applied in other lineages or to other taxa (Heckman et al. 2001).

This is not surprising, because a range of errors can affect phylogenetic methods for estimating evolutionary rates. First, inadequate sampling of taxa and short
sequences can result in an inaccurate tree topology. Second, a series of potential errors involves the use of the fossil record. These potential errors include misdiagnosis of fossil taxa, uncertainty in dating fossils, inaccurate positioning of relevant fossils on a phylogenetic tree, improper dating of the stem lineage of an extant group, and the minimum age estimates that fossils provide (Soltis et al. 2002). Moreover, other factors that can bias estimates of rate may include a history of mitochondrial introgression, nonneutrality, or rate heterogeneity among lineages (Mindell and Thacker 1996). This diversity of potential errors, in any particular case, is likely to result in either an overestimate or underestimate of the rate of evolution.

The first estimate for a rate of evolution for an avian mitochondrial genome was from *Anser* and *Branta* geese (Shields and Wilson 1987). Using restriction fragment length polymorphisms (RFLPs) of mitochondrial genomes, Shields and Wilson recorded 9% sequence difference between *Anser* and *Branta* geese. Knowing that the oldest fossils from each genus were dated at 4 to 5 Myr, Shields and Wilson (1987) proposed a mean rate of divergence of 2%/Myr (9/4.5 = 2) for the entire mitochondrial genome. Quinn (1992) later estimated that the *Branta* mitochondrial DNA (mtDNA) control region (CR) evolved at a rate of 20.8%/Myr. This was based on the finding that the 5′ end of the CR of two *Branta* subspecies evolved 10.4 times faster than the entire mitochondrial genome (2%/Myr [Shields and Wilson 1987]). This relative difference was suggested because the CRs of these two subspecies had 13.5% sequence difference, accompanied by 1.3% RFLP difference between their entire mitochondrial genomes. This was the basis on which Quinn (1992) calculated a 10.4 times difference in rates between the entire molecule and the CR itself (13.5/1.3 = 10.4). Subsequently, a large number of authors have adopted Quinn’s relative rate estimate of 20.8%/Myr as a general rate for the CR of avian species (for example Baker and Marshall 1997).

Not withstanding all the above difficulties, until recently, the phylogenetic approach has represented the only method for estimating evolutionary rates and thereby timing divergent events. However, using a CR rate of evolution calculated from ancient DNA, we report here changes through space and time of two highly divergent mitochondrial lineages of Adélie penguins. Furthermore, we estimated the time to the most recent common ancestor (tMRCA) of mitochondrial lineages, using both the phylogenetic rate typically adopted for bird species (i.e., 20.8%/Myr) and the ancient DNA rate specific to Adélie penguins (i.e., 53%/Myr to 143%/Myr). These two rates resulted in very different times of origin of the lineages. We discuss the relative merits of these rates and suggest that the time of divergence of lineages, as estimated using ancient DNA data, is more consistent with the glacial history of Antarctica.

Materials and Methods
Sample Collection

Blood sampling of Adélie penguins from 16 populations (table 1) was conducted in Antarctica during the austral summers of 1996 to 2001. Subfossil Adélie
penguin bones were collected from 17 locations (table 2) along the coast of the Ross Sea, which included relic nesting sites in the vicinity of presently occupied rookeries, as well as from abandoned rookeries. We dug for subfossil bones using a stratigraphic method of excavation (Baroni and Orombelli 1994). Subfossil bone samples were collected and stored frozen, then returned to the laboratory where they were kept at −20°C. A total of 380 sequences from modern and 96 sequences from ancient Adélie penguin samples have been included in the present study from our previous work (Lambert et al. 2002). These sequences can be retrieved at the GenBank (accession numbers AF474792 to AF474997).

DNA Extraction and Sequencing

Modern DNA was extracted from 177 whole-blood samples following standard methods (Sambrook, Fritsch, and Maniatis 1989). The hypervariable region 1 (HVR-1) of the mtDNA control region was amplified using PCR and primers specific to Adélie penguins. For the modern DNA samples, we amplified a 640-bp fragment using the primers L-tRNAGlu (5′-CCCGCTTGCTYTCTCAGA GTC-3′) and H-A650 (5′-CTGACATAGGAACCAG AGGCGC-3′) (Ritchie and Lambert 2000; Roeder, Ritchie, and Lambert 2002). Samples were amplified using a Hybaid Omn-E thermal cycler at 94°C for 10 s, 50°C or 61°C for 10 s, and 72°C for 25 s for 30 cycles. All resulting PCR products were purified using High Pure PCR purification columns (Roche), sequenced using the ABI PRISM® BigDye™ Terminator Cycle Sequencing Kit (Applied Biosystems), and analyzed on a ABI 377A automated sequencer. DNA sequences were aligned using ClustalW (Thompson, Higgins, and Gibson 1994) and can be retrieved at the GenBank accession numbers AY525167–AY525326.

Summary Statistics

The rate of each type of nucleotide substitution relative to G→T was estimated under a general time-reversible model (GTR) with maximum likelihood (Rodríguez et al. 1990) using PAUP* version 4.0 (Swofford 2002). Haplotypic (h) and nucleotide (π ± SE) diversity, number of segregating sites (S), and nucleotide differences (k) were estimated in DNAsp version 3.5 (Rozas and Rozas 1999). To test for neutrality, we calculated Tajima’s D statistic (Tajima 1989), the difference between S and π, which are both estimates of the expected number of substitutions per site.

Phylogenetic Analyses

A phylogenetic analysis of all sequences using the neighbor-joining (NJ) method (Saitou and Nei 1987) was conducted using a HKY85 substitution model (Hasegawa, Kishino, and Yano 1985) in PAUP* version 4.0. The phylogenetic tree was rooted with two sequences from chinstrap penguins (P. antarctica) and one sequence from a gentoo penguin (P. papua). We assessed the level of support for the major features of the NJ tree in the data set by resampling with a bootstrap procedure of 1,000 replicates. Support for the monophyly of lineages was also assessed by simulating the uncertainty of the tree (τ) given the data set (X) in a Bayesian framework. The analysis began with the NJ tree, and the posterior probability densities were estimated with a Metropolis-coupled Markov chain Monte Carlo (MCMC) integration algorithm using MrBayes version 2.01 (Huelsenbeck and Ronquist 2001).

Time to a Most Recent Common Ancestor

The $t_{MRCA}$ was estimated for the modern HVR-1 sequences (352 bp) by simulating the uncertainty of the genealogy in a Bayesian framework, using the 95% highest probability density (HPD) (Box and Tiao 1992) intervals for $μ_e$ using a population model of constant growth. In addition, we also estimated divergence times given the phylogenetic rate of nucleotide change of 0.208 substitutions/site/Myr (s/s/Myr). In both cases, the posterior probability density of the $t_{MRCA}$ was estimated from

Table 1

<table>
<thead>
<tr>
<th>Population</th>
<th>n</th>
<th>h</th>
<th>π</th>
<th>k</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Welch Island</td>
<td>21</td>
<td>1.0000</td>
<td>0.0205</td>
<td>21</td>
<td>0</td>
</tr>
<tr>
<td>2 Gardner Island</td>
<td>21</td>
<td>1.0000</td>
<td>0.0186</td>
<td>21</td>
<td>0</td>
</tr>
<tr>
<td>3 Torgersen Island</td>
<td>16</td>
<td>0.9917</td>
<td>0.0205</td>
<td>16</td>
<td>0</td>
</tr>
<tr>
<td>4 Cape Bird</td>
<td>122</td>
<td>0.9969</td>
<td>0.0393</td>
<td>98</td>
<td>24</td>
</tr>
<tr>
<td>5 Cape Royds</td>
<td>38</td>
<td>0.9915</td>
<td>0.0312</td>
<td>21</td>
<td>8</td>
</tr>
<tr>
<td>6 Cape Crozier</td>
<td>29</td>
<td>0.9901</td>
<td>0.0465</td>
<td>34</td>
<td>4</td>
</tr>
<tr>
<td>7 Beaufort Island</td>
<td>22</td>
<td>0.9870</td>
<td>0.0381</td>
<td>18</td>
<td>4</td>
</tr>
<tr>
<td>8 Franklin Island</td>
<td>16</td>
<td>0.9917</td>
<td>0.0411</td>
<td>13</td>
<td>3</td>
</tr>
<tr>
<td>17 Inexpressible Island</td>
<td>27</td>
<td>0.9915</td>
<td>0.0360</td>
<td>23</td>
<td>4</td>
</tr>
<tr>
<td>18 Adélie Cove</td>
<td>23</td>
<td>1.0000</td>
<td>0.0421</td>
<td>18</td>
<td>5</td>
</tr>
<tr>
<td>21 Edmonson Point</td>
<td>23</td>
<td>1.0000</td>
<td>0.0438</td>
<td>20</td>
<td>3</td>
</tr>
<tr>
<td>22 Cape Wheatstone</td>
<td>10</td>
<td>1.0000</td>
<td>0.0534</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>23 Cape Hallett</td>
<td>29</td>
<td>0.9975</td>
<td>0.0513</td>
<td>10</td>
<td>19</td>
</tr>
<tr>
<td>24 Cape Adare</td>
<td>100</td>
<td>0.9401</td>
<td>0.0599</td>
<td>38</td>
<td>62</td>
</tr>
<tr>
<td>25 Port Martin</td>
<td>30</td>
<td>0.9662</td>
<td>0.0255</td>
<td>29</td>
<td>1</td>
</tr>
<tr>
<td>26 Balleny Islands</td>
<td>30</td>
<td>0.9577</td>
<td>0.0245</td>
<td>29</td>
<td>1</td>
</tr>
<tr>
<td>Pooled</td>
<td>557</td>
<td>0.9973</td>
<td>0.0380</td>
<td>437</td>
<td>120</td>
</tr>
</tbody>
</table>

NOTE.—Sample size (n), haplotypic diversity (h), nucleotide diversity (π), and the number of Antarctic (A) or Ross Sea (RS) lineages record at each site. The number assigned to each population corresponds to their position in figure 1.
a distribution of likely trees, which were calculated as a product of a likelihood function $\Pr(X | \tau, \mu, Q)$, where $Q$ is a GTR substitution model, the prior probabilities of the coalescent process ($\tau, N_e, \mu$, and $N_c$ (effective population size), and $Q$ (Drummond et al. 2002). The posterior probability densities were approximated using MCMC integration in MEPI version 1.0 (Drummond et al. 2002). The prior probabilities for $N_e$ were allowed to vary between $10^5$ and $10^5$ individuals. The estimated evolutionary rate for Adélie penguin HVR-1 sequences, and their clocklike behavior, has previously been established at 0.96 s/s/Myr (95% HPD interval 0.53 to 1.43) (Lambert et al. 2002). In the first analysis, $\mu_e$ was allowed to vary between 0.53 and 1.43, and for the second analysis, $\mu_e$ was fixed at a point estimate of 0.208 s/s/Myr. The latter rate has been estimated from the divergence of Anser and Branta species of geese (Shields and Wilson 1987; Quinn 1992) and is a generally accepted phylogenetic rate for the HVR-1 of avian species. The Markov chains were $10^5$ steps long, and the first $5 \times 10^5$ iterations were discarded as the burn-in time, until a steady state was reached.

Results

The HVR-1 sequences from the modern Adélie penguins were typically 594 bp long and were aligned to 96 ancient DNA sequences 352 bp long. These sequences are homologous to the CR sequences for which Quinn (1992) estimated an evolutionary rate of 20.8%/Myr. For all modern samples ($n = 557$), the homologous 352 bp region showed unequal base frequencies ($f_A = 0.31$, $f_T = 0.30$, $f_C = 0.19$, and $f_G = 0.20$ from the H-strand sequence) and revealed 440 haplotypes among 557 sequences ($h = 0.998$). The relative rates for each substitution type was biased towards transitions ($A \rightarrow C = 1.11$, $A \rightarrow G = 21.52$, $A \rightarrow T = 0.33$, $C \rightarrow G = 1.49$, and $C \rightarrow T = 15.99$, relative to $G \rightarrow T = 1.0$ from the heavy strand sequence) over 193 segregating sites, of which 47 were singleton changes, and $\pi = 0.043$ ($\pm$SE 0.002). Tajima’s $D$ statistic was $-1.712$, and there was no significant deviation from neutrality ($P > 0.05$).

Two Mitochondrial DNA Lineages

The most prominent feature of any phylogenetic analysis of HVR-1 sequence data for Adélie penguins is the presence of two monophyletic lineages of Adélie penguins (fig. 2). One lineage was recorded mainly in the Ross Sea and is designated the Ross Sea (RS) lineage, whereas the other was present at all locations around the Antarctic continent and was consequently designated the Antarctic (A) lineage (fig. 1A). There were, on average, 27.7 nucleotide differences between sequences from the $A$ and $RS$ lineages (the overall $d = 0.081 \pm SE 0.012$, and the net $d = 0.057$), significantly more than the average number of differences among sequences within each lineage ($k_A = 7.1$ and $k_RS = 9.8$). The two lineages shared 49 polymorphic sites, but there were no fixed differences between them. Sequences from the $A$ lineage had 98 polymorphic sites that were monomorphic among the $RS$ sequences, whereas sequences from the $RS$ lineage had 46 polymorphic sites that were all monomorphic among the $A$ lineage sequences. When the two outgroup species, the chinstrap and gentoo penguins, were compared with the Adélie penguins sequences, the proportion of nucleotide differences were 0.305 ($\pm$SE 0.022) and 0.294 ($\pm$SE 0.027), respectively. Furthermore, there was a 100 bp insertion/deletion event when the HVR-1 sequences from Adélie and gentoo penguins were compared with those from chinstraps.

The monophyly of each lineage was well supported. There was 100% bootstrap support for the split, and the Metropolis-coupled MCMC algorithm found that the monophyly of each lineage was 100% credible. A sequence from one lineage never clustered with a sequence from the other lineage in the full posterior distribution of likelihood trees. Plots of the log probability of the observed data through time showed that the Markov chain reached stability after about 8,000 iterations, and the first 50,000 iterations were discarded as the burn-in time (data not shown). To check for convergence, a second run was conducted that began with a random tree, and again this showed no topological variance in the monophyly of the $A$ and $RS$ lineages.

The Geographic Pattern of mtDNA Lineages

Populations in the Ross Sea have a higher nucleotide diversity ($\pi = 0.042 \pm SE 0.006$) than those sampled from other Antarctic locations ($\pi = 0.019 \pm SE 0.001$), because of the presence of both lineages in the Ross Sea and the absence of one lineage outside of that region. A feature of the lineage proportions is the decreasing frequency of the $RS$ lineage with increasing latitude. The geographic distribution of each lineage is presented in figure 1A. In figure 1A, the frequency cline has been divided into

Table 2
The Locations in Antarctica Where Subfossil Bones Were Sampled

<table>
<thead>
<tr>
<th>Location</th>
<th>n</th>
<th>Calibrated Years BP</th>
<th>A</th>
<th>RS</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 Cape Crozier</td>
<td>14</td>
<td>310–523</td>
<td>11</td>
<td>3</td>
</tr>
<tr>
<td>6 Cape Bird</td>
<td>10</td>
<td>440–481</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>7 Beaufort Island</td>
<td>3</td>
<td>275</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>9 Marble Point</td>
<td>2</td>
<td>4,580</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>10 Dunlop Island</td>
<td>6</td>
<td>2,662–5,997</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>11 Cape Roberts</td>
<td>1</td>
<td>3,091</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>12 Cape Ross</td>
<td>10</td>
<td>3,514–4,467</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>13 Depot Island</td>
<td>1</td>
<td>2,911</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>14 Cape Day</td>
<td>1</td>
<td>3,364</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>15 Cape Hickey</td>
<td>5</td>
<td>2,513–2,965</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>16 Prior Island</td>
<td>9</td>
<td>750–3,888</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>17 Inexpressible Island</td>
<td>18</td>
<td>2,328–6,082</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td>18 Adélie Cove</td>
<td>1</td>
<td>5,418</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>19 Northern Foothills</td>
<td>8</td>
<td>3,580–4,270</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>20 Gondwana Station</td>
<td>1</td>
<td>6,424</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>21 Edmonson Point</td>
<td>3</td>
<td>1,086–1,162</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>23 Cape Hallett</td>
<td>3</td>
<td>500–607</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Pooled</td>
<td>96</td>
<td>275–6,424</td>
<td>68</td>
<td>28</td>
</tr>
</tbody>
</table>

Note.—The sample sizes ($n$), the range of assigned 14C-age and the number of Antarctic (A) or Ross Sea (RS) lineages recorded at each site are also presented. Numbers assigned to each location refer to their position in figure 1.

The locations in Antarctica where subfossil bones were sampled.
ancient DNA has previously been recovered, and HVR-1 sequences were determined from 96 subfossil bone samples (Lambert et al. 2002). The geographic distribution of the ages for each subfossil bone and its respective lineage are presented in figure 1B. Younger bones tended to be recovered from more southern locations, and the largest numbers of old bones were concentrated in the Terra Nova Bay area. DNA was recovered from 15 samples dated to between 5,706 and 6,082 years BP from Dunlop and Inexpressible islands and revealed similar numbers of the A (n = 8) and RS (n = 7) lineages. There was a significant difference between the relative frequencies of the two lineages when the pooled modern Ross Island and Terra Nova populations were compared with the relic Dunlop and Inexpressible islands populations ($\chi^2 = 4.634, P = 0.037$). There was no significant difference between the relative frequencies of the two lineages recovered from the less than 2,500 years and the 2,500 to 5,000 years BP time periods when compared with the modern populations in the same region ($P > 0.05$). The average divergence between the A and RS lineages was estimated for sequences that were older than 2,000 years (mean age = 4,200 years BP) at 0.074 (±SE 0.010). This result is slightly lower than the same comparison among the modern samples and may indicate a detectable amount of evolutionary change between the two lineages through time.

The Posterior Distribution of Times to a Common Ancestor

The estimated number of independent samples in the Markov chain, calculated by dividing the chain length by the integrated autocorrelation time, was 744 when the rate varied between 0.53 and 1.43 s/s/Myr (assuming a constant population) and 972 when the rate was fixed at 0.205 s/s/Myr. There was a significant difference between the two posterior probability densities of $t_{MRCA}$ (figs. 3 and 4). Using the rate derived from ancient DNA sampling of Adélie penguins, the MCMC approach estimated the $t_{MRCA A-RS}$ at 75 kyr BP (95% HPD interval 37 to 122 kyr BP). In contrast, using the phylogenetic rate, the same approach estimated the $t_{MRCA A-RS}$ at 308 kyr BP (95% HPD interval 227 to 391 kyr BP). The median estimate of the $t_{MRCA A-A}$ using the rate derived from ancient DNA sampling of Adélie penguins was 30 kyr BP (95% HPD interval 16 to 52 kyr BP) and $t_{MRCA RS-RS}$ was 32 kyr BP (95% HPD interval 18 to 52 kyr BP). In contrast, the median estimates of the $t_{MRCA A-A}$ using the phylogenetic rate were 127 kyr BP (95% HPD interval 89 to 176 kyr BP) and the $t_{MRCA RS-RS}$ were 143 kyr BP (95% HPD interval 121 to 186 kyr BP). These estimated divergence times differ from those previously presented in (Lambert et al. 2002) because the latter represented only estimates based on our best estimate of the rate of evolution and the sequence divergence between lineages.

Discussion

The Origin of Two Lineages

The present study has revealed the existence of two distinct and genetically diverse mtDNA lineages of Adélie
penguins. We report that these groups of haplotypes are unevenly distributed through time and space around Antarctica. The fact that the Antarctic (A) lineage is distributed around the continent, whereas the Ross Sea (RS) lineage is restricted to the Ross Sea, is consistent with two ancestral refugia that were subject to population bottlenecks. The evidence presented of a cline in A/RS frequencies in the Ross Sea suggests subsequent expansion and secondary admixture. An evolutionary rate estimated for HVR-1 using Adélie penguin ancient DNA sequences of 0.96 s/s/Myr (0.53 to 1.43 s/s/Myr) suggests that the time to a most recent common ancestor of these two lineages \((t_{\text{MRCA A-RS}})\) is 75 kyr BP (37 to 122 kyr BP). The latter estimate is for a constant population. Simulations for an exponentially growing population resulted in similar estimates (data not shown). This \(t_{\text{MRCA A-RS}}\) is the best estimate of the age of the sequence that gave rise to all modern haplotypes of both lineages. Given that there was almost certainly a level of existing polymorphisms in the population at the time of separation of the lineages, this estimate represents the maximum age of any period of lineage sorting. Lineage sorting results from a population bottleneck in which a diverse set of haplotypes is reduced in number. A number of extreme models, consistent with these data, can be envisaged. First, the period of sorting might have begun after the onset of the last glacial cycle. An ancestral population of mtDNA haplotypes would have been dramatically reduced so that only the two ancestors of the A and RS lineages would have survived. Over the last glacial cycle, these two lineages would have accumulated differences as Adélie penguins expanded in population size and haplotype diversity, starting around the onset of the deglacial period. Second, lineage sorting may have occurred late in the recent glacial cycle, when ice conditions were at their most severe and not long before the subsequent expansion associated with the deglacial warming. In the latter scenario, there would be a significant temporal separation between the \(t_{\text{MRCA}}\) and the sorting event itself.

Using the ancient DNA rate of evolution calculated for Adélie penguins, the estimated time to a most recent common ancestor of both lineages \((t_{\text{MRCA A-RS}})\) is consistent with our understanding of the glacial history of Antarctica. The Late Pleistocene climate in Antarctica has been well documented; for example, temperature records are available from ice cores taken from Vostok on the East Antarctic Ice Sheet (fig. 3). These records date back to 420 kyr BP (Petit et al. 1999), and provide a framework to test hypotheses about the relationship between environmental change and the history of Adélie penguins. A recent date of divergence of the lineages (approximately 75 kyr BP) coincides with the middle of the last glacial cycle. At that time, Adélie penguins may have had limited opportunities to breed on the Antarctic mainland, and our results suggest they were separated into refugia. Certainly
at the time of the last glacial maximum, ice-free areas would have been rare, if present at all in Antarctica (Colhoun et al. 1992). The isolation of the Antarctic continent from other major landmasses would have limited the ability of Adélie penguins to progressively move north as ice-free areas were reduced and perhaps totally eliminated. Moreover, the expansion of the sea-ice would have meant that breeding grounds were far removed from the open sea, even if ice-free areas were present on the continent. Regular visiting of such feeding grounds— a necessity during the breeding season when adults are raising chicks—would be energetically expensive, if even physically possible. Such sea ice covering would have been extensive. Benthic sedimentation of $^{13}$C-rich organic material from sea-ice algae shows that at the last glacial maximum the summer sea ice may have extended as far as the winter sea ice now reaches (fig. 1A) (Cooke and Hays 1982; Gersonde and Zielinski 2000). The latter extends from hundreds to thousands of kilometers from the existing coastline and specifically the Ross Sea would have been completely covered at that time. These large-scale disruptions to gene flow between ice-age Adélie penguin refugia are likely to have persisted as recently as 17 to 20 kyr BP (Baroni and Orombelli 1994).

In contrast, the phylogenetic rate (0.208 s/s/Myr) suggests that the $t_{MRCA_{A-RS}}$ was 308 kyr BP and that, consequently, the $A$ and $RS$ lineages persisted through two or three glacial cycles (fig. 3). According to this scenario, Adélie penguin population sizes would have remained large enough over the last two or three cycles for both lineages to survive. Since each glacial cycle is likely to have resulted in some form of population bottleneck— given that there would have been a loss of ice-free nesting areas— the above scenario would seem less likely than the explanation based on the higher evolutionary rate calculated directly from Adélie penguin ancient DNA. The latter suggested a divergence during the last glacial cycle and therefore requires the lineages to survive only a single ice-age event.

The Expansion of the Two Lineages

Using the ancient DNA rate, we estimated $t_{MRCA}$ of modern haplotypes of each of the $A$ and $RS$ lineages separately. This resulted in very similar estimates in the case of each lineage ($t_{MRCA_{A-A}} = 30$ kyr BP [16 to 52 kyr BP] and $t_{MRCA_{RS-RS}} = 32$ kyr BP [18 to 52 kyr BP]) (fig. 4). We suggest that these two lineages began to diversify some time after these estimates because there will inevitably be a time lag between the $t_{MRCA}$ and any dramatic increase in the number of haplotypes. In contrast, the phylogenetic rate estimates the $t_{MRCA}$ of all modern $A$ haplotypes to be 127 kyr BP (111 to 192 kyr BP) and all modern $RS$ haplotypes to be 143 kyr BP (121 to 186 kyr BP). This suggests that there was a dramatic increase in the number of haplotypes of both lineages and that the number of haplotypes accumulated over the last ice age. However, this period was characterized by a general contraction in the distribution rather than expansion of many species (Roberts 1998). As we have discussed above, in the case of Adélie penguins, this period was probably one that was associated with a reduction in numbers of Adélie penguins in Antarctica.

Ice Age Refugia

The restricted distribution of the $RS$ lineage suggests a refugium that was geographically close to, or adjacent to, the Ross Sea. Moreover, the high relative frequency of the $RS$ lineage at sites near the entrance to the Ross Sea, compared with lower frequencies nearer to the permanent ice shelf, suggests that after deglaciation, the $RS$ lineage moved down the coast following the retreat of the Ross Ice Shelf (fig. 1). In contrast, because of the Antarctic-wide distribution of the $A$ lineage, it is difficult to infer a possible location of any refugium. However, it is interesting to note that the $A$ lineage was well established in the Ross Sea by 6,082 years BP, and, in fact, at that date, it was in approximately equal proportions to the $RS$ lineage at Terra Nova Bay in the Ross Sea (fig. 1). This would suggest that both lineages quickly established themselves in the Ross Sea, and our ancient DNA results show them to be sympatric at many locations in past times. Our study also shows, however, that the proportions of each lineage, at individual sites, have changed during this Holocene expansion period.

Rate Heterogeneity and the Molecular Clock

The observation that the amount of sequence difference between any two taxa does not have a precise linear relationship to the time to their common ancestor has plagued the molecular clock hypothesis for many years (Gillespie 1991). Hence, a degree of heterogeneity in evolutionary rates is inevitable. A variety of mechanisms have been proposed to explain such rate heterogeneity among species or groups of species, including, for example, the metabolic rate hypothesis (Martin and Palumbi 1993), the body temperature hypothesis (Prager et al. 1974), and the generation time hypothesis (Kohne 1970). Moreover, the rate of evolution is not a deterministic process and can vary along a lineage, and nucleotide changes have been described using a Poisson model (Gillespie 1991). Rate heterogeneity along a species lineage can be caused by changes in the underlying mutation rate or the fixation of nearly neutral alleles as the effective population size fluctuates (Ohta 1987). Therefore, it has been recognized that a rate of evolution estimated using fossils calibration dates of millions of years might not be transferable to other species and to more recent (or older) time periods.

Most studies that use the molecular clock as a dating tool rely on the assumption that, first, there is a constant rate at which the clock “ticks” and that, second, there are available relatively precise estimates of the true rate of evolution for any particular species. The majority of studies are restricted to adopting a rate of evolution previously established from a related group of species. A rate of evolution is usually deemed a good approximation to the true rate if there have been two or more studies of independent taxa that have provided corroborating estimates. For example, Klucka and Zink (1997) used the level of mtDNA sequence divergence between 35 sister species
of oscine passerine and a clock rate of 2%/Myr to refute the conventional model that many North American bird populations speciated during the glacial cycles of the Late Pleistocene. Klicka and Zink (1997) suggested that a rate of 2%/Myr for mtDNA was a reasonable estimate because it was corroborated among a diverse array of studies, including, for example, Shields and Wilson (1987), Nunn et al. (1996), and Tarr and Fleischer (1993). However, our study of Adélie penguins has been the first to employ a rate that was directly estimated from our study species, independent of both the fossil record and a phylogeny of closely related species. Hence, we were able test how different our results would be from one where we adopted the conventional CR rate for birds. We recorded significantly different $\mu_{\text{MRCA}}$ using rate specific to Adélie penguins. Importantly, this had significant implications for biogeographic explanations of the origin of two distinct mtDNA lineages found in Adélie penguins.

Finally, ancient DNA has been widely used to investigate a variety of issues, from changing levels of genetic variation (e.g., Nielsen, Hansen, and Loeschcke 1997), to analyses of the phylogeny of extinct species (e.g., Höss et al. 1996), to investigations of ancient diets (Poinar et al. 1998). Our research on ancient DNA of Adélie penguins suggests a new role in evolutionary biology; namely the timing of divergent events and tracking changes in the proportion of lineages over space and time. This approach, when used to detect the time to the MRCA of the A and RS lineages, is more consistent with the known glacial history of Antarctica and requires fewer assumptions than does a scenario based on an adopted phylogenetic rate of mitochondrial DNA evolution. Hence, our results indicate that complex biogeographical narratives based on evolutionary rates adopted from other species could be seriously misleading. As an increasing number of ancient DNA studies are reported, with consequently better estimates of evolutionary rates, this will enable a more accurate understanding of rate heterogeneity and consequently the genetic changes that have accompanied major global environmental events, such as the ice ages in Antarctica.

Acknowledgments

This research was made possible by a grant from the Marsden Fund of New Zealand and Centres of Research Excellence Fund. We also acknowledge support from Massey University, the Italian Antarctic Research Programme, Antarctica New Zealand, and the U.S. Coast Guard. P.A.R. acknowledges a Massey University doctoral scholarship. For assistance we thank P. Barrett, L. Davis, A. Drummond, S. Eyton, K. Kerry, J. Macdonald, P. Metcalf, J. Robins, L. Shepherd, C. Vleck, and E. Young. This paper represents publication No. 97 from the Allan Wilson Centre for Molecular Ecology and Evolution.

Literature Cited


Axel Myer, Associate Editor

Accepted August 26, 2006