Mitochondrial DNA Diversity and Variability in the Adélie Penguin of Antarctica

Gillian Claire Gibb

A thesis presented in partial fulfilment of the requirements for a degree of Master of Science in Genetics at Massey University, Palmerston North, New Zealand

2003

Cape Bird, Antarctica 2002

Abstract

In Antarctica, there are two distinct lineages of Adélie penguin (*Pygoscelis adeliae*) characterised by 8.3% divergence in mitochondrial DNA hypervariable region I (mt DNA HVR I). These two lineages are known as the *Antarctic* and *Ross Sea* lineages (*A* and *RS* respectively). This study aims to characterise aspects of mutation and variation as seen in HVR I of the Adélie penguin, by sequencing the DNA of individuals from different locations around Antarctica.

The geographic distribution of the two lineages was examined in greater detail. A dramatic decrease in the RS lineage was discovered on the edge of the Ross Sea region of Antarctica. Because the two lineages have different geographic distributions, and are separated by 8.3% sequence divergence, this study also investigated the possibility that these two lineages were in fact cryptic species. Sequencing of mt DNA and microsatellite genotyping proved that individuals of the two lineages mate randomly and produce offspring.

Recently, a rate of evolution based on serially preserved DNA from Adélie penguins was estimated at 0.96 substitutions/site/Million years. (0.53-1.43 s/s/Myr). This rate is four to seven times higher than previous avian control region evolution rates estimated by phylogenetic methods, and is more akin to rates of mutation determined by pedigree studies in other species such as humans. In the light of this higher direct estimate of the rate of evolution in Adélie penguins, this study also begins to determine a rate of mutation in Adélie penguins based on pedigree analysis. No new mutations were found, however three cases of inherited single point heteroplasmy were detected. The inclusion of heteroplasmy in mutation rate calculation is also addressed.

One of the arguments as to why pedigree studies find a higher rate of mutation than phylogenetic studies is that pedigree studies preferentially find mutations at 'hot spots' in the DNA sequence. This study also seeks to characterise the distribution of variable sites in hypervariable region I in relation to the two mt DNA lineages, and

also to geographic location. While the exact sites of variation differ between the two lineages, it was seen that the regions where variation was high or low is very similar in both lineages. This could be due to underlying physical constraints on DNA sequence variation.

Looking towards future work in Adélie penguin mt DNA and an expansion of the studies undertaken here, the complete mitochondrial genome of the Adélie penguin was determined. This now provides the opportunity to estimate rates of change in the entire Adélie penguin mitochondrial genome, using ancient DNA from the extremely well preserved sub-fossil bones in Antarctica.

Acknowledgements

Firstly, and most importantly I wish to thank my supervisor, Professor Dave Lambert for the opportunity to work on this project, and to visit Antarctica. You see the bigger picture, and I appreciate being able to be part of your vision. I have enjoyed working on this project very much.

This work was made possible by a Marsden Fund of New Zealand grant to Dave Lambert. I would like to acknowledge Massey University for financial support from a Molecular Genetics Research Scholarship.

I am indebted to Pete Ritchie for all the work he has done previously on Adélie penguins. You have been very patient with me, and had so much advice every time I had a problem. This project would not exist without the work you had done before!

Cheers to Craig Millar for help collecting samples down in Antarctica, and advice while writing up. Thanks also to Greg Arnold and Tom Parsons for statistical advice.

To everyone in the ME lab, past and present, for making a wonderful working environment, and for helping me get through this with some form of sanity. Thanks to Pete Ritchie, Hillary Miller, Lara Shepherd, Jennie Hay, Jenn Anderson, Gwilym Haynes, Leon (sigh) Huynen, Olly Berry, Jo Chapman, Quannah Hudson, Amy Roeder and Niccy Aitken. You've all been great.

A big thank you to everyone in David Penny's lab for help and advice in whole genome sequencing. Especially of course to Trish McLenachan for her enthusiasm, excitement and so much advice.

To Craig, Kerry, BJ and Malcolm for good friendship and an amazing time on the ice. Go penguins!

Last but certainly not least, a huge thank you to all my family and friends for a life outside this thesis! Thanks for your support and understanding. I could not have done this without you.

Preface

Many people were involved in collecting and analysing the Adélie penguin blood samples used in this thesis. Their contributions are listed below.

Blood samples from Cape Adare, Balleny Islands and Port Martin, Antarctica were collected in the austral summer of 2000/2001 by David Lambert (Massey University), John MacDonald and Peter Metcalf (Auckland University).

Adult Adélie penguin blood samples were collected from Cape Bird, Antarctica by Craig Millar (Auckland University), Peter Ritchie, Lara Shepherd (Massey University) and Bruce Thomas (Landcare Research, Nelson) in November 2001.

Adult and chick Adélie penguin blood samples were collected from Cape Bird by Craig Millar and myself in December 2001 and January 2002.

25 additional blood samples from adult and chick Adélie penguins from Cape Bird were kindly donated by Fiona Hunter (Plant and Animal Sciences, University of Sheffield).

7 DNA samples from Mawson and Davis were donated by Knowles Kerry (Australian Antarctic Division, Kingston, Tasmania). 16 DNA samples from the Antarctic Peninsula were donated by Carol Vleck (Iowa State University).

DNA from Davis, Mawson and the Antarctic Peninsula was sequenced by Peter Ritchie. I extracted, sequenced, sexed and genotyped DNA from all other blood samples used in this study.

All DNA sequencing was performed by Lorraine Berry (Massey University Sequencing Facility), as was the genotyping of T series samples. I genotyped the 25 samples donated by Fiona Hunter, with the kind assistance of Danielle Hubbard at the Equine Blood Typing Facility, Massey University.

I performed all post-sequencing analysis on samples used in this study.

Contents Page

List of Figures

List of Tables

List of Abbreviations

A		adenine
A	Ala	Alanine
A		*Antarctic* (mt DNA lineage)
aa		amino acid
ATPase	6, 8	Adenine triphosphate synthase subunit 6, 8
BAL		Balleny Islands
bp		base pairs
C		cytosine
C	Cys	Cysteine
°C		degrees Celsius
CA		Cape Adare
CB		Cape Bird
CCD		central conserved domain
CI		confidence interval
CO	I, II, III	cytochrome oxidase subunit I, II, III
CR		control region
Cyt*b*		Cytochrome *b*
D	Asp	Aspartic
D-Loop		displacement loop
DMP		Davis, Mawson and Antarctic Peninsula
DNA		deoxyribonucleic acid
E	Glu	Glutamic
F	Phe	Phenylalanine
G		guanine
G	Gly	Glycine
H		heavy (strand)
H	His	Histidine
HSP		heavy strand promoter
HVR	I, II	Hypervariable region I, II
hX		hypoxanthine
I	Ile	Isoleucine
K	Lys	Lysine
k_a		rate of evolution
kb		kilobase
km		kilometer
k_s		rate of nucleotide substitution
kya		thousand years ago
L		light (strand)
L	Leu	Leucine
LGM		last glacial maximum
LR		long repeat
LSP		light strand promoter
μ		rate of mutation

M	Met	Methionine
mt		mitochondrial
Myr		million years
n		sample size
N	Asn	Asparagine
NADH	1-6	Nicotinamide adenine dinucleotide dehydrogenase subunits 1-6
nt		nucleotide
O_H		origin of heavy strand replication
O_L		origin of light strand replication
P	Pro	Proline
PCR		polymerase chain reaction
PM		Port Martin
Q	Gln	Glutamine
R	Arg	Arginine
RNA		ribonucleic acid
rRNA		ribosomal RNA
RS		*Ross Sea* (mt DNA lineage)
S	Ser	Serine
s/s/Myr		substitutions per site per million years
SSR		simple short repeat
T		thymine
T	Thr	Threonine
TAS		termination associated sequences
tRNA		transfer RNA
U		uracil
V	Val	Valine
W	Trp	Tryptophan
Y	Tyr	Tyrosine
yr BP		years before present
12S		12S rRNA subunit
16S		16S rRNA subunit

Chapter One

A Study of Mitochondrial DNA Diversity and Variability Using Adélie Penguins

1.1 The Mitochondrial Genome

This thesis is a study of mutation rates and processes in the mitochondrial control region of Adélie penguins (*Pygoscelis adeliae*), together with haplotype variation in the mitochondrial genome. The mitochondrial (mt) genome, and especially its control region has become a popular tool for resolving questions in evolutionary biology (Baker and Marshall 1997; Brown 1985). This is because of its specific properties that include maternal inheritance, a probable lack of recombination and a high rate of mutation (Stoneking 2000). These properties mean mitochondrial DNA has a much higher nucleotide diversity than the nuclear genome. In addition, its small size and high copy number makes the mt genome much easier to use.

The Adélie penguin is an extensively researched species. Many aspects of lifestyle, breeding and population biology, in addition to its genetics have been studied (Ainley 2002). The control region of the mitochondrial genome, and many nuclear microsatellite markers have been described (Lambert et al. 2002; Ritchie and Lambert 2000; Roeder et al. 2001). It is relatively straightforward to collect DNA samples from Adélie penguins in order to answer questions of population genetics and evolutionary studies. In addition, the cold dry climate of Antarctica, coupled with the

breeding biology of the Adélie penguin has allowed the extraction of ancient DNA from serially preserved bones found beneath living and abandoned Adélie penguin colonies (Lambert et al. 2002; Ritchie 2001). This allows the exciting addition of another dimension to genetic studies.

In this chapter a summary of the sources of mt DNA variability will be discussed. This is followed by overviews of the current literature on rates of mitochondrial genome evolution, the role of rate heterogeneity and variation in calculating these estimates and the uses of a highly variable mitochondrial control region in population genetics. The Adélie penguin as a model for studying mt DNA variability is also examined. The aims of this thesis will then be discussed in greater detail.

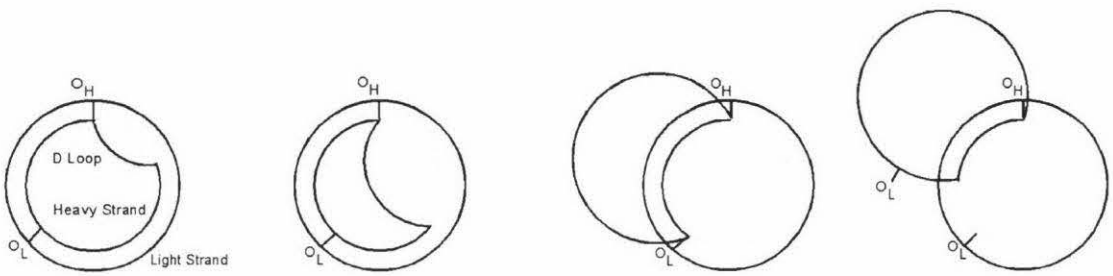
1.2 A Molecular Basis for the High Nucleotide Diversity of the Mt Genome

Mitochondrial DNA (mt DNA) has been observed to have a very high rate of sequence mutation relative to nuclear DNA (Brown, Jr. and Wilson 1979). The high rate of mt DNA mutation is thought to be partially a byproduct of the biochemical processes that govern DNA repair and replication of the mitochondrial genome (Avise 1991; Brown 1985). The proximity of the mt genome to the source of oxidative phosphorylation which takes place in the inner mitochondrial cell membrane, coupled with the fact that mt DNA lacks histones is part of the reason for such a high rate of mutation in the mitochondrial genome.

Another part of the reason for such a high rate of mt DNA mutation lies in its method of replication. The mitochondrial genome utilises a method called asymmetric replication (Figure 1.1a). In replication, synthesis of the daughter H strand begins in the displacement loop (D-Loop) in the control region by displacing the parental H strand. This parental H strand is now single stranded, and remains so until two thirds of the daughter H strand is synthesised. At this point the daughter L strand begins replication in the opposite direction. The consequence of this is that the parental L strand is never single stranded. However with replication taking up to two hours, cytochrome *b* on the parental H strand can be single stranded for approximately 80 minutes (Clayton 1982). During this single stranded phase, spontaneous deamination

of cytosine (C) to uracil (U) and adenine to hypoxanthine (hX) through hydrolytic attack can occur. If it is not repaired, when the daughter L strand is synthesised, U will pair with adenine (A) and hX with C, causing guanine (G) to be replaced by A and thymine (T) by C (Figure 1.1b). Hence, there is a high level of transitions in the H strand.

a)



b)

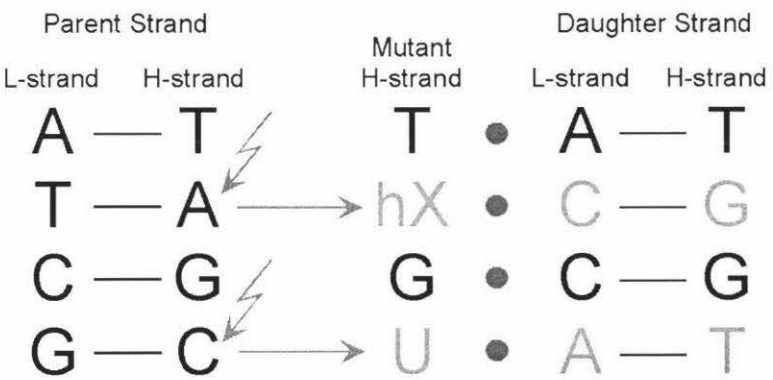


Figure 1.1
a.) Asymmetric replication of the mitochondrial genome. The black lines represent the parent molecule, grey lines the daughter molecule. O_L is the origin of light strand replication, O_H the origin of heavy strand replication.
b.) Deamination by hydrolytic attack of A and C to hX and U respectively causes high levels of A-G and C-T transitions on the H strand

1.2.1 The Mitochondrial Control Region

The control region (CR) is a noncoding segment of the mt genome that contains the promoters for transcription (light and heavy strand promoters: LSP and HSP respectively), the origin of heavy strand replication (O_H) and the displacement loop (D Loop) in vertebrates, as well as termination associated sequences (TAS), F, E, D and

C boxes and conserved sequence block 1 (CSB1) (Chang and Clayton 1985; Clayton 1982; Randi and Lucchini 1998). In most avian species including Adélie penguins the CR spans the region between the genes for tRNA-Glu and tRNA-Phe. Based on the distribution of variable sites and differing nucleotide frequencies, the control region is divided into three regions (Brown et al. 1986; Saccone, Pesole and Sbisa 1997). The 5' peripheral domain is known as hypervariable region I (HVR I). This is followed by a central conserved domain (CCD), then a second hypervariable region (HVR II). These three regions are sometimes also known as domains I, II and III respectively (Baker and Marshall 1997). The control region is often the most variable region of the mt genome, although in some species cytochrome *b* is more variable (Ruokonen and Kvist 2002). Most of the variability is in the two hypervariable regions, which show greater length variation and nucleotide variability than the central conserved domain (Brown et al. 1986; Ruokonen and Kvist 2002). The Adélie penguin CR is 1758 bp long, with HVR I being 560 bp, the CCD 200 bp and HVR II 998 bp long (Ritchie and Lambert 2000). The length of HVR II is only slightly smaller than the complete CR of many avian species (Baker and Marshall 1997).

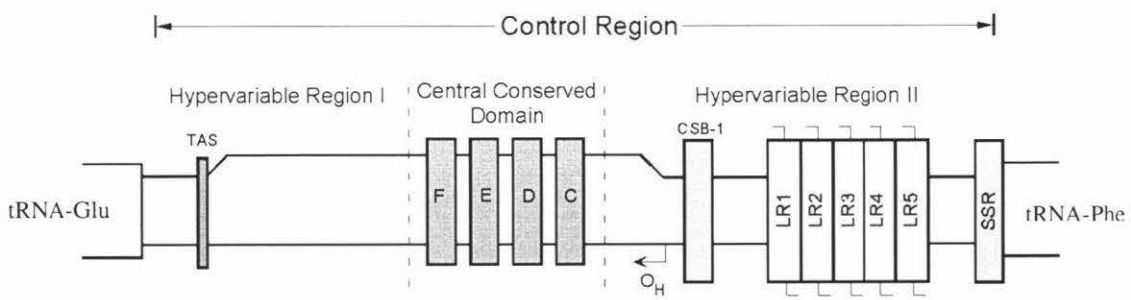


Figure 1.2. A diagram of the mitochondrial control region and flanking genes in Adélie penguins. OH, origin of heavy strand replication; LR1-5, 81bp large repeats 1-5 each containing putative sites for L and H strand promotion; SSR, a simple sequence repeat of 4 nucleotides. For additional abbreviations see text. Figure adapted from Ritchie and Lambert (2000).

1.2.2 Heteroplasmy and Bottlenecks

Another component of the diversity of mt DNA types is the presence of heteroplasmy and bottlenecks. Mutations in the mitochondrial genome occur at the level of the individual mitochondrion, not at the level of a cell. Somatic cells have between 10

and 10,000 mt DNA copies per cell, and oocytes have about 100,000 copies (Piko and Taylor 1987). When mutant forms of mt DNA arise, more than one mitochondrial type is then present in a single cell. This is known as heteroplasmy. Varying levels of heteroplasmy have been recorded in different tissues of a single individual as somatic mutations occur throughout the life of the individual (Bendall, Macaulay and Sykes 1997). When mutations occur in the oocyte, the variation can be passed on to the next generation. Transmission of only some oocyte variants of mt DNA to the next generation is known as the mitochondrial DNA bottleneck (Bergstrom and Pritchard 1998).

Mitochondrial DNA does not replicate until the 100 cell stage of the blastocyst. Because of this, the mt genomes within the oocyte are randomly sorted into each cell as they divide. This random sorting means only a small proportion of the mt DNA pool will be separated into the precursor germ cells to be transmitted to the next generation. In some species, such as humans, the mt DNA bottleneck appears to be very severe, as complete shifts from one genotype to another in a single generation have been detected (Parsons et al. 1997; Sigurðardóttir et al. 2000). Single point heteroplasmy, as well as length heteroplasmy has been observed between generations in many species (for example Parsons et al. 1997; Taylor and Breden 2002), although it does not appear to be common in humans (Sigurðardóttir et al. 2000).

Studies that have examined mitochondrial bottlenecks include Bendall et al. (1996), who investigated heteroplasmy in humans. Using DNA from twin studies, they found four heteroplasmic point mutations in 180 twin pairs. Changes in heteroplasmy within mother/offspring pairs were used to estimate developmental bottleneck size ranges. Their data suggested a bottleneck of between three and 20 segregating units, supporting the hypothesis of a tight generational bottleneck for mitochondria in humans.

Although studies in humans seem to suggest heteroplasmy is a rare observation (Bendall et al. 1996; Sigurðardóttir et al. 2000), studies in other species would perhaps indicate it is more widespread than previously thought. In other mammals such as cattle and mice the effective size of the bottleneck appears slightly larger (reviewed in Bendall, Macaulay and Sykes 1997). In guppies, individuals may

possess up to nine different sized mitochondrial haplotypes, and comparison of related individuals provided no evidence for a mt DNA bottleneck (Taylor and Breden 2002). As sequencing detection techniques improve and the cost of sequencing multiple cloned mitochondrial sequences decreases, detection of mitochondrial heteroplasmy is likely to increase.

1.3 Rates of Mutation and Evolution

1.3.1 The Neutral Theory of Molecular Evolution

In 1983, Kimura published a landmark work in the theory of neutral evolution. Briefly, neutral theory states that at the molecular level, most variability in species comprises variant mutant alleles that are selectively neutral, whose fate is determined mostly by random genetic drift (Kimura 1983). This theory describes how the rate of evolution (k_a) is related to the intergenerational rate of mutation (μ). The formula

$$k_a = N.\ \mu.\ P_{fix}$$

explains this relationship, where N is the number of haploid genomes and P_{fix} is the probability of fixation. Since any mt DNA sequence in a population could become fixed in a future generation, only a very small fraction of observable intergenerational mutations will be fixed, hence

$$P_{fix} = 1/N$$

Together, these two formulae give

$$k_a = N.\ \mu.\ 1/N$$

$$k_a = \mu$$

This means that under neutrality, the rate of evolution is equal to the rate of mutation and so one is an indirect measure of the other (Kimura 1983, summarised by Loewe, 1997).

Traditionally, the rate of mutation has been calculated indirectly through phylogenetic studies that use sequence variation between two related taxa calibrated to the fossil record to estimate the rate of evolution. Until recently, it has been very difficult to directly calculate the intergenerational rate of mutation. With the advent of high

throughput sequencing this has become feasible. Now it is possible to independently calculate k_a and μ and experimentally investigate Kimura's neutral theory of molecular evolution.

1.3.2 Calculating a Rate of Evolution

Also termed the 'relative branch length method' (Ruvolo 1996), phylogenetic studies estimate k_a, the rate of evolution along an ancestral lineage. Phylogenetic studies have in the past used restriction fragment length polymorphism (RFLP) and lately sequence variation between closely related species, for example humans and chimpanzees, to derive a phylogenetic tree using such methods as maximum parsimony, neighbour joining or maximum likelihood. Branch lengths are then calibrated by some external measure, such as species divergence dates from the palaeontological fossil record. This information can then be used to calculate a rate of nucleotide substitution (k_s) (Shields and Wilson 1987; Stoneking et al. 1992).

1.3.3 The Rate of Evolution in Birds

Shields and Wilson (1987) used this method to calculate a rate of substitution for the mitochondrial genome in birds. Using RFLP analysis, they calculated the extent of sequence divergence between species of the geese genera *Anser* and *Branta*. The midpoint root of the phylogenetic tree showed 9% sequence difference between the two genera, and the oldest fossil records from each genus were dated about 4-5 million years ago. By dividing the sequence difference by the age of the oldest fossils (9%/4.5 Myr), Shields and Wilson estimated a mean rate of divergence of 2%/Myr for the whole mitochondrial genome in geese. The authors argued that the similarity of their results and those from mammals was evidence for the validity of their calculations.

The results of Shields and Wilson (1987) were further extrapolated by Quinn (1992) to calculate the rate of substitution of the mt DNA control region. Quinn estimated that the control region evolves approximately 10.4 times faster than the overall mt

genome. This estimate was multiplied by the Shields and Wilson rate of 2%/Myr to give a rate of substitution for the control region of 20.8%/Myr (0.208 s/s/Myr). This result has subsequently been used as an estimate of the substitution rate of the mt DNA control region for all bird species (e.g. Baker and Marshall 1997).

However, such phylogenetic analyses do have inherent difficulties. They assume that subsequent to the point of divergence the two lineages evolved at the same rate, i.e., that the substitution rate is the same in *Anser* and *Branta*, or humans and chimpanzees (e.g., Stoneking et al. 1992). Another problem is the possibility of multiple substitutions occurring at the same site (especially in the rapidly changing mt DNA control region), and consequently these need to be corrected for (Howell, Kubacka and Mackey 1996).

Stoneking et al. (1992) attempted to correct these problems by the use of intraspecific calibration. Intraspecific calibration calculates sequence divergence *within* a species and avoids the shortcomings of comparing two species. Instead, the calibration is achieved by using monophyletic clusters of mt DNA types specific to defined geographic locations. Stoneking et al. wished to calculate the rate of mt DNA control region substitution in humans. To do this, the authors used samples from people living in Papua New Guinea, a location essentially colonised once, with a relatively well defined colonization time, and little back-migration (parameters to avoid errors in rate estimation). Using calculations of 'within group' and 'between group' mt DNA diversity, Stoneking et al. calculated a substitution rate of 11.8%/Myr (0.118 s/s/Myr) for the mt control region. This estimate is similar to that calculated by previous phylogenetic estimates in humans (Vigilant et al. 1991), and similar to the avian rate calculated in geese (Quinn 1992).

1.3.4 Calculating a Rate of Mutation.

The rate of mutation (μ) of a DNA sequence is defined as the mutation rate per individual per generation (Sigurðardóttir et al. 2000) and is then converted to substitutions per site per million years for comparison. This can be directly estimated through pedigree studies. Pedigree studies compare DNA sequences of closely

related individuals (mother-child, grandmother-grandchild and sibling-pair comparisons). The mutation rate is then estimated from the proportion of sequence changes observed.

Many key studies in the area of mutation estimation are based on human sequences, however the underlying arguments remain the same no matter which species is used. Since the onset of high throughput DNA sequencing techniques, directly estimating the rate of mutation through sequence comparison has become feasible. Use of this method in human pedigree studies has lead to estimates of mutation rate in the mt DNA control region ranging from 0 to 2.5 s/s/Myr (Parsons and Holland 1998).

Unfortunately, as can be seen from the summary in Table 1.1, this method has not lead to a consistent estimate of mutation rate. This is perhaps not surprising since these studies each use slightly different methods. Pedigree study estimates are on average much higher than the previous indirect estimates calculated by phylogenetic methods, but also cover a range that includes the phylogenetic estimates. This variation stems in part from debate about whether or not to include instances of heteroplasmy, the small size of some samples, the possibility of recombination and the importance of mutational hot spots.

Early results included those of Howell et al. (1996), who found two mutations in 81 transmission events. This gave a mutation rate of approximately 1 mutation per 40 generations (0.75 s/s/Myr). On its own, this study would appear to be an outlier, especially considering the small size of the study, and the fact that the four pedigree families chosen all contained the Leber hereditary optic neuropathy (LHON) disease. This is a mt DNA disease with pathogenic mutations that impair mitochondrial respiratory-chain function. It is entirely plausible that this abnormal mitochondrial metabolism may alter the rate of mutation in another part of the mitochondrial genome (i.e., the control region) (Jazin et al. 1998).

Table 1.1 Summary of studies into the rate of mutation

Study	Region Sequenced	Point muts. observed	No. of transmissions	Rate (s/s/Myr)
Howell et al. (1996)	HVR I & II	2	81	0.75
Bendall et al. (1996)	HVR I	4 heteroplasmies	180 twin pairs	0.06-1.35 (0.564 midpoint)
Parsons et al. (1997)	HVR I & II	10	327	2.5 (95% CI 1.2-4.0)
Jazin et al. (1998)	HVR II	0	208	<0.46 (99% CI 0.0-1.52)
Jazin et al. (1998)	HVR II pooled rate	7	804	1.17 (99% CI 0.15-2.2)
Parsons et al. (1998)	HVR I & II pooled rate	17	1065	1.35 (95% CI 0.72-1.98)
Sigurðardóttir et al. (2000)	HVR I & II	3 subs 3 hets	705	0.32 (95% CI 0.065-0.97)
Sigurðardóttir et al. (2000)	HVR I & II pooled rate	14	1221	0.852 (95% CI 0.46-1.42)
Heyer et al. (2001)	HVR I & II	4	508	0.39 (95% CI 0.113-0.917)
Heyer et al. (2001)	HVR I & II pooled rate	18	1729	0.517 (95% CI 0.34-0.737)
Howell et al. (2003)	HVR I & II pooled Howell et al. studies	3	263	1.02 (99.5% CI 0.09-3.97)
Howell et al. (2003)	HVR I & II pooled rate	28	2633	0.95 (99.5% CI 0.53-1.57)

However, soon after the high mutation rate report of Howell et al., a series of other reports were published. One of the larger studies undertaken was that of Parsons et al. (1997). They studied pedigrees from four different sources (the Armed Forces DNA Identification Laboratory (AFDIL), Oxford British families, CEPH pedigree cell lines and Old Order Amish pedigree lines); a large number of mitochondrial lineages totalling 357 individuals. Sequences from these individuals were used in the direct measurement of the intergenerational substitution rate of the human mitochondrial DNA control region. A total of 327 generational events were measured, resulting in the detection of 10 instances of substitution. This is also a high mutation rate, about one in 33 generations, or 2.5 s/s/Myr. Again, this is roughly 20 times higher than that predicted by phylogenetic analysis (Stoneking et al. 1992).

Jazin et al. (1998) argued that the high rate found by Parsons et al (1997) may be the result of mutational hot spots and a higher mutation rate in certain family lines (because of the disease state). Their own study, and that of Soodyall et al (1997) found no homoplasmic mutation events in 288 and 108 meiotic events respectively. Jazin et al. suggest pooling the results for HVR II to provide a more accurate result (see Jazin et al. and references therein). This yielded a mutation rate of 7/804 events, or 1.17 s/s/Myr (99%CI=0.15-2.2), a range that includes the phylogenetic rate estimate (~0.15) at its lower end, suggesting the two methods of estimating k_a and μ do not produce significantly different results.

Parsons et al. (1998) responded that Jazin et al.'s (1998) conclusions were due to the use of selective data and to inappropriate methods of analysis. They argue that by restricting their pooled results to only the HVR II, much published data on the control region was omitted. By pooling all published data on both the HVR I and HVR II (see Parsons et al 1998. and references therein), Parsons et al. achieved a resulting rate of 1.35 s/s/Myr (95% CI=0.72-1.98). Furthermore, they argue this result is conservative as it assumes in studies of only one hypervariable region that there were no additional mutations present in the other hypervariable region.

1.4 Mutational Hot Spots, Rate Heterogeneity

It has been shown that there is great mutation rate heterogeneity along the control region. In humans, it has been noted that roughly half the positions in HVR I are almost invariable (Meyer, Weiss and von Haeseler 1999). Although thousands of control region sequences have been determined in different species, knowledge of substitution patterns in the control region is far from complete. Control region evolution is very complex, among other things because base composition is not uniform, transitions occur more frequently than transversions, and substitution rates vary among sites. Meyer et al. (1999) have begun to unravel this complexity by looking at the pattern of nucleotide substitution in the human control region. They argued that accurate modelling of substitution rate variation is needed especially in a population genetics context. They suggested that inappropriate modelling may lead to a misinterpretation of data, for example mutational hot spots could mimic population expansion.

The question has been asked if mutational hot spots can account for some of the discrepancy between phylogenetic and pedigree analyses of mutation rate (e.g.Jazin et al. 1998; Stoneking 2000). Meyer et al. (1999) state their estimates of nucleotide substitution rates for each site in the two hypervariable regions will be useful in interpreting the conflict in mutation rate estimates for HVR I and II. In addition, there are benefits to sequence analysis in general as it allows the refinement of phylogenetic models and more precise interpretation of population sequence data.

Other authors have investigated the possibility that highly polymorphic sites represent mutational hot spots rather than old sites rooted early in the phylogenetic tree, and the degree to which these can account for the higher rates of substitution seen in pedigree studies. Gurven (2000) analysed linkage disequilibrium patterns in mt DNA, on the expectation that hot spots would show little or no disequilibrium as they can be interpreted as having randomly expressed patterns. Gurven's results suggested many polymorphic sites were in fact hot spots. Stoneking (2000) compared rates of mt DNA mutations occurring in germ line cells, heteroplasmic cells and somatic cancer cells (using HVR I and II nt site rates estimated by Meyer et al. 1999), and found that all three types of mutation occurred preferentially at hypervariable sites. This

supports the view that these represent mutational hot spots, not ancient sites rested early in the phylogenetic tree.

The analysis by Stoneking (2000) deserves further discussion. Not only did Stoneking show that mutations occur preferentially at hypervariable sites, but also that the rates of evolution are very similar for sites where germline, somatic and heteroplasmic mutations were observed. Stoneking suggested that this could be interpreted in two ways. Either the processes resulting in mutation are the same in germ line and somatic cells, or some percentage of the supposed germline mutations may actually be somatic mutations occurring early in prenatal development. The early somatic mutations would then drift to homoplasmy, with heteroplasmy reflecting an intermediate stage. Therefore supposed germline mutations are in fact somatic, hence the same observed rate. This second hypothesis could also be part of the answer as to why pedigree studies find a much higher rate of mutation than phylogenetic studies. Further work needs to be carried out to find out how mutations are transmitted to subsequent generations. It is however interesting to note that Sigurðardóttir et al.'s (2000) study of Icelandic populations was designed to rule out somatic mutations, and has one of the lower estimates of mutation rate (but still higher than estimates of evolutionary rate using phylogenetic methods).

Sigurðardóttir et al. (2000) agreed that there appears to be substantial heterogeneity of mutation rate across the control region, but argued that it is actually irrelevant to the estimation of mutation rate. The authors state that directly counting mutations by pedigree studies will give an unbiased estimate of the average mutation rate for the whole region, no matter how much rate heterogeneity there is within that region. This could be contested only if some sites mutate so quickly that back mutations occur within a pedigree, which seems unlikely. Howell et al. (1996) noted that rate heterogeneity is a problem for phylogenetic studies, which may consistently underestimate the average mutation rate unless they are able to accurately correct for multiple mutations at the same site. An accurate mutation rate estimation using phylogenetic methods requires a good understanding of rate heterogeneity.

Heyer et al. (2001) agree that there is rate heterogeneity in the control region, and propose a model where sites vary at one of three rates: slow, moderate or fast. They

argue that sites seen to mutate in pedigree studies are generally those that mutate at a faster rate. This is because the limited number of maternal transmissions studied means the time window is short, favouring those sites that mutate quickly. Slow and medium sites would predominate in phylogenetic rate estimates. Heyer et al. suggest this is because back mutations are more likely to occur at fast sites, masking their contribution to phylogenetic rate estimates.

1.5 Mt DNA Variability as a Tool in Population Studies

In addition to using mt DNA variation to determine rates of mutation and evolution, these data can also be applied to a diverse range of evolutionary problems. For example, mitochondrial DNA variation has been used to study theories of population migration and makeup through space and time. Because of its high polymorphism, mt DNA is an excellent marker for population studies on a smaller time scale than nuclear DNA.

Variation in mt DNA has been used many times to track population movement, especially migrations of modern humans. Matisoo-Smith et al. (1998) presented an alternative approach to questions about Polynesian settlement and mobility by looking at mt DNA control region variation in the Polynesian rat (*Rattus exulans*). The Polynesian rat was transported throughout Remote Oceania in the colonising canoes of the ancestral Polynesians. DNA phylogenies derived from mt DNA control region sequences of diverse Polynesian rats were used to test hypotheses on the degree of interaction within Polynesia. The results were more consistent with a general pattern of multiple contacts rather than isolation between Polynesian populations in the pacific.

Mitochondrial DNA can also be used to study population bottlenecks, such as those due to population isolation and fragmentation during a glacial maximum. A study by Wenink et al. (1994) of mt DNA HVR I looked at genetic variability in two shorebirds, the turnstone (*Arenaria interpres*) and the dunlin (*Calidris alpina*). By looking at population genetic structuring, Wennink et al. could hypothesize on their respective phylogeographic histories. Dunlins were shown to have considerable

variation in their sequences with strong genetic structuring subdividing the population. The variation could be explained by strong natal phylopatry combined with a high mutation rate in HVR I. The genetic subdivision was postulated to have evolved through population isolation and fragmentation into refugia during the Pleistocene. In contrast, the turnstones were shown to have very little HVR I variation and no definitive population structure. Wenink et al. (1994) postulated that this is the result of an expansion from a refugial population that was bottlenecked in the recent past.

1.6 The Adélie Penguin Model

Adélie penguins represent a unique opportunity to study rates of mutation, HVR I variation and phylogeography in a species other than human. The unique situation of Adélie penguins in Antarctica has provided the opportunity for a different approach to phylogenetic analyses of evolution rate. Beneath the densely populated living and abandoned Adélie penguin rookeries in Antarctica are large deposits of sub-fossil bones that have been continuously preserved for thousands of years (Baroni and Orombelli 1994). The cold and dry climate of Antarctica has meant these bones are good sources of ancient DNA (Ritchie 2001). This makes it an ideal situation for 'serial sampling', using ancient DNA (bones) and modern DNA (living penguins) to follow mt DNA sequence through time (Lambert et al. 2002). From a comparison of ancient and living taxa it is possible to directly estimate rates of nucleotide sequence evolution (Lambert et al. 2002).

Lambert et al. (2002) have been able to calculate a rate of evolution for the mt DNA hypervariable region 1 of Adélie penguins using the methods mentioned above in conjunction with two population based analyses: Markov Chain Monte Carlo analysis (MCMC) (Drummond et al. 2002) and a regression analysis (Drummond and Rodrigo 2000). In this way, Lambert et al. were able to calculate that the HVR I evolves at approximately 0.96 s/s/Myr (highest posterior density (HPD) interval 0.53 to 1.43). This estimate is approximately five times higher than previous indirect phylogenetic estimates (Quinn 1992), and is more akin to rate estimates estimated by pedigree analysis (e.g. Howell et al. 2003; Parsons et al. 1997).

As interesting as comparing evolutionary rates in one species to mutation rates in another might be, it would be more appropriate to compare this new rate of evolution to a rate of mutation in Adélie penguins. Unlike many bird species, Adélie penguins are well suited to calculating a rate of mutation through pedigree analysis. Pedigree analyses require two factors to calculate a rate of mutation: a large number of mt DNA transmissions (or generations) and the ability to identify siblings or the mother of an individual. Adélie penguins live most of their life amongst the pack ice off shore from Antarctica, but come ashore to breed in large colonies during the short austral summer (Ainley 2002). Both parents take turns sitting on the nest and going out to sea to fish. Because both parents are present on the nest at one time during the changeover it is possible to identify the whole family group at once. Adélie penguins raise one or two chicks, which stay at the nest site until they are about 22 days old (Ainley 2002; Spurr 1975). After this time chicks from neighbouring nests group together in crèches allowing both parents to feed at the same time (Ainley 2002). By the time the chicks are this age, it becomes difficult to distinguish family groups. Compared to many other bird species, whose nests are often widespread, hard to access and well camouflaged, it is relatively easy to collect DNA samples from Adélie penguins belonging to many families over a short period of time. This makes the calculation of a mutation rate from pedigree studies feasible. Indeed, this is a difficult task to achieve in many species other than birds as well.

Previous studies of Adélie penguin populations around the Antarctic have shown that there are two mitochondrial lineages comprising this species. These have been designated the *Antarctic* (*A*) and *Ross Sea* (*RS*) lineages (Ritchie 2001). These are so named because the *A* lineage has been recorded throughout Antarctica, while the *RS* lineage appears to be limited to the Ross Sea region. The two highly divergent lineages show 8.3% sequence divergence (Lambert et al. 2002). Using a MCMC approach, Drummond et al. (2002) have determined that the time to the most recent common ancestor of the two lineages was 75 kyr BP (95% CI 37 –122 kyr BP).

This estimate is consistent with what is known of the glacial history of Antarctica. It has been suggested that the two lineages are likely to have evolved as two isolated ancestral populations subjected to population bottlenecks during the last glaciation

(Ritchie et al. in press.). The estimated time to the most recent common ancestor of the two lineages places their divergence during the late Pleistocene, in the middle of the last glacial cycle (Ritchie et al. in press.). At this time ice covered a much greater area of Antarctica, and Adélie penguins were almost certainly greatly restricted in areas for breeding sites on the Antarctic mainland. Indeed, populations may have been mainly reduced to breeding on small offshore islands (Ainley 2002). It is under these conditions that the two lineages are expected to have evolved.

1.6.1 Looking to the Future

Rand (2001) called for "additional studies of the accumulation of mt DNA mutations in diverse organisms (e.g. Denver et al. 2000; Parsons et al. 1997). Such studies done in organisms with very different patterns of germ line differentiation will provide important comparative data on the fate of mt DNA mutations through the mitochondrial population booms and bottlenecks of extended germ lines." The work done previously on Adélie penguins, and that conducted in this thesis will help in understanding the problems put forward by Rand (2001).

The main focus of this thesis will be an analysis of the variation and diversity found in HVR I of the Adélie penguin. Looking towards the future, many of the analyses presented in the introduction and in the body of this thesis are also applicable to regions of the mt genome other than HVR I. The first step in performing analyses on part or all of the mt genome is knowledge of the complete mt genome sequence.

1.7 Thesis objectives

The objectives of this Thesis are as follows:

1) To examine the range of the *Antarctic* and *Ross Sea* lineages in the vicinity of the Ross Sea region of Antarctica (Chapter 3)

2) To determine if Adélie penguins of the *Antarctic* and *Ross Sea* lineages are or are not one species (Chapter 3)

3) To calculate a preliminary rate of mutation in the mitochondrial hypervariable region I for the Adélie penguin (Chapter 4)

4) To analyse the nucleotide diversity and variability in hypervariable region I as it relates to the two lineages and their geographic distribution (Chapter 5)

5) To sequence the entire mitochondrial genome of an Adélie penguin (Chapter 6)

Chapter Two

Methods

2.1. Blood Collection

100 Adélie penguin blood samples were collected from Cape Adare (71°18'S/170°09'E), 30 from the Balleny Islands (66°55'S/163°20'E) and 30 from Port Martin (66°49'S/141°24'E). 5 blood samples were collected from Mawson (67°34'S/62°55'E), 2 from Davis (68°34'S/77°52'E) and 16 from the Antarctic Peninsula (64°46'S/64°5'W). 25 blood samples were collected from Cape Bird (77°13'S/166°28'E). See preface for acknowledgement of collectors.

377 blood samples from Cape Bird, Antarctica were collected in the austral summer of 2001/2002. In early November blood samples were collected from adult birds showing typical pair behaviour. The birds were bled, banded, weighed and the nest site recorded. In late December and early January blood samples were collected from chicks belonging to pairs identified and bled earlier. Further pairs were also identified with one or preferably two chicks on the nest and all family members were bled.

2.1.1 Family identification and sample collection

A nest was identified by the presence of two adult birds on a nest site with one or two chicks exchanging ritual greeting and pair bonding as one bird returned from feeding

and before the other bird left to feed. Pairing was also confirmed through behaviour as both adults were seen to defend the nest site when approached by a human. The nest site was recorded, one adult and both chicks were uplifted from the nest site, and the remaining adult observed to stay and defend the nest site. Adults were either picked up by hand or restrained with a hand net or shepherds crook if agitated and chicks picked up by hand. The chicks were kept confined in a small black bag for warmth and protection while the adult penguin was being bled.

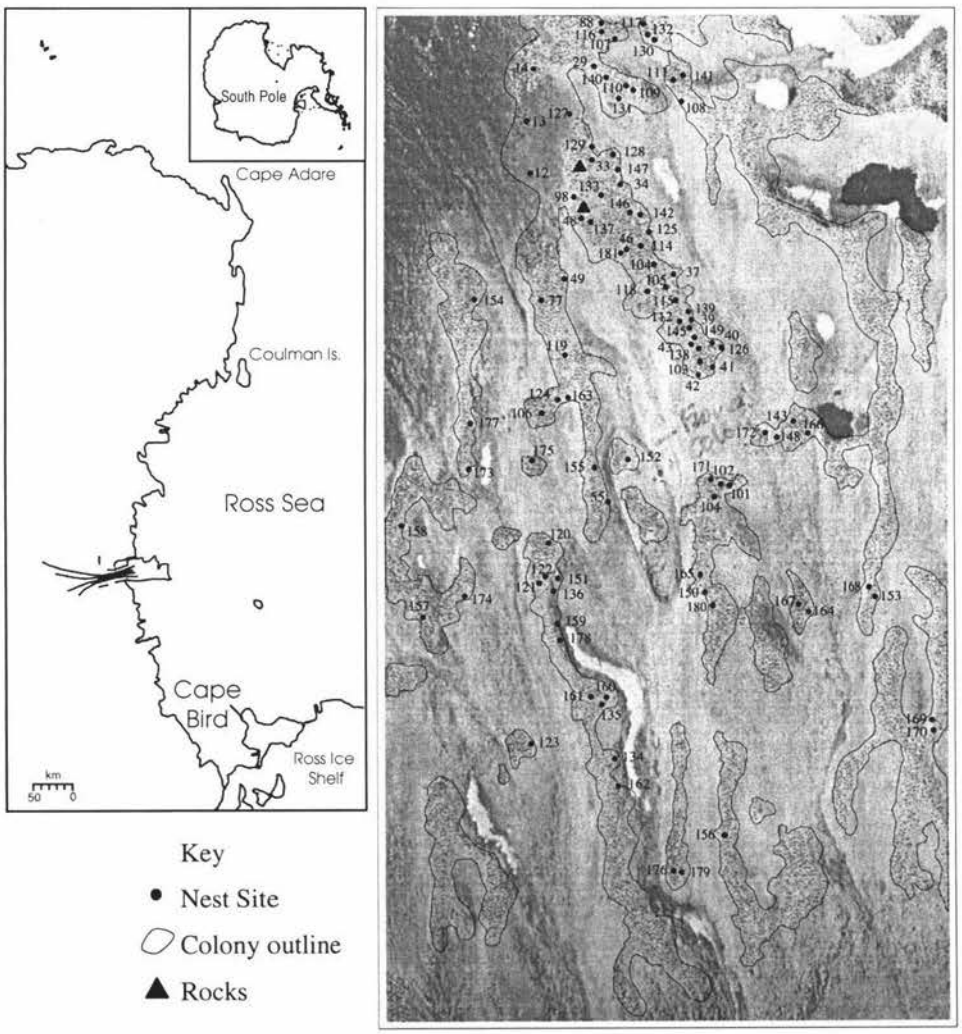


Figure 2.1 Map showing location of nest sites at Cape Bird, Antarctica used in the pedigree mutation study. The small black dots on the aerial photograph are Adélie penguins. The GPS readings of the two rocks are 77°12'54.0"S/166°27'1.6"E and 77°12'55.3"S/166°26'58.9"E from the top of the figure respectively.

The adult was weighed using a 5 or 10 kg hand scale and a numbered band was placed securely around the upper left flipper. The right foot was swabbed with ethanol and 100 µL of blood was taken from the outside of the leg intertarsal vein, or from the foot intradigital vein using a 1 mL syringe with 25 gauge needle washed in TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA). The blood was immediately put into 1 mL lysis buffer (Seutin, White and Boag 1991) and stored at 4°C for return to the lab. The chicks were then bled in the same manner, and all three birds returned to the nest, as the other adult penguin was uplifted. This bird was then weighed, banded and bled in the same manner as the first adult, and returned to the nest site.

2.2 Extraction of DNA from Blood

2.2.1 Method 1

DNA was extracted from blood samples using a phenol/chloroform method (based on Sambrook, Fritsch and Maniatus 1989) as follows. Four µL of blood was dissolved in 400 µL extraction buffer (100 mM Tris-HCl pH 8.0, 100 mM NaCl, 1 mM EDTA), 10 µL 10% sodium-dodecyl-sulphate (SDS) and 20 µL 10 mg/mL proteinase-K, and incubated at 50°C for at least two hours, until completely digested. Proteins and cellular debris were removed by adding 400 µL tris-saturated phenol and rocking for 30 minutes at room temperature. Samples were centrifuged for 1 minute at 16 000 g and the lower phenol layer removed and discarded. 400 µL phenol: chloroform: isoamyl alcohol (25:24:1) was added and the sample was rocked and centrifuged as before. The lower layer was again removed. 400 µL chloroform: isoamyl alcohol (24:1) was added and the sample rocked for 5 minutes and centrifuged as before. The upper aqueous layer containing DNA was then removed and transferred to a new 1.7 mL tube.

DNA was precipitated by addition of 1 mL of cold 100% ethanol and 40 µL 3 M sodium acetate pH 5.2, then placed at -20°C for 30 minutes before being centrifuged for 30 minutes at 16 000 g. The supernatant was removed, and the pellet washed with 200 µL 70% ethanol and spun at 16 000 g for a further 10 minutes. The supernatant

was removed, and the pellet allowed to air dry. When dry, the pellet was resuspended in 100 µL of TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA). Purified DNA samples were kept at 4°C while in use, and then stored at -20°C.

2.2.2 Method 2

DNA was extracted from blood by phenol/chloroform method with ethanol precipitation as described in method 1 (2.2.1), except the phenol step, and the chloroform: isoamyl (24:1) step were omitted, and after the addition of 400 µL phenol: chloroform: isoamyl the sample was rocked for 60 minutes, until the upper DNA layer was clear and the lower phenol layer was cloudy.

2.2.3 Method 3

DNA was extracted from 25 blood samples donated by Fiona Hunter using the Bio-Rad Instagene matrix as per the manufacturers instructions (Bio-Rad Laboratories, Inc.).

2.3.1 Sexing of Adélie Penguins Using Three PCR-Based Methods

Three avian sex specific primer sets were trialed in sexing Adélie penguins of known gender. These were the forward primer P8 (5'-CTCCCAAGGATGAGRAAYTG-3') with its reverse P2 (5'-TCTGCATCGCTAAATCCTTT-3') (Griffiths et al. 1998), W1 (5'-CCTTTAAACAAGCTGTTAAAGCA-3') with K7 (5'-AGAGAAAACAAG ATCTGTGGGA-3') and K1 (5'-GCAGAGGTGAATGCTCAATGG-3') with K2 (5'-CGCCATGTACTCATAGACAGG-3') (Huynen, Millar and Lambert 2002). 10 µL PCR reactions were set up containing 1x PCR Buffer (10 mM Tris-HCl, 50 mM KCl pH 8.3), 2.5 mM $MgCl_2$, 100 µM of each dNTP, 1 µg/µL BSA, 20 ng of each primer and 1U Ampli*Taq* (PE Biosystems) with 1 µL extracted DNA and overlaid with a drop of mineral oil. PCR amplification conditions in a Hybaid Omn-E Thermal Cycler were 94°C for 2 minutes followed by 10 cycles of 94°C denaturation for 20

seconds, 55°C annealing for 20 seconds, 72°C elongation for 30 seconds, then 25 cycles of 94°C for 20 seconds, 50°C for 20 seconds, and 72°C for 30 seconds.

The P2/P8 product contains a digestion site for the restriction enzyme *Hae* III. 5 μL of the PCR product was digested with 0.2 μL *Hae* III (10 U/μL), 0.6 μL 10x React II and 0.2 μL H_2O for 55 minutes at 37°C. PCR products were separated by electrophoresis in 1% MS: 1% LE agarose (Boehringer Manheim) in TBE (45 mM Tris-borate, 1 mM EDTA), stained with ethidium bromide and visualised under UV light. Samples were run alongside a 1 Kb ladder to determine size of DNA fragments.

2.4 Genotyping of Adélie Penguin Families

2.4.1 Method 1

Samples donated by Fiona Hunter were genotyped for the six microsatellite DNA loci RM3, RM6, AM3, AM12, AM13 (Roeder et al. 2001) and HrU2 (Ellegren 1992; Primmer, Moller and Ellegren 1996) (Appendix B, Table 3). In each case the reverse primer had been 5'-end labelled with 6-FAM (AM3, AM12 and HrU2) or HEX (RM3, RM6 and AM13) phosphoramidite dye. Six separate 10 μL PCR reactions were run for each sample containing 10 pmol/μL of each primer, 200 μM of each dNTP, 1 μg/μL BSA, 2.5 mM $MgCl_2$, 1.5 U Ampli*Taq* and 1 μL extracted DNA. This was overlaid with a drop of mineral oil and placed in a Hybaid Omn-E PCR Thermal Cycler. PCR conditions were 94°C for 2 minutes followed by 30 cycles of 94°C for 45 seconds, 56°C for 50 seconds and 72°C for 60 seconds. Samples were separated on a 1% MS: 1% LE agarose gel and the relative concentration of DNA in each band was estimated.

Because of the different size ranges and labelling colour of the six products, loci were pooled before genotyping. This meant the six loci for each sample could be analysed in two lanes. RM3, RM6 and HrU2 were grouped together, as were AM3, AM12 and AM13. Different volumes of each PCR product were added depending on the relative

concentration of DNA in each, so that the final volume was not more that 10 μL. If necessary, dH_2O was added to bring the final volume up to 10 μL.

Samples were genotyped using an ABI Prism 377 Sequencer. 0.5 μL of the pooled PCR product was mixed with 2.5 μL formamide, 0.5 μL loading buffer (with 5% blue dextran) and 0.5 μL of an internal size standard, ROX (Applied Biosystems). Samples were denatured for 5 minutes at 95°C, placed immediately on ice and 2 μL loaded on to a polyacrylamide gel (BioWhittaker Molecular Applications). After running according to the manufacturers instructions, the data was analysed using GeneScan Analysis 3.1 and loci were scored by eye.

2.4.2 Method 2

T series families were genotyped using Adélie penguin specific microsatellite primers designed in this laboratory (Haynes 2002). The three microsatellite loci XVI G4, XVI H11 and XVIII B2 were used (Appendix B, Table 3). The forward primer for each locus was 5'-end labelled. XVI G4 was labelled with NED, XVI H11 with 6-FAM and XVIII B2 with VIC phosphoramidite dye (Applied Biosystems). For each sample, the three primer pairs were multiplexed into one 10 μL PCR reaction. Each PCR reaction contained 2x Buffer, 1.5 mM $MgCl_2$, 0.4 μg/μL BSA, 400 μM each dNTP, 2 M Betaine, 1U Taq, 0.4 mM XVI G4, 0.35 mM XVI H11, 0.3 mM XVIII B2 and 1μL extracted DNA. PCR was performed in a Bio-Rad iCycler. Thermal cycling conditions were as follows: 94°C for 2 minutes, 30 cycles of 95°C for 30 seconds, 58°C for 20 seconds, 72°C for 20 seconds, finishing with one cycle of 72°C for 5 minutes.

The multiplex PCR products were separated on a 1% MS: 1% LE agarose gel, then the products were diluted 1:2 with dH_2O and sent for genotyping at the Massey University Sequencing Facility on an ABI Prism 377 Sequencer (Applied Biosystems) according to the manufacturers instructions. Data analysis was with GeneScan Analysis 3.1 (Applied Biosystems) and loci were scored by eye.

2.5 PCR of Mitochondrial DNA

2.5.1 Amplification of the mt DNA control region by PCR

The control region HVR I was amplified using primers L-tRNA-Glu (5'-CCCGCTTG GCTTYTCTCCAAGGTC-3') and AH530 (5'-CTGATTTCACGTGAGGAG ACCG-3') (Ritchie 2001). PCR reactions consisted of 1 μL extracted DNA with 200 μM each dNTP, 1x PCR Buffer (10 mM Tris-HCl, 50 mM KCl pH 8.3), 0.4 μg/μL BSA, 1.5 mM $MgCl_2$, 0.4 μM of each primer and 0.5 U Ampli*Taq* (Applied Biosystems), made up to 25 μL with dH_2O and overlaid with a drop of mineral oil. PCR amplification conditions in a Hybaid Omn-E Thermal Cycler were as follows: 94°C for 2 minutes followed by 30 cycles of 94°C denaturation for 10 seconds, 60°C annealing for 10 seconds, and 72°C extension for 30 seconds.

2.5.2 Long-range PCR of the mitochondrial genome

Mitochondrial DNA was amplified in fragments longer than 5 kb in order to avoid amplifying nuclear copies of mt DNA. Three long-range PCRs were carried out using the Expand™ Long template PCR System (Roche) (Figure 2.2), with the resulting products encompassing the whole mitochondrial genome. Primer pairs were 16SFwdlg and Av13026R, Co310647F and tPro16137R, Av13063FtLeu and 16S3797R (Appendix B, Table 1). 50 μL PCR reactions contained 1x Buffer, 0.5 mM dNTPs, 3.75 U taq, 0.4 μM of each primer and 10 μL of extracted DNA. Cycling conditions were as follows: 94°C for 2 minutes followed by 10 cycles of 93°C for 30 seconds, 54°C for 30 seconds and 68°C for 10 minutes, then 26 cycles of 93°C for 30 seconds, 54°C for 30 seconds and 68°C for 10 minutes plus 20 seconds per cycle, then one cycle of 68°C for 5 minutes. Primer pair Co310647F-tPro16137R had an elongation time of 7 minutes not 10.

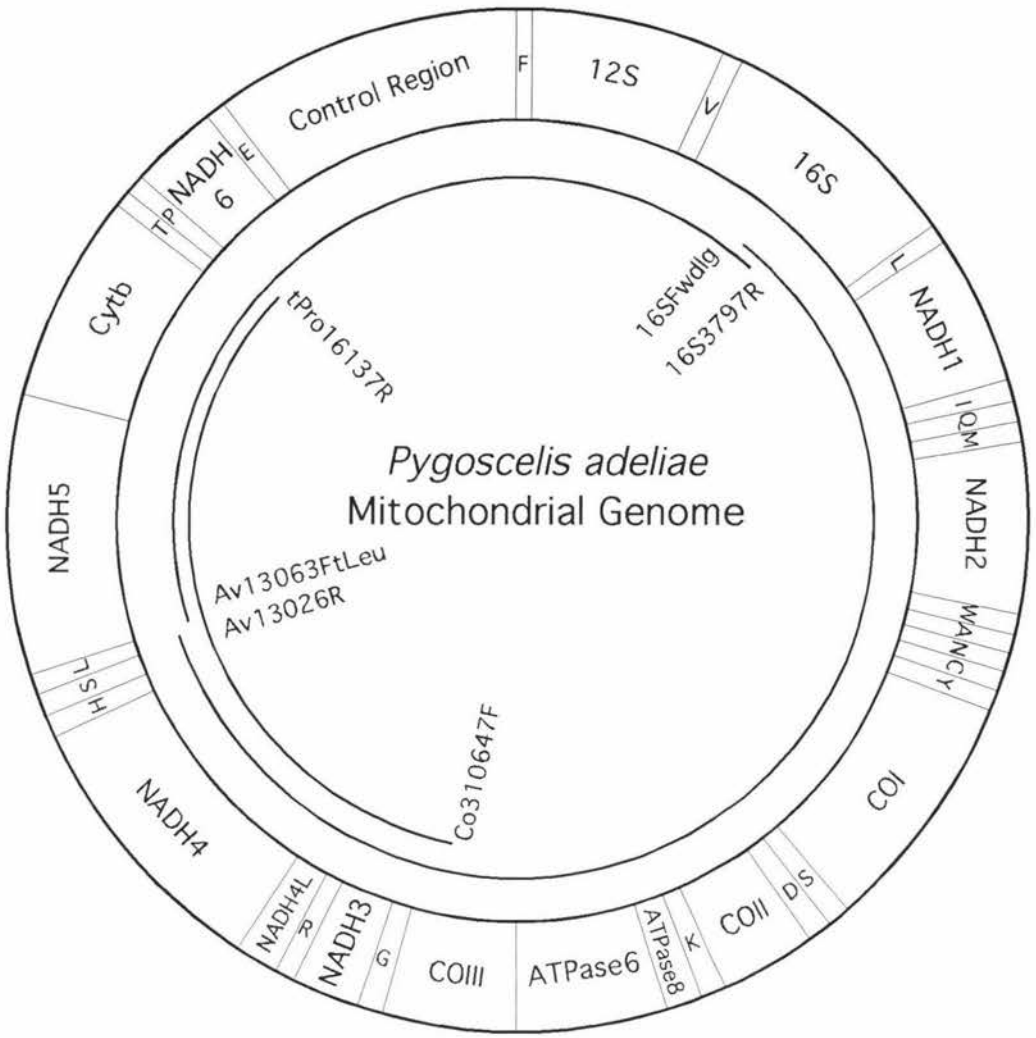


Figure 2.2 Diagram of the Adélie penguin mitochondrial genome showing location of the long-range PCR products and their primers

2.5.2 Long-range PCR of the mitochondrial genome

Mitochondrial DNA was amplified in fragments longer than 5 kb in order to avoid amplifying nuclear copies of mt DNA. Three long-range PCRs were carried out using the Expand™ Long template PCR System (Roche) (Figure 2.2), with the resulting products encompassing the whole mitochondrial genome. Primer pairs were 16SFwdlg and Av13026R, Co310647F and tPro16137R, Av13063FtLeu and 16S3797R (Appendix B Table 1). 50 µL PCR reactions contained 1x Buffer, 0.5 mM dNTPs, 3.75 U taq, 0.4 µM of each primer and 10 µL of extracted DNA. Cycling

conditions were as follows: 94°C for 2 minutes followed by 10 cycles of 93°C for 30 seconds, 54°C for 30 seconds and 68°C for 10 minutes, then 26 cycles of 93°C for 30 seconds, 54°C for 30 seconds and 68°C for 10 minutes plus 20 seconds per cycle, then one cycle of 68°C for 5 minutes. Primer pair Co310647F-tPro16137R had an elongation time of 7 minutes not 10.

2.5.3 Determination of primers for short-range PCR of the whole mt genome

Primers were from the database held by the Allan Wilson Centre for Molecular Ecology and Evolution (Massey University, Palmerston North, New Zealand). Primers in this database have been designed from conserved regions of the mt DNA genomes of mammals and birds, with 0-5 degenerate sites per primer to optimise their usefulness in diverse species. The FASTA search in the GCG program (Wisconsin Package, version 10.0) was used to search the primer database. Initially the database was searched using the mt DNA genome of the little blue penguin (*Eudyptula minor*) (Slack et al. 2003), but as sequence from the Adélie penguin was elucidated this was used to find primers to span unsequenced regions. When no primers were available in a region, new primers were designed using Oligo®4.03 (National Biosciences Inc.). The primers used in short-range PCR of the whole mitochondrial genome can be found in Appendix B, Table 2.

2.5.4 Short-range PCR from long-range products

PCR was carried out in 20 µL reactions containing 1 µL of the appropriate long range PCR product, 1x PCR Buffer (10 mM Tris-HCl, 50 mM KCl, pH 8.3) 1.25 mM $MgCl_2$, 1x Q-Buffer (Quiagen), 0.25 mM dNTPs, 0.5 µM of each primer and 0.5 U Ampli*taq* (Applied Biosystems). PCR cycling conditions in a BioRad iCycler were as follows. 2 minutes at 94°C, then 35 cycles of 94°C for 30 seconds, 50°C for 30 seconds, 72°C for 1 minute per Kb sequence then 72°C for 2 minutes. Products were run out on a 1 % MS: 1 % LE agarose gel. Size was approximated with reference to a 1 Kb ladder, and concentration by reference to a Low DNA Mass Ladder (Invitrogen™).

2.5.5 Cleanup of PCR products

PCR products were purified using DNA purification columns (HighPure™) as per the manufacturers instructions (Roche). If necessary, PCR products were excised from agarose gel before purification as described by the manufacturer.

2.6 DNA Sequencing

2.6.1 Sequencing and cleanup of the target control region sequence.

Sequencing of the PCR product was carried out directly using the BigDye Terminator Cycle Sequencing kit (V1.0 and 3.0 Applied Biosystems). 10 μL reaction mixes were used consisting of 5 μL of the PCR product, 4 μL of the BigDye terminator ready reaction mix and 1 μL of the primer AH530 (10 μM). Samples were overlaid with a drop of mineral oil before sequencing was carried out in a Hybaid Omn-E Thermal Cycler. Conditions were as per the manufacturer's instructions (25 cycles of 96°C for 10 seconds, 50°C for 5 seconds and 60°C for 4 minutes).

The sequenced products were cleaned up as follows using an isopropanol method. The products (without the mineral oil) were placed in a new 0.6 μL tube then 10 μL of H_2O and 80 μL of 75% isopropanol were added. This was mixed thoroughly and allowed to stand for at least 15 minutes. These were then spun for 20 minutes at 16 000 g and the supernatant removed, taking care not to disturb the pellet. A second wash with 250 μL 75% isopropanol was then performed and the products spun at 16 000 g for 5 minutes. The supernatant was removed and the pellet air dried. The samples were sequenced at the Massey University Sequencing Facility on an ABI Prism 337 automated sequencer (Applied Biosystems) according to the manufacturers instructions.

2.6.2 Sequencing of short-range PCR products

PCR products were sequenced using Big Dye Terminator Cycle Sequencing Kit (V3.1 Applied Biosystems). 20 μL reactions were used consisting of 4 μL 5x Sequencing buffer, 4 μL Big Dye terminator ready reaction mix, 3.2 μL of primer (1μM) and (template length/20) ng of DNA. PCR cycling conditions in a Biorad iCycler were as per the manufacturer's instructions (96°C for 1 minute, then 25 cycles of 96°C for 10 seconds, 50°C for 5 seconds and 60°C for 4 minutes). All PCR products were sequenced in both directions.

Cleanup of the sequenced products was by ethanol precipitation as follows. The products were moved to new 0.6 mL tubes, to which 2 μL 3 M sodium acetate (NaOAc), 2 μL 125 mM EDTA and 50 μL 100% ethanol was added. This was vortexed, allowed to stand for at least 15 minutes, then spun at 16 000 g for 20 minutes. The supernatant was removed immediately, taking care not to disturb the pellet. 250 μL 70% ethanol was added and spun again for 5 minutes. This was removed, and the pellet allowed to air dry. The samples were sequenced at the Massey University Sequencing Facility on an ABI Prism 337 automated sequencer (Applied Biosystems) according to the manufacturers instructions.

2.7 Analysis of DNA Sequences

2.7.1 Analysis of control region sequences

The resulting HVR I sequences were aligned in Sequencher™ 3.1.1 (Gene Codes Corp.) and checked by eye for computer error in base assignment. Those sequences used in the pedigree mutation study were aligned by family. The sequences belonging to the mother and offspring were compared for possible sequence anomalies that could be mutations. If anomalies were detected, sequences were re-sequenced to confirm or deny the anomalies.

2.7.2 Determination of lineage

Sequences were transferred from Sequencher™ 3.1.1 (Gene Codes Corp.) to PAUP*4.0b (Swofford 2002) and aligned with reference to Adélie penguin sequences of known mitochondrial lineage and Chinstrap penguin outgroups (Ritchie 2001). Sequences were designated as *Antarctic* or *Ross Sea* lineage depending on where they fell in a UPGMA tree constructed from the alignment using PAUP*4.0b.

2.7.3 Analysis of HVR I nucleotide variability

Analysis of HVR I nucleotide variability was from sequences aligned in PAUP*4.0b (Swofford 2002) using Microsoft® Excel 98 (Microsoft Corporation), CA Cricket Graph III version 1.5.1 (Computer Associates International, Inc.) and MEGA2 (Kumar et al. 2001; Kumar, Tamura and Nei 1993).

2.7.4 Alignment of whole mitochondrial genome sequences

Short range PCR products were aligned using Sequencher™ 3.1.1 (Gene Codes Corp.) and a consensus sequence created. Mitochondrial protein-coding genes encoded by the H strand were aligned at the amino acid level with reference to an alignment of genes from other species (Slack et al. 2003) using Se-Al version 1.0α1 (http://evolve.zoo.ox.ac.uk/software/Se-Al/main.html). The two rRNAs and 22 tRNAs were aligned using Se-Al 1.0α1 and checked using secondary structure models (e.g., http://www.rna.icmb.utexas.edu/ and Kumazawa and Nishida 1993).

Chapter Three

Two Mitochondrial Lineages, One Species: The distribution of *A* and *RS* Adélie Penguins in Antarctica

3.1 Introduction

The two mitochondrial DNA lineages, separated by 8.3% sequence divergence (Lambert et al. 2002), have different geographical distributions around Antarctica. The *Ross Sea* lineage is found only in the Ross Sea region, while the *Antarctic* lineage is found throughout Antarctica (Ritchie 2001; Ritchie et al. in press.). Because of this, it would be expected that the ice age refuge of *RS* lineage Adélie penguins would be in or adjacent to the Ross Sea. During the last glacial maximum, the grounding line of the West Antarctic Ice Sheet (WAIS) was just south of Cape Wheatstone (Conway et al. 1999), and fast sea ice extended further north of this, making most of the Ross Sea uninhabitable to Adélie penguins. Therefore, the ice age refuge of the *RS* lineage is most likely located somewhere further north or northwest along the coast from the Ross Sea. The location of the ice age refugia of the *Antarctic* lineage is much more difficult to determine, as this lineage is distributed throughout Antarctica. It is interesting to note, however that both lineages were well established in the Ross Sea by 6082 yr BP (Ritchie 2001). This indicates both lineages moved quickly to establish themselves in the Ross Sea as the ice retreated in the Holocene period.

While the relative abundance of *RS* lineage individuals, compared to *A* lineage individuals, is known at colonies within the Ross Sea (Ritchie 2001), this study aims

to determine the extent of the *RS* lineage at colonies in the far northwest of the Ross Sea at Cape Adare, as well as in the Balleny Islands to the north of Cape Adare, and at Port Martin in Adélie Land further west along the coast of Antarctica (Figure 3.1). By doing this, it is hoped a clearer picture of the distribution of the two lineages will appear, and information will be gained about the site of origin of at least the *Ross Sea* lineage.

There is 8.3 % sequence divergence between the *A* and *RS* lineages of Adélie penguin (Lambert et al. 2002). This is greater than the level of divergence reported between HVR I of some species and subspecies. For example, avian species in the genus *Polioptila* (gnatcatchers) are separated by between 2.02% and 3.4% sequence divergence (Ruokonen and Kvist 2002). The combination of the different geographic ranges of the two Adélie penguin lineages and the high level of sequence divergence has lead to the question of whether the two lineages may in fact represent cryptic species. In order to address this suggestion, this research aims to answer the following questions. First, do the two lineages pair together in a colony? Second, is the proportion of *A* x *RS* pairings seen the same as is expected under Hardy-Weinberg equilibrium? Third, are these pairings producing viable offspring?

Specifically, in this chapter, the following studies were carried out

1) Samples from three locations in the northwest of the Ross Sea were analysed and designated as either *A* or *RS* to determine the extent of the *Ross Sea* lineage in locations bordering the Ross Sea and the proportion of *A* and *RS* lineages at individual sites.

2) HVR I DNA sequence samples from families of Adélie penguins at Cape Bird were designated as *A* or *RS* and the proportion of *A* x *RS* pairings was calculated.

3) A selection of pairs and their offspring representing *A* x *A*, *A* x *RS* and *RS* x *RS* pairings were genotyped to confirm putative family groups identified in the field.

3.2 Results.

3.2.1 Proportions of each lineage at different locations

A total of one hundred and sixty blood samples from Cape Adare, the Balleny Islands and Port Martin were sequenced to determine the extent of the *RS* lineage outside the Ross Sea. Previous sequencing of samples in the Ross Sea region (Cape Royds, Cape Crozier, Cape Bird, Beaufort Island, Franklin Island, Inexpressible Island, Adélie Cove, Edmonson Point, Cape Wheatstone and Cape Hallett) has shown that the frequency of the Ross Sea lineage increases with decreasing latitude. In other words, the further north along the Victoria coast, the greater the frequency of the *Ross Sea* lineage (Ritchie 2001).

At Cape Adare, 62% of the 100 blood samples were *RS* lineage, and 38% were *A.* This follows the clinal trend seen along the Victoria coast, as the proportion of *RS* penguins is greater than at Capes Hallett and Wheatstone further south along the coast (Figure 3.1). In contrast, samples collected from both the Balleny Islands and Port Martin contained only one *RS* lineage at each location (30 blood samples were taken from each site), or 3% (Table 3.1). This is a sudden and dramatic decrease in the frequency of the *RS* lineage upon leaving the Ross Sea region.

Table 3.1 Location, number of samples and lineage of Adélie penguins from three locations, compared to the Cape Bird, Ross Island population. Penguin DNA from these four locations was sequenced as part of this Thesis

Location	Number of Samples	*Antarctic* Lineage	*Ross Sea* Lineage	Proportion *RS* Lineage
Cape Bird	208	162	46	22%
Cape Adare	100	38	62	62%
Balleny Islands	30	29	1	3.3%
Port Martin	30	29	1	3.3%

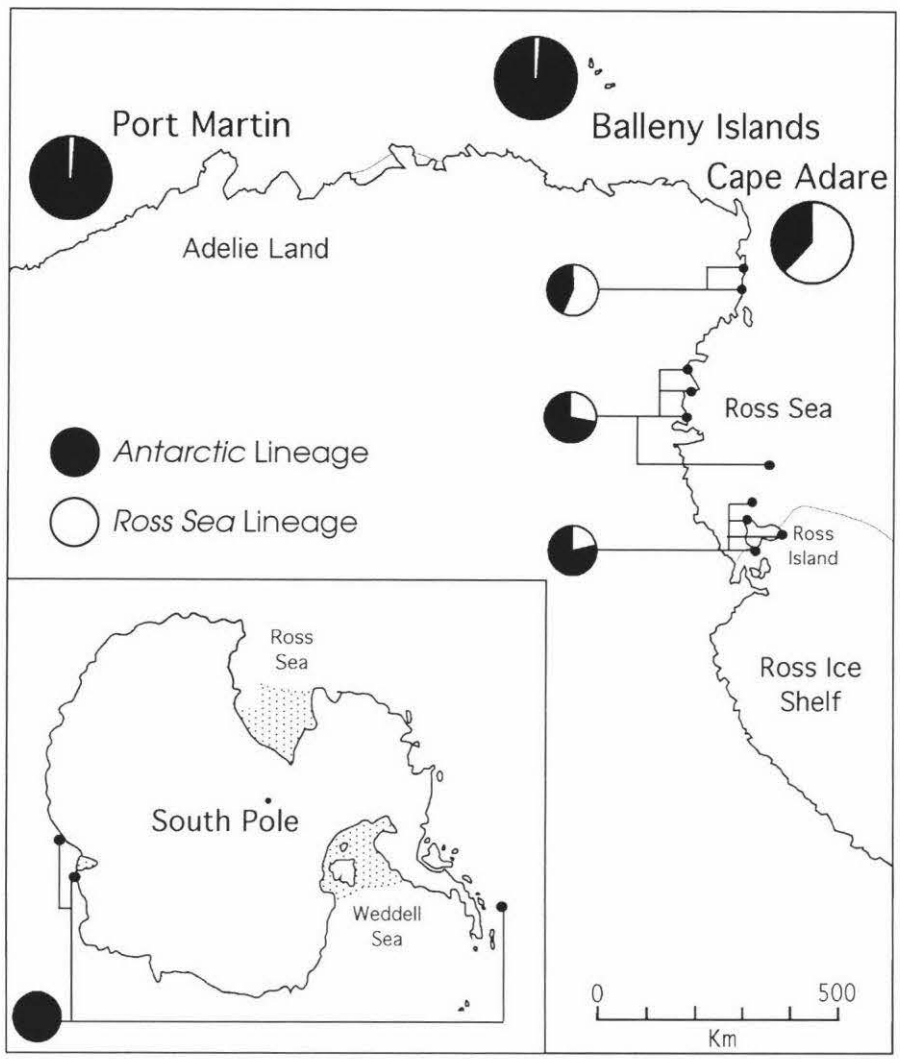


Figure 3.1 Proportions of *Antarctic* and *Ross Sea* lineages at locations around Antarctica. The four smaller pie diagrams are adapted from Ritchie (2001)

3.2.2 Family samples to determine if the two lineages are one species

Blood samples were collected from 101 Adélie pairs at Cape Bird (3 pairs were found to be incorrect families – see Table 4.1). The mitochondrial control region was sequenced, the sequences aligned in PAUP and designated as *A* or *RS* based on sequence segregation. The Adélie penguins were also sexed to determine which penguin was male and which female in each pair (see Section 4.2.1).

3.2.3 Genotyping confirms A and RS pairs produce viable offspring

Five families were genotyped to confirm the offspring were related to both parents. Of the five families genotyped, one was an *A* x *A* pairing, one *RS* x *RS,* and three families were *A* x *RS* pairings. In all five cases maternity and paternity of both *A* and *RS* parents was shown to be correctly assigned in these families (Table 3.2).

3.2.4 *Are the two lineages mating randomly?*

If Adélies of *A* and *RS* lineage were mating randomly, it would be expected that the birds would form pairs in Hardy-Weinberg equilibrium. Of the 202 birds, 156 were *A* lineage, and 46 were *RS* (77% and 23% respectively). 58 pairs were *A* x *A*, 40 pairs were *A* x *RS* and 3 were *RS* x *RS*. This gave observed frequencies of 0.574, 0.396 and 0.030 respectively. Assuming random mating, the expected frequencies for the same pairings would be 0.596, 0.352 and 0.052 respectively (Table 3.3). Chi squared tests of this data showed a 99.2 % likelihood of random pairing with respect to mitochondrial lineage. In addition, an overview of where *A* x *A*, *RS* x *RS* and *A* x *RS* pairs were nesting in the colony showed no obvious segregation of the two lineages (Figure 3.2).

Table 3.2 Microsatellite DNA genotypes of individual Adélie penguins, belonging to putative families, together with their mitochondrial lineage.
a.) Families genotyped for loci reported in Roeder et al. (2001). Results from locus AM3 are not shown as all individuals were homozygous for the same allele
b.) Families genotyped for loci reported in Haynes (2002)

a.)

DNA Sample	Sex	Mt Lineage	Microsatellite Locus				
			RM3	RM6	HrU2	AM12	AM13
286	Male	*RS*	221/225	170/172	125/125	148/148	117/125
311	Female	*RS*	221/221	170/170	125/125	148/148	117/123
93 C-16	Chick	*RS*	221/225	170/170	125/125	148/148	117/123
94 1-16	Chick	*RS*	221/225	170/170	125/125	148/148	117/117
94 2-16	Chick	*RS*	221/221	170/170	125/125	148/148	117/117
249	Male	*A*	221/221	171/171	123/125	145/148	122/126
209	Female	*A*	221/225	170/170	123/123	148/152	122/124
93 1-61	Chick	*A*	221/221	170/171	123/125	148/148	124/126
93 2-61	Chick	*A*	221/221	170/171	123/125	148/148	122/122
94 C-57	Chick	*A*	221/225	170/171	123/125	148/152	122/124
274	Male	*A*	225/227	170/170	123/123	145/148	117/122
223	Female	*RS*	221/221	170/170	125/125	148/148	123/128
93 C-17	Chick	*RS*	221/225	170/170	123/125	148/148	122/128
94 1-14	Chick	*RS*	221/227	170/170	123/125	148/148	117/123
94 2-14	Chick	*RS*	221/227	170/170	123/125	148/148	117/128

b.)

DNA Sample	Sex	Lineage	Locus		
			XVI G4	XVI H11	XVIII B2
102 T728	Male	*RS*	192/194	195/238	93/99
102 T726	Female	*A*	187/196	233/258	91/93
102 T727	Chick	*A*	187/192	233/238	91/93
102 T729	Chick	*A*	192/196	233/238	93/99
118 T784	Male	*RS*	187/194	196/237	86/88
118 T783	Female	*A*	193/200	233/237	86/88
118 T785	Chick	*A*	194/200	196/233	86/88
118 T786	Chick	*A*	187/200	196/237	86/88

Table 3.3 Observed and expected frequency of *A* and *RS* pairs

	A x *A*	*A* x *RS*	*RS* x *RS*
No. of pairs	58	40	3
Observed frequency	0.574	0.396	0.030
Expected frequency	0.596	0.352	0.052

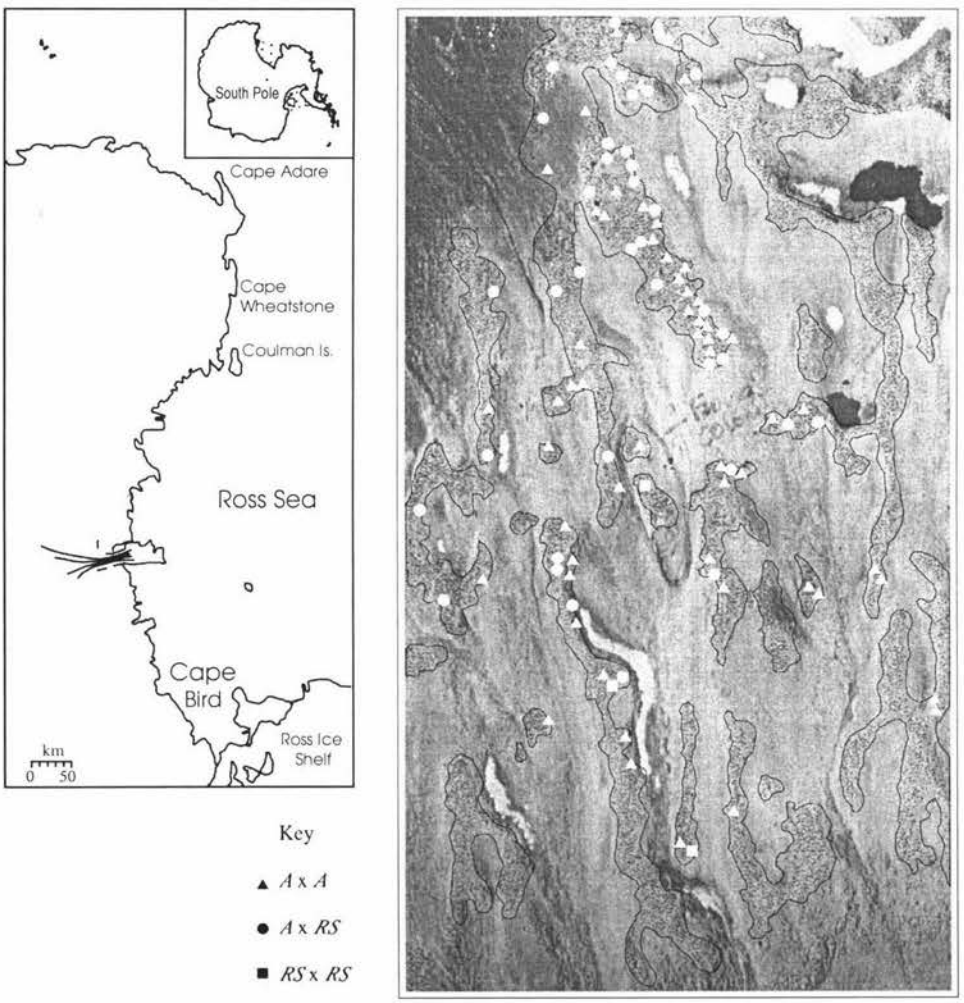


Figure 3.2 Aerial photograph showing location of nest sites sampled at Cape Bird, Antarctica. The mt lineages of Adélie penguin pairs at each nest site are shown. All other information can be found in Figure 2.1.

3.3 Discussion

3.3.1 Dramatic decrease in Ross Sea lineage upon leaving the Ross Sea region

Given the number of Adélie penguin mitochondrial DNA sequences belonging to the *Antarctic* or *Ross Sea* lineages, the highest proportion of *RS* lineage in a colony occurs at Cape Adare at the mouth of the Ross Sea. Further north and west of this location at the Balleny Islands and Port Martin the majority of Adélie penguins are *A* lineage and very few *RS* lineage penguins are seen. Around the rest of Antarctica, only *A* lineage Adélie penguins have been recorded (Ritchie 2001; Ritchie et al. in press.).

3.3.2 Implications for location of RS ice age refugia

The location of any ice age refuge for the Ross Sea lineage would have to fulfil the following criteria. Colonies require ice free beaches with plenty of small stones and pebbles for nest building. Extensive fast ice in front of the colony (more than 10 km between the shore and the sea or pack ice) must have gone by mid November, but pack ice must still be nearby, covering between about 15 to 85% of the sea area as it is required for Adélie penguin fishing (Ainley 2002). If the ice is denser than this, Adélie penguins must walk over the ice, at a speed much slower and less energy efficient than when they swim (Culik and Wilson 1991; Wilson et al. 1991). If the ice is less dense than 15% sea coverage, Adélie penguins do not thrive, as their prey is closely associated with the pack ice (Ainley 2002). In addition, pack ice must still be nearby in February as Adélie penguins inhabit the pack ice while moulting (Ainley 2002).

During the last glacial maximum, about 18, 600 yr BP, the West Antarctic Ice Shelf was grounded just south of Cape Wheatstone (Conway et al. 1999) (Figure 3.2). Sea ice would have covered the sea further north of here, as well as around much of the coast of mainland Antarctica. While it is unknown where Adélie Penguin colonies were during the LGM, it is presumed Adélie penguins lived much further north than they do now, and consequently not on the mainland. Colonies would have had to be smaller and more isolated than they are currently, due to the small number of islands

available between mainland Antarctica and more temperate landmasses. Candidate locations include South Georgia, Heard Island, South Shetland Islands, South Sandwich Islands, and Lützow-Holm Bay (Ainley 2002). Cape Adare in Victoria Land was the only ice free part of mainland Antarctica in the Ross Sea region (Baroni and Orombelli 1994), and it is interesting to note that this is also the location of the largest proportion of the *RS* lineage (Figures 3.1, 3.3).

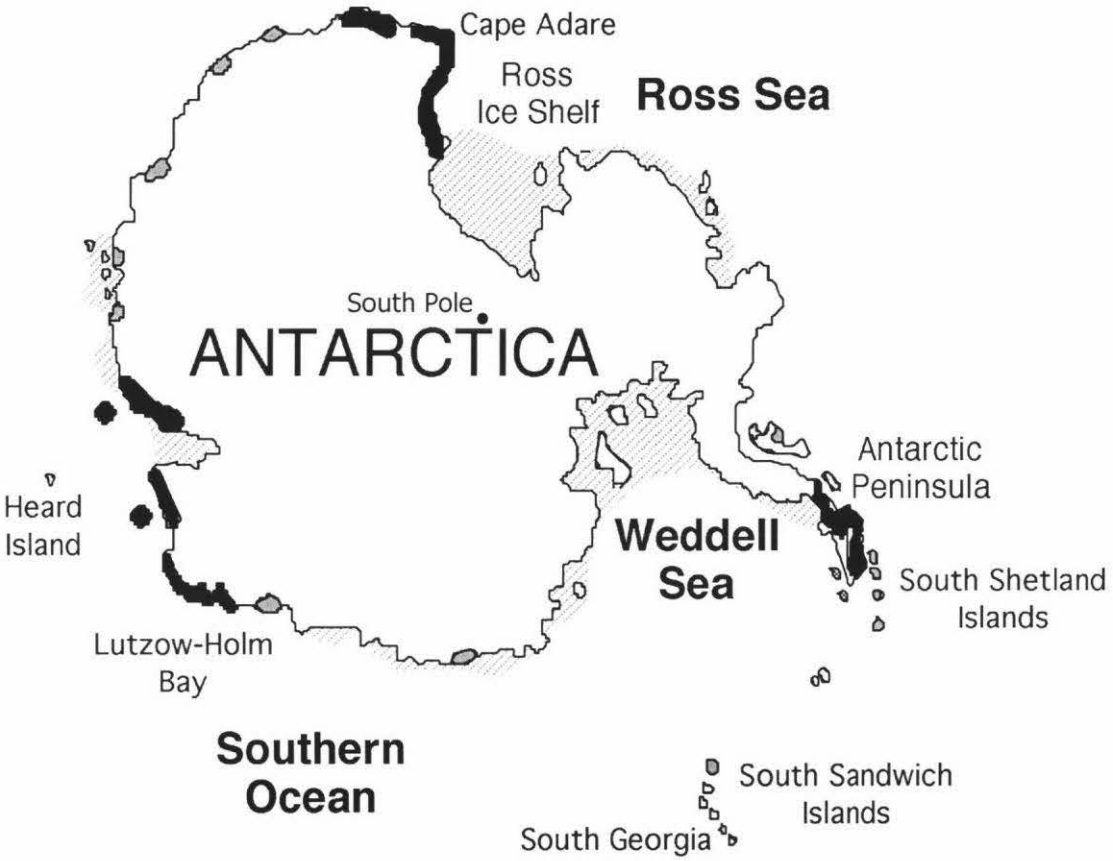


Figure 3.3 Map showing possible Adélie penguin colony refuges during the last glacial maximum (see 3.3.2 for details). Dark and medium shading represents present day high and low colony densities respectively. The light grey represents current permanent ice shelves.

3.3.3 A possible escape from the constraints of the ice age

Typical nesting behaviour of Adélie penguins includes building large pebble nests on rocky beaches in ice-free areas. Observations of penguins on Coulman Island in the Ross Sea has lead to the suggestion that some Adélie penguins may have been able to counteract some of the effects of the last ice age that covered rocky beaches in ice. Adélie penguins on Coulman Island have been observed to nest directly on the ice shelves fringing the island. Their nests are built from a very small number of stones gleaned from rocky debris left in ice-cored moraines. While these Adélie penguins still require easy access to the sea, it is possible that nesting sites were not as restricted during the last ice age as commonly assumed (D. Lambert and C. Baroni, pers. comm.). A similar situation has been observed in the Weddell Sea, where Adélie penguins nest on sea ice in close proximity to Emperor penguin (*Aptenodytes forsteri)* colonies. Emperor penguins regurgitate gizzard stones they have picked up from the ocean floor. In large colonies, there are enough stones lying around that Adélie penguins are able to use them for nesting material (Ainley 2002).

3.3.4 Fossil evidence of Adélie penguin colonies in the past

Fossilised Adélie penguin remains have been found at both current and abandoned Adélie colonies, and dating of these remains is a way to estimate the age of the penguin colonies (Baroni and Orombelli 1994). From carbon-dated bones, it is apparent that Adélie penguins recolonised the Ross Sea soon after the ice retreated. Ancient DNA from Adélie penguin bones at locations throughout the Ross Sea has been sequenced. From ancient mitochondrial control region sequences, it is clear that both the *A* and *RS* lineages were well established in the Ross Sea by 6082 yr BP. At that date, the two lineages have been found at relatively equal proportions at Terra Nova Bay in the Ross Sea (Ritchie 2001; Ritchie et al. in press.).

To date, no Adélie penguin bones have been found that pre-date the LGM. Indeed, at some sites around Antarctica, only much younger fossilised bones have been found. It is hard to know in some cases if this means Adélie penguins were not present at a location before this time, or, as is more likely, older bones have simply not been found. In the case of Cape Adare, bones have not been found that date earlier than

650 (± 90) yr BP (Baroni and Orombelli 1991), but evidence suggests Cape Adare has been ice free for the last 35 000 years (Ainley 2002). Unless some unknown factor made Cape Adare uninhabitable to penguins until about 700 years ago, it would appear this is a gross underestimate of the age of the Adare colony. In addition, population data from this study would suggest Cape Adare, or a nearby location such as Duke of York Island as the ice age refugia of the Ross Sea lineage.

If Cape Adare is indeed the ice age refugia of the *RS* lineage, it is perhaps a little surprising that so few *RS* lineage penguins have been found at both the Balleny Islands to the north and Port Martin to the west. Perhaps the relative abundance of new nesting sites opening up in the Ross Sea induced many more Adélie penguins of both lineages to migrate south at the end of the last ice age, rather than west. As has been stated before, both lineages were established in the Ross Sea relatively quickly after the end of the last ice age. Currently, 30% of the world's Adélie penguins breed in the Ross Sea (Ainley 2002).

3.3.5 Lineages are not reproductively isolated species

In the introduction of this chapter three criteria were set out as tests to determine if the two lineages of Adélie penguins separated by 8.3% sequence divergence were reproductively isolated species. The results presented in this chapter clearly refute these criteria. *Antarctic* and *Ross Sea* lineage penguins were seen to pair together in a colony, and produce chicks. In addition, *A* and *RS* Adélie penguins did not preferentially mate with their own lineage, but form a single, randomly mating population in Hardy Weinberg equilibrium. Additional studies (Haynes, Gibb and Lambert submitted) have also shown no nuclear genetic differentiation was detected between the two lineages in the Cape Adare population. There are two main possibilities why 8.3% sequence divergence in the mitochondrial DNA is not accompanied by species divergence. One, it may not be long enough since the split of the two lineages for nuclear genetic divergence to occur. Two, the two lineages are currently not geographically isolated from each other, but form single interbreeding populations in the Ross Sea region. If the geographic isolation of the two lineages from each other that is suspected to have occurred during the last glacial maximum

was still in effect today, it is entirely possible the two lineages may have evolved into two separate species.

Summary of Main Findings

1) The highest proportion of the *RS* lineage was found to be at Cape Adare. At colonies to the north and west of this location only 3% of the Adélie penguins were *RS* lineage.

2) Genetic analysis of family groupings of Adélie penguins from Cape Bird has shown the two mt DNA lineages comprise a single species, and do not represent cryptic species.

Chapter Four

The Rate of Mutation in Adélie Penguin HVR I

4.1 Introduction

Recently a rate of evolution for mitochondrial DNA hypervariable region I in Adélie penguins has been directly calculated at 0.96 s/s/Myr using serially preserved penguin bones dating back thousands of years (Lambert et al. 2002). This rate is notable in that it is approximately five times higher than previous phylogenetic estimates of avian evolution rate (Quinn 1992). Indeed, it is in much better agreement with recent estimates of mutation rate calculated by pedigree analysis in humans (Howell et al. 2003; Parsons et al. 1997). While it is interesting to compare the rate of evolution in Adélie penguins to the mutation rate in humans, it would be more appropriate to compare rates within the same species. Hence in the light of the new higher estimate of mt DNA control region evolution reported for Adélie penguins, it is desirable to estimate the rate of mutation of the HVR I in Adélie penguins by a pedigree related analysis, in order to compare it to the rate of evolution.

The results laid out in the following chapter represent the results from the first year of a three year study to determine the rate of mutation in the control region of Adélie penguins by pedigree analysis. By comparing HVR I sequences between female Adélie penguins and their offspring, preliminary estimates can be made for the rate of mutation in the mitochondrial control region of Adélie penguins.

As with approximately 50% of the world's bird species, Adélie penguins are not sexually dimorphic and it is almost impossible to distinguish females from males using external features (Ainley 2002). In order to confirm the sex of the parent Adélie penguins used in the pedigree analysis, a molecular based sexing test was developed to easily distinguish the two sexes in Adélie penguins.

In this chapter, the following studies were carried out:

1) A PCR based test to sex Adélie penguins was developed from general avian molecular sexing tests

2) Hypervariable region I was sequenced for Adélie penguins belonging to family groups identified in the field. These sequences were aligned by family and scanned for mutations

3) Based on the results from the Hypervariable region I sequences, a preliminary rate of mutation was calculated for HVR I

4.2 Results

4.2.1 Sexing of family samples by 3 PCR based methods

Three primer sets were trialed in sexing Adélie penguins of known sex (Figure 4.1a). The primer set K1/K2 (Huynen, Millar and Lambert 2002) produced no easily discernable difference in product length between male and female Adélie penguins. Primer sets P2/P8 (Griffiths et al. 1998) and K7/W1 (Huynen, Millar and Lambert 2002) produced PCR products that were discernibly longer in females than in males, however the difference was only marginal. P2/P8 products contained a digestion site for the restriction enzyme *Hae* III that selectively cut a fragment from the Z chromosome version. This resulted in two clearly visible products in females and one in males. The P2/P8 primer set, with subsequent digestion was consequently chosen to sex birds in the family study (Figure 4.1b).

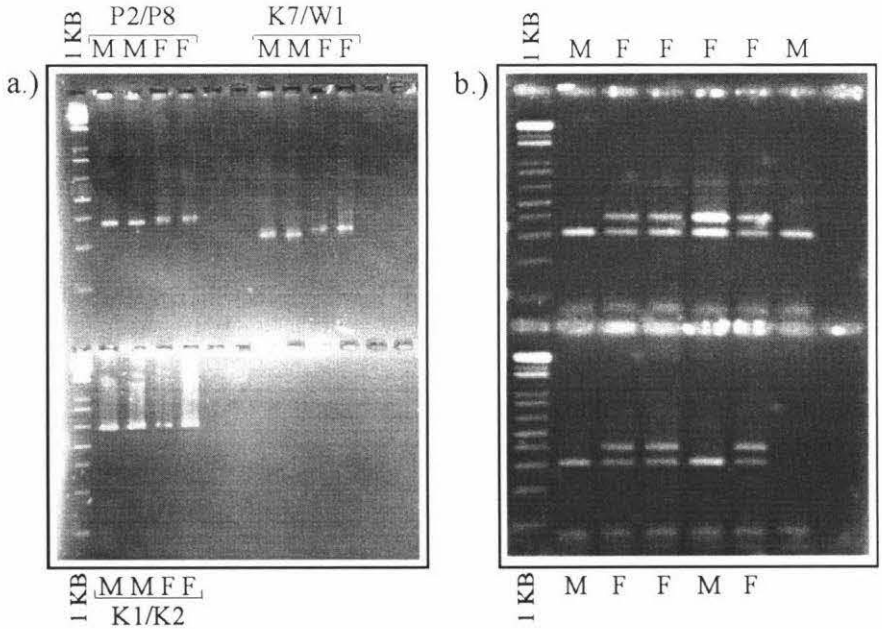


Figure 4.1.
a.) Gel picture showing sexing trials of three primer pairs using Adélie penguins of known sex.
b.) After digestion with *Hae* III, primer pair P2/P8 produces two bands in females and one in males. 1 KB = 1Kb plus DNA ladder (GibcoBRL)

4.2.2 Pedigree analysis

The mitochondrial control region was sequenced in 402 Adélie penguins from 104 families and the sequence of the offspring compared to the mother, as mt DNA is maternally inherited (Figure 4.2). Ninety-nine of these families were sampled in the 2001/2002 summer season, and five families were from samples collected in the summers of 1993 and 1994 (Fiona Hunter, pers. comm.). All birds were from the Cape Bird colony, Ross Island. In three families the chick's sequence contained multiple substitutions when compared to the mother, and these cases were assumed to represent incorrect determination of maternity. One of these families had one chick, and as such the presence of mutated mt DNA could not be confirmed. As the other two families had two chicks with HVR I sequence identical to each other, these four mt DNA transmissions are assumed to contain no mutations. The 101 remaining families resulted in 187 transmissions of mitochondrial DNA from mother to chick. In 98 of these families, the HVR I sequence of at least one offspring was identical to that of the mother. In the other three families, while the mitochondrial sequence was identical between mother and chicks, single point heteroplasmy was detected in both generations (Table 4.1).

Table 4.1 Summary of pedigree analysis showing details of mitochondrial DNA transmissions

Type of family	No. of families	No. of putative transmissions	
		Confirmed	Denied
All chicks' sequences identical to mother	96	179	-
Chicks essentially identical with single point heteroplasmy	3	6	-
One chick same as mother, other different	2	2	2
All chicks' sequences different to both parents	3	4	1

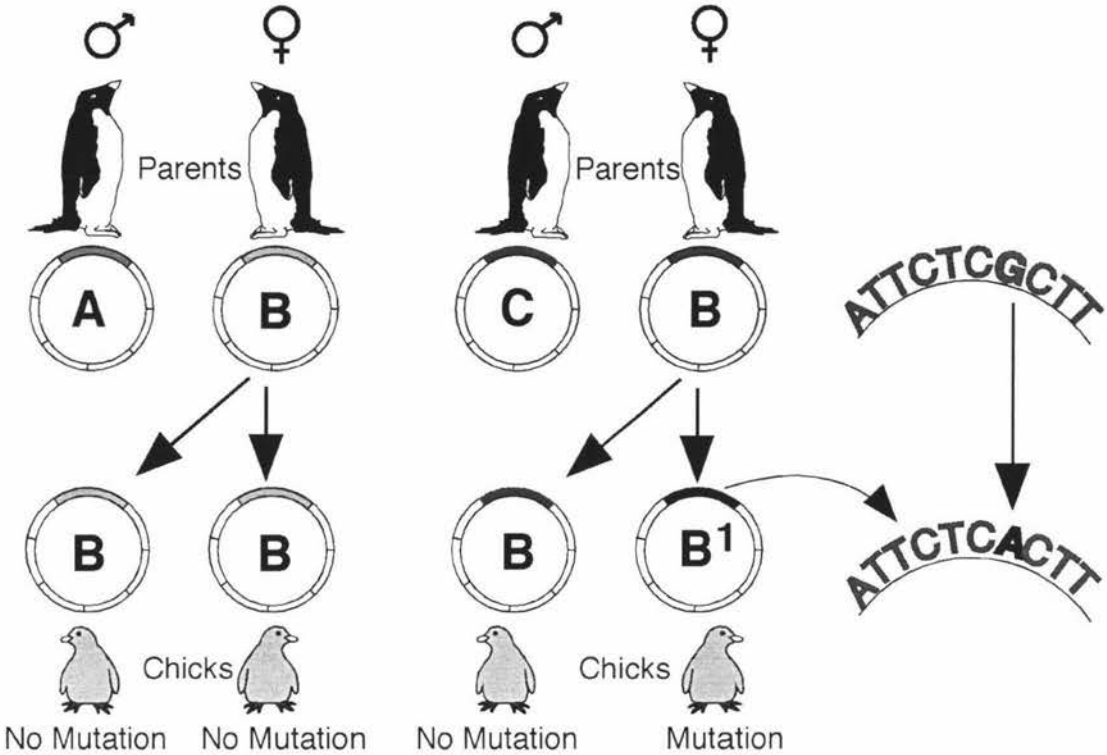


Figure 4.2 Diagram outlining the basis of pedigree analyses in detecting mutations. On the left, the chicks' sequence is identical to the mother's and no mutation is detected. On the right, one chick's sequence differs from the mother's at a single nucleotide location, revealing a putative mutation.

4.2.3 Verification of maternal transmission of mt DNA

In 104 family samples, there were 194 putative transmissions of mitochondrial DNA from mother to chick (Table 4.1). In seven cases at five nests, neither of the adult birds identified at the nest was the mother of one or both of the chicks. Where a male and a female adult bird was identified on the nest, but the mitochondrial sequence of the chick differs from that of the resident female, this could be the result of one of three factors. First, there was human error in assigning an adult to a nest, second there was egg dumping by one female in another's nest and third, a chick may have moved from one nest into a neighbour's in the initial stages of crèching. In three families with incorrect parentage, two with two chicks and one with one chick (i.e., five transmissions of mt DNA), the chicks' mitochondrial DNA sequence contained multiple differences when compared to the adult birds (10 to 43 differences over 560 bp). At one of these nests both adult birds were sexed as male and is likely to be an example of incorrect determination of parentage in the field. Four out of these five transmissions can be used in the mutation rate estimation, as siblings are available for comparison. The other two cases of incorrect parentage involved families with two chicks, where one of the chicks had HVR I sequence that differed at multiple positions to the resident female's HVR I while the second chick's sequence was identical to that of the mother.

4.2.4 Inherited single point heteroplasmy detected in three families

In three families, putative heteroplasmy was detected at three different single base loci (Figure 4.3). In all three cases, the heteroplasmy was seen in both the mother and the two chicks. All cases of heteroplasmy were resequenced in both directions from new PCR product amplified with high fidelity Taq to confirm the double peak in the electropherogram. Although these families were not cloned to confirm the presence of two mitochondrial types, other Adélie penguins with equivalently patterned electropherograms have been cloned and heteroplasmy confirmed (P. Ritchie, pers. comm.).

a.) b.) c.)

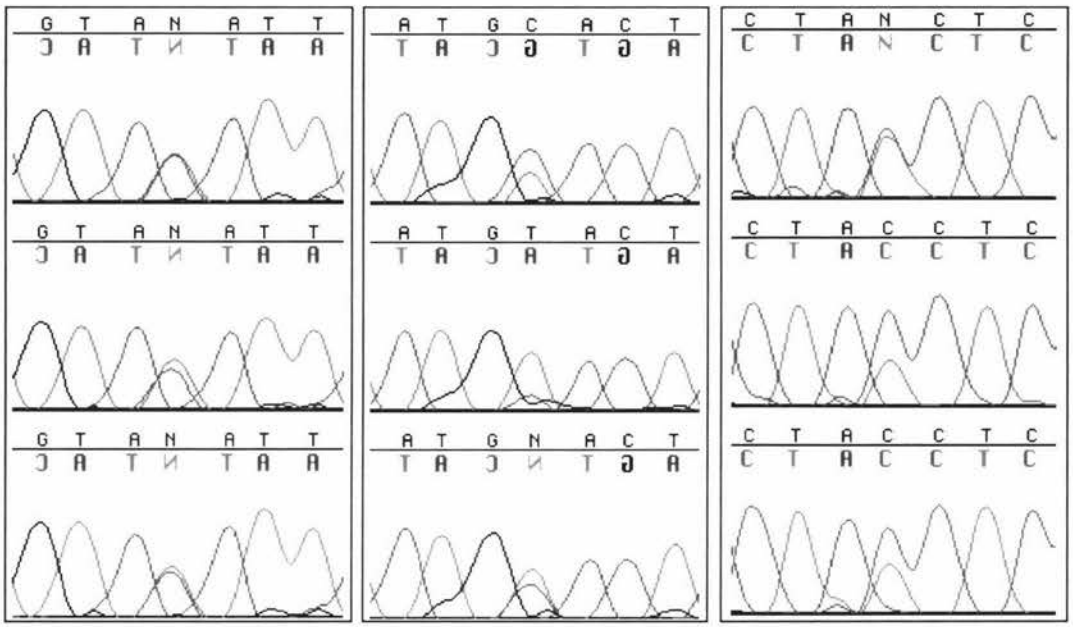


Figure 4.3 Heteroplasmy detected in three families. In each case the mothers sequence is first, and the two chicks below.
a.) Nest 116, relatively equal proportions of T and C were seen at site 15626 of the genome in both the mother and the two chicks, as determined from peak heights on the electropherogram.
b.) Nest 139, single point heteroplasmy was detected at site 15645 between nucleotides T and C in both the chicks and the mother. In the mother, C was present in significantly higher levels than T, while in one chick T was present in higher levels than C. In the other chick levels of T and C were relatively equal.
c.) Nest 178, significant levels of T and C were seen at site 15975 of the genome with similar levels of T and C in the mother, and C slightly higher than T in both chicks.

4.2.5 Calculating a preliminary rate of mutation in Adélie penguins

The samples collected for this family study are the first of a three year study. However initial analyses can be performed, and preliminary estimates made for the rate of mutation in HVR I. Mitochondrial control region sequences were analysed in 402 Adélie penguins from 104 families. These 560 bp sequences corresponded to 191 generations; i.e. 191 transmissions of mitochondrial DNA from mother to offspring. No fixed substitutions were found in the 191 transmissions, however three cases of transmitted heteroplasmy were observed.

If a rate of mutation is calculated that includes the three cases of putative heteroplasmy, the rate is 5.610 s/s/Myr. This is much higher than previous estimates of substitution rate. Moreover, the heteroplasmy was seen in both the mother and the chicks. This indicates that the mutation event that lead to the heteroplasmy occurred in a previous generation. There is no way of knowing how many generations ago the mutation event occurred, and as such any rate calculated using these cases of heteroplasmy is upwardly biased by such a low number of generations. Therefore, it would be unwise to include these heteroplasmic events in a calculation of a rate of mutation.

Given no substitutions in 191 transmissions over 560 bp, it could be said the rate of mutation in the control region of the Adélie penguin is 0 s/s/Myr. However, the number of generations (or transmissions) used to calculate this rate is very low. Mutations are known to occur; therefore more generations need to be sampled in order to detect mutations. As such, it would perhaps be better to use these figures to calculate an 'upper limit' for the rate of mutation. Using a Poisson distribution, there is 95% confidence that the current upper limit for the rate of mutation in Adélie penguins is 5.602 s/s/Myr (Figure 4.4). As the number of generations sampled increases, this upper limit will decrease.

4.2.6 How many samples need to be sequenced in order to obtain a rate of mutation in the realm of previous studies?

The rate of evolution in Adélie penguins has recently been estimated at 0.96 s/s/Myr (95% HPD 0.53-1.43) (Lambert et al. 2002). If the rate of mutation is equal to the rate of evolution, as expected by Kimura's neutral theory (Kimura 1983), how many generations must be sequenced to exclude 0 from a 95% confidence interval around 0.96 s/s/Myr? Using a Poisson distribution with a 95% confidence interval, more than 1115 generations (i.e. 1115 transmissions of mt DNA) are needed.

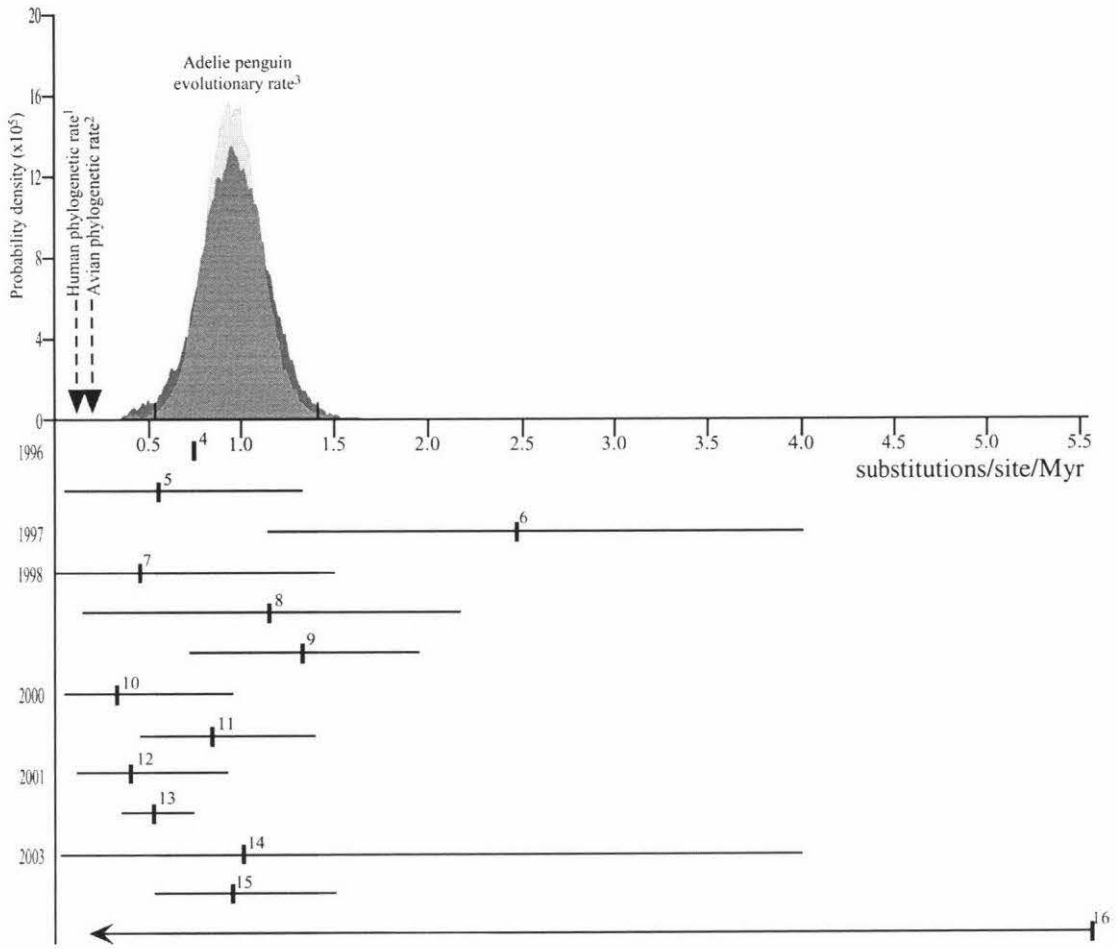


Figure 4.4 Graphic summary of pedigree studies' estimates of the rate of mutation. This is compared to the phylogenetic rates of evolution calculated in birds and humans, and the new evolution rate recently calculated for Adélie penguins. The posterior probability density graphs of the Adélie penguin evolutionary rate are adapted from Lambert et al. (2002). Pale grey shading represents the 95% highest posterior density (HPD) interval of the Adélie penguin evolutionary rate. Horizontal bars represent 95% confidence intervals unless stated otherwise, with vertical bars representing midpoint values. For more details, see Table 1.1. 1 Stoneking (1992); 2 Quinn (1992); 3 Lambert (2002); 4 Howell (1996) no confidence intervals given; 5 Bendall (1996); 6 Parsons (1997); 7, 8 99% CI's Jazin (1998); 9 Parsons (1998); 10, 11 (Sigurðardóttir et al. 2000); 12, 13 (Heyer et al. 2001); 14, 15 99.5% CI's(Howell et al. 2003); 16 Present study.

4.3 Discussion

4.3.1 The rate of mutation in Adélie penguins

Currently, the rate of HVR I mutation in Adélie penguins has an upper limit of 5.602 s/s/Myr. This was calculated after finding no mutations in 191 transmissions of mt DNA. In addition, it was calculated that more than 1115 transmissions are needed to exclude 0 from a 95% confidence interval, if the actual rate of mutation is equal to the rate of evolution calculated from serially preserved ancient DNA. Although the present study represents approximately 20%of the transmissions needed to be tested, the number of transmissions of mt DNA required in the Adélie penguin study is higher than that published in many of the human pedigree studies (Table 1.1).

There are a number of reasons for this discrepancy in required sample size between studies. A major reason lies in the length of a generation. In order to compare mutation rates between species, the rate must be based on a common time scale, such as the million years used here. The original pedigree rate is calculated per generation, and then converted to a rate per million years. A human generation is calculated at between 20 and 30 years (the discrepancy in some human studies is based on the number of years chosen), while an Adélie penguin generation is about five years (Ainley 2002). Because of this difference in generation length, 4-6 more Adélie penguin generations are needed to get a similar rate per million years.

In addition to requiring more transmissions of mt DNA, a study of mutation rate in Adélie penguins also requires more sequences. This is because only the immediate maternal generation is known. Many human pedigree studies only sequence a small number of individuals, as genealogical records allow relationships to be traced back a number of generations. For example, Sigurðardóttir (2000) sequenced the mt control region in 272 individuals related by 705 transmissions of mt DNA. In contrast, DNA from 402 Adélie penguins was sequenced, resulting in 191 mt DNA transmissions. This is because it is essentially impossible to know Adélie penguin grandparents or cousins; only the immediate mother-chick transmission can be known.

4.3.2 Heteroplasmy – to include or not to include?

An important part of the estimation of mutation rate is the presence of heteroplasmy. It must be remembered that mutations in mitochondrial DNA occur at the level of the individual mitochondrion, and there are approximately 200 000 per mammalian oocyte (Bendall et al. 1996). However, the fact that the mitochondrial sequence of an individual can be seen to change in one generation implies very rapid genotype shifts in humans. These observations seem to indicate the presence of a relatively strong bottleneck effect between generations (Bendall et al. 1996).

In this study of 191 transmission events in Adélie penguins, three cases of putative single point heteroplasmy were detected in three different lineages. Length heteroplasmy in the interrupted poly C sequence, as well as in other sites of the mt genome, has also been detected in the Adélie penguin (data not shown and P. Ritchie, pers. comm.). The point heteroplasmy seen in this study was present in varying levels in both mother and offspring. In two of the cases, proportions of each nucleotide as determined from electropherogram data appeared stable, while in the other case levels varied but didn't disappear. This may indicate the presence of a less severe bottleneck effect than that seen in humans, where the presence of constant heteroplasmy levels seen over generations has not been widely observed (Parsons et al. 1997; Sigurðardóttir et al. 2000).

Different studies have chosen to include and exclude observed heteroplasmy in their estimation of mutation rate. Parsons et al. (1997) detected one case of heteroplasmy in an Amish lineage where the grandmother appeared fixed for C and the grandchild fixed for T. Examination of the mother and siblings (of the grandchild) showed the mother was heteroplasmic at approximately 0.7T: 0.3C, and the five siblings had all reverted to C. This heteroplasmic result was included in their estimate of mutation rate. Further low levels of heteroplasmy were suggested in other lines, but firm conclusions could not be drawn on the basis of electropherogram data.

Sigurðardóttir et al (2000) however argue for a conservative estimate of mutation rate and the omission of heteroplasmy in its calculation. In their study of 26 large Icelandic pedigrees, Sigurðardóttir et al. found three substitutions and three cases of

heteroplasmy in 705 mt DNA transmission events. Sigurðardóttir et al. argued against the inclusion of the instances of heteroplasmy in calculating their mutation rate, because they could not prove that the heteroplasmy had not occurred as the result of somatic mt DNA mutations during the life of the individual. In other words, although the individuals display somatic heteroplasmy, they may be homoplasmic in the germ line. Furthermore, they argue that because the processes by which heteroplasmy is resolved in future generations is still not well understood, it should be removed from calculations of mutation rate. Because of this, their estimate of mutation rate is at the lower end, however they concede theirs is an underestimate as heteroplasmy may resolve in favour of the 'new' base.

In all three cases of putative heteroplasmy detected in this study of Adélie penguins, heteroplasmy was seen in varying proportions in both the mother and the offspring. This indicates that the mutation leading to heteroplasmy occurred in the germ line, and is not a result of somatic mutation during the life of the individual (as postulated by Sigurðardóttir et al. 2000). This is also the case in other studies such as that of Parsons et al. (1997) who detected varying levels of single point heteroplasmy in different generations of the same family. Because the heteroplasmy was seen in both generations of Adélie penguins, it has not been included in mutation rate estimation in this study as the mutation that caused the heteroplasmy would have occurred in an earlier generation. Until the detection of heteroplasmy improves, and an understanding of its resolution in future generations in better understood, it is perhaps more prudent to exclude it from calculations of mutation rate.

Another possible cause of observed sequence change or heteroplasmy is mt DNA recombination. It has been suggested that mutations could be the result of rare recombinations between maternal and paternal mt DNA (Eyre-Walker, Smith and Smith 1999). Stoneking (2000) presents logical arguments to refute this possibility. The mutations seen in pedigree analyses are almost all single point mutations. Most paternal and maternal mt DNA sequences differ at several nucleotide positions in the control region, so it would have to be an unusual method of recombination to only ever include one different nucleotide. The recombination would always have to occur over short stretches - single genes, or perhaps even single nucleotide recombinations to produce this pattern. This same argument can be used to refute the suggestion that

heteroplasmy is also a result of paternal and maternal mitochondria in one individual (paternal leakage).

4.3.3 No evidence for paternal transmission of mitochondrial DNA

The data presented in this study support a number of the points made by Stoneking (2000). In all seven cases of the Adélie penguin chicks' mitochondrial DNA differing significantly from that of the resident female, not once did the chicks' sequence match that of the resident male. Even if the resident male were not the father, it is much more likely that the chick's mitochondrial sequence is the same as its absent mother than the same as its absent father. In other words, in 104 families with 194 putative transmissions, no evidence of paternal mitochondrial transmission was detected.

4.4 Summary of Main Findings

1) No mutations were found in 191 transmissions of mt DNA. Based on this finding, the rate of mutation in Adélie penguins is estimated at less than 5.602 s/s/Myr.

2) Three cases of heteroplasmy were found, but not included in the calculation of a rate of mutation as it was seen in the DNA of both generations sequenced. This may indicate the presence of a less severe bottleneck effect than seen in humans where transmission of heteroplasmy is not detected very often (Sigurðardóttir et al. 2000).

Chapter Five

A comparison of the two lineages with respect to nucleotide variation and diversity in HVR I

5.1 Introduction

Many blood samples have been collected from Adélie penguins of both mitochondrial lineages from various locations around Antarctica and the sequence of hypervariable region I (HVR I) determined (see Chapters 3 and 4). This gives the unique opportunity to analyse nucleotide variability along HVR I in relation to both the *Antarctic (A)* and *Ross Sea (RS)* lineages and geographical location. Hypervariable region I is the most variable region of mt DNA in Adélie penguins (Ritchie 2001). The two lineages are separated by 8.3% sequence divergence (Lambert et al. 2002). The lineages diverged approximately 75 kya, and each lineage expanded about 32 kya (Ritchie et al. in press.), consistent with the glacial history of Antarctica.

This chapter is an analysis of the nucleotide substitution variation seen in hypervariable region I of the mitochondrial control region. In particular, this chapter seeks to explore the following questions:

1) Does the pattern of nucleotide variation along HVR I differ between the *A* and *RS* lineages?

2) Are there differences in nucleotide variation among populations within each lineage?

3) Is there a pattern to the nucleotide variation, or is it random?

4) Can mutational hot spots be detected in HVR I?

To explore these questions, the following studies were carried out:

1) Variation along HVR I was analysed in the two lineages, and among populations within each lineage. Non-majority plots were used as a basis for this analysis.

2) The number of variable sites in each lineage was analysed in 10 and 50 bp intervals to look for patterns in the regions of HVR I that vary.

3) The locations of the nucleotide sites that define the *A/RS* lineage split, regions of HVR I with known function and the location of heteroplasmic sites determined in Chapter 4 were also examined as part of the analysis.

5.2 Results

5.2.1 Summary statistics

In the present study, a total of 368 adult Adélie penguins from four locations around Antarctica were sequenced for the 560 bp HVR I. Specifically, 30 individuals were from Port Martin (29 *A*, 1 *RS*), 30 from the Balleny Islands (29 *A*, 1 *RS*), 100 from Cape Adare (38 *A*, 62 *RS*) and 208 from Cape Bird (162 *A*, 46 *RS*). Chicks from Cape Bird were excluded, as their sequences were identical to their mothers'. 23 individuals from the Antarctic Peninsula, Mawson and Davis that had been sequenced previously (Ritchie 2001) were added to the dataset. These three populations were combined for analysis as all comprised only *Antarctic* lineage Adélie penguins.

Base compositions for a 352 bp portion of the HVR I L-strand have previously been reported (Ritchie 2001). Base frequencies for the complete HVR I were calculated to compare to the shorter 352 bp region reported previously. To compare the two lineages, base frequencies for the complete 560 bp HVR I were calculated for the L-strand of each lineage (Table 5.1).

Table 5.1. Average base composition (%) of the Adélie penguin HVR I

	A	T	C	G
Antarctic	29.7	29.7	23.1	17.5
Ross Sea	30.0	29.0	23.8	17.2
Combined	29.8	29.5	23.3	17.4
Previous 352 bp (Ritchie 2001)	30	31	20	19

5.2.2 Heteroplasmic sites

Three families sequenced as part of the pedigree analysis in chapter 4 were found to have single point heteroplasmy. The three nucleotide positions where heteroplasmy was found are sites 60, 79 and 409 of the control region (positions 15626, 15645 and 15975 of the genome). All three are *Antarctic* lineage and fall on sites where no variation was found in the other individuals sequenced as part of this thesis.

5.2.3 The sites that define the A/RS lineage split.

There are 29 sites in HVR I that define the split of the two mitochondrial lineages, however there are no fixed differences between the two lineages. These sites are positions 15627, 15651, 15652, 15660, 15710,15717, 15729, 15738, 15740, 15746, 15750, 15751, 15805, 15849, 15850, 15852, 15876, 15900, 15925, 15926, 15930, 15949, 16030, 16032, 16034, 16035, 16041, 16069 and 16081 of the genome. Eight of these sites fall within the 352 bp region already analysed and as such have been reported previously (Lambert et al. 2002; Ritchie 2001). These sites are shown in Figure 5.1.

5.2.4 Non-majority plot analysis

Non-majority plots measure the proportion of the non-majority nucleotide at each base position in a set of sequences. By definition, the less frequent nucleotide cannot have a proportion greater than one half. Non-majority plots comparing the *A* and *RS* lineages appear to show a different pattern of nucleotide variation across hypervariable region I (Figure 5.1). A comparison of nucleotide variation within each lineage in different populations around Antarctica appears to show similar variation regardless of location (Figure 5.2). The larger bars in the graphs represent bases where there are two variants with significant proportions in a population. Here it can be seen that more large bars are at different nucleotide sites when comparing the two lineages, and are at the same nt sites when comparing one lineage at different locations around Antarctica.

Most variable sites in HVR I have only two variants, for example A and G, or C and G. A very small proportion of variable sites have three variants, e.g. A, T and C. In the Antarctic lineage (n=281), 122 sites varied in the 560 bp HVR I, and two of these had three variants (positions 15786 and 15965 of the mt genome). In the Ross Sea lineage (n=110) 120 sites varied and three of these had three variants (positions 15730, 15912 and 15953). There does not appear to be a pattern to the location of these sites (data not shown). In all cases only one sample had the third variant. No site had all four variants.

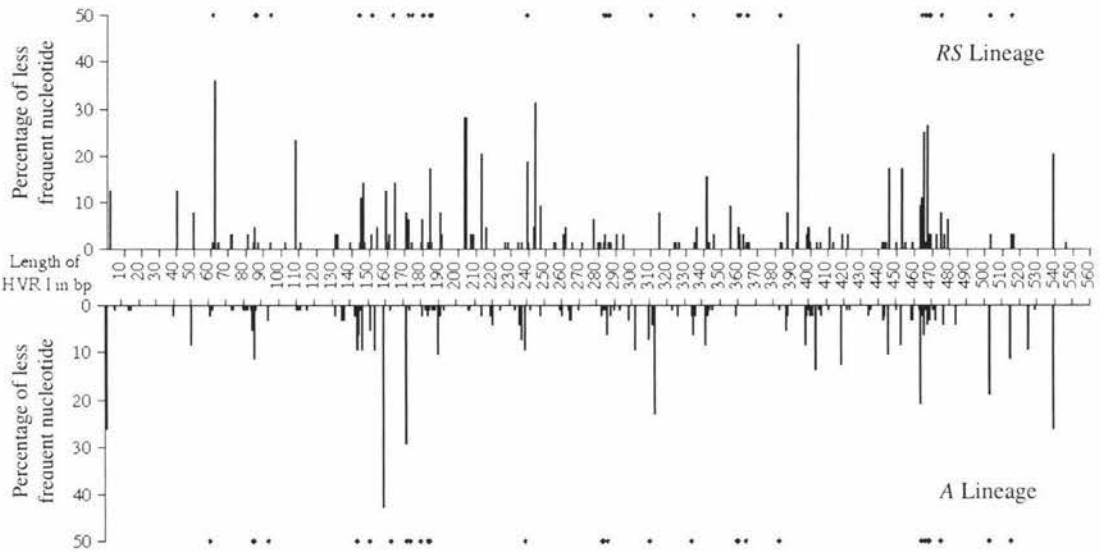


Figure 5.1. Non-majority plots of the *A* and *RS* lineages show a different pattern of nucleotide variation across HVR I in the two lineages. The small crosses along the top and bottom of the graph represent the sites that define the *A/RS* lineage split

Figure 5.2. (Overleaf) Non-majority plots of the two lineages from different regions around Antarctica. *A* lineages from various locations are more similar to each other than to *RS* lineages from the same region and vice versa

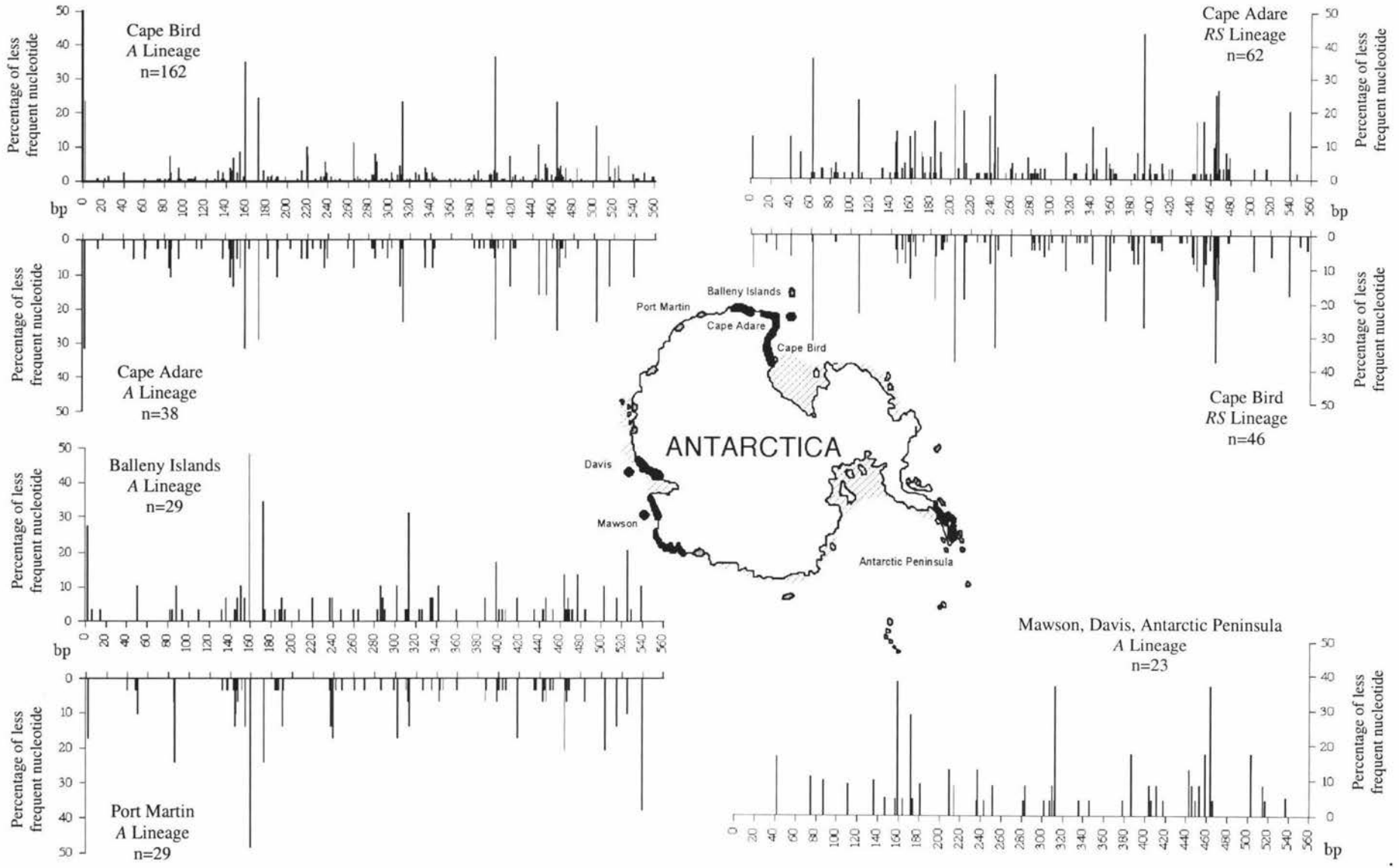
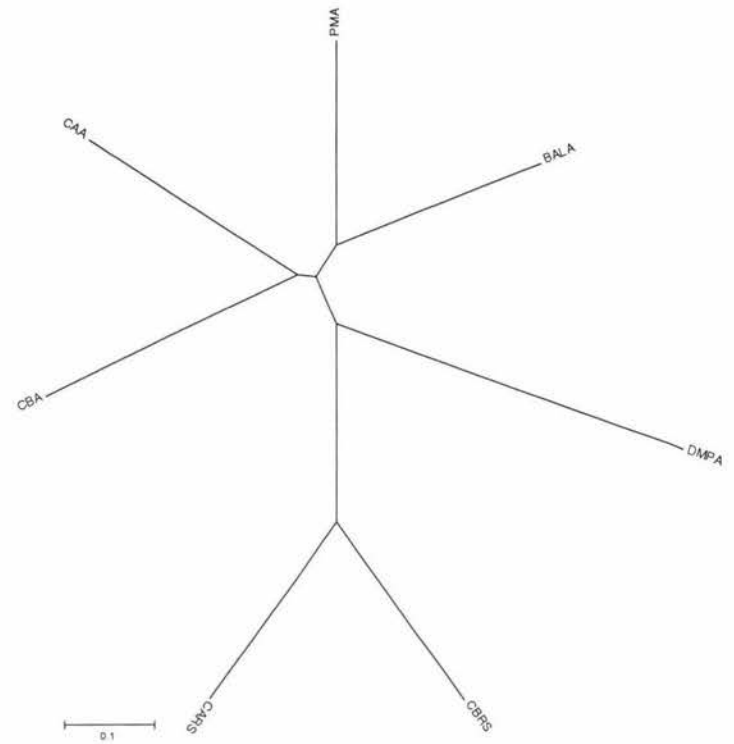

Cape Bird
A Lineage
n=162
Cape Adare
A Lineage
n=38
Balleny Islands
A Lineage
n=29
Port Martin
A Lineage
n=29
Cape Adare
RS Lineage
n=62
Cape Bird
RS Lineage
n=46
Mawson, Davis, Antarctic Peninsula
A Lineage
n=23
Percentage of less frequent nucleotide
bp
Balleny Islands
Port Martin
Cape Adare
Cape Bird
ANTARCTICA
Davis
Mawson
Antarctic Peninsula

Table 5.2 enumerates the pattern seen in Figures 5.1 and 5.2. Pairwise comparisons between lineages and populations were carried out using those sites where the non-majority nucleotide was greater than 5%. From the table it can be seen that a much higher proportion of variable sites are shared within the *Antarctic* or *Ross Sea* lineage than between them, even at the same geographic location. This pattern disappears when all sites rather than just those over 5% are included (Table 5.2 lower diagonal). Cluster diagrams of the data in Table 5.2 show the separation of the two lineages for sites with greater than 5% variation (Figure 5.3a), while separation is much less clear when all sites are added (Figure 5.3b).

Table 5.2. Pairwise comparison of variation within and between populations of the *A* and *RS* lineages.
Upper diagonal: The number of variable sites over 5% in a non-majority plot that are shared between two populations, as a proportion of the total variable sites over 5% in those two populations.
Lower diagonal: The same as upper diagonal, except using all sites, not just those over 5%.
Abbreviations: CB, Cape Bird; CA, Cape Adare; PM, Port Martin; BAL, Balleny Islands; DMP, Davis, Mawson and Antarctic Peninsula; *A, Antarctic* lineage; *RS, Ross Sea* lineage. Shaded cells are the pairwise comparisons between lineages. Sample size is given in brackets after each location.

	CB *A* (162)	CB *RS* (46)	CA *A* (38)	CA *RS* (62)	PM *A* (29)	BAL *A* (29)	DMP *A* (23)
CB *A*	-	**0.120**	0.415	**0.104**	0.371	0.416	0.257
CB *RS*	**0.304**	-	**0.169**	0.511	**0.222**	**0.190**	**0.127**
CA *A*	0.389	**0.284**	-	**0.197**	0.447	0.392	0.231
CA *RS*	0.357	0.382	**0.338**	-	**0.235**	**0.200**	**0.113**
PM *A*	0.327	**0.272**	0.411	**0.331**	-	0.525	0.227
BAL *A*	0.300	**0.258**	0.342	**0.308**	0.489	-	0.213
DMP *A*	0.219	**0.164**	0.257	**0.187**	0.226	0.234	-

a.)



b.)

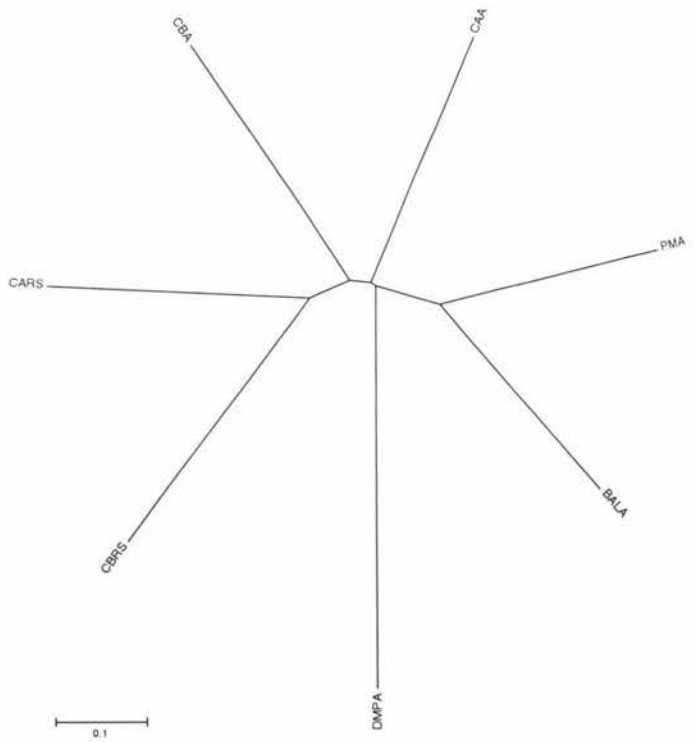


Figure 5.3 Cluster diagrams of the data in Table 5.2
a) The relationship of non-majority sites over 5% between different populations.
b) The relationship of all non-majority sites in different populations
Abbreviations are the same as in Table 5.2

5.2.5 *Distribution of variable sites in hypervariable region I*

The number of variable sites in non-overlapping 50 bp intervals was plotted for each lineage in order to analyse the distribution of variable sites along HVR I. From Figure 5.4 the two lineages can be seen to have a very similar distribution of polymorphic sites per 50 bp across HVR I. Both show a distinctive decrease in variation for the first and last 50 bp in HVR I, from an average of 12.5/50 variable sites in the central portion to 4.25/50 variable sites at the beginning and end of HVR I.

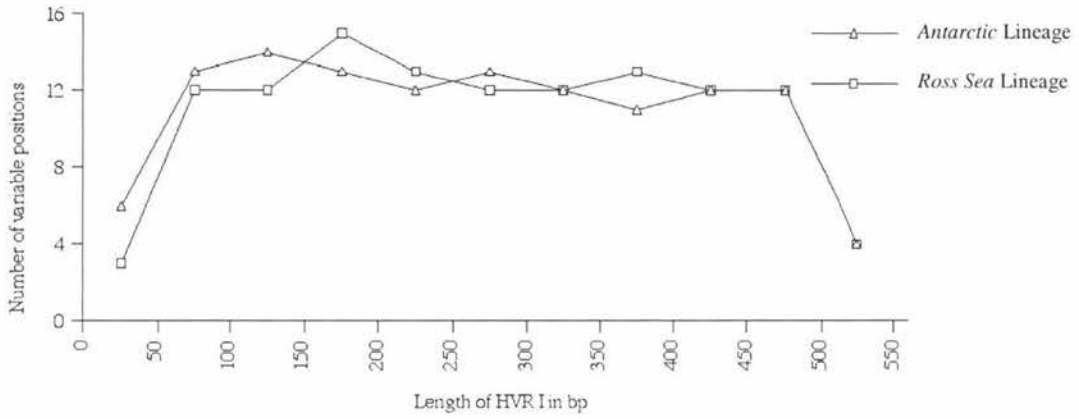


Figure 5.4 Distribution of the variable sites in HVR I. The number of variable sites within each lineage has been plotted in 50 bp intervals

The distribution of variable sites along HVR I was also plotted in non-overlapping 10 bp intervals (Figure 5.5). The reduced variability seen at the ends of HVR I on the 50 bp interval graph is still present on the 10 bp interval graph. In addition, the pattern is very similar in both lineages, with variation being high or low in a particular region. The amount of variation seen along HVR I follows a cyclic pattern, with both lineages peaking between 70-90 bp, 170-190 bp, 280-290 bp, 390-410 bp and 460-470 bp. There are also slightly smaller peaks between 140-150 bp, 230-240 bp, 340-350 bp, and 440-450 bp.

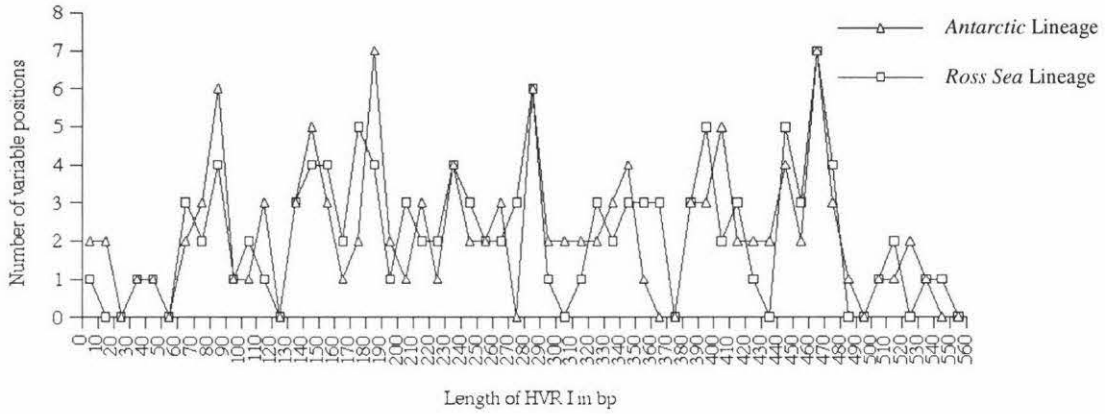


Figure 5.5 Distribution of the variable sites in HVR I. The number of variable sites within each lineage has been plotted in 10 bp intervals.

The sites defining the A/RS lineage split were also plotted on the 10 bp interval graph (Figure 5.6). From this it can be seen that regions with more variation within each lineage are also regions where more of the sites that define the *A/RS* split are found.

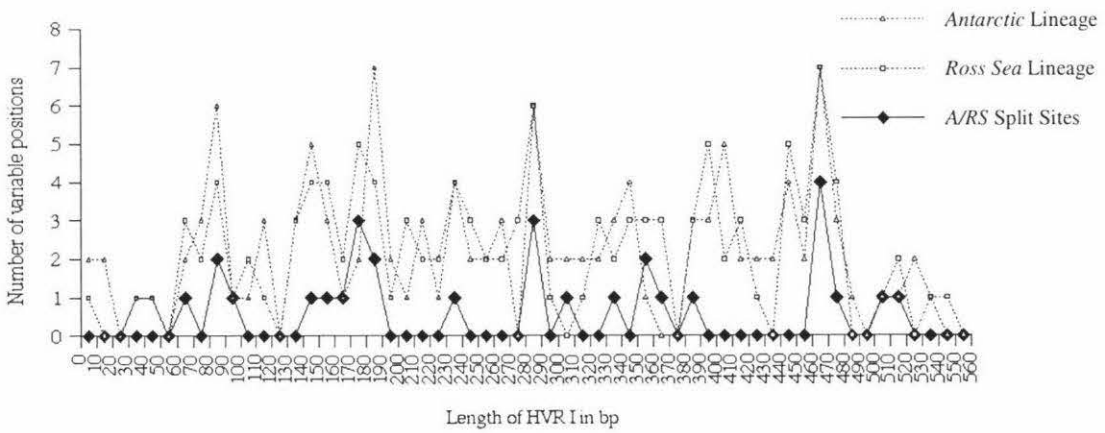


Figure 5.6 Distribution of variable sites over 10 bp intervals. Comparison of the *A* and *RS* lineages with the sites defining the *A/RS* lineage split.

5.2.6 HVR I features

There are not many conserved features in hypervariable region I, however two have been identified in the Adélie penguin (Ritchie and Lambert 2000). These are the termination associated sequences (TAS) and dC-terminus of the D-Loop. These two features have been putatively identified at positions 110-120 and 105 of HVR I respectively (positions 15676-15686 and 15617 of the genome). The non-majority plots in Figure 5.1, show there is very little variation in this region. Indeed, no variation was seen at the putative dC-terminus in this study, and the variation there is between positions 110-120 is only present in a few individuals.

5.3 Discussion

5.3.1 Nucleotide composition

This chapter has looked at nucleotide diversity and variability along HVR I in comparison both between and within the two lineages. The two lineages have almost identical base compositions. Analysis of the whole of HVR I in Adélie penguins has made the lack of L-strand guanine more apparent (17.4%), although it is still higher than that of many other species (Baker and Marshall 1997; Ruokonen and Kvist 2002). This is a typical feature of vertebrate mt DNA, where the GC asymmetry between the two strands has resulted in one strand being low in G (light strand) and the other high (heavy strand) (Baker and Marshall 1997; Saccone, Pesole and Sbisa 1991). In addition to the three regions of the control region (see Chapter 1.2) being delineated by nucleotide variability, they also have variable base composition. Generally, HVR I is AC rich, the central conserved domain CT rich and HVR II AT rich (Baker and Marshall 1997). As can be seen from this study, HVR I in Adélie penguins is AT rich, rather than AC as was seen in many other avian studies (Baker and Marshall 1997; Ruokonen and Kvist 2002). This feature of the Adélie penguin HVR I is also seen in other penguin species, which have very similar nucleotide compositions (Ritchie 2001). An AT rich HVR I is also more common in mammals (Sbisà et al. 1997).

5.3.2 HVR I variation within and between the two lineages

Non-majority plots show different patterns of nucleotide variation between the *A* and *RS* lineages. Those sites with a large proportion of each variant occur in different locations in the two lineages. In addition, when comparing sequences from different populations within a lineage, the number of shared sites with a high proportion of each variant is greater, even from opposite sides of the continent. Populations within the *A* or *RS* lineage are more similar to each other than to populations of the other lineage, even when the *A* and *RS* lineage individuals are in the same geographical location. The Davis, Mawson and Antarctic Peninsula populations are slightly different from the other *A* lineage populations in the Ross Sea region. It is unclear, however, if this is due to the small sample sizes, which required the three populations to be pooled, or if it is an effect of the distance of these populations both from each other and from the Ross Sea region, or a combination of both. DMP samples still had more similarity to the other *A* lineage populations than to the *RS* populations (Table 5.2, Figure 5.3a). This pattern is not nearly so clear when all sites were analysed, rather than just those with greater than 5% variation (Table 5.2, Figure 5.3b). This may be because HVR I is so variable that when all variable sites are included, even those sites where the variation is present in only one individual, this leads to the pattern becoming indistinct and clouded.

It is interesting that different populations of a lineage have similar proportions of the non-majority nucleotide at particular sites. One explanation for this observation is that there is significant gene flow between populations around Antarctica. In this case, the structure seen in the non-majority plot analysis would be indicative of a single large population spanning the region around the Ross Sea, and indeed also spreading around the whole of Antarctica. This theory is supported by microsatellite DNA studies that indicate little geographical structuring of alleles throughout Antarctica (Roeder et al. 2001). The authors suggested that in addition to high penguin dispersal, lack of differentiation between populations may also be explained by the very large sizes of Adélie penguin colonies. This would consequently cause stability of gene and genotype frequencies within populations. Observational evidence is contradictory, suggesting Adélie penguins have high natal phylopatry in

the short term, although in the longer term Adélie penguin populations appear to be able to quickly exploit newly exposed terrain as it becomes available (Ainley 2002). In addition, this theory is contradicted by the results seen in Chapter 3, where the *Ross Sea* lineage is found only in the Ross Sea region, and is very rare in populations bordering the Ross Sea.

5.3.3 The flaws of non-majority plot analyses

While non-majority plots can be useful for looking at variation along a sequence, and showing how many samples in a population have that variation, they are unable to distinguish between hot spots that mutate with great frequency and those sites with old mutations that occurred early in the phylogenetic tree (Figure 5.7).

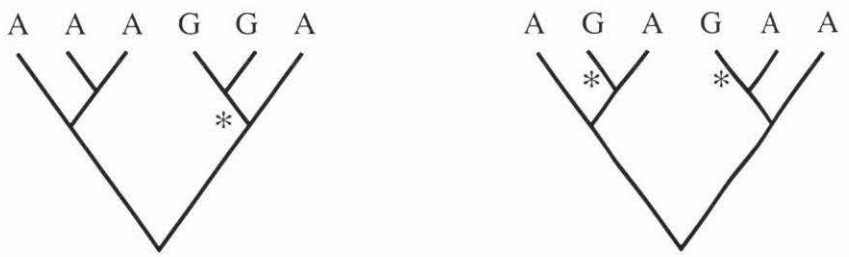


Figure 5.7 Two alternate scenarios where two different nucleotides (A and G) are present at one site in a sequence. The percent of the non-majority nucleotide in both diagrams is 33%, but the number of mutation events that lead to this percent in each scenario is different. Asterisks (*) represent mutation events.

The eye is drawn to the peaks in a non-majority plot but these are not necessarily indicative of patterns of nucleotide variation along the control region. It is desirable to have a clearer picture of the distribution of variable sites along the control region that is unrelated to the frequency of a particular mutation. To achieve this, the number of variable sites in non-overlapping 50 (Figure 5.4) and 10 bp (Figure 5.5) segments was plotted for each lineage.

5.3.4 A pattern to the variation along HVR I

When the variation in each lineage was plotted in 50 bp intervals, the distribution of variable sites in the two lineages is very uniform along HVR I. Except for the first and last 50 bp of HVR I, both lineages have an average of 25% sequence variation along HVR I (*Antarctic* lineage 12.4/50 variable sites, *Ross Sea* lineage 12.5/50 variable sites). When this is compared to the distribution of variable sites in other avian genera over 50 bp (see Baker and Marshall 1997; Marshall and Baker 1997; Ruokonen and Kvist 2002), Adélie penguin variation in HVR I seems more constant than that of other genera. This may be a factor of the large number of samples used in this study compared to previous studies, or it may be that the variation along HVR I is more homogeneous in Adélie penguins than other avian species.

Another feature of the sequence variation over 50 bp intervals is the lack of variation at both the beginning and end of HVR I. As can be seen in Figure 5.4, the first and last 50 bp intervals have about a third of the variation of the rest of HVR I. This may indicate that although the area between the end of tRNA-Glu and the beginning of CCD is defined as HVR I (Ruokonen and Kvist 2002), the ends of this sequence have less variability than the central portion. When the sequence variation was examined in 10 bp intervals, the lack of variability in the first and last 50 bp was still obvious. Studies of other avian species suggest a similar pattern, although as the variation in the middle of HVR I is not as uniform, this pattern is not as pronounced (Baker and Marshall 1997; Ruokonen and Kvist 2002).

When the sequence variation is examined in smaller intervals, a very different pattern emerges. Over 10 bp intervals (Figure 5.5), the variation along HVR I is not constant. The number of variable sites varies between 0% and 70% per 10 bp. Interestingly, the pattern of variation in the *A* and *RS* lineages follows each other very closely. Where one lineage has little variation, so does the other, and where one has a lot of variation, the other does as well. This is in contrast to the picture from the non-majority plots where it would appear the two lineages have different patterns of variation. When the distribution of variable sites is analysed by region and independent of the number of individuals with that variation (this is the size of the bars in a non-majority plot) it appears the two lineages are very similar. Other studies of variation have only used

50 bp interval graphs, not smaller regions such as 10 bp intervals (Baker and Marshall 1997; Ruokonen and Kvist 2002).

The sites where heteroplasmy was found were all sites where no previous variation was found. Looking at the variation in the 10 bp interval graph, the three sites fall in regions of different variation. There is no apparent pattern to the three occurrences of heteroplasmy, but they may be more an indication of the high levels of variability seen in HVR I. The sites where the TAS and dC-terminus occur had little variation, and that variation was present in only a few individuals. This may indicate some functional constraint on DNA variation in this region.

When the sites in HVR I that define the *Antarctic - Ross Sea* lineage split are also plotted in 10 bp intervals (Figure 5.6), it can be seen that they too fall in regions of higher variability. This suggests there may be underlying physical constraints on which sites can vary that is independent of the differences between the two lineages. The variation seen over 10 bp intervals has a cyclic pattern. High peaks appear at regular intervals with low valleys in between. When looking at the distribution of high peaks in both lineages they are spaced fairly evenly. The five largest peaks (taking both lineages together) are spaced about 100 bp apart, with slightly lesser peaks in between.

This cyclic pattern of high followed by low variation could be related to the secondary structure of the DNA when the D loop is single stranded during replication. A series of stem loop structures would allow for regions of constrained bases in a stem, followed by less constrained bases in the loop. Those bases in the loop region would be more likely to vary than those in the stem. In this scenario, variability and mutational hot spots are not related exactly to single bases in HVR I, but to short stretches of constrained and unconstrained DNA structure.

5.4 Summary of Main Findings

1) Analysis of the variation in HVR I by various methods has shown that populations within a lineage are more similar to each other than to populations of the other lineage. This observation is more pronounced at sites where more than 5% of the population has the lesser variant.

2) Despite the observation that variation along HVR I is different between the two lineages, comparison of the regions of HVR I where variation occurs showed the two lineages are very similar. In addition, the amount of variation along HVR I follows a cyclic pattern. This may suggest an underlying structural constraint on which regions of HVR I may vary, regardless of lineage.

3) There may also be constraint on variation around the region where the Termination Associated Sequences and dC-terminus of the D-Loop occur, as little variation was seen in this region in either lineage, and the variation that was there was present in only a few individuals.

Chapter Six

Looking to the Future: the Complete Mitochondrial Genome of the Adélie Penguin

6.1 Introduction

So far this thesis has been a study of variation in the control region of the mitochondrial genome, and what questions that can answer. This chapter looks at the question 'Where to from here?' More and more people are now looking to whole mitochondrial genome analyses to better answer questions in areas such as phylogenetics, rate variation and DNA evolution. I wish to extend the analyses of control region variability undertaken in this thesis to the whole mitochondrial genome. The first step in these analyses is to determine the mt DNA sequence of an Adélie penguin.

Hence, the aim of this chapter is to first determine the primers required, and then to sequence the entire mitochondrial genome of an Adélie penguin.

6.2 Results

6.2.1 The mt genome of the Adélie penguin

The complete mitochondrial genome of the Adélie penguin *Pygoscelis adeliae* is 17486 bp long. Gene order and orientation can be seen in Figure 6.1. The list of primers used to sequence the genome can be found in Appendix B.

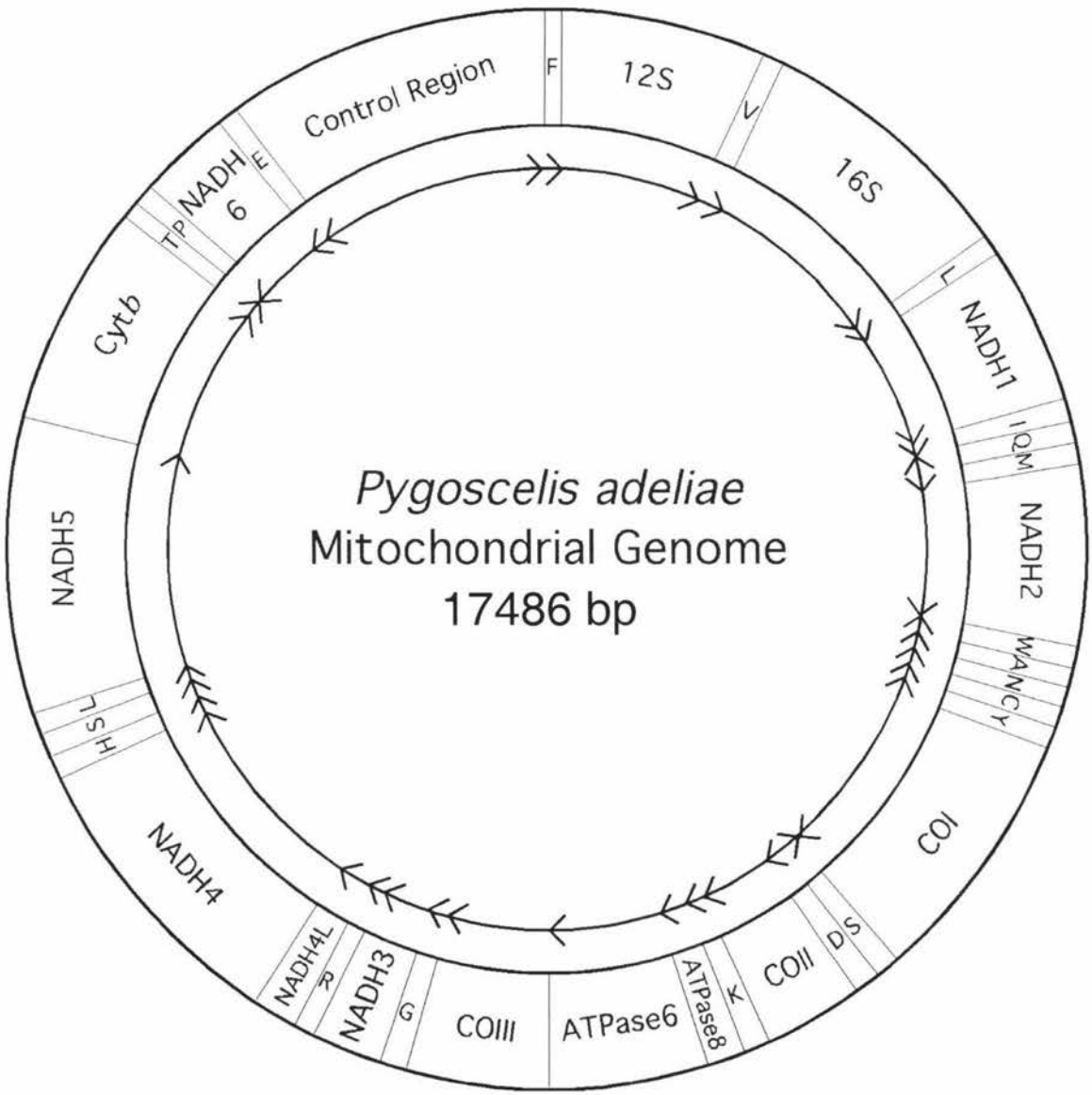


Figure 6.1. The mitochondrial genome of the Adélie penguin. The tRNA genes are listed by their amino acid one-letter codes. Arrows on the inner circle represent the gene orientation.

6.2.2 Mitogenomic features

Based on segregation of the hypervariable region I sequence in a UPGMA tree, the Adélie penguin (sample number PA1) whose sequence is reported here is from the *Antarctic* lineage.

The genome length of 17486 bp is subject to slight variability among mitochondria due to length heteroplasmy found in the control region. At the 3' end of the control region is a tetranucleotide microsatellite $(CAAA)_n$. Previously published sequence from the control region reported 30 simple short repeats (SSR) (Ritchie and Lambert 2000). The mitochondrial genome of the individual sequenced here contains 69 SSR. HVR II also contains a long repeat 81 bp in length. The sequence reported here contains four complete 81 bp repeats followed by one incomplete 79 bp repeat. This is the same number of repeats as reported previously (Ritchie and Lambert 2000). However when sequencing PCR products from this region, one primer pair produced a product that contained three complete 81 bp repeats and one incomplete 79 bp repeat. As all primers were amplifying from the same long range PCR template, it is most likely this is a result of inaccurate primer binding or PCR error. Nevertheless, the possibility remains that this indicates some heteroplasmic length variation in repeat number within an individual.

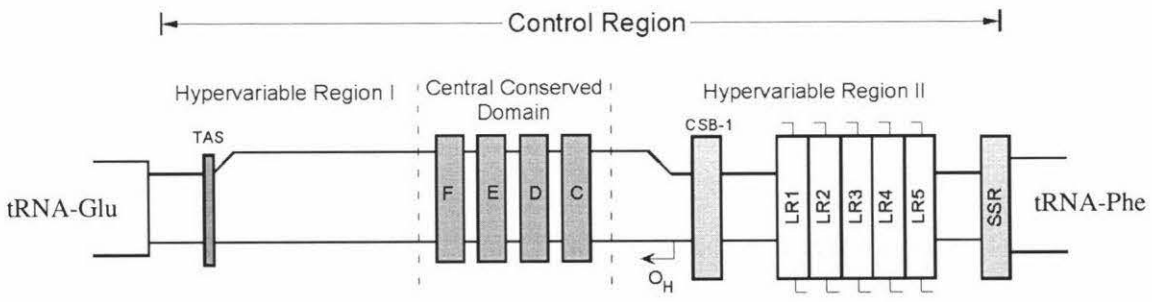


Figure 6.2. The mitochondrial control region of the Adélie penguin, showing the location of the five long repeats (LR 1-5) and the tetranucleotide simple short repeats (SSR). All other abbreviations used here are summarised in Figure 1.2.

The Adélie penguin has the standard gene order as originally identified in the chicken (Desjardins and Morais 1990). A summary of the mitochondrial genomes of other species mentioned here can be found in Slack et al. (2003). The protein-coding genes

are similar to those of other birds in terms of length, start and stop codons. Like the little blue penguin (*Eudyptula minor*), the Adélie penguin's NADH4 gene has an incomplete TA stop codon instead of the T seen in other neognaths and NADH6 has one less amino acid than most other birds. As is seen in many other bird and some turtle species, the Adélie penguin has an extra nucleotide in the gene NADH3 at position 9714 of the genome, and it is thought this gene is not translated (Mindell, Sorenson and Dimcheff 1998). The Adélie penguin also has no spacer between the genes for tRNA-Ser-AGY and tRNA-Leu-CUN. This is unlike most other birds where there is a one nucleotide overlap between the two genes, however this has been detected in the chicken.

A summary of the organisation of the Adélie penguin mitochondrial genome can be found in Table 6.1. This table follows the genome annotation as described in Slack et al. (2003). The length of protein coding genes is given excluding stop codons in both amino acids and base pairs. Stop codons are included in the intergenic spacer following the gene (for example the spacer following NADH1 is one nucleotide, therefore the tRNA-Ile gene following it starts on the first G of the stop codon AGG). This allows overlaps in coding sequence to be clearly identified. The start and stop positions for each gene also refer to coding sequence, except for the control region which is defined as being the genome region between the end of tRNA-Glu and the beginning of tRNA-Phe.

Table 6.1 Organisation of the mitochondrial genome of *Pygoscelis adeliae*

Gene/ Region	Start Position	End Position	Following Spacer	Start Codon	Stop Codon	Length bp	Length aa	Coding Strand
tRNA-Phe	1	69	-	-	-	69	-	H
12SrRNA	70	1046	-	-	-	977	-	H
tRNA-Val	1047	1118	-	-	-	72	-	H
16SrRNA	1119	2728	-	-	-	1610	-	H
tRNA-Leu-UUR	2729	2802	4	-	-	74	-	H
NADH1	2807	3781	1	ATG	AGG	975	325	H
tRNA-Ile	3783	3852	9	-	-	70	-	H
tRNA-Gln	3862	3932	1 overlap	-	-	71	-	L
tRNA-Met	3932	4001	-	-	-	70	-	H
NADH2	4002	5039	1	ATG	TAG	1038	346	H
tRNA-Trp	5041	5110	1	-	-	70	-	H
tRNA-Ala	5112	5180	2	-	-	69	-	L
tRNA-Asn	5183	5255	6	-	-	73	-	L
tRNA-Cys	5262	5328	1 overlap	-	-	67	-	L
tRNA-Tyr	5328	5398	1	-	-	71	-	L
COI	5400	6947	6 overlap	GTG	AGG	1548	516	H
tRNA-Ser-UCN	6942	7017	4	-	-	76	-	L
tRNA-Asp	7022	7090	2	-	-	69	-	H
COII	7093	7773	4	ATG	TAA	681	227	H
tRNA-Lys	7778	7847	1	-	-	70	-	H
ATPase8	7849	8013	7 overlap	ATG	TAA	165	55	H
ATPase6	8007	8687	2	ATG	TAA	681	227	H
COIII	8690	9472	1	ATG	T-	783	261	H
tRNA-Gly	9474	9542	-	-	-	69	-	H
NADH3	9543	9891	7	ATC	TAA	349	116	H
tRNA-Arg	9899	9967	1	-	-	69	-	H
NADH4L	9969	10262	4 overlap	ATG	TAA	294	98	H
NADH4	10259	11635	2	ATG	TA-	1377	459	H
tRNA-His	11638	11707	-	-	-	70	-	H
tRNA-Ser-AGY	11708	11773	-	-	-	66	-	H
tRNA-Leu-CUN	11774	11844	-	-	-	71	-	H
NADH5	11845	13659	8	ATG	TAA	1815	605	H
Cytb	13668	14807	7	ATG	TAA	1140	380	H
tRNA-Thr	14815	14884	14	-	-	70	-	H
tRNA-Pro	14899	14968	11	-	-	70	-	L
NADH6	14980	15495	3	ATG	TAG	516	172	L
tRNA-Glu	15499	15568	-	-	-	70	-	L
CR	15569	17486	-	-	-	1918	-	H

6.3 Discussion

6.3.1 The Adélie penguin has one of the longest mt genomes sequenced to date

This chapter has reported the complete mitochondrial genome of the Adélie penguin. At 17486 base pairs in length, the Adélie penguin has one of the longer avian mt genomes sequenced to date. Mitochondrial genomes range in length from 16591 bp (*Struthio camelus*) to 18674 bp (*Buteo buteo*) (Haring et al. 2001) with the majority being less than 17611 bp (little blue penguin) (Slack et al. 2003). Many of the longer genomes (e.g. *B. buteo* and *Falco peregrinus*) are long due to a rearrangement of mt genes that resulted in the formation of a second non-coding section designated the pseudo control region. This is not the case in the Adélie penguin, which has the standard avian gene order as first identified in the chicken (Desjardins and Morais 1990).

The length of the Adélie penguin mt genome is due in part to the length of the control region that is 1918 bp long and contains both a long and short repeated sequence in the second hypervariable region (Ritchie and Lambert 2000). Variations on long and short repeats have been seen in the control region of other avian species. For example, the little blue penguin has a 79 bp sequence repeated 9 times and a 7 bp sequence repeated 20 times (Slack et al. 2003). *F. peregrinus* also contains two different types of repeats, while the control region of *B. buteo* has seven different types of repeats (Haring et al. 2001).

6.3.2 The accuracy of genome sequencing from PCR templates

Sorensen et al. (1999) investigated the accuracy of PCR-based mt DNA sequencing as compared to a cloning strategy. They compared their PCR-based mt genome sequences from a variety of species to cloned sequences in a database. Based on their findings, they concluded sequences from a PCR-based approach, in conjunction with alignment to known sequences, proved more accurate than previously published cloned sequences (see also Waddell et al. 1999). However, they attribute some of the differences to errors in interpreting and transcribing raw data (in the cloned sequence) than systematic differences in replication error. When comparing the Adélie penguin

mt genome sequence to an alignment of mitochondrial genes in other species, very few anomalies were found. Checking the raw sequence data proved that these anomalies were errors in interpretation of the sequence data, rather than errors in PCR or sequencing.

6.3.3 Looking to the future

Knowledge of the complete mitochondrial genome will allow specific genes to be targeted for particular analyses. Different genes may be more informative for different studies. In 1999, Sorensen et al. published general PCR primers for the entire avian mt genome. This was to replace the inertia of many studies that looked at cytochrome *b* due to the availability of universal primers. They hoped more studies would now choose gene regions that were the most informative for that particular study, rather than be limited by primer availability. In particular, adequate resolution of higher-level relationships in birds and other taxa requires large datasets of a number of different genes, and even whole mitochondrial genomes (Sorenson et al. 1999).

There are many examples of future analyses that are now possible due to the sequencing of the complete mitochondrial genome of the Adélie penguin. Some of these are discussed in more detail in Chapter 7.

6.4 Summary of Main Findings

1.) This chapter reports the complete mitochondrial sequence of an *Antarctic* lineage Adélie penguin.

2.) As part of this, a database of primers useful in sequencing further Adélie penguin mt genomes has been compiled.

Chapter Seven

Summary and Discussion of Future Work

7.1 Summary of Findings

In the introduction of this thesis, objectives were put forward. Here, the main findings of this research are summarised as they relate to each objective (*italicised*).

1) To examine the range of the Antarctic and Ross Sea lineages in the vicinity of the Ross Sea region of Antarctica (Chapter 3)

A sudden and dramatic decrease in the *Ross Sea* lineage was seen upon leaving the Ross Sea region. The highest proportion of the *Ross Sea* lineage in a colony was observed at Cape Adare where 62% of Adélie penguins were *Ross Sea* lineage. The colonies at Port Martin and the Balleny islands to the north and west of the Ross Sea are only 3% *Ross Sea* lineage.

2) To determine whether the Antarctic and Ross Sea lineages are or are not one species (Chapter 3)

This study has proven that the *Antarctic* and *Ross Sea* lineages are one species. Adélie penguins in mated pairs were classified as *A* or *RS* based on hypervariable region I sequences. *A* x *RS* pairs and their offspring were genotyped and found to produce viable offspring. In addition, the proportion of *A* x *A*, *A* x *RS* and *RS* x *RS*

pairs was determined, and found to be in Hardy-Weinberg equilibrium with a Chi squared test showing 99.2% likelihood of random mating.

3) To calculate a preliminary rate of mutation in the mitochondrial hypervariable region I for the Adélie penguin (Chapter 4)

The DNA from Adélie penguins in 104 families was sequenced, corresponding to 191 transmissions of mt DNA. No mutations were found, resulting in an upper limit to the rate of mutation of 5.602 s/s/Myr (95% CI). Three cases of single point heteroplasmy were discovered, but were not included in the mutation rate calculation as the heteroplasmy was detected in both the mother and the chicks, indicating it arose in a previous generation.

4) To analyse the nucleotide diversity and variability in hypervariable region I as it relates to the two lineages and their geographic distribution (Chapter 5)

The amount of variation at different sites along HVR I differs between the two lineages, however when sites were grouped in contiguous 10 bp regions, the two lineages showed very similar patterns in the number of variable sites. The pattern of variation along HVR I also appears to be cyclic in nature. Less variation was seen in the sites where the TAS and dC-terminus are thought to occur, suggesting a constraint on sequence variation in this region.

5) To sequence the entire mitochondrial genome of an Adélie penguin (Chapter 6)

The entire mitochondrial genome of the Adélie penguin was sequenced and found to be 17486 bp in length. This is one of the longer bird mitochondrial genomes sequenced to date.

7.2 Future Work

7.2.1 Distribution of the Antarctic and Ross Sea mitochondrial lineages

While the proportions of the two lineages are known in the Ross Sea, and now to the north and west, it is unknown which of the two lineages penguins belong to in Marie Byrd land to the east of the Ross Sea. Loville Bluff colony in this region numbers about 61,000 pairs (Ainley 2002). On one side of Marie Byrd land, colonies in the Ross Sea have varying proportions of the two lineages. In the other direction, Adélie penguins on the Antarctic Peninsula are only *Antarctic* lineage. Sequencing HVR I mt DNA of Adélie Penguins from Marie Byrd land to the east of the Ross Sea would help to complete the picture of the geographic distribution of the two lineages.

Sequencing HVR I mt DNA of further old bones from the Ross Sea region will be useful in elucidating the proportions of the two lineages in the Ross Sea region in the past. This would be especially helpful at Cape Adare where the highest proportion of *Ross Sea* lineage Adélie penguins is found today. In the Antarctic summer season of 2002/2003 ancient bone samples were excavated from sites at Cape Adare, and it is hoped these bones will prove useful in further studies of the two lineages (D. Lambert, pers. comm.). A more complete picture of the range of the Antarctic and *Ross Sea* lineage both in the present day and in the past would facilitate understanding of the location of possible refugia of the Adélie penguin during the last glacial maximum.

7.2.2 An accurate measure of mutation rate in HVR I

Further blood samples from Adélie penguins and their chicks are required to get a more precise measure of the mutation rate of the mt DNA hypervariable region I. The DNA samples sequenced as part of this thesis are the first in a three-year study. The 2002/2003 summer season in Antarctica was the second year of this study, and has resulted in a further 882 blood samples from 221 families. The DNA from these blood samples is currently being sequenced and analysed for mutations.

7.2.3 The inheritance of heteroplasmy

Based on evidence from this study, similar proportions of two mt DNA haplotypes are passed from one generation to the next. This indicates that the mt DNA bottleneck in Adélies is more relaxed than in humans. As further samples are sequenced, more cases of heteroplasmy may be discovered which will add to the understanding of the mt DNA bottleneck in Adélie penguins. In addition to sequencing, HVR I DNA samples from pedigree individuals known to have heteroplasmy can be cloned to get a better estimate of the proportion of each haplotype in an individual. Looking at how levels of each haplotype change from parent to chick will increase the understanding of the transmission of mt DNA haplotypes from one generation to the next in Adélie penguins. As part of a study into the inheritance of heteroplasmy, the female Adélie penguins known to be heteroplasmic could be found again, and the DNA of any further offspring they have sequenced. All adult birds in the study were banded and the nest sites where they were captured marked. Adélie penguins return to the same area to nest each year, so baring the death of the individual, re-locating heteroplasmic adult females should be possible.

7.2.4 Identifying site-specific substitution rates

This study compared the pattern of nucleotide variability in HVR I between the two lineages. As part of further work in this field, the dataset of HVR I sequences available could be used to calculate site-specific substitution rates. Several different approaches have been developed to do this. Meyer and von Haeseler (2003) have designed a maximum likelihood framework for identifying site-specific rates based on empirical pairwise theory. In testing this method, Meyer and von Haeseler estimated site-specific rates for the complete mitochondrial genome of humans. This same method, or one developed by a number of different authors (for example Excoffier and Yang 1999; Gurven 2000; Meyer, Weiss and von Haeseler 1999; Pesole and Saccone 2001; Wakeley 1993) could be used to calculate site specific substitution rates in the Adélie penguin HVR I. If enough complete Adélie penguin mt genomes are sequenced (see Section 7.2.6) this analysis can be extended to the entire mitochondrial genome.

7.2.5 Secondary structure in HVR I

In Chapter 6, the variation along HVR I was seen to have a cyclic pattern. It was hypothesised that this might be a result of secondary structure in hypervariable region I. There are a number of DNA 2° structure folding programs available that could be used to determine the secondary structure of HVR I. This would determine if secondary structure can explain the pattern of nucleotide variability seen in the *Antarctic* and *Ross Sea* lineages. Programs available to do this include MFOLD (Mathews et al. 1999) and LoopDLoop (Gilbert 1992).

7.2.6 The complete mt genome as a basis for further analyses

There is 8.3% sequence difference between hypervariable region I of the two lineages of Adélie penguin (Lambert et al. 2002). This corresponded to 29 sites in the 560 bp HVR I (Chapter 5.2.3). Lambert et al. (2002) note in sequencing the cytochrome *b* gene 1 difference was detected between the *Antarctic* and *Ross Sea* lineages in this region. Now that one complete mitochondrial genome has been sequenced, primers are available to sequence many Adélie penguin mitochondrial genomes of both lineages. It will then be possible to examine the differences between the *A* and *RS* lineage as they occur throughout the rest of the genome.

Traditionally, whole mt genome sequences have been used to clarify phylogenetic relationships between different species and groups of species (Haring et al. 2001; Lin et al. 2002; Slack et al. 2003). The complete mitochondrial genome of the Adélie penguin will help analyses in this regard as well. However, perhaps more exciting is the opportunities it opens up in sequencing and analysing ancient Adélie penguin mt DNA.

The complete mitochondrial sequence of a modern Adélie penguin has provided a template from which primers can be designed to sequence mt DNA from ancient Adélie penguin bone samples. Ancient DNA samples are fragmented, and long PCR fragments will be difficult to obtain. Ritchie (2001) was able to amplify 1600 bp of mt DNA from bones over 400 years old, and with 35% of bones under 2000 years old

amplifying a 663-1042 bp sequence. 45% of bones over 2000 years old consistently amplified a 390 bp HVR I sequence. Using the modern Adélie penguin sequence, primers can be designed to amplify fragments of approximately 200-300 bp in length. When a number of modern Adélie penguin genomes have been sequenced, alignments will show where regions of conserved sequence lie. These would be the most advantageous regions for designing primers to sequence ancient DNA. In addition, a complete modern sequence will be beneficial for comparison when aligning ancient mt DNA fragments.

Lambert et al. calculated a rate of sequence evolution for hypervariable region I of the mitochondrial genome using ancients DNA from serially preserved Adélie penguin bones (Lambert et al. 2002). Now that it is possible to sequence further regions of ancient mitochondrial DNA, this same method can be applied to other regions of the mitochondrial genome in order to calculate rates of evolution for those regions. Quinn (1992) calculated that the mitochondrial control region of the snow goose evolves 10 times as fast as the rest of the mt genome, based on HVR I sequence analysis and an RFLP study of the entire mt genome (Quinn, Shields and Wilson 1991). Using the method of Lambert et al. (2002) to calculate a rate of evolution from serially preserved ancient DNA will allow a comparison of rates between regions, and with other studies.

7.3 Concluding Remarks

The Adélie penguin is a remarkable species whose unique history and population structure lends itself to a number of interesting molecular studies. The high variability of hypervariable region I and the presence of two highly divergent mitochondrial lineages in Adélie penguins with distinct geographical distribution, in conjunction with the ease of high throughput DNA sequencing has opened up many new avenues for DNA analysis. This is demonstrated by the complementary array of mt DNA analyses presented in this thesis.

References

Ainley DG (2002) The Adélie Penguin: Bellwether of Climate Change. Columbia University Press, New York

Avise JC (1991) Ten unorthodox perspectives on evolution prompted by comparative population genetic findings on mitochondrial DNA. Annual Review Genetics 25:45-69

Baker AJ, Marshall HD (1997) Mitochondrial control region sequences as tools for understanding evolution. In: Mindell DP (ed) Avian Molecular Evolution and Systematics. Academic Press, San Diego, pp 51-82

Baroni C, Orombelli G (1991) Holocene raised beaches at Terra Nova Bay, Victoria Land, Antarctica. Quaternary Research 36:157-177

Baroni C, Orombelli G (1994) Abandoned penguin rookeries as Holocene paleoclimatic indicators in Antarctica. Geology 22:23-26

Bendall KE, Macaulay V, Sykes B (1997) Variable levels of a heteroplasmic point mutation in individual hair roots. American Journal of Human Genetics 61:1303-1308

Bendall KE, Macaulay VA, Baker JR, Sykes BC (1996) Heteroplasmic point mutations in the human mtDNA control region. American Journal of Human Genetics 59:1276-1287

Bergstrom CT, Pritchard J (1998) Germline bottlenecks and the evolutionary maintenance of mitochondrial genomes. Genetics 149:2135-2146

Brown WM (1985) The mitochondrial genome of animals. In: Macintyre RJ (ed) Molecular Evolutionary Genetics. Plenum, New York, pp 95-130

Brown WM, Gadaleta G, Pepe G, Saccone C, Sbisa E (1986) Structural conservation and variation in the D-Loop containing region of vertebrate mitochondrial DNA. Journal of Molecular Biology 192:503-511

Brown WM, Jr. GM, Wilson AC (1979) Rapid evolution of animal mitochondrial DNA. Proceedings of the National Academy of Sciences of the United States of America 76:1967-71

Chang D, Clayton DA (1985) Identification of primary transcription start sites of mouse mitochondrial DNA: Accurate in vitro initiation of both heavy- and light-strand transcriptions. Molecular Cell Biology 6:1446-53

Clayton DA (1982) Replication of animal mitochondrial DNA. Cell 28:693-705

Conway H, Hall BL, Denton GH, Gades AM, Waddington ED (1999) Past and future grounding-line retreat of the West Antarctic Ice Sheet. Science 286:280-283

Culik B, Wilson RP (1991) Energetics of under-water swimming in Adélie penguins (*Pygoscelis adeliae*). Journal of Experimental Biology B 161:285-291

Denver DR, Morris K, Lynch M, Vassilieva LL, Thomas WK (2000) High direct estimate of the mutation rate in the mitochondrial genome of Caenorhabditis elegans. Science 289:2342-2344

Desjardins P, Morais R (1990) Sequence and gene organization of the chicken mitochondrial genome. A novel gene order in higher vertebrates. Journal of Molecular Biology 212:599-634

Drummond A, Rodrigo AG (2000) Reconstructing genealogies of serial samples under the assumption of a molecular clock using serial-sample UPGMA. Molecular Biology and Evolution 17:1807-1815

Drummond AJ, Nicholls GK, Rodrigo AG, Solomon W (2002) Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. Genetics 161:1307-1320

Ellegren H (1992) Polymerase chain reaction (PCR) of microsatellites - a new approach to studies of genetic relationships in birds. Auk 109:886-895

Excoffier L, Yang ZH (1999) Substitution rate variation among sites in mitochondrial hypervariable region I of humans and chimpanzees. Molecular Biology and Evolution 16:1357-1368

Eyre-Walker A, Smith NH, Smith JM (1999) How clonal are human Mitochondria? Philosophical Transactions of the Royal Society of London Series B-Biological Sciences 266:477-483

Gilbert D (1992) loopDloop: A Macintosh program for visualising RNA secondary structure. ftp://iubio.bio.indiana.edu/molbio/loopdloop/:

Griffiths R, Double MC, Orr K, Dawson RJG (1998) A DNA test to sex all birds. Molecular Ecology 7:1071-1075

Gurven M (2000) How can we distinguish between mutational "hot spots" and "old sites" in human mtDNA samples? Human Biology 72:455-471

Haring E, Kruckenhauser L, Gamauf A, Riesing MJ (2001) The complete sequence of the mitochondrial genome of *Buteo buteo* (Aves, Accipitridae) indicates an early split in the phylogeny of raptors. Molecular Biology and Evolution 18:1892-1904

Haynes GD (2002) Population Genetics of Adélie Penguin in Antarctica. Institute of Molecular BioSciences. Massey University, Palmerston North

Haynes GD, Gibb GC, Lambert DM (submitted) Genetic divergence without speciation: random mating between divergent mitochondrial lineages of Adélie Penguin.

Heyer E, Zietkiewicz E, Rochowski A, Yotova V, Puymirat J, Labuda D (2001) Phylogenetic and familial estimates of mitochondrial substitution rates: Study of control region mutations in deep- rooting pedigrees. American Journal of Human Genetics 69:1113-1126

Howell N, Kubacka I, Mackey DA (1996) How rapidly does the human mitochondrial genome evolve? American Journal of Human Genetics 59:501-509

Howell N, Smejkal CB, Mackey DA, Chinnery PF, Turnbull DM, Herrnstadt C (2003) The pedigree rate of sequence divergence in the human mitochondrial genome: There is a difference between phylogenetic and pedigree rates. American Journal of Human Genetics 72:659-670

Huynen L, Millar CD, Lambert DM (2002) A DNA test to sex ratite birds. Molecular Ecology 11:851-856

Jazin E, Soodyall H, Jalonen P, Lindholm E, Stoneking M, Gyllensten U (1998) Mitochondrial mutation rate revisited: hot spots and polymorphism. Nature Genetics 18:109-110

Kimura M (1983) The Neutral Theory of Molecular Evolution. Cambridge University Press, Cambridge

Kumar S, Koichiro T, Jakobsen IB, Nei M (2001) MEGA2: Molecular Evolutionary Genetics Analysis software. Bioinformatics 17:1244-1245

Kumar S, Tamura K, Nei M (1993) Manual for MEGA: molecular evolutionary genetic analysis software., IMEG, The Pennsylvania State University, University Park.

Kumazawa Y, Nishida M (1993) Sequence evolution of mitochondrial tRNA genes and deep-branch animal phylogenetics. Journal of Molecular Evolution 37:380-398

Lambert DM, Ritchie PA, Millar CD, Holland B, Drummond AJ, Baroni C (2002) Rates of evolution in ancient DNA from Adélie penguins. Science 295:2270-2273

Lin Y, McLenachan P, Gore A, Phillips M, Ota R, Hendy M, Penny D (2002) Four new mitochondrial genomes and the increased stability of evolutionary trees of mammals from improved taxon sampling. Molecular Biology and Evolution 19:2060-2070

Loewe L, Scherer S (1997) Mitochondrial eve: The plot thickens. Trends in Ecology & Evolution 12:422-423

Marshall HD, Baker AJ (1997) Structural conservation and variation in the mitochondrial control region of fringilline finches (Fringilla spp) and the greenfinch (Carduelis chloris). Molecular Biology and Evolution 14:173-184

Mathews D, Sabine J, Zuker M, Turner D (1999) Expanded sequence dependence of thermodynamic parameters provides robust prediction of RNA secondary structure. Journal of Molecular Biology 288:911-940

Matisoo-Smith E, Roberts RM, Irwin GJ, Allen JS, Penny D, Lambert DM (1998) Patterns of prehistoric human mobility in Polynesia indicated by mtDNA from the Pacific rat. Proceedings of the National Academy of Sciences of the United States of America 95:15145-15150

Meyer S, von Haeseler A (2003) Identifying site-specific substitution rates. Molecular Biology and Evolution 20:182-189

Meyer S, Weiss G, von Haeseler A (1999) Pattern of nucleotide substitution and rate heterogeneity in the hypervariable regions I and II of human mtDNA. Genetics 152:1103-1110

Mindell DP, Sorenson MD, Dimcheff DE (1998) An extra nucleotide is not translated in mitochondrial ND3 of some birds and turtles. Molecular Biology and Evolution 15:1568-1571

Parsons TJ, Holland MM (1998) Mitochondrial mutation rate revisited: hot spots and polymorphism - Response. Nature Genetics 18:110-110

Parsons TJ, Muniec DS, Sullivan K, Woodyatt N, AllistonGreiner R, Wilson MR, Berry DL, et al (1997) A high observed substitution rate in the human mitochondrial DNA control region. Nature Genetics 15:363-368

Pesole G, Saccone C (2001) A novel method for estimating substitution rate variation among sites in a large dataset of homologous DNA sequences. Genetics 157:859-865

Piko L, Taylor KD (1987) Amounts of mitochondrial DNA and abundance of some mitochondrial gene transcripts in early mouse embryos. Developmental Biology 123:364-374

Primmer CR, Moller AP, Ellegren H (1996) A wide-range survey of cross-species microsatellite amplification in birds. Molecular Ecology 5:365-378

Quinn TW (1992) The genetic legacy of Mother Goose- phylogenetic patterns of lesser snow goose *Chen caerulescens caerulescens* maternal lineages. Molecular Ecology 1:105-117

Quinn TW, Shields GF, Wilson AC (1991) Affinities of the Hawaiian goose based on two types of mitochondrial DNA data. Auk 108:585-593

Rand D (2001) The units of selection on mitochondrial DNA. Annual Review of Ecology and Systematics 32:415-48

Randi E, Lucchini V (1998) Organisation and evolution of the mitochondrial DNA control region in the avian genus *Alecoris*. Journal of Molecular Evolution 47:449-462

Ritchie PA (2001) The Evolution of the Mitochondrial DNA Control Region in the Adélie Penguins of Antarctica. Institute of Molecular BioSciences. Massey University, Palmerston North, New Zealand.

Ritchie PA, Lambert DM (2000) A repeat complex in the mitochondrial control region of Adélie penguins from Antarctica. Genome 43:613-618

Ritchie PA, Millar CD, Gibb GC, Baroni C, Lambert DM (in press.) Ancient DNA enables timing of the Pleistocene origin and Holocene expansion of two lineages of Adélie penguin lineages in Antarctica. Molecular Biology and Evolution:

Roeder AD, Marshall RK, Mitchelson AJ, Visagathilagar T, Ritchie PA, Love DR, Pakai TJ, et al (2001) Gene flow on the ice: genetic differentiation among Adélie penguin colonies around Antarctica. Molecular Ecology 10:1645-1656

Ruokonen M, Kvist L (2002) Structure and evolution of the avian mitochondrial control region. Molecular Phylogenetics and Evolution 23:422-432

Ruvolo M (1996) A new approach to studying modern human origins: hypothesis testing with coalescence time distributions. Molecular Phylogenet Evol 5:202-219

Saccone C, Pesole G, Sbisa E (1991) The Main Regulatory Region of Mammalian Mitochondrial-DNA - Structure-Function Model and Evolutionary Pattern. Journal of Molecular Evolution 33:83-91

Saccone C, Pesole G, Sbisa E (1997) Mammalian mitochondrial D-Loop region structural analysis: Identification of new conserved sequences and their functional and evolutionary implications. Gene 205:125-140

Sambrook J, Fritsch EF, Maniatus T (1989) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, New York.

Sbisà E, Tanzariello F, Reyes A, Pesole G, Saccone C (1997) Mammalian mitochondrial D-loop region structural analysis: identification of new conserved sequences and their functional and evolutionary implications. Gene 205:125-140

Seutin G, White BN, Boag PT (1991) Preservation of avian blood and tissue samples for DNA analysis. Canadian Journal of Zoology 69:82-90

Shields GF, Wilson AC (1987) Calibration of Mitochondrial DNA Evolution in Geese. Journal of Molecular Evolution 24:212-217

Sigurðardóttir S, Helgason A, Gulcher JR, Stefansson K, Donnelly P (2000) The mutation rate in the human mtDNA control region. American Journal of Human Genetics 66:1599-1609

Slack KE, Janke A, Penny D, Arnason U (2003) Two new avian mitochondrial genomes (penguin and goose) and a summary of bird and reptile mitogenomic features. Gene 302:43-52

Soodyall H, Jenkins T, Mukherjee A, DuToit E, Roberts DF, Stoneking M (1997) The founding mitochondrial DNA lineages of Tristan da Cunha islanders. American Journal of Physical Anthropology 104:157-166

Sorenson MD, Ast JC, Dimcheff DE, Yuri T, Mindell DP (1999) Primers for a PCR-based approach to mitochondrial genome sequencing in birds and other vertebrates. Molecular Phylogenetics and Evolution 12:105-114

Spurr EB (1975) Orientation of Adélie penguins on their territories. Condor 77:335-337

Stoneking M (2000) Hypervariable sites in the mtDNA control region are mutational hotspots. American Journal of Human Genetics 67:1029-1032

Stoneking M, Sherry ST, Redd AJ, Vigilant L (1992) New approaches to dating suggest a recent age for human mtDNA ancestor. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences 337:167-175

Taylor JS, Breden F (2002) The inheritance of heteroplasmy in guppies. Journal of Fish biology 60:1346-1350

Vigilant L, Stoneking M, Harpending H, Hawkes K, Wilson AC (1991) African populations and the evolution of human mitochondrial DNA. Science 253:1503-1507

Waddell PJ, Cao Y, Hasegawa M, Mindell DP (1999) Assessing the Cretaceous superordinal divergence times within birds and placental mammals by using whole mitochondrial protein sequences and an extended statistical framework. Systematic Biology 48:119-137

Wakeley J (1993) Substitution Rate Variation among Sites in Hypervariable Region-1 of Human Mitochondrial-DNA. Journal of Molecular Evolution 37:613-623

Wenink PW, Baker AJ, Tilanus MGJ (1994) Mitochondrial control region sequences in two shorebird species, the turnstone and the dunlin, and their utility in population genetic studies. Molecular Biology and Evolution 11:22-31

Wilson RP, Culik B, Adelung D, Coria N, Spairani H (1991) To slide or stride: when should Adélie penguins (*Pygoscelis adeliae*) toboggan? Canadian Journal of Zoology 69:221-225

Appendix A

Animal ethics and Antarctic permits

Permission to sample blood from Adélie penguins from Antarctica was approved by the Massey University Ethics Committee, protocol numbers 99/160 and 00/161. Approval to restrain, take blood from Adélie penguins and enter Specially Protected Areas (SPA) and Sites of Special Scientific Interest (SSSI) was given by Antarctica New Zealand permit numbers 00/7 and 01/7.

Appendix B

Table 1 Primers used in long-range PCR of the complete Adélie penguin mitochondrial genome

Primer Name	Sequence (5'-3')
16SFwdlg	AATTAGGGTT TACGACCTCG ATGTTGGATC AGG
Av13026R	CTTGGAKTTG CACCAAGRTD VTTGGTTCCT AAGACCA
Co310647F	TTTGAAGCAG CAGCCTGATA YTG
tPro16137R	ARAATRCCAG CTTTGGGAGT TGG
Av13063FtLeu	TGGTCTTAGG ARCCATCTAT CTTGGTGCAA MTCCAAGT
16S3797R	CGACCTGGAT TTCTCCGGTC TG

Table 2 Primers used in short-range PCR and sequencing of the complete Adélie penguin mitochondrial genome

Primer Name	Sequence (5'-3')
tPhe1278F	GGCACTGAAG ATGCCAAGAT GGTA
H1861R	TCGATTATAG AACAGGCTCC TC
Av175312SF	AAACTGGGAT TAGATACCCC ACTAT
16S2919R	GTTGAGCTTT GACGCACTC
Av224612SF	GAGGTAAGTC GTAACAAGGT AAG
16S3797R	CGACCTGGAT TTCTCCGGTC TG
BatH2157	CCATAGGGTC TTCTCGTCTT
L3450F	GAAGACCCTG TGGAACTTGA A
16S3787Fb	CGATTAACAG TCCTACGTG
1.4LP	AATGGRCCTG CDGCRTATTC
BatL1348	CWCARACWAT YTCYTATGAA GT
tMet5201R	CCATCATTTT CGGGGTATGG
tMet5197F	GGTCAGCTAA TTAAGCTATC G
H-WANCY	CCAAAAATCT GTGGTTCAAT TCCTCTTC
ND25583R	CCTTGGAGGA CTTCTGGGAA
L-WANCY	CCAAAGGCCT TCAAAGCCTT AAATAAGAG
CoI7101RH	CCTAGGATAG AGGAAACACC TGC
Av6838FCoI	CGTTACCGCC CATGCCTTCG T
H7662R	AGGAAGATGA AGCCYAGAGC TCA
CoI7546F	GTCGGAATAG ACGTAGACAC
CoII8377Rb	ATAGGAGATG AGGCGTCTTG
tSer8171FHL	AAGGAATCGA ACCCCCTTTA RCTGCTTTCA AGCCA
CoII8796R	CATGAGTGGA GAACATCTCC
UP7371	GGYCATCAAT GATAYTGAAG CTA
Av9881Atp6R	GGCTTAGTAG GAGGACGAAG
Atp69874FH	GCAGTAGCCA TAATCCAAGC
Av10529RCoIII	CCRATGATAA CATGGAGBCC GTGGAA
Av12138ND4R	ATTGGAGCTT CTACGTGGGC TT
L10647F	TTTGAAGCAG CAGCCTGATA YTG

Av11168FtArg	AGACAGTTGR TTTCGRCTCA ACA
Av11492ND4F	AACYTNAATC THCTACAATG CTAA
Av12717ND4R	GGTCCTCGTT GTGTGGTTAG TA
ND412312FH	GGTGCACTAA TAACCAGCTC AATCTG
Av13026tLeuR-LR	CTTGGAKTTG CACCAAGRTD VTTGGTTCCT AAGACCA
ND412788FH	CTCAAACACA CGAGAACACC
Av14050ND5R	GTTGAGATAT GGAGGAAGGC
ND513993FH	GCCTTCTCCA CATCAAG
Cytb14946Rb	AGGGGGAGGG GAGGTCAATT AGG
Av14617ND5F	CACACCCTAA CTCAACCAAA ACA
Cytb15783R	GGAATAGGAT TAGGACGGAG
HS106031Fcytb	CCCGCACACA TCAAACCAGA A
tPro16152	CTCCAGCTCC CAAAGCTG
AH530	CTGATTTCAC GTGAGGAGAC CG
L-tRNAGlu	CCCGCTTGGC TTYTCTCCAA GGTC
H-CentCR3'	CAAAGTGCAT CAGCGCGGAG ATGA
L-CentralCR5'	CCATTTTCGT CCGTGATCGC GGCAT
H-RepAnchor3'	TGTTGAGAGG GTGGTGGGCT TGAT
L-CentCR3'	TCATCTCCGC GCTGATGCAC TTTG
L-RepAnchor3'	CATTGATAAA CTACCACCCA ACCG
H-12SAves	CTGCTGAGTA CCCGTGGGGG TGTGGC

Table 3 Primers used in Microsatellite DNA Genotyping

Primer Name	Sequence (5'-3')
AM3 F	AGGAAAGAAG TAACTGAAGC AG
AM3 R	CATCTTCCCA CAGAAGAAAC
AM12 F	AAAAACCCAA CACAACAAAC
AM12 R	CCCAAGAAGA GATTTGTGAG
AM13 F	TTTTCCCATC TCTCTCCTG
AM13 R	CAGTTTTCAA CAATCCTTCC
RM3 F	AATCAGGCTC CAAGGTCA
RM3 R	ATGCAAGTGA CACAAAGG
RM6 F	CAGGAGGCTTT GAGACAA
RM6 R	CTGTTTACAT CCGATGCA
HrU2 F	CATCAAGAGA GGGATGGAAA GAGG
HrU2 R	GAAAAGATTA TTTTTCTTTC TCCC
XVI G4 F	CAGGAGAACA GTTGGAGAGG CAG
XVI G4 R	AGACAGGGTA AGGGGAGGAT GG
XVI H11 F	GCTAGGGCAA GGGACAGAC
XVI H11 R	CCTGTGGACA AGGATGCAG
XVIII B2 F	AGTAGCCAGC TCCTCCAGGT
XVIII B2 R	CTGGTGAGTA GTTGGGACCA G

Appendix C

Manuscripts

Haynes GD, Gibb GC, Lambert DM (submitted) Genetic divergence without speciation: random mating between divergent mitochondrial lineages of Adélie Penguin.

Ritchie PA, Millar CD, Gibb GC, Baroni C, Lambert DM (in press.) Ancient DNA enables timing of the Pleistocene origin and Holocene expansion of two lineages of Adélie penguin lineages in Antarctica. Molecular Biology and Evolution

Molecular Ecology 2003, 0, 000-000

SHORT COMMUNICATION

Genetic divergence without speciation: random mating between highly divergent mitochondrial lineages of Adélie penguin

G. D. HAYNES*, G. C. GIBB* and D. M. LAMBERT*[1]

**Allan Wilson Centre for Molecular Ecology and Evolution, Institute of Molecular BioSciences, Massey University, Private Bag 102-904, Auckland, New Zealand*

Running title: Genetic divergence without speciation

2, 401 words

Correspondence: David Lambert. FAX: +64-6-3505866; email: D.M.Lambert@Massey.ac.nz

Abstract

The mitochondrial control region sequences of the Antarctic and Ross Sea lineages of Adélie penguins are separated by an average of 8.3% difference. This high level of divergence brings into question the specific status of the lineages. Using microsatellite DNA genotyping markers, we tested for any genetic discontinuities between these two forms at sympatric sites in Antarctica. We also examined field data to test for random mating between males and females of individuals belonging to the two lineages. Our data show a highly genetically variable species, where the two mitochondrial types mate at random and produce offspring. This example illustrates that large levels of genetic divergence can be contained within a single species and brings into question the widespread use of phylogenetic methods to detect species in nature.

Key words: genetic divergence, speciation, random mating, Adélie penguin, microsatellite, control region

Introduction

Genetic analysis of populations is frequently employed to resolve biological questions not answerable by other means. Examples include discerning levels of genetic structure between populations (eg. Kyle *et al.*, 2001), identifying source populations (eg. Vernesi *et al.*, 2002) and resolving cryptic species and subspecies complexes (eg. Biruni and Actander, 2000; Gomez *et al.*, 2001; Holder *et al.*, 1999).

Previously, two distinct mitochondrial lineages have been reported in Adélie penguins (*Pygoscelis adeliae*), based on sequence differences in hypervariable region I (HVRI) of the mitochondrial control region. These lineages form two clades, separated by an average of 8.3% sequence difference. One lineage is found almost exclusively in the Ross Sea region of Antarctica, and is designated the *Ross Sea* (*RS*) lineage. In contrast, the *Antarctic* (*A*) lineage has been found at all localities investigated around Antarctica, (Lambert *et al.*, 2002; Ritchie *et al.*, in press).

Two lines of evidence suggest that these distinctive lineages evolved in allopatry during the Last Glacial Maximum (LGM), and have since come into secondary contact. First, the frequency of *RS* individuals is highest near the entrance to the Ross Sea, and decreases as the latitude of Adélie penguin colonies increases. This clinal distribution of lineages is consistent with the Adélie penguins colonising the Antarctic continent from separate refugia, after the retreat of ice shelved at the end of the LGM made the continent suitable for colonisation (Ritchie *et al.*, in press). Second, the rate of sequence evolution for the Adélie penguin HVRI region calculated by Lambert *et al.* (2002), at 0.96 substitutions/site/million years (s/s/Myr) places the time of divergence between lineages at 37-122kyrBP. This corresponds to the middle of the last glacial cycle, when Adélie penguin colonies would have been isolated by a major expansion of the polar ice caps (Ritchie *et al.*, in press).

The 8.3% sequence difference between the *A* and *RS* lineages is greater than the level of sequence divergence reported between the control regions of some species and subspecies (Polziehn and Strobeck., 2002; Ruber *et al.*, 2001; Waters and Wallis,

2001). This leads to the question of whether or not the Adélie penguins are a single species, or if they comprise a complex of distinct but morphologically cryptic species. Such cryptic species have been observed in sympatric populations of millipede (Bond and Sierwald, 2002), rotifer (Gomez *et al.*, 2002), and bonefish (Colborn *et al.* (2002), and between sibling species of cichlid fish (Ruber *et al.*, 2001). A study by Roeder *et al.* (2001) reported genetic homogeneity at nuclear loci between Adélie penguin colonies, but no direct comparison was made of the nuclear loci of the two mitochondrial lineages.

Five microsatellite loci were isolated from a partial genomic library and enriched for microsatellite DNA sequences. Ninety-eight Adélie penguins from the Cape Adare colony were characterised as *A* or *RS*, and for their microsatellite alleles content. Statistical analyses were performed to test for nuclear differentiation between the *A* and *RS* individuals, which would indicate the presence of a reproductively isolated cryptic species. In addition, ninety-nine mated pairs of Adélie penguin from the Cape Bird colony, and their chicks, were characterised as *A* or *RS* lineage. The frequency of *RS* and *A* pairings was then tested as more or less frequent than expected at random.

Methods and Materials

Sample Collection

Blood samples were taken from Adélie penguins at the Cape Adare colony during the Austral summer of 2000/2001, and the Cape Bird colony during the Austral summer of 2001/2002. Cape Adare lies 71°18'S, 170°09'E, and Cape Bird 77°30'S, 162°10'E. Genomic DNA was extracted from blood by phenol-chloroform extraction with ethanol precipitation (Sambrook *et al*, 1989).

Library Construction

A partial genomic library enriched for microsatellite sequences was constructed from Adélie penguin DNA following a protocol similar to Armour *et al.* (1994). Genomic DNA was from three invidivuals was pooled and digested with NdeII. DNA fragments 300-650 base pairs in size were ligated to SAU linkers. Fragments containing microsatellite DNA sequences were selectively hybridised to $3mm^2$ nylon membranes saturated with synthetic (AAAG)n/(TTTC)n, (GATA)n/(CTAT)n, (CA)n/(GT)n and (GA)n/(CT)n target sequences. The membranes were washed with 1xSSC (150mM sodium chloride, 15mM sodium citrate) and 0.1%SDS. The enriched DNA was then stripped from the membranes (100µl 50mM KOH/0.01%SDS, followed by 100µl 50mM TrisHCl pH 7.5/0.01%SDS) and PCR amplified using SAULA as a primer.

The SAU linkers were then removed from the amplicons by complete digestion with NdeII. The DNA was then purified (Highpure Kit, Roche), ligated into pUC18 plasmids cut with *Bam*HI, and used to transform Max Efficiency DH5α competent cells (Life Technologies). The cells were grown on a selective medium containing

0.4μg/ml IPTG, 20μg/ml X-gal and 100μg/ml ampicillan. Recombinant clones were identified by blue white selection.

Isolation of microsatellite sequences

Approximately 1900 positive clones were cultured and screened for the presence of microsatellite sequences by hybridisation with ^{32}P-labelled (AAAG)n/(TTTC)n, (GATA)n/(CTAT)n, (CA)n/(GT)n and (GA)n/(CT)n oligonucleotides. Plasmids were extracted and sequenced from 60 of the 132 positive clones. Primers were designed using Primer3 (Rozen and Skaletsky, 2000). Primers and microsatellite DNA loci are detailed in Table 1.

Genotyping

The HVRI sequence of each penguin was amplified according to Lambert et al., 2002. Sequences were aligned using Sequencher™ version 4.1.2 (Gene Codes Corporation) and analysed using PAUP (Sinauer Associates). HVRI sequence from a chinstrap penguin (Pygoscelis antarctica) was used as an outgroup (Ritchie 2001). Individuals were designated as A or RS lineage, according to their position on a Neighbour Joining Tree.

Microsatellite DNA loci were PCR amplified from the Cape Adare penguins using fluorescently labelled forward primers. The forward primers were labelled with 6-FAM (XVIH11), VIC (XVIIIB2), PET (XIXB3) or NED (XVIG4, XXG12a) (Applied Biosystems). Three separate PCRs were used to amplify the microsatellites: (1) 2x buffer, 0.8mM $MgCl_2$, 0.4mg/ml BSA, 0.5U Taq DNA polymerase, 400μM each dNTP, 1M betaine, 0.5μM XVIG4, 0.45μM XVIH11 and 0.35μM XVIIIB2; (2)

2x buffer, 0.8mM $MgCl_2$, 0.4mg/ml BSA, 0.5U Taq DNA polymerase, 400μM each dNTP, 1M betaine, 0.35μM XIXB3; (3) 1x buffer, 0.8mM $MgCl_2$, 0.4mg/ml BSA, 0.5U Taq DNA polymerase, 400μM each dNTP, 1M betaine, 0.4μM XXG12a. PCR cycling conditions were 1 cycle of 94°C for 2 minutes, 30 cycles of 95°C for 30 seconds, 58°C for 20 seconds, 72°C for 20 seconds, 1 cycle of 72°C for 5 minutes. Samples were genotyped using an ABIPrism 377 sequencer. Genotype data was analysed using GENESCAN version 3.1 software (Perkin Elmer). Genotypes were scored by eye. Only alleles that could be scored unambiguously were included in the analysis.

Collection of mating preference data

Blood samples were collected from mated pairs and their chicks at the Cape Bird colony. Mated pairs were identified by display behaviour. The sex of each pair was confirmed using the genetic sex tests described by Griffiths *et al.* (1998), with the additional step of digesting the PCR products with *Hae* III. This made scoring the sexes clearer (L. Huynen, personal communication). The mitochondrial sequences of the mothers and chicks were compared to confirm maternity. In addition, all members of 4 families were characterised for loci XVIG4, XVIH11 and XVIIIB2.

Data analysis

For the Cape Adare samples, GENEPOP v. 3.1c was used to measure the number of microsatellite alleles at each locus, calculate the observed and expected number of homozygous and heterozygous individuals for each locus, and to test for significant departure from linkage equilibrium (option 2.1), Hardy-Weinberg equilibrium (option 1.3), genic (single locus) homogeneity (option 3.1) and genotypic (whole genotype)

157 homogeneity (option 3.3). Results with a probability of ≤ 0.05 were considered
158 significant. In addition, a Doh assignment test was performed using the calculators at
159 www.biology.ualberta.ca/jbrzusto/Doh (Paetkau *et al.*, 1995), and a Chi-squared
160 probability of random assignment calculated.
161
162 For the Cape Bird samples, the frequency of observed *A-A*, *RS-A* and *RS-RS* pairing
163 was compared to the frequencies of such pairings expected under Hardy-Weinberg
164 Equilibrium. A Chi-squared probability of random pairing was calculated.
165

Results

Allele frequencies and linkage equilibrium

Ninety-eight individuals from the Cape Adare colony were genotyped for the five microsatellite loci. Thirty-seven of these penguins were *A* lineage, and 61 were *RS* lineage. The frequencies of alleles at each locus are reported on Table 2. The number of alleles at each locus ranged from 5 (XXG12a) to 31 (XVIH11). No pairs of loci showed significant departure (p-value<0.05) from linkage equilibrium, with probability values (p-values) for linkage equilibrium calculated between 0.056 (XVIH11 & XVIIIB2) and 1.000 (XVIH11 & XVIG4) (data not shown).

Hardy-Weinberg equilibrium

No significant departure from Hardy-Weinberg equilibrium was detected at any of the loci, indicating that the population is mating randomly. Hardy-Weinberg values are recorded in Table 2.

Genetic differentiation

No significant departures from genetic homogeneity were detected in either the genotypic or genic tests. Results from these tests are recorded in Table 3. The assignment test correctly assigned 46.9% of individual as *RS* or *A* lineage. The chi-squared probability of random assignment was calculated at 0.60062.

Mating preference

Ninety-nine mated pairs of Adélie penguin from Cape Bird, together with their offspring, were sequenced for the HVRI region, and determined to be either *RS* or *A* lineage. Ninety-six of the 99 families showed true maternity between the chicks and

their putative mothers (ie. not adopted from neighbouring nests). Of the 96, there were 55 *A-A* pairings, 36 *A-RS* pairings and 5 *RS-RS* pairings. The Chi-squared tests yielded a 1.0 probability of random pairing with respect to the mitochondrial lineages.

RS-A pairings were found in two of the four Cape Bird Adélie penguin families genotyped for loci XVIG4, XVIH11 and XVIIIB. In both these families, the paternity and maternity of chicks had been correctly assigned (ie. chicks were not a result of extra-pair copulation or adoption) (data not shown).

Discussion

If the two lineages represented biologically distinct species, then three predictions could be made: (1) nuclear genetic differentiation would be expected between the two lineages; (2) an excess of homozygote individuals would be detected in the overall population (Wahlund effect); (3) there would be preferential mating, where *A* lineage individuals would preferentially mate with other *A* lineage individuals, and the *RS* with the *RS*. In relation to (1) above, the failure of the exact tests to detect significant departure from genetic homogeneity, and of the assignment tests to class individuals as *RS* or *A* lineage more frequently than by random, is indicative of genetic homogeneity. No significant excess of homozygotes was detected in the Hardy-Weinberg test. The field data indicate that mating did not significantly depart from random with respect to the two lineages. Moreover, *RS* and *A* lineage individuals produce viable offspring together (although the fertility of these offspring has not been tested). The failure of any of these expectations to be met provides strong support that the two lineages of Adélie penguin represent a single species.

Genetic divergence without speciation

This study demonstrates that large amounts of genetic divergence can accumulate within a species without speciation. High levels of intraspecific genetic differentiation have also been reported in gorillas (Garner and Ryder, 1996), orangutans (Warren et al., 2001), kob (Birungi and Arctander, 2000), wildebeest, hartebeest and topi (Arctander at al., 1999). In all but one of these examples, however, the deep phylogenetic sequence divisions were between geographically separate populations, which had been divided into accepted or nominal subspecies. Highly divergent (9.8%) mitochondrial lineages were detected in Uganda kob (Birungi and Arctander, 2000),

224 but random mating between these lineages was not tested, so the authors could not
225 eliminate the possibility of a cryptic species.
226
227 The genetic divergence between the A and RS lineages is likely to have evolved in
228 allopatric Adélie penguin populations during the LGM (Ritchie et al., in press). These
229 allopatric populations were presumably small enough so that mitochondrial lineages
230 intermediate between A or RS did not persist. The populations have since come into
231 secondary contact and interbred. Similar scenarios have been observed in other
232 species after the LGM. Lake Ciscoes (Coregonus artedi) in North America, for
233 example, exhibit secondary contact and interbreeding between a pair of genetically
234 distinct populations that were separated during the LGM (Turgeon and Bernztchez,
235 2001).
236
237 It is likely that the two lineages of Adélie penguins remained reproductively coherent,
238 despite their separation during the LGM, for two reasons. (1) Insufficient time passed
239 for reproductive incompatibility to evolve as a by-product of random genetic drift; (2)
240 no shift in ecology. The large genetic divergence between the *RS* and *A* lineage Adélie
241 penguins is likely to be a result of the high rate of molecular evolution of the HVRI
242 locus in Adélie penguins. Lambert *et al.* (2002) estimated this rate to be 2 to 7 times
243 greater than the phylogenetic rate of 0.208s/s/Myr, which is generally accepted for
244 birds.
245
246 This study also emphasises that DNA sequence differences alone are not sufficient to
247 delimit species. This is because there is no correlation between speciation and genetic
248 differentiation. In this study it was shown that large amounts of genetic divergence

249 can be maintained within a single species. Speciation can also occur with very little or
250 no genetic change at neutral loci, especially when a species changes some aspect of its
251 ecology (eg. Piertney *et al.*, 2001; Bleiweiss, 2001; Beheregaray and Sunnucks 2001).
252 It is therefore necessary to use number of phylogenetic characters when delimiting
253 species. Mitochondrial DNA sequences, nuclear DNA allele frequencies and mate
254 recognition were employed in this study. Other studies of cryptic species have used
255 similar characters, as well as morphological and/or ecological information (eg. Bond
256 and Sierwald, 2002; Wilcox *et al* 1997; Gomez *et al.*, 2002; Colborn *et al.*, 2001;
257 Beheregaray and Sunnucks, 2001). In studies that use only mitochondrial sequence
258 data, it is not possible to determine the species status of allopatric populations (eg.
259 Birungi and Arctander, 2000).
260

Acknowledgements

We would like to thank John Macdonald and Peter Metcalf from the University of Auckland for their assistance in collecting blood samples. We would also like to acknowledge Peter Ritchie, Leon Huynen and Oliver Berry for their technical advice and assistance. Funding for this research was provided by a Marsden Fund of New Zealand grant to D.M.L. This paper is publication number # from the Allan Wilson Centre for Molecular Ecology and Evolution.

268 **References**

269

270 Arctander P, Johansen C, Coutellec-Vreto M (1999) Phylogeography of three closely
271 related African bovids (tribe Alcelaphini). *Molecular Biology and Evolution*
272 **16**, 1724-1739.

273 Armour JAL, Neumann R, Gorbert S, Jeffreys AJ (1994) Isolation of human simple
274 repeat loci by hybridisation selection. *Human Molecular Genetics* **3**, 599-605.

275 Beheregaray LB, Sunnucks P (2001) Fine-scale genetic structure, estuarine
276 colonization and incipient speciation in the marine silverside fish *Odontesthes*
277 *argentinensis*. *Molecular Ecology* **10**, 2849-2866.

278 Birungi J, Arctander P (2000) large sequence divergence of mitochondrial DNA
279 genotypes of the control region within populations of the African antelope,
280 kob (*kobus kob*). *Molecular Ecology* **9**, 1997-2008.

281 Bleiweiss R (2001) Mimicary on the QT(L): genetics of speciation in *Mimulus*.
282 *Evolution* **55**, 1706-1709.

283 Bond JE, Sierwald P (2002) Cryptic speciation in the *Anadenbolus excisus* millipede
284 species complex on the island of Jamaica. *Evolution* **56**, 1123-1135.

285 Colborn J, Crabtree RE, Shaklee JB, Feiler EP, Bowen BW (2001) The evolutionary
286 enigma of bonefishes (*Albula spp*): cryptic species and ancient separations in a
287 globally distributed shorefish. *Evolution* **55**, 807-820.

288 Garner KJ, Ryder OA (1996) Mitochondrial DNA diversity in gorillas. *Molecular*
289 *Phylogenetics and Evolution* **6**, 39-48.

290 Gomez A, Serra M, Carvalho GR, Lunt DH (2002) Speciation in ancient cryptic
291 species complexes: evidence from the molecular phylogeny of *Brachionus*
292 *plicatilis* (Rotifera). *Evolution* **56**, 1431-1444.

293 Holder MT, Erdmann MV, Wilcox TP, Caldwell RL, Hills DM (1999) Two living
294 species of coelacanths? *PNAS* **96**, 12616-12620.
295 Kyle CJ, Robitaille JF, Strobeck C (2001) Genetic variation and structure of fisher
296 (*Martes pennanti*) populations across North America. *Molecular Ecology* **10**,
297 2341-2347.
298 Lambert DM, Ritchie PA, Millar CD, *et al.* (2002) Rates of evolution in Ancient
299 DNA from Adélie penguins. *Science* **295**, 2270-2273.
300 Piertney SB, Summers R, Marquiss M (2001) Microsatellite and mitochondrial DNA
301 homogeneity among phenotypically diverse crossbill taxa in the UK.
302 *Proceedings of the Royal Society of London. Series B.* **268**, 1511-1517.
303 Polziehn EO, Stroback Q (2002) A phylogenetic comparison of red deer and wapiti
304 using mitochondrial DNA. *Molecular Phylogenetics and Evolution* **22**, 342-
305 356.
306 Ritchie PA, Millar CD, Gibb GC, Baroni C, Lambert DM (In press) Ancient DNA
307 enables timing of the Pleistocene origin and Holocene expansion of two
308 Adélie penguin lineages in Antarctica. *Molecular Biology and Evolution*
309 Roeder AD, Marshall RK, Mitchelson AJ, *et al.* (2001`) Gene flow on the ice: genetic
310 differentiation among Adélie penguin colonies around Antarctica. *Molecular*
311 *Ecology* **10**, 1645-1656.
312 Rozen S, Skaletsky HJ (2000) Primer3 on the WWW for general users and for
313 biologist programmers. In: *Bioinformatics Methods and Protocols: Methods in*
314 *Molecular Biology* (eds. Krawetz S, Misener S), pp 365-386. Humana Press,
315 Totowa, NJ. (Code available at http://www-genome.wi.mit.edu/
316 genome_software/other/primer3.html.)
317 Ruber L, Meyer A, Sturmbauer C, Verheyen E (2001) Population structure in two
318 sympatric species of the lake Tanganyika cichlid tribe Ereymodini: evidence

for introgression. *Molecular Ecology* **10**, 1207-1225.

Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning a Laboratory Manual, Second Edition* Cold Spring harbour Laboratory Press, New York.

Turgeon J, Bernatchez L (2001) Clinal variation at microsatellite loci reveals historical secondary intergradation between glacial races of *Coregonus artedi* (Teleostei: Coregonninae). *Evolution* **55**, 2274-2286.

Vernesi C, Pecchioli E, Caramelli D, *et al.* (2002) The genetic structure of natural and reintroduced roe deer (*Capreolus capreolus*) populations in the Alps and central Italy, with reference to the mitochondrial DNA phylogeography of Europe. *Molecular Ecology* **11**, 1285-1297.

Warren KS, Verschoor EJ, Langenhuijzen S, *et al.* (2001) Speciation and Intrasubspecific variation of Bornean orangutans, *Pongo pygmaeus pygmaeus*. *Molecular Biology and Evolution* **18**, 472-480.

Waters JM, Wallis GP (2000) Mitochondrial DNA phylogenetics of the *Galaxias vulgaris* complex from South Island, New Zealand: rapid radiation of a species flock. *Journal of Fish Biology* **2001**, 1166-1180.

Wilcox TP, Hugg L, Zeh JA, Zeh DW(1997) Mitochondrial DAN sequencing reveals extreme genetic differentiation in a cryptic species complex of neotropical pseudoscorpions. *Molecular Phylogenetics and Evolution* **7**, 208-216.

Table 1. Microsatellite loci and primers isolated from Adélie penguins

Locus	Primer Sequence (5'-3')	Repeat Array	Product Size
XVIG4	F CAGGAGAACAGTTGGAGAGGCAG R AGACAGGGTAAGGGGAGGATGG	$(CA)_{18}$	179-241 bp
XVIH11	F GCTAGGGCAAGGGACAGAC R CCTGTGGACAAGGATGCAG	$(CTAT)_{11}(GTAT)(CTAT)_{10}(CCAT)_9$	165-277 bp
XVIIIB2	F AGTAGCCAGCTCCTCCAGGT R CTGGTGAGTAGTTGGGACCAG	$(CA)_{16}$	84-102 bp
XIXB3	F TGATGATTCCCATTGCCATA R TCAAGTCAGGAGTATTGCCATTT	$(CA)_{11}$	122-132 bp
XXG12a	F GGTCCTGGATTCCCTTGACT R GCCAACTGCTGGAGAGACC	$(CA)_7AA(CA)_2CG(CA)_4$	109-117 bp

Table 2. Microsatellite allele frequencies in Adélie penguins at Cape Adare

Locus	Allele	Frequency		Locus	Allele	Frequency		Locus	Allele	Frequency	
		A	*RS*			*A*	*RS*			*A*	*RS*
XVIH11	1	0	0.01	*XVIH11*	*24*	0.032	0.041	*XVIG4*	10	0.206	0.1
	2	0	0.01		25	0.032	0.031		11	0.088	0.118
	3	0.016	0.01		26	0	0.02		12	0.059	0.055
	4	0.065	0.02		27	0.016	0		13	0.015	0.045
	5	0.016	0.051		28	0.048	0.02		14	0.015	0.009
	6	0.081	0.112		29	0.016	0		15	0.015	0
	7	0.145	0.133		30	0	0.01		16	0.015	0.009
	8	0.081	0.122		31	0	0.01		17	0.015	0
	9	0.065	0.102						18	0	0.009
	10	0.081	0.031	Obs Het		*28*	44		19	0	0.009
	11	0.032	0.01	*Exp. Het*		*29.4427*	*45.6908*				
	12	0.016	0.01	*Obs Hom.*		*3*	*5*	Obs Het		*29*	*45*
	13	0.016	0	*Exp. Hom.*		*1.5574*	*3.3093*	*Exp. Het*		*29.2154*	*49.2243*
	14	0.032	0					*Obs Hom.*		*4*	*9*
	15	0	0.051	*XVIG4*	1	0	0.009	*Exp. Hom.*		*3.7846*	*4.7757*
	16	0.065	0.01		2	0	0.027				
	17	0.032	0.031		3	0.29	0.045	*XVIIIB2*	1	0	0.025
	18	0.016	0.01		4	0	0.009		2	0.149	0.123
	19	0.016	0.01		5	0.59	0.118		3	0.108	0.074
	20	0.016	0.092		6	0.074	0.064		4	0.095	0.107
	21	0.016	0		7	0.132	0.136		5	0	0.008
	22	0	0.01		8	0.147	0.136		6	0.27	0.434
	23	0.048	0.031		9	0.132	0.1		7	0.162	0.123

Table 2. (Continued). Microsatellite allele frequencies in Adélie penguins at Cape Adare

Locus	Allele	Frequency		Locus	Allele	Frequency	
		A	*RS*			*A*	*RS*
XVIIIB2	8	0.162	0.057	*XXG12a*	1	0.015	0
	9	0.041	0.041		2	0.269	0.298
	10	0	0.008		3	0.507	0.423
	11	0.014	0		4	0.015	0.029
					5	0.194	0.25
Obs Het		*29*	*45*				
Exp. Het		*31.1233*	*46.6529*	Obs Het		*22*	*34*
Obs Hom.		*8*	*16*	*Exp. Het*		*21*	*35.1165*
Exp. Hom.		*5.8767*	*14.3471*	*Obs Hom.*		*11*	*18*
				Exp. Hom.		*12*	*16.8835*
XIXB3	1	0	0.03				
	2	0.758	0.68				
	3	0.197	0.25				
	4	0.045	0.02				
	5	0	0.01				
	6	0	0.01				
Obs Het		*12*	*22*				
Exp. Het		*12.9077*	*23.9192*				
Obs Hom.		*21*	*28*				
Exp. Hom.		*20.0923*	*26.0808*				

Table 3. Exact test for Hardy-Weinberg equilibrium and genetic differentiation between the *Ross Sea* and *Antarctic* lineage Adélie penguins at the Cape Adare colony.

Locus	Hardy-Weinberg Equilibrium				Genic Differentiation		Genotypic Differentiation	
	Antarctic lineage		*Ross Sea* lineage					
	P-value	**S.E.**	***P-value***	**S.E.**	***P-value***	**S.E.**	***P-value***	**S.E.**
XVIH11	0.1293	0.0288	0.2138	0.01256	*0.3104*	0.0112	0.0928	0.0223
XVIG4	0.6988	0.0212	0.7298	0.02039	0.8162	0.0067	0.0917	0.017
XVIIIB2	0.4353	0.0153	0.14429	0.01013	0.1758	0.008	0.3276	0.028
XXG12a	0.816	0.00074	0.58322	0.00851	0.5049	0.0072	0.9059	0.0029
XIXB3	0.6327	0.0044	0.4891	0.00886	0.4503	0.0067	0.0658	0.0077
combined across all loci	*0.649*		*0.0659*		*0.4334*		*0.9807*	
combined across all loci, both lineages	*0.193903*							

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, Institute of Molecular BioSciences, Massey University, Private Bag 102 904, Auckland, New Zealand

†Allan Wilson Centre for Molecular Ecology and Evolution, School of Biological Sciences, University of Auckland, Private Bag 92019, Auckland, New Zealand

‡Dipartmento Scienze della Terra, Università di Pisa, and Consiglio Nazionale Ricerche, Centro Studio Geologia Strutturale, Via Santa Maria, 53, 56126, Pisa, Italy

Address for correspondence and reprints: David M. Lambert, Allan Wilson Centre for Molecular Ecology and Evolution, Institute of Molecular BioSciences, Massey University, Private Bag 102 904, Auckland, New Zealand. Email: D.M.Lambert@massey.ac.nz.

RUNNING TITLE: Pleistocene origin and Holocene expansion of Adélie penguins

Word count: 5,251

Including 4 Figures and 2 Tables

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 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.

Key words: molecular clock, phylogeography, control region, Bayesian inference, rate heterogeneity.

Introduction

Evolutionary biology, over the past 100 years, has focused on first determining phylogenetic relationships amongst 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 impacted on the evolution of diverse biota. Such vicariant events are typically associated with periods of global environmental change. The Pleistocene epoch (2-Myr BP - 10-kyr BP) was, for example, characterized by a series of global ice ages that were remarkably synchronous (Williams 1989). Over the last 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 (figure 1) was uninhabitable to Adélie penguins due to 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.

Owing to 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). Perhaps 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 include misdiagnosis of fossil taxa, degrees of uncertainty in dating fossils, the accurate positioning of relevant fossils on a phylogenetic tree, proper dating of the stem lineage of an extant group and finally fossils offer only minimum age estimates (Soltis et al. 2002). Moreover, other factors that can bias estimates of rate may include a history of mitochondrial introgression, non-neutrality 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 over or under-estimate 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 polymorphism (RFLP) of mitochondrial genomes Shields and Wilson recorded 9% sequence difference between *Anser* and *Branta* geese. Knowing that that oldest fossils from each genus were dated at 4-5 million years (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% / Myrs, Shields and Wilson 1987). This relative difference was suggested because the control regions 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 difference in rates between the entire molecule and the control region 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 (t_{MRCA}) 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 - 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 1996-2001. Sub-fossil 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 sub-fossil bones using a stratigraphic method of excavation (Baroni and Orombelli 1994). Sub-fossil 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 – AF474997.

DNA extraction and sequencing

Modern DNA was extracted from 177 whole blood samples following standard methods (Sambrook, Fritsch, and Maniatus 1989). The hypervariable region 1 (HVR-1) of the mtDNA control region was amplified using the PCR and primers specific to Adélie penguins. For the modern DNA samples we amplified a 640 bp fragment using the primers L-tRNA^{Glu} (5′-CCCGCTTGGCTTYTCTCCAAGGTC-3′) and H-A650 (5′-CTGACATAGGAACCAGAGGCGC-3′) (Ritchie and Lambert 2000) (Roeder, Ritchie, and Lambert 2002). Samples were amplified using a Hybaid Omn-E

153 thermal cycler at 94°C 10 sec, 50°C or 61°C 10 sec and 72°C 25 sec for 30
154 cycles. All resulting PCR products were purified using High Pure PCR
155 purification columns (Roche), sequenced using the ABI PRISM® BigDye™
156 Terminator Cycle Sequencing kit (Applied Biosystems) and analyzed on a
157 ABI 377A automated sequencer. DNA sequences were aligned using
158 CLUSTAL W (Thompson, Higgins, and Gibson 1994) and can be retrieved at
159 the GenBank accession numbers AF000000 to AF000000.
160
161 Summary statistics
162 The rate of each type of nucleotide substitution relative to G↔T, was
163 estimated under a general time-reversible model (GTR) with maximum
164 likelihood (Rodríquez et al. 1990) using PAUP*4.0 (Swofford 2002).
165 Haplotypic (h) and nucleotide ($\pi \pm$ S.E.) diversity, and number of segregating
166 sites (S) and nucleotide differences (k) were estimated in DNASP v. 3.5 (Rozas
167 and Rozas 1999). To test for neutrality we calculated Tajima's (Tajima 1989) D
168 statistic the difference between S and π, which are both estimates of the
169 expected number of substitutions per site.
170
171 Phylogenetic analyses
172 A phylogenetic analysis of all sequences using the neighbor joining (NJ)
173 method (Saitou and Nei 1987) was conducted using a HKY85 substitution
174 model (Hasegawa, Kishino, and Yano 1985) in PAUP*4.0. The phylogenetic
175 tree was rooted with two sequences from chinstrap penguins (*P. antarctica*)
176 and one sequence from a gentoo penguin (*P. papua*). We assessed the level of
177 support for the major features of the NJ tree in the data set by re-sampling
178 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 v. 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 b.p.) 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 s/s/Myr. In both cases the posterior probability density of the t_{MRCA} was estimated from a distribution of likely trees, which were calculated as a product of a likelihood function $\Pr(X \mid \tau, \mu_e, Q)$ where Q is a GTR substitution model, the prior probabilities of the coalescent process $(\tau, \mid N_e)$, μ_e, N_e (effective population size) and Q (Drummond et al. 2002). The posterior probability densities were approximated using MCMC integration in MEPI v. 1.0 (Drummond et al. 2002). The prior probabilities for N_e were allowed to vary between 10^5 and 10^7 individuals. The estimated evolutionary rate for Adélie penguin HVR-1 sequences, and their clock-like behavior, has previously been established at 0.96 substitutions/site/Myr (s/s/Myr) (95%, HPD, interval 0.53-1.43) (Lambert et al. 2002). In the first analysis μ_e was allowed to vary between 0.53 and 1.43 and for the second analysis μ_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;

205 Quinn 1992) and is a generally accepted phylogenetic rate for the HVR-1 of
206 avian species. The Markov chains were 10^7 steps long and the first 5×10^5
207 iterations were discarded as the burn-in time, until a steady state was
208 reached.
209

210 ## Results

211

212 The HVR-1 sequences from the modern Adélie penguins were typically
213 594 bp long and were aligned to 96 ancient DNA sequences 352 bp long.
214 These sequences are homologous to the CR sequences for which (Quinn 1992)
215 estimated an evolutionary rate of 20.8% / Myr. For all modern samples (n =
216 557), the homologous 352 bp region showed unequal base frequencies (f_A =
217 0.31, f_T = 0.30, f_C = 0.19, and f_G = 0.20 from the H-strand sequence) and
218 revealed 440 haplotypes among 557 sequences (h = 0.998). The relative rates
219 for each substitution type was biased towards transitions (A↔C = 1.11, A↔G
220 = 21.52, A↔T = 0.33, C↔G = 1.49, C↔T = 15.99, relative to G↔T = 1.0 from the
221 heavy strand sequence) over 193 segregating sites, of which 47 were singleton
222 changes, and π = 0.043 (± S.E. 0.002). Tajima's D statistic was –1.712 and there
223 was no significant deviation from neutrality ($P > 0.05$).
224

225 Two mitochondrial DNA lineages

226 The most prominent feature of any phylogenetic analysis of HVR-1
227 sequence data for Adélie penguins is the presence of two monophyletic
228 lineages of Adélie penguins (figure 2). One lineage was recorded mainly in
229 the Ross Sea and is designated the Ross Sea (*RS*) lineage, whereas the other
230 was present at all locations around the Antarctic continent and was

231 consequently designated the Antarctic (*A*) lineage (figure 1*A*). There were on
232 average 27.7 nucleotide differences between sequences from the *A* and *RS*
233 lineages (overall $d = 0.081 \pm$ S.E. 0.012, and the net $d = 0.057$), significantly
234 more than the average number of differences among sequences within each
235 lineage ($k_A = 7.1$ and $k_{RS} = 9.8$). The two lineages shared 49 polymorphic sites
236 but there were no fixed differences between them. Sequences from the *A*
237 lineage had 98 polymorphic sites that were monomorphic among the *RS*
238 sequences, whilst sequences from the *RS* lineage had 46 polymorphic sites
239 that were all monomorphic among the *A* lineage sequences. When the two
240 outgroup species, the chinstrap and gentoo penguins, were compared with
241 the Adélie penguins sequences, the proportion of nucleotide differences were
242 0.305 (± S.E. 0.022) and 0.294 (± S.E. 0.027), respectively. Furthermore, there
243 was a 100 bp insertion-deletion event when the HVR-1 sequences from Adélie
244 and gentoo penguins were compared with those from chinstraps.
245
246 The monophyly of each lineage was well supported. There was 100%
247 bootstrap support for the split and the Metropolis-coupled MCMC algorithm
248 found that the monophyly of each lineage was 100% credible. A sequence
249 from one lineage never clustered with a sequence from the other lineage in
250 the full posterior distribution of likelihood trees. Plots of the log probability
251 of the observed data through time showed that the Markov chain reached
252 stability after about 8000 iterations and the first 50000 iterations were
253 discarded as the burn in time (data not shown). To check for convergence, a
254 second run was conducted that began with a random tree, and again this
255 showed no topological variance in the monophyly of the *A* and *RS* lineages.
256

257 The geographic pattern of mtDNA lineages

258 Populations in the Ross Sea have a higher nucleotide diversity ($\pi = 0.042 \pm$
259 S.E. 0.006) than those sampled from other Antarctic locations ($\pi = 0.019 \pm$ S.E.
260 0.001), owing to the presence of both lineages in the Ross Sea, and the absence
261 of one lineage outside of that region. A feature of the lineage proportions is
262 the decreasing frequency of the *RS* lineage with increasing latitude. The
263 geographic distribution of each lineage is presented in figure 1*A*. In figure 1*A*
264 the frequency cline has been divided into regional groups according to known
265 locations of the grounded ice shelf during the retreat of the Ross Ice sheet
266 since the last glacial maximum. These five regional groupings were
267 designated for convenience. These comprise: *Ross Island* (Capes Bird, Royds
268 and Crozier, and Beaufort Island), *Terra Nova* (Franklin and Inexpressible
269 Islands, Adélie Cove and Edmonson Point), *Adare* (Capes Wheatstone, Hallett
270 and Adare), *Balleny* (Balleny Islands and Port Martin), and the *non-Ross Sea*
271 locations (Torgersen, Welsh and Gardner Islands). The number of the *A* and
272 *RS* lineages in each of these regional groups are presented in table 1 and their
273 relative frequencies in figure 1*A*. The existence of a cline in *A* / *RS*
274 proportions is generally supported by a series of contingence χ^2-test that were
275 performed to compare the relative proportions of each lineage between
276 regions. This test showed a significant difference in all comparisons between
277 regional groupings ($p < 0.001$), excluding the comparison of *Ross Island* with
278 *Terra Nova* ($\chi^2 = 2.243$, $p = 0.191$) and *Balleny* with *non-Ross Sea* regional groups
279 ($\chi^2 = 2.034$, $p = 0.99$).
280
281 Ancient DNA samples

DNA has previously been recovered and HVR-1 sequences determined from 96 sub-fossil bone samples (Lambert et al. 2002). The geographic distribution of the ages for each sub-fossil bone and their respective lineage are presented in figure 1*B*. 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 fifteen samples dated to between 5706-6082 yr BP from Dunlop and Inexpressible islands, and revealed similar number 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 <2500 and 2500 – 5000 yrs 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 2000 years (mean age = 4,200 yr BP) at 0.074 (± S.E. 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 - 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} (figures 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 intervals 37 - 122-kyr BP). In contrast, using the phylogenetic rate the same approach estimated the $t_{MRCA\ A\ -\ RS}$ at 308-kyr BP (95% HPD intervals 227 - 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 16 – 52-kyr BP) and $t_{MRCA\ RS\ -\ RS}$ was 32-kyr BP (95% HPD 18 – 52-kyr BP). In contrast, the median estimates of the $t_{MRCA\ A\ -\ A}$ using the phylogenetic rate were 127-kyr BP (95% HPD 89 – 176-kyr BP) and $t_{MRCA\ RS\ -\ RS}$ were 143-kyr BP (95% HPD 121 – 186-kyr BP). These estimated divergence times differ from those previously presented in (Lambert et al. 2002) since 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, while the Ross Sea (*RS*) form 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

334 sequences of 0.96 s/s/Myr (0.53 - 1.43 s/s/Myr) suggests that the time to a
335 most recent common ancestor of these two lineages ($t_{MRCA\ A\ -\ RS}$) is 75-kyr BP
336 (37 - 122-kyr BP). The latter estimate is for a constant population.
337 Simulations for an exponentially growing population resulted in similar
338 estimates (data not shown). This $t_{MRCA\ A\ -\ RS}$ is the best estimate of the age of
339 the sequence that gave rise to all modern haplotypes of both lineages. Given
340 that there was almost certainly a level of existing polymorphisms in the
341 population at the time of separation of the lineages, this estimate represents
342 the maximum age of any period of lineage sorting. Lineage sorting results
343 from a population bottleneck in which a diverse set of haplotypes is reduced
344 in number. A number of extreme models, consistent with these data, can be
345 envisaged. First, the period of sorting might have begun after the onset of the
346 last glacial cycle. An ancestral population of mtDNA haplotypes would have
347 been dramatically reduced so that only the two ancestors of the *A* and *RS*
348 lineages would have survived. Over the last glacial cycle these two lineages
349 would have accumulated differences and Adélie penguins expanded in
350 population size and haplotype diversity, starting around the onset of the
351 deglacial period. Second, lineage sorting may have occurred late in the recent
352 glacial cycle, when ice conditions were at their most severe and not long
353 before the subsequent expansion associated with the deglacial warming. In
354 the latter scenario there would be a significant temporal separation between
355 the t_{MRCA} and the sorting event itself.
356
357 Using the ancient DNA rate of evolution calculated for Adélie penguins,
358 the estimated time to a most recent common ancestor of both lineages ($t_{MRCA\ A\ -}$
359 $_{RS}$) is consistent with our understanding of the glacial history of Antarctica.

360 The Late Pleistocene climate in Antarctica has been well documented, for
361 example temperature records are available from ice cores taken from Vostok
362 on the East Antarctic Ice Sheet (figure 1). These records date back to 420-kyr
363 BP (Petit et al. 1999), and provide a framework to test hypotheses about the
364 relationship between environmental change and the history of Adélie
365 penguins. A recent date of divergence of the lineages (approximately 75-kyr
366 BP) coincides with the middle of the last glacial cycle. At that time, Adélie
367 penguins may have had limited opportunities to breed on the Antarctic
368 mainland and our results suggest they were separated into refugia. Certainly
369 at the time of the last glacial maximum, ice-free areas would have been rare, if
370 present at all in Antarctica (Colhoun et al. 1992). The isolation of the Antarctic
371 continent from other major landmasses would have limited the ability of
372 Adélie penguins to progressively move north, as ice-free areas were reduced
373 and perhaps totally eliminated. Moreover, the expansion of the sea-ice would
374 have meant that breeding grounds were far removed from the open sea, even
375 if ice-free areas were present on the continent. Regular visiting of such
376 feeding grounds - a necessity during the breeding season when adults are
377 raising chicks - would be energetically expensive, if even physically possible.
378 Such sea ice covering would have been extensive. Benthic sedimentation of
379 ^{13}C-rich organic material from sea-ice algae shows that at the last glacial
380 maximum the summer sea-ice may have extended as far as the winter sea-ice
381 now reaches (figure 1A) (Cooke and Hays 1982; Gersonde and Zielinki 2000).
382 The latter extends from hundreds to thousands of kilometers from the
383 existing coastline and specifically the Ross Sea would have been completely
384 covered at that time. These large-scale disruptions to gene flow between ice

385 age Adélie penguin refugia are likely to have persisted as recently as 17 - 20
386 kyr BP (Baroni and Orombelli 1994).
387
388 In contrast, the phylogenetic rate (0.208 s/s/Myr) suggests that the $t_{MRCA\ A\text{-}RS}$
389 was 308-kyr BP and that consequently, the *A* and *RS* lineages persisted
390 through two or three glacial cycles (figure 3). According to this scenario,
391 Adélie penguin population sizes would have remained large enough over the
392 last two-three cycles for both lineages to survive. Since each glacial cycle is
393 likely to have resulted in some form of population bottleneck – given that
394 there would have been a loss of ice-free nesting areas – the above scenario
395 would seem less likely than the explanation based on the higher evolutionary
396 rate calculated directly from Adélie penguin ancient DNA. The latter
397 suggested a divergence during the last glacial cycle and therefore requires the
398 lineages to survive only a single ice age event.
399
400 The expansion of the two lineages
401 Using the ancient DNA rate we estimated t_{MRCA} of modern haplotypes of
402 each of the *A* and *RS* lineages separately. This resulted in very similar
403 estimates in the case of each lineage ($t_{MRCA\ A\text{-}A}$ = 30-kyr BP (16 - 52-kyr BP) and
404 $t_{MRCA\ RS\text{-}RS}$ = 32-kyr BP (18 - 52-kyr BP) (figure 4). We suggest that these two
405 lineages began to diversify some time after these estimates since there will
406 inevitably be a time lag between the t_{MRCA} and any dramatic increase in the
407 number of haplotypes. In contrast, the phylogenetic rate estimates the t_{MRCA} of
408 all modern *A* haplotypes to be 127-kyr BP (89 - 176-kyr BP) and all modern *RS*
409 haplotypes to be 143-kyr BP (121 - 186-kyr BP). This suggests that there was a
410 dramatic increase in the number of haplotypes of both lineages and that the

411 number of haplotypes accumulated over the last ice age. However, this
412 period was characterized by a general contraction in the distribution rather
413 than expansion of many species (Roberts 1998). As we have discussed above,
414 in the case of Adélie penguins, this period was probably one that was
415 associated with a reduction in numbers of Adélie penguins in Antarctica.
416
417 Ice age refugia
418 The restricted distribution of the *RS* lineage suggests a refugium that was
419 geographically close to, or adjacent to, the Ross Sea. Moreover, the high
420 relative frequency of the *RS* lineage at sites near the entrance to the Ross Sea,
421 compared with lower frequencies nearer to the permanent ice shelf, suggests
422 that after deglaciation the *RS* lineage moved down the coast following the
423 retreat of the Ross Ice Shelf (figure 4). In contrast, because of the Antarctic-
424 wide distribution of the *A* lineage, it is difficult to infer a possible location of
425 any refugium. However, it is interesting to note that the *A* lineage was well
426 established in the Ross Sea by 6082 yr BP and, in fact, at that date it was in
427 approximately equal proportions to the *RS* lineage at Terra Nova Bay in the
428 Ross Sea (figure 1). This would suggest that both lineages quickly established
429 themselves in the Ross Sea and our ancient DNA results show them to be
430 sympatric at many locations in past times. Our study also shows however,
431 that the proportions of each lineage, at individual sites, have changed during
432 this Holocene expansion period.
433
434 Rate heterogeneity and the 'molecular clock'
435 The observation that the amount of sequence difference between any two
436 taxa does not have a precise linear relationship to the time to their common

437 ancestor, has plagued the 'molecular clock' hypothesis for many years
438 (Gillespie 1991). Hence, a degree of heterogeneity in evolutionary rates is
439 inevitable. A variety of mechanisms have been proposed to explain such rate
440 heterogeneity among species or groups of species, including, for example, the
441 metabolic rate hypothesis (Martin and Palumbi 1993), the body temperature
442 hypothesis (Prager et al. 1974) and the generation time hypothesis (Kohne
443 1970). Moreover, the rate of evolution is not a deterministic process and can
444 vary along a lineage, and nucleotide changes have been described using a
445 Poisson model (Gillespie 1991). Rate heterogeneity along a species lineage
446 can be due to changes in the underlying mutation rate or the fixation of nearly
447 neutral alleles as the effective population size fluctuates (Ohta 1987).
448 Therefore, it has been recognized that a rate of evolution estimated using
449 fossils calibration dates of millions of years, might not be transferable to other
450 species and to more recent (or older) time periods.
451
452 Most studies that use the 'molecular clock' as a dating tool, rely on the
453 assumption that, first, there is a constant rate at which the 'clock ticks', and
454 second, that there are available relatively precise estimates of the true rate of
455 evolution for any particular species. The majority of studies are restricted to
456 adopting a rate of evolution previously established from a related group of
457 species. A rate of evolution is usually deemed a good approximation to the
458 true rate if there have been two or more studies of independent taxa that have
459 provided corroborating estimates. For example, Klicka and Zink (1997) used
460 the level of mtDNA sequence divergence between 35 sister species of oscine
461 passerine and a clock rate of 2% / Myr to refute the conventional model that
462 many North American birds populations speciated during the glacial cycles

463 of the Late Pleistocene. Klicka and Zink suggested that a rate of 2% / Myr for
464 mtDNA was a reasonable estimate, because it was corroborated among a
465 diverse array of studies, including, for example, Shields and Wilson (1987),
466 Nunn et al. (1996) and Tarr and Fleischer (1993). However, our study of
467 Adélie penguins has been the first to employ a rate that was directly
468 estimated from our study species, independent of both the fossil record and a
469 phylogeny of closely related species. Hence, we were able test how different
470 our results would be from one where we adopted the conventional CR rate
471 for birds. We recorded significantly different t_{MRCA} using rate specific to
472 Adélie penguins. Importantly, this had significant implications for
473 biogeographic explanations of the origin of two distinct mtDNA lineages
474 found in Adélie penguins.
475
476 Finally, ancient DNA has been widely used to investigate a variety of
477 issues, from changing levels of genetic variation (e.g. Nielsen, Hansen, and
478 Loeschcke 1997), to analyses of the phylogeny of extinct species (e.g. Höss et
479 al. 1996) to investigations of ancient diets (Poinar et al. 1998). Our research on
480 ancient DNA of Adélie penguins suggests a new role in evolutionary biology;
481 namely the timing of divergent events and tracking changes in the proportion
482 of lineages over space and time. This approach, when used to detect the time
483 to the MRCA of the *A* and *RS* lineages, is more consistent with the known
484 glacial history of Antarctica and requires fewer assumptions than does a
485 scenario based on an adopted phylogenetic rate of mitochondrial DNA
486 evolution. Hence, our results indicate that complex biogeographical
487 narratives based on evolutionary rates adopted from other species could be
488 seriously misleading. As an increasing number of ancient DNA studies are

489 reported, with consequently better estimates of evolutionary rates, this will
490 enable a more accurate understanding of rate heterogeneity and consequently
491 the genetic changes which have accompanied major global environmental
492 events, such as the ice ages in Antarctica.

Acknowledgements

This research was made possible by a grant from the Marsden Fund of New Zealand. 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. 00 from the Allan Wilson Centre for Molecular Ecology and Evolution.

LITERATURE CITED

AINLEY, D. G., R. E. LARESECHE, and W. J. SLADEN. 1983. Breeding Biology of Adélie Penguins. University of California Press, Berkeley.

BAKER, A. J., and H. D. MARSHALL. 1997. Mitochondrial control region sequences as tools for understanding evolution. Pp. 51-82 *in* D. P. Mindell, ed. Avian Molecular Evolution and Systematics. Academic Press, San Diego.

BARONI, C., and G. OROMBELLI. 1994. Abandoned penguin rookeries as Holocene paleoclimatic indicators in Antarctica. Geology **22**:23-26.

BOX, G. E. P., and G. C. TIAO. 1992. Bayesian inference in statistical analysis. Wiley, New York.

COLHOUN, E. A., M. C. G. MABIN, D. A. ADAMSON, and R. M. KIRK. 1992. Antarctic ice volume and contribution to sea-level fall at 20,000 yr BP from raised beaches. Nature **358**:316-319.

COOKE, D. W., and J. D. HAYS. 1982. Estimates of Antarctic ocean seasonal sea-ice cover during glacial intervals. Antarctic Geoscience: Symposium on

Antarctic Geology and Geophysics. University of Wisconsin Press, Madison, Wisconsin.

DENTON, G. H., J. G. BOCKHEIM, S. C. WILSON, and M. STUIVER. 1989. Late Wisconsin and early Holocene glacial history, inner Ross embayment, Antarctica. Quaternary Research **31**:151-182.

DRUMMOND, A., G. K. NICHOLLS, A. G. RODRIGO, and W. SOLOMON. 2002. Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. Genetics **161**: 1307-1320.

GERSONDE, R., and U. ZIELINKI. 2000. The reconstruction of late Quaternary Antarctic sea-ice distribution — the use of diatoms as a proxy for sea-ice. Palaeo **162**:263-286.

GILLESPIE, J. H. 1991. The Causes of Molecular Evolution. Oxford University Press, Oxford, UK.

HASEGAWA, M., H. KISHINO, and T. YANO. 1985. Dating the human-ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. **22**:160-174.

HECKMAN, D. S., D. M. GEISER, B. EIDELL, R. L. STAUFFER, N. L. KARDOS, and S. B. HEDGES. 2001. Molecular evidence for the early colonization of land by fungi and plants. Science **293**:1129-1133.

HELD, C. 2001. No evidence for slow-down of molecular substitution rates at subzero temperatures in Antarctic serolid isopods (Crustacea, Isopoda, Serolidae). Polar Biol. **24**:497-501.

HÖSS, M., A. DILLING, A. CURRANT, and S. PÄÄBO. 1996. Molecular phylogeny of the extinct ground sloth *Mylodon darwinii*. Proc. Natl. Acad. Sci. USA **93**:181-185.

546 HUELSENBECK, J. P., and F. RONQUIST. 2001. MRBAYES: Bayesian inference of
547 phylogenetic trees. Bioinformatics **17**:754-755.
548 IMBRIE, J., E. A. BOYLE, S. C. CLEMANS, A. DUFFY, W. R. HOWARD, G. KUKLA, J.
549 KUTZBACK, D. G. MARTINSON, A. MCINTYRE, A. C. MIX, B. MOLFINO, J. J.
550 MORLEY, L. C. PETERSON, N. G. PISIAS, W. L. PRELL, M. E. RAYMO, N. J.
551 SHACKLETON, and J. R. TOGGWEILER. 1992. On the structure and origin
552 of major glaciation cycles 1. linear responses to Milankovitch forcing.
553 Paleooceanography **7**:701-738.
554 KLICKA, J., and R. M. ZINK. 1997. The importance of recent Age Ages in
555 speciation: A failing paradigm. Science **277**:1666-1669.
556 KOHNE, D. E. 1970. Evolution of higher-organism DNA. Q. Rev. Biophys.
557 **33**:327-375.
558 LAMBERT, D. M., P. A. RITCHIE, C. D. MILLAR, B. HOLLAND, A. J. DRUMMOND,
559 and C. BARONI. 2002. Rates of evolution in ancient DNA from Adélie
560 penguins. Science **295**:2270-2273.
561 MARTIN, A. P., and S. R. PALUMBI. 1993. Body size, metabolic rate, generation
562 time, and the molecular clock. Proc. Natl. Acad. Sci. USA **90**:4087-4091.
563 MINDELL, D. P., and C. E. THACKER. 1996. Rates of molecular evolution:
564 phylogenetic issues and applications. Annu. Rev. Ecol. Syst. **27**:279-303.
565 NIELSEN, E. E., M. M. HANSEN, and V. LOESCHCKE. 1997. Analysis of
566 microsatellite DNA from old scale samples of Atlantic salmon *Salmo*
567 *salar*: a comparison of genetic composition over 60 years. Mol. Ecol. **6**:
568 487-492.
569 NUNN, G. B., J. COOPER, P. JOUVENTIN, C. J. R. ROBERTSON, and G. G.
570 ROBERTSON. 1996. Evolutionary relationships among extant albatrosses

(Procellariiformes: Diomedeidae) established from complete cytochrome-B gene sequences. Auk **113**:784-801.

OHTA, T. 1987. Very slightly deleterious mutations and the molecular clock. J. Mol. Evol. **26**:1-6.

PETIT, J. R., J. JOUZEL, D. RAYNAUD, N. I. BARKOV, J.-M. BARNOLA, I. BASILE, M. BENDER, J. CHAPPELAZ, M. DAVIS, G. DELAYGUE, M. DELMOTTE, V. M. KOTLYAKOV, M. LEGRAND, V. Y. LIPENKOV, C. LORUS, L. PÉPIN, C. RITZ, E. SATLZMAN, and M. STIEVENARD. 1999. Climate and atmospheric history of the past 420, 000 years from the Vostok ice core, Antarctica. Nature **399**:429-436.

POINAR, H. N., M. HOFREITER, W. G. SPAULDING, P. S. MATIN, B. A. STANKIEWICZ, H. BLAND, R. P. EVERSHED, G. POSSNERT, and S. PÄÄBO. 1998. Molecular Coproscopy: Dung and diet of the extinct Ground Sloth *Nothrotheriops shastensis*. Science **281**:402-406.

PRAGER, E. M., A. H. BRUSH, R. A. NOALN, M. NAKANISHI, and A. C. WILSON. 1974. Slow evolution of transferrin and albumin in birds according to micro-complement fixation analysis. J. Mol. Evol. **3**:243-262.

QUINN, T. W. 1992. The genetic legacy of Mother Goose - phylogeographic patterns of lesser snow goose *Chen caerulescens caerulescens* maternal lineages. Mol. Ecol. **1**:105-117.

RAND, D. M. 1994. Thermal habit, metabolic rate and the evolution of mitochondrial DNA. Trends Ecol. Evol. **9**:125-131.

RITCHIE, P. A., and D. M. LAMBERT. 2000. A repeat complex in the mitochondrial control region of Adélie penguins from Antarctica. Genome **43**:613-618.

596 ROBERTS, N. 1998. The Holocene: An Environmental History. Blackwell
597 Publishing, Oxford, U.K.
598 RODRÍQUEZ, F., J. L. OLIVER, A. MARÍN, and J. R. MEDINA. 1990. The general
599 stochastic model of nucleotide substitution. J. Theor. Biol. **142**:485-501.
600 ROEDER, A. D., P. A. RITCHIE, and D. M. LAMBERT. 2002. New DNA markers
601 for penguins. Cons. Genetics **3**: 341-344.
602 ROZAS, J., and R. ROZAS. 1999. DnaSP version 3: an integrated program for
603 molecular population genetics and molecular evolution analysis.
604 Bioinformatics **15**:174-175.
605 SAITOU, H., and M. NEI. 1987. The neighbor-joining method: A new method
606 for reconstructing phylogenetic trees. Mol. Biol. Evol. **4**:406-425.
607 SAMBROOK, J., E. F. FRITSCH, and T. MANIATUS. 1989. Molecular Cloning: A
608 Laboratory Manual. Cold Spring Harbor Laboratory Press, New York.
609 SHIELDS, G. F., and A. C. WILSON. 1987. Calibration of the mitochondrial DNA
610 evolution in geese. J. Mol. Evol. **24**:212-217.
611 SOLTIS, P. S., D. E. SOLTIS, V. SAVOLAINEN, P. R. CRANE, and T. G.
612 BARRACLOUGH. 2002. Rate heterogeneity among lineages of
613 tracheophytes: Integration of molecular and fossil data and evidence
614 for molecular living fossils. Proc. Natl. Acad. Sci. USA **99**:4430-4435.
615 STONEKING, M., S. T. SHERRY, A. J. REDD, and L. VIGILANT. 1992. New
616 approaches to dating suggest a recent age for human mtDNA ancestor.
617 Phil. Trans. R. Soc. Lond. B **337**:167-175.
618 SWOFFORD, D. L. 2002. PAUP*. Phylogenetic Analysis Using Parsimony (*and
619 Other Methods) Version 4.0b10. Sinauer Associates, Sunderland,
620 Massachusetts.

621 TAJIMA, F. 1989. Statistical methods for testing the neutral mutation
622 hypothesis by DNA polymorphism. Genetics **123**:585-595.
623 TARR, C. L., and R. C. FLEISCHER. 1993. Mitochondrial-DNA variation and the
624 evolutionary relationships in the Amakihi complex. Auk **110**:825-831.
625 THOMPSON, J. D., D. G. HIGGINS, and T. J. GIBSON. 1994. CLUSTAL W:
626 improving the sensitivity of progressive mutliple sequence alignment
627 through sequence weighting, position-specific gap penalties and
628 weight matrix choice. Nucleic Acid. Res. **22**:4673-4680.
629 WILLIAMS, M. A. J., DUNKERLEY, D., DEDECKKER, P., KERSHAW, P., and
630 CHAPPELL, M. 1989. Quaternary Environments. Arnold, London.
631

632 **List of figure legends**

633

634 FIG 1. – (*A and B)* The locations on Antarctica where 16 modern populations
635 of Adélie penguins and the 17 relic colonies containing of sub-fossil remains,
636 were sampled. Circled numbers correspond to the sites listed in tables 1 and
637 2. The area marked by dots show the two largest marine based ice shelves,
638 both fed from the West Antarctic Ice Sheet. The dashed line in the insert
639 marks the grounding line of the Ross Ice Shelf at the last glacial maximum
640 (LGM) and during its retreat 7,600 yr BP. Each pie diagram indicates the
641 relative frequencies of the *A* (black) and *RS* (white) lineages in each regional
642 grouping. (*C, D and E*) The sub-fossil sampling locations on the coast of the
643 Ross Sea, and the ages and lineages for all 96 ancient DNA sequences
644 recovered from three different time periods.

645

646 FIG 2. – The neighbor-joining tree of the 557 modern and 96 ancient Adélie
647 penguin HVR-1 sequences and the outgroup is represented by the chinstrap
648 and gentoo penguins. The monophyly of Antarctic (*A*) and Ross Sea (*RS*)
649 Adélie penguin lineages was well supported by the bootstrap analysis (100%)
650 and Bayesian inference (1.0). The dark circles indicate the positions of the
651 three t_{MRCA} estimates.

652

653 FIG 3. – The temperature changes over the last 420-kyr BP from present day
654 at Vostok on the East Antarctic Ice Sheet are from (Petit et al. 1999). The
655 interglacial periods are interpreted from the temperature profile. Above the
656 temperature graph are the posterior probability densities for the times to the

most recent common ancestor using the 'Adélie penguin' rate (left) and the avian 'phylogenetic' rate (right). Arrows indicate the median value of divergence times of each distribution.

FIG 4. – The temperature changes on the East Antarctic Ice Sheet over the last 200-kyr BP from (Petit et al. 1999) and the posterior probability densities for the times to the most recent common ancestor of each mitochondrial lineage. The t_{MRCA} for the *A* and *RS* lineages were estimate separately using both the 'Adélie penguin' rate (left) and the 'phylogenetic' rate (right). Arrows indicate the median value of divergence times of each distribution.

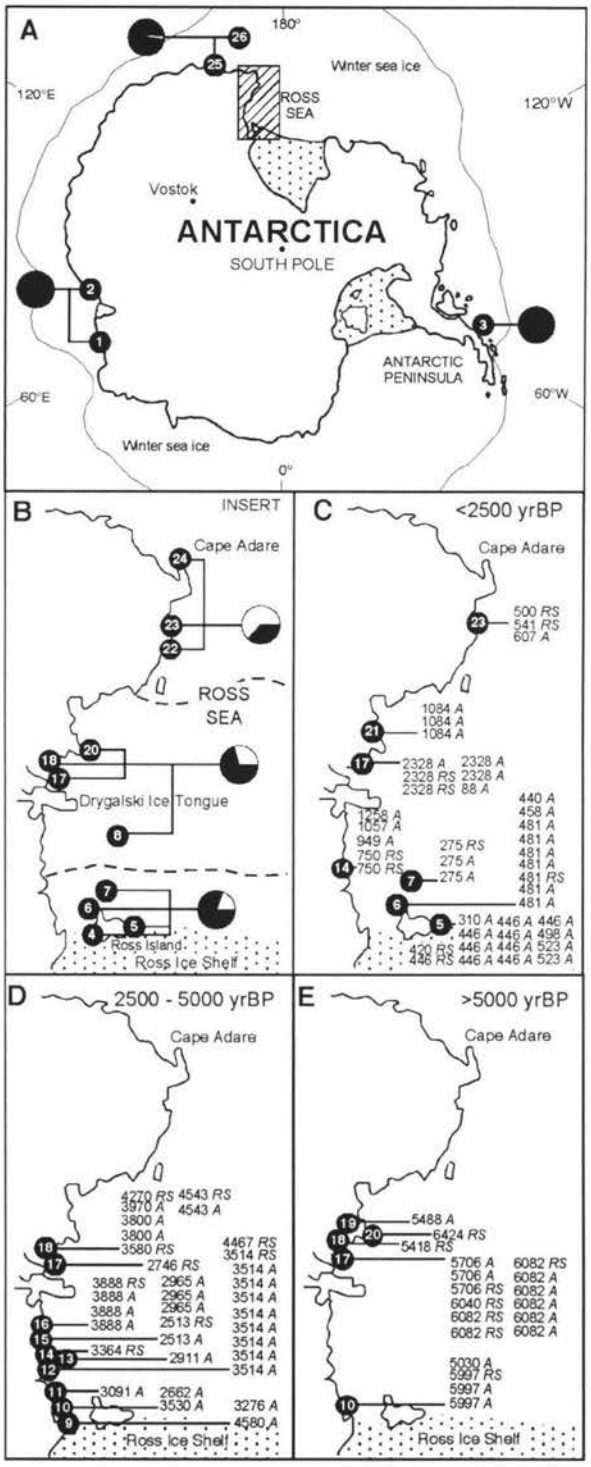

A
180°
Winter sea ice
120°E
120°W
ROSS SEA
Vostok
ANTARCTICA
SOUTH POLE
ANTARCTIC PENINSULA
60°E
60°W
Winter sea ice
0°
B
INSERT
Cape Adare
ROSS SEA
Drygalski Ice Tongue
Ross Island
Ross Ice Shelf
C
<2500 yrBP
Cape Adare
500 RS
541 RS
607 A
1084 A
1084 A
1084 A
2328 A 2328 A
2328 RS 2328 A
2328 RS 88 A
1258 A
1057 A
949 A
750 RS
750 RS
275 RS
275 A
275 A
440 A
458 A
481 A
481 A
481 A
481 A
481 RS
481 A
481 A
310 A 446 A 446 A
446 A 446 A 498 A
420 RS 446 A 446 A 523 A
446 RS 446 A 446 A 523 A
D
2500 - 5000 yrBP
Cape Adare
4270 RS 4543 RS
3970 A 4543 A
3800 A
3800 A
3580 RS
2746 RS
3888 RS 2965 A
3888 A 2965 A
3888 A 2965 A
3888 A 2513 RS
2513 A
3364 RS
2911 A
3091 A 2662 A
3530 A
4467 RS
3514 RS
3514 A
3514 A
3514 A
3514 A
3514 A
3514 A
3514 A
3276 A
4580 A
Ross Ice Shelf
E
>5000 yrBP
Cape Adare
5488 A
6424 RS
5418 RS
5706 A 6082 RS
5706 A 6082 A
5706 RS 6082 A
6040 RS 6082 A
6082 RS 6082 A
6082 RS 6082 A
5030 A
5997 RS
5997 A
5997 A
Ross Ice Shelf

Fig. 1.

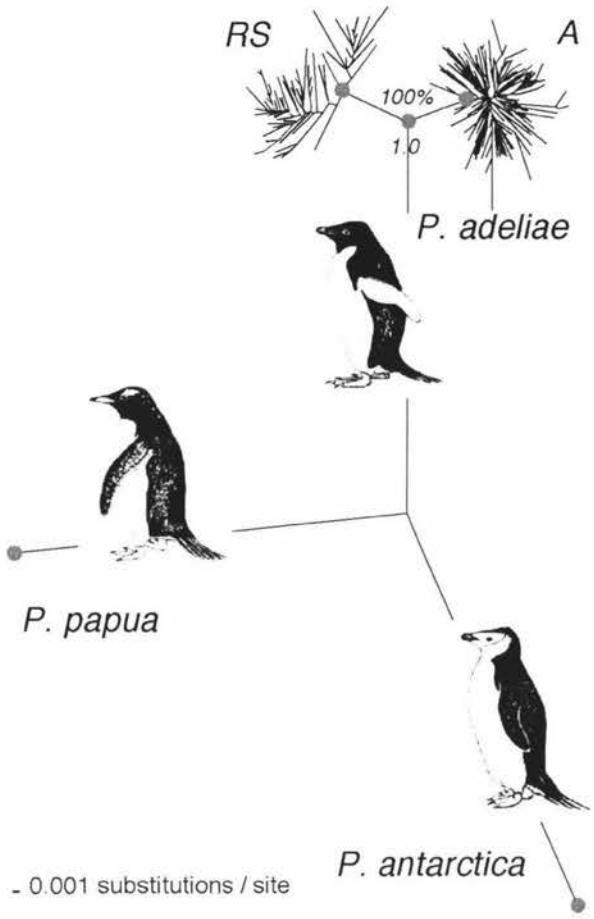

RS
A
100%
1.0
P. adeliae
P. papua
P. antarctica
0.001 substitutions / site

Fig. 2.

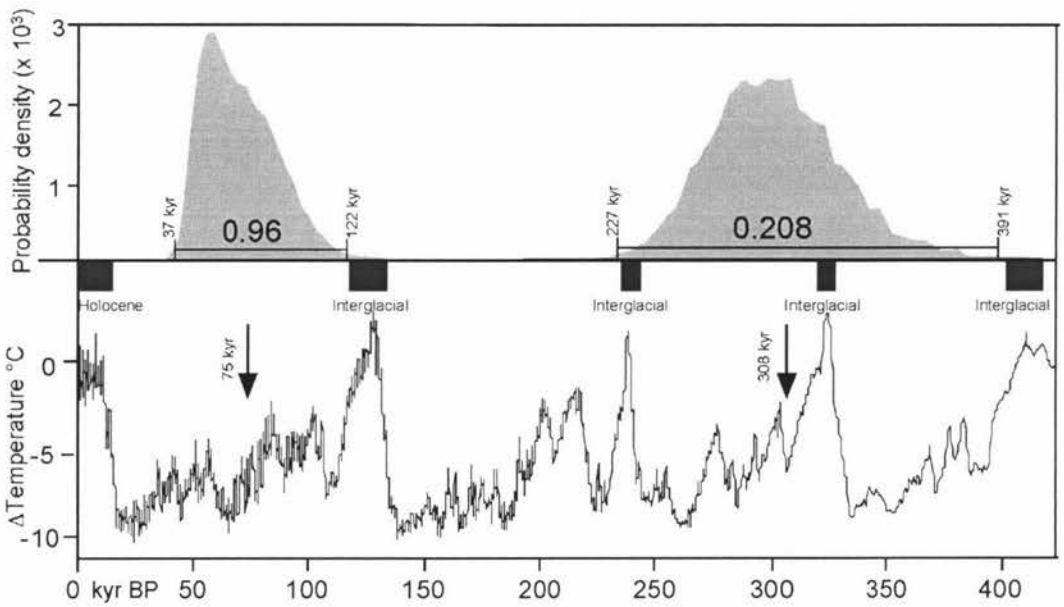

Probability density (x 10³)
3
2
1
37 kyr
0.96
122 kyr
227 kyr
0.208
391 kyr
Holocene
Interglacial
Interglacial
Interglacial
Interglacial
ΔTemperature °C
0
-5
-10
75 kyr
308 kyr
0 kyr BP
50
100
150
200
250
300
350
400

Fig. 3.

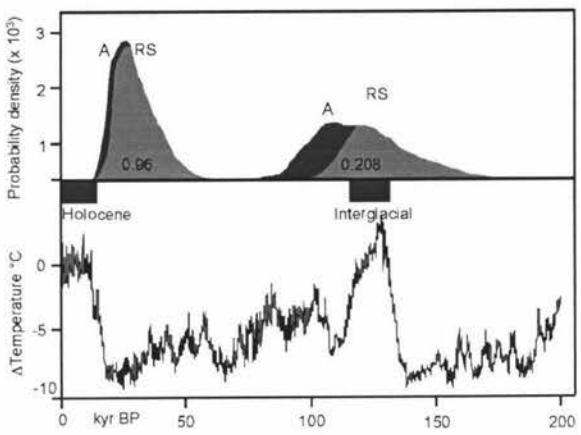

Probability density (x 10³)
3
2
1
A
RS
0.96
A
RS
0.208
Holocene
Interglacial
ΔTemperature °C
0
-5
-10
0
kyr BP
50
100
150
200

Fig. 4.

Table 1

The locations in Antarctica where modern penguin populations were sampled.

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

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.

Table 2

The locations in Antarctica where sub-fossil bones were sampled.

	Location	*n*	Calibrated yr BP	*A*	*RS*
5	Cape Crozier	14	310-523	11	3
6	Cape Bird	10	440-481	9	1
7	Beaufort Island	3	275	2	1
9	Marble Point	2	4580	2	0
10	Dunlop Island	6	2662-5997	5	1
11	Cape Roberts	1	3091	1	0
12	Cape Ross	10	3514-4467	8	2
13	Depot Island	1	2911	0	1
14	Cape Day	1	3364	0	1
15	Cape Hickey	5	2513-2965	4	1
16	Prior Island	9	750-3888	6	3
17	Inexpressible Island	18	2328-6082	10	8
18	Adélie Cove	1	5418	0	1
19	Northern Foothills	8	3580-4270	6	2
20	Gondwana Station	1	6424	0	1
21	Edmonson Point	3	1086-1162	3	0
23	Cape Hallett	3	500-607	1	2
	Pooled	96	275-6424	68	28

NOTE. — The sample sizes (n), the range of assigned ^{14}C-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.