The Evolution of the Mitochondrial DNA Control Region in the Adélie Penguins of Antarctica

Peter A. Ritchie

A Thesis Presented for the Degree of Doctor of Philosophy
in Molecular BioSciences at Massey University
February, 2001

The Adélie Penguins of Antarctica

Cape Bird, 1998

Abstract

Rates of evolution have been important components of many phylogenetic and population genetic studies. In Antarctica, underlying active and abandoned Adélie penguin (*Pygoscelis adeliae*) rookeries, are well preserved subfossil bones. This thesis aimed to estimate a rate of mitochondrial (mt) DNA evolution by sampling sequences from serially preserved subfossil bones through time, and compare these to sequences from modern populations. During this study four field expeditions were undertaken in Antarctica during the austral summers, resulting in the collection of over 400 blood and 329 subfossil bone samples. This thesis research showed that the Adélie penguin mtDNA control region is unusually long (1768 b.p.) and contains a repeat complex at the 3'-end. A phylogeny of 17 penguin species (suborder Sphenisciformes) was constructed from *rns* and *rnl* mtDNA gene sequences. The characteristics (i.e. heteroplasmy and length variation) of control region sequences from penguin species were plotted onto this phylogeny as a mechanism for investigating their evolution. DNA sequence variation from a total of 381 modern Adélie penguin samples revealed the presence of two distinct maternal lineages (7.1% net sequence difference). One lineage is present in all locations around Antarctica (*A*) sampled and the other was recorded only in Ross Sea populations (*RS*). The phylogeographic pattern of the *A* and *RS* lineages suggests Adélie penguins were restricted to distinct ice-age refugia during the last glacial cycle. Ancient DNA was extracted from 80 subfossil bones (^{14}C dates ranged from 310-6082 years before present) from 16 locations on the coast of the Ross Sea. The ancient DNA from these frozen subfossil Adélie penguins is extraordinarily well preserved. Using both modern and ancient DNA sequences a rate of mtDNA control region evolution was determined. This estimated rate is five times higher than previous estimates for the avian mtDNA control region. This rate was then used to time the divergence of the two lineages *A* and *RS*, and showed they split 83 kyr BP during the last glacial cycle. Adélie penguins appear to have endured dramatic changes to their habitat during the ice-ages of the Late Pleistocene.

Acknowledgments

I would first like to thank my supervisor David Lambert for his continued support and encouragement throughout my graduate studies. It was his broad and innovative thinking (and persistence) that started and finished this research project. I will always be thankful for everything he has so generously given.

We really have done the 'hard yards up the middle' on this project.

This project was made possible by a grant from the Marsden Fund of New Zealand to David Lambert (96-MAU-ALS-0030). Additional support was from Antarctica New Zealand, the US Coastguard, and the Programme Nazionale di Ricerche in Antartide (Italy). I wish to acknowledge the support of a Doctoral Scholarship from Massey University and New Zealand Post for their Antarctic scholarship.

Craig Millar. He is one of my great friends and has always helped me whenever I needed it. I don't even know where to begin when I think of all the things for which I need to acknowledge Craig. I just look forward to another ten years of doing it all again!

I need to thank all the kind people who have help me along the way. Thanks to Alexei Drummond, Allan Rodrigo and Roald Forsberg for all their helpful discussions and for making available a pre-release of their software. I am grateful to David Penny and Peter Lockhart for their encouragement and most importantly their interest in the science. I wish to thank Barbara Holland who constructed all the medium networks and calculated a rate from them. Rick Ward kindly allowed me to work in his laboratory at Oxford University, and Alan Cooper was such a great help and so hospitable - I am in debt to both of them.

Thanks to my colleagues in the Molecular Ecology Laboratory. For help with this thesis I am thankful to Amy Roeder, Lara Shepherd, Jennie Hay, Oliver Berry, Hilary Miller, Deputy-head boy Dr Leon Huynen, Richelle Marshall, Quanah Hudson (da Q-dogg), Joanne Chapman, Alex Quinn, Patricia Stapleton, Gillian Gibb, Niccy Aitken, Lee Davies, Judith Robins and Ed Minot. Cheers to Peter Cleary, Rex Hendry, Davie Robertson, Paul Barrett and David Ainley for their help and assistance in Antarctica. I need to especially acknowledge Carlo Baroni who hosted me at Terra Nova Station and worked so hard with us on this project – I am in your debt Carlo!

The following people kindly assisted with the collection of Adélie penguin blood samples: Euan Young, David Lambert, Craig Millar, Judy Clark, Knowles Kerry, Carol Vleck. The following kindly supplied me with penguin samples: Graeme Elliot, Kath Walker and Peter Moore (Department of Conservation, NZ); Boris Culik (Institut für Meereskunde an der Universität Kiel, Germany); Kerri-

Anne Edge, Allan Baker, Cindy Hull, Corey Bradshaw, Janier González, Gerry Kooyman. Kathleen Newman and Liz Grant provided illustrations.

I must thank all my great friends who are such an important part of my life. They never realised how they kept my sanity intact throughout this thesis. Jonny Horrox (you are always the legend!), Felix Collins (you're a great man), James Bower (amazing - you came from Mad Ave in Glendowie, but we still have so much in common), Reece Fowler, Graham Franklin, Wayne Linklater, Elissa Cameron, Tarmo Põldma, Grant Blackwell, Alex Quinn, and Oliver Berry (you're a great guy, even though you're an Aussie). I have really enjoyed my friendship with Melissa and Amy, even though they both suppressed their very dodgey sides.

I will never been able to thank to Anna Grant enough. I couldn't have achieved all of this work without her. Thankyou Anna for your friendship and for everything you are.

Finally, it has been my mother, Beatrice Ritchie, who has given me so much. Without her support and belief in me I would never have achieved so much in life. My appreciation for her will never cease. Thanks to my family Anne, Ja'rome, Simon, Valary and Philip – you're all so great to have around. This thesis is dedicated to my father who passed away and never managed to see what he wanted me to achieve in life.

There are so many unusual paradoxes in life. I returned to New Zealand because my father passed away and I started this PhD. I should be in Paris now, handing in another dissertation. However, I can't begin to describe the immense enjoyment I have had from this thesis. From such a defining moment in my life came such an opportunity.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

A	adenine
A	Antarctic (mtDNA lineage type I)
AMS	accelerator mass spectrometry
BSA	bovine albumin serum
BER	base excision repair
b.p.	base pairs
C	cytosine
^{14}C	carbon-14
°C	degree Celsiuis
CR	control region
d	DNA sequence distance
D	dextro
g	grams
G	guanine
h	haplotypic diversity
H	heavy
HVR-1	hypervariable region I
HVR-2	hypervariable region II
kb	kilo bases
L	light (mtDNA) or levo (amino acids)
L	likelihood distances
LRT	likelihood ratio test
m	metres
M	moles
ML	maximum likelihood
MP	maximum parsimony
mt	mitochondrial
mtDNA	mitochondrial DNA
n	sample size
π	nucleotide diversity
NJ	neighbor-joining
P	probability
RRT	relative rate test
RS	Ross Sea (mtDNA lineage type II)
s/s/Myr	subsitutions/site/million years
T	thymine
ti	transition
tv	transversion
µl	microlitres
UPGMA	unweighted paired-group method with arithmetic means
UV	ultra-violet
yr	years
yr BP	years before present

Animal Ethics and Antarctic Permits

All blood sampling of Adélie penguins from Antarctica was performed under approval of the Massey University Ethics Committee, protocol numbers 96/146 and 99/160. Approval to restrain, take blood from Adélie penguins, and to enter Specially Protected Areas (SPA) and Sites of Special Scientific Interest (SSSI) was given by *Antarctica New Zealand* permit numbers 96/4, 97/5, 98/7, 99/9.

Chapter One

Rates of Mitochondrial DNA Evolution

1.1 Estimating Substitution Rates

Genealogical reconstruction is used by biologists as a central tool to explain the evolutionary diversification of life on earth. The advent of nucleotide and protein sequencing techniques has had a profound impact on this mode of explanation. In particular, DNA sequence data have allowed rates of substitution per unit time to be estimated. Rate estimates have been used to time the global expansion of *Homo sapiens*, to show changes in species distributions during the periodic ice-ages of the Pleistocene, and to describe major explosions of species diversity during the Cambrian.

An average rate of nucleotide substitution is typically calculated by measuring the sequence difference among extant taxa, and dividing this by the age of the most recent common ancestor. This age is usually estimated by reference to a known age of fossil material. Shields and Wilson (1987) were the first to calibrate a rate of evolution for mitochondrial DNA (mtDNA) in birds, using two genera of geese. Their study used restriction fragment length polymorphisms (RFLPs) to propose a phylogeny of two subspecies from the genus *Branta* (the brant and Canada goose)

and three subspecies from the genus *Anser* (the white-fronted goose, Ross' goose and the snow goose). The midpoint root of the phylogenetic tree represented approximately 9% sequence difference between the two genera, and the oldest fossils from each genus were dated at 4-5 million years (Myr) (Figure 1.1). Dividing 9% by 4.5 Myr gave a mean rate of divergence of 2%/Myr, i.e. 0.02 substitutions/site/Myr (s/s/Myr). These authors compared this rate to a similar value suggested for mammals.

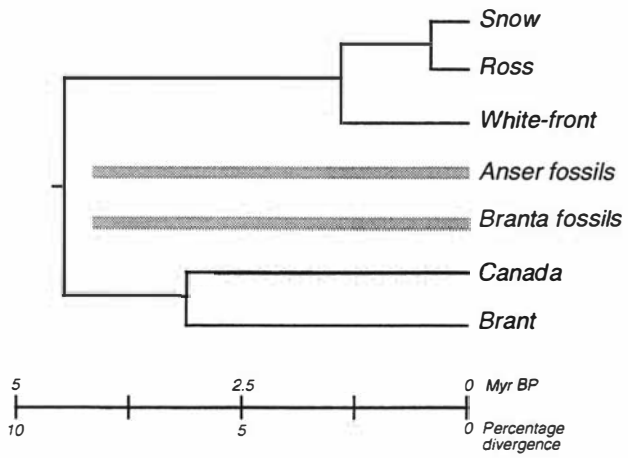


Figure 1.1. The avian mtDNA molecular clock calibration from geese by Shields and Wilson (1987). The thicker greyed lines indicate the oldest fossils (4-5 million years old) found in Mongolia and California. The bottom scale indicates both time in millions of years before present (Myr BP) and the percentage sequence difference.

Quinn (1992) determined the DNA sequence of a 178 base pair (b.p.) portion of the 5'-end (part of the hypervariable region I, HVR-1) of the mtDNA control region to investigate the phylogeography of the snow goose. An earlier RFLP study of the entire mtDNA by Quinn et al. (1991) showed that there was 1.3% sequence difference between two *Branta* subspecies, and Quinn's (1992) study revealed 13.5% corrected sequence difference in the control region between the same two subspecies. Therefore, Quinn (1992) calculated that the control region evolves (13.5/1.3 =) 10.4 times faster than the overall mtDNA. This rate difference was then multiplied by the Shields and Wilson (1987) rate of substitution for the entire mt genome (10.4 x 2%) to arrive at a rate of 20.8%/Myr (0.208 s/s/Myr) for the control region. Subsequent authors (e.g. Baker and Marshall, 1997) adopted Quinn's rate for the control region in populations of bird species. Wenink et al. (1996) extrapolated this rate further in dunlins (*Calidris alpina*) to include a slower-evolving portion of the control region (the central conserved domain, CCD). These authors argued that the CCD was 2.36 times more conserved than the flanking 5'-

end. Therefore, using Quinn's earlier estimate, an average rate across both regions was calculated as: (20.8 + (20.8/2.36))/2 = 14.8%/Myr (0.148 s/s/Myr). Quinn's (1992) result is currently the commonly used estimate for substitution rates of the avian mt control region.

In recent years a controversy has arisen over the rate of substitution in the human mtDNA control region (Jazin et al., 1998). Longitudinal sampling of humans over three generations spanning 40 years (yr) (Parsons et al., 1997) has provided a mutation rate estimate (2.6 s/s/Myr) which is much higher than those calculated using a phylogenetic approach (e.g. 0.118 s/s/Myr). Furthermore, Denver et al. (2000) recently reported a direct measurement of the rate of mutation from 74 *Caenorhabditis elegans* mt genomes, over 214 generations that followed inbred single-progeny descent. This study revealed a mutation rate for the entire mtDNA that was two orders of magnitude higher than predicted from indirect phylogenetic methods. It is apparent from these studies that direct measurements of the mtDNA mutation rate do not match indirect measurements of substitution. The phylogenetic and pedigree studies of maternally inherited mtDNA have used very similar approaches to estimating a rate (Sigurðardóttir et al., 2000). If two sequences are compared from two species (the phylogenetic approach), at some point in history they both originated from the same ancestral sequence. The difference between the two sequences is the number of mutations that have accumulated in each lineage since their divergence, provided selection, back mutations, and multiple mutations at the same site are negligible. Explanations for the conflicting estimates include: selection and drift (Parsons et al., 1997); heterogeneity of the rate amongst sites (Jazin et al., 1998); non-reporting of zero rate estimates (Macaulay et al., 1997); and sampling of older subjects who may have accumulated mutations during ageing (Michikawa et al., 1999).

The estimation of a substitution rate for a measurably evolving population is another issue. Differences in evolutionary rate can arise from variations in the mutation rate and/or variation in the fixation rate (Mindell and Thacker, 1996). It has been known for many years that in a population, the rate of substitution (k) of an allele is a product of the population size ($2N$ generations), mutation rate per individual (μ), and probability of fixation (Kimura and Ohta, 1971). Mutations may have arisen with a high frequency, but these will not be demonstrated by the

variability in existing alleles because of selection. On the other hand, according to the neutral theory (Kimura, 1983) the substitution rate is simply equal to the mutation rate ($k = \mu$). However, there will always be a large number of transient neutral alleles in a large population due to random genetic drift (Figure 1.2). Hence, a natural population of mtDNA haplotypes evolves as a combination of mutation and stochastic lineage turnover.

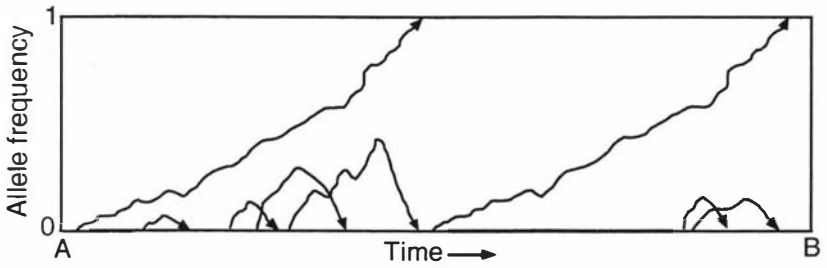


Figure 1.2. Changes in the frequency of neutral mutations over time. Some mutations reach fixation while some are transient in the population. If a sample taken at time A were compared to one taken at time B, the number of mutations in this time period would be recorded as two. In reality eight mutations had occurred during this period of time. Adapted from Nei (1987, p.383).

Mitochondrial DNA is maternally transmitted and therefore has only one-quarter the effective population size of nuclear alleles. In natural populations this clonal feature of mtDNA means haplotypes are highly susceptible to extinction, especially when there are changes in demographic parameters (Avise, 1984). If every mother in a population simply transmitted her mtDNA to a single daughter and so on, the number of haplotypes would be equivalent to the mutation rate. Natural populations are never deterministic, and females show variability in their contribution to the next generation. If the production of daughters by mothers in a population is considered a random process with non-overlapping generations, then it can be modelled according to a Poisson distribution (Avise, 1984). Therefore, over an increasing number of generations the probability of survivorship for independent founding lineages in a population will decrease and asymptote. Furthermore, population size, family size, and expansion or decline of populations all dramatically affect the probability of survival or extinction of lineages. For example, consider the effect of population size under a Poisson distribution where the mean equals one. In a population of two females there is a probability of 0.16 for the survival of a founding lineage after ten generations, whereas in a population size of 10,000 females such a low probability is not reached until around 100,000 generations. Avise et al. (1987) described this stochastic lineage turnover in mtDNA

trees as "self-pruning" events. The mutation rate in matrilines is not equivalent to the rate of substitution in evolving populations.

Rates of substitution, either calibrated over many millions of years or measured directly in maternal lines, are often used as the rate of change for an evolving population. Given the conflict between direct and indirect measurements, and the theoretical ideas underlying coalescent theory (Hudson, 1990), there should be attempts to match the measurement of substitution rates with the time frame of their application. It is clear that a substitution rate needs to be estimated for a population of mtDNA sequences over a time period of ten to one hundred thousand years.

1.2 Ancient DNA: The Opportunity to Measure Evolution

Ancient DNA technology (Pääbo, 1989) potentially offers the opportunity to measure mtDNA changes from serially preserved samples spanning thousands of years. This approach is not constrained to postulate divergence times among extant taxa. In contrast to the phylogenetic method of estimating rates, serial sampling of populations through time can reveal both changes in nucleotides and haplotype frequency. The approach of sampling sequences at known timepoints, measures the successive divergence of one lineage from another (Figure 1.3). Such serial sampling of mt lineages requires large sample sizes of both modern and ancient sequences. An estimate of the substitution rate includes mutation and lineage turnover. Figure 1.3 is an example of these processes when five related lineages are sampled at various points in time. First, lineage E diverged from lineage A through the mutational steps of lineages B, C and D; the rate of substitution is estimated from sampling these cumulative mutations. Secondly, if lineage E was present at the same time as B, C and D, but these latter three lineages were not detected in the modern population, then the rate is the result of lineage turnover. In this latter case, different combinations of lineages B, C, D and E at each timepoint may give a negative rate of divergence from lineage A. However, a large sample of lineages would provide an average rate applicable to an evolving population.

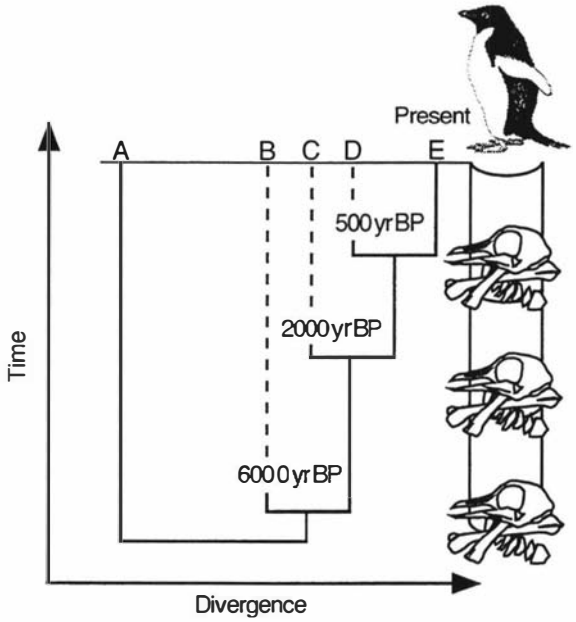


Figure 1.3. The principle of measuring mtDNA substitutions from serial sampling. Lineage E has diverged from lineage A over time. Serial sampling of ancient sequences allows the measurement of a mutational pathway from lineage A to E through lineages B, C and D.

To date there has only been a handful of ancient DNA studies reporting non-contemporaneous population samples of DNA sequences. Thomas et al. (1990) investigated temporal changes in a 135-b.p. portion of the control region from three subspecies of the Panamint kangaroo rat (*Dipodomys panamintinus*). Samples spanning a 78-year period from 63 modern and 49 museum specimens showed the maintenance of similar levels of genetic variation over time. The small sample size, a complex of subspecies, and the short span of sampling times precluded a substitution rate from being estimated in this study. Temporal variability of a 233-b.p. domain in the apocytochrome b (*cob*) gene over an 11,000 year period has been analysed from the European rabbit (*Oryctolagus cuniculus*) (Hardy et al., 1995). The study by Hardy et al. revealed stability of haplotypes over time until the Middle Ages. The sudden appearance of a new haplotype in samples several hundred years old, was suggested to be the result of humans moving rabbits with this haplotype into new locations as societies developed on the continent. Leonard et al. (2000) compared seven ancient DNA sequences from the control region (178 b.p.) and *cob* gene (258 b.p.) of the brown bear (*Ursus arctos*), to previously published data sets of 317 control region and 166 *cob* sequences. Sampling of ^{14}C-dated material spanning 14 to 34 kyr before present (BP), revealed haplotypes never seen in the modern populations. These authors concluded that populations were genetically more diverse 14,000 yr ago. These three studies, as well as Bailey et al.

(1996) and Wang et al. (2000), has been informative but none has estimated a rate of change for their respective DNA sequences over time.

To obtain a rate estimate over time for a vertebrate population, a study needs to target a DNA sequence that is likely to evolve over a period of thousands of years; probably the best candidate is the HVR-1 from the mtDNA control region. Furthermore, the ideal situation would include a large number of extant individuals and a large number of undisturbed, and well preserved, hard tissue samples that have been deposited continuously over thousands of years.

1.3 Antarctica: The 'Ideal Situation'

The Adélie penguins (*Pygoscelis adeliae*) of Antarctica represent an ideal model for serial sampling from the last 13,000 years. This species represents approximately 70% of the total avian biomass of the Antarctic region (Everson, 1977), with a total population size estimated at 2.6 million breeding pairs and ten million immature birds (Croxall, 1985; Woehler, 1993). Adélie penguins nest in colonies on the relatively small proportion of the Antarctic coastline that is ice-free during the austral summer. The Adélie penguins have been present in large numbers along the coast of the Ross Sea for at least the last 7,000 yr (Baroni and Orombelli, 1994).

Breeding success of Adélie penguin individuals improves with age, and mature females produce two eggs during the summer months. Therefore a reproducing female produces on average one daughter each season. However, of the chicks that hatch, one-quarter are estimated to die before fledging, a large proportion of these deaths occurring before a crèche forms (Ainley et al., 1983). Most juvenile deaths are the result of desertion of the nest by young parents, followed by predation of eggs and chicks by the Antarctic skua (*Catharacta maccormickii*).

The pristine and stable Antarctic environment has led to significant deposits of well-preserved penguin remains in the ornithogenic soils underlying colonies (Ugolini, 1972). Stratified penguin guano underlying existing and abandoned (relic) rookeries on the coast of the Ross Sea contains subfossil bones and feathers. These have been ^{14}C-dated to as old as 13,000 yr BP (Baroni, 1994). Antarctica is the coldest and driest continent on earth. Mean monthly temperatures in Antarctica are

always below 0°C. During winter, temperatures at high latitudes resemble those in a commercial freezer and consequently are ideal for slowing down biological processes of decomposition. For example, old freeze dried seal carcasses are a common feature of the Antarctic landscape (Barwick and Balham, 1967). The combination of large extant penguin populations and extensive deposits of well-preserved subfossil bones spanning thousand of years, make this an ideal situation to sample population-level changes in mtDNA.

1.4 Thesis Outline

The aim of my thesis research was to estimate the rate of evolution for mtDNA HVR-1 in Adélie penguins over a period of thousands of years. To achieve this I have:

1. Reviewed the features of the mitochondrial genome and the processes that underlie mutations (Chapter Two)
2. Sequenced the Adélie penguin mtDNA control region; identified the HVR-1 and designed specific PCR primers (Chapter Three).
3. Established a phylogeny for penguin species and identified a suitable outgroup for the Adélie penguin HVR-1 sequences. This also allowed a comparative analysis of control regions across all penguin species (Chapter Four).
4. Characterised the DNA sequence variation in the HVR-1 from a large sample of modern Adélie penguins, across numerous Antarctic populations (Chapter Five).
5. Extracted and sequenced DNA from a large number of ^{14}C-dated subfossil bones (Chapter Six).
6. Presented an estimated substitution rate for the mtDNA control region that utilised both modern and ancient DNA samples (Chapter Seven).

1.5 References

Ainley, D.G., LaReseche, R.E., and Sladen, W.J. 1983. Breeding Biology of Adélie Penguins. University of California Press, Berkeley.

Avise, J.C., Arnold, J., Ball, R.M., Bermingham, E., Lamb, T., Neigel, J.E., Reeb, C.A., and Saunders, N.C. 1987. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. Annual Review of Ecology and Systematics **18:** 489-522.

Avise, J.C., Neigel, J.E., and Arnold, J. 1984. Demographic influences on mitochondrial DNA lineage survivorship in animal populations. J. Mol. Evol. **20:** 99-105.

Bailey, J.F., Richards, M.B., Macaulay, V.A., Colson, I.B., James, I.T., Bradley, D.G., Hedges, R.E.M., and Sykes, B.C. 1996. Ancient DNA suggests a recent expansion of European cattle from a diverse wild progenitor species. Proc. Roy. Soc. Lond B **263:** 1467-1473.

Baker, A.J., and Marshall, H.D. 1997. Mitochondrial control region sequences as tools for understanding evolution. *In* Avian Molecular Evolution and Systematics. *Edited by* D. P. Mindell. Academic Press, San Diego. pp. 51-82.

Baroni, C. 1994. Notes on late-glacial retreat of the Antarctic ice sheet and Holocene environmental changes along the Victoria Land coast. Memoirs of the National Institute of Polar Research, Special Issue **50:** 85-107.

Barwick, R.E., and Balham, R.W. 1967. Mummified seal carcases in a deglaciated region of South Victoria Land, Antarctica. N. Z. J. Geol. Geophys. **11:** 165-180.

Croxall, J.P. 1985. The Adélie penguin (*Pygoscelis adeliae*). Biologist **32:** 165-170.

Denver, D.R., Morris, K., Lynch, M., Vassilieva, L.L., and Thomas, W.K. 2000. High direct estimate of the mutation rate in the mitochondrial genome of *Caenorhabditis elegans*. Science **289:** 2342-2344.

Everson, I. 1977. The Southern Ocean: The living resources of the Southern Ocean. Food & Agriculture Organisation report GCO/SO/77/1 Southern Oceans Fishery Survey Program, Rome, Italy.

Hardy, C., Callou, C., Vigme, J.-D., Casane, D., Dennebouy, N., Mounolou, J.-C., and Monnerot, M. 1995. Rabbit mitochondrial diversity from prehistoric to modern times. J. Mol. Evol. **40:** 227-237.

Hudson, R.R. 1990. Gene genealogies and the coalescent process. Oxford Survey of Evolutionary Biology **7:** 1-44.

Jazin, E., Soodyall, H., Jalonen, P., Lindholm, E., Stoneking, M., and Gyllensten, U. 1998. Mitochondrial mutation rate revisited: hot spots and polymorphism. Nat. Genet. **18:** 109-110.

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

Kimura, M., and Ohta, T. 1971. Theoretical Aspects of Population Genetics. Princeton University Press, Princeton.

Leonard, J.A., Wayne, R.K., and Cooper, A. 2000. Population genetics of Ice Age brown bears. Proc. Natl. Acad. Sci. USA **97:** 1651-1654.

Macaulay, V.A., Richards, M.B., Forster, P., Bendall, K.E., Watson, E., Sykes, B., and Bandelt, H.-J. 1997. mtDNA mutation rates – no need to panic. Am. J. Hum. Genet. **61:** 983-986.

Michikawa, Y., Mazzucchelli, F., Bresolin, N., Scarlato, G., and Attardi, G. 1999. Aging-dependent large accumulation of point mutations in the human mtDNA control region for replication. Science **286:** 774-778.

Mindell, D.P., and Thacker, C.E. 1996. Rates of molecular evolution: phylogenetic issues and applications. Annu. Rev. Ecol. Syst. **27:** 279-303.

Nei, M. 1987. Molecular Evolutionary Genetics. Columbia University Press, New York.

Pääbo, S. 1989. Ancient DNA: Extraction, characterization, molecular cloning, and enzymatic amplification. Proc. Natl. Acad. Sci. USA **86:** 1939-1943.

Parsons, T.J., Muniec, D.S., Sullivan, K., Woodyatt, N., Alliston-Greiner, R., Wilson, M.R., Berry, D.L., Holland, K.A., Weedn, V.W., Gill, P., and Holland, M.M. 1997. A high observed substitution rate in the human mitochondrial DNA control region. Nat. Genet. **15:** 363-368.

Quinn, T., Shields, G.F., and Wilson, A.C. 1991. Affinities of the Hawaiian goose based on two types of mitochondrial DNA data. Auk **108:** 585-593.

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.

Shields, G.F., and Wilson, A.C. 1987. Calibration of the mitochondrial DNA evolution in geese. J. Mol. Evol. **24:** 212-217.

Sigurðardóttir, S., Helgason, A., Gulcher, J.R., Stefansson, K., and Donnelly, P. 2000. The mutation rate in the human mtDNA control region. Am. J. Hum. Genet. **66:** 1599-1609.

Thomas, W.K., Pääbo, S., Villablanca, F.X., and Wilson, A.C. 1990. Spatial and temporal continuity of kangaroo rat populations shown by sequencing mitochondrial DNA from museum specimens. J. Mol. Evol. **31:** 101-112.

Ugolini, F.C. 1972. Ornithogenic soils of Antarctica. *In* Antarctic Terrestrial Biology. *Edited by* G. Llano. Antarctic Research Series, American Geophysical Union. pp. 181-193.

Wang, L., Oota, H., Saitou, N., Jin, F., Matsushita, T., and Ueda, S. 2000. Genetic structure of a 2,500-year-old human population in China and its spatiotemporal changes. Mol. Biol. Evol. **17:** 1396-1400.

Wenink, P.W., Baker, A.J., Rösner, H.-U., and Tilanus, M.G.J. 1996. Global mitochondrial DNA phylogeography of Holarctic breeding dunlins (*Calidris aplina*). Evolution **50:** 318-330.

Woehler, E.J. 1993. The distribution and abundance of Antarctic and subantarctic penguins. SCAR, Cambridge.

Chapter Two

The Mitochondrial Genome: Maternal Inheritance and Mutations

2.1 Introduction

The mitochondrial (mt) genome evolves as a combination of mutation and stochastic lineage extinction. The latter issue was dealt with largely in Chapter One. This chapter discusses a number of issues relating to mutations in the mt genome that are of significance to the thesis generally. For example, there has been debate on the source of the mt genome's high mutation rate and recent studies have linked this to the transient single-stranded feature of replication (Gissi et al., 2000). Moreover, the mt genome is unusual because: (1) it follows maternal inheritance; (2) there can be several variant genomes within a single individual; (3) a genotype can change over one generation. Many of these features have been studied in mammals but are applicable to avian species. The biological properties of the mt genome are shared by most vertebrate species, and make the estimation of a mutation rate of general importance.

2.2. Maternal Inheritance

A variety of mechanisms are thought to exist which exclude the male mitochondria from being inherited in eukaryotes (Lima-de-Faria, 1983). The flagellum midpiece of the mammalian spermatozoon contains around 50 tightly packed mitochondria, and this tail is incorporated into the oocyte cytoplasm. In mammals, dilution of sperm, oxidative damage and proteolytic destruction have all been suggested as mechanisms of preventing incorporation of male mitochondria (Ankel-Simon and Cummins, 1998). The fact that paternal leakage was found to occur in interspecies crosses of *Mus musculus* and *M. spretus* (Gyllensten et al., 1991; Kaneda et al., 1995) casts doubt on the theories of dilution or oxidative damage. More likely is some form of signal associated with the male germ cell that targets it for destruction. In bovines, sperm mitochondria undergo degenerative changes and disappear by the four-cell stage. Multivesiculated bodies and lysosomes surround the mitochondria targeting them for disintegration (Sutovsky and Schatten, 2000). Sutovsky et al. (1998) showed that during spermatogenesis sperm are tagged by ubiquitin, after which disulfide bond cross-linking temporarily masks the ubiquitin molecules. In cells, ubiquitin is used to tag proteins for destruction. The carboxyl end of an 8.5-k.Da. ubiquitin molecule is covalently attached to lysine amino acid side chains of a damaged protein. A ubiquitin-protein complex is formed and acts as a recognition factor for the cell to deliver the complex to proteasomes to be destroyed by proteolysis. Sutovsky et al. propose that upon arrival in the oocyte, the disulfide bonds on the sperm midpiece could be relieved and the mitochondria coated with recognisable ubiquitin. The mitochondrial of the sperm midpiece are thus recognised, surrounded by lysosomes and/or proteasomes, and destroyed.

2.2.1. *Heteroplasmy and the Oocyte Bottleneck*

Somatic cells contain anywhere between 10 and 1,000 copies of mtDNA, whereas an oocyte possesses around 100,000 copies (Pikó and Taylor, 1987). When a mutant arises in the pool of mtDNA, its presence along with the wild type mtDNA is termed heteroplasmy. Heteroplasmy is not uncommon in the rapidly evolving mtDNA control region, resulting from either point mutations and/or variation of repetitive elements. The level of heteroplasmy varies across tissues types in the body. Jazin et al. (1996) showed that in humans the brain contained higher levels of

heteroplasmy than blood, and linked it with the nonmitotic nature of brain tissue. Furthermore, Calloway et al. (2000) reported slightly higher levels of heteroplasmy in muscle tissue than in the heart, brain or blood. All of these studies, as well as Michikawa et al. (1999), report that heteroplasmy and point mutations increase throughout ageing.

Uniparental transmission of cytoplasmic genomes has been observed throughout most animal and plant species (Birky, 1995). The transmission of only oocyte copies of mtDNA to the next generation is termed the mtDNA "bottleneck" (Bergstrom and Pritchard, 1998). Because there is no mtDNA replication until the 100 cell stage of the blastocyst, mt genomes are randomly assorted into each cell as they divide. This means that early sequestering of precursor germ cells in a developing embryo will cause the selective transmission of a small pool of mtDNA to the next generation. Evidence of this bottleneck comes from rapid shifts in genotype frequencies of heteroplasmic mtDNA over a few generations (Hauswirth and Laipis, 1982; Bendall et al., 1996).

2.3 Mutations in the Mitochondrial Genome

2.3.1. The Metabolic Rate Hypothesis

The mt genome evolves fairly rapidly at the sequence level in comparison to many nuclear genes (Wilson et al., 1985). There is a strong bias for transitional (purine to purine A↔G and pyrimidine to pyrimidine T↔C) over transversional (purine to/or from pyrimidine G↔T, G↔C, A↔T and A↔C) changes (Brown et al., 1982). Variables such as body size, generation time and metabolic rate have all been reported to correlated with mtDNA mutation rates (Rand, 1994). However, the factors affecting the rate of mutations are the subject of some debate. The most informative study on the variability of mutation rates would be a direct comparison of mutations in pedigree lines, from a range of species with differing metabolism, body size and generation time. However, to date all studies have focused on phylogenetic comparisons among groups of species.

Dominating much of the debate on rates of mutation have been the results of Martin and Palumbi (1993). They plotted percentage sequence difference per million years

from RFLPs, against species body mass, revealing a negative correlation. From these results, Martin and Palumbi proposed that species with a higher metabolic rate and a short generation time will accumulate changes in mtDNA sequences more rapidly, when compared to those with a slower metabolism. There are two pivotal points in this theory: first, that the cell metabolism generates oxygen free radicals resulting in oxidative damage to mtDNA and consequently mutations; secondly, there is an accurate estimate of the point of divergence for the taxa compared. Recent studies into the processes of mtDNA mutations deserve attention to be drawn to the former point, oxidative damage.

2.3.2. Oxidative Damage in the Mitochondria

During oxidative phosphorylation in the mitochondria, oxygen is reduced to water. Escaping from this reaction are superoxide anions ($^{\bullet}O_2^-$) which are pacified by superoxide dismutase into the relatively stable hydrogen peroxide ($^{\bullet}O_2^- + {}^{\bullet}O_2^- + 2H^+ \rightarrow H_2O_2 + O_2$). However, highly reactive hydroxyl radicals ($^{\bullet}OH$) are formed via the reduction of H_2O_2 by a metal ion ($Fe^{2+} + H_2O_2 + H^+ \rightarrow Fe^{3+} + H_2O + {}^{\bullet}OH$), in a process known as a Fenton reaction (Breen and Murphy, 1995). In addition, hydroxyl radicals can be formed nonenyzmatically through exposure to ionising radiation (Téoule, 1978). Nitric oxide, which is important in many physiological actions, has also been identified as a damaging free radical ($NO^{\bullet}$ and $NO_2^{\bullet}$). These reactive oxygen and nitrogen species cause oxidative base damage to DNA and as a result have been implicated in mutagenesis, carcinogenesis and ageing (Halliwell and Gutteridge, 1999). The mitochondrial genome is particularly vulnerable to this kind of oxidative insult as it is close to the electron transport chain and lacking protective histones.

The most common oxidative lesion in mtDNA is 8-oxo-deoxyguanine (8-oxo-G), where a $^{\bullet}OH$ has been added to the eighth position of the purine ring structure in guanine. The steady state levels of 8-oxo-2'-deoxyguanosine (oxo8dG), the released 8-oxo-G base, are about 10-fold higher in the mitochondria than in the nucleus (Richter et al., 1988). Interestingly, while 8-oxo-G does not inhibit replication, DNA polymerases will insert either a dC or dA opposite this adduct (Shibutani et al., 1991). Pinz et al. (1995) showed that 27% of the time polymerase-γ caused a G→T transversion via the misincorporation of dA. In contrast, transitional changes

(G↔A and C↔T) are much more common than transversions in mtDNA (Brown et al., 1982; Wilson et al., 1985; Thomas and Beckenbach, 1989). To resolve this paradox a clearer understanding of mutagenesis and DNA repair of the mitochondrial genome is needed.

2.3.3. DNA Damage and Repair Mechanisms

DNA is damaged in many different ways, but commonly involves spontaneous degradation or chemical attack (Lindahl, 1993). In response to DNA damage the cells will repair and correct mistakes. The primary mode of correction relies on recognition of the opposing base and inserting the appropriate corresponding nucleotide (ie. G:C or A:T). DNA repair mechanisms are generally grouped into three classes: Base-Excision Repair (BER); Nucleotide-Excision Repair (NER); and Mismatch Repair (MMR) (Table 2.1). In the mitochondria, the BER pathway has been studied most extensively, there is no evidence for the existence of NER, and MMR is associated with BER and is beginning to be better understood (Bogenhagen, 1999). The absence of the NER explains why bulky pyrimidine dimers caused by exposure to UV-irradiation are not repaired, a result that led to the dogma that mitochondria were repair-deficient (Sawyer and van Houten, 1999). The likely fate of a genome with a bulky DNA lesion, which impairs replication or transcription, is destruction, possibly by Endonuclease G (Ikeda and Ozaki, 1997).

Table 2.1. The DNA repair pathways, their targets and corresponding proteins found in the mitochondria. AP = apurinic/apyrimidinic

Repair Process	Lesions	Enzymes
Base-excision repair (BER)	AP sites & Glycosylase products	AP endonuclease, DNA pol-γ, mtDNA ligase
Nucleotide-excision repair (NER)	Pyrimidine dimers & bulky chemical adducts	None
Mismatch repair (MMR)	Base•Base mispairing & insertions/deletions	DNA pol-γ or mut-S homologues

BER involves the recognition of a damaged base by *N*-glycosidases which then catalyse the hydrolytic cleavage of the *N*-glycosidic bond (sugar-base bond) freeing the base from the sugar-phosphate group (Lindahl et al., 1977). Following base removal, apurinic/apyrimidinic (AP) endonuclease hydrolytically cleaves the phosphodiester bond of the remaining sugar-phosphate, leaving a 3'-OH terminus.

DNA polymerase fills the remaining gap and DNA ligase completes the repair by sealing the nick in the phosphate backbone. The use of AP endonuclease, DNA polymerase and DNA ligase in BER is generic, however, different *N*-glycosidase enzymes recognise and remove specific base damage. The first mitochondrial repair protein to be described was uracil-DNA glycosidase (Domena et al., 1988). The removal of uracil is particularly common in a genome. If a cytosine loses its amine group (NH_2) due to spontaneous hydrolytic deamination, it becomes a uracil, bonds with adenine and causes a C→T transition (Figure 2.2) (Coulondre et al., 1978). Domena et al. (1988) estimated that there are approximately 2 uracils and 100 AP sites expected to form in the mitochondrial DNA of each cell per day. A group of *N*-glycosidases that recognise oxidative damage (MTH1, 8-oxoG glycosidase, and mtMYH), in particular all of those acting on the common adduct 8-oxoG, have recently been characterised (Croteau et al., 1997; Parker et al., 2000). These proteins form specific binding complexes with A, G or T paired with 8-oxoG, A/G mismatches, and 2-hydroxyadenine paired with A, G, T, C or 8-oxoG. These proteins also remove 8-oxoGTP from the dNTP pool and possess glycosidase/AP lyase activity.

In addition to BER there are two other points to mention about the mtDNA repair process. First, although bulky alkylation damage to the mitochondria genome is not repaired, small alkyl (CH_3 – methyl and CH_2CH_3 – ethyl) lesions found on the O^6 position of guanine are easily repaired by alkylguanine transferase (Satoh et al., 1988). Secondly, Thyagarajan et al. (1996) detected the presence of homologous recombination activity on plasmid DNA substrates from crude mammalian mitochondrial extracts and suggested this process could be involved in mtDNA repair. These authors showed inhibition of homologous recombination by treatment with anti-*rec*A antibodies, concluding that this activity was mediated by a homolog of the bacterial strand-transferase protein. This form of recombination repair activity had previously been demonstrated in yeast mitochondria and plant chloroplasts (Thyagarajan et al., 1996). These findings should be followed up by a study that uses highly purified mitochondrial extracts to isolate a product of the *rec*A gene, and demonstrates whether or not this product acts directly on the mitochondrial genome.

2.3.4. *Predictions of the Metabolic Rate Hypothesis*

In many cases mutations are the result of DNA damage followed by the insertion of a different base at the repaired position. Given the current knowledge of repair mechanisms in the mitochondria and the efficiency, with which oxidative damage is repaired, how can mutations in mtDNA and their rate of accumulation be explained? Doubt has been cast upon the predictions of the metabolic rate hypothesis since mtDNA are well equipped to deal with this form of insult as previously described (Croteau et al., 1999). Moreover, based on first principles, the prediction of this theory should be more G→T transversions than transitions. Gissi et al. (2000) conducted a rate analysis of both ribosomal RNA, and first and second positions of Heavy-strand (H-strand) protein coding genes, of 35 complete mtDNA sequences across 15 mammalian orders. Most analyses rely on accurate divergence time estimates. Gissi et al. avoided this, however, by using relative-rate methods with unambiguous outgroup species within their sample. Surprisingly, the rate of the fastest evolving order never exceeded 1.8 times that of the slowest evolving order, and they found no significant correlation between rates of mutation and body size, generation time or metabolic rate. Furthermore, Slowinski and Arbogast (1999) re-analysed Martin and Palumbi's data taking into account rate variation among sites using a Jukes-Canter gamma correction model. The calculated gamma parameter (the average $\alpha = 0.140$) demonstrates a strong rate variation among sites, and once this variation is corrected for, there is no significant correlation between body size and rate of sequence evolution across poikilotherms and endotherms.

2.3.5 *Alternatives to the Metabolic Rate Hypothesis*

Tanaka and Ozawa (1994) present a mutational spectrum from sequences of 43 complete human mitochondrial genomes and show that C→T and A→G transitions were 9- and 1.8-fold more frequent on the H-strand than on the L-strand, respectively. These authors examined 4-fold degenerate sites[1] of the 12 H-strand encoded genes, finding a positive correlation ($P < 0.05$) between the frequency of C and the distance from the origin of L-strand replication. The implications of these findings is that asymmetric replication has a strong influence on mutagenesis.

[1] The third codon positions of Leu-CTN, Val-GTN, Ser-TCN, Pro-CCN, Thr-ACN, Ala-GCN, Arg-CGN and Gly-GGN.

The process of asymmetric replication is an unusual characteristic of the mitochondrial genome. Figure 2.1 shows how this type of replication proceeds. Synthesis of a daughter H-strand begins from the displacement loop (D-loop) in the control region. The parental H-strand is displaced and has a transient single-stranded phase while the L-strand is being copied. In mammals this single-stranded state lasts until two-thirds of the daughter H-strand has been synthesised. At this point replication of the daughter L-strand begins, extending in the opposite direction. As a result the parental L-strand never has a single-stranded phase. The synthesis of a new mitochondrial genome is catalysed by DNA polymerase-γ (Bolden et al., 1977). Pol- γis found both in the nucleus and mitochondria, and shows a high fidelity with exonucleolytic proofreading of errors (Kunkel and Soni, 1988). Under optimal conditions, pol- γdisplays fidelity of less than one error for every 2.6×10^5 bases polymerised.

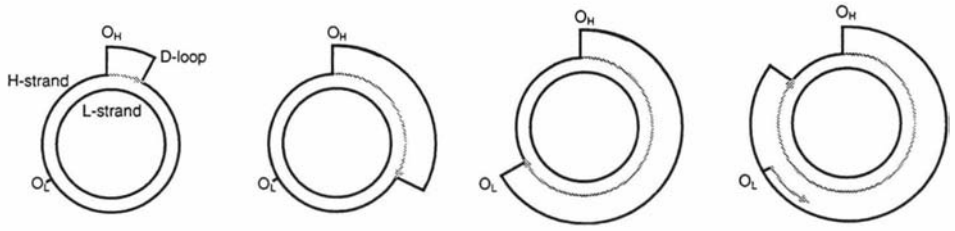


Figure 2.1. Replication proceeds asymmetrically in the mitochondrial genome. The O_H and O_L are the origins of heavy and light strand replication respectively. The dark lines indicate the parent strands and the grey lines the daughter strands

Clayton (1982) reported that the replication of mtDNA is slow, requiring around two hours for the synthesis of both strands. Moreover, the *cob* gene on the parental H-strand remains single-stranded for almost 80 minutes, whereas the parental L-strand is never single-stranded. Tanaka and Ozawa (1994) suggest that the spontaneous deamination of cytosine to uracil is the most common source of mutation. Cytosine will lose its amine group through hydrolytic disassociation when not paired with guanine far more rapidly than while double-stranded, and will subsequently pair with adenine (Lindahl and Nyberg, 1974). Accumulations of C→T mutations have been shown to occur on non-transcribed strands during their unprotected single-stranded phase (Fix and Glickman, 1987; Beletskii and Bhagwat, 1996; Francino and Ochman, 1997).

Reyes et al. (1998), following the previously mentioned studies, showed that in a data set comprising 25 mammalian genomes there was a positive correlation between the duration of the H-strand single-stranded phase and the degree of variability of a gene. Two genes, however, *atp8* and *cob*, do not show this strong correlation. Reyes et al. suggest that their unusual characteristics (*atp8* being a fast evolving gene and *cob* being abnormally constrained) explain their deviation from this relationship. This study reinforces a mechanism of spontaneous deamination on C and A during H-strand replication, moreover, it provides an explanation for the asymmetric and biased base composition observed in mtDNA. The rapid deamination of C and A would cause a high number of C→T and A→G transitions (Figure 2.2), and therefore help to explain the Brown et al. (1982) finding of a high transition to transversion substitution ratio in mtDNA.

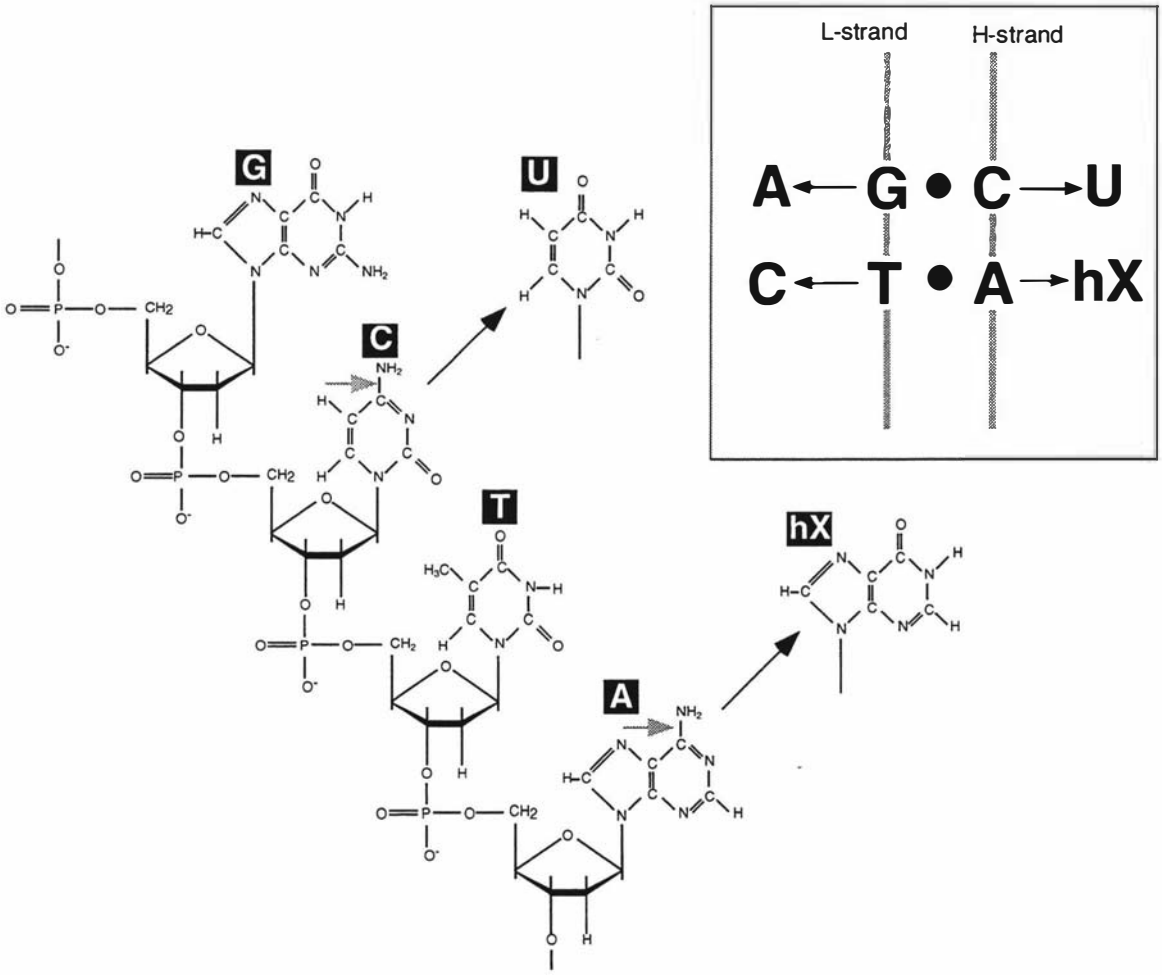


Figure 2.2. The mutational pathway proposed for mtDNA. Cytosine (C) and adenine (A) are deamination by hydrolytic attack to form uracil (U) and hypoxanthine (hX) respectively. These products U and hX will pair with A and C respectively and create a transitional change. T = thymine and G = guanine.

If damaging events occur during the transient single-stranded phase of H-strand replication and are not repaired (by for example uracil-DNA glycosidase), then polymerase-γ will most likely pair the damaged base with a new nucleotide. The damaged base could either be repaired by a glycosidase, using the newly changed undamaged base and inserting a new 'corrected' base, or during the next round of replication the newly inserted base would be transmitted to the daughter molecule. If this mutant type is present in the germ line it is be likely to be transmitted to the next generation. The high mutation rate is not simply due to the presence of damaging oxygen free radicals. The primary assertion in this chapter is that the unique property of mtDNA asymmetric replication explains the higher mutation rate in the mt genome compared with the nuclear genome.

2.4 Summary Remarks

A new mtDNA haplotype arises through mutation and a haplotype will become extinct if its lineage is not transmitted to the next generation. Paternal mtDNA is actively discarded along with the sperm soon after it enters the oocyte and mtDNA is maternally inherited. As a consequence mtDNA has one-quarter the effective population size as a Mendelian locus and haplotype extinction can occur rapidly. Secondly, mutations arise more frequently in the mitochondrial genome than in comparable nuclear gene because mtDNA replication has a long single-stranded phase. The exposed strand is highly susceptible to hydrolytic attack and therefore results in a high frequency of transitional mutations. This process of mutation would predict a similar, or even universal, mutation rate across a broad range of vertebrate species. This prediction makes any study that attempts to estimate a rate of evolution relevant to more than just the species from which it was measured.

2.5 References

Ankel-Simons, F., and Cummins, J.M. 1998. Misconceptions about the mitochondria and mammalian fertilization: implications for theories on human evolution. Proc. Natl. Acad. Sci. USA **93:** 13859-13863.

Beletskii, A., and Bhagwat, A.A. 1996. Transcription-induced mutations: Increase in C to T mutations in the nontranscribed strand during transcription in *Escherichia coli*. Proc. Natl. Acad. Sci. USA **93:** 13919-13924.

Bendall, K.E., Macaulay, V.A., Baker, J.R., and Sykes, B.C. 1996. Heteroplasmic point mutations in the human mtDNA control region. Am. J. Hum. Genet. **59:** 1276-1287.

Bergstrom, C.T., and Pritchard, J. 1998. Germline bottlenecks and the evolutionary maintenance of mitochondrial genomes. Genetics **149:** 2135-2146.

Birky, C.W. 1995. Uniparental inheritance of mitochondrial and chloroplast genes: mechanisms and evolution. Proc. Natl. Acad. Sci. USA **92:** 11331-11338.

Bogenhagen, D.F. 1999. Repair of mtDNA in vertebrates. Am. J. Hum. Genet. **64:** 1276-1281.

Bolden, A., Noy, G.P., and Weissbach, A. 1977. DNA polymerase of mitochondria is a γ-polymerase. J. Biol. Chem. **252:** 3351-3356.

Breen, A.P., and Murphy, J.A. 1995. Reactions of oxyl radicals with DNA. Free Radical Biol. Med. **18:** 1033-1077.

Brown, W.M., Prager, E.M., Wang, A., and Wilson, A.C. 1982. Mitochondrial DNA sequences in primates: tempo and mode of evolution. J. Mol. Evol. **18:** 225-239.

Calloway, C.D., Reynolds, R.L., George L Herrin, J., and Anderson, W.W. 2000. The frequency of heteroplasmy in the HVII region of mtDNA differs across tissue types and increases with age. Am. J. Hum. Genet. **66:** 1384-1397.

Clayton, D.A. 1982. Replication of animal mitochondrial DNA. Cell **28:** 693-705.

Coulondre, C., Miller, J.H., Farabaugh, P.J., and Gilbert, W. 1978. Molecular basis of base substitution hotstops in *Escherichia coli*. Nature **274:** 775-780.

Croteau, D.L., Rhys, C.M.J., Hudson, E.K., Dianov, G.L., Hansford, R.G., and Bohr, V.A. 1997. An oxidative damage-specific endonuclease from rat liver mitochondria. J. Biol. Chem. **272:** 27338-27344.

Croteau, D.L., Stierum, R.H., and Bohr, V.A. 1999. Mitochondrial DNA repair pathways. Mutat. Res. **434:** 137-148.

Domena, J.D., Timmer, R.T., Dicharry, S.A., and Mosbaugh, D.W. 1988. Purification and properties of mitochondrial uracil-DNA glycosylase from rat liver. Biochemistry **27:** 6742-6751.

Fix, D.F., and Glickman, B.W. 1987. Asymmetic cytosine deamination revealed by spontaneous mutational specificity in an Ung^- strain of *Escherichia coli*. Mol. Gen. Genet. **209:** 78-82.

Francino, M.P., and Ochman, H. 1997. Strand asymmetries in DNA evolution. Trends Gen. **13:** 240-245.

Gissi, C., Reyes, A., Pesole, G., and Saccone, C. 2000. Lineage-specific evolutionary rate in mammalian mtDNA. Mol. Biol. Evol. **17:** 1022-1031.

Gyllensten, U., Wharton, D., Josefsson, A., and Wilson, A.C. 1991. Paternal inheritance of mitochondrial DNA in mice. Nature **352:** 255-257.

Halliwell, B., and Gutteridge, J.M.C. 1999. Free Radicals in Biology and Medicine. Oxford University Press, Oxford.

Hauswirth, W.W., and Laipis, P.J. 1982. Mitochondrial DNA polymorphism in a maternal lineage of Holstein cows. Proc. Natl. Acad. Sci. USA **79:** 4686-4690.

Ikeda, S., and Ozaki, K. 1997. Action of mitochondrial endonuclease G on DNA damaged by L-ascorbic acid, peplomycin, and *cis*-diamminedichloroplatinum (II). Biochem. Biophys. Res. Commun. **235:** 291-294.

Jazin, E.E., Cavelier, L., Eriksson, I., Oreland, L., and Gyllensten, U. 1996. Human brain contains high levels of heteroplasmy in the noncoding regions of mitochondrial DNA. Proc. Natl. Acad. Sci. USA **93:** 12382-12387.

Kaneda, H., Haysahi, J.-I., Takahama, S., Taya, C., Fischer-Lindahl, K., and Yonekawa, H. 1995. Elimination of paternal mitochondrial DNA in intraspecific crosses during early mouse embryogenesis. Proc. Natl. Acad. Sci. USA **92:** 4542-4546.

Kunkel, T.A., and Soni, A. 1988. Exonucleolytic proofreading enhances the fidelity of DNA synthesis by chick embryo DNA polymerase-γ. J. Biol. Chem. **263:** 4450-4459.

Lima-de-Faria, A. 1983. Molecular evolution and organization of the chromosome. Elsevier, Amsterdam.

Lindahl, T. 1993. Instability and decay of the primary structure of DNA. Nature **362:** 709-715.

Lindahl, T., Ljungquist, S., Wolfgang, Siegert, Nyberg, B., and Sperens, B. 1977. DNA *N*-glycosidases. J. Biol. Chem. **252:** 3286-3294.

Lindahl, T., and Nyberg, B. 1974. Heat-induced deamination of cytosine residues in deoxyribonucleic acids. Biochemistry **13:** 3405-3410.

Martin, A.P., and Palumbi, S.R. 1993. Body size, metabolic rate, generation time, and the molecular clock. Proc. Natl. Acad. Sci. USA **90:** 4087-4091.

Michikawa, Y., Mazzucchelli, F., Bresolin, N., Scarlato, G., and Attardi, G. 1999. Aging-dependent large accumulation of point mutations in the human mtDNA control region for replication. Science **286:** 774-778.

Parker, A., Gu, Y., and Lu, A.-L. 2000. Purification and characterization of a mammalian homolog of *Escherichia coli* MutY mismatch repair protein from calf liver mitochondria. Nucleic Acids Res. **28:** 3206-3215.

Pikó, L., and Taylor, K.D. 1987. Amounts of mitochondrial DNA and abundance of some mitochondrial gene transcripts in early mouse embryos. Dev. Biol. **123:** 364-374.

Pinz, K.G., Shibutani, S., and Bogenhagen, D.F. 1995. Action of mitochondrial DNA polymerase γ at sites of base loss or oxidative damage. J. Biol. Chem. **270:** 9202-9206.

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

Reyes, A., Gissi, C., Pesole, G., and Saccone, C. 1998. Asymmetrical directional mutation pressure in the mitochondrial genome of mammals. Mol. Biol. Evol. **15:** 957-966.

Richter, C., Park, J.-W., and Ames, B.N. 1988. Normal oxidative damage to mitochondrial and nuclear DNA is extensive. Proc. Natl. Acad. Sci. USA **85:** 6465-6467.

Satoh, M.S., Huh, N.-h., Rajewsky, M.F., and Kuroki, T. 1988. Enzymatic removal of O^6-ethylguanine from mitochondrial DNA in rat tissues exposed to *N*-ethyl-*N*-nitrosourea *in vivo*. J. Biol. Chem. **263:** 6854-6856.

Sawyer, D.E., and van Houten, B. 1999. Repair of DNA damage in mitochondria. Mutat. Res. **434:** 161-176.

Shibutani, S., Takeshita, M., and Grollman, A.P. 1991. Insertion of specific bases during DNA synthesis past the oxidation-damage base 8-oxodG. Nature **349:** 431-434.

Slowinski, J.B., and Arbogast, B.S. 1999. Is the rate of molecular evolution inversely related to body size? Syst. Biol. **48:** 396-399.

Sutovsky, P., Moreno, R., and Schatten, G. 1998. Sperm mitochondrial ubiquitination and a model explaining the strictly materal mtDNA inheritance in mammals. Mol. Biol. Cell **9:** Suppl. 309a.

Sutovsky, P., and Schatten, G. 2000. Paternal contributions to the mammalian zygote: fertilization after sperm-egg fusion. Int. Rev. Cytol. **195:** 1-65.

Tanaka, M., and Ozawa, T. 1994. Strand asymmetry in human mitochondrial DNA mutations. Genomics **22:** 327-335.

Téoule, R. 1978. Radiation-induces degradation of the base component in DNA and related substances - final products. *In* Effects of Ionizing Radiation of DNA: Physical, Chemical and Biological Aspects. *Edited by* J. Hüttermann, W. Köhnlein and R. Téoule. Springer-Verlag, Berlin. pp. 171-203.

Thomas, W.K., and Beckenbach, A.T. 1989. Variation in salmonid mitochondrial DNA: Evolutionary constraints and mechanisms of substitution. J. Mol. Evol. **29:** 233-245.

Thyagarajan, B., Padua, R.A., and Campbell, C. 1996. Mammalian mitochondria possess homologous DNA recombination activity. J. Biol. Chem. **271:** 27536-27543.

Wilson, A.C., Cann, R.L., Carr, S.M., M George, J., Gyllensten, E.B., Helm-bychowski, K.M., Higuchi, R.G., Palumbi, S.R., Prager, E.M., Sage, R.D., and Stoneking, M. 1985. Mitochondrial DNA and two perspectives on evolutionary genetics. Biol. J. Linn. Soc. **26:** 375-400.

Chapter Three

A Repeat Complex in the Mitochondrial Control Region of Adélie Penguins[1]

3.1 Introduction

The mitochondrial (mt) control region (CR) is responsible for transcription and replication of the mitochondrial genome (Taanman, 1999). Three internal CR portions have been recognised, the 5'-peripheral domain (Right or I Domain), the Central Conserved Domain (*c.* 200 b.p.), and the 3'-peripheral domain (Left or II Domain) (Upholt and Dawid, 1977; Southern et al., 1988; Saccone et al., 1991). In the avian CR, the 5'-peripheral domain contains the Termination Associated Sequences (TAS); the Central Conserved Domain contains the F, E, D, and C boxes; and the 3'-peripheral domain contains the origin of heavy strand replication (O_H), Conserved Sequence Block (CSB) 1, and the Heavy (H) and Light (L) strand Transcriptional Promoter (HSP/LSP) sites (L'Abbé et al., 1991; Quinn and Wilson, 1993; Randi and Lucchini, 1998). Transcription and replication of the

[1] This chapter is published as: Ritchie PA, Lambert DM (2000) A repeat complex in the mitochondrial control region of Adélie penguins from Antarctica. *Genome* **43**, 613-618.

mt genome are intimately connected. After transcription has begun at the LSP site, mitochondrial processing endoribonuclease (RNase MRP), and possibly endonuclease G (Côté and Ruiz-Carrillo, 1993), are thought to cleave the L-strand transcript at sequences corresponding to the CSBs (Chang and Clayton, 1985). This transcript forms an RNA primer that initiates replication of the H-strand at a site designated the O_H (Walberg and Clayton, 1981). It is known that in humans as H-strand synthesis displaces the L-strand at the WANCY tRNA cluster approximately two-thirds of its way around the molecule, this free L-strand is primed by a mt DNA primase and L-strand synthesis begins (Wong and Clayton, 1986).

Sequence data from mtCRs has revealed the presence of repeat DNA sequences in many different species, from nematodes (*Caenorhabditis elegans*, Okimoto et al., 1992) to sheep (*Ovis ovis*, Zardoya et al., 1995). Lunt et al. (1998) surveyed repeat sequences in the CRs of a range of species and showed that they varied in number and ranged in size from 3 b.p. to as large as 777 b.p. To date, several mechanisms have been proposed to explain the generation and persistence of repetitive DNA in the mitochondrial CR. These include slippage during replication (Levinson and Gutman, 1987; Madsen et al., 1993), or rarer recombination events (Lunt and Hyman, 1997). Consistent with these models, repetitive DNA in the CR is found in areas where transcription and replication are initiated (ie. near the TAS in the 5'-peripheral domain or between CSB-1 and the tRNA in the 3'-peripheral domain). This chapter reports on the mtCR of the Adélie penguin *Pygoscelis adeliae* (Aves: Sphenisciformes) from Antarctica, which is one of the largest avian mtCR found to date, due to the presence of two classes of repeat units in the 3'-peripheral domain.

3.2 Materials and methods

Blood samples from Adélie penguins were collected during the austral summer of 1997/1998 at Cape Bird, Antarctica (77°14'S, 166°28'E), then transported in liquid nitrogen and stored at –80°C. Total genomic DNA was isolated from whole blood using proteinase K – SDS (sodium dodecyl sulfate) dissolution, followed by extraction with phenol-chloroform and ethanol precipitation (Sambrook et al., 1989). The Adélie penguin control region was amplified using the Polymerase

Chain Reaction (PCR) with primers designed to the flanking *nad6* (5'-ACTAAACCAATTACCCCATAATA-3' L-nad6-gb) and *rns* (5'-CTGCTGAGTACCCGTGGGGGTGTGGC-3' H-rns-pen) genes. These PCR primers were designed using *nad6* gene sequences from GenBank (accession no.: chicken, X52392; duck, L16769; ostrich, Y12025; goose, L22477; and gnatcatcher, AF027829) and *rns* gene sequences determined for the penguin species *Pygoscelis adeliae, Eudyptes pachyrhynchus, Megadyptes antipodes* and *Eudyptula minor*. To design the H-strand primer the 5'-end of the *rns* gene was sequenced using the PCR primers 5'-AGCATGGCACTGAAGATGC-3' (L-tRNA-P, modified from Tarr, 1995) and 5'-ATAGTGGGGTATCTAATCCCA-3' (H-rnsA, modified from Palumbi, 1995). A 260 b.p. portion of the WANCY tRNA cluster was amplified and sequenced in the Adélie penguin, yellow-eyed penguin (*M. antipodes*), red-billed gull (*Larus novaehollandiae scopulinus*), Chatham Island taiko (*Pterodroma magentae*) and kakapo (*Strigops habroptilus*), using the primers 5'-CCAAAGGCCTTCAAAGCCTTAAATAAGAG-3' (L-WANCY), 5'-CCAAAAA TCTGTGGTTCAATTCCTCTTC-3' (H-WANCY); designed from GenBank sequences (chicken, X52392; ostrich, Y12025). All mitochondrial gene nomenclature in this thesis follows that of the Organelle Genome Database (GOBASE) available at http://alice.bch.umontreal.ca/genera/gobase/gobase.html (Korab-Laskowska et al., 1998).

All PCRs contained 0.4 µM of each primer, 200 µM of each dNTP, 1.5 mM $MgCl_2$, 10 mM Tris-HCl pH 8.3, 50 mM KCl and 0.5 U *Taq* DNA polymerase (AmpliTaq, PE Biosystems). The PCR-products were purified using High Pure™ PCR Product Purification columns (Boehringer Mannheim). Purified PCR-products were sequenced with internal primers using an ABI PRISM® BigDye™ Terminator Cycle Sequencing Kit (PE Applied Biosystems) and run on an ABI 377A automated sequencer, according to the manufacturer's recommendations. Sequences for the CR can be found at GenBank accession number AF272143 and for the tRNA WANCY at AF272144 and AF272145. The UNIX-based software by GCG (Genetics Computer Group, Inc.) was used to predict the single-stranded secondary structures of the CR 5'-end.

To verify that these CR sequences were of mitochondrial origin, and did not represent nuclear copies, a large portion (*ca.* 6.0 k.b.) of the Adélie penguin

mitochondrial genome was amplified and sequenced. The long PCR template was generated using the Expand™ Long Template PCR System (Boehringer Mannheim) using two PCR-primers designed to conserved portions of the *rnl* gene 5'-TGATTGCGCTACCTTCGCACGGTTAGGATACC-3' (H-ExpPeng-rnl) and the *cob* gene 5'-CCATTCCACCCCTACTACTCCACAAAAGA-3' (L-ExpPeng-cob). All PCRs contained 0.3 μM of each primer, 500 μM of dNTPs, PCR Buffer 3 (50 mM Tris-HCl pH 9.2, 16 mM $(NH_4)_2SO_4$, 2.25 mM $MgCl_2$, 2% DMSO and 0.1% Tween 20) and 2.6 U Expand™ Long Template enzyme mix. The resulting PCR-product was purified using High Pure™ PCR Product Purification columns (Boehringer Mannheim), and the CR was amplified and sequenced as above. The same sequences were obtained across the control region for both the long and short PCR products indicating that both have been amplified from the same template. In addition, no double peaks on the electropherograms were encountered any during sequencing.

3.3 Results and discussion

3.3.1. The Organisation of the Adélie mt Genome

The CR of this Adélie penguin is 1758 b.p. in length and is bound by the tRNAs *trnE(uuc)* and *trnF(gaa)*, and the genes *nad6* and *rns* (Figure 3.1)[2]. The position of the flanking genes, tRNAs and the control region is the same as that determined for chicken (*Gallus gallus*, Desjardins and Morais, 1990) (Figure 3.1b). A recent study by Mindell et al. (1998) showed there are two types of gene arrangement in birds; the common *cob*/*nad6*/control region/*rns* order (e.g. the gene arrangement of the chicken) and the less frequent order of; *cob*/control region/*nad6*/a non-coding region/*rns*. If the avian phylogeny that Mindell et al. (1998) presented is correct, then the latter, less frequent gene rearrangement, has occurred as separate independent events in Picidae, Cuculidae, Passeriformes suboscines and Falconiformes. The results of Mindell et al. (1998) indicate that these mt gene rearrangements are likely to be due to specific processes that result in the reoccurrence of this condition.

[2] The abbreviations correspond to: *nad6*, encodes the product NADH dehydrogenase subunit 6; *trnE(uuc)*, transfer RNA Glutamic Acid; *trnF(gaa)*, transfer RNA Phenylalanine; *rns*, encodes the small subunit ribosomal RNA (12S).

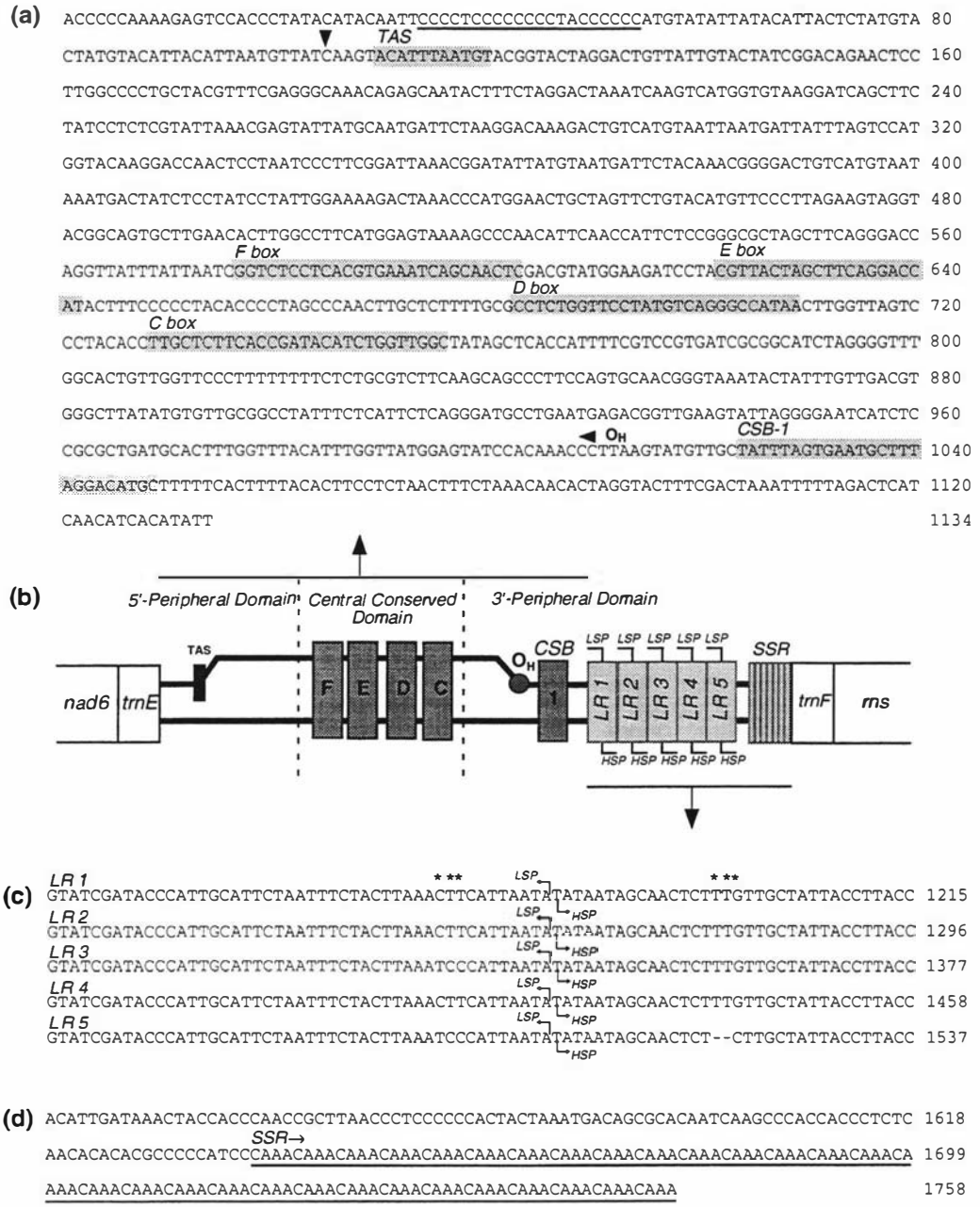


Figure 3.1. The L-strand sequence and a schematic representation of the Adélie penguin control region (CR). (a) The first 1143 b.p. (comprising the 5'-perpherial domain, Central Conserved Domain, and part of the 3'-peripheral domain) of the CR. Underlined is the interrupted poly-C sequence, marked with an arrow head is the possible dC terminus of the D-loop and highlighted are the Termination Associated Sequences (TAS). Also highlighted are the F, E, D and C boxes and Conserved Sequence Block (CSB)-1. (b) A schematic representation of the CR, which shows the flanking genes and portions of the CR sequence represented in this figure. The origin of heavy strand replication (O_H) is estimated to be 12 nt from CSB-1 and is represented as a circle. (c) The 81 b.p. large repeat (LR 1, 2, 3, 4 and 5). The repeat units are aligned with each other, the stars show the variable positions and the ⌟ symbols represent the Light (L) and Heavy (H) Strand Promoter (LSP/HSP) sites. (d) The 92 b.p. sequence after the LRs and underlined is the simple sequence repeat (SSR); a tetranucleotide microsatellite (CAAA).

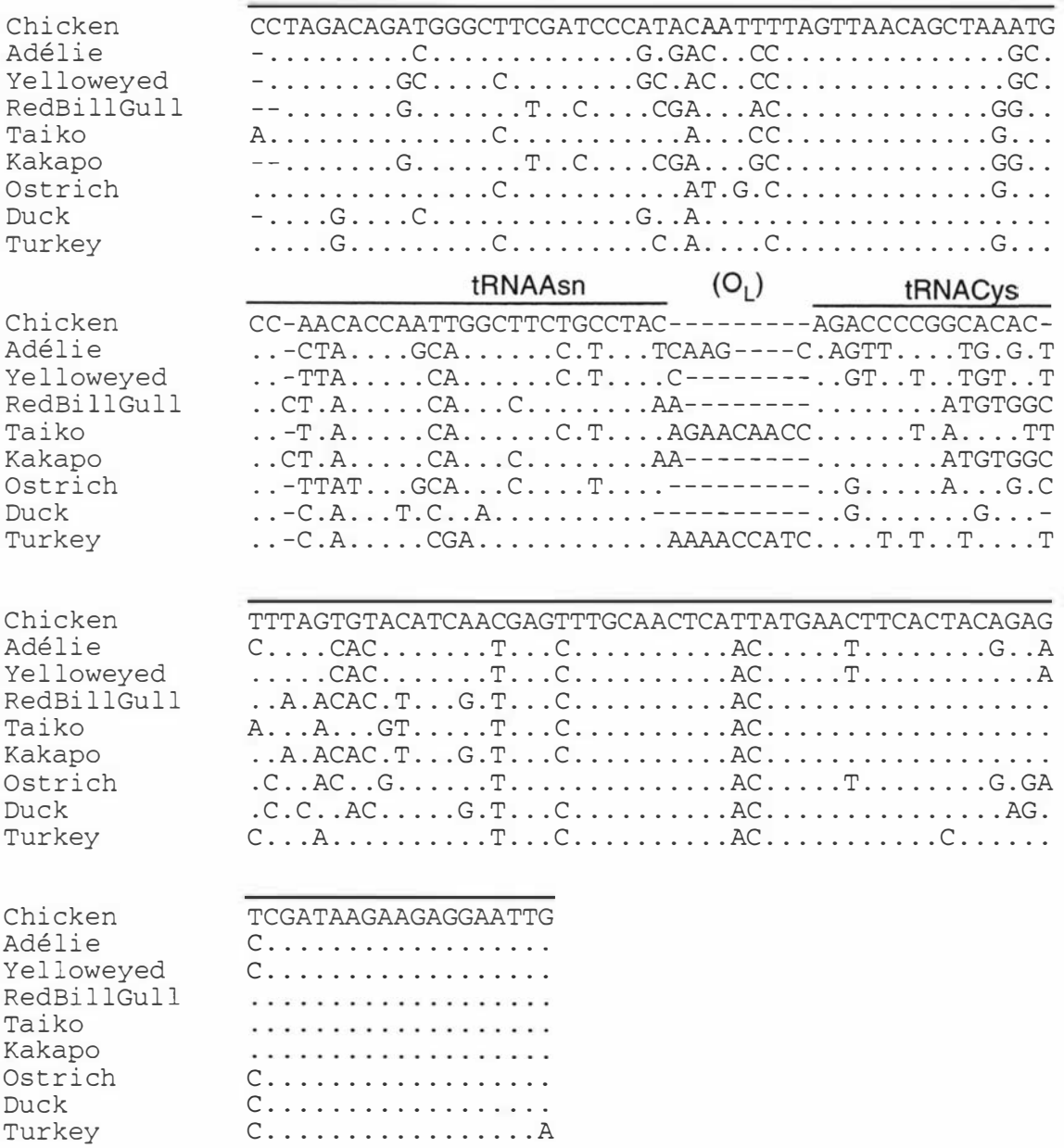


Figure 3.2. An alignment of sequences from the tRNA's Asn and Cys from the WANCY tRNA cluster. The chicken, ostrich, duck and turkey sequences were taken from GenBank. The position of the origin of light strand replication in humans is shown (O_L). In these avian taxa there are no 50 b.p. sequences between the two tRNA that correspond to known recognition sequences for mtDNA primase.

In mammals the WANCY tRNA cluster includes an ~50 nucleotide (nt) sequence that is capable of forming a stable single-stranded secondary structure, that is known to be involved in priming the synthesis of the nascent L-strand. However, this sequence is conspicuously absent from the WANCY tRNA cluster of birds (Seutin et al., 1994). This begged the questions of (1) whether the O_L is located at a different position, yet undiscovered, in birds, and (or) (2) whether there are other unknown recognition sequences on the mt genome involved in initiation of L-strand synthesis (Shadel and Clayton, 1997). A 250 b.p. portion of the WANCY tRNA cluster was determined for five bird species. Like other avian species, these species possess no mammalian like sequences responsible for priming the O_L in this region of the mt genome (Figure 3.2).

3.3.2. The Organisation of the Adélie Penguin CR

The internal organisation of the Adélie penguin CR was compared to that found in other avian CRs (Quinn and Wilson, 1993; Randi and Lucchini, 1998) and found them to be very similar. The 5'-peripheral domain is 560 b.p., the CCD is 200 b.p. (as defined by Saccone et al., 1991), and the 3'-peripheral domain is 998 b.p. long. The 3'-peripheral domain of the Adélie penguin is larger than the entire 909 b.p. CR of the cow (*Bos taurus*, Anderson et al., 1982), and only slightly smaller than the CRs of many avian species (Baker and Marshall, 1997). The overall nucleotide composition of the Adélie penguin CR is A: 30.4%; T: 30.3%; C: 25.5%; and G: 13.8%. These values show that there is a A + T (60.8%) to G + C (39.2%) asymmetry in this sequence. It has been shown that the CCD differs from the two peripheral domains in being lower in L-strand adenine content (Ramirez et al., 1993). The reduced adenine content of the L-strand of the CCD is also the case in the Adélie penguin with A: 20.7%; T: 30.8%; C: 31.3%; and G: 17.1%, which almost corresponds to a A + T to G + C symmetry: 51.5% to 48.5%, respectively.

The 5'-peripheral domain contains an interrupted poly-C sequence (5'-CCCCTCCCCCCCCTACCCCCC-3') which forms a potential single stranded secondary structure (Figure 3.1a and 3.3). Zardoya and Meyer (1998) reported a similar motif in the CR of the African side-necked turtle (*Pelomedusa subsrufa*), and noted its similarity to the CSB-2 motif (5'-CAAACCCCCCCWMCCCC-3') identified in the 3'-peripheral domain of mammals by Walberg and Clayton (1981)

and Saccone et al. (1991). A survey of GenBank shows that a similar motif is present in the 5'-peripheral domain of other avian species, including Struthioniformes, Galliformes, and Falconiformes. In the lesser snow goose (*Anser caerulescens caerulescens*) and the chicken, this motif forms a potentially stable hairpin structure, because it is followed by an interrupted poly-G sequence (Quinn and Wilson, 1993). Although this interrupted poly-C sequence is a conserved feature across many species and is close to one identified TAS (Nass, 1995), a role for this sequence has never been determined.

H-strand synthesis originates within the first 40 nt downstream from CSB-1 and temporarily stalls shortly after initiation. The 3'-end(s) of the nascent H-strand terminates close to the TAS to form a stable triplex displacement-loop (D-loop) structure. Nass (1995) has identified the 5' and 3'-ends of the D-loop from the chicken and the TAS (5'-ATATATTTACA-3') at 81 b.p. to the right of *trnE(uuc)*. She went on to show that this sequence is present in other avian species and associated with a putative secondary structure. Furthermore, the 5'-end of the chicken D-loop was shown to begin 11-12 nt downstream of the CSB-1, is 781 (±1) nt in length, and has two nascent H-strand 3'-termini which exhibit dG and dC 3'-end nt. Presented in Figure 3.3 is the putative secondary structure of the first 164 b.p. of the 5'-peripheral domain of the Adélie penguin, which includes the interrupted poly-C sequence and the putative TAS sequence and dC 3'-terminus. Based on an inferred location of the O_H and TAS, this d-loop would correspond to 893 nt in length.

3.3.3. Tandem repeats in the 3'-peripheral domain

The most striking feature of the Adélie penguin CR is the presence of two types of tandemly repeated sequences, a large repeat (LR) and a simple sequence repeat (SSR) in the 3'-peripheral domain (Figure 3.1c and 3.1d). These two repeats contribute to the extraordinary size of this CR. Considering the base composition of this domain, the first 365 nucleotides are 58.4% A + T to 41.6% G + C, followed by the large repeats at 69.7% A + T to 30.3% G + C, then the LR/SSR spacer sequence is 46.0% A + T to 54.0% G + C, and lastly the SSR is 75% A to 25% C. These values show that the LRs have a significant A + T bias, a feature known to cause helix instability and which is common at origins of replication.

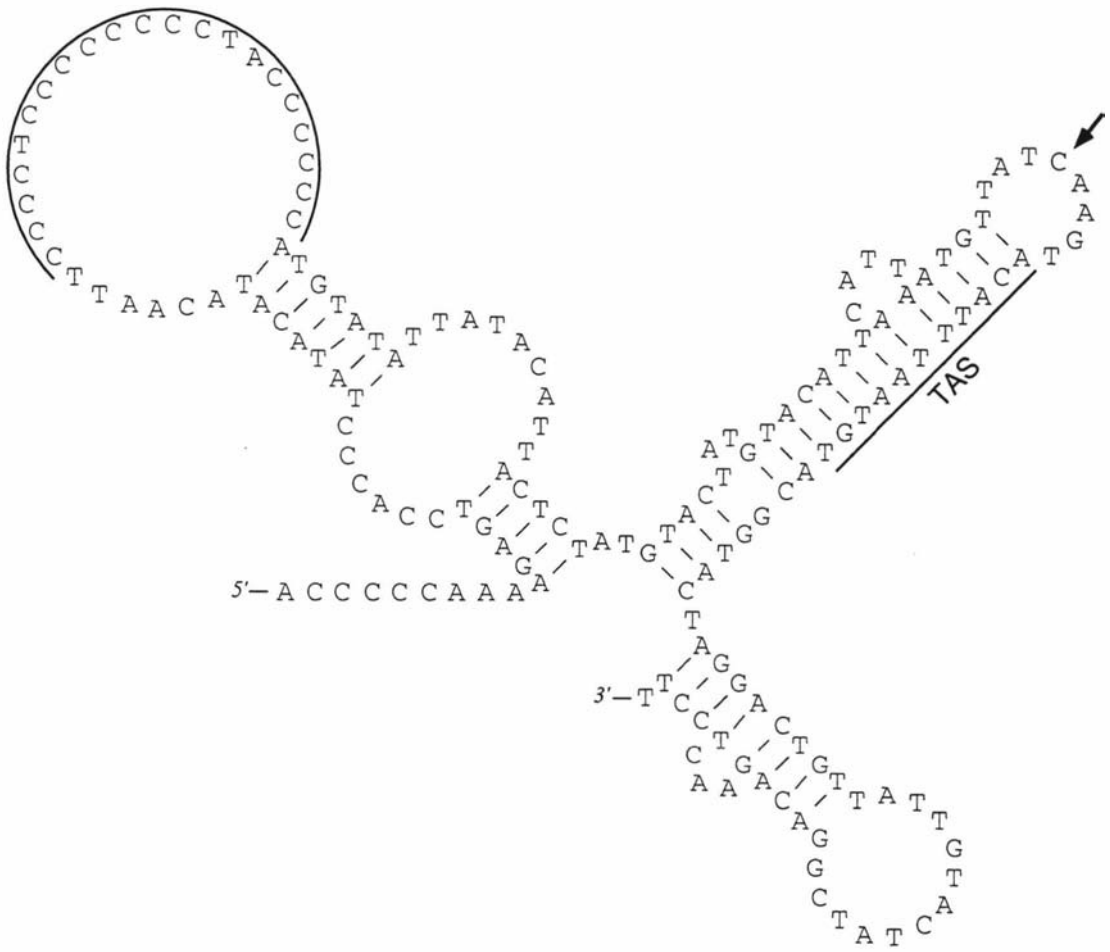


Figure 3.3. The secondary structure of the first 164 b.p. of the 5'-peripheral domain of the Adélie penguin control region (CR). Outlined is the interrupted poly-C sequence and the putative termination associated sequences (TAS), the arrow corresponds to the dC terminus of the chicken D-loop.

The SSRs comprise 30 perfect tetranucleotide microsatellite repeats consisting of $(dC\text{-}dA\text{-}dA\text{-}dA)_{30}\bullet(dG\text{-}dT\text{-}dT\text{-}dT)_{30}$. This microsatellite is located at the extreme 3'-end of the CR, is contiguous with the tRNA and since it is downstream from the HSP site, is most likely to be transcribed. Interestingly, the same tetranucleotide microsatellite has also been found in the ratite, *Rhea americana* (Härlid et al., 1998), and similar repeats containing a CA-type motif have been found on the L-strand in other avian species (Berg et al., 1995). Like other microsatellites, there is good evidence that SSR are formed and variation maintained via slipped-strand mispairing during mitochondrial replication (Levinson and Gutman, 1987; Madsen et al., 1993; Trinh and Sinden, 1993).

The LR region comprises four complete 81 b.p. repeats, followed by a repeat missing two base pairs (i.e. 5'-81-81-81-81-79-3'). Alignment of the repeats with each other reveal substitutional differences at positions 1171, 1172 and 1173 (C↔T), at position 1199 (C↔G), and insertions-deletions (indels) at positions 1197-1198 with respect to LR-1 in the data set (Fig. 1c.). Avian species differ from many vertebrates by having a single major transcriptional promoter in the CR that has a bi-directional capacity for both the L- and H-strand (L'Abbé et al., 1991). All these repeat units contain a putative HSP/LSP sequences, initially identified in the chicken (5'-GTATAATATATATACA-3'). The first two repeats (LR 1, LR 2) and the fourth repeat (LR 4) are identical. The third repeat (LR 3) and the last repeat (LR 5) are each unique and differ from the other repeats (LR 1, 2 and 4) by the substitutions described above and from each other by an indel event and a substitution. When folded, the LR repeats form a potential hairpin structure (Figure 3.4). In the European rabbit (*Oryctolagus cuniculus*, CR = 2437 b.p.) a primer extension analysis detected families of H-strand RNA species which originated from the CR LSP. These different sized RNA transcripts were attributed to a tandem repeat (153 b.p.) in which each motif contained a LSP site (Dufresne et al., 1996). Based on that study it is likely that the Adélie penguins too, produce multiple RNA transcripts, each originating from promoter sites in the LRs. However, in contrast to rabbits, there would be variable length transcripts for both the L- and H-strand RNA, since avians possess a bi-directional promoter.

To explain the changes in 5'-peripheral domain repeat numbers Buroker et al. (1990) proposed an illegitimate elongation model that involved slipped

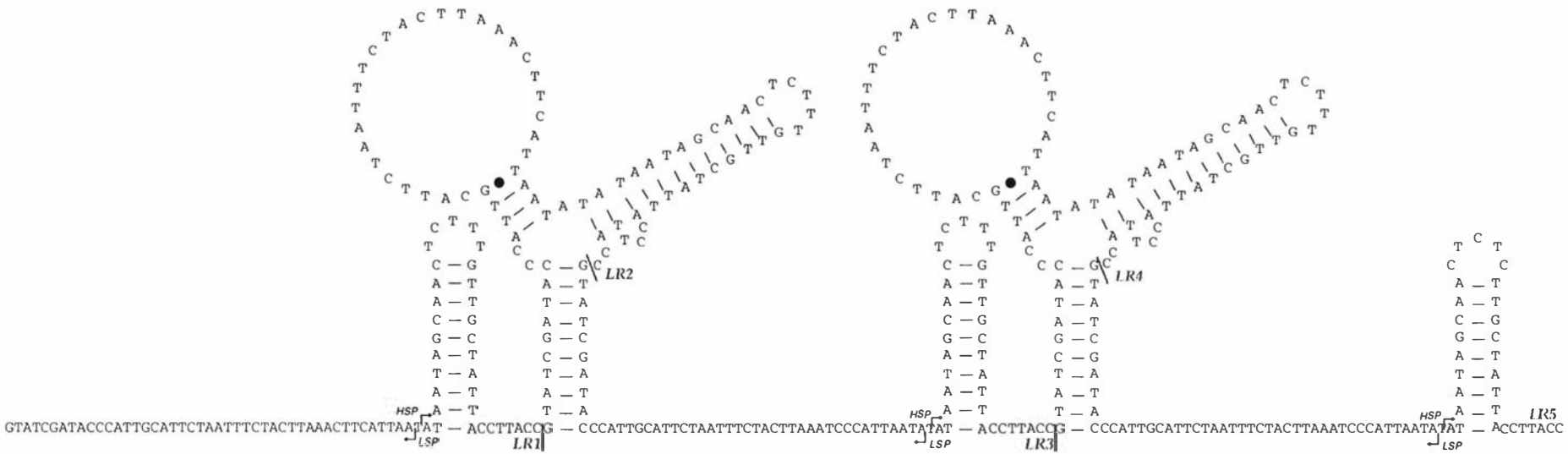


Figure 3.4. The secondary structure of the 81 b.p. repeat region. Shown are the LSP/HSP and the position of each repeat. The hairpin directly following the LSP/HSP is most likely involved in transcriptional enhancer recognition. These large structures only form with the presence of the other repeats. This could lead to RNA primer-template binding competition.

strand mispairing and non-homologous recombination. A similar model could be proposed to explain the LRs present in the 3'-peripheral domain of Adélie penguins. If, during replication, there is a dynamic competitive equilibrium between the large RNA primers and the H-strand for pairing with the L-strand, either at the beginning or end of replication, then partial displacement of a repeat unit is possible. A hairpin structure in the H-strand would result in a repeat loss, whereas a hairpin formation in the RNA primer would result in a repeat gain. The daughter molecule would form a heteroduplex with the parental molecule, which would be resolved during the next round of replication.

In summary, this chapter presented the mtCR sequence from the Adélie penguin and identified its internal organisation, including the conserved motifs (TAS, the F, E, D and C boxes, CSB-1 and the LSP/HSP). This control region is larger than those from other avian species, due to the presence of a large 81 b.p. repeat sequence and a $(dC\text{-}dA\text{-}dA\text{-}dA)_{30}$ microsatellite in the 3'-peripheral domain. The 81 b.p. repeat contains both the LSP and HSP and may therefore produce a variety of RNA transcripts. These results showed that avian mtCRs, like those in mammals, can vary greatly in size due to the presence of different repeat complexes.

3.3 References

Anderson, S., de Bruijn, M.H., Coulson, A.R., Eperon, I.C., Sanger, F., and Young, I.G. 1982. Complete sequence of bovine mitochondrial DNA: conserved features of the mammalian mitochondrial genome. J. Mol. Biol. **156:** 683-717.

Baker, A.J., and Marshall, H.D. 1997. Mitochondrial control region sequences as tools for understanding evolution. *In* Avian Molecular Evolution and Systematics. *Edited by* D. P. Mindell. Academic Press, San Diego. pp. 51-82.

Berg, T., Moum, T., and Johansen, S. 1995. Variable number of simple tandem repeats make birds of the order Ciconiiformes heteroplasmic in their mitochondrial genomes. Curr. Genet. **27:** 257-262.

Buroker, N.E., Brown, J.R., Gilbert, T.A., O'Hara, P.J., Beckenbach, A.T., Thomas, W.K., and Smith, M.J. 1990. Length heteroplasmy of Sturgeon mitochondrial DNA: An illegitimate elongation model. Genetics, **124:** 157-163.

Chang, D.D., and Clayton, D.A. 1985. Priming of human mitochondrial DNA replication occurs at the light-strand promoter. Proc. Natl. Acad. Sci. USA **82:** 351-355.

Côté, J., and Ruiz-Carrillo, A. 1993. Primers for mitochondrial DNA replication generation by Endonuclease G. Science, **261:** 765-769.

Desjardins, P., and Morais, R. 1990. Sequence and gene organization of the chicken mitochondrial genome. J. Mol. Biol. **212:** 599-634.

Dufresne, C., Mignotte, F., and Guéride, M. 1996. The presence of tandem repeats and the initiation of replication in rabbit mitochondrial DNA. European J. Biochem. **235:** 593-600.

Härlid, A., Janke, A., and Árnason, U. 1998. The complete mitochondrial genome of *Rhea americana* and early avian divergences. J. Mol. Evol. **46:** 669-679.

Korab-Laskowska, M., Rioux, P., Brossard, N., Littlejohn, T.G., Gray, M.W., Lang, B.F., and Burger, G. 1998. The Organelle Genome Database Project (GOBASE). Nucleic Acids Res. **26:** 139-144.

L'Abbé, D., Duhaime, J.-F., Lang, B.F., and Morais, R. 1991. The transcription of DNA in chicken mitochondria initiates from one major bidirectional promoter. J. Biol. Chem. **266:** 10844-10850.

Levinson, G., and Gutman, G.A. 1987. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol. Biol. Evol. **4:** 203-221.

Lunt, D.H., and Hyman, B.C. 1997. Animal mitochondrial DNA recombination. Nature, **387:** 247.

Lunt, D.H., Whipple, L.E., and Hyman, B.C. 1998. Mitochondrial DNA variable number of tandem repeats (VNTRs): utility and problems in molecular ecology. Mol. Ecol. **7:** 1441-1455.

Madsen, C.S., Ghivizzani, S.C., and Hauswirth, W.W. 1993. *In vivo* and *in vitro* evidence for slipped mispairing in mammalian mitochondria. Proc. Natl. Acad. Sci. USA **90:** 7671-7675.

Mindell, D.P., Sorenson, M.P., and Dimcheff, D.E. 1998. Multiple independent origin of mitochondrial gene order in birds. Proc. Natl. Acad. Sci. USA **95:** 10693-10697.

Nass, M.M.K. 1995. Precise sequence assignment of replication origin in the control region of chick mitochondrial DNA relative to 5' and 3' D-loop ends, secondary structure, DNA synthesis, and protein binding. Curr. Genet. **28:** 401-409.

Okimoto, R., Mcfarlane, J.L., Clary, D.O., and Wolstenholme, D.R. 1992. The mitochondrial genomes of two nematodes, *Caenorhabditis elegans* and *Ascaris suum*. Genetics, **130:** 471-498.

Palumbi, S.P. 1995. Nucleic Acids II: The Polymerase Chain Reaction. *In* Molecular Systematics. *Edited by* D. M. Hills, C. Moritz and B. K. Mable. Sinauer Associates, Inc, Massachusetts. pp. 205-247.

Quinn, T.W., and Wilson, A.C. 1993. Sequence evolution in and around the mitochondrial control region in birds. J. Mol. Evol. **37:** 417-425.

Ramirez, A., Savoie, P., and Morais, R. 1993. Molecular characterization and evolution of a duck mitochondrial genome. J. Mol. Evol. **37:** 296-310.

Randi, E., and Lucchini, V. 1998. Organization and evolution of the mitochondrial DNA control region in the avian genus *Alectoris*. J. Mol. Evol. **47:** 449-462.

Saccone, C., Pesole, G., and Sbisá, E. 1991. The main regulatory region of mammalian mitochondrial DNA: structure-function model and evolutionary patterns. J. Mol. Evol. **33:** 83-91.

Sambrook, J., Fritsch, E.F., and Maniatus, T. 1989. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Lab. Press, NY.

Seutin, G., Lang, B.F., Mindell, D.P., and Morais, R. 1994. Evolution of the WANCY region in amniote mitochondrial DNA. Mol. Biol. Evol. **11:** 329-340.

Shadel, G.S., and Clayton, D.A. 1997. Mitochondrial DNA maintenance in vertebrates. Ann. Rev. Biochem. **66:** 409-435.

Southern, S.O., Southern, P.J., and Dizon, A.E. 1988. Molecular characterization of a cloned mitochondrial genome. J. Mol. Evol. **28:** 32-42.

Taanman, J.-W. 1999. The mitochondrial genome: structure, transcription, translation and replication. Biochim. Biophys. Acta **1410:** 103-123.

Tarr, C.L. 1995. Primers for amplification and determination of mitochondrial control-region sequences in oscine passerines. Mol. Ecol. **4:** 527-529.

Trinh, T.Q., and Sinden, R.R. 1993. The influence of primary and secondary DNA structure in deletion and duplication between direct repeats in *Escherichia coli*. Genetics, **134:** 409-422.

Upholt, W.B., and Dawid, I.B. 1977. Mapping of mitochondrial DNA of individual sheep and goats: rapid evolution in the D loop region. Cell, **11:** 571-583.

Walberg, M.W., and Clayton, D.A. 1981. Sequence and properties of the human KB cell and mouse L cell D-loop regions of mitochondrial DNA. Nucleic Acids Res. **9:** 5411-5421.

Wong, T.W., and Clayton, D.A. 1986. DNA primase of human mitochondria is associated with structural RNA that is essential for enzymatic activity. Cell, **45:** 817-825.

Zardoya, R., and Meyer, A. 1998. Cloning and characterization of a microsatellite in the mitochondrial control region of the African side-necked turtle, *Pelomedusa subrufa*. Gene, **216:** 149-153.

Zardoya, R., Villalta, M., Lopez-Perez, M.J., Garrido-Petierra, A., and Montoya, J. 1995. Nucleotide sequence of the sheep mitochondrial D-loop and its flanking tRNA genes. Curr. Genet. **28:** 94-96.

Chapter Four

The Evolution of the Mitochondrial DNA Hypervariable Region in Penguins

4.1 Introduction

The hypervariable region I (HVR-1) comprises the 5'-end (domain I) and half of the central conserved domain in the mitochondrial DNA control region. HVR-1 sequences have been widely used as tools to study both intraspecific population structure and the relationships among recently divergent taxa (e.g. Viligant et al., 1991; Saccone et al., 1991; Saunders and Edwards, 2000). This is because the control region evolves rapidly, a result of a high number of substitutional changes, insertion-deletion events (indels), a variable number of repeative sequences and heteroplasmy. Therefore, HVR-1 sequences from deeply split taxa are often saturated and indels tend to make alignments extremely difficult (Sbisà et al., 1997). The rapid evolution of the HVR-1 represents an useful sequence not only for intraspecific studies of Adélie penguins, but also for a comparative study of penguin species. The fossil record suggests penguins are a relatively old avian suborder (Sphenisciformes), yet the divergence of some species could have been

relatively recent. Consequently, a broad range of evolutionary changes should be evident in the control region of this avian group.

The 17 extant species recognised in the suborder Spheniscifformes have been assigned to six genera: *Aptenodytes*, *Pygoscelis*, *Spheniscus*, *Eudyptes*, *Megadyptes* and *Eudyptula* (Stonehouse, 1975). All species are listed in Table 4.1. While penguins occur from Antarctica to the equatorial Galápagos Islands and comprise around 80% of the avian biomass in the southern oceans (Williams, 1995), the majority of species occur in the temperate and sub-Antarctic regions. All species come ashore to breed and moult, with nesting sites ranging from Antarctic ice to tropical lava flows. Penguins breed during the summer, with the exception of the Fiordland crested penguin which begins its breeding cycle during the winter (Warham, 1974a).

These flightless diving birds represent a monophyletic group that is thought to have evolved from a fully flighted ancestor (Stonehouse, 1975). The oldest known penguin fossil, dated at 50-60 Myr BP, was found in Waipara, New Zealand, and is believed to have been a specialist marine diver (Fordyce et al., 1986). A number of anatomical novelties differentiate penguins from other taxa in the class Aves. These include the loss of flight feathers but retention of wings, the reduction of wing articulation accompanied by wing flattening, and a well-developed fat layer. Penguins are believed to share their closest common ancestry with the Gaviiformes (Olson, 1985) or Procellariiformes (Livezey, 1989; Ho et al., 1976). Most of the 38 fossilised penguin species have been uncovered in New Zealand, but their distribution spans Antarctica, South America, Australia and Africa (Fordyce and Jones, 1990). Some penguin species are purported to have gone extinct as recently as the Holocene (Harrison, 1984). No fossil penguins have ever been found in the Northern Hemisphere, although fossils of "giant, flightless penguin-like birds" ascribed to the order Pelecaniformes, have been found in the North Pacific (Olson and Hasegawa, 1979). While the body masses of the extant species range from 1-35kg, the more diverse extinct species included the massive *Anthropornis grandis*, estimated to have weighed 81kg (Livezey, 1989). Even more unusual was a giant late Eocene penguin species found on Seymour Island which had a long dagger-like bill, possibly used to spear prey (Olson, 1985). In contrast, all extant penguin

species have short bills and feed mostly on shoaling marine species such as euphausids or small fish.

The phylogeny of penguin species has received little attention despite being such a widely recognised group of birds. Only a few studies have attempted to address this deficiency (e.g. O'Hara, 1989; Edge, 1996; Paterson et al., 1997). This chapter describes the use of DNA sequences from both the small subunit ribosomal RNA (*rns*) and large subunit ribosomal RNA (*rnl*) genes to reconstruct a phylogeny of the spheniscids. Based on this phylogenetic analysis, HVR-1 sequence variation among penguin species is investigated, and the evolution of both length variation and heteroplasmy are discussed.

Table 4.1. The penguin species with their common names, and the sample sizes for sequencing of each mitochondrial DNA region are shown. Small subunit ribosomal RNA = *rns*, large subunit ribosomal RNA = *rnl*, and hypervariable region I = HVR-1. Recognised subspecies: two gentoo subspecies *P. p.* (*papua* and *ellsworthii*); two rockhopper subspecies *E. c.* (*chrysocome* and *moseleyi*); six little blue/fairy subspecies *E. m.* (*minor*, *iredalei*, *novaehollandiae*, *albosignata*, *chathamensis* and *variabilis*); and two king subspecies *A. p.* (*patagonicus* and *halli*). The collection details for each species are listed in Appendix Table B.2. The thesis follows the taxonomy of Williams (1995).

Species	Common Name	*rns*	*rnl*	HVR-1
Pygoscelis adeliae	Adélie	21	3	381
P. antarctica	Chinstrap	5	1	3
P. papua	Gentoo	2	1	1
Eudyptes chrysolophus	Macaroni	1	1	1
E. schlegeli	Royal	1	3	4
E. pachyrhynchus	Fiordland crested	3	1	6
E. robustus	Snares crested	1	1	1
E. sclateri	Erect-crested	3	2	5
E. chrysocome	Rockhopper	3	2	4
Megadyptes antipodes	Yellow-eyed	5	1	3
Eudyptula minor	Little Blue/Fairy	2	1	1
Spheniscus demersus	Black footed	1	1	1
S. magellanicus	Magellanic	1	1	1
S. humboldti	Peruvian	1	1	6
S. mendiculus	Galápagos	1	1	1
Aptenodytes patagonicus	King	1	1	1
A. forsteri	Emperor	2	1	5

4.2 Methods

4.2.1. Sampling and DNA Extraction

Table 4.1 lists the penguin species and their sample sizes for this study. Total genomic DNA was extracted from blood or muscle tissue following Sambrook et al. (1989). Each sample was dissolved in 400 µl of extraction buffer (10 mM Tris-HCl pH 8.0, 50 mM NaCl, 10 mM EDTA, 2% SDS) containing proteinase K and incubated at 55°C overnight. The digested tissue was purified by standard methods of phenol-chloroform extraction. Nucleic acids were then precipitated with 2.5 volumes of 100% ethanol following by a 70% ethanol wash. The precipitant was pelleted in a microfuge, dried and resuspended in 200 µl of ultra-pure H_20.

4.2.2. PCR and DNA Sequencing of the rns and rnl genes

The polymerase chain reaction (PCR) was used to amplify two segments from the *rns* and *rnl* genes using the PCR primers L-12SA 5'-AAACTGGGATTAGATACCCCACTAT-3' and 5'-CTTCCGGTACACTTACC TTGTTACGAC-3' H-12SB2, and L-16SA2 5'-CGACTGTTTACCAAAAACAT AGCC-3' and H-16SB2 5'-CCGGTCTGAACTCAGATCACGT-3', respectively (modified from Kocher et al., 1989; Palumbi, 1995). Double stranded amplifications were performed in 25 µl volumes containing 10 mM Tris-HCl pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$, 200 µM of each dNTP, 1 µM of each primer, 10-100 ng genomic DNA, and 1 unit of *Taq* Polymerase (AmpliTaq, PE Biosystems). PCRs were carried out in a Hybaid thermal cycler at 94°C/10 sec, 58-62°C/10 sec, and 72°C/35 sec for 25 cycles. PCR-products were purified using High Pure™ PCR Product Purification columns (Boehringer Mannheim). Sequencing reactions were performed in 10 µl volumes using a DyeDeoxy Terminator Sequencing Kit (Applied Biosystems). Cycle sequencing reactions were carried out in a Hybaid thermal cycler at 96°C/10 s and 60°C/4 min for 25 cycles. Removal of excess DyeDeoxy terminators from the reaction mix was achieved by ethanol/NaOAc precipitation followed by a 70% ethanol wash. The precipitate was dried in a concentrator and resuspended in 5 µl of formamide/50 mM EDTA pH 8.0. Samples were denatured at 90°C, transferred to ice, immediately loaded on a

Applied Biosystems 377 automated sequencer and run according to the instructions of the manufacturer. All templates were sequenced in both directions. Orthologous DNA sequences were aligned and managed in Sequencher ver 3.1.1. (Gene Codes Corp.).

4.2.3. Phylogenetic Analyses

Maximum parsimony (MP), maximum likelihood (ML) and neighbor-joining (NJ) methods were employed in the phylogenetic analyses. Outgroup taxa were selected from four orders closely related to Sphenisciformes: The red-throated loon, *Gavia stellata* (Gaviiformes, GenBank accession no. AF173578); horned grebe, *Podiceps auritus* (Podicipediformes, AF173567); brown pelican, *Pelecanus occident* (Pelecaniformes, AF173570); and great shearwater, *Puffinus gravis* (Procellariiformes, AF173572). All phylogenetic analyses were carried out in PAUP* 4.0b4a (Swofford, 2000). To accommodate rate variation among sites, the shape parameter (α) of a gamma distribution (Γ) was estimated using ML (Yang, 1994). The MP analysis used the heuristic search procedure, with 100 random replicates. The robustness of the phylogeny was assessed with a non-parametric bootstrap analysis using 1000 replicates (Felsenstein, 1985). Several different character-weighting schemes were introduced since weighted parsimony can perform better than unweighted (Hillis et al., 1994). Analyses were performed considering: (1) transitions (ti) and transversions (tv) equally; (2) *a priori* weighting of ti and tv (ti:tv) 5:1 according to the observed ratio between the ingroup and outgroup and within the ingroup (ratio = 5:1, Figure 4.1); (3) additional ti:ts ratios of 3:1, 7:1 and 10:1; and (4) tv only. In addition, a successive-approximation approach to weighting of characters was performed according to their "cladistic reliability" (Farris, 1969; Williams and Fitch, 1989). Reweighting was based on both the consistency index and the mean rescaled index of the most parsimonious tree and a base weight of 1.0. Neighbor-joining analyses (Saitou and Nei, 1987) were performed and 500 bootstrap replicates were used to assess the sampling error in the phylogenetic estimate. Three methods that accounted for multiple substitutions were used to estimate evolutionary distances: Jukes-Cantor (Jukes and Cantor, 1969), Kimura's two-parameter model (Kimura, 1980), and Tamura-Nei's method (Tamura, 1992). In addition, Unweighted Paired-Group Method of Arithmetic Means (UPGMA) analyses were conducted also using PAUP* 4.0b4a.

4.2.4. PCR and DNA Sequencing of the Hypervariable Region I

Primers to the *nad6* L-strand (5'-ACTAAACCAATTACCCCATAATA-3') (Chapter Three) and the *trnE* H-strand (5'-GTTCCTGTGGTTGAAGTAACA-3') of Adélie penguins were used to amplify DNA flanking the mtDNA control region in all penguin species. All PCRs contained 0.4 µM of each primer, 200 µM of each dNTP, 1.5 mM $MgCl_2$, 10 mM Tris-HCl pH 8.3, 50 mM KCl, and 1 unit of *Taq* DNA polymerase (AmpliTaq, PE Biosystems). Thermal cycling conditions were 94°C/10 sec, 50°C/10 sec, and 72°C/25 sec for 30 cycles. PCR products were purified and sequenced as described above. These DNA sequences were aligned and conserved regions were used to design two penguin specific L-strand PCR-primers for *trnE* (see Table 4.3 in the Results section). A primer (5'-ATGCCGCGATCACGGACGAAAATGG) designed to the H-strand of the central conserved domain of the Adélie penguin control region sequence (presented in Chapter Three) was used in conjunction with one of the *trnE* primers, to amplify and sequence (as above) the control regions of all other penguins species, except emperor and king penguins. These sequences were aligned and two primers, spanning a shorter region, were designed to the conserved D box which lies close to the variable 5'-end on the periphery of the central domain (Table 4.3).

To distinguish between nuclear derived control region sequences and mtDNA heteroplasmy, a long-PCR fragment of 6.0 kb that spanned the mitochondrial genome from *cob* to *rnl*, was amplified from Fiordland crested penguins and sequenced. The long-PCR template was generated using the Expand™ Long Template PCR System (Boehringer Mannheim) using two PCR primers designed to conserved portions of the *rnl* gene 5'-TGATTGCGCTACC TTCGCACGGTTAGGATACC-3' (H-ExpPeng-rnl) and the *cob* gene 5'-CCATTCCACCCCTACTACTCCACAAAAGA-3' (L-ExpPeng-cob). All PCR reactions contained 0.3 µM of each primer, 500 µM of each dNTP, PCR Buffer 3 (50 mM Tris-HCl pH 9.2, 16 mM $(NH_4)_2SO_4$, 2.25 mM $MgCl_2$, 2% DMSO and 0.1% Tween 20) and 2.6 units of Expand™ Long Template enzyme mix. The resulting PCR product was purified using High Pure™ PCR Product Purification columns (Boehringer Mannheim), and the control region was amplified and sequenced as above.

4.3 Results

4.3.1. DNA Sequence Variation in the rns and rnl Genes

A total of 985 nucleotide sites were included in the analyses of *rnl* and *rns* sequences, of which 159 (16.1%) were variable and 122 (12.4%) were parsimony informative. On average (including outgroup taxa), the nucleotide (π) composition for both the *rnl* and *rns* was $\pi_A = 0.31$, $\pi_T = 0.21$, $\pi_C = 0.27$, and $\pi_G = 0.21$. This base composition was stationary across the data set ($\chi^2 = 8.54$, d.f. = 60, $P = 1.0$). More variation was observed in the *rns* data set (84 of the 411 b.p. were variable, 20.4%), compared with the *rnl* data set (75 positions in 574 b.p. were variable, 13.1%). Whenever multiple representatives from the same species were sequenced, identical results were obtained. Using the ML, tree the shape parameter of a gamma distribution (Γ) was estimated at 0.3553.

The plot of the absolute number of ti versus tv both within the ingroup and between the ingroup and outgroup is shown in Figure 4.1. The overall ratio of ti to tv (τ) is 5:1 (i.e. 5 ti to every 1 tv). However, ti appear to have reached saturation in the extreme comparisons of the ingroup, and between the ingroup and outgroup (Figure 4.1). An analysis of the substitutional spectrum for the ingroup and between the ingroup and outgroup is presented in Figure 4.2. The L-strand sequence shows C↔T ti predominate, while A↔C changes are the most frequent tv and G↔T the least common of all changes. A detailed pairwise comparison showing percentage sequence differences and number of ti and tv for all ingroup comparisons, is presented in Table 4.2. When the substitutional spectrum of the ingroup is compared with the outgroup (Figure 4.2) there is an overall increase in the number of observations; however, there is a disproportionate increase in A↔C changes compared with all other tv. The percent sequence differences for pairwise comparisons ranged from 7% to <1%. None of the taxonomic comparisons revealed identical sequences. Number of substitutions among members of the ingroup, with respect to both ti and tv, ranged from one (Fiordand vs Snares) to 74 (Adélie vs Galápagos). When considering tv only the substitutional differences range from as low as zero (for nine pairwise comparisons) to as high as 21 (king penguins compared with yellow-eyed penguins).

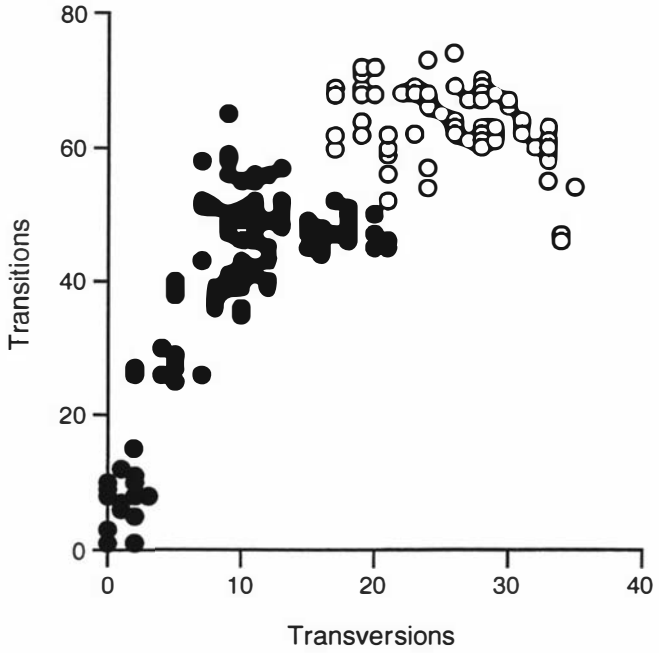


Figure 4.1. The number of transversions versus the number of transitions in the combined *rnl* and *rns* data set. The filled circles are ingroup comparisons and open circles are ingroup to outgroup comparisons.

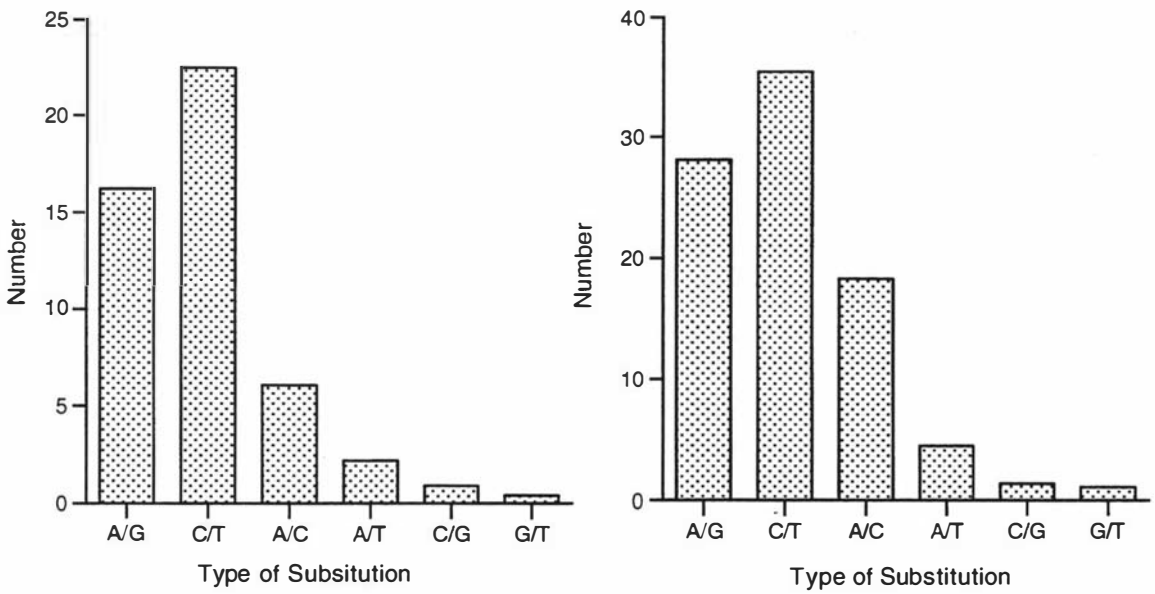


Figure 4.2. The type and average number of substitutions among the ingroup taxa (left) and between the ingroup and outgroup taxa (right) for L-strand sequences. Transitions (A/G and C/T) occur more frequently that transversions (A/C, A/T, C/G and G/T). A/C changes have increased disproportionately to other transversions from the ingroup to outgroup comparison.

Table 4.2. Pairwise comparisons of ingroup taxa for their percentage sequence difference (lower diagonal) and number of transitions compared with transversions (upper diagonal) in the combined *rnl* and *rns* data set.

	Ade	Chi	Gen	Mag	Bla	Per	Gal	Mac	Sna	Fio	Roy	Ere	Roc	Yel	Emp	Kin	Lit
Ade	-	27/2	30/4	59/9	59/9	58/9	65/9	52/13	55/11	56/11	52/11	57/13	50/11	56/12	49/15	47/15	55/10
Chi	0.03	-	26/2	52/7	52/7	51/7	58/7	48/11	50/9	49/9	47/9	50/11	48/9	49/10	45/15	48/15	51/8
Gen	0.04	0.03	-	50/9	50/9	49/9	56/9	48/13	50/11	51/11	47/11	50/13	46/11	49/12	47/17	52/17	51/10
Mag	0.07	0.06	0.06	-	3/0	9/0	5/2	40/10	37/8	38/8	38/8	36/10	39/8	40/9	46/16	47/16	39/5
Bla	0.07	0.06	0.06	0.00	-	10/0	8/2	39/10	36/8	37/8	37/8	35/10	38/8	41/9	47/16	48/16	38/5
Per	0.07	0.06	0.06	0.01	0.01	-	8/2	39/10	36/8	37/8	37/8	35/10	38/8	39/9	44/16	45/16	40/5
Gal	0.08	0.07	0.07	0.01	0.01	0.01	-	43/12	40/10	41/10	41/10	39/12	42/10	42/11	50/18	51/18	43/7
Mac	0.07	0.06	0.06	0.05	0.05	0.05	0.06	-	10/2	11/2	1/2	8/3	10/2	26/4	50/20	47/20	45/12
Sna	0.07	0.06	0.06	0.05	0.05	0.05	0.05	0.01	-	1/0	9/0	6/1	8/0	27/5	48/18	46/18	42/10
Fio	0.07	0.06	0.07	0.05	0.05	0.05	0.05	0.01	0.00	-	10/0	7/1	9/0	28/5	49/18	47/18	43/10
Roy	0.07	0.06	0.06	0.05	0.05	0.05	0.05	0.00	0.01	0.01	-	7/1	9/0	25/5	50/18	47/18	43/10
Ere	0.07	0.06	0.07	0.05	0.05	0.05	0.05	0.01	0.01	0.01	0.01	-	12/1	26/7	47/20	45/20	40/12
Roc	0.06	0.06	0.06	0.05	0.05	0.05	0.06	0.01	0.01	0.01	0.01	0.01	-	29/5	48/18	46/18	46/10
Yel	0.07	0.06	0.06	0.05	0.05	0.05	0.06	0.03	0.03	0.04	0.03	0.04	0.04	-	46/21	45/21	41/11
Emp	0.07	0.06	0.07	0.07	0.07	0.06	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	-	15/2	48/15
Kin	0.07	0.07	0.07	0.07	0.07	0.06	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.02	-	49/15
Lit	0.07	0.06	0.07	0.05	0.05	0.05	0.05	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.07	0.07	-

4.3.2. The Phylogenetic Relationships of Penguins

The MP (Figure 4.3), ML (Figure 4.4), UPGMA (Figure 4.5) and NJ (Figure 4.6) methods all recovered similar trees, although the bootstrapped MP tree had three polytomies. Different weighting schemes employed in the MP analysis recovered similar topologies. A successive-approximation approach to the weighting of characters recovered a tree of the same topology as the MP tree in Figure 4.3.

In all phylogenetic analyses the penguins were monophyletic relative to the four outgroup suborders. These analyses also showed a good correspondence between the currently recognised taxonomy of penguins and their phylogenetic relationships (i.e. all six genera were monophyletic). The *Aptenodytes* species (king and emperor) were the most basal clade to the outgroup species, and nested between this group and the *Eudyptes* and *Spheniscus* groups were the three *Pygoscelis* species. To investigate the significance of monophyly in the ingroup, 100 replicates of a topology-dependent permutation tail probability test (T-PTP, Faith, 1991) were performed. In the present study, the T-PTP showed a significant difference from the MP tree of 473 steps, while the permuted data ranged from 708-732 steps ($P = 0.01$).

A comparison of all trees showed disagreement on the placement of taxa within the genus *Spheniscus*; for example, the NJ tree ((magellanic, Galápagos), black footed, Peruvian) compared to the ML tree ((Peruvian Galápagos), black footed, magellanic). The total distance between the magellanic and black footed penguins (three transitions) is much less than the distance for the other five possible comparisons among the magellanic, black footed, Galápagos and Peruvian penguins (average of ten substitutions). Other data sets (morphology and *cob* sequences) show Galápagos and Peruvian penguins are more closely related to each other than to the other *Spheniscus* species (Edge, 1996). A comparative plot of tv versus ti for members of the genus *Spheniscus* and all other penguin taxa, and a plot for the genus *Eudyptes* and all other taxa, were generated (data not presented). The plot of *Spheniscus* shows an obvious gap between the within genus and between genera comparisons. Transitions quickly approach a plateau for all comparisons between genera. In contrast, a smaller gap is apparent in *Eudyptes*. Hence the combination of deep and shallow splits within the ingroup may

inadvertently group unrelated taxa. A non-random deletion (i.e. of taxonomic groups) followed by an MP analysis showed that the removal of *Eudyptes*, *Megadyptes* and *Eudyptula* species gave a topology for the *Spheniscus* ((Galápagos, Peruvian), (black footed, magellanic)) comparable to that recovered from other studies using the *cob* gene and morphological characters.

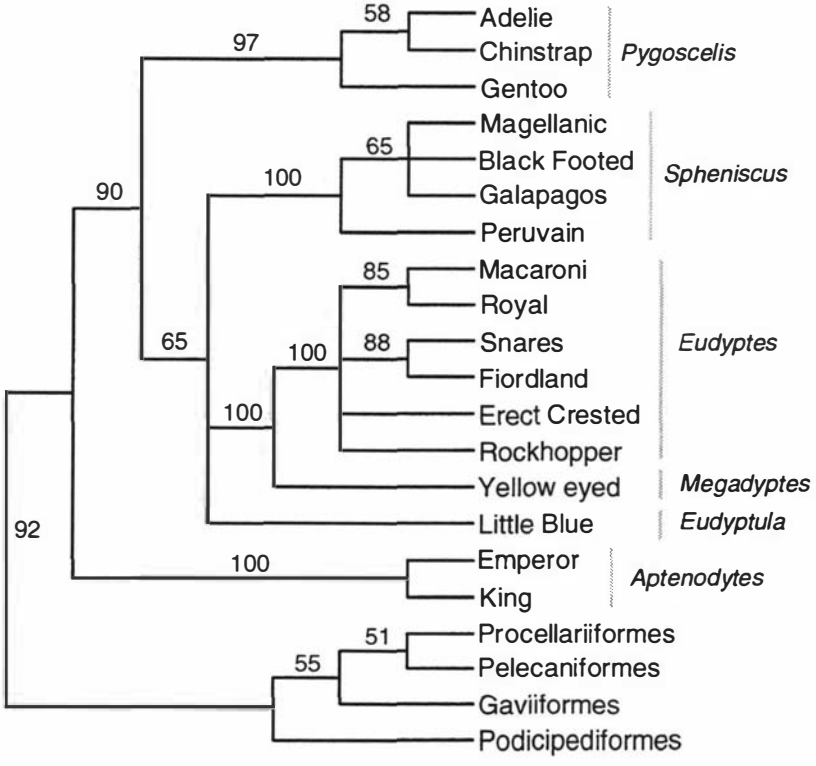


Figure 4.3. The bootstrapped maximum parsimony tree showing relationships among penguin taxa and the outgroup. The numbers represent the proportion of bootstraps that supported each node. There were ten most-parsimonious trees of 473 steps recovered (consistency index = 0.6047, retention index = 0.6825).

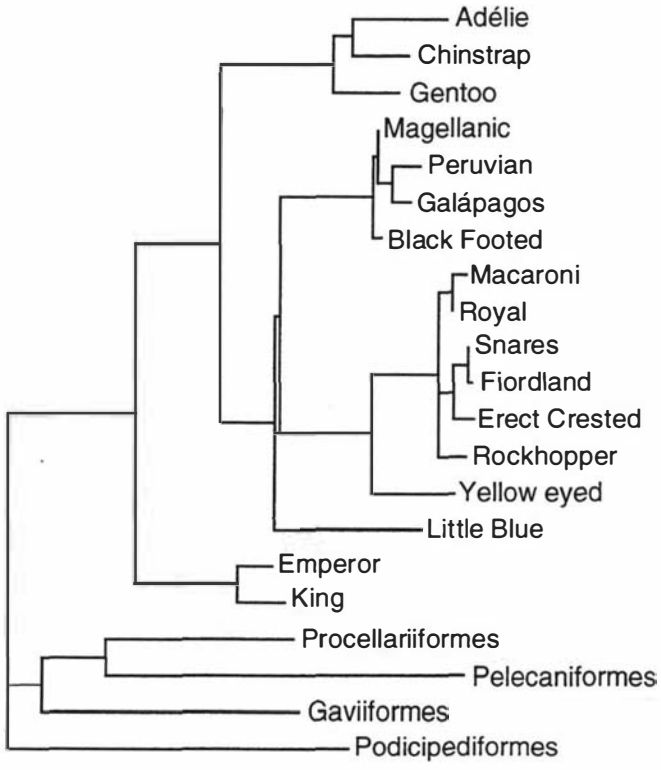


Figure 4.4. The maximum-likelihood tree showing relationships among penguin taxa and the outgroup. When unconstrained this topology has a -Ln(L_i) score of 3941.2 and when a molecular clock is enforced -Ln(L_0) = 3963.6. A likelihood ratio test of these two *L* values shows a significant difference ($P = 0.0008$).

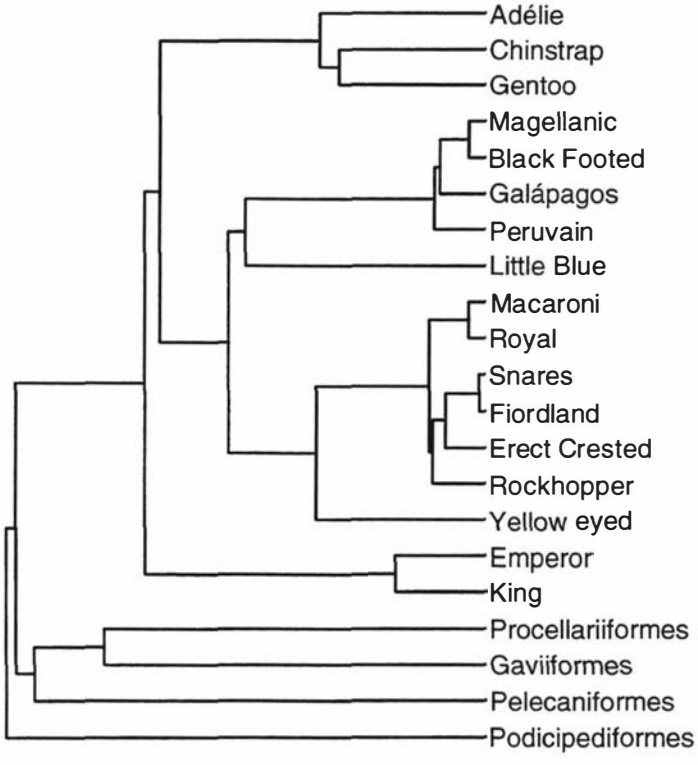


Figure 4.5. The UPGMA tree showing genetic distance among extant penguin taxa and the outgroup

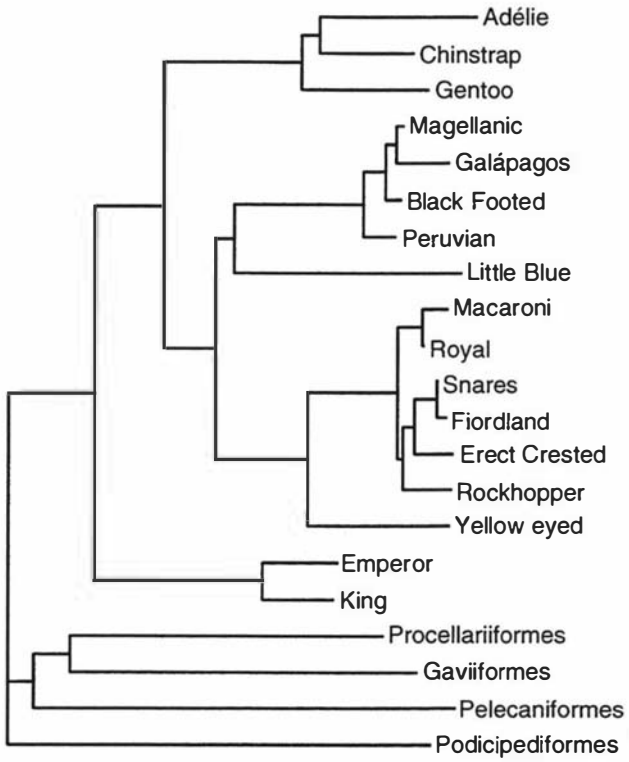


Figure 4.6. The neighbor-joining tree showing relationships among penguin taxa and the outgroup. This topology has a minimum evolution score of 0.43114 (i.e. sum of all branch lengths). None of the nodes had negative branch lengths. Overall the branch lengths to outgroup taxa were all > 0.04, while the longest ingroup branch was to the little blue penguin (0.0204), and the shortest to the Snares penguin (0.000083).

4.3.3. DNA Sequences from the Hypervariable Region I

A total of 519 sites were included in the analysis (which includes the ~100 b.p. insertion), of which 288 (55.5%) were variable and 206 (40%) were phylogenetically informative. On average, the base composition was $\pi_A = 0.31$, $\pi_T = 0.27$, $\pi_C = 0.24$, and $\pi_G = 0.18$, and was stationary across all taxa in the data set ($\chi^2 = 47.93$, d.f. = 42, $P = 0.245$). A ML tree (-Ln(L) = 3312.3) was used to estimate the shape parameter (α) at 2.07 for the gamma distribution (Γ). There was intraspecific variation for every case where multiple representatives of the same species where sequenced. A comparison of ti and tv (Figure 4.7 and Table 4.4) revealed an almost 1:1 relationship between many of the pairwise comparisons, hence ti appear to have reached a plateau. The averages for each type of substitution in the L-strand sequence are presented in Figure 4.8. Overall, there are on average 80 ti and 54 tv. There are a similar number of C↔T and A↔G ti, while all types of tv are common, except C↔G, which is probably due to the low G+C content of this sequence. All pairwise comparisons of the percentage sequence difference as well as the number of ti and tv for all ingroup comparisons, are presented in Table 4.4. Differences in

sequences range from 37.5% (the black footed penguin compared with gentoo penguin) to 4% (e.g. the Galápagos with Peruvian penguins). For every species comparison both ti and tv differences were noted.

4.3.4. Heteroplasmy and Insertion-deletion Events

The sizes of the HVR-1 fragments amplified and sequenced for each species are presented in Table 4.3. Samples from emperor and king penguins could not be amplified using these primers. These are the most basal species from the ribosomal genes and presumably have nucleotide substitutions in the primer sites that prevented them from being amplified. The majority of the HVR-1 fragments were 653-663 b.p. long, although in Adélie and gentoo penguins the PCR fragments were 767 and 763 b.p. respectively. These two species differed from the others by a ~100 b.p. insertion near the central conserved domain and just inside the 5′-domain. DNA sequence alignment was partially confounded by the presence of small indels across all species.

Ten of the species sequenced showed presumptive mitochondrial heteroplasmy (Table 4.3). In these species, sequencing with the L-strand primer resulted in two sequences appearing in electropherograms directly after a poly-C stretch (see Appendix Figure B.4). This interrupted poly-C sequence is at positions 34-55 in the Adélie penguin control region and is thought to be involved in the termination of the displacement loop (D-loop) during replication (Chapter Three). Direct sequencing of the 5′-end of long-PCR fragments from the Fiordland crested penguins shows the same double sequence pattern. This result reinforces the suggestion of heteroplasmy and excludes the possibility of preferential amplification of a nuclear mitochondrial segment. With the exception of the poly-C region, sequencing with the H-strand primer resulted in unambiguous electropherograms. The presence or absence of heteroplasmy and the ~100 b.p. insertion (see Table 4.3) was plotted on to the bootstrapped ribosomal MP tree (Figure 4.9), and revealed neither HVR-1 character has a monophyletic origin within penguins.

Table 4.3. Details of mtDNA control region primers designed for penguins. The sample sizes(n), primer combinations, sizes of amplified fragments, and the presence/absence of presumptive heteroplasmy are shown. Primer sequences: A (L-*trnE*, 5'-CCCGC TTGGCTTYTCTCCAAGGTC), B (H-Dbox, 5'-CTGACATAGGAACCAGAGGCGC), C (L-*trnE*, 5'-CCTGCTTGGCTTTTYTCCAAGACC), D (H-Dbox, 5'-CTGACCGAGGAACCAGA GGCGC).

Penguin Species	n	PCR Primer Combinations	Size (b.p.)	Presumptive Heteroplasmy
Adélie	290	A/B	767	No
Chinstrap	3	A/B	662	Yes
Gentoo	1	A/B	763	No
Macaroni	1	C/D	658	No*
Royal	4	C/D	659	No*
Fiordland crested	6	C/D	662	Yes
Snares crested	1	C/D	662	Yes
Erect-crested	5	C/D	660	Yes
Rockhopper	4	C/D	663	Yes
Yellow-eyed	3	C/D	670	Yes
Little/Fairy	1	C/D	655	Yes
Black footed	1	C/D	653	No
Magellanic	1	C/D	654	Yes
Peruvian	6	C/D	657	Yes
Galápagos	1	C/D	655	Yes
King	1	-	-	-
Emperor	5	-	-	-

*There were 4-5 double peaks in these electropherograms.

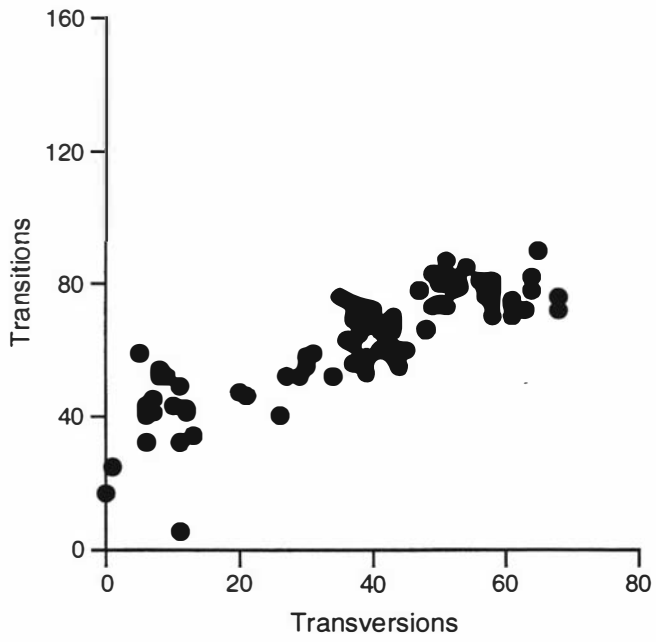


Figure 4.7. The number of transversions compared with the number of transitions in the HVR-1 data set from 15 penguin species.

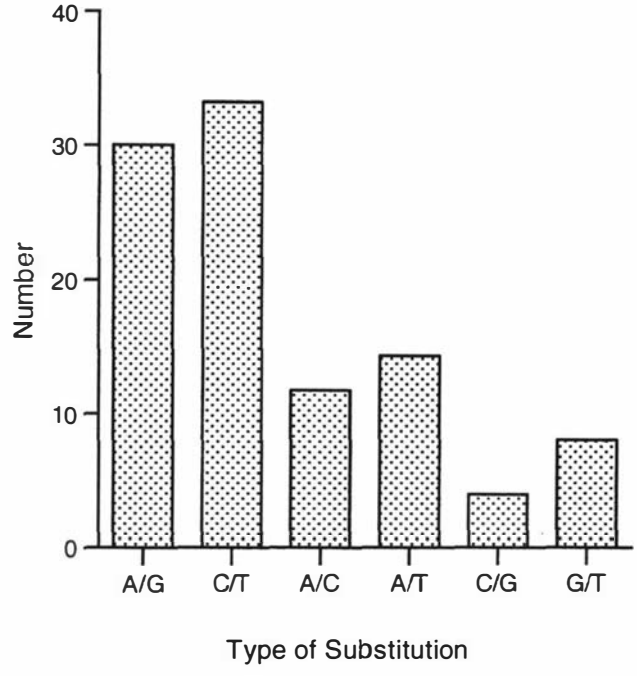


Figure 4.8. The type and number of substitutions in the HVR-1 sequences. Transitions (A/G and C/T) occur more frequently that transversions (A/C, A/T, C/G and G/T). Overall there is an average of 63 transitions and 38 transversions.

Table 4.4. Pairwise comparisons of penguin taxa for their percentage sequence difference (lower diagonal) and number of transitions compared with the number of transversions (upper diagonal) in the HVR-1 data set.

	Ade	Gen	Chi	Mag	Roy	Fio	Sna	Roc	Ere	Yel	Bla	Mac	Hum	Gal	Lit
Ade	-	70/37	47/20	75/58	83/49	73/51	82/50	78/47	74/50	80/50	74/58	80/50	70/61	70/61	85/54
Gen	0.215	-	46/21	82/64	75/58	76/57	78/58	78/58	81/56	72/63	78/64	70/58	72/68	76/68	90/65
Chi	0.168	0.168	-	81/58	79/53	80/50	87/51	78/52	73/49	80/53	77/57	82/52	75/61	73/61	81/58
Mag	0.332	0.366	0.359	-	67/41	61/42	70/43	68/41	60/41	60/44	32/6	67/39	40/6	43/6	71/39
Roy	0.330	0.335	0.341	0.270	-	43/10	49/11	52/8	54/8	52/27	66/41	5/11	67/38	65/38	75/36
Fio	0.311	0.336	0.335	0.257	0.133	-	25/1	34/13	42/6	58/30	55/44	41/7	58/39	57/39	53/39
Sna	0.330	0.342	0.355	0.282	0.150	0.065	-	42/12	45/7	59/31	60/45	53/8	72/40	71/40	58/38
Roc	0.314	0.345	0.336	0.273	0.151	0.118	0.134	-	32/11	55/30	66/43	52/9	73/38	67/38	56/37
Ere	0.311	0.345	0.314	0.252	0.156	0.120	0.129	0.107	-	40/26	58/43	59/5	69/37	62/37	52/34
Yel	0.325	0.339	0.343	0.260	0.198	0.220	0.224	0.212	0.165	-	66/48	52/29	66/43	61/43	63/42
Bla	0.346	0.375	0.362	0.099	0.281	0.260	0.275	0.286	0.264	0.299	-	67/41	41/12	41/12	67/43
Mac	0.326	0.322	0.346	0.264	0.040	0.120	0.152	0.153	0.160	0.200	0.283	-	63/36	63/36	76/35
Hum	0.326	0.349	0.351	0.115	0.263	0.243	0.280	0.279	0.264	0.273	0.139	0.248	-	17/0	70/37
Gal	0.326	0.359	0.346	0.122	0.258	0.241	0.278	0.263	0.247	0.260	0.138	0.248	0.042	-	69/37
Lit	0.348	0.390	0.359	0.276	0.280	0.232	0.241	0.236	0.217	0.264	0.291	0.280	0.270	0.267	-

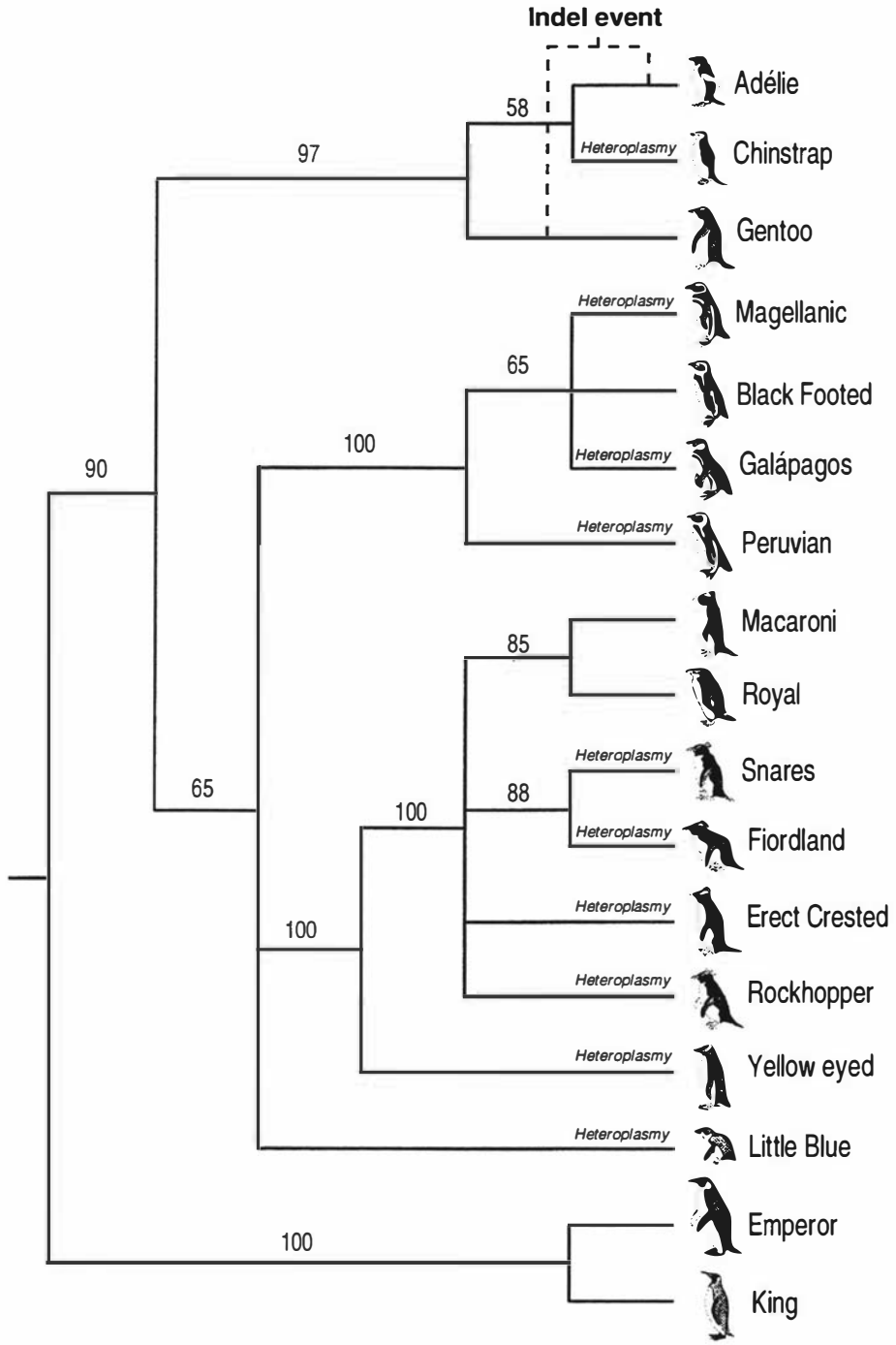


Figure 4.9. The reconstructed phylogeny of heteroplasmy in ten species and the ~100b.p. indels event recorded in the gentoo and Adélie penguins. This tree topology is based on the bootstrapped MP analysis of the ribosomal genes.

4.4 Discussion

4.4.1 *Phylogeny and Sequence Diversity*

The phylogenetic analysis of the two mtDNA ribosomal genes shows that penguins are a monophyletic group. This result was not unexpected. Based upon morphological characters and the fossil record penguins have always been considered to have arisen from a common origin. The tree topologies at the generic level presented here are similar to analyses performed on morphology (O'Hara, 1989), morphometrics (Livezey, 1989) and *cob* gene sequences (Edge, 1996). The king and emperor penguins (*Aptenodytes* spp.) are the most basal species, which is in agreement with the findings of Sibley and Ahlquist (1990).

The six penguin genera are quite distinct from each other. Some ingroup comparisons had a similar number of transversional differences as ingroup to outgroup comparisons. This pattern is consistent with other mtDNA sequences from penguins. For example, Edge (1996) presented a ML tree (based on *cob* sequences) that placed the *Aptenodytes* spp. as basal; however, a MP tree in the same study showed the pygoscelid species as the basal group. The bootstrap proportions for the difference placement of these genera was <50% for both ML and MP methods suggesting the topology was not robust. In contrast, the ribosomal DNA data presented here showed high bootstrap support for the basal nodes (>90%) and supported the first split to the *Aptenodytes* spp. Using the raw data of Edge (1996), a plot of ti vs tv was constructed (see Appendix B.5). This graph reveals a similar, but more pronounced pattern than that found for the ribosomal data. There is a steep increase in ti that quickly reaches a plateau at the same level as the outgroup taxa. This sudden saturation could result in ambiguities for the deeper branches in a penguin phylogeny. Because of the contrasting depths of splits within the penguins, a combination of gene sequences with different rates of evolution will be better at resolving different parts of the tree.

The shallow sequence divergence within both the eudyptids and spheniscid species was perhaps not surprising given the close morphological resemblance of species within these groups. The eudyptid species all share crests or tufts of feathers on their heads. Fiordland and Snares penguins are sometimes difficult to distinguish,

apart from slight differences in their bills, head stripes, and yellow plumes (Warham, 1974b; Warham, 1974a). Considering the ribosomal gene sequence data only a single ti separates this species pair, while HVR-1 sequences showed 25 ti and one tv. Simpson (1976) raised the question of whether royal and macaroni penguins were a different species or subspecies. The royal penguins are distinguished from the other eudyptids by a white coloured chin and side of the head. However, an occasional morph with a white coloured head has turned up in macaroni penguin colonies, and royal penguins without the distinctive white coloured patterns have also been seen. For these species the ribosomal DNA sequences show three substitutions, while 16 (4%) are seen between HVR-1 sequence comparison. Among individuals of the royal penguin species, 1.3% sequence difference was detected. The average sequence divergence within Adélie penguins was the largest at 4.1%. When Adélie penguin HVR-1 sequences were compared with their closest relatives, the mean sequence divergence was 24.8% with gentoo penguins and 30% with chinstrap penguins (Table 4.4). Thus, there was more intraspecific variation in Adélie penguins than between the interspecific comparison of macaroni and royal penguins.

The ribosomal DNA sequence of the spheniscids revealed the absolute number of substitution does not always correspond to the placement of taxa on the phylogenetic trees. Morphology and analyses of *cob* and HVR-1 sequence concur with the ML and UPGMA analysis of *rnl* and *rns*; magellanic and black footed are more closely related to each other than to either the Peruvian or Galápagos species. Interestingly, the magellanic, Peruvian and Galápagos penguins have all been known to hybridise both in captivity and in the wild (Williams, 1995).

Numerous fossilised penguin species have been recovered from the late Tertiary period, however, only two species (*Pygoscelis tyreei* and *Aptenodytes ridgeni*) are considered part of the modern fauna (Simpson, 1972). These two late Pliocene species were recovered from boulders on Motunau Beach in the South Island of New Zealand, and considered closely related to the modern gentoo and emperor penguins, respectively. Unfortunately the absence of HVR-1 data from the emperor and king penguins, precluded using these fossil records to estimate a phylogenetic rate of substitution for the control region in the pygoscelid and aptenodytid species.

4.4.2. The Evolution of Heteroplasmy and Indels in the HVR-1

Ten of the penguin species showed presumptive mtDNA heteroplasmy. This heteroplasmy is most likely due to length differences and appears at a common position in the control region for all penguin species. Various forms of heteroplasmy have been described in mtDNA, the most common are differences at particular nucleotide positions among genomes (e.g. Lightowlers et al., 1997; Chinnery et al., 2000). Other forms of heteroplasmy are the result of a variable number of repetitive sequences (Lunt et al., 1998). The type of length heteroplasmy that appears in penguins has not yet been described for any other vertebrate species. The most interesting aspect of this heteroplasmy is that it may be associated with the termination of, or extension from, the D-loop during replication. It could be that the daughter H-strand begins to extend and then prematurely terminates at the Termination Associated Sequences (TAS). This type of heteroplasmy could be explained if the extending strand skipped on the secondary structure as it reaches the point of termination. Alternatively, slippage could also occur on these complex secondary structures when replication of the daughter H-strand is re-initiated from the D-loop (see Figure 3.3 from Chapter Three).

In Figure 4.9 the presence or absence of length heteroplasmy has been plotted onto the MP phylogenetic tree. It appeared from the tree topology that there were either multiple gains or losses of heteroplasmy in penguins. For example, heteroplasmy was detected in the chinstrap penguin, but not in either the gentoo or Adélie penguins. These latter taxa branch before and after the chinstrap respectively. There are also three other losses of heteroplasmy across the tree (includes black footed, macaroni and royal penguins). The samples sizes in this study are too small to detect the prevalence of heteroplasmy throughout a species. Moreover, with a small sample size the presence of heteroplasmy is conclusive, but a negative result is not. Length heteroplasmy of this type may be the result of a particular sequence type arising through point mutations at the area near the D-loop termination. If a particular sequence array arose (e.g. increased G + C content) at important positions, it might tend to reduce slippage events and hence heteroplasmy. The rapid elimination of haplotypes at the oocyte bottleneck of mtDNA inheritance, could quickly remove heteroplasmy from a maternal lineage in a few generations (Chinnery et al., 2000). Longitudinal sampling of HVR-1 sequences within species

may show that heteroplasmic states become homoplasmic over a short evolutionary time.

There is a 100 b.p. indel in the HVR-1 of gentoo and Adélie penguins compared with all the other penguin species sampled. Unfortunately, the basal king and emperor penguins were not represented in this study so polarity (e.g. ancestral or derived) could not be ascribed to this character. If both the emperor and king penguins possessed the indel then it is more likely that it was lost (i.e. a deletion event) in all other penguins, including the close relatives of the gentoo and Adélie; the chinstrap. In Chapter Three illegitimate elongation was identified to explain the presence of 81 b.p.-repeats in the 3'-end of the Adélie penguin mtDNA control region. Similarly, illegitimate extension of the D-loop could explain the presence of the large indel in the gentoo and Adélie penguins. Intraspecific comparisons of the 3'-end among Adélie penguins revealed a variable number of tandem repeats (Chapter Three). In contrast to the 3'-end repetitive sequences, the 5'-end indel in Adélie and gentoo penguins appears to evolve over long periods of time. In summary, given that Adélie and gentoo penguins share the same length HVR-1 sequence it is therefore the most appropriate species to use as the outgroup sequence to Adélies in the following chapters.

4.5 References

Chinnery, P.F., Thorburn, D.R., Samuels, D.C., White, S.L., Dahl, H.-H.M., Turnbull, D.M., Lightowlers, R.N., and Howell, N. 2000. The inheritance of mitochondrial DNA heteroplasmy: random drift, selection or both? Trend. Genet. **16:** 500-505.

Edge, K.-A. 1996. Parental investment in penguins: a phylogenetic and experimental approach. Ph.D. Thesis, University of Otago, New Zealand.

Efron, B. 1982. The jacknife, the bootstrap, and other resampling plans. CBMS-NSF Regional Conference in Applied Mathematics, Monograph 38, Society of Industrial and Applied Mathematics, Philadelphia.

Faith, D.P. 1991. Cladistic permutation tests for monophyly and nonmonophyly. Syst. Zool. **40:** 366-375.

Farris, J.S. 1969. A successive approximations approach to character weighting. Syst. Zool. **18:** 374-385.

Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution **39:** 783-791.

Fordyce, R.E., Jones, C.M., and Field, B.D. 1986. The world's oldest penguin? Geological Society of New Zealand **74:** 56.

Fordyce, R.E., and Jones, C.M. 1990. Penguin history and new fossil material from New Zealand. *In* Penguin Biology. *Edited by* L. S. Davis, and J. T. Darby. Academic Press, San Diego.

Harrison, C. 1984. Holocene penguin extinction. Nature, **310:** 545.

Hillis, D.M., Huelsenbeck, J.P., and Cunningham, C.W. 1994. Application and accuracy of molecular phylogenies. Science, **264:** 671-677.

Ho, C.Y.-K., Prager, E.M., Wilson, A.C., Osuga, D.T., and Feeney, R.E. 1976. Penguin evolution: protein comparisons demonstrate phylogenetic relationship to flying aquatic birds. J. Mol. Evol. **8:** 271-282.

Jukes, T.H., and Cantor, C.R. 1969. Evolution of protein molecules. *In* Mammalian Protein Metabolism. *Edited by* H. N. Munro. Academic Press, New York. pp. 21-132.

Kimura, M. 1980. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. **16:** 111-120.

Kocher, T.D., Thomas, W.K., Meyer, A., Edwards, S.V., Pääbo, S., Villablanca, F.X., and Wilson, A.C. 1989. Dynamics of mitochondrial DNA evolution in animals: Amplification and sequencing with conserved primers. Proc. Natl. Acad. Sci. USA **86:** 6196-6200.

Lightowlers, R.N., Chinnery, P.F., Turnball, D.M., and Howell, N. 1997. Mammalian mitochondrial genetics: heredity, heteroplasmy and disease. Trend. Genet. **13:** 450-455.

Livezey, B. 1989. Morphometric patterns in recent and fossil penguins (Aves, Sphenisciformes). J. Zool. **219:** 269-307.

Lunt, D.H., Whipple, L.E., and Hyman, B.C. 1998. Mitochondrial DNA variable number of tandem repeats (VNTRs): utility and problems in molecular ecology. Mol. Ecol. **7:** 1441-1455.

O'Hara, R. 1989. An estimate of the phylogeny of living penguins (Aves: Spheniscidae). Amer. Zool. **29:** 11A.

Olson, S.L. 1985. The fossil record of birds. *In* Avian Biology. *Edited by* D. S. Farner, J. R. King and K. C. Parkers. Academic Press, New York. pp. 79-252.

Olson, S.L., and Hasegawa, Y. 1979. Fossil counterparts of giant penguins from the North Pacific. Science, **206:** 688-689.

Palumbi, S.P. 1995. Nucleic Acids II: The Polymerase Chain Reaction. *In* Molecular Systematics. *Edited by* D. M. Hills, C. Moritz and B. K. Mable. Sinauer Associates, Inc, Massachusetts. pp. 205-247.

Paterson, A.M., Wallis, G.P., and Gray, R.D. 1995. Penguins, petrels, and parsimony: does cladistic analysis of behaviour reflect seabird phylogeny. Evolution **49:** 974-989.

Saccone, C., Pesole, G., and Sbisá, E. 1991. The main regulatory region of mammalian mitochondrial DNA: structure-function model and evolutionary patterns. J. Mol. Evol. **33:** 83-91.

Saitou, H., and Nei, M. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4:** 406-425.

Sambrook, J., Fritsch, E.F., and Maniatus, T. 1989. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, New York.

Saunders, M.A., and Edwards, S.V. 2000. Dynamics and phylogenetic implications of mtDNA control region sequences in new world jays (Aves: Corvidae). J. Mol. Evol. **51:** 97-109.

Sbisà, E., Tanzariello, F., Reyes, A., Pesole, G., and 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.

Sibley, C.G., and Ahlquist, J.E. 1990. Phylogeny and Classification of Birds. Yale University Press, New Haven.

Simpson, G.G. 1972. Pliocene penguins from North Canterbury, New Zealand. Records of the Canterbury Museum **9:** 159-182.

Simpson, G.G. 1976. Penguins: Past and Present, Here and There. Yale University Press, New Haven, Connecticut.

Stonehouse, B. 1975. Introduction: the Spheniscidae. *In* The Biology of Penguins. *Edited by* B. Stonehouse. The MacMillan Press Ltd, London. pp. 1-15.

Swofford, D.L. 2000. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Sinauer Associates, Sunderland, Massachusetts.

Tamura, K. 1992. Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+C-content biases. Mol. Biol. Evol. **9:** 678-687.

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

Warham, J. 1974a. The Fiordland crested penguin *Eudyptes pachyrhynchus*. Ibis **116:** 1-27.

Warham, J. 1974b. The breeding biology and behaviour of the Snares crested penguin. J. Roy. Soc. NZ **4:** 63-108.

Williams, P.L., and Fitch, W.M. 1989. Finding the minimal change in a given tree. *In* The Hierarchy of Life. *Edited by* B. Fernholm, K. Bremer and H. Jörnfall. Elsevier Press, Amsterdam. pp. 453-470.

Williams, T.D. 1995. The Penguins: Spheniscidae. Oxford University Press, Oxford.

Yang, Z. 1994. Maximum likelihood phylogenetics estimation from DNA sequences with variable rates over sites: approximate methods. J. Mol. Evol. **39:** 306-314.

Chapter Five

The Two Maternal Lineages of Adélie Penguins: An Ice-Age Legacy

5.1 Introduction

The rapidly evolving and maternally inherited mtDNA has been used extensively to study gene flow and population-genetic structure in species (Avise, 1992; Avise, 2000). Genetic structuring of mtDNA can result from processes acting within and among populations, and/or historical events such as habitat fragmentation, range expansion and colonisation (Templeton, 1995). Marked genetic structure has been correlated with limited dispersal ability and morphological clines, for example in the flightless brown kiwi (Baker et al., 1995) and red-winged blackbirds (Ball et al., 1988). However, cryptic population structure has been reported in many species including migratory humpback whales (Baker et al., 1994) and holarctic-breeding dunlins (Wenink et al., 1996). In these cases, behavioural characteristics such as sex-biased dispersal or natal philopatry were suggested to have had the greatest influence on structuring. For many terrestrial species, phylogeographic structuring and recent speciation events have been linked to bottlenecks and founder events caused by disruptive Pleistocene ice-ages (Avise, 1989; Avise and Walker, 1997;

Hewitt, 2000). The goal of many intraspecific mtDNA studies has been to understand what factors result in either the absence of genetic diversity, or promote genetic divergence among populations.

A range of factors could influence the population-genetic structure of Adélie penguins, including their behaviour, population size, circum-Antarctic distribution and the glacial history of the continent. The total population size of Adélie penguins is estimated to be around 15 million (Williams, 1996). They are known to breed as far south as Cape Royds, Ross Island (77°33'S, 166°10'E), and as far north as Bouvetøya (54°26'S, 3°24'E). All of the known breeding locations for Adélie penguins are within the Antarctic convergence associated within the limits of the pack-ice (Fraser et al., 1992). The Ross Sea region harbours the highest densities of breeding pairs, with the largest rookery of 282,307 pairs situated at Cape Adare (Taylor et al., 1990). Two abiotic characteristics are highly correlated with the presence of Adélie colonies. The first is a stony ice-free area, since Adélies use small stones to build their nests. The second is ready access to open sea, as Adélies feed mainly on krill and small fish. In addition, areas of open water surrounded by ice, known as coastal polynya, have been implicated as an important feature contributing to the distribution of many Arctic (Falk and Møller, 1997) and Antarctic (Ainley, 2001) bird species. Antarctic coastal polynya are often created by katabatic winds (cold dense air literally falling down slopes such as glaciers due to gravity) flowing off the continent and pushing pack-ice out to sea (Gordon and Comiso, 1988).

Like most penguin species, Adélies are colonial breeders. A regular annual breeding cycle is important since the Antarctic has such a short summer window. They are socially monogamous, although extra-pair copulations in exchange for nest building material (small pebbles) is a common feature at colonies (Hunter and Davis, 1998). Generally, Adélie penguins lay two eggs each season. Breeding usually begins in the austral spring and chicks fledge 5-6 months later during February or March. Two to three year old individuals wander through their natal rookery and engage in low-key territorial displays (Sladen, 1953). Ainley et al. (1983) reported significant numbers beginning to breed at 5-7 years in males and 4-6 years in females at Cape Crozier. Based on mark-recapture data, 96% of those

precipitated with 2.5 volumes of 100% ethanol followed by a 70% ethanol wash. The precipitant was pelleted in a microfuge, dried and resuspended in 200 µl of TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA).

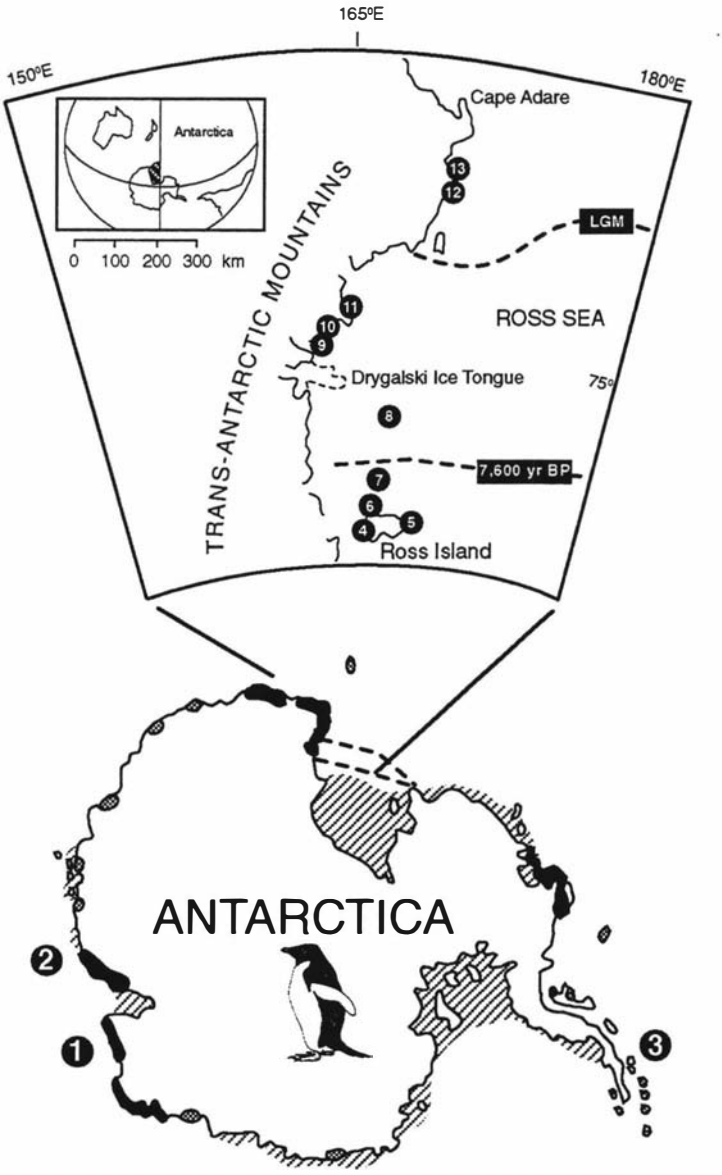


Figure 5.1. The distribution of active Adélie penguin rookeries in Antarctica (black and grey indicate areas of high and low density respectively). The location of the permanent ice at two different times during the Pleistocene is indicated by dotted lines, LGM = Last Glacial Maximum. The sites of collection, with latitude and longitude coordinates, are given: **1.** Welch Island (67°33'S, 62°55'E); **2.** Gardner Island (77°52'S, 68°34'E); **3.** Torgersen Island (64°46'S, 64°05'W); **4.** Cape Royds (77°33'S, 166°10'E); **5.** Cape Crozier (77°14'S, 166°28'E); **6.** Cape Bird (77°30'S, 162°10'E); **7.** Beaufort Island (76°56'S, 167°03'E); **8.** Franklin Island (76°07'S, 168°15'E); **9.** Inexpressible Island (74°53'S, 163°45'E); **10.** Adélie Cove (74°45'S, 164°00'E); **11.** Edmonson Point (74°19'S, 165°04'E); **12.** Cape Wheatstone (72°37'S, 170°14'E); **13.** Cape Hallett (72°19'S, 170°12'E).

The hypervariable region I (HVR-1) was amplified using the primers specific to Adélie penguins (reference numbers correspond to the sequence reported in Ritchie and Lambert, 2000) AH530 (5'- CTGATTTCACGTGAGGAGACCG-3'), AH432 (5'-GCTGGTCCTTGTACCATGGAC-3'), AL93 (5'-CATTTAATGTACGT ACTAGGA C-3'), and L-tRNA^{Glu} (5'-CCCGCTTGGCTTYTCTCCAAGGTC-3'). Amplifications from 1 µl of the extracts were conducted in 25 µl volumes with 1 unit of Ampli*Taq* (PE Biosystems), 10 mM Tris-HCl 8.0, 50 mM KCl, 1.5 mM $MgCl_2$, 0.4 µM of each primer and 200 µM of each dNTP. PCR products were purified with the QIAquick PCR purification kit before direct sequencing using the Big Dye Terminator sequencing kit (PE Biosystems) and analysis on a 377A automated sequencer (PE Biosystems).

To verify that amplified sequences were of mitochondrial origin, rather than nuclear copies, a large portion (*ca.* 6.0 kb) of the Adélie penguin mitochondrial genome was amplified and sequenced. The long PCR template was generated using the Expand™ Long Template PCR System (Boehringer Mannheim) using two PCR-primers designed to conserved portions of the *rnl* gene 5'-TGATTGCGCTACCTTCGCACGGTTAGGATACC-3' (H-ExpPeng-rnl) and the *cob* gene 5'-CCATTCCACCCCTACTACTCCACAAAAG A-3' (L-ExpPeng-cob). All PCR reactions contained 0.3 µM of each primer, 500 µM of dNTPs, PCR Buffer 3 (50 mM Tris-HCl pH 9.2, 16 mM $(NH_4)_2SO_4$, 2.25 mM $MgCl_2$, 2% DMSO and 0.1% Tween 20) and 2.6 units of Expand™ Long Template enzyme mix. The resulting PCR-product was purified using High Pure™ PCR Product Purification columns (Boehringer Mannheim), and the control region was amplified and sequenced as above.

Sequences were aligned in CLUSTAL V and managed in Sequencher ver. 3.1.1 (GeneCodes, Incorp.). Base frequencies and pairwise comparisons of sequence differences were determined using PAUP*4.0b4a (Swofford, 2000). To accommodate rate variation among sites the shape parameter (α) of a gamma distribution (Γ) was approximated. An estimate of α was obtained by calculating the frequency distribution of parsimony-type changes required at each site across a neighbor-joining tree, and fitting a negative binomial distribution to these frequencies (Sullivan et al., 1995). Haplotypic diversity was calculated as $h = (n/(n-1))(1-\sum f_i^2)$, where f_i is the frequency of the ith haplotype in a sample of n individuals.

Nucleotide diversity was calculated as $\pi = (n/(n\text{-}1))(\sum f_i f_j p_{ij})$, where f_i and f_j are the frequencies of the ith and jth haplotypes in a sample of size n, and p_{ij} is the estimated sequence divergence between the ith and jth sequence (Nei, 1987).

Using Jukes-Cantor corrected distances, a test for geographic structuring was performed using an Analysis of Molecular Variation (AMOVA) (Excoffier et al., 1992) with Arlequin ver. 2.0. An AMOVA calculates a Φ_{ST} statistic that is analogous to an F_{ST} measurement of fixation. A significance test of the Φ_{ST} was performed using 1,000 random permutations of samples drawn from the total population. The significance level (0.05) was adjusted using a sequential Bonferroni procedure (Rice, 1989). A Mantel's permutation test (Mantel, 1965) was used to compare matrices of distance between populations (kilometres) and the Φ_{ST} statistics. The relationships among haplotypes were determined using a split decomposition graph in SPLITS TREE 2.4 (Huson, 1997), a minimum spanning network (Kurskal, 1956), and neighbor-joining (Saitou and Nei, 1987) in PAUP* 4.0b4a. A UPGMA tree was constructed using HKY85 corrected maximum likelihood distances and demographic history inferred using a likelihood evaluation and a *skyline plot* in GENIE ver. 1.0 (Pybus et al., 2000).

5.3 Results

5.3.1. Summary Statistics

A total of 381 individuals from 13 Antarctic collection sites were sequenced during this study and analyses were conducted on a 352 b.p. portion of the HVR-1 L-strand sequence. This smaller portion of the HVR-1 control region was chosen for compatibility with the results in the next chapter on ancient DNA. The 352 b.p. region showed asymmetric base frequencies ($\pi_A = 0.30$, $\pi_T = 0.31$, $\pi_C = 0.20$, and $\pi_G = 0.19$) and strong bias towards transitional mutations (frequencies C↔T = 0.65, G↔A = 0.35 and transversions < 0.01) over 153 variable sites. There were 153 variable sites, of which 124 were informative for parsimony. Using a neighbor-joining tree, a maximum likelihood estimate of the purine/pyrimidine transition ratio (κ) was calculated as 1.46 and the transition/transversion ratio (τ) as 47.3. The distribution of the number of parsimony-type changes across the neighbor-joining tree had a

mean of 3.153 (σ = 12.817), and the gamma correction parameter was approximated at 0.4366 (the distribution is presented in Appendix B.6).

There were a total of 287 haplotypes detected in the pooled sample of all populations. The sample sizes, number of haplotypes, haplotypic diversity (*h*), nucleotide diversity (π) and range of sequence differences for each population are presented in Table 5.1. Populations in the Ross Sea have a higher nucleotide diversity (average π = 0.0423 ± 0.0067) than those sampled from other Antarctic locations (average π = 0.0198 ± 0.0011). The haplotypic diversity within each population is also high (0.9973 from the pooled sequences). Figure 5.2. shows the number of individuals with a unique haplotype (n = 249) along with the number of individuals (ranging from 2-9) that share a haplotype. Tajima's (1989) test for selective neutrality (D statistic) was applied to the sequences and showed no significant deviation from neutrality in any of the populations.

Table 5.1. The sample size (*n*), number of haplotypes (no. H), haplotypic diversity (*h*), nucleotide diversity (π) and the range of sequence dissimilarity among pairwise comparisons for all populations sampled. The asterisk (*) indicates sequences that were not collected as part of this thesis research.

Population	*n*	no. H	*h*	π	range
Cape Bird	123	104	0.9969	0.0393	0- 0.11
Cape Royds	38	33	0.9915	0.0312	0-0.10
Cape Crozier	29	26	0.9901	0.0465	0-0.10
Beaufort Island	22	20	0.9870	0.0381	0-0.09
Franklin Island	16	15	0.9917	0.0411	0-0.11
Inexpressible Island	27	24	0.9915	0.0360	0-0.10
Adélie Cove*	16	16	1.0000	0.0421	0.003-0.10
Edmonson Point*	13	13	1.0000	0.0438	0.003-0.09
Cape Wheatstone*	10	10	1.0000	0.0534	0.003-0.09
Cape Hallett	29	28	0.9975	0.0513	0-0.11
Torgersen Island	16	15	0.9917	0.0205	0-0.032
Welch Island	21	21	1.0000	0.0205	0.004- 0.03
Gardner Island	21	21	1.0000	0.0186	0.003-0.03
Pooled	381	287	0.9973	0.0380	0-11.3

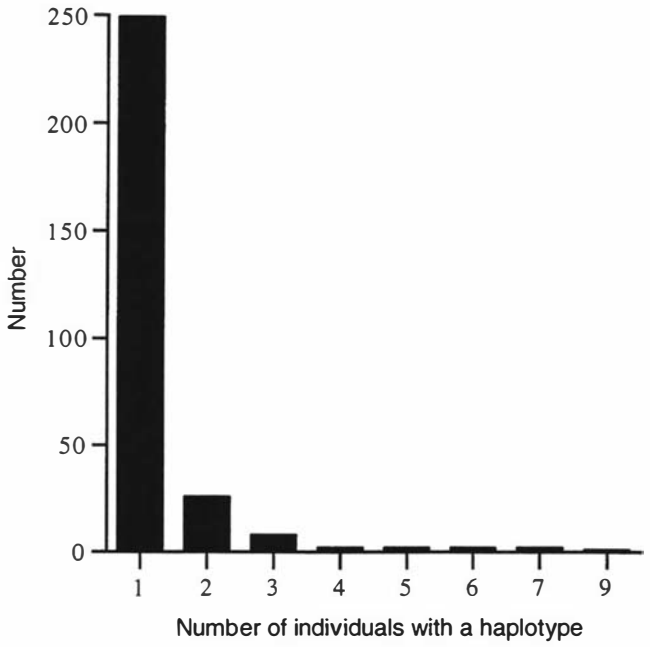


Figure 5.2. The frequency distribution of individuals in which any one haplotype is present (ie. 249 haplotypes are unique to an individual; one haplotype is present in nine individuals).

5.3.2. Population Subdivision

The overall ϕ_{ST} (= 0.029, P = 0.002) was low, indicating that only 2.9% of the variation was due to population differences and hence there is very little population structuring. Table 5.2 shows the nucleotide differences between populations (expressed as ϕ_{ST} values) and whether or not each result is significantly different from randomly drawn individuals from the two populations pooled. Using a sequential Bonferroni adjustment of the significance level, Cape Wheatstone shows significant ϕ_{ST} statistics when compared with the three populations outside the Ross Sea. Under the same criterion Gardner Island shows a significant difference when compared to Cape Hallett. Figure 5.3 is a plot of geographic distance (kilometres) versus the ϕ_{ST} value for each population comparison. The Mantels test showed no significant association between the ϕ_{ST} statistic and geographic distance matrices (r^2 = 0.025, P = 0.19).

Table 5.2. Pairwise comparisons of the Φ_{ST} values among populations of Adélie penguins. The lower diagonal shows the Φ_{ST} values, an expression of nucleotide difference between populations. The upper diagonal represents probabilities of the observed Φ_{ST} value occurring by chance in 1,000 random permutations. The bolded numbers show the significant comparisons using a sequential Bonferroni adjustment at the 0.05 level. The first three letters from the name of each population are presented.

	Bir.	Roy.	Cro.	Bea.	Fra.	Ine.	Adé.	Edm.	Whe.	Hal.	Tor.	Wel.	Gar.
Bir.	-	0.310	0.087	0.676	0.896	0.858	0.996	0.760	0.007	0.038	0.023	0.969	0.016
Roy.	0.000	-	0.040	0.303	0.433	0.250	0.516	0.525	0.095	0.334	0.011	0.320	0.007
Cro.	0.015	0.050	-	0.385	0.490	0.374	0.459	0.238	0.003	0.004	0.083	0.916	0.027
Bea.	-0.010	0.004	-0.002	-	0.815	0.643	0.968	0.747	0.014	0.141	0.035	0.627	0.017
Fra.	-0.019	-0.007	-0.007	-0.027	-	0.967	0.950	0.799	0.032	0.212	0.120	0.999	0.039
Ine.	-0.011	0.006	0.000	-0.016	-0.030	-	0.892	0.516	0.013	0.081	0.120	0.999	0.033
Adé.	-0.026	-0.013	-0.006	-0.038	-0.043	-0.028	-	0.882	0.086	0.285	0.083	0.949	0.037
Edm.	-0.019	-0.016	0.010	-0.031	-0.037	-0.016	-0.044	-	0.093	0.424	0.018	0.339	0.019
Whe.	0.172	0.080	0.282	0.175	0.160	0.205	0.144	0.111	-	0.262	**0.000**	**0.000**	**0.000**
Hal.	0.032	-0.003	0.093	0.028	0.019	0.047	0.008	-0.008	0.011	-	0.003	0.027	**0.000**
Tor.	0.058	0.108	0.026	0.060	0.043	0.031	0.057	0.097	0.387	0.171	-	0.999	0.026
Wel.	-0.017	0.000	-0.015	-0.009	-0.058	-0.062	-0.030	0.003	0.308	0.079	-0.036	-	0.999
Gar.	0.055	0.102	0.031	0.067	0.053	0.045	0.063	0.099	0.418	0.175	0.029	-0.045	-

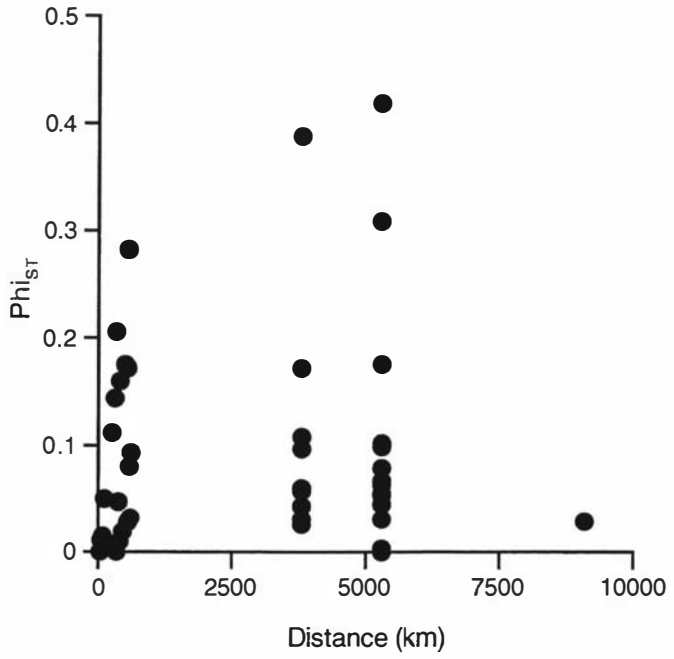


Figure 5.3. The isolation by distance graph. The ϕ_{ST} statistics are plotted against geographic distance (kilometres) for every population comparison.

5.3.3. Relationships Among Haplotypes

A mismatch distribution of all pairwise comparisons of percentage sequence difference is presented in Figure 5.4; this includes chinstrap and gentoo penguins, which were used as outgroup species. The sampled Adélie penguin sequences differ from their two closest relatives by 25-33%. The most striking feature of this distribution is the bimodality of sequence differences across all sampled Adélies; one mode is at 2% and the other at 8% sequence difference.

The split decomposition graph, minimum spanning network and neighbor-joining tree are presented in Figure 5.5 (a), (b) and (c) respectively. The most prominent feature of these analyses is the presence of two major lineages. Split decomposition graphs of a large number of taxa are often unresolved (Bandelt and Dress, 1992). However, the single line connecting the two lineages in the present data set indicates very little conflict, or even none, among sequence comparisons for this split. Thirteen mutually compatible sites were used to construct the minimum spanning network in Figure 5.5 (b); eight of these positions supported the two

lineage split and five other positions supported the data being split further into seven subgroups. The outgroup sequences from gentoo and chinstrap penguins fall in the middle of this two lineage split, according to the eight sites used to differentiate the lineages on the network. A bootstrapped NJ tree, with 1,000 replicates and rooted with both the gentoo and chinstrap penguins, shows 100% support for the two lineage split.

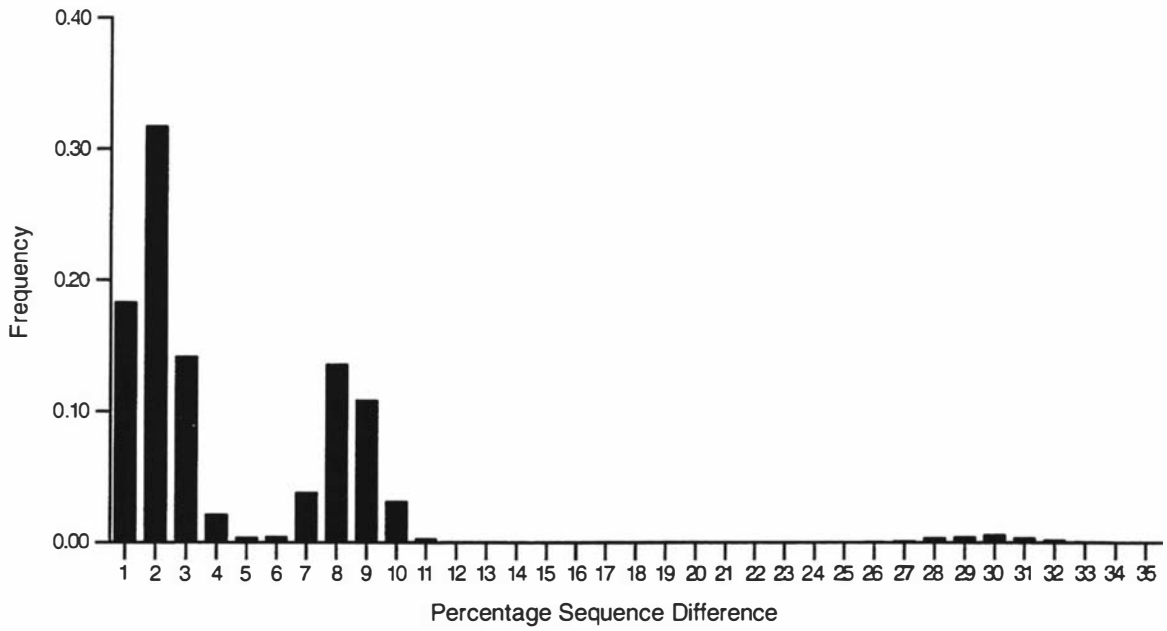


Figure 5.4. Mismatch distribution of all pairwise comparisons of Adélie penguins and to their two closest relatives the gentoo and chinstrap penguins

Two checks were performed to control for the possibility of a nuclear rather than mitochondrial origin for the sequence data. First, the same sequences were obtained across the control region for both the long and short PCR product, indicating that both had been amplified from the same template. Secondly, sequence data obtained from 640 b.p. of the *cob* mtDNA gene (n=10) shows a fixed difference between these two HVR-1 types, as a single G↔A transition. This single difference is a silent substitution at the third position of a serine amino acid (base pair number 15,214 in the chicken mitochondrial genome).

The geographic distribution of each lineage is presented in Figure 5.6. One lineage was recorded exclusively in the Ross Sea and is designated the Ross Sea (*RS*) lineage, whereas the other was present at all locations around the Antarctic continent and is designated the Antarctic (*A*) lineage. A feature of the lineage

distribution is the decreasing frequency of the *RS* lineage with increasing latitude. In Figure 5.6 the frequency cline has been divided into regional groups according to known locations of the grounded ice shelf during the retreat of the Ross Ice sheet since the Last Glacial Maximum (LGM) 17-20 kyr BP. These four groupings comprised Ross Island (Capes Bird, Royds and Crozier, and Beaufort Island), Terra Nova (Franklin and Inexpressible Islands, Adélie Cove and Edmonson Point), Hallet (Capes Wheatstone and Hallett) and the non-Ross Sea (Torgersen, Welsh and Gardner Islands). The frequencies of the *A* and *RS* lineages in each regional group are presented in Table 5.3 and Figure 5.6. A contingence χ^2-test was performed to compare the relative numbers of each lineage between regions (Table 5.3). This test showed a significant difference in all comparisons at the 99% confidence level, excluding the comparison of Ross Island with Terra Nova.

Table 5.3. The *P* values for the contingence χ^2-tests among the four regional groupings and the frequency of each lineage within each group.

	Ross Island	Terra Nova	Hallett	Frequencies
Ross Island	-			*A* = 0.81, *RS* = 0.19
Terra Nova	0.492	-		*A* = 0.82, *RS* = 0.18
Hallett	0.000	0.000	-	*A* = 0.36, *RS* = 0.64
non-Ross Sea	0.000	0.001	0.000	*A* = 1.00, *RS* = 0.00

The average uncorrected sequence divergence within each lineage, including each haplotype only once, is d_A = 0.0214 ± 0.0026 (n = 224) and d_{RS} = 0.0254 ± 0.0035 (n = 63); between these two lineages $d_{A/RS}$ = 0.0833 ± 0.0111. Standard errors were calculated in MEGA ver 2.0 (Kumar et al., 2001) using a bootstrap procedure with 500 replicates. To estimate the net distance between the two lineages, adjustment was made for within lineage polymorphism $d_{adjusted} = d_{A/RS} - 0.5\ (d_A + d_{RS})) = 0.0705 \pm 0.0123$ (Nei, 1972; Wilson et al., 1985).

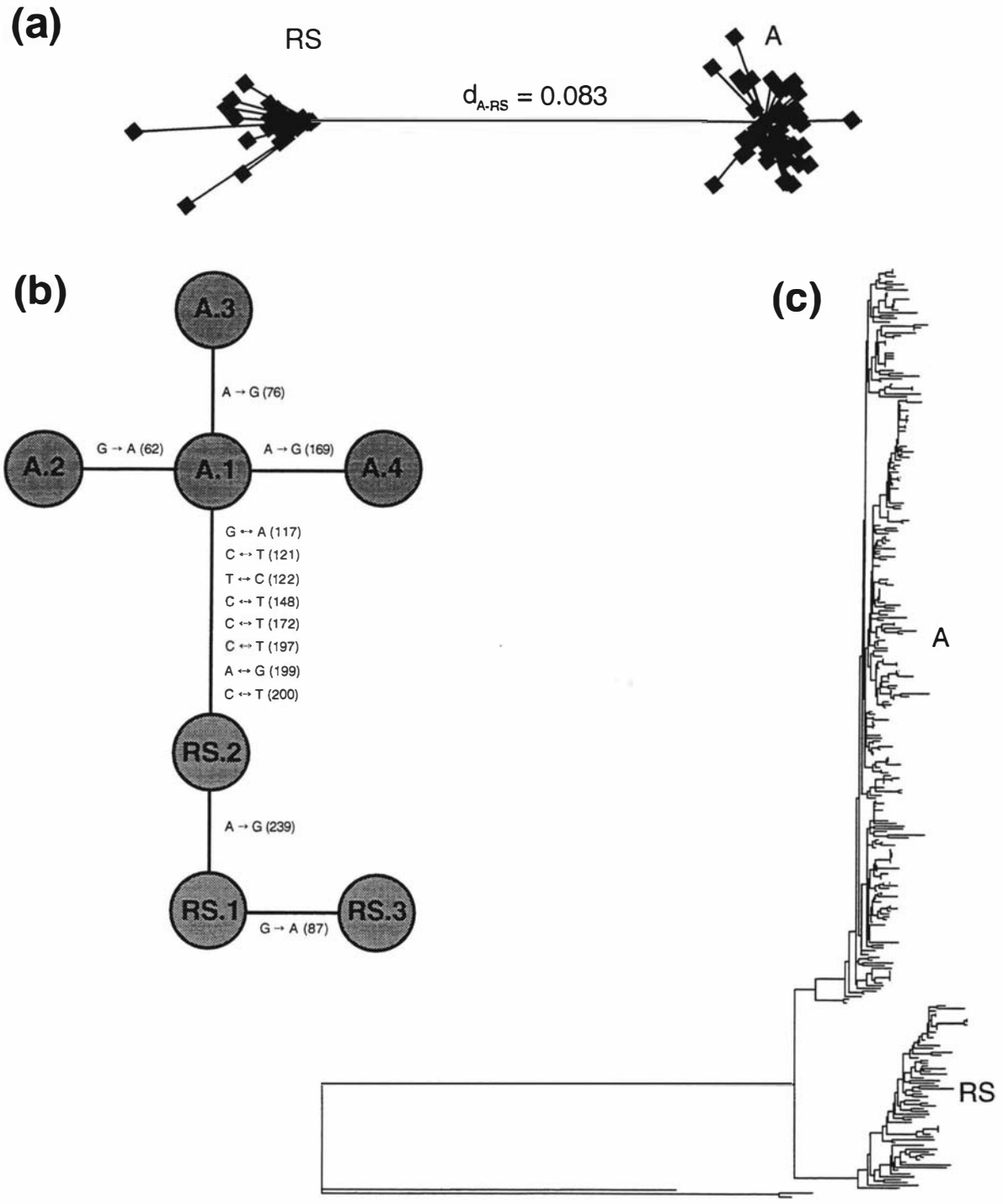


Figure 5.5. Split decomposition graph (a), minimum spanning network (taken from Lambert et al., Appendix D) (b), and a neighbor-joining tree (c). These analyses all show the clear split between lineages A and RS. The minimum spanning network shows seven subgroups (lineage A 1-4 and RS 1-3), with the defining substitutions on each connecting branch. The bracketed numbers represent the position in the HVR-1 data set for each transitional change. The neighbor-joining tree is rooted with two outgroup species, gentoo and chinstrap penguins.

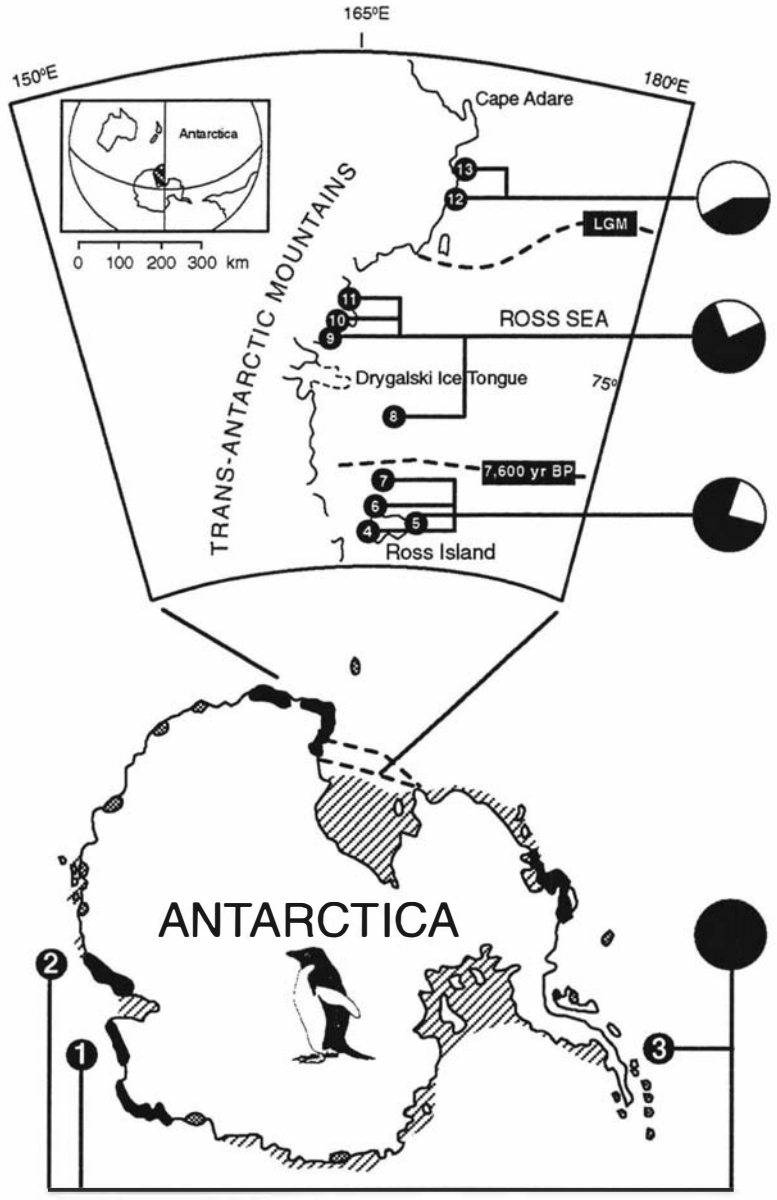


Figure 5.6. The relative frequencies of the *A* (black) and *RS* (white) lineages at sampled populations around Antarctica (see Table 5.3). The numbers correspond to populations identified in Figure 5.1. The location of the grounded ice at two different times during the Pleistocene is indicated by dashed lines, LGM = Last Glacial Maximum.

5.3.4. Population History

The population history of each lineage (*A* and *RS*) was inferred from reconstructed genealogies. This analysis assumes that as a population expands there is an increase in the survival rate of new lineages. A likelihood approach can be used to distinguish whether the observed genealogy (i.e. the number of coalescent events) is better explained by one of two demographic models: constant population size or exponential growth. First, a Kolmorgorov-Smirnov (KS) goodness-of-fit test was performed, which showed no significant difference between the observed (*A* and *RS* lineages) and the two expected (modelled) distributions (Table 5.4). Secondly, a maximum likelihood evaluation of each lineage showed that an exponential growth model better explained population history (Table 5.4). A likelihood ratio test (LRT) indicated a highly significant difference between maximum likelihood estimates for constant and exponential models. Furthermore, the confidence limits for ρ (exponential growth rate r divided by μ) never included zero. A comparison of model parameters (summarised as $\alpha = \rho\theta$) between lineages can suggest either a larger effective population size or higher exponential growth rate (r). The values $\alpha_A = 92.97$ and $\alpha_{RS} = 83.48$ (which are independent of the substitution rate) show a difference between lineages; however, the confidence limits for θ and ρ overlap. Presented in Figure 5.7 and 5.8 are the *skyline plots* for lineages *A* and *RS* respectively. These plots show changes in the estimated effective population size over time.

Table 5.4. The maximum likelihood estimates on the demographic models of constant population size and exponential growth. Presented are analyses of each lineage, treated separately and combined. The Kolmorgorov-Smirnov (KS) goodness-of-fit test critical values at 0.05 are 0.16322 (*A* lineage), 0.07725 (*RS* lineage) and 0.06941 (*A* + *RS* lineage). The model parameters are $\theta = N^*\mu$ and ρ = the exponential growth rate/μ. The confidence limits (C.L.) of the latter parameter are given.

Lineage	Model	KS-test	ML value	θ	ρ	ρ C.L.
A	Constant	0.05753	33.671	0.3300		
	Exponential	0.05727	90.855	0.5832	159.41	134.53 — 182.51
RS	Constant	0.03990	57.712	0.1720		
	Exponential	0.03986	80.111	0.4894	170.54	130.96 — 207.60
A + *RS*	Constant	0.05452	-100.45	0.4795		
	Exponential	0.05423	-30.471	0.7680	118.01	103.51 — 130.71

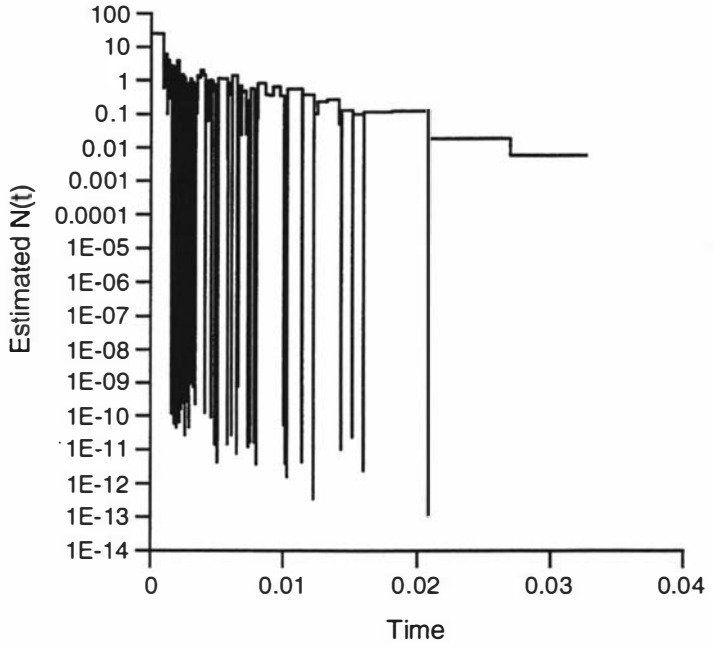


Figure 5.7. The *skyline plot* for the A lineage. The *y*-axis is logarithmic. Time on the *x*-axis is measured in units of substitutions. The *y*-axis represents the changes in the effective population size taken from the observed coalescent intervals in the genealogy. An increase in the number of coalescent intervals corresponds to an increasing population size. Maximum likelihood estimates for this lineage show that an exponential growth model better explains the history of this population (ρ confidence limits = 134.53 – 182.51).

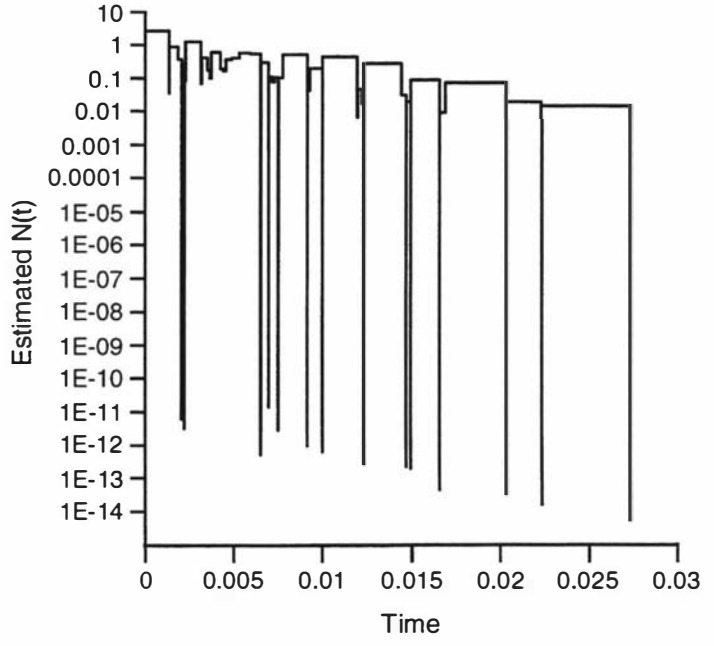


Figure 5.8. The *skyline plot* for the RS lineage. The *y*-axis is logarithmic. Maximum likelihood estimates for this lineage show that an exponential growth model better explains the history of this population (ρ confidence limits = 130.96-207.60).

5.4 Discussion

This chapter sought to characterise the level of diversity in Adélie penguin populations, and to describe and interpret present population structuring with respect to the glacial history of Antarctica. Adélie penguins have high levels of haplotypic (h = 0.997) and nucleotide (π = 0.038) diversity compared with other bird species. Baker and Marshall (1997) summarised the studies of mtDNA control region variation of bird species and showed that diversity ranged from the knot (*Calidris canutus*) h = 0.449 and π = 0.01, to the gray-crowned babbler (*Pomatostomus temporalis*) h = 0.973 and π = 0.037. A high level of genetic diversity has also been detected in a microsatellite DNA study of Adélie penguins; heterozygosity levels from seven loci ranged from 0.25-0.9 (Roeder et al., 2001). Populations with high h and π are reasoned to have been large for a long period of time, whereas a large h and small π indicate a population bottleneck followed by rapid growth (Grant and Bowen, 1998). Adélie penguins have a large population size, estimated at 15 million individuals. The high level of haplotypic diversity and heterozygosity suggest that this population size has been sustained over a reasonable period of time. However, labelling any h or π value as 'high' or 'low' involves arbitrary classification.

Felsenstein (1992) was the first to point out that more information about a population history could be inferred from a genealogy than from pairwise or segregating sites methods. Subsequently, Nee et al. (1995) proposed that the history of a population could be estimated by plotting time as a percentage of substitutions, against a logarithmic transformation of the number of lineages. The lineages-through-time plot can estimate parameters such as exponential growth or constant population size over the period covered by the coalescence. Genealogies with deep branches, indicating limited population growth for a long period of time, show a downward concavity on the graph. Conversely, many shallow branches indicate a rapidly expanding population history and show an upward concavity. Nee et al. (1995) constructed a lineages-through-time plot for humpback whale mtDNA control region data, and demonstrated that the sequences were drawn from a population that had maintained a constant size. Subsequent developments on these techniques produced mid-depth methods (Pybus et al., 1999) and maximum likelihood parameter estimates (Pybus et al., 2000). Most of these later developments focused on populations of rapidly evolving pathogens. In this study,

an estimate of the demographic history of each lineage was obtained using a maximum likelihood estimator. This likelihood approach revealed that each Adélie penguin lineage (*A* and *RS*) has undergone rapid expansion during the period covered by the coalescence events. A similar conclusion was not evident in the high levels of nucleotide diversity detected in each lineage.

5.4.1. Population Subdivision

In this study, the ϕ_{ST} values showed little significant genetic structuring between most pairs of populations. The Cape Wheatstone population provided almost all of the significant comparisons, possibly due to a combination of small sample size (n=10) and the high frequency of the *RS* lineage recorded at this site. Considering the high level of haplotypic diversity within each population, the sample sizes may have been too small to show any existing structure between populations using ϕ_{ST} statistics. A study of five microsatellite DNA loci from Adélie penguins also showed very little genetic structuring (the overall $F_{ST} = 0.0007$) (Roeder et al., 2001). However, despite the lack of subdivision indicated by the ϕ_{ST}, genetic structuring was evident from the distribution of two deeply split lineages (*A* and *RS*) around the Antarctic continent. The *A* lineage was found in all locations around Antarctica and the *RS* lineage was only recorded in the Ross Sea, indicating restricted long-distance movement.

There are several possible explanations for the geographic structure to mtDNA lineages, yet lack of structure to microsatellite alleles, among Adélie penguin populations. First, there is structure to these populations but insufficient time has elapsed for genetic differentiation of nuclear alleles. Since, Adélie penguins have a large population size, genetic drift and the formation of genetic structure will proceed relatively slowly. However, mtDNA has one-quarter the effective population size of a diploid locus and the geographic distribution of the *A* and *RS* lineages may represent the more rapid fixation of mitochondrial lineages. Second, there may be a male biased gene flow. Females who perpetuate a mtDNA lineage may remain at their natal colonies, whereas males, who do not transmit their mtDNA, disperse outside the Ross Sea. Nuclear alleles are continually being exchanged among populations via male migration. There is evidence of Adélie penguins returning to their natal areas (Ainley et al., 1983), but little information on

their dispersal and even less on the comparison of sexes. Last, the *RS* lineage may not be detected in locations other than the Ross Sea because it is present at such low frequencies. If there is a low level of migration among populations then there is a high probability of *RS* lineage extinction, shortly after arrival at a location. If the *RS* mitochondrial lineage is moving outside of the Ross Sea, it may take a long period of time to become established at colonies distant from the Ross Sea or a very large sample size may be necessary for it to be detected.

5.4.2. The Origin of Two Lineages

The two divergent mtDNA lineages (*A* and *RS*) present in the Ross Sea Adélie penguin populations could have arisen under several different conditions. First, in a population that has maintained a very large effective size over a long time period, there is a high probability of old (i.e. highly divergent) lineages having survived (Avise et al., 1984; Rogers and Harpending, 1992). Second, the lineages could be the result of recent contact between two populations having evolved in allopatry. Third, some form of "intrinsic isolation" between sympatric populations could have given rise to divergent lineages (Avise et al., 1987). For example, divergent mtDNA lineages have been attributed to the maintenance of female social groupings in macaque monkeys (Hoelzer et al., 1994), and host-specific races of females in the common cuckoo (Gibbs et al., 2000). A mixture of highly divergent mtDNA lineages in populations is most often attributed to secondary admixture between previously isolated populations (Avise, 2000). Ice-age refugia account for deeply split mtDNA lineages in a number of high-latitude species, especially in species from the Northern Hemisphere (Quinn, 1992; Taberlet et al., 1992; Monehan, 1994; Bensch and Hasselquist, 1999). The Quaternary environment (beginning *ca.* 3.5-1.8 Myr BP) has been characterised by frequent and rapid changes in the world's climate, including large-scale glacial cycles (Williams et al., 1998). Within this Epoch the Pleistocene showed very regular cyclic glacial and interglacial periods. Over the last 900 kyr these glacial cycles lasted nearly 100 kyr, the end of each cycle punctuated by a short warm period (Imrie et al., 1992). The expansion of the Antarctic ice sheets during the late Pleistocene most likely resulted in a discontinuous distribution of Adelie penguin populations around the Antarctic continent for up to 100 kyr at any one time. Large-scale disruptions to gene flow

between ice-age Adélie penguin populations would have persisted as recently as 15-20 kyr BP.

The entire Ross Sea embayment would have been uninhabitable to Adélie penguins due to the periodic expansion of the Ross Ice Shelf during the Pleistocene. The grounding line of the Ross Ice Shelf is currently about 900 km from the edge of the ice shelf at Ross Island. This ice shelf is fed mainly by ice streams falling ~2000m from the West Antarctic Ice Sheet through subglacial valleys, in some cases moving at a rate of 100 m/year (Joughin et al., 1999). This large shelf of ice is held in place by the buttressing effect of the Eastern coast as it swings "like a door", hinged north of Roosevelt Island, to fill half the Ross Sea (Denton et al., 1989). Lower sea levels experienced during glacial periods would encourage the advance of the Ross Ice Shelf grounding line, through thickening and subsequent backfilling. Moreover, the procession of the ice shelf further into the Ross Sea would eventually meet with additional glacial outlets along the Northern Scott coast (e.g. the David Glacier) and the Borchgrevink coast, increasing the overall input on the Eastern edge. At the last glacial maximum the grounding line was as far north as Coulman Island and the shelf front stretched from Possession Island across to the coast north of Roosevelt Island (Denton et al., 1989). At that time Ross Island would have been almost 900 km inland and surrounded by a huge grounded ice sheet.

At the time of the last glacial maximum, ice-free areas would have been rare, if present at all, in Antarctica (Colhoun et al., 1992). The isolation of Antarctica from other major landmasses would have limited the ability of Adélie penguins to simply move north, as the southern latitudes became uninhabitable. Moreover, the expansion of the sea-ice would have meant there was a long distance to travel to feed in the sea when an ice-free area was occupied on the continent. Benthic sedimentation of ^{13}C-rich organic material from Antarctic sea-ice algae has been used to indicate the extent of the sea-ice during the last glacial maximum (Gibson et al., 1999). Studies have shown that the summer sea-ice may have extended as far as the winter sea-ice now reaches (Cooke and Hays, 1982; Gersonde and Zielinski, 2000). Ice-free areas would have been present on some of the Peri- or Subantarctic islands, or if ice-areas were present on the continent they may have been surrounded by open waters formed by polynya.

The cline of the *RS* and *A* lineage frequencies in the Ross Sea suggests a refugia for the *RS* may have been north of the Ross Sea, and that after deglaciation the *RS* lineage moved down the coast as the Ross Ice Shelf retreated. The candidate ice-free areas (i.e. refugia) are Cape Adare (71°17'S, 182°24'E) and the Balleny Islands (66°55'S, 163°20'E). Cape Adare currently has the largest population of Adélie penguins, and the East Antarctic Ice Sheet has never overrun the surrounding mountains in the last 6-8 Myr (Armienti and Baroni, 1999). The *A* and *RS* lineage have many shallow branching events, typical of a population having undergone rapid expansion. As the Ross Ice Shelf retreated, the population from which *RS* originated may have moved southwards to reach Ross Island; while the *A* lineage could have rapidly expanded around Antarctica to occupy the majority of the current breeding locations. The centre-of-origin of the *A* lineage is more difficult to determine, as there is no obvious indicative cline in frequencies outside the Ross Sea. Interestingly, the population history of emperor penguins may contrast with that of the Adélie penguins. Emperors are a high-latitude Antarctic penguin species and make their nests on ice, travelling over 200 km across sea ice to sites (William, 1995). Not being bound to ice-free areas, this species would simply have moved with the advancing ice during the Pleistocene glacial cycles. A survey of mtDNA variation in emperors would not be expected to show such a distinctive two-lineage pattern.

The deeply split lineages and history of rapid population growth are consistent with the influence of glacial ice expansion and contraction through the late Pleistocene and the Holocene. Adélie penguins had limited opportunity to breed on the Antarctic mainland during the last 100 kyr of the Pleistocene and were separated into two allopatric populations. The effects of the last ice-age on Adélie penguin population size is more difficult to determine. A bottleneck signal (i.e. low nucleotide diversity) is not obvious from the summary statistics. The multitude of branching events in the mtDNA genealogy as a result of rapid population expansion, in addition to the length of time since the last glacial maximum, may obscure any record of a severe population bottleneck. Populations in ice-age refugia most likely had a smaller effective population size than present populations. They would have remained under these conditions for a long period of time and as a consequence mitochondrial lineages sorted into two types (*A* and *RS*). However, one parameter missing from this discussion is an accurate estimate of the

substitution rate (k). An estimated evolutionary rate would enable a date to be assigned to the split of the two lineages. This date would give a more precise link to the glacial cycles and their effect on population structure and genetic diversity. This issue will be further explored in the final chapter in which the evolving lineages are measured.

5.5 References

Ainley, D.G. 2001. The Adélie Penguin: Bellwether of Climate Change. Columbia University Press, New York (In press).

Ainley, D.G., LaReseche, R.E., and Sladen, W.J. 1983. Breeding Biology of Adélie Penguins. University of California Press, Berkeley.

Armienti, P., and Baroni, C. 1999. Cenozoic climate change in Antarctica recorded by volcanic activity and landscape evolution. Geology **27:** 617-620.

Avise, J.C. 1989. Gene trees and organismal histories: a phylogenetic approach to population biology. Evolution **43:** 1192-1208.

Avise, J.C. 1992. Molecular population structure and the biogeographic history of a regional fauna: a case history with lessons for conservation biology. Oikos **63:** 62-76.

Avise, J.C. 2000. Phylogeography, the history and formation of species. Harvard University Press, Cambridge, Massachusetts.

Avise, J.C., Arnold, J., Ball, R.M., Bermingham, E., Lamb, T., Neigel, J.E., Reeb, C.A., and Saunders, N.C. 1987. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. Annual Review of Ecology and Systematics **18:** 489-522.

Avise, J.C., Neigel, J.E., and Arnold, J. 1984. Demographic influences on mitochondrial DNA lineage survivorship in animal populations. J. Mol. Evol. **20:** 99-105.

Avise, J.C., and Walker, D. 1997. Pleistocene phylogeographic effects on avian populations and the speciation process. Proceedings of the Royal Society London Series B **265:** 457-463.

Baker, A.J., Daugherty, C.H., Colbourne, R., and McLennan, J.L. 1995. Flightless brown kiwis of New Zealand possess extremely subdivided population structure and cryptic species like small mammals. Proc. Natl. Acad. Sci. USA **92:** 8254-8258.

Baker, A.J., and Marshall, H.D. 1997. Mitochondrial control region sequences as tools for understanding evolution. *In* Avian Molecular Evolution and Systematics. *Edited by* D. P. Mindell. Academic Press, San Diego. pp. 51-82.

Baker, C.S., Slade, R.W., Bannister, J.L., Abernethy, R.B., Weinrich, M.T., Lien, J., Urban, J., Corkeron, P., Calmabokidis, J., Vasquez, O., and Palumbi, S.R. 1994.

Hierarchical structure of mitochondrial DNA gene flow among humpback whales *Megaptera novaeagliae* worldwide. Mol. Ecol. **3:** 313-327.

Ball Jr, R.M., Freeman, S., James, F.C., Bermingham, E., and Avise, J.C. 1988. Phylogeographic population structure of red-winged blackbirds assessed by mitochondrial DNA. Proc. Natl. Acad. Sci. USA **85:** 1558-1562.

Bandelt, H.J.-., and Dress, A.W.M. 1992. Split decomposition: a new and useful approach to phylogenetic analysis of distance data. Mol. Phylogenet. Evol. **1:** 242-252.

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

Bensch, S., and Hasselquist, D. 1999. Phylogeographic population structure of great reed warblers: an analysis of mtDNA control sequences. Biol. J. Linn. Soc. **66:** 171-185.

Colhoun, E.A., Mabin, M.C.G., Adamson, D.A., and Kirk, R.M. 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 Hays, J.D. 1982. Estimates of Antarctic ocean seasonal sea-ice cover during glacial intervals. In Antarctic Geoscience: Symposium on Antarctic Geology and Geophysics. University of Wisconsin Press, Madison, Wisconsin.

Denton, G.H., Bockheim, J.G., Wilson, S.C., and Stuiver, M. 1989. Late Wisconsin and early Holocene glacial history, inner Ross embayment, Antarctica. Quat. Res. **31:** 151-182.

Edwards, S.V. 1993. Long-distance gene flow in a cooperatively breeder detected in genealogies of mitochondrial DNA sequences. Proceedings of the Royal Society of London, Biological Sciences B **252:** 177-185.

Excoffier, L., Smouse, P.E., and Quattro, J.M. 1992. Analysis of molecular variance inferred from metric distance among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics **131:** 479-491.

Falk, K., and Møller, S. 1997. Breeding ecology of the Fulmar *Fulmarus glacialis* and the Kittiwake *Rissa tridactyla* in high-arctic northeastern Greenland, 1993. Ibis **139:** 270-281.

Felsenstein, J. 1992. Estimating effective population size from samples of sequences: inefficiency of pairwise and segregating sites as compared to phylogenetic estimates. Genetics Research Cambridge **59:** 139-147.

Frazer, W.R., Trivelpiece, W.Z., Ainley, D.G., and Trivelpiece, S.G. 1992. Increases in Antarctic penguin populations: reduced competition with whales or loss of sea ice due to environmental warming. Polar Biology **11:** 525-531.

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

Gibbs, H.L., Sorenson, M.D., Marchetti, K., Brookes, M.d.L., Davis, N.B., and Nakamura, H. 2000. Genetic evidence for female host-specific races of the common cuckoo. Nature **407:** 183-185.

Gibson, J.A.E., Trull, T., Nichols, P.D., Summons, R.E., and McMinn, A. 1999. Sedimentation of ^{13}C-rich organic matter from Antarctic sea-ice algae: a potential indicator of past sea-ice extent. Geology **27:** 331-334.

Gordon, A.L., and Comiso, J.C. 1988. Polynyas in the Southern Oceans. Scientific American **June:** 70-77.

Grant, W.A., and Bowen, B.W. 1998. Shallow population histories in deep evolutionary lineages of marine fishes: insights from sardines and anchovies and lessons for conservation. The Journal of Heredity **89:** 415-426.

Hewitt, G. 2000. The genetic legacy of the Quaternary ice ages. Nature **405:** 907-913.

Hoelzer, G.A., Dittus, W.P.J., Askley, M.V., and Melnick, D.J. 1994. The local distribution of highly divergent mitochondrial DNA haplotypes in toque macaques (*Macaca sinica*) at Polonnaruwa, Sri Lanka. Mol. Ecol. **3:** 451-458.

Hunter, F.M., and Davis, L.S. 1998. Female Adélie penguins acquire nest material from extrapair males after engaging in extrapair copulations. The Auk **115:** 526-528.

Huson, D.H. 1997. SplitsTree: A Program for Analyzing and Visualizing Evolutionary Data. University of Bielefeld, Germany.

Imbrie, J., Boyle, E.A., Clemans, S.C., Duffy, A., Howard, W.R., Kukla, G., Kutzback, J., Martinson, D.G., McIntyre, A., Mix, A.C., Molfino, B., Morley, J.J., Peterson, L.C., Pisias, N.G., Prell, W.L., Raymo, M.E., Shackleton, N.J., and Toggweiler, J.R. 1992. On the structure and origin of major glaciation cycles 1. linear responses to Milankovitch forcing. Paleooceanography **7:** 701-738.

Joughin, I., Gray, L., Bindschadler, R., Price, S., Morse, D., Hulbe, C., Mattar, K., and Werner, C. 1999. Tributaries of West Antarctic ice streams revealed by RADARSAT interferometry. Science **286:** 283-286.

Kruskal, J.B. 1956. On the shortest spanning subtree of the graph and the travelling salesman problem. Proc. Amer. Math. Soc. **7:** 48-57.

Kumar, S., Tamura, K., Jakobsen, I., and Nei, M. 2001. MEGA 2: Molecular Evolution Genetics Analysis software. Bioinformatics ***Submitted.***

Mantel, N. 1967. The detection of disease clustering and a generalised regression approach. Cancer Research **27:** 209-220.

Moneham, T.M. 1994. Molecular genetic analysis of Adélie penguin populations, Ross Island, Antarctica. M.Sc. thesis, University of Auckland, Auckland.

Nee, S., Holmes, E.C., Rambaut, A., and Harvey, P.H. 1995. Inferring population history from molecular phylogenies. Phil. Trans. Roy. Soc. Lond. B **349:** 25-31.

Nei, M. 1972. Genetic distance between populations. American Naturalist **106:** 283-292.

Nei, M. 1987. Molecular Evolutionary Genetics. Columbia University Press, Columbia, USA.

Pybus, O.G., Holmes, E.C., and Harvey, P.H. 1999. The mid-depth method and HIV-1: a practical approach for testing hypotheses of viral epidemic history. Mol. Biol. Evol. **16:** 953-959.

Pybus, O.G., Rambaut, A., and Harvey, P.H. 2000. An integrated framework for the inference of viral population history from reconstructed genealogies. Genetics **155:** 1429-1437.

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.

Radok, U. 1985. The Antarctic ice. Scientific American **253:** 98-105.

Rice, W.R. 1989. Analyzing tables of statistical tests. Evolution **43:** 223-225.

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

Roeder, A.D., Marshall, R.K., Mitchelson, A.J., Visagathilagar, T., Ritchie, P.A., Love, D.R., Pakai, T., McPartlan, H., Murray, N.D., Robinson, N.A., Kerry, K.R., and Lambert, D.M. 2001. Extraordinary genetic homogeneity among Adélie penguin colonies around Antarctica. Mol. Ecol. **In Press.**

Rogers, A.R., and Harpending, H.C. 1992. Population growth makes waves in the distribution of pairwise genetic differences. Mol. Biol. Evol. **9:** 552-569.

Saitou, H., and Nei, M. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4:** 406-425.

Sambrook, J., Fritsch, E.F., and Maniatus, T. 1989. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, New York.

Seutin, G., White, B.N., and Boag, P.T. 1991. Preservation of avian blood and tissue samples for DNA analyses. Canadian Journal of Zoology **69:** 82-90.

Sladen, W.J.L. 1953. The Adélie Penguin. Nature **171:** 952-955.

Sullivan, J., Holsinger, K.E., and Simon, C. 1995. Among-site rate variation and phylogenetic analysis of 12S rRNA in sigmodontine rodents. Mol. Biol. Evol. **11:** 261-277.

Swofford, D.L. 2000. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Sinauer Associates, Sunderland, Massachusetts.

Taberlet, P., Meyer, A., and Bovet, J. 1992. Unusual mitochondrial DNA polymorphism in two local populations of blue tit *Parus caeruleus*. Mol. Ecol. **1:** 27-36.

Tajima, F. 1989. Statistical methods for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **123:** 585-595.

Taylor, R.H., Wilson, P.R., and Thomas, B.W. 1990. Status and trends of Adélie penguin populations in the Ross Sea region. Polar Record **26:** 293-304.

Templeton, A.R., Routman, E., and Phillips, C.A. 1995. Separating population structure from population history: A cladistic analysis of the geographical distribution of mitochondrial DNA haplotypes in the tiger salamander, *Ambystoma tigrinum*. Genetics **140:** 767-782.

Wenink, P.W., Baker, A.J., Rösner, H.-U., and Tilanus, M.G.J. 1996. Global mitochondrial DNA phylogeography of Holarctic breeding dunlins (*Calidris alpina*). Evolution **50:** 318-330.

Williams, M.A.J., Dunkerley, D., DeDeckker, P., Kershaw, P., and Chappell, M. 1989. Quaternary Environments. Arnold, London.

Williams, T.D. 1995. The Penguins: Spheniscidae. Oxford University Press, Oxford.

Wilson, A.C., Cann, R.L., Carr, S.M., George, M., Gyllensten, U.B., Helm-Bychowski, K.M., Higuchi, R.G., Palumbi, S.R., Prager, E.M., Sage, R.D., and Stoneking, M. 1985. Mitochondrial DNA and two perspectives on evolutionary genetics. Biol. J. Linn. Soc. **26:** 375-400.

Chapter Six

Ancient DNA from Cryopreserved Adélie Penguins in Antarctica

6.1 Introduction

The aim of this thesis research was to estimate a rate of evolution for the mtDNA control region, utilising both modern and ancient DNA sequences. The previous chapter dealt with the genetic parameters of the hypervariable region I (HVR-1) within and among modern populations of Adélie penguins. In order to estimate a rate of evolution, DNA sequences need to be serially sampled through time. Presented in this chapter are HVR-1 sequences of DNA extracted from a large number of ancient subfossil Adélie penguins. Many of these penguins died thousands of years ago, but their DNA still remains intact and amenable to genetic analyses.

6.1.1. The Stability of DNA

DNA has limited stability and will quickly degrade without the constant attention of cell repair mechanisms. However, small amounts of DNA can remain intact, if

protected under the right conditions, for thousands and even tens of thousands of years. The ability to retrieve old or ancient DNA with the polymerase chain reaction (PCR) from organisms long after death, has given rise to a new field of research: molecular archaeology (Pääbo et al., 1989). The recovery of ancient DNA from preserved materials has allowed genetic analyses of, for example, extinct Neanderthal species (Krings et al., 1997), Holocene pocket gopher populations (Hadly et al., 1998) and even pedigrees from the remains of a Tsar Nicholas II and his family (Ivanov et al., 1996). The sensitivity of PCR to amplify from a few remaining templates has allowed ancient DNA retrieval, however, this sensitivity has also caused a series of contamination problems (Handt et al., 1994; Austin et al., 1997).

Ancient DNA is highly modified, fragmented and surrounded by compounds which are inhibitory to enzymatic manipulation (Pääbo, 1989; Pääbo, 1990). Hydrolytic attack and oxidative damage are the two main modes of spontaneous DNA decay (Lindahl, 1993) (Figure 6.1). The most common target of hydrolytic disassociation is the liable *N*-glycosidic bond (base-sugar) at each purine base, a process known as depurination. For example, in one living human cell up to 10,000 spontaneous depurination events are predicted to occur each day (Lindahl and Nyberg, 1972). DNA fragmentation most likely follows depurination, as breakage can easily occur at 3'-phosphodiester bonds on the DNA backbone through β-elimination. Hydrolysis will also deaminate cytosine and adenine to form uracil and hypoxanthine, respectively. Oxidative damage to DNA is usually a direct result of ionising radiation, or free radicals created in water by irradiation (Lindahl, 1993). DNA can never escape damaging free radical $^{\bullet}OH$ species; even in its A-form (dehydrated) water molecules are still present in grooves along the double helix. Amplification by PCR can not proceed if DNA is heavily damaged (Höss et al., 1996), and if DNA is moderately damaged it can cause PCR to generate erroneous copies of the original template (Pääbo et al., 1990; Kamiya et al., 1994).

6.1.2. Factors Affecting the Instability of DNA

Cold temperatures, high ionic strength mediums, desiccating conditions and exclusion of oxygen have all been implicated in retarding the degradation of DNA's primary structure (Lindahl, 1993; Waite et al., 1997). Aspartic acid racemization

proceeds at a similar rate to depurination of DNA, and has been a useful indicator for the preservation of DNA in fossil specimens (Bada et al., 1994; Poinar et al., 1996). In an organism the functional levo (L)-isomeric form of an amino acid is dominant, the non-functional dextro (D)-isomer is actively discarded. However, after death transformations between the L- and D-forms proceed back and forth until both reach equilibrium. This process of amino acid racemization, like depurination, is strongly affected by temperature. Results from amino acid racemization and the recovery of ancient DNA suggest that the processes of nucleotide decomposition are significantly slower in a cold environment (Poinar et al., 1996).

The amount and type of oxidised DNA adducts in ancient specimens, has been assessed using mass spectrometry (Höss et al., 1996). The highest concentration of oxidised bases are from the thymine derivatives, 5-hydroxyhydantoin (5-OH-Hyd) and 5-hydroxy-5-methyl-hydantoin (5-OH-5-MeHyd), as well as 8-hydroxyguanine (8-OH-Gua). A hydantoin is a 5-member stable ring, formed through saturation and then condensation of the 6^{th} carbon atom from a pyrimidine ring, and in the case of 5-OH-5-Hyd the CH_3 group is also lost (Breimer and Lindahl, 1985). A consequence of this modification is the inability to pair with any other nucleotide, which means that polymerase enzymes will stall when posed with this adduct. Höss et al. (1996) found that specimens from cooler climates show the least oxidative damage and contained DNA that could be readily amplified by PCR. Poinar and Stankiewicz (1999) went on to show that a low D/L ratio of aspartic acid (<0.12) and alanine (<0.01), as well as, minimal levels of hydantoins (<1.72 nmol/mg), all correlate with the recovery of well preserved ancient DNA. Most importantly, these authors showed that ancient DNA is better preserved in a cold environment.

Studies on the preservation of DNA suggest that in a cold and dry environment like Antarctica, DNA should remain largely intact over long periods of time. In this chapter the quality of DNA found in cryopreserved subfossil bones is discussed. In the previous chapter sequencing results from modern Adélie penguin populations, showed the presence of two deeply split mtDNA lineages (*A* and *RS*) in the Ross Sea. A cline of the relative frequencies of these lineages along the coast suggested the invasion by two allopatric ice-age populations after the last deglaciation. Using

ancient DNA sequences from C^{14}-dated bones, the relative invasion times for these populations is determined.

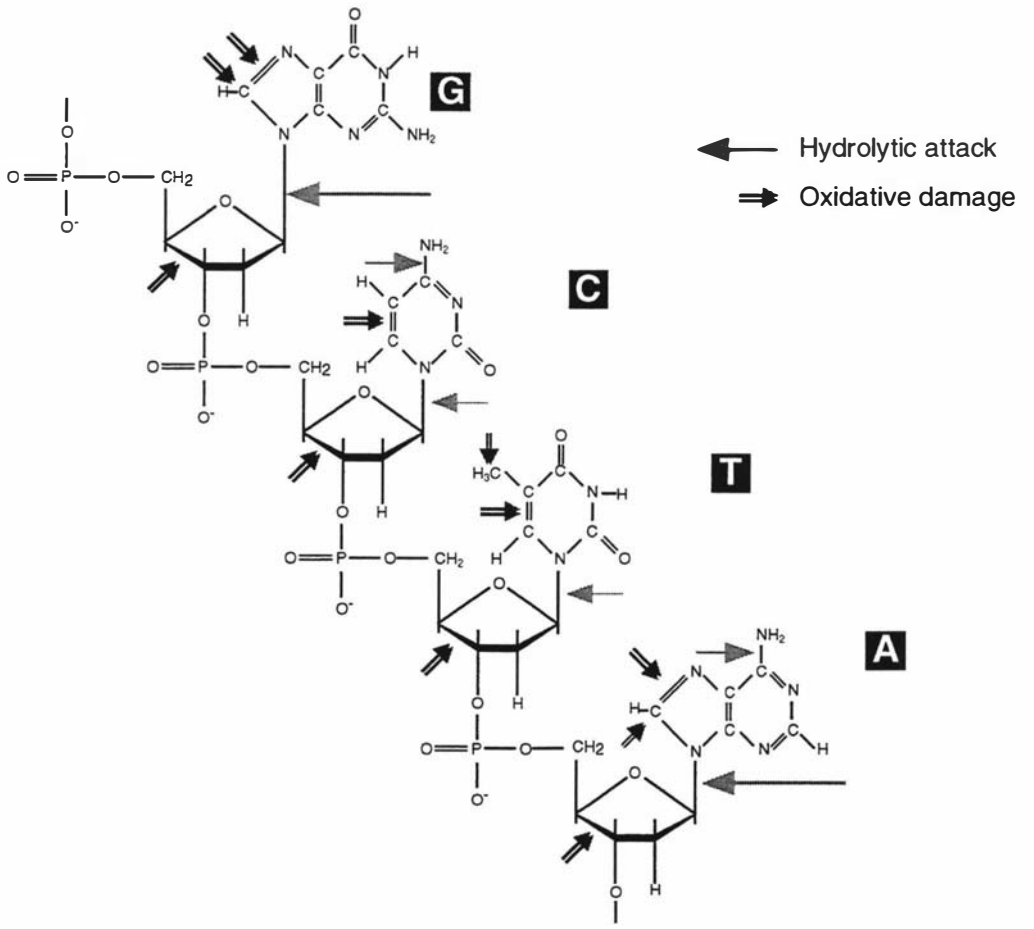


Figure 6.1. The degradation of a DNA molecule results from hydrolytic attack and oxidative damage. The single arrows indicated the bonds liable for hydrolytic disassociation and the double arrows indicate the sites susceptible to oxidative damage. G = guanine, C = cytosine, T = thymine, and A = adenine. Adapted from Lindahl (1993).

6.2 Materials and Methods

Subfossil Adélie penguin bones were collected from 16 locations along the coast of the Ross Sea (Figure 6.2). Samples were collected from abandoned nesting sites both in the vicinity of presently occupied rookeries and in relict rookeries. Subfossil bones were excavated using a stratigraphic method in which pits from 1-2 to 6 m^2 were excavated, layer by layer, from the top. This method allowed the identification and separation of individual remains within the same layer, excluded top layer contamination, and identified wedges of sediments reworked by periglacial processes.

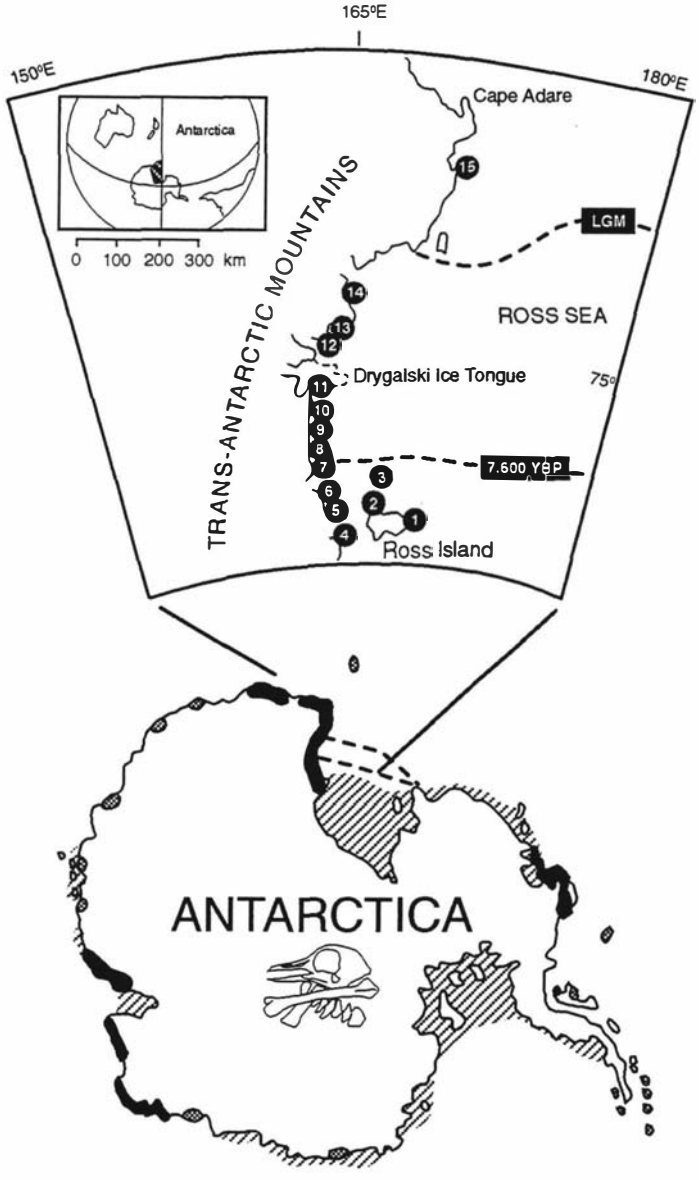


Figure 6.2. The distribution of abandoned Adélie penguin rookeries where subfossil bones were sampled (sfb = number of subfossil bones collected), with latitude and longitude coordinates: **1.** Cape Crozier (77°14'S, 166°28'E), sfb 54; **2.** Cape Bird (77°30'S, 162°10'E), sfb 39; **3.** Beaufort Island (76°56'S, 167°03'E), sfb 20; **4.** Spike Cape (77°18'S, 163°33'E), sfb 2; **5.** Dunlop Island (77°14'S, 163°28'E), sfb 22; **6.** Cape Roberts (77°02'S, 163°10'E) sfb 17; **7.** Cape Ross (76°43'S, 162°59'E), sfb 19; **8.** Depot Island (76° 42' S, 162° 57 E), sfb 3; **9.** Cape Day (76°14'S, 162°47'E), sfb 7; **10.** Cape Hickey (76°05'S, 162°38'E), sfb 5; **11.** Prior Island (75°41'S, 162°52'E), sfb 12; **11.** Cape Irizar (75°33'S, 162°57'E), sfb 9; **12.** Inexpressible Island (74°53'S, 163°45'E), sfb 37; **13.** Northern Foothills (74°45'S, 164°00'E), sfb 63; **13.** Gondawa Station (74°38'S, 164°12'E), sfb 4; **14.** Edmonson Point (74°19'S, 165°04'E), sfb 11; **15.** Cape Hallett (72°19'S, 170°12'E), sfb, 5. The location of the permanent ice at different times are indicated by dotted lines, LGM = Last Glacial Maximum. The black and grey indicate areas of high and low penguin density respectively.

Subfossil bone samples were collected and stored frozen, then returned to the laboratory where they were kept at –20°C. All DNA extractions and setting up of PCRs from subfossil material were conducted in a dedicated ancient DNA laboratory, which underwent regular decontamination with UV-irradiation and hypochloride treatment. Mock DNA extractions and blank PCRs were continually screened for contamination.

Approximately 0.5 g of each bone was powdered in a sterile 'coffee-grinder' and then decalcified in 20 mls of 500 mM EDTA pH 8.0 at room temperature overnight. The supernatant was removed after centrifugation and the sediment resupended in 5 ml of extraction buffer (10 mM Tris-HCl pH 8.0, 1 mM NaCl), 500 µl of 10% SDS, 30 µl of 200 mg/ml Dithiothreitol (DTT) and 50 µl of 50 mg/ml Proteinase-K (or Collagenase A). This solution was incubated overnight at 50°C. The samples were extracted twice with Tris-saturated phenol and once with chloroform:isoamyl (24:1). This extracted solution was then concentrated to 200 µl on a Vivaspin-30 (Viva Science, U.K.) membrane. This column is a mechanical sieve, which retains DNA (above 30,000 MW) while small molecules and chemicals passed through. One hundred microlitres of the sample was then purified using QIAGEN DNA mini kit and stored at 4°C. In this final step DNA was bound to silica particles (using a salt bridge) utilising the free negatively charged oxygen on the phosphate backbone, and foreign compounds were then washed off the column with buffered ethanol.

The HVR-1 was amplified using the primers specific to Adélie penguins (reference numbers correspond to the sequence reported in Chapter Three) AH530 (5'-CTGATTTCACGTGAGGAGACCG-3'), AH432 (5'-GTGTTCAAGCTCTGCCGTACC-3'), AL305 (5'-GGACCAGCTCCTAATCCCTTCG-3'), AH271 (5'-GCTGGTCCTTGTACCATGGAC-3'), AL93 (5'-CATTTAATGTACGTACTAGGA C-3'), and L-tRNA^{Glu} (5'-CCCGCTTGGCTTYTCTCCAAGGTC-3'). Amplifications were carried out in 25 µl volumes containing 1µl of extracted DNA, 1 unit of AmpliTaq (PE Biosystems), 1.5 mM $MgCl_2$, 2 mg/ml bovine albumin serum (BSA), 0.4 pmol/µl of each primer and 200 µM of each dNTP. Samples were amplified in a separate laboratory using a Hybaid Thermal Cycler at 94°C/10 sec, 50°C/10 sec, and 72°C/25 sec for 35-40 cycles. PCR products were purified with the QIAquick PCR purification kit before direct sequencing using the Big Dye Terminator

sequencing kit (PE Biosystems) and analysed on a 377A automated sequencer (PE Biosystems).

The age of selected bones was determined using an Accelerator Mass Spectrometry (AMS) to measure the relative abundance of the cosmogenic radioisotope ^{14}C and its daughter ^{14}N formed by β-emission. The analysis was performed at the Institute of Nuclear and Geological Sciences in Upper Hutt, NZ. A half-life convention of 5730 years for ^{14}C was used. In addition to these direct ^{14}C-dated bones, a total of 99 radiocarbon ages were used to assign dates to bones by dating strata from which they were isolated (see Appendix C).

6.3 Results

A total of 329 subfossil Adélie penguin bones were collected during the course of this study. DNA extractions were attempted from 110 of these samples, 95 yielded DNA from which 80 unambiguous DNA sequences were obtained. Mock extractions and PCR blanks were continually screened and no evidence of contaminant was seen. A portion of the DNA extractions were conducted twice. All PCR targets were amplified twice, on separate occasions and each amplicon sequenced in both directions. Six of the 80 positive DNA samples were extracted, amplified and sequenced in an independent laboratory at the University of Auckland. All controls yielded results consistent with authentic ancient DNA.

The level of sequencing error that results from *Taq* polymerase inserting the wrong nucleotide across from a damaged nitrogenous base was estimated with reference to a control sequence (the *rns* gene). Sequencing errors can occur when, for example, deaminated cytocine forms uracil, which is, in turn, mis-read by *Taq* polymerase as thymine and consequently induces an apparent C to T transition. Approximately, 313 b.p. of the *rns* gene was sequenced in ancient material (n=18). Modern Adélie penguins showed no intraspecific variation for this sequence (n=21). All ancient DNA sequences from the *rns* gene were also identical to modern sequences. Hence the background damage induced error rate in this study was assumed to be low.

The DNA extracted from these ancient bones exhibited an autofluoresence (a blue-green colour) in a ≤600 b.p. range when exposed to UV-light in an agarose gel. This autofluorescence has been identified as degraded collagen I (~94 kDa), the major organic component of bone (Scholz et al., 1998). Furthermore, these authors showed that a gelatin solution (collagen-related glutins) inhibited PCR at concentrations greater than 500 ng/100 µl. Collagen was removed either during DNA extraction by using Collagenase A (a collagenase suitable for osteoblasts) as a substitute for Proteinase-K, or during post-DNA extraction clean-up with the silica columns.

All DNA extraction were a dark brown colour. This colouration may be due to the presence of melanoidins (brown nitrogenous polymers and copolymers, i.e. sugars that become covalently attached to proteins), the final products of Maillard reactions, and fulvic and humic acids (Tuross, 1994). Condensation of reactive sugars and amino compounds result in products like humus (soil-derived degradation products) and the browning is a result of an increase in the molecular weight of proteins (Namiki, 1988). During the extraction procedures, both buffer saturated phenol and phenol:chloroform:isoamyl were tested. It became apparent that saturated phenol mixed well with the digested solution and drawing more dark brown matter away from the aqueous layer. Phenol:chloroform:isoamyl (25:24:1) did not tend to mix well and remained clear even after 30 mins, hence never cleaning the aqueous solution. The contaminating matter in these bones was presumed to be resisting the hydrophobic property of chloroform. These brown compounds were persistent during all stages of the purification procedure, even after washing the DNA across a silica column. These compounds carried through to the purified DNA and inhibited *Taq* polymerase until diluted out at 1:100. The addition of BSA to a final concentration of between 1-5 mg/ml was essential for the function of any enzymatic reactions (Figure 6.3).

The ^{14}C-dated Adélie penguin bones and their calibrated ages are listed on Table 6.1. The information used to assign ^{14}C-dates for each ancient DNA sequence is presented in Appendix C. The dates assigned to each sample are listed in Table 6.2, and ranged from 275 to 6082 yr BP. The reservoir effect was accounted for by subtracting the age average of ^{14}C dated samples collected before the 1950-thermonuclear bomb pulse. Organic material from Antarctica contains 'old' carbon

which has entered the food chain from upwellings after being trapped in deep ocean for hundreds of years, or from carbon released from glacial meltwaters (Berkman et al., 1998). This long residence time of carbon before entering the food chain means the radiocarbon age of a sample is much older than the true age. Petri and Baroni (1997) have quantified the ΔReservoir (ΔR) for Antarctic penguins at 688 ± 55 yr. Since the error varies with time the calibration program Calib 4.2 (Stuiver et al., 2000) was used to estimate calibrated ages.

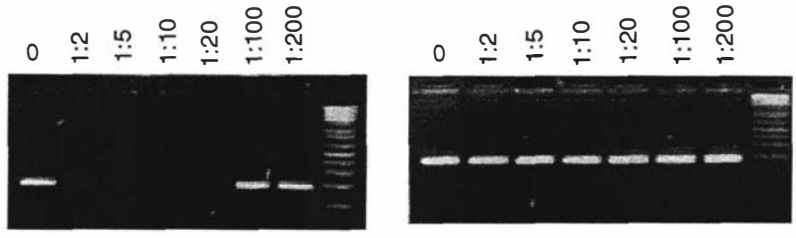


Figure 6.3. The inhibition of PCR by ancient DNA extracts. Each PCR contained a modern DNA template and different concentrations of the ancient DNA extract (0 to 1:100 dilution). The gel on the left shows how ancient DNA extracts strongly inhibited PCR. The gel on the right shows how high concentrations of BSA prevented the co-extracted inhibitors from affecting the performance of *Taq* polymerase.

The most remarkable result obtained during this study was the amplification of up to 1600 b.p. of the mitochondrial control region from bones over ~450 years old. Thirty five percent of the younger bones (under 2000 yr old) consistently amplified a 663-1042 b.p. sequence. Forty five percent of bones over 2000 yr old consistently allowed amplification of a 390 b.p. sequence (Figure 6.4, Table 6.2). All of these results were also independent verified at a laboratory at the University of Auckland. In total, 82% of sub-fossil bones could be sequenced for this shorter fragment. Histological sectioning and X-rays of ^{14}C-dated bones revealed generally well-preserved internal structures.

Table 6.1. Subfossil bones for which direct radiocarbon ages were obtained. All material was AMS dated at the Institute of Geological and Nuclear Sciences (NZA). The calibrated age (Cal. Age) is the radiocarbon age adjusted for the reservoir effect with a ΔR of 688 ± 55 using the program Cal. 4.2.

Extraction	Location	Carbon Dating Number	Radiocarbon Age (yr BP ± error)	Cal. Age (yr BP)
PE18	Cape Crozier	NZA-8818	1475 ± 59	440
PE24	Cape Crozier	NZA-10306	1380 ±60	310
PE131	Cape Crozier	NZA-8819	1503 ± 57	461
PE132	Cape Crozier	NZA-8820	1616 ±62	523
PE37	Cape Crozier	NZA-8821	1564 ± 55	498
PE48	Cape Bird	NZA-9182	1534 ± 55	481
PE42	Beaufort Island	NZA-12287	1313 ± 55	275
PE64	Inexpressible Island	NZA-12286	6358 ± 55	6082

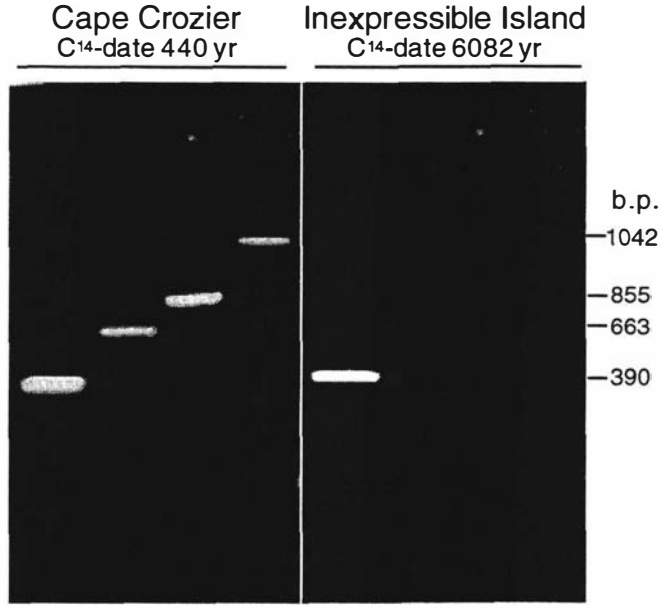


Figure 6.4. The length of amplified control region sequences from subfossil bone samples of different ages. These PCRs used the L-tRNA^{Glu} primer and additional primers designed from the control region sequence presented in Chapter Three. Cape Crozier = PE18 and Inexpressible Island = PE64.

Table 6.2. The age assigned of each subfossil bone that ancient DNA was extracted from, the length of sequence determined, their lineage (*A* or *RS*) and the maximum size fragment (in b.p.) obtaining using PCR. The asterisk (*) indicates bones that direct radiocarbon dates were obtained. All other dates were assigned to samples using the radiocarbon dating information presented in Appendix C.

	Location	Extraction	Cal. yr BP	Sequence	Lineage	PCR
1.	Adelie Cove	PE110	5418	179	*RS*	179
2.	Beaufort Island	PE44	275	390	*RS*	663
3.	Beaufort Island	PE43	275	390	*A*	663
4.	Beaufort Island	PE42*	275	390	*A*	663
5.	Cape Bird	PE48*	481	390	*A*	663
6.	Cape Bird	PE33	481	390	*A*	663
7.	Cape Bird	PE32	481	390	*A*	663
8.	Cape Bird	PE31	481	390	*A*	663
9.	Cape Bird	PE29	481	390	*RS*	663
10.	Cape Bird	PE10	481	390	*A*	663
11.	Cape Bird	PE07	481	390	*A*	663
12.	Cape Crozier	PE132*	523	390	*A*	663
13.	Cape Crozier	PE131*	461	390	*A*	663
14.	Cape Crozier	PE38	446	390	*A*	663
15.	Cape Crozier	PE37*	498	390	*A*	663
16.	Cape Crozier	PE30	446	390	*A*	663
17.	Cape Crozier	PE25	446	390	*A*	663
18.	Cape Crozier	PE24*	310	390	*RS*	663
19.	Cape Crozier	PE23	446	390	*A*	663
20.	Cape Crozier	PE22	446	390	*A*	1042
21.	Cape Crozier	PE21	446	390	*RS*	1042
22.	Cape Crozier	PE20	446	390	*A*	663
23.	Cape Crozier	PE18*	440	390	*A*	1687
24.	Cape Crozier	PE15	420	390	*RS*	1042
25.	Cape Crozier	PE13	523	390	*A*	663
26.	Cape Day	PE130	3364	390	*RS*	390
27	Cape Hallett	PE126	607	390	*A*	390
28.	Cape Hallett	PE125	500	390	*RS*	390
29.	Cape Hickey	PE89	2965	390	*A*	390
30.	Cape Hickey	PE88	2965	390	*A*	390
31.	Cape Hickey	PE87	2965	390	*A*	390
32.	Cape Hickey	PE74	2513	390	*RS*	390
33.	Cape Hickey	PE73	2513	390	*A*	390
34.	Cape Roberts	PE112	3091	390	*A*	390
35.	Cape Ross	PE85	3514	390	*A*	390
36.	Cape Ross	PE84	3514	390	*RS*	390
37.	Cape Ross	PE83	3514	390	*A*	390

Table 6.2. (continued)

	Location	Extraction	Cal. yr BP	Sequence	Lineage	PCR
38.	Cape Ross	PE82	3514	390	*A*	390
39.	Cape Ross	PE81	3514	390	*A*	390
40.	Cape Ross	PE79	3514	390	*A*	390
41.	Cape Ross	PE78	3514	270	*A*	270
42.	Cape Ross	PE77	3514	390	*A*	390
43.	Cape Ross	PE75	3514	120	*A*	120
44.	Depot Island	PE122	2911	390	*A*	390
45.	Dunlop Island	PE121	5997	390	*RS*	390
46.	Dunlop Island	PE120	5997	390	*A*	390
47.	Dunlop Island	PE119	2662	390	*A*	390
48.	Dunlop Island	PE118	3530	245	*A*	245
49.	Edmonson Point	PE135	1086	390	*A*	390
50.	Historic sample	PE39	88	390	*A*	390
51	Inexpressible Island	PE103	2328	390	*A*	390
52.	Inexpressible Island	PE104	2328	390	*RS*	390
53.	Inexpressible Island	PE107	5706	245	*RS*	245
54.	Inexpressible Island	PE106b	5706	120	*A*	120
55.	Inexpressible Island	PE55	2328	390	*A*	390
56.	Inexpressible Island	PE56	2328	270	*A*	270
57.	Inexpressible Island	PE64*	6082	390	*A*	390
58.	Inexpressible Island	PE65	6082	390	*A*	390
59.	Inexpressible Island	PE66	6082	390	*A*	390
60.	Inexpressible Island	PE69	6082	390	*A*	390
61.	Inexpressible Island	PE70	6082	390	*RS*	390
62.	Inexpressible Island	PE71	6082	290	*RS*	290
63.	Inexpressible Island	PE72	6082	120	*RS*	120
64.	Inexpressible Island	PE63	6040	260	*RS*	260
65.	Inexpressible Island	PE51	2328	390	*RS*	390
66.	Inexpressible Island	PE106a	2746	120	*RS*	120
67.	Inexpressible Island	PE67	6082	260	*A*	260
68.	Northern Foothills	PE123	4270	390	*A*	390
69.	Northern Foothills	PE116	4543	390	*A*	390
70.	Northern Foothills	PE115	4543	390	*RS*	390
71.	Northern Foothills	PE113	3970	390	*A*	390
72.	Prior Island	PE101	1057	390	*A*	663
73.	Prior Island	PE100	949	390	*A*	663
74.	Prior Island	PE99	750	390	*RS*	663
75.	Prior Island	PE98	1258	390	*A*	663
76.	Prior Island	PE97	750	390	*RS*	663
77.	Prior Island	PE95	3888	120	*A*	120
78.	Prior Island	PE93	3888	390	*A*	390
79.	Prior Island	PE91	3888	120	*A*	120
80.	Prior Island	PE90	3888	270	*RS*	270

For the majority of ancient DNA extractions a 352 b.p. region of the HVR-1 was obtained (primer sequences were excluded from the 390 b.p. fragment). However, for a small number of samples, sequences of only 180 b.p. or 270 b.p. were able to be amplified. The HVR-1 region of ancient DNA sequences possesses asymmetric base frequencies (π_A = 0.30, π_T = 0.31, π_C = 0.20, and π_G = 0.19) and strong bias towards transitional mutations (frequencies C↔T = 0.65, G↔A = 0.35 and transversions < 0.01). These results are very similar to the pattern detected in the modern samples.

A neighbor-joining tree was constructed for the 80 ancient DNA sequences using PAUP 4.0* (Swofford, 2000) and is presented in Figure 6.5. Overall there was 4.5% sequence difference among the 80 ancient DNA sequences, with 55 sequences from the *A* lineage (69%) and 25 from the *RS* lineage (31%). Thirteen samples dated to between 5706-6082 yr BP from Dunlop and Inexpressible islands are from the *A* (n=7) and *RS* (n=6) lineages (Table 6.2). These indicate that both lineages were present on the Ross Sea coast at similar frequencies at least as six thousand years ago. The average uncorrected sequence divergence was calculated for sequences that were older than 2000 years (mean age = 4,200 yr BP). Within each lineage, including each haplotype only once, is d_A = 0.0227 ± 0.0027 (n = 23) and d_{RS} = 0.0268 ± 0.0052 (n = 14); between these two lineages $d_{A/RS}$ = 0.0746 ± 0.0108. Standard errors were calculated in MEGA ver 2.0 (Kumar et al., 2001) using a bootstrap procedure with 500 replicates. To estimate the net distance between the two lineages, adjustment was made for within lineage polymorphism $d_{adjusted} = d_{A/RS} - 0.5 (d_A + d_{RS})) = 0.06114 \pm 0.0101$.

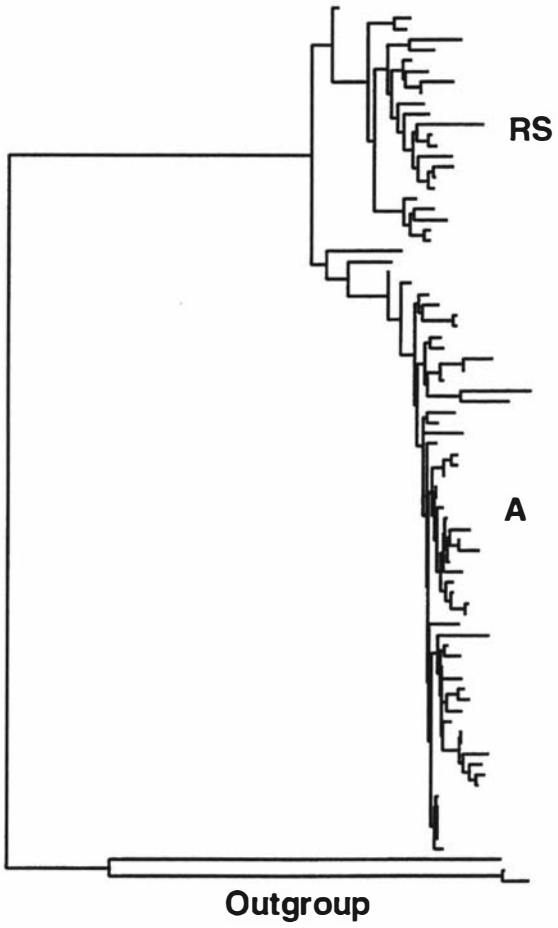


Figure 6.5. A neighbor-joining tree of the 80 ancient DNA sequences. This tree was rooted with the chintrap and gentoo penguins.

6.4 Discussion

6.4.1. Survival and Recovery of Ancient DNA

Subfossil Adélie penguin bones from Antarctica contained an extraordinary quality of ancient DNA. Well-preserved DNA was interpreted as the ability to amplify long sequences of DNA, from material dated at hundreds or thousands of years old. The largest amplicon (~1600 b.p.) was derived from a ~450 yr old subfossil bone. The maximum size PCR-fragment obtained is not typically reported in ancient DNA studies, however, most authors report amplifying and sequencing 100-300 b.p. fragments (Rollo, 1998). Previously, the longest reported PCR fragment was from the study of Hagelberg et al. (1991), who showed a faint amplification of ~800 b.p. sequence from a medieval (1200 A.D.) human skeleton. These authors could

not make any direct correlation between DNA preservation and burial conditions, since significant variability in preservation was reported from the same site.

The well-preserved DNA recovered from subfossil Adélie penguin bones, suggest the cold and arid conditions of Antarctica have dramatically slowed the decomposition of nucleic acids. Antarctica is colder than any other continent or any island, recording the lowest temperature ever (-89.6°C) at Vostok station. The averaged monthly temperatures from several Antarctic locations, as well as those recorded on a temperate island are shown in Figure 6.6. Cold and dry conditions slow the process of depurination and oxidation of DNA (Höss et al., 1996; Greenwood et al., 1999). Future studies of amino acid racemization and base oxidation in Antarctic subfossil material, will most likely show a low rate consistent with the long length of sequences able to be recovered

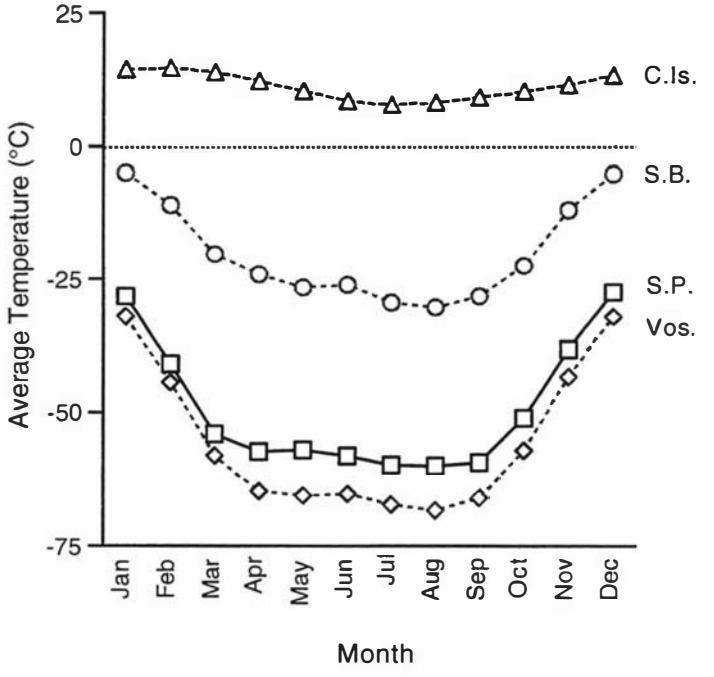


Figure 6.6. The mean monthly temperatures for four southern hemisphere locations. C.Is. = the temperate Chatham Islands (mean 11.1°C). Three Antarctic locations: S.B. = Scott Base (mean -20°C); S.P. = South Pole (mean -49.3°C); Vos. = Vostok (mean -55°C). These temperature were recorded from automatic weather stations. Raw data was obtained from the Space Science and Engineering Center, University of Wisconsin Madison (ftp://ice.ssec.wisc.edu/).

In this study, there was a negative correlation between the age of a bone and the length of sequence able to be amplified. For example, for bones dated to 275-1258 yr BP 663-1023 b.p. sequences were routinely amplified, and 180-390 b.p. sequences

were consistently amplified from bones dated between 1086-6082 yr BP. This age correlation is unusual since many studies contend there is no relationship between the age of a bone and DNA preservation (e.g. Höss et al., 1996). There has not been a detailed study of dated subfossil bone samples compared with length of amplifiable sequences. Many ancient DNA studies have a limited sample size, since large quantities of material from which to extract DNA are not readily available. Therefore the number of points on which an age to sequence length correlation can be made is limited. Furthermore, in many situations fragmentation of DNA may occur too rapidly for such comparisons to be made. Most of the endogeneous DNA in an ancient sample is thought to fragment within the first few months or years after death of an organism (Pääbo, 1989). The most common mode of DNA fragmentation is breakage at 3'-phosphodiester bonds following depurination. The cold temperatures of Antarctica may have slowed depurination, causing DNA fragmentation to be slower than in other environments. Data from this thesis indicates gradual fragmentation of DNA over the time period sampled. Longer fragments may exist in ancient DNA extractions but modified thymines (hydantoins) prevent PCR from making double-stranded copies, since polymerases stall at these lesions. Unfortunately, there has not being a detailed study of ancient DNA template size, thymine damage and the size range of sequences able to be amplified. The large number of samples and range of amplicon sizes recovered from ancient Adélie penguin samples make them ideal for this type of investigation.

6.4.2. Ancient Adélie DNA and the Deglaciation

DNA sequences from radiocarbon dated subfossil bone material revealed the times *A* and *RS* lineages were present in the Ross Sea. These dates can be compared with the timing of deglaciation after the last glacial maximum (~20 kyr BP) and hence the arrival of penguins into the area. Ice cores from the Taylor Dome (77°47.7'S, 158°43.1'E) have provided a measure of temperature change in the Ross Sea embayment, spanning the last glacial maximum to the Holocene (Mayewski et al., 1996; Steig et al., 1998). These results were obtained by measuring changes in the concentration of a heavy hydrogen isotope, deuterium (δD), serially through each core. In a colder climate there is a decrease in the amount of ^2H_2O deposited as glacial ice. Using δD measurements Steig et al., (1998) showed that temperatures in the Ross Sea increased suddenly around 14.6 kyr BP (Figure 6.7). Interestingly, this

finding mimics the ice core record from Greenland more closely than records from inland Antarctica. Steig et al. (1998) suggested that during global warming after the last glacial maximum, the coastal margins of Antarctica were strongly influenced by Northern Hemisphere heat transferred through deep Atlantic currents.

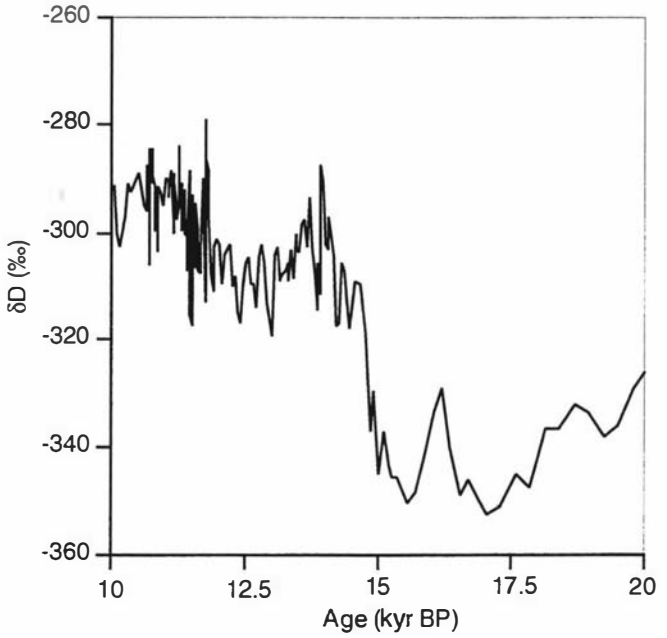


Figure 6.7. A proxy for temperature, the deuterium (δD) profile from the Taylor Dome ice core covering 10-20 kyr BP shows changing climate between 14-15 kyr BP.

Warming of the coastal margins in Antarctica occurred at ~14.6 kyr BP, however the retreat of the voluminous Ross Ice Shelf lagged several thousand years behind this temperature change. Conway et al. (1999) proposed that the onset of deglaciation was some time after 11 kyr BP and the retreat of the Ross Ice Shelf completed around 7-8 kyr BP. This is based on several lines of evidence. (1) Proglacial lakes formed by grounded ice damming the Taylor Valley date to 9,420 yr BP. The arrival of marine molluscs in the proglacial lakes has been ^{14}C-dated to 7,550 yr BP, and indicates when the grounding line retreated, opening the valley to a marine environment. (2) The offloading of glacial ice from land, results in isostatic uplift and the sea level appears to drop. These authors used dated shells, sealskins and other remains to construct a relative sea level curve from raised beaches on the Scott Coast. (3) Support is also taken from the height of the Hatherton Glacier that is strongly effected by the grounding line, as well as (4) the differing accumulation rates of ice on Roosevelt Island. (5) Lastly, storm patterns around the Taylor Dome

changed as the Ross Ice Shelf retreated and the modern weather pattern became established around 6 kyr BP (Morse et al., 1998).

The time of deglaciation is indicative of when two previously allopatric Adélie populations may have resumed contact after being separated during the last glacial cycle. Radiocarbon dated penguin remains from the Terra Nova Bay area indicate occupation began around 7.5 kyr BP (Baroni and Orombelli, 1994). The discovery of *A* and *RS* lineages at approximately equal frequencies in the Ross Sea from 6082 years ago, suggests a simultaneous invasion from two ice-age populations after the retreat of the Ross Ice Shelf. A rapid influx of both lineages of penguins from ice-age regufia is consistent with the evidence for rapid population expansion of Adélie penguins presented in Chapter Five. This illustrates the relationship between the evolution of the Antarctic continent and its remarkable biota.

6.5 References

Austin, J.J., Ross, A.J., Smith, A., Fortey, R.A., and Thomas, R.H. 1997. Problems of reproducibility — does geologically ancient DNA survive in amber preserved insects. Proc. Roy. Soc. Lond B **264:** 467-474.

Bada, J.L., Wang, X.S., Poinar, H.N., Pääbo, S., and Poinar, G.O. 1994. Amino acid racemization in amber-entombed insects – implications for DNA preservation. Geochimica et Cosmochimica Acta **58:** 3131-3135.

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

Berkman, P.A., Andrews, J.T., Björck, S., Colhoun, E.A., Emslie, S.D., Goodwin, I.D., Hall, B.L., Hart, C.P., Hirakawa, K., Igarashi, A., Ingólfsson, O., Löpez-Martínez, J., Lyons, W.B., Mabin, M.C.G., Quilty, P.G., Taviani, M., and Yoshida, Y. 1998. Circum-Antarctic coastal environment shifts during the Late Quaternary reflected by emerged marine deposits. Antarctic Sci. **10:** 345-362.

Breimer, L., and Lindahl, T. 1985. Thymine lesions produced by ionizing radiation in double-stranded DNA. Biochem. **24:** 4018-4022.

Bucala, R., Model, P., and Cerami, A. 1984. Modification of DNA by reducing sugars: A possible mechanism for nucleic acid aging and age-related dysfunction in gene espression. Proc. Natl. Acad. Sci. USA **81:** 105-109.

Conway, H., Hall, B.L., Denton, G.H., Gades, A.M., and Waddington, E.D. 1999. Past and future grounding-line retreat of the West Antarctic ice sheet. Science **286:** 280-283.

Greenwood, A.D., Capelli, C., Possnert, G., and Pääbo, S. 1999. Nuclear DNA sequences from late Pleistocene megafauna. Mol. Biol. Evol. **16:** 1466-1473.

Hadly, E.A., Kohn, M.H., Leonard, J.A., and Wayne, R.K. 1998. A genetic record of population isolation in pocket gophers during the Holocene climate change. Proc. Natl. Acad. Sci. USA **95:** 6893-6896.

Hagelberg, E., and Clegg, J.B. 1991. Isolation and characterisation of DNA from achaeological bone. Proc. Roy. Soc. Lond B **244:** 45-50.

Handt, O., Höss, M., Krings, M., and Pääbo, S. 1994. Ancient DNA: methological challenges. Experientia **50:** 524-528.

Höss, M., Jaruga, P., Zastawny, T., Dizdaroglu, M., and Pääbo, A. 1996. DNA damage and DNA sequence retrieval from ancient tissues. Nucleic Acids Res. **24:** 1304-1307.

Ivanov, P.L., Wadhams, M.J., Roby, R.K., Holland, M.M., Weedn, V.W., and Parson, T.J. 1996. Mitochondrial DNA sequence heteroplasmy in the Grand Duke of Russia Georgij Romanov establishes the authenticity of the remains of Tsar Nicholas II. Nat. Genet. **12:** 417-420.

Kamiya, H., Murata-Kamiya, N., Lin, P.K.T., Brown, D.M., and Ohtsuka, E. 1994. Nucleotide incorporation opposite degenerate bases by *Taq* DNA polymerase. Nucleosides and Nucleotides **13:** 1483-1492.

Krings, M., Stone, A., Schmitz, R.W., Krainitzki, H., Stoneking, M., and Pääbo, S. 1997. Neandertal DNA sequences and the origin of modern humans. Cell **90:** 19-30.

Kumar, S., Tamura, K., Jakobsen, I., and Nei, M. 2001. MEGA 2: Molecular Evolution Genetics Analysis software. Bioinformatics ***Submitted***.

Lindahl, T. 1993. Instability and decay of the primary structure of DNA. Nature **362:** 709-715.

Lindahl, T., and Nyberg, B. 1972. Rate of depurination of native deoxyribonucleic acids. Biochem. **11:** 3610-3618.

Mayewski, P.A., Twickler, M.S., Whitlow, S.I., Meeker, L.D., Yang, Q., Thomas, J., Kreutz, K., Grootes, P.M., Morse, D.L., Steig, E.J., Waddington, E.D., Saltzman, E.S., Whung, P.-Y., and Taylor, K.C. 1996. Climate change during the last deglaciation in Antarctica. Science **272:** 1636-1638.

Morse, D.L., Waddington, E.D., and Steig, E.J. 1998. Ice age storm trajectories inferred from radar stratigraphy at Taylor Dome, Antarctica. Geophys. Res. Letters **25:** 3383-3386.

Namiki, M. 1988. Chemistry of Maillard reactions: Recent studies of the browning reaction mechanism and the development of antioxidants and mutagens. Adv. Food Res. **32:** 115-184.

Pääbo, S. 1989. Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification. Proc. Natl. Acad. Sci. USA **86:** 1939-1943.

Pääbo, S. 1990. Amplifying ancient DNA. *In* PCR Protocols: A Guide to Methods and Applications. *Edited by* M. A. Innis, D. H. Gelfand, J. J. Snisky and T. J. White. Academic Press, San Diego. pp. 159-166.

Pääbo, S., Higuchi, R.G., and Wilson, A.C. 1989. Ancient DNA and the polymerase chain reaction: the emerging field of molecular archaeology. J. Biol. Chem. **264:** 9709-9712.

Pääbo, S., Iriwn, D.M., and Wilson, A.C. 1990. DNA damage promotes jumping between templates during enzymatic amplification. J. Biol. Chem. **265:** 4718-4721.

Petri, A., and Baroni, C. 1997. *Penguin*, a Macintosh application for entry and presentation of radiocarbon-dated samples. Radiocarbon **39:** 61-65.

Poinar, H.N., Höss, M., Bada, J.L., and Pääbo, S. 1996. Amino acid racemization and the preservation of ancient DNA. Science **272:** 864-866.

Poinar, H.N., and Stankiewicz, B.A. 1999. Protein preservation and DNA retrieval from ancient tissues. Proc. Natl. Acad. Sci. USA **96:** 8426-8431.

Rollo, F. 1998. Ancient DNA: problems and perspectives for molecular microbial palaeoecology. *In* Advances in Molecular Ecology. *Edited by* G. R. Carvalho. IOS Press, Oxford, UK.

Scholz, M., Giddings, I., and Pusch, C.M. 1998. A Polymerase Chain Reaction inhibitor of ancient hard and soft tissue DNA extracts is determined as human collagen type I. Anal. Biochem. **259:** 283-286.

Steig, E.J., Brook, E.J., White, J.W.C., Sucher, C.M., Bender, M.L., Lehman, S.J., Morse, D.L., Waddington, E.D., and Clow, G.D. 1998. Synchronous climate changes in Antarctica and the North Atlantic. Science **282:** 92-95.

Stuiver, M., and Reimer, P.J. 2000. Radiocarbon calibration program. Quaternary Isotope Laboratory, University of Washington, Washington.

Swofford, D.L. 2000. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Sinauer Associates, Sunderland, Massachusetts.

Tuross, N. 1995. The biochemistry of ancient DNA in bones. Experientia **50:** 530-535.

Waite, E.R., Child, A.M., Craig, O.E., Collins, M.J., Gelsthorpe, K., and Brown, T.A. 1997. A preliminary investigation of DNA stability in bone during artificial diagenesis. Bull. Soc. géol. France **168:** 547-554.

Chapter Seven

Conclusions: Modern and Ancient Adélie Penguins

7.1 Introduction

The hypervariable region I (HVR-1) of the mtDNA control region was sequenced for a total of 381 modern and 80 ancient samples of Adélie penguins. The modern samples were derived from 13 populations around Antarctica and the ancient samples from 16 locations on the coast of the Ross Sea. This chapter presents an analysis of the sequence divergence between modern and ancient samples and discusses recent estimates of the rate of evolution for the HVR-1. The two lineages of Adélie penguins (*A* and *RS*) are deeply split, separated by 6% sequence difference (Chapter Five). These lineages indicate a population history of Adélie penguins that is closely associated with the glacial changes in Antarctica during the ice-ages of the Late Pleistocene. A rate of evolution for the HVR-1 provides an opportunity to understand the origin of these lineages. Moreover, timing their divergence can provide a view of how Adélie penguins endure global climate change while inhabiting the most extreme continent on earth, Antarctica.

7.2 Measurably Evolving Lineages

7.2.1. A Discussion of the Estimated Rate of Evolution

Any method that measures a rate of evolution from DNA sequences assumes they are evolving at the same rate. Therefore the HVR-1 data set was tested for its clock-like structure. Both a relative-rate test (RRT; Wu and Li, 1985) and a likelihood ratio test (LRT; Felsenstein, 1981) were performed between the two lineages *A* and *RS*. The RRT examines the distance from one reference taxon to two others. A group of taxa are considered to pass a RRT when there is no significant difference between the two distance measurements (i.e. $d_{AB} \sim d_{AC}$, where d is the distance and A, B and C are three different sequences). A RRT conducted across multiple samples can estimate the proportion of deviant sequences within a data set. A relative rate test was performed on 800 combinations of three taxa compiled using a Basic program and analysed with the software SPLITS TREE ver. 2.4 (Huson, 1997). In each case, one individual from lineage *RS* was compared to two individuals from lineage *A*. Almost all comparisons in lineage *A* passed a RRT (89%). The significance levels for the RRT are presented graphically in Figure 7.1.

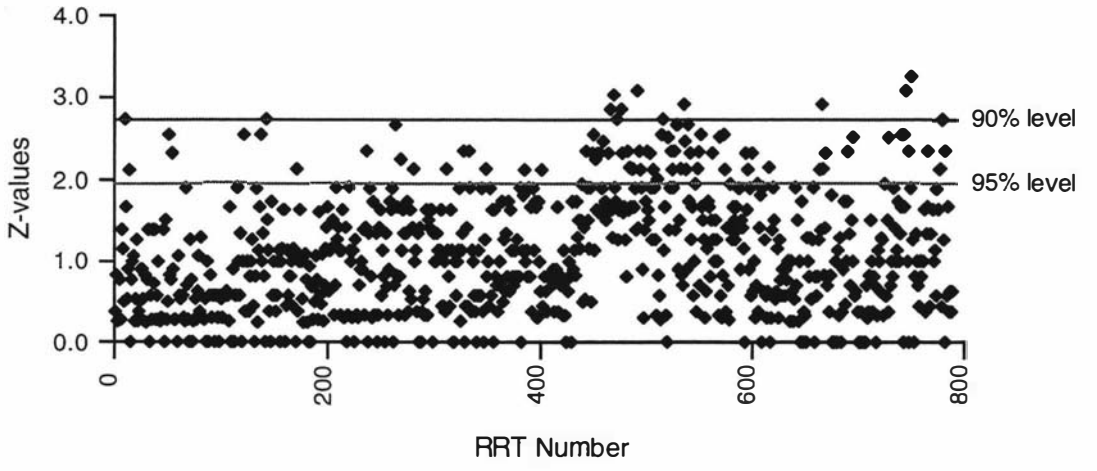


Figure 7.1. The significance values for 800 relative rate tests (RRT). Each test was performed on random combinations of three sequences between the two lineages. Z-normal values were calculated as Δ/standard error.

The LRT looks for a significant difference between maximum likelihood distances (L) from a tree topology, both constrained and not constrained to a molecular clock (ie. $2\Delta = \log L_{\text{no clock}} - \log L_{\text{clock}}$). A UPGMA tree was constructed using HKY85

corrected maximum likelihood distances using the *RS* lineage as the root to the *A* lineage, and the likelihood values for the constrained and unconstrained tree were estimated in PAUP 4.0* (Swofford, 2000). A χ^2-test of the 2Δ value, with the number of sequences (n) as the degrees of freedom (n-2), shows no significant difference ($P = 0.065$) at the 95% level between the unconstrained maximum likelihood distances and a molecular clock. Therefore the two lineages can be considered as diverging from each other in a clock-like manner.

Studies of rapidly evolving pathogens (e.g. dengue virus and HIV-1) which involve repeated sampling of viral populations over many years have used both pairwise distance and maximum likelihood methods to estimate a rate of substitution (Shankarappa et al., 1999; Rambaut, 2000). The Adélie penguin HVR-1 data set comprises a large number of sequences with few informative sites; hence the computational time needed to search all non-polytomious maximum likelihood trees is enormous. Therefore maximum likelihood methods were not employed on this data set. For a large and complex data set, a median network is the most appropriate form in which to represent a set of relationships among haplotypes (Posada and Crandall, 2001). Moreover, a standard method for estimating a rate using reconstructed lineages (Li et al., 1988; outlined in Figure 7.2) can be modified to fit these more complex networks.

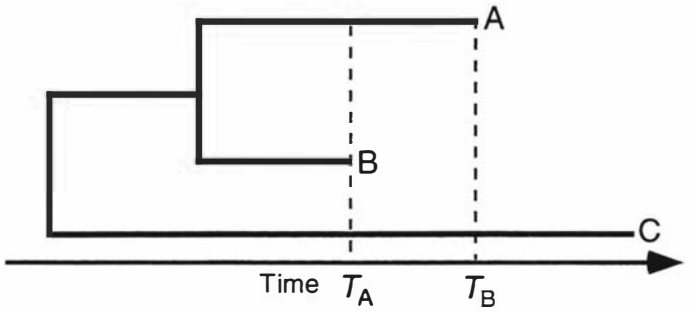


Figure 7.2. The rate of evolution can be calculated by comparing two sequences (A and B), isolated at different points in time and independent of each other, with an outgroup sequence (C). The rate along lineage A = $(d_{AC} - d_{BC})/(T_A - T_B)$, where the distance between A and C and between B and C is the divergence between time T_A and T_B. Adapted from Rambaut (2000).

Median networks are graphs constructed by partitioning a data set into groups character by character, and if represented in multi-dimensional space include all the most parsimonious trees (Bandelt et al., 1995). Such networks are constructed using binary characters, applicable to HVR-1 population data where transitions are most

frequent at all sites (e.g. A and G can be coded as 0 and 1 respectively). The presence of more than two bases at a single position (i.e. a rare transversional event) can be denoted as a pair or triplet of binary characters. The nodes on a median network represent haplotypes that are either sampled or inferred as ancestral intermediates (Bandelt et al., 2000). A rate can be estimated by counting the number of changes required to minimally connect the edges for all ancient sequences to a designated origin (root) on the graph, then subtracting this value from the value obtained using the same process for the modern sequences (Lambert et al., *Appendix D*). Considering only independent ancient and modern lineages, their difference ($\Delta_{M\text{-}A}$ = modern diversity – ancient diversity) can then be transformed into the average substitution rate (mean number of substitutions per lineage per unit time).

To estimate a rate in Adélie penguins, sequences and sites with any missing data were excluded from the analysis. This new data set contained 318 modern and 67 ancient 193 b.p. DNA sequences, representing 152 haplotypes. Lambert et al. (*Appendix D*) constructed a minimum spanning tree (Kruskal, 1956; Bandelt et al., 1999) using 13 mutually compatible sites which divided the sequences into seven subgroups. In that study median networks were then constructed for each of the seven subgroups and the root was designated as the haplotype(s) which connected the outgroup. The outgroup was assigned to the midpoint between the two lineages (*A* and *RS*), a point which connects the sequences to their closest relative, the gentoo penguin. The summed $\Delta_{M\text{-}A}$ of all subgroups was 151 substitutions. This is the total number of changes that have occurred since the ancient sequences. This $\Delta_{M\text{-}A}$ total was then averaged for the mean age of bone samples (2437 yr BP), the number of sites used in the analysis, and the number of modern samples, thereby providing a rate estimate of 1.01 s/s/Myr. A delete-one jackknife procedure (Efron, 1982) was then used to determine the confidence intervals of 0.89 – 1.16 s/s/Myr (Lambert et al., *Appendix D*).

A second approach to estimating the rate used the distance measure proposed by Shankarappa et al. (1999) and Leitner and Albert (1999). In its simplest form, when a constant mutation rate and population size can be assumed, time of separation (t) of every pair of samples is plotted against their pairwise sequence difference (d). A linear regression is fitted to these plotted points ($d = \beta t + \alpha$) using the least-squares

method and the expected (average) rate of change per unit time is given by the slope (β). The *y*-axis intercept (α) has been interpreted as the ancestral divergence, or the amount of sequence difference before sampling began (Leitner and Albert, 1999). An HVR-1 rate estimate was obtained as part of the serial sample UPGMA (sUPGMA) algorithm when implemented in the software Pebble 0.4α (Drummond and Rodrigo, 2000). Using this approach on an uncorrected data set representing 373 unique-within-timepoint sequences, the rate of change was calculated to be 0.92 s/s/Myr (Lambert et al., *Appendix D*). A parametric bootstrap that compares the actual data to a model of simulated data gave confidence limits of –1.18 to 6.17; whereas a delete-half jackknife procedure gave 95% confidence intervals of –0.42 to 2.35.

7.2.2. Comparing the Estimated Rates for the Control Region

The two above methods, averaging $\Delta_{M\text{-}A}$ across seven median networks and the sUPGMA algorithm, have estimated the HVR-1 substitution rate in Adélie penguins as five times higher than a rate calibrated from a phylogenetic approach (0.208 s/s/Myr: Quinn, 1992). This phylogenetic estimate for the HVR-1 has been considered applicable to many bird species (eg. Wennick et al., 1996; Baker and Marshall, 1997; Bensch and Hasselquist, 1999). However, the confidence intervals for the sUPGMA estimate included 0.208 s/s/Myr as well as zero substitutions. The negative value on these intervals may reflect divergence from the outgroup sequences by stochastic lineage extinction and does not involve mutation (see page 5 of Chapter One).

There are three possible explanations for the higher rate found in Adélie penguins, compared with the phylogenetic rate estimated for avian species (Quinn, 1992). First, a rate calibrated for one species has no relevance to another species. For example, there is variability in the mutation rates among species due to the process through which mutations arise, a prediction of the metabolic rate hypothesis (see Chapter Two). However, Reyes et al. (1998) and Gissi et al. (2000) proposed that the type and rate of mutations observed in mtDNA result from the dynamics of DNA replication in the mitochondria, and rejected metabolism, generation time and body size as explanations (reviewed in Chapter Two). If their theory is correct, as long as the mtDNA of avian species replicated in the same manner, base line mutation rates

would be relatively universal. An estimated rate of mutation from one species could therefore be exported to another, if the mtDNA in both species is assumed to be evolving in a clock-like manner.

Second, differences in rate estimates between indirect (phylogenetic) and direct (pedigree) analyses are the subject of much debate in human mtDNA studies of the HVR-1 and HVR-2. A directly measured mutation rate from family material is more than twenty-fold higher than substitution rate estimated from the fossil record (see Table 7.1). Parsons et al. (1997) suggested that drift and selection might account for the lower mutation rate over longer time periods than that detected in pedigree studies. Mutations may arise at a high frequency but are eliminated from the population due to high rates of lineage extinction. Sigurðaròttir et al. (2000) recently estimated a mutation rate using pedigrees of Icelandic families and reported a lower rate (0.32 s/s/Myr) compared with other studies. Jazin et al. (1998) disagreed with the interpretation of the high rate detected by Parsons et al., arguing their averaged rate was based on 'hot spots'. If a few sites in the control region have a high mutation rate compared to all other positions, they should not be used as the mutation rate across all sites. Parson and Holland (1998) responded by pointing out that the bootstrap analysis Jazin et al. included roughly 50% of all nucleotide positions in the HVR-1 and HVR-2. Hence, it was not surprising that the sites of observed substitutions could be categorised as 'hot spot'. Moreover, a recent study by Meyer et al. (1999) estimated the average discrete gamma value (a measurement of rate per site using maximum likelihood) for each nucleotide position using several thousand human HVR-1 and HVR-2 sequences. Their results showed that the sites at which Parsons et al. detected mutations did not have a significantly higher gamma value than most of the other sites in the HVR-1. The different rates estimated in pedigree and phylogenetic studies is still a debate that is largely unresolved. Therefore, the avian phylogenetic rate estimated over millions of years may not be comparable to the mutation rate in a population over 6000 yr.

Third, the estimated rate for the avian control region using a phylogenetic approach measures the number of mutations that have accumulated along two lineages since they diverged from their most recent common ancestor (Sigurðaròttir et al., 2000). This approach can not take into account stochastic lineage extinction because it only measures the differences between two lineages that have survived since their most

recent common ancestor. The rate estimate presented in this study is a population rate, a measurement of both mutations and lineage turnover. Phylogenetic and population approaches measure slightly different parameters. However, a comparable population estimate has been reported from a study of humans. Lundstrom et al. (1992) used a coalescent model to estimate the substitution rate for the HVR-1. These authors knew the effective population size of the Amerindian tribal group, and using both a least-squares and a likelihood approach estimated the rate of change for both transitions and transversions at 1.1 s/s/Myr. This result is very similar to the estimated rate for Adélie penguins.

Table 7.1. Estimated evolutionary rates for the mtDNA control region hypervariable regions (HVR) I and II, using different approaches and different species. The rate is presented as substitutions/site/Myr. For the ancient DNA estimate: a. is from median networks; b. is from the sUPGMA algorithm.

Estimate	Species	Reference	Rate	C.I.	HVR
Pedigree	Humans	Howell et al., 1996	2.60		I & II
	Humans	Parsons et al., 1997	2.50		I & II
	Humans	Sigurðaròttir et al., 2000	0.32	0.065-0.97	I & II
Ancient DNA	Adélie	Lambert et al.	1.01[a]	0.740-1.05	I
	Adélie	Lambert et al.	0.92[b]	-0.420-2.35	I
Coalescent Model	Humans	Lundstrom et al., 1992	1.100		I
Phylogenetic	Geese	Quinn, 1992	0.208	-	I
	Humans	Vigilant et al., 1991	0.151		I & II
	Humans	Stoneking et al., 1992	0.118	0.058-0.18	I & II
	Humans	Hassegawa et al., 1993	0.098	0.000-0.22	
	Humans	Tamura and Nei, 1993	0.087	0.025-0.15	I & II

A potential problem for the rate estimated from Adélie penguins is a small sample size of contemporaneous sequences. Every haplotype in the living population must be sampled in order to make the most accurate comparison between modern and ancient samples. If all the variants in the modern population are not sampled, then differences to the ancient population could be an artefact of a small sample size. One way to circumvent this problem would be to show changes in some 'clear' phylogenetic structure through time. A strong feature of the Adélie HVR-1 data is that the net sequence divergence (ie. adjusted for within lineage polymorphisms) between the modern *A* and *RS* Adélie penguin lineages is 7% (±1.2), while that

between older ancient bone samples (mean 4,200 yr BP; n = 37) is 6.1% (±1). Hence, these two lineages are measurably evolving and have accumulated 0.9% sequence divergence over this time period. The two lineages are the most phylogenetically distinct feature of the HVR-1 data set and it could be their presence that has allowed a rate to be estimated. It is interesting to note that the sUPGMA estimation procedure applied to each lineage (*A* and *RS*) separately gives 0.1 and 0.4 s/s/Myr respectively, compared with 0.9 s/s/Myr overall.

7.2.3. The Ice-Age Legacy: Estimating the Divergence Time

A rate of change for the HVR-1 has been estimated and the two lineages *A* and *RS* appear to have diverged from each other in a clock-like fashion. Therefore, the timing of separation of the two lineages can be established. The rate estimated from ancient and modern samples assumed that the population size and mutation rate remained constant over time. This assumption must also be made when applying this rate to the timing of divergence between two lineages. However, during the last glacial cycle Adélie penguin populations were almost certainly smaller and thus, the rate of lineage extinction was probably higher than over the last 6000 yr. Consequently, the 'population rate' will underestimate the divergence time. On the other hand, the phylogenetic rate can not take into account lineage turnover and is only an estimate of the mutations that have accumulated between two extant lineages.

The total average sequence difference (ie. unadjusted for within lineage polymorphisms) between the two lineages is 8.3% (±1.1), when each haplotype is counted only once. Using the estimated substitution rate of 1.0 s/s/Myr, the point of divergence for lineage *A* and *RS* would have been 83 kyr BP. The much slower rate (0.208 s/s/Myr) proposed by Quinn (1992) places the point of divergence at 400 kyr BP. Given the extensive knowledge of the past climatic and glaciological conditions in Antarctica, the interpretation of the recent history of Adélie penguins proposed in Chapter Five can now be given a temporal component and examined. An historical interpretation of this species may shed light on the applicability of each rate estimate.

7.3 The Ice-Ages of Antarctica

7.3.1. Indicators of Past Climate Change

Past climate trends are inferred in several different ways. Most commonly used is the measurement of abundance of oxygen isotopes (^{18}O and ^{16}O) from preserved calcitic tests (outer coverings) of selected marine Foraminifera (Williams et al, 1998). The deviation of the isotopic ratios from that found in modern oceans ($\delta^{18}O$) indicates the climatic conditions when the sample was deposited. Water molecules formed by the abundant ^{16}O are comparatively light, and therefore evaporate more easily and travel longer distances than those formed by the heavier ^{18}O. When moist air cools the less abundant and heavy $H_2{}^{18}O$ preferentially condenses and the atmosphere becomes $H_2{}^{16}O$ rich. A colder climate will leave oceans isotopically heavy, displaying a positive $\delta^{18}O$ value. Although Shackleton (1967) showed that during the huge ice-ages of the Quaternary large amounts of ^{16}O were stored in glacial ice. This lighter isotope travels further inland across ice sheets than ^{18}O during the cooler glacial cycles. The large-scale entrapment of ^{16}O in the ice sheets caused major changes in the abundance of ^{16}O and ^{18}O in the oceans. Shackleton (1967) proposed $\delta^{18}O$ values from Quaternary forams could also be used as a proxy ice-volume measure. The analysis of $\delta^{18}O$ values in Quaternary marine deposits showed the timing and subdivision of glacial cycles in terms of ice-volumes.

There are several other elements that can be measured to reveal past climate conditions, especially those trapped in ice cores. Cosmic radiation in the upper atmosphere constantly produces the isotope Beryllium (^{10}Be), which is washed out attached to aerosol particles. If precipitation were lower than present day levels, the concentration of ^{10}Be in the ice core would also be lower than the surface ice. Dust in the atmosphere from arid regions rises when the climate is cooler. The atmospheric concentration of greenhouse gases, such as CO_2 and CH_4, decrease as oceans absorb them during cool periods. Air bubbles become trapped in dense ice as it accumulates, providing 'snap shots' of the atmosphere hundreds of thousands of years ago.

Measurements of $\delta^{18}O$ from the Quaternary are also confounded because oxygen is a component of CO_2 and is absorbed into the oceans during a cold climate. There

has been more success with measuring the concentration of the heavy hydrogen isotope (^2H), deuterium (D), as a proxy for local temperature. If a sample of D is measured at a place where snow begins to fall on ice and then compared to the amount in modern seawater (δD), it serves as a fairly good proxy for temperature. The same principle applies to ^{18}O. Water molecules containing D are heavier than those with ^1H, hence they are not able to travel as far inland in cooler conditions. During the ice-ages δD values (measured in parts per thousand, per ml, and expressed as ‰) are much lower than during any of the interglacial periods.

7.3.2. The Great Ice-ages of Antarctica

The Cenozoic era (70 Myr BP) of geological history has been one characterised by a gradual decrease in global temperatures. At the beginning of this era, the Palaeocene epoch (55-70 Myr BP), Antarctica was ice-free, sea levels were much higher and forests grew to high latitudes. Continental drift and the subsequent break-up of Gondwanaland isolated Antarctica in its present polar position. A consequence of this continental movement was the establishment of the circum-polar ocean current (*ca.* 22 Myr BP) and an unrestricted atmospheric circulation. The Antarctic convergence would have reduced the movement of sensible heat (a transfer of heat between the ocean and atmosphere) from the tropics to the southern latitudes. Major changes in the ocean currents that transfer heat around the globe are among the factors suggested to have contributed to the fall in the earth's temperature. As Antarctica cooled, the East Antarctic ice sheet formed around 12-14 Myr BP. However, a marked cooling at the onset of the Quaternary (2.5-5 Myr BP), led to a huge accumulation of ice on Antarctica. This period has been characterised by dramatic fluctuations in global temperatures that resulted in massive ice-ages lasting approximately 100 kyr interspersed by short warm periods. The Serbian mathematician, Milutin Milankovtich (1879-1958), was the first to develop the theory that changes in the earth's orbit were the key to the onset of glacial cycles.

Glacial variability over the last half-million years has been characterised by 23, 41 and 100 kyr BP cycles, which have been associated with astronomical cycles of precession, obliquity and eccentricity of the earth's orbit (Imbrie et al., 1992). Throughout the major part of the Late Pleistocene the earth experienced ice-age

conditions. During this epoch, relatively short periods of warmer conditions arose approximately every 100 kyr. For the last 10,000 yr (Holocene) the earth has been in one of these warmer periods. Data retrieved from ice coring in Antarctica and Greenland has now provided quantitative details on climatic changes through the late Pleistocene glacial and interglacial cycles. The most important records have been retrieved from Vostok (78°28'S and 106°48'E), which is located on the East Antarctic Ice Sheet at an altitude of 3,488 m above sea level. This remarkable achievement has extended measurements of $\delta^{18}O$, δD, Na, dust, CO_2 and CH_4 over four glacial-interglacial cycles covering 420 kyr (Petit et al., 1999).

The change of temperature at Vostok, calculated from the heavy hydrogen isotope deuterium, is presented in Figure 7.3. Glacial periods are characterised by stages of decreasing temperature until reaching a glacial maximum, on average 9°C cooler than at present. Following each maximum is a sudden rise in temperature to warmer interglacial conditions. Over the last 420 kyr there have been five short interglacial (warm) periods and four long glacial periods which terminated at 324, 238, 129 and 14.6 kyr BP. The interglacial periods following the termination of each glacial cycle are short and ice-age conditions return after 5-10 kyr. The Holocene differs from other interglacial periods by exhibiting stable temperatures persistently for over 11 kyr. Although the previous interglacial stages were relatively short, they were significantly warmer and more erratic than the Holocene. Comparing the previous interglacial period (midpoint at ~129 kyr BP) with the Holocene, it was ~2°C warmer and began cooling after about 4 kyr, however, overall this relatively warm period persisted over a similar time period as the Holocene.

Pressure differences between the contents of air bubbles in the Vostok ice cores and the present atmosphere indicate that the ice on inland Antarctica was several hundred meters lower during the glacial periods (Jouzel et al., 1993). These results emphasise the sensitivity of ice caps to a decrease in the precipitation experienced during glacial periods. The massive accumulation of ice around the globe at the last glacial maximum locked up the world's fresh water and decreased the global sea levels by around 120 m. Changes in the Southern Hemisphere ice volumes were never as dramatic as in the Northern Hemisphere; the largest ice volume increment in Antarctic was the expansion of marine-grounded ice shelfs (e.g. the Ross Ice Shelf) and at the coastal margins. The expanding continental ice would have

overrun many of the present coastal ice-free areas. Colhoun et al. (1992) have estimated that during the last glacial maximum the ice around the coastal margins of Antarctica was on average 500 m thicker and 30 km more advanced than at present. Penguin colonies that formed on the Antarctic coast at the previous interglacial period would have undergone dramatic changes as glacial ice advanced during the ice-age.

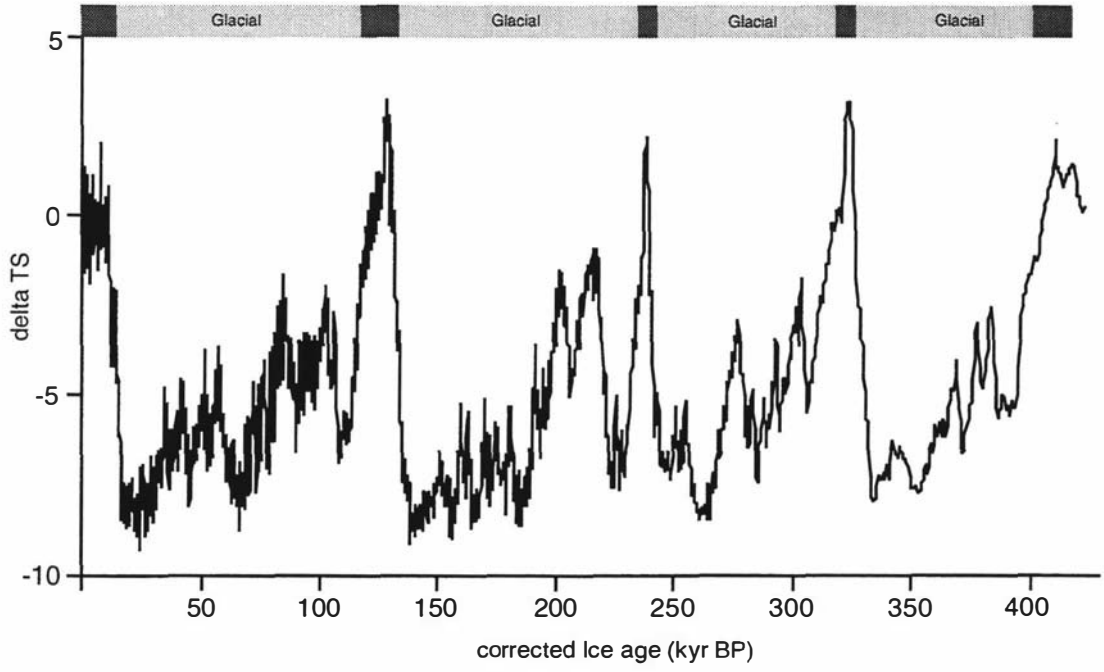


Figure 7.3. Change of temperature from present day at Vostok from ice core records covering 420 kyr BP. Temperatures are calculated as $\Delta T_I = (\Delta\delta D_{ice} - 8\Delta\delta^{18}O_{sw})/9$, for which $\Delta\delta^{18}O_{sw}$ is the global average change from the present day value of seawater $\Delta\delta^{18}O$. Raw data has been taken from Petit et al. (1999) and the delta temperature series (delta TS) is measured in °C. The upper bar represents an interpretation of the glacial and interglacial periods.

7.3.3. Estimating the Time of Divergence

The results presented in Chapter Five indicated that Adélie penguins were split into two isolated populations for a substantial period of time. This resulted in complete lineage sorting into two distinct mtDNA types (*A* and *RS*). The Ross Ice Shelf was one of the most significant features of Antarctica during the last glacial cycle, an impenetrable barrier to nesting sites in the Ross Sea area. When the ice shelf receded the penguins began to inhabit the Ross Sea area and present day populations show admixture of the previously isolated populations. At the end of Chapter Five a rate of substitution was needed to time the allopatric event and make sense of the two lineages in light of the glacial history of Antarctica. Using

the substitution rate of 1.0 s/s/Myr, a point of divergence for lineage *A* and *RS* is estimated to be ~83 kyr BP. Presented in Figure 7.4 is the Vostok temperature profile from the last interglacial period until present day and the proposed divergence time between the two lineages. This divergence date suggests that the circum-polar Adélie penguins were separated into allopatric populations during the last glacial cycle.

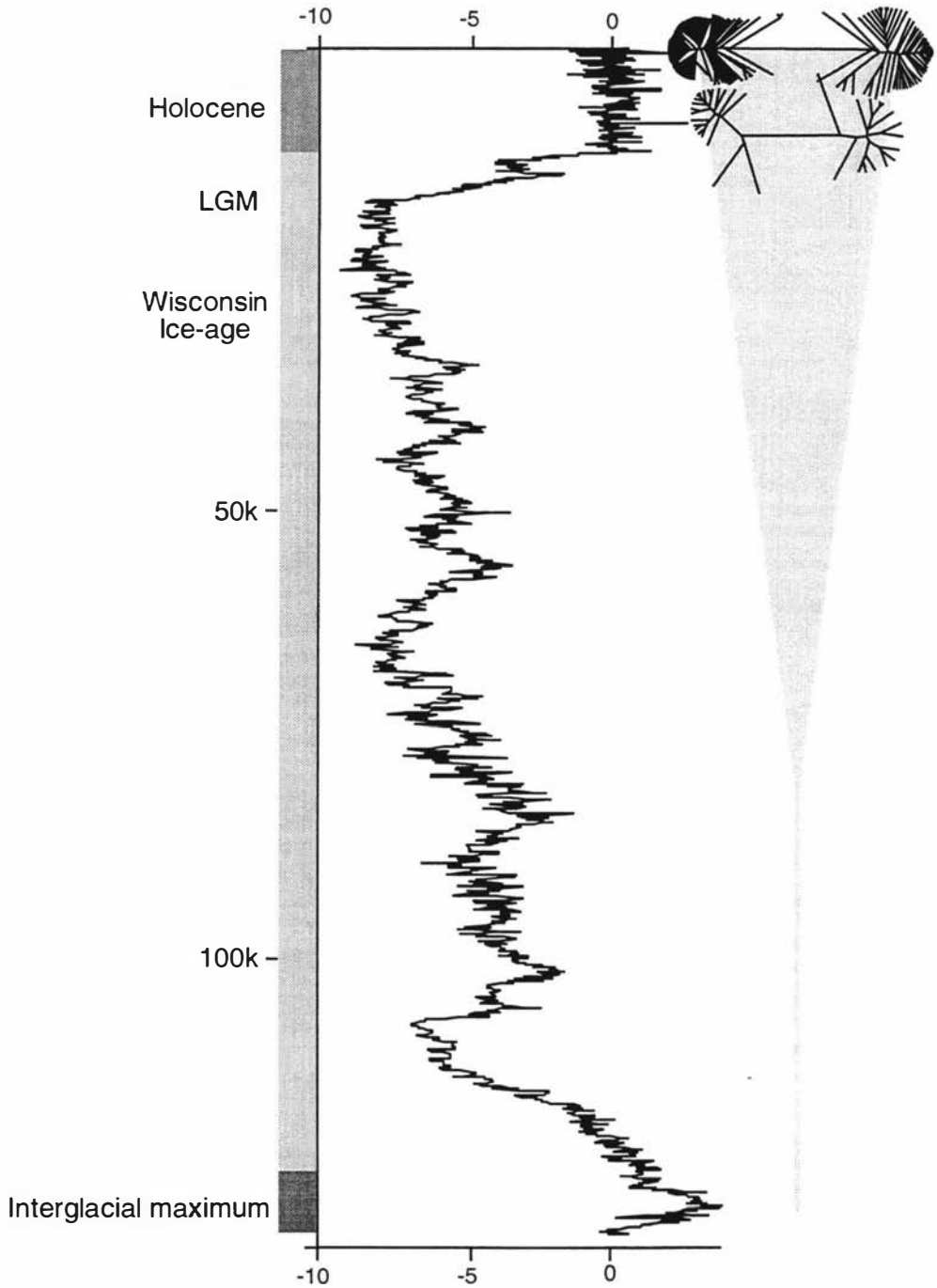


Figure 7.4. Temperature profile from the Vostok ice core covering 125 kyr BP and the proposed divergence time of the *A* and *RS* mtDNA lineages. Using a estimated rate of 1.0 s/s/Myr this split happened 83 kyr BP and was most likely associated with the changes in ice volume during the last glacial cycle.

The much slower phylogenetic rate (0.208 s/s/Myr) proposed by Quinn (1992) places the point of divergence of the two lineages at 400 kyr BP. If this date were correct there must have been an unusual event that occurred to split Adélie penguins into two isolated populations four glacial cycles before present. Such an event was never repeated over the last four subsequent glacial cycles. The population size must have also remained large enough over these four glacial periods, to reduced the chances of complete lineage sorting of the either the *A* or *RS* type from Adélie penguins. The other alternative is that Adélie penguins have maintained a large population size over hundreds of thousand of years. Hence, the two lineages represent two very old haplotypes that have survived in the populations for 400 kyr. The conclusion drawn from trying to explain a slower substitution rate seems unlikely.

The most straight forward explanation is that a substitution rate of 1.0 s/s/Myr is a reasonable estimate and the two mtDNA lineages are the result of separation into ice-age refugia during the last glacial cycle. Contact between these populations was then re-established after the onset of the deglaciation. Every ice-age Adélie penguins endure from dramatic changes to their habitat, and are split into isolated populations. With such limited nesting space in Antarctica, small global climate changes have a big impact on a species in such an extreme environment. If the world's climate became warmer than at present, Adélie penguins might be a species that would thrive with new opportunities.

7.4 Concluding Remarks

The Adélie penguins and their well-preserved subfossil material represent a remarkable opportunity to understand the processes of evolution over thousands of years. Moreover, if the mutation rates for the mitochondrial genome are similar in many different species then Adélie penguins may represent an ideal model for evolutionary genetics. Phylogeographic patterns that classically fit Avise's (2000) models, along with some of the best-preserved ancient DNA known, may one day make this species a distinguished example in biology. There are likely to be many more unique biological examples yet to be discovered in Antarctica. This continent possesses a biota whose precarious existence is intrinsically connected to its journey of geological evolution.

7.5 References

Avise, J.C. 2000. Phylogeography, the history and formation of species. Harvard University Press, Cambridge, Massachusetts.

Baker, A.J., and Marshall, H.D. 1997. Mitochondrial control region sequences as tools for understanding evolution. *In* Avian Molecular Evolution and Systematics. *Edited by* D. P. Mindell. Academic Press, San Diego. pp. 51-82.

Bandelt, H.-J., Forster, P., and Röhl, A. 1999. Median-joining networks for inferring intraspecific phylogenies. Mol. Biol. Evol. **16:** 37-48.

Bandelt, H.-J., Forster, P., Sykes, B.C., and Richards, M.B. 1995. Mitochondrial portraits of human populations using median networks. Genetics **141:** 743-753.

Bandelt, H.-J., Macaulay, V., and Richards, M. 2000. Median networks: speedy construction and greedy reduction, one simulation, and two case studies from human mtDNA. Mol. Phylogenet. Evol. **16:** 8-28.

Bensch, S., and Hasselquist, D. 1999. Phylogeographic population structure of great reed warblers: an analysis of mtDNA control sequences. Biol. J. Linn. Soc. **66:** 171-185.

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

Drummond, A., and Rodrigo, A.G. 2000. Reconstructing genealogies of serial samples under the assumption of a molecular clock using serial-sample UPGMA (sUPGMA). Mol. Biol. Evol. **17:** 1807-1815.

Felsenstein, J. 1981. Evolutionary trees and DNA sequences: a maximum likelihood approach. J. Mol. Evol. **17:** 368-376.

Gissi, C., Reyes, A., Pesole, G., and Saccone, C. 2000. Lineage-specific evolutionary rate in mammalian mtDNA. Mol. Biol. Evol. **17:** 1022-1031.

Hassegawa, M., Di Rienzo, A., Kocher, T.D., and Wilson, A.C. 1993. Toward a more accurate time scale for the human mitochondrial DNA tree. J. Mol. Evol. **37:** 347-354.

Howell, H., Kubacka, I., and Mackey, D.A. 1996. How rapidly does the human mitochondrial genome evolve. Am. J. Hum. Genet. **59:** 501-509.

Huson, D.H. 1997. SplitsTree: A Program for Analyzing and Visualizing Evolutionary Data. University of Bielefeld, Germany.

Imbrie, J., Boyle, E.A., Clemans, S.C., Duffy, A., Howard, W.R., Kukla, G., Kutzback, J., Martinson, D.G., McIntyre, A., Mix, A.C., Molfino, B., Morley, J.J., Peterson, L.C., Pisias, N.G., Prell, W.L., Raymo, M.E., Shackleton, N.J., and Toggweiler, J.R. 1992. On the structure and origin of major glaciation cycles 1. linear responses to Milankovitch forcing. Paleooceanography **7:** 701-738.

Jazin, E., Soodyall, H., Jalonen, P., Lindholm, E., Stoneking, M., and Gyllensten, U. 1998. Mitochondrial mutation rate revisited: hot spots and polymorphism. Nat. Genet. **18:** 109-110.

Jouzel, J., Barkov, N.I., Barnola, J.M., Bender, M., Chappellaz, J., Genthon, C., Kotlyakov, V.M., Lipenkov, V., Lorius, C., Petit, J.R., Raynaud, D., Raisbeck, G., Ritz, C., Sowers, T., Stievenard, M., Yiou, F., and Yiou, P. 1993. Extending the Vostok ice-core record of palaeoclimate of the penultimate glacial peroid. Nature **364:** 407-412.

Kruskal, J.B. 1956. On the shortest spanning subtree of the graph and the travelling salesman problem. Proc. Amer. Math. Soc. **7:** 48-57.

Leitner, T., and Albert, J. 1999. The molecular clock of HIV-1 unveiled through analysis of a known transmission history. Proc. Natl. Acad. Sci. USA **96:** 10752-10757.

Li, W.-H., Tanimura, M., and Sharp, P.M. 1988. Rates and dates of divergence between AIDS virus nucleotide sequences. Mol. Biol. Evol. **5:** 313-330.

Lundstrom, R., Tavare, S., and Ward, R.H. 1992. Estimating substitution rate from molecular data using the coalescent. Proc. Natl. Acad. Sci. USA **89:** 5961-5965.

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

Parsons, T.J., Muniec, D.S., Sullivan, K., Woodyatt, N., Alliston-Greiner, R., Wilson, M.R., Berry, D.L., Holland, K.A., Weedn, V.W., Gill, P., and Holland, M.M. 1997. A high observed substitution rate in the human mitochondrial DNA control region. Nat. Genet. **15:** 363-368.

Petit, J.R., Jouzel, J., Raynaud, D., Barkov, N.I., Barnola, J.-M., Basile, I., Bender, M., Chappelaz, J., Davis, M., Delaygue, G., Delmotte, M., Kotlyakov, V.M., Legrand, M., Lipenkov, V.Y., Lorus, C., Pépin, L., Ritz, C., Satlzman, E., and Stievenard, M. 1999. Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica. Nature **399:** 429-436.

Posada, D., and Crandall, K.A. 2001. Intraspecific gene genealogies: tree grafting into networks. Trends Ecol. Evol. **16:** 37-45.

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.

Rambaut, A. 2000. Estimating the rate of molecular evolution: incorporating non-contemporaneous sequences into maximum likelihood phylogenies. Bioinformatics **16:** 395-399.

Reyes, A., Gissi, C., Pesole, G., and Saccone, C. 1998. Asymmetrical directional mutation pressure in the mitochondrial genome of mammals. Mol. Biol. Evol. **15:** 957-966.

Shackleton, H.J. 1967. Oxygen isotope analyses and Pleistocene temperature records. Nature **215:** 15-17.

Shankarappa, R., Margolick, J.B., Gange, S.J., Rodrigo, A.G., Upchurch, D., Farzadegan, H., Gupta, P., Rinaldo, C.R., Learn, G.H., He, I., Huang, X.-L., and Mullins, J.I. 1999. Consistent viral evolutionary dynamics associated with the progression of HIV-1 infection. J. Virol. **73:** 10489-10502.

Sigurðardóttir, S., Helgason, A., Gulcher, J.R., Stefansson, K., and Donnelly, P. 2000. The mutation rate in the human mtDNA control region. Am. J. Hum. Genet. **66:** 1599-1609.

Stoneking, M., Sherry, S.T., Redd, A.J., and Vigilant, L. 1992. New approaches to dating suggest a recent age for human mtDNA ancestor. Phil. Trans. Roy. Soc. Lond. B **337:** 167-175.

Swofford, D.L. 2000. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Sinauer Associates, Sunderland, Massachusetts.

Tamura, K., and Nei, M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol. Biol. Evol. **10:** 512-526.

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

Wenink, P.W., Baker, A.J., Rösner, H.-U., and Tilanus, M.G.J. 1996. Global mitochondrial DNA phylogeography of Holarctic breeding dunlins (*Calidris alpina*). Evolution **50:** 318-330.

Williams, D., Dunkerley, D., DeDeckker, P., Kershaw, P., and Chappell, M. 1998. Quaternary Environments. Arnold, London.

Wu, C.-I., and Li, W.-H. 1985. Evidence for higher rates of nucleotide substitution in rodents than in man. Proc. Natl. Acad. Sci. USA **82:** 1741-1745.

Appendix A – Field Diaries, Antarctica (K030)

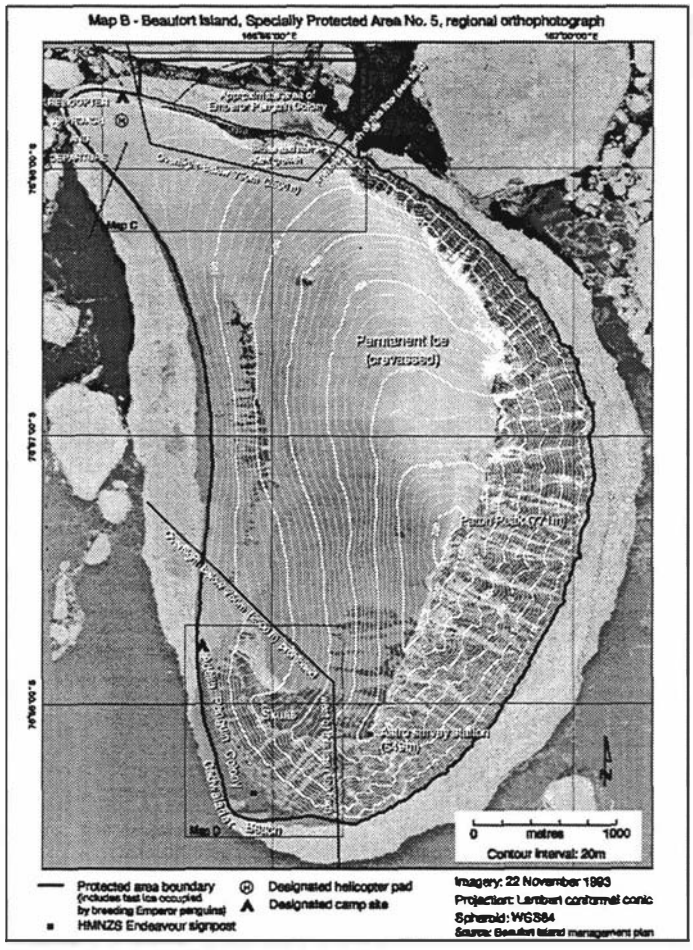
Season ONE: 1996/1997
Personnel: Peter Ritchie and Paul Barrett

Date	Activity
2/4-1-97	Arrived at Scott Base on RNZAF C-130 Hercules
3-4-97	Field training with Davie Robertson
5 -1-97	Transported to Cape Bird by RNZAF Helicopter Hut orientation/unpacked equipment
6 -1-97	Reconnaissance of the Cape Bird Northern, Middle and Southern Adélie penguin colonies Determined that because of the small size and geological structure of the middle and southern colonies, they would probably not yield good stratified subfossil material.
7-1-97	Close survey of Northern colony
8 -1-97	Excavated samples from Cape Bird Sites
9 -1-97	Excavated samples from Cape Bird Sites
10 -1-97	Excavated samples from Cape Bird Sites
11 -1-97	Bad weather
12 -1-97	Bad weather
13 -1-97	Worked site 3A
14 -1-97	Took surface samples
15 -1-97	Caught radio-tagged individuals (studied by K122) and took blood samples.
16 -1-97	Reconnaissance at Cape Crozier — Day Trip on RNZAF Helicopter Caught in bad weather over Mt Erebus. Emergency land at Cape Bird with crew and engineers from Scott Base. Unscheduled overnight stay.
17 -1-97	Weather cleared and returned to Scott Base
18 -1-97	Returned to New Zealand on USAF C-130 Hercules

Season TWO: 1997/1998
Personnel: Peter Ritchie and Paul Barrett

Date	Activity
2-1-98	Arrived at Scott Base on Air National Guard C-130 Hercules
3-1-98	Field training day with Davie Robertson
5-1-98	Cape Crozier on USAP Astar. Weather clear and light winds. Stayed in USAP hut, shared facility with two members of S031
6-1-98	Visited the small Crozier Colony. This trip passed behind Post Office Hill. There were steep ice runs, but recent snow made them easy to cross. Did not need crampons. Colony is small and did not appear to be a good site to visit for subfossil Adélie bones. Returned with several surface samples
7-1-98	Temperature 0°C and 2 kts winds. Survey of the main Cape Crozier Adélie Colony. Beach appears to be geologically dynamic and unstable. There are, however, good cut-a-ways near the waters edge with some good stratification.
8-1-98	Temperature 0°C and 12 kts SW Main sampling day choose sites closest to the beach — these are presumed oldest. SITE A: crumbled and fell apart quickly, very wet and smelt SITE B: largest samples (approximately 10). Tended to be drier and accessible. Never seemed to find whole skeleton. SITE C: straight edges and easy to cut away. Dry and accessible. Never seemed to find whole skeletons. Possible slumping of the edge.
9-1-98	Returned to sample SITE B, this really is an excellent site. Found bone standing upright through the permafrost. The permafrost begins at 0.5m and impossible to dig through. Left 24hrs to thaw before attempting deeper sampling.
10-1-98	Returned to sample SITE B.
11-1-98	Helicopter pick-up time moved forward one day. Return to SITE B for additional sampling.

12-1-98	Helicopter turned around due to bad weather — snow and very low visibility.
13-1-98	Moved to Cape Bird on K04. Good weather unlimited visibility
14-1-98	Sampled northern colony, the same sites as the 1996/97 seasons
15-1-98	Sampled northern colony, the same sites as the 1996/97 seasons
16 -1-98	Sampled northern colony, the same sites as the 1996/97 seasons
17 Jan	Sampled northern colony, the same sites as the 1996/97 seasons
18 Jan	Sampled northern colony, the same sites as the 1996/97 seasons
19 -1-98	Moved to Beaufort Island on US Coast Guard Dolphine. Joined by Davie Robertson. Landed and established camp on the Northern end. (See Figure)
20 -1-98	Traversed island via the western coast to access the southern colony. This was the most direct and safest access route. We used ropes, harnesses, crampons and ice axes. The good weather meant much of the snow and ice was soft. Caught 20 Adélie penguins with a hand net and took a small blood sample. Surveyed beach and found good subfossil bone samples at the northern end of the colony
21 -1-98	Sampled the northern end of Beaufort Island for subfossil bones
22 -1-98	Returned to Scott Base by US Coast Guard Dolphine.
23 -1-98	Returned to NZ on Air National Guard C-130 Hercules



. **Figure B.1** A map of Beaufort Island (made available by ICAIR)

Season THREE: 1998/1999
Personnel: Peter Ritchie and Craig Millar

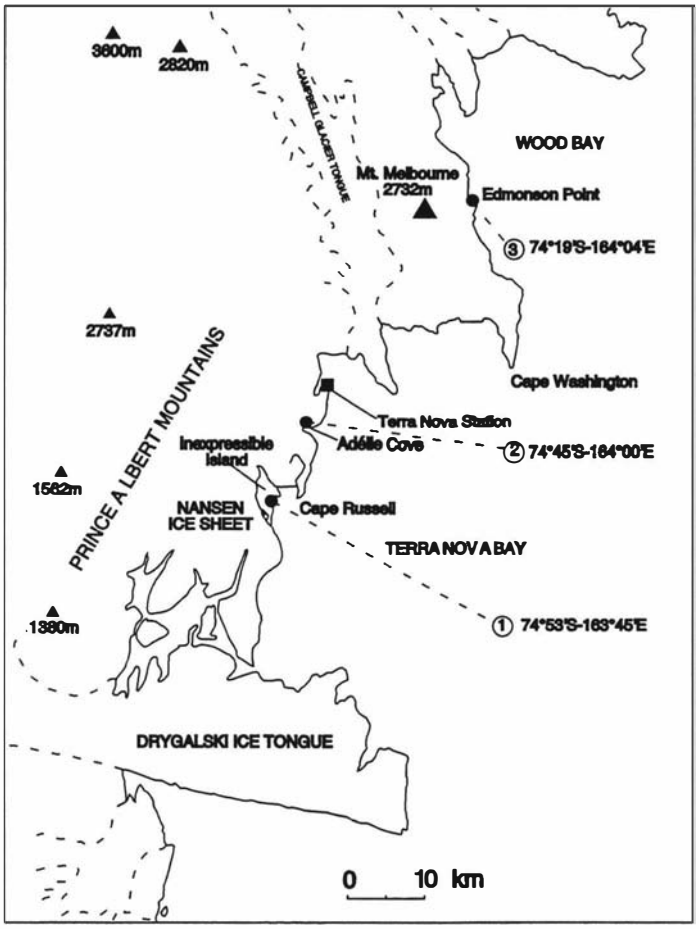
Date	Activity
9-12 -98	Arrived in CHCH to collect clothing kit
10-12-98	Travel to Hobart (via Melbourne). Flight NZ191 departed at 0645 hrs arriving in Hobart at 1255hrs. Stayed at Blue Hills Motel, Sandy Bay, Hobart (Tel. 61-3-62-231777, Fax 61-3-62-233995)
12-12-98	USCG Polar Sea docked at Macquarie Wharf 3 at 1000hrs
15-12-98	Boarded USCG Polar Sea at 0730hrs for 1000hrs departure. Departure was delayed until 1500hrs. 4-5 Days transit to Adelie Land Coast, Antarctica
21-12-98	Arrived at Antarctic Coast. US party serviced Automatic Weather Stations Pack-ice between Dumont D'Urville and Cape Adare substantial, therefore proceeded on NE course behind the Balleny Islands, turning South around Scott Island (67°26'S-182°24'E) and headed into the Ross Sea. Site 1 Cape Adare (71°17'S-170°14E) cancelled due to ice conditions
31-12-98	Site 2 Cape Hallett (72°19'S 170°16'E) for K024 put-in. Ice anchored in sea ice 5 miles off-shore. Pack was thick and LCVP landings unfeasible We arrived on the first helicopter flight in • 8/10 ice cover bwt Hallett and Polar Sea • Helicopter crew single pilot and avtech • Lift off at 1000hrs • Sampled blood from 30 Adelie penguins • Weather overcast with some isolated snow showers Site 2 Cape Wheatstone (72°17'S-170°14'E), At 1900hrs we flew to Wheatstone 20 miles from Polar Sea • 7/10 ice cover bwt Wheatstone and Polar Sea • Full helicopter crew (2 pilots, 1 avtech) • Small penguin colony perched next to large cliffs, potential rock falls • The size of the area restricted landing to the sea-ice in front of the colony 40 metres offshore. Accessed land by navigating through ice cracks • Sampled blood from 10 individuals • Two hours total ground time
1-1-99	Site 2 Cape Jones (73°17'S-169°13'E) - cancelled Site 3 Cape Washington (74°39'S-165°25'E) - cancelled Site 4 Cape Russell (74°54'S-163°54'E) - cancelled
2-1-99	Site 5 Franklin Island (76°05'S-168°19'E) • Launched 23 miles from site while steaming South • Full helicopter crew (2 pilots, 1 avtech) • 1330hr start • Overcast day with isolated snow showers • Sea ice only at south end of Island. • Adelie penguin colony on southern end on a small platform several metres above sea level. The rest of the island consisted of high cliffs. • Sampled blood from 20 individuals. • Two hours total ground time • Returned to Polar Sea. Status: steaming south at a position now south of Franklin Isl. Polar Sea arrived in Ross Island area, remained bwt Beaufort Island and Cape Bird overnight.
3-1-99	Arrived at Cape Bird by helicopter at 1000hrs
4-1-99	Set up centrifuge and sampling equipment. Began catching adult Adélie penguins and taking blood samples (n=15) for 'IBD Virus' testing.
5-1-99	Finished blood sampling for adult Adélie penguins (n=5) for IBD Virus testing.
6-1-99	Began excavation of C. Bird sub-fossil sites. Collected samples and photographed soil profiles
7-1-99	Excavation of C. Bird sub-fossil sites and collected samples
8-1-99	Excavation of C. Bird sub-fossil sites and collected samples

9-1-99	Excavation of C. Bird sub-fossil sites and collected samples
10-1-99	Bad weather – no helicopter operations
11-1-99	Helicopter pick-up (US Coast Guard) and move to Cape Royds. Set up centrifuge and sampling equipment using a borrowed generator from K062. Caught adult Adélie penguins and took blood samples (n=20) for 'IBD Virus' testing. Weather closed in. Helicopter pick-up at 1700hrs and return to Scott Base
12-1-99	Returned to New Zealand on Air National Guard C-130 Hercules

Season FOUR: 1999/2000
Personnel: Peter Ritchie, Craig Millar, Carlo Baroni and Peter Metcalf

Date	Activity
Cape Bird - Peter Ritchie, Craig Millar and Peter Metcalf	
29-12-99	Arrived at Scott Base on Air National Guard C-130 Hercules
3-1-00	Flew to Cape Bird
4-1-00	Sampled guano at North Cape Bird
5-1-00	Sampled guano at North Cape Bird
6-1-00	Sampled guano at North Cape Bird
7-1-00	Ritchie and Millar rtn to Scott Base
8-1-00	Survey of Southern Cape Bird
9-1-00	Survey of Middle Cape Bird
10-1-00	Metcalf rtn to Scott Base
Terra Nova Bay – Peter Ritchie, Craig Millar and Carlo Baroni	
15-1-00	Flew to TNB – Ritchie & Millar Picked up at McMurdo @ 1800hrs Refueled at Marble Pt, joined with 2nd helicopter 2:30 hrs total flight time
16-1-00	Flew to Inexpressible Island penguin colony – Ritchie & Millar Collected 60 Adélie blood samples, position northern area of colony Sample number T01-T60
17-1-00	Flew to Adélie Cove – Ritchie & Millar Collected 60 Adélie blood samples Sample numbers T61-120
18-1-00	All day close helicopter support – pilot Frank Ross – Ritchie & Millar • Inexpressible Island Adélie blood samples T121-T160 • Edmonson Pt Adélie blood samples T161-T200
19-1-00	Flew to Adélie Cove – Ritchie & Millar Collected Adélie blood samples T201-240 Cold and windy rtn to TNB at 1200hrs
20-1-00	Flew to Edmonson Pt – Ritchie & Millar Collected Adélie blood samples T241-T300
21-1-00	At TNB station
22-1-00	Flew to Inexpressible Island – Baroni, Ritchie & Millar Sampled subfossil Adélie bones from abandoned rookeries on Seaview beach Sampled subfossil Adélie bones abandoned rookeries on Southern Promontory
23-1-00	At TNB station
24-1-00	At TNB station
25-1-00	Dropped off by helicopter on Northern foothills (between TNB and Adélie Cove) – Baroni, Ritchie & Millar Walked north returning to TNB and sampling abandoned rookeries, including Camp Icardo.
26-1-00	Flew to Inexpressible Island – Baroni, Ritchie & Millar Sampled abandoned Adélie rookeries on Seaview beach

	Sampled abandoned Adélie rookeries on Southern Promontory OVERNIGHT in 3 person shelter ('Mellon') South of Seaview beach, approximately 60min walk
27-1-00	Baroni, Ritchie & Millar Sampled abandoned Adélie rookeries on Seaview beach Sampled abandoned Adélie rookeries on Southern Promontory RTN to TNB
28-1-00	Flew South to Cape Hickey and Cape Day - Baroni, Ritchie & Millar Close helicopter support Survey and photographed
29-1-00	Flew South to Prior Island and Cape Irizar - Baroni, Ritchie & Millar Close helicopter support Surveyed and photographed Rtn to TNB and flew to Gondawa Station (disused Germany base) Rtn to TNB and sampled abandoned Adélie behind the base
30-1-00	Flew to Edmonson Pt - Baroni, Ritchie & Millar Sampled abandoned rookeries
31-1-00	En-route to Scott Base (x2 helicopters) Landed to surveyed and photographed sites: Cape Ross, Cape Roberts, and Spike Cape. Re-fueled at Marble Pt, then off loaded at Scott Base
1-2-00	Rtn to New Zealand on USAF C-141 StarLifter



Appendix B – Sequence Data and Analyses

This thesis follows the standardised mitochondrial gene names of the Organelle Genome Database (Korab-Laskowska, M., Rioux, P., Brossard, N., Littlejohn, T.G., Gray, M.W., Lang, B.F., and Burger, G. 1998. The Organelle Genome Database Project (GOBASE). Nucleic Acids Res. 26: 139-144.) and are listed in Table B.1. and shown in Figure B.1.

Table B.1. The mitochondrial gene nomenclature follows that of Korab-Laskowska et al. (1998), the product of each gene and the strand (H or L) that encodes these products. The transfer RNAs presented are those of the chicken.

Gene	Product Name	Coding Strand
atp6	ATP synthase F0 subunit 6	H
atp8	ATP synthase F0 subunit 8	H
cob	apocytochrome b	H
cox1	cytochrome c oxidase subunit 1	H
cox2	cytochrome c oxidase subunit 2	H
cox3	cytochrome c oxidase subunit 3	H
nad1	NADH dehydrogenase subunit 1	H
nad2	NADH dehydrogenase subunit 2	H
nad3	NADH dehydrogenase subunit 3	H
nad4	NADH dehydrogenase subunit 4	H
nad4L	NADH dehydrogenase subunit 4L	H
nad5	NADH dehydrogenase subunit 5	H
nad6	NADH dehydrogenase subunit 6	L
rnl	large subunit ribosomal RNA	H
rns	small subunit ribosomal RNA	H
trnA (ugc)	transfer RNA Alanine	L
trnC (gca)	transfer RNA Cysteine	L
trnD (guc)	transfer RNA Aspartic Acid	H
trnE (uuc)	transfer RNA Glutamic Acid	L
trnF (gaa)	transfer RNA Phenylalanine	H
trnG (ucc)	transfer RNA Glycine	H
trnH (gug)	transfer RNA Histidine	H
trnI (gau)	transfer RNA Isoleucine	H
trnK (uuu)	transfer RNA Lysine	H
trnL (uar)	transfer RNA Leucine	H
trnM (cau)	transfer RNA Methionine	H
trnN (guu)	transfer RNA Asparagine	L
trnP (ugg)	transfer RNA Proline	L
trnQ (uug)	transfer RNA Glutamine	L
trnR (ucg)	transfer RNA Arginine	H
trnS (ncu)	transfer RNA Serine	H
trnT (ugu)	transfer RNA Threonine	H
trnV (uac)	transfer RNA Valine	H
trnW (uca)	transfer RNA Tryptophan	H
trnY (gua)	transfer RNA Tyrosine	L

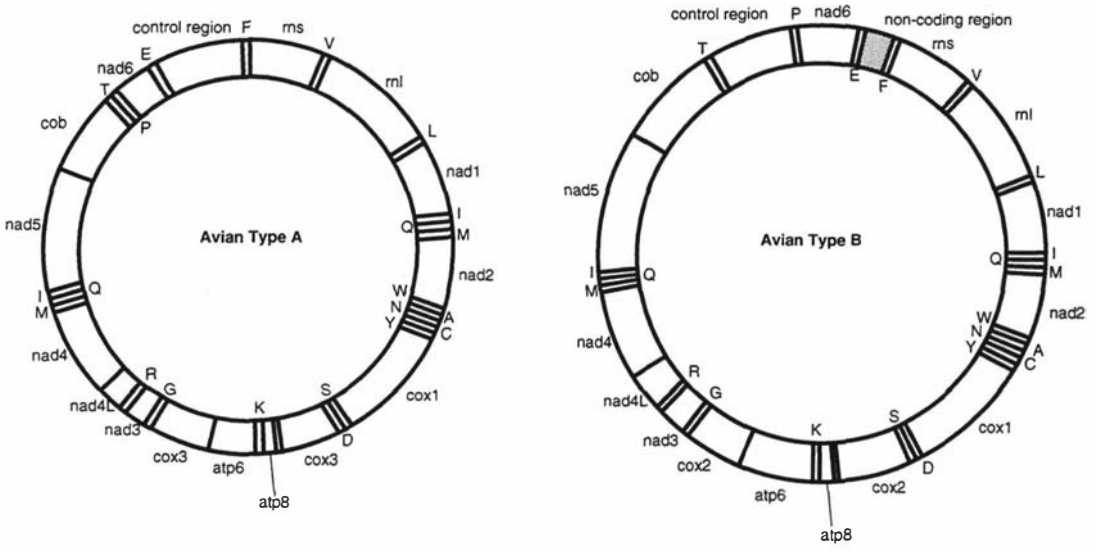
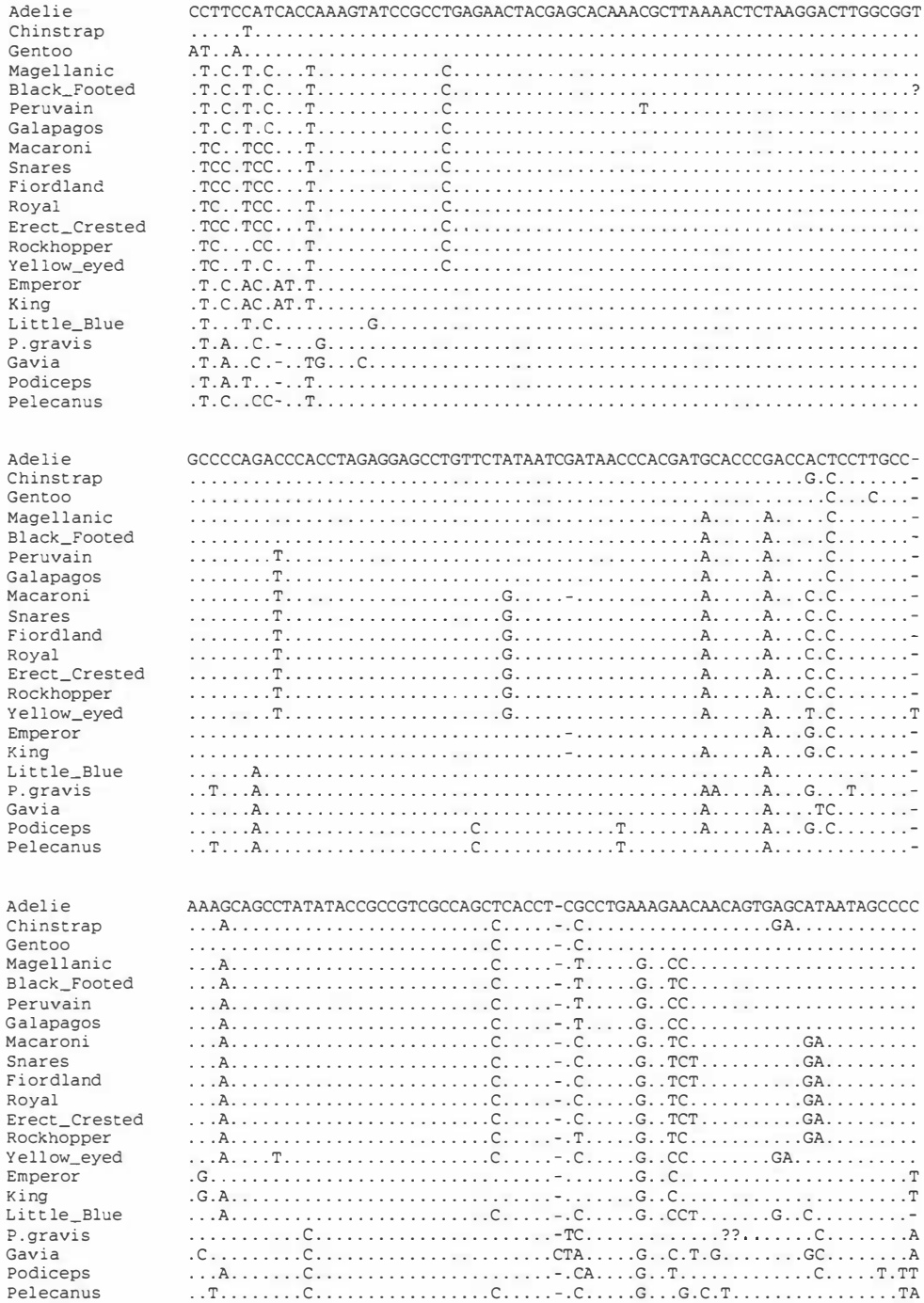


Figure B.1. The two types of mitochondrial genome organisation found in birds. The length of this molecule is approximately 16.5 kilo bases, comprising 13 protein-coding genes, 2 rRNA single-coding genes, 22 transfer RNA genes, a control region and in some species a single non-coding (type B). Gene names and their products are listed in Table B.1.

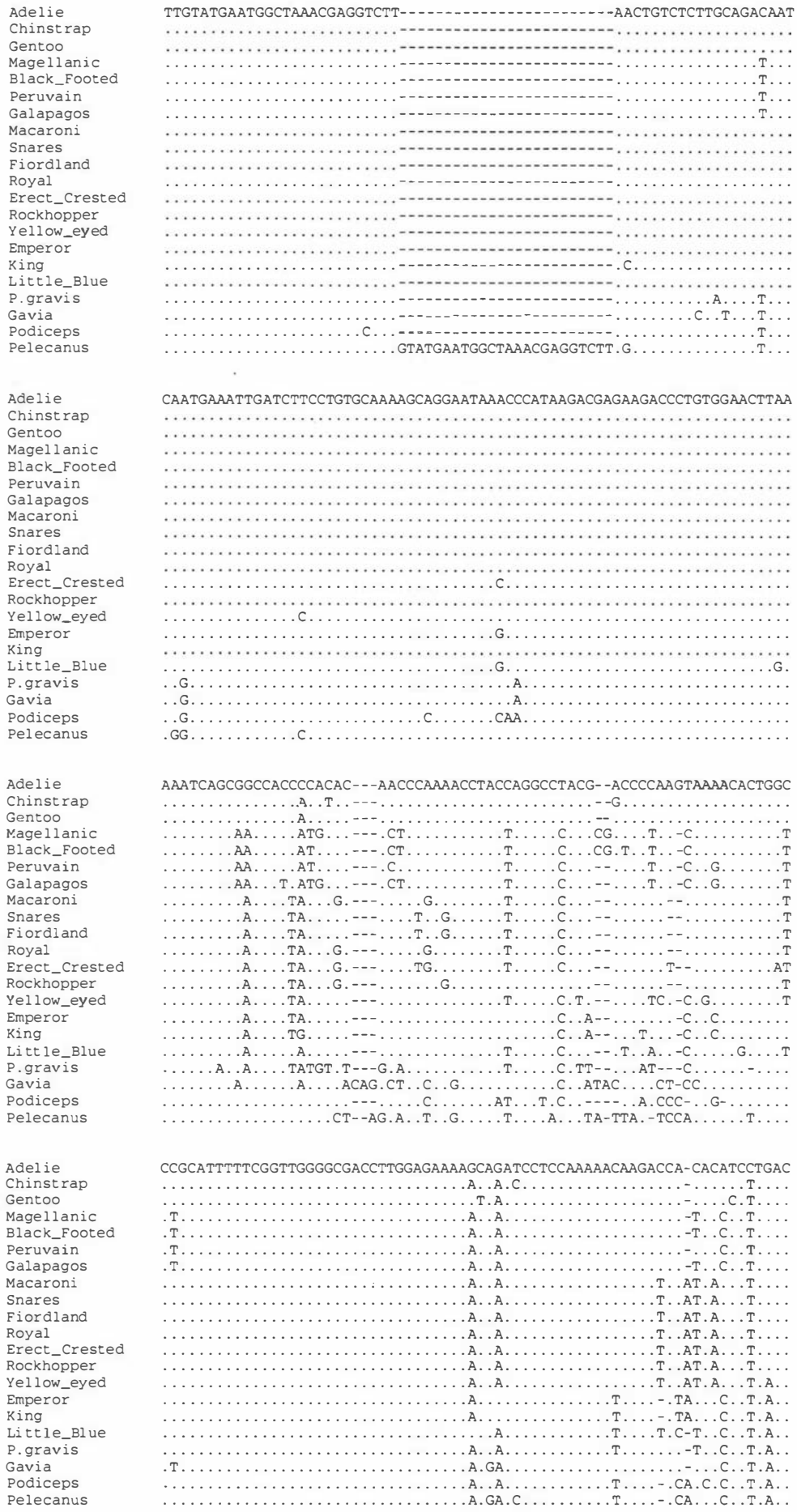
Table B.2. The penguin species from Chapter Four and the location from where tissue or blood samples were collected. The asterisk indicate specimens that where provide from the collection of Dr Allan Baker at the Royal Ontario Museum (ROM), Canada. All specimens where identified to the species level.

Species	Collection Locality	Collector
Pygoscelis adeliae	Cape Bird, Antarctica	P. Ritchie
P. antarctica	Ardley – King George Island	C. Bradshaw & J. González
P. papua	Ardley – King George Island	C. Bradshaw & J. González
Eudyptes chrysolophus	Unknown	ROM*
E. schlegeli	Macquarie Island	C. Hull
E. pachyrhynchus	Dusky Sound, NZ	?
E. robustus	The Snares, NZ	ROM*
E. sclateri	Antipodes Island	G. Elliot & K. Walker
E. chrysocome	Campbell Island	P. Moore
Megadyptes antipodes	Otago Peninsula	J. Darby
Eudyptula minor	Auckland Museum	Unknown
Spheniscus demersus	Unknown	ROM*
S. magellanicus	Punta Tombo, Argentina	ROM*
S. humboldti	Pan de Azúcar Island, Northern Chile	B. Culik
S. mendiculus	Bartolome and Fernandina, Galapagos Islands	ROM*
Aptenodytes patagonicus	South Georgia, Antarctica	ROM*
A. forsteri	Cape Crozier, Antarctica	G. Kooyman

Figure B.2. The *rns* and *rnl* sequence data from Chapter Four.



```
Adelie          AA---CCCCGCTAACAAGACAGGTCAAGGTATAGCCTATGGGGTGGAAGAAATGGGCTACATTTTCTAAA
Chinstrap       ..---.....................................C.....................G
Gentoo          ..---.....G..........G.........C.....................C.......G
Magellanic      .C---AT......G.....................A.....................G
Black_Footed    .C---AT......G.....................A.....................G
Peruvain        .C---AT......G.....................A.....................G
Galapagos       .C---AT......G.....................T.....A.....................G
Macaroni        .C---...GT...G.....................C....A.....................?
Snares          .C---.-T.....G.....................C.....................G
Fiordland       .C---.-T.....G.....................C.....................G
Royal           .C---.-......G.....................C.....................G
Erect_Crested   .C---.-......G.....................C.....................G
Rockhopper      .C---.-T.....G.....................C.....................G
Yellow_eyed     .C---...G-...G.....................................................G
Emperor         .G---..A......T.....................C.....................
King            .G---.........T.....................C.....................
Little_Blue     ------T......G.....................A.....................G
P.gravis        -CAGT-A.....................C..A.A.C.....................G.
Gavia           .CAGA.AT.....................C.....AC.....................G.
Podiceps        ----AA...A...G.........A...........AC......C......G.............A.....
Pelecanus       -----..A.....G.....................AA.........G.....................

Adelie          ATAGATAACCT-CACGGAAAGGGACATGAAAT-TGCCCCTGGAAGGCGGATTTAGCAGTAAGGTGGGACA
Chinstrap       .........T.-.......G.............-......C.............................
Gentoo          .....C...T.-.......G.............-......C.....................A...A..T.
Magellanic      .....C..TT.-.......G...........C-C.T....AA.....................A.C....T.
Black_Footed    .....C..TT.-.......G...........C-C.T....AA.....................A.C....T.
Peruvain        .....C..TT.-.......G...........C-C.T....AA.....................A.C....T.
Galapagos       .....C..TT.-.......G...........C-C.T....AA.....................A.CA...T.
Macaroni        .....C...T.-.....................C-CAT...CA.....................A......T.
Snares          .....C...T.-.....................C-C.T...CA.....................A.C....T.
Fiordland       .....C...T.-.....................C-C.T...CA.....................A.C....T.
Royal           .....C...T.-.....................C-CAT...CA.....................A......T.
Erect_Crested   .....C...T.-.....................C-C.T...CA.....................A.C....T.
Rockhopper      .....C...T.-.....................C-C.T...CA.....................A.C....T.
Yellow_eyed     .....C...T.T.....................-C.T...CA.....................A.A....T.
Emperor         G....C..TTC-.......G.............-C.T....A.....A.............A.C......
King            .....C..T.C-....A......G........-C.T....A.....A.............A.C......
Little_Blue     .......CTT.-T...A..G............TC.T....A.....................A.C....T.
P.gravis        .....A..TT--T...A......G.......-C.A.....A.....................A.A....T.
Gavia           .....A..----....A......GTG.....-C.A.........................A.C......
Podiceps        .....AG.A.--A...........TG.....ACAAT....A.....................A.C..A...
Pelecanus       T....A....C-....A..GA..G.......-C.C....C.....................A.A......

Adelie          ATAAAG-CCCACTTTAAGCCGGCTCTGA-GGTACGTACATACCGCCGGAGGTGATGCCTGCCCAGTGACA
Chinstrap       ....G.-............T...C....-..?.............A.....................
Gentoo          ......-............T...C....-..C.............A.....................
Magellanic      ......-...C......A.T...C....-..C.............A.....................
Black_Footed    ......-...C......A.T...C....T..C.............A.................?...
Peruvain        ......-...C......A.T...C....-..C.............A.....................
Galapagos       ......-...G......A.T...C....-..C.............A.....................
Macaroni        ......-...C......A.....C....-..C.............A.....................
Snares          ......-...C......A.....C....-..C.............A.....................
Fiordland       ......-...C......A.....C....-..C.............A.....................
Royal           ......-...C......A.....C....-..C.............A.....................
Erect_Crested   ......-...C......A.....C....-..C.............A.....................
Rockhopper      ......-...C......A.....C....-..C.............A.....................
Yellow_eyed     ...G..-...T......A.....C....-..C.............A.....................
Emperor         ...GT.-...C........T...C....-..C.............A.....................
King            ....T.-...C............C....-..C.............A.....................
Little_Blue     ......-...T......A.....C....-..C.............A.....................
P.gravis        ......G...T........T...C...G-A.C.............A.....................
Gavia           ......-...C.......TT...C...G-..C.............A.....................C
Podiceps        ...C..-T..C.....T...C...G-..C.............A.....................
Pelecanus       ...G..-...C......A.T.......G-A.C.............A.....................

Adelie          CT-CTGTTCAACGGCCGCGGTATCCTAACCGTGCGAAGGTAGCGCAATCAATTGTCCCATAAATCGAGAC
Chinstrap       ..-.....................................................................
Gentoo          ..-.C...................................................................
Magellanic      ..-.....T...............................................................
Black_Footed    ..-.....T...............................................................
Peruvain        ..-.....T...............................................................
Galapagos       ..-.....T...............................................................
Macaroni        .C-.A...T...............................................................
Snares          .C-.A...T...............................................................
Fiordland       .C-.A...T...............................................................
Royal           .C-.A...T...............................................................
Erect_Crested   .C-.A...T...............................................................
Rockhopper      .C-.A...T...............................................................
Yellow_eyed     .C-.A...T...............................................................
Emperor         .C-A....................................................................
King            .C-A....................................................................
Little_Blue     ..T.....T...............................................................
P.gravis        TC-A....T...............................................................
Gavia           .C-AC...T...............................................................
Podiceps        T--AC...T...............................................................
Pelecanus       TC-A....T...............................................................
```


Adelie TTGTATGAATGGCTAAACGAGGTCTT----------AACTGTCTCTTGCAGACAAT
Chinstrap
Gentoo
Magellanic
Black_Footed
Peruvain
Galapagos
Macaroni
Snares
Fiordland
Royal
Erect_Crested
Rockhopper
Yellow_eyed
Emperor
King
Little_Blue
P.gravis
Gavia
Podiceps
Pelecanus
Adelie CAATGAAATTGATCTTCCTGTGCAAAAGCAGGAATAAACCCATAAGACGAGAAGACCCTGTGGAACTTAA
Chinstrap
Gentoo
Magellanic
Black_Footed
Peruvain
Galapagos
Macaroni
Snares
Fiordland
Royal
Erect_Crested
Rockhopper
Yellow_eyed
Emperor
King
Little_Blue
P.gravis
Gavia
Podiceps
Pelecanus
Adelie AAATCAGCGGCCACCCCACAC---AACCCAAAACCTACCAGGCCTACG--ACCCCAAGTAAAACACTGGC
Chinstrap
Gentoo
Magellanic
Black_Footed
Peruvain
Galapagos
Macaroni
Snares
Fiordland
Royal
Erect_Crested
Rockhopper
Yellow_eyed
Emperor
King
Little_Blue
P.gravis
Gavia
Podiceps
Pelecanus
Adelie CCGCATTTTTCGGTTGGGGCGACCTTGGAGAAAAGCAGATCCTCCAAAAACAAGACCA-CACATCCTGAC
Chinstrap
Gentoo
Magellanic
Black_Footed
Peruvain
Galapagos
Macaroni
Snares
Fiordland
Royal
Erect_Crested
Rockhopper
Yellow_eyed
Emperor
King
Little_Blue
P.gravis
Gavia
Podiceps
Pelecanus

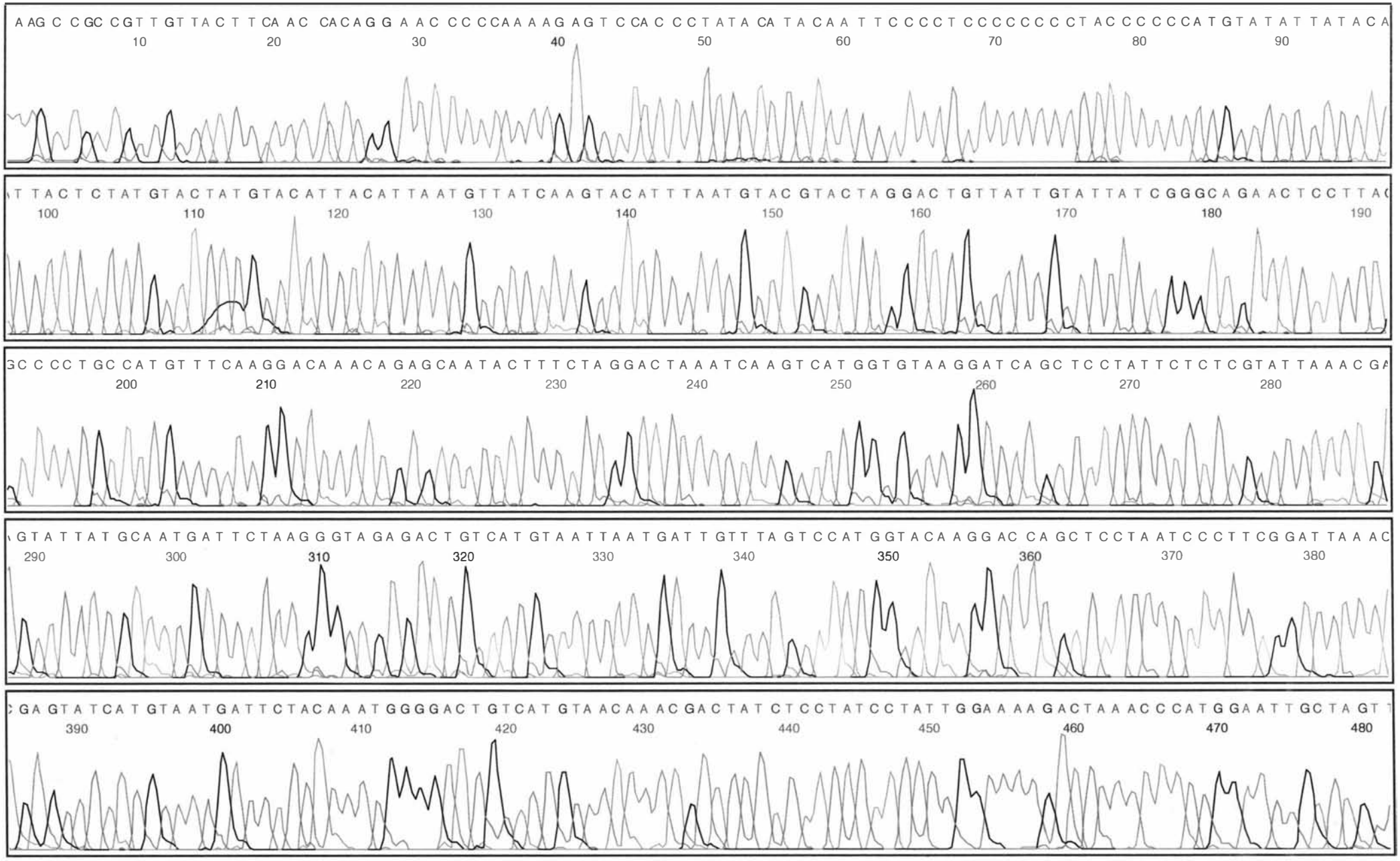
```
Adelie          CAAGAACAACCCCTCAACGTACCAACAGCAACCAGACCCAATATAATT--GATCAATGGACCAAGCTACC
Chinstrap       .....G.....T.........T.....................--..C.................
Gentoo          .....G.................T.....T.....................--..C.................
Magellanic      .....G.........G......T.....................--..C.................
Black_Footed    .....G.........G......T.....................--..C.................
Peruvain        .....G.........G......T.....................--..C.................
Galapagos       .....G.........G......T.....................--..C.................
Macaroni        ....................G.T.....................--....................
Snares          ....................G.T.....................--....................
Fiordland       ...........T........G.T.....................--....................
Royal           ....................G.T.....................--....................
Erect_Crested   ....................G.T.....................--..C.................
Rockhopper      ....................G.T.....................--....................
Yellow_eyed     ..........T..........T.....T.....................--..C.................
Emperor         .....G.G..A...........T.....T.....................--..C.................
King            .....G.G..A...........T..T..T.....................--..C.................
Little_Blue     .....GT.............G.T.....................C....--..C.................
P.gravis        .....G....A...........T..T..T.............C....--T...................
Gavia           .....G....A....G......T.....T.C.........GC.C..G.CT..C.................
Podiceps        .G..G.....TA.C...A....T.....T.............C....--...T.................
Pelecanus       .....G.................T.....T....C...............--..C.................

Adelie          CCAGGGATAACAGCGCAATCCCCTCCAAGAGCCCATATCGACAAGGGGGTTTACGACCTCGATGTTGGAT
Chinstrap       ............................................................................
Gentoo          ............................................................................
Magellanic      ....................T.....................G...A.....................
Black_Footed    ....................T.....................G...A.....................
Peruvain        ..........................................G.........................
Galapagos       ....................T.....................G...A.....................
Macaroni        ............................G.......G.....................................
Snares          ............................G.......G.....................................
Fiordland       ............................G.......G.....................................
Royal           ............................G.......G.....................................
Erect_Crested   ............................G.......G.....................................
Rockhopper      ............................G.......G.....................................
Yellow_eyed     ............................G.....................................
Emperor         ..................................T.........G.....................
King            ..........................................G.....................
Little_Blue     ..........................................G.....................
P.gravis        ....................T.....................G...A.....................
Gavia           ....................T...T.....................A.....................
Podiceps        ..........................................G.....................
Pelecanus       ......G.............T...T.....................A.....................

Adelie          CAGGACACCCTAATGGTGCAGCCGCTATTAAGGGCTCGTTTGTTCAACGATTAACAGTCCTACGTGATCT
Chinstrap       ..................................T.....................................
Gentoo          ..................................T.....................................
Magellanic      .......T..........................T.....................................
Black_Footed    .......T..........................T.....................................
Peruvain        .......T..........................T..A..................................
Galapagos       .......T..........................T..A..................................
Macaroni        .......T..........................T.....................................
Snares          .......T..........................T.....................................
Fiordland       .......T..........................T.....................................
Royal           .......T..........................T.....................................
Erect_Crested   .......T..........................T.....................................
Rockhopper      .......T..........................T.....................................
Yellow_eyed     .......T..........................T.....................................
Emperor         .......T..........................T...................T.................
King            .......T..........................T.....................................
Little_Blue     .......T..........................T.....................................
P.gravis        .......T....G..............C......T...................T.................
Gavia           .......T..........................T.....................................
Podiceps        .......T..........................T...................T.................
Pelecanus       .......T..........................T.....................................

Adelie          GAGTT
Chinstrap       .....
Gentoo          .....
Magellanic      .....
Black_Footed    .....
Peruvain        .....
Galapagos       .....
Macaroni        .....
Snares          .....
Fiordland       .....
Royal           .....
Erect_Crested   .....
Rockhopper      .....
Yellow_eyed     .....
Emperor         .....
King            .....
Little_Blue     .....
P.gravis        .....
Gavia           .....
Podiceps        .....
Pelecanus       .....
```

Figure B.3. The HVR-1 sequence data for penguins analysed in Chapter Four. Presented are the portion of sequences that were alignable and therefore analysed.

```
Adelie        GGTCCTGAAGCTAGTAACGTAGGATCTTCCATACGTCGAGTTGCTGATTTCACGTGAGGAGACCGATTAA
Gentoo        ..........................................................C....CC..
Chinstrap     ..........................................................C....C...
Magellanic    ..G.....C.......G.A............C.........................A..C.AA.....
Royal         ..G.........G..CG.A.G.......T.G....C..G...................A..T.AA.....
Fiordland     ..G............CG.A?G.......T......C..G...................A..C.AA.....
Snares        ..G............CG.A.G.......T......C..G...................A..C.A......
Rockhopper    ..G............CG.ACG.......T......C..G...................A..T.AA..C..
Erect         ..G............CG.A.G.......T......C..G...................A..T.AA.....
Yelloweyed    ..G.........G..CGTA.........T......C..G...................A..T.AA.....
Blackfoot     ..?.....C.....?.?.ACG.......T.....??.?.??.................A..C.AA.....
Macaroni      ..G.........G..CG.A.G.......T.G....C..G...................A..T.AA..C..
Humbolt       ..G.....C.......G.AC............C.........................A..CTAA.....
Galapagos     ..G.....C.......G.AC............C.........................A..CTAA.....
LittleBlue    ..G.............G.A.G.......T......C..G...................A..T.AA..C..

Adelie        TAAATAACCTGGTCCCTGAAGCTAGCGCCCGGAGAA-TGGTTGAATGTTGGGCTTTTACTCCATGAAGGC
Gentoo        ....................A.....A.........-.........A.C..A..C.C.......T.....
Chinstrap     ....................................-.........A.C..A.ACAC.......T.....
Magellanic    ..G.......AA.......C...G..A...A-....A.TA...G..AA.AA..GCA.T......T..A..
Royal         ..G......CA..T......A..G......C.....A..A...G..AC.AA..GC.CC.A....TG..TT
Fiordland     ..G......CA..T......A..G......C.....A..A...G..AC.AA..GCA.C.A....TG..TT
Snares        ..G......CA..T......A..G......C....GA..A...G..A..AA..GCA.C.A....TG..TT
Rockhopper    ..G......CA..T......A..G......C.....A..A...G..A..AA..GC.CC.A....TG..TT
Erect         ..G......CA..T......A..G......C....GA..A...G..A..AA..GCA.C.A....TG..TT
Yelloweyed    ..G......CA..T......A..G..A.T.C....GA..A...G..A..AA..GC..C.A....TG..TT
Blackfoot     .?G?......AA.......C...G..A...A-....G.TA...G..AA.AA..ACA.T?.....T..AA.
Macaroni      ..G......CA..T......A..G......C.....A..A...G..AC.AA..GCACC.A....TG..CT
Humbolt       ..G.......AA.......CA..G..A.-.C.....G.TA...G..AA.AA..GCA.T......T..AA.
Galapagos     ..G.......AA.......CA..G..A.-.C.....A.TA...G..AA.AA..GCA.T......T..AA.
LittleBlue    ..G.......AAGG.........G......C.....G.TA...G..A..AA..GC-.C......T...AT

Adelie        CAAGTGTTCAAGCACTGCCGTACCTACTTCTAAGGGAA-CATGTACAGAACTAGCAGTTCCATGGGTTTA
Gentoo        ..G..AC..............T.CC.GCT..C..T.CT..A.CG....G......-..........C..
Chinstrap     ....C.CCT..............T.-CT..CA...G-..A.C.....G....-..A..T......C..
Magellanic    .....AGCA.G....AC.....T.ATT-C..GCT--.GAAGGA......TTATCAGAA..-.....C...
Royal         ..G.CAAC.......G.T....TACT.-.G...T--G.TA..AC...T.CTATCAC.A..-......C..
Fiordland     .....AGC...A...G......TACT.-.G...T--..CA..AC...T.TAATTACAA..-.........
Snares        ..G..AA....A...G.T....TACGT-.A...T--G.TA..AC...T.CAGTTATAAC.-.........
Rockhopper    ......AG.......A.......ACGT-.AA.GC--.GCAG.AC...C.--GTTACAA.TT.........
Erect         ....CAAC.......A.........T.-.A.GGC--.GCAGCAC...C.CAGTTA-AAG.T.........
Yelloweyed    ..G..AAC.......A.......TAT.-CAC.GC--.GAG..AC...A.CGATA.CAAC.T.........
Blackfoot     ..?.CA?.A..A...GC...?.T.ATT-C..GCT--.GAAGGA......CTATCA.AA..-....A....
Macaroni      ..G..AAC.......G.T....TACT.-.G...T--..TA..AC...T.TTATCACAA..-......C..
Humbolt       .....AGCAG....TAT.....T.AT.-....CT--..TAGGTC.....TTACCA-AAG...........
Galapagos     ...A.AACAG....TAC.....T.AT.-....CT--..CAGATC.....TTATCA-AGG.T.........
LittleBlue    .....AGC..GA..TAT.....T....-CAC.GATA.GCA...C...C.C.GTTA.AAG.-.....C...

Adelie        GTCTTT-TCCAATAGGATAGGAGATAGTCATTTATTACATGACAGTCCCCGTTTGTAGAATCATTACATA
Gentoo        .....C-.....G-.A..............G.G..........AT.TTA.CC-....C-.......G.
Chinstrap     ......C.....G-AA.C........----------------------------------------------
Magellanic    ...C.---....CG...C..A.....A.....A-----------------------------------
Royal         ...C..-..T.CCG.A.C........A..G..C-----------------------------------
Fiordland     ...C.C-..T.T.G...C....A.C.A.T.CC------------------------------------
Snares        ...C.C-..T.T.G...C..A...C.A.T.CC------------------------------------
Rockhopper    ...C.--..T.T.G...C..A.A.C.ACT.C.------------------------------------
Erect         ...C.?-....T.G.A.C..A.A...A.T.C.------------------------------------
Yelloweyed    ...C..-.....CG.A.C..A.....A.....------------------------------------
Blackfoot     .C.C.--..A.C-G...C..A.....A....C------------------------------------
Macaroni      ...C..-..T.CCG.A.C........A..G..------------------------------------
Humbolt       ...C---..T...G.A.C..A.....A.....A----------------------------------
Galapagos     ...C---..T...G.A.C..A.....A.....A----------------------------------
LittleBlue    ...C.C-.-T..CG...C..AGAGC.A.T.CCA----------------------------------

Adelie        ATATCCGTTTAATCCGAAGGGATTAGGAGTTGGTCCTTGTACCATGGACTAAATAATCATTAATTACATG
Gentoo        GCC.T......G..TT.G.T.TC......C.AA.....A.......ACT..GGC.G......GC......
Chinstrap     ----------------------------------------------------------.........C.....A
Magellanic    ----------------------------------------------------------.-------.ACC...A
Royal         ----------------------------------------------------------------.GC....A
Fiordland     ----------------------------------------------------------.---------A.....
Snares        ----------------------------------------------------------.---------A.....
Rockhopper    ----------------------------------------------------------.---------.GC....A
Erect         ----------------------------------------------------------.---------.GC....A
Yelloweyed    ----------------------------------------------------------.---------.GC....A
Blackfoot     ----------------------------------------------------A--------CAAC...A
Macaroni      ----------------------------------------------------C---------.GC.....
Humbolt       ----------------------------------------------------------.---------.A.C....
Galapagos     ----------------------------------------------------------.---------.A.C...A
LittleBlue    ----------------------------------------------------------.-------.G.T...A
```

```
Adelie          ACAGTCTTTGTCCTTAGAATCATTGCATAATACTC-GTTTAATAC-GAGAGGAT-AGAAGCTGATCCTTA
Gentoo          G...ATC..A....-...-........G.G.T.C.-.....GCC.T.....TT.T..G.....G......
Chinstrap       ....GTC..A....-...-........G.G...C.-.....G.C.--TA.....T....A...G......
Magellanic      ...TAA.AAT...AC.AGT........C.GAC...-.........GA.A.AAGAACTG.CT.A....C..
Royal           ...TAGCA.T....C.-G.C....A...CTC..C.-.....G.C.-A.AGTATAGC--.CT.A.......
Fiordland       ....AA.ACT..AC..A.C-....A...TCC..C.-.....G.C.---A...TAT..T.CT.A.......
Snares          ....GA.ACT.TAC..A.C-....A..CTCC..C.-.....G.C.---AGA.TAT..T.CT.A.......
Rockhopper      ...TAA.ACT........C-....A...TCC..C.-.....GCC.--.AG..G.A-.G.CT.A.......
Erect           ...TGA.ACT......A.C-........TCC....-.....G.C.---AG.ATAT..T.C..A.......
Yelloweyed      ...TA.CAAT......A.C-...C....T.A....-.....G.-.AAT.TAATAG--G....A.....C.
Blackfoot       ...TGA.AAT.?.CC.A.T........C..AC.C.-.?,...C..GA.A..AGAATT..CT.A....CC.
Macaroni        ...TAGCA.T....C...C-....A...CTC..C.-.....G.-.CA..GTATAG--C.CT.A.......
Humbolt         ...CAG.AAT...AC...T.........CGA..C.C.........GA.A.AAGAATTG.T..A....CC.
Galapagos       ...TAA.AAT...AC...T.........C.G..C.C.........GA.A.AAGAATT..T..A.....C.
LittleBlue      .....A.CAT.T.C....T...C.A.....C.T..-.....G...G--A.AATAGTA.....A....A..


Adelie          CACCATGACTTGATTTAGTCCTAGAAAGTATTGCTCTGTTTGCCCTCGAAACGTAG-CAGGGGC-CAAGG
Gentoo          ...........A.A.A.-......GT..C...A.....C.G.T...T....G?AT.A......AT.G.?A
Chinstrap       ............A..........T..C.C....T.A.C.....CT.TGGTA..A-...A..A--C...
Magellanic      TT....T...CAG..A.-....T.........A...CA..AAT...T.....AC..-...TTC.-....A
Royal           ..T...T.T..AG..A.-....C.........A....AC..AT...T.....AC..-...CAAG-....A
Fiordland       T.....T....AG..A.-....C.........A....AC..AT..CT.....AC..-...TAAG-....A
Snares          ......T....AG..A.-....C.........A....AC..AT..CT.....AC..-...TAAG-....A
Rockhopper      ......T....AG..A.-....C.........A....A.C..T...T.....AC..-...TCAG-....A
Erect           ...T..T..C.A...A.-....T.........A....A...AT..CT.....AC..-...TAAG-....A
Yelloweyed      T.....T...CAG..A.-....T.........A...CAA..AT..CT.....A...-...T..A-G..AC
Blackfoot       TT....T...CA...A.-....T.........A...CA..AAT...T.....AC..-...?TCT-....A
Macaroni        T.....T.TC.AG..A.-....C.........A....AC..AT...T.....AC..-...TAAG-....A
Humbolt         TT....T....AG..A.-....T.........A...CA..AAT...T.....AC.A-...CTCT-T...A
Galapagos       TT....T....AG..A.-....T.........A...CA..AAT...T.....AC.A-...CTCT-T...A
LittleBlue      ..T...T....AG..A.-...CT.........A....A.CAAT.TCTA..G.A..T-TC.TATA-.G..A


Adelie          AG--TTCTGTCCGATAGTACAATAACA--
Gentoo          .TGA....TG..C.C...G..G...T.--
Chinstrap       .AAGA...A...-.G.............--
Magellanic      GT-ACCTCA...A..TA.T.T.G.C..--
Royal           .C-A....AC..A..TAC...T..T..--
Fiordland       .C-A....AC..A..TAC...GA.T..--
Snares          .C-A..T.AC..A.CTAC...GA.T..--
Rockhopper      .C-A....AC..A..TAC...T.GT..-A
Erect           .C-A....AC..A..TAC...T..CT.--
Yelloweyed      CA-T...CAC..A.CCCAC..T...T.TG
Blackfoot       .T-ACC.CA...A..TA.T.T.A.C..--
Macaroni        .C-A....AC..A..TAC...C..T..--
Humbolt         .T-ACC.CA...A..TA.T.C.A.C..--
Galapagos       .T-ACC.CA...A..TA.C.C.A.C..--
LittleBlue      .C-A..---...ATGTTA..CT..CA?T?
```

Figure B.4. The following three pages are the electropherograms for the HVR-1 sequences from three penguin species: Adélie (Pa01.L-glu-t*2); Chinstrap (20*A15.L-tGlu) and Royal (23*6-R.L-tGlu). These DNA sequences show the result for possible heteroplasmy that may be associated with the termination of the D-loop in the control region.


A AG C C GC C GT T GT TA C T T CA AC CA C A G G A A C C C C CA A A A G AG T C CA C C C T A T A CA T A C A A T T C C C C T C C C C C C C C T AC C C C C C A T G T A T A T T A T A C A
10
20
30
40
50
60
70
80
90
T T A C T C T A T G T A C T A T G T A C A T T A C A T T A A T G T T A T C A A G T A C A T T T A A T G T A C G T A C T A G G A C T G T T A T T G T A T T A T C G G G C A G A A C T C C T T A C
100
110
120
130
140
150
160
170
180
190
G C C C C T G C C A T G T T T C A A G G A C A A A C A G A G C A A T A C T T T C T A G G A C T A A A T C A A G T C A T G G T G T A A G G A T C A G C T C C T A T T C T C T C G T A T T A A A C G A
200
210
220
230
240
250
260
270
280
G T A T T A T G C A A T G A T T C T A A G G G T A G A G A C T G T C A T G T A A T T A A T G A T T G T T T A G T C C A T G G T A C A A G G A C C A G C T C C T A A T C C C T T C G G A T T A A A C
290
300
310
320
330
340
350
360
370
380
G A G T A T C A T G T A A T G A T T C T A C A A A T G G G G A C T G T C A T G T A A C A A A C G A C T A T C T C C T A T C C T A T T G G A A A A G A C T A A A C C C A T G G A A T T G C T A G T
390
400
410
420
430
440
450
460
470
480

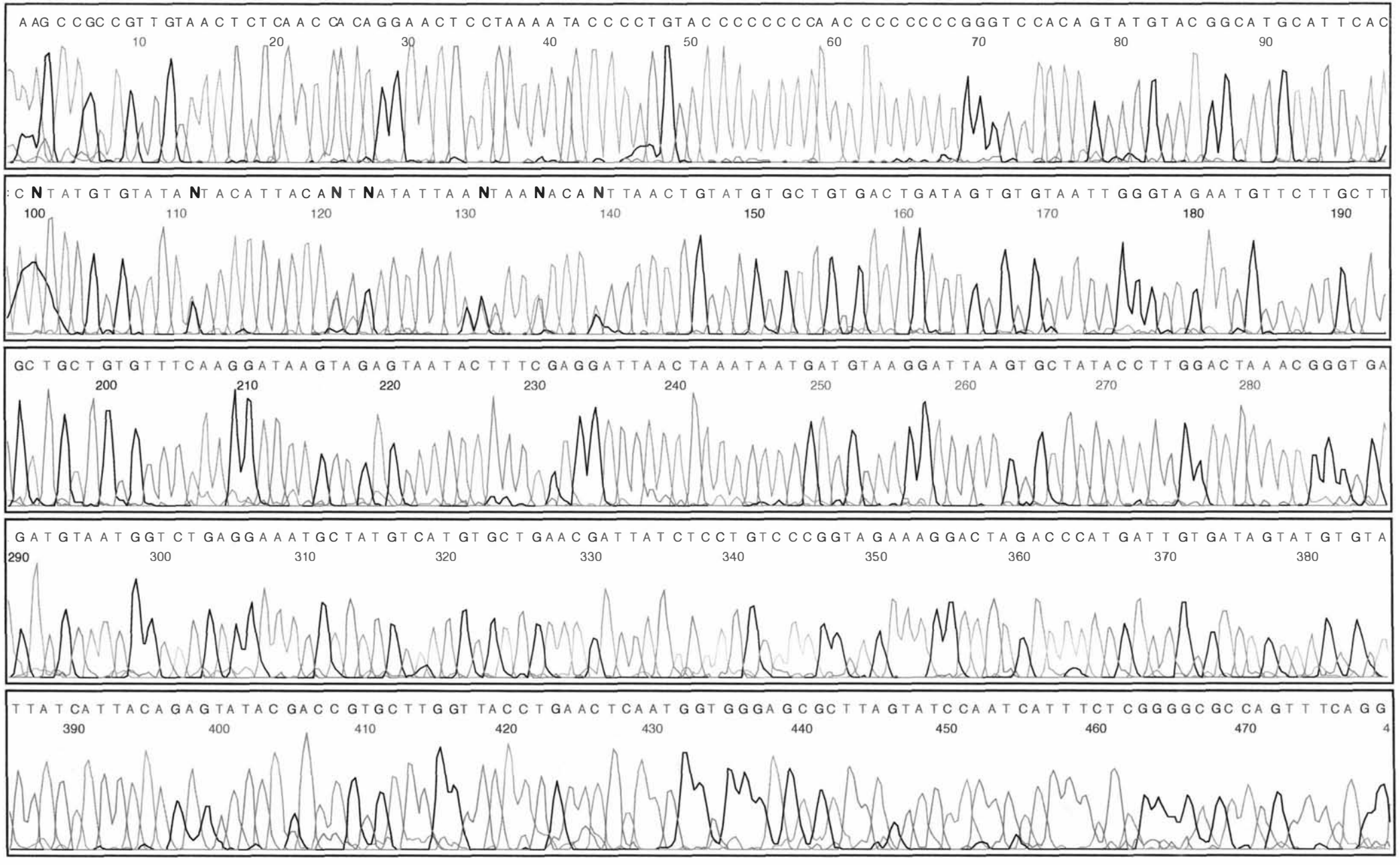
ABI PRISM™

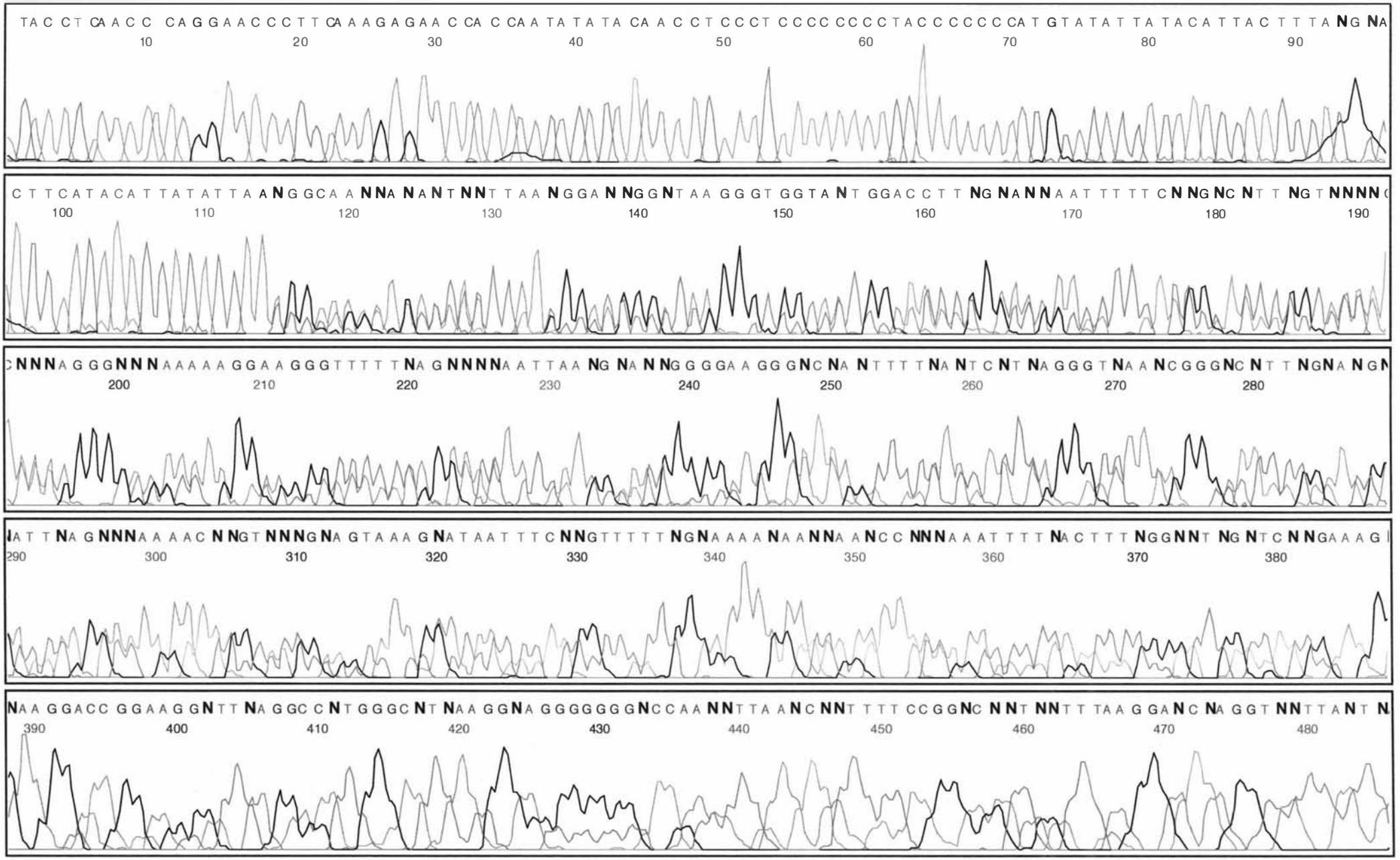
Model 377
Version 3.2

23 6-R.L-tGlu
Loaded 1ul onto the sequencer
6-R.L-tGlu
Lane 23

Signal G:138 A:108 T:79 C:105
DT {BD Set Any-Primer}
377XL 48CM Matrix
Points 1350 to 9400 Base 1: 1350

Thu, 15 Jun 2000 11:03 am
Wed, 14 Jun 2000 11:09 am
Spacing: 12.74





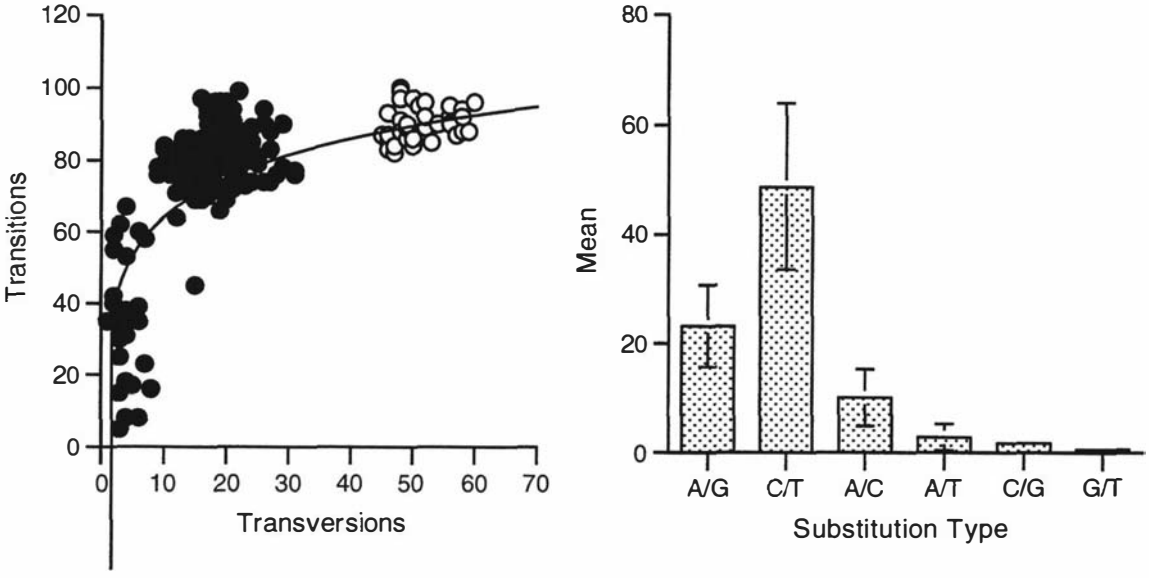


Figure B.5. The graph on the left is transversion against transitions for *cob* gene sequences, among all penguins (ingroup) comparisons (filled circles) and ingroup to outgroup comparisons (open circles) with a line of best fit. On the right is the type and number of substitutions in the (with standard deviations) from ingroup comparisons only. C/T transitions are predominant substitution in this gene. Raw data was taken from Edge (1996).

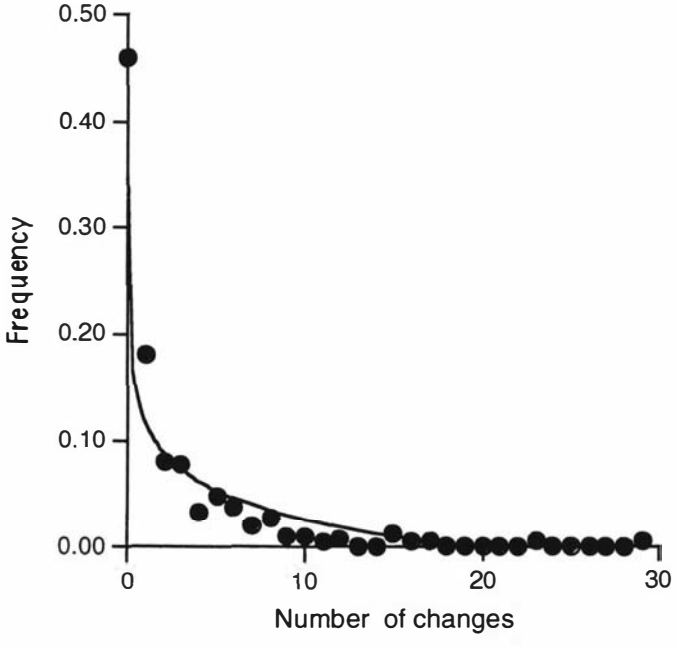


Figure B.6. The number of parsimony-type changes required for all 353 sites across a neighbor-joining tree, constructed from the 381 Adélie penguin HVR-1 sequences in Chapter Five. This distribution has a mean of 2.152975 and the variance = 12.817441. A negative binomial distribution was fitted to this frequency distribution and α approximated at 0.4366. This estimate was using as the shape parameter for a gamma (Γ) distribution in Chapter Five.

Appendix C – Radiocarbon Dating of Sub-fossil Bones

Table C.1. Radiocarbon dates of penguin remains from Victoria Land, Antarctica. 14C dates were supplied by: Geochron Lab.-Krurger Enterprise Inc., Cambridge Massachusetts (conventional and AMS, GX-); IsoTrace Radiocarbon Lab., Toronto (AMS, TO-); NOSAMS, Woods Hole Oceanographic Institution (AMS, OS-); The Institute of Geological and Nuclear Sciences, Lower Hutt, NZ (AMS, NZA-). T = sampled from the top of a horizon, B = sampled from the bottom of a horizon, and I = intermediate of the horizon. All ages were calibrated using Stuiver et al. (1998).

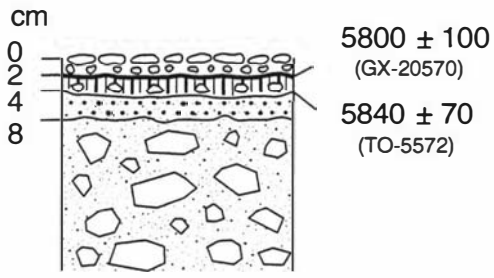
	Sample	Lat.-Long.	Code (collection/lab)	Radiocarbon Age (yr BP)	Calibrated age (yr BP)
1.	Cape Hallett, guano, site S1 (T), 2.5 m	72°19'S, 170°12'E	991231.01 OS-27676	1620 ± 45	525
2.	Cape Hallett, guano, site S1 (B), 2.5 m	72°19'S, 170°12'E	991231.04 OS-27677	1650 ± 30	541
3.	Cape Hallett, guano, site S2 (B), 3 m	72°19'S 170°12'E	991231.08 OS-27678	1700 ± 80	607
4.	Edmonson Pt., guano, site S4 (T)	74°19'S, 165°04'E	1000204.03 OS-26825	2070 ± 35	924
5.	Edmonson Pt., guano, Site S4 (B)	74°19'S, 165°04'E	1000204.04 0S-26415	2200 ± 35	1054
6.	Edmonson Pt., guano, site S2 (T)	74°19'S, 165°04'E	1000130.02 0S-27681	2230 ± 40	1086
7.	Edmonson Pt., guano, site S2 (B)	74°19'S, 165°04'E	1000130.03 0S-27682	2290 ± 35	1166
8.	Edmonson Pt., guano, site S1 (B), 40 m	74°19'S, 165°04'E	880118.02 GX-14076	2870 ± 155	1783
9.	Gondwana St., guano, site S1, 15 m	74°38'S, 164°12'E	870119.03 GX-13620	4615 ± 85	3894
10.	Gondwana St., guano, site S2 (T), 15 m	74°38'S, 164°12'E	941111.01 TO-4964	4670 ± 80	3970
11.	Gondwana St., guano, site S2 (B) 15 m	74°38'S, 164°12'E	941111.03 TO-4965	9220 ± 100	8977
12.	Terra Nova Station, guano & remains, 18 m	74°41'S, 164°06E	860209.01 GX-12760	5770 ± 60	5448
13.	Terra Nova Station, guano, 19 m	74°41'S, 164°06E	890204.106 GX-15494	4585 ± 105	3858
14.	Terra Nova Station, guano, 19 m	74°41'S, 164°06E	890204.107 GX-15495	4915 ± 105	4339
15.	Terra Nova Station, guano, (T), 33 m	74°41'S, 164°06E	941122.02 TO-5573	5100 ± 70	4543
16.	Terra Nova Station, guano, (B), 33 m	74°41'S, 164°06E	941122.05 GX-21408	5307 ± 310	4826
17.	Terra Nova Station, guano, (T), 40 m	74°41'S, 164°06E	970115.01 GX-25109am	4880 ± 40	4270
18.	Terra Nova Station, guano, (base of T), 40 m	74°41'S, 164°06E	970115.05 GX-25110am	5040 ± 50	4489
19.	Terra Nova Station, guano, (B), 40 m	74°41'S, 164°06E	970115.07 OS-27674	6480 ± 45	6201
20.	Terra Nova Station, guano, (B), 57 m	74°41'S, 164°06E	971224.03 GX-25111am	5240 ± 40	4795
21.	Icarus Camp, Northern Foothills, guano, (B), 40 m	74°42'S, 164°07'E	860113.01 GX-12754	4290 ± 50	3471
22.	Icarus Camp, Northern Foothills, guano, (B), 25 m	74°42'S, 164°07'E	860113.03 GX-12755	4495 ± 135	3727
23.	Icarus Camp, Nothern Foothills, guano, (B), 50 m	74°42'S, 164°07'E	870118.01 GX-13619	4495 ± 95	3727
24.	S. Icarus Camp, Northern Foothills, guano, S3/94	74°42'S, 164°07'E	941102.08 GX-20569	4555 ± 85	3825

	(T), 60 m				
25.	S. Icarus Camp, Northern Foothills, guano, S3/94 (T), 60 m	74°42'S, 164°07'E	941102.09 GX-21405	4080 ± 105	3260
26.	S. Icarus Camp, Northern Foothills, guano, S3/94 (B), 60 m	74°42'S, 164°07'E	941102.10 TO-4961	7360 ± 70	7210
27.	S. Icarus Camp, Northern Foothills, guano, S4/94 (T),35 m	74°42'S, 164°07'E	941102.13 GX-21406	4205 ± 150	3385
28.	Icarus Camp, Northern Foothills, guano, S5/94 (T), 50 m	74°42'S, 164°07'E	941112.01 GX-20574	4625 ± 120	3906
29.	Icarus Camp, Northern Foothills, guano, S5/94 (B), 50 m	74°42'S, 164°07'E	941112.05 TO-4966	10180 ± 150	10274 !!!
30.	Icarus Camp, Northern Foothills, guano, S5/94 (B), 50 m	74°42'S, 164°07'E	941112.05 OS-27673	6150 ± 45	5862
31.	N. Adélie Cove, guano, (B), 52 m	74°44'S, 164°06E	870120.03 GX-13621	6855 ± 195	6614
32.	N Adélie Cove, guano, Site S1/88 (B), 39 m	74°44'S, 164°06E	870120.04 GX-13622	6860 ± 110	6619
33.	N Adélie Cove, guano, Site S1/88 (B), 39 m	74°44'S, 164°06E	880211.01 GX-14098	7065 ± 250	6844
34.	N Adélie Cove, guano, Site S1/88 (T), 39 m	74°44'S, 164°06E	880211.01 GX-20568	6945 ± 95	6709
35.	N Adélie Cove, guano, Site S1/88 (B), 39 m	74°44'S, 164°06E	941102.02 TO-4960	8030 ± 70	7786
36.	N Adélie Cove, guano, Site S2/94 (T), 40 m	74°44'S, 164°06E	941102.04 GX-21404	5825 ± 125	5488
37.	N Adélie Cove, guano, Site S2/94 (B), 40 m	74°44'S, 164°06E	941102.06 TO-5571	8490 ± 90	8280
38.	Inexpressible Is., remains, 23 m (from Dr T. Chinn)	74°53'S, 163°45'E	ccc NZ-7037a	2530 ± 60	1375
39.	Inexpressible Is., guano, 40 m	74°53'S, 163°45'E	870106.02 GX-13607	2900 ± 90	1814
40.	Inexpressible Is., guano, Site P1, 20.7 m	74°53'S, 163°45'E	870107.03 GX-13613	3010 ± 220	1927
41.	Inexpressible Is., guano, Site P1, 20.7 m	74°53'S, 163°45'E	870107.01 GX-13608	5360 ± 90	4858
42.	Inexpressible Is., guano, Site P1, 20.7 m	74°53'S, 163°45'E	870107.04 GX-13614	5945 ± 340	5599
43.	Inexpressible Is., guano, Site P1, 20.7 m	74°53'S, 163°45'E	870107.04 GX-13615	6335 ± 110	6004
44.	Inexpressible Is., guano, Site P2, 19.7 m	74°53'S, 163°45'E	870107.02a GX-13610	5440 ± 85	4975
45.	Inexpressible Is., guano, Site P2, 19.7 m	74°53'S, 163°45'E	870107.02b GX-13611	5530 ± 100	5172, 5168, 5134
46.	Inexpressible Is., guano, Site P2, 19.7 m	74°53'S, 163°45'E	870107.02c GX-13612	6235 ± 110	5915
47.	Inexpressible Is., guano, Site P3, 6.2 m	74°53'S, 163°45'E	870107.06 GX-13616	3340 ± 85	2328
48.	Inexpressible Is., guano, close to Site P3, 6.2 m	74°53'S, 163°45'E	941101.04 TO-4959	3890 ± 80	2988
49.	Inexpressible Is., guano, Site P4, 14 m	74°53'S, 163°45'E	870107.07 GX-13617	3675 ± 90	2746
50.	Inexpressible Is., guano, Site P4, 14 m	74°53'S, 163°45'E	860116.02 GX-12757	4190 ± 80	3373

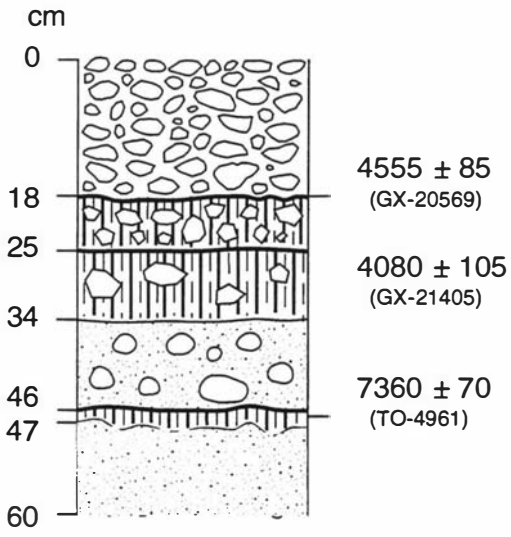
51.	Inexpressible Is., guano, Site S2/94 (T), > 50 m	74°53'S, 163°45'E	941101.06 GX-21403	6040 ± 195	5706
52.	Inexpressible Is., bone	74°53'S, 163°45'E	29 PE64 NZA-12286	6358 ± 55	6082
53.	Cape Irizar, guano, Site S2/96 (T), > 120 m	75°33'S, 162°57'E	961216.01 GX-23275am	4520 ± 65	3801
54.	Cape Irizar, guano, Site S2/96 (B) > 120 m	75°33'S, 162°57'E	961216.04 GX-23276am	5310 ± 60	4828
55.	Prior Is., guano, Site S1 (T), 17 m	75°41'S, 162°52'E	901217.01 GX-16933	1845 ± 75	686
56.	Prior Is., guano, Site S1 (B) 17 m	75°41'S, 162°52'E	901217.02 GX-16932	1860 ± 75	702
57.	Prior Is., guano, Site S2 (T), 18.5 m	75°41'S, 162°52'E	901217.03 GX-16931	1910 ± 75	750
58.	Prior Is., guano, Site S2 (I), 18.5 m	75°41'S, 162°52'E	901217.04 GX-16930	2105 ± 75	949
59.	Prior Is., guano, S2 (I), 18.5 m	75°41'S, 162°52'E	901217.05 GX-16929	2205 ± 75	1057
60.	Prior Is., guano, Site S2 (B), 18.5 m	75°41'S, 162°52'E	901217.06 GX-16928	2385 ± 80	1258
61.	Prior Is., guano, Site S3 (B), > 100 m	75°41'S, 162°52'E	961216.07 GX-23277a	5825 ± 55	5488
62.	Prior Is., guano, Site S4 (T), > 120 m	75°41'S, 162°52'E	1000129.01 OS-27679	4610 ± 65	3888
63.	Prior Is., guano, Site S4, (B), > 120 m	75°41'S, 162°52'E	1000129.02 OS-27680	5080 ± 40	4521
64.	N. Cape Day, guano, Site S1 (Base), 18 m	76°14'S, 162°47'E	901226.01 GX-16923	4180 ± 90	3364
65.	N. Cape Day, guano, close to Site S1 (T), 18 m	76°14'S, 162°47'E	941128.08 TO-4972	4360 ± 50	3574
66.	N Cape Day, guano, Site S2 (Base), 17 m	76°14'S, 162°47'E	901230.03 GX-16910	4230 ± 85	3421
67.	C. Hickey W, guano, Site CH1 (T), 45.5 m	76°05'S, 162°38'E	970117.01 GX-23278am	3485 ± 55	2513
68.	C. Hickey W, guano, Site CH1 (lower), 45.5 m	76°05'S, 162°38'E	970117.06 OS-27675	7540 ± 60	7375
69.	C. Hickey W, guano, Site CH1 (B), 45.5 m	76°05'S, 162°38'E	970117.04 GX-23279a	10005 ± 85	9946, 9908, 9859
70.	C. Hickey W, guano, close to Site CH1, (B), 40 m	76°05'S, 162°38'E	901217.13 GX-16925	11325 ± 360	11894, 11842, 11714
71.	C. Hickey W, guano, close to Site CH1 (B), 40 m	76°05'S, 162°38'E	901217.13 OS-27351	12250 ± 50	13118, 13103, 13022
72.	C. Hickey W, guano, close to Site CH1 (B), 40 m	76°05'S, 162°38'E	921012.01 GX-18843	13070 ± 360	13838
73.	C. Hickey W, guano, Site CH3 (B), 46 m	76°05'S, 162°38'E	970117.10 GX-23280	4265 ± 55	3456
74.	Peninsula c/o Depot Is., guano, Site S4 (T), 16.5 m	76° 42' S, 162° 57 E	941128.05 TO-4971	3320 ± 50	2317
75.	Peninsula c/o Depot Is., guano, Site S4 (T), 16.5 m	76° 42' S, 162° 57 E	901228.03 GX-16915	3825 ± 150	2911
76.	Cape Ross, guano, Site S2 (T), 24.5 m	76°43'S, 162°59'E	941115.09 GX-21407	4135 ± 125	3330
77.	Cape Ross, guano, Site S2 (B), 24.5 m	76°43'S, 162°59'E	901227.02 GX-16921	4555 ± 90	3825
78.	Cape Ross, guano, Site S9 (T), 31.7 m	76°43'S, 162°59'E	941115.10 TO-4967	4310 ± 60	3514
79.	Cape Ross, guano, Site S9 (B), 31.7 m	76°43'S, 162°59'E	901230.02 GX-16911	4315 ± 90	3525
80.	Cape Ross, guano, Site	76°43'S,	901230.01	4465 ± 90	3692

	S8 (B), 31 m	162°59'E	GX-16912		
81.	Cape Ross, guano, Site	76°43'S,	901229.03	4570 ± 90	3836
	S7 (B), 31 m	162°59'E	GX-16913		
82.	Cape Roberts, guano, Site	77°02'S,	941109.11	3955 ± 155	3091
	S1/94 (T), 14 m	163°10'E	GX-20572		
83.	Cape Roberts, guano, Site	77°02'S,	941109.13	4150 ± 60	3341
	S1/94 (B), 14 m	163°10'E	TO-4963		
84.	Spike Cape, guano, Site	77°18'S,	941109.06	5800 ± 100	5467
	S1/94 (T), 22 m	163°33'E	GX-20570		
85.	Spike Cape, guano, Site	77°18'S,	941109.08	5840 ± 70	5544
	S1/94 (B), 22 m	163°33'E	TO-5572		
86.	Dunlop Is., guano, Site	77°14'S,	941124.09	3985 ± 175	3144
	S1/94 (T), 18.8 m	163°28'E	GX-21409		
87.	Dunlop Is., guano, Site	77°14'S,	941124.10	3690 ± 70	2754
	S1/94, 18.8 m	163°28'E	TO-5575		
88.	Dunlop Is., guano, Site	77°14'S,	941125.01	3530 ± 60	2662
	S2/94 (T), 19.5 m	163°28'E	TO-5576		
89.	Dunlop Is., guano, Site	77°14'S,	941125.06	5630 ± 70	5286
	S2/94 (B), 19.5 m	163°28'E	TO-4968		
90.	Dunlop Is., guano, Site	77°14'S,	941125.12	6330 ± 80	5997
	S3/94, 19.5 m	163°28'E	TO-5577		
91.	Dunlop Is., guano, Site	77°14'S,	941125.12	5750 ± 70	5438
	S3/94, 19.5 m	163°28'E	TO-5578		
92.	Beaufort Is.,	76°56'S,	2B-4 PE42	1313 ± 55	275
	Penguin bone	167°03'E	NZA-12287		
93.	Cape Crozier	77°14'S,	4B3 PE	1475 ± 59	440
	Penguin bone	166°28'E	NZA-8818		
94.	Cape Crozier	77°14'S,	1A-4 PE	1534 ± 57	481
	Penguin bone	166°28'E	NZA-9182		
95.	Cape Crozier	77°14'S,	1B3 PE	1380 ± 60	310
	Penguin bone	166°28'E	NZA-10306		
96.	Cape Crozier	77°14'S,	9B3 PE	1503 ± 57	461
	Penguin bone	166°28'E	NZA-8819		
97.	Cape Crozier	77°14'S,	11B3 PE	1616 ± 62	523
	Penguin bone	166°28'E	NZA-8820		
98.	Cape Crozier	77°14'S,	13B3 PE	1564 ± 57	498
	Penguin bone	166°28'E	NZA-8821		
99.	Cape Bird, remains	77°30'S,		7070 ± 180	6848
		162°10'E	NZ-5590		

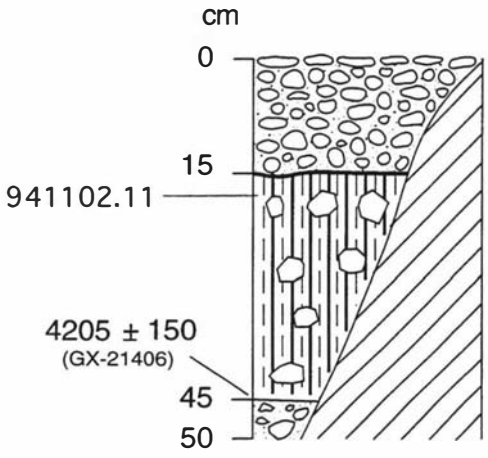
Figure C.1. Soil profiles from the locations sampled for subfossil bones. The age of each horizon is taken from Table C.1. These drawing were prepared by Carlo Baroni, Università di Pisa, Italy.



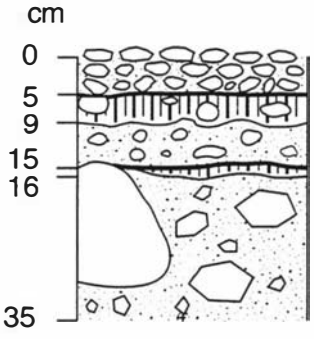
Spike Cape



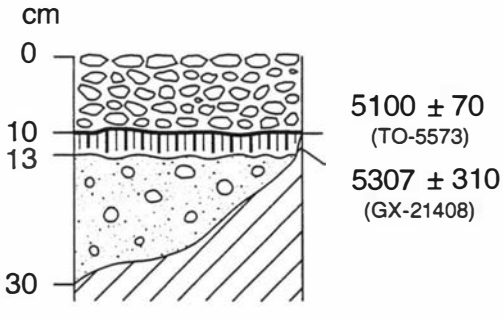
Northern Foothills S of
C. Icaro, S3/94
60 m asl
74° 42' 50" S
164° 06' 30" E



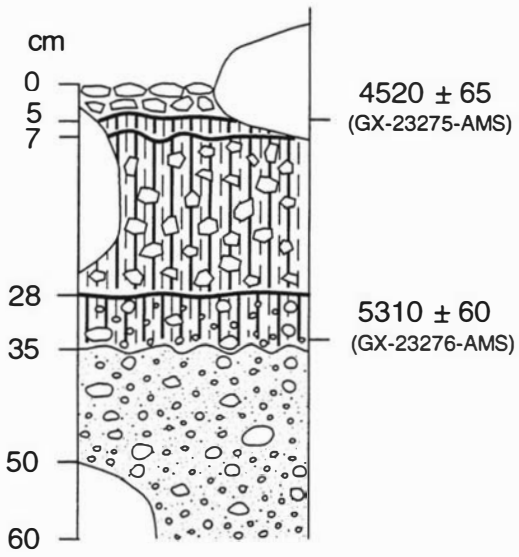
Northern Foothills (100 m S of
Campo Icaro)
S4/94 ca 35 m asl
164°7'10"E
74°42'46"S



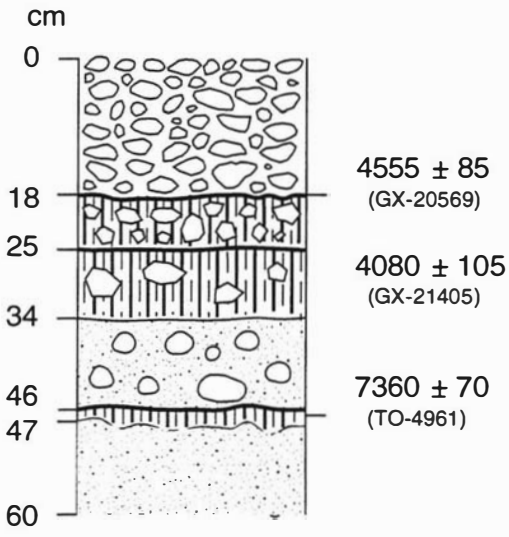
Terra Nova Station, 40m



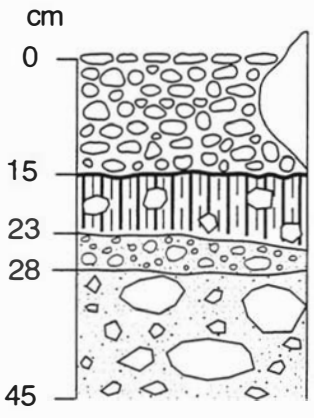
Terra Nova Station
74° 41' 37" S
164° 06' 39" E



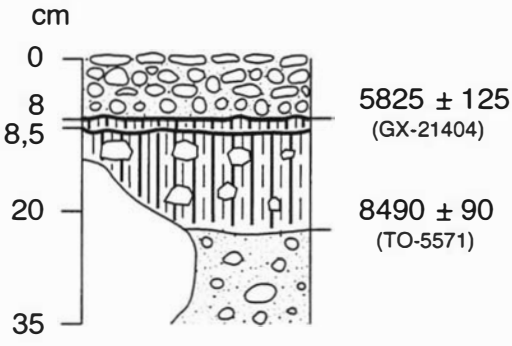
C. Irizar 120 m
S2/96
75° 33' 47" S
162° 57' 20" E



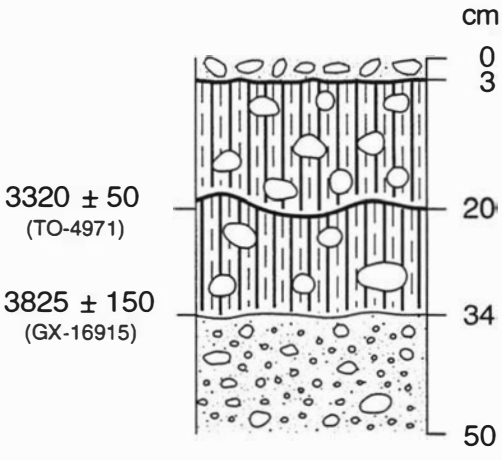
Northern FoothillsS of
C. Icaro, 60 m asl
74° 42' 50" S
164° 06' 30" E



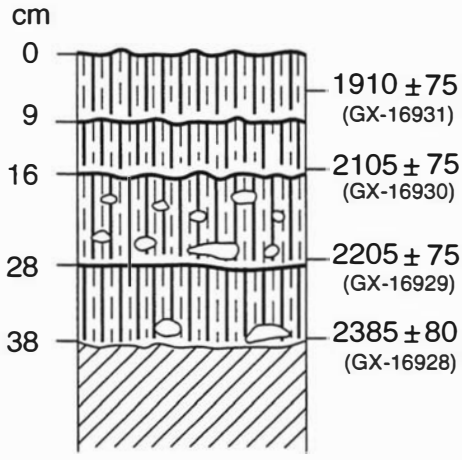
N Adélie Cove, S1/88
Northern Foothills
74° 44' 11" S
164° 06' 57" E



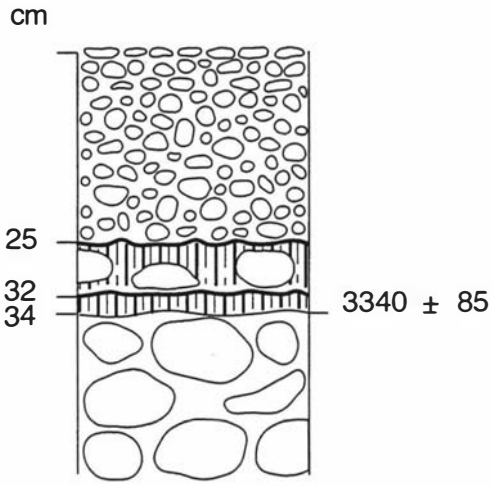
Northern Foothills
N Adélie Cove S2/94 (40 m)
74° 44' 07" N
164° 06' 55" E



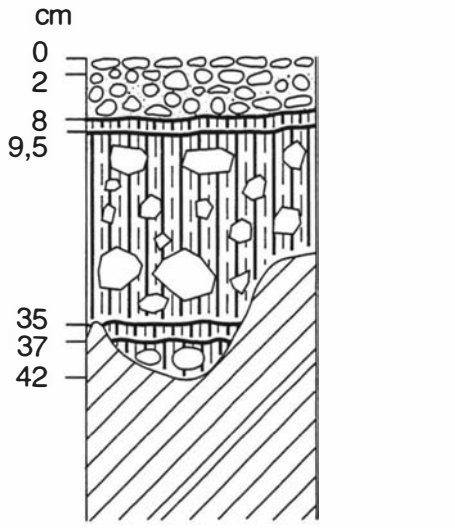
Penisola c/o Depot Island site 4
76° 42' 03"S
162° 57'E



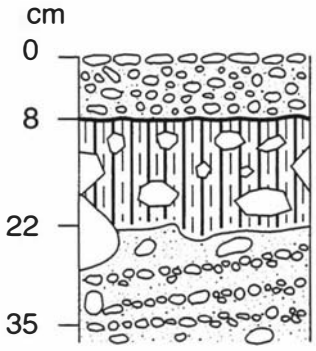
Prior Island S2/90 18 m asl
75° 41' 33" S
162° 52' 38" E



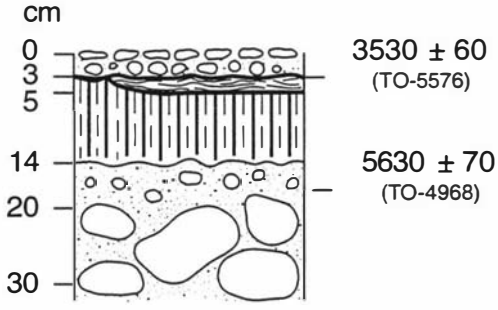
Inexpresible Island
Site P3



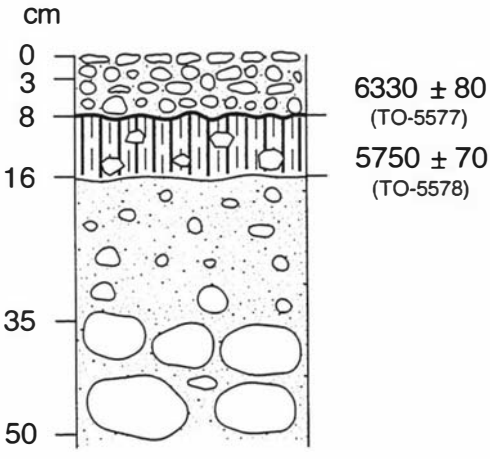
North. Foothills C. Icaro S5/95
74° 42' 42" S
164° 07' 16" E



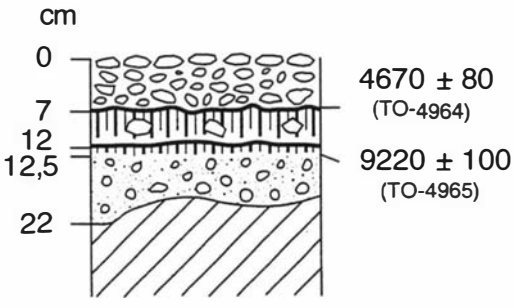
Cape Roberts
77° 02' 22" S
163° 10' 49" E



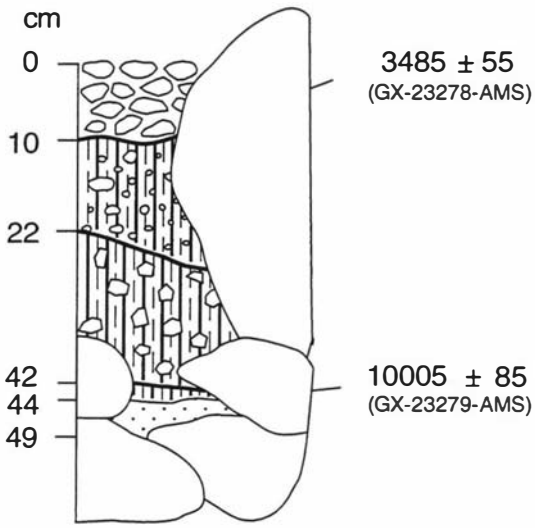
Dunlop Island Site 2
(19.5 m)
77° 14' 28" S
163° 28' 59" E



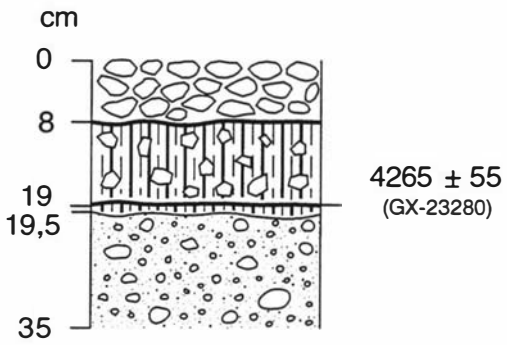
Dunlop Island Site 3 (19.5 m)
77° 14' 26" S 163° 28' 53" E



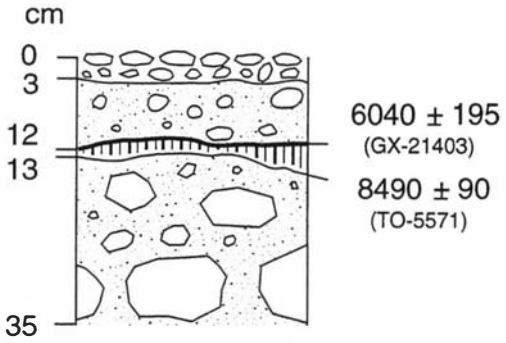
Gondwana St. S2/94 (15 m)
74° 38' 16" S
164° 12' 54" E



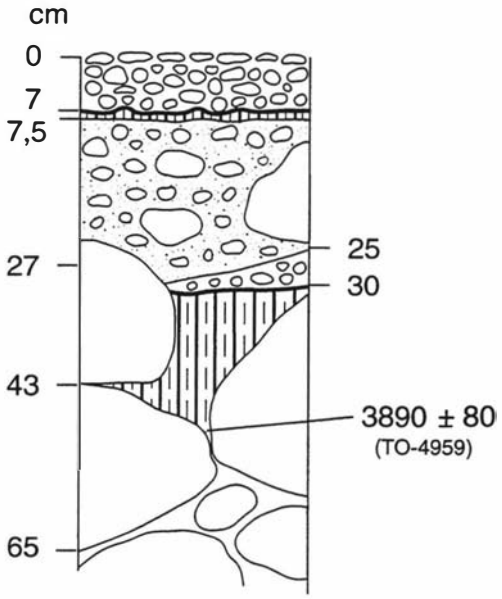
Cape Hickey site CH-1
76° 05' 24" S
162° 38' 09" E



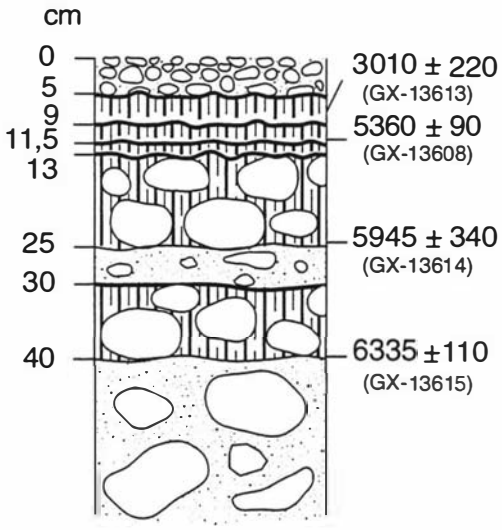
Cape Hickey site CH-3 (46 m)
76° 05' 24" S
162° 38' 09" E



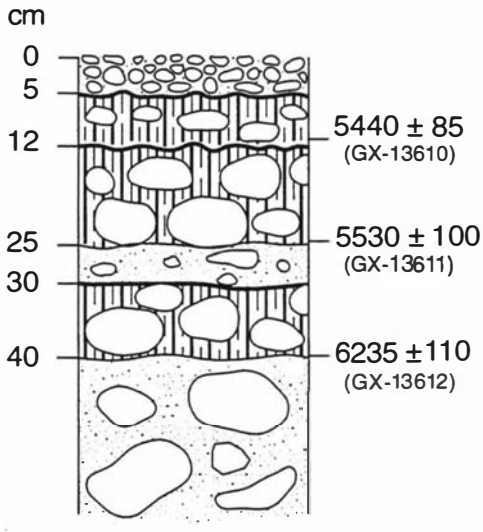
Inexpressible Island, Promontory S of
Seaview Bay, S2/94
above marine limit (ca 50m)
163° 44' 00" E
74° 54' 40" S



Inexpressible Island Seaview Bay
(6.2 m) few m S of P3
74° 54' 21" S
163° 43' 42" E



Inexpressible Island
Site P1



Inexpressible Island
Site P2

Appendix D – Manuscripts

A repeat complex in the mitochondrial control region of Adélie penguins from Antarctica

Peter A. Ritchie and David M. Lambert

Abstract: We have determined the nucleotide sequence of the entire mitochondrial control region (CR) of the Adélie penguin (*Pygoscelis adeliae*) from Antarctica. Like in most other birds, this CR region is flanked by the gene *nad6* and transfer (t)RNA trnE(uuc) at the 5′ end and the gene *rns* and trnF(gaa) at the 3′ end. Sequence analysis shows that the Adélie penguin CR contains many elements in common with other CRs including the termination associated sequences (TAS), conserved F, E, D, and C boxes, the conserved sequence block (CSB)-1, as well as the putative light and heavy strand promoters sites (LSP–HSP). We report an extraordinarily long avian control region (1758 bp) which can be attributed to the presence, at the 3′ peripheral domain, of five 81-bp repeat sequences, each containing a putative LSP–HSP, followed by 30 tetranucleotide microsatellite repeat sequences consisting of $(dC–dA–dA–dA)_{30}$. The microsatellite and the 81-bp repeat reside in an area known to be transcribed in other species.

Key words: Aves, microsatellite, evolution, D-loop, TAS, WANCY.

Résumé : Les auteurs ont déterminé la séquence nucléotidique de toute la région de contrôle (CR) mitochondriale chez le pingouin Adélie (*Pygoscelis adeliae*) de l'Antarctique. Comme chez la plupart des autres oiseaux, la région CR est bordée du côté 5' par le gène *nad6* ainsi que par le gène codant pour l'ARN(t) trnE(uuc) et du côté 3′ par les gènes *rns* et trnF(gaa). Une analyse de séquences a montré que la région CR du pingouin Adélie contient plusieurs éléments en commun avec d'autres régions CR dont les séquences TAS (« termination associated sequences »), les boîtes conservées F, E, D et C, le motif CSB (« conserved sequence block »)-1, de même que les sites promoteurs potentiels LSP–HSP (« light and heavy strand promoters »). Les auteurs rapportent une région de contrôle exceptionnellement longue (1758 pb) pour un oiseau. Cette grande taille est due à la présence, dans le domaine périphérique en 3′, de cinq copies d'un motif de 81 pb, chacune contenant une séquence LSP–HSP suivie d'un microsatellite tétranucléotidique $(dC–dA–dA–dA–dA)_{30}$. Le microsatellite et le motif répété de 81 pb sont situés dans une région connue comme étant transcrite chez d'autres espèces.

Mots clés : Aves, microsatellite, évolution, D-loop, TAS, WANCY.

[Traduit par la Rédaction]

Introduction

The mitochondrial (mt) control region (CR) is responsible for transcription and replication of the mitochondrial genome (Taanman 1999). Three internal CR portions have been recognized; the 5′-peripheral domain (right domain), the central conserved domain (ca. 200 bp), and the 3′-peripheral domain (left domain) (Upholt and Dawid 1977; Southern et al. 1988; Saccone et al. 1991). In the avian CR, the 5′-peripheral domain contains the termination associated sequences (TAS), the central conserved domain contains the F, E, D, and C boxes, and the 3′-peripheral domain contains the origin of heavy strand replication (O_H), conserved sequence block (CSB)-1, and the heavy (H) and light (L) strand transcriptional promoter (HSP–LSP) sites (L'Abbé et al. 1991; Quinn and Wilson 1993; Randi and Lucchini 1998). Transcription and replication of the mt genome are intimately connected. After transcription has begun at the LSP site, mitochondrial processing endoribonuclease (RNase MRP), and possibly endonuclease G (Côté and Ruiz-Carrillo 1993), are thought to cleave the L-strand transcript at sequences corresponding to the CSBs (Chang and Clayton 1985). This transcript forms an RNA primer that initiates replication of the H-strand at a site designated the O_H (Walberg and Clayton 1981). It is known that in humans, as H-strand synthesis displaces the L-strand at the WANCY tRNA cluster, approximately two thirds of its way around the molecule, this free L-strand is primed by a mtDNA primase and L-strand synthesis begins (Wong and Clayton 1986).

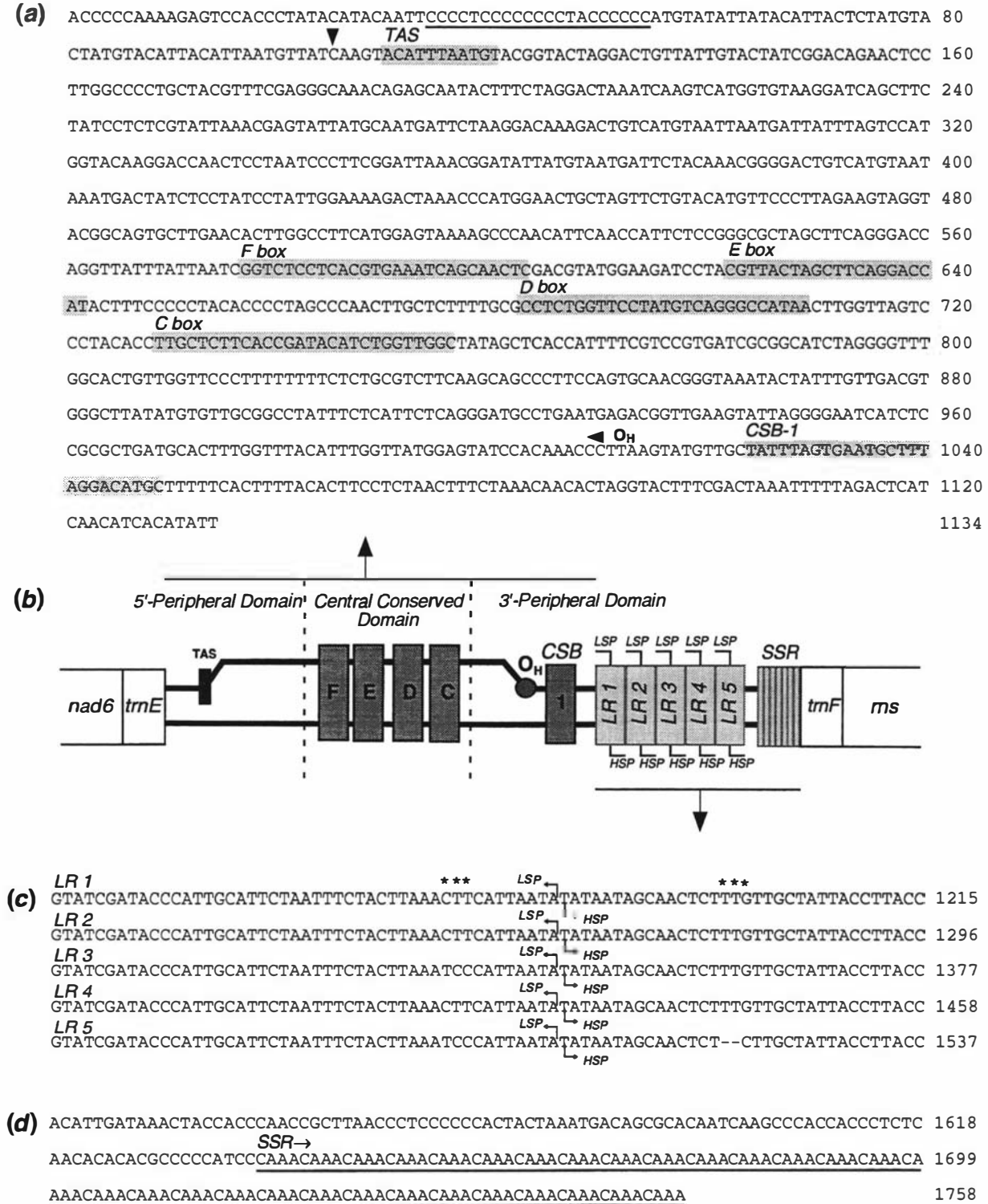
Sequence data from mtCRs has revealed the presence of repeat DNA sequences in many different species, from nematodes (*Caenorhabditis elegans*, Okimoto et al. 1992) to sheep (*Ovis ovis*, Zardoya et al. 1995). Lunt et al. (1998) surveyed repeat sequences in the CRs of a range of species and showed that they varied in number and ranged in size from 3 bp to as large as 777 bp. To date, several mechanisms

Corresponding Editor: P.B. Moens.

Received November 11, 1999. Accepted February 23, 2000. Published on the NRC Research Press web site on June 19, 2000.

P.A. Ritchie and D.M. Lambert.[1] Molecular Ecology, Institute of Molecular BioSciences, Massey University, Private Bag 11–222, Palmerston North, New Zealand.

[1]Author to whom all correspondence should be addressed (e-mail: D.M.Lambert@massey.ac.nz).



have been proposed to explain the generation and persistence of repetitive DNA in the mitochondrial CR. These include slippage during replication (Levinson and Gutman 1987; Madsen et al. 1993) or rarer recombination events (Lunt and Hyman 1997). Consistent with these models, repetitive DNA in the CR is found in areas where transcription and replication are initiated (i.e., near the TAS in the 5′-peripheral domain or between CSB-1 and the tRNA in the 3′-peripheral domain). We report here on the mtCR of the Adélie penguin *Pygoscelis adeliae* (Aves: Sphenisciformes)

Fig. 1. The L-strand sequence and a schematic representation of the Adélie penguin control region (CR). (*a*) The first 1143 bp (comprising the 5′-perpherial domain, central conserved domain, and part of the 3′-peripheral domain) of the CR. Underlined is the interrupted poly-C sequence, marked with an arrowhead is the possible dC terminus of the D-loop, and highlighted are the termination associated sequences (TAS). Also highlighted are the F, E, D, and C boxes and consersed sequence block (CSB)-1. (*b*) A schematic representation of the CR, which shows the flanking genes and portions of the CR sequence represented in this figure. The origin of heavy strand replication (O_H) is estimated to be 12 nt from CSB-1 and is represented as a circle. The abbreviations correspond to: *nad6*, encodes the product NADH dehydrogenase subunit 6; trnE(uuc), transfer RNA glutamic acid; trnF(gaa), tRNA phenylalanine; *rns*, encodes the small subunit ribosomal RNA (12S). (*c*) the 81-bp large repeat (LR-1, -2, -3, -4, and -5). The repeat units are aligned with each other, the stars show the variable positions, and the ← symbols represent the light (L) and heavy (H) strand promotor (LSP–HSP) sites. (*d*) The 92-bp sequence after the LR and underlined is the simple sequence repeat (SSR); a tetranucleotide microsatellite (CAAA).

from Antarctica, and show that this is one of the largest avian mtCRs found to date, due to the presence of two classes of repeat units in the 3′-peripheral domain.

Materials and methods

Blood samples from Adélie penguins were collected during the austral summer of 1997–1998 at Cape Bird, Antarctica (lat 77°14′S, long 166°28′E), then transported in liquid nitrogen and stored at –80°C. Total genomic DNA was isolated from whole blood using standard proteinase K – SDS (sodium dodecyl sulfate) dissolution, followed by extraction with phenol–chloroform and ethanol precipitation (Sambrook et al. 1989). The Adélie penguin control region was amplified using the polymerase chain reaction (PCR) with primers designed to the flanking *nad6* (5′-ACTAAACCAATTACCCCATAATA-3′ L-*nad6*-gb) and *rns* (5′-CTGCTGAGTACCCGTGGGGGTGTGGC-3′ H-*rns*-pen) genes. These PCR primers were designed using *nad6* gene sequences from GenBank (Acc. Nos.: chicken, X52392; duck, L16769; ostrich, Y12025; goose, L22477; and gnatcatcher, AF027829) and *rns* gene sequences determined for the penguin species *Pygoscelis adeliae, Eudyptes pachyrhynchus, Megadyptes antipodes*, and *Eudyptula minor*. To design the H-strand primer, we sequenced the 5′ end of the *rns* gene using PCR primers 5′-AGCATGGCACTGAAGATGC-3′ (L-tRNA-P, modified from Tarr 1995) and 5′-ATAGTGGGGTATCTAATCCCA-3′ (H-*rns*A, modified from Palumbi 1995). A 260-bp portion of the WANCY tRNA cluster was amplified and sequenced in both the Adélie and yellow-eyed (*M. antipodes*) penguins using the primers 5′-CCAAAGGCCTTCAAAGCCTTAAATAAGAG-3′ (L-WANCY), 5′-CCAAAAATCTGTGGTTCAATTCCTCTTC-3′ (H-WANCY); designed from GenBank sequences (chicken, X52392; ostrich, Y12025). All mitochondrial gene nomenclature in this work follows that of the Organelle Genome Database (GOBASE) available at http://alice.bch.umontreal.ca/genera/gobase/gobase.html (Korab-Laskowska et al. 1998).

All PCRs contained 0.4 μM of each primer, 200 μM of each dNTP, 1.5 mM $MgCl_2$, 10 mM Tris–HCl pH 8.3, 50 mM KCl, and 0.5 U AmpliTaq DNA polymerase (PE Biosystems, N.J.). The PCR products were purified using High Pure™ PCR Product Purification columns (Boehringer Mannheim, Mannheim). Purified PCR products were sequenced with internal primers using an ABI PRISM® BigDye™ Terminator Cycle Sequencing Kit (PE Biosystems, N.J.) and run on an ABI 377 automated sequencer, according to the manufacturer's recommendations. Sequences for the CR can be found at GenBank Acc. No. AF272143 and for the tRNA WANCY at AF272144 and AF272145. To predict the single-stranded secondary structures of the CR 5′-end we used the UNIX-based software by GCG (Genetics Computer Group, Inc., Wis).

To verify that these CR sequences were of mt origin, and did not represent nuclear copies, a large portion (ca. 6.0 kb) of the Adélie penguin mt genome was amplified and sequenced. The long PCR template was generated using the Expand™ Long Template PCR System (Boehringer Mannheim) using two PCR primers designed to conserved portions of the *rsl* gene 5′-TGATTGCGCTACCTTCGCACGGTTAGGATACC-3′ (H-ExpPeng-*rsl*) and the *cob* gene 5′-CCATTCCACCCCTACTACTCCACAAAAGA-3′ (L-ExpPeng-*cob*). All PCRs contained 0.3 μM of each primer, 500 μM of dNTPs, PCR buffer 3 (50 mM Tris–HCl pH 9.2, 16 mM $(NH_4)_2SO_4$, 2.25 mM $MgCl_2$, 2% DMSO, and 0.1% Tween–20), and 2.6 U Expand™ Long Template enzyme mix. The resulting PCR product was purified using High Pure™ PCR Product Purification columns (Boehringer Mannheim), and the CR was amplified and sequenced as above. We obtained the same sequences across the control region for both the long and short PCR products indicating that both have been amplified from the same template. In addition, we did not encounter any double peaks on the electropherograms during sequencing.

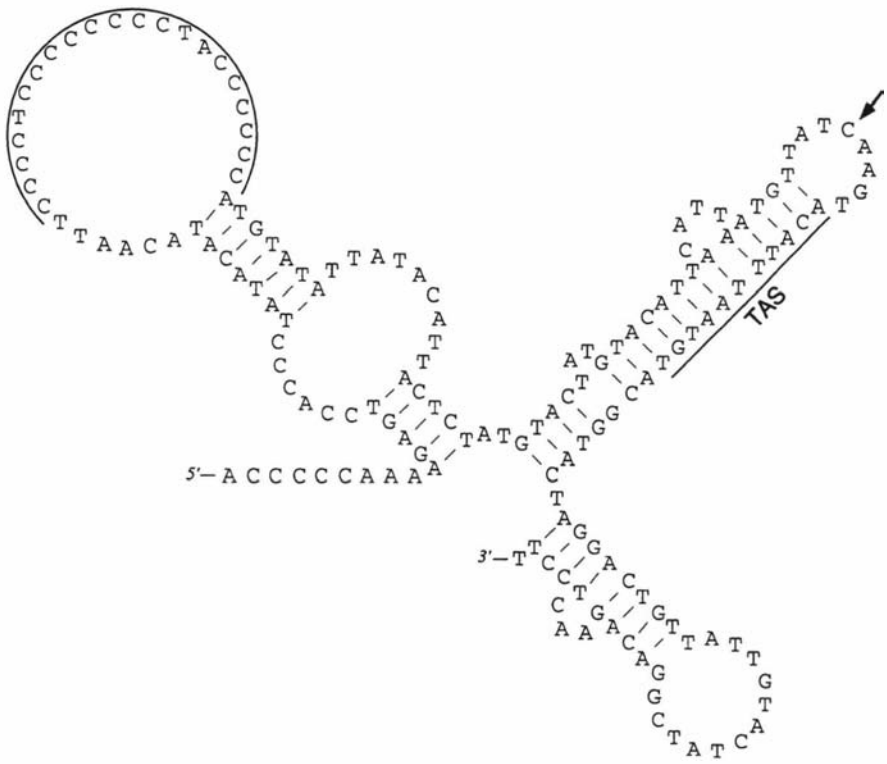
Results and discussion

The organization of the Adélie mt genome

The CR of this Adélie penguin is 1758 bp in length and is bound by the tRNAs trnE(uuc) and trnF(gaa), and the genes *nad6* and *rns* (Fig. 1). The position of the flanking genes, tRNAs, and the control region is the same as that determined for chicken (*Gallus gallus*, Desjardins and Morais 1990) (Fig. 1*b*). A recent study by Mindell et al. (1998) showed there are two types of gene arrangement in birds; the common *cob*/*nad6*/control region/*rns* order (e.g., the gene arrangement of the chicken) and the less frequent order of; *cob*/control region/*nad6*/a non-coding region/*rns*. If the avian phylogeny that Mindell et al. (1998) presented is correct, then the latter, less frequent gene rearrangement has occurred as separate independent events in Picidae, Cuculidae, Passeriformes suboscines, and Falconiformes. The results of Mindell et al. (1998) indicate that these mt gene rearrangements are likely to be due to specific processes that result in the reoccurrence of this condition.

In mammals the WANCY tRNA cluster includes an ~50-nucleotide (nt) sequence that is capable of forming a stable single-stranded secondary structure, which is known to be involved in priming the synthesis of the nascent L-strand. However, this sequence is conspicuously absent from the WANCY tRNA cluster of birds (Seutin et al. 1994). This begged the questions of (1) whether the O_L is located at a different position, yet undiscovered, in birds; and (or) (2) whether there are other unknown recognition sequences on the mt genome involved in initiation of L-strand synthesis (Shadel and Clayton 1997). We have determined the sequence for a 250-bp portion of the WANCY tRNA cluster for the Adélie and yellow-eyed penguins. Like other avian species, the Adélie and yellow-eyed penguins possess no

Fig. 2. The secondary structure of the first 164 bp of the 5′-peripheral domain of the Adélie penguin control region (CR). Outlined is the interrupted poly-C sequence and the putative termination associated sequence (TAS); the arrow corresponds to the dC terminus of the chicken D-loop.



mammalian-like sequences responsible for priming the O_L in this region of the mt genome (data not shown).

The organization of the Adélie penguin CR

We compared the internal organization of the Adélie penguin CR to that found in other avian CRs (Quinn and Wilson 1993; Randi and Lucchini 1998) and found them to be very similar. The 5′-peripheral domain is 560 bp, the CCD is 200 bp (as defined by Saccone et al. 1991), and the 3′-peripheral domain is 998 bp long. The 3′-peripheral domain of the Adélie penguin is larger than the entire 909-bp CR of the cow (*Bos taurus*, Anderson et al. 1982), and only slightly smaller than the CRs of many avian species (Baker and Marshall 1997). The overall base composition of the Adélie penguin CR is A, 30.4%; T, 30.3%; C, 25.5%; and G, 13.8%. These values show that there is a A + T (60.8%) to G + C (39.2%) asymmetry in this sequence. It has been shown that the CCD differs from the two peripheral domains in being lower in L-strand adenine content (Ramirez et al. 1993). The reduced adenine content of the L-strand of the CCD is also the case in the Adélie penguin with A, 20.7%; T, 30.8%; C, 31.3%; and G, 17.1%, which almost corresponds to an A + T to G + C symmetry: 51.5% to 48.5%, respectively.

The 5′-peripheral domain contains an interrupted poly-C sequence (5′-CCCCTCCCCCCCCTACCCCCC-3′) which forms a potential single-stranded secondary structure (Figs. 1*a* and 2). Zardoya and Meyer (1998) reported a similar motif in the CR of the African side-necked turtle (*Pelomedusa subsrufa*), and noted its similarity to the CSB-2 motif (5′-CAAACC-CCCCCWMCCCC-3′) identified in the 3′-peripheral domain of mammals by Walberg and Clayton (1981) and Saccone et al. (1991). A survey of GenBank shows that a similar motif is present in the 5′-peripheral domain of other avian species, including Struthioniformes, Galliformes, and Falconiformes. In the lesser snow goose (*Anser caerulescens caerulescens*) and the chicken, this motif forms a potentially stable hairpin structure, because it is followed by an interrupted poly-G sequence (Quinn and Wilson 1993). Although this interrupted poly-C sequence is a conserved feature across many species and is close to one identified TAS (Nass 1995), a role for this sequence has never been determined.

H-strand synthesis originates within the first 40 nt downstream from CSB-1 and temporarily stalls shortly after initiation. The 3′ end(s) of the nascent H-strand terminate close to the TAS to form a stable triplex displacement-loop (D-loop) structure. Nass (1995) has identified the 5′ and 3′ ends of the D-loop from the chicken and the TAS (5′-ATATAT-TTACA-3′) at 81 bp to the right of trnE(uuc). She went on to show that this sequence is present in other avian species and associated with a putative secondary structure. Furthermore, the 5′ end of the chicken D-loop was shown to begin 11–12 nt downstream of the CSB-1, is 781 (±1) nt in length, and has two nascent H-strand 3′-termini which exhibit dG and dC 3′-end nt. Presented in Fig. 2 is the putative secondary structure of the first 164 bp of the 5′-peripheral domain of the Adélie penguin, which includes the interrupted poly-C

sequence and the putative TAS sequence and dC 3′ terminus. Based on an inferred location of the O_H and TAS, this D-loop would correspond to 893 nt in length.

Tandem repeats in the 3′-peripheral domain

The most striking feature of the Adélie penguin CR is the presence of two types of tandemly repeated sequences, a large repeat (LR), and a simple sequence repeat (SSR) in the 3′-peripheral domain (Fig. 1*c* and 1*d*). These two repeats contribute to the extraordinary size of this CR. Considering the base composition of this domain, the first 365 nucleotides are 58.4% A + T to 41.6% G + C, followed by the large repeats at 69.7% A + T to 30.3% G + C, then the LR–SSR spacer sequence is 46.0% A + T to 54.0% GC, and last, the SSR is 75% A to 25% C. These values show that the LRs have a significant A + T basis, a feature known to cause helix instability and which is common at origins of replication.

The SSRs comprise 30 perfect tetranucleotide microsatellite repeats consisting of $(dC–dA–dA–dA)_{30}$·(dG–dT–dT–dT)$_{30}$. This microsatellite is located at the extreme 3′ end of the CR, is contiguous with the tRNA, and since it is downstream from the HSP site, is most likely to be transcribed. Interestingly, the same tetranucleotide microsatellite has also been found in the ratite, *Rhea americana* (Härlid et al. 1998), and similar repeats containing a CA-type motif have been found on the L-strand in other avian species (Berg et al. 1995). Like other microsatellites, there is good evidence that SSRs are formed and variation maintained via slipped-strand mispairing during mitochondrial replication (Levinson and Gutman 1987; Madsen et al. 1993; Trinh and Sinden 1993).

The LR region comprises four complete 81-bp repeats, followed by a repeat missing two base pairs (i.e., 5′-81–81–81–81–79-3′). Alignment of the repeats with each other reveals substitutional differences at positions 1171, 1172, and 1173 (CT), at position 1199 (CG), and insertions–deletions (indels) at positions 1197–1198 with respect to LR-1 in the data set (Fig. 1c). Avian species differ from many vertebrates by having a single major transcriptional promoter in the CR which has a bidirectional capacity for both the L- and H-strand (L'Abbé et al. 1991). All these repeat units contain a putative HSP–LSP sequences, initially identified in the chicken (5′-GTATAATATATATACA-3′). The first two repeats (LR-1, LR-2) and the fourth repeat (LR-4) are identical. The third repeat (LR-3) and the last repeat (LR-5) are each unique and differ from the other repeats (LR-1, -2, and -4) by the substitutions described above and from each other by an indel event and a substitution. When folded, the LR repeats form a potential hairpin structure. In the European rabbit (*Oryctolagus cuniculus*, CR = 2437 bp), a primer extension analysis detected families of H-strand RNA species which originated from the CR LSP. These different-sized RNA transcripts were attributed to a tandem repeat (153 bp) in which each motif contained a LSP site (Dufresne et al. 1996). Based on that study, it is likely that the Adélie penguins, also, produce multiple RNA transcripts, each originating from promoter sites in the LRs. However, in contrast to rabbits, there would be variable length transcripts for both the L- and H-strand RNA, since avians possess a bidirectional promoter.

To explain the changes in 5′-peripheral domain repeat numbers, Buroker et al. (1990) proposed an illegitimate elongation model which involved slipped-strand mispairing and nonhomologous recombination. A similar model could be proposed to explain the LRs present in the 3′-peripheral domain of Adélie penguins. If, during replication, there is a dynamic competitive equilibrium between the large RNA primers and the H-strand for pairing with the L-strand, either at the beginning or end of replication, then partial displacement of a repeat unit is possible. A hairpin structure in the H-strand would result in a repeat loss, whereas a hairpin formation in the RNA primer would result in a repeat gain. The daughter molecule would form a heteroduplex with the parental molecule which would be resolved during the next round of replication.

We have sequenced the mtCR of the Adélie penguin and identified its internal organization, including the conserved motifs (TAS, the F, E, D, and C boxes, CSB-1, and the LSP–HSP). This control region is larger than those from other avian species, due to the presence of a large 81-bp repeat sequence and a $(dC–dA–dA–dA)_{30}$ microsatellite in the 3′-peripheral domain. The 81-bp repeat contains both the LSP and HSP and may therefore produce a variety of RNA transcripts. This study shows that avian mtCRs, like those in mammals, can vary greatly in size due to the presence of different repeat complexes.

Acknowledgements

This research is supported by a grant to DML from the Marsden Fund of New Zealand (96-MAU-ALS-0030). We are grateful to Antarctica New Zealand, and the U.S. Coast Guard cutter Polar Sea WAGB-11 for logistic support in Antarctica. Thanks to Paul Barrett and Craig Millar for assistance with field collections, and to Kerryn Slack for kindly sharing unpublished sequences. We also thank Craig Millar, Amy Roeder, James Bower, and Leon Huynen for comments on this manuscript. PAR acknowledges the support of a Massey University doctoral scholarship.

References

Anderson, S., de Bruijn, M.H., Coulson, A.R., Eperon, I.C., Sanger, F., and Young, I.G. 1982. Complete sequence of bovine mitochondrial DNA: Conserved features of the mammalian mitochondrial genome. J. Mol. Biol. **156**: 683–717.

Baker, A.J., and Marshall, H.D. 1997. Mitochondrial control region sequences as tools for understanding evolution. *In* Avian Molecular Evolution and Systematics. *Edited by* D.P. Mindell. Academic Press, San Diego, Calif. pp. 51–82.

Berg, T., Moum, T., and Johansen, S. 1995. Variable number of simple tandem repeats make birds of the order Ciconiiformes heteroplasmic in their mitochondrial genomes. Curr. Genet. **27**: 257–262.

Buroker, N.E., Brown, J.R., Gilbert, T.A., O'Hara, P.J., Beckenbach, A.T., Thomas, W.K., and Smith, M.J. 1990. Length heteroplasmy of Sturgeon mitochondrial DNA: An illegitimate elongation model. Genetics, **124**: 157–163.

Chang, D.D., and Clayton, D.A. 1985. Priming of human mitochondrial DNA replication occurs at the light-strand promoter. Proc. Natl. Acad. Sci. U.S.A. **82**: 351–355.

Côté, J., and Ruiz-Carrillo, A. 1993. Primers for mitochondrial DNA replication generation by Endonuclease G. Science, **261**: 765–769.

Desjardins, P., and Morais, R. 1990. Sequence and gene organization of the chicken mitochondrial genome. J. Mol. Biol. **212**: 599–634.

Dufresne, C., Mignotte, F., and Guéride, M. 1996. The presence of tandem repeats and the initiation of replication in rabbit mitochondrial DNA. Eur. J. Biochem. **235**: 593–600.

Härlid, A., Janke, A., and Árnason, U. 1998. The complete mitochondrial genome of *Rhea americana* and early avian divergences. J. Mol. Evol. **46**: 669–679.

Korab-Laskowska, M., Rioux, P., Brossard, N., Littlejohn, T.G., Gray, M.W., Lang, B.F., and Burger, G. 1998. The Organelle Genome Database Project (GOBASE). Nucleic Acids Res. **26**: 139–144.

L'Abbé, D., Duhaime, J.-F., Lang, B.F., and Morais, R. 1991. The transcription of DNA in chicken mitochondria initiates from one major bidirectional promoter. J. Biol. Chem. **266**: 10 844 – 10 850.

Levinson, G., and Gutman, G.A. 1987. Slipped-strand mispairing: A major mechanism for DNA sequence evolution. Mol. Biol. Evol. **4**: 203–221.

Lunt, D.H., and Hyman, B.C. 1997. Animal mitochondrial DNA recombination. Nature, **387**: 247.

Lunt, D.H., Whipple, L.E., and Hyman, B.C. 1998. Mitochondrial DNA variable number of tandem repeats (VNTRs): Utility and problems in molecular ecology. Mol. Ecol. **7**: 1441–1455.

Madsen, C.S., Ghivizzani, S.C., and Hauswirth, W.W. 1993. *In vivo* and *in vitro* evidence for slipped mispairing in mammalian mitochondria. Proc. Natl. Acad. Sci. U.S.A. **90**: 7671–7675.

Mindell, D.P., Sorenson, M.P., and Dimcheff, D.E. 1998. Multiple independent origin of mitochondrial gene order in birds. Proc. Natl. Acad. Sci. U.S.A. **95**: 10 693 – 10 697.

Nass, M.M.K. 1995. Precise sequence assignment of replication origin in the control region of chick mitochondrial DNA relative to 5′ and 3′ D-loop ends, secondary structure, DNA synthesis, and protein binding. Curr. Genet. **28**: 401–409.

Okimoto, R., Mcfarlane, J.L., Clary, D.O., and Wolstenholme, D.R. 1992. The mitochondrial genomes of two nematodes, *Caenorhabditis elegans* and *Ascaris suum*. Genetics, **130**: 471–498.

Palumbi, S.P. 1995. Nucleic acids II: The polymerase chain reaction. *In* Molecular Systematics. *Edited by* D.M. Hills, C. Moritz, and B.K. Mable. Sinauer Associates, Inc., Sunderland, Mass. pp. 205–247.

Quinn, T.W., and Wilson, A.C. 1993. Sequence evolution in and around the mitochondrial control region in birds. J. Mol. Evol. **37**: 417–425.

Ramirez, A., Savoie, P., and Morais, R. 1993. Molecular characterization and evolution of a duck mitochondrial genome. J. Mol. Evol. **37**: 296–310.

Randi, E., and Lucchini, V. 1998. Organization and evolution of the mitochondrial DNA control region in the avian genus *Alectoris*. J. Mol. Evol. **47**: 449–462.

Saccone, C., Pesole, G., and Sbisá, E. 1991. The main regulatory region of mammalian mitochondrial DNA: Structure-function model and evolutionary patterns. J. Mol. Evol. **33**: 83–91.

Sambrook, J., Fritsch, E.F., and Maniatis, T. 1989. Molecular Cloning: A Laboratory Manual. 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Seutin, G., Lang, B.F., Mindell, D.P., and Morais, R. 1994. Evolution of the WANCY region in amniote mitochondrial DNA. Mol. Biol. Evol. **11**: 329–340.

Shadel, G.S., and Clayton, D.A. 1997. Mitochondrial DNA maintenance in vertebrates. Annu. Rev. Biochem. **66**: 409–435.

Southern, S.O., Southern, P.J., and Dizon, A.E. 1988. Molecular characterization of a cloned mitochondrial genome. J. Mol. Evol. **28**: 32–42.

Taanman, J.-W. 1999. The mitochondrial genome: Structure, transcription, translation and replication. Biochim. Biophys. Acta, **1410**: 103–123.

Tarr, C.L. 1995. Primers for amplification and determination of mitochondrial control-region sequences in oscine passerines. Mol. Ecol. **4**: 527–529.

Trinh, T.Q., and Sinden, R.R. 1993. The influence of primary and secondary DNA structure in deletion and duplication between direct repeats in *Escherichia coli*. Genetics, **134**: 409–422.

Upholt, W.B., and Dawid, I.B. 1977. Mapping of mitochondrial DNA of individual sheep and goats: Rapid evolution in the D loop region. Cell, **11**: 571–583.

Walberg, M.W., and Clayton, D.A. 1981. Sequence and properties of the human KB cell and mouse L cell D-loop regions of mitochondrial DNA. Nucleic Acids Res. **9**: 5411–5421.

Wong, T.W., and Clayton, D.A. 1986. DNA primase of human mitochondria is associated with structural RNA that is essential for enzymatic activity. Cell, **45**: 817–825.

Zardoya, R., and Meyer, A. 1998. Cloning and characterization of a microsatellite in the mitochondrial control region of the African side-necked turtle, *Pelomedusa subrufa*. Gene, **216**: 149–153.

Zardoya, R., Villalta, M., Lopez-Perez, M.J., Garrido-Petierra, A., and Montoya, J. 1995. Nucleotide sequence of the sheep mitochondrial D-loop and its flanking tRNA genes. Curr. Genet. **28**: 94–96.

Evolution in action: ancient DNA from frozen Adélie penguins bones in Antarctica

D.M. Lambert*§, P.A. Ritchie*§, C.D. Millar†, B. Holland¶ & C. Baroni‡

* *Institute of Molecular BioSciences, Massey University, Private Bag 11-222, Palmerston North, New Zealand*

§ *These authors contributed equally to the work*

† *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*

¶ *Institute of Fundamental Sciences, Massey University, Private Bag 11-222, Palmerston North, New Zealand*

In Antarctica, large numbers of well-preserved sub-fossil bones of Adélie penguins (*Pygoscelis adeliae*) underlie their existing and abandoned nesting colonies. Penguin remains have been radiocarbon dated up to 13,000 yr BP and we show that frozen bones harbour some of the best-preserved ancient DNA yet discovered. We report DNA sequence variation in the mitochondrial hypervariable region I (HVRI) of 80 Carbon-14 dated bones collected from 18 Antarctic locations. These were compared with sequences from 380 living Adélie penguins from 13 Antarctic colonies. Using ancient and modern penguin mitochondrial haplotypes, we demonstrate DNA sequence evolution through time. In addition, we estimate the rate of evolution of the HVRI using two population-based analyses: median networks and a distance approach. Our calculated rates are five times higher than previous indirect phylogenetic estimates, and demonstrate the power of this unique situation for the study of evolution.

Until recently, estimates of rates of nucleotide sequence evolution have been limited to comparisons among living taxa. Such rates have typically been calculated by measuring the genetic distances between extant forms, and calibrating this against postulated divergence times from a most recent common ancestor. Such divergence times are usually themselves estimated by reference to the age of fossil material[1]. Typically the time frame for such calibrations span millions of years. For example, using a study of

restriction fragment length polymorphisms, Shields and Wilson[2] measured the divergence between two groups of geese at 8% and calibrated this against fossil evidence, suggesting a common ancestor 4-5 million years ago. Consequently, the entire avian mt genome is regarded to evolve at a rate of 2% per million years, similar to the value commonly accepted for mammals[3]. This value of 0.02 substitutions / site / million years (s/s/Myr) was then used to calculate the rate of substitution for a portion of the HVRI, estimated at 0.208 s/s/Myr, on the basis that it evolves 10.4 times faster than the entire mitochondrial genome[4,5]. This latter rate lies within the accepted range of substitution rates for mammalian HVRI (0.025 – 0.26 s/s/Myr), estimated from phylogenetic methods[6,7]. Such indirect estimates inevitably have limited precision and hence lack reliability.

Recently, there has been controversy over the rate of change of the mitochondrial control region[8-10]. For example, using samples from known pedigrees, Parsons and coworkers[11] measured a rate of 2.5 s/s/Myr for the HVRI in humans. This value is similar to that recorded in other human pedigree studies[12] (1.2 s/s/Myr – 2.7 s/s/Myr) and is more than an order of magnitude higher than those previously suggested from phylogenetic studies. Jazin et al.[13] suggested this high rate was due to mutational hot spots, and argued that it is not supported by other studies. However, this debate has not been resolved[14,15]. For example, using a long-term series of *Caenorhabditis elegans* mutation accumulation lines, Denver et al.,[16] recently reported a mutation rate of mitochondrial protein coding regions that is two orders of magnitude higher than previous indirect estimates. Such rates for family lines might be expected to be very different from those calculated using phylogenetic methods because a combination of drift and selection is likely to eliminate variants over time[17]. This controversy, in conjunction with the inherent inaccuracy of the phylogenetic approach to rate estimation, raises the question of what evolutionary rates might be for intermediate time frames (eg tens of thousands of years).

The advent of ancient DNA technology[18], in principle, offers a new opportunity to directly and accurately estimate the rate of nucleotide substitutions using population approaches. However, to date there have been no examples of populations comprising a sufficient number and distribution of ancient samples of known ages. Because of the particular aspects of their life history and the extreme Antarctic environment, Adélie penguins represent an ideal model for such a study. This remarkable species represents approximately 80% of the total avian biomass of Antarctica in the summer months[19], and has an estimated population size of over 15 million[20]. During the austral summer, Adélie penguins nest in distinct colonies in ice free areas along a small proportion of the Antarctic coastline (Fig. 1). Colonies are characterised by extremely high densities, together with high mortality[21]. These factors have led to large deposits of sub-fossil

bones that have been continuously preserved in the cold Antarctic environment for at least 7,786 yr BP (TO-4960 Supplementary Information)[22].

In order to measure evolutionary rates in Adélie penguins, we sequenced the mitochondrial HVRI from both modern samples and ancient bone material. A major feature of the HVRI of populations of living Adélie penguins (n = 380) is the presence of two mtDNA lineages that have distinct geographic distributions. Type *A* (Antarctica) is present at all locations around the continent that we have sampled, while the *RS* (Ross Sea) lineage appears restricted to the Ross Sea (Fig. 1). These lineages each have high haplotype diversity (h_A = 0.995; and h_{RS} = 0.995) and average within-lineage sequence differences of 2.0% (*A* type) and 2.6 (*RS* type). These lineages show an average of 8.3% sequence divergence, and 6% when adjusted for within-lineage polymorphism.

The same two lineages occurred in 80 sub-fossil bones of Adélie penguins preserved within soils below penguin colonies[23,24]. Soil horizons are composed of droppings, feathers, egg fragments and other penguin remains mixed with sand, gravel and pebbles (Fig. 2). Abandoned penguin nesting sites are common landscape features along the Antarctic coasts[25,22]. For example, in Victoria Land 15 relict colonies are known. In this study penguin guano or other remains, from both occupied and abandoned colonies, were radiocarbon dated. Ages were assigned to nucleotide sequences from bones, either because the bones themselves were directly dated, or strata from which they were isolated were aged. A total of 99 radiocarbon ages were used for this purpose (supplementary information). We show that both mitochondrial lineages recorded from the modern populations were also present in the Ross Sea area 6082 yr BP. (Fig. 2).

The ancient DNA extracted from the cryopreserved Adélie penguin bones was of extraordinary quality. The polymerase chain reaction enabled amplification up to 1600 base pairs (bp) of the mitochondrial control region from younger radiocarbon dated bones (e.g. 523 yr BP). These younger bones consistently amplified a 663-1042 bp sequence (35% of bones younger than 2000 years). Even older bones (eg. 6082 yr BP) consistently amplified a 390bp sequence (45% of those older than 2000 years). In summary, 82% of sub-fossil bones could be sequenced for this shorter fragment. In addition, single copy nuclear loci were routinely amplified from bone samples, suggesting DNA of extreme quality. These results support the suggestion that low temperatures and arid environments are major factors in retarding post-mortem degradation of DNA[26].

Samples of ancient and modern populations of Adélie penguins provide a unique opportunity to estimate evolutionary rates of change over thousands of years. Both

median networks[27] and serial sample unweighted pair-group method of arithmetic averages (sUPGMA)[28] were used to estimate the rate. First, the evolutionary rate of change in the HVRI of Adélie penguins was measured by calculating the rate of change in bone samples, using the relationships among individual haplotypes in median networks. Such networks are appropriate for molecular data sets like that of Adélie penguins that are characterised by few phylogenetically informative sites and a broad spectrum of mutation rates[29]. Median networks are ideal because the nodes represent haplotypes that are either sampled or inferred as ancestral intermediates[29]. We employed median networks of the HVRI to infer ancient Adélie mitochondrial sequences that are ancestral to sequences from living individuals. Fig. 3 represents median networks generated from the HVRI data set. Sequences and sites with a large number of missing values were excluded from the analysis. This data set contained 318 modern and 67 ancient samples representing 152 haplotypes, 193 sites were used in the analysis. Thirteen mutually compatible sites define the phylogenetic relationship among the seven subgroups shown in Fig.3a. These 13 sites are completely consistent with the tree (Fig.3a), without requiring any parallel changes. Median networks were then constructed for each of the seven subgroups. The rate was estimated by calculating, first, the minimum number of substitutions required to account for the diversity of the modern samples, i.e. the number of edges in the median network needed to minimally connect modern samples. Second, from this connecting network, we subtracted the number of edges that account for substitutions in the ancient sequences. This provided the number of extra changes needed to explain the modern samples. For our data set, 151 extra changes were required. This value was averaged for the age of bone samples (2437 years), the number of sites used in the analysis (193) and the number of modern samples (318). This provided a rate estimate of 1.01 s/s/Myr. A modified jackknife procedure was used to estimate the 95% confidence interval of 0.85 – 1.16 s/s/Myr. This provides evidence for molecular evolution of Adélie penguins over this time period. In order to eliminate the possibility that this result is due to the different sizes of modern and ancient samples, we took 20 random sub-samples of 67 modern sequences and compared these to the 67 bone sequences. There was a significant difference between the number of changes required to explain the modern samples (mean 60.30 changes, with standard error 1.23) in comparison to the ancient samples (53 changes).

The second approach to rate estimation involved a general regression of the expected number of substitutions per nucleotide site (μ), against time, between serially preserved Adélie penguin samples. The sUPGMA approach of Drummond and Rodrigo[28] involves two steps. First, estimating the rate using a regression approach, independent of tree topology. Second, the construction of an adjusted distance matrix, and subsequent clustering using UPGMA. Serial sample UPGMA (sUPGMA)[28] allows the

reconstruction of a phylogeny in which tips terminate at different times, but constrains tips sampled at the same time to terminate at identical distances from the root. We have used this approach, as implemented in Pebble 0.4α[28] to estimate the rate of HVRI evolution and to present the relationships among all sub-fossil bone sequences and an equivalent representative number of modern samples (Fig. 4). We calculated a rate of change in the HVRI of 0.92 s/s/Myr. Using a delete-half jackknife on the unique-within-timepoint sequences, the 95% confidence intervals were –0.42 to 2.35.

We also investigated a maximum likelihood approach as a potential third estimator of the rate of change of the HVRI but without a well-defined single tree this method is difficult to apply. This approach, implemented in TipDate[30], uses a maximum likelihood tree of sequences, and consequently estimates μ, the rate of evolution over the entire tree using the differences between the ages of sub-fossil bones and modern samples. In contrast to sUPGMA, maximum likelihood approaches rely on tree topology in order to estimate evolutionary rates[30]. Our population data resulted in a large number of equally likely trees and therefore large variation in rates and extreme confidence intervals.

A strong feature of the data we present, is that the average sequence divergence between the modern *A* and *RS* Adélie penguin lineages is 6.0%, while that between older ancient bone samples (mean 4,200 yr BP; n = 35) is 5.1%. Hence these two lineages have accumulated 0.9% sequence divergence over this time period. Moreover, mitochondrial HVRI sequences from Adélie penguins are evolving in a clock-like manner in that 89% of all samples belonging to the *A* and *RS* lineages passed a relative rate test[31] (n = 800) and a likelihood ratio test[32] ($p > 0.05$). Assuming an evolutionary rate of 1.01 s/s/Myr, as estimated by median networks, the sequence divergence of 6.0% between the *A* and *RS* lineages, dates their separation to 59-60,000 yr BP. Using the sUPGMA rate of 0.92 s/s/Myr the same divergence dates at 65,000 years (Fig. 4). Both these estimates indicate divergence of the two lineages in the middle of the last glacial cycle[33,34]. This is consistent with the fact that at the Last Glacial Maximum there were few, if any, ice free areas in the Ross Sea, and Adélie penguins are likely to have been restricted to refugia on sub-Antarctic islands. In contrast, the slower phylogenetic rate of divergence, 0.208 s/s/Myr[5] would suggest that the lineages were separated 280-300,000 yr BP and consequently that their distinctness has been maintained through at least three glacial cycles. This is only possible given an absence of population bottlenecks over that period. However, this is highly unlikely given the wide-scale displacement of penguins from ice-free breeding areas at each 120 kyr glacial cycle[34].

Given that we have detected a significant difference between the genetic composition of ancient and modern penguin populations, this study represents the first demonstration of molecular evolution in a vertebrate species over geological time. We show that Adélie penguin sub-fossil bones comprise a large number of ancient mitochondrial lineages, some of which are extinct, while others, or direct ancestors of them, are still represented in the living populations (Fig. 4). In addition, to measure the rate of HVRI evolution, two successful approaches were employed. Although these were methodologically very different, they result in generally concordant estimates. These rates are approximately five times higher than the rate commonly accepted from phylogenetic studies. While other studies have attempted to document changes in populations over time, these have either lacked a large number of population samples[35], or they have employed ancient samples over only short time frames[36]. We suggest that the mitochondrial evolutionary rates presented here represent the best estimates to date, and that they are relevant to a broad range of evolutionary and phylogeographic studies. Our findings illustrate the potentials of this remarkable biological situation to study both the tempo and mode of evolution, and more generally highlight the importance of Antarctica as a natural laboratory for international science.

Methods

DNA extraction, PCR and sequencing

DNA from blood samples of 380 Adélie penguins was isolated using standard procedures[37]. Approximately 0.5g of each sub-fossil bone was ground and decalcified in 20mls of 0.5 M EDTA pH 8.0 overnight. The supernatant was removed after centrifugation and the sediment digested in SDS/Proteinase-K at 50°C overnight. The samples were extracted twice with Tris-saturated phenol and once with chloroform/isoamyl, followed by concentration on a Vivaspin-30 (Viva Science, U.K.) membrane. Fifty microlitres of the sample was purified using QIAGEN DNA mini kit and stored at 4°C. The HVRI mitochrondrial control region was amplified using the primers specific to the Adélie penguin (reference numbers correspond to the sequence deposited in GenBank, accession no. AF272143) AH530 (5'-CTGATTTCACGTGAGGAGACCG-3'), AH432 (5'-GTGTTCAAGCTCTGCCGTACC-3'), AL305 (5'-GGACCAGCTCCTAATCCCTTCG-3'), AH271 (5'-GCTGGTCCTTGTACCATGGAC-3'), AL93 (5'-CATTTAATGTACGTACTAGGAC-3'), and L-tRNA^{Glu} (5'-CCCGCTTGGCTTYTCTCCAAGGTC-3'). Amplifications from 1µl of the extracts was conducted in 25µl volumes using AmpliTaq (PE Biosystems), 1.5 mM $MgCl_2$, 2 mg/ml bovine serum albumin, 0.4 pmol/µl of each primer and 200 µM of each dNTP. PCR products were purified with the QIAquick PCR purification kit before direct sequencing using the Big Dye Terminator sequencing kit (PE Biosystems) and analysed on a 377A automated sequencer (PE Biosystems). Samples from the two major lineages were re-sequenced from a long PCR fragment to check for preferential amplification from a nuclear copy of the control region as described in Ritchie and Lambert[38]. Additional evidence that one of these lineages is not a nuclear copy comes from our finding that a 640 bp sequence from the apocytochrome b (*cob*) gene (n=10) shows a

fixed transitional substitution between the two lineages at the 3rd position of a Serine (corresponding to position 15214 of the chicken mtDNA).

Ancient DNA and sequence verification

All DNA extractions from sub-fossil material were conducted in a dedicated laboratory facility that underwent regular decontamination. Extraction and PCR blanks were continually screened and a proportion of the DNA extractions were conducted twice. All PCR targets were amplified twice, on separate occasions and each amplicon sequenced in the opposite direction to the first. Six ancient Adélie penguin bones were extracted in an independent laboratory at the University of Auckland and sequenced. In each case identical sequences were retrieved.

In our ancient samples, we estimated the level of sequencing error that results from *Taq* polymerase inserting the wrong nucleotide across from a damaged nitrogenous base. This results, for example, when deaminated cytocine forms uracil, which is, in turn, mis-read by *Taq* polymerase as thymine and consequently induces an apparent C to T transition. To control for this, we sequenced 313 base pairs of the small subunit ribosomal RNA (*rns*) gene in ancient material (n=18). Modern Adélie penguins show no intraspecific variation for this sequence (n=21). All ancient DNA sequences from the *rns* gene were also identical to modern sequences. Hence the background damage induced error rate in this study was assumed to be low.

Sub-fossil bone excavations and radiocarbon dating

We collected sub-fossil Adélie penguin bone samples from abandoned nesting sites both in the vicinity of presently occupied rookeries and in relict (fossil) rookeries in the Ross Sea area. We dug for sub-fossil bones using an accurate stratigraphic method of excavation. Pits from 1-2 to 6 m^2 were excavated, layer by layer, from the top. This method allowed us to identify and separate different individual remains within the same layer. In addition, we were able to exclude contamination from the top, to identify wedges of sediments reworked by periglacial processes, and to prevent mixing of materials from different layers.

Due to the depletion of ^{14}C in sea water and to the up-welling of deep, old oceanic water, radiocarbon ages of penguin remains require adjustment to calibrate for this "reservoir effect"[39,40]. The apparent age of Antarctic samples is older by 1000 to 1300 yrs than their true age. Several authors apply a correction of 1050, 1200, 1300 yrs, in order to compare the Antarctic dates with other sets of data obtained elsewhere in the world. However, since that the error varies with time, we used the calibration program Calib 4.2[41] to calibrate our data, as accurately as possible. We introduced a ΔReservoir (ΔR) value of 688 ± 55 to apply a correction factor appropriate for penguins[42], considering that the reservoir effect differs with the species considered. The parameter ΔR is required by Calib 4.2 and represents the constant difference in reservoir age between a regional part of the ocean and the world ocean. The indicated value derives from the weighted mean of seven ΔR parameters obtained from radiocarbon ages of penguin remains of known historical age.

Carbon dates were supplied by Geochron Laboratory - Krueger Enterprise Inc., Cambridge, Massachusetts (conventional and AMS, GX-), IsoTrace Radiocarbon Laboratory, Toronto (AMS, TO-) and NOSAMS, Woods Hole Oceanographic Institute (AMS, OS-) and the New Zealand Institute of Geological and Nuclear Sciences, Lower Hutt, New Zealand (AMS, NZA).

Acknowledgements

This research was made possible by a grant from the Marsden Fund to DML. We also gratefully acknowledge support from Massey University, Pisa University, the Italian Antarctic Research Programme, the National Science Foundation, Antarctica New Zealand and the US Coastguard. We especially thank A. Drummond and A.G Rodrigo for advice and for making available the pre-release of Pebble 0.4α software. For helpful discussions regarding data analysis and/or comments on the manuscript, we thank: R. Forsberg, J. Heine, M. Hendy, P. Lockhart, V.E. Neall and D. Penny. For technical assistance we are grateful to A. Cooper, S. Eyton, R. Marshall, L. Matisoo-Smith, J. Robins, R. Ward and L. Shepherd. For assistance in Antarctica and/or with sample collections we thank: P. Barrett, L. Davis, K. Kerry, C. Vleck, and E.C. Young. P.A.R. acknowledges the support of a Massey University doctoral scholarship. We thank K. Newman for assistance with graphics. B.R.H. would like to acknowledge A. Dress and the Program for Scientific and Technological Exchange between New Zealand and Germany. B.R.H also thanks V. Moultan (STINT grant) and K.Huber.

List of figure legends

Figure 1. Distribution of Adélie penguin breeding colonies during the austral summer in Antarctica. Black shading indicates areas of highest density. The current distributions of permanent ice shelves are shown by dotted parallel lines. The extent of the Ross Ice Shelf grounding line at different times is shown by dotted lines. The sites of collection, with latitude and longitudinal coordinates, are given, together with the number of both blood (b) and sub-fossil bone (sfb) samples: **1.** Welch Island (67°33'S, 62°55'E), b 21; **2.** Gardner Island (77°52'S, 68°34'E), b 2; **3.** Torgersen Island (64°46'S, 64°05'W), b 17; **4.** Cape Royds (77°33'S, 166°10'E), b 38; **5.** Cape Crozier (77°14'S, 166°28'E), b 29, sfb 54; **6.** Cape Bird (77°30'S, 162°10'E), b 122, sfb 39; **7.** Beaufort Island (76°56'S, 167°03'E), b 22, sfb 20; **8.** Franklin Island (76°07'S, 168°15'E), b 16; **9.** Spike Cape (77°18'S, 163°33'E), sfb 2; **10.** Dunlop Island (77°14'S, 163°28'E), sfb 22; **11.** Cape Roberts (77°02'S, 163°10'E) sfb 17; **12.** Cape Ross (76°43'S, 162°59'E), sfb 19; **13.** Depot Island (76° 42' S, 162° 57 E), sfb 3; **14.** Cape Day (76°14'S, 162°47'E), sfb 7; **15.** Cape Hickey (76°05'S, 162°38'E), sfb 5; **16.** Prior Island (75°41'S, 162°52'E), sfb 12; **17.** Cape Irizar (75°33'S, 162°57'E), sfb 9; **18.** Inexpressible Island (74°53'S, 163°45'E), b 21, sfb 37; **19.** Adélie Cove (74°45'S, 164°00'E), b 12; **20.** Northern Foothills (74°45'S, 164°00'E), sfb 63; **21.** Gondawa Station (74°38'S, 164°12'E), sfb 4; **22.** Edmonson Point (74°19'S, 165°04'E), b 100, sfb 11; **23.** Cape Wheatstone (72°37'S, 170°14'E), b 10; **24.** Cape Hallett (72°19'S, 170°12'E), b 29, sfb, 5. The relative proportions of the Antarctic (A) and Ross Sea (RS) mitochondrial lineages among living populations are shown in relation to the retreat of the Ross Ice Shelf after the Last Glacial Maximum (LGM).

Figure 2. Three soil profiles from abandoned Adélie penguin colonies on Inexpressible Island, Antarctica (collection site 17, Fig. 1) from which sub-fossil bones were obtained. Details of stratigraphy are shown: **1.** surface horizon composed of pebbles accumulated from penguin nests, **2.** organic horizon of guano, **3.** sandy gravel with guano and **4.** sandy gravel and boulders. Carbon-14 dates from penguin remains, with marine calibration [40] are given at a range of depths, together with the position in the profile of bones, their sample numbers and mt lineages.

Figure 3. Median networks of HVRI mtDNA sequences of ancient and modern Adélie penguin haplotypes. **a,** Overall organisation of subgroups of median networks. **b-h,** the seven subgroups of Adélie penguins studied. Double-headed arrows indicate a change at one of the 13 mutually compatible sites used to define the subgroups. Modern samples (●), sub-fossil bones (●), and hypothetical intermediates (●) are shown. Parallel lines indicate a change at the same site. Small circles represent 1-5 sequences, medium circles 6-20 sequences and large circles 20+ sequences. Blue circles indicate the point of connection between each of the subgroups.

Figure 4. Relationships among ancient (●) and living (●) HVRI mtDNA haplotypes of Adélie penguins using sUPGMA analysis. Numbers indicate the collection site of modern and ancient samples (as detailed in Fig. 1). Calibrated ages of ancient bones (yr BP) are shown.

1. Cann, R. L., Stoneking, M. & Wilson, A. C. Mitochondrial DNA and human evolution. *Nature* **325**, 31-36 (1987).
2. Shields, G. F. & Wilson, A. C. Calibration of the mitochondrial DNA evolution in geese. *J. Mol. Evol.* **24**, 212-217 (1987).
3. Brown, W. M., George, M. & Wilson, A. C. Rapid evolution of animal mitochondrial DNA. *Proc. Natl. Acad. Sci. USA* **76**, 1967-1971 (1979).
4. Quinn, T. W. The genetic legacy of Mother Goose phylogeographic patterns of lesser snow goose *Chen caerulescens caerulescens* maternal lineages. *Mol. Ecol.* **1**, 105-117 (1992).
5. Baker, A. J. & Marshall, H. D. in *Avian Molecular Evolution and Systematics* (ed. Mindell, D. P.) (Academic Press, San Diego, 1997).
6. Vigilant, L., Stoneking, N., Harpending, H., Hawkes, K. & Wilson, A. C. African populations and the evolution of human mitochondrial DNA. *Science* **253**, 1503-1507 (1991).
7. Stoneking, M., Sherry, S. T., Redd, A. J. & Vigilant, L. New approaches to dating suggest a recent age for human mtDNA ancestor. *Phil. Trans. Roy. Soc. B* **337**, 167-175 (1992).
8. Sigurðardóttir, S., Helgason, A., Gulcher, J. R., Stefansson, K. & Donnelly, P. The mutation rate in the human mtDNA control region. *Am. J. Hum. Genet.* **66**, 1599-1609 (2000).
9. Howell, N., Kubacka, I. & Mackie, D. A. How rapidly does the human mitochondrial genome evolve? *Am. J. Hum. Genet.* **59**, 501-509 (1996).
10. Macaulay, V. A. et al. mtDNA mutation rates - no need to panic. *Am. J. Hum. Genet.* **61**, 983-986 (1997).
11. Parsons, T. J. et al. A high observed substitution rate in the human mitochondrial DNA control region. *Nat. Genet.* **15**, 363-368 (1997).
12. Bendall, K. E., Macaulay, V. A., Baker, J. R. & Sykes, B. C. Heteroplasmic point mutations in the human mtDNA control region. *Am. J. Hum. Genet.* **59**, 1276-1287 (1996).
13. Jazin, E. et al. Mitochondrial mutation rate revised: hot spots and polymorphism. *Nat. Genet.* **18**, 109-110 (1998).
14. Parsons, T. J. & Holland, M. M. Response to Jazin et al. *Nat. Genet.* **18**, 110 (1998).
15. Meyer, S., Weiss, G. & von Haeseler, A. Pattern of nucleotide substitution and rate heterogeneity in the hypervariable regions I and II of human mtDNA. *Genetics* **152**, 1103-1110 (1999).
16. Denver, D. R., Morris, K., Lynch, M., Vassilieva, L. L. & Thomas, W. K. High direct estimate of the mutation rate in the mitochondrial genome of *Caenorhaditis elegans*. *Science* **289**, 2342-2344 (2000).
17. Nei, M. *Molecular Evolutionary Genetics* (Columbia University Press, New York, 1987).
18. Pääbo, S. Ancient DNA. *Sci. Amer.* **November, 1993**, 60-66 (1993).
19. Everson, I. (FAO Southern Oceans Fishery Survey Programme GLO/SO/77/1, Rome, 1977).
20. Williams, T. D. *The Penguins: Spheniscidae* (Oxford University Press, Oxford, UK, 1995).

21. Ainley, D. G., LeResche, R. E. & Sladen, W. J. L. *Breeding Biology of the Adélie Penguin* (University of California Press, Berkeley, Los Angeles and London, 1983).

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

23. Heine, J. C. & Speir, T. W. Ornithogenic soils of the Cape Bird Adélie penguin rookeries, Antarctica. *Polar Biol.* **10**, 89-99 (1989).

24. Ugolini, F. C. in *Antarctic Terrestrial Biology* (ed. Llano, G. A.) 181-193 (American Geophysics Research, Washington D. C., 1972).

25. Baroni, C. & Orombelli, G. Holocene raised beaches at Terra Nova Bay, Victoria Land, Antarctica. *Quat. Res.* **36**, 157-177 (1991).

26. Höss, M., Jaruga, P., Zastawny, T., Dizdaroglu, M. & Pääbo, A. DNA damage and DNA sequence retrieval from ancient tissues. *Nucleic Acids Res.* **24**, 1304-1307 (1996).

27. Bandelt, H.-J., Forster, P., Sykes, B. C. & Richards, M. B. Mitochondrial portraits of human populations using median networks. *Genetics* **141**, 743-753 (1995).

28. Drummond, A. & Rodrigo, A. G. Reconstructing genealogies of serial samples under the assumption of a molecular clock using serial-sample UPGMA (sUPGMA). *Mol. Biol. Evol.* **17**, 1807-1815 (2000).

29. Bandelt, H.-J., Macauley, V. & Richards, M. Median networks: Speedy construction and greedy reduction, one simulation, and two case studies from human mtDNA. *Mol. Phyl. Evol.* **16**, 8-28 (2000).

30. Rambaut, A. Estimating the rate of molecular evolution: Incorporating non-contemporaneous sequences into maximum likelihood phylogenies. *Bioinformatics* **16**, 395-399 (2000).

31. Wu, C.-I. & Li, W.-H. Evidence for higher rates of nucleotide substitution in rodents than in man. *Proc. Natl. Acad. Sci. USA* **82**, 1741-1745 (1985).

32. Felsenstein, J. Evolutionary trees and DNA sequences: a maximum likelihood approach. *J. Mol. Evol.* **17**, 368-376 (1981).

33. Jouzel, J. et al. Vostok ice core: a continuous isotope temperature record over the last climate cycle (160,000 years). *Nature*, 403-408 (1987).

34. Petit, J. R. et al. Climate and atmospheric history of the past 420,000 years from the Vostock ice core, Antarctica. *Nature* **399**, 429-436 (1999).

35. Leonard, J. A., Wayne, R. K. & Cooper, A. Population genetics of Ice Age brown bears. *Proc. Natl. Acad. Sci. USA* **97**, 1651-1654 (2000).

36. Thomas, W. K., Pääbo, S., Villablanca, F. X. & Wilson, A. C. Spatial and temporal continuity of kangaroo rat populations shown by sequencing mitochondrial DNA from museum specimens. *J. Mol. Evol.* **31**, 101-112 (1990).

37. Sambrook, J., Fritsch, E. F. & Maniatus, T. *Molecular Cloning: A Laboratory Manual* (Cold Spring Habor Labatory Press, New York, USA, 1989).

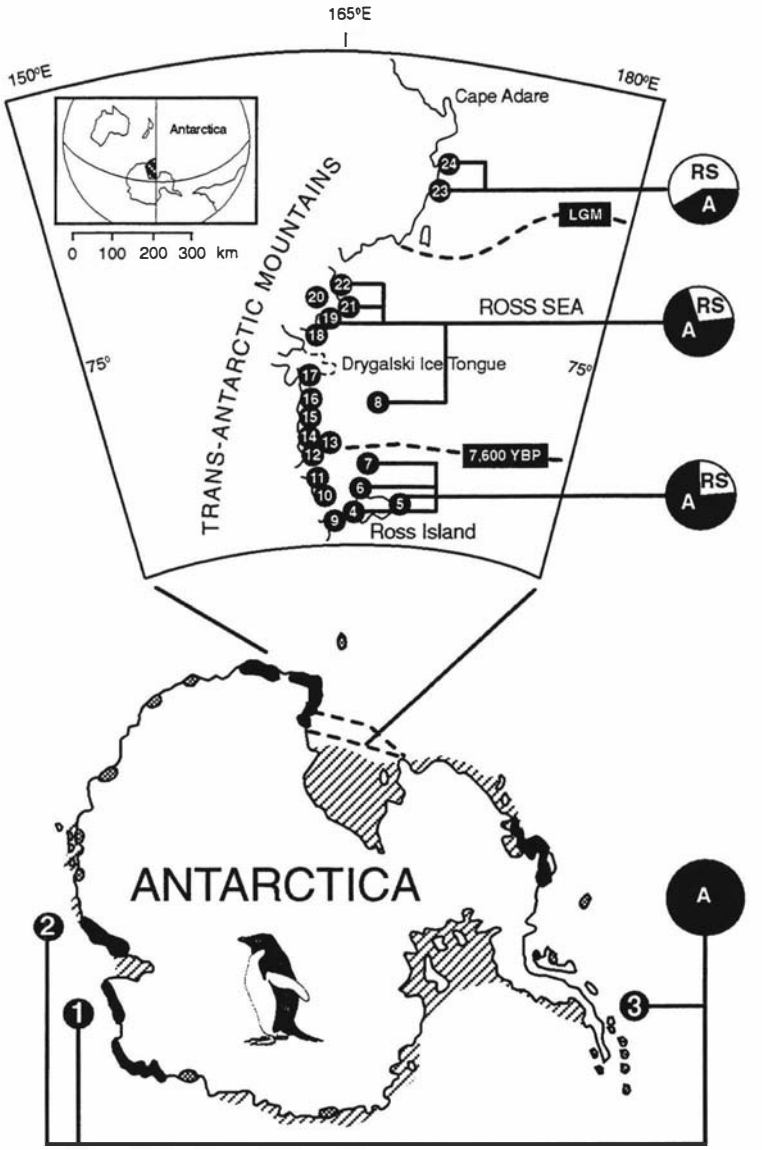
38. Ritchie, P. A. & Lambert, D. M. A repeat complex in the mitochondrial control region of Adélie penguins from Antarctica. *Genome* **43**, 613-618 (2000).

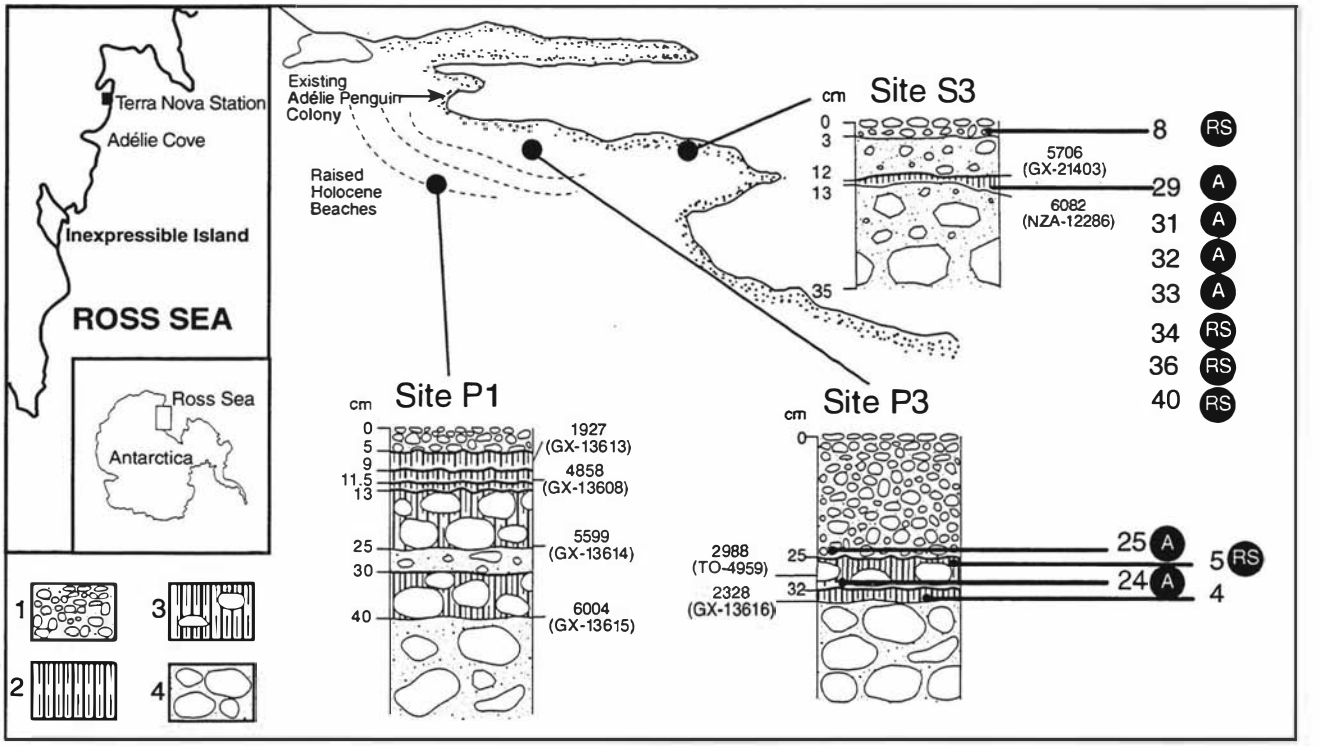
39. Harkness, D. D. Radiocarbon dates from Antarctica. *Brit. Ant. Sur. Bull.* **47**, 43-59 (1979).

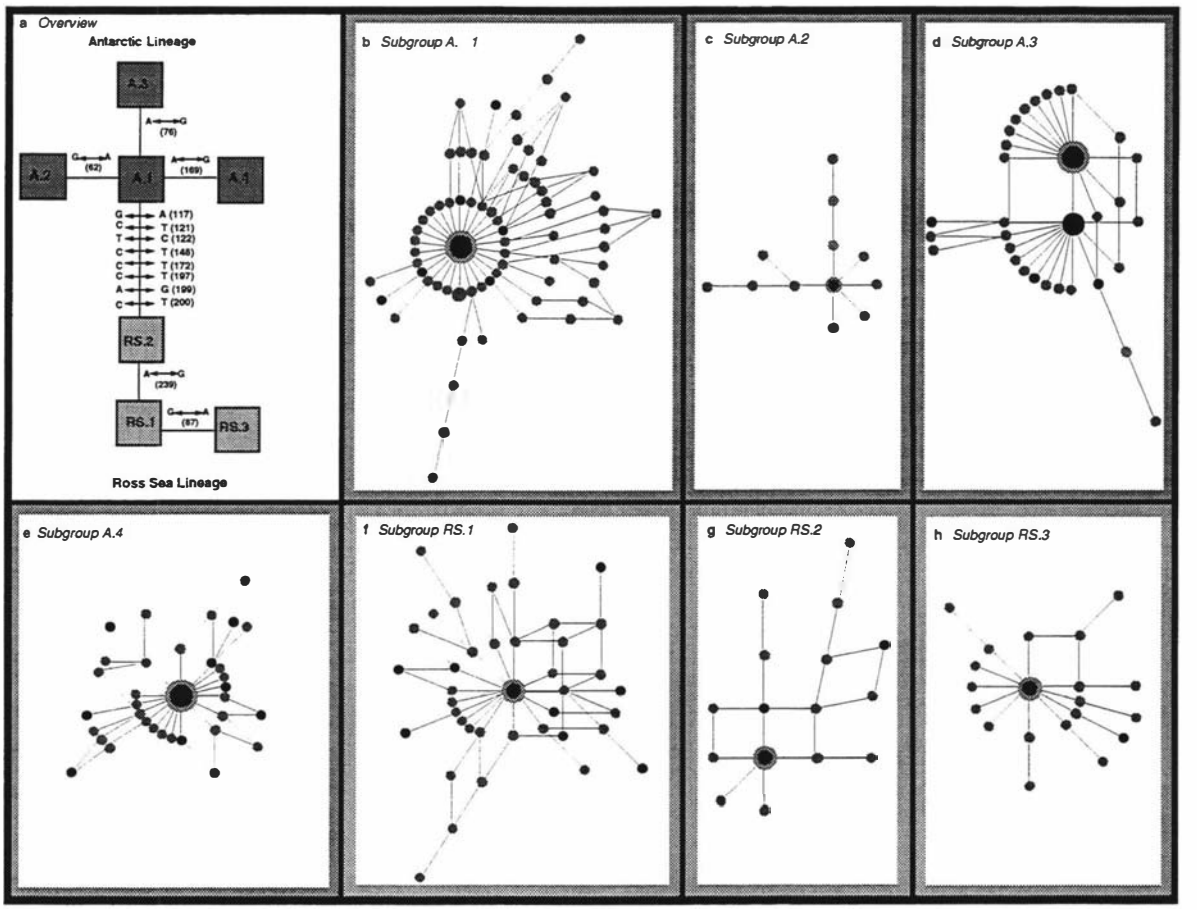
40. Stuiver, M., Pearson, G. W. & Braziunas, T. Radiocarbon age calibration of marine samples back to 9000 cal yr B.P. *Radiocarbon* **28**, 980-1021 (1986).

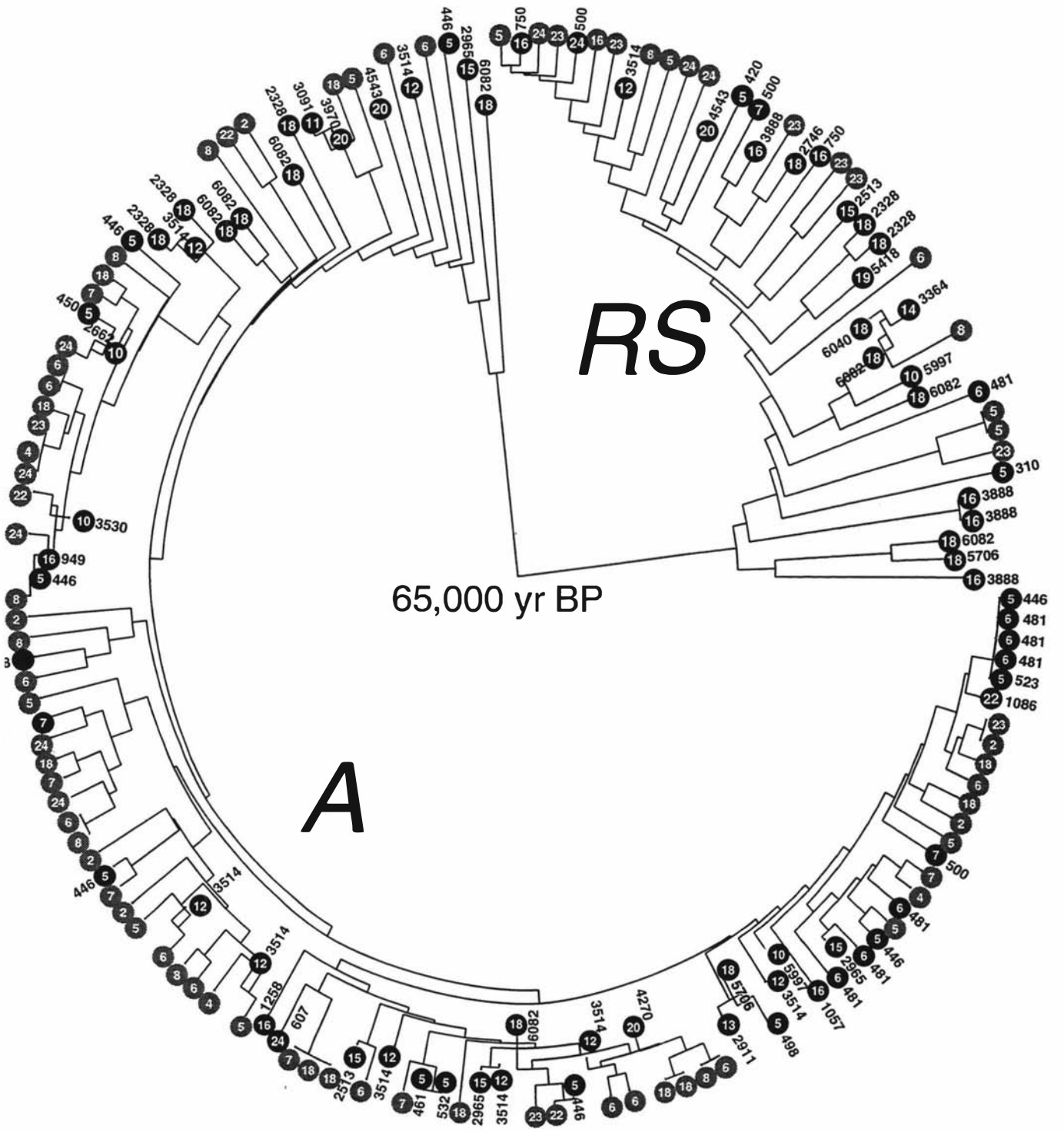
41. Stuiver, M. & Reimer, P. J. (Quaternary Isotope Laboratory, University of Washington, Washington, 2000).

42. Petri, A. & Baroni, C. 'Penguin', a Macintosh application for entry and presentation of radiocarbon-dated samples. *Radiocarbon* **39**, 61-65 (1997).


165°E
150°E
180°E
Antarctica
0 100 200 300 km
Cape Adare
24
23
RS
A
LGM
TRANS-ANTARCTIC MOUNTAINS
22
20
21
19
18
ROSS SEA
RS
A
75°
Drygalski Ice Tongue
75°
17
16
15
14
13
12
8
7,600 YBP
7
11
6
10
5
4
9
RS
A
Ross Island
ANTARCTICA
A
2
1
3


Terra Nova Station
Adélie Cove
Inexpressible Island
ROSS SEA
Ross Sea
Antarctica
Existing
Adélie Penguin
Colony
Raised
Holocene
Beaches
Site S3
cm
0
3
12
13
35
5706
(GX-21403)
6082
(NZA-12286)
8 RS
29 A
31 A
32 A
33 A
34 RS
36 RS
40 RS
Site P1
cm
0
5
9
11.5
13
25
30
40
1927
(GX-13613)
4858
(GX-13608)
5599
(GX-13614)
6004
(GX-13615)
Site P3
cm
0
25
32
2988
(TO-4959)
2328
(GX-13616)
25 A
24 A
5 RS
4
1
2
3
4


a Overview
Antarctic Lineage
A.3
A→G (76)
A.2
G→A (62)
A.1
A→G (169)
A.1
G→A (117)
C→T (121)
T→C (122)
C→T (148)
C→T (172)
C→T (197)
A→G (199)
C→T (200)
RS.2
A→G (239)
RS.1
G→A (87)
RS.3
Ross Sea Lineage
b Subgroup A. 1
c Subgroup A.2
d Subgroup A.3
e Subgroup A.4
f Subgroup RS.1
g Subgroup RS.2
h Subgroup RS.3


RS
A
65,000 yr BP