Microsatellite Evolution and Population Genetics of Ancient and Living Adélie penguins in Antarctica.

Lara Dawn Shepherd

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

2001

Errata

Page	
24	**Table 2.1** The length of the clone sequences that are known are as follows: TP500 121 bp RM3 223 bp RM6 169 bp FhU2 75 bp
78	**Table 4.4** Sequences are listed 5' to 3'
79	**Table 4.5** Sequences are listed 5' to 3'

Abstract

Microsatellites are widely used as genetic markers for examining a variety of biological questions. Despite their widespread use, little is known about the processes by which they evolve. An accurate understanding of these processes is essential for their correct use as population genetic markers. In this study, microsatellite loci from both living and cryopreserved (AMS ^{14}C dated at up to 6424 years BP ±80) Antarctic Adélie penguins (*Pygoscelis adeliae*) were examined in order to gain insights into temporal population genetics and the evolution of microsatellite loci.

Firstly, ancient DNA extracted from Adélie penguin subfossil bones was found to be extremely well-preserved and readily allowed the amplification of single-copy nuclear microsatellite DNA. Genotyping six microsatellite loci in ancient and living samples from three populations of Adélie penguins in the Terra Nova Bay region allowed a comparison of genetic change over time. Although the ancient sample sizes were limiting, several statistical tests indicated that the ancient and living populations from Inexpressible Island were genetically distinct. In addition, differentiation was also inferred between the three ancient populations that were examined, which is in contrast to the lack of differentiation found between the living populations. These genetic changes may be a result of population expansion out of ice-age refugia since the Last Glacial Maximum.

To study microsatellite evolution over a substantial time period, up to 500 living and 100 cryopreserved Adélie penguins were genotyped at six microsatellite loci. No novel electromorph alleles were detected in the ancient samples. Numerous alleles were sequenced from four of these loci in both Adélie penguins and several other species of penguin (Spheniscidae). Analysis of these sequences provided an insight into the mutational processes occurring at these loci. In particular, these allele sequences revealed extensive size homoplasy, both within Adélie penguins and between penguin species. At one locus, variation in the flanking region allowed discrimination between the mechanisms proposed for length change at microsatellite loci. Slippage was the most plausible mechanism for length change. In this same locus, instability was observed in the

region bordering the repeat tract with a transversional bias predominating. This bias may be a caused by inaccurate DNA replication resulting from structural features of DNA.

Acknowledgements

First, and foremost I want to thank my supervisor Professor Dave Lambert for giving me the opportunity to join this project at the last minute. Your infectious enthusiasm and guidance throughout this research has been greatly appreciated. The past year has been a very enjoyable introduction to the world of scientific research.

This research was funded by a Marsden Fund of New Zealand grant to Dave Lambert (96-MAU-ALS-0300). I would like to acknowledge financial support from a Massey University Masterate Scholarship, IMBS Postgraduate Scholarship, Massey University Affinity Card Alumni Scholarship, New Zealand Federation of University Women Manawatu Branch Scholarship (twice) and a Coombs Memorial Bursary.

I would also like to thank everyone in the Molecular Ecology lab: Amy Roeder, Dr Ritchie, Leon Huynen (yeah, yeah), Olly Berry, Hilary Miller, Jo Chapman, Jennie Hay, Niccy Aitken, Gillian Gibb, Steve Sarre and Quanah Hudson for all their help and advice throughout the year, and for creating such an enjoyable atmosphere in which to work.

I am indebted to Danielle at the Equine Blood typing centre for letting me assist her with microsatellite genotyping on their very new, very expensive, machine.

I would like to thank Professor Carlo Baroni for all his help on this project and for providing the soil profiles used in chapter three. Thanks also to Pete Ritchie and Kathleen Newman for providing maps.

This project could not have been done without assistance from a large number of people - so thanks to everyone who contributed to this project, for example by collecting samples or extracting DNA. These contributions are outlined in the following preface.
Lastly, I want to thank Leon Perrie, my parents and George for all the encouragement and assistance they have given me whilst doing this thesis. Your support is greatly appreciated (although I may not always show it!).

Preface

The research undertaken for this thesis was part of a much larger research project studying Adélie penguins. Other research being undertaken includes an examination of population structure in living Adélie penguins (Roeder et al., in press), calculation of the mutation rate of the mitochondrial control region using ancient and living samples (Ritchie, 2001) and the development of nuclear and mitochondrial genetic markers for penguins (Roeder et al., submitted). Consequently much of the work discussed in this thesis was carried out in collaboration with others.

The contributors to various aspects of this project are given below.

Sample collection

Adélie blood samples and subfossil bones were collected by Peter Ritchie, Paul Barrett (Massey University) and Craig Miller (University of Auckland). Subfossil bones were also collected by Carlo Baroni (University of Pisa). Blood or tissue samples from other penguin species were provided by the following people: Graeme Elliot, Kath Walker and Peter Moore (Department of Conservation); Boris Culik (Institut für Meereskunde an der Universität Kiel, Germany); Allan Baker (Royal Ontario Museum); Cindy Hull (University of Tasmania); Corey Bradshaw (Otago University); Janier González, Gerry Kooyman (Scripps Institution of Oceanography); John Darby and Ian Mclean (Otago Museum).

DNA extractions

The Adélie blood samples were extracted by Richelle Marshall, Peter Ritchie, Amy Roeder and Sarah Eyton (Massey University). DNA was extracted from the Adélie subfossil bone samples, PE7 to PE136, by Peter Ritchie. I extracted samples PE137 to PE196. The tissue and blood samples from other species were extracted by Amy Roeder and Peter Ritchie (Massey University), and Kerri-Anne Edge (Otago University). I re-extracted DNA from the Chinstrap and Emperor samples used in this thesis.

Microsatellite Genotyping

540 genotypes from living Adélie penguin are referred to in this thesis. I genotyped the 98 living samples from the Terra Nova Bay region (Inexpressible Island, Northern Foothills and Edmonson Point). I also genotyped all of the ancient samples. The remaining samples were genotyped by Richelle Marshall, Amy Roeder (Massey University), Amanda Mitchelson and Helen McPartlan (Victorian Institute of Animal Sciences).

DNA Sequencing

I did all the microsatellite DNA sequencing in both Adélie and other penguin species.

Data Analysis

I conducted all the analyses reported in this thesis.

Table of Contents

CHAPTER FOUR

A Study of Microsatellite Evolution in Ancient and Living Adélie Penguins.

CHAPTER FIVE

Summary and Discussion of Future Work

APPENDIX A

List of Figures

List of Tables

List of Abbreviations

A	Adenine
AMS	Accelerator mass spectrometry
bp	base pairs
BSA	bovine serum albumin
C	cytosine
^{14}C	Carbon-14
DNA	Deoxyribonucleic acid
G	guanine
IAM	infinite allele model
indel	insertion/deletion
KAM	K-allele model
kb	kilobases
LGM	last glacial maximum
M	moles
n	number of samples
p	Probability
SSCP	Single-stranded conformation polymorphism
SMM	stepwise mutation model
T	thymine
TPM	two phase model
UCO	unequal crossing-over
yrs BP	years before present

CHAPTER ONE

Introduction

1.1 A study of microsatellite evolution in Adélie penguins using ancient DNA.

An essential component of evolutionary biology involves understanding the way that mutations occur in DNA. However, the rare occurrence of mutations makes it difficult to gain information about the mechanisms of DNA mutation events. Ancient DNA provides a method by which to study mutational mechanisms over a significant time period. Adélie penguins are an ideal species for such a study because large numbers of subfossil bones exist in Antarctica. In addition, Antarctica has a climate particularly well-suited to the preservation of DNA. Microsatellite DNA sequences are ideal subjects for such studies because they are short in length, and thus easily amplified from degraded samples, and they have very high mutation rates. An understanding of microsatellite evolution is important because it is required for the application of microsatellite data to population genetic questions.

In this chapter the current literature on microsatellite evolution will be reviewed, followed by overviews about Adélie penguins and ancient DNA studies. The specific aims of this study will then be discussed in greater detail.

1.2 Microsatellite DNA.

Genetic markers

The development of molecular techniques has allowed scientists to routinely address numerous biological questions involving behaviour, life history and evolutionary relationships (Avise, 1994). A number of genetic markers have proven useful when investigating such questions. These include allozymes, RFLPs (Restriction Fragment

Length Polymorphism), mitochondrial DNA, and markers based on tandemly repeated DNA.

The tandemly repeated DNA routinely used as genetic markers can be divided into two main categories based on the size of the repeat unit: (1) *minisatellites* in which a unit of 15 - 70 bp is repeated giving arrays of 0.5-3.0 kb in length; and (2) *microsatellites* where a 1-6 base motif is repeated to give short blocks of 20-100 bp (McDonald and Potts, 1997). Microsatellites are often classified as perfect, imperfect, interrupted or compound based on the repeat array sequence. Perfect microsatellites consist of uninterrrupted stretches of repeat units, e.g. CACACACACA; imperfect microsatellites contain point mutations within one or more repeat units in an otherwise perfect repeat array, e.g. CACACA**T**ACACA. Interrupted microsatellites have a small number of inserted base pairs that do not fit the repeat structure, e.g. CACACA**TTT**CACA; compound microsatellites consist of adjacent repeated arrays, each containing different repeat types, e.g. CACACA**TAATAATAA.**

Microsatellite DNA

According to Goldstein and Pollock (1997), microsatellite DNA loci are becoming the principal genetic markers used in a wide range of studies. They are extremely versatile and have been used to examine questions ranging from individual-specific, such as identity and gender (e.g. Hagelberg *et al.*, 1991), to parentage and relatedness (e.g. Primmer *et al.*, 1995), population genetic structure (e.g. Estoup *et al.*, 1995a) and even phylogenetic relationships (e.g. Bowcock *et al.*, 1994). Microsatellites have also been used widely for the investigation of genetic disorders and for forensic analyses.

Microsatellites have many advantages over other markers. They are predominately selectively-neutral, and are abundant in eucaryote genomes where they are essentially randomly distributed (Amos, 1999a). In addition, microsatellites are often highly polymorphic as a result of their rapid mutation rates. Estimates of microsatellite mutation rates range from 10^{-2} to 10^{-5} events per locus per generation which is far higher than the rates of point mutations (10^{-9} to 10^{-10}) (Hancock, 1999). The majority of microsatellite mutations involve changes in allele length. Therefore, alleles can be easily amplified by PCR using primer sequences that flank the repeat region and then

identified by their electrophoretic mobility. Also, the use of PCR often allows the typing of small and degraded samples. In contrast, it is not feasible to amplify most minisatellites by PCR because the large size of minisatellite loci often exceeds PCR limitations (McDonald and Potts, 1997). Also, the size difference between many minisatellite alleles may not be detected on a gel because they are beyond the resolving power of most gel systems.

Microsatellite evolution

The inference of microsatellite data in evolutionary and population genetic studies is dependent upon accurate models of microsatellite evolution. The development of such models necessitates a thorough understanding of the mutational processes occurring at microsatellite loci. However, despite the widespread use of microsatellites as genetic markers, their mutational processes remain little understood (Amos, 1999b; Estoup and Cornuet, 1999; Ellegren, 2000).

Several mechanisms, which are not necessarily mutually exclusive, have been proposed to explain the length mutations observed at microsatellite loci. The predominant mechanism appears to be slipped strand mispairing ('slippage') during replication of a single DNA double helix (Levinson and Gutman, 1987). Slippage is believed to occur when the elongating (newly-synthesised) DNA strand dissociates from the template strand and misaligns in the microsatellite region. When the strands reanneal and synthesis continues, a small number of repeats will be either lost or gained on the elongating strand (Figure 1.1a). Whether a gain or loss occurs depends upon whether the misannealment results in a loop structure on the elongating strand, which will produce a longer product, or on the template strand, in which case the resulting product will be shorter (Hancock, 1999). Many of these errors are recognised and corrected by exonucleoytic proofreading and mismatch repair enzymes (Eisen, 1999). Observed mutation rates therefore represent the balance between the generation of allele length changes by slippage and the correction of some of these errors by the mismatch repair system (Schlötterer, 1998).

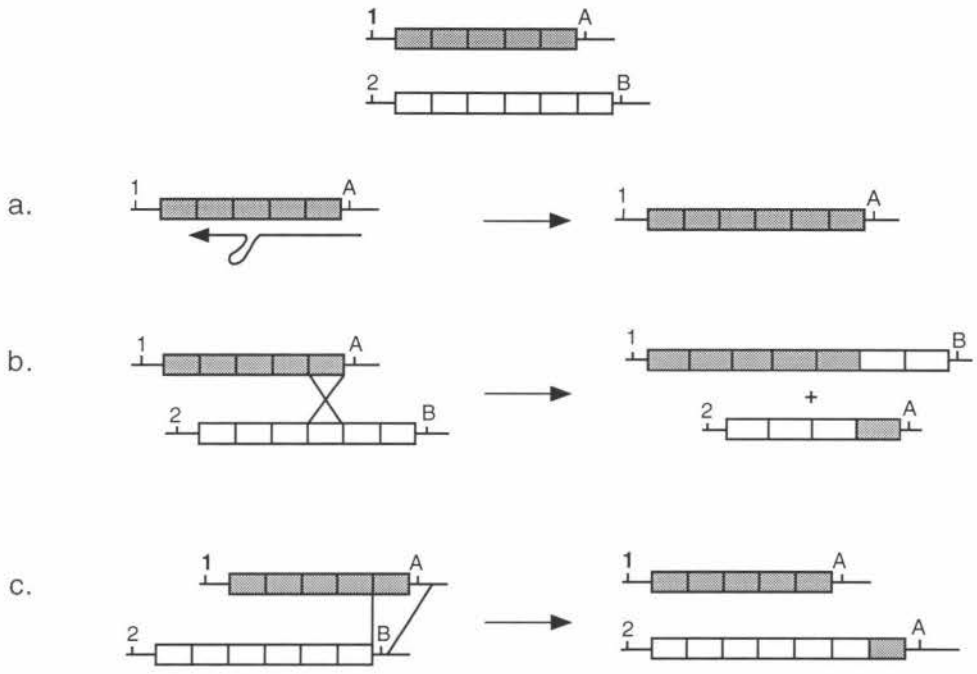

Figure 1.1 Possible mechanisms for length change mutations at microsatellite loci (adapted from Armour, 1996). The progenitor alleles are at the top.

a. Slippage occurs when the elongating strand slips in relation to the template strand.

b. Unequal crossing over (UCO) involves DNA exchange between misaligned DNA strands. It also results in the exchange of flanking markers.

c. Unequal gene conversion occurs when repeats are transferred from one DNA strand to another.

The other mechanisms proposed to operate at microsatellite loci involve recombination between DNA molecules. Recombination could cause microsatellite length change in two ways; by unequal crossing-over ('UCO') or by gene conversion. Unequal crossing-over (Figure 1.1b) involves recombination between misaligned DNA strands and results in an insertion in one DNA strand and a deletion on the other (Smith, 1976). Under this model the high mutation rate of microsatellites is believed to result from the presence of repeats which increase the chance that the two chromosomes will not be correctly aligned during recombination. Unequal gene conversion (Figure 1.1c) involves the transfer of information unidirectionally from one allele to another by recombination (Szostak, *et al.*, 1983).

Support for slippage as the primary mechanism of microsatellite mutation

Considerable support, from a number of different studies, exists for slippage being the primary mechanism of microsatellite evolution. Genetic evidence suggests that the stability of microsatellites is not affected by mutations in genes necessary for UCO such as *rad52* in *Saccharomyces cerevisiae* (Henderson and Petes, 1992). In contrast, microsatellite mutation rates have been observed to increase when genes involved in DNA replication error repair are defective (Strand *et al.*, 1993). This is consistent with the slippage model because it requires DNA replication (Eisen, 1999). If UCO is the main mechanism of microsatellite evolution then mutations should be more common in meiotic cells than mitotic cells because recombination is more frequent during meiosis than mitosis. However, microsatellite mutation rates have been observed to be similar in both mitotic and meiotic cells in *S. cerevisae* (Strand *et al.*, 1993). Also, recombination is expected to result in the exchange of regions flanking the repeat but this has not been observed in the majority of cases studied (Morral *et al.*, 1993; Kunst and Warren, 1994; but see Makova *et al.*, 2000).

Although the majority of studies support slippage as the primary mechanism, the occurrence of rare recombination events cannot be excluded. For example, Estoup *et al.* (1995b) sequenced a new mutation in the haploid male offspring of a diploid honeybee. Sequencing of the unique and maternal alleles revealed a complex rearrangement of the core sequence and the possible involvement of recombination. There are also some observations of microsatellite behaviour which are not consistent with slippage alone but may be better explained by recombination. For example, the highly mutable disease alleles of trinucleotide repeat disorders, such as Fragile X syndrome, have been demonstrated to comprise large expansions and occasional contractions (reviewed in Rubinsztein *et al.*, 1999). Large deletions have also been observed in yeast (Wierdl *et al.*, 1997). In addition, studies also suggest that mutations in di- and trinucleotide repeats may occur in a polar fashion (Kunst and Warren, 1994; Eichler *et al.*, 1995a; Eichler *et al.*, 1995b; Suzuki *et al.*, 2001). This similarity with minisatellites may reflect a possible role of gene conversion in microsatellites (Hancock, 1999). Finally, an *in vivo* study has provided evidence that recombination rates increase when microsatellites are present (Wahls *et al.*, 1990).

Mutational models

The use of microsatellite data for studies of population genetics and evolution requires a model of microsatellite mutation. Several models have been applied to microsatellite DNA markers. The *infinite allele model* (IAM), where each new mutation creates a new electrophoretically distinguishable allele not already present in the population, was originally developed for use with allozyme data (Kimura and Crow, 1964). In contrast, the *stepwise mutation model* (SMM) assumes that slippage will result in the addition or subtraction of a single repeat unit and that mutations occur symmetrically, with gains of repeats balancing losses on average (Kimura and Ohta, 1978). Thus alleles may mutate towards alleles that are already present in the population, thereby producing homoplasies i.e. alleles identical in state, but not descent (Estoup *et al.*, 1995b). Under the SMM it is predicted that closely related alleles will have a similar number of repeats. Closely related to the SMM is the *two phase model* (TPM) introduced by DiRienzo *et al.* (1994) which allows mutation steps of slightly larger sizes. A fourth model, the *K-allele model* (KAM), has also been considered for microsatellite loci (Crow and Kimura, 1970). Under this model there are K possible allelic states and each allele has a constant probability of mutating to any other of the K-1 alleles (The IAM is simply a form of the KAM where K is infinite). Under all these mutational models, except the IAM, homoplasy is expected to occur. Homoplasy has been shown to have an effect upon the parameters used to describe population structure (e.g. Angers *et al.*, 2000). For example, it may reduce genetic distances and therefore mask population structure. In this study, the term 'electromorph allele' will be used to describe PCR products of a given size determined by electrophoretic mobility, whilst 'sequence allele' will refer to electromorphs exhibiting the same sequence. 'Allele' will be used when either term is appropriate.

There are several methods by which the suitability of these models for use with microsatellite population data has been assessed. A direct method involves examination of sequence alleles from pedigree studies that differ by a single mutational event. These most strongly support the SMM or TPM (e.g. Weber and Wong, 1993; Primmer *et al.*, 1996a). However, most conclusions from this direct method have been based on averaging mutations across a large number of loci because the mutation rate per locus is

generally low. This pooling of results may conceal different mutation properties between loci (Estoup and Cornuet, 1999).

A second approach involves comparing observed electromorph allele distributions in populations with theoretical data generated under different models of microsatellite evolution (Shriver *et al.*, 1993; Estoup *et al.,* 1995a). Conclusions from such statistical tests of mutation models appear to be contradictory or inconclusive, which suggests that mutation processes occurring at these loci may be more complex than those assumed for the theoretical model being tested (Estoup and Angers, 1998).

Other factors influencing microsatellite mutation

There are several other factors, that the classical forms of the above models do not take into account, that have been proposed to affect microsatellite evolution. Studies of these factors, which are discussed below, have often produced conflicting results, making it difficult to incorporate them into mutational models.

Pedigree studies by Weber and Wong (1993) suggested that microsatellites with tetranucleotide repeat motifs had a higher mutation rate than those with dinucleotide repeats. However, the opposite was found by Chakraborty *et al.* (1997) who observed an inverse relationship between mutation rate and size of repeat unit. The conflicting results of these two studies may be explained by the confounding effect of repeat number (Schlötterer, 2000). A recent study comparing loci with the same number of di- and tetranucleotide repeats found that the mutation rate of the dinucleotide repeats was significantly higher (Lee *et al.*, 1999). In addition, the base composition of the repeat unit appears to correlate with levels of polymorphism, with most polymorphic tri- and tetranucleotide repeats being AT-rich (Gastier *et al.*, 1995; Sheffield *et al.*, 1995). Polymorphism levels of perfect microsatellites have also been observed to increase with an increase in the average number of repeats (Goldstein *et al.*, 1995; Wierdl *et al.* 1997; Schlötterer *et al.*, 1998). This has been suggested to occur because there is greater opportunity for slippage to occur on longer repeat tracts (Primmer and Ellegren 1998). Another factor influencing microsatellite polymorphism appears to be the purity of the repeat array. Interrupting bases appear to lower the level of polymorphism, probably by reducing the chance of misalignment (Jin *et al.*, 1996).

The above factors may explain the observed differences in mutation rate between loci and between alleles at a locus, as well as differences between the same locus in different species.

In addition to the above factors, asymmetry of microsatellite mutations is not considered in the classical mutational models. Observed mutations in pedigree studies have suggested that there is a mutational bias towards gain of repeats (Weber and Wong, 1993; Amos *et al.*, 1996; Primmer *et al.*, 1996a; Twerdi *et al.*, 1999). However, these studies involved hypervariable or very long microsatellites, and this mutational bias may not be present in all types of microsatellites.

If a mutation bias towards the gain of repeats does exist, then theoretically it would result in microsatellite arrays expanding to very large sizes. However, this is not observed as most microsatellite loci contain only a few tens of repeats. Several mechanisms have been proposed to account for size constraints. A few studies have assumed that selection prevents infinite growth of the microsatellite (Bowcock *et al.*, 1994; Garza *et al.*, 1995), although the only direct evidence of selection restricting allele length is from human diseases associated with trinucleotide expansions (Estoup and Angers, 1998). More recently it has been proposed that substitutions causing imperfection in the repeat array decrease the mutation rate (Bell and Jurka, 1997; Krugylak et al,. 1998) and may even lead to degeneration of the array (Taylor *et al.*, 1999a).

However, Xu *et al.* (2000) studied large numbers of mutations in human tetranucleotide loci and suggested that there is no mutational bias i.e. the overall expansion rate does not differ from contractions. They propose that alleles shorter than a certain critical repeat length are prone to expansion whereas longer alleles tend to contract.

Several variants of the SMM have been developed and take into account the following factors: an upper boundary to the size of microsatellites (Nauta and Weissing, 1996), the increase in mutation rate with repeat size (Falush and Iwasa, 1999), the occurrence of point mutations (Krugylak *et al.*, 1998), multistep mutations and directionally biased mutations (Kimmel and Chakraborty, 1996; Kimmel *et al.*, 1996).

Although there have been many studies examining microsatellite evolution, the mechanisms involved are still not fully understood. Slippage appears to be the primary mechanism, although several studies have suggested a possible role of recombination that may indicate mutations are far more complex than originally thought.

In this study, the evolution of microsatellites in penguins (Spheniscidae) will be examined using two different methods. Firstly, microsatellite allele sequences from a number of penguin species will be mapped on to an independent phylogeny. Secondly, microsatellite electromorph alleles from ancient and living Adélie penguins (*Pygoscelis adeliae*) will be amplified and compared.

1.3 A phylogenetic study of microsatellite evolution in penguins.

Penguins (Spheniscidae) are a monophyletic group consisting of 17 living species which live exclusively in the Southern Hemisphere (Williams, 1995). The common and scientific names of the penguins are given in Appendix A. Individuals from the penguin species have previously been genotyped at 6 microsatellite loci (Roeder *et al.*, submitted). Sequencing of alleles at some of these loci, followed by the mapping of sequence variation on to an independent phylogenetic tree provides a method by which to examine microsatellite evolution.

1.4 Ancient DNA of Antarctic Adélie penguins provides a new perspective in analyzing microsatellite evolution.

The advent of molecular genetic techniques has enabled investigations into population genetics and evolution. Until recently, these analyses of past events have been limited to inferences made from present day genetic patterns. The development of ancient DNA technology (reviewed below in section 1.5) has provided a new method by which to examine microsatellite evolution. In this study subfossil bones from Antarctic Adélie penguins (*Pygoscelis adeliae*) will be used to examine microsatellite evolution and temporal population genetics. Essential to this study are well-preserved subfossil samples. Antarctica provides an ideal environment for the preservation of ancient DNA because it is cold, dry and the climate has remained stable since the last glaciation (Baroni and Orombelli, 1994).

Adélie penguins have a circum-Antarctic distribution and an estimated minimum breeding population of 2.47 million pairs (Fraser and Trivelpiece, 1996). They form dense breeding colonies in ice-free areas along the coastline during summer months. For example, the colonies on Ross Island total approximately 250 000 breeding pairs with individual colony sizes recorded as ranging from around 4 000 to over 150 000 pairs. Although colonies are large, they occupy only 3% of the Antarctic coastline. Natal philopatry is believed to be a dominant feature of these colonies: an examination of natal return to the Cape Crozier colony on Ross Island over five years estimated that 96% of penguins that survive to adulthood return to their natal colonies to breed (Ainley *et al.*, 1983). Consequently, significant deposits of stratified subfossil bones have been accumulating at colony sites over a considerable amount of time. Carbon-14 radiometric dating has indicated that some of these sites have been occupied for the last 8 000 years (Spellerberg, 1970; Heine and Speir, 1989). The high level of natal philopatry, combined with the small area of coastline that is free of ice and thus suitable for breeding, assures the geographical separation of colonies (Roeder *et al.*, in press). However, recent genetic evidence suggest that Adélie penguins do not exhibit genetic differentiation between populations, indicating that there may be significant levels of gene flow between them and/or the large effective population sizes has resulted in little genetic drift (Roeder *et al.*, in press).

The Adélie penguin subfossil deposits, in combination with the DNA-preserving climate of Antarctica, provide an ideal situation in which to examine population genetics and the molecular evolution of microsatellites over a significant time frame. Although the populations may not be genetically isolated, and thus new alleles may be introduced into particular populations by migration, this does not pose a problem for such a study because numerous samples from living Adélie penguins have been collected from throughout their range.

1.5 Ancient DNA

The first report of the retrieval of ancient DNA sequences (Higuchi *et al.*, 1984) led to much excitement at the prospect of adding a temporal aspect to studies of molecular genetics and evolution. This initial study involved cloning a DNA sequence from a relatively young (150 year old), well preserved, museum specimen. Several studies

followed and led to the conclusion that ancient DNA is often highly degraded which may result in a low cloning efficiency and errors in cloned sequences (Pääbo, 1989).

The development of the polymerase chain reaction (PCR) provided a major technological advance to the field of ancient DNA research by allowing the amplification of specific sequences from small amounts of intact DNA template. PCR has allowed the amplification of DNA sequences from a range of ancient samples, e.g. bone (Höss and Pääbo, 1993) and coprolite (Poiner *et al.*, 1998) dating back to 50 000 years BP in some cases (Höss and Pääbo, 1993; Krings *et al.*, 1997; Adcock *et al.*, 2001). However, the degradation of DNA imposes a limit to the age of samples from which DNA can be retrieved.

DNA degradation

Within living organisms DNA decays spontaneously (undergoes random chemical changes), but the majority of these mutations are corrected by DNA repair mechanisms. After death, DNA repair ceases, but the decay of DNA continues. Both *in vitro* experiments under controlled conditions and biochemical analyses of ancient samples have demonstrated that DNA is a chemically unstable molecule that decays mainly through hydolysis and oxidation (Pääbo, 1989; Lindahl, 1993; Tuross, 1994; Höss *et al.*, 1996).

The most common type of hydrolytic damage is depurination. This process involves the cleavage of phosphodiester bonds between purine nucleotides and sugar residues and results in the release of free bases (adenine and guanine). These baseless sites are susceptible to DNA strand breakages that leads to the fragmentation of the DNA molecule. Oxidation also causes DNA strand breaks, as well as destroying the ring structure of nucleotide bases. Oxidation results from both direct modification by ionising radiation as well as damage from free radicals released through the interaction of ionising radiation with water molecules.

Lindahl and Nyberg (1972) calculated a spontaneous depurination rate using modern bacterial DNA in solution, and extrapolated these results to set an age limit of 100 000 years for specimens in which endogenous DNA could survive. However, the results of

experiments using artificial conditions should be approached with caution because the environmental conditions to which the ancient DNA has been exposed may be far more complex. For example, several studies have suggested that DNA degradation does not occur at a linear rate. Studies of archeological specimens led Pääbo (1989) to propose that DNA undergoes an initial phase of rapid decay immediately after death, with this rate then decreasing - the 'plateau effect'.

Although DNA has been demonstrated to decay spontaneously over time, numerous publications have provided evidence that the environmental conditions in which a sample is preserved have a greater effect on DNA preservation than the age of the sample (Colson *et al.*, 1997; Poiner, *et al.* 1996; *Höss et al.,* 1996; Tuross, 1994). The few studies that have examined the effect of individual environmental factors on DNA degradation have concluded that favourable conditions for DNA preservation include dryness, low temperature, neutral or slightly alkaline pH and the absence of microorganisms (Burger *et al.*, 1999).

The relative influence of each of the above factors is difficult to assess but several researchers stress the importance of low temperatures during burial. According to Höss *et al.* (1996), a decrease in temperature of 20°C is estimated to reduce the rate of several chemical reactions involved in DNA decay 10- to 25-fold. This claim is supported by the fact that some of the oldest substantiated reports of ancient DNA have come from 50 000 year old woolly mammoth remains preserved in Arctic permafrost (e.g. Hagelberg *et al.*, 1994; Höss *et al.*, 1996).

The degradation of DNA is also influenced by the type of tissue in which it is preserved. Initial ancient DNA studies used soft tissue specimens (Higuchi *et al.*, 1984; Pääbo, 1985) but more recently bones and teeth have become more popular sources of DNA. This is because, on average, the molecular weight of ancient DNA extracted from bone exceeds that of soft tissues. It has been proposed that the binding of DNA to hydroxyapatite mineral surfaces in bone stabilises the DNA molecules (Tuross, 1994).

Ancient DNA Sequences

By far the majority of ancient DNA studies in vertebrates have focused on

mitochondrial DNA sequences. With the low yield of ancient DNA, typically 1-5% of the DNA able to be extracted from modern tissues (O'Rourke *et al.*, 1996), mitochondrial DNA sequences, which are present in several hundred to thousands of copies per cell, are more easily retrieved than single-copy nuclear genes. In addition, mitochondrial sequences have been used extensively for systematic and population genetic studies in extant species. Therefore, a large data set of modern mitochondrial sequences exists for comparison with ancient sequences.

Nuclear DNA sequences are desirable because they are a complement to mtDNA data and provide further, independently inherited, characters for phylogenetic and population genetic analysis (Cooper, 1997). However, nuclear single-copy sequences, present in only two copies per cell in diploid organisms, are more difficult to amplify from degraded samples. It was believed that sequencing of single-copy genes from ancient samples would not produce consistent or reliable results (Landweber, 1999), but improved extraction methods have led to numerous studies employing single-copy sequences (e.g. Faerman *et al.*, 1995; Zierdt *et al.*, 1996; Burger *et al.*, 1999). Nevertheless, the number of samples that are expected to be able to yield such sequences is likely to be smaller than those from which mitochondrial DNA can be amplified. The majority of ancient nuclear DNA sequences reported have been from recent (Holocene-aged) samples. One notable exception is the retrieval and characterisation of both multiple and single copy nuclear DNA sequence from 13 000 year old mammoth samples (Greenwood *et al.*, 1999).

The short length of nuclear microsatellite sequences make them advantageous for ancient DNA studies (Schmerer *et al.*, 1999). In addition, these sequences are rapidly evolving making them particularly useful for studies of evolution and population genetics (Cooper, 1997).

Contamination and genotyping errors associated with low DNA quantities

The polymerase chain reaction (PCR) has become the method of choice for ancient DNA research because its extreme sensitivity allows amplification of small quantities of target DNA. However, this sensitivity means that retrieving true ancient sequence may be difficult. Contaminating DNA may be amplified in an ancient DNA extract,

especially if there are few endogenous ancient DNA molecules present (Austin *et al.*, 1997).

In addition, a number of studies have identified genotyping errors associated with amplifying nuclear DNA microsatellite loci from trace amounts of DNA (Navidi *et al.*, 1992; Taberlet *et al.*, 1996; Gagneux *et al.*, 1997; Goossens *et al.*, 1998). Under such conditions two main types of error have been shown to occur. The most frequent error, allelic dropout, occurs when only one allele in a heterozygote is amplified and is a consequence of unequal amplification of alleles (Taberlet *et al.*, 1996). Allelic dropout results in the incorrect typing of a heterozygous individual as homozygous.

The second type of genotyping error results from the tandem repeat structure of microsatellites. This structural feature results in the amplification of artifact products, called stutter or shadow bands that are usually one or two repeat units smaller than the microsatellite allele. This phenomenon is characteristic of microsatellite loci, in particular dinucleotide loci, and is believed to be a result of enzyme slippage during PCR (Foucault *et al.* 1996). If slippage occurs during the first few cycles of a PCR in a sample where the DNA quantity is low, then a greater quantity of artifact product may be generated than the true allele, resulting in a 'false allele' and a corresponding incorrect genotype.

Reliable amplification and genotyping of ancient DNA samples

As the field of ancient DNA has matured, a number of criteria have been developed by researchers to reduce the risk of contamination and to verify sequences as truely ancient (Cooper and Poiner, 2000). These criteria are summarised below.

A dedicated work area. Prior to amplification all ancient DNA research should be carried out in a dedicated, physically isolated, laboratory to avoid contamination with modern DNA, in particular PCR products (Austin *et al.*, 1997).

Controls. Negative controls during both the extraction and PCR set-up are essential for detecting contamination (Hummel and Herrmann, 1994).

Appropriate molecular behaviour. If single-copy nuclear DNA is able to be amplified then mitochondrial DNA sequence should also be able to be amplified. In addition, an inverse relationship should be demonstrated between the efficiency of amplification and sequence length (Cooper and Poiner, 2000). The DNA sequences should also make phylogenetic sense. That is, they should be similar to sequences from the same or closely related species (Yang, 1997).

Replication. Sequences should be able to be reproduced from the same, as well as different, extractions from a sample. In addition, samples should also be independently re-extracted and sequenced in another laboratory (Austin *et al.*, 1997).

Preservation indicators. Histological preservation, but not overall appearance, of bone appears to be indicative of DNA preservation (Hagelberg *et al.*, 1991). Also, apparent correlations between both the level of hydantoins (Höss *et al.*, 1996) and the rate of amino acid racemisation (Poiner *et al.*, 1996) with DNA hydrolysis may allow indirect assessment of DNA survival.

All the above criteria have been designed to reduce the risk of contamination. However, apart from replication, these criteria will not aid in detection of microsatellite genotyping errors. The quantity of extracted ancient DNA determined using conventional methods, such as fluorometry, may not give an accurate indication of whether the samples may be at risk from genotyping errors for the following three reasons: a) the DNA quantity is often too low; b) large amounts of DNA may be present but be too degraded to amplify; c) much of the DNA present may be microbial DNA (Taberlet and Luikart,1999).

One strategy to increase genotyping accuracy is to perform several independent amplifications from each sample (the 'multiple-tubes approach') (Navidi *et al.*, 1992). Taberlet *et al.* (1996) suggested a procedure based on the multiple-tubes approach (Figure 1.2) which they propose obtains reliable genotypes with a confidence level of 99%.

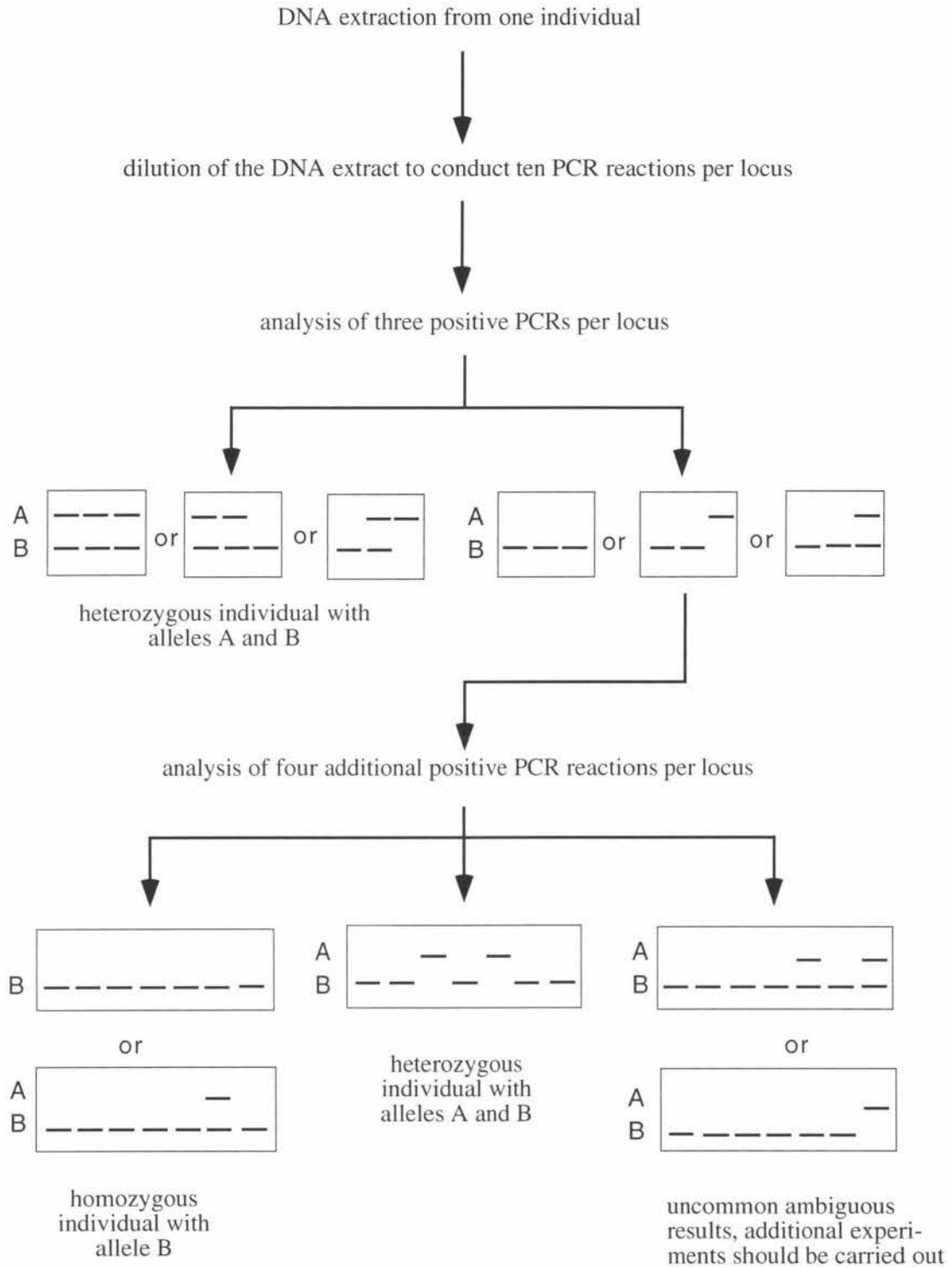

Figure 1.2 Flow diagram of the multiple-tubes approach procedure (adapted from Taberlet *et al.*, 1996). Note that for an individual that is homozygous a total of seven PCR reactions must be conducted to reach a 99% confidence interval.

Questions addressed by ancient DNA research

Many areas of research have benefited from the direct historical perspective provided by ancient DNA including the systematics of extinct species (e.g. Cooper *et al.*, 1992), conservation genetics (e.g. Cooper *et al.*, 1996) and the ecology of extinct plants and animals (e.g. Poiner *et al.*, 1998; Hofreitner *et al.*, 2000). Some of the least explored, but perhaps the most exciting, areas of research where ancient DNA has great potential to be of use are the evolution of DNA sequences and populations (Wayne *et al.*, 1999).

Mutations occurring in physically isolated populations over time provide a direct measure of tempo and mode of mutations. Rates of evolution for the mitochondrial control region have been estimated using ancient DNA from humans (Hauswirth *et al.*, 1994) and cattle (Bailey *et al.*, 1996). However, neither population was likely to have been isolated and gene flow was likely to have provided new alleles in both cases (Wayne *et al.*, 1999).

Ancient DNA also provides a method with which to examine temporal population genetics, including changes in gene frequency over time (Villablanca, 1994). The majority of these studies have used museum specimens, and thus have been limited to examining changes over the last 100 years. Studies of the San Clemente loggerhead shrike (Mundy *et al.*, 1997) and the greater prairie chicken (Bouzet *et al.*, 1998) have demonstrated loss of genetic diversity in post-bottleneck extant populations relative to pre-bottleneck museum specimens. Two recent studies have examined population genetics over a much longer period. Leonard *et al.* (2000) examined mitochondrial sequence in Pleistocene brown bear samples. These results suggest that genetic diversity was greater in bear populations 36 000 years ago but by 15 000 years ago was similar to the present day diversity. Wang *et al.* (2000) studied human population change by sequencing mitochondrial DNA from bones collected in a single region of China. A comparison of sequences from the present with 2 000 and 2 500 year old sequences indicated that the genetic backgrounds of the populations living at each of these time periods were distinctly different.

Ultimately, ancient DNA may provide a means by which to test existing theoretical models that reconstruct population genetic change from present-day patterns of genetic

diversity (Wayne *et al.*, 1999). Essential to studies of molecular evolution and population genetics over time are a large number of well-preserved samples spanning a considerable time period.

1.6 Research objectives

The objectives of this thesis are as follows:

• To reliably amplify and characterise ancient single-copy microsatellite DNA from subfossil Adélie penguin bones by adhering to the 'criteria of authenticity' developed by researchers in the ancient DNA field.

• To examine temporal population genetic change at three Adélie penguin colonies by comparing microsatellite DNA data from ancient and living Adélie penguins.

• To investigate mutational mechanisms occurring at microsatellite loci by comparing electromorph alleles from ancient and living Adélie penguins.

• To explore the evolution of four microsatellite loci on a broader scale by sequencing alleles from a number of species of the Spheniscidae, and for one locus mapping the sequences on to a phylogeny.

CHAPTER TWO

The Extraordinary Preservation of Adélie Penguin Ancient DNA.

2.1 Introduction

Studies of temporal population genetics and evolution require large numbers of well-preserved specimens. Unfortunately, ancient DNA is often highly degraded and thus is difficult, or impossible to amplify.

The observation that a cold, dry environment is particularly favourable for DNA preservation (e.g. Burger *et al.*, 1999) suggests that the Antarctic is likely to be an excellent source of ancient DNA. This prediction will be tested by characterising nuclear microsatellite DNA from Antarctic Adélie penguin subfossil bones ranging in age from 300 to over 6000 yrs BP. Additionally, the histology of representative bone specimens will be examined to test their usefulness as a predictor of DNA preservation.

Essential to retrieving ancient DNA are successful extraction and amplification techniques. The components of the PCR amplification mix have been suggested to influence the product yield and/or specificity of amplification (Hummel *et al.,* 1996). For example, bovine serum albumin (BSA) has previously been demonstrated to be essential for PCR amplification from Adélie subfossil bone extractions (Ritchie, 2001). In this chapter the influence of the type of Taq DNA Polymerase used for amplification was investigated. In addition, the amplification success rate of microsatellite loci was compared both with sample age and the length of amplifiable mitochondrial DNA sequence. Also, the prediction that an increase in the number of PCR cycles gives an increase in the quantity of PCR product is discussed in relation to genotyping errors.

2.2 Materials and Methods.

Sample collection and ^{14}C dating

A total of 329 subfossil Adélie bone samples were excavated from 17 sites along the Ross Sea coast (see preface). The sites selected (Figure 2.1) were deserted nesting areas that were either in abandoned rookeries or near occupied rookeries. Pits one to six m^2 in area were dug in order to locate and collect bones. Care was taken to keep the stratigraphic layers intact in order to avoid cross-contamination between layers. The depth of each sample was recorded. After collection, bones were kept frozen until being returned to the laboratory where they were stored at -20°C.

^{14}C ages were determined from eight bones as well as a number of guano layers. Analyses were performed at the Institute of Nuclear and Geological Sciences in Upper Hutt, New Zealand, using Accelerator Mass Spectrometry (AMS). The Calib. 4.2 program (Stuiver and Reimer, 2000) was used to correct for atmospheric ^{14}C fluctuations, and a half-life for ^{14}C of 5730 years was used. The radiocarbon ages of the samples were adjusted to take into account the 'reservoir effect'. This is the incorporation of depleted ^{14}C from seawater into organic material, which results in an upward biasing of the age of the samples (i.e. the samples appear older). This effect is particularly elevated in Antarctica because 'old' carbon is introduced through glacial meltwater and the upwelling of deep, old oceanic water (reviewed in Berkman *et al.*, 1998). This was taken into account when applying the calibration program. The program requires the input of a ΔR value, which is an approximation of reservoir age of the Southern Ocean. A ΔR value of 688 ± 55 was used so that the correction was appropriate for penguins (Petri and Baroni, 1997). This value was calculated from radiocarbon ages of seven penguin remains, including Adélie penguins, of known historical age (reviewed in Petri and Baroni, 1997).

Bones not directly carbon dated were assigned ages from dated guano or penguin bones from the same stratigraphic level.

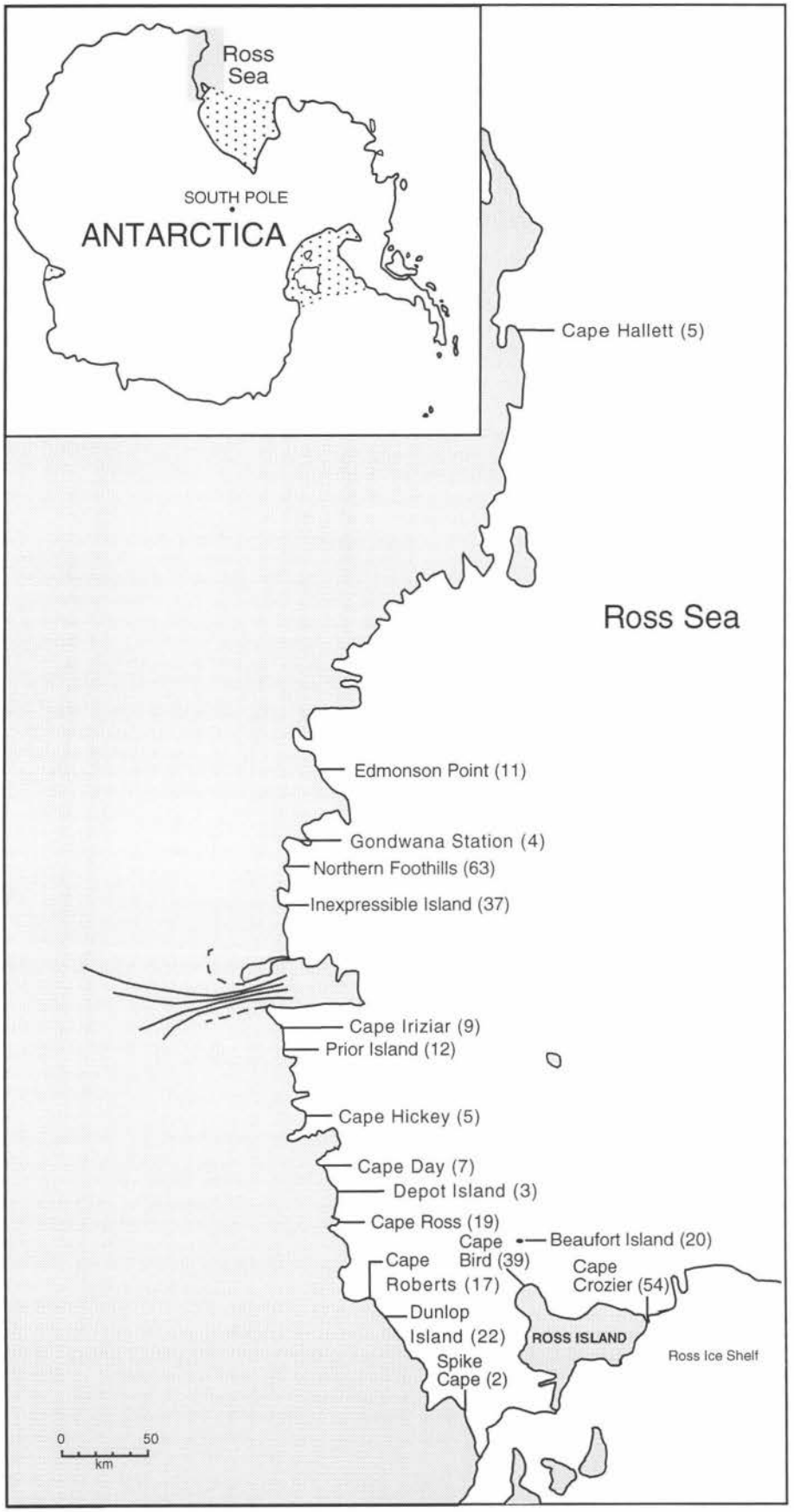

Figure 2.1 Location of Adélie penguin subfossil bone collection sites. The number of subfossil bones collected from each site is given in brackets.

Independent indicators of DNA preservation

Adélie penguin subfossil bone sections were prepared for microscopic analyses at the Pathobiology section, Institute of Veterinary, Animal and Biomedical Sciences, Massey University. Subfossil bones were decalcified before being processed using a standard technique (Ross *et al.*, 1995), then sectioned with a rotary microtome.

The histological preservation of six bones approximately 500 years old and three older bones (with ^{14}C calibrated ages of 3220±65 yrs BP, 6082 yrs BP and 6424 yrs BP) was examined. Bones were selected for sectioning that were intact and robust. Preservation of the bone sections was assessed using the histological index of Hedges and Millard (1995). This index categorises the extent of survival of the original bone structure.

DNA extraction

DNA extractions were attempted from 123 ancient Adélie bones. A negative control was included with each batch of extractions (6-10 samples). All ancient DNA extractions were undertaken in a separate and dedicated laboratory to minimise the risk of contamination. Benches and equipment were regularly cleaned with bleach and UV-irradiated to remove foreign DNA.

Bones were selected, with tarso-metatarsus, tibio-tarsus, femur, radius or humerus bones being preferred. The bone surface was shaved using a Dremel grinder in a fumehood to remove surface contamination. The Dremel grinder was then used to cut a sample of 0.5-1.5 g from the centre of the bone. Samples were homogenized in a coffee grinder (cleaned rigorously with ethanol between each sample). The high concentration of calcium ions present in the samples were removed by chelation with 0.5 M EDTA pH 8.0 (15 ml added and rotated overnight at room temperature in a 50 ml tube). Samples were then centrifuged for 20 minutes at 1000g in a swing bucket centrifuge and the calcium-rich supernatent discarded.

The decalcified bone material was then digested overnight at 50° C by rotation in 5 ml of extraction buffer (10 mM Tris-HCl pH 8.0, 1 mM NaCl), 500 µl of 10% sodium-dodecyl-sulphate (SDS, $C_{12}H_{25}O_4SNa$), 30 µl of 200 mg/ml dithiothreitol (DTT, $C_4H_{10}O_2S_2$) and 50 µl of 50 mg/ml proteinase-K. Digestion was followed by centrifugation for 20 minutes at 1000g in a swing bucket centrifuge, and the aqueous solution was transferred to a new 15 ml tube.

Proteins were removed by adding 5 ml of Tris-saturated phenol (C_6H_6O with 10 mM Tris HCl, pH 8.0, 1mM EDTA) to the sample and rotating at room temperature for 30 minutes. Samples were then centrifuged for 5 minutes at 1000g and the top aqueous layer containing the DNA was removed and transferred to a new 15 ml tube. The phenol addition, centrifugation, and removal of the aqueous phase was then repeated. 5 ml of chloroform:isoamyl alcohol ($CHCl_3$: $C_5H_{12}O$, 24:1) was added to the extracted DNA solution and rotated for 5 minutes at room temperature. Following centrifugation at 1000g for 5 minutes the top aqueous layer (containing the DNA) was added to a new 15 ml tube.

The extracted solutions were then diluted by the addition of MilliQ H_2O to make a total volume of 15 ml. Up to 4 ml at a time of diluted extract was then concentrated to 200 µl by spinning at 1000g through a 4 ml Vivaspin concentrator column (Viva Science, U.K.) for 20 minutes. The membrane in these columns retains molecules above 30 000 MW, such as DNA, while smaller molecules pass through. This 200 µl extract was stored at -20°C. 50 µl of the extracted solution was diluted to 200 µl with MilliQ H_2O and purified by using the QIAamp DNA mini kit (QIAGEN) following the manufacturers' instructions. Purified DNA samples were kept at 4°C while in use, then stored at -20°C.

Microsatellite primers and PCR amplifcation

Eight nuclear microsatellite loci were examined for this study (Table 2.1). Six of the microsatellite loci used in this study (AM3, AM13, AM12, TP500, RM3, and RM6) were isolated from Adélie genomic libraries (Roeder *et al.*, in press). The FhU2 locus was isolated from the pied flycatcher, *Ficedula hypoleuca,* and the HrU2 locus from the

Table 2.1 Primers used for amplifying microsatellite DNA from ancient Adélie bones.

Locus Name	Oligonucleotide sequence (5' to 3')	Repeat sequence from cloned allele
AM3	F - AGGAAAGAAGTAACTGAAGCAG	$(AT)_4N_5\ T_8$
	R - CATCTTCCCACAGAAGAAAC	
AM12	F - AAAAACCCAACACAACAAAC	$(CA)_{13}$
	R - CCCAAGAAGAGATTTGTGAG	
AM13	F - TTTTCCCATCTCTCTCTCCTG	$A_{11}N_{33}(GT)_9(GC)_4$
	R - CAGTTTTCAACAATCCTTCC	$(GT)_5$
TP500	F - GGACACAGGCAGCCAC	$(CA)_{14}$
	R - GGGAGTGGTATGGCTGG	
RM3	F - AATCAGGCTCCAAGGTCA	$(CA)_{10}$
	R - ATGCAAGTGACACAAAGG	
RM6	F - CAGGAGGCTTTGAGACAA	$(CA)_{10}$
	R - CTGTTTACATCCGATGCA	
HrU2	F - CATCAAGAGAGGGATGGAAAGAGG	$(CT)_5(CA)_2CT$
	R - GAAAAGATTATTTTTCTTTCTCCC	$(CA)_6$
FhU2	F - GTGTTCTTAAAACATGCCTGGAGG	$(GA)_4$
	R - GCACAGGTAAATATTTGCTGGGCC	

swallow, *Hirundo rustica* (Ellegren 1992; Primmer *et al.* 1996b). Primers to all these loci were manufactured by Sigma Genesis and the reverse primer was 5' -end labelled with either 6-FAM or HEX phosphoramidite dye. Specifically, HrU2, AM12, TP500, and AM3 were labelled with 6-FAM; and FhU2, RM6, AM13, and RM3 were labelled with HEX.

Polymerase chain reaction (PCR) amplification of microsatellite loci was performed in 10 µl volumes containing 0.5 units of Amplitaq™ DNA polymerase (PE Applied

Biosystems), 200 µM of each dNTP, 0.8 pmol of each primer, 1.5 mM $MgCl_2$, PCR reaction buffer (100 mM Tris-HCl pH 8.3, 500 mM KCl; PE Biosystems), 2 µg/µl bovine serum albumin (BSA) and 1 µl of extracted DNA. The PCR mix for the TP500 locus also contained 10% DMSO (BDH Laboratory Supplies). For each PCR setup a negative control (containing no DNA) and a positive control (with DNA extracted from blood samples added in another laboratory) was included. Each reaction was overlain with mineral oil and then transferred to another laboratory, where PCR was performed in a Hybaid Omnigene Thermal Cycler. Thermal cycling conditions consisted of an initial denaturation step of 94°C for 4 minutes, followed by 40 cycles of denaturation at 94°C for 45 seconds, annealment at 50°C (FhU2), 56°C (AM13, AM12, AM3, RM3, HrU2), 60°C (RM6), or 62°C (TP500) for 50 seconds and elongation at 72°C for one minute. Multiplex PCR was also trialled using the same PCR conditions as for single locus PCR. Loci were grouped together if they amplified under the same thermal cycling conditions and the size range of electromorph alleles labelled with the same dye did not overlap. Following amplification, PCR products were visualised using ethidium bromide ultraviolet-fluorescence. 3 µl of each PCR product was mixed with 3 µl of loading dye (10% w /v bromophenol blue, 15% w /v Ficoll, 1X TA buffer). Samples were then run on an agarose gel (1% MS, 1% LE) containing 1 µl ethidium bromide in a minigel chamber (Horizon®58 GibcoBRL) with 1X TA (10 mM Tris, 11.4 mM acetic acid). A 1 Kb plus DNA ladder (GibcoBRL) was run beside the samples to determine if the observed DNA was in the correct size range. The relative intensity of the PCR products was also estimated so that when pooled for genotyping they could be diluted to roughly the same concentration. After electrophoresis, gels were photographed using a UV transillumintor (Insta Doc™ System, Bio-Rad).

The effect of AmplitaqGold™ (PE Applied Biosystems) on product yield, compared to standard Amplitaq™, was investigated at the TP500 and AM3 loci for a subset of samples. The PCR reaction mixes were set up as master mixes using the same components as described above (except for the Taq). The master mixes were then divided in half before the addition of 0.5 units per sample of either Amplitaq or AmplitaqGold. Thermal cycling conditions were the same as described above except that the Amplitaq Gold samples underwent a pre-PCR step of 94°C for 10 minutes. PCR

products were run adjacently on a gel with mass ladder and the intensity of product observed.

Sequences of varying length from the mitochondrial control region had previously been amplified from the same extracted samples (Ritchie, 2001). For some samples a maximum amplifiable sequence length was determined (i.e. the amplification of larger fragments was attempted but not successful). These samples were used in a comparison of maximum amplifiable mitochondrial sequence length with the number of amplifiable microsatellite loci.

Automated genotyping

Samples were genotyped using an ABI Prism 377 Sequencer. Single locus amplification products were pooled when samples had high band intensity on an agarose gel, to make a total volume of 10 µl. This involved combining PCR products, the volume of each depending on their relative estimated concentration, so that the size range of electromorph alleles from different microsatellite loci labelled with the same dye were not overlapping. This allowed each sample to be genotyped across six loci in two lanes (usually TP500, AM3 and AM13 in a lane and HrU2, RM6 and RM3 in another lane). Poorly amplifying single locus amplification products and multiplex PCR products were not pooled but used directly for genotyping.

PCR products (0.5 µl) were mixed with 2.5 µl of formamide, 0.5 µl loading buffer (containing 5% blue dextran) and 0.5 µl of an internal size standard, ROX (PE Applied Biosystems). Samples were then denatured for 5 minutes at 95°C, quenched on ice and 2 µl loaded onto a 5% polyacrylamide gel (BioWhittaker Molecular Applications). The sequencer was run according to the manufacturer's instructions, and the data was analysed using GeneScan Analysis 3.1.

Authentification of results

In this study, seven extractions were repeated (i.e. the same bones were re-extracted on independent occasions) and genotyped to test for consistency of results. Another six

extractions had previously been repeated, and mitochondrial DNA amplified and sequenced at an independent laboratory at the University of Auckland (Ritchie, 2001).

In this study, an examination of the occurrence of genotyping errors, such as allelic dropout and the generation of false alleles (see Chapter One, section 1.5), was undertaken by performing independent PCR reactions for a subset of samples. Samples selected to have amplifications repeated had low band intensity on agarose. For the RM6 locus a total of 41 amplifications were performed for 17 samples (average of 2.4 amplifications/sample); for the RM3 locus 38 amplifications were done on 16 samples (an average of 2.4 amplifications/sample); and at the TP500 locus a total of 55 amplifications were performed from 26 samples (an average of 2.1 amplifications/sample). The samples examined covered a range of radiocarbon ages.

Statistical analysis

To test for a relationship between the number of microsatellite loci amplified and both the sample age and the maximum length of mitochondrial sequence amplified, Spearman's rank-order correlations were performed. This is a nonparametric statistical test that does not make assumptions about the distribution of the data. The test was performed using the program SPSS (SPSS, 1997).

2.3 Results

^{14}C dates

The radiocarbon-dated Adélie bones and their ages are shown in Table 2.2. Samples, whose ages were inferred, are listed in Appendix B. These ages ranged from 275 yrs BP to 6424 yrs BP

Histology

The younger bones (c. 500 yrs BP) that were examined were extremely well preserved and assigned a value of 5 on the histological index. Hedges *et al.* (1995) define category

Table 2.2 Ancient Adélie bones that were directly radiocarbon-dated. The calibrated age takes into account the reservoir effect. Adapted from Ritchie (2001).

Extraction	Location	Carbon Dating Number	Radiocarbon Age (yrs BP ± S.E.)	Cal. Age (yrs BP)
PE 4B3	Cape Crozier	NZA-8818	1475 ± 59	440
PE1B3	Cape Crozier	NZA-10306	1380 ± 60	310
PE9B3	Cape Crozier	NZA-8819	1503 ± 57	461
PE11B3	Cape Crozier	NZA-8820	1616 ± 62	523
PE13B3	Cape Crozier	NZA-8821	1564 ± 55	523
PE48	Cape Bird	NZA-9182	1534 ± 55	481
PE42	Beaufort Island	NZA-12287	1313 ± 55	275
PE64	Inexpressible Island	NZA-12286	6358 ± 55	6082

5 as having over 95% of the bone intact and being virtually indistinguishable from fresh bone. The histological sections of the older bones were not successful because they disintegrated during the decalcification step.

DNA amplification

All extractions that were repeated gave consistent results. In addition, no negative extraction controls exhibited signs of contamination.

The FhU2 locus was only genotyped for 23 ancient samples because they were all monomorphic for a 122 bp band (it was also monomorphc for 49 living samples). The success of AM12 amplification was poor, only 17 /34 samples that were younger than 500 years old amplified, so it was omitted.

Multiplex PCR was not successful for the vast majority of samples trialled because the poorer amplifying locus tended not to amplify. For example, when AM13 and AM12 were combined, AM12 often did not amplify at all, despite amplifying successfully independently.

The comparison of Ampilitaq Gold and Amplitaq suggests that there are significant differences between these two types of Taq Polymerase when amplifying ancient Adélie penguin samples (Figure 2.2). Fewer ancient samples could be amplified using AmplitaqGold, especially when they contained low quantites of DNA (i.e. when only faint bands were obtained with Amplitaq).

The success rate of microsatellite amplifications at each locus versus sample age are presented in Figure 2.3. These graphs demonstrate that there is considerable variation in

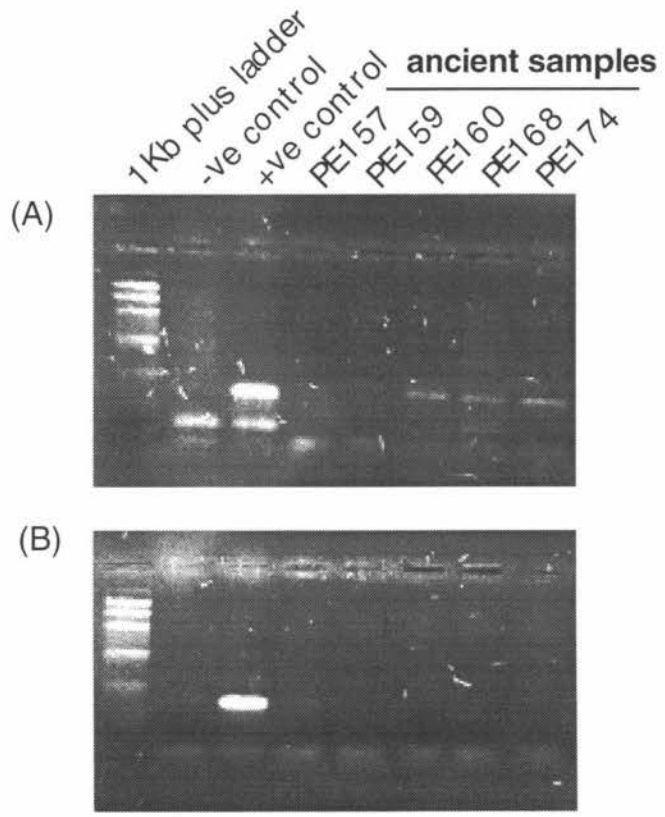

Figure 2.2 PCR products from amplifications of the TP500 locus using standard Amplitaq Polymerase (A) and AmplitaqGold Polymerase (B). For the ancient samples PCR products were only produced with Amplitaq (samples PE160, PE168 and PE174).

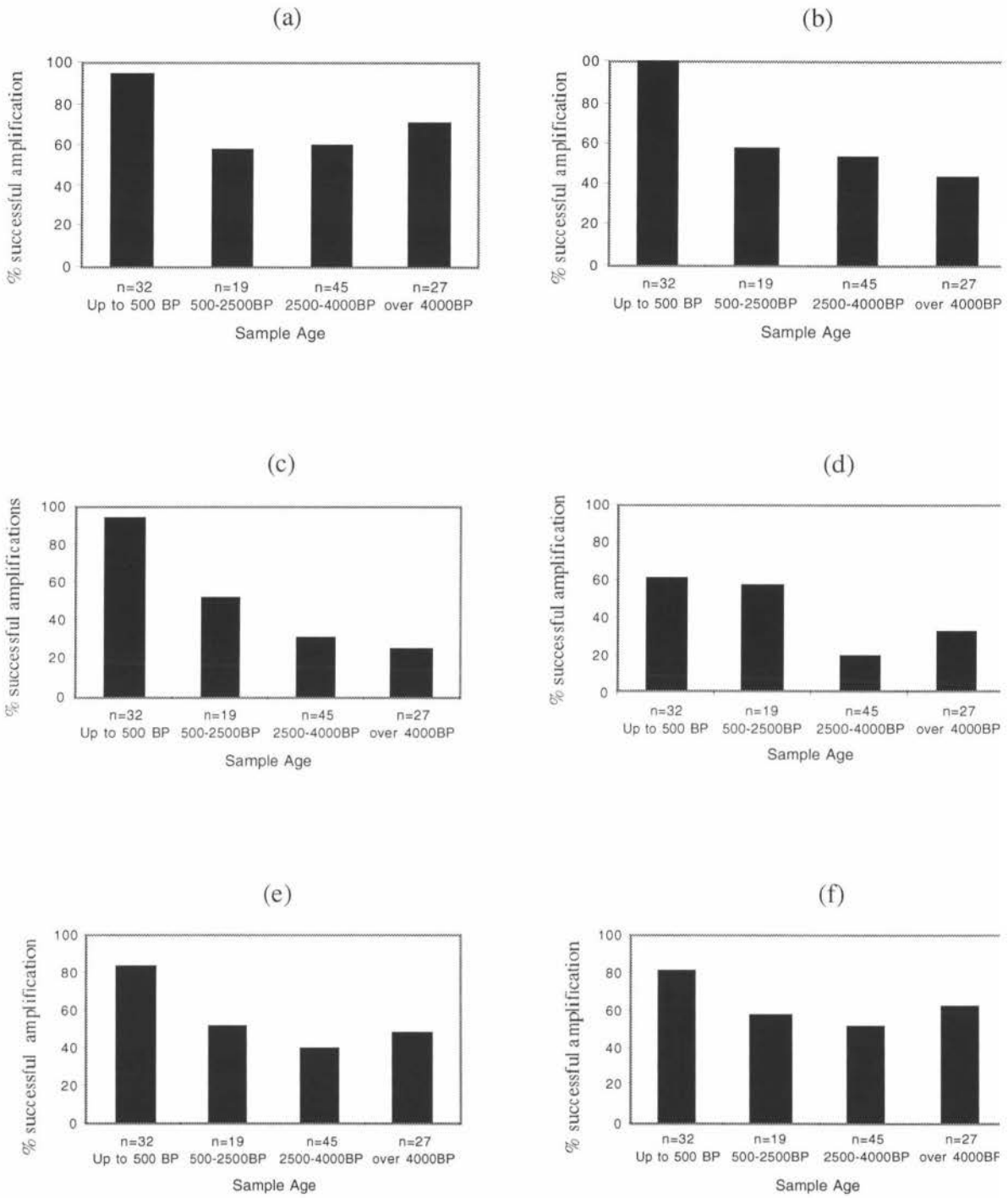

Figure 2.3 Success rate of the amplification of six microsatellite loci in ancient Adélie penguin samples of varying ages. (a) TP500 locus (b) RM6 locus (c) RM3 locus (d) AM13 locus (e) AM3 locus (f) HrU2 locus.

amplification success of ancient samples between both microsatellite loci and sample age. The differences in amplification success between loci are not related to the length of alleles. For example, the RM3 locus, with alleles around 220 bp in length, amplified in a greater number of samples than the AM13 locus which has alleles around 100 bp shorter. A scattergraph was constructed to display the relationship between the number of microsatellite loci able to be amplified and sample age (Figure 2.4). Although the points appear scattered, a Spearman's rank-order correlation suggests a significant negative correlation between these two variables ($r_s = -0.411$, $p<0.001$). That is, there is a tendency for fewer microsatellite loci to amplify with increasing sample age.

A scattergraph was also constructed to compare the number of microsatellite loci amplified and the maximum length of mitochondrial sequence amplified (Figure 2.5). A highly

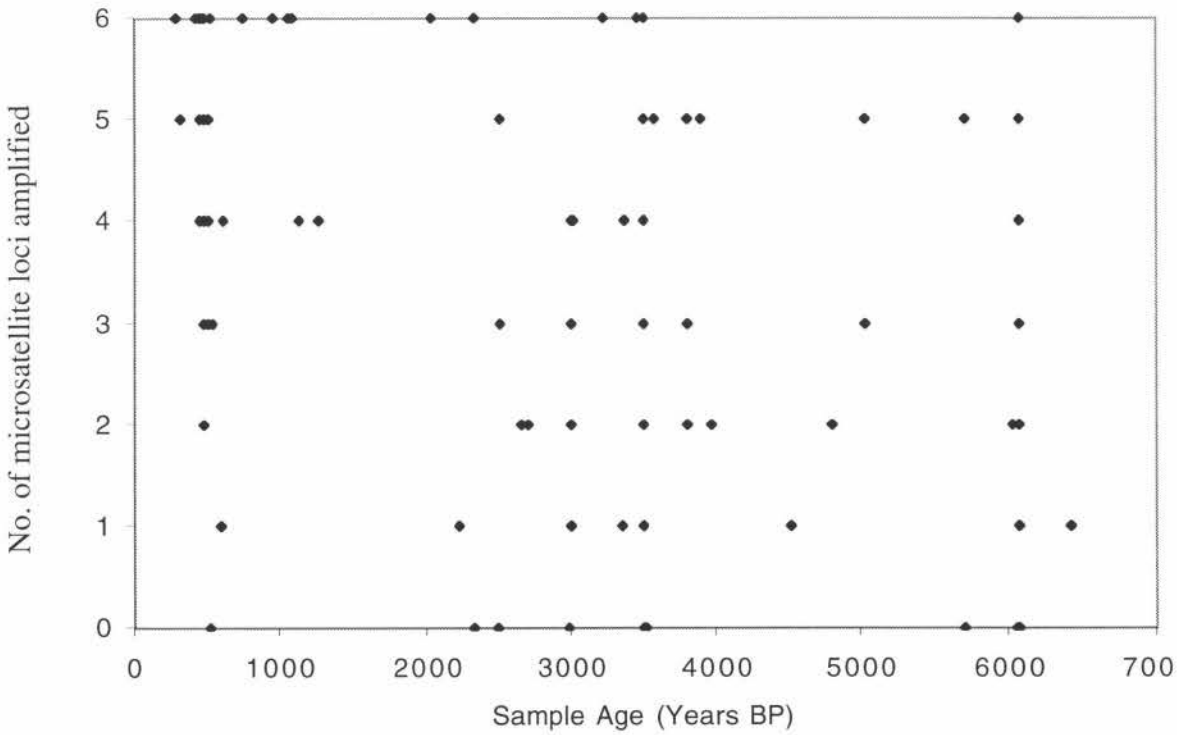

Figure 2.4 Scattergraph of the number of microsatellite loci amplified versus sample age for 123 ancient Adélie DNA samples.

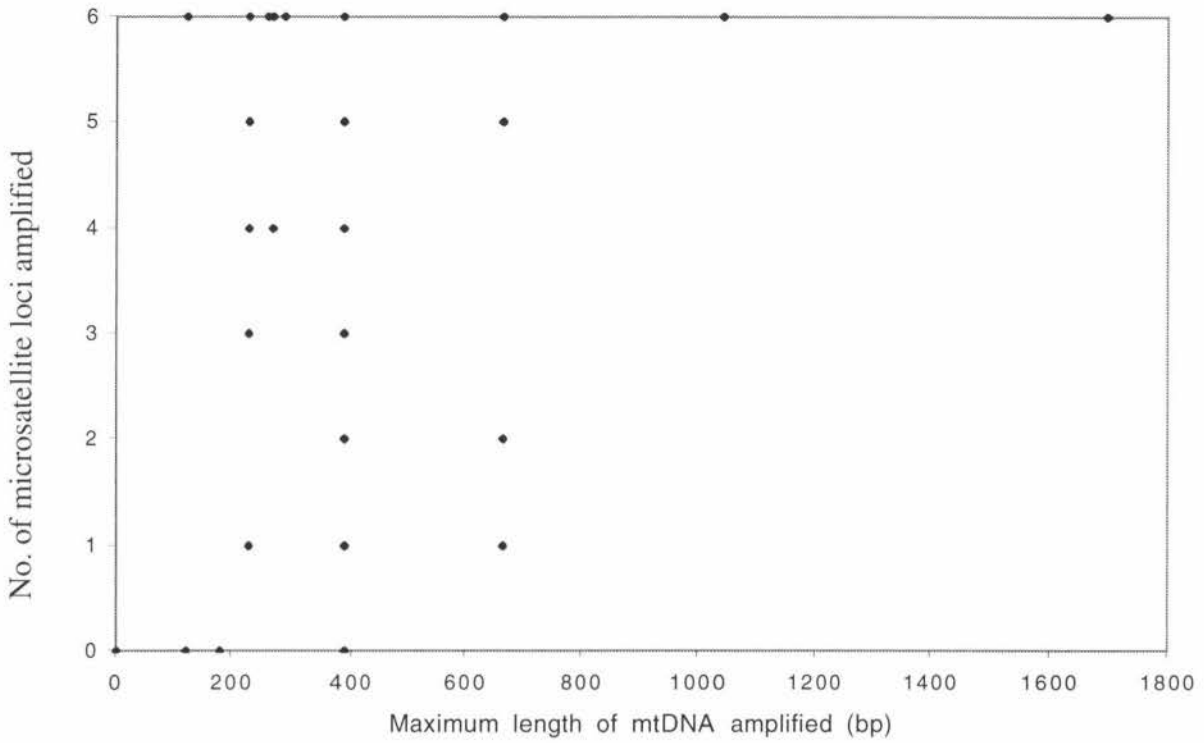

Figure 2.5 Scattergraph of the number of microsatellite loci able to be amplified versus the maximum length of mitochondrial DNA able to be amplified from ancient Adélie DNA samples (n=49).

significant positive association is suggested by a Spearman's rank-order correlation (r_s = 0.635, $p < 0.0001$). That is, the longer the fragment of mitochondrial DNA that can be amplified from a sample, the more likely a large number of nuclear microsatellite loci will also be able to be amplified.

Assessment of genotyping errors

Of the three loci examined for genotyping errors, only one sample, at the TP500 locus demonstrated a genotyping error. This sample, PE 74, with a calibrated radiocarbon date of 2513 yrs BP, initially appeared homozygous at this locus, but a second amplification revealed it was a heterozygote (Figure 2.6).

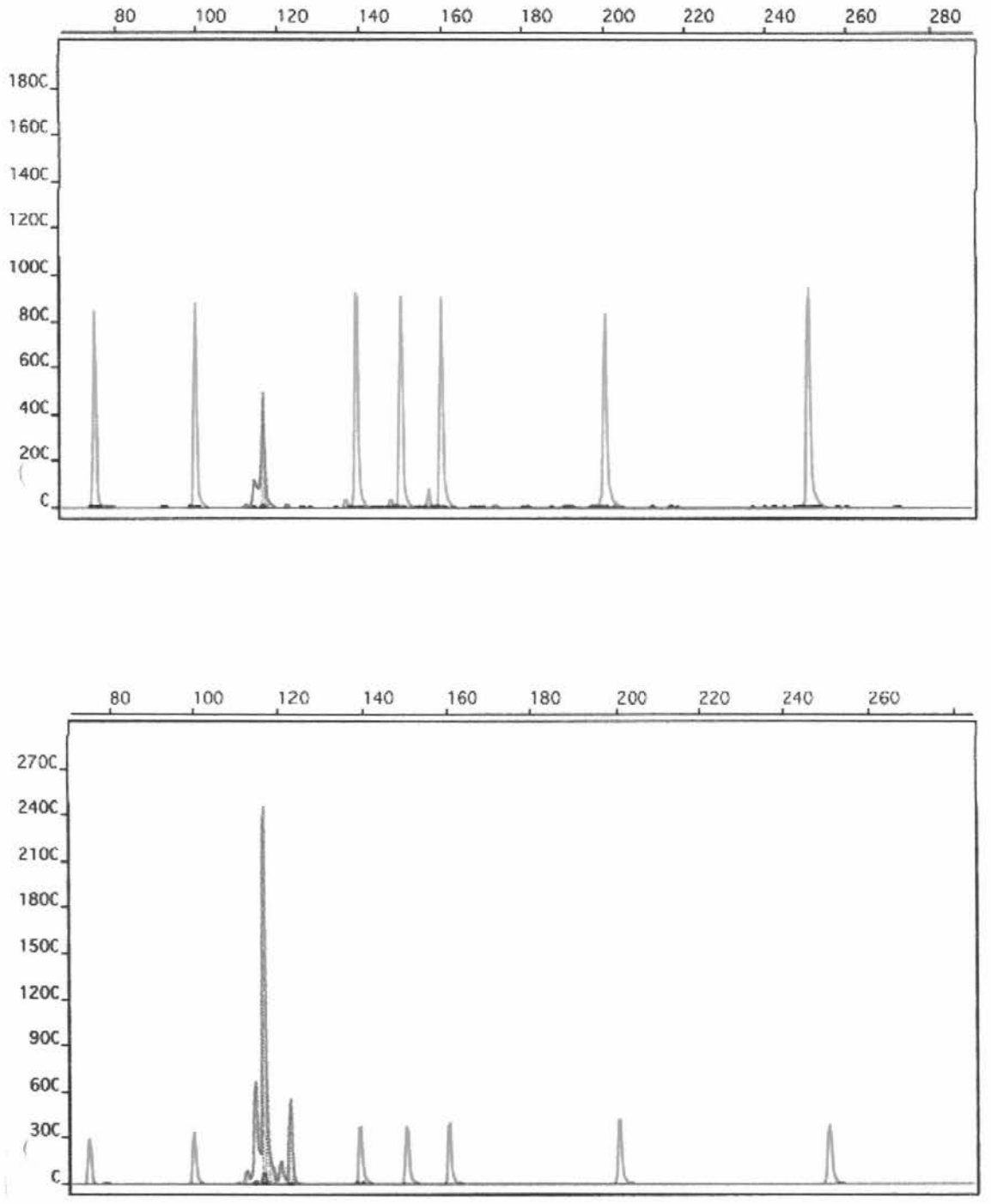

Figure 2.6 Electropherograms demonstrating allelic dropout at the TP500 locus in sample PE74 from Cape Hickey. The first amplification (top picture) suggests that PE74 is homozygous for an allele 114 bp in size (blue peak). The second amplification (bottom picture) shows that an allele 124 bp in size is also present. The red peaks are the size standard.

2.4 Discussion

Ancient Adélie DNA Preservation

No other DNA study has been as successful in amplifying single-copy nuclear sequences (e.g. 100% success rate for amplification at the RM6 locus in Adélie penguin samples up to 500 years old). Ramos *et al.* (1996) observed a 71.4% nuclear microsatellite amplification success rate from bone and tooth samples up to 5000 years old. However, samples were only selected from which mitochondrial DNA had previously been successfully amplified. In addition, repeated analysis revealed genotyping errors in 97% of samples. Zierdt *et al.* (1996) amplified single-copy nuclear DNA from 1200-1500 year old bone samples with a 22.2% success rate. However, they also reported allelic dropout, inferred from a deficiency of heterozygotes.

The remarkable preservation of Adélie penguin ancient DNA is likely to result from the low temperatures and dry conditions present in Antarctica. These conditions slow the rate of hydrolytic and oxidative damage (Greenwood *et al.*, 1999) and, in addition, low temperatures reduce micro-organism infestation (Burger *et al.*, 1999).

The weak negative correlation between the number of microsatellite loci amplified and sample age was unexpected because many studies have reported there to be no relationship between the age of a sample and its preservation (e.g. Höss *et al.*, 1996; Burger *et al.*, 1999). A negative correlation was also observed, but not quantified, for these samples between the length of mitochondrial sequence able to be amplified and sample age (Ritchie, 2001). One reason that a correlation between preservation and age has not been observed previously might be because significant numbers of samples preserved in the same conditions, and of varying ages, are not commonly available. Most comparisons of preservation have been between samples preserved in different environments. Therefore, environmental factors were also exerting an influence. The Adélie subfossil samples are plentiful and the conditions in which they are preserved are relatively constant, making them ideal for studies of the types and rate of DNA damage.

Independent indicators of DNA preservation

The main purpose of indicators of DNA survival is to verify the authenticity of ancient DNA, i.e. that the preservation conditions of a sample are compatible with DNA preservation. Several studies that examined the histological preservation of human and/or cow bones found a correlation with ancient DNA presence (Hagelberg *et al.*, 1991; Colson *et al.*, 1997). However, histological preservation does not seem to be an accurate predictor of ancient DNA survival in Adélie subfossils because many of the bones were soft and friable, and thus unsuitable for sectioning. Yet, they yielded DNA of sufficient quality and quantity to enable PCR of the studied microsatellite loci.

The level of amino acid racemisation (Poiner *et al.*, 1996), the presence of hydantoins (Höss *et al.*, 1996) and the preservation of proteins (Poiner and Stankiewicz, 1999) have all been correlated to the presence of DNA in ancient samples. These techniques could be included in future studies of Adélie ancient DNA to support the authenticity of ancient DNA found in these samples.

The strong positive correlation between the length of mitochondrial DNA (mtDNA) able to be amplified and the number of microsatellite loci amplified in this study suggests that the presence of mtDNA may indicate if nuclear DNA is able to be retrieved. For example, in all samples where 1 kb of mtDNA was able to be amplified, DNA sequence from all six microsatellite loci also amplified (n=5).

Accurate amplification of ancient DNA.

The negative extraction and amplification controls demonstrated no signs of contamination, and several mitochondrial amplifications were independently extracted and amplified in another laboratory (University of Auckland). These results suggest that the procedures undertaken to prevent contamination were successful.

In this study, although not involving a large number of repeat amplifications, genotyping appears to be producing reliable results because all but one of the repeated genotypes produced consistent results. According to Schmerer *et al.* (1999) genotyping errors are related to both the quality and quantity of DNA. That the ancient Adélie DNA

is very well preserved is suggested not only by the success of microsatellite amplification, but also by the ability to amplify mitochondrial DNA fragments up to 1.6 kb (Ritchie, 2001). Further support comes from the relatively small number of PCR cycles performed. Only 40 PCR cycles were used to amplify the microsatellite loci, compared to the 60 or more cycles used by many ancient DNA studies (e.g. Zierdt *et al.* (1996) used a total of 90 cycles).

According to Rameckers *et al.* (1997) if a large number of PCR cycles are required to produce amplification products (i.e. more than 50), then the initial number of target template molecules is estimated to be within the range of 1-10. When amplifications are performed from such low target numbers, genotyping errors can become a problem. By reducing the number of PCR cycles to 40, the majority of samples with very low target numbers and thus those most at risk of genotyping errors, will not amplify. Restricting amplification in this way may result in more accurate genotyping. This method may be more practical when using ancient DNA for population studies, because implementing the multiple-tubes approach (Navidi *et al.*, 1992) for large numbers of samples is expensive and time consuming when multiple loci are involved (see Figure 1.2).

The findings of the Taq Polymerase trial were not in agreement with a study by Hummel *et al.* (1996) who concluded that the use of PE Applied Biosystems AmplitaqGold enzyme yielded greater amounts of PCR product from ancient DNA. They also found that the specificity of amplification increased and that the size of the stutter bands decreased, thus allowing easier scoring of microsatellite alleles. They suggested that this is a consequence of the amplification reaction being able to reach a temperature that favours specific primer annealing, thus decreasing primer dimers. In this study, the AmplitaqGold reaction yielded less product than Amplitaq for ancient samples. However, it did appear to decrease primer-dimer formation in the negative and positive control reactions compared with Amplitaq. Primer-dimers were not present in the amplifications of ancient samples for either type of Taq Polymerase. Contrary to Hummel *et al.* (1996), the use of AmplitaqGold™ in this study did not decrease stutter bands on the electropherograms. In conclusion, standard Amplitaq appears to amplify ancient DNA samples better than AmplitaqGold, despite the latter being commonly used for ancient DNA studies.

2.5 Concluding Remarks

1. Adélie penguin subfossil DNA from Antarctica is generally extremely well preserved, indicated by the single copy nuclear DNA loci and the long fragments of mitochondrial DNA that can be readily amplified. In addition, genotyping errors, a common problem when working with nuclear microsatellites from ancient DNA, do not occur often in amplifications from Adélie penguin ancient DNA. This extraordinary preservation is likely to be a consequence of the cold, dry conditions in Antarctica.

2. A weak negative correlation was observed between the success of microsatellite amplification and sample age. Though intuitive, this finding was contrary to previous studies, and may result from the large number of samples of varying ages preserved under the same conditions that were analysed.

3. The length of mitochondrial DNA able to be amplified was found to be a fairly accurate indicator of success of microsatellite amplification. However, histological preservation, which has been suggested in other studies to be an indicator of DNA preservation, was not found to be useful as an indicator of the presence of DNA in ancient Adélie penguin bones.

4. The Taq polymerase enzyme commonly used for amplifying DNA from ancient samples, AmplitaqGold, was found in this study to produce less PCR product than standard Amplitaq.

CHAPTER THREE

Temporal Genetic Change in Adélie Penguin Colonies.

3.1 Introduction

The advent of molecular genetic techniques has enabled a range of original and precise investigations into the genetics of natural populations. However, until recently, the extension of this research has been limited by the fact that inferences about past events had to be made from present day genetic patterns. Ancient DNA provides a unique opportunity to investigate genetic change in populations over time. This area of ancient DNA research has been described as perhaps the most intriguing, yet most neglected (Tessier and Bernatchez, 1999; Wayne *et al.,* 1999) facet of recent studies. The majority of population genetic analyses that have included ancient DNA have been restricted to examining change over the last 100 years because museums have been the best source of large numbers of well-preserved samples (e.g. Nielson *et al.*, 1997; Nielson *et al.*, 1999). Those studies that have examined older samples have typically used mitochondrial DNA as a genetic marker (e.g. Wang *et al.*, 2000). This chapter will examine the feasibility of detailing population genetic change over a significant geological time period using nuclear microsatellite loci. The large, well-preserved deposits of Antarctic Adélie penguin bones are ideal for such a study.

The aims of this chapter are to examine the population history of three Adélie penguin colonies. Specifically the study was aimed at determining whether the ancient samples from those colonies exhibit genetic differentiation. In addition, the research aims to reveal if changes in gene frequencies have occurred within colonies over time. The three populations, all from the Terra Nova Bay region, were selected because they are

presently occupied and moderate numbers of subfossil bones had been collected at these sites.

3.2 Materials and Methods

Sample collection

Subfossil bone samples were collected from the Inexpressible Island, Northern Foothills and Edmonson Point sites (Figure 3.1) as described in Chapter Two (section 2.2). Blood samples were taken from adult birds in these same colonies after capturing them with a hand net (see preface). A 23-gauge syringe was used to remove approximately 300 µl of blood from the intertarsal or jugular vein. Blood was temporarily stored in 0.2 M Na_2EDTA before being placed in lysis buffer (Seutin *et al.*, 1991). Following transportation to the University of Auckland, the blood samples were stored at -4°C.

DNA extraction

DNA was extracted from ancient subfossil bone samples as described in Chapter Two (section 2.2). Genomic DNA was extracted from blood samples at the University of Auckland using the instagene kit (Biorad) following the manufacturer's instructions. DNA was transported to Massey University and stored at 4°C.

PCR amplification and genotyping

Six microsatellite loci (AM3, AM13, TP500, RM3, RM6, HrU2) were amplified and genotyped, as described in Chapter Two (section 2.2), for both the living and ancient DNA samples (these genotyping results are in Appendices B and C).

Statistical analysis

The microsatellite loci were tested for genotypic linkage disequilibrium using GENEPOP v3.1c (Raymond and Rousset, 1995). All possible combinations of loci were analysed for each population as well as a global test for all pairs of loci across all

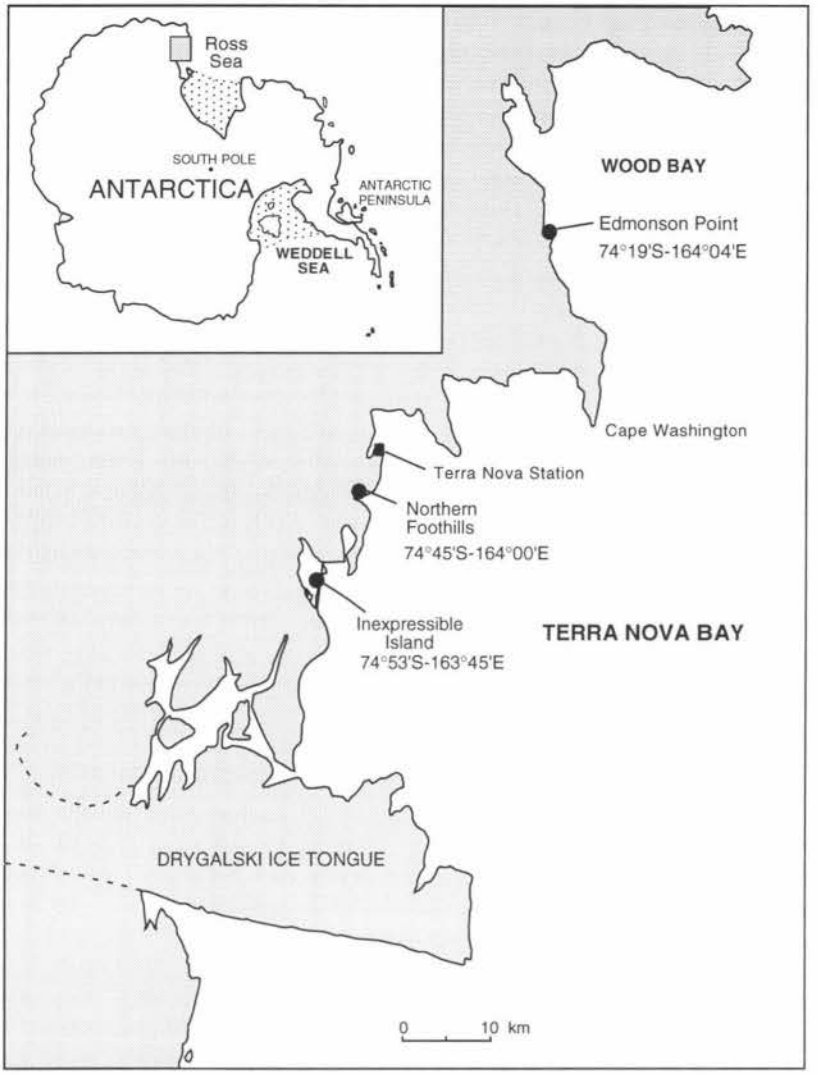

Figure 3.1 The location of Adélie penguin colonies where blood and subfossil bone samples were collected for population genetic analysis.

populations (Fisher's method). Markov chain parameters (Guo and Thomson, 1992) were set at 1000 dememorisation steps, 100 batches, and 1000 iterations/batch. The genetic diversity of each of the Adélie penguin populations was measured by the number of electromorph alleles per locus, and the observed and expected heterozygosity. Conformity with Hardy-Weinberg equilibrium predictions at each locus was tested with a Markov chain method (Guo and Thompson, 1992). The number of steps in the Markov chain was set at 1000 with 100 batches and 1000 iterations/batch. The above tests were conducted using GENEPOP v3.1c (Raymond and Rousset, 1995).

Genic analyses were chosen to examine population differentiation rather than genotypic analyses because gene frequencies change more slowly than than genotypic arrays. Genotypic arrays are only useful over short time-scales, such as in the study of migration or parentage, because they are usually destroyed within a single round of sexual recombination (Sunnocks, 2000). Exact tests of heterogeneity (Raymond and Rousset, 1995) were performed using Genepop v3.1c. Tests were performed separately between all living populations and between all ancient populations, as well as between the temporal samples from each colony. To minimise the chance of making a type 1 error, the sequential Bonferroni correction was applied to adjust the levels of significance for the multiple tests that were performed (Rice, 1989). To investigate the effect of small sample sizes and missing data (as is present in the ancient samples) on the exact test, the living data set was made similar to the ancient data set by randomly manually removing samples and loci. For example, to make the Inexpressible Island living data similar to the ancient samples from the same location the living data set was randomly reduced from 28 to 22 samples. The number of loci amplified was also made the same as in the ancient data set e.g. the number of genotypes at the TP500 locus was decreased from 22 to 20. The exact test was then performed on this simulated data set. This bootstrapping procedure was repeated 25 times.

The extent of differentiation between all living populations, between all ancient populations, and between temporal samples from each colony was then investigated by calculating F_{ST} values (Wright, 1951; Weir and Cockerham, 1984) and R_{ST} values (Slatkin 1995; Michalakis and Excoffier, 1996). In addition, overall F_{ST} and R_{ST} values were calculated for the living populations and for the ancient populations. These values measure the reduction in heterozygosity of a population resulting from random genetic drift. If the populations being compared are in Hardy-Weinberg equilibrium and have the same electromorph allele frequencies then F_{ST} and R_{ST} values = 0. Significant differentiation is present between populations when F_{ST} and $R_{ST} > 0.05$. F_{ST} is based on the IAM model of evolution and R_{ST} is based on the SMM. Both F_{ST} and R_{ST} values were computed because it is unknown which, if either, of these underlying mutation models the loci in this study follow.

F_{ST} values were calculated using FSTAT v.2.8 (Goudet, 1999). 95% confidence

intervals for F_{ST} were also calculated in FSTAT by using a bootstrapping procedure. R_{ST}, was calculated using the program R_{ST}CALC (Goodman, 1997). This program was used because it corrects for differences in sample size between populations. R_{ST}CALC was also used to calculate 95% confidence intervals for R_{ST} by bootstrapping with 1000 iterations. The AM13 locus genotypes were removed from calculations of F_{ST} and R_{ST} values involving the Northern Foothills ancient population. This was because samples from this population could not be genotyped at this locus and both the FSTAT and R_{ST}CALC programs would not calculate F_{ST} and R_{ST} values when this locus was included.

To further investigate differentiation, individuals were classified according to their composite genotype by an assignment test. This test assigns each individual to the population in which its composite genotype has the highest probability of occurring (Paetkau *et al.*, 1995; Paetkau *et al.*, 1997; Wasser and Strobeck, 1998). The greater the level of differentiation between populations the greater the number of individuals assigned to the correct population. The genotypic data from all loci were used to perform this assignment test using the calculators at http://www.biology.ualberta.ca/jbrzusto/Doh.php. Assignment tests were performed between living populations, between ancient populations and between living and ancient samples after pooling living and ancient populations into separate groups. Assignment tests were also performed between the two time periods (ancient and living) for each colony.

3.3 Results

Genetic diversity within Adélie penguin populations

Genotyping of 125 samples was attempted for six microsatellite loci. These samples were from three locations and were divided into modern (blood) samples and ancient (bone) samples of varying ages. Figure 3.2 shows soil profiles and the location and ^{14}C dates of some of the subfossil samples used in this chapter. The success rate of amplifying the ancient samples varied considerably, resulting in very small sample sizes for some loci e.g. the RM3 locus could only be amplified in three ancient Northern Foothills samples. Therefore, for some loci, a smaller number of electromorph alleles

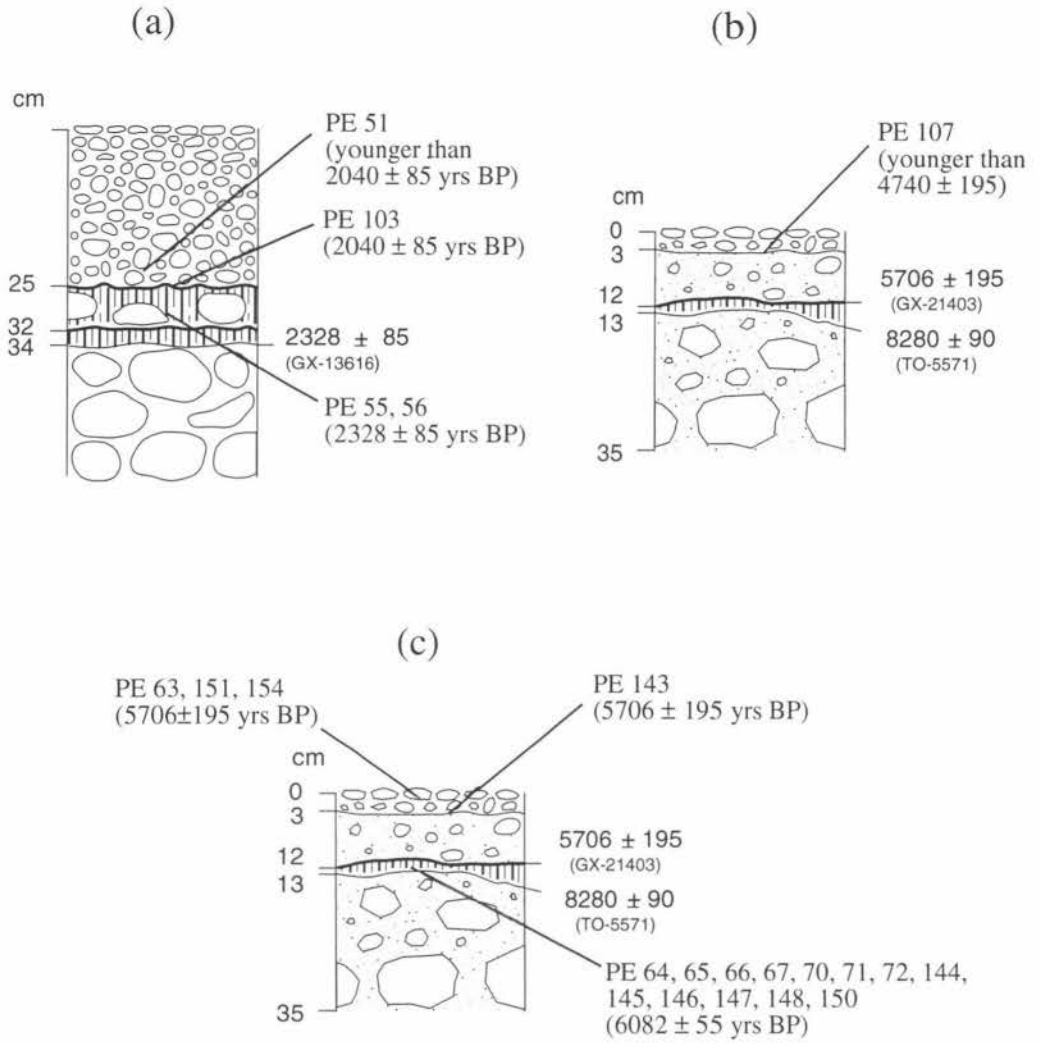

Figure 3.2 Soil profiles and the approximate locations of subfossil bones from several of the sites used in this study. The calibrated ages of subfossil bones and guano layers are given. (a) Inexpressible Island. Site P3. 6.2 metres above sea level (a.s.l). (74°54'21"S, 163°43'42"E). (b) Inexpressible Island. Site S2/94. Promontory south of Seaview Bay. 50m a.s.l. (74°54'40"S, 163°43'42"E). (c) Inexpressible Island. Site S3. Promontory south of Seaview Bay. 50m a.s.l. (74°54'40"S, 163°44'00"E). Note that the ages for site C are based on a correlation with site B.

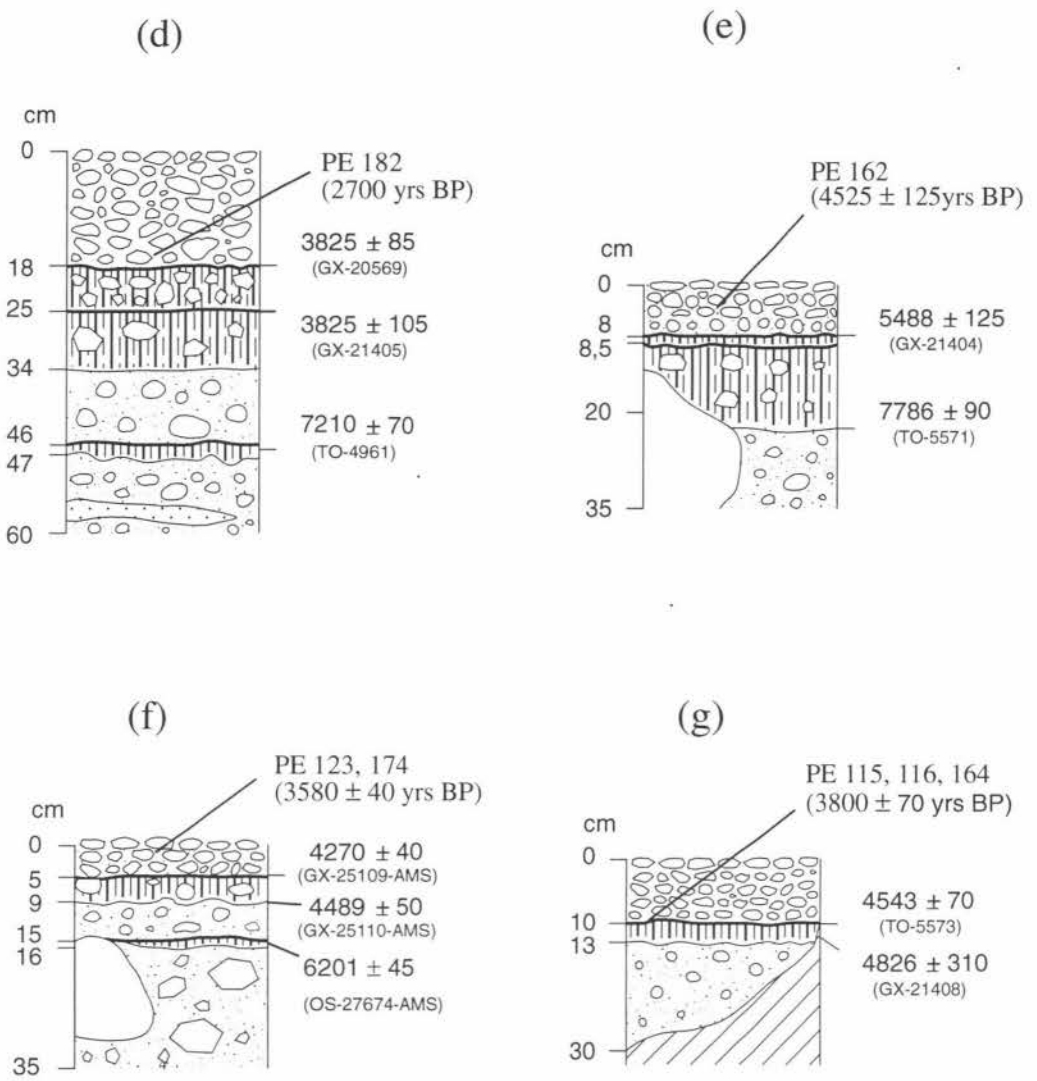

Figure 3.2 continued. Soil profiles and the approximate locations of the subfossil bones from several of the sites used in this study. The calibrated ages of subfossil bones and guano layers are given. (d) Northern foothills. S3/94. South of Cape Icaro. 60m a.s.l. (74°42'50"S, 164°06'30"E). (e) Northern Foothills. S2/94. North Adélie Cove. 40m a.s.l. (74°44'07"S, 164°06'55"E). (f) Northern Foothills. Terra Nova Station. 40m a.s.l. (74°41'37"S, 164°06'39"E). (g) Northern Foothills. Terra Nova Station. 33m a.s.l. (74°41'37"S, 164°06'39"E).

were amplified for the ancient samples, with electromorph alleles that were rare in the living samples not detected. The electromorph allele frequencies and expected and observed heterozygosities at each locus for each location and time period are shown in Table 3.1.

The analysis of genotypic linkage disequilibrium indicated that none of the loci demonstrated disequilibrium (p-values all > 0.05). This is consistent with the results of Roeder *et al.* (in press) who examined linkage between the same loci except for HrU2 and TP500. This study indicates that HrU2 and TP500 also do not show linkage disequilibrium.

Deviation from Hardy-Weinberg equilibrium was tested at each locus in each population (Table 3.1). Three tests demonstrated a significant departure from Hardy-Weinberg equilibrium (with p-values < 0.05). These tests were all from different populations (one from each of the modern (AM13 locus) and ancient (AM3) Inexpressible Island populations and one from the modern Northern Foothills population (AM3)). However, the application of the sequential Bonferroni correction reduced the number of significant tests to one (the AM3 locus of the modern Northern Foothills population).

Genetic analysis between Adélie penguin populations of a given time period

Fisher's exact test of genic differentiation (Table 3.2) showed that, following the application of the Bonferroni correction, there was no significant difference in electromorph allele frequency distributions between living populations. F_{ST} and R_{ST} values also indicate that there is no significant heterogeneity between living colonies, with F_{ST} values ranging from -0.006 to 0.006, and R_{ST} values from -0.025 to - 0.003 (Table 3.3). Assignment tests were used as another, largely independent, measure of differentiation. These methods determine if individuals can be assigned correctly to the colony from which they were collected. They can hence be used to determine whether living populations could be distinguished (Figure 3.3). Assignment of the living samples to their colony of origin did not differ greatly from random ($\chi^2 = 3.945$, p=0.4135), which is in agreement with the results of the exact test, F_{ST} and R_{ST} values.

Table 3.1 Electromorph allele frequencies, observed (H_O) and expected (H_E) heterozygosities of each population and the probability (p) that it is not in Hardy-Weinberg equilibrium. Values that are significant at the 5% level after the sequential Bonferroni correction are shown in bold type.

Locus	Electro -morph	Inexpressible Is.		Northern Foothills		Edmonson Pt.	
		modern	ancient	modern	ancient	modern	ancient
		n=28	n=20	n=28	n=9	n=30	n=8
TP500	108					0.033	
	112	0.089	0.025	0.179	0.167	0.133	0.188
	114	0.339	0.300	0.161	0.444	0.183	0.250
	116	0.107	0.150	0.232	0.278	0.183	0.250
	118	0.179	0.175	0.054		0.083	0.125
	120	0.054	0.025	0.071		0.067	
	122	0.089	0.025	0.179	0.056	0.15	0.125
	124	0.125	0.225	0.125	0.056	0.15	0.062
	126	0.018	0.075			0.017	
	H_O	0.8214	0.8193	0.8481	0.732	0.8718	0.8583
	H_E	0.8571	0.750	0.8214	0.5555	0.8667	0.7777
	p	0.5172	0.3545	0.6408	0.1930	0.5994	0.7252
RM6		n=27	n=14	n=26	n=5	n=32	n=7
	170		0.036	0.020		0.031	
	172	0.815	0.964	0.740	0.700	0.797	0.929
	174	0.148		0.160	0.300	0.172	0.071
	176	0.037		0.080			
	H_O	0.3186	0.0714	0.4217	0.4667	0.3398	0.1429
	H_E	0.2593	0.0714	0.3461	0.2000	0.4062	0.1429
	p	0.1923		0.1992	0.3327	0.6988	
RM3		n=28	n=9	n=27	n=3	n=32	n=3
	219	0.018					
	221	0.750	0.944	0.852	0.833	0.812	1.000
	223	0.018				0.016	
	225	0.196	0.056	0.148	0.167	0.172	
	227	0.018					
	H_O	0.4052	0.1111	0.2572	0.3333	0.3149	
	H_E	0.2857	0.1111	0.2963	0.3333	0.3125	
	p	0.2736		1.0000		1.0000	

Table 3.1 continued.

Locus	Electro -morph	Inexpressible Is.		Northern Foothills		Edmonson Pt.	
		modern	ancient	modern	ancient	modern	ancient
AM13		n=28	n=12	n=27		n=31	n=3
	109			0.019			
	111	0.018	0.042	0.019		0.016	
	112	0.018				0.016	
	113		0.042				
	114	0.018				0.016	
	118	0.036	0.167			0.065	
	119	0.054	0.167	0.019		0.145	
	120	0.018		0.093		0.016	
	121			0.074		0.032	0.167
	122	0.018		0.037		0.032	
	123	0.036		0.019		0.081	0.167
	124	0.107	0.125	0.148		0.097	
	125	0.089	0.083	0.111		0.081	
	126	0.143	0.042	0.148		0.129	
	127	0.161	0.208	0.185		0.129	0.500
	128	0.179	0.125	0.037		0.113	0.167
	129			0.074		0.016	
	130	0.107		0.019		0.016	
	H_O	0.9000	0.8949	0.9022		0.9175	0.8000
	H_E	0.8889	0.8333	0.8889		0.7742	0.6667
	p	0.0289	0.5039	0.3236		0.5300	0.6034
AM3		n=28	n=16	n=28	n=3	n=29	n=6
	174	0.839	0.933	0.857	1.000	0.820	0.917
	175	0.089	0.069	0.107		0.180	0.083
	176	0.071		0.036			
	H_O	0.2877	0.1287	0.2571	0.200	0.3128	0.1667
	H_E	0.2500	0.0000	0.0714	0.6667	0.3103	0.1667
	p	0.3037	0.0345	**0.0001**	1.0000	1.0000	
HrU2		n=28	n=17	n=27	n=8	n=32	n=7
	116		0.059				0.071
	118			0.019			0.071
	120			0.037			
	122	0.125	0.088	0.037	0.188	0.047	
	124	0.554	0.647	0.519	0.688	0.516	0.286
	126	0.321	0.147	0.352	0.125	0.359	0.571
	128		0.059	0.019		0.078	
	130			0.019			
	H_O	0.5851	0.5615	0.6149	0.5083	0.6061	0.6264
	H_E	0.6071	0.4706	0.6667	0.3750	0.5000	0.5714
	p	0.1354	0.1672	0.7463	0.2811	0.2241	0.3487

Table 3.2 Results of the genic differentiation tests within each time period (first two categories), between living and ancient samples (third category), and between temporal samples of each colony (last three categories), for each microsatellite locus. The standard error is given in brackets. Results that are significant after the application of the sequential bonferroni correction are shown in bold type.

	TP500	RM6	RM3	AM13	AM3	HrU2	Overall
Between Living Populations	0.218 (0.011)	0.367 (0.009)	0.807 (0.007)	0.043 (0.005)	0.159 (0.005)	0.109 (0.010)	0.069 (df = 12)
Between Ancient Populations	0.443 (0.011)	0.024 (0.002)	0.644 (0.003)	0.280 (0.007)	0.146 (0.003)	0.018 (0.002)	**0.018** (df = 12)
All Living vs. All Ancient	0.126 (0.007)	0.244 (0.006)	0.579 (0.007)	0.302 (0.112)	0.511 (0.086)	0.086 (0.006)	0.161 (df = 12)
Inexpress. Island (Living vs. Ancient)	**0.006** (0.002)	0.018 (0.001)	0.692 (0.006)	0.288 (0.009)	0.399 (0.004)	0.046 (0.004)	**0.003** (df = 12)
Northern Foothills (Living vs. Ancient)	0.254 (0.006)	0.514 (0.005)	1.000 (0.000)	-	0.327 (0.005)	0.194 (0.006)	0.478 (df = 10)
Edmonson Point (Living vs. Ancient)	0.947 (0.002)	0.635 (0.004)	0.614 (0.004)	0.644 (0.009)	0.675 (0.002)	0.044 (0.003)	0.632 (df = 12)

Therefore, in summary, none of the above tests provided any evidence for genetic structure between the living Adélie penguin populations. In contrast, the overall exact test when applied to the ancient population comparisons suggests the presence of significant heterogeneity (p=0.018, Table 3.2). This finding was also supported by the formation of the simulated ancient data sets. A graph of the p-values of exact tests performed on these 25 simulated data sets is shown in Figure 3.4. The small sample sizes and/or missing data do appear to have a large effect on the results of the exact test because the simulated data generated a wide range of p-values (from 0.027-0.961). If the small sample sizes and/or missing data had little effect on the p-value calculation, then the range of values would be expected to be much smaller. However, despite this

Table 3.3 Pairwise F_{ST}, R_{ST} values and estimates of 95% confidence intervals for these values between living populations, between ancient populations and between the two time periods for pooled samples and for each colony. An asterisk denotes populations in which the AM13 locus was removed before the calculation. Significant F_{ST} and R_{ST} estimates are shown in bold type.

Comparison	F_{ST}	95% CI	R_{ST}	95% CI
Living				
Inexpressible Is. vs Northern FH	0.006	-0.013-0.018	-0.011	-0.015-0.058
Inexpressible Is. vs. Edmonson Pt.	-0.001	-0.008-0.006	-0.003	-0.005-0.099
Northern FH vs. Edmonson Pt.	-0.006	-0.012-0.002	-0.003	-0.0112-0.058
Overall	-0.001	-0.009-0.005	-0.025	-0.005-0.542
Ancient				
Inexpressible Is. vs Northern FH *	0.039	-0.018-0.136	0.003	-0.039-0.181
Inexpressible Is. vs. Edmonson Pt.	0.040	-0.042-0.130	**0.053**	-0.049-0.354
Northern FH vs. Edmonson Pt. *	**0.117**	0.010-0.197	-0.026	-0.054-0.253
Overall	0.038	0.008-0.093	-0.026	-0.046-0.138
Living-Ancient				
Living vs. Ancient	0.006	0.002-0.013	0.015	-0.000-0.089
Inexpressible Island	0.027	0.008-0.050	**0.055**	-0.015-0.215
Northern FH *	0.019	-0.039-0.044	**0.085**	-0.026-0.309
Edmonson Point	0.008	-0.13-0.029	-0.008	-0.042-0.258

large effect, the p-value of the actual ancient samples (0.018) falls outside the range (0.027-0.961). The assignment test of the ancient samples into the three populations (Figure 3.3) showed that the correct assignment was significantly better than random ($\chi^2 = 5.587$, p=0.0159), therefore also supporting the presence of differentiation.

In contrast to the exact test and assignment test, F_{ST} and R_{ST} estimates did not, in

1. Living	Assignment test classification			
Source population	Inexpressible Island	Northern Foothills	Edmonson Point	Total
Inexpressible Island	**13**	6	9	28
Northern Foothills	9	**11**	8	28
Edmonson Point	9	13	**10**	32

$\chi^2 = 3.945$, p=0.4135

2. Ancient	Assignment test classification			
Source population	Inexpressible Island	Northern Foothills	Edmonson Point	Total
Inexpressible Island	**11**	9	2	22
Northern Foothills	1	**5**	4	7
Edmonson Point	2	2	**3**	10

$\chi^2 = 5.587$, p=0.0159

Figure 3.3 Results of assignment tests for 1) living samples from the three colonies 2) ancient samples from the three colonies. The number of individuals correctly assigned is shown in bold type. χ^2 values and p-values for each assignment test are shown below each table.

3. Overall	Classification	
Source	Living	Ancient
Living	**57**	31
Ancient	16	**23**

$\chi^2 = 8.936$, p=0.0028

4. Inexpressible Island	Classification	
Source	Living	Ancient
Living	**22**	6
Ancient	6	**16**

$\chi^2 = 13.687$, p=0.0002

5. Northern Foothills	Classification	
Source	Living	Ancient
Living	**16**	12
Ancient	6	**4**

$\chi^2 = 0.972$, p=0.3243

6. Edmonson Point	Classification	
Source	Living	Ancient
Living	**25**	7
Ancient	4	**3**

$\chi^2 = 10.267$, p=0.0013

Figure 3.3 continued. Results of assignment tests for between the two time periods 3) overall, 4) for Inexpressible Island, 5) Northern Foothills and 6) Edmonson Point. The number of individuals correctly assigned is shown in bold type.χ^2 values and p-values for each assignment test are shown below each table.

general, indicate the presence of structure among ancient colonies. Overall F_{ST} and R_{ST} values between all ancient populations were not significant (0.038 and -0.026 respectively). However, both F_{ST} and R_{ST} values were considerably higher for comparisons between ancient samples than between living samples and one F_{ST} value (Northern Foothills vs. Edmonson Point) and one R_{ST} value (Inexpressible Island vs.

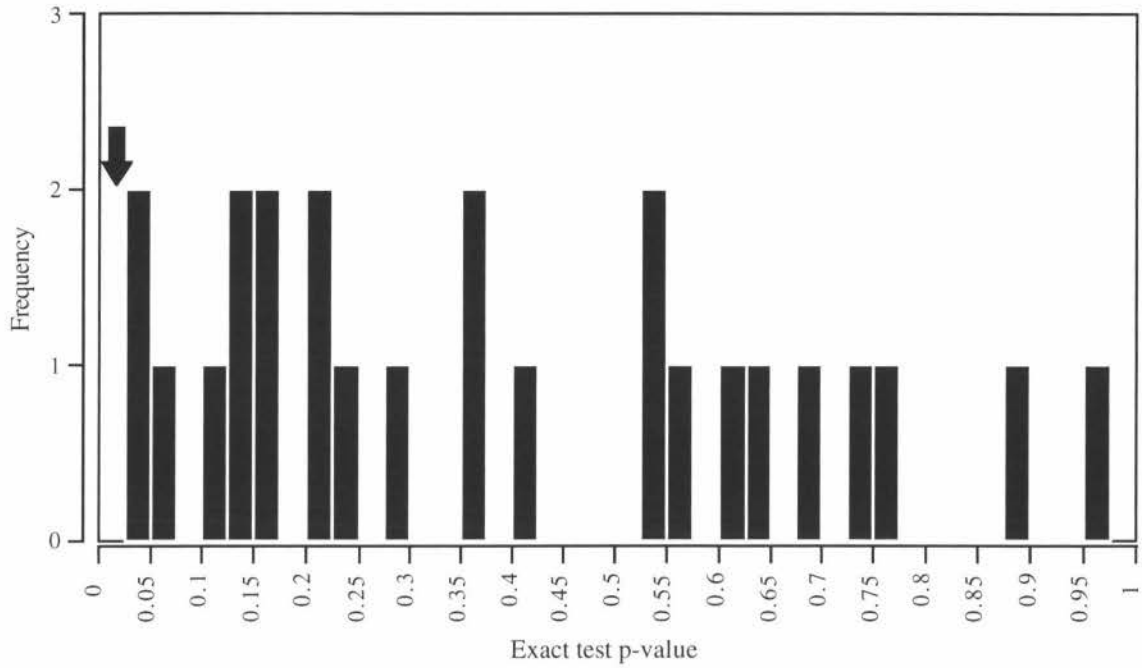

Figure 3.4. Graph of the p-values from the exact tests of simulated ancient microsatellite data (n=25). The p-values range from 0.027 to 0.961. The arrow indicates where the p-value (0.018) for the actual ancient samples falls in this distribution (ie. outside of the range of the simulated data).

Edmonson Point) were significant ($p>0.05$, Table 3.3). It should be noted that confidence intervals for the F_{ST} and R_{ST} values of the ancient samples are far larger than for the living samples. This is likely to be a result of the smaller sample sizes of the ancient populations. All of the confidence intervals exceed the 0.05 threshold level, suggesting that there is no certainty that the F_{ST} and R_{ST} values are not significant. Therefore, whilst all of the statistical tests of the living populations produced concordant results, measures of genetic structure between ancient Adélie penguin populations were ambiguous with an exact test and assignment test, but not F_{ST} and R_{ST} values, supporting the presence of structure.

Genetic analysis of Adélie penguins over time

Exact tests (Table 3.2), F_{ST} and R_{ST} values (Table 3.3) indicated a lack of genetic differentiation over time when samples from each time period were pooled i.e. when the total number of ancient samples were compared to the total number of living samples. However, assignment of these pooled samples to ancient and living categories resulted

in 63% of samples being assigned correctly (Figure 3.3); this result is significantly better than random ($\chi^2 = 8.936$, $p = 0.0028$).

Significant heterogeneity between temporal samples of the Inexpressible Island penguin colony was inferred from the exact tests, both by the TP500 locus data and averaged over all loci. The R_{ST} value, but not the F_{ST} value, for the comparison of the past and present populations on Inexpressible Island was also significant (p-value = 0.055), although the 95% confidence intervals have a large range and extend below the 0.05 threshold. This high variance is not unexpected and has been suggested to be a major drawback of statistics such as R_{ST} which incorporate allele size differences (Slatkin, 1995). Assignment of individuals at this colony as either ancient or living was correct 76% of the time ($\chi^2 = 13.687$, $p=0.0002$). Collectively, these results suggest that the genetic structure of this population has changed over time. Figure 3.5 illustrates the electromorph allele frequencies at each locus in both living and ancient samples from Inexpressible Island.

The exact test and F_{ST} values were not significant for the comparison between the past and present populations of the Northern Foothills colony. In addition, the assignment of individuals did not differ significantly from a random-based assignment ($\chi^2 = 0.972$, $p=0.3243$). In contrast, the R_{ST} value was significant (p-value = 0.085) although the 95% confidence intervals have a large range and extend below the 0.05 threshold.

The exact tests, F_{ST} and R_{ST} values also indicated that there was no significant differentiation between living and ancient populations from Edmonson Point. However, the assignment test resulted in individuals being assigned to their correct time period more often than expected by chance ($\chi^2 = 10.267$, $p=0.0013$).

3.4 Discussion

The lack of genetic differentiation among the living populations of Terra Nova Bay, as indicated by Fisher's exact test, assignment tests and genetic distance data, is in agreement with the previous study of Adélie penguin population genetics (Roeder *et al.*, in press). The reasons proposed in Roeder *et al.* (in press) for this lack of heterogeneity

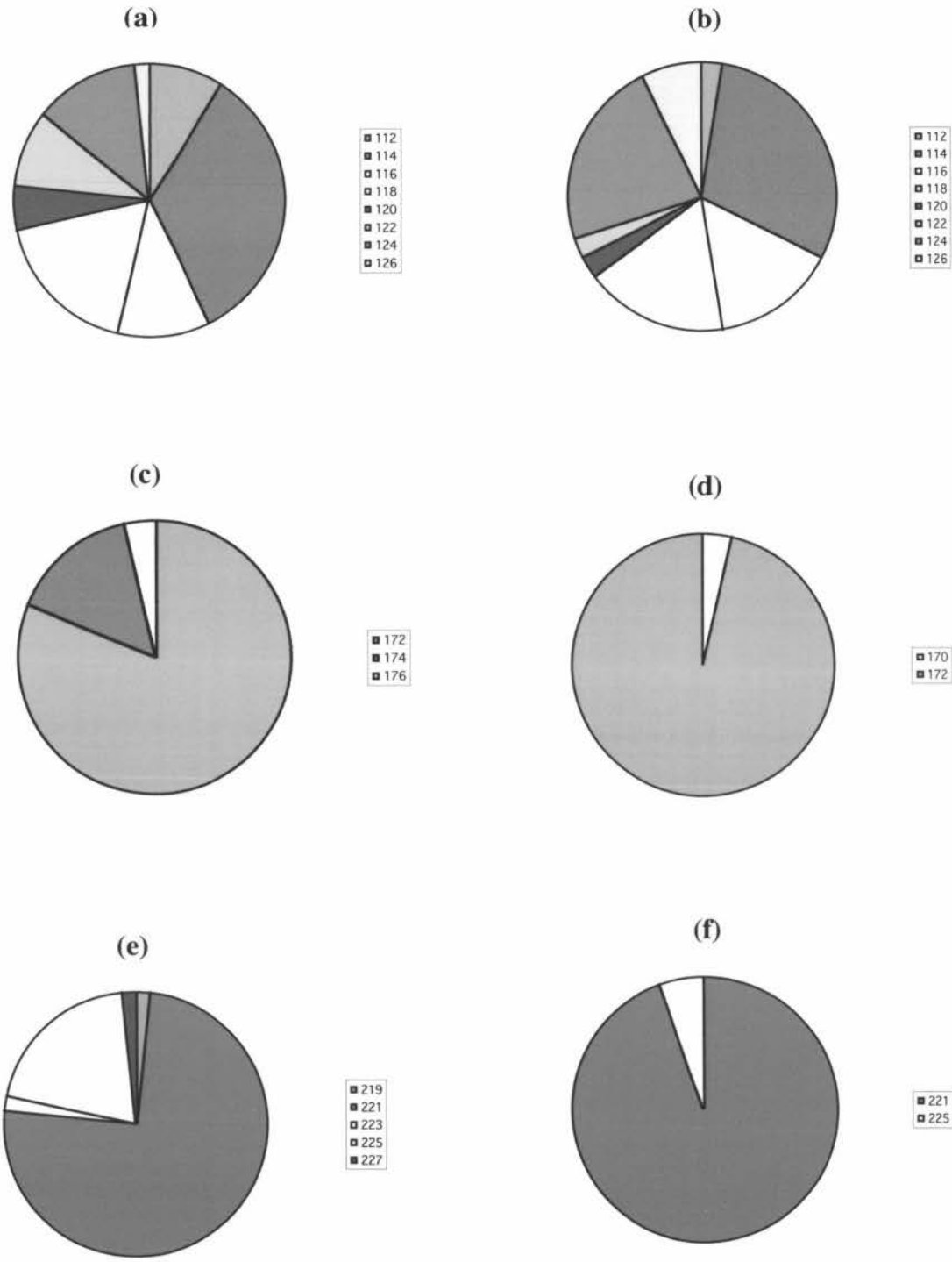

Figure 3.5 Electromorph allele frequencies at the TP500 locus (a) living (n=28) and (b) ancient (n=20), RM6 locus (c) living (n=27) and (d) ancient (n=14), RM3 locus (e) living (n=28) and (f) ancient (n=9) in populations of Adélie penguins from Inexpressible Island.

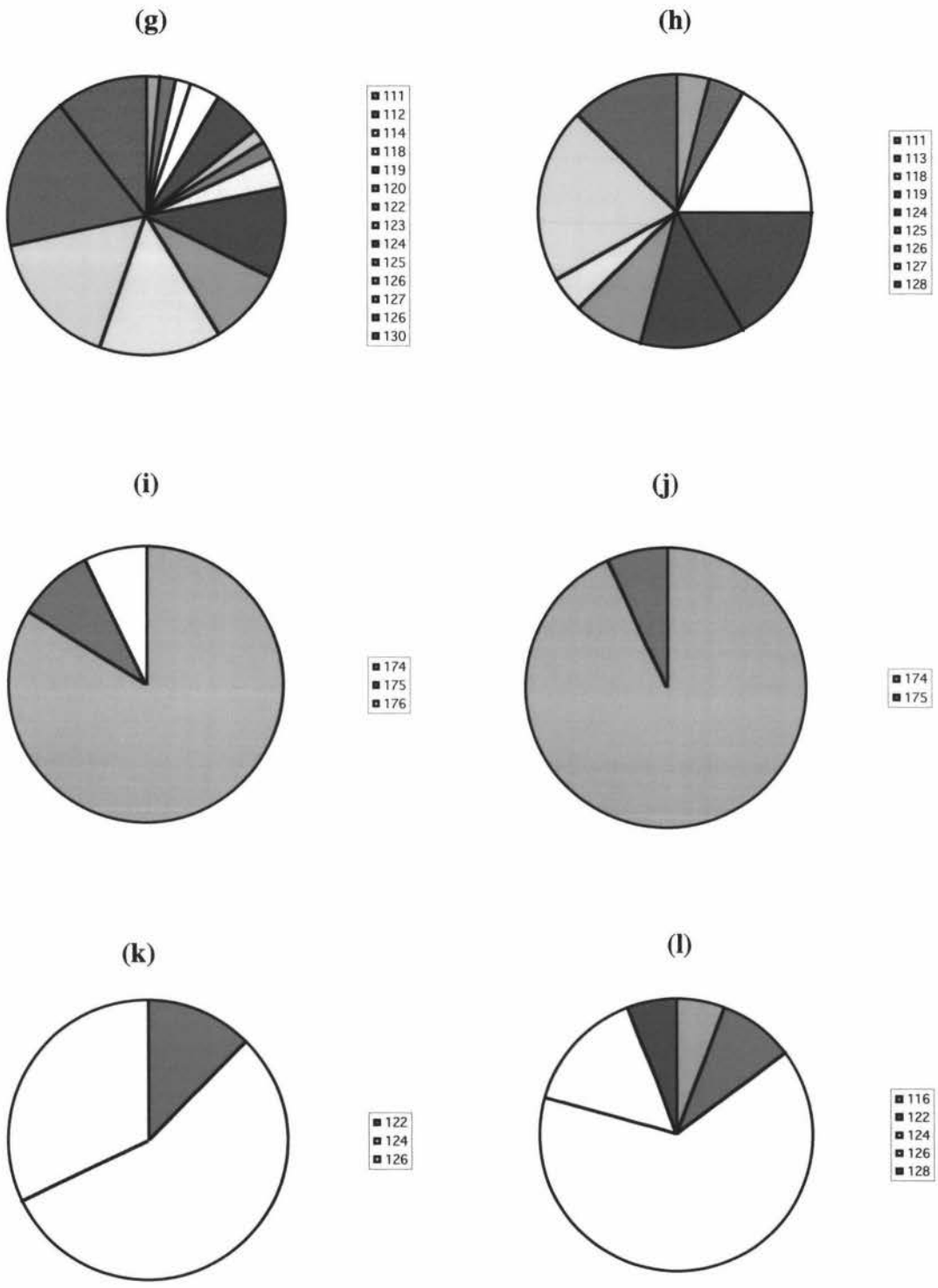

Figure 3.5 continued. Electromorph frequencies at the AM13 locus (g) living (n=28) and (h) ancient (n=12), AM3 locus (i) living (n=28) and (j) ancient (n=16), HrU2 locus (k) living (n=28) and (l) ancient (n=17) in populations of Adélie penguins from Inexpressible Island.

include a high level of dispersal and/or large population sizes resulting in stable gene and genotype frequencies.

The degree of structure between ancient populations is not so clear. An exact test between the ancient populations indicates that there is a significant level of structure present. The exact test p-value for the ancient data falls below the distribution of values calculated from simulated data sets (permuted from the living data sets to contain an equivalent amount of data as the ancient populations), suggesting that this significance is not just an artifact of the small sample size and missing data present in the ancient data set. However, only 25 simulated data sets were created because of the need to do these calculations by hand. A more accurate result would be achieved if a greater number of data sets were constructed (e.g. 1000). However, a specialised computer program would be required to calculate such a large number of data sets. An assignment test was able to accurately assign ancient individuals to their population of origin, which also suggests that the populations are genetically distinct.

In contrast, the F_{st} values and R_{st} values of overall relationships among the ancient populations do not suggest differentiation between them. The non-significance of these results may be caused by either the small sample sizes and missing data, because these tests are not powerful enough to detect low levels of differentiation, or because assumptions of the underlying models of these measures, in particular the model of evolution, are not fulfilled in this data set. However, the presence of differentiation cannot be rejected at a 95% confidence interval.

It should be noted that the ancient samples are not all from a single time period. The average age of samples from each colony varies by four thousand years. Therefore, the detection of differentiation between the ancient populations by the exact and assignment tests may represent temporal, not just spatial, differences.

Tests for differences in the genetic composition of colonies over time also gave variable results. Both an exact test, assignment test and R_{st} value suggest that the two temporal groups of samples from Inexpressible Island (i.e. ancient and modern) are genetically distinct. In contrast, only one of the four statistical tests indicated

differentiation between the two time periods for both the Northern Foothills and Edmonson Point samples; an R_{st} value and an assignment test respectively. This result is not unexpected because the Inexpressible Island ancient samples are, on average, at least a thousand years older than samples from the other two sites. These results may indicate that Adélie colonies in the greater Terra Nova Bay region prior to 5000 yrs BP (e.g. the Inexpressible Island ancient population) were genetically differentiated from modern populations. The level of differentiation may have decreased over time until it reached the modern level. Consistent with this is that there is less support for the Northern Foothills and Edmonson Point samples, which average have average ages of 4000 and 1000 years respectively, being differentiated from the modern samples.

Holocene climatic change: possible implications for Adélie penguin population genetics

Such temporal change in the genetic structure of the Adélie populations may be a reflection of climatic change during the Holocene. During the last ice age the majority of the Ross Sea region was occupied by a large marine based ice shelf (Ingólfsson *et al.*, 1998). This ice shelf is believed to have reached as far north as Coulman Island about 20 000 yrs BP (see Figure 3.6) (Licht *et al.*, 1996). During this time the lack of areas of open water during the summer, which Adélie penguins require, is likely to have resulted in a reduction in the number and size of their colonies. Consequently, Adélie penguins may have been restricted to refugia in which they were subject to quite different demographic processes than the large and expansive modern populations. Genetic drift in small refugial populations and low levels of dispersal between the scattered colonies may have led to inter-populational genetic differentiation. Following the Last Glacial Maximum (LGM) (25 000-18 000 yrs BP), Adélie penguin populations may have expanded (both in numbers and geographically). The timing and pattern of deglaciation has been inferred from a number of sources including glacial landforms, raised beaches and orthinographic soils from penguin rookeries (reviewed in Ingólfsson et al, 1998). These studies suggest that the minimum age for the retreat of the ice sheet past the Drygalski ice tongue, which forms the southern border of Terra Nova Bay, is approximately 11 400

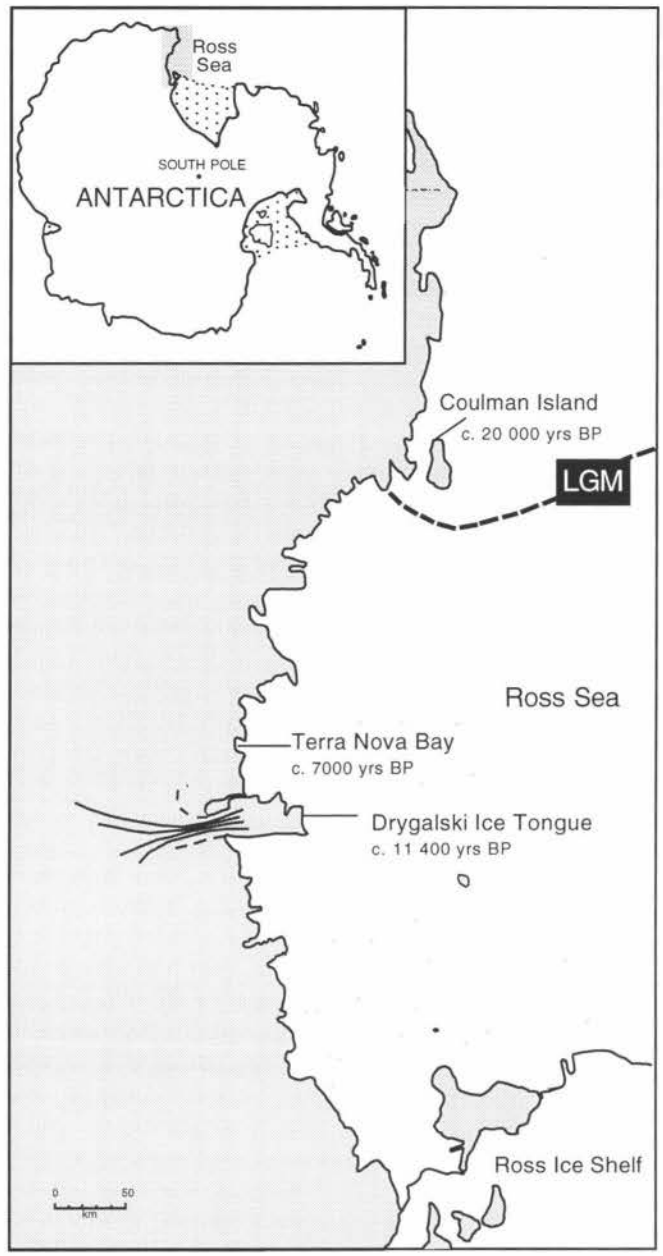

Figure 3.6. The Ross Sea region with estimated ages for deglaciation. The location of the Ross Ice Shelf grounding line at the Last Glacial Maximum (LGM) is indicated by a dotted line.

yrs BP. However, the Terra Nova Bay region is thought to have been occupied by ice for considerably longer. Glacial drift deposits indicate that Terra Nova Bay was occupied by a small ice shelf that was fed by glaciers draining the East Antarctic ice sheet (Baroni and Orombelli, 1991). Radiocarbon dates from penguin rookeries provide a minimum

date of 7000 yrs BP for when Terra Nova Bay probably became free of ice (Baroni and Orombelli, 1991). Greater dispersal of Adélie penguins following deglaciation may have reduced genetic differentiation between populations, while larger population sizes may have given rise to a more temporally 'stable' structure. This scenario may explain why the Inexpressible Island ancient population, with the oldest samples in this study, appears strongly differentiated from modern populations, whereas there is less support for the other younger populations (i.e. Northern Foothills and Edmonson Point) being differentiated from modern samples.

However, no firm conclusions can be made for several reasons. In particular, data from the ancient samples contains a large amount of missing loci and consist of only small numbers of samples. In addition, and perhaps more importantly, samples that have been called members of a particular ancient 'population' actually range in age over several thousand years. Therefore, they have not really been taken from a single ancient population.

This study suggests that investigating population genetics over long time periods using nuclear microsatellites is not straightforward. This is despite the ancient DNA used in this study being extremely well preserved. It should be pointed out that the samples used here were not collected specifically for this study, but instead to cover a considerable geographic range rather than a large number of bones of similar age from a few sites. Ideally a temporal population genetic study should involve a large number of bones from a single site from the same stratigraphic level to provide a 'snapshot in time'. For example, for follow-ups of this study, a greater number of samples from Inexpressible Island should be collected that are approximately 6000 yrs in age. The younger samples, which may mask differences that have occurred over time, could then be removed from the data set. Ideally, large numbers of samples from the same colony but separated into several different ages may provide a means by which to examine population change more closely.

3.5 Concluding Remarks

1. All the methods employed in this study suggest that there is no differentiation

between the three living populations of Adélie penguins in the greater Terra Nova Bay region.

2. An exact test of differentiation and an assignment test indicate the presence of genetic differentiation between the ancient populations.

3. Exact and assignment tests, as well as an R_{st} value, indicate that Adélie penguins populations from the two temporal time periods at Inexpressible Island are genetically distinct. This genetic change is suggested to be an indirect consequence of climatic change since the LGM. There is less support for differentiation between the temporal samples of the Northern Foothills and Edmonson Point colonies. Only one of the four statistical tests used was significant for each of these colonies.

4. The measures of genetic structure employed in this study often gave ambiguous results. Discrepancies between measures of genetic structure have been previously noted (Valsecchi *et al.*, 1997). The inconsistent results in this study may be a consequence of the small and unequal sample sizes, missing data, the level of differentiation being too low for the measures to detect and/or unrealistic models of microsatellite evolution.

CHAPTER FOUR

A Study of Microsatellite Evolution in Ancient and Living Adélie Penguins.

4.1 Introduction

Microsatellites are increasingly becoming the marker of choice for population genetic studies (Jarne and Lagoda, 1996; Estoup and Angers, 1998). However, their use in estimating numerous population parameters (e.g. genetic differentiation, number of migrants per generation, etc.) depends upon the mutational model of evolution that is assumed (Estoup and Angers, 1998). It is widely believed that current theoretical models do not accurately portray microsatellite evolution because there is insufficient knowledge of the complex mutational processes occurring at these loci (Estoup and Cornuet, 1999; Chambers and MacAvoy, 2000). In this study, mutational processes at microsatellite loci will be explored. In addition, homoplasy, which has recently become the focus of many microsatellite studies because of its possible influence on population genetic parameters, will be examined.

In the majority of studies that have employed microsatellites, alleles have been identified by length alone (via electrophoretic mobility) without detailed nucleotide sequence information. Recently a number of researchers have sequenced microsatellite alleles and have found that differences in repeat number are not the only form of variation between microsatellite alleles at a locus (e.g. Grimaldi and Crouau-Roy, 1997; Taylor *et al.*, 1999b). Alleles may also differ by interruptions in the repeat, as well as nucleotide substitutions and indels in the flanking sequences. This additional variation may result in the products of distinct evolutionary lineages being categorised as alleles of the same size. This process is known as size homoplasy. Size homoplasy has been found to be a common feature of microsatellite loci and has been observed in

individuals from the same population (e.g. Viard *et al.*, 1998; Makova *et al.*, 2000), from different populations of the same species (e.g. Orti *et al.*, 1997; Taylor *et al.*, 1999b) and between species (e.g. Garza and Freimer, 1996; Angers and Bernatchez, 1997). Several features of microsatellite alleles result in them being particularly prone to size homoplasy. Firstly, there is some evidence to suggest that microsatellites evolve in a stepwise fashion (see Chapter One, section 1.2), therefore restricting possible mutational outcomes. Secondly, the high mutation rates increase the probability of parallel mutations occurring in different lineages (Taylor *et al.*, 1999b).

Specifically, the aims of this chapter are to explore microsatellite evolution and homoplasy on three scales. Firstly, microsatellite evolution was examined by genotyping ancient Adélie subfossil bones and comparing the electromorphs they contain with those that are presently found in living populations. Secondly, alleles from both living and ancient Adélie penguins were examined through sequence analysis. Lastly, microsatellite evolution was examined on a broader scale by sequencing alleles from a number of penguin species and, at one locus, these sequences were mapped onto an independently derived phylogeny.

4.2 Materials and Methods

Sample collection

The comparison of living and ancient Adélie electromorphs involved all of the ancient samples that were described in Chapter Two (section 2.2). The living samples included the Terra Nova Bay samples from chapter three as well as the 442 individuals described in Roeder *et al.* (in press). These living samples cover much of the distribution of Adélie penguins in Antarctica (Figure 4.1).

The blood or tissue samples from the other penguin species were supplied, as described in the preface.

DNA extraction and microsatellite genotyping

All the ancient DNA samples described in Chapter Two (section 2.2) were also used in

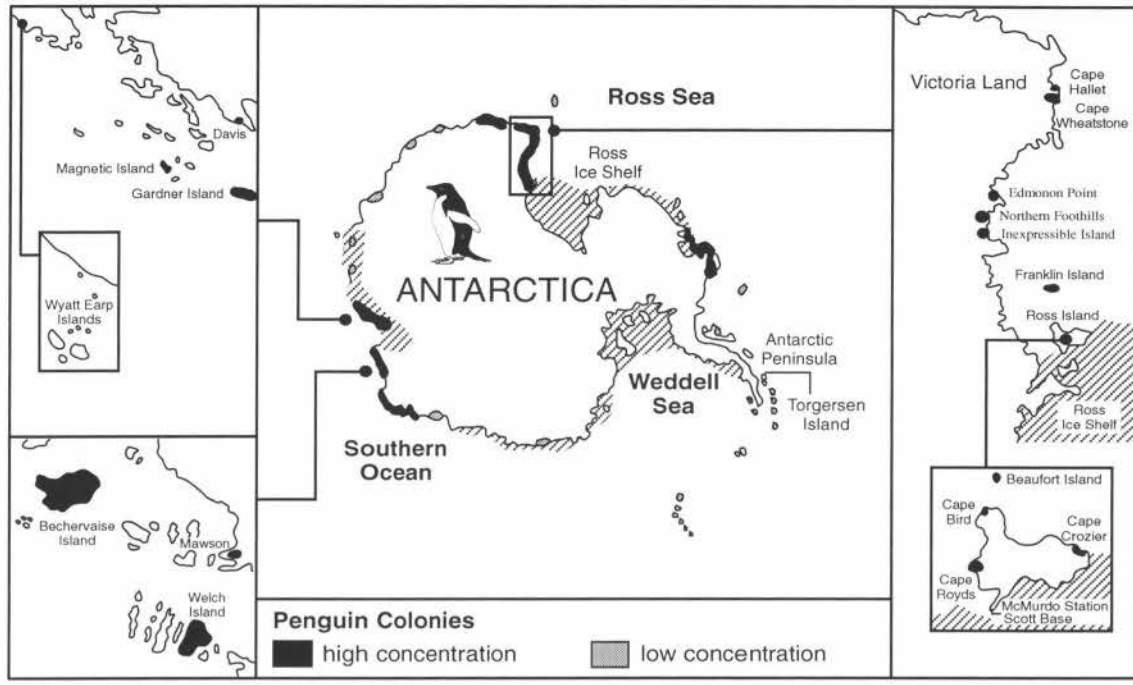

Figure 4.1 Location of Adélie penguin breeding colonies and blood sample collection sites around Antarctica (adapted from Roeder *et al.*, in press). The samples collected cover much of the distribution of Adélie penguins.

this part of the study (see Figure 2.1 for the distribution of these samples). Terra Nova Bay Adélie blood samples were extracted as described in Chapter Three (section 3.2). DNA from blood samples of the remaining Adélie samples and blood or tissue from other penguin species was extracted using a protocol modified from Sambrook *et al.* (1989). About 100 mg of ground tissue or 4 µl of blood was added to 400 µl of extraction buffer (10 mM Tris-HCl pH 8.0, 1 mM NaCl), 10 µl of 10% SDS and 10 µl of 10 mg/ml proteinase-K. This mixture was incubated overnight in a rotisserie at 55°C.

Proteins were removed by adding 400 µl of phenol to each sample. They were rocked for 30 minutes followed by centrifugation at 13 000g for 5 minutes and removal of the phenol. This phenol step was then repeated. Next, 400 µl of chloroform:isoamyl alcohol (C:I, 24:1) was added and the rocking and centrifugation steps were repeated. Finally, the chloroform:isoamyl was removed.

The DNA was precipitated by adding 1 ml of 100% ethanol and 40 µl of 3 M NaOAc pH 5.2 to each sample and placing at -20°C for 1 hour. The samples were then

centrifuged for 15 minutes at 13 000g and the supernatant removed. The pelleted DNA was then washed by vortexing with 500 µl of 70% ethanol. This was followed by centrifugation at 13 000g for 15 minutes. The ethanol was removed and the pellet dried under vacuum for 10 minutes. Finally, the pellet was resuspended overnight in 200 µl of 10 mM Tris-HCl pH 8.0.

Genotyping of the ancient and Terra Nova Bay samples was performed using an Applied Biosystems 377 sequencer as described in Chapter Two (section 2.2). The remaining living Adélie samples and the samples from other penguin species were genotyped using the autoradiographic technique described in Roeder *et al.* (in press). Because two genotyping techniques were used, and the results were to be compared, a representative of each allele at each locus was genotyped using both techniques. The sizes of the alleles genotyped with an automated sequencer were adjusted so that they were equivalent to the alleles genotyped by the autoradiographic technique.

Determination of allele sequence

Alleles from the RM3, AM3, AM13 and FhU2 loci were sequenced. Samples that had been genotyped as homozygotes were directly sequenced, whereas each allele of a heterozygote was cloned before sequencing. PCR products of both homo- and heterozygotes were purified by using the High Pure PCR product purification kit (Roche) following the manufacturer's instructions. Samples were concentrated to 10 µl under vacuum at 45°C. PCR products were quantified by running 1 µl with 2 µl Low DNA Mass™ ladder (GibcoBRL®) on a 1% MS, 1% LE agarose gel.

Cloning of PCR products

Following purification and concentration, alleles of heterozygotes were cloned using the pGEM®-T Easy vector system (Promega). PCR products were ligated for one hour at room temperature in 10 µl volumes containing 3 units of T4 DNA ligase, 2X rapid ligation buffer, 50 ng pGEM®-T Easy vector, 10 ng PCR product and MilliQ H_2O. DH5α cells were made competent by a procedure similar to that described by Pope and Kent (1996). Firstly, a liquid culture was obtained by inoculating 2 ml of autoclaved

L-broth (LB) (containing 1% bacto-peptone, 1% NaCl and 0.5% bacto-yeast extract) with a single colony, and incubating at 37°C overnight on a shaker. 50 ml LB was then inoculated with 750 µl of the overnight culture of DH5α cells. They were grown at 37°C on a shaker until log phase (~4 hours). The cells were pelleted by centrifugation at 800g for 5 minutes. The supernatant was then removed, the pellet gently vortexed and the cells resuspended in 20 ml 0.1M $CaCl_2$, and then placed on ice for 30 minutes. The cells were then centrifuged as above and resuspended in 2 ml 0.1M $CaCl_2$. Competent cells were stored at 4°C for 24 hours before being transformed. Transformation involved combining 100 µl of competent cells with 3 µl ligation reaction and placing on ice for 5 minutes. The cells were then spread onto agar plates (containing 1% bacto-peptone, 1%NaCl, 0.5% bacto-yeast extract, 1.5% bacto-agar and 100 µg/ml amplicillin), pre-warmed to 37°C, and then grown overnight at 37°C.

Six colonies were selected from each plate and screened by PCR amplification using the standard M13 primers whose sequences are complementary to sequences flanking the insertion site of the vector. Each PCR reaction mix was inoculated with cells from a single colony. Cells from the same colonies were grown in 2 ml LB broth containing ampicillin (100 µg/ml) on a shaker overnight at 37°C. PCR products were sized by running on a 1% LE/1% MS agarose gel. DNA was extracted from the cultures that contained inserts of the correct size by using a High Pure plasmid isolation kit (Roche) and following the manufacturer's instructions.

Following extraction, the samples were screened by an *Eco*R1 digest. 5 µl of plasmid was digested at 37°C for two hours with 1 µl *Eco*R1 enzyme, 1 µl of 10X buffer and 3 µl MilliQ H_2O. After digestion the samples were run on a 1% LE/1% MS agarose gel. Samples in which the insert was present were then sequenced.

DNA sequencing and purification

ABI Prism cycle sequencing reactions contained 4 µl terminator ready reaction mix (Big Dye termination mix for AM3, AM13 and FhU2, and dGTP termination mix for RM3) (PE Applied Biosystems), 3.2 µl of either a forward or reverse primer (10 µM), 20-30 ng PCR product, and made up to a total volume of 10 µl with MilliQ H_2O. After

being overlain with mineral oil, reactions were performed in a Hybaid Omni-E Thermal Cycler. Thermocycling conditions consisted of 25 cycles of 96°C for 10 seconds, 50°C for 5 seconds and 60°C for 4 minutes. The products were purified using an isopropanol clean-up. This involved placing each sample (minus the mineral oil) in a new eppendorf tube and adding 10 µl MilliQ H_2O and 80 µl of 75% isopropanol. These were vortexed, precipitated at room temperature for 15 minutes then centrifuged for 20 minutes at 13 000g to pellet the DNA. The supernatent was then carefully removed with a pipette and discarded. A further 250 µl of 75% isopropanol was added, the samples vortexed and centrifuged for 5 minutes at 13 000g. The isopropanol was again removed and the pellet dried under vacuum at 45°C for 10 minutes. Samples were then stored at -20°C.

Samples were sequenced at the Massey University sequencing facility on an Applied Biosystems ABI Prism 377, according to the manufacturer's instructions.

DNA sequence analysis

The microsatellite DNA sequence data was aligned using Sequencher™ 3.1.1(Gene Codes Corp.). At the RM3 locus the repeat region was aligned manually because Sequencher was unable to align these sequences. This alignment was constructed by allowing indels to account for the variation in repeat number and reducing the number of base substitutions in the repeat.

4.3 Results

The evolution of Adélie microsatellites over time

A comparison of the electromorphs present in ancient and living Adélie penguins at six loci revealed no novel electromorphs in the ancient samples (see Table 4.1). However, several electromorphs were present in living Adélie penguins that were not detected in ancient samples (e.g. electromorphs 168 and 180 at the RM6 locus). These electromorphs were uncommon in the living samples so it is likely that they were not found in the ancient samples because fewer of these samples were genotyped. The allele discovery curves for each of the six loci for both ancient and living samples (Figure 4.2) flatten, suggesting that the variation at each locus has been sampled.

Table 4.1. The electomorphs present in living and ancient Adélie penguin samples at six loci. Electromorph frequencies are given and n is the number of samples genotyped.

Locus	Electromorph	Frequency living	Frequency ancient
TP500		n=280	n=111
	106	0.002	0.000
	108	0.018	0.009
	110	0.012	0.000
	112	0.113	0.176
	114	0.241	0.252
	116	0.173	0.171
	118	0.099	0.081
	120	0.066	0.050
	122	0.173	0.135
	124	0.095	0.122
	126	0.007	0.005
RM6		n=519	n=96
	168	0.002	0.000
	170	0.006	0.010
	172	0.842	0.844
	174	0.130	0.140
	176	0.012	0.005
	180	0.002	0.000
RM3		n=517	n=70
	219	0.025	0.007
	221	0.777	0.779
	223	0.004	0.014
	225	0.159	0.179
	227	0.031	0.021
	231	0.002	0.000
AM13		n=504	n=52
	109	0.001	0.000
	110	0.002	0.000
	111	0.014	0.009

Table 4.1. continued.

Locus	Electromorph	Frequency living	Frequency ancient
	112	0.012	0.009
	114	0.005	0.009
	115	0.001	0.009
	118	0.007	0.075
	119	0.085	0.321
	120	0.041	0.047
	121	0.011	0.019
	122	0.016	0.000
	123	0.043	0.047
	124	0.176	0.094
	125	0.112	0.000
	126	0.114	0.085
	127	0.097	0.023
	128	0.167	0.094
	129	0.018	0.028
	130	0.076	0.009
	132	0.001	0.000
	135	0.001	0.000
AM3		n=522	n=79
	173	0.001	0.000
	174	0.789	0.873
	175	0.169	0.127
	176	0.042	0.000
HrU2		n=123	n=94
	116	0.004	0.021
	118	0.004	0.011
	120	0.012	0.021
	122	0.110	0.090
	124	0.512	0.585
	126	0.352	0.245
	128	0.028	0.016
	130	0.028	0.005
	134	0.004	0.005

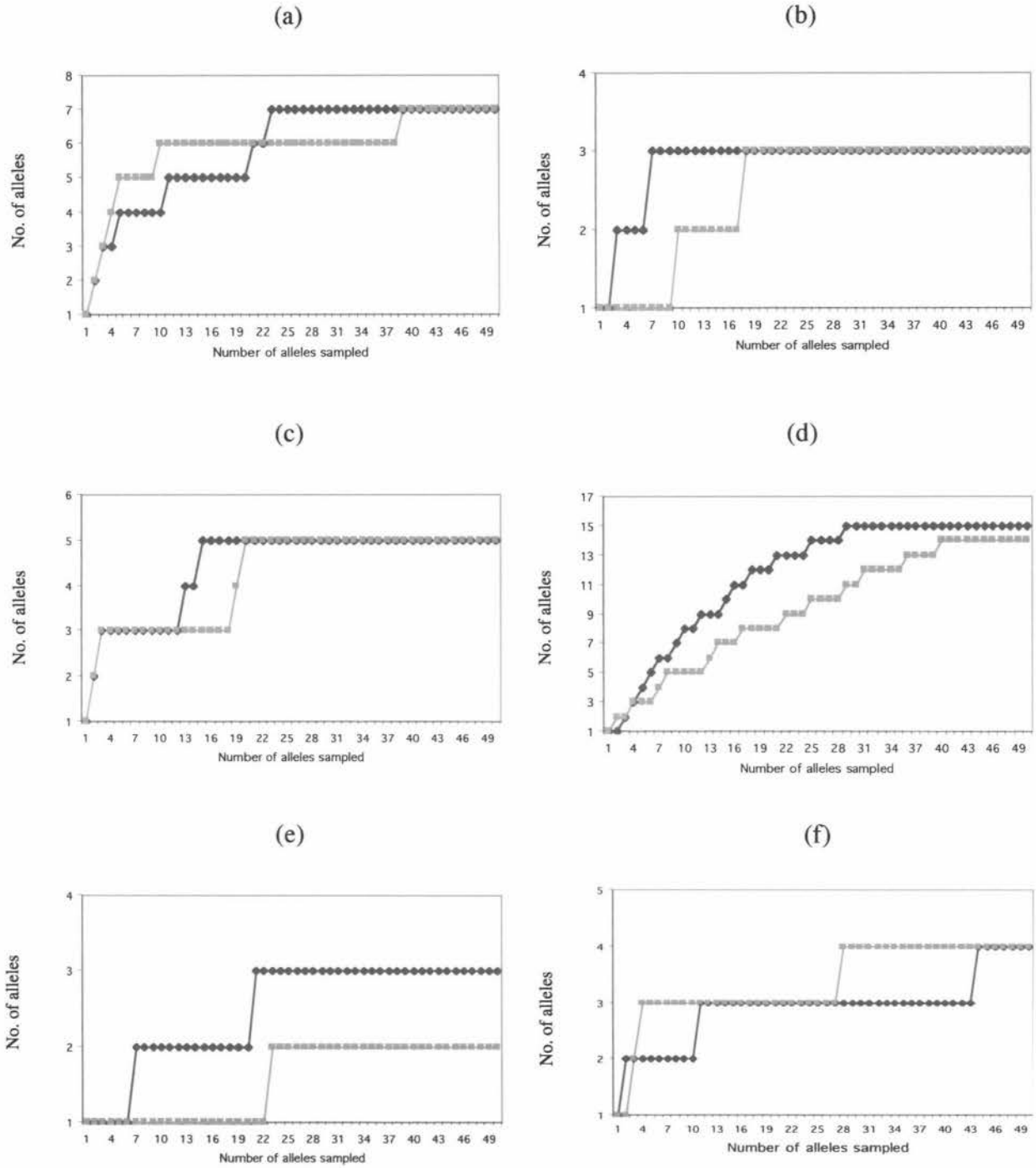

Figure 4.2 Allele discovery curves for six microsatellite loci: (a) TP500, (b) RM6, (c) RM3, (d) AM13, (e) AM3 and (f) HrU2. The blue diamonds represent the living samples and the red squares represent the ancient samples. The X-axis values represent t cumulative number of alleles sampled and the Y-axis indicates the cumulative number o new alleles sampled.

The RM3 locus

A total of 31 DNA sequences were obtained for four different electromorphs at the RM3 locus in Adélie penguins, resulting in a total of 15 different sequence alleles. Eight of the 'homozygotes' that were sequenced exhibited heteroplasmy (double peaks) (see Figure 4.3), suggesting that more than one sequence was present. These sequences were not included in the analysis.

Within each Adélie electromorph, sequence alleles differed by substitutions in the repeat region and/or in the flanking sequence (see Table 4.2). The Adélie alleles can be grouped into two genealogical groups based on two sites where mutations have occurred (Figure 4.4). Group One contains five 221 bp alleles, one 225 bp and one 227 bp allele and is defined by $(TA)_2$ at the beginning of the microsatellite repeat and T at character 45. Group Two contains five 221 bp alleles, two 223 bp alleles and a 225 bp allele and has a single TA at the beginning of the microsatellite and a C (or in one case an A) at character 45. Allele 221j appears to be intermediate between the two groups because it has a flanking region from each group, i.e. a single TA at the beginning of the microsatellite and a C at character 45. The population of origin for each of these sequence alleles is shown in Figure 4.5. It should be noted that three of these sequence alleles (221c, 221h and 221i) are unique to the ancient samples.

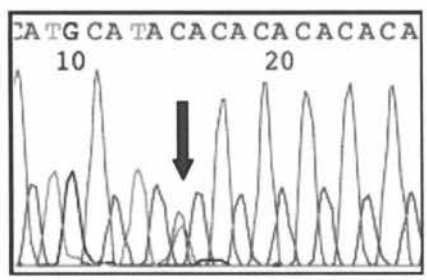

Figure 4.3 An example of sequence heteroplasmy (arrowed) in an RM3 'homozygote' This individual probably contains two different alleles - one allele with a thymine at tl site and the other with a cytosine. The majority of RM3 'homozygotes' in which heteroplasmy was observed had double peaks at the site pictured above (immediately adjacent to the repeat tract) and at character 45.

Table 4.2. Polymorphic sites in Adélie RM3 alleles (sequences are listed 5' to 3'). Nucleotides identical to the consensus sequence (CS) are labelled with a dot. n is the number of sequenced alleles. Only the repeat region for allele 227 is included because the flanking region sequenced poorly.

Allele	n	Repeat region	14	24	31	33	39	45	68	69	71	84	92	99	104	105
CS		$(TA)_n(CA)_n(C)_n$	A	G	A	T	T	C	A	A	T	G	T	T	G	T
221a	3	$(TA)_2(CA)_9(C)_2$	.	.	.	.	.	T	.	.	.	.	.	.	.	.
221b	1	$(TA)_2(CA)_9(C)_2$	G	.	.	.	.	T	.	.	.	.	.	.	.	.
221c	1	$(TA)_2(CA)_9(C)_2$	.	.	.	G	.	T	.	.	G	.	A	.	.	.
221d	2	$(TA)_2(CA)_7(C)_6$	.	.	.	.	.	T	.	.	.	.	.	.	.	.
221e	1	$(TA)_2(CA)_7(C)_6$	.	.	.	.	.	T	.	.	.	C	.	.	.	.
221f	5	$(TA)(CA)_{10}(C)_2$	.	.	.	.	.	.	.	.	.	.	.	.	.	.
221g	1	$(TA)(CA)_{10}(C)_2$	.	.	.	.	.	A	.	.	.	.	.	.	.	.
221h	1	$(TA)(CA)_{10}(C)_2$	.	.	.	.	.	.	.	.	.	C	.	A	.	A
221i	1	$(TA)(CA)_9(C)_4$	.	.	.	.	.	.	.	.	.	.	.	.	.	.
221j	1	$(TA)(CA)_9(C)_4$	.	.	.	.	.	T	.	.	.	.	.	.	.	.
221k	1	$(TA)(CA)_9(C)_4$	.	.	.	.	.	.	.	.	G	.	.	.	.	.
223	2	$(TA)(CA)_{11}(C)_2$	.	.	.	.	.	.	.	.	.	.	.	.	T	.
225a	1	$(TA)_2(CA)_{11}(C)_2$	.	.	G	.	.	T	.	.	.	.	.	.	.	.
225b	1	$(TA)(CA)_2(GA)(CA)_8C_4$	.	.	.	.	C	.	G	G	.	.	.	.	.	.
227	1	$(TA)(CA)_{13}(C)_2$														

Figure 4.6 shows the possible relationships between the Adélie penguin RM3 sequence alleles using a parsimony criterion to minimise the overall number of mutational steps. This diagram is based on the assumption that the substitution rate is lower than the occurrence of length changes. Allele 221a is considered to be the ancestral allele in this diagram because it is closest to the alleles that were sequenced in Chinstrap penguins (the closest available relative to Adélie penguins). Allele 221a differs from the most common Chinstrap allele by one base pair (a 1 bp transition in the flanking region).

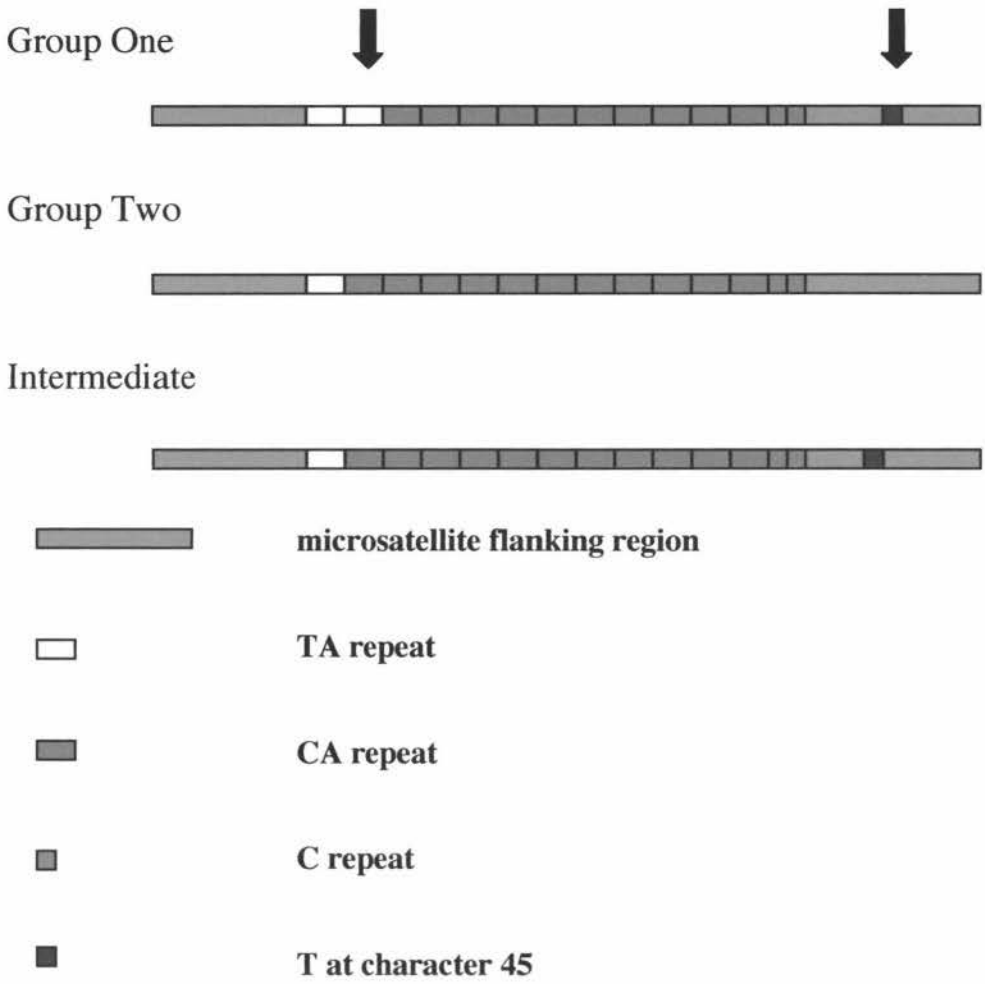

Figure 4.4 Schematic diagram of a representative allele from each of the two groups of RM3 sequence alleles, as well as the allele (221j) that appears intermediate between the two groups. The defining characters of the two groups are the number of TA repeats before the CA repeat and the nucleotide at character 45.

A further 22 copies of two electromorph alleles at the RM3 locus were sequenced in 11 other penguin species (Table 4.3). Apart from the Chinstrap penguin (and Gentoo penguin which was genotyped in Roeder et al., submitted, although not sequenced in this study), all of these other penguin species were monomorphic for an electromorph of 211 bp at this locus. However, the sequence of this electromorph demonstrated a number of substitutional differences between species in both the flanking sequence and repeat region. These alleles can be separated into three groups based on differences in the number of cytosine residues in the repeat region (Figure 4.7). As in Adélie penguin alleles, the number of cytosine nucleotides present at the end of the microsatellite repeat region is two, four or six. Mapping the three forms of the RM3 repeat region on

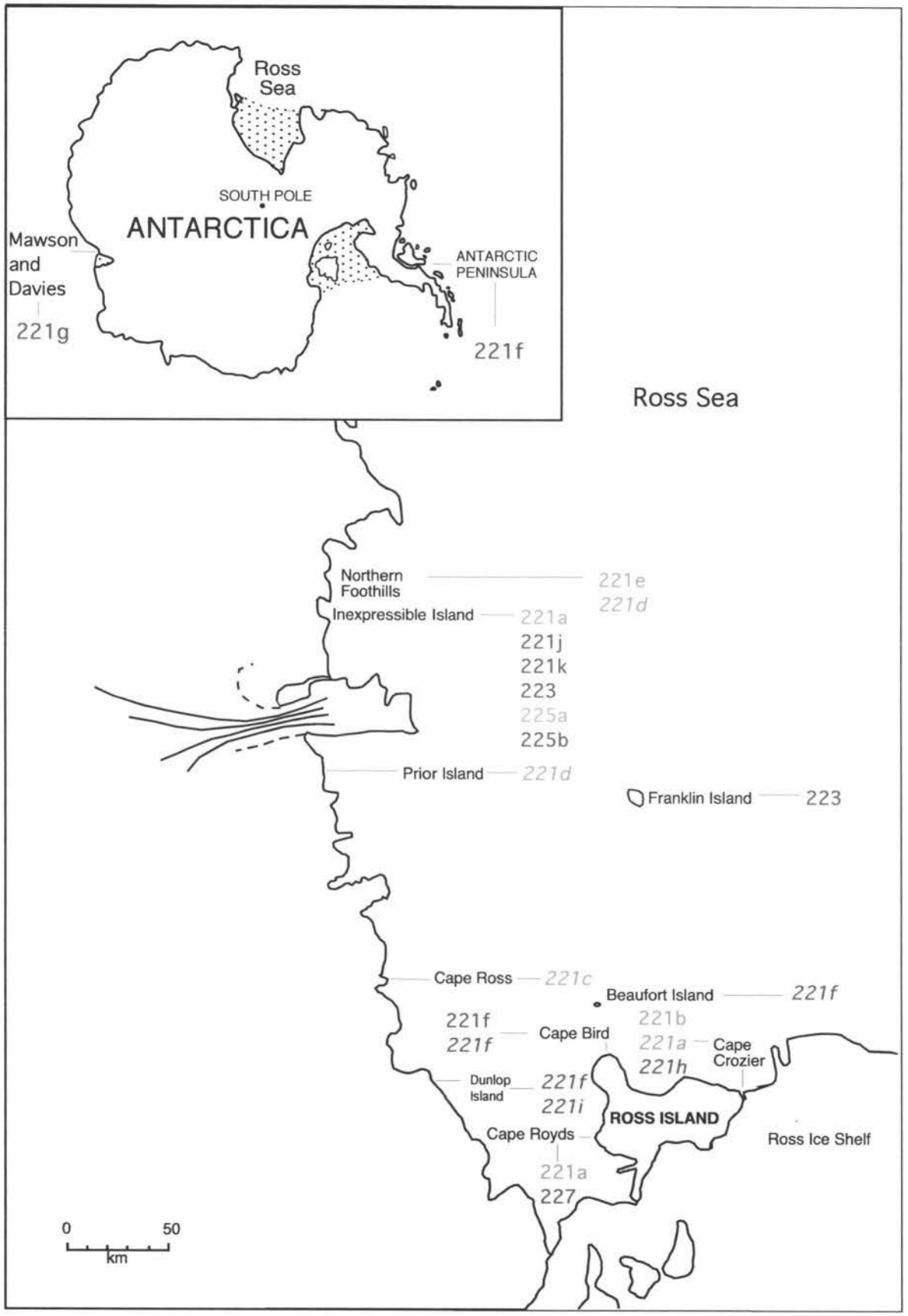

Figure 4.5. The distribution of RM3 alleles around Antarctica. Group One alleles are in red and Group Two alleles are in blue. Alleles from ancient samples are shown in italics.

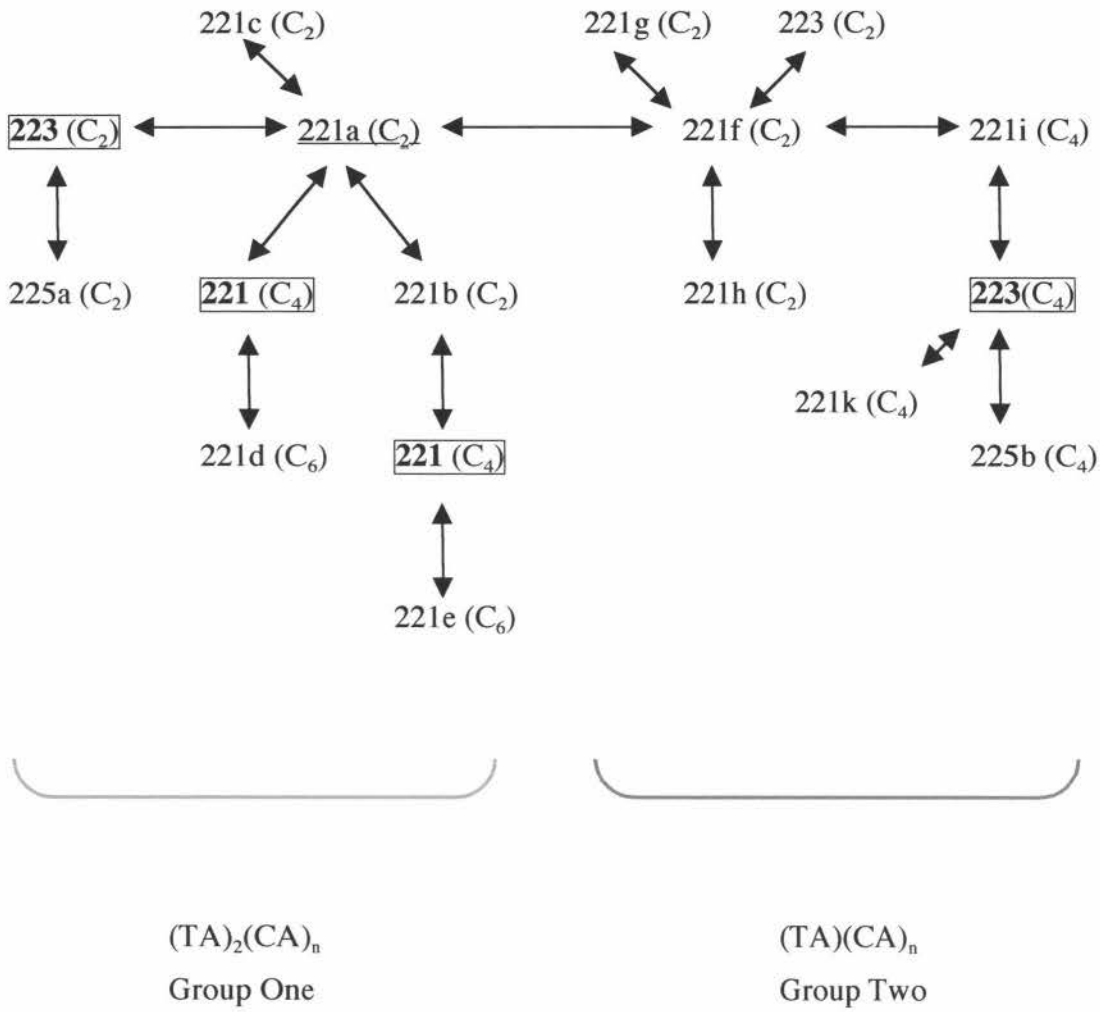

Figure 4.6 A diagram illustrating the possible relationships between the Adélie penguin RM3 alleles sequenced in this study. A parsimony criterion was used to minimise the number of overall mutational steps. This diagram is based on the assumption that the substitution rate is lower than the occurrence of length changes. Each arrow represents either a change in a single repeat unit or, when the same size alleles are joined, one or more substitutions in the flanking region. Hypothetical intermediates of the repeat region are shown in boxes and the number of cytosine nucleotides at the end of the $(CA)_n$ repeat tract is in brackets. Allele 221a is underlined to indicate that it is the likely ancestor.

Table 4.3. Polymorphic sites at the RM3 locus in penguin species excluding Adélie penguins (sequences are listed 5' to 3'). Nucleotides identical to the consensus sequence (CS) are labelled with a dot. n is the number of sequenced alleles. The Chinstrap alleles are 221 bp in length and the alleles in the remaining penguin species are 211 bp.

Allele	n	Repeat region	14	15	23	28	35	39	40	46	49	55	61	68	86	88	91	94	101	105	106	110	124
Consensus		$(TA)_n(CA)_n(C)_n$	A	A	A	A	A	G	T	C	T	G	T	G	G	A	T	T	T	T	T	A	T
Chinstrap 1	3	$(TA)_2(CA)_9(C)_2$	.	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.	.	.	.	C
Chinstrap 2	1	$(TA)_2(CA)_9(C)_2$	.	G	.	.	C	.	G	T	.	.	.	.	C	.	.	.	.	.	.	.	C
Emperor	1	$(TA)(CA)_4(C)_6$	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.
Little Blue	2	$(TA)(CA)_5T(C)_5$	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.
African	1	$(TA)(CA)_6(C)_2$	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Humbolt	1	$(TA)(CA)_5(C)_4$	.	.	.	.	.	.	.	.	.	.	.	.	.	G	C	.	.	.	.	.	.
Galapagos	1	$(TA)(CA)_6(C)_2$	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Erect-cr.	1	$(TA)(CA)_4(C)_6$	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Snares-cr.	1	$(TA)(CA)_5(C)_4$	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Fiordland-cr.	2	$(TA)(CA)_5(C)_4$	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Rockhopper 1	1	$(TA)(CA)_5(C)_4$	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Rockhopper 2	1	$(TA)(CA)_6(C)_2$	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Yellow-eyed 1	2	$(TA)(CA)_6(C)_2$	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Yellow-eyed 2	1	$TA(CA)_4TA(C)_4$	.	.	.	.	.	C	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.
Yellow-eyed 3	1	$TA(CA)_4TA(C)_4$	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Yellow-eyed 4	1	$(TA)(CA)_6(C)_2$	C	.	G	.	.	.	.	.	.	C	C	T	.	.	.	C	.	.	C	C	.
Yellow-eyed 5	1	$(TA)(CA)_4(C)_$	.	.	.	.	.	.	.	.	.	.	C	T	.	.	.	C	C	.	C	C	.

to a penguin phylogeny (Figure 4.7) demonstrates that they are not clustered together but are scattered across the tree.

The AM3 locus

At the AM3 locus two copies of each of the three electromorph alleles (174, 175 and 176) present in Adélie penguins were successfully sequenced. Each of these six sequences came from an individual from a different population. One electromorph allele was also sequenced in each of the following species of penguin: Rockhopper, Fiordland-crested, Chinstrap, Gentoo and Yellow-eyed penguins. The Adélie penguin AM3 microsatellite repeat consists of two discrete regions where length mutations have occurred (see Table 4.4). An AT unit is repeated either four or five times, and a poly-T repeated eight or nine times. In the other penguin species, there were no length variation in the alleles that were sequenced. However, the poly-T region was shorter in the other penguin species than in Adélie penguins with only six T repeats.

In Adélie penguins, the only electromorphs that were identical in sequence were the two 174 bp alleles. The remaining pairs of alleles differed by substitutions in either the flanking or repeat sequences (see Table 4.4). The 176 bp electromorphs had identical flanking sequences, but differed by a single substitution within the AT repeat. The 175 bp alleles differ by a 1 bp substitutions in the flanking region.

The FhU2 locus

The FhU2 locus was sequenced in seven penguin species (Table 4.5). This locus had been previously amplified in 15 diverse bird species, including Adélie penguins (Primmer and Ellegren, 1998). The Adélie penguin sequence amplified in this study was identical to the sequence recorded by Primmer and Ellegren (1998). The only length variation observed in any of the penguin species was a 2 bp deletion in the flanking region of the Yellow-eyed penguin sequence. Therefore, there was no length variation in the microsatellite repeat in the penguin species examined. Three different single base pair substitutions were observed. In relation to the consensus sequence, the Emperor penguin sequence had two different transitions, and the Adélie sequence had a single base pair transversion.

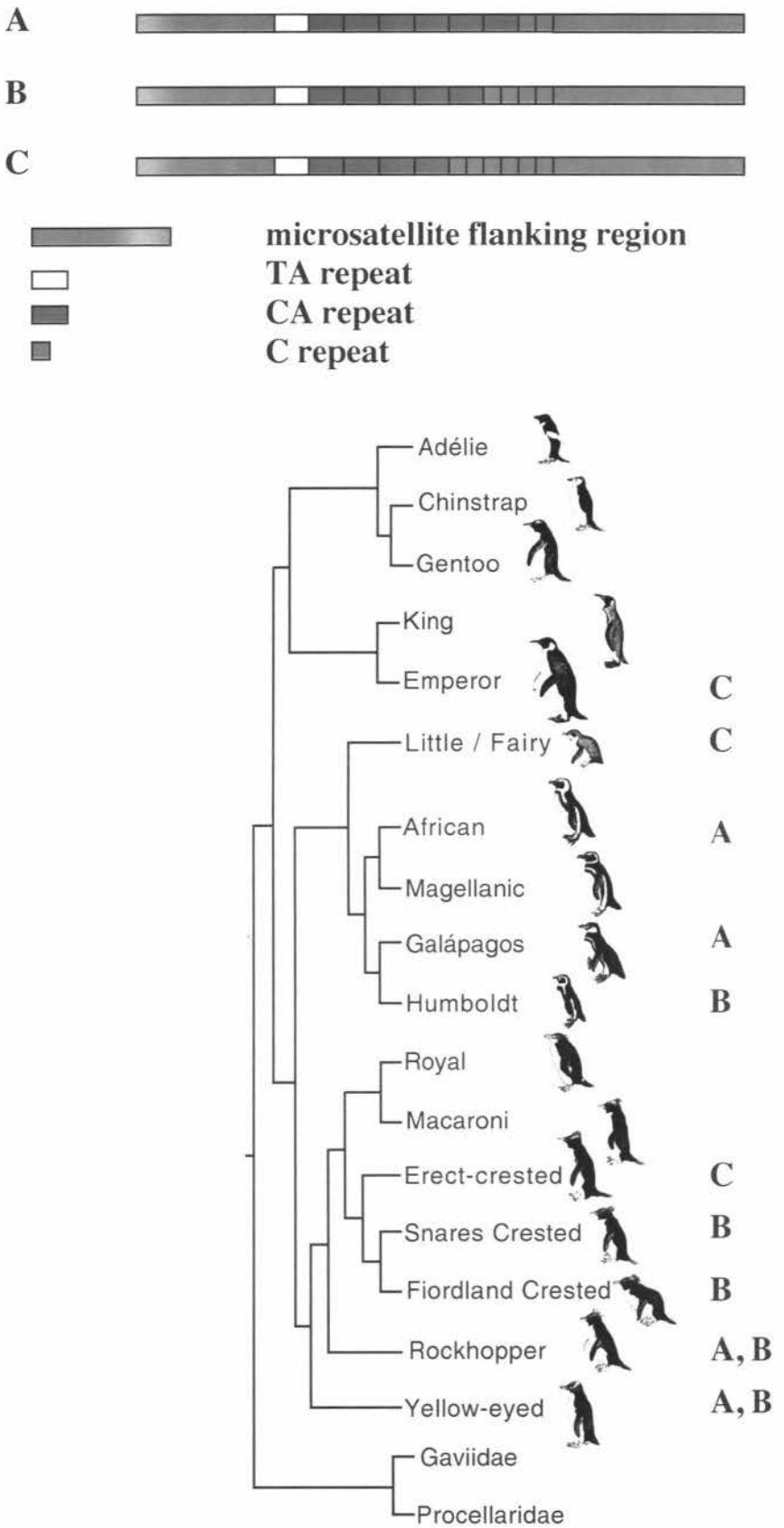

Figure 4.7 Top. Schematic representation of the three forms of the 211 bp allele at the RM3 locus. Bottom. The distribution of these three allele forms on a penguin phylogeny. The phylogeny is from Roeder et al. (submitted).

Table 4.4 Polymorphic sites in penguin AM3 alleles. Nucleotides identical to the consensus sequence (CS) are labelled with a dot. n is the number of sequenced alleles.

Species	Allele length	n	1	7	3 1	3 7	repeat region	4 4
CS	174		A	T	C	C	$(AT)_n$	GGT_n
Adélie	174	2	.	.	.	.	$(AT)_4$	T_8
Adélie	175a	1	.	.	.	.	$(AT)_4$	T_9
Adélie	175b	1	.	C	.	.	$(AT)_4$	T_9
Adélie	176a	1	.	.	.	.	$(AT)_5$	T_8
Adélie	176b	1	.	.	.	.	$(AT)_4AA$	T_8
Gentoo	174	1	.	.	T	.	$(AT)_4$	$GG\ T_6$
Chinstrap	174	1	.	.	T	T	$(AT)_4$	$GG\ T_6$
Fiordland-crested	174	1	.	.	T	.	$(AT)_4$	$GG\ T_6$
Rockhopper	174	1	.	.	T	.	$(AT)_4$	$GG\ T_6$
Yellow-eyed	174	1	C	.	T	.	$(AT)_4$	$GG\ T_6$

The AM13 locus

The AM13 locus was the most polymorphic locus of those examined in this study, with 21 electromorph alleles. This locus demonstrated a high level of heterozygosity. Therefore, there were few electromorph homozygotes. Five of these electromorph homozygotes were directly sequenced in Adélie penguins, and one each from Galapagos, African and Yellow-eyed penguins. The electropherograms demonstrated heteroplasmy suggesting that two sequences were present for each of these 'homozygotes', i.e. heterozygotes had actually been sequenced. Therefore, despite the

Table 4.5. Polymorphic sites in penguin FhU2 alleles. Nucleotides identical to the consensus sequence (CS) are labelled with a dot. Dashes represent deletions. One allele was sequenced from each species.

Species	1	3	3 1	3 2	repeat region	5 3
CS	T	C	C	T	$(GA)_4$	A
Adélie	C	.	.	.	.	.
Emperor	.	A	.	.	.	G
Humboldt	.	.	.	.	.	.
Snares-crested	.	.	.	.	.	.
Fiordland-crested	.	.	.	.	.	.
Rockhopper	.	.	.	.	.	.
Yellow-eyed	.	.	–	–	.	.

AM13 locus already having a large number of electromorph alleles, it seems that an even greater number of sequence alleles may be present.

4.4 Discussion

Conclusions about several aspects of microsatellite evolution, including the occurrence of homoplasy, mutational mechanisms, microsatellite formation and flanking sequence instability, can be drawn from this data.

Extensive allele size homoplasy of penguin microsatellite alleles

The results of sequencing microsatellite alleles in this study suggest that genetic diversity at these loci is not purely from length changes in the microsatellite repeat regions. This observation is in agreement with those from other species such as humans

(Grimaldi and Crouau-Roy, 1997), mice (Makova *et al.*, 2000), honeybees (Estoup *et al.*, 1995b) and cichlid fishes (van Oppen *et al.*, 2000). Size homoplasy was observed at three of the loci examined (RM3, AM3 and FhU2 loci), and was also likely to be present at the AM13 locus. Homoplasy at the RM3 locus was found to occur among alleles from the same population of Adélie penguins, between populations of Adélie penguins, as well as between penguin species. In addition, the observation that a number of the RM3 'homozygotes' had double peaks suggests that homoplastic alleles may also be present within individuals.

Variation within the flanking regions of the Adélie RM3 locus allowed the reconstruction of possible relationships between alleles. This reconstruction suggested that alleles of the same size are not necessarily more closely related to each other than to alleles in non-adjacent size classes (e.g. the two 225 bp alleles at the RM3 locus were members of different genealogical groups).

The occurrence of homoplasy amongst alleles sequenced in this study may mean that although no new microsatellite electromorphs were found in the ancient Adélie subfossil bones, unique sequence alleles may still be present. At the RM3 locus three sequence alleles were unique to ancient samples (221c, 221h and 221i). However, the number of samples sequenced was so small that it cannot be concluded that these alleles are not present in living Adélie penguins.

Mutational mechanisms

Discrimination between the mutational mechanisms that have been proposed to result in length change at microsatellite repeats requires variation in the flanking sequences. Recombinational mechanisms, e.g. unequal crossing over and gene conversion, are expected to result in the exchange of flanking sequences, whereas slippage would not. Sequence alleles of the RM3 locus show variation in the flanking region following the repeat tract and, although only a small length of sequence was recovered from the flanking region preceeding the repeat, the TA repeat at the beginning of the microsatellite region can be considered a marker. Polymorphism in these two regions define two genealogical groups of sequence alleles. The separation of the sequence alleles into two groups suggests that the majority of mutations at the RM3 locus appear

to result from slippage rather than recombination. This is consistent with other studies of microsatellites with sufficient flanking region variation to discriminate between these mutational mechanisms (e.g. Orti *et al.*, 1997; Makova *et al.*, 2000). However, one allele at the RM3 locus has an intermediate sequence (221j). It may have originated in three possible ways: via recombination, by a back mutation in either the TA repeat or at character 45, or it may represent an ancestral intermediate mutational stage between the two groups. Sequencing a longer length of flanking region before the microsatellite repeat may allow discrimination between these hypotheses. It should also be noted that even if this allele was formed by recombination there is no evidence that this was accompanied by a change in the number of repeat units.

Microsatellite formation

Little is known about the process by which microsatellites form. It is widely assumed that random point mutations expand repeat arrays beyond the minimum number of repeats on which slippage can act (Schlötterer, 2000). The threshold number of nucleotides required for microsatellite expansion is suggested to be eight to ten (i.e. four to five dinucleotide repeat units) (Messier *et al.*, 1996; Rose and Falush, 1998). However, more recent studies have suggested that there is no critical number of repeats above which microsatellites become hypervariable, with shorter microsatellites having the potential to expand, albeit with a lower probability than longer microsatellites (Pupko and Grauer, 1999; Zhu *et al.*, 2000; Noor *et al.*, 2001).

Microsatellite evolution at the FhU2 locus in 15 bird species, including Adélie penguins, was discussed by Primmer and Ellegren (1998). They suggest that Adélie penguins possess the ancestral repeat form of $(GA)_4$, and that this is not a sufficient number of repeats to allow slippage to occur. In passerines, where repeat length polymorphism is present, the minimum total repeat length was 14 bp, with this extended polymorphic state possibly evolving twice independently in the passerine group. In this study, FhU2 alleles were sequenced from other penguin species, and it was revealed that these species also contain the ancestral number of repeats. Therefore, this locus appears to have maintained its stability across the diversification of penguin species. The two repeat regions of the AM3 locus also contain close to eight nucleotides each.

There is little length variation in either of the repeat regions, suggesting that there are not enough repeat units for slippage to occur often.

In contrast to the stability of the FhU2 and AM3 loci, expansion appears to have occurred at the RM3 locus in some species of penguin. The majority of penguin species are monomorphic for a 211 bp electromorph allele which contains $(CA)_4C_6$, $(CA)_5C_4$ or $(CA)_6C_2$ repeats. However, the repeat region has expanded in the ancestor of the extant *Pygoscelis* lineages because Gentoo, Chinstrap and Adélie penguins all have longer alleles. This locus is also polymorphic in the three *Pygoscelis* species which is in agreement with the positive relationship between repeat number and polymorphism that has been observed in other studies (e.g. Primmer and Ellegren, 1998).

A comparison of the flanking region sequence provides information concerning whether initial increases in repeat number have occurred via slippage or base-pair substitution. Base-pair substitution in the flanking region resulting in the formation of an additional repeat unit would cause the flanking region to change in sequence. Substitution may have occurred in repeat regions of both the AM3 and RM3 loci. The poly-T region of the AM3 locus has undergone two transversions from G→T in the Adélie lineage, thus increasing the number of T repeats from six to eight. Therefore, at this locus substitution appears to have increased the number of adjacent T nucleotides to the minimum slippage threshold of eight nucleotides suggested by Rose and Falush (1998). In Adélie penguins the number of T nucleotides varies from eight to nine with this variation probably caused by slippage rather than substitution. In the RM3 locus a C→A transversion may have occurred several times thus increasing the number of repeat units (although the ancestral state, and thus the direction of mutation, is unknown).

Instability at the end of the RM3 microsatellite sequence

An interesting observation from Table 4.3 is the variability in the number of cytosine nucleotides at the end of the Adélie penguin RM3 microsatellite repeat. These cytosines are always present in multiples of two, suggesting that they are the result of a transversion between an A and a C in the microsatellite repeat, rather than the result of slippage during replication of the cytosine nucleotides. The direction of mutation, i.e.

whether an A→C or C→A, is unknown because the ancestral state is uncertain. This change in the number of cytosine nucleotides appears to have occurred independently in both groups of alleles because the flanking region mutation at character 45 is only likely to have occurred once. This variation in the number of cytosine nucleotides is also present in the other penguin species resulting in three forms of allele (A,B,C). These other penguin species are monomorphic for an electromorph allele of 211 bp that further supports the hypothesis that change in the number of these cytosine nucleotides has been generated by substitution rather than slippage, since slippage would be expected to generate allele length change.

A mutation at character 124, separating all Adélie and Chinstrap alleles from those in the other penguin species, suggests that the variation in the number of cytosine nucleotides is homoplastic between the *Pygoscelis* group (i.e. Adélie and Chinstraps) and the remaining penguin species. The mutation at character 124 would have initially occurred in a 2C, a 4C or a 6C allele, with the remaining two states subsequently regenerating in parallel.

The scattered distribution pattern of the 2C, 4C and 6C alleles over the phylogeny of the non-*Pygoscelis* penguins (Figure 4.7) is also likely to be a result of homoplasy. However, an absence of phylogenetically informative flanking region variation means that ancestral polymorphism cannot be ruled out. In addition, only a small number of individuals from each species were sampled, so the possibility exists that all forms of the 211 allele are present in each species but that not all were detected.

The data suggests that substitutions occur more often at the 3' end of the RM3 CA repeat than in the middle of the repeat region. This is in agreement with the findings of Grimaldi and Crouau-Roy (1997) who observed a relatively unstable flanking region at one end of a $(CA)_n$ microsatellite from the human MHC region. Brohede and Ellegren (1999), who examined a large set of $(CA)_n$ microsatellites in sheep and cattle, suggest that instability is restricted to a 5-10 bp region at the border of the repeat. However, neither of the above two studies noted a bias in the substitutions towards any particular nucleotide base, as appears to be present at the RM3 locus.

Several models, all associated with mismatch repair during either replication or recombination, have been proposed to account for this observed instability in the border region (Brohede and Ellegren, 1999). Firstly, an increased chance of substitution in this border region could be produced if there is an increased chance that slippage will occur at the end of a repeat tract. Several studies have reported the occurrence of polarity in slippage mutations (see Chapter One, section 1.2). Secondly, DNA loops may form at the repeat border during recombination between two different length microsatellite alleles. These loops may form when the migrating recombination joint reaches the end of the microsatellite tract where the sequence is not homologous (i.e. repeat units are present on one strand, and flanking region on the other). In both of the models proposed above, the DNA loops formed are either excised or filled in by DNA mismatch repair systems. If the mismatch repair enzymes which fill in these loops have a lower fidelity compared to DNA polymerases involved in standard replication, then it would result in a higher substitution rate (Brohede and Ellegren, 1999). Consistent with this hypothesis is the observation that there is significant variation in polymerase fidelity (Kunkel *et al.*, 1992).

Neither of the models proposed by Brohede and Ellegren (1999) provide an explanation for why the same mutation, an A↔C transversion, has apparently occurred independently at two positions at the RM3 locus in each Adélie allele group, as well as in different penguin species. However, if the mutations that have occurred are A→C transversions then an explanation may be provided by crystallographic studies of DNA duplexes. According to Timsit (1999), accurate DNA replication depends not only upon polymerase fidelity but also the DNA sequence, although the latter's role is less well understood. Analysis of the crystal structure of $(CA)_n$ tracts reveals that structural alterations, rather than slippage due to looped out bases, may explain the high rate of replication errors at these loci. Crystallographic, NMR and biochemical studies have shown that $(CA)_n$ tracts may exhibit shifted base-pairing during replication (reviewed in Timsit and Moras, 1996). This occurs when the nucleotide bases pair with their direct 5' neighbours on the opposite strand, rather than their Watson-Crick complements. The bases form a run of consecutive A·G and C·T mispairs on the major groove side, while the base-pairing is unaltered in the minor groove (the major and minor grooves are the two grooves in the sugar-phosphate backbone produced by the stacking of bases in the DNA helix - Figure 4.8).

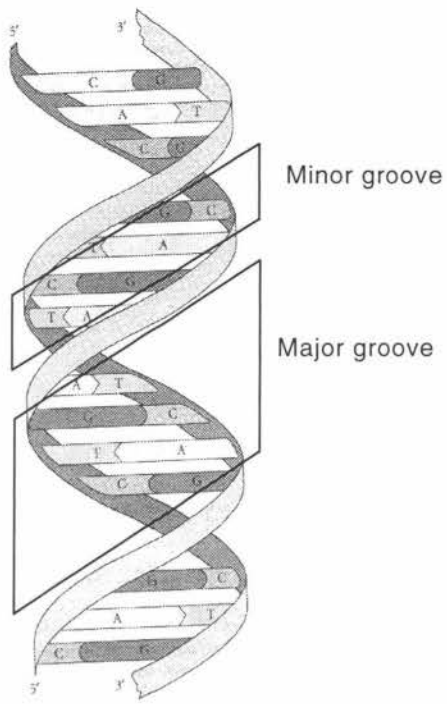

Figure 4.8. Diagram illustrating the position of the major and minor grooves on the DNA double helix. (The DNA drawing is from Futuyma, 1998).

The apparently normal geometry in the minor groove may mislead correction enzymes, which have been shown to detect replication errors through the incorrect geometry they cause in the minor groove of the nascent duplex. Timsit (1999) suggests that base substitution mutations may be produced if the nascent template-primer strands realign prior to a subsequent nucleotide incorporation (Figure 4.9). The base substitution is templated by the next template base. This base substitution may escape correction because the strand has the correct geometry in the minor groove. Under this model, the bias towards A→C transversions at the RM3 locus is explained by the base substitution being templated by the adjacent guanine nucleotide (Figure 4.9).

4.5 Concluding Remarks

1. No novel electromorph alleles were found in the Adélie penguin ancient samples despite the examination of a large number of both living and ancient genotypes.

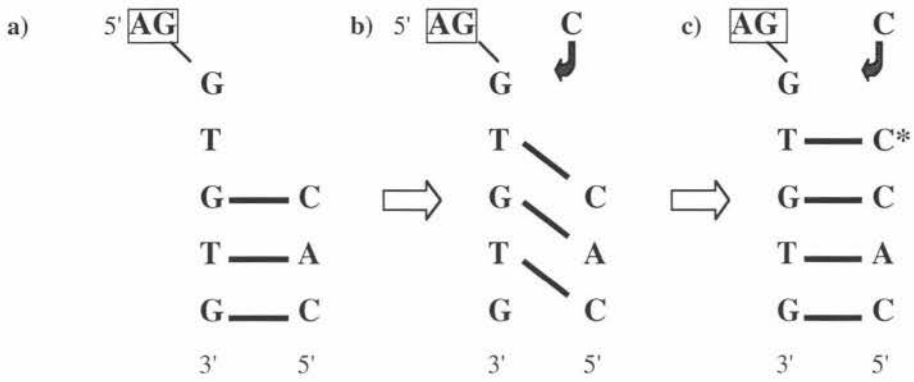

Figure 4.9. A pathway for an A→C transversion in an RM3 allele.

a) A template (left-strand)-primer duplex with normal base-pairing. The flipped out template bases (non-repeat tract) are boxed.
b) The base-pairing may shift within a repeat tract without movement of the bases. The thymine template base that should be paired with an incoming dATP forms a T•C mismatch. Consequently, the next base is considered a template for dNTP incorporation, in this case a guanine is the template base for an incoming dCTP.
c) The DNA strands realign prior to nucleotide incorporation and the incoming dCTP interacts with the thymine template base forming a T•C mismatch (the newly incorporated cytosine is marked with asterisk).

Novel sequence alleles from the RM3 locus were found in the ancient samples. However, the low number of sequences from living Adélie penguins meant that it could not be concluded these sequence alleles were truly unique to the ancient samples.

2. Homoplasy is a common feature of the loci that were sequenced. For a given electromorph allele a number of sequence alleles are often 'hidden', even within a population. These sequence alleles are not detected using conventional electrophoretic methods. This observation, and its possible impacts on this study

(particularly, Adélie population genetics and microsatellite evolution over time), will be discussed further in the following chapter.

3. The majority of length mutations at the microsatellite loci investigated in this study are consistent with generation by slippage rather that recombination. However, the occurrence of a large number of mutations in the microsatellite flanking regions of the loci examined suggests that the mutational processes occurring at these loci may be considerably more complex than simple DNA slippage.

4. Sequence from the AM3 locus suggests that initial microsatellite formation has occurred via substitutions creating a sufficient number of repeat units on which slippage has then occurred.

5. The RM3 locus exhibits instability in the region bordering the repeat tract. In particular, an A↔C transversional bias appears to be present. If this bias is an A→ C substitution, then it is proposed to be a product of inaccurate DNA replication caused by structural features of the DNA.

CHAPTER FIVE

Summary and Discussion of Future Work

5.1 Synopsis of Major Findings

The major findings of this thesis are summarised below in relation to the research aims (italicised) from Chapter One (section 1.6):

• *To reliably amplify and characterise ancient single-copy microsatellite DNA from subfossil Adélie penguin bones by adhering to the 'criteria of authenticity' developed by researchers in the ancient DNA field.*

In this study, single-copy microsatellite DNA was readily, and reliably, extracted from a large number of ancient Adélie penguin subfossil bones. The extraction process and PCR set-up were carried out in a dedicated ancient DNA laboratory. Controls and replication were routinely used to check for contamination. The excellent preservation of DNA in these bones is likely to be a consequence of the cold, dry, undisturbed Antarctic environment.

• *To examine temporal population genetic change at three Adélie penguin colonies by comparing microsatellite DNA data from ancient and living Adélie penguins.*

All the statistical tests used in this study suggest that there is no genetic differentiation between the three populations of living Adélie penguins from the Terra Nova Bay region. In contrast, both an exact test and an assignment test indicated the presence of differentiation between ancient Adélie penguins from these sites. These tests also suggest that the ancient and living Inexpressible Island populations have genetically diverged over time, as do an R_{ST} value. There is also some support for the younger

Northern Foothills and Edmonson Point temporal populations being genetically differentiated. However, to make these conclusions more robust, more ancient samples of similar ages are required.

• *To investigate mutational mechanisms occurring at microsatellite loci by comparing electromorph alleles from ancient and living Adélie penguins.*

A comparison of six loci present in ancient and living Adélie penguins revealed that there were no novel electromorph alleles in the ancient samples. However, there is a possibility that novel sequence alleles exist in the ancient samples. Limited sequence analysis revealed several unique sequence alleles in the ancient samples. However, insufficient sequence data from living Adélie penguins means that no firm conclusions can be drawn.

• *To explore the evolution of four microsatellite loci on a broader scale by sequencing alleles from a number of species of the Spheniscidae, and for one locus mapping the sequences on to a phylogeny.*

The sequencing of alleles from several microsatellite loci in both Adélie and other penguin species resulted in the discovery of substantial size homoplasy. This finding suggests size homoplasy may be a common feature of microsatellite loci.

Microsatellite flanking region variation at the RM3 locus permitted an examination of the mutational mechanisms proposed to produce length change at microsatellite loci. Slippage is considered the most plausible mechanism for the generation of length change at this locus in Adélie penguins. Sequence from the AM3 locus in a number of penguin species suggests that substitutions initially create a sufficient number of adjacent repeat units on which slippage can occur.

A transversional bias was observed in the region bordering the microsatellite repeat tract of the RM3 locus. This bias is suggested to be a consequence of inaccurate DNA replication resulting from structural features of the DNA.

5.2 Future Work

Population genetics

There is still a considerable amount of future work that could be done on Adélie penguin population genetics using living and ancient material. Firstly, more genotyped samples are required from ancient subfossils to investigate whether population differentiation is present between the ancient populations. These samples should be approximately the same age and preferably the data generated from them should contain no missing elements.

Secondly, a greater number of loci would increase the resolving power of population differentiation measures (reviewed in Estoup and Angers, 1998). Therefore, more loci should be used in any future studies. There is little discussion in the literature of the ideal number of loci to use, although Takezaki and Nei (1996) claim that R_{ST} estimates will have large standard errors unless at least thirty loci are considered.

Thirdly, sequence alleles, rather than electromorph alleles, could be used to provide more accurate estimates of population differentiation and structure. The finding of high levels of size homoplasy at the loci examined in this study is likely to have an impact on the analysis of population structure. Population structure was examined using electromorph alleles rather than sequence alleles. Therefore, the number of electromorph alleles, heterozygosities and genetic diversities within and between populations was likely to be underestimated in Chapter Three because only allele size was considered. However, the effect of homoplasy on population structure is more difficult to predict (Rousset, 1996). Previous studies have demonstrated that size homoplasy does not necessarily mask population structure (Angers *et al.*, 2000). The effect of homoplasy on population parameters depends upon the fraction of electromorph alleles shared between populations (Viard *et al.*, 1998). When few electromorph alleles are shared between populations, homoplasy will not have a great influence upon measures of population genetic structure. However, if a large proportion of electromorph alleles is shared between populations, then sequencing may reveal different alleles in different populations and thus increase the measured population structure. Studies that have investigated the effect of microsatellite homoplasy on

population structure have examined populations that exhibited a high level of differentiation when electromorph alleles alone were considered (Viard *et al.*, 1998; Angers *et al.*, 2000). Moreover, both these studies examined only a single locus. The one study (Sunnocks *et al.*, 2000) that has examined multiple loci, obtained inconsistent results from different loci when studying population differentiation using electromorph alleles. Most loci had fixed differences between populations (i.e. supported differentiation). However, some loci had electromorph alleles common in both populations. Sequencing one such locus demonstrated substantial homoplasy.

In this study the Adélie penguin populations shared all of the common electromorph alleles for all six loci and, when electromorph alleles alone were considered, there was no genetic differentiation between living populations. Therefore, genetic differentiation may be present between living Adélie penguin populations and is masked at all of the loci examined. Alternatively, there may be no genetic differentiation even when sequence alleles are considered and if homoplasy is present at some or all loci, it does not mask estimates of differentiation. In any case, estimates of population structure made with sequence alleles rather than electromorph alleles are likely to be more accurate because the number of allelic states is higher (Estoup and Angers, 1998).

Sequencing by cloning the large number of samples required for population studies is a time-consuming and expensive method. A technique that has been successfully used to distinguish between different microsatellite sequence alleles in other studies, and which may be suitable for this study, is single-stranded conformation polymorphism (SSCP) (Angers *et al.*, 2000; Sunnocks *et al.*, 2000). SSCP is a fast and inexpensive method to assay sequence variation (Orita *et al.*, 1989). SSCP is particularly sensitive for short DNA fragments, such as those used in this study. For example, single base pair differences are resolved 99% of the time in 100-300 bp fragments (Sunnocks *et al.*, 2000). The SSCP technique would be ideal for assaying allele sequence variation present in these populations.

Microsatellite evolution

Microsatellite evolution is another aspect of this research that warrants further investigation. Microsatellite evolution over time could be further studied by searching

for novel alleles in the ancient samples. If novel alleles were to be confirmed in ancient samples it is important to consider what conclusions might be drawn from them about microsatellite evolution. This type of study is not a direct investigation of microsatellite evolution because the direct mutational steps cannot be determined as in parent-offspring pedigree analyses. However, new mutations may reveal information about broader aspects of evolution. For example, if several new alleles were found in the ancient samples and they were all shorter than the alleles from living penguins, then this would suggest the presence of mutational directionality where the mutations were biased towards an increase in length.

There are two methods that could be used to search for new alleles in the ancient samples. Firstly, a greater number of loci could be investigated. Although microsatellites have a rapid mutation rate relative to other DNA sequences, it may still not result in the occurrence of many mutations over the time period being examined (about 5000 years). Therefore, the examination of a large number of microsatellite loci, i.e. thirty to fifty would increase the chances of finding new mutations. The initial selection of additional loci would be important to consider. Information that has been gained concerning the lifecycles of microsatellites may be a useful tool in predicting which loci may have the highest probability of having unique alleles in the ancient samples. Numerous studies have suggested that loci with a small number of repeats (e.g. the AM3 and FhU2 locus in this study) and interrupted loci have lower mutation rates (Weber and Wong, 1993; Xu *et al.*, 2000). These loci are less likely to have changed over the time period being examined. Thus, future work in this area should concentrate on loci with long stretches of uninterrupted repeats (Bill Amos, pers. comm). A long repeat length for dinucleotide microsatellite in birds is about 17 repeat units (estimated from Figure 4.1 in Amos, 1999b). The majority of loci used in this study were considerably shorter than this length (see Chapter Two, Table 2.1). However, if only long microsatellites were to be examined and novel alleles were found, the inferences they provide for microsatellite evolution could not be extrapolated to all microsatellite loci.

An alternative method by which to further investigate microsatellite evolution using the ancient DNA of Adélie penguins is to examine in greater detail the loci from this study, by investigating allele sequence, rather than electromorph allele size, in a large number

of samples. Again, SSCP would be an ideal technique to sample the sequence variation present in these samples. If a novel SSCP variant is detected in the ancient samples then it can be sequenced. This can be achieved by sequencing a band excised from an SSCP gel (Orti *et al.*, 1997).

5.3 Concluding Remarks

The research presented in this thesis demonstrates that ancient DNA in Adélie penguin subfossil bones is extremely well preserved and therefore provides an ideal situation for the study of the evolution of both populations and nuclear DNA sequences. Several studies have claimed that ancient single-copy sequence, such as nuclear microsatellite DNA, would not be able to be reliably amplified from sufficient numbers of samples to study population genetics (e.g. Ramos *et al.*, 1996; Landweber, 1999). However, this study provides a tantalising glimpse of the possibilities for examining temporal population genetics on a larger scale.

Furthermore, the excellent preservation of these ancient samples means that, despite finding no new electromorph alleles in this study, they still offer an excellent opportunity to study microsatellite evolution over time by either examining the sequence (via SSCP) of alleles and/or examining a greater number of loci.

Appendix A

Table of the common and scientific names of the penguin species used in this study.

Common Name	Scientific Name
Adélie	*Pygoscelis adeliae*
Chinstrap	*P. antarctica*
Gentoo	*P. papua*
King	*Aptenodytes patagonicus*
Emporer	*A. forsteri*
Little Blue/Fairy	*Eudyptula minor*
African	*Spheniscus demersus*
Magellanic	*S. magellanicus*
Galapagos	*S mendiculus*
Humboldt	*S. humboldti*
Royal	*Eudyptes schlegeli*
Macaroni	*E. chrysolophus*
Erect-crested	*E. sclateri*
Fiordland-crested	*E. pachyrhynchus*
Snares-crested	*E. robustus*
Rockhopper	*E. chrysocome*
Yellow-eyed	*Megadyptes antipodes*

Appendix B

Results of genotyping six loci from ancient Adélie subfossil bone samples together with their inferred ages. Grey boxes indicate unsuccessful amplification attempts.

DNA samples		Locus					
	C14 (yrs BP)	TP500	RM6	RM3	AM13	AM3	HrU2
Northern Bird							
PE7	481±57	124/114	174/172			174/174	
PE8	481±57		174/172	225/221			
PE9	481±57	122/112	174/172	223/221	128/127		
PE10	481±57	116/114	172/172	221/221	128/124		
PE27	481±57	122/116	172/172	223/221	128/124	174/174	122/122
PE28	481±57	122/118	172/172	221/221	120/119	174/174	126/124
PE29	481±57	116/114	174/172	225/221	127/119	174/174	124/122
PE31	481±57	124/114	172/172	225/221		174/174	124/124
PE32	481±57	124/114	172/172	225/221	127/127	174/174	124/124
PE33	481±57	124/114	172/172	225/221		174/174	124/122
PE47	481±57	114/112	172/172	227/219	127/123	174/174	
PE48	481±57	114/112	172/172	221/221		174/174	124/124
PE49	440±59	118/114	172/172	221/221		174/174	124/122
PE189	440±59	124/112		225/221	118/118	174/174	124/124
Western Crozier							
PE19, PE132	523±62	114/112	174/174	221/221	127/121	174/174	126/124
PE14	440±59	122/112	172/172	221/221			
PE15	440±59	114/112	174/172	221/221	128/127	174/174	124/124
PE17	498±57		172/172	221/221		175/174	
PE18	440±59	122/112	172/172	221/221	127/127	175/174	124/122
PE20	c. 446	124/112	172/172	225/221	119/119	174/174	124/122
PE21	c. 446	120/114	172/172	221/221	125/124	174/174	124/124
PE22	c. 446	118/116	172/172	225/221	118/118	174/174	124/118
PE23	c. 446	122/114	172/172	225/221		174/174	124/124
PE24	310±60	122/116	174/172	225/221	126/120		124/124
PE25	c. 446	116/114	172/172	225/221			126/124
PE30	c. 446	124/114	172/172	225/221	123/119	175/175	124/124
PE35	c. 446	124/114	172/172	225/221	126/119	175/174	124/124
PE36, PE131	461±57	114/112	172/172	221/221	129/125	175/174	124/124
PE37	498±57	124/114	172/172	225/221		174/174	126/124
PE38	c. 446	120/114	172/172			174/174	126/124
Beaufort Island							
PE41	275±55	118/116	172/172	225/221	130/125	174/174	124/124
PE42	275±55	114/114	174/174	227/221	128/124	174/174	126/124
PE43	275±55	118/114	172/172	221/221	126/126	175/174	124/122
PE44	275±55	124/116	172/172	227/221	128/119	174/174	126/124
PE45	275±55	124/114	172/172	225/221	123/119	174/174	124/124
Inexpressible Island							
PE51, PE104	2040±85	114/112	172/170	221/221	128/118	174/174	126/124
PE55	2328±85	116/116	172/172	221/221	125/118	174/174	126/126
PE56	2328±85	116/116	172/172	221/221	119/118	174/174	124/124
PE63	5706±195	124/112					128/128
PE64	6082±55	122/114	172/172	225/221	125/119	174/174	124/124
PE65	6082±55	124/122	172/172	221/221	128/126	174/174	124/116

PE66	6082±55	124/122	172/172		127/124	174/174	124/116
PE67	6082±55	114/112	172/172				
PE70	6082±55	118/108	172/172	221/221		174/174	124/122
PE71	6082±55	122/122			118/111	175/175	
PE72	6082±55	120/114	172/172		128/124	174/174	
PE50,PE103	2040±85	116/112	172/172	221/221	127/124	174/174	126/124
PE60,PE107	4740±195	122/112	172/172		127/127	174/174	126/124
PE143	6082±55	116/112	172/172	221/221		174/174	124/122
PE144	6082±55						124/124
PE145	6082±55	116/114	172/172		127/113	174/174	124/122
PE146	6082±55	122/112	172/172	221/221	119/119	174/174	124/124
PE147	6082±55						124/124
PE148	6082±55	122/112					124/124
PE150	5706±195	112/112				174/174	
PE151	5706±195	112/112					
PE154	5706±195	122/114				174/174	124/124
Cape Hickey							
PE73	2513±55	120/116	172/172		125/123	174/174	126/124
PE74	2513±55	124/118	172/172				126/124
PE87	3456±55	124/116	174/172	221/221	127/120	174/174	124/124
PE88	3456±55	122/112	174/172	221/221	127/126	175/175	124/116
PE89	3456±55	124/118	172/172	221/221	120/115	174/174	126/124
Cape Ross							
PE75	3514±60						126/124
PE76	3514±60	116/116					124/124
PE77	3514±60	114/112	172/172	225/221		174/174	124/124
PE78	3514±60	122/118	172/172	225/221		174/174	128/126
PE79	3514±60	126/114	174/172	221/221	127/127	175/174	124/124
PE81	3514±60	124/122	174/172				126/124
PE82	3514±60	124/122	174/172	221/221			126/124
PE83	3514±60	112/108	172/172				124/124
PE85	3514±60		172/172	221/221	127/127	174/174	126/126
PE140	3514±60		172/172				
PE156	3514±60			221/221			
PE159	3514±60	116/116	172/172			174/174	130/122
PE160	3514±60	124/116	172/172	225/221	127/119	174/174	124/122
Prior Island (Upper)							
PE90	3888±65	114/112	172/172		125/120	174/174	126/124
Prior Island (Lower)							
PE97	750±75	118/112	172/172	221/221	126/125	174/174	124/124
PE98	1258±80	122/116	172/172		129/126		124/120
PE99	750±75	122/112	172/172	221/221	125/124	174/174	124/120
PE100	949±75	122/114	172/172	221/221	124/112	174/174	124/120
PE101	1057±75	122/118	172/172	225/221	129/125	174/174	126/124
PE102	1085±80	120/112	172/172	221/221	127/126	174/174	126/126
Cape Roberts							
PE112	3370±80	120/120	174/174				
Gondwana Station							
PE113	3370±80	122/114	172/172			174/174	126/122
PE169	3970±80	120/114	174/174				
PE170	6424±80	114/114					
Northern Foothills							
PE115	3800±70	114/112					126/124
PE116	3800±70	122/114	174/172				124/124
PE123	3580±40	114/114	172/172	221/221		175/174	122/122
PE157	4795±40	116/116					124/124
PE162	4525±125	116/116					
PE164	3800±70	116/114	172/172	225/221		174/174	126/124
PE174	3580±40	114/112	174/174	221/221		175/174	124/124
PE182	c. 2700	114/114					124/122
PE190	?	124/112	172/172				124/124
Dunlop Island							
PE118	2230±60						126/126
PE119	2230±60						134/126
PE120	5030±80	120/114	174/172	221/221		174/174	126/124
PE121	5030±80	124/122	172/172	221/221		174/174	126/124
PE163,168	5030±80	124/118	172/172			175/175	
PE165	5030±80	124/112	172/172	221/221		175/174	124/124

Peninsula/Depot Island							
PE122	3020±50	122/116	172/170	225/221			126/126
Cape Hallet							
PE125	c. 500	118/116	172/172			174/174	124/120
PE126	607±80	124/114	174/172			174/174	126/126
PE138	541±45	124/120	172/172				
PE139	c. 600	114/114					
Cape Day							
PE129	c. 3000	118/116					
PE130	3364	116/112					
PE166	c. 3000	118/114		225/225		174/171?	
PE171	c. 3000	120/114	176/174			174/174	
PE172	c. 3000	116/112	172/172	221/221		174/174	
PE173	c. 3000	118/112					124/124
Cape Iriziar							
PE134	3220±65	116/114	174/172	221/221	124/119	175/175	126/124
Edmonson Point							
PE135	1086±40	122/122	172/172	221/221	127/127	174/174	126/126
PE183	1126±40	116/112	172/172	221/221	123/121	174/174	126/124
PE184	1086±40	114/112	172/172	221/221	128/127	174/174	126/124
PE185	1126±40	118/116	174/172			175/174	124/116
PE186	1126±40	118/114	172/172				124/118
PE191	1126±40	124/112					
PE194	1126±40	116/114	172/172			174/174	126/126
PE195	1126±40	116/114	172/172				126/126

Appendix C

Results of genotyping six loci from Adélie penguin samples from the Terra Nova Bay collection sites.

DNA samples	Locus					
Inexpressible Island	TP500	RM6	RM3	AM13	AM3	HrU2
T01	118/114	172/172	221/219	127/127	174/174	126/124
T02	116/114	174/172	225/221	128/119	174/174	126/126
T03	124/116	172/172	225/221	122/114	174/174	126/124
T04	124/114	176/172	221/221	130/119	176/174	124/124
T05	124/116	172/172	221/221	128/118	174/174	126/124
T06	120/118	172/172	221/221	130/124	174/174	124/122
T07	124/114	172/172	223/221	130/124	174/174	126/124
TO8	124/118	174/172	227/221	125/112	174/174	124/122
T09	124/118		221/221	128/123	176/174	124/122
T10	118/114	172/172	225/221	128/128	174/174	124/122
T11	116/114	172/172	221/221	126/126	174/174	124/124
T12	122/122	174/172	221/221	125/120	175/174	124/124
T13	118/112	172/172	225/225	127/127	175/174	126/126
T14	124/122	174/174	221/221	128/111	175/174	124/124
T15	118/114	172/172	225/221	126/126	176/174	124/122
T17	114/112	172/172	221/221	130/119	174/174	126/124
T18	114/114	172/172	221/221	130/126	174/174	126/124
T19	114/114	172/172	225/225	128/126	174/174	126/124
T20	114/114	174/172	225/225	125/124	175/175	124/124
T21	122/114	172/172	221/221	127/124	174/174	126/124
T22	122/114	172/172	221/221	127/124	174/174	126/126
T23	116/114	172/172	221/221	124/118	174/174	126/126
T24	120/112	172/172	221/221	128/126	174/174	126/124
T25	118/114	172/172	221/221	130/123	174/174	124/124
T26	118/112	176/174	221/221	127/127	174/174	124/124
T27	126/114	172/172	221/221	128/127	176/174	124/122
T28	120/116	172/172	221/221	126/125	174/174	126/124
T29	118/112	174/172	225/221	128/125	174/174	124/122
Northern Foothills						
T61	116/116	172/172	221/221	126/124	175/175	124/124
T62	124/114	176/174	221/221	127/124	174/174	124/122
T64	118/116	172/172	221/221	130/127	174/174	126/124
T65	124/118	172/172	221/221	127/125	174/174	126/124
T66	122/118	172/172	221/221	127/127	174/174	126/124
T67	116/114	172/172	225/221	126/126	174/174	124/122
T68	122/112	174/172	221/221	122/109	174/174	
T69	122/114	176/174	221/221	124/119	174/174	128/124
T70	122/116	176/172	225/221	125/122	174/174	126/124
T71	122/120	172/172	221/221	127/127	176/174	126/126
T73	124/124	174/172	225/221	126/120	174/174	124/118
T74	114/112	172/172	225/221	125/120	174/174	124/124
T79	122/120	172/172	221/221	124/111	174/174	126/124
T80	116/116	172/172	221/221	127/124	174/174	124/124
T82	120/114	172/172	221/221		174/174	126/124
T85	116/112	172/172	225/221	128/120	174/174	124/120
T86	114/112	172/172	221/221	129/120	174/174	130/124
T88	122/122	172/172	221/221	128/121	174/174	126/124
T89	120/116	174/174	221/221	124/121	174/174	126/124
T90	124/114	172/172	225/221	129/125	174/174	126/126

T93	122/112	172/172	221/221	126/120	174/174	124/124
T94	116/112	174/172		129/124	174/174	124/124
T95	124/114	172/172	221/221	126/125	176/174	124/120
T96	116/112	174/172	221/221	126/121	175/175	126/124
T99	124/122	176/172	225/221	127/123	174/174	126/126
T101	116/112	172/172	225/221	129/125	175/175	126/124
T102	112/112	172/172	221/221	127/124	174/174	126/124
T106	116/114	172/170	221/221	126/121	174/174	124/124
Edmonson Point						
T167	120/120	172/172	225/221	122/119	174/174	126/124
T168	124/116	174/172	221/221	128/124	174/174	124/124
T169	124/114	172/172	225/221	119/119	174/174	126/124
T170	122/112	172/172	225/221	123/112	174/174	124/122
T171	124/122	174/172	223/221	128/123	174/174	126/124
T172	118/114	172/172	221/221	127/126	175/174	124/122
T173		174/172	221/221	123/119	174/174	124/124
T174	120/112	172/172	225/221	126/126	174/174	122/122
T175	124/122	172/172	221/221	126/126	000/174	126/124
T176	124/114	174/172	221/221	127/118	175/174	124/122
T177	122/120	172/170	225/225	126/120	174/174	126/126
T178	116/112	172/170	221/221	121/114	174/174	124/124
T179	116/114	174/172	221/221	127/127	175/174	126/126
T180		172/172	221/221	125/119	174/174	126/124
T181	122/116	172/172	221/221	118/118	000/174	126/126
T182	122/112	174/172	221/221	126/111	174/174	128/126
T184	122/118	172/172	221/221	124/118	175/174	128/124
T185	124/118	172/172	221/221	128/128	175/174	126/126
T186	114/114	172/172	221/221	128/127	174/174	126/126
T187	114/112	172/172	225/221	125/125	175/174	128/124
T188	122/108	172/172	221/221	127/121	000/174	124/124
T189	122/116	174/172	221/221	129/125	174/174	124/124
T191	116/114	172/172	221/221	126/125	175/174	124/124
T192	114/112	174/172	221/221	123/119	175/174	126/124
T193	116/116	172/172	221/221	124/122	174/174	124/124
T194	124/114	172/172	225/221	123/119	174/174	124/124
T195	124/112	172/172	225/221	130/128	175/174	126/126
T196	116/116	172/172	225/221	127/124	175/175	124/124
T197	118/108	172/172	221/221	128/127	174/174	128/124
T198	116/114	174/172	221/221	126/126	174/174	126/124
T199	124/112	174/172	221/221	124/119	174/174	128/124
T200	126/118	174/172	225/221		174/174	126/126

Appendix D

Animal ethics and Antarctic permits

Approval to sample blood from Antarctic Adélie penguins was granted by the Massey University Ethics Committee (protocol 96/146 and 99/160). Blood sampling from Adélie penguins and entering Specially Protected Areas (SPA) and Sites of Special Scientific Interest (SSSI) was approved by Antarctica New Zealand (permit numbers 96/4, 97/5, 98/7, 99/9).

References

Adcock, G. J., Dennis, E. S., Easteal, S., Huttley, G. A., Jermiin, L. S. and Peacock, W. J. (2001). "Mitochondrial DNA sequences in ancient Australians: implications for modern human origins." *Proceedings of the National Academy of Sciences of the USA* **98**(2): 537-542.

Ainley, D. G., LeReseche, R. E. and Sladen, W. J. (1983). *Breeding Biology of the Adélie Penguin.* Berkeley, University of California Press.

Amos, W. (1999a). A comparative approach to the study of microsatellite evolution. *Microsatellites: Evolution and Applications.* D. B. Goldstein and C. Schlötterer (eds) Oxford, Oxford University Press**:** 66-79.

Amos, W. (1999b). Two problems with the measurement of genetic diversity and genetic distance. *Genetics and the Extinction of Species.* L. F. Landweber and A. P. Dobson (eds) Princeton, Princeton University Press**:** 75-100.

Amos, W., Sawar, S. J., Feakes, R. W. and Rubinsztein, D. C. (1996). "Microsatellites show mutational bias and heterozygote instability." *Nature Genetics* **13**: 390-391.

Angers, B. and Bernatchez, L. (1997). "Complex evolution of a salmonid microsatellite locus and its consequences in inferring allelic divergence from size information." *Molecular Biology and Evolution* **14**(3): 230-238.

Angers, B., Estoup, A. and Jarne, P. (2000). "Microsatellite size homoplasy, SSCP, and population structure: a case study in the freshwater snail *Bulinus truncatus*." *Molecular Biology and Evolution* **17**(12): 1926-1932.

Armour, J. A. L. (1996). Tandemly repeated minisatellites: generating human diversity via recombinational mechanisms. *Human Genome Evolution.* M. Jackson, T, Strachan and G.A. Dover (eds).Oxford, BIOS Scientific Publishers.

Austin, J. J., Smith, A. B. and Thomas, R. H. (1997). "Palaeontology in a molecular world: the search for authentic ancient DNA." *Trends in Ecology and Evolution* **12**(8): 303-306.

Avise, J. C. (1994). *Molecular Markers, Natural History and Evolution.* New York, Chapman and Hall.

Bailey, J. F., Richards, M. B., Macaulay, V. A., Colson, I. B., James, I. T., Bradley, D. G., Hedges, R. M. and Sykes, B. C. (1996). "Ancient DNA suggests a recent expansion of European cattle from a diverse wild progenitor species." *Proceedings of the Royal Society of London B* **263**: 1467-1473.

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

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

Bell, G. I. and Jurka, J. (1997). "The length distribution of perfect dimer repetitive DNA is consistent with its evolution by an unbiased single-step mutation process." *Journal of Molecular Evolution* **44**: 414-421.

Berkman, P. A., Andrews, J. T., Bjorck, S., Colhoun, E. A., Emslie, S. D., Goodwin, I. D., Hall, B. L., Hart, C. P., Hirakawa, K., Igarashi, A., Ingolfsson, O., Lopez-Martinez, J., Lyons, W. B., Mabin, M. C. G., Quilty, P. G., Taviani, M. and Yoshida, Y. (1998). "Circum-Antarctic coastal environmental shifts during the Late Quaternary reflected by emerged marine deposits." *Antarctic Science* **10**(3): 345-362.

Bouzet, J. L., Lewin, H. A. and Paige, K. N. (1998). "The ghost of genetic diversity past: historical DNA analysis of the greater prairie chicken." *The American Naturalist* **152**(1): 1-6.

Bowcock, A. M., Ruiz-Linares, A., Tomfohrde, J., Minch, E., Kidd, J. R. and Cavalli-Sforza, L. L. (1994). "High resolution of human evolutionary trees with polymorphic microsatellites." *Nature* **368**: 455-457.

Brohede, J. and Ellegren, H. (1999). "Microsatellite evolution: polarity of substitutions within repeats and neutrality of flanking sequences." *Proceedings of the Royal Society London B* **266**: 825-833.

Burger, J., Hummel, S., Herrmann, B. and Henke, W. (1999). "DNA preservation: a microsatellite-DNA study on ancient skeletal remains." *Electrophoresis* **20**: 1722-1728.

Chakraborty, R., Kimmel, M., Stivers, D. N., Davison, J. and Deka, R. (1997). "Relative mutation rates at di-, tri- and tetranucleotide microsatellite loci." *Proceedings of the National Academy of Sciences of the USA.* **94**: 1041-1046.

Chambers, G. K. and MacAvoy, E. S. (2000). "Microsatellites: consensus and controversy." *Comparative Biochemistry and Physiology Part B* **126**: 455-476.

Colson, I. B., Bailey, J. F., Vercauteren, M., Sykes, B. C. and Hedges, R. E. M. (1997). "The preservation of ancient DNA and bone diagenesis." *Ancient Biomolecules* **1**(2): 109-117.

Cooper, A. (1997). Studies of avian ancient DNA: from Jurassic Park to modern island extinctions. *Avian Molecular Evolution and Systematics.* D. P. Mindell (ed). San Diego, Academic Press.

Cooper, A., Mourer-Chauvire, C., Chambers, G. K. and Haeseler, A. V. (1992). "Independent origins of New Zealand moas and kiwis." *Proceedings of the National Academy of Sciences of the USA.* **89**: 8741-8744.

Cooper, A. and Poiner, H. (2000). "Ancient DNA : do it right or not at all." *Science* **289**(5482): 1139.

Cooper, A., Rhymer, J., James, H. F., Olsen, S. L., McIntosh, C. E., Sorenson, M. D. and Fleisher, R. C. (1996). "Ancient DNA and island endemics." *Nature (London)* **381**: 484.

Crow, J. F. and Kimura, M. (1970). *An Introduction to Population Genetics Theory.* New York, Harper and Row.

DiRienzo, A., Peterson, A. C., Garza, J. C., Valdes, A. M., Slatkin, M. and Freimer, N. B. (1994). "Mutational processes of simple sequence repeat loci in human populations." *Proceedings of the National Academy of Sciences of the USA* **91**: 3166-3170.

Eichler, E. E., Hammond, H. A., Macpherson, J. N., Ward, P. A. and Nelson, D. L. (1995a). "Population survey of the human FMR1 CGG repeat substructure suggests biased polarity for the loss of AGG interruptions." *Human Molecular Genetics* **4**: 2199-2208.

Eichler, E. E., Kunst, C. B., Lugenbeel, K. A., Ryder, O. A., Davison, D., Warren, S. T. and Nelson, D. L. (1995b). "Evolution of cryptic FMR1 CGG repeat." *Nature Genetics* **11**: 301-307.

Eisen, J. A. (1999). Mechanistic basis for microsatellite instability. *Microsatellites: Evolution and Application.* D. B. Goldstein and C. Schlötterer (eds). Oxford, Oxford University Press**:** 34-48.

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

Ellegren, H. (2000). "Heterogeneous mutation processes in human microsatellite DNA sequences." *Nature Genetics* **24**: 400-402.

Estoup, A., Garnery, L., Solignoc, M. and Cornuet, J. M. (1995a). "Microsatellite variation in honey bee (*Apis mellifera* L.) populations : hierachical genetic strructure and test of the infinite allele and stepwise mutation models." *Genetics* **140**: 679-695.

Estoup, A., Taillez, A., Cornuet, J. M. and Solignac, M. (1995b). "Size homoplasy and mutational processes of interrupted microsatellites in two bee species, *Apis mellifera* and *Bombus terrestris* (Apidae)." *Molecular Biology and Evolution* **12**(6): 1074-1084.

Estoup, A. and Angers, B. (1998). Microsatellites and minisatellites for molecular ecology: theoretical and empirical considerations. *Advances in Molecular Ecology*. G. R. Carvalho (ed), Amsterdam, IOS Press**:** 55-86.

Estoup, A. and Cornuet, J. M. (1999). Microsatellite evolution: inferences from population data. *Microsatellites: Evolution and Application*. D. B. Goldstein and C. Schlötterer (eds). Oxford, Oxford University Press**:** 49-65.

Faerman, M., Filon, D., Kahlia, G., Greenblatt, C. L., Smith, P. and Oppenheim, A. (1995). "Sex identification of archaeological human remains based on amplification of X and Y amelogenin alleles." *Gene* **167**: 327-332.

Falush, D. and Iwasa, Y. (1999). "Size-dependent mutability and microsatellite constraints." *Molecular Biology and Evolution* **16**(7): 960-966.

Foucault, F., Praz, F., Jaulin, C. and Amor-Gueret, M. (1996). "Experimental limits of PCR analysis of $(CA)_n$ repeat alterations." *Trends in Genetics* **12**(11): 450-452.

Fraser, W. R. and Trivelpiece, W. Z. (1996). "Factors controlling the distribution of seabirds: winter-summer heterogeneity in the distribution of Adélie penguin populations." *Antarctic Research Series* **70**: 257-272.

Futuyma, D. (1998). *Evolutionary Biology*. Massachussetts, Sinauer Associates, Inc.

Gagneux, P., Boesch, C. and Woodruff, D. S. (1997). "Microsatellite scoring errors associated with noninvasive genotyping based on nuclear DNA amplified from shed hair." *Molecular Ecology* **6**: 861-868.

Garza, J. C. and Freimer, N. B. (1996). "Homoplasy for size at microsatellite loci in humans and chimpanzees." *Genome Research* **6**: 211-217.

Garza, J. C., Slatkin, M. and Freimer, N. B. (1995). "Microsatellite allele frequencies in humans and chimpanzees, with implications for constraints on allele size." *Molecular Biology and Evolution* **12**(4): 594-603.

Gastier, J. M., Pulido, J. C., Sunden, S., Brody, T., Buetow, K. H., Murray, J. C. et al., (1995). "Survey of trinucleotide repeats in the human genome: assessment of their utility as genetic markers." *Human Molecular Genetics* **4**: 1829-36.

Goldstein, D. B., Linares, A. R., Cavelli-Sforza, L. L. and Feldman, M. W. (1995). "An evaluation of genetic distances for use with microsatellite loci." *Genetics* **139**: 463-471.

Goldstein, D. B. and Pollock, D. D. (1997). "Launching microsatellites: a review of mutation processes and methods of phylogenetic inference." *Journal of Heredity* **88**: 335-342.

Goodman, S. J. (1997). "RSTCalc: a collection of computer programs for calculating estimates of genetic differentiation from microsatellite data and determining their significance." *Molecular Ecology* **6**: 881-885.

Goossens, B., Waits, L. P. and Taberlet, P. (1998). "Plucked hair samples as a source of DNA: reliability of dinucleotide microsatellite genotyping." *Molecular Ecology* **7**: 1237-1241.

Goudet, J. (1995). "FSTAT (vers. 1.2): a computer program to calculate F-statistics." *Journal of Heredity* **86**: 485-486.

Goudet, J. (1999). FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.8). Updated from Goudet (1995)

Grimaldi, M. C. and Crouau-Roy, B. (1997). "Microsatellite allelic homoplasy due to variable flanking sequences." *Journal of Molecular Evolution* **44**(336-340).

Greenwood, A. D., Capelli, C., Possnert, G. and Pääbo, S. (1999). "Nuclear DNA sequences from late Pleistocene megafauna." *Molecular Biology and Evolution* **16**(11): 1466-1473.

Guo, S. W. and Thomson, E. A. (1992). "Performing the exact test of Hardy-Weinberg proportion for multiple alleles." *Biometrics* **48**: 361-372.

Hagelberg, E., Bell, L. S., Allen, T., Boyle, A., Jones, S. J. and Clegg, J. B. (1991). "Analysis of ancient bone DNA: techniques and application." *Philosophical Transactions of the Royal Society of London B* **333**: 339-407.

Hagelberg, E., Quevedo, S., Turbon, D. and Clegg, J. (1994). "DNA from ancient mammoth bone." *Nature* **370**: 333-4.

Hancock, J. M. (1999). Microsatellites and other simple sequences: genomic context and mutational mechanisms. *Microsatellites: Evolution and Applications*. D. B. Goldstein and C. Schlötterer (eds). Oxford, Oxford University Press: 1-9.

Hauswirth, W. W., Dickel, C. D., Rowold, D. J. and Hauswirth, M. A. (1994). "Inter- and intrapopulational studies of ancient humans." *Experimentia* **50**: 585-591.

Hedges, R. E. M. and Millard, A. R. (1995). "Measurements and relationships of diagenetic alteration of bone from three archaelogical sites." *Journal of Archaelogical Science* **22**: 201-9.

Heine, J. C. and Speir, T. W. (1989). "Orthinographic soils of the Cape Bird Adélie penguin rookeries, Antarctica." *Polar Biology* **10**: 89-99.

Henderson, S. T. and Petes, T. D. (1992). "Instability of simple sequence DNA in *Saccharomyces cerevisae*." *Molecular and Cellular Biology* **12**: 2749-2757.

Higuchi, R., Bowman, B., Freiberger, M., Ryder, O. and Wilson, A. (1984). "DNA sequence from the quagga, an extinct member of the horse family." *Nature* **312**: 282-284.

Hofreitner, M., Poiner, H. N., Spaulding, W. G., Bauer, K., Martin, P. S., Possnert, G. and Pääbo, S. (2000). "A molecular analysis of ground sloth diet through the last glaciation." *Molecular Ecology* **9**: 1975-1984.

Höss, M., Jaruga, P., Zastawny, T. H., Dizdaroghu, M. and Pääbo, S. (1996). "DNA damage and DNA sequence retrieval from ancient tissues." *Nucleic Acid Research* **24**(7): 1304-1307.

Höss, M. and Pääbo, S. (1993). "DNA extraction from Pleistocene bones by a silica-based purification method." *Nucleic Acid Research* **21**(16): 3913-3914.

Hummel, S., Burger, J., Rameckers, J., Lassens, C., Schmerer, W. and Herrmann, B. (1996). "AmpliTaq Gold improves short tandem repeat amplifications of highly degraded DNA." *Amplifications* **1**(14): 5-6.

Hummel, S. and Herrmann, B. (1994). General aspects of sample preparation. *Ancient DNA*. B. Herrmann and S. Hummel (eds), New York, Springer-Verlag.

Ingólfsson, O., Hjort, C., Berkaman, P. A., Björck, S., Colhoun, E., Goodwin, I. D., Hall, I. D., Hall, B., Hirakawa, K., Melles, M. and Prentice, M. L. (1998). "Antarctic Glacial History since the Last Glacial Maximum: an overview of the record on land." *Antarctic Science* **10**(3): 326-344.

Jarne, P. and Lagoda, P. J. L. (1996). "Microsatellites, from molecules to populations and back." *Trends in Ecology and Evolution* **11**(10): 424-429.

Jin, L., Macaubas, C., Hallmayer, J., Kimura, A. and Mignot, E. (1996). "Mutation rate varies among alleles at a microsatellite locus: phylogenetic evidence." *Proceedings of the National Academy of Sciences of the USA* **93**: 15285-15288.

Kimmel, M. and Chakraborty, R. (1996). "Measures of variation at DNA repeat loci under a general stepwise mutation model." *Theoretical Population Biology* **50**: 345-367.

Kimmel, M., Chakraborty, R., Stivers, D. N. and Deka, R. (1996). "Dynamics of repeat polymorphisms under a forward-backward mutation model: within- and between- population variability at microsatellite loci." *Genetics* **143**: 549-555.

Kimura, M. and Crow, J. F. (1964). "The number of alleles that can be maintained in a finite population." *Genetics* **49**: 725-738.

Kimura, M. and Ohta, T. (1978). "Stepwise mutation model and the distribution of allelic frequencies in a finite population." *Proceedings of the National Academy of Sciences of the USA* **75**: 2868-72.

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

Krugylak, S., Durrett, R. T., Schug, M. D. and Aquadro, C. F. (1998). "Equilibrium distribution of microsatellite repeat length resulting from a balance between slippage events and point mutations." *Proceedings of the National Academy of Sciences of the USA* **95**: 10774-10778.

Kunkel, T. A. (1992). "Biological asymmetries and the fidelity of eukaryotic DNA replication." *BioEssays* **14**(5): 303-308.

Kunst, C. B. and Warren, S. T. (1994). "Cryptic and polar variation of the fragile X repeat could result in predisposing normal alleles." *Cell* **77**: 853-61.

Landweber, L. F. (1999). Something old for something new: the future of ancient DNA in conservation biology. *Genetics and the Extinction of Species*. L. F.

Landweber and A. P. Dobson (eds), Princeton, Princeton University Press: 163-186.

Lee, J. S., Hanford, M. G., Gnova, J. L. and Farber, R. A. (1999). "Relative stabilites of dinucleotide and tetranucleotide repeats in cultured mammalian cells." *Human Molecular Genetics* **8**: 2567-2572.

Leonard, J. A., Wayne, R. K. and Cooper, A. (2000). "Population genetics of Ice Age brown bears." *Proceedings of the National Academy of Sciences of the USA* **97**(4): 1651-1654.

Levinson, G. and Gutman, G. A. (1987). "Slipped-strand mispairing: a major mechanism for DNA sequence evolution." *Molecular Biology and Evolution* **4**: 203-221.

Licht, K. J., Jennings, A. E., Andrews, J. T. and Williams, K. M. (1996). "Chronology of late Wisconsin ice retreat from the western Ross Sea, Antarctica." *Geology* **24**(3): 223-226.

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

Lindahl, T. and Nyberg, B. (1972). "Rate of depurination of native deoxyribonucleic acid." *Biochemistry* **11**.

Makova, K. D., Nekrutenko, A. and Baker, R. J. (2000). "Evolution of microsatellite alleles in four species of mice (Genus *Apodemus*)." *Journal of Molecular Evolution* **51**: 166-172.

McDonald, D. B. and Potts, W. K. (1997). DNA microsatellites as genetic markers at several scales. *Avian Molecular Evolution and Systematics*. D. P. Mindell (ed). San Diego, Academic Press.

Messier, W., Li, S.-H. and Stewart, C. (1996). "The birth of microsatellites." *Nature* **381**: 483.

Michalakis, Y. and Excoffier, L. (1996). "A generic estimation of population subdivision using distances between alleles with special reference for microsatellite loci." *Genetics* **142**(1061-1064).

Morral, N., Nunes, V., Casals, T., Chillon, M., Gimenez, J., Bertranpetit, J. and Estivill, X. (1993). "Microsatellite haplotypes for cystic fibrosis: mutation frameworks and evolutionary tracers." *Human Molecular Genetics* **2**(7): 1015-1022.

Mundy, N. I., Winchell, C. S., Burr, T. and Woodruff, D. S. (1997). "Microsatellite variation and microevolution in the critically endangered San Clemente Island loggerhead shrike (*Lanius ludovicianus mearnsi*)." *Proceedings of the Royal Society of London Series B* **264**: 869-875.

Nauta, M. J. and Weissing, F. J. (1996). "Constraints on allele size at microsatellite loci: implications for genetic differentiation." *Genetics* **143**: 1021-1032.

Navidi, W., Arnheim, N. and Waterman, M. S. (1992). "A multiple-tubes approach for accurate genotyping of very small DNA samples by using PCR: statistical considerations." *American Journal of Human Genetics* **50**: 347-359.

Nielson, E. E., Hansen, M. M. and Loeschcke, V. (1997). "Analysis of microsatellite DNA from old scale samples of Atlantic salmon *Salmo salar*: a comparison of genetic composition over 60 years." *Molecular Ecology* **6**: 487-492.

Nielson, E. E., Hansen, M. M. and Loeschcke, V. (1999). "Genetic variation in time and space: microsatellite analysis of extinct and extant populations of Atlantic salmon." *Evolution* **53**(1): 261-268.

Noor, M. A. F., Kliman, R. M. and Machado, C. A. (2001). "Evolutionary history of microsatellites in the obscura group of *Drosophila*." *Molecular Biology and Evolution* **18**(4): 551-556.

Orita, M., Iwahana, H., Kamazawa, H., Hayashi, K. and Sekiya, T. (1989). "Detection of polymorphism of human DNA by gel electrophoresis as single-strand conformation polymorphisms." *Proceedings of the National Academy of Sciences of the USA.* **86**: 2766-2770.

O'Rourke, D. H., Carlyle, S. W. and Parr, R. L. (1996). "Ancient DNA: methods, progress, and perspectives." *American Journal of Human Biology* **8**: 557-571.

Orti, G., Pearse, D. E. and Avise, J. C. (1997). "Phylogenetic assessment of length variation at a microsatellite locus." *Proceedings of the National Academy of Sciences of the USA* **94**: 10745-10749.

Pääbo, S. (1985). "Molecular cloning of ancient Eygptian mummy DNA." *Nature* **314**: 644-645.

Pääbo, S. (1989). "Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification." *Proceedings of the National Academy of Science of the USA* **86**: 1939-1943.

Paetkau, D., Calvert, W., Stirling, I. and Strobeck, C. (1995). "The molecular basis and evolutionary history of a microsatellite null allele in bears." *Molecular Ecology* **4**: 347-354.

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

Petri, A. and Baroni, C. (1997). "Penguin, a macintosh application for entry and presentation of radiocarbon-dated samples." *Radiocarbon* **39**(1): 61-65.

Poiner, H. N., Hoss, M., Bada, J. L. and Paabo, S. (1996). "Amino acid racemization and the preservation of ancient DNA." *Science* **272**: 864-6.

Poiner, H. N., Hofreitner, M., Spaulding, W. G., Martin, P. S., Stankiewicz, B. A., Bland, H., Evershad, R. P., Possnert, G. and Pääbo, S. (1998). "Molecular coproscopy: dung and diet of the extinct ground sloth *Nothrotheriops shastensis*." *Science* **281**: 402-406.

Poiner, H. N. and Stankiewicz, B. A. (1999). "Protein preservation and DNA retrieval from ancient tissues." *Proceedings of the National Academy of Sciences of the USA* **96**: 8426-8431.

Pope, B. and Kent, H. M. (1996). "High efficiency 5 min transformation of *Escherichia coli*." *Nucleic Acid Research* **24**(3): 536-537.

Primmer, C. R. and Ellegren, H. (1998). "Patterns of molecular evolution in avian microsatellite mutations." *Molecular Biology and Evolution* **15**(8): 997-1008.

Primmer, C. R., Moller, A. P. and Ellegren, H. (1995). "Resolving genetic relationships with microsatellite markers: a parentage testing system for the swallow *Hirundo rustica*." *Molecular Ecology* **4**: 493-498.

Primmer, C. R., Ellegren, H., Saino, N. and Moller, A. P. (1996a). "Directional evolution in germline microsatellite mutations." *Nature Genetics* **13**: 391-393.

Primmer, C. R., Moller, A. P. and Ellegren, H. (1996b). "A wide-range survey of cross-species microsatellite amplification in birds." *Molecular Ecology* **5**: 365-378.

Pupko, T. and Graur, D. (1999). "Evolution of microsatellites in the yeast *Saccharomyces cerevisiae*: role of length and number of repeated units." *Journal of Molecular Evolution* **48**: 313-316.

Rameckers, J., Hummel, S. and Herrmann, B. (1997). "How many cycles does a PCR need? Determinations of cycle numbers depending on the number of targets and the reaction efficiency factor." *Naturwissenschaften* **84**: 259-262.

Ramos, M. D., Lalueza, C., Girbau, E., Perez-Perez, A., Quevedo, S., Turbon, D. and Estivill, X. (1996). "Amplifying dinucleotide microsatellite loci from bone and tooth samples of up to 5000 years of age: more inconsistency than usefulness." *Human Genetics* **96**: 205-212.

Raymond, M. and Rousset, F. (1995). "GENEPOP (version 1.2): population genetics software for exact tests and ecumenism." *Journal of Heredity* **86**(248-249).

Rice, W. R. (1989). "Analyzing tables of statistical tests." *Evolution* **43**(1): 223-225.

Ritchie, P. (2001). The Evolution of the Mitochondrial DNA Control Region in the Adélie Penguins of Antarctica., Massey University.*Unpublished PhD thesis.*

Roeder, A. D., Marshall, R. K., Mitchelson, A. J., Visagathilagar, T., McPartlan, H. C., Murray, N. D., Kerry, K. R., Robinson, N. A. and Lambert, D. M. "Gene flow on the ice: genetic differentiation among Adélie penguin colonies around Antarctica." *Molecular Ecology*, in press.

Roeder, A. D., Ritchie, P. A. and Lambert, D. M. "New DNA markers for penguins." *submitted for publication.*

Rose, O. and Falush, D. (1998). "A threshold size for microsatellites." *Molecular Biology and Evolution* **15**(5): 613-615.

Ross, M. H., Romrell, L. J. and Kaye, G. I. (1995). *Histology: a text and atlas.* Maryland, Williams and Wilkins

Rousset, F. (1996). "Equilibrium values of measures of population subdivision for stepwise mutation processes." *Genetics* **142**: 1357-1362.

Rubinsztein, D., Amos, B. and Cooper, G. (1999). "Microsatellite and trinucleotide-repeat evolution: evidence for mutational bias and different rates of evolution in different lineages." *Philosophical Transactions of the Royal Society London B* **354**: 1095-1099.

Schlötterer, C. (1998). "Are microsatellites really simple sequences?" *Current Biology* **8**: R132-R134.

Schlötterer, C. (2000). "Evolutionary dynamics of microsatellite DNA." *Chromosoma* **109**(6): 365-371.

Schlötterer, C., Ritter, R., Harr, B. and Brem, G. (1998). "High mutation rate of a long microsatellite allele in *Drosophila melanogaster* provides evidence for allele-specific mutation rates." *Molecular Biology and Evolution* **15**(10): 1269-1274.

Schmerer, W. M., Hummel, S. and Herrmann, B. (1999). "Optimized DNA extraction to improve reproducibility of short tandem repeat genotyping with highly degraded DNA as target." *Electrophoresis* **20**: 1712-1716.

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

Sheffield, V. C., Weber, J. L., Buetox, K. H., Murray, J. C., Even, D. A., Wiles, K. et al., (1995). "A collection of tri- and tetranucleotide repeat markers used to generate high quality, high resolution human genome-wide linkage maps." *Human Molecular Genetics* **4**: 1837-44.

Shriver, M. D., Jin, L., Chakraborty, R. and Boerwinkle, E. (1993). "VNTR allele frequency distributions under the stepwise mutation model: a computer simulation approach." *Genetics* **134**: 983.

Slatkin, M. (1995). "A measure of population subdivision based on microsatellite allele frequencies." *Genetics* **139**: 457-462.

Smith, G. P. (1976). "Evolution of repeated DNA sequences by unequal crossover." *Science* **191**: 528-35.

Spellerberg, I. F. (1970). "Abandoned penguin rookeries near Cape Royds, Ross Island, Antarctica and ^{14}C dating of penguin remains." *New Zealand Journal of Science* **13**(3): 380-385.

SPSS (1997). SPSS version 8.0.0, SPSS Inc.

Strand, M., Earley, M., Grouse, G. F. and Petes, T. D. (1993). "Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair." *Nature* **365**: 274-6.

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

Sunnocks, P. (2000). "Efficient genetic markers for population biology." *Trends in Ecology and Genetics* **15**(5): 199-203.

Sunnocks, P., Wilson, A. C. C., Beheregaray, L. B., Zenger, K., French, J. and Taylor, A. C. (2000). "SSCP is not so difficult: the application and utility of single-

stranded conformation polymorphism in evolutionary biology and molecular ecology." *Molecular Ecology* **9**: 1699-1710.

Suzuki, A., Maruno, A., Tahira, T. and Hayashi, K. (2001). "Polar alteration of short tandem repeats (STRs) in mammalian cells." *Mutation Research* **474**: 159-168.

Szostak, J. W., Orr-Weaver, T. L. and Rothstein, R. J. (1983). "The double-strand-break model for DNA recombination." *Cell* **33**: 25-35.

Taberlet, P., Griffen, S., Goossens, B., Questiau, S., Manceau, V., Escaravage, N., Waits, P. and Bouvet, J. (1996). "Reliable genotyping of samples with very low DNA quantities using PCR." *Nucleic Acids Research* **24**(16): 3189-3194.

Taberlet, P. and Luikart, G. (1999). "Non-invasive genetic sampling and individual identification." *Biological Journal of the Linnean Society* **68**: 41-55.

Takezaki, N. and Nei, M. (1996). "Genetic distances and reconstruction of phylogenetic trees from microsatellite data." *Genetics* **144**: 389-399.

Taylor, J. S., Durkin, J. M. H. and Breden, F. (1999a). "Death of a microsatellite: a phylogenetic perspective on microsatellite interruptions." *Molecular Biology and Evolution* **16**(4): 567-572.

Taylor, J. S., Sanny, J. S. P. and Breden, F. (1999b). "Microsatellite allele size homoplasy in the Guppy (*Poecilia reticulata*)." *Journal of Molecular Evolution* **48**(245-247).

Tessier, N. and Bernatchez, L. (1999). "Stability of population structure and genetic diversity across generations assessed by microsatellites among sympatric populations of landlocked Atlantic salmon (*Salmo salar* L.)." *Molecular Ecology* **8**: 169-179.

Timsit, Y. (1999). "DNA structure and polymerase fidelity." *Journal of Molecular Biology* **293**: 835-853.

Timsit, Y. and Moras, D. (1996). "Cruciform structures and functions." *Quarterly Reviews of Biophysics* **29**: 279-307.

Tuross, N. (1994). "The biochemistry of ancient DNA in bone." *Experimentia* **50**: 530-535.

Twerdi, C. D., Boyer, J. C. and Farber, R. A. (1999). "Relative rates of insertion and deletion mutations in a microsatellite sequence in cultured cells." *Proceedings of the National Academy of Sciences of the USA* **96**(2875-2879).

Valsecchi, E., Palsbøll, P., Hale, P., Glockner-Ferrari, D., Ferrari, M., Clapham, P., Larsen, F., Mattila, D., Sears, R., Siguronsson, J., Brown, M., Cockeron, P. and Amos, B. (1997). "Microsatellite genetic distances between oceanic populations of the humpback whale (*Megaptera noveagliae*)." *Molecular Biology and Evolution* **14**(4): 355-362.

van Oppen, M. J. H., Rico, C., Turner, G. F. and Hewitt, G. M. (2000). "Extensive homoplasy, nonstepwise mutations, and shared ancestral polymorphism at a complex microsatellite locus in Lake Malawi cichilids." *Molecular Biology and Evolution* **17**(4): 489-498.

Viard, F., Franck, P., Dubois, M. P., Estoup, A. and Jarne, P. (1998). "Variation of size homoplasy across electromorphs, loci and populations in three invertebrates." *Journal of Molecular Evolution* **47**: 42-51.

Villablanca, X. (1994). Spatial and temporal aspects of populations revealed by mitochondrial DNA. *Ancient DNA*. B. Herrmann and S. Hummel (eds) New York, Springer-Verlag.

Wahls, W. P., Wallace, L. J. and Moore, P. D. (1990). "The Z-DNA motif $d(TG)_{30}$ promotes reception of information during gene conversion events while stimulating homologous recombination in human cells in culture." *Molecular and Cellular Biology* **10**: 785-93.

Wang, L., Oota, H., Saitou, N., Jin, F., Matsushita, T. and Ueda, S. (2000). "Genetic structure of a 2500-year-old human population in China and its spatiotemporal changes." *Molecular Biology and Evolution* **17**(9): 1396-1400.

Wasser, P. and Strobeck, C. (1998). "Genetic signatures of interpopulation dispersal." *Trends in Ecology and Evolution* **13**: 43-44.

Wayne, R. K., Leonard, J. A. and Cooper, A. (1999). "Full of sound and fury: the recent history of ancient DNA." *Annual Review of Ecology and Systematics* **30**: 457-477.

Weber, J. and Wong, C. (1993). "Mutation of human short repeats." *Genetics* **2**: 1123-1128.

Weir, B. S. and Cockerham, C. C. (1984). "Estimating F-statistics for the analysis of population structure." *Evolution* **38**: 1358-1370.

Wierdl, M., Dominska, M. and Petes, T. D. (1997). "Microsatellite instability in yeast: dependence on the length of the microsatellite." *Genetics* **146**: 769-779.

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

Wright, S. (1951). "The genetic structure of populations." *Annals of Eugenics* **15**: 323-354.

Xu, X., Peng, M., Fang, Z. and Xu, X. (2000). "The direction of microsatellite mutations is dependent upon allele length." *Nature Genetics* **24**: 386-399.

Yang, H. (1997). "Ancient DNA from Pleistocene fossils: preservation, recovery, and utility of ancient genetic information for Quaternary research." *Quaternary Science Reviews* **16**: 1145-1161.

Zhu, Y., Strassman, J. E. and Queller, D. C. (2000). "Insertions, substitutions, and the origin of microsatellites." *Genetical Research* **76**: 227-236.

Zierdt, H., Hummel, S. and Hermann, B. (1996). "Amplification of human short tandem repeats from medieval teeth and bone samples." *Human Biology.* **68**(2): 185-199.