Evolutionary genetics of neutral and immune related loci in North American thinhorn sheep (*Ovis dalli* spp)

By

Kirsty Worley

Department of Animal and Plant Sciences

University of Sheffield

A thesis submitted for the degree of Doctor of Philosophy

September 2005

ABSTRACT

The regulation of genetic variation in natural populations is via both neutral and selective processes. Signatures of the neutral processes of drift and gene flow can be found within non-coding regions of the genome, while natural selection acts on variation within coding genes that confer changes in fitness. The affects of neutral and selective processes are examined in thinhorn sheep (*Ovis dalli* spp), a rare example of a North American large mammal that occupies most of its native range. There are currently two recognised thinhorn subspecies (*O. d. dalli* and *O. d. stonei*), the validity of which remains uncertain.

Microsatellites reveal significant genetic structure throughout the thinhorn species range (F_{ST}=0.16). At least eight regional genetic groups can be defined, the limits of which are delineated by mountain range topology. Strong isolation-by-distance is evident (r=0.75, P<0.001), suggesting limited dispersal and philopatry within the species. Analysis of mitochondrial DNA reveals that sheep survived Pleistocene glaciations in four refugia. In addition to the well known refugia of eastern Beringia and areas south of the Laurentide and Cordilleran ice sheets I show evidence of two smaller refugia, providing support for the presence of the disputed 'ice-free corridor' through eastern British Columbia. MtDNA also reveals a genetic signal confirming the past hybridization of thinhorn and bighorn sheep.

Patterns of variation in linked microsatellite loci show no evidence of natural selection acting on three genes involved in the thinhorn immune region. Although Watterson tests suggest balancing selection is acting in all genes, evidence for selection is confounded by population structure. Concordantly, the translated coding sequences of thinhorn interferon gamma, natural resistance associated macrophage protein and prion protein have low genetic diversity. In contrast, the major histocompatibility complex locus DRB3 shows significant evidence of overdominance through both an excess of nonsynonymous substitution and trans-species polymorphism.

ACKNOWLEDGEMENTS

This project was possible due to the funding provided by the University of Sheffield, and the governments of the Yukon and Northwest Territories, Canada. I would firstly like to thank my supervisor Dave Coltman for allowing me to take on this study and providing support both in Sheffield and Edmonton during its completion. Additional thanks go to Jon Slate for support in Sheffield in the absence of Dave. Thanks also to Alasdair Veitch and Jean Carey both for providing samples and showing me that sheep are more than bits of horn in envelopes. Additional samples were kindly supplied by Helen Schwanje and Steve Arthur.

My time in Sheffield wouldn't have been the same without everyone in TAB lab and beyond who are too numerous to mention here (I want to try and keep this short!). Friday nights will never be the same again! However, special thanks must go to Debs Dawson, Melissa Gunn and Iain Barr, long time office mates of B54 for sharing coffee, cake, chocolate and conversation throughout this project. Lab support came from Andy Krupa, TAB lab certainly wouldn't be the same without him. Thanks also to Jake Gratten for providing several tubes of colourless liquid that were invaluable for this venture. Kelly Green was a fantastic summer helper, whose lab abilities were matched only by her Thursday night dancing skills! Finally I must make a special mention to Matt Hale, the finest housemate and friend to be had in Sheffield, whose culinary abilities are matched by only the finest Indian restaurants and with whom I spent many nights discussing this study and more over several bottles of gin. During my nine months at the University of Alberta I was made welcome by Curt Strobeck and members of his lab.

Finally, a very special thanks to my family for supporting me throughout my studies and especially to Paul, who in addition to reading parts of this thesis out of choice (and therefore expanding his knowledge on all things sheep related), tolerated my absence in Canada for so long because of it.

CONTENTS

PREFACE

The data included in this thesis are published or in submission at peer review journals as three manuscripts, details are given below.

Worley K, Strobeck C, Arthur S, Carey, J, Schwantje, H, Veitch, A, Coltman, DW. (2004) Population genetic structure of North American thinhorn sheep (*Ovis dalli*). *Molecular Ecology* **13**, 2545-2556.

The genotyping of microsatellites, analysis of data and composition of this manuscript were completed solely by me.

Loehr J, **Worley K**, Grapputo A, Carey J, Veitch A, Coltman DW. Evidence for cryptic glacial refugia from North American mountain sheep mitochondrial DNA (2005) *In press at; Journal of Evolutionary Biology*

This manuscript was written by John Loehr as part of a PhD at the University of Jyväskylä, Finland. I contributed to the study by choosing appropriate samples, carrying out the DNA sequencing, editing and alignment and writing associated methods.

Worley K, Carey J, Veitch A, Coltman DW. Detecting the signature of selection on immune genes in highly structured populations of wild sheep (*Ovis dalli*). *In press at; Molecular Ecology*

The genotyping of microsatellites, DNA sequencing, data analysis and composition of this manuscript were completed solely by me.

CHAPTER 1. INTRODUCTION

1.1. Mechanisms generating genetic variation

All phenotypic and functional variation between populations and species result from evolutionary processes. In addition to obvious differences between organisms there are many more silent differences in their genetic make up. Small differences between organisms can accumulate over time to produce distinct species or populations adapted to diverse environments. Understanding processes that are responsible for the production and maintenance of genetic variation is therefore a central subject in biology.

All genetic differences between individuals are produced by the processes of mutation and recombination. Both processes occur spontaneously during DNA replication and result in sequences which differ from those of the parental type. Mutations occur on several levels, deletions, insertions, inversions or substitutions of chromosomes, genes and nucleotides are all possible. Recombination during meiosis produces genetic variation by creating new combinations of pre-existing parental genes in offspring. Mutation and recombination occurring within somatic cells is not heritable and is therefore of no evolutionary significance, while events occurring during meiosis are passed on to offspring. All mutations arise randomly and most have no effect on the fitness of organisms. The fitness consequences of mutation and recombination are dependent on the location of change within the genome. Most mutations occur in non-coding DNA and therefore have no affect on fitness. Moreover, the majority of mutational events in gene coding regions are also selectively neutral (Kimura, 1968). Mutation rates are not equal across the genome, some chromosomes and gene regions are presumed to have higher mutation rates than others, and there is also an effect of neighbouring nucleotides on the rate of mutation (Ellegren *et al.*, 2003). This can lead to heterogeneity of genetic variation across the genome (Sunyaev *et al.*, 2003). Once genetic polymorphisms are present

within populations their frequencies are regulated by both random neutral processes and adaptive forces.

1.2. Consequences of neutral processes for genetic variation

Several neutral processes lead to either an increase or decrease in levels of variation in polymorphic genes. In finite populations the stochastic process of random genetic drift acts to alter allele frequencies in variable gene regions (Kimura, 1955). In small populations allele frequencies of offspring differ from those present in parents by chance alone. Over many generations the resultant fluctuations in gene frequencies can result in the gradual drift of allele frequencies to fixation or complete loss. The effects of genetic drift are especially apparent in bottlenecked populations due to the reduced population size. Random genetic drift acts on neutral or slightly deleterious mutations, in the latter this can lead to a mutation load in small populations, increasing the probability of extirpation (Puurtinen *et al.*, 2004; Kimura *et al.*, 1963). The decrease in genetic variation expected by genetic drift is opposed by the production of new variation through mutation.

The movement of migrants between populations also opposes the effects of drift in small populations. As random drift occurs by chance, levels of neutral genetic variation vary between populations. Movement of migrants into new populations can introduce new alleles to the recipient population, increasing levels of genetic variation. The opposite occurs when small numbers of individuals colonise new regions. Founder populations have low levels of genetic variation as only a subset of the original genetic variation is represented in the new location.

The effects of genetic drift and gene flow differ between populations depending on many factors including size and degree of spatial isolation. Patterns of neutral genetic diversity within and between populations can therefore reveal much about population demography and structure. As most genomic DNA can be considered selectively

neutral there are many markers available for studies on these areas. Among the most useful are non-coding repeats such as microsatellites.

1.2.1. Microsatellites

Regions of non-coding DNA account for a significant proportion of total DNA in mammalian species. Although some regions of non-coding DNA appear to be involved in the regulation and transcription of proximal genes (Vandendries *et al.*, 1996; Nikolajczyk *et al.*, 1996), the majority of non-coding DNA has no known functionality. A property of non-coding DNA is that it often consists of tandem repeat units, one type of which, microsatellites, comprise between two and five base pair repeats and are of particular use in studies of population demography and structure.

Microsatellites are found scattered throughout the genomes of all eukaryotic organisms, including vertebrates, insects and plants. Mutation rates at these loci are among the highest of all loci at approximately 10^{-2} to 10^{-5}, allowing the accumulation of a large number of alleles (Henderson & Petes, 1992). Mutation at microsatellite loci occurs by the addition or loss of repeat units, most frequently by the process of slippage during replication (Kimura & Ohta, 1978). There is a bias for the addition of repeats (Primmer *et al.*, 1996; Yamada *et al.*, 2002) limited by an increase in mutation frequency in long microsatellites (Sibly *et al.*, 2001; Yamada *et al.*, 2002). Microsatellite evolution is therefore cyclical (Falush & Iwasa, 1999; Zhu *et al.*, 2000). The mutation rate of microsatellites varies with nucleotide repeat type (Kruglyak *et al.*, 2000), position within the microsatellite (Brohede & Ellegren, 1999), and the number of nucleotides within each repeat unit (Xu *et al.*, 2005).

Microsatellites are frequently conserved between species; primers developed in one species will often amplify the same microsatellite locus in related species. For example, many primers developed in cattle were found to amplify in sheep and deer species (Slate *et al.*, 1998). This allows the study of evolutionary relationships

between closely related species in addition to relationships between populations within species. However, using microsatellite primers across species can lead to null alleles being recorded due to point mutations within binding sites. Microsatellites are also more polymorphic in the species from which they are initially cloned (Ellegren *et al.*, 1997).

1.3. Consequences of adaptive processes for genetic variation

In addition to the neutral processes of genetic drift and gene flow, the adaptive process of natural selection affects levels of genetic variation. Natural selection acts on mutations that have consequences for fitness, increasing the frequency of beneficial mutations and eliminating those that are disadvantageous. Over many generations natural selection produces phenotypes that are optimally adapted to their environment. The fitness associated with each allele and the associated genetic variation varies throughout the generations due to the non-static nature of the environment. Natural selection can act in several different ways. Fitness can decrease away from a mean phenotype (stabilising selection), increase towards an extreme phenotype (directional selection), or be directed towards two different niches in heterogeneous environments (disruptive selection). In order for natural selection to increase overall fitness in a population a degree of genetic variability must first be present. In the absence of natural selection assumptions can be made regarding the degree of genetic variation expected in populations as a result of mutation rates and recombination alone (Kimura, 1968). Deviations from models of neutrality can therefore provide evidence of natural selection (Otto, 2000; Nielsen, 2001). Conversely, regions of the genome under the influence of natural selection can not be used in empirical studies which rely on the effects neutral processes have on genetic variation (such as population demography) (van Oosterhout *et al.*, 2004).

There are many studies examining evidence of natural selection on organisms such as *Drosophila* and humans (Wall *et al.*, 2002; Glinka *et al.*, 2003; Stajich & Hahn, 2005), where gene coding regions are characterised and genetic markers are

available. However, it remains important to examine patterns of genetic variation in non-human natural populations to gain a more general understanding of how natural selection and neutral processes regulate genetic variation.

1.4. The mammalian immune system

Perhaps the best-characterised gene regions implicated in non-neutral variation are those involved in the vertebrate innate immune response (Hedrick, 1999; Hughes & Nei, 1989). Examining genetic variation at immune loci allows the investigation of evolutionary consequences of disease, while comparing allelic variation between populations can reveal disease resistance alleles. Four important regions related to disease resistance are included in this thesis; brief outlines of their functions are given below.

1.4.1. The major histocompatibility complex

The vertebrate MHC comprises a group of linked genes that in humans (where it is known as the HLA) number over 100, almost half of which have some role in immunity (Flajnik & Kasahara, 2001). The four megabase region is the most well characterised part of the vertebrate genome. The primary role of genes within the MHC is the recognition of foreign antigens and their subsequent presentation to T cells (Benjamini *et al.*, 1996).

MHC genes are divided into three regions (or classes). Class I genes are expressed continually in most nucleated cells. Class I molecules bind peptides produced by the intracellular degradation of proteins and display them on the cell surface. In most cases peptides are host derived and elicit no immune response. However, foreign peptide complexes are recognised and lysed by antigen specific cytotoxic (CD8+) T-lymphocytes (Harding & Unanue, 2000). Class I MHC genes play an important part in the immune response to intracellular pathogens, especially those of viral origin (Benjamini *et al.*, 1996).

Class II MHC genes encode cell surface glycoprotein receptors. Unlike class I genes, class II genes are expressed only in immunocompetent cells such as macrophages and B-lymphocytes (Benjamini *et al.*, 1996). After internalisation of extracellular pathogenic organisms by immunocompetent cells, vesicles containing the foreign antigens fuse to specialised endosomal compartments (Figure 1.1). Pathogens are lysed in these compartments and resultant antigenic peptides are bound to specific class II receptors. Receptor-peptide complexes are then presented at the cell surface where they are recognised by circulating helper (CD4+) T-cells (Harding & Unanue, 2000). The combined response of class I products presenting intracellular peptides to T cells and exogenously derived peptides, such as bacterial products or viral capsid proteins presented by class II proteins is a major component of the vertebrate adaptive immune response.

The class III MHC region is located between class I and class II regions on the chromosome. Class III genes are the most well conserved of the mammalian MHC region. As such it is has been proposed to be the primordial immune complex (Saltercid & Flajnik, 1995). Class III genes encode several proteins involved in the immune response including complement components C4 and the cytokine tumour necrosis factor (TNF) (Flajnik & Kasahara, 2001). The region also contains many genes not involved in immunity.

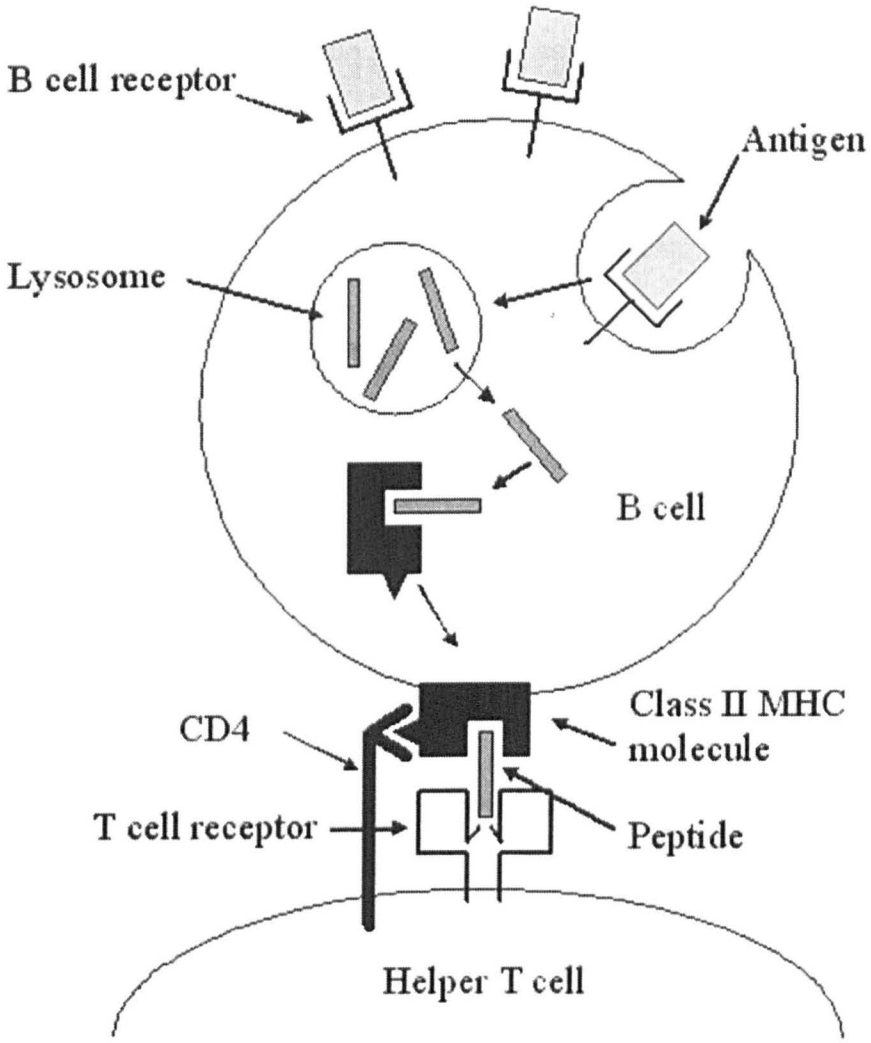


Figure 1.1. A cartoon illustrating activation of the mammalian immune system following the detection of foreign antigenic peptides by MHC class II molecules.

1.4.2. Interferon gamma (IFNG)

Cytokines are soluble proteins which play a role in viral defence and contribute towards the regulation on the acquired immune response. The cytokine IFNG is mainly secreted by activated NK and T cells (Boehm *et al.*, 1997). The subsets of T helper cells (Th_1 and Th_2) vary in their production of cytokines. Th_1 cells are associated with cell-mediated immunity and preferentially secrete IFNG upon induction by intracellular pathogens. Th_2 cells are associated with antibody-mediated mechanisms and favour the production of IL-4 and IL-5 upon induction by nematode infection. The production of IFNG feeds back to further stimulate cell-mediated immune responses through NK cells and the activation of macrophages, while antagonised by other cytokines such as IL-4 (Boehm *et al.*, 1997). There are IFNG receptors on most nucleated cells. Among the first elucidated functions of IFNG was the up-regulation of genes leading to an elevation in the expression of MHC class I and class II molecules (Reiter, 1993). Other effects of IFNG include those relating to cell proliferation and apoptosis.

1.4.3. Natural resistance associated macrophage protein (NRAMP)

NRAMP is expressed in the membranes of macrophages and monocytes where it plays a role in the immune response to several intracellular pathogens (Canonne-Hergaux *et al.*, 1999), including the mycobacteria *Salmonella* and the protozoan *Lieshmania.* After phagocytosis of invading pathogens the NRAMP protein is directed towards the phagosome membrane where it effects microbial replication (Gruenheid *et al.*, 1997), possibly by removing iron from the phagosome (Fleming & Andrews, 1998). NRAMP also has a regulatory role towards interleukin-1β (IL-1β), MHC class II molecules and tumour necrosis factor (TNFα).

1.4.4. Prion protein (PrP)

The PrP gene encodes for a prion protein implicated in the occurrence of several prion diseases, including chronic wasting disease (CWD) and scrapie. Scrapie is a neurodegenerative disease of sheep which ultimately leads to death. Although the exact cause of the disease is not understood, a major component is the host prion protein (Prusiner, 1982). Initial study suggests that the disease is hereditary and the result of a recessive allele. When challenged by a Scrapie brain homogenate, the disease incubation time is under the control of the PrP gene. The ovine PrP gene is located on chromosome 13, and is composed of three exons (Tranulis, 2002). Pathogenesis of all prion diseases involves conversion of normal, cellular PrP into a protease-resistant, pathogenic isoform called PrPSc. The conversion to PrPSc involves change in secondary structure; it is impacts on these structural changes that may link polymorphisms to disease.

1.5. North American mountain sheep

1.5.1. Species range and habitat limitations

Mountain sheep are among the most widely distributed genera of ungulates, with species found throughout Europe, Asia and North America. It is postulated that ancestors of North American wild sheep became isolated from Asian types during the Pleistocene (Valdez & Krausman, 1999). During this time massive ice sheets covered much of northern North America. The volume of ice caused sea levels to drop by 150m, resulting in the creation of the Bering land bridge connecting Siberia with Alaska. During glacial maxima the surrounding area (Beringia) remained ice free, creating a refugium suitable for sheep habitation throughout the ice age. Retreating ice caused a rise in sea levels, resulting in the loss of the Bering land bridge, isolating animals between the two continents. Mountain sheep have been present in North American for 2,000,000 years. The oldest fossil evidence comes

from Yukon deposits, which bear a close resemblance to extant species (Valdez & Krausman, 1999).

There are currently around 185,000 mountain sheep in North America. Sheep are split into two species, bighorn (*O. canadensis*) and thinhorn (*O. dalli*), which differ in their morphology and species range. A defining characteristic of thinhorn are their more widely curled horns. Thinhorn also possess smaller cranial and body measurements than bighorn (Geist, 1971). It is likely that the ancestors of extant North American wild sheep inhabited all suitable range throughout the continent. The current hypothesis suggests that morphological differences between extant species evolved in isolation (Valdez & Krausman, 1999), thinhorn to the north of the ice and bighorn to the south, although the evolutionary relationship between species remains a debated topic. There is no contemporary overlap in the ranges of thinhorn and bighorn sheep.

The present day geographic range of thinhorn sheep encompasses much of mountainous northwestern North America, from western Alaska to the Mackenzie River in the east, and to the Pine Pass of the Rocky Mountains of northeastern British Columbia in the southeast (Figure 1.2). There are an estimated 120,000 individuals present (Valdez & Krausman, 1999; Veitch & Simmons, 1999; Barichello *et al.*, 1989), over half of which can be found in Alaska. The current census estimate is representative of a hypothesised ancestral population size, the number of animals remaining approximately constant since colonisation. The most recent censuses estimate a total of 22,000 animals in the Yukon (Barichello *et al.*, 1989), 7,000 to 26,000 in the Northwest Territories (where data is scarce) (Veitch & Simmons, 1999), 12,000 in British Columbia and approximately 74,000 in Alaska (Valdez & Krausman, 1999).

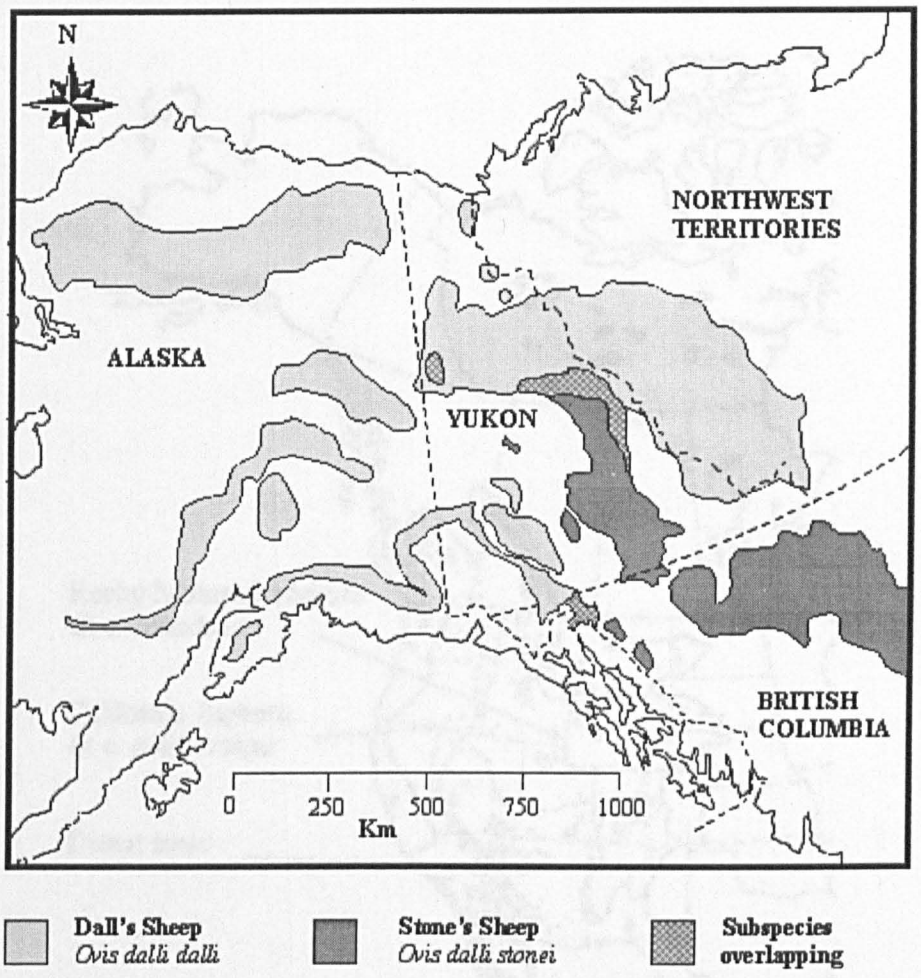


Figure 1.2. The species range of thinhorn across northern North America. Alaska and the Northwest Territories contain only Dall's sheep, while the British Columbia range contains only Stone's. Subspecies meet across central Yukon.

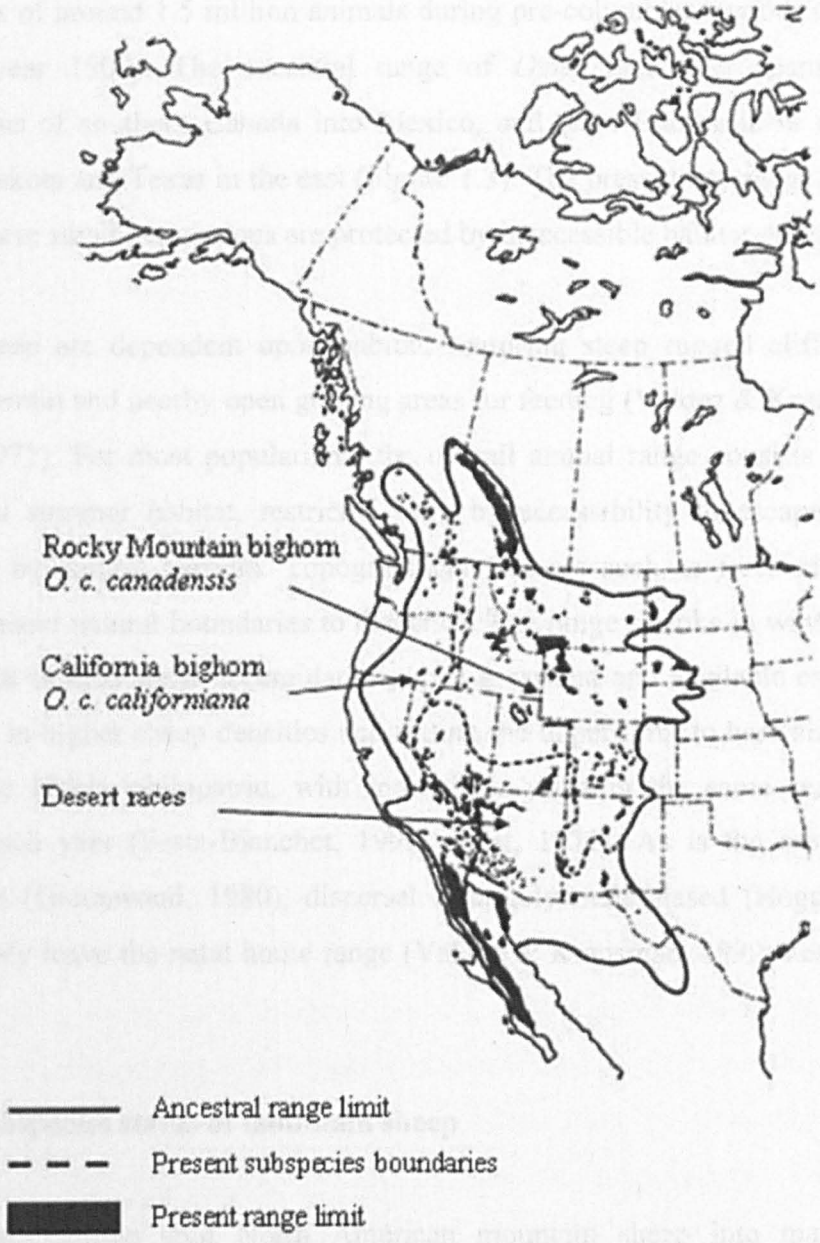


Figure 1.3. Current distribution of North American populations of bighorn sheep (*O. canadensis* spp). Figure adapted from Valdez & Krausman (1999).

Bighorn sheep are less numerous than their sister species, with current census estimates across the species range of less than 65,000. This is much reduced from estimates of around 1.5 million animals during pre-columbian periods (which ended in the year 1500). The ancestral range of *Ovis canadensis* spans the Rocky Mountains of southern Canada into Mexico, and from California in the west into North Dakota and Texas in the east (Figure 1.3). The present day range is reduced to areas where small populations are protected by inaccessible habitat or by refuges.

Wild sheep are dependent upon habitat, requiring steep rugged cliffs for use as escape terrain and nearby open grazing areas for feeding (Valdez & Krausman, 1999; Geist, 1971). For most populations, the overall annual range consists of relatively unlimited summer habitat, restricted only by accessibility to escape terrain and bounded by natural barriers. Topographical features such as forested valleys and rivers present natural boundaries to dispersal. The range shrinks in winter to smaller areas with limited snow accumulation, open grassland and available escape terrain, resulting in higher sheep densities and setting the upper limit to herd size. Mountain sheep are highly philopatric, with individuals utilising the same seasonal home ranges each year (Festa-Bianchet, 1991; Geist, 1971). As is the case with most mammals (Greenwood, 1980), dispersal is largely male-biased (Hogg, 2000) and ewes rarely leave the natal home range (Valdez & Krausman, 1999; Festa-Bianchet, 1991).

1.5.2. Subspecies status of mountain sheep

Early classification split North American mountain sheep into many different species, which do not hold up to modern species definitions. Not all forms are reproductively isolated, but rather allopatrically distributed throughout North America. Thinhorn are currently classified in two putative subspecies predominantly on the basis of coat colour. The more abundant white Dall's sheep (*O. dalli dalli*) comprise over 50% of extant wild sheep in North America and can be found across much of the species range, except the far south. The darker Stone's sheep (*O. d.*

stonei) are less numerous and are found in the Yukon and northern British Columbia only (Figure 1.2). The proportion of each subspecies in the Yukon is not uniform, with more Stone's present in the southern ranges than in the north (Figure 1.2), although there are around six times more Dall's sheep in the territory (Barichello *et al.*, 1989). The taxonomic validity of the subspecies status in thinhorn and other wild sheep species is debatable. Genetic analyses of desert bighorn reveals that currently recognised subspecies do not provide a full explanation for phylogenetic boundaries (Ramey, 1993; Gutierrez-Espeleta, 1999). Rather, it is probable that differences between individuals are representative of variation in one species over a geographical range.

There are many colour variants within each thinhorn subspecies. All Alaskan Dall's sheep are pure white while those from southwestern Yukon and the Mackenzie Mountains of the Northwest Territories occasionally have grey colouration on their flanks. All Stone's sheep from the Pelly ranges of the Yukon are dark but are on average lighter than those from British Columbia. In regions of subspecies overlap many sheep are classified as 'fannin' or saddle-backed. They often show intermediate colouration such as light flanks and darker tails or heads and are presumed by many to be the result of introgression between subspecies (Barichello *et al.*, 1989). The existence of fannin sheep as a separate subspecies is invalidated due to the continuous nature of thinhorn coat colouration across subspecies. Instead they are currently recognised as a variant of Stone's sheep. Due to the wide spectrum of coat colour variants within each subspecies, colouration is a weak diagnostic tool in defining thinhorn subspecies. Other differences between subspecies include Stone's sheep having generally longer horns than those of Dall's, although the former shows more 'flare' in the horns curl.

As in thinhorn there are recognised bighorn subspecies, although the number of these classifications is now generally recognised to be lower than that originally suggested. Presently there are three recognised groupings, Rocky Mountain bighorn (*O. canadensis canadensis*), California bighorn (*O. c. californiana*) and desert races (*O.*

c. nelsoni, *O. c. weemsi*, *O. c. mexicana* and *O. c. cremnobates*). All subspecies are patchily distributed (Figure 1.3). The delineation of subspecies in both species remains an unresolved issue.

1.5.3. Wild sheep and disease

1.5.3.1. Likely parasites and diseases of wild sheep populations

There a number of parasites known to infect wild sheep populations, including bacteria, viruses and nematodes. Among the most severe in North American wild sheep is the bacterium *Pasteurella*. Infection causes respiratory distress and can result in death through severe pneumonia. This is most severe in lambs, where high rates of *Pasteurella* infection can result in the death of up to 95% within a month of birth (Watson & Davies, 2002; Monello *et al.*, 2001). The strains of *Pasteurella* most detrimental to wild sheep can be carried by domestic sheep, goats, horses, cattle and llamas with little or no ill effect (Rudolph *et al.*, 2003).

Wild sheep are highly susceptible to lungworm infection (*Protostrongylus* spp.). Sheep can ingest snails, the hosts for lungworm larvae, while grazing. Larvae then penetrate the intestinal wall, maturing once they reach the lungs. Both *Parelaphostrongylus odocoilei* and *Protostrongylus stilesi* are known to be present in Dall's sheep of the Mackenzie Mountains (NT) (Kutz *et al.*, 2001) where they are associated with haemorrhages and localized muscle inflammation. Adults lay their eggs within the hosts lungs, their larvae enter the air passages upon hatching where they are coughed up and swallowed. Lungworm larvae are excreted in faecal pellets and seek the host snail. Large numbers of the parasite causes respiratory stress and can create lesions in the lungs and bronchial passages that predispose the host sheep to pneumonia through bacterial invasion. Many sheep dying through infection by *Pasteurella* also have heavy lungworm infections.

Other presumed parasites of wild sheep are less widespread. One example is bluetongue, a viral disease spread through gnats that bite an infected animal and then

spread the virus to subsequent hosts. Very few cases of bluetongue have been documented in North American wild sheep, all in bighorn (Noon *et al.*, 2002). The impacts of this disease on wild sheep populations are not known but it is suspected that transmission of this disease would be easy between infected individuals of different species, especially if they were to be in contact for any amount of time during summer months when gnats are more prevalent. Other common diseases with unknown impact on wild sheep numbers include chronic sinusitis, psoroptic scabies, and parainfluenza-3 virus.

1.5.3.2. History of disease in North American sheep populations

Since the late nineteenth century the numbers of bighorn sheep have been subject to rapid population decline coinciding with human colonisation of their range. Road and dam construction, changes in land use and the introduction of domestic livestock are all responsible for this extirpation. The most important cause of death is the spread of novel disease from domestic sheep and cattle. Contact with domestic sheep frequently results in the decline of resident wild sheep populations due to acute mortality from pneumonia, often associated with the bacterium *Pasteurella multocida.* Mortality is usually associated with a suite of secondary respiratory pathogens. *In vitro* studies show that wild sheep are more susceptible to these respiratory pathogens than domestic sheep (Foreyt *et al.*, 1996; Foreyt, 1989).

Although thinhorn sheep are also susceptible to diseases of domestic sheep (Foreyt *et al.*, 1996), their numbers and range have remained approximately stable since colonisation. Man has had no significant impact on the species through introduced disease and few translocations have been made between populations. This is mostly due to the more harsh northern climates preventing the historical introduction of domestic livestock.

1.5.3.3. Fitness effects of disease in wild ungulate populations

There are likely to be fitness differences between ungulates infected and those uninfected by parasites. Soay sheep highly parasitised by intestinal nematodes are

less likely to survive over winter than those with lower levels of parasitism (Coltman *et al.*, 1999). Moreover, sheep temporarily relieved of their worm burden show a significant increase in over winter survival (Gulland *et al.*, 1993). As thinhorn sheep are also affected by nematode parasites, it could be hypothesised that there will be fitness effects of infection in this species. In contrast, although the chronic stress produced by lungworm infection has been hypothesised to increase lamb mortality, no evidence of this is present in a population of bighorn sheep (Goldstein *et al.*, 2005). Any parasite mediated fitness differences can provide means for selection to act on immune genes.

1.5.4. Why study the genetics of wild sheep populations?

Patterns in the genetic variation of both neutral and selected regions of the genome are described in many study systems, the latter especially in model organisms such as *Drosophila* and in non-model human populations. Given this why should we investigate levels of genetic variation in wild sheep? There are several reasons why we should be interested in doing this, some of which are outlined below. We know little about the demography of natural populations of mountain ungulates, such as the distances which young males disperse from their natal groups. Examining the genetic similarity between spatially differentiated populations can help answer this question. An especially interesting reason to study genetic variation within and between North American mountain sheep species is the uncertainty surrounding colonisation history and species status. A thorough examination of neutral genetic differentiation may be the key needed to resolve these issues. An insight into patterns of genetic diversity within wild sheep immune genes is of particular interest because of the impact introduced parasites have had on bighorn populations. Thinhorn represent a species with a stable population size and species range from which we can investigate a role for selection in the immune genes of ancestral wild sheep populations.

The genetic study of all wild sheep species is made easier by the wealth of genetic information collected from domestic sheep. As a result there are many genetic markers available throughout the genome.

1.6. General methods

1.6.1. Sample sites

Samples from over 900 thinhorn sheep from 24 sampling areas across the species range are used in this study (Figure 1.4). Boundaries of sample area sites (as described in Table 1.1) are defined from game management zones or units in the Northwest Territories (NWT) and northern British Columbia (BC), and by groups of adjacent management subzones from the Yukon Territory (YK). Alaskan sample areas comprise individuals collected from the same mountain range, with a maximum distance separating samples between 20km (in the Central Alaskan Range: CAR) and 140km (in the Gates of the Arctic Reserve: GAAR). Sample areas are divided into seven contiguous mountain blocks defined in Table 1.1. Mountain blocks consist of continuous upland regions separated by unsuitable sheep habitat such as major river valleys or lowland forest. Populations also vary with regard to the subspecies of the sheep sampled. NWT and Alaskan populations comprise only Dall's sheep, while Yukon populations contain herds of both Dall's and Stone's sheep. British Columbia individuals are exclusively Stone's.

1.6.2. Collection methods

The majority of samples were collected between 1994 and 2002 and comprise flakes of horn produced by drilling horn material to insert metal identification plugs (horn "corings") of hunted rams taken at kill registration or compulsory inspection. Alaskan samples comprise whole blood taken from both rams and ewes during the period 1999 to 2002. Between 25 and 42 individuals are sampled in each area (mean

of 38). Horn corings are stored dry in sealed paper envelopes, while whole blood samples stored at -20°C in EDTA.

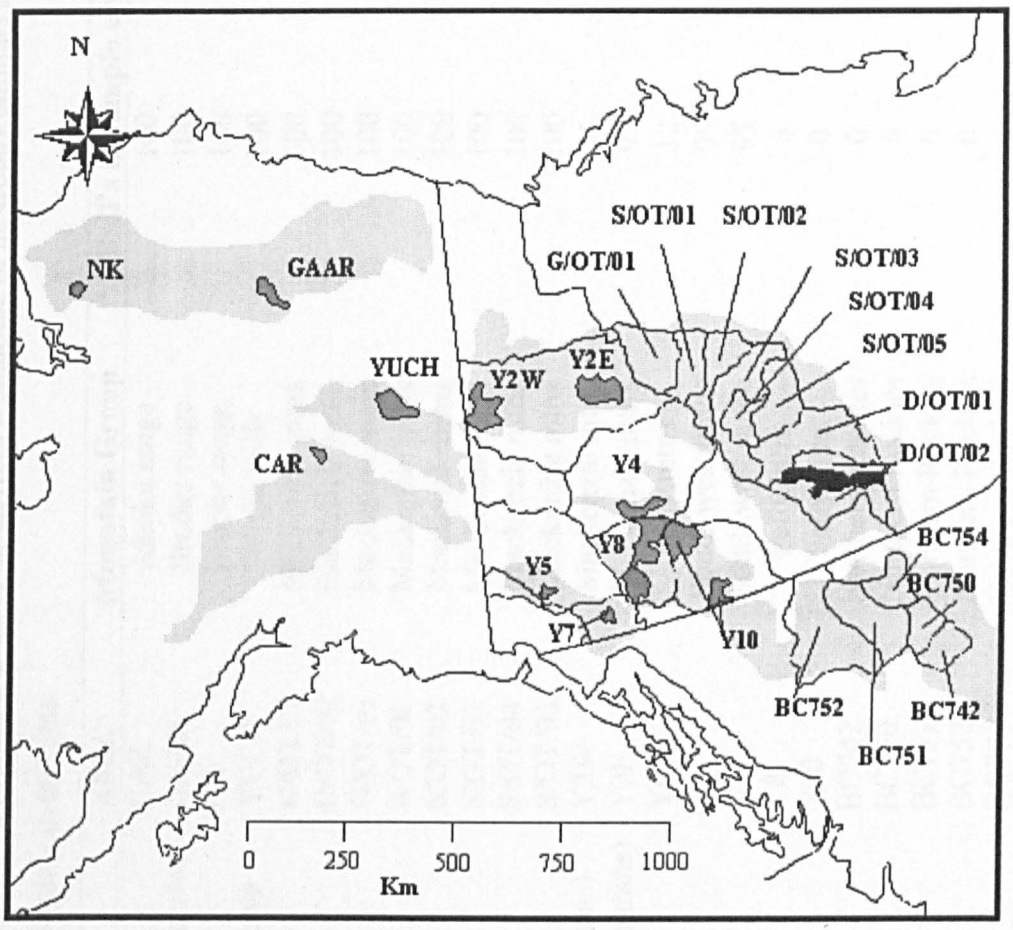


Figure 1.4. North American *Ovis dalli* population locations included in this study. Sampling locations in Yukon and Alaska are within the black shaded areas. Populations in the Northwest Territories and British Columbia cover all the area within the labelled region. Abbreviations for sample area names are described in Table 1.1. Lightly shaded regions represent the full extent of mountain ranges.

Table 1.1. Sample areas are defined from game management zones in British Columbia, the Yukon and Northwest Territories and from geographic clusters in Alaska. The percentage of Dall's in the sample areas is taken from a database of all trophy animals and as so is reported by hunters. Fannin sheep are reported as Stone's in this data.

Region	Sample Area	Abbr.	Mountain Group	% Dall's in sample area
Alaska	Central Alaskan Range	CAR	Alaska range	100
	Gates of the Arctic National Park Preserve	GAAR	Brooks range	100
	Noatak National Preserve	NK	Brooks range	100
	Yukon-Charley Rivers National Preserve	YUCH	White Hills	100
Northwest	Game management zone D/OT/01	D/OT/01	Mackenzie mtns	100
Territories	Game management zone D/OT/02	D/OT/02	Mackenzie mtns	100
	Game management zone G/OT/01	G/OT/01	Mackenzie mtns	100
	Game management zone S/OT/01	S/OT/01	Mackenzie mtns	100
	Game management zone S/OT/02	S/OT/02	Mackenzie mtns	100
	Game management zone S/OT/03	S/OT/03	Mackenzie mtns	100
	Game management zone S/OT/04	S/OT/04	Mackenzie mtns	100
	Game management zone S/OT/05	S/OT/05	Mackenzie mtns	100
Yukon	Game management zone 2 (Ogilvie mtns)	Y2W	Mackenzie mtns	55
Territories	Game management zone 2 (Wernecke mtns)	Y2E	Mackenzie mtns	93
	Game management zone 4	Y4	Pelly mtns	12
	Game management zone 5	Y5	South west range	99
	Game management zone 7	Y7	South west range	95
	Game management zone 8	Y8	Pelly mtns	6
	Game management zone 10	Y10	Pelly mtns	0
British	Game zone 742	BC742	Northern Rockies	0
Columbia	Game zone 750	BC750	Northern Rockies	0
	Game zone 751	BC751	Northern Rockies	0
	Game zone 752	BC752	Northern Rockies	0
	Game zone 754	BC754	Northern Rockies	0

1.7. Thesis aims

In this thesis I will use thinhorn as a model species to investigate both neutral and selective genetic variation within a natural system. Specifically, I use a range of molecular methods with the aim of answering questions in four areas relating to population structure and the presence of variation within genes involved in the vertebrate immune response;

i) Given the strong natal philopatry and dependence on suitable terrain in thinhorn sheep, what is the extent of population structure throughout the species range?

There are several studies of neutral genetic variation in bighorn sheep (Luikart & Allendorf, 1996; Boyce *et al.*, 1999; Forbes & Hogg, 1999). However, most populations have undergone some degree of bottleneck or are derived from few founders of a translocated stock. This limits the number of studies where data are representative of a native wild species. Even with the numerous translocations all bighorn studies report significant evidence of genetic structure. In contrast thinhorn have not had a bottlenecked history. I therefore predict there to be increased levels of neutral variation and a more pronounced genetic structure in the latter species. A previous genetic study shows limited genetic variation in thinhorn allozymes (Sage & Wolff, 1986). Given the higher resolution of microsatellites I expect to show more pronounced structure than was reported there. As outlined above the taxonomic status of thinhorn sheep subspecies are in doubt. Neutral genetic variation can also aid the resolution of this topic.

Studies of population structure across entire species ranges have been conducted on several species, including polar bear (*Ursus maritimus*) (Paetkau *et al.*, 1999), while there are numerous published studies of population structure on smaller scales. It is possible to investigate population structure when levels of microsatellite variation are low. For example, genetic differentiation between Svalbard reindeer (*Rangifer*

tarandus) populations separated by only 45km is possible even with an extremely low mean microsatellite allelic diversity of 2.4 per locus (Cote *et al.*, 2002). Microsatellites can successfully elucidate levels of classification within species, for example the relationships between breeds of Spanish domestic sheep (Arranz *et al.*, 1998). The stringent reliance on suitable habitat required by mountain sheep predicts significant population structure across the thinhorn species range. I expect a greater degree of population structure in wild sheep than is present in plains living mammals such as bison (Polziehn *et al.*, 1996) and caribou (Jepsen *et al.*, 2002) where the effective population size is many times greater, especially for the latter species.

ii) Given the extent of ice-sheets covering suitable sheep habitat during the last glaciation events, where were the refugia for North American wild sheep during the last ice-age?

Descriptions of the effects of glaciation are common amongst studies of species phylogeography. Indeed, there may not have been any phylogeography in European mammals before the last glaciation (Hofreiter *et al.*, 2004). The presence of ice limits organisms to smaller refugia of suitable habitat which increases following retreat of the ice leaving a fingerprint of colonisation history in their DNA. Examining genetic variation in mitochondrial DNA allows the reconstruction of historical population demography. I expect that mitochondrial DNA will reveal the locations of glacial refugia of North American mountain sheep. In addition genetic variation can reveal the species relationship between thinhorn and bighorn sheep.

Elucidating the evolutionary relationship between known species is a widespread use of genetic data. Here I aim to reconstruct a species phylogeny to investigate the validity of mountain sheep species and subspecies. However, the degree of genetic differentiation between putative species should be investigated solely with no regard to other differences between them. This will be considered here.

iii) What evidence is there for natural selection acting on microsatellites linked to immune regions of thinhorn sheep?

There are relatively few studies showing evidence of non-neutral genetic variation in linked microsatellite loci. The majority consider variation in human populations with the aim of identifying genes under the influence of selection (Payseur *et al.*, 2002; Storz *et al.*, 2004). Many genes identified as being under selection form part of the immune response, and especially the MHC region. As such the MHC is an obvious candidate region for examining the effects of selection on genetic variation. Populations that have been affected by recent disease may be expected to show evidence of natural selection at neutral loci linked to MHC and other immune genes. Thinhorn sheep have been through no known disease related bottlenecks, but as with all wild populations they are affected by various pathogens (Kutz *et al.*, 2001). In addition there are known associations between parasite load and genotype at immune genes in natural and farmed domesticated ungulate species otherwise unaffected by serious disease outbreaks (Paterson *et al.*, 1998; Goldmann *et al.*, 1998; Coltman *et al.*, 2001; Sander *et al.*, 2004). I therefore predict there to be evidence of non-neutrality within microsatellites linked to thinhorn immune genes.

Caution is required when interpreting results of studies aiming to identify natural selection. For example, populations of African buffalo (*Syncerus caffer*) have declined following disease epidemics, leading to the assumption of reduced diversity within the MHC region. In fact no such reduction in variation is present. Rapid re-colonisation by animals originating from outside the affected area explains the lack of any loss in MHC diversity. Subsequence sampling includes these new residents incorrectly assumed to be survivors of disease (Wenink *et al.*, 1998). In addition, factors other than selection, such as population substructure, can cause deviations from expectations of neutral genetic variation (Muirhead, 2001). Despite these issues, successful identification of genes under selection have been conducted on several wild populations (Paterson, 1998; Huang & Yu, 2003).

iv) Is there evidence of natural selection in the nucleotide sequences of sheep immune genes, and if so, is it reflective of findings from linked microsatellites?

The presence of balancing selection is a common phenomenon within the MHC regions of many species. Even though infectious diseases are thought to be the main selective force maintaining diversity within the MHC (Jeffery & Bangham, 2000), selection is present in species with very different incidences of observed disease-related mortality (Hedrick, 1999). Given these findings, I expect to find evidence of balancing selection within the thinhorn MHC region, especially given a known incidence of balancing selection within the MHC regions of other sheep species (Paterson, 1996; Gutierrez-Espeleta *et al.*, 2001).

Sequence variation in thinhorn immune genes will be compared to those of other species differing in disease status and population demography. This will allow a discussion of the effects, if any, these have on sequence variability.

Neutral loci in linkage disequilibrium with those under the influence of selection are also expected to show evidence of non-neutrality. As such I predict that evidence of selection in sequence data should mirror that in immune linked microsatellites. In contrast, a correlation of non-linked microsatellite and sequence variation indicates the importance of drift and migration over selection. This was observed in Atlantic salmon (*Salmo salar*) populations (Landry & Bernatchez, 2001). The degree of any difference in genetic variation between non-linked microsatellites and immune sequences indicates the strength of selection in the immune region. There are known differences between these levels of genetic variation in a bighorn sheep population, although not as great as that expected after a recent disease related bottleneck (Boyce *et al.*, 1997).

PART I:
VARIATION AT NEUTRAL LOCI

CHAPTER 2. POPULATION GENETIC STRUCTURE OF NORTH AMERICAN THINHORN SHEEP (*OVIS DALLI* SPP.)

2.1. INTRODUCTION

2.1.1. Factors affecting population genetic structure

Many species show some degree of geographic stratification of genetic diversity. At selectively neutral loci, population genetic structure is primarily determined by the interplay between genetic drift and the rate of gene flow between geographically separated subpopulations (Slatkin, 1987). In theory, relatively little migration is required for the homogenisation of allele frequencies across populations at equilibrium under the island model (Wright, 1978). However, as most species have a dispersal range considerably less than that of the species range, population structure often reflects a pattern of isolation-by-distance (Slatkin, 1993) where genetic differences increase with geographic distance.

The degree of population structure present within populations is affected by many factors. The effects of neutral drift are more evident in small populations, so a greater magnitude of genetic structure is expected between small isolated populations than between populations of a larger effective size. In addition the magnitude of gene flow between populations is influenced by several social and environmental factors. Among mammals the most common types of social structure are polygynous mating systems and female philopatry (Greenwood, 1980). This is consistently associated with high levels of within-group heterozygosity (Storz, 1999) and reinforced by sex biased gene flow. Environmental factors limiting or facilitating gene flow can be divided into those present today and those ancient processes such as ice age glaciations, whose fingerprints of population bottlenecks and geographical separation can still be observed. In the present day gene flow may be reduced by the presence of unsuitable terrain and facilitated by corridors of suitable habitat.

Due to all these factors there are differences in scale and pattern of population structure across different taxa. Many signatures of structure can be predicted based on life history traits. Large carnivores, such as wolves and bears, generally have larger potential dispersal distances and therefore decreased levels of structure than mountain ungulates (Forbes & Hogg, 1999). Forest or plains dwelling ungulates show low to moderate levels of genetic structure (Polziehn *et al.*, 2000; Wilson & Strobeck, 1999; Broders *et al.*, 1999) as they are able to utilise larger areas of continuous suitable habitat and may conduct long distance migration. Animals with more stringent habitat preference criteria, such as mountain sheep, may be expected to exhibit more pronounced population structure. The combination of limited gene flow and small population size should cause genetic differences to accumulate rapidly between geographically separated populations.

Quantifying levels of genetic structure can give an insight into the colonisation history of a species (Perez *et al.*, 2002), rates of dispersal (Waser & Strobeck, 1998) and the effects of environmental barriers on gene flow (Carmichael *et al.*, 2001; Paetkau *et al.*, 1999; Kyle & Strobeck, 2002). In trophy game animals, such as thinhorn sheep (*Ovis dalli* spp.), knowledge of genetic structure may also have practical implications for harvest management and for forensic analyses of suspected illegal hunting (Primmer *et al.*, 2000).

2.1.2. Population structure in wild sheep species

Perhaps unsurprisingly, previous studies on mountain sheep have demonstrated considerable genetic structure. Mitochondrial DNA analyses in desert bighorn (*Ovis canadensis nelsoni*) ewes showed significant differentiation at the level of the home range group (Boyce *et al.*, 1999; Ramey, 1995). Studies using nuclear markers have found significant genetic distances between Rocky Mountain bighorn (*O. c. canadensis*) populations sampled across the species range (Forbes & Hogg, 1999), and on a much finer scale between recently reintroduced populations (Fitzsimmons *et al.*, 1997).

Human activity is likely to have influenced the genetic structure of bighorn sheep and other mountain ungulates. Translocations, population declines due to unregulated hunting and habitat loss, and the transmission of novel pathogens from domestic sheep have severely impacted bighorn sheep population structure (Luikart & Allendorf, 1996). These factors make it difficult to representatively sample the native range of the bighorn sheep. In contrast, thinhorn occupy most of their historic range in approximately ancestral numbers. Cases of domestic livestock introduction in the mostly remote areas inhabited by thinhorn are few, having little impact on wild sheep numbers. In addition there have been limited recorded translocations either into or out of populations. Thinhorn sheep are therefore an ideal model in which to study mountain ungulate population structure as it evolved prior to anthropogenic influence.

2.1.3. Outline

This chapter details a study using ungulate-derived microsatellite markers (Slate *et al.*, 1998; de Gortari *et al.*, 1997) aiming to quantify levels of genetic variation in thinhorn sheep populations across the species range. Specifically, I aim to investigate the following hypotheses:

i) There is an abundance of neutral genetic variation present within thinhorn populations. Previous genetic study on thinhorn utilising allozyme markers was unable to find significant levels of genetic variation (Sage & Wolff, 1986). Due to the increased mutation rate of microsatellites I expect to find substantial levels of variation. Female philopatry, such as is present in wild sheep, is expected to increase the degree of population structure relative to other forms of social system (Storz, 1999).

ii) Thinhorn populations from across different regions of the species range are genetically distinct. As mountain sheep are known to exhibit natal philopatry and sightings of long distance migration events are negligible, I expect to find evidence

of significant substructure across the species range. From these data I also aim to examine evidence of barriers to gene flow, environmental and otherwise, and their role in the maintenance of population structure.

iii) There are genetic differences between the two putative thinhorn subspecies of Dall's and Stone's sheep. The legitimacy of classifying thinhorn into two subspecies is an issue much debated. Throughout this and following chapters I will use a variety of genetic data with the aim of clarifying this issue.

2.2. MATERIALS AND METHODS

2.2.1. Samples

A total of 919 individuals from all 24 sample areas (Figure 1.2) were included in this study. Sample areas comprised a minimum of 25 individuals with a mean of 38 per region (Table 2.1).

2.2.2. Molecular techniques

Genomic DNA was extracted from approximately 0.5ml of horn material per sample using a tissue extraction kit (Qiagen, Crawley, West Sussex) with modified protocol (Appendix A). Blood samples were extracted using a phenol chloroform technique (Sambrook *et al.*, 1989) from 200µl of blood. Twelve dinucleotide microsatellite loci developed in domestic sheep and cattle (Table 2.1) were amplified for each individual. Loci were chosen based on levels of variability, chromosomal location and previous successful use in bighorn sheep (Coltman *et al.*, 2003a; Coltman *et al.*, 2002).

Each polymerase chain reaction (PCR) was carried out in 10µl reactions. Reactions contained 2µl of DNA template, 80 µmol of each primer, 0.16mM dNTP's, 2 mM $MgCl_2$, and 0.5 units *Taq* polymerase (Sigma, Gillingham, Dorset, or Bioline, London). Additionally, the quality of some products was improved by the addition of 25 mg ml^{-1} bovine serum albumin (BSA). The PCR profile consisted of 35 cycles of 30s each at 94°C and 54°C followed by 40s at 72°C. Cycles were preceded by 5 min at 94°C and terminated with 10 min at 72°C. PCR products were genotyped using an ABI 377 sequencer and analysed using the software GENESCAN and GENOTYPER (Applied Biosystems, Foster City, CA).

2.2.3. Data analyses

2.2.3.1. Genetic diversity within sample areas

Genetic variability was quantified within each sample area by the number of alleles per locus (A) and expected heterozygosity (H_E) as calculated by GENETIX v.4.01 (Belkhir, 1996). GENETIX was also used to determine these values for each locus. Homogeneity of genetic variation between the 24 sample areas was tested using Wilcoxon signed rank tests blocks. Departures from Hardy Weinberg equilibrium (HWE) were examined using exact tests (Guo & Thompson, 1992) using a Markov chain as implemented by GENEPOP V.3.3 (Raymond & Rousset, 1995). Loci were combined using Fisher's method to examine departure from equilibrium for each area, with the significance interpreted after sequential Bonferroni correction for multiple comparisons (Rice, 1989). The presence of linkage disequilibrium between loci within sample areas was also tested using exact tests with GENEPOP software.

2.2.3.2. Genetic differentiation between sample areas

Exact tests were performed for allele frequency differences between all pairs of areas. Genetic divergences between sample areas were quantified using the distance statistics F_{ST}, Nei's unbiased genetic distance, D_S (Nei, 1978) and the distance $(\delta\mu)^2$ based on allele size (Goldstein *et al.*, 1995). D_S has been shown to fare better at fine-scale population differentiation, while $(\delta\mu)^2$ has proved more useful for examining relationships between more distinct populations and for estimating evolutionary times (Pactkau *et al.*, 1997). F_{ST} is insensitive when migration rates are low. D_S was therefore used in preference to F_{ST} when constructing isolation by distance plots and trees. F_{ST} was calculated with GENEPOP, other distance measures were calculated using the software SPAGeDi v.1.0 (Hardy & Vekemans, 2002). Significance of F_{ST} was tested with 10,000 permutations using GENETIX. PHYLIP 3.5 (Felsenstein, 1993) was used together with the program TREEVIEW 1.5 (Page, 1996) to construct an unrooted tree from the D_S distance matrix.

Determining the origin of individuals using microsatellite markers is useful in many applications of conservation, such as identifying the source population in investigations of illegal hunting (Manel *et al.*, 2002). In this instance assignment tests were conducted to test the reliability of placements to correct source sample areas and to identify any possible migrants in the dataset (Waser & Strobeck, 1998; Rannala & Mountain, 1997). There are many ways of identifying the origin of individuals using genotype data. Assignment tests were implemented using the Bayesian methodology (Rannala & Mountain, 1997) of GENECLASS (Cornuet *et al.*, 1999) with both no simulation and simulation of sampling from a population size of 10,000 individuals, with minimum probability of belonging to the population set at 0.05, 0.01 and 0.001. The Bayesian methodology used is more powerful than methods based on frequencies (Maudet *et al.*, 2002; Manel *et al.*, 2002; Cornuet *et al.*, 1999). GENECLASS determines the probability of a genotype originating from each population in turn, before assigning it to populations where the probability is above the minimum threshold set, or to no population if the threshold is not exceeded. An individual was considered correctly assigned when it could be excluded from all but one population at each level of probability. Both analyses used the 'leave one out' method, removing the individual to be assigned before assigning it based on remaining allele frequencies of each population. Individuals that were cross-assigned were examined as putative migrants. Samples that were 100 times more likely to be assigned to a population other than the source were taken as being migrants.

The Bayesian methodology of STRUCTURE v.2.0 (Pritchard *et al.*, 2000) was implemented to determine the level of genetic substructure in the dataset independently of sampling areas. To estimate the number of subpopulations (K), five independent runs of K = 1-20 were carried out at 100 000 MCMC repetitions. The most probable number of populations was taken using the log-likelihood of K. Individuals were then assigned to each subpopulation, based on the highest percentage membership (q). Sample locations of these individuals were then plotted

on a map of the species range to examine the relative geographical position of subpopulations.

2.2.3.3. Isolation-by-distance

The relationship between genetic and geographic distances was examined to assess isolation by distance (Slatkin, 1993). Geographic distances between pairs of areas were calculated from linear distances between mean latitude and longitude positions of samples from each area and plotted against the genetic distances D_s and $(\delta\mu)^2$. The effects of mountain block and colour polymorphism on genetic distance was investigated by controlling for the effects of geographic distance by partial Mantel tests (10,000 permutations) as calculated by the R-PACKAGE version 4.0 (Casgrain & Legendre, 2001). To assess the effects of subspecies and mountain block on genetic distance independently of geographic distance, a general linear model of the residuals from linear regressions of genetic on geographic distances was conducted.

2.3. RESULTS

2.3.1. Genetic diversity and tests of disequilibrium

Twelve locus microsatellite profiles were recorded for 919 thinhorn sheep. To assess the accuracy of profiles obtained from horn material, a subset of 50 individuals were retyped. In all cases identical genotypes were returned (total of 600 genotypes from 12 loci), showing the profiles to be repeatable. Overall, the rate of missing data was 1.9%.

Between four (*BMC1222*) and 22 (*BM848*) alleles were found at the 12 loci (mean 12.33, SE 0.43). Expected heterozygosities per locus ranged from 0.229 (*MAF209*) to 0.922 (*BM848*) (Table 2.2). At less diverse loci several alleles were present in Mackenzie and southern YK populations, while one allele was fixed in Alaska and BC. Populations had a mean number of alleles per locus (*A*) between 3.97 and 7.50 (Table 2.1). Measures of genetic variation did not differ between areas (A: Z=0.1065, P=0.9152; H_O: Z=-0.0152, P=0.9879), although the highest measures of genetic diversity were observed in areas from the Mackenzie and Yukon (Pelly and south west) mountain ranges. F_{IS} for all but one area was positive, ranging from -0.028 to 0.119. Several significant departures from HWE due to heterozygote deficit were found, with every locus showing deviation in at least one area (see Tables 2.1 and 2.2 for the distribution). The repeatability of microsatellite genotypes, low rate of missing data and consistent levels of HWE disequilibrium across loci suggest that heterozygote deficiencies did not arise from PCR artefacts.

Tests for genotypic disequilibrium revealed no significant association after correcting for multiple comparisons. When uncorrected, linkage disequilibrium was suggested between *BM1225* and *FCB266* (χ^2=71.35, d.f.=48, P=0.016). Physical linkage between these markers is unlikely as they are located on separate chromosomes in domestic sheep (Maddox *et al.*, 2001) (16 and 25 respectively).

Table 2.1. Genetic variability estimates from populations included in the study (mean number of alleles per locus (A), observed (H_O) and expected heterozygosities (H_E), and mean estimates of F_{IS} (Weir & Cockerham, 1984). Also included are the numbers of loci showing significant deviation from Hardy Weinberg Equilibrium at each population (at $P<0.05$). Global departure across all loci from HWE is indicated by *. Significance of F_{IS} is indicated $P<0.05$*, $P<0.01$**.

Sample area	N	% Dall's in sample	H_O	H_E	Loci not in HWE	A	F_{IS}
CAR	25	100	0.58	0.58	1	5.17	0.020
GAAR	34	100	0.55	0.60	3*	5.17	0.107**
NK	25	100	0.63	0.60	0	3.97	-0.028
YUCH	36	100	0.57	0.59	4*	5.25	0.050*
D/OT/01	40	100	0.62	0.65	0	6.92	0.055**
D/OT/02	42	100	0.54	0.57	5*	6.92	0.070**
G/OT/01	40	100	0.60	0.62	1	6.92	0.047*
S/OT/01	40	100	0.62	0.63	2	7.50	0.040
S/OT/02	40	100	0.64	0.65	3	7.33	0.024
S/OT/03	40	100	0.60	0.65	1	7.08	0.096**
S/OT/04	40	100	0.59	0.66	1	7.25	0.119**
S/OT/05	39	100	0.64	0.67	2	7.42	0.061*
Y2W	40	93	0.62	0.67	4	6.58	0.087**
Y2E	40	100	0.63	0.66	4	6.58	0.057*
Y4	40	36	0.59	0.66	4*	7.17	0.116**
Y5	40	100	0.56	0.58	3	5.67	0.051*
Y7	40	100	0.55	0.59	5	6.17	0.064**
Y8	40	38	0.57	0.61	2	6.00	0.070**
Y10	40	32	0.60	0.64	0	6.33	0.070**
BC742	40	0	0.48	0.49	2	4.42	0.042
BC750	40	0	0.51	0.54	2	5.17	0.071**
BC751	40	0	0.47	0.52	2	5.58	0.104**
BC752	38	0	0.47	0.48	1	5.00	0.043
BC754	40	0	0.46	0.51	2	4.83	0.101**

Table 2.2. Comparison of the genetic diversity found at the 12 microsatellite markers used (number of alleles over all populations (A), size range of alleles and mean observed (H_O) and expected heterozygosities (H_E) of all sample areas). Also included are the number of populations out of 24 that show significant deviation from HWE (P<0.05). Global departure from HWE is indicated by *.

Locus	A	Allele size range (bp)	H_o	H_E	Areas deviating from HWE	Reference
AE16	9	82-98	0.63	0.66	6*	(Penty *et al.*, 1993)
BM1225	13	225-265	0.63	0.68	6*	(Bishop *et al.*, 1994)
BM4505	11	239-275	0.48	0.55	9*	(Bishop *et al.*, 1994)
BM4513	16	131-161	0.71	0.73	1	(Bishop *et al.*, 1994)
BM848	22	201-243	0.75	0.83	10*	(Bishop *et al.*, 1994)
BMC1222	4	288-294	0.19	0.23	1	(de Gortari *et al.*, 1997)
CP26	19	123-165	0.69	0.72	3	(Ede *et al.*, 1995)
FCB266	13	82-106	0.63	0.65	4	(Buchanan & Crawford, 1993)
MAF209	5	109-119	0.20	0.21	1	(Buchanan & Crawford, 1992)
MAF36	13	84-116	0.72	0.74	5	(Swarbrick *et al.*, 1991)
TGLA126	11	112-144	0.54	0.54	5	(Georges & Massey, 1992)
TGLA387	12	129-153	0.68	0.70	3	(Georges & Massey, 1992)

2.3.2. Genetic differentiation

2.3.2.1. Measures of genetic distance

Allele frequencies between pairs of sample areas differed significantly at up to all twelve loci (Table 2.3). Significant genic differentiation (all with $P<0.0001$) was recorded between all but two pairs of areas (BC751 and BC754; $\chi^2=20.867$, d.f.=24, and G/OT/01 and S/OT/03: $\chi^2=35.288$, d.f.=24, $P=0.0642$) when tests were combined across loci. Global F_{ST} was 0.160 with pairwise F_{ST} ranging from -0.0034 to 0.3575 (Table 2.3). The greatest levels of differentiation were observed between Brooks Range and BC populations, whereas comparisons within the BC range and Mackenzie Mountains showed little genetic structure. Patterns in the statistics D_S and $(\delta\mu)^2$ paralleled those found in F_{ST}. Correlation between D_S and F_{ST} was high (r=0.953), whereas that of D_S and $(\delta\mu)^2$ (r=0.693) and F_{ST} and $(\delta\mu)^2$ (r=0.741) were moderate. An unrooted neighbour-joining tree constructed based on D_S revealed that sample areas on the same mountain range clustered together (Figure 2.1).

Table 2.3. Genetic distance matrix of pairwise F_{ST} (Weir & Cockerham, 1984) as calculated by GENEPOP (Raymond & Rousset, 1995) below the diagonal. All F_{ST} values are significant at <0.01 except those marked *. Values above the diagonal are the number of significantly differentiated loci (from the 12 used) between pairwise sample areas.

	CAR	GAAR	NK	YUCH	D/OT/01	D/OT/02	G/OT/01	S/OT/01	S/OT/02	S/OT/03	S/OT/04
CAR		11	11	11	10	11	11	11	12	10	10
GAAR	0.140		8	11	12	12	12	12	12	12	12
NK	0.201	0.099		11	12	12	12	12	12	12	12
YUCH	0.112	0.148	0.191		12	12	12	11	11	12	12
D/OT/01	0.157	0.138	0.216	0.141		10	8	9	9	7	6
D/OT/02	0.164	0.167	0.240	0.150	0.065		11	11	10	10	10
G/OT/01	0.150	0.160	0.215	0.115	0.036	0.083		4	5	3	7
S/OT/01	0.140	0.130	0.207	0.124	0.037	0.066	0.023		3	4	5
S/OT/02	0.135	0.127	0.193	0.124	0.043	0.048	0.030	0.013		5	5
S/OT/03	0.144	0.145	0.210	0.121	0.018	0.064	*0.003	0.023	0.025		4
S/OT/04	0.128	0.119	0.188	0.115	0.017	0.056	0.026	0.008	0.019	0.013	
S/OT/05	0.114	0.138	0.198	0.134	0.036	0.058	0.061	0.044	0.026	0.043	0.029
Y2W	0.119	0.155	0.186	0.108	0.096	0.113	0.081	0.066	0.052	0.083	0.069
Y2E	0.104	0.108	0.166	0.068	0.057	0.095	0.031	0.035	0.043	0.033	0.033
Y4	0.125	0.152	0.206	0.137	0.104	0.114	0.117	0.094	0.089	0.107	0.083
Y5	0.168	0.196	0.234	0.180	0.210	0.253	0.217	0.201	0.203	0.207	0.182
Y7	0.197	0.203	0.245	0.243	0.220	0.259	0.239	0.225	0.207	0.215	0.203
Y8	0.182	0.167	0.232	0.160	0.132	0.187	0.155	0.129	0.144	0.138	0.120
Y10	0.179	0.183	0.220	0.169	0.132	0.165	0.152	0.136	0.150	0.136	0.116
BC742	0.289	0.281	0.342	0.276	0.200	0.251	0.231	0.222	0.244	0.209	0.202
BC750	0.266	0.261	0.324	0.261	0.171	0.225	0.205	0.196	0.214	0.181	0.179
BC751	0.264	0.266	0.338	0.259	0.176	0.221	0.211	0.204	0.219	0.190	0.182
BC752	0.298	0.297	0.358	0.283	0.210	0.254	0.242	0.227	0.246	0.217	0.210
BC754	0.265	0.273	0.345	0.260	0.182	0.229	0.215	0.207	0.225	0.196	0.185

Table 2.3. continued

	S/OT/05	Y2W	Y2E	Y4	Y5	Y7	Y8	Y10	BC742	BC750	BC751	BC752	BC754
CAR	12	10	9	11	11	12	11	11	11	11	11	11	11
GAAR	12	11	11	12	12	11	11	11	12	12	12	12	12
NK	12	12	11	12	12	12	12	12	12	12	12	12	12
YUCH	12	12	10	11	12	12	11	11	12	12	12	12	12
D/OT/01	7	11	11	10	11	12	11	11	11	11	11	11	11
D/OT/02	11	12	11	10	11	12	12	12	11	10	10	10	11
G/OT/01	10	12	10	11	12	12	11	11	12	11	11	11	12
S/OT/01	11	12	9	12	12	12	12	12	12	12	12	12	12
S/OT/02	8	11	11	11	12	12	12	12	12	12	12	12	12
S/OT/03	8	11	10	11	12	12	11	11	12	11	11	11	12
S/OT/04	8	11	9	11	12	12	11	11	12	12	11	11	12
S/OT/05		12	12	9	12	12	11	11	12	11	11	11	12
Y2W	0.082		10	12	12	12	12	12	12	12	12	12	12
Y2E	0.068	0.059		12	11	12	12	12	11	11	11	11	11
Y4	0.094	0.106	0.104		11	11	10	8	11	11	11	11	11
Y5	0.204	0.178	0.173	0.101		10	11	12	11	10	11	11	11
Y7	0.200	0.199	0.219	0.138	0.123		9	10	12	12	12	11	12
Y8	0.136	0.156	0.133	0.086	0.107	0.153		8	12	12	11	11	12
Y10	0.137	0.147	0.134	0.054	0.121	0.192	0.082		12	11	11	11	12
BC742	0.219	0.250	0.217	0.215	0.283	0.301	0.170	0.172		6	8	7	10
BC750	0.195	0.219	0.194	0.182	0.259	0.272	0.154	0.146	0.025		5	5	4
BC751	0.197	0.230	0.205	0.192	0.263	0.278	0.156	0.147	0.047	0.022		6	0
BC752	0.224	0.250	0.229	0.206	0.282	0.310	0.152	0.151	0.044	0.029	0.026		5
BC754	0.198	0.235	0.209	0.196	0.266	0.291	0.162	0.152	0.047	0.020	*-0.003	0.034	

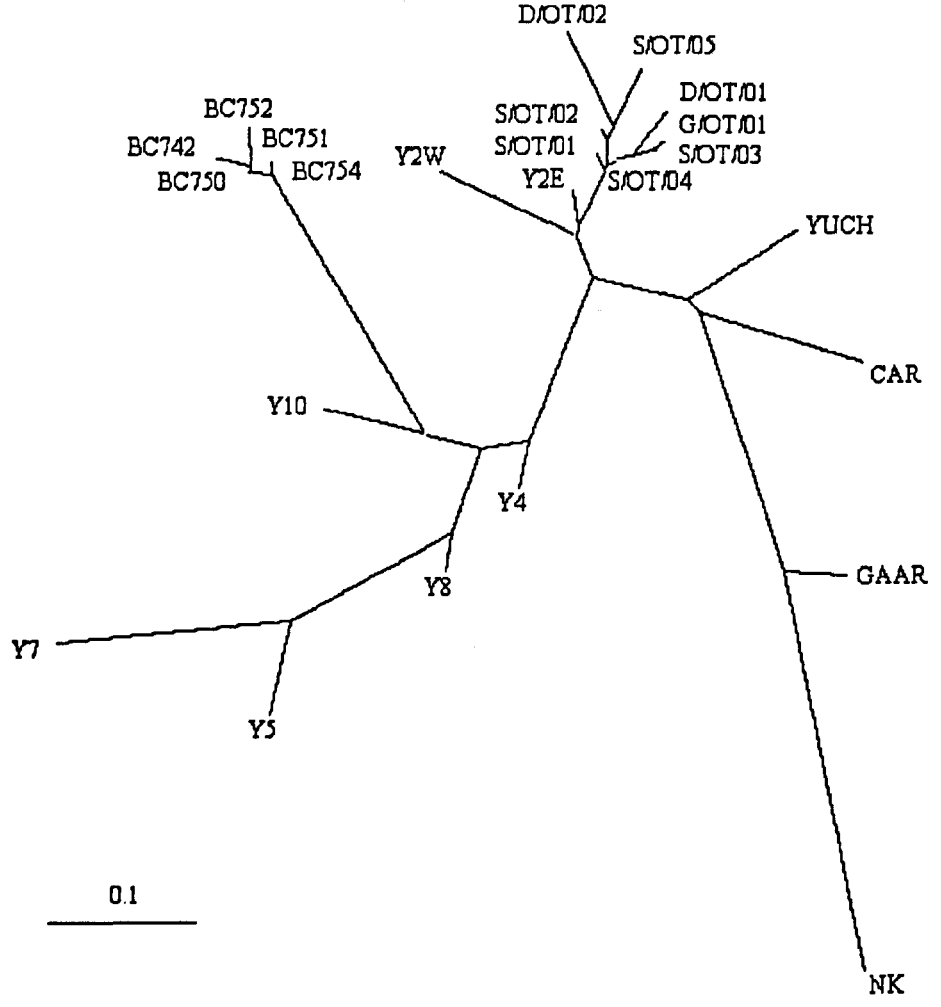


Figure 2.1. Neighbour joining tree of genetic distances (D_S) between sample areas of thinhorn sheep (*O. dalli*) constructed using TREEVIEW (Page, 1996).

2.3.2.2. Traditional assignment tests

Overall, there were high percentages of correct assignment to source sample areas (71.27%). Correct assignment to one of seven mountain blocks (defined in table 1.1) was extremely high at 97.71%. The highest correct assignments to sample areas came from the Alaska range, the Brooks range, and the Tanana mountains of Alaska, and the south west and southern Pelly ranges of Yukon (Table 2.4). The lowest correct assignments were seen within BC.

Most cross assignments were made to geographically adjacent populations. We recorded only one mis-assignment from an Alaskan mountain block to another region, an individual from YUCH assigning to Y2W. However this was only 1.6 times more likely to be the source than the Alaskan population. The BC range individuals assigned with lower success on a population level, yet all except one individual assigned to another BC population. This individual assigned to S/OT/01 in northern Northwest Territories with more than 100 times higher likelihood than the source population. No other individual cross-assigned outside of a mountain range with this level of likelihood. When the assignment is not forced to the population with the highest likelihood, by means of simulating a larger population size, with various threshold probabilities, the number of assignments to the source population falls (Table 2.5).

Assignment plots of likelihood with Dall's and Stone's populations show the extent of the cross assignments (Figure 2.2). Most of the individuals cross assigning to Stone's from Dall's are those from Yukon 5 and Yukon 7 reflecting the geographical rather than colour split. There is one significant cross-assignment from Yukon 7 to Stone's when Yukon Dall's populations are grouped with other Dall's populations.

Table 2.4. GENECLASS assignment matrix of individuals assigned to the most likely population. Individuals cross assigning with a log likelihood difference of more than 4.6, equating to more than 100 times more likely as the source population are indicated (*). There is variation in the ability to assign individuals to the correct source sample area between regions and mountain ranges.

To below	CAR	GAAR	NK	YUCH	D/OT/01	D/OT/02	G/OT/01	S/OT/01	S/OT/02	S/OT/03	S/OT/04	S/OT/05
CAR	25											
GAAR		32	1									
NK		2	24									
YUCH				35							1	
D/OT/01					28		4	1	1	5	1	2 (1*)
D/OT/02					1	37			1			
G/OT/01					1		23		3	14	1	1
S/OT/01					1			21	7		6	1
S/OT/02					1		2	12	19	2	2	3
S/OT/03					7		9	1	3	19	5	
S/OT/04					1		1	3	4		19	2
S/OT/05						4			1		4	29
Y2W				1					1			1
Y2E							1	2			1	
Y5												
Y7												
Y4												
Y8												
Y10												
BC742						1						
BC750												
BC751												
BC752												
BC754												

Table 2.4. continued from previous page. Shaded cells represent correct assignments. Groups of sample areas contained within solid lines comprise those within the same mountain range.

	S/OT/05	Y2W	Y2E	Y5	Y7	Y4	Y8	Y10	BC742	BC750	BC751	BC752	BC754
CAR				1									
GAAR													
NK													
YUCH			2										
D/OT/01	2 (1*)												
D/OT/02													
G/OT/01	1		2										
S/OT/01	1	1	2								1*		
S/OT/02	3		2										
S/OT/03													
S/OT/04	2		2			1							
S/OT/05	29	2											
Y2W	1	36											
Y2E		1	30										
Y5				39	2 (1*)		1						
Y7					38	1							
Y4						31	1	4					
Y8						6 (1*)	36						
Y10							1	36					
BC742									28	3		5 (1*)	3
BC750						1			3	22	4	2	4
BC751							1		3	2	9	5	16
BC752									4	4	8	23	1
BC754									2	9	18	3	16

Table 2.5. Summary of assignment success with simulation of 10,000 individuals, with threshold probabilities set at 0.05, 0.01 and 0.001. Assignment to source with highest likelihood is the output of GENECLASS with no simulation. Where individuals assign to source plus others, the threshold set for the simulation is exceeded for more than one population.

Sample Area	N	Source population is most likely	Assigned to source area only			Assigned to no sample area			Assigned to area(s) other than source		
			0.05	0.01	0.001	0.05	0.01	0.001	0.05	0.01	0.001
CAR	25	25	16	12	8	5	3	1	1	0	0
GAAR	34	32	22	23	21	8	4	0	1	0	1
NK	25	24	12	6	2	2	0	0	2	2	0
YUCH	36	35	23	16	8	2	1	0	3	2	1
D/OT/01	40	28	2	2	3	6	4	1	4	0	0
D/OT/02	42	37	4	4	0	6	3	0	6	2	2
G/OT/01	40	23	0	0	0	5	3	1	6	2	1
S/OT/01	40	21	2	2	0	6	5	2	5	2	2
S/OT/02	40	19	1	2	1	5	3	1	3	3	2
S/OT/03	40	19	3	1	2	4	3	0	4	1	1
S/OT/04	40	19	8	2	2	6	3	0	5	1	1
S/OT/05	39	29	11	12	7	5	2	0	4	1	1
Y2W	40	36	17	13	7	3	1	1	0	0	0
Y2E	40	30	7	4	3	4	2	0	1	3	0
Y4	40	31	17	16	12	8	2	0	1	1	1
Y5	40	39	15	9	5	7	3	1	1	2	0
Y7	40	38	24	23	17	5	1	0	2	1	0
Y8	40	36	18	9	1	8	2	0	2	4	1
Y10	40	36	19	16	4	5	2	1	2	0	1
BC742	40	28	0	0	0	2	1	0	5	2	2
BC750	40	22	9	11	6	6	1	0	1	2	1
BC751	40	9	1	1	1	5	2	2	3	1	1
BC752	38	23	2	1	0	2	0	0	3	3	2
BC754	40	16	0	0	0	2	0	0	6	4	0

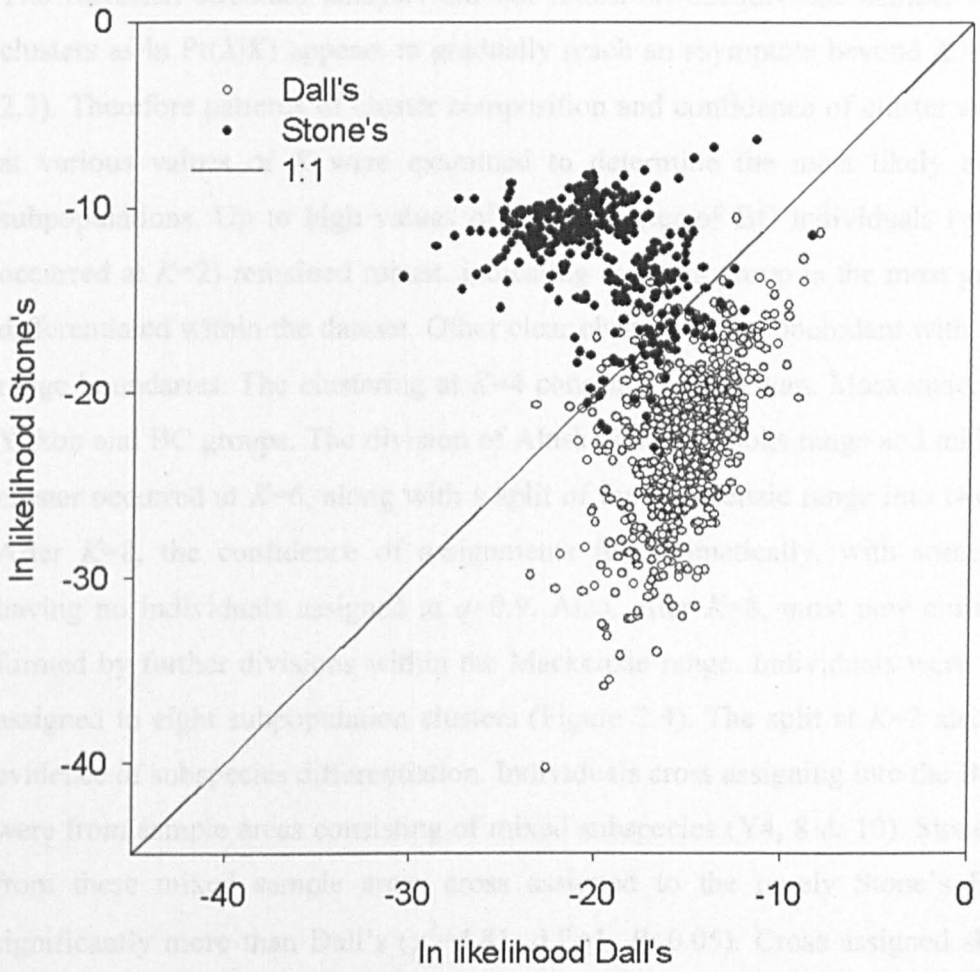


Figure 2.2. The likelihood of assigning to the correct subspecies based on microsatellite genotype. There is some cross-assignment between subspecies, mostly in the direction of mis-assignment to Dall's sheep.

2.3.2.3. STRUCTURE analyses

The Bayesian structure analysis did not return an unequivocal number of genetic clusters as ln Pr($X|K$) appears to gradually reach an asymptote beyond K =8 (Figure 2.3). Therefore patterns of cluster composition and confidence of cluster assignment at various values of K were examined to determine the most likely number of subpopulations. Up to high values of K the cluster of BC individuals (which first occurred at K=2) remained robust, indicating that this group is the most genetically differentiated within the dataset. Other clear clusters were concordant with mountain range boundaries. The clustering at K=4 consisted of Alaskan, Mackenzie, southern Yukon and BC groups. The division of Alaska into a Brooks range and mid Alaskan cluster occurred at K=6, along with a split of the Mackenzie range into two groups. After K=8, the confidence of assignments fell dramatically, with some clusters having no individuals assigned at q>0.9. Also, after K=8, most new clusters were formed by further divisions within the Mackenzie range. Individuals were therefore assigned to eight subpopulation clusters (Figure 2.4). The split at K=2 also showed evidence of subspecies differentiation. Individuals cross assigning into the BC cluster were from sample areas consisting of mixed subspecies (Y4, 8 & 10). Stone's sheep from these mixed sample areas cross assigned to the purely Stone's BC areas significantly more than Dall's (χ^2=4.81, d.f.=1, P<0.05). Cross assigned sheep had lower confidences of assignment, q, than sheep originating from BC.

2.3.2.4. Finer scale geographical barriers to gene flow

The largest region of unsuitable habitat within any of the continuous mountain blocks is the Nahanni River valley in the southern Northwest Territories. A sub-set of sample areas in this region (samples from S/OT/03, D/OT/01 and D/OT/02 with known kill site coordinates, n=114) were re-analysed using GENECLASS and STRUCTURE methods to determine if this indeed comprised an environmental barrier to gene flow. Traditional assignment tests returned a high confidence of correctly placing samples to either north or south of the river valley (Table 2.6) when assignment is forced to the most likely grouping. There was however no statistical support for this (Table 2.7) when assignment was unforced. When no *a priori* source

locations were made (using STRUCTURE), the most likely number of populations was two, largely reflecting the north – south divide across the Nahanni River valley (Table 2.8).

Table 2.6. Assignment success of Nahanni samples when forced to most likely source

	Assigned to;		
Origin	**North**	**South**	**% correct**
North	81	6(1*)	93.1
South	1	26	96.3
Total			93.9

* indicates a sample mis-assigned with likelihood of 100 times more than to source sample area.

Table 2.7. Assignment success of Nahanni samples with $P>0.05$

	Assigned to;			
Origin	**North**	**South**	**None**	**% correct**
North	72	2	13	82.8
South	12	8	7	29.6
Total				70.2

Table 2.8. Proportion of individuals from each sample area in each of the two STRUCTURE clusters

Sample Area	**Cluster 1**	**Cluster 2**
D/OT/02	0.937	0.063
D/OT/01	0.076	0.924
S/OT/03	0.072	0.928

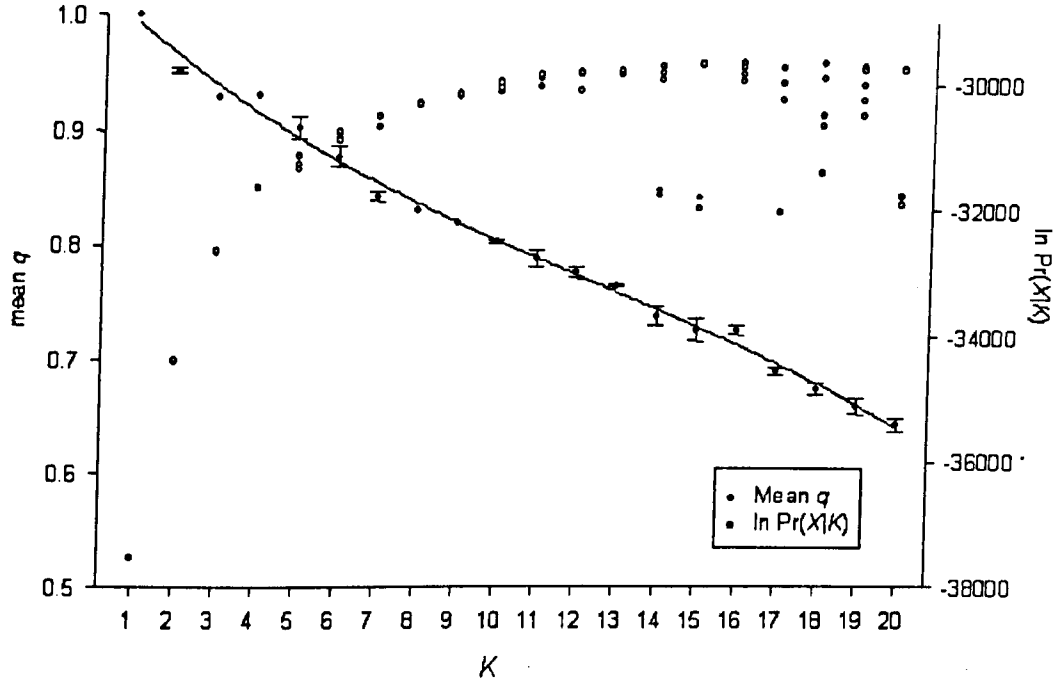


Figure 2.3. Likelihood plot of structure results (MCMC with 100 000 repetitions) (Pritchard *et al.*, 2000). Mean q is the mean confidence assignment of all individuals to their most likely cluster, indicating robustness of assignment (shown with standard error). ln Pr(X|K) is the log likelihood for each value of K, the number of simulated clusters. The most likely K is that where ln Pr(X|K) is maximised. The plot illustrates the difficulty in deciding on the most likely number of subpopulations in the dataset due to isolation by distance.

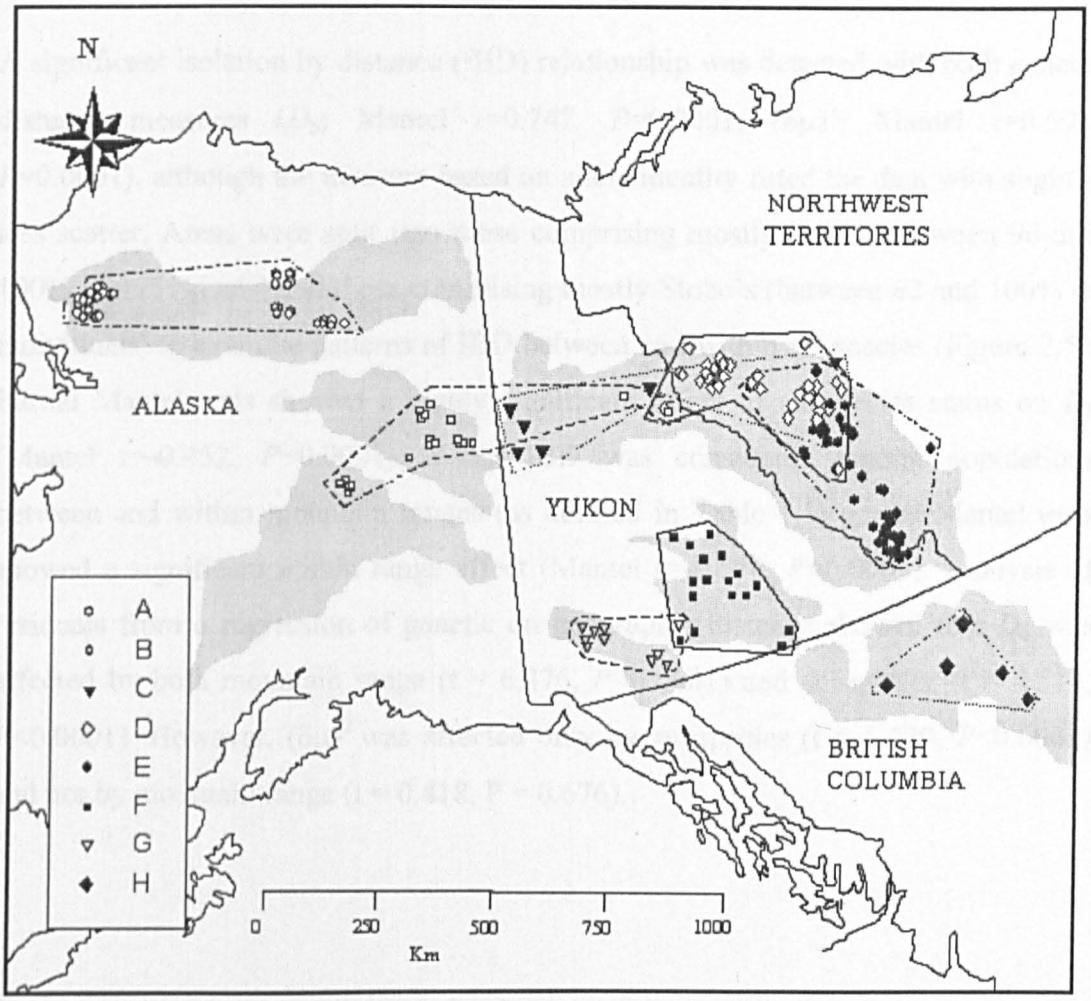


Figure 2.4. Geographic locations of genetically similar clusters of individuals as assigned by STRUCTURE at K=8 (individuals with q>0.9 only). Boundaries of each cluster are defined by lines surrounding all individuals within each. Due to the close sampling sites of some animals, one point on the map does not equal one sheep, but a variable number and is only intended as a guide to the geographical limitations of the cluster. Mountain range limitations are included in lighter shading to indicate obvious barriers to gene flow.

2.3.3. Isolation-by-distance

A significant isolation by distance (IBD) relationship was detected with both genetic distance measures (D_S: Mantel r=0.747, P=0.0001; $(\delta\mu)^2$: Mantel r=0.593, P=0.0001), although the measure based on allele identity fitted the data with slightly less scatter. Areas were split into those comprising mostly Dall's (between 90 and 100% of individuals) and those comprising mostly Stone's (between 62 and 100% of individuals) to examine patterns of IBD between and within subspecies (Figure 2.5). Partial Mantel tests showed a highly significant effect of subspecies status on D_S (Mantel r=-0.452, P=0.0001). When IBD was considered among populations between and within mountain ranges (as defined in Table 1.1) partial Mantel tests showed a significant within range effect (Mantel r=-0.570, P=0.0001). Analysis of residuals from a regression of genetic on geographic distance showed that D_S was affected by both mountain range (t = 6.376, P<0.0001) and subspecies (t = 4.511, P<0.0001). However, $(\delta\mu)^2$ was affected only by subspecies (t = 6.179, P<0.0001) and not by mountain range (t = 0.418, P = 0.676).

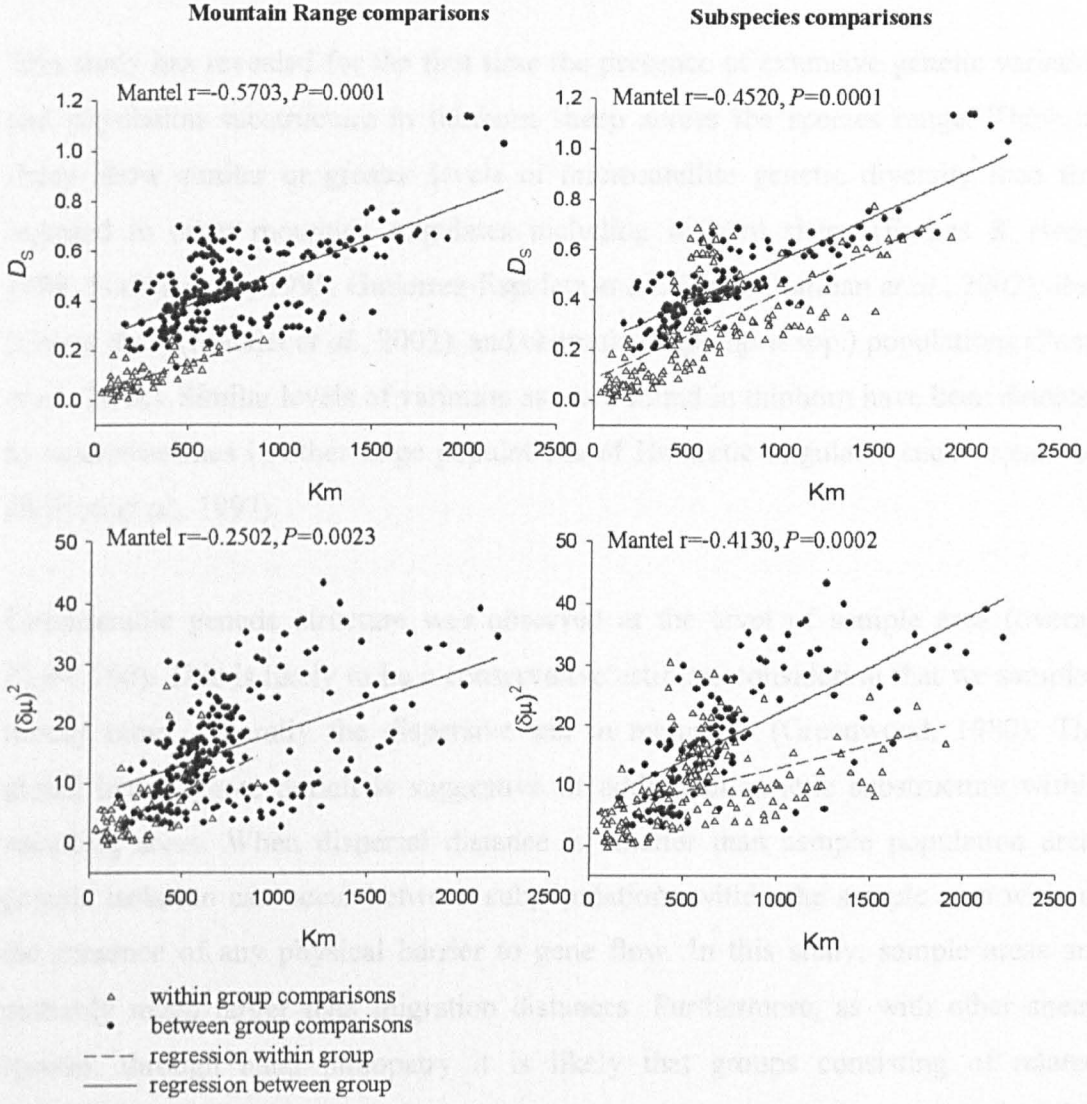


Figure 2.5. Isolation by distance relationships between population pairs (D_S and $(\delta\mu)^2$) by mountain group and putative subspecies. Two measures of genetic distance are plotted (D_S and $(\delta\mu)^2$). Both show significant isolation by distance (P=0.0001). Results of partial Mantel tests for the effect of mountain range and subspecies are shown.

2.4. DISCUSSION

This study has revealed for the first time the presence of extensive genetic variation and population substructure in thinhorn sheep across the species range. Thinhorn sheep show similar or greater levels of microsatellite genetic diversity than that reported in other mountain ungulates including bighorn sheep (Forbes & Hogg, 1999; Forbes *et al.*, 1995; Gutierrez-Espeleta *et al.*, 2000; Coltman *et al.*, 2002), ibex (*Capra ibex*) (Maudet *et al.*, 2002), and chamois (*Rupicapra* spp.) populations (Perez *et al.*, 2002). Similar levels of variation as were found in thinhorn have been detected by microsatellites in other large populations of Holarctic ungulates such as caribou (Wilson *et al.*, 1997).

Considerable genetic structure was observed at the level of sample area (overall F_{ST}=0.160). This is likely to be a conservative estimate considering that we sampled mostly rams, generally the dispersive sex in mammals (Greenwood, 1980). The global heterozygote deficit is suggestive of additional genetic substructure within sampling areas. When dispersal distance is smaller than sample population area, genetic isolation can occur between subpopulations within the sample area without the presence of any physical barrier to gene flow. In this study, sample areas are probably much larger than migration distances. Furthermore, as with other sheep species, through natal philopatry it is likely that groups consisting of related individuals inhabit adjacent ranges within each sample area (Coltman *et al.*, 2003b). The distribution of suitable habitat to specialists such as mountain sheep is also likely to be easily fragmented by forest encroachment or other environmental obstacles to contemporary migration. These factors have likely created a widespread "Wahlund effect" at the level of sample area.

Genetic distances between areas were of a similar magnitude to those reported previously in bighorn (Forbes & Hogg, 1999) and other mountain ungulates (Maudet *et al.*, 2002; Perez *et al.*, 2002) but higher than those reported from carnivores (Paetkau *et al.*, 1999; Kyle & Strobeck, 2002) over similar geographic distances.

This is expected due to the lower levels of migration and smaller home ranges characteristic of wild sheep populations. Genetic differentiation is evident between thinhorn sheep less than 40km apart show, a similar scale of differentiation to that reported in desert bighorn sheep using both mtDNA (Boyce *et al.*, 1999) and microsatellites (Gutierrez-Espeleta *et al.*, 2000).

Although I show the presence of fairly strong genetic differentiation and a robust pattern of isolation by distance among areas, the delineation of sampling areas may partly be an artefact of wildlife management and political boundaries. The Bayesian method (Pritchard *et al.*, 2000) is robust to bias introduced by sample area boundaries, as populations are not defined *a priori*. However, the results of the STRUCTURE analyses were not straightforward to interpret. This may be partly because the algorithm implemented by STRUCTURE is not well-suited for situations where there is isolation by distance (Pritchard & Wen, 2002). The traditional area-based analysis demonstrated significant differentiation between all but two pairs of areas, whereas STRUCTURE found inconsistent results with $K > 13$ regardless of the length of burn-in. The decision to choose the relatively small number of eight subpopulations based on higher values of q was supported by the clear geographic boundaries of these clusters. Moreover after K=8 further population subdivisions occurred within clearly defined groups of mountain blocks. The lower values of q at higher K reflect mixed membership in multiple groups, which is a likely consequence of finer scale isolation by distance relationships (Pritchard & Wen, 2002). The genetic subpopulation boundaries at K=8, which mostly correspond to mountain range topology, were also supported by the traditional measures of distance between pre-defined populations (Figure 2.1).

The magnitude of genetic structure varied greatly across the species range. The Mackenzie Mountains of NWT and the Ogilvie and Wernecke Mountains of the Yukon (areas Y2W and Y2E, respectively) form part of the same upland system, hence the grouping of ten sample areas into the Mackenzie range. Genetic distances between areas in the Mackenzie range were relatively low. This is a mostly

continuously inhabited area of the species range, and there may be higher rates of migration in this region due to the lack of major habitat breaks. There have been individual documented cases of young ram and nursery group sheep movement within the Mackenzie Mountain range (Simmons, 1982), although data on their fate are scarce. While the Mackenzie range was mostly weakly structured, the D/OT/02 sample area is an exception. The Nahanni National Park, where no hunting can take place, divides this area into two subunits. The park includes the large South Nahanni River with a rugged canyon that appears to present a major barrier to gene flow (Tables 2.6 and 2.8). This may explain the relatively large genetic distances between this area and the rest of the Mackenzie range. The presence of further fine-scale genetic structuring across the species range as a result of geographical features remains untested.

Traditional assignment tests reflect the differing magnitude in genetic structure across the species range. The high rate of assignment within the same eight regional groupings as defined by STRUCTURE analyses reinforces the importance of unsuitable valley habitats in limiting thinhorn gene flow. Tests reveal the present dataset to be ineffective for forensics cases where assignment is required beyond the level of mountain range with the possible exception of Alaskan and southern Yukon populations.

Thinhorn sheep population structure will also be influenced by patterns of range expansion following the recession of the Laurentide and Cordilleran ice sheets after the last Pleistocene glaciation (Sage & Wolff, 1986). At the last glacial maximum the southern Yukon and BC were under ice, whereas the Mackenzie Mountains, west-central Yukon and much of Alaska were largely unglaciated (Dyke *et al.*, 2002). The current BC populations may have therefore been founded by colonisation of a small number of Stone's sheep migrating southward from refugia in south-central Yukon. Alternately, BC thinhorn populations could be relicts from small ice-free refugia that existed in the northern Rocky Mountains in the seam between the Laurentide and Cordilleran ice sheets (Catto *et al.*, 1996). Either scenario is consistent with the lower

levels of observed genetic variation (Table 2.1) and genetic homogeneity (Figure 2.1) of BC thinhorn populations.

Higher levels of genetic variation and structure were evident in the southern Yukon sample areas. If these areas were mostly ice-free during the last period of glaciation, then higher levels of variation would be consistent with the relative age of these populations. Pronounced population structure and long-branch lengths separating these populations (Figure 2.1) may therefore reflect bottleneck-like effects of re-colonization from multiple refugia in this region. However, it may also be a consequence of reduced contemporary gene flow due to post-Pleistocene forest encroachment and more recent anthropogenic habitat fragmentation.

Alaskan areas were highly genetically differentiated. Genetic distance measures between Alaskan areas were greater than expected from geographic distance alone, rates of correct assignment were very high and there are very high q values for individuals in the two clusters based in the region (Figure 2.4). This is consistent with topographical features of the region that may limit gene flow. Distances through suitable sheep habitat in Alaska are larger than the linear distance between sampling locations due to large areas of unsuitable habitat, which create semi-insular populations. The Alaskan interior lowlands and the Yukon River valley separate the Central Alaska and Brooks ranges, leading to large genetic distances between the GAAR/NK and YUCH sample areas in Alaska (Table 2.2, Figure 2.1) and discrimination of these groups by STRUCTURE (Figure 4). The range covered by YUCH is relatively continuous with that of the West Ogilvie Mountains (Y2W), however, providing easier route of gene flow between YUCH and Y2W. This is also reflected in the STRUCTURE analyses, with spread of the mid Alaskan based cluster into northern Yukon. I also note that genetic distances may be upwards biased between Alaskan populations, and genetic diversity decreased as a result of the sampling procedure in these regions. Ewes collected close together are likely to be more related than the hunted ram samples from elsewhere. However, I suggest that

even with this factor, genetic distances between Alaskan ranges are higher than those between other mountain groups.

After accounting for subspecies and mountain range effects, thinhorn sheep population structure is well described by an isolation by distance pattern (Figure 2.5) which reflects hypothetical regional equilibrium between gene flow and drift (Hutchison & Templeton, 1999). The gradient of the overall isolation by distance plot is greater than many carnivore populations, but of a similar magnitude to bighorn sheep (Forbes & Hogg, 1999). This positive relationship is typically caused by limited dispersal. Wild sheep are highly philopatric with little migration from the natal region, and a strong association with winter range. Low rates of gene flow among thinhorn sheep populations are suggested by a large degree of population differentiation and the relatively steep gradient of the isolation by distance plot.

Differences in pelage colour and skull measurements delineate the classification of *O. d. dalli* and *O d. stonei* subspecies (Ramey, 1993). Genetic data also provide evidence of genetic differentiation between subspecies (Figure 2.5). Genetic distances between areas consisting of the same subspecies were smaller than distances between areas of different subspecies after correcting for geographic distance. However, genetic distances are quite variable and there is considerable overlap between the within and between subspecies comparisons, a trend also reported in bighorn sheep (Forbes & Hogg, 1999). Additional evidence of genetically differentiated subspecies comes from assignment and STRUCTURE analyses. A subspecies assignment plot (Figure 2.2) shows relatively little cross-assignment between the two groups. STRUCTURE analyses reveal that Stone's sheep from mixed sample areas cluster more with BC samples than with areas comprising pure Dall's. In these mixed areas sheep have lower assignment confidences at K=2 than in areas of one subspecies, reflecting their mixed ancestry. This provides evidence of a zone of introgression whereby 'Stone's' alleles have been introduced from the South. Genetic evidence is supported by pelage patterns with many sheep from southern Yukon showing an intermediate colour. These so called fannin sheep are presumed to

be the result of cross breeding between the two subspecies. In thinhorn therefore genetic evidence of subspecies status is concordant with pelage colour.

Figure 2.5 shows a greater effect of subspecies than mountain range on $(\delta\mu)^2$ than on D_S. As measures of allele size variance may better reveal distinctions between more deeply divergent populations (Forbes & Hogg, 1999; Forbes *et al.*, 1995; Paetkau *et al.*, 1997), this suggests that the genetic difference between subspecies is more ancient than the division resulting from mountain block. In all attempts to describe genetic differentiation between subspecies the geographic differences in genotype must be taken into account. This was done using general linear models of genetic distance with both mountain block and subspecies status as factors. The results support a more ancient division of subspecies due to the non-significance of mountain range in models of $(\delta\mu)^2$.

In summary, microsatellite analyses of thinhorn sheep populations have demonstrated considerable genetic structure across the species range. The pattern of genetic differentiation is broadly consistent with isolation by distance. However, subspecies differentiation, colonization from multiple glacial refugia and the effects of mountain range topology complicate this pattern. Microsatellite data provide some support for the subspecies status of Dall's and Stone's sheep.

CHAPTER 3. PHYLOGEOGRAPHY OF NORTH AMERICAN MOUNTAIN SHEEP

3.1. INTRODUCTION

3.1.1. Ice-age refugia

The distribution of many species that currently inhabit temperate and boreal environments has been significantly shaped by patterns of glaciation since the Quaternary period 2.4 million years ago (Hewitt, 2000). During periods of glaciation the landscape became highly fragmented, restricting populations within species to geographically isolated refugia (Hewitt, 2000; Hewitt, 1996). This process has been attributed as one of the major catalysts for the development of biodiversity in present temperate and arctic environments (Pielou, 1991; Willis & Whittaker, 2000). While isolation in major glacial refugia has been found to account for much genetic and morphological diversity (Ehrich *et al.*, 2000; Holder *et al.*, 2000; Brunhoff *et al.*, 2003; Flagsted & Røed, 2003; Dobeš *et al.*, 2004; Galbreath & Cook, 2004), very little evidence is available to assess the evolutionary importance of smaller refugia. Smaller refugia are often difficult to identify due to a lack of fossil evidence or sufficient detail from studies of glacial limits.

Although examples are rare, genetic evidence has helped in the verification of 'cryptic' refugial sites, and a reconstruction of their paleoecology. Plants of the *Packera* genus probably survived in southwestern Alberta, Canada in small ice-free areas between the Laurentide and Cordilleran ice sheets or in nunataks (small ice free 'islands' extending above the ice sheets) in the same area (Golden & Bain, 2000). Mitochondrial DNA (mtDNA) evidence supports the survival of the endemic Norwegian lemming *Lemmus lemmus* in Scandinavian refugia (Federov & Stenseth, 2002). There is strong evidence for the existence of ice-free areas in the northwest of the Canadian arctic archipelago since congruence can be found among molecular evidence for refugial populations of rock ptarmigan *Lagopus mutus* (Holder *et al.*,

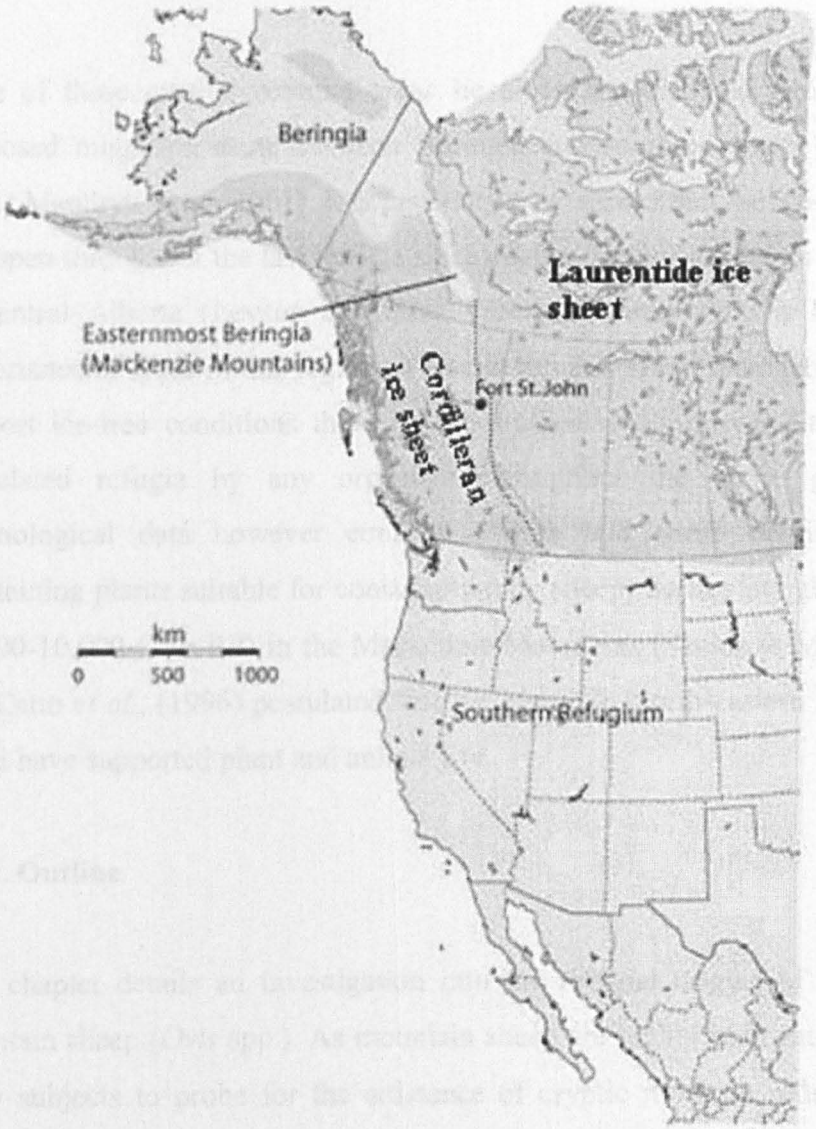
1999), the arctic plants *Dryas integrifolia* (Tremblay & Schoen, 1999) and *Saxifraga oppositifolia* (Abbot *et al.*, 2000). The collared lemming *Dicrostonyx groenlandicus* also survived in this region (Federov & Stenseth, 2002), which has been independently verified by a lemming parasite *Paranoplocephla arctica* (Wickström *et al.*, 2003).

Small populations may have their genetic signal of refugial origin erased if post-glacial migration from other refugial populations is great enough. This may explain the lack of signal from *Rangifer tarandus*, which was hypothesised to have survived glaciation in the Canadian arctic archipelago (Flagsted & Røed, 2003), and in *Saxifraga oppositifolia* which may have survived in northern Norway (Gabrielsen *et al.*, 1997). The signal of past refugial separation may be stronger in species with limited dispersal abilities (Federov & Stenseth, 2002).

3.1.2. Glacial history of regions currently inhabited by North American wild sheep

The species of thinhorn and bighorn sheep are recognised based on morphological evidence (Ramey, 1993; Cowan, 1940). Their divergence is currently thought to be a result of long-term separation by glacial activity, thinhorn in Beringia to the north of invading ice sheets and bighorn to the south (Figure 3.1) (Pielou, 1991; Geist, 1999). However, recent geological evidence shows that sheep currently inhabit four regions known to have been ice-free during Wisconsinan glaciations (a period lasting approximately from 70,000-10,000 years ago). In addition to the two major refugia of Beringia and southern North America, portions of the Mackenzie Mountains of the Northwest Territories (also known as 'Easternmost Beringia') also remained ice-free (Dyke & Prest, 1987; Duk-Rodkin & Hughes, 1991), as well as a region further south in northeastern British Columbia (Catto, 1996).

Figure 3.1. Glacial coverage of North America during glacial maximum (approximately 21,000 years ago). The two major refugia were found in Beringia and areas of southern North America. Easternmost Beringia was probably cut off from the rest of Beringia during glacial maximum (Catto, 1996). Catto *et al.*, (1996) found areas of unglaciated terrain near Fort St. John. Glacial limits are based on Mandryk *et al.,* (2001) and Duk-Rodkin (1999).



The glacial history of northern North America is complex. Easternmost Beringia was influenced by Laurentide and montane glaciation, the ice-free area appearing to have been effectively cut off from other parts of Beringia during much of the last ice age (Catto, 1996). Northeastern British Columbia was a crossroads for Laurentide and Cordilleran ice sheets and montane glaciers (Figure 3.1). From investigations near Fort St. John, British Columbia Catto *et al.*, (1996) found that coalesence of ice sheets did not occur during the last ice age. Instead a temporally and geographically shifting ice-free zone existed. In addition to this numerous isolated foothills of the Rocky Mountains remained ice-free during glacial maximum.

Both of these smaller refugial areas lie along the disputed 'ice-free corridor', a proposed migration route between Beringia and southern North America (Pielou, 1991; Mandryk *et al.*, 2001). It is generally now agreed that the ice-free corridor was not open throughout the last ice age since coalescence of ice sheets has been verified in central Alberta (Levson & Rutter, 1996), however little is known about the importance of areas in this region as glacial refugia. Other than geologic evidence to support ice-free conditions there are no studies to confirm habitation of the two postulated refugia by any organism throughout the entire glaciated period. Palynological data however confirm a herb and shrub dominated vegetation (containing plants suitable for consumption by sheep) during late glacial periods (ca. 12,000-10,000 C yr BP) in the Mackenzie Mountains (Szeicz & MacDonald, 2001) and Catto *et al.*, (1996) postulated that the region in North-eastern British Columbia could have supported plant and animal life.

3.1.3. Outline

This chapter details an investigation into the refugial origins of North American mountain sheep (*Ovis* spp.). As mountain sheep are highly philopatric they are ideal study subjects to probe for the existence of cryptic montane refugia. In addition, mountain sheep are superbly adapted to glacial conditions, and currently inhabit

areas within sight of large ice sheets (Geist, 1971). Specific hypotheses and aims are detailed below.

i) Divergence of bighorn and thinhorn sheep can be dated to glaciations during the last ice age. This finding would support the widely held belief that the two species evolved in isolation south and north of the ice sheets respectively. To investigate this hypothesis a region of mitochondrial DNA will be compared between individuals from across the range of both species, from which approximate dates of divergence will be calculated.

ii) Thinhorn sheep inhabited one major refugium in Beringia and two minor glacial refugia in Eastern Beringia (Mackenzie Mountains) and northern British Columbia. The current distribution of thinhorn sheep encompasses areas known to have remained ice-free during the height of the last ice-age (Beringia), in addition to two regions postulated to have been free from ice throughout this period and many regions covered by ice. Mitochondrial DNA will be analysed for evidence of support for smaller cryptic sheep refugia.

iii) Stone's sheep survived the glaciations in different refugia from those of northern Dall's sheep. Morphological differences between thinhorn sheep subspecies are hypothesised to have arisen in geographical isolation such as that provided by differential glacial refugia. I will use mitochondrial DNA to test for genetic differences between subspecies.

3.2. MATERIALS AND METHODS

3.2.1 Samples

Thinhorn sheep were sampled from eight regions across the species range (Figure 3.2), as identified in Chapter 2. Regions are based largely on mountain range limitations (Figure 3.2), a divide supported by differences in microsatellite genotype (Chapter 2). Three locations were sampled in Alaska (with sample region abbreviation and sample size given in parentheses); Central Alaska Range (AR, 3), Yukon Charley Rivers National Preserve (YU, 4), Brooks Range of northern Alaska (NA, 4). Five geographic regions were sampled in Canada; Ogilvie Mountains (OG, 9), Mackenzie Mountains, Northwest Territories (NT, 16), Pelly Mountains, central Yukon (CY, 18), southwest Yukon (SW, 12) and British Columbia (BC, 22).

Across sample regions there are major differences in ram colour morphology. All thinhorn sheep from Alaskan populations are predominantly white, while very few sheep in SW and NT have some grey colour on their flanks. Approximately 35-40% of OG sheep have dark colouration on their flanks, the rest of the population being completely white or white with dark tails. All CY sheep are dark in coloration, but are on average lighter than sheep in BC (Sheldon, 1911). If the mean value of darkness for these populations is calculated OG is lightest, followed by CY and BC is darkest. In all of these populations a great deal of individual variation is also prevalent, such that the darkest sheep in OG look similar to the lighter sheep in CY, and a similar relationship exists between CY and BC.

Bighorn sheep were sampled from two regions incorporating two subspecies; Rocky Mountain (*O. canadensis canadensis*) and desert bighorn (*O. c. nelsoni*). Eleven Canadian Rocky Mountain samples (OCC) were obtained from the Alberta Department of Wildlife. An additional sample from the same region was obtained from Genbank (accession number AY091486, (Hiendleder *et al.*, 2002). All desert bighorn samples (OCN, 124) were obtained from Genbank (accession numbers

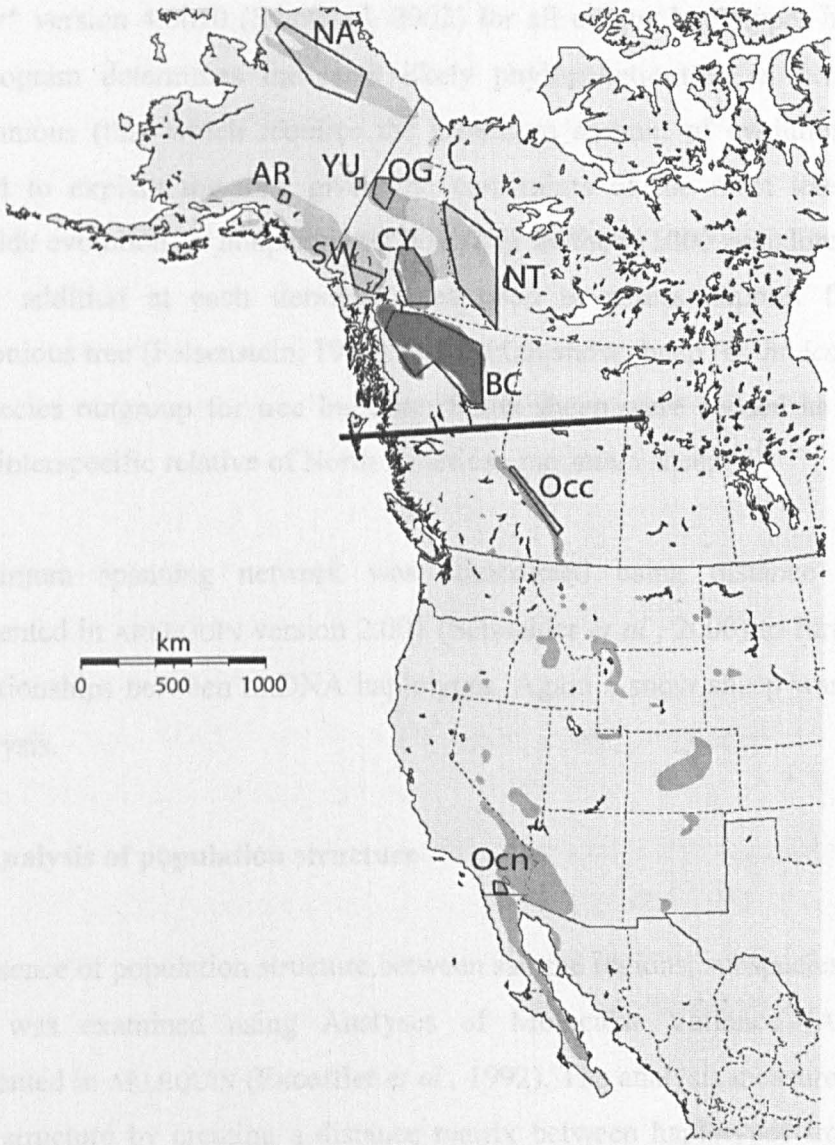
AF076911-AF076917, (Boyce *et al.*, 1999) and originated from southern California (Figure 3.2).

3.2.2. Molecular methods

As with the majority of thinhorn samples, Rocky Mountain bighorn samples comprised horn core samples stored dry in sealed paper envelopes. Genomic DNA extraction was carried out on these samples following the protocol outlined in Appendix A.

The human primers L15999 (5'-ACCATCAACACCCAAAGCTGA-3') and H16498 (5'-CCTGAAGTAGGAACCAGATG-3') were used to amplify a 604 base pair section of the sheep mitochondrial (mtDNA) control region. Polymerase chain reactions (PCR) were carried out in 50μl reactions containing 5μl DNA template, 0.4μM of each primer, 100μM dNTP's, 1.5mM $MgCl_2$ and 1 unit *Taq* polymerase (Bioline, London). PCR profiles comprised 32 cycles of 94°C denaturing for 30s, 50°C annealing for 30s, and extension at 72°C for 40s. Cycles were preceded by 3 min at 94°C and terminated with 5 min at 72°C. PCR products were gel purified (Qiagen, Crawley, West Sussex, UK) and quantified before sequencing. Purified PCR products were sequenced using primers as above and run on an ABI 3730. Editing and alignment was carried out using the software SeqScape (Applied Biosystems, Foster City, USA).

Figure 3.2. Distribution of mountain sheep (*Ovis* spp.) in North America and regions sampled for mitochondrial DNA. Shaded areas represent approximate boundaries of sheep habitation. The solid line divides putative species: thinhorn (*O. dalli*) to the north and bighorn (*O. canadensis*) to the south. For thinhorn sheep darkness of shaded areas represent relative darkness of coat pelage. All populations except OG, CY and BC are almost exclusively comprised of white coloured sheep (see methods).



3.2.3. Phylogeny reconstruction

Measures of genetic diversity (both haplotype and nucleotide) within sample regions were calculated in ARLEQUIN. MODELTEST version 3.06 (Posada & Crandall, 1998) was used to determine which model of nucleotide evolution was most likely for producing the sheep mitochondrial sequences present. The program determined that the Kimura three parameter model (Kimura, 1981) was the most likely model to account for nucleotide variation in the control region of mountain sheep mtDNA. Phylogenetic trees were constructed using neighbour joining methods implemented in PAUP* version 4.0b10 (Swofford, 2002) for all unique haplotypes in the dataset. The program determines the most likely phylogenetic tree as being the most parsimonious (that which requires the minimum amount of evolutionary change) required to explain the data given the constraints of the most likely model of nucleotide evolution. A nonparametic bootstrap method (1000 pseudoreplicates with random addition at each iteration) was used to assess support for the most parsimonious tree (Felsenstein, 1985). A Siberian snow sheep (*O. nivicola*) was used as a species outgroup for tree building. Snow sheep were chosen as they are the closest interspecific relative of North American mountain sheep.

A minimum spanning network was constructed using distance methods as implemented in ARLEQUIN version 2.001 (Schneider *et al.*, 2000) to further examine the relationships between mtDNA haplotypes. Again a snow sheep was included in the analysis.

3.2.4. Analysis of population structure

The presence of population structure between sample regions, subspecies and species groups was examined using Analyses of Molecular Variance (AMOVA) as implemented in ARLEQUIN (Excoffier *et al.*, 1992). The analysis measures population genetic structure by creating a distance matrix between haplotypes from which to measure hierarchical components of genetic diversity. As it measures haplotype

variation there is no variation within individuals. As hypotheses are tested using a permutation model no assumption of normality is required. Bighorn sheep subspecies comparisons were not made since sampling from bighorn sheep was less extensive. A previous detailed mitochondrial DNA analysis of bighorn subspecies has been made by Ramey (1993).

3.2.5. Calculations of divergence dates

Haplotype divergence times were calculated with the program MDIV (Nielsen & Wakely, 2001). The program employs a Bayesian framework using the Hasegawa-Kishino-Yano nucleotide substitution model (HKY 1985) to correct for multiple hits, producing an estimation of the times to most recent common ancestor (TMRCA) and divergence for any pair of populations. These estimates differ since haplotypes can begin to diverge before a single population diverges into two or more populations, and hence TMRCA can exceed the time of population divergence. Using both estimates together it is possible to estimate the time that haplotypes began to diverge within populations in addition to the time when populations separated or experienced secondary contact.

To calculate the time since the divergence of haplotypes a molecular clock of 24% per locus per million years was used. This is the rate which has been estimated for domestic and wild sheep from the complete length of the mtDNA control region (Hiendleder *et al.*, 2002). Since the control region evolves too rapidly for comparison based on fossil evidence of separate species, an indirect method was used to generate this figure. Hiendleder *et al.*, (2002) first estimated the protein-coding sequence distance between the most divergent lineages of sheep (*O. canadensis* and *O. aries*). This was then referenced against a molecular divergence of the complete mtDNA genome for cow and sheep, and then further referenced using fossil evidence of a divergence of the *Ruminantia* and *Cetacea* 60 million years ago. A more conservative estimate of divergence was considered using a much faster mutation

rate estimate of 38% per locus per million years, the rate estimated using cattle (*Bos* spp.) mtDNA from a 240-bp portion of the control region (Troy *et al.*, 2001).

3.2.6. Population growth

To probe for evidence of population expansion Tajima's *D* (Tajima, 1989) and Fu's *F* test (Fu, 1997) statistics were calculated. Both Tajima's *D* and Fu's *F* statistics measure the deviation of allele frequencies from the distribution expected under neutrality; *D* from the number of nucleotide polymorphisms between sequences, *F* from the number of derived nucleotide variants between sequences observed only once in a sample. Both selection and demography result in deviations from neutrality. Here I am examining variation in mtDNA and can assume no selection is present. Therefore deviation from the null model can be attributed to non-equilibrium population demographics. Population growth can be inferred through an excess of rare alleles, reflected in $D/F<0$ while population bottlenecks can be inferred from $D/F>0$ reflecting an excess of alleles of an intermediate frequency.

More in depth analysis of population expansion was carried out using mismatch distribution analysis (MDA) using ARLEQUIN. MDA produces an age expansion parameter (τ) which is a relative measure of the time (in generations) since population expansion and is useful to date the initiation of rapid population growth (Rogers & Harpending, 1992; Schneider & Excoffier, 1999). The MDA parameters θ_0 and θ_1 were also calculated. These are relative to the expected values of haplotype differences within populations at the start and end of a model of sudden population expansion. These values are then compared to actual values from the data to test the significance of any evidence for population expansions from thinhorn and bighorn sheep haplotypes.

3.3. RESULTS

3.3.1. Measures of sequence diversity

From the total of 223 thinhorn and bighorn mtDNA sequences there were 64 unique haplotypes defined by a total of 90 polymorphic sites. This included 84 transitions and 6 transversions. The mean number of pairwise differences between sequences was 17.31 (SD±7.72) and nucleotide diversity was 0.035 (SD±0.017). Population specific results of diversity are presented in table 3.1. Within thinhorn sheep, both the least and most variable haplotype diversity was present in Alaskan regions. Every individual sequenced in the Brooks and Central Alaskan Ranges had a unique mtDNA haplotype, whereas the YU haplotype was fixed in all individuals. Low haplotype diversity extended into the Ogilvie Range (OG). There was otherwise a wide range of diversities present between sample regions. Sequence diversities can not be compared directly due to the different sample sizes from each region, but it would appear that sequence diversity for bighorn sheep lies within the range of that seen within thinhorn regions. In contrast to high haplotype diversities in Alaskan regions, the highest nucleotide diversity was observed in the Central Yukon.

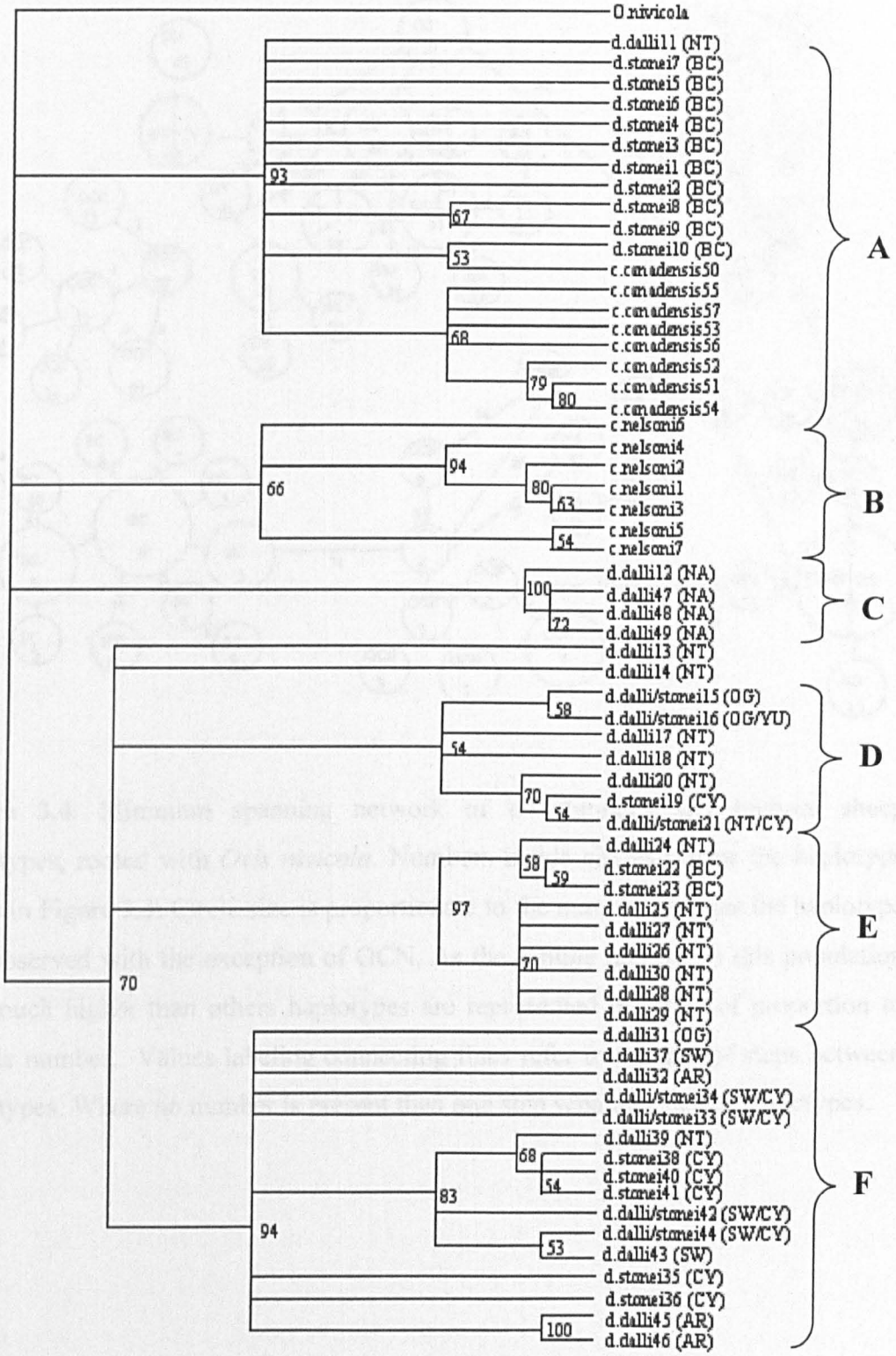
Table 3.1. Genetic diversity for eight thinhorn and two bighorn sheep populations based on mtDNA control region. Number of samples (N), number of unique haplotypes, haplotype diversity (*h*), nucleotide diversity (π), number of segregating sites (S) (* transversion was observed in the sample region) and mean number of pairwise differences (*d*) between haplotypes.

	AR	**YU**	**NA**	**OG**	**NT**	**CY**	**SW**	**BC**	**OCC**	**OCN**
N	3	4	4	9	16	18	12	22	11	124
Unique haplotypes	3	1	4	3	14	12	6	12	8	7
Haplotype diversity *h*	1.000	0	1.000	0.722	0.975	0.948	0.894	0.896	0.946	0.773
(±95%CI)	(0.272)		(0.177)	(0.097)	(0.035)	(0.033)	(0.054)	(0.045)	(0.054)	(0.025)
Nucleotide diversity π	0.017	0	0.004	0.020	0.024	0.027	0.011	0.013	0.008	0.013
(±95%CI)	(0.014)		(0.003)	(0.011)	(0.013)	(0.014)	(0.006)	(0.007)	(0.005)	(0.007)
Number of segregating sites S	11*	0	3	22*	47*	35*	14	39	11*	20*
Mean number of pairwise	7.33	0	1.67	10.89	11.24	12.69	5.14	6.97	3.02	6.56
differences *d* (SD)	(4.73)		(1.22)	(5.47)	(5.39)	(6.00)	(2.68)	(3.40)	(1.70)	(3.12)

3.3.2. Mountain sheep phylogeny

There was extensive polyphyly between thinhorn and bighorn sheep from both a haplotype tree (Figure 3.3) and minimum spanning network (Figure 3.4) Most mitochondrial haplotypes of sheep sampled can be placed into one of six groups on a phylogenetic tree (groups A to F Figure 3.3). Only two of these groupings comprised samples from only one subspecies of bighorn or thinhorn, group B being solely desert bighorn, group C entirely Dall's sheep. All Rocky Mountain bighorn cluster closely with BC Stone's sheep (Group A). The remaining groupings comprise only thinhorn sheep from Alaskan, Yukon and Northwest Territory sample regions. Figure 3.4 again shows groupings of similar haplotypes. In this figure there appears to be approximately seven groups mostly mirroring those present in the previous figure, the extra group coming from the division of tree group A into bighorn and BC thinhorn.

Figure 3.3. (overleaf) Consensus phylogram of thinhorn and bighorn haplotypes. Haplotypes are identified according to species and subspecies followed by the haplotype number and in parentheses the population of origin as per Figure 3.1.


O.nivicola
d.dalli11 (NT)
d.stonei7 (BC)
d.stonei5 (BC)
d.stonei6 (BC)
d.stonei4 (BC)
d.stonei3 (BC)
d.stonei1 (BC)
d.stonei2 (BC)
d.stonei8 (BC)
d.stonei9 (BC)
d.stonei10 (BC)
c.canadensis50
c.canadensis55
c.canadensis57
c.canadensis53
c.canadensis56
c.canadensis52
c.canadensis51
c.canadensis54
c.nelsoni6
c.nelsoni4
c.nelsoni2
c.nelsoni1
c.nelsoni3
c.nelsoni5
c.nelsoni7
d.dalli12 (NA)
d.dalli47 (NA)
d.dalli48 (NA)
d.dalli49 (NA)
d.dalli13 (NT)
d.dalli14 (NT)
d.dalli/stonei15 (OG)
d.dalli/stonei16 (OG/YU)
d.dalli17 (NT)
d.dalli18 (NT)
d.dalli20 (NT)
d.stonei19 (CY)
d.dalli/stonei21 (NT/CY)
d.dalli24 (NT)
d.stonei22 (BC)
d.stonei23 (BC)
d.dalli25 (NT)
d.dalli27 (NT)
d.dalli26 (NT)
d.dalli30 (NT)
d.dalli28 (NT)
d.dalli29 (NT)
d.dalli31 (OG)
d.dalli37 (SW)
d.dalli32 (AR)
d.dalli/stonei34 (SW/CY)
d.dalli/stonei33 (SW/CY)
d.dalli39 (NT)
d.stonei38 (CY)
d.stonei40 (CY)
d.stonei41 (CY)
d.dalli/stonei42 (SW/CY)
d.dalli/stonei44 (SW/CY)
d.dalli43 (SW)
d.stonei35 (CY)
d.stonei36 (CY)
d.dalli45 (AR)
d.dalli46 (AR)
A
B
C
D
E
F

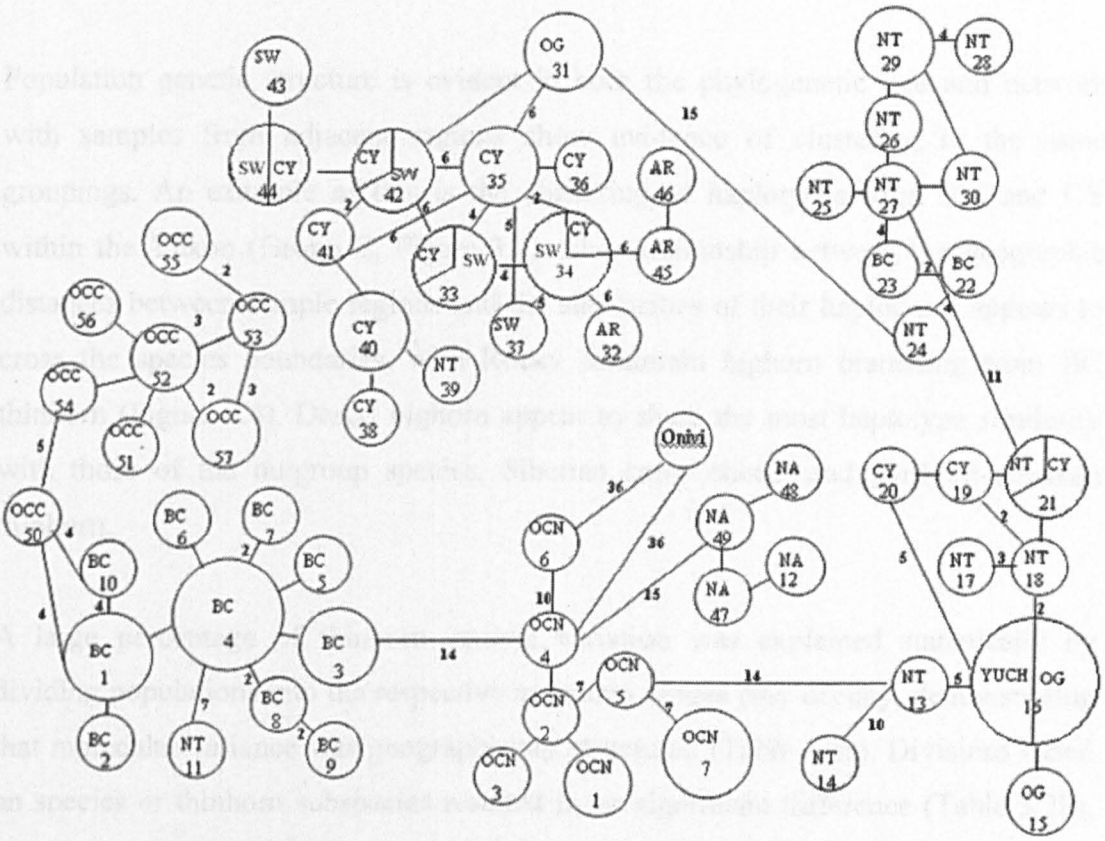


Figure 3.4. Minimum spanning network of 64 thinhorn and bighorn sheep haplotypes, rooted with *Ovis nivicola*. Numbers inside circles mirror the haplotype labels in Figure 3.3. Circle size is proportionate to the number of times the haplotype was observed with the exception of OCN. As the sample number in this population was much higher than others haplotypes are represented in terms of proportion to sample number. Values labeling connecting lines refer to number of steps between haplotypes. Where no number is present then one step separates linked haplotypes.

3.3.3. Population structure

Population genetic structure is evident in both the phylogenetic tree and network with samples from adjacent regions show evidence of clustering in the same groupings. An example of this is the clustering of haplotypes from SW and CY within the Yukon (Group E, Figure 3.3). The relationship between the geographic distances between sample regions and the similarities of their haplotypes appears to cross the species boundaries, with Rocky Mountain bighorn branching from BC thinhorn (Figure 3.4). Desert bighorn appear to share the most haplotype similarity with those of the outgroup species, Siberian snow sheep, and northern Alaskan thinhorn.

A large percentage of thinhorn genetic variation was explained statistically by dividing populations into the respective mountain ranges they occupy, demonstrating that molecular variance was geographically structured (Table 3.2a). Divisions based on species or thinhorn subspecies resulted in no significant difference (Table 3.2b), rather there were significant genetic differences within species groups and sample regions. Pairwise genetic distance (Φ_{ST}) revealed significant differentiation of all regions except the samples groups OG/AR, CY/AR and OG/YU (Table 3.3). Most of these non-significant comparisons came from regions of adjacent mountain ranges.

3.3.4. Divergence times between regions and species

The shortest divergence times between pairwise groups were present between the thinhorn regions of OG and YU and CY and SW, estimated at around 6,500 years and 11,300 years respectively (Table 3.4). These results were in agreement with Φ_{ST} results (Table 3.3) showing no significant genetic distances between the same regions. The longest time to group divergence was recorded between the thinhorn regions of NA and SW at 379,000 years. Divergence within bighorn subspecies was approximately 142,000 years ago.

Table 3.2a. Analysis of molecular variance (AMOVA) of thinhorn and bighorn sheep sample regions based on mountain range and species.

Molecular variance of thinhorn and bighorn sheep					
Criteria for division	**Components**	**% of total variation**	**Fixation indices**		***p***
Mountain range	Among sample regions (i.e. mountain ranges)	67.5	Φ_{ST}	0.67	<0.001
	Within sample regions	32.5			
Species	Among species groups	9.4	Φ_{CT}	0.09	0.18
	Among sample regions within species groups	58.9	Φ_{SC}	0.65	<0.001
	Within sample regions	31.7	Φ_{ST}	0.63	<0.001

Table 3.2b. AMOVA of thinhorn sheep across sample regions based on mountain range and subspecies.

Molecular variance of thinhorn sheep only					
Criteria for division	**Components**	**% of total variation**	**Fixation indices**		***p***
Mountain range	Among sample regions (i.e. mountain ranges)	67.5	Φ_{ST}	0.67	<0.001
	Within sample regions	32.5			
Subspecies	Among subspecies groups	-11.1	Φ_{CT}	-0.11	0.18
	Among sample regions within subspecies groups	69.8	Φ_{SC}	0.63	<0.001
	Within sample regions	41.3	Φ_{ST}	0.53	<0.001

Table 3.3. Pairwise Φ_{ST} for thinhorn and bighorn sample regions calculated using ARLEQUIN. Significance of genetic distances are indicated (*p<0.05, ** p<0.001, ^{ns} p>0.05).

	AR	YU	NA	OG	NT	CY	SW	BC	OCC
YU	0.845*								
NA	0.868*	0.959*							
OG	0.396^{ns}	0.107^{ns}	0.664*						
NT	0.443**	0.279*	0.625**	0.188*					
CY	0.151^{ns}	0.504*	0.641**	0.281**	0.373**				
SW	0.428**	0.826**	0.858**	0.576**	0.574**	0.085*			
BC	0.762**	0.750**	0.745**	0.663**	0.610**	0.675**	0.800**		
OCC	0.867**	0.892**	0.874**	0.741**	0.679**	0.701**	0.863**	0.462**	
OCN	0.782**	0.707**	0.678**	0.676**	0.663**	0.708**	0.781**	0.637**	0.681**

There was a confused picture of divergence between species with the shortest time between OCC and thinhorn being the BC group at 68,500 years while that between the more geographically distant OCN and thinhorn was 167,000 years between haplotypes of the distant NA thinhorn region, of a similar magnitude to the divergence date within bighorn subspecies. To be conservative the divergence time was also calculated using a much faster mutation rate calculated for cattle (Troy *et al.*, 2001). The faster mutation rate resulted in a divergence time of 43,305 years before present between OCC and BC thinhorn groupings.

Using the mutation rate of 24% calculated for sheep the estimated Time to Most Recent Common Ancestor (TMRCA) for BC and OCC was about 457,000 (Table 3.4). A similar TMRCA was found for most populations surveyed with the comparison AR/SW showing the least time at approximately 144,500. An anomalous finding was a relatively short TMRCA for populations at the opposite ends of mountain sheep distribution. Thinhorn sheep from northern Alaska and desert bighorn sheep from the southern United States shared a common ancestor around 253,000 years ago. This mirrors the haplotype network, where desert bighorn haplotypes showed the least number of steps to Siberian snow sheep.

In all but one pairwise comparison the TMRCA exceeded that of the group divergence date. In other comparisons the difference between these dates ranged from 14,000 years between these dates (AR/YU) to almost 389,000 between OCC and BC thinhorn.

Table 3.4. Estimates of divergence and TMRCA between pairwise sample regions. Below diagonal: MDIV estimates of divergence times for population pairs. Numbers in parentheses are 95% confidence intervals of the estimate. Divergence times before present are presented below credibility intervals. Above diagonal: MDIV standardised estimate and the number of years before present calculated for most recent common ancestor. The times in years were obtained based on a mutation rate for the control region (24% per million years) and an average generation time of 3 years.

	AR	YU	NA	OG	NT	CY	SW	BC	OCC	OCN
AR	-	4.51 237,245	4.88 376,167	4.54 283,750	1.96 309,823	3.37 333,638	2.09 144,504	3.66 438,914	4.90 430,026	2.72 423,583
YU	4.24 (0.76-9.74) 223,042	-	9.50 239,974	8.83 253,029	2.23 311,152	4.57 308,237	5.46 277,266	3.87 349,711	5.6 272,708	2.77 282,771
NA	4.18 (0.64-9.74) 322,208	9.59 (2.10-34.2) 242,247	-	5.45 319,336	1.97 348,341	3.81 402,332	6.43 426,155	3.97 407,855	4.86 294,258	2.27 253,306
OG	0.97 (0.35-4.88) 60,625	0.23 (0.06-1.95) 6,531	4.16 (0.76-9.74) 243,750	-	2.34 297,694	4.40 326,070	4.93 296,750	3.97 402,040	5.1 348,633	3.25 341,504
NT	1.16 (0.28-4.87) 183,365	0.28 (0.19-4.86) 39,156	1.21 (0.34-4.86) 213,956	0.30 (0.17-1.94) 38,644	-	2.74 309,580	2.40 301,687	2.64 399,011	2.54 355,865	1.55 341,283
CY	0.81 (0.38-4.85) 77,836	3.86 (0.58-11.7) 260,349	2.08 (0.52-4.86) 219,646	0.95 (0.45-4.87) 70,274	1.39 (0.43-4.73) 156,803	-	5.77 355,468	3.75 445,649	3.97 411,991	3.05 427,358
SW	1.32 (0.33-4.85) 91,266	4.16 (0.84-9.74) 211,250	5.72 (1.12-9.76) 379,099	1.72 (0.69-4.88) 103,729	1.32 (0.48-4.86) 165,998	0.18 (0.00-1.86) 11,336	-	4.39 463,687	5.74 434,237	3.90 418,438
BC	1.87 (0.55-4.84) 224,254	1.06 (0.46-4.86) 95,786	2.14 (0.64-9.70) 219,852	1.74 (0.56-4.85) 235,059	0.32 (0.28-4.86) 48,434	1.19 (0.58-4.81) 141,501	2.10 (0.79-4.84) 221,891	-	4.99 457,417	2.78 378,268
OCC	3.87 (0.90-7.79) 339,633	4.7 (0.86-7.82) 228,880	4.30 (0.35-5.87) 260,352	4.08 (0.43-7.78) 278,906	1.93 (0.22-2.93) 270,401	3.05 (0.79-5.82) 316,517	4.48 (0.62-7.81) 338,917	0.75 (0.36-1.94) 68,567	-	1.86 256,477
OCN	1.94 (0.20-4.92) 302,115	7.84 (4.1-19.24) 177,625	1.5 (0.54-9.72) 167,383	2.12 (0.82-5.86) 222,766	1.10 (0.62-2.60) 242,201	1.57 (0.84-9.68) 219,964	2.86 (1.10-9.70) 306,854	1.34 (0.56-4.18) 182,331	1.03 (0.22-4.86) 142,027	-

3.3.5. Evidence of population growth

Tajima's *D* test of neutrality produced no significant evidence of population growth although a trend (*D*=-1.37, *P*=0.06) was evident in the BC sample region that hinted of population growth (Table 3.5). Fu's *F* test is a more sensitive detector of population growth (Fu, 1997) and showed significant evidence of population expansion in the thinhorn sample regions NT, NA as well as the bighorn sample region OCC (Table 3.5).

Mismatch distribution analysis revealed evidence of population expansion in the same four regions as identified above by Tajima and Fu's tests of neutrality (Figure 3.5). In the cases of NA and BC this expansion appears to have occurred relatively recently. In addition to these regions, AR and SW also showed limited evidence of population growth not identified by former tests. In contrast OCN clearly showed no growth pattern in all tests, with highly significant evidence of stability ($P<0.001$). The thinhorn region OG showed almost significant evidence of population stability; θ_0 and θ_1 (used to estimate population size at start and end of growth) were identical, and no clear growth wave was evident in the graph (Figure 3.5). The difference in θ_0 and θ_1 for CY was small also indicating little population growth.

Table 3.5. Indicators of historical population demography are shown by Tajima's *D*, Fu's *F* together with significance values. The demographic parameters τ, θ_0, and θ_1 are also given together with 95% confidence intervals. $P(SSD_{obs})$ measures the probability of deviation from a hypothesis of sudden expansion according to the mismatch distribution analysis. Years since expansion initiation are estimated from τ. No statistics were calculated for YU as all haplotypes within this region were identical.

	AR	NA	OG	NT	CY	SW	BC	OCC	OCN
Tajima's *D* (*P* value)	0.00 (0.67)	0.17 (0.73)	1.71 (0.99)	-0.83 (0.20)	0.10 (0.87)	0.46 (0.73)	-1.37 (0.06)	-0.85 (0.23)	2.17 (0.99)
Fu's *F* (*P* value)	0.81 (0.41)	-2.18 (0.01)	7.75 (0.99)	-3.85 (0.03)	0.07 (0.48)	1.08 (0.72)	-0.82 (0.40)	-2.97 (0.02)	11.42 (0.99)
τ	11.72 (6.49-30.72)	1.84 (0.00-3.80)	7.00 (2.53-11.25)	11.83 (4.91-26.83)	3.00 (0.98-25.58)	7.55 (3.82-12.38)	2.00 (0.00-5.07)	3.21 (1.27-4.24)	12.27 (5.02-23.26)
θ_0	0.00 (0.00-17.00)	0.00 (0.00-1.35)	2.10 (0.00-2.84)	4.86 (0.00-11.50)	12.32 (0.00-29.2)	0.00 (0.00-2.16)	1.46 (0.00-7.70)	0.00 (0.00-2.61)	0.00 (0.00-10.48)
θ_1	841.09 (24.06 7013.6)	3776.3 (3761.7-6946.3)	2.10 (0.22-59.13)	21.60 (12.41-191.19)	14.61 (3.31-310.61)	13.22 (9.04-3301.97)	38.71 (4.80-6773.3)	5010.60 (32.70-9225.63)	21.23 (16.49-156.40)
$P(SSD_{obs})$	0.29	0.44	0.07	0.59	0.19	0.41	0.11	0.16	<0.001
Years since expansion	42,825	6724	25,584	43,238	10,964	27,595	7309	11,732	44,846

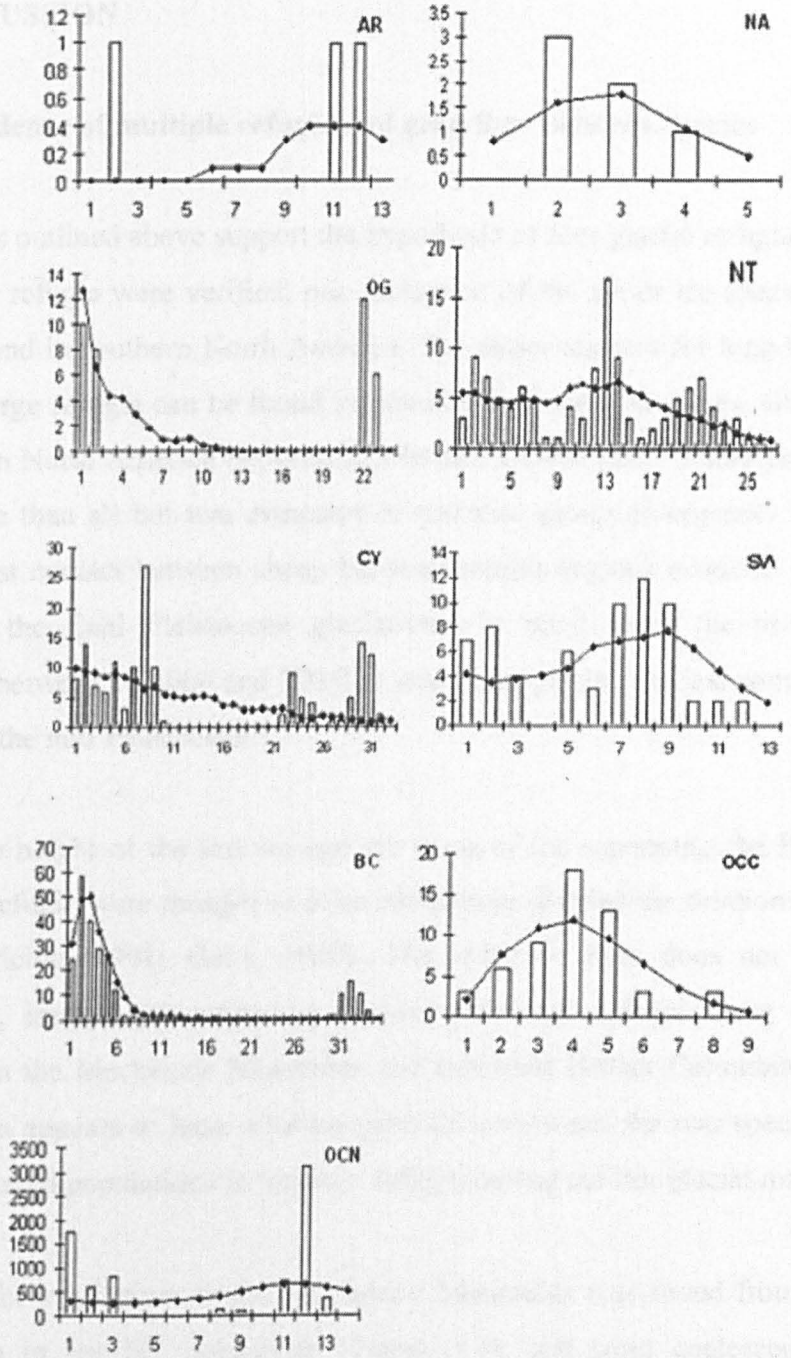


Figure 3.5. Mismatch distribution for sequences from thinhorn and bighorn sheep populations. Bars represent observed values and lines expected values for a model of sudden population expansion. Y axis: number of pairs, X axis: number of pairwise differences between haplotypes.

3.4. DISCUSSION

3.4.1. Evidence of multiple refugia and gene flow between species

The results outlined above support the hypothesis of four glacial refugia. As expected two major refugia were verified; one northwest of the major ice sheets in Beringia, and a second in southern North America. The major support for long-term isolation in these large refugia can be found in population divergence times. Glacial maxima occurred in North America between 22,000 and 14,000 years before present, a more recent date than all but two estimates of pairwise group divergence. This suggests that the last contact between sheep between sample regions occurred prior and not following the final Pleistocene glaciations. In most cases the divergence was estimated between 100,000 and 350,000 years ago, placing the last contact of sample regions to the mid Pleistocene.

During the height of the last ice age the mass of ice separating the Beringian and Southern refugia were thought to have completely divided the thinhorn and bighorn species (Pielou, 1991; Geist, 1999). The evidence here does not support this hypothesis, instead the existence of two additional refugia along the 'ice-free corridor' in the Mackenzie Mountains and northeast British Columbia is probable. This region appears to have allowed gene flow between the two species, and may have harboured populations in 'cryptic' refugia during the last glacial maximum.

Evidence for a refugium in the Mackenzie Mountains was found from a divergent haplogroup in the NT population (Figure 3.4), and from coalescence times of adjacent populations that predate glacial maximum. Contact between NT, CY, OG and BC appears to have happened before glacial maximum (Table 3.4), and geological evidence supports the existence of an ice-free area (Figure 3.1) in this region suitable for sheep habitation. Since divergence estimates can only be approximate there is still a possibility that gene flow also occurred after glacial maximum. Both mtDNA and microsatellite evidence suggests that gene flow has

occurred between OG and NT (Worley *et al.*, 2004). Although this refugium may not have been completely separated from Beringia, it is evident that it has served to allow considerable differentiation of haplotypes within it, supporting a hypothesis of continual long-term habitation of the Mackenzie Mountains (Figures 3.3 and 3.4).

The BC thinhorn sheep population also had divergence times with adjacent populations that predated glacial maximum. Interestingly, this population appears to be a product of hybridisation between species since thinhorn and bighorn haplotypes were present. In fact most haplotypes in this thinhorn sheep population were more similar to a bighorn sheep haplotype found in the southern Canadian Rocky Mountains than to thinhorn sheep. This finding contradicts Geist's (1971) hypothesis of thinhorn range expansion from a Beringian refugium into British Columbia. A second explanation may be that the BC population was colonised by bighorn haplotypes after ice sheets retreated. Evidence against this comes from an estimated divergence time between BC thinhorn and bighorns that predates glacial maximum by about 50,000 years. Even using the much more conservative mutation rate of 38% per locus per million years (Troy *et al.*, 2001) the divergence estimate predated glacial maximum by over 20,000 years.

Colour morphology also suggests that a post-glacial colonisation of BC by bighorn sheep is unlikely. Thinhorn in BC share similar coat pelage with sheep further north in CY and both of these populations differ greatly from the coat colour of bighorn sheep. If BC were initially colonised by bighorn sheep then the following scenario would have occurred. After ice retreated bighorn sheep occupied the region first. When populations were established, sheep from CY migrated south and interbred with bighorn sheep resulting in a change in morphology to resemble CY thinhorn sheep. However, post Wisconsinan glaciation gene flow between these populations is generally not supported. No CY haplotypes were found in BC, and nuclear DNA microsatellite evidence also showed a high degree of population structure between BC and CY suggestive of limited gene flow (Worley *et al.*, 2004). It is still possible that small number of migrants introduced the present coat colour to the BC

population, and a strong selective advantage lead to it spreading throughout the population.

A third explanation is most congruent with the evidence available. Although parts of the range occupied by BC thinhorn remained ice-free during glacial maximum (and possibly the entire Wisconsinan glaciation) (Catto *et al.*, 1996) no large ice free areas have been identified in the southern Canadian Rocky Mountains currently inhabited by bighorn (Levson & Rutter, 1996). Mitochondrial DNA from bighorn sheep support a scenario in which the southern Canadian Rockies were recolonised after ice retreated (Luikart & Allendorf, 1996). Thus a probable explanation for the pattern we have found is that contact occurred between bighorn and thinhorn prior to glacial maximum. As ice advanced the BC thinhorn population survived glaciation in small ice-free areas described by Catto *et al.*, (1996), while bighorn populations in the Canadian Rockies were overcome by ice and went extinct. The star shaped pattern in the haplotype network for BC (Figure 3.4) indicates a possible population bottleneck (Slatkin & Hudson, 1991). Indicators of nuclear genetic diversity were also relatively low for BC populations, lending some support to a bottleneck hypothesis (Worley *et al.*, 2004). Results should be interpreted with some caution since extinction events, genetic drift or a selective sweep may have altered the composition of haplotypes in populations since the end of the ice ages and could confound the results.

3.4.2. Historic population demography

During the last ice age overall habitat available to sheep was limited by the extent of ice sheets. With the retreat of ice new territory was made available allowing some populations to expand. Evidence of rapid population growth was conflicting with the use of different detection methods. Tajima's D statistic identified only BC as showing evidence of population growth while the more powerful Fu's F statistic identified three regions (OCC, NA, and NT) but not that identified by the former statistic. The cause of this conflict is unknown as both tests use similar methods to identify non-neutrality. All four regions identified as having evidence of population

expansion had extensive glacial coverage of the habitat they now occupy (Figures 3.1 and 3.2). Initiation of population expansion was roughly dated to the time period during and following glacial maximum (22,000-14,000 years ago) for all regions experiencing growth. The exception is NT, which appears to have experienced expansion prior to glacial maximum (Table 3.5). This may hint that glacial maximum occurred earlier in this region than in others. Temporal estimation of expansion events based on a single locus must however be interpreted with caution (Hillis *et al.*, 1996).

The signal of population growth was relatively weak for CY, SW and AR. This is surprising since much of the current habitat occupied by these populations was apparently covered by the Cordilleran ice sheet (Figure 3.1). One possible explanation for this is that sheep in these regions survived the last ice age in areas adjacent to ice sheets, losing habitat at the same time as it was gained with the melting of glaciers and therefore resulting in no net increase in population size. There was no evidence for rapid population growth for OG. The MDA graph for OG showed no signs of a wave in the distribution of pairwise genetic differences that would indicate rapid population growth (Figure 3.5). This result seems to be consistent with the lesser effect of glaciation in this region, where ice covered only a portion of the Ogilvie range. Finally, rapid post glacial population growth for OCN could be ruled out, a result which is not surprising since the southern United States remained unglaciated (Table 3.5, Figures 3.1 and 3.5).

Significant population structure was observed in all populations except for the pairs OG/YU, SW/CY, CY/AR and an exceptionally low Φ_{ST} value was observed for SW/CY (Table 3.3). In these same population pairs significant population structure was found from microsatellite analysis. It seems likely that at least in the case of OG/YU and SW/CY non-significant mtDNA results are caused by contact as glaciers retreated. Divergence estimates of 7,000 and 11,000 years respectively were recorded for these population pairs, and the estimate for rapid expansion initiation was about 11,000 years ago for CY (Table 3.4). The CY/AR non-significant result may also be

explained by colonisation of CY by some AR haplotypes, however it may also be partly an artefact of low sample size from AR since divergence was estimated at 78,000 years.

Comparing differences between the magnitude of gene flow revealed by mtDNA and microsatellites enables inferences to be made relating to the extent of sex-biased migration. Gene flow between thinhorn sample regions is evident from both mtDNA and microsatellites (Chapter 2). Migration rates between sample regions estimated from maternally inherited mtDNA are on average five times higher than those estimated from bi-parentally inherited microsatellites (data not shown). Moreover, I was able to identify likely migrants between sample areas using assignments tests on data from nuclear markers (Chapter 2). Both results reflect the presence of male-biased dispersal in thinhorn. Comparisons between migration rates estimated from nuclear and mitochondrial markers in African buffalo (*Syncerus caffer*) are similar to those from thinhorn, with estimated male migration rates up to twenty times those of females (van Hooft *et al.*, 2003).

3.4.3. Mountain sheep taxonomy

3.4.3.1. Species level

Polyphyly was found between thinhorn and bighorn sheep. There was evidence for recent contact between these species, and AMOVA at the species level produced non significant results (Table 3.2a). This confirms earlier unpublished work by Ramey (1993), also reporting paraphyly. Although morphological evidence suggests that glacial vicariance resulted in long term species separation (Pielou, 1991; Geist, 1999), this is clearly not the case. Species classification is currently based on analysis of skull morphology (Cowan, 1940). Fewer significant differences between species were found upon reanalysis of the same data (Ramey, 1993). At the molecular level thinhorn and bighorn sheep do not appear to fulfil species criteria since their mtDNA lineages are not reciprocally monophyletic (Moritz, 1994). Basing species on mtDNA monophyly may be overly stringent (Crandall *et al.*, 2000). Instead species

status may be better based on criteria of historic and current genetic and ecological exchangeability. Crandall *et al.*, (2000) suggested that evidence for inexchangeability can support species division, while evidence for exchangeability supports synonymisation. Results indicate that historic gene flow has occurred, thus the species have demonstrated historic exchangeability. It should be determined if genetic exchange currently takes place before a decision can be made regarding species synonymisation. This seems unlikely given the spatial separation of BC and OCC (Figure 3.2). Ecological exchangeability must also be assessed. Some evidence for ecological inexchangeability may be found in differences in skull and colour morphology. However, at their closest geographic point of contact the habitat and life history traits of these species are very similar indicating exchangeability.

3.4.3.2. Thinhorn subspecies

Paraphyly was evident between thinhorn subspecies (Figures 3.3 and 3.4). MtDNA therefore does not support the current thinhorn subspecies delineations. Although examples of concordance between mtDNA clades and recent intraspecific differentiation can be found (Holder *et al.*, 2000), this research can be added to the long list of studies where taxonomic divisions for species and subspecies of recent origin are not supported by monophyletic mtDNA clades (Talbot & Shields, 1996; Holder *et al.*, 1999; Burbrink *et al.*, 2000; Brunner *et al.*, 2001; Flagsted & Røed, 2003).

Current taxonomy groups thinhorn populations into Stone's or Dall's sheep and there is limited genetic support for this classification in nuclear DNA (Worley *et al.*, 2004), where microsatellite markers showed significant subspecies differentiation by grouping BC and CY together as Stone's sheep and all others as Dall's sheep. However, mitochondrial DNA, together with the clinal nature of colour morphology, does not support this. In addition OG differs in colour morphology from Dall's sheep populations, yet it is geographically and genetically isolated from other dark sheep populations. Overall, genetic evidence confirms that the OG, CY and BC populations have been launched on independent evolutionary paths, isolation has occurred

geographically and genetically for long time periods, and colour morphology has diverged. The current Stone's sheep subspecies designation does not appear useful to describe the evolutionary history of these populations, and it should therefore be recommended that these populations be managed as separate entities.

3.4.4. Overall conclusions

Pleistocene glaciations are a major catalyst of the evolution of diversity (Pielou, 1991; Avise & Walker, 1998; Holder *et al.*, 2000), and data here demonstrate the importance of looking beyond well known major glacial refugia to smaller ice-free areas to establish their evolutionary importance. Small refugia have allowed for hybridization between species as well as stimulation of diversification processes. As might be expected, the signature left by isolation in small glacial refugia is often a genetic bottleneck (Federov & Stenseth, 2002), and results from British Columbia show this same trend. The Mackenzie Mountain population on the other hand appeared to have maintained high levels of genetic diversity. This may be due to larger refugial size, or possibly post glacial admixture of divergent haplotypes. High genetic polymorphism was also maintained in *Packera contermina* despite nunatak isolation, which may have been a result of hybridisation or life history traits specific to the species (Golden & Bain, 2000). Small refugial populations commonly show morphological differentiation due to random drift or directional selection while in isolation (Clark *et al.*, 2001; Federov & Stenseth, 2002). In one proposed refugial population (BC) morphological differentiation was evident, while the other had no obvious indications of differentiation (NT).

The interpretation of phylogeographic histories presented here are made using the traditional approach of inferring causation for associations between observed patterns of variation and the geographical distribution of populations. An alternative method using statistical phylogeography, the inference of species history from statistical tests of hypotheses and demographic parameters (Knowles & Madison, 2002) is not considered. The traditional approach is suggested to allow the over-interpretation of

data, which may be misleading (Knowles, 2004). Explanations for patterns in genetic variation here are in many cases supported by statistical knowledge of current population structure in identifying indisputable evidence of glacial refugia. It is also suggested that traditional interpretive methods may be unable to accurately define past events in species with complex histories. The population demography of thinhorn in interglacial periods is unknown, and the use of a slowly evolving gene region to investigate phylogeography may have included the effects of previous glaciations, in addition to the most recent from which conclusions are made here. However, estimates of divergence between statistically supported populations support the locations of four recent refugia. The additional problem that a gene's genealogy may not reflect the specie's genealogy due to selection is a lesser problem in this data set as it involves relatively neutral mitochondrial DNA.

These findings are the first to provide evidence that an organism could have survived Wisconsinan glaciation in refugia in northern British Columbia and the Mackenzie Mountains, and demonstrate the importance of mtDNA as a tool to further knowledge about glacial refugia and ice age paleoenvironments and paleoecology. Evidence exists to support usage of the ice-free corridor as a refugium. This finding is of great relevance to reconstructing the evolutionary histories of other North American species, which may have been affected by similar processes. Although previous research has resulted in consensus that the environment of the ice-free corridor was very severe, and probably unsuitable to at least human habitation (Catto *et al.*, 1996; Catto, 1996; Mandryk *et al.*, 2001), evidence from mountain sheep provides new insights into the paleoecology of the region, demonstrating that portions of the corridor were probably inhabited by a large mammal for the duration of the last ice age.

PART II:
GENETIC VARIATION IN THE THINHORN IMMUNE REGION

CHAPTER 4. GENETIC VARIATION AT IMMUNE LINKED MICROSATELLITE LOCI

4.1. INTRODUCTION

4.1.1. Identifying natural selection

Evidence for selection at the molecular level is most directly obtained by applying statistical tests of neutrality to patterns of sequence variation in coding regions (Kreitman & Akashi, 1995; Kreitman, 2000; Nielsen, 2001; Fu, 1997; Tajima, 1989). However, it requires a great deal of gene isolation and *de novo* sequencing to apply this approach to multiple genes in non-model taxa. A solution lies in the use of neutral linked markers such as microsatellites which are also affected by selection through genetic hitchhiking (Slatkin, 1995a). Recently this methodology has been implemented in large-scale genome scans of many loci to identify regions of high linkage disequilibrium expected in the presence of selection (Payseur *et al.*, 2002; Kauer *et al.*, 2003; Vigouroux *et al.*, 2002). As selection is locus-specific, only those markers linked to genes under selection are affected. In contrast, genetic drift and population level effects on variation, such as inbreeding and fluctuating population size, are likely to affect all loci equally due to shared demographic history. Selective hitchhiking can therefore be differentiated from neutral variation via the identification of markers with levels of variation outside neutral expectations. This can be accomplished in non-model taxa using heterologous markers such as cross-amplifying microsatellites linked to functional genes.

The idea of identifying outlier loci under the influence of selection originated with the F_{ST}-based test of Lewontin and Krakauer (1973). Methods for differentiating between neutrality and selection are not without problems. Lewontin and Krakauers' original method has been criticised due to its sensitivity to population structure and history (Nei & Maruyama, 1975; Robertson, 1975), which recent adaptations have attempted to improve upon (Vitalis *et al.*, 2001; Beaumont & Nichols, 1996;

Bowcock *et al.*, 1991; Schlotterer, 2002). These are useful tools in species where many microsatellite loci are available. An alternative concept of examining selection is to test for neutrality (Watterson, 1978) on a smaller number of loci where selection is presumed likely based on a prior knowledge of the functional significance of the linked gene. Genes involved in the immune response are excellent candidates for this type of analysis because of known associations with fitness (Hedrick, 2002; Jeffery & Bangham, 2000).

4.1.2. Expectations of selection

Balancing selection may be expected in genes involved in pathogen resistance as increased allelic diversity can result in increased resistance to a more varied array of pathogens. This is especially well documented in the MHC region (Hedrick, 1999), suggested in the human NRAMP region (Blackwell & Searle, 1999) but unreported in IFNG. Natural selection can result in both a decrease or increase in genetic variation compared to levels expected under neutrality (Maynard Smith & Haigh, 1974; Watterson, 1978; Hudson *et al.*, 1987). However, we can make specific predictions regarding the expected levels of allelic diversity within and between populations relative to neutrality assuming balancing selection is acting on immune loci.

Balancing selection generates an increase in the proportion of heterozygotes that opposes the effects of inbreeding on selected loci. Reduced values of F_{IS} and Watterson's statistic F_A (see methods) at immune linked loci relative to neutrality are therefore predicted. Conversely, an increased F_{IS} and reduced within population diversity are expected following positive directional selection, such as a selective sweep following disease outbreak (Tajima, 1989; Barton, 2000; Fay & Wu, 2000). Balancing selection predicts a reduction in the degree of genetic differentiation (F_{ST}) between populations (Muirhead, 2001). In addition migrants carrying rare alleles are at a selective advantage as they increase the genetic diversity of the recipient population. This results in an increase of the effective migration rate (Schierup *et al.*,

2001) via the homogenisation of between population allelic distributions. Again, this is in opposition to expectations of positive directional selection, where population specific selection against maladaptive alleles increases the degree of differentiation between populations (Lewontin & Krakauer, 1973).

Microsatellites linked to advantageous alleles are expected to show a reduction of their natural variability (Maynard Smith & Haigh, 1974; Slatkin, 1995a). This effect can be identified from a multilocus screen to distinguish those microsatellites which are potentially affected by selection. Any such screen must account for differences in mutation rate and coalescence time between microsatellite loci. One such method is the statistic ln RV (Schlotterer, 2002). The statistic is based on a ratio of the variance in repeat number of microsatellite alleles (see section 4.2.6 for more details), and has proven able to identify several loci under selection in human populations.

The presence of selection is associated with high levels of linkage disequilibrium (Wiehe & Slatkin, 1998; Barton, 1995; Kohn *et al.*, 2000; Kim & Nielsen, 2004). Linkage disequilibrium is likely to occur between closely associated loci, but is rapidly lost with increasing distance between markers due to recombination. As such, extremely fine-scale genome maps are required to identify regions of selection (Wiener *et al.*, 2003). Using linkage disequilibrium to identify regions of selection must be carried out with caution. Demographic factors, such as population substructure (Wall *et al.*, 2002), can also lead to its maintenance over large distances. Despite this, the use of linkage disequilibrium in identifying selection has proven successful (Gilad *et al.*, 2002; Kohn *et al.*, 2000; Rosenblum *et al.*, 2004).

The ability of tests to identify regions of selection is dependent on several factors. Migration acts to obscure the ability to identify selection (Beaumont & Balding, 2004), while it is more easily recognised in regions where a relatively strong selection coefficient is present. The latter is particularly relevant to studies of the immune system. In populations with recent histories of serious disease selection coefficients on genes in the immune system may be higher than in populations where

selective agents are less virulent. Identifying selection may therefore prove more productive in species with known histories of disease. Bighorn sheep have been severely impacted by the pathogen *Pasteurella haemolytica* (Foreyt, 1989) and may therefore be expected to exhibit evidence of selection at immune related genes. Balancing selection is indeed evident within nucleotide sequences of the MHC region of bighorn (Gutierrez-Espeleta *et al.*, 2001). The strength of selection is not strong, with neutral processes dominating the maintenance of genetic variation (Boyce *et al.*, 1997). In contrast to the disease history of bighorn, there are no known serious outbreaks of disease in the history of thinhorn. As such, the strength of selection at immune related genes is expected to be lower in the latter species.

There have been only a limited number of previous studies to utilise linked microsatellites with the aim of identifying regions under natural selection within immune genes. This is not surprising since the high density linkage maps required for genome scanning methods are not available in non-model organisms. No work has been conducted on species of North American wild sheep. Methods of analysing microsatellites linked to candidate genes have included the following; the identification of outlying levels of genetic variation (Huang & Yu, 2003; van Haeringen *et al.*, 1999), most often using tests for excess heterozygosity; significance tests of non-neutrality (Paterson, 1998), such as Watersons' statistic; and the presence of linkage disequilibrium over relatively long distances (Kohn *et al.*, 2000). These methods are in addition to the identification of associations between microsatellite alleles and disease resistance (Obexer-Ruff *et al.*, 2003; Paterson *et al.*, 1998; Coltman *et al.*, 2001). All but one of these studies concerns genetic variation within the MHC region, the most well documented immune region in the mammalian genome.

4.1.3. Outline

In this chapter I employ microsatellites linked to three regions involved in the vertebrate immune response: the major histocompatibility complex (MHC),

interferon gamma (IFNG) and natural resistance macrophage protein (NRAMP) to examine the nature of genetic variation in natural populations of thinhorn sheep. Several recent studies on selection make use of coalescent simulations to identify outlying loci under the influence of selection (Payseur *et al.*, 2002; Storz *et al.*, 2004; Stajich & Hahn, 2005). However, here I use empirical data from presumably neutral, non-linked microsatellite markers (Chapter 2; Worley *et al.*, 2004) to provide a background level of neutral genetic variation. Levels of genetic variation between non-linked and immune-linked microsatellites are compared in a candidate gene approach aiming to identify natural selection in thinhorn sheep. Specifically, I aim to investigate the following hypotheses;

i) Patterns of variation within immune linked microsatellites reflect expectations of selection. I expect there to be evidence of balancing selection at the MHC region of thinhorn sheep (as with most other species investigated). Balancing selection is a common expectation of immune genes, due to the following nonexclusive mechanisms. Firstly, under expectations of negative frequency dependent selection genotypes with rare alleles are at a strong selective advantage as few pathogens are adapted to them. There are known associations between specific MHC alleles and parasite resistance that support this hypothesis (Paterson, 1998). Secondly, heterozygous individuals may be more resistant than homozygotes due to their presumed ability to respond to a wider array of pathogens. As genes of the MHC region are a central component of the immune response, interacting with other immune region genes, balancing selection is particularly logical. Balancing selection is also expected within the interferon gamma region. As the two components of this gene (T_1 and T_2 subsets) are involved in the response to two very different pathogens (intracellular and extracellular) we may expect that individuals that are heterozygous are better able to respond to both sets of pathogens. Balancing selection is expected in NRAMP for similar reasons. There are few known alleles of the human NRAMP, but the two most common are associated with either autoimmune or infectious disease (Blackwell & Searle, 1999). Thus the maintenance of both alleles would confer a selective advantage.

ii) Evidence of past selective events will be weaker in thinhorn sheep, where there is no recent evidence of disease, than in species with known disease histories. Selection acting to increase diversity at immune regions should be strong in species with recent histories of disease. Whenever possible I contrast the signatures of selection in thinhorn with those present in the limited number of comparable studies. This will allow an evaluation of the patterns of variation in species with different disease and life histories.

iii) There will be evidence of significant linkage disequilibrium between immune linked microsatellite loci within the same gene region. The degree of linkage disequilibrium between loci is high in gene regions affected by natural selection. As such linkage disequilibrium has been used as a means of identifying regions of positive selection (Kim & Nielsen, 2004). I aim to test the association between loci within and adjacent to each immune region and use results to evaluate evidence of selection.

4.2. MATERIALS AND METHODS

4.2.1. Samples

Horn and blood samples from 922 sheep from 24 sample areas (Figure 1.4) across the species range were included in this chapter. I have shown widespread population substructure within sample areas (Chapter 2). In order to minimise the effects of population substructure on locating selection, analyses of immune linked microsatellite data were performed at the level of sample area.

4.2.2. Microsatellite loci

Samples were genotyped over microsatellite loci within or adjacent to the MHC, IFNG and NRAMP gene regions (Figure 1). The MHC microsatellites included OLADRB (Schwaiger *et al.*, 1993) and OLADRBps (Blattman & Beh, 1992), both located within the class II region. Primer sequences for both were modified to improve the product quality from thinhorn samples (OLADRB-F, 5'-TGTGCAGCG-GCGAGGTGAGCG-3', OLADRB-R, 5'-CGTACCCAGA$^{T}/_{G}$TGAGTGAAGTATC-3'; OLADRBps-F, 5'-CATGGGTTTCATCCCTGAGT-3', OLADRBps-R, 5'-CTC-CTGTCTTGTCATCTCTACGA-3'). The microsatellite oMHC1 is located within the class I MHC region (Groth & Wetherall, 1994). In addition, two microsatellites that flank the MHC region, BMS468 (Stone *et al.*, 1995) and BM1258 (Bishop *et al.*, 1994) were screened. Two IFNG linked microsatellites were genotyped. The tetranucleotide microsatellite IFNG1 is situated in intron 1 of the IFNG gene (Schmidt *et al.*, 1996). No recombination has been observed between this locus and the microsatellite KP6 (Paterson & Crawford, 2000). The NRAMP linked microsatellites OVINRAO1 and OVINRAO2 are positioned at the 3' end of the gene (Pitel *et al.*, 1996). Samples were also genotyped over the NRAMP flanking markers ARO28 (Avraham *et al.*, 1993) and VIAS-S5 (Primmer & Matthews, 1993).

Sheep included in this chapter have previously been genotyped over twelve presumably neutral dinucleotide microsatellite loci (Chapter 2) developed in domestic sheep and cattle.

4.2.3. Microsatellite genotyping

Each polymerase chain reaction (PCR) was carried out in 10μl reactions containing 2μl of DNA template, 80 μmol of each primer, 0.16mM dNTP's, and 0.5 units *Taq* polymerase (Bioline, London). Concentrations of $MgCl_2$ varied with locus from 0.5 to 1.5mM. The PCR profile comprised 35 cycles of 30s at 94°C, 30s annealing at either 63°C (OVINRAO1, OLADRBps, BM1258), 60°C (IFNG), 57°C (BMS468), 54°C (OarKP6, ARO28, Vias-s5) or 50°C (OLADRB, oMHC1, OVINRAO2) followed by 40s at 72°C. Cycles were preceded by 5 min at 94°C and terminated with 10 min at 72°C. PCR products were genotyped using an ABI 3730 sequencer and analysed using the software GENEMAPPER (Applied Biosystems, Foster City, CA).

Initial observations of a high rate of excess homozygosity at OVINRAO1 in BC populations suggested a null allele at measurably high frequency. BC samples and a proportion of all others were therefore typed with the NRAMP1 primer set which amplifies the same microsatellite locus (Bussmann *et al.*, 1998). Many OVINRAO1 genotypes in BC were therefore extrapolated from the NRAMP1 genotype. OVINRAO1 was initially chosen over the NRAMP1 primer because it amplified more reliably. As a result of using the less reliable NRAMP1 primer, sample sizes are low in some BC sample areas (BC751, BC752 and BC754) where null OVINRAO1 genotypes could not be verified using NRAMP1. Where this is the case these regions are omitted from some analyses to avoid introducing bias in the data.

A high frequency null allele was also present in the loci OLADRB, BMS468 and BM1258. It was not possible to correct this difficulty despite several attempts at redesigning primer sequences. As a result all three loci were not used for analyses.

4.2.4. Within population diversity and evidence of selection

Observed and expected heterozygosities were calculated for linked loci within each sample area using GENETIX version 4.01 (Belkhir, 1996). Hardy Weinberg equilibrium was examined using exact tests as implemented by GENEPOP version 3.3 (Raymond & Rousset, 1995) combining sample areas using Fisher's method. All statistics have previously been calculated for twelve non-linked loci (Chapter 2) enabling a comparison between marker types to be made. The statistic F_{IS} was calculated (using GENEPOP) at each locus both over all samples and within each sample area. Comparisons were made between values from linked and non-linked loci to identify any hypothesised evidence of selection.

Watterson's homozygosity test of neutrality (Watterson, 1978) was used to test for selection at each locus and in each sample area. The test statistic (F_A) is the expected frequency of homozygotes under Hardy-Weinberg equilibrium, where $F_A = \sum_{i=1}^{k} p_i^2$.

Significance is tested against distributions of values expected for populations under neutrality given a number of alleles (k) and population size (2n) (Ewens, 1972). The distributions of expected values of F_A are available for up to k=40 and 2n=500 at http://allele5.biol.berkeley.edu/homozygosity/homozygosity.html. The test remains valid for subdivided populations. The normalised deviate of Watterson's test statistic, $F\mathrm{nd}_A$, was calculated for each locus (Salamon *et al.*, 1999)

$$F\mathrm{nd}_A = \frac{F_{A\,\mathrm{observed}} - F_{A\,\mathrm{expected}}}{\left[\mathrm{Var}\left(F_{A\,\mathrm{expected}}\right)\right]^{1/2}}$$

This allows loci to be compared across populations differing in allele number and population size. A positive value of $F\mathrm{nd}_A$ is indicative of positive selection, a negative value results from balancing selection. The mean value of $F\mathrm{nd}_A$ over m populations was calculated at all loci. The standard error of $F\mathrm{nd}_A$ is represented by S_A,

$$S_A^2 = \frac{\sum\left(F\mathrm{nd}_A - \overline{F\mathrm{nd}_A}\right)^2}{m-1}$$

The overall test of selection at a locus comprises a two-tailed test on mean Fnd_A,

$$z = \overline{Fnd_A}\left(\frac{\sqrt{m}}{S_A}\right)$$

4.2.5. Evidence of selection among sample areas

The overall mean genetic differentiation, as measured by F_{ST} (Weir & Cockerham, 1984), was compared between linked and non-linked loci. Under neutral expectations a homogeneous distribution of F_{ST} across loci is expected as drift and gene flow should affect all loci equally. Differences between neutral and putative selected loci can therefore be attributed to selection. To test for differences between values of pairwise F_{ST} at immune linked regions relative to neutral markers Mantel tests were conducted on distance matrices. The direction of any deviation from equality in the relationship was examined by conservative sign tests with reduced degrees of freedom to account for the non-independence of data points.

Isolation-by-distance relationships (Slatkin, 1993) of immune linked loci were compared with those of non-linked markers. Pairwise geographic distances were calculated from linear distances between mean latitudinal and longitudinal locations of sample areas and plotted against the corrected pairwise F_{ST} value, $F_{ST}/(1-F_{ST})$. Significance of the isolation-by-distance relationship was tested by Mantel tests as part of GENEPOP. The slope of the regression line of each plot was calculated constraining the relationship through the origin. This value was then compared between linked and non-linked loci.

4.2.6. Methods based on genome scanning

Schlotterer's statistic ln RV (2002) was calculated across all linked and non-linked loci in each sample area. The statistic is based on the variance in repeat number (V), a good estimator of microsatellite variability (Slatkin, 1995b):

$$E[V] = 4N_e\mu$$

where N_e is the effective population size of a diploid population and μ is the mutation rate. As we cannot compare variances between loci directly given the different mutation rates between loci, a ratio of variances (RV) is used. This is independent of the mutation rate;

$$E[RV] = E\left[\frac{V_{\text{Pop1}}}{V_{\text{Pop2}}}\right] \cong \frac{4N_{e\text{Pop1}}\mu}{4N_{e\text{Pop2}}\mu}$$

Simulations have shown that the delta method of estimating RV provides a better fit to data (Schlotterer, 2002);

$$E[RV] \cong \frac{4N_{e\text{Pop1}}\mu}{4N_{ePop2}\mu}\left(1+\frac{V(V_{\text{Pop2}})}{(4N_{\text{ePop2}}\mu)^2}\right) \cong \frac{4N_{e\text{Pop1}}\mu}{4N_{\text{ePop2}}\mu}\left(1+\frac{1}{12}\right)$$

therefore, a good estimator of RV can be calculated as follows;

$$E[RV] = E\left[\frac{V_{\text{Pop1}}}{V_{\text{Pop2}}}\right]\left(1+\frac{1}{12}\right)$$

There remain differences in coalescence times between loci. This results in a range of observed values of RV. The distribution of RV values follows a normal distribution with the identification of outliers being the basis of tests of selection. Each locus can be assigned a probability of neutrality from the density function of the standardised normal distribution. The test was expanded to enable pairwise comparisons between all sample areas over all 19 loci (total number of tests = 5244).

4.2.7. Linkage disequilibrium

Linkage disequilibrium between loci was tested using the LINKDOS program as part of GENEPOP, and significance tested using the Markov Chain method of Guo & Thompson (1992). Results for each sample area were combined using Fishers methods to test overall significance of linkage disequilibrium. Linkage disequilibrium calculated within each sample area was used to estimate the correlation between locus pairs within each gene region.

4.3. RESULTS

4.3.1. Genetic variation and evidence of selection within sample areas

Microsatellites contained between 1 (IFNG) and 23 (OLADRBps) alleles (Table 4.1). Expected heterozygosities of the polymorphic loci ranged from 0.353 (OVINRAO2) to 0.811 (KP6). At non-linked neutral markers we previously observed widespread overall heterozygote deficit, a result of sampling over genetically differentiated subpopulations within each sample area (Chapter 2). We observed a similar pattern in linked loci, with deviations from Hardy-Weinberg expectations present at all but NRAMP loci (Table 4.1).

The locus OLADRB was omitted after preliminary analyses due to several lines of evidence that suggested null alleles. Firstly, the levels of heterozygosity observed here were much lower than previously reported in domestic sheep (Paterson, 1998). Secondly, although deviations from Hardy Weinberg equilibrium were reported in other non-linked loci, the magnitude of this deviation was much greater in OLADRB (H_E-H_O = 0.132). Thirdly, the overall value of F_{IS} was more than ten times greater than that for other MHC linked loci (Table 4.1). This great deficit of heterozygotes in the dataset strongly implicates non-amplified alleles. The strong deficit of heterozygotes leading to highly significant deviation from Hardy Weinberg expectations was also present in BM1258 and BMS468, and hence these loci were both dropped due to suspected null alleles at high frequencies.

Values of F_{IS} were positive for all linked loci, significantly so in KP6 and OVINRAO1 (Table 4.1). The mean F_{IS} for the MHC linked microsatellites OLADRBps and oMHC1 was lower than the overall average for unlinked markers (Table 4.1). However, most values for linked loci were within the range observed at individual non-linked loci (Figures 4.1.A & 4.2.A). Overall, linked loci show a reduction in mean F_{IS} compared to non-linked loci (Mann Whitney U=13, P=0.013). This result should be treated with caution due to the non-independence of data

points. However, there were no loci that consistently fell outside the range of neutral expectations across sample areas (Figure 4.3), and there is considerable overlap in errors for estimates of mean F_{IS}. In all plots loci were ordered from low to high heterozygosity. This was done as some previous studies have identified significant relationships between measures of diversity and population genetic statistics (Paetkau *et al.*, 1997). There was no correlation here between non-linked F_{IS} and locus heterozygosity as measured by mean H_E (Figure 4.1.A; r=0.17, P=0.59).

Watterson's test results for linked loci are summarised in Table 4.2. There was significant deviation from neutrality in at least one sample area for each locus. The greatest deviation was observed at the IFNG linked locus KP6, at which seven out of 24 populations deviated from neutrality. At all loci deviations from neutrality were in the direction of increased heterozygosity (lower F statistics) resulting from more equal allele frequency distributions than expected under neutrality. There were many more negative values of $F\text{nd}_A$ than positive, and the overall mean value of $F\text{nd}_A$ from all linked loci was significantly negative.

Watterson's tests were also conducted on the allele frequencies of all non-linked markers. As with the linked markers most loci had negative values of $F\text{nd}_A$ (Figure 4.1.B). Moreover, loci with higher levels of variability show reduced values of mean $F\text{nd}_A$ (Figure 4.2.B, correlation; r=-0.63, P=0.03). It is possible that NRAMP linked markers showed some evidence for lower values of $F\text{nd}_A$ than those observed in non-linked loci with comparable diversity, but other linked markers show values in line with expectations.

Table 4.1. Characteristics of the linked microsatellite loci used in this study. The locus OLADRB has been omitted due to the presence of null alleles. OVINRAO1 genotypes are corrected for a null allele by additional genotyping at NRAMP1 and conversion to the equivalent OVINRAO1 genotype. Heterozygosities are given as the mean over sample regions. Hardy-Weinberg probabilities (as calculated by exact test in GENEPOP), F_{IS} and F_{ST} values are combined for populations using Fishers method (* represents significant values of F_{IS}).

Locus	N	Allele Size Range	*k*	H_O	H_E	HWE Prob.	F_{IS}	F_{ST}	Reference
OLADRB	892	148-238	14	0.311	0.443	<0.001	0.302*	0.353	(Schwaiger *et al.*, 1993)
OLADRBps	900	238-286	23	0.741	0.755	<0.001	0.020	0.118	(Blattman & Beh, 1992)
oMHC1	838	178-200	11	0.574	0.588	0.025	0.028	0.185	(Groth & Wetherall, 1994)
IFNG1	200	122	1	n/a	n/a	n/a	n/a	n/a	(Schmidt *et al.*, 1996)
KP6	897	182-212	16	0.751	0.811	<0.001	0.076*	0.108	(Paterson & Crawford, 2000)
OVINRAO1	734	218-238	9	0.560	0.585	0.787	0.046*	0.124	(Pitel *et al.*, 1996)
OVINRAO2	836	300-306	3	0.344	0.353	0.999	0.021	0.157	(Pitel *et al.*, 1996)
ARO28	914	139-145	3	0.308	0.300	0.980	-0.010	0.171	(Avraham *et al.*, 1993)
Vias-s5	908	106-114	5	0.382	0.374	0.065	-0.012	0.297	(Primmer & Matthews, 1993)
BMS468	770	118-170	12	0.337	0.639	<0.001	0.474*	0.179	(Stone *et al.*, 1995)
BM1258	876	109-131	11	0.367	0.609	<0.001	0.407*	0.249	(Bishop *et al.*, 1994)
Non-linked loci	921			0.570	0.597	<0.001	0.066*	0.160	

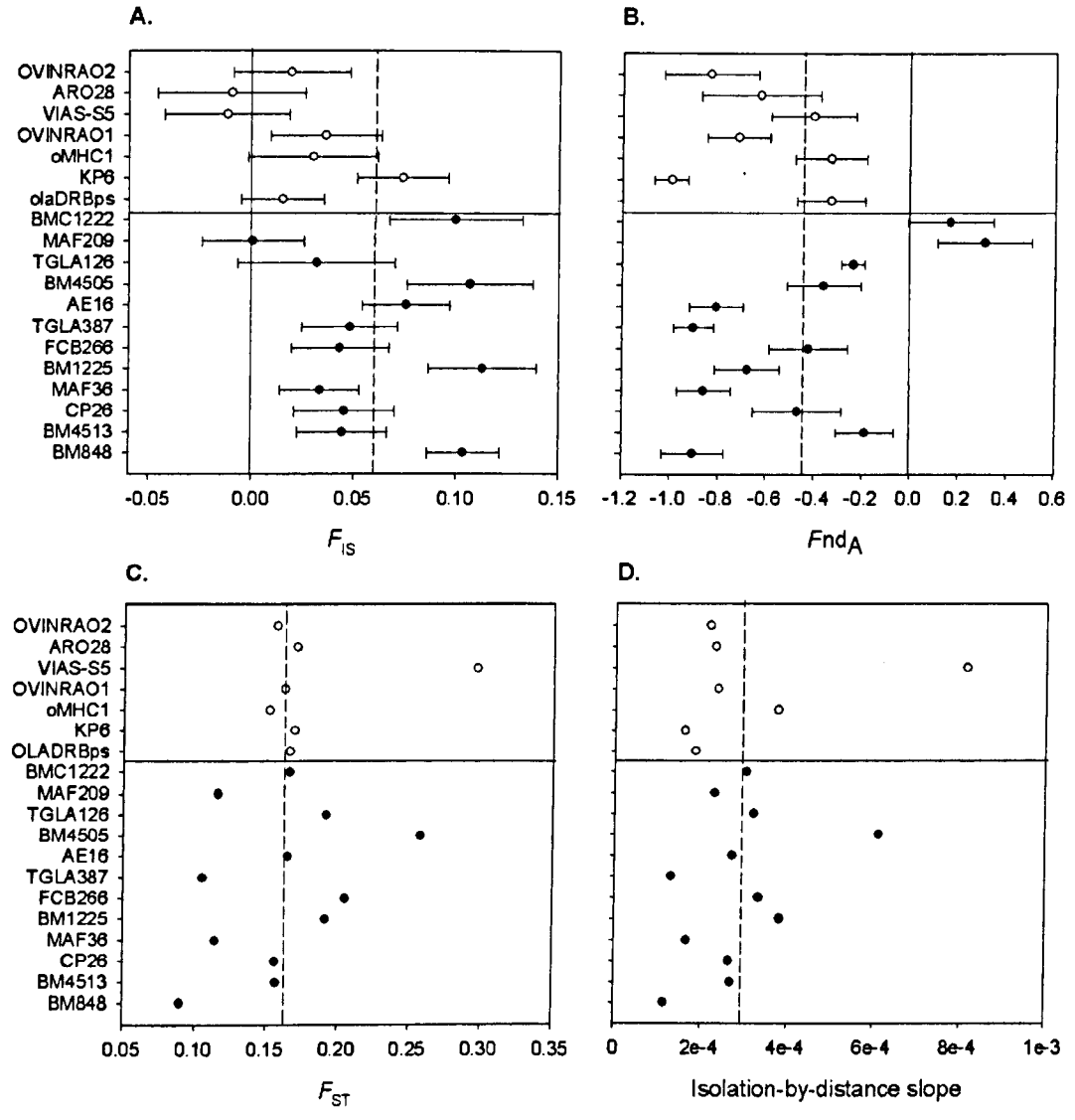


Figure 4.1. Comparisons of the genetic characteristics at all linked and non-linked loci. In all plots linked loci are shown in white above the horizontal line and non-linked loci are shown in black below it. Vertical dashed lines indicate overall values for non-linked loci. Loci are ordered from low to high variability (as measured by mean number of alleles per sample area), with OLADRBps being the most heterozygous of the linked loci. Standard errors for point values of F_{IS} and Fnd_A were calculated from values from all sample areas.

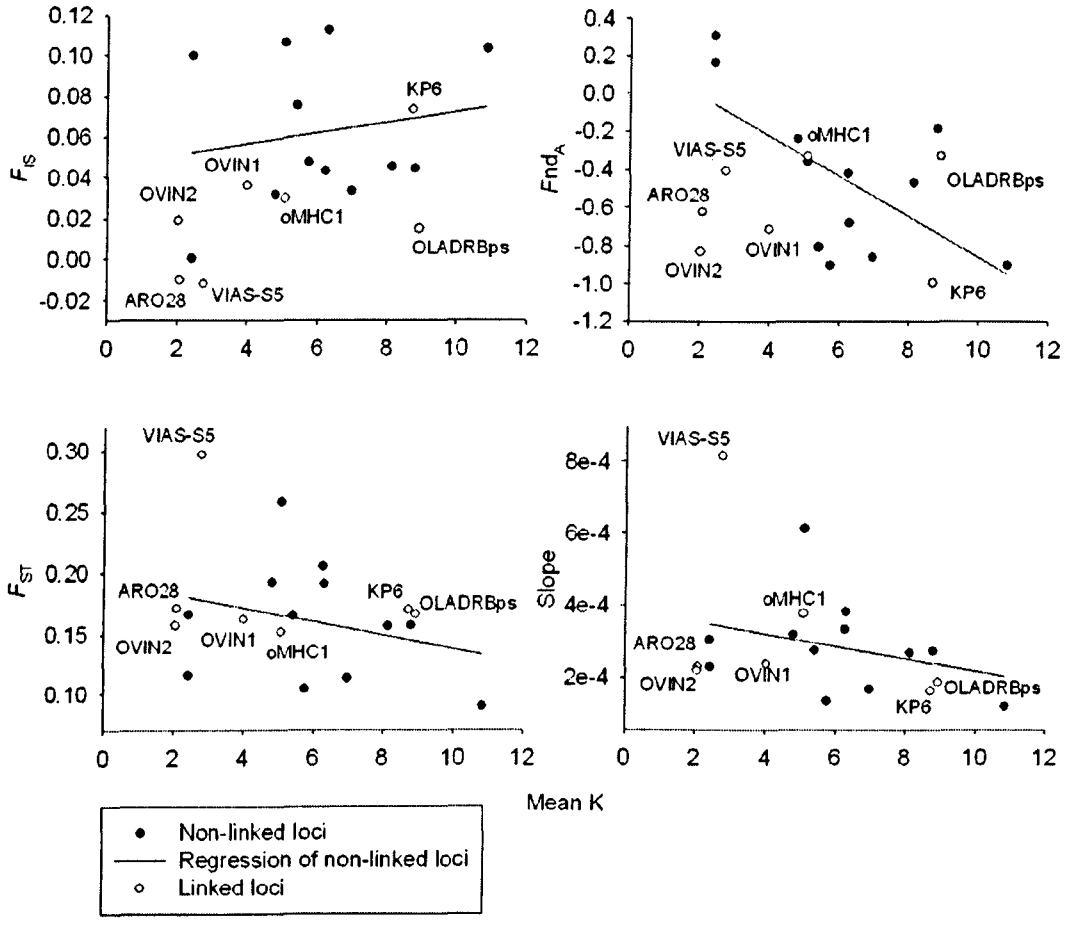


Figure 4.2. Plots of the relationships between microsatellite diversity (as given by mean allele number per sample area) and various genetic measures reveal that correlations present in non-linked loci are reflected in immune linked loci. Only the relationship between diversity and Fnd_A is statistically significant (t=2.19, P=0.05). Similar plots of F_{ST} and isolation-by-distance slope reflect a positive correlation between these two variables.

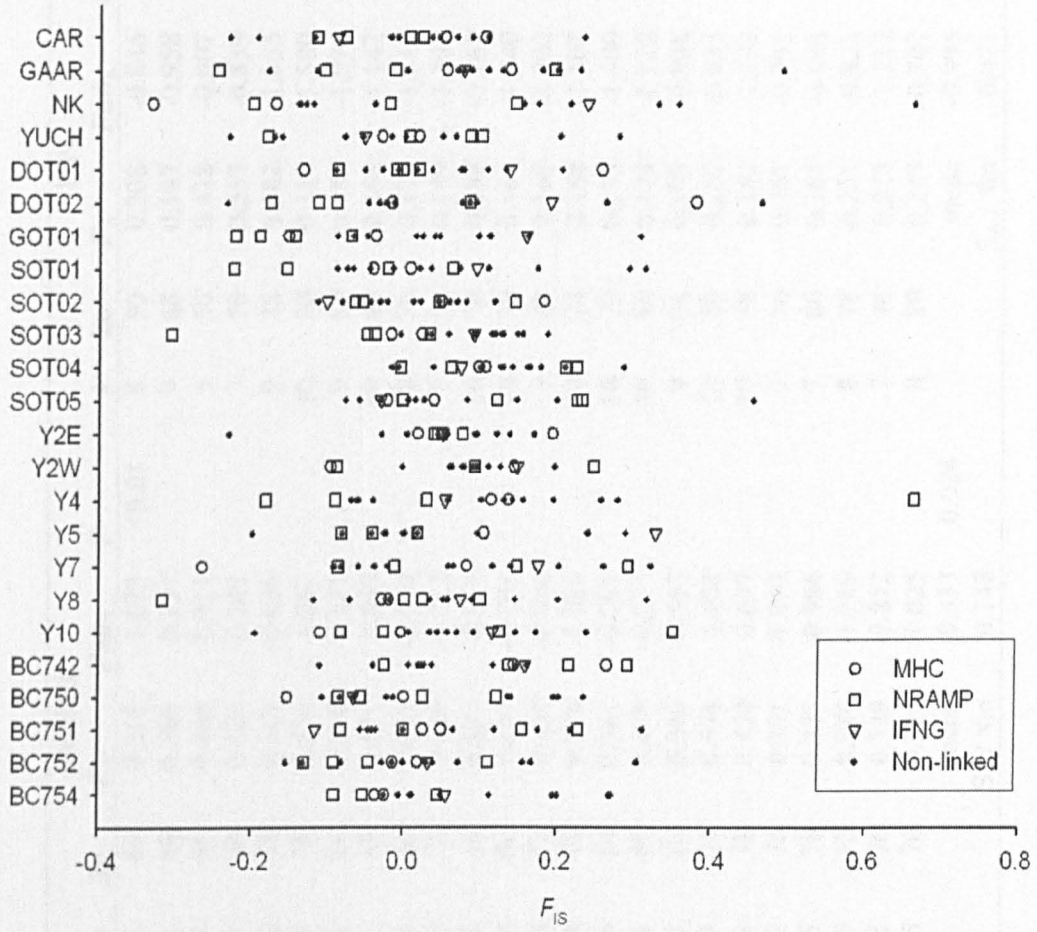


Figure 4.3. The distribution of F_{IS} values over all sample areas. Values for any given gene region are not consistently outwith those of non-linked loci across sample areas. In addition, the distribution of values is non-uniform across sample locations.

Sample area	OLADRBps					oMHC1					KP6				
	k	2n	F_A	Fnd_A	*P*	*k*	2n	F_A	Fnd_A	*P*	*k*	2n	F_A	Fnd_A	*P*
CAR	9	50	0.175	-0.966		5	50	0.215	-1.630	<0.01	8	50	0.208	-0.846	
GAAR	9	68	0.297	0.253		4	68	0.565	0.134		9	68	0.187	-0.958	
NK	5	50	0.395	-0.321		4	50	0.466	-0.354		5	50	0.438	-0.007	
YUCH	8	70	0.372	0.598		5	66	0.424	-0.249		7	70	0.252	-0.839	
D/OT/01	10	76	0.358	1.226		6	74	0.322	-0.616		9	78	0.182	-1.055	
D/OT/02	7	84	0.307	-0.503		3	76	0.876	1.241		10	80	0.119	-1.580	<0.01
G/OT/01	9	78	0.294	0.082		6	68	0.370	-0.221		9	80	0.182	-1.070	
S/OT/01	7	80	0.230	-1.064		5	68	0.471	0.059		10	80	0.158	-1.142	<0.05
S/OT/02	9	76	0.251	-0.362		4	68	0.435	-0.659		8	76	0.175	-1.317	<0.05
S/OT/03	10	80	0.269	0.134		5	76	0.439	-0.212		9	80	0.159	-1.307	<0.05
S/OT/04	10	78	0.217	-0.443		5	62	0.327	-0.887		10	76	0.188	-0.768	
S/OT/05	9	80	0.252	-0.351		4	80	0.487	-0.401		10	78	0.166	-1.040	
Y2E	10	78	0.249	-0.075		8	72	0.304	-0.076		9	80	0.169	-1.202	<0.05
Y2W	7	76	0.312	-0.370		4	74	0.379	-1.026		10	74	0.158	-1.108	
Y4	11	76	0.211	-0.238		6	64	0.361	-0.209		10	72	0.133	-1.400	<0.01
Y5	10	80	0.340	0.951		4	68	0.499	-0.273		10	80	0.124	-1.522	<0.001
Y7	8	74	0.160	-1.455	<0.01	3	62	0.546	-0.567		9	74	0.195	-0.906	
Y8	8	80	0.201	-1.087		7	74	0.534	1.488		10	80	0.202	-0.633	
Y10	7	80	0.266	-0.776		7	78	0.470	0.657		10	78	0.143	-1.298	<0.05
BC742	12	74	0.226	0.296		7	76	0.351	-0.053		7	76	0.261	-0.791	
BC750	11	80	0.215	-0.221		5	78	0.323	-0.984		7	80	0.281	-0.649	
BC751	10	80	0.164	-1.064		5	72	0.280	-1.249		8	78	0.251	-0.621	
BC752	8	72	0.189	-1.169		5	76	0.339	-0.872		7	76	0.223	-1.112	
BC754	10	80	0.166	-1.046		5	76	0.315	-1.025		8	80	0.243	-0.707	
			mean	-0.332	0.018			mean	-0.333	0.024			mean	-0.995	<0.01
			$S_A / \sqrt{m}$	0.141				$S_A / \sqrt{m}$	0.148				$S_A / \sqrt{m}$	0.071	

	OVINRAO1					OVINRAO2					ARO28					VIAS-S5				
Sample area	***k***	**2n**	F_A	Fnd_A	***P***	***k***	**2n**	F_A	Fnd_A	***P***	***k***	**2n**	F_A	Fnd_A	***P***	***k***	**2n**	F_A	Fnd_A	***P***
CAR	3	50	0.382	-1.476		2	50	0.520	-1.609		2	50	0.507	-1.688	0.05	3	50	0.449	-1.078	
GAAR	3	68	0.391	-1.480		2	64	0.805	0.080		2	68	0.943	0.901		2	68	0.673	-0.737	
NK	2	32	0.695	-0.406		2	48	0.920	0.869		1	50	1	NA		2	50	0.635	-0.900	
YUCH	5	68	0.368	-0.640		2	66	0.537	-1.556		2	70	0.632	-0.991		4	70	0.522	-0.143	
D/OT/01	5	74	0.458	-0.075		2	78	0.521	-1.685		2	78	0.796	-0.024		2	80	0.928	0.767	
D/OT/02	6	74	0.462	0.415		2	74	0.500	-1.800	<0.05	2	80	0.840	0.238		2	80	0.883	0.495	
G/OT/01	7	78	0.432	0.601		2	70	0.515	-1.704		2	78	0.837	0.222		2	78	0.975	1.057	
S/OT/01	5	64	0.345	-0.776		2	80	0.508	-1.771	<0.05	3	78	0.731	0.410		2	72	0.895	0.596	
S/OT/02	5	74	0.317	-1.010		2	64	0.504	-1.772	<0.05	3	80	0.742	0.461		3	76	0.852	1.106	
S/OT/03	6	76	0.350	-0.420		2	72	0.519	-1.683		2	80	0.561	-1.448		3	80	0.501	-0.895	
S/OT/04	5	78	0.354	-0.779		2	66	0.502	-1.771	<0.05	2	82	0.567	-1.423		2	82	0.501	-1.821	<0.05
S/OT/05	5	82	0.289	-1.216		2	78	0.512	-1.742		2	80	0.501	-1.810	<0.05	3	80	0.546	-0.639	
Y2E	4	78	0.309	-1.460	<0.05	2	70	0.500	-1.793	<0.05	2	80	0.500	-1.816	<0.025	2	80	0.638	-0.985	
Y2W	4	78	0.499	-0.322		2	66	0.566	-1.380		2	78	0.501	-1.805	<0.05	4	78	0.630	0.459	
Y4	3	78	0.531	-0.719		2	74	0.829	0.187		2	80	0.525	-1.665		2	80	0.553	-1.499	
Y5	4	80	0.478	-0.454		2	76	0.900	0.614		2	78	0.521	-1.686		3	80	0.509	-0.849	
Y7	3	78	0.496	-0.914		2	80	0.951	0.909		2	82	0.500	-1.827	<0.025	3	82	0.543	-0.661	
Y8	2	80	0.570	-1.393		2	76	0.683	-0.700		2	80	0.520	-1.697		4	80	0.718	0.972	
Y10	4	70	0.526	-0.118		2	80	0.711	-0.541		2	76	0.772	-0.165		3	80	0.455	-1.153	
BC742	3	44	0.437	-1.115		2	58	0.715	-0.447		2	80	0.728	-0.442		3	78	0.553	-0.592	
BC750	3	44	0.408	-1.289		2	70	0.680	-0.702		2	80	0.840	0.238		3	80	0.448	-1.194	
BC751	3	12	0.431	NA		2	62	0.775	-0.094		2	80	0.861	0.365		3	80	0.613	-0.261	
BC752	3	6	0.389	NA		2	74	0.829	0.187		2	80	0.975	1.054		3	80	0.454	-1.159	
BC754	1	2	1.000	NA		3	76	0.642	-0.081		2	80	0.840	0.238		3	80	0.550	-0.620	
			mean	-0.717	<0.001			mean	-0.833	<0.001			mean	-0.624	0.011			mean	-0.406	0.021
			$S_A / \sqrt{m}$	0.131				$S_A / \sqrt{m}$	0.196				$S_A / \sqrt{m}$	0.247				$S_A / \sqrt{m}$	0.176	

Table 4.2. (previous page) Watterson test results from linked microsatellite markers. Values are given for the number of alleles per locus in each sample area (*k*), sample size of alleles (2n), the homozygosity statistic F_A (Watterson, 1978), and the normalised deviate of this statistic (Fnd_A) amended for comparison over sample populations (Salamon *et al.*, 1999). Significance of deviation from neutrality (*P*) is given for each value of F_A and for the overall mean Fnd_A for all five loci.

4.3.2. Evidence of selection from between sample regions

I have reported a moderate level of genetic differentiation among sample areas included in this study from the analysis of non-linked loci (overall mean population F_{ST}=0.16). Overall genetic differentiation calculated from polymorphic linked microsatellites showed less variation but was of a similar magnitude (Table 4.1). F_{ST} values for individual linked loci fell within the range of values observed for unlinked loci (Figure 4.1.C). There was also no correlation between locus variability and F_{ST} (Figure 4.2.C, r=-0.28, *P*=0.38).

There were significant relationships between pairwise F_{ST} at neutral microsatellites and those at all three immune regions (Mantel tests, MHC (OLADRBps and oMHC1): Z=14.63, *P*<0.001; IFNG (KP6): Z=11.30, *P*<0.001; NRAMP (OVINRAO1 and OVINRAO2): Z=15.29, *P*=0.002). Plots of the relationships between pairwise F_{ST} values from neutral and linked markers show deviations from expectations of neutrality. At least twice as many points lie below the 1:1 relationship than above it for MHC class II (OLADRBps) and IFNG (KP6) gene regions (Figure 4.4). In contrast there are more points above the 1:1 line than below for NRAMP associated loci (estimates of linked F_{ST} are based on a combined dataset of NRAMP loci in linkage disequilibrium; OVINRAO1, OVINRAO2 and VIAS-S5). Due to the non-independence of data points the significance of any departures from neutrality can only be tested using highly conservative methods. Despite this the

class II MHC linked locus OLADRBps (sign test with d.f.=23, P=0.035) and the IFNG linked locus KP6 (P=0.003) exhibit significant reductions in differentiation from those expected under neutrality. Deviations in other gene regions are non-significant (oMHC1, P=0.405; NRAMP: P=0.210). In contrast to the overall trend of reduced genetic differentiation at MHC and IFNG linked markers, individuals from a central Alaskan (YUCH) sample area show an increase in genetic distance at linked relative to non-linked loci (Figure 4.4).

The highly significant isolation-by-distance relationship observed in non-linked microsatellites (Chapter 2) remained in linked loci (Figure 4.5; all P<0.01). After constraining the slopes of all isolation-by-distance relationships through the origin values for linked loci fell within the range of those between non-linked loci (Figure 4.1.D) and approximate values expected given the non-significant relationship with genetic heterozygosity (Figure 4.2.C).

4.3.3. Genome scanning methods

The distribution of values of Schlotterer's statistic ln RV (Schlotterer, 2002) approximated a normal distribution as expected (Figure 4.6). 6.8% of the 5244 comparisons showed significant deviation from the distribution (P<0.05), slightly above the level expected from error alone. The distribution of significant tests was non uniform across loci (Figure 4.7) and sample area comparisons. In more than a third of pairwise comparisons the locus TGLA126 showed evidence of significant departure from neutrality. Moreover, at the level of P<0.001, 46 out of the 59 significant tests came from this locus. All of the significant tests of TGLA126 came from pairwise comparisons that included sample areas in Alaska or southwestern Yukon. A similar pattern was seen with all other loci. For example, significant deviations from neutrality at ARO28 all included comparisons with the sample areas GAAR (northern Alaska) and Y8 (southern Yukon) while those at oMHC1 all included the sample area D/OT/02 (southern Mackenzie range).

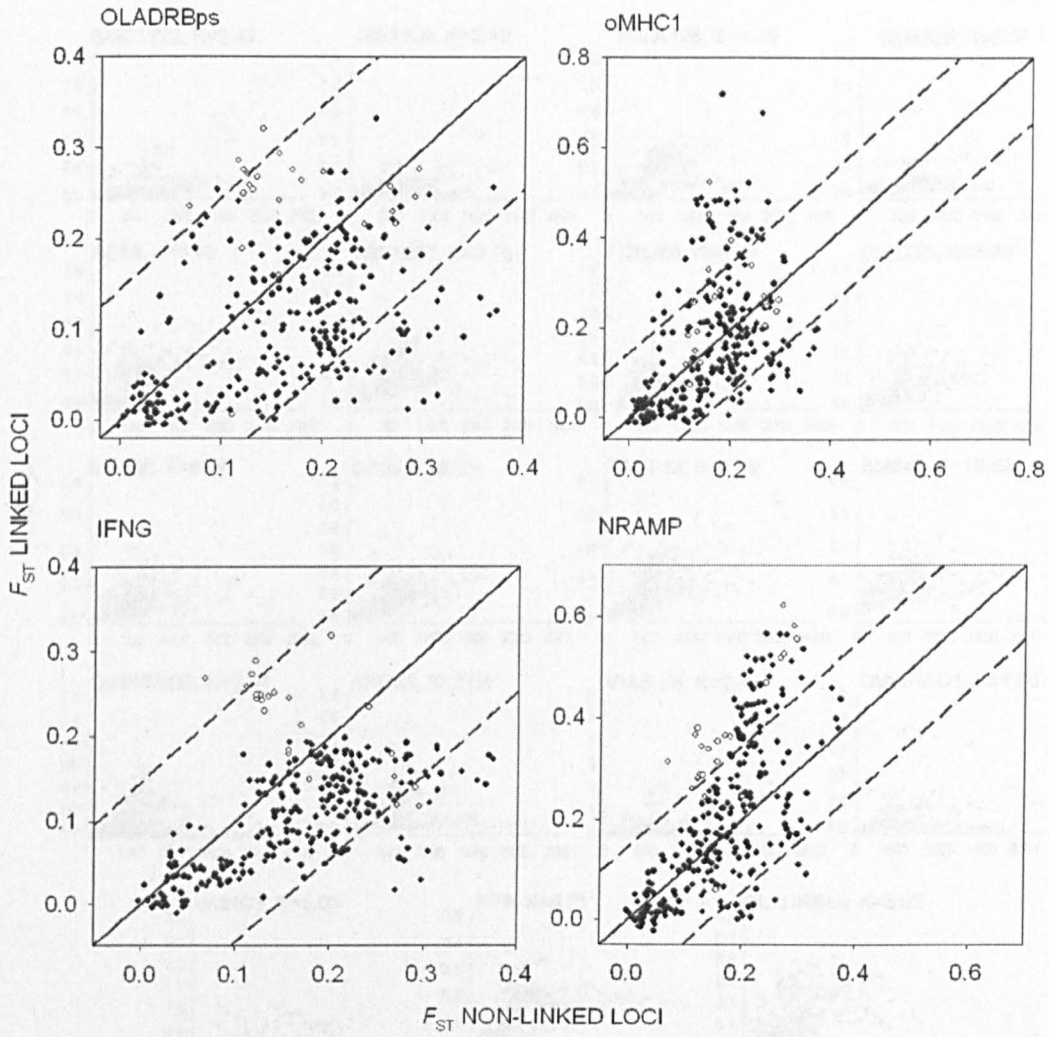


Figure 4.4. A comparison of genetic distances calculated from pairwise populations at linked to non-linked loci for each of the three immune regions. The solid line represents the 1:1 relationship between F_{ST} values calculated from non-linked and linked microsatellites expected under a null model of no selection. The extent of deviation from neutrality is indicated by the dotted lines (two standard deviations of the estimated F_{ST} of all non-linked loci). Estimated F_{ST} of non-linked loci is calculated from all twelve markers. Genetic distances of linked markers are estimated from single loci (OLADRBps, oMHC1 and KP6) or from two loci (NRAMP: OVINRAO1, OVINRAO2). Open data points include pairwise comparisons with the central Alaskan sample area YUCH, which exhibits a greater degree of genetic differentiation at immune linked loci (especially IFNG and NRAMP) than expected under neutrality.

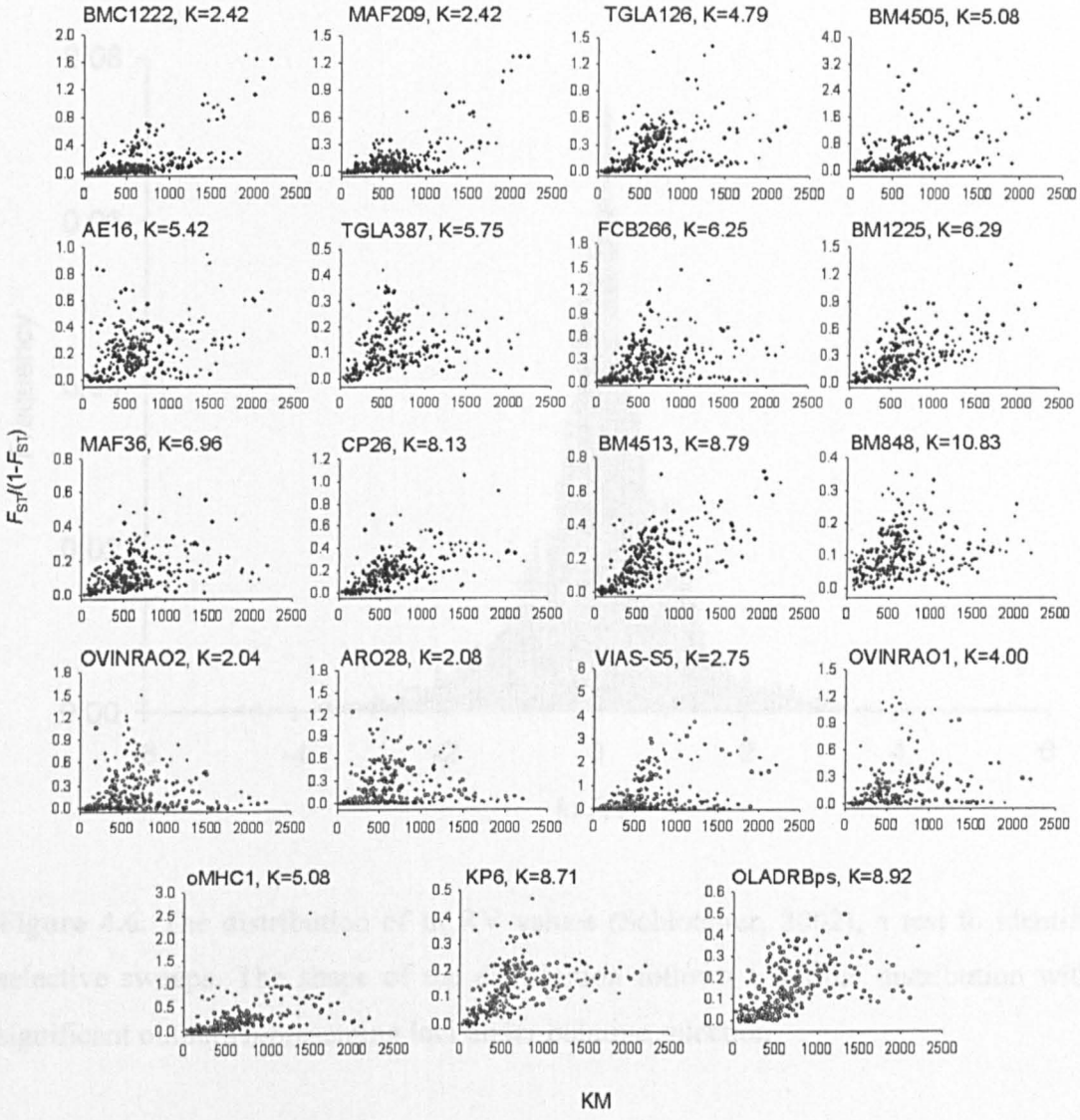


Figure 4.5. Isolation-by-distance plots of all nineteen loci included in this study reveal a bias in reporting increased genetic distance when using markers with low variability. The patterns present in linked markers do not appear to differ from the general trend seen in non-linked loci. K indicates the mean number of alleles within sample areas.

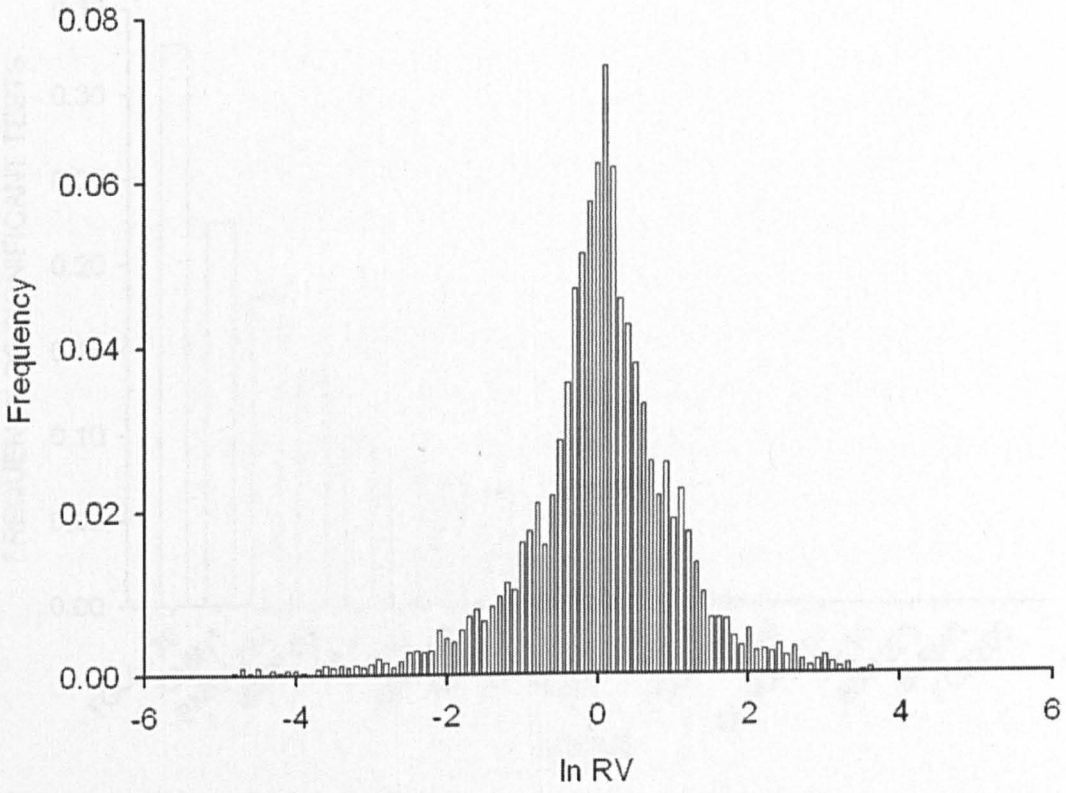


Figure 4.6. The distribution of ln RV values (Schlotterer, 2002), a test to identify selective sweeps. The shape of the distribution follows a normal distribution with significant outliers representing loci under putative selection.

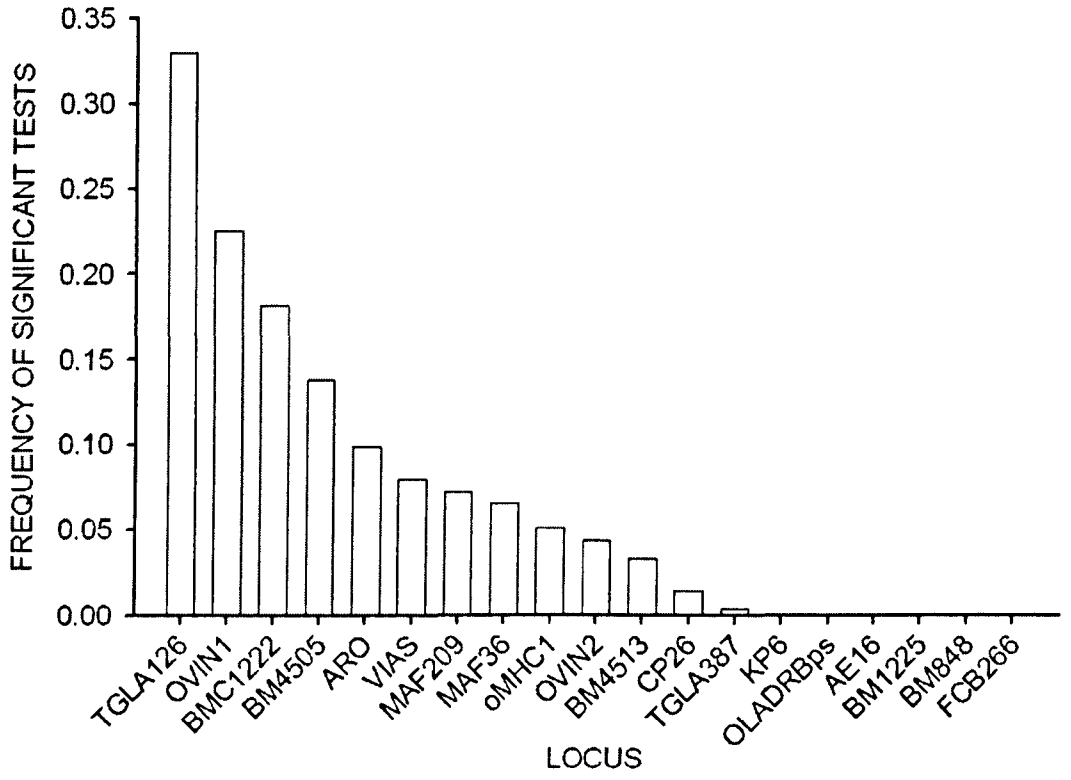


Figure 4.7. The distribution of significant deviations from neutrality using the statistic ln RV reveals that not all loci are equal. The locus TGLA126 exhibits more significant tests of non-neutrality than any other marker. The frequency of significant tests comes from values of ln RV calculated from pairs of sample areas (23 independent tests per locus).

4.3.4. Linkage disequilibrium

Significant linkage disequilibrium was observed between the locus pairs OLADRB and OLADRBps and oMHC1 and OLADRBps within the MHC region of chromosome 20 and between the NRAMP linked markers OVINRAO1 and OVINRAO2 and OVINRAO1 and VIAS-S5 (Figure 4.8). No assessment of the linkage disequilibrium present at IFNG was possible due to fixation of the microsatellite IFNG1 (Table 4.3). Due to the presence of null alleles in OLADRB, association with other loci was tested using data from heterozygotes alone. To examine the effect this has on accurate confidence levels of linkage I also recalculated linkage disequilibrium between other locus pairs using only heterozygote genotypes (Table 4.3). When only heterozygotes were used significance was lost at some comparisons including OVINRAO1 and VIAS-S5. Coefficients of linkage disequilibrium were moderate in all cases. When only heterozygotes were used for calculations, the correlation coefficient increased in many comparisons.

The significant linkage disequilibrium observed over all sample areas is not as apparent when sample areas are considered on an individual basis (Table 4.4). In addition, the number of significant tests of linkage disequilibrium between loci on the same chromosome does not exceed levels found between loci on different chromosomes.

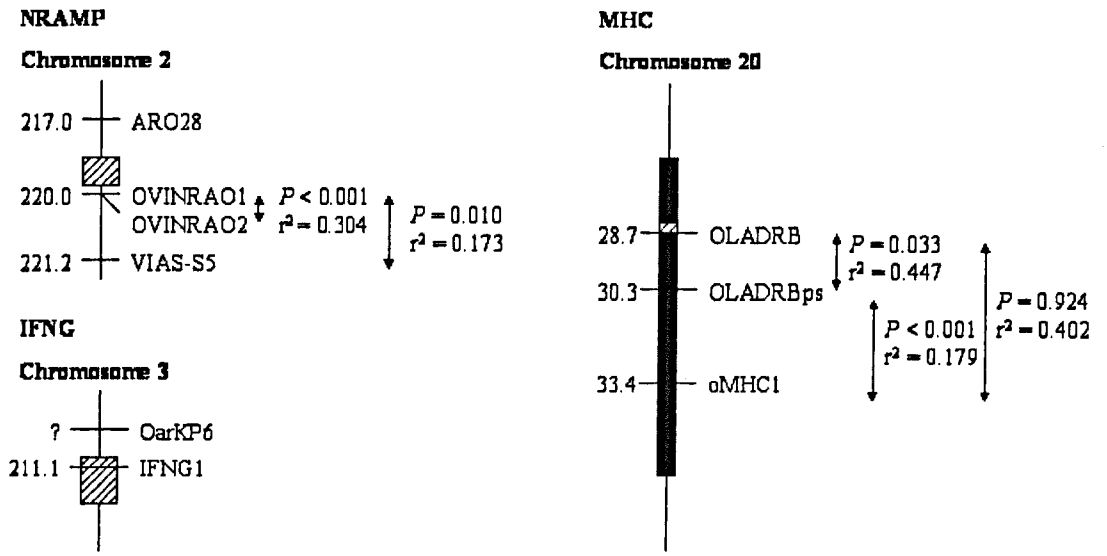


Figure 4.8. A linkage map of gene regions examined in this study. Values to the left of the figures indicate the map distances of microsatellite markers along each chromosome and are taken from SMC sex averaged maps of ARKdb (www.thearkdb.org). There is no given map distance for KP6, but no recombination has been observed between this locus and IFNG1. Approximate locations of coding regions in relation to microsatellite loci are indicated by hash filled rectangles. r^2 is the mean correlation coefficient of linkage disequilibrium between microsatellite pairs, while *P* is the overall probability of association. Only heterozygote genotypes were included when calculating these values for the null locus OLADRB. Within each immune region significant linkage disequilibrium is present between all markers. The associated correlation coefficients are moderate and variable.

Table 4.3. The effect of omitting homozygotes from analyses of linkage disequilibrium. In all cases of comparison with OLADRB no homozygotes were used due to the presence of null alleles at this locus. *calculating the correlation was only possible in one sample area. Significant linkage is highlighted by underlining values.

Loci comparison	All data		Heterozygotes only	
	P	r^2	*P*	r^2
OVINRAO1-OVINRAO2	<0.001	0.304	Not possible	
OVINRAO1-ARO28	0.676	0.154	Not possible	
OVINRAO1-VIAS-S5	0.010	0.173	0.693	0.185
OVINRAO2-ARO28	0.945	0.138	Not possible	
OVINRAO2-VIAS-S5	0.536	0.145	1.000	0
ARO28 – VIAS-S5	0.096	0.153	Not possible	
OLADRBps-oMHC1	<0.001	0.179	<0.001	0.282
OLADRBps-OLADRB	0.033	0.447	0.037	0.565
oMHC1-OLADRB	0.924	0.402	0.770	0.537
OVINRAO1-KP6	0.896	0.170	0.992	0.235
OVINRAO1-OLADRBps	0.117	0.183	0.797	0.258
OVINRAO1-oMHC1	0.316	0.190	0.479	0.337
OVINRAO1-OLADRB	0.914	0.356	0.854	0.284
OVINRAO2-KP6	0.726	0.152	0.444	0.368*
OVINRAO2-OLADRBps	0.033	0.155	0.574	0.292*
OVINRAO2-oMHC1	0.385	0.157	1.000	0.252*
OVINRAO2-OLADRB	0.631	0.359	1.000	0
KP6-OLADRBps	0.844	0.146	0.986	0.217
KP6-oMHC1	0.049	0.151	0.913	0.245
KP6-ARO28	0.281	0.145	0.211	0.661
KP6-VIAS-S5	0.605	0.145	0.984	0.391
KP6-OLADRB	0.998	0.379	0.969	0.499
OLADRBps-ARO28	0.359	0.148	0.713	0.284
OLADRBps-VIAS-S5	0.096	0.162	0.999	0.261
oMHC1-ARO28	0.710	0.130	1.000	0.289*
oMHC1-VIAS-S5	0.255	0.141	0.999	0.320
OLADRB - ARO28	0.997	0.262	Not possible	
OLADRB - VIAS-S5	0.981	0.319	0.824	0.568

Table 4.4. Linkage disequilibrium present within NRAMP and MHC loci at the level of sample areas. All probabilities below 0.1 are included. n.s., non significant, n.p., not possible given a small sample size or locus monomorphism.

Loci compared	CAR	GAAR	NK	YUCH	D1	D2	G1	S1	S2	S3	S4	S5
OVINRAO1-OVINRAO2	n.s.	<0.001	n.p.	<0.001	0.033	0.052	0.031	n.p.	n.p.	n.s.	n.s.	n.s.
OVINRAO1-ARO28	n.s.	n.s.	n.p.	n.s.	0.014	n.s.	n.p.	n.p.	n.p.	n.s.	n.s.	n.s.
OVINRAO1-VIAS-S5	n.s.	n.s.	n.s.	0.073	n.s.	n.s.	n.p.	n.p.	n.p.	0.048	0.098	n.s.
OVINRAO2-ARO28	n.s.	n.s.	n.p.	n.s.	n.s.	n.s.	n.s.	n.s.	0.074	n.s.	n.s.	n.s.
OVINRAO2-VIAS-S5	n.s.	n.s.	n.p.	0.066	n.s.	n.s.	n.p.	n.s.	0.078	n.s.	n.s.	n.s.
ARO28-VIAS-S5	n.s.	n.s.	n.p.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	0.051	0.031	n.s.
OLADRB-OLADRBps	n.s.	<0.001	n.s.	0.035	n.s.	n.s.	n.s.	n.p.	n.s.	n.s.	n.s.	n.s.
OLADRB-oMHC1	n.s.	n.s.	0.065	0.012	n.s.	n.s.	n.s.	n.s.	n.s.	n.p.	n.s.	n.s.
OLADRBps-oMHC1	n.s.	n.s.	n.s.	<0.001	<0.001	n.s.	0.005	n.s.	n.s.	0.004	0.049	n.s.
OVINRAO1-KP6	n.s.	0.037	n.s.	n.s.	0.006	n.s.	n.s.	<0.001	n.s.	n.s.	n.s.	0.045
OVINRAO1-OLADRBps	n.s.	0.016	n.s.	0.003	0.072	0.001	n.s.	n.s.	n.s.	0.034	n.s.	n.s.
OVINRAO1-oMHC1	n.s.	0.082	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	0.003	n.s.	0.028	n.s.
OVINRAO1-OLADRB	n.s.	n.s.	n.s.	n.s.	n.s.	0.002	0.052	n.s.	n.s.	n.s.	n.s.	n.s.
OVINRAO2-KP6	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	0.080	n.s.	0.095
OVINRAO2-OLADRBps	n.s.	0.012	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
OVINRAO2-oMHC1	n.s.	0.025	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
OVINRAO2-OLADRB	n.s.	n.s.	n.p.	0.022	n.s.	n.p.	n.s.	n.s.	n.s.	0.046	n.s.	n.s.
KP6-OLADRBps	0.073	n.s.	0.084	n.s.	n.s.	n.s.	<0.001	0.008	n.s.	n.s.	0.057	n.s.
KP6-oMHC1	n.s.	n.s.	<0.001	n.s.	0.001	n.s.	0.001	n.s.	0.074	n.s.	n.s.	n.s.
KP6-ARO28	n.s.	n.s.	n.p.	n.s.	n.s.	n.s.	0.053	n.s.	n.s.	n.s.	n.s.	0.095
KP6-VIAS-S5	n.s.	n.s.	0.019	0.001	0.055	n.s.	0.012	0.001	n.s.	n.s.	n.s.	n.s.
KP6-OLADRB	n.s.	n.s.	0.005	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
OLADRBps-ARO28	n.s.	n.s.	n.p.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	0.003	0.022
OLADRBps-VIAS-S5	n.s.	n.s.	0.098	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	0.020	n.s.
oMHC1-ARO28	n.s.	n.s.	n.p.	n.s.	n.s.	n.s.	n.s.	0.065	n.s.	n.s.	n.s.	n.s.
oMHC1-VIAS-S5	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	0.006
ARO28-OLADRB	n.p.	n.s.	n.p.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.

Table 4.4. continued

Loci compared	Y2E	Y2W	Y4	Y5	Y7	Y8	Y10	BC742	BC750	BC751	BC752	BC754
OVINRAO1-OVINRAO2	0.007	0.059	0.075	<0.001	<0.001	<0.001	0.002	n.s.	n.s.	n.s.	n.s.	n.s.
OVINRAO1-ARO28	0.094	n.s.	n.s.	n.s.	0.014	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
OVINRAO1-VIAS-S5	0.025	0.004	0.034	n.s.	n.s.	0.024	n.s.	0.002	n.s.	0.042	n.s.	0.059
OVINRAO2-ARO28	n.s.	n.s.	n.s.	n.s.	0.023	n.s.	n.s.	n.s.	n.s.	n.s.	n.p.	0.054
OVINRAO2-VIAS-S5	n.s.	0.044	n.s.	n.s.	n.s.	n.s.	0.011	n.s.	n.s.	n.s.	0.028	n.s.
ARO28-VIAS-S5	n.s.	0.086	n.s.	0.071	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	0.021
OLADRB-OLADRBps	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	0.091	n.s.	n.s.	n.s.
OLADRB-oMHC1	n.s.	<0.001	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
OLADRBps-oMHC1	<0.001	<0.001	n.s.	n.s.	n.s.	0.000	0.018	<0.001	n.s.	n.s.	n.s.	<0.001
OVINRAO1-KP6	n.s.	n.s.	n.s.	0.007	0.061	n.s.	0.001	n.s.	n.s.	n.s.	n.p.	n.p.
OVINRAO1-OLADRBps	0.007	n.s.	n.s.	n.s.	n.s.	n.s.	<0.001	0.094	n.s.	n.s.	n.p.	n.p.
OVINRAO1-oMHC1	n.s.	<0.001	n.s.	n.s.	n.s.	n.s.	<0.001	0.048	n.s.	0.092	n.p.	n.p.
OVINRAO1-OLADRB	n.s.	n.s.	n.p.	n.s.	n.s.	n.s.	n.s.	n.p.	n.s.	n.p.	n.p.	n.p.
OVINRAO2-KP6	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	0.026	0.056	n.s.	n.s.	n.s.	n.s.
OVINRAO2-OLADRBps	0.004	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	0.004	0.003	n.s.	n.s.	n.s.
OVINRAO2-oMHC1	n.s.	n.s.	n.s.	n.s.	n.p.	0.501	0.010	n.s.	n.s.	n.s.	n.s.	n.s.
OVINRAO2-OLADRB	n.s.	n.s.	n.p.	n.p.	n.s.	0.022	n.s.	n.s.	n.s.	n.s.	n.p.	n.s.
KP6-OLADRBps	0.052	n.s.	n.s.	0.015	0.053	0.072	0.057	n.s.	n.s.	n.s.	n.s.	n.s.
KP6-oMHC1	n.s.	n.s.	<0.001	n.s.	n.s.	n.s.	<0.001	0.049	n.s.	n.s.	0.053	n.s.
KP6-ARO28	n.s.	n.s.	n.s.	0.007	0.082	<0.001	0.032	0.041	n.s.	n.s.	n.s.	n.s.
KP6-VIAS-S5	0.060	n.s.	<0.001	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	0.063	n.s.	n.s.
KP6-OLADRB	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	0.026	0.018
OLADRBps-ARO28	n.s.	n.s.	n.s.	n.s.	0.008	0.001	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
OLADRBps-VIAS-S5	0.061	n.s.	n.s.	n.s.	0.054	n.s.	n.s.	n.s.	n.s.	<0.001	n.s.	n.s.
oMHC1-ARO28	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
oMHC1-VIAS-S5	n.s.	n.s.	0.033	n.s.	n.s.	n.s.	0.063	n.s.	n.s.	0.055	n.s.	n.s.
ARO28-OLADRB	n.s.	<0.001	n.p.	n.s.	n.s.	n.p.	n.s.	n.s.	n.s.	n.s.	n.p.	n.s.
OVINRAO1-KP6	n.s.	n.s.	0.080	n.s.	n.s.	n.s.	n.s.	n.s.	n.p.	n.p.	n.p.	n.s.

4.4. DISCUSSION

Most recent studies identifying adaptive variation in natural populations concentrate on sequence variation (Gutierrez-Espeleta *et al.*, 2001; Landry & Bernatchez, 2001; Aguilar *et al.*, 2004; Huang & Yu, 2003; Jarvi *et al.*, 2004) and differences in allele frequencies within single populations (Paterson, 1998). Studies of selection within single populations do not face the confounding effects of population genetic structure. Here I report the results of an analysis of genetic variation within microsatellites linked to thinhorn immune genes in a structured network of subpopulations. I account for the effects of demographic history by comparing levels of genetic variation at linked microsatellites to a background of genetic structure empirically estimated from presumably neutral markers.

4.4.1. Evidence of non-neutral variation in immune-linked microsatellites

In common with most mammals (Greenwood, 1980) thinhorn sheep exhibit strong female natal philopatry (Festa-Bianchet, 1991). In addition there is no long distance male migration. The population structure that results is evident in neutral microsatellite markers (Chapter 2) and is mirrored in immune-linked loci. The presence of genetic substructure precludes any direct measures of heterozygote excess at linked microsatellite markers. Instead evidence of selection was examined by comparing population genetic statistics between linked and non-linked markers.

Balancing selection lowers values of F_{IS} relative to those at neutral loci. There appears to be an overall reduction in values of F_{IS} at the MHC linked microsatellites OLADRBps and oMHC1 compared to the overall value for non-linked loci (Table 4.1). However, when values from non-linked loci are presented individually no linked outliers are present (Figure 4.1.A). Figure 4.1.A does not reflect the positive trend between F_{IS} and locus diversity. Against this background the F_{IS} value reported for OLADRBps appears to be lower than expected under conditions of neutrality (Figure 4.2.A). No linked locus exhibits an outlying value of F_{IS} when comparisons

are made within each sample area separately (Figure 4.4). Given the underlying population substructure this level of comparison is more appropriate than global comparisons. I therefore conclude there to be no evidence for selection in comparisons of the statistic F_{IS}.

Initial Watterson's tests indicate deviation from neutrality at microsatellites linked to all three immune regions that are consistent with expectations of balancing selection (Table 4.2). However, evidence for selection only becomes widely significant when the global statistic $F\text{nd}_A$ is used in preference to the traditional sample area approach using F_A. The lack of outliers in comparisons of Watterson's statistics between linked and non-linked loci supports a conclusion of limited deviation from neutrality in thinhorn immune genes (Figures 4.1.B and 4.2.B). The majority of studies reporting Watterson's tests are limited to human populations. The few exceptions concern MHC sequences (Hambuch, Lacey, 2002; Landry, Bernatchez, 2001) and linked microsatellite alleles (Paterson, 1998; Huang & Yu, 2003; Landry & Bernatchez, 2001; Hambuch & Lacey, 2002), all reporting evidence of balancing selection. Only the study of Huang and Yu (2003) of structured wild mouse populations makes use of a comparison between MHC-linked and neutral microsatellite loci. Significant evidence of balancing selection was observed in all loci, echoing patterns of variation in thinhorn. Both studies show how balancing selection may be inferred incorrectly if Watterson's tests are interpreted against unknown levels of background population structure.

The identification of loci with outlying levels of variation is highly subjective in populations where there is underlying genetic substructure. To further complicate interpretations of selection I report a significant correlation between the number of alleles at a locus (K) and values of the statistic $F\text{nd}_A$ (Figure 4.2.B). As balancing selection leads to an increase in the number of alleles this correlation is perhaps intuitive. Regardless, the correlation acts to illustrate why understanding underlying associations between population genetic statistics is critical when they are to be used in the field of adaptive variation. Problems can arise in comparative investigations

(such as in this study) where there is no overlap in allelic diversity between the chosen neutral and linked loci. In these cases there may be an inherent bias to incorrectly report selection, balancing or otherwise.

The increase in allelic diversity associated with balancing selection can lead to low levels of genetic differentiation between populations (Schierup *et al.*, 2000; Fay & Wu, 2000). Migrants carrying rare alleles are also at a selective advantage, thereby increasing the effective migration rate (Schierup *et al.*, 2001) and further reducing genetic differentiation through the homogenisation of allelic distributions between populations. The presence of balancing selection can therefore be recognised by identifying outliers in comparisons of pairwise F_{ST} estimates between loci (Hedrick, 1999; Muirhead, 2001). Compared to the overall level of genetic differentiation at non-linked loci a reduction in F_{ST} was observed in the class II MHC linked marker OLADRBps and the IFNG locus KP6 (Table 4.1). The class I linked marker oMHC1 does not show the hypothesised reduction in overall F_{ST}. Instead I observed an increased F_{ST} in relation to neutral loci. A lack of evidence of selection in the locus oMHC1 was also present in domestic sheep (Paterson, 1998). This could reflect the reduced importance of selection in class I MHC genes compared to those in the class II region.

Figure 4.4 illustrates that the reduced levels of F_{ST} at OLADRBps and KP6 in comparison to neutral loci are also present between pairwise sample areas. In addition there appears to be an increased level of differentiation between sample areas at the locus oMHC1 and within the NRAMP region. Higher F_{ST} at linked loci can occur as a result of selection for reduced allelic diversity, such as purifying selection against deleterious alleles (Gilad *et al.*, 2002). Conservative statistical tests were however unable to support evidence of selection at either NRAMP of MHC class I loci. Against the overall trend for lower level F_{ST} at MHC class II and IFNG linked loci, a population from central Alaska exhibits relatively high genetic differentiation. This trend has previously been reported at MHC loci (Landry & Bernatchez, 2001). It can arise when pathogen driven directional selection acts

differentially between populations via local adaptation (Slatkin & Wiehe, 1998). Alternative alleles may confer differential fitness to individuals in different environments. The signature of selection in these cases is the increase of genetic differentiation between populations. I am unable to test this hypothesis here, but support would come from the identification of pathogens or environmental conditions unique to the sample area YUCH, neither of which seems likely. Drawing any conclusions about selection from the 1:1 expected relationship of neutral to immune linked measures of F_{ST} (Figure 4.4) could be misleading. The distribution of F_{ST} values is bound at zero and has no upper limit. This may lead to a clustering of data points at lower levels of differentiation. Figure 4.4 shows some evidence of this, where in most cases a clustering of data points can be seen near the origin.

As with all genetic distance measures, F_{ST} shows a negative relationship with genetic diversity (Paetkau *et al.*, 1997). The overall heterozygosity of non-linked loci on which comparisons of F_{ST} were made is 0.597. The corresponding heterozygosities of the immune linked markers KP6 and OLADRBps exceed this level (Table 4.1). Differential heterozygosities between loci provide a non-adaptive explanation for the apparent deviation from neutrality at linked markers. In common with analyses of F_{IS} and Watterson's test statistics, the overall putative evidence of balancing selection from F_{ST} is lost when comparisons are made with individual non-linked loci (Figure 4.2.C). Given all analyses of population differentiation I again conclude a lack of strong evidence supporting the hypothesis of balancing selection acting on immune linked microsatellites. This reflects the results of several other mammalian studies, where little evidence of selection is present in comparisons of between population genetic diversity (Boyce *et al.*, 1997; Bernatchez & Landry, 2003).

Regional adaptive variation can leave signals in isolation-by-distance relationships (Storz & Dubach, 2004). Significant isolation-by-distance is present at all thinhorn loci irrespective of linkage to immune coding regions (Figure 4.5). This supports the predominance of non-selective forces over selection at maintaining genetic variation.

In addition, there is no evidence of differential isolation-by-distance slopes between linked and non-linked loci (Figures 4.1.D and 4.2.D).

The statistic ln RV uses expectations of reduced variability in order to locate regions affected by the selective sweep of novel beneficial mutations (Schlotterer, 2002). It revealed no evidence of selection at immune linked loci, although the locus TGLA126 was highlighted as a possible region of positive selection. As most ovine gene regions remain unmapped I cannot draw any conclusions about the linkage of TGLA126 with a coding region under selection. However, the locus is located on chromosome 16, approximately 8cM from the growth hormone receptor (distance from www.thearkdb.org).

Grouping genetically distinct subpopulations, each with different allele frequency distributions, may result in an even allele frequency distribution in the global sample not present in the original subpopulations. This would result in lower values of the test statistics F_{IS}, F_A and F_{ST} and ultimately to a bias in reporting balancing selection. As most 'evidence' of balancing selection in thinhorn comes from global statistics where sample areas have been combined, I conclude there to be few, if any, signatures of balancing selection at immune-linked microsatellites in thinhorn sheep.

4.4.2. Linkage disequilibrium

As expected I found evidence of strong linkage disequilibrium (LD) between closely mapped microsatellite loci within both the MHC and NRAMP gene regions. LD appears to be stronger within the MHC region compared to that surrounding the NRAMP gene. Within the MHC region, significant LD was recorded between markers up to 3cM apart, whereas in the NRAMP region LD was only present between markers up to approximately 1cM. As LD is inflated in regions of selection, I might infer that if selection is present it is stronger within the MHC than within NRAMP. However, no conclusions can be made regarding the presence of selection at either region due to the lack of a reference region for comparison against immune

microsatellites. The presence of LD between microsatellites within both regions does show that the identification of regions of selection using them should be possible, as linkage with coding regions is highly likely.

The MHC region is known to exhibit extensive levels of LD (Huttley *et al.*, 1999). However, the correlation coefficient associated with LD between oMHC1 and OLADRBps was low suggesting recombination has occurred between class I and II regions. There is evidence for recombination within the MHC from simulations of the shape of phylogenetic trees (Schierup *et al.*, 2001). For example, recombination rates in non-human primates are estimated at ~1cM per 1Mb (Satta *et al.*, 1999), while those in humans have been estimated at ~2.4cM per 1Mb (Payseur *et al.*, 2002). Where recombination rates and mutation rates are approximately equal the power of tests of selection is reduced (Wall, 1999). Recombination within the thinhorn MHC region may have obscured evidence of selection in linked microsatellite loci. Pooling genetically distinct subpopulations can also obscure evidence of LD, and hence selection (Fay & Wu, 2000). However, as with tests of neutrality, measures of global LD here are stronger than those made on a sample area basis.

Correlation coefficients associated with the locus OLADRB are high (Figure 4.8), although this may largely be a result of using heterozygotes alone to calculate LD. Although the effects of using only heterozygotes to calculate LD are uncertain, I observed two general trends in the data. In addition to an increase in correlation coefficients, the associated probabilities of LD are reduced (Table 4.3). In cases where only two or three alleles are present, removing homozygotes removes the ability to infer linkage disequilibrium at all.

4.4.3. Comparison with previous studies

As methods for examining selection using linked neutral markers are relatively novel, there are few previous studies available for comparison. As with thinhorn

sheep, heterozygote excess has been reported at MHC linked microsatellites in domestic cattle breeds (van Haeringen *et al.*, 1999). Unlike this study Watterson's tests of neutrality were non-significant in all cases. As the authors make no comparison between linked and non-linked loci they were unable to conclude whether patterns of variability were due to balancing selection or population bottleneck resulting from domestication, as both are expected to increase heterozygosity. Paterson's (1998) study of Soay sheep shows that it is possible to identify regions of balancing selection from Watterson's tests in equilibrium populations. MHC flanking markers in Soay show no evidence of selection, indicating that balancing selection is confined to within the immune complex. Although no reliable thinhorn MHC flanking markers were typed, the NRAMP flanking marker ARO28 shows no evidence of linkage disequilibrium with microsatellites closer to the coding region (Figure 4.8), but does show deviation from neutrality both globally and within several sample areas in Watterson's tests (Table 4.2). This comparison between studies again illustrates the differential success in identifying regions of selection between equilibrium and non-equilibrium populations.

Although all the populations above are likely affected by an array of pathogens (some of which may be associated with mortality), none has had any history of a severe disease related bottleneck. I cannot therefore compare the strength of selection pressures caused by severe disease outbreak with the magnitude of signatures of selection in linked microsatellites. I can however draw inferences from cases of drug resistance. A study of microsatellites within the warfarin-resistant gene in rats has revealed a deficit in the number of heterozygotes that is consistent with directional selection (Kohn *et al.*, 2000). Moreover, the number of loci within the resistance gene exhibiting deviation from Hardy-Weinberg equilibrium increases with the proportion of warfarin resistant animals in the population. The associated selection coefficient is estimated at 0.6. Although this value is high, statistical support for directional selection could only be confirmed through the fine-scale linkage disequilibrium mapping of 26 microsatellites within ~32cM of the resistance gene. It

is estimated that up to 18% of selection coefficients acting in natural populations are of the magnitude of that present in the rat warfarin-resistance gene. This figure is however an estimate and it remains difficult to assess values of the selection coefficient from levels of variability (Schierup *et al.*, 2001). It is unlikely that any selection coefficient acting on the thinhorn immune genes NRAMP and IFNG is so high, and therefore not surprising that an analysis of few candidate linked microsatellites would identify regions of natural selection within these genes.

Selection coefficients within the MHC region have been reported above 0.5 in an otherwise monomorphic mammalian population (Aguilar *et al.*, 2004). However, this level of selection seems to lie outside the range of values otherwise reported in loci under balancing selection within populations polymorphic at neutral microsatellite loci (Sauermann *et al.*, 2001; Garrigan & Hedrick, 2001; Thursz *et al.*, 1997; Satta *et al.*, 1994). I have shown that thinhorn exhibit extensive variability at neutral microsatellite loci (Chapter 2). Together with the lack of recent widespread disease I therefore expect selection coefficients within the MHC region to be much lower than 0.5. The reliable identification of balancing selection requires the selection coefficient to be in the order of twenty times that of the migration rate (Beaumont & Balding, 2004). Although the differential between the rates of migration and the selection coefficient in thinhorn is expected to be higher than this, the low selection pressures I expect at the MHC would still make any reliable identification of balancing selection difficult.

4.4.4. Conclusions

Examining evidence of selection in sub-structured populations is always problematic. Microsatellite data from thinhorn MHC, IFNG and NRAMP show empirically the extent of these difficulties. Without prior knowledge of neutral variation traditional tests of neutrality suggest the presence of balancing selection in all regions under study. After accounting for the effects of population structure widespread evidence of selection within sample areas is rejected. Limited evidence based on genetic

distances between sample areas is suggestive of balancing selection in MHC class II genes and IFNG, as hypothesised, although tests of selection are not statistically robust. There remain difficulties regarding the ability of statistical methods to identify natural selection from microsatellite data. In this study conservative tests were used in many cases, with the possible loss of evidence of weak selection.

As hypothesised significant linkage disequilibrium (LD) was present within both the MHC and NRAMP regions of thinhorn sheep, although the strength of locus interactions was not high. In opposition to the actions of selection generating associations between loci, high mutation rates in non coding loci and recombination between microsatellites can both act to obscure the effects of microsatellite hitchhiking, such as LD (Wiehe, 1998). Although the mutation rate within thinhorn immune regions is unknown, low coefficients of LD were present between several linked loci. The presence of recombination this suggests may add to the lack of evidence of selection in thinhorn from microsatellite loci.

Finally, the lack of evidence supporting the presence of balancing selection (particularly within the MHC region) suggests that any selection that may be present is associated with a low selection coefficient. Success in identifying selection in thinhorn remains to be compared to that possible from bighorn sheep, where higher selection pressures from pathogens may be hypothesised. However, I expect that unless the selection coefficients are very high, microsatellites would not produce evidence of selection from across groups of genetically structured populations.

CHAPTER 5. SEQUENCE VARIATION AT THINHORN IMMUNE GENES

5.1. INTRODUCTION

5.1.1. Locating evidence of selection from sequence polymorphisms

Analyses of microsatellite variability revealed no conclusive evidence for selection acting on any of the candidate genes (Chapter 4). There may however be significant evidence of selection located in sequence data. Microsatellite loci, although linked to immune genes, are not the targets of selection. Recombination, together with high rates of mutation, may have obscured evidence of selection within such loci. Instead, analysis of variation within coding sequences may more directly reveal signatures of selection. In addition, it has been argued that tests of selection based on allele frequencies (as are necessary for microsatellite data) can be weak and unreliable due to strong assumptions regarding population demography (Nielsen, 2001).

Separating the effects of selection and neutral drift must also be considered when identifying selection from sequence variation. Moreover, selection may be acting differently in different subsets of nucleotides (Kreitman & Akashi, 1995). In the same way as outliers to microsatellite variability can be compared against the neutral expectations of non-linked loci, sequence variation can be compared against quantitative expectations arising from the neutral theory of evolution. The first test of neutrality specifically for sequence data was developed by Hudson *et al.*, (1987) and is based on the infinite sites neutral model. The neutral theory suggests that levels of polymorphism within a species are proportional to the amount of divergence at the gene between closely related species. The test statistic has specific assumptions regarding population demography, which if violated can lead to misleading 'evidence' of selection. Tests within populations rather than between species are also susceptible to the effects of population structure. Tajima's *D* test is one such popular example (Tajima, 1989). Due to the difficulty of distinguishing demography from

genealogy, the results of all tests of selection based on variability at one locus must be interpreted with caution (Nielsen, 2001).

Rather than comparing empirical sequence variation to expectations of the neutral theory, comparisons can be made between levels of variation in different classes of mutation within the same gene. Such tests do not rely on assumption relating to population demography as variability is compared within individuals. Synonymous substitutions can be regarded as essentially neutral, as they do not change the amino acid sequence (and hence the protein product). Theory therefore suggests that the rate of synonymous substitution (d_S) is equal to the mutation rate. In contrast, substitutions in nonsynonymous sites are under considerable selective pressure and the relative rate of substitutions (d_N) can reveal much about the role of natural selection. Deleterious mutations in nonsynonymous sites can result in non-functional protein products. As a consequence lower substitution rates are expected than those in synonymous sites ($d_N/d_S<1$). An excess of beneficial mutations at nonsynonymous sites produces the opposite trend ($d_N/d_S>1$). Testing the significance of any deviation from a $d_N=d_S$ relationship is a simple test of positive natural selection. Under models of balancing selection the number of nonsynonymous substitutions is expected to be higher than the number of synonymous substitutions (Hughes & Nei, 1988). Recent progress has been made on methodologies to test the significance of any relationship, including the implementation of a maximum likelihood approach (Bustamante *et al.*, 2003). Advances in significance testing enable comparisons to be made between divergent species and can account for differences in transition/transversion rates and multiple substitutions at the same site (Yang & Nielsen, 2000).

All loci included in this chapter show some association between sequence polymorphism and the ability to respond to pathogens. Evidence of natural selection is evident in all chosen loci across several species. Signatures of both directional and balancing selection are expected at immune related loci, corresponding to the association of specific alleles to disease resistance or the ability to respond to a more

varied array of pathogens. A summary of the known actions of selection in all chosen loci is outlined below.

5.1.2. Evidence of selection at immune genes

5.1.2.1. Major histocompatibility complex

Functional MHC molecules are formed by non-covalent associations between alpha and beta chain peptides encoded by *A* and *B* genes (Figure 5.1). The functionally important region of both genes is exon 2, which encodes the peptide binding region (PBR) of both domains in class II MHC molecules. The specificity of MHC proteins for foreign peptides is determined by sequence variation contained within alleles of the PBR of both class I and class II genes. In many vertebrates, including sheep, the most polymorphic region of MHC genes is the DR beta molecule (DRB). Within this molecule most polymorphism is restricted the DRB3 gene. As a result the most variable region of the MHC (and of the vertebrate genome) is exon 2 of the DRB3 gene, where there are currently almost a hundred characterised domestic sheep alleles. The microsatellite OLADRB is located at the 3' end of DRB3 exon 2.

Nucleotide diversity (π) within the human MHC is at least forty times more than that of other parts of the genome (Gaudieri *et al.*, 2000) suggesting the predominance of selection over neutral processes in maintaining variation. Such high levels of nucleotide diversity are evidence of balancing selection acting on the region (Hedrick, 1999; Paterson, 1998; Aguilar *et al.*, 2004). Balancing selection is also expected to lead to an excess of nonsynonymous sequence substitutions, the presence of trans-species sequence haplotypes and significant selection tests on linked non-coding repeat regions (Huang & Yu, 2003; Aguilar *et al.*, 2004; Paterson, 1998). Many studies show the presence of all these patterns within variation of the MHC (Paterson, 1998; Aguilar *et al.*, 2004; Landry & Bernatchez, 2001; Jarvi *et al.*, 2004; Garrigan & Hedrick, 2001; Hambuch & Lacey, 2002; Gutierrez-Espeleta *et al.*, 2001; Wilson *et al.*, 2003).

There have been relatively few studies attempting to test associations between individual genotype and disease resistance in wild vertebrates. However, in both salmon and Soay sheep higher than average parasite resistance has been correlated to specific alleles at an MHC locus (Langefors *et al.*, 2001; Paterson, 1998). Evidence correlating MHC heterozygosity with parasite resistance can be found in several fish species including stickleback (*Gasterosteus aculeatus*), where populations exhibiting more diversity in MHC genes also show more genetically diverse parasites. Individuals with an intermediate number of alleles show the minimal parasite loads (Wegner *et al.*, 2003).

5.1.2.2. Interferon gamma

IFNG is encoded by a single gene of 1.2kb mRNA length in humans and in its active form comprises a 34kDa homodimer (Boehm *et al.*, 1997). Mutations in the IFNG receptor result in recurrent bacterial infections (van Deventer, 2000). The IFNG gene is involved in disease resistance in humans (Davis *et al.*, 2000; Jenkins *et al.*, 2000; Pokorny *et al.*, 2001), horses (Horin *et al.*, 2004) and sheep (Coltman *et al.*, 2001) amongst others. Although nucleotide substitution rates are lower than in the MHC, one domestic sheep sequence is implemented in conferring disease resistance. Moreover, linked microsatellites show allelic correlations to disease resistance in both sheep (Coltman *et al.*, 2001) and goats (Obexer-Ruff *et al.*, 2003). Despite involvement in disease resistance there is no empirical evidence of natural selection within IFNG from any species.

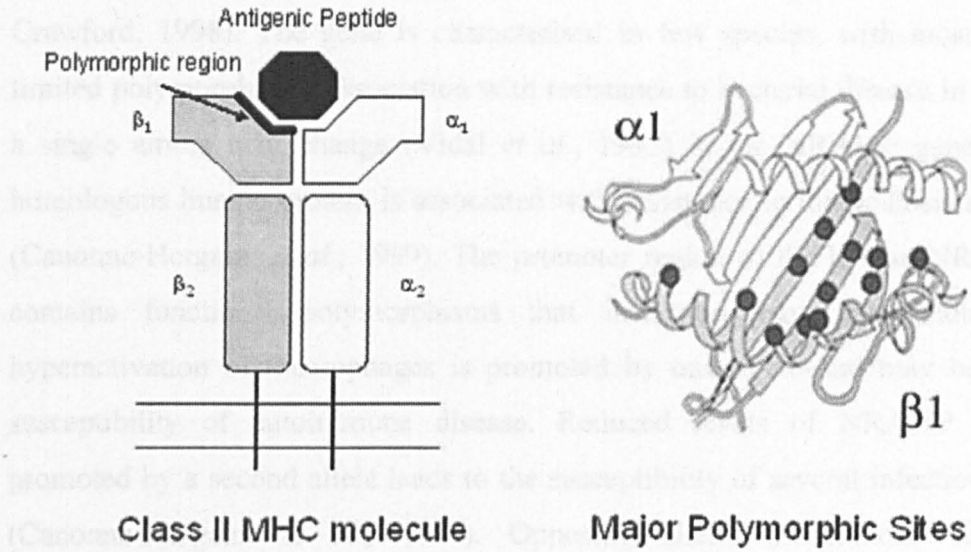


Figure 5.1. An illustration showing the structure of human MHC class II molecules and positions of major known nucleotide polymorphisms contained within them. Most polymorphic sites are located within the protein binding region (PBR) of subunit β_1 (illustrated here in filled circles). Figure of molecular structure adapted from Austyn & Wood, (1993).

5.1.2.3. Natural resistance associated macrophage protein

The NRAMP gene encodes a protein of 548 amino acids in length (Matthews & Crawford, 1998). The gene is characterised in few species, with most exhibiting limited polymorphism. Association with resistance to bacterial disease in mice is via a single amino acid change (Vidal *et al.*, 1995) in the NRAMP gene while the homologous human protein is associated with resistance to tuberculosis and leprosy (Canonne-Hergaux *et al.*, 1999). The promoter region of the human NRAMP gene contains functional polymorphisms that influence gene expression. Chronic hyperactivation of macrophages is promoted by one allele and may be linked to susceptibility of autoimmune disease. Reduced levels of NRAMP expression promoted by a second allele leads to the susceptibility of several infectious diseases (Canonne-Hergaux *et al.*, 1999). Opposing allelic associations suggest the maintenance of polymorphism in humans is through balancing selection (Blackwell & Searle, 1999).

There are species differences in the associations observed between NRAMP polymorphism and disease susceptibility. Known associations to tuberculosis resistance in human populations are not present in deer (Mackintosh *et al.*, 2000) or cattle (Kumar *et al.*, 1999), but associations with other pathogens are probable in these species. NRAMP nucleotide substitution sites are present between cattle breeds (Ables *et al.*, 2002), where there is an association between a 3' noncoding region and disease resistance (Barthel *et al.*, 2001). Preliminary studies to investigate variation in the NRAMP gene of bighorn sheep identified a degree of polymorphism in linked microsatellites. Early indications suggest that one allele confers resistance to *Pasteurella multilocida* infection (Rudolph *et al.*, 2001). However, as with IFNG there have been no specific tests for selection within the NRAMP region.

5.1.2.4. Prion protein

The development of scrapie occurs following contact with prion infectious agents. Development of disease is closely linked to nucleotide variants in exon 3 of the host PrP gene, which although not strictly an immune gene, encodes the host prion

protein. Indeed, mouse PrP knockout mutants show complete resistance to prion disease (Bueler *et al.*, 1992). The ovine exon is 771 base pairs in length, containing at least 14 amino acid polymorphisms (Seabury *et al.*, 2004), three of which are acutely linked to the occurrence of scrapie: A136V, R154H and Q171R/H (Baylis & Goldmann, 2004; Caplazi *et al.*, 2004). These generate five commonly observed alleles: ARQ, ARR, AHQ, ARH and VRQ (Baylis & Goldmann, 2004). ARR and AHQ are associated with scrapie resistance while ARQ, ARH and VRQ are associated with susceptibility. In addition, there are subtle effects of specific pairs of alleles (genotypes), and different sheep breeds vary in the assortment of the five alleles that they predominantly encode.

5.1.3. Outline

This chapter details a study aiming to examine sequence variation within four immune related genes on thinhorn sheep. Specifically I aim to explore the following objectives;

i) Characterisation of MHC, IFNG, NRAMP and PrP coding regions of thinhorn sheep. None of these regions has been studied before in this species, while only MHC genes are characterised in any wild sheep species.

ii) Test patterns of variation within all four chosen thinhorn immune genes for evidence of selection. I expect to find strong evidence of balancing selection in the MHC gene DRB3 as is common is this region. I also expect to find evidence of balancing selection within IFNG and NRAMP, although weaker than that expected in the MHC. The lack of scrapie in thinhorn sheep predicts little if any variation within the PrP gene.

iii) Compare levels of variation in thinhorn immune genes to those of other species. Sequences will be collated from GenBank databases and compared to those characterised in this chapter. I expect to find similar levels of variation for all genes

except PrP (where lower levels of variation are expected) to those present in other sheep species.

5.2. MATERIALS AND METHODS

5.2.1. Sequencing samples

Due to the lower sample sizes required for sequence characterisation and logistical constraints, samples chosen for sequencing were representative of the eight regions defined by a Bayesian population structure analysis (Chapter 2) rather than the 24 sample areas. All samples chosen had a high confidence of belonging to one of these regions (P>0.9). Where possible (i.e. IFNG and NRAMP), samples that were homozygous at the closest linked microsatellite locus were sequenced. A small subset of only ten samples was chosen for the characterisation of prion protein sequences.

In addition to thinhorn samples, several bighorn sheep were also sequenced over MHC, IFNG and NRAMP gene regions. All individuals originated from the Ram Mountain region of the Canadian Rocky Mountains, a long term monitored study population (Coltman *et al.*, 2002).

5.2.2. Direct sequencing

5.2.2.1. Major histocompatibility complex

Exon 2 of the class II MHC gene DRB3 was amplified in 22 thinhorn samples from across the species range and two bighorn samples. I used primers modified from those of Sigurdardottir *et al.* (1991) (LA31-K, 5'-ATCCTCTCTCTGCAGCACAT-TTCCT-3', LA32-K, 5'-TCACCTCGCCGCTGCACA-3') which amplified a 249 base pair fragment.

5.2.2.2. Interferon gamma

Two fragments of IFNG were amplified in 35 thinhorn samples (mean 4.4 per STRUCTURE cluster) and one bighorn sample using the primer pairs IFNG_F3 and IFNG_R3, IFNG_F7 and IFNG_R7a (Jake Gratten, pers. comm.). The first set

amplifies a fragment encompassing exons 2, 3 and surrounding partial intronic sequence. The second primer set amplifies exon 4 and surrounding intronic regions. A total of 1256 base pairs of IFNG sequence were used for analyses.

5.2.2.3. Natural resistance associated macrophage protein

A continuous 980 base pair NRAMP fragment was amplified in 43 thinhorn and six bighorn by two primer sets designed using primer 3 (Rozen & Skaletsky, 2000) from aligned GenBank sequences of the region (NRAMP-2F, 5'-CTCTCCTCTGGC-TGACCATC-3', NRAMP-2R, 5'-CACGATGGTGATGAGGACAC-3', NRAMP-3F, 5'-GTGGGAGATCCAGACTCCTG-3', NRAMP-3R, 5'-CCGAAGGTCAAA-GCCATTAT-3'). The complete NRAMP amplicon comprised exons 5, 6 and 7 and separating introns; the most variable regions in numerous species.

5.2.2.4. Prion Protein

A fragment of exon 3 from thinhorn prion protein was amplified in ten thinhorn by primers designed from a GenBank sequence from bighorn sheep (accession number AF166334) (PrP-F, 5'-GGATCCTGGTTCTCTTTGT-3', PrP-R, 5'-AGAGATGA-GGAGGATCACAG-3'). A total of 603 out of the total length of 771 base pairs were used for analysis, including all residues associated with scrapie resistance and susceptibility (Tranulis, 2002).

5.2.2.5. Common amplification protocol

The low quality of DNA extracted from horn material results in the failure of larger volume PCRs. As a result five different 10µl volume reactions were used to amplify each gene product. The same protocol was used to amplify bighorn DNA, originally extracted from tissue samples using a phenol chloroform method (see Appendix for further information). Each reaction comprised 2µl of DNA template, 40 µmol of each primer, 0.16mM dNTP's, 1.5mM $MgCl_2$ (2mM for IFNG_F7 and IFNG_R7a) and 0.5 units *Taq* polymerase (Bioline, London). The PCR profile comprised forty cycles of 94°C for 30s, annealing for 30s (DRB; 55°C, IFNG_3; 58°C, IFNG_7; 60°C, NRAMP-2; 57°C, NRAMP-3; 57°C, PrP; 50°C) and 72°C for 40s. Cycles were

preceded by 5 min at 94°C and terminated with 10 min at 72°C. PCR products from the same individual were pooled and cleaned using a standard gel purification kit (Qiagen, Crawley, West Sussex, UK). The concentration of the purified PCR product was estimated by running 1µl alongside a ladder of a known concentration (Φx174 RF DNA Hae III, ABgene, Epsom, UK) in 1.5% agarose. Approximately 30ng DNA was cycle sequenced following a standard ABI protocol. After precipitation sequences were run on an ABI 3730 automated sequencer.

5.2.3. Cloning and sequencing exon 2 of the MHC gene DRB

Gel purified PCR product of heterozygous samples was ligated into pGEM-T vector (Promega, Southampton, UK) in 10µl reactions containing approximately 10ng insert DNA. Reactions were incubated overnight at 4°C before transformation into JM109 *E. coli* by heat shock (see Appendix for further methods). Cells were plated onto Amp/IPTG/X-Gal agar and cultured for 24 hours before picking positive clones. At least eight positive clones were grown further in LB glycerol overnight for each sample before plasmid miniprep (Promega, Southampton, UK) and sequencing with both M13 (-21) forward and LA32-K reverse primers. Following precipitation, sequences were run on an ABI 3730 automatic sequencer.

All sequences were edited and aligned using the software SeqScape (Applied Biosystems, Foster City, USA). Polymorphic sites were identified in alleles from all immune regions.

5.2.4. Measures of sequence diversity

To examine the relative diversity of the MHC region of the thinhorn immune system compared to other ungulate species a total of 462 additional DRB sequences were collected from GenBank (Table 5.1, full GenBank accession numbers in Appendix). Relative genetic diversity within taxa was examined by calculating Kimura 2-parameter distances between sequence alleles with standard error measured by 1000

replicates. Bootstrapped (2000 replications) neighbour-joining trees were constructed using Jukes-Cantor distances for all immune genes. All unique alleles from other species currently on the GenBank database were included (IFNG; 19, NRAMP; 12). Additional trees were constructed for thinhorn DRB alleles using only synonymous or non-synonymous substitutions.

Genetic distances between thinhorn DRB alleles were calculated in MEGA (Kumar *et al.*, 2001) using the Kimura 2-parameter model. The relative rate of synonymous (d_S) and nonsynonymous (d_N) substitutions in DRB alleles within all taxa was estimated in MEGA (Kumar *et al.*, 2001), using the distance model of Nei & Gojobori (1986) and applying a Jukes-Cantor correction for multiple substitutions at a site. Deviation from neutral expectations (in thinhorn only) was tested by comparing the difference between these rates by using t tests with an infinite number of degrees of freedom,

$$t = \frac{d}{s(d)}$$

where $s(d)$ is the standard error of d,

$$s(d) = \left[Var\left(\bar{d}_s\right) + Var\left(\bar{d}_N\right)\right]^{1/2}$$

Analyses were also conducted on the 16 putative peptide binding codons and 67 non-peptide binding codons of DRB alleles. Again relative rates of synonymous and non-synonymous substitutions were calculated. Variation in variable residues across the DRB coding region were examined and related to the putative protein binding region.

Within each immune region the mean nucleotide diversity across the length of the sequence was calculated. HKA tests (Hudson *et al.*, 1987) were used to test for positive selection within and between pairs of immune genes. Both the above were calculated using DnaSP (Rozas & Rozas, 1999). The HKA test uses an infinite sites model to compare levels of polymorphism within a species to the divergence at a gene between species. It can also be used to compare polymorphism and divergence between two genes within the same species or to compare two parts of the same gene with different functions. Where there is no selection the expectation is that levels of

polymorphism are proportional to divergence. In this study any evidence of selection found should be taken cautiously, as population structure violates several assumptions of the test statistic.

5.3. RESULTS

5.3.1. Sequence polymorphism of DRB3 exon 2

5.3.1.1. Polymorphism within *O. dalli*

Seventeen thinhorn DRB3 sequence alleles were characterised from 22 thinhorn samples (11 homozygotes and 11 heterozygotes), with 57 out of the 249 bases showing polymorphism (Figure 5.2a). Alleles have been deposited in GenBank under the accession numbers AJ920396-AJ920412. Comparison of sequences to those from other species available in GenBank showed that three of the seventeen thinhorn alleles, *O. dalli**2, 4 and 6, correspond to alleles 20, 18 and 2 (Gutierrez-Espeleta *et al.*, 2001) in the sister species bighorn sheep (*O. canadensis*). An additional thinhorn allele (*O. dalli**8) is shared with the *O. aries* allele DRB1*0322 (Kostia *et al.*, 1998). There was great amino acid diversity between sequences with the seventeen thinhorn nucleotide alleles corresponding to sixteen amino acid sequences in the translated protein (Figure 5.1b, sequences *O. dalli** 12 and *O. dalli**13 are identical with respect to amino acid sequence).

The distribution of sequence alleles showed evidence of geographic structure. Seven unique haplotypes were present in northern Alaska, two were unique to central Alaska, two to the Mackenzie ranges of the Northern Territories and one to populations from British Columbia (Table 5.1). The remaining five haplotypes were found in more than one mountain range. A neighbour joining tree of thinhorn DRB3 alleles reflects this geographical structuring in topology (Figure 5.3A). Sequences found in more than one sample region were largely from neighbouring mountain ranges. Tree topology varied among trees drawn from different nucleotide classes as did the degree of geographical structure revealed (Figure 5.3B and C). The topologies of the total tree (A) and that drawn from nonsynonymous sites (C) were very similar. A tree constructed from synonymous sites (B) appeared very different and showed increased geographical structuring of alleles.

Nucleotide diversity (π) across the entire 249 base pair fragment length of thinhorn was 0.0519±0.0019. A significant excess of nonsynonymous substitutions was evident in DRB3 sequences (Table 5.2, $d_N/d_S = 1.925$, $P<0.01$), the magnitude of the excess increasing further in the putative protein binding region ($d_N/d_S = 6.456$, $P<0.001$). Codons not involved in protein binding did not show a difference in substitution rate ($d_N/d_S = 1.308$, $P=0.801$). The excess of nonsynonymous substitutions in the putative PBR was reflected in increased levels of amino acid diversity (Figure 5.4); 88% of PBR codons were polymorphic compared to 27% of other residues. Tajima's test of neutrality was non-significant, indicating the absence of selection ($D = 0.189$, $P>0.1$).

```
                                                                                                    50
O. dalli*1          GAGTATCATAAGAGCGAGTGTCGTTTCTCCAACGGGACGGAGCGGGTGCG
O. dalli*2          ......GC..............A.....T.....................
O. dalli*3          ......AC.....AA.....................................
O. dalli*4          ......GC.....................T.....................
O. dalli*5          ......GC..............A.....T.....................
O. dalli*6          ......T.....G.........A.....T.........C...........
O. dalli*7          ......................A..............C...........
O. dalli*8          ......................A.....T.....................
O. dalli*9          .......G.....................T.....................
O. dalli*10         ............................T.......................
O. dalli*11         ...................................................
O. dalli*12         ...................................................
O. dalli*13         ...................................................
O. dalli*14         ......G..............T....T.T.....................
O. dalli*15         ......??....???.......A.....T.....................
O. dalli*16         ......................A.....T.....................
O. dalli*17         ......??....???......??..............C...........
O. canadensis*1     ......T.....G.........A.....T.........C...........
O. canadensis*2     ......GC.....................T.....................

                                                                                                   100
O. dalli*1          GTTCCTGGACAGATACTTCTATAATGGAGAAGAGTACGCGCGCTTCGACA
O. dalli*2          ..................................A...T...........
O. dalli*3          .....................................T...........
O. dalli*4          .........A...........................T...........
O. dalli*5          ..................................A...T...........
O. dalli*6          ...G...C......T..A.AC.............AC..T...........
O. dalli*7          ...G..............C................T..T...........
O. dalli*8          .....................................T...........
O. dalli*9          ...................................................
O. dalli*10         ...................................................
O. dalli*11         ...................................................
O. dalli*12         ...................................................
O. dalli*13         ...................................................
O. dalli*14         ...................................................
O. dalli*15         .....................................T...........
O. dalli*16         .....................................T...........
O. dalli*17         ...G.................................T.............
O. canadensis*1     ...G...C......T..A.AC.............AC..T...........
O. canadensis*2     .........A...........................T...........
```

Figure 5.2a. Nucleotide sequences of the 17 thinhorn and two bighorn DRB3 alleles characterised in this thesis. Bold type indicates nucleotides within the PBR.

```
                                                                   150
O. dalli*1         GCGACTGGGGCGAGTACCGAGCGGTGGCCGAGCTGGGGCGGCGGAGCGCC
O. dalli*2         ............................................GA....
O. dalli*3         A...........................................GA....
O. dalli*4         A...........................................GA....
O. dalli*5         ............................................GA....
O. dalli*6         ...............T...G......A..C...A.....A.GA.GA....
O. dalli*7         ...............T...G......................C.GC....
O. dalli*8         ...............T...G......................C.GA....
O. dalli*9         ..........................................C.GA....
O. dalli*10        ...................G......................C.GA....
O. dalli*11        ..................................................
O. dalli*12        ...................G......................C.GA....
O. dalli*13        .............................................A....
O. dalli*14        ...................G..............................
O. dalli*15        ...............T...G......................C.GA....
O. dalli*16        ...............?...G........................G.....
O. dalli*17        ..........................................C.GC....
O. canadensis*1    ...............T...G......A..C.A.A.....A.GA.GA....
O. canadensis*2    A.........................................C.GA....

                                                                   200
O. dalli*1         GAGTACTGGAACAGCCAGAAGGAGCTCCTGGAGCGGAGGCGGGCCGAGGT
O. dalli*2         A....................................A.......A.T..
O. dalli*3         A.......................A............A.......A.T..
O. dalli*4         A.......................A............A.......A.T..
O. dalli*5         A.................................A..C............
O. dalli*6         ...C....................A.........A..A............
O. dalli*7         .......AC..............CT........A.C...A......C...
O. dalli*8         A......................CT........A.C...A..A...C...
O. dalli*9         A.................................................
O. dalli*10        A...................................GC........CC..
O. dalli*11        ..................................A..C............
O. dalli*12        A...................................GC........CC..
O. dalli*13        ..................................A..C............
O. dalli*14        ....................................GC........C...
O. dalli*15        A......................CT........A.C...A..?.......
O. dalli*16        A......................CT........A.C...A..A...C...
O. dalli*17        .......................C..........................
O. canadensis*1    ...C....................A.........A..A............
O. canadensis*2    A.......................A............A.......A.T..
```

```
                                                                    249
O. dalli*1          GGACACGTACTGCAGACACAACTACGGGGTCATTGAGAGTTTCACTGTG
O. dalli*2          ..............................T............G....
O. dalli*3          ..............................GG..............
O. dalli*4          ..............................GG..............
O. dalli*5          ..............................GG..............
O. dalli*6          .......GTG.............T......T..............
O. dalli*7          ....................................G....
O. dalli*8          ..............................GG..............
O. dalli*9          ..............................GG..............
O. dalli*10         ..............................GG..............
O. dalli*11         ................................................
O. dalli*12         ..............................GG..............
O. dalli*13         ................................................
O. dalli*14         ..............................??...........?....
O. dalli*15         ..............................GG..............
O. dalli*16         ..............................??...........?....
O. dalli*17         ................................................
O. canadensis*1     .......GTG.............T......G..............
O. canadensis*2     ..............................GG..............
```

```
O. dalli*1          EYHKSECRFSNGTERVRFLDRYFYNGEEYARFDSDWGEYRAVAELGRR
O. dalli*2          ..A....H.F.................NV.................P
O. dalli*3          ..T.K.......................V...N.............P
O. dalli*4          ..A......F.........E.........V...N.............P
O. dalli*5          ..A....H.F.................NV.................P
O. dalli*6          ..Y.G..H.F.......L.H.FYT....TV........F...TQQ.QE
O. dalli*7          .......H.........L.....H....FV........F.........P
O. dalli*8          .......H.F.................V........F.........P
O. dalli*9          ..R......F.....................................P
O. dalli*10         .....................................P
O. dalli*11         .....................................
O. dalli*12         .....................................P
O. dalli*13         .....................................
O. dalli*14         ..D....C.F...........................
O. dalli*15         ..?.?..H.F.................V........F.........P
O. dalli*16         .......H.F.................V........X.........
O. dalli*17         ..?.?..?.........L..........F.................P
O. canadensis*1     ..Y.G..H.F.......L.H.FYT....TV........F...TQQ.QE
O. canadensis*2     ..A......F.........E.........V...N.............P
```

```
O. dalli*1          SAEYWNSQKELLERRRAEVDTYCRHNYGVIESFTV
O. dalli*2          D.K...........K..N...........F...S.
O. dalli*3          D.K.......I...K..N...........G.....
O. dalli*4          D.K.......I...K..N...........G.....
O. dalli*5          D.K..........QT..............G.....
O. dalli*6          D..H......I..QK......V.......F.....
O. dalli*7          A...Y....DF..S...A..............S.
O. dalli*8          D.K......DF..S..TA...........G.....
O. dalli*9          D.K..........................G.....
O. dalli*10         D.K...........A..A...........G.....
O. dalli*11         .............QT..................
O. dalli*12         D.K...........A..A...........G.....
O. dalli*13         N............QT..................
O. dalli*14         ............LEA..A...........?...?.
O. dalli*15         D.K......DF..S..?............G.....
O. dalli*16         G.K......DF..S..TA...........?...?.
O. dalli*17         A........D.......................
O. canadensis*1     D..H......I..QK......V.......V.....
O. canadensis*2     D.K.......I...K..N...........G.....
```

Figure 5.2b. Protein translations of all thinhorn and bighorn DRB3 sequence alleles.

Table 5.1. Distribution of DRB3, IFNG and NRAMP sequence alleles across the species range. Regions are based on clusters from STRUCTURE analyses of neutral microsatellites, NA = northern Alaska, CA = central Alaska, OG = Ogilvie range, NT-N = Northwest Territories (north), NT-S = Northwest Territories (south), PY = Pelly range, SW = Southwest Yukon, BC = British Columbia. Sample region identification in parentheses represent codes shown in Chapter 2 (Figure 2.4). Values in the table represent the number of alleles found within each sample region.

	Sample Region							
Allele	**NA (A)**	**CA (B)**	**OG (C)**	**NT-N (D)**	**NT-S (E)**	**PY (F)**	**SW (G)**	**BC (H)**
*DRB*1*		1	1	3		1		1
*DRB*2*	2							
*DRB*3*	2							
*DRB*4*						2	1	3
*DRB*5*		2	1				1	
*DRB*6*	2							
*DRB*7*		2						1
*DRB*8*				1	6	1		
*DRB*9*	1							
*DRB*10*		2						
*DRB*11*	1							
*DRB*12*	1							
*DRB*13*	1							
*DRB*14*		1						
*DRB*15*				1				
*DRB*16*				1				
*DRB*17*								1
*IFNG*1*	8							
*IFNG*2*	20	4			4	1	4	6
*IFNG*3*	2	8	2	4	2	3	2	2
*NRAMP*1*	10	10	8	6	8	2	5	4
*NRAMP*2*						8	2	15
*NRAMP*3*						2	1	5

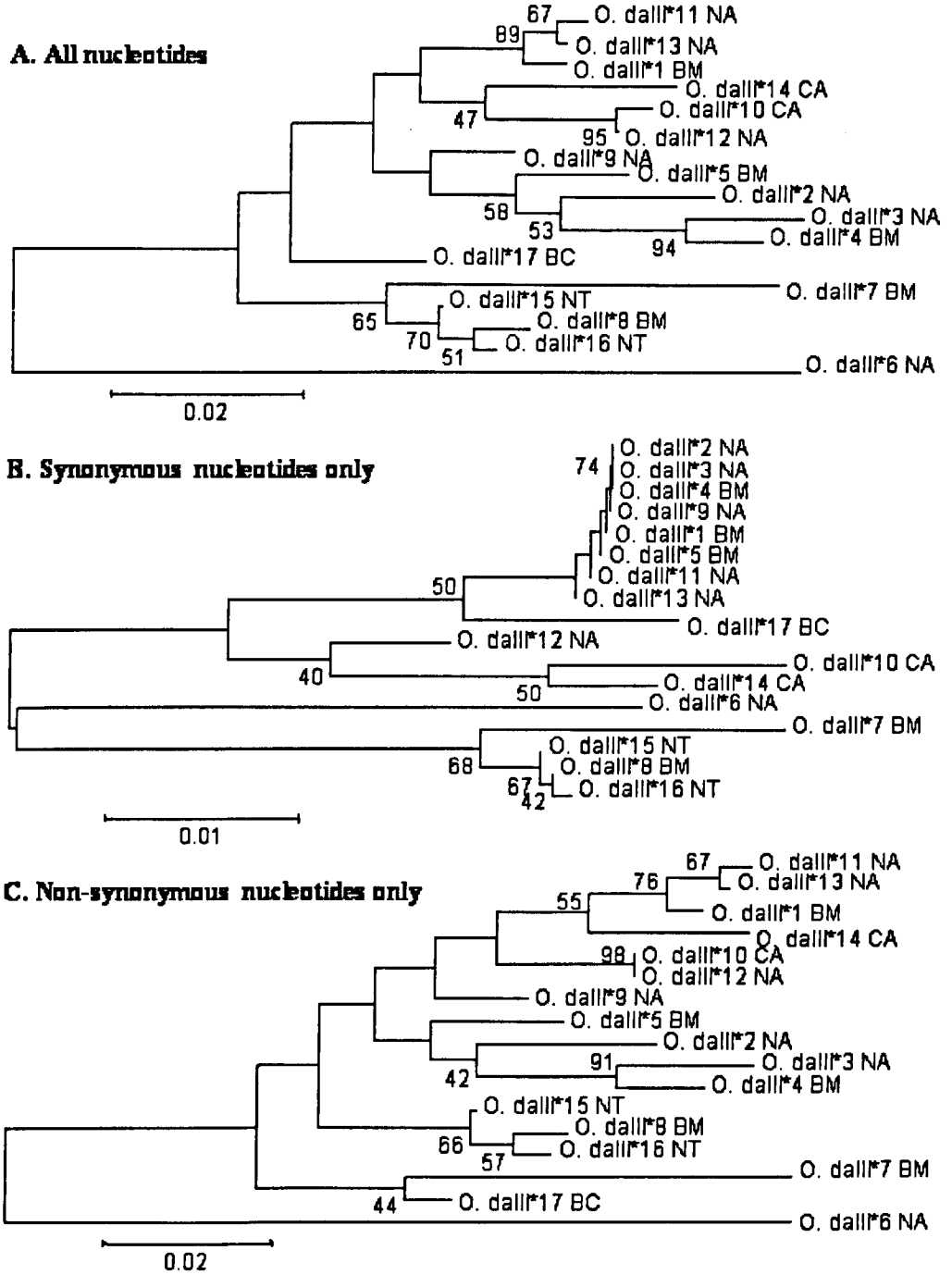

A. All nucleotides
O. dalli*11 NA
O. dalli*13 NA
O. dalli*1 BM
O. dalli*14 CA
O. dalli*10 CA
O. dalli*12 NA
O. dalli*9 NA
O. dalli*5 BM
O. dalli*2 NA
O. dalli*3 NA
O. dalli*4 BM
O. dalli*17 BC
O. dalli*7 BM
O. dalli*15 NT
O. dalli*8 BM
O. dalli*16 NT
O. dalli*6 NA
67
89
47
95
58
53
94
65
70
51
0.02
B. Synonymous nucleotides only
O. dalli*2 NA
O. dalli*3 NA
O. dalli*4 BM
O. dalli*9 NA
O. dalli*1 BM
O. dalli*5 BM
O. dalli*11 NA
O. dalli*13 NA
O. dalli*17 BC
O. dalli*12 NA
O. dalli*10 CA
O. dalli*14 CA
O. dalli*6 NA
O. dalli*7 BM
O. dalli*15 NT
O. dalli*8 BM
O. dalli*16 NT
74
50
40
50
68
67
42
0.01
C. Non-synonymous nucleotides only
O. dalli*11 NA
O. dalli*13 NA
O. dalli*1 BM
O. dalli*14 CA
O. dalli*10 CA
O. dalli*12 NA
O. dalli*9 NA
O. dalli*5 BM
O. dalli*2 NA
O. dalli*3 NA
O. dalli*4 BM
O. dalli*15 NT
O. dalli*8 BM
O. dalli*16 NT
O. dalli*7 BM
O. dalli*17 BC
O. dalli*6 NA
67
76
55
98
42
91
66
57
44
0.02

Figure 5.3. (Previous page) A comparison of *O. dalli* DRB3 gene trees produced from including different classes of nucleotide. In all cases Jukes Cantor distances are used with support provided by bootstrapping with 2000 replicates. Nodes with more than 40% support are shown. Letters following thinhorn alleles encode the geographic locations of the respective sequences, NA = Northern Alaska, CA = Central Alaska, BC = British Columbia, NT = Northern Territories, BM = alleles found between mountain ranges.

Table 5.2. Rate of non-synonymous (d_N) and synonymous (d_S) substitutions at the MHC gene DRB3. *P* is the probability of the significance of the ratio d_N/d_S.

Position	Codon number	d_N	d_S	d_N/d_S	*P*
Antigen binding	16	0.312	0.048	6.456	<0.001
Non-antigen binding	67	0.043	0.033	1.308	0.801
All	83	0.077	0.040	1.947	<0.010

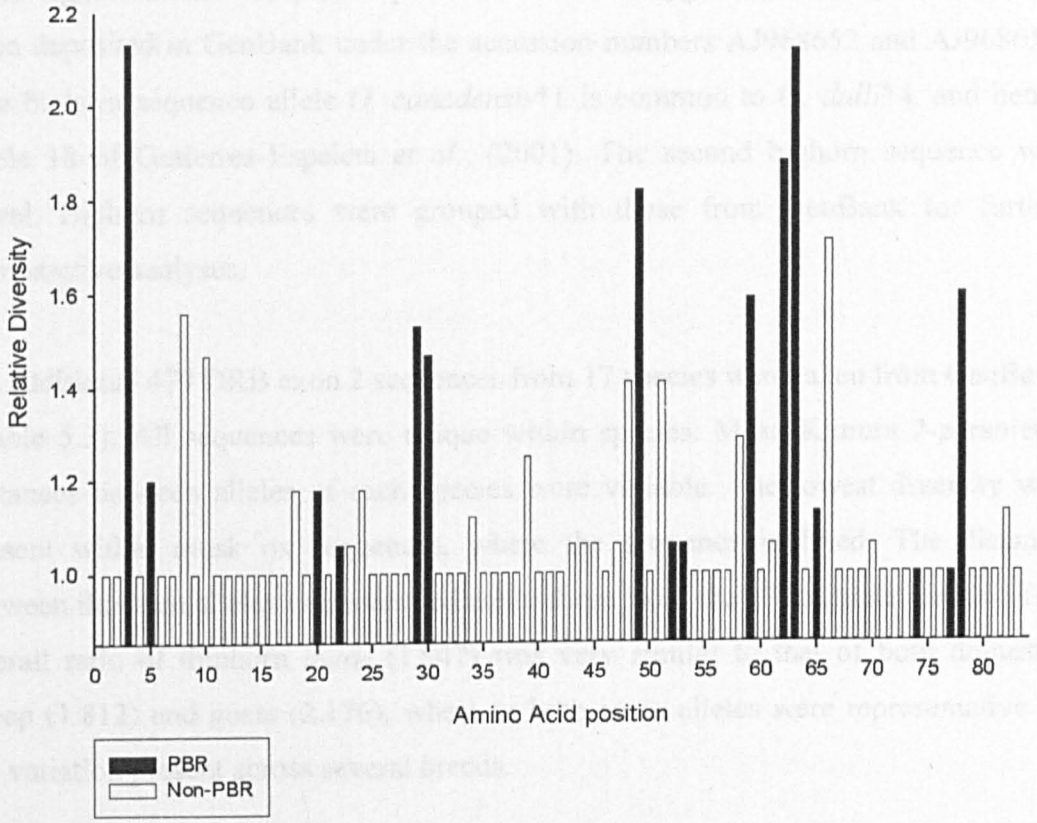


Figure 5.4. Amino acid diversity is variable across the seventeen thinhorn DRB3 alleles. Relative diversity is a measure of mean polymorphism at a codon where different residues are numbered from one upwards, where one signifies the most frequent residue. A relative diversity of one indicates monomorphism at a residue site. Shaded bars represent locations of the putative PBR region, indicating the increased diversity at these sites.

5.3.1.2. Nucleotide substitutions between species

Both bighorn DRB3 samples sequenced were homozygotes, their sequences having been deposited in GenBank under the accession numbers AJ968652 and AJ968653. The bighorn sequence allele *O. canadensis**1 is common to *O. dalli**4, and hence allele 18 of Gutierrez-Espeleta *et al.*, (2001). The second bighorn sequence was novel. Bighorn sequences were grouped with those from GenBank for further comparative analyses.

An additional 479 DRB exon 2 sequences from 17 species were taken from GenBank (Table 5.3). All sequences were unique within species. Mean Kimura 2-parameter distances between alleles of each species were variable. The lowest diversity was present within musk ox sequences, where the sequence is fixed. The distance between thinhorn alleles was intermediate to those from the 17 ungulate species. The overall ratio of thinhorn d_N/d_S (1.947) was very similar to that of both domestic sheep (1.812) and goats (2.176), where in both cases alleles were representative of the variation present across several breeds.

A gene tree of all available DRB sequences was very complex and showed alleles from most species in Table 5.3 to be polyphyletic (data not shown). The exceptions were *Alces alces* and *Rupicapra* spp, the former branching from within *Cervus elaphus*, the second falling within a clade comprised entirely of other Caprinae. A second tree comprising a subset of 73 alleles showed thinhorn alleles occurring throughout those of other sheep and goat species (Figure 5.5). There was some separation between these sequences and those of cattle species. Two main groupings of thinhorn sequences alleles (Figure 5.5, A and B) are present, separated by cattle sequences and a large degree of genetic distance, although these are not supported as clades by bootstraps. Further interpretation can not be made on the topology of the reduced DRB sequence tree, as its appearance has been simplified from that of the complete tree, which shows a much greater distribution of sequences from individual taxa.

5.3.1.3. Association between sequence and microsatellite alleles

I was unable to fully infer associations between DRB3 sequences and alleles at the locus OLADRB due to the presence of null alleles at the microsatellite. By only taking individuals that were homozygous for sequence alleles eight putative associations were made and referenced with heterozygote sequences (Table 5.4). I assumed that at least one copy of the sequence allele was associated with the amplifying microsatellite allele. Comparisons did not make any assumptions about two different OLADRB alleles (in this case one of which was null) associating with only one sequence as was shown to be the case in a study of domestic sheep (Paterson, 1996). In all but one case each sequence allele could be associated with a different microsatellite genotype. In two cases two sequence alleles showed an association to a common microsatellite allele; both DRB*4 and DRB*5 relating to OLADRB allele of size 213 and an assumed association between the microsatellite allele 163 and sequences DRB*10 and either/both DRB*11 and DRB*12. No association could be made between DRB3 sequence and genotype at the microsatellite OLADRBps, which is located further from the coding region.

Table 5.3. Mean genetic distances between DRB3 sequence alleles within species calculated using different genetic distance measures and nucleotide class. A Jukes Cantor distance measure with Nei & Gojobori (1986) correction was used to calculate d_S and d_N. Standard error was calculated by 1000 replications of the data.

Species	Common Name	N	Kimura 2-parameter (SE)		d_S (SE)		d_N (SE)		d_N/d_S
Ovis dalli	Thinhorn sheep	17	0.0674	(0.0097)	0.0395	(0.0157)	0.0768	(0.0145)	1.9470
Ovis canadensis	Bighorn sheep	24	0.0854	(0.0115)	0.0375	(0.0142)	0.1003	(0.0172)	2.6727
Ovis aries	Domestic sheep	94	0.0798	(0.0108)	0.0495	(0.0134)	0.0898	(0.0165)	1.8124
Capra hircus	Domestic goat	18	0.0936	(0.0117)	0.0493	(0.0137)	0.1073	(0.0185)	2.1755
Capra aegagrus	Bezoar goat	36	0.0928	(0.0116)	0.0629	(0.0180)	0.1024	(0.0185)	1.6267
Capra pyrenaica	Spanish Ibex	6	0.1193	(0.0156)	0.0770	(0.0230)	0.1332	(0.0244)	1.7307
Rupicapra pyrenaica	Pyrenean Chamois	13	0.0467	(0.0088)	0.0096	(0.0057)	0.0580	(0.0146)	6.0444
Rupicapra rupicapra	Alpine Chamois	15	0.0433	(0.0089)	0.0088	(0.0044)	0.0540	(0.0133)	6.1519
Bos taurus	Domestic cattle	66	0.0839	(0.0110)	0.0373	(0.0132)	0.0981	(0.0180)	2.6330
Bos indicus	Auroch	59	0.0902	(0.0120)	0.0396	(0.0132)	0.1063	(0.0202)	2.6874
Bison bison	Wood bison	18	0.0777	(0.0114)	0.0371	(0.0129)	0.0901	(0.0185)	2.4298
Bubalus bubalis	Water buffalo	26	0.1186	(0.0130)	0.1270	(0.0278)	0.1162	(0.0195)	0.9149
Syncerus caffer	African buffalo	10	0.0930	(0.0125)	0.0774	(0.0241)	0.0977	(0.0191)	1.2624
Alces alces	Moose	11	0.0191	(0.0059)	0.0094	(0.0087)	0.0222	(0.0101)	2.3602
Ovibos moschatus	Musk Ox	1	n/c	n/c	n/c	n/c	n/c	n/c	
Damaliscus pygargus	Blesbok	23	0.0655	(0.0105)	0.0224	(0.0100)	0.0790	(0.0173)	3.5290
Cervus elaphus	Red deer	44	0.1092	(0.0120)	0.0687	(0.0152)	0.1222	(0.0188)	1.7786

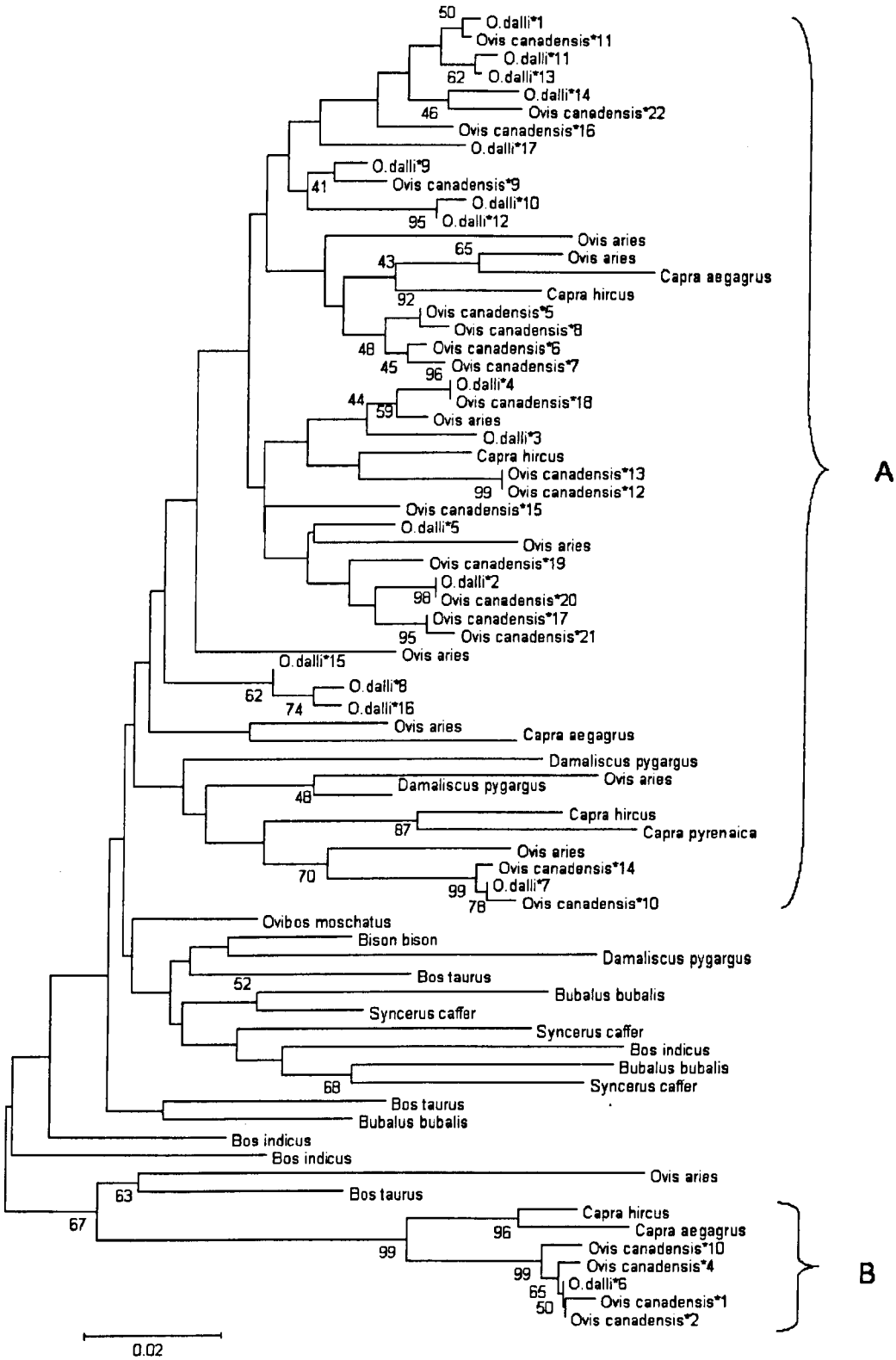

O.dalli*1
Ovis canadensis*11
O.dalli*11
O.dalli*13
O.dalli*14
Ovis canadensis*22
Ovis canadensis*16
O.dalli*17
O.dalli*9
Ovis canadensis*9
O.dalli*10
O.dalli*12
Ovis aries
Ovis aries
Capra aegagrus
Capra hircus
Ovis canadensis*5
Ovis canadensis*8
Ovis canadensis*6
Ovis canadensis*7
O.dalli*4
Ovis canadensis*18
Ovis aries
O.dalli*3
Capra hircus
Ovis canadensis*13
Ovis canadensis*12
Ovis canadensis*15
O.dalli*5
Ovis aries
Ovis canadensis*19
O.dalli*2
Ovis canadensis*20
Ovis canadensis*17
Ovis canadensis*21
Ovis aries
O.dalli*15
O.dalli*8
O.dalli*16
Ovis aries
Capra aegagrus
Damaliscus pygargus
Ovis aries
Damaliscus pygargus
Capra hircus
Capra pyrenaica
Ovis aries
Ovis canadensis*14
O.dalli*7
Ovis canadensis*10
A
Ovibos moschatus
Bison bison
Damaliscus pygargus
Bos taurus
Bubalus bubalis
Syncerus caffer
Syncerus caffer
Bos indicus
Bubalus bubalis
Syncerus caffer
Bos taurus
Bubalus bubalis
Bos indicus
Bos indicus
Ovis aries
Bos taurus
Capra hircus
Capra aegagrus
Ovis canadensis*10
Ovis canadensis*4
O.dalli*6
Ovis canadensis*1
Ovis canadensis*2
B
0.02

Figure 5.5. (previous) Neighbour-joining tree of thinhorn and selected other ungulate DRB alleles. The scale refers to Jukes-Cantor distance and values indicate bootstrap significance (from 2000 replicates). Numbers following thinhorn and bighorn indicate the alleles as submitted to GenBank from this study and that of Gutierrez-Espeleta *et al.*, (2001) respectively.

Table 5.4. Associations between DRB3 sequence alleles and genotypes at the null microsatellite locus OLADRB. Values represent the total number of sequence copies associated with each microsatellite locus. Counts from sequence homozygotes are separated from heterozygotes, the latter are shown in parentheses.

Sequence	OLADRB genotype							
allele	148	163	167	169	201	213	215	234
DRB*1			(1)			(3)		(3)
DRB*2					2			
DRB*3							2	
DRB*4						4(2)		
DRB*5						2(2)		
DRB*6	2							
DRB*7				2(1)				
DRB*8						(1)		6(1)
DRB*9						(1)		
DRB*10		2						
DRB*11		(1)						
DRB*12		(1)						
DRB*13						(1)		
DRB*14			(1)					
DRB*15								(1)
DRB*16								(1)
DRB*17				(1)				

5.3.2. Variation in the interferon gamma gene

The combined 1256 base pair amplified region of thinhorn IFNG contained only two polymorphic base positions; a rare synonymous substitution in exon 4 restricted to northern Alaska and a polymorphism in intron 3 common throughout sample areas (Table 5.1, Figure 5.6). The fixed sequence for exons 2 and 3 and surrounding non-coding sequence has been deposited in GenBank under the accession number AJ920413 (see Appendix for sequence), while the three alleles of the variable exon 4 region were deposited under AJ920414-AJ920416. The common coding sequence (Figure 5.6; *O. dalli**2 and *O. dalli**3) is identical to that of a domestic sheep haplotype (GenBank accession number X52640, allele C) whereas the Northern Alaskan (GAAR) nucleotide sequence (Figure 5.6; *O. dalli**1) variant is novel. The bighorn sequence was identical to thinhorn allele two, and hence *O. aries* allele C.

A neighbour joining tree of GenBank ungulate exonic IFNG sequences (intronic sequences were not available for most species) rooted with a human sequence showed clustering of thinhorn and domestic sheep sequences (Figure 5.7). The topology appears to reflect evolutionary relationships between taxa, with good bootstrap support for all major species branches. No tests of selection could be conducted on thinhorn intraspecific variation due to the lack of polymorphism between alleles. Mean nucleotide diversity (π) between thinhorn alleles was 0.0011±0.0002.

```
                                                                      55
O.dalli*1    CTCAGGTTACTGTCTTACACACCAAGCATGAGAGCCATTGGCTATAGTAACTCAT
O.dalli*2    ...........................................................
O.dalli*3    ...........................................................

                                                                     110
O.dalli*1    CTGATACTTATAGCAAAGACTAAGTACTAAGCGATAGAAGTGACAGAGACCCTGG
O.dalli*2    ...........................................................
O.dalli*3    ..............................A..............................

                                                                     165
O.dalli*1    GGATCATCTAAACAATCTCATACTTAGTGCAGAAAACAGAGATCTGGAGAGCTTC
O.dalli*2    ...........................................................
O.dalli*3    ...........................................................

                                                                     220
O.dalli*1    AGTGACTGGCTTAAAGTCAGATGTATTGCACACACAGAATTCACACCCACATTTT
O.dalli*2    ...........................................................
O.dalli*3    ...........................................................

                                                                     275
O.dalli*1    TCTGTCACTCCACTTCTGGTTTTTTTTCACTATATAATTTTCAGATCCCTGAAGT
O.dalli*2    ...........................................................
O.dalli*3    ...........................................................

                                                                     330
O.dalli*1    GATAAATTTATGGATATGTGATGTGCTCCATCCTGTTACAGCACAACTGTCAATT
O.dalli*2    ...........................................................
O.dalli*3    ...........................................................

                                                                     385
O.dalli*1    GAATTGTTATAATTTTAGTCTTTATCACTGAAGAAACCAACATTACATATTAAAG
O.dalli*2    ...........................................................
O.dalli*3    ...........................................................

                                                                     440
O.dalli*1    TTCATTACTGCTAGTGAAAATAATTTGTTTTAAAGAATACTTTGATTTTCTTGGG
O.dalli*2    ...........................................................
O.dalli*3    ...........................................................

                                                                     495
O.dalli*1    CCTAGACAGCAGAATATCCTAATGACTCATATGCTTGAAATTAATTTTGCTGTTT
O.dalli*2    ...........................................................
O.dalli*3    ...........................................................

                                                                     550
O.dalli*1    TCTTTCCCAATAGGTGGATGATCTGCAGATCCAGCGCAAAGCCATCAATGAACTC
O.dalli*2    ...........................................................
O.dalli*3    ...........................................................
```

```
                                                                      605
O.dalli*1    ATCAAGGTGATGAATGACCTGTCACCAAAATCTAACCTCAGAAAGCGGAAGAGAA
O.dalli*2    ........................G.............................
O.dalli*3    ........................G.............................

                                                  640
O.dalli*1    GTCAGAATCTCTTTCGAGGCCGGAGAGCATCAATG
O.dalli*2    ....................................
O.dalli*3    ....................................
```

Figure 5.6. Interferon gamma exon 4 and surrounding intronic sequences. Exonic sequence is highlighted.

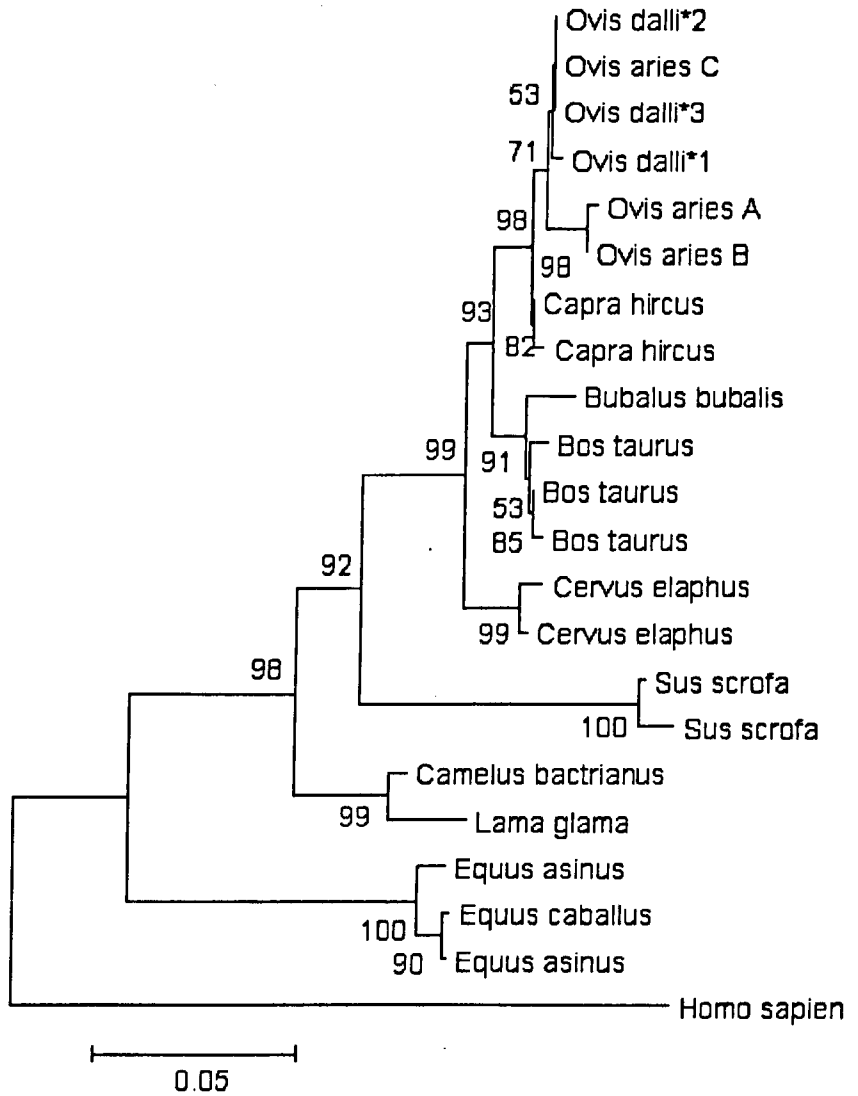


Figure 5.7. A bootstrapped neighbour joining tree of IFNG exons 2 to 4 constructed using MEGA and including all available ungulate species currently on GenBank. Genetic distance was calculated using Jukes Cantor method and bootstrap support was provided by 2000 replications.

All individuals with the sequence allele *O. dalli*1* shared a homozygote 184 KP6 genotype. Both microsatellite and sequence alleles were present only in northern Alaska. Although several other microsatellite alleles corresponded to only one sequence variant, few samples were available for a conclusion of linkage disequilibrium between markers to be made.

5.3.3. Variation in the NRAMP gene

The thinhorn NRAMP exonic sequence was fixed and differed from that of domestic sheep at one synonymous nucleotide position in exon 7. Four intronic polymorphisms were present, combining to give three haplotypes (Figure 5.8). Sequences for these haplotypes were deposited in GenBank under the accession numbers AJ920417-AJ920419. All available unique ungulate exons 5 through 7 were used to construct an NRAMP neighbour joining tree using Jukes Cantor distances. A mouse sequence was included as an outgroup (Figure 5.9). As with IFNG, the tree reflects the evolutionary relationships between taxa.

Only incomplete sequences were obtained for bighorn samples as the first primer set amplifying the region surrounding exon 5 failed to amplify a product. The bighorn exon 7 sequence was identical to that of thinhorn. Intronic alleles *O. dalli*2* and *O. dalli*3* were present in addition to a third intronic variant novel to bighorn. The novel sequence was the most prevalent, accounting for almost half the total number of copies.

The NRAMP*3 sequence was only present in combination with the OVINRAO1 microsatellite allele 218 and was restricted to the southern species range (Table 5.1). To test the robustness of this apparent association between sequence and microsatellite alleles, several OVINRAO1 heterozygotes with 218 alleles were sequenced. In all cases the sequence genotypes were also heterozygote, and all contained an NRAMP*3 allele. NRAMP sequence polymorphism shows geographical structure. Alaskan, northern Yukon and NWT samples share the same

fixed sequence while all three haplotypes are present in the southern range (Table 5.1).

5.3.4. Variation in the thinhorn PrP gene

Thinhorn PrP contained only one polymorphic nucleotide throughout the 603 base pair length of partial exon 3. Sequences have been deposited in GenBank under the accession numbers AJ969015 and AJ969016. Seventeen out of the twenty gene copies amplified were shared with a bighorn sequence (accession number AY769957). The rarer allele was common to that of several partial domestic sheep sequences (including that of AJ567984). Sequences differed at a synonymous nucleotide at the third base position of codon 103 (this position is adjusted for the total length of exon 3). The genotype of the three crucial scrapie related codons of both thinhorn alleles is ARQ (positions 136, 154 and 171 respectively), a form associated with disease susceptibility.

```
O.dalli*1     GTGCCCCGCACTCTCCTCTGGCTGACCATCGAGCTAGCCATCGTGGGCTCGGACA
O.dalli*2     ............................................................
O.dalli*3     ............................................................
O.canadensis  ............................................................

O.dalli*1     TGCAGGAAGTCATTGGCACAGCTATTGCGTTCAGTCTGCTCTCAGCCGGACGGTA
O.dalli*2     ............................................................
O.dalli*3     ............................................................
O.canadensis  ............................................................

O.dalli*1     CTCCAGAGGGGCCCCAGGTCTCCAGCCCAGGGCTGGGAACAACTGCCGCTTTCTC
O.dalli*2     ............................................................
O.dalli*3     ............................................................
O.canadensis  ............................................................

O.dalli*1     CCCAGGCCTTCTATTTCCCCCTGACCACTTCATCCTGCAGTCATCTCTAACTCAG
O.dalli*2     ............................................................
O.dalli*3     ............................................................
O.canadensis  ............................................................

O.dalli*1     TTATTGCCCAATCTGCCAGATGGGGAAACAGAGACCCAGATATGTAAAGCGACTT
O.dalli*2     ............................................................
O.dalli*3     ............................................................
O.canadensis  ............................................................

O.dalli*1     GCTGCAAATCAGAGGCTTCAGTACTCAGCCCCAAGCTCCCTACCACGCTGGTCCC
O.dalli*2     ............................................................
O.dalli*3     ............................................................
O.canadensis  ............................................................

O.dalli*1     CTGCCCAGCATGGAGGTGCCATTTTAGGTCTCCAGCACCCCGGCTCCCCTCTCCC
O.dalli*2     ............................................................
O.dalli*3     ............................................................
O.canadensis  ............................................................

O.dalli*1     TCAGCCGTCTGGGGCCCCTTGAGTGCCTGCTCCCTGGAGGCCCCATTCCACAGTC
O.dalli*2     ............................................................
O.dalli*3     ............T...............................................
O.canadensis  ............................................................
```

```
O.dalli*1     TCCAGCCCTGAGGGTGGGAGATCCAGACTCCTGAACCAGGCTAGGCTGACCCAGG
O.dalli*2     ....................................................
O.dalli*3     ....................................................
O.canadensis  ....................................................

O.dalli*1     CCACTCTGGTTTCAGAATCCCACTCTGGGGTGGTGTCCTCATCACCATCGTGGAC
O.dalli*2     ....................................................
O.dalli*3     ....................................................
O.canadensis  ....................................................

O.dalli*1     ACTTTCTTCTTCCTCTTCCTCGATAACTACGGTGGGTGTGTGCCCCTCATCCCCA
O.dalli*2     ....................................................
O.dalli*3     ....................................................
O.canadensis  ....................................................

O.dalli*1     GCGGGACTTGGGTGAGGAGGGGGCTGGCAATGAAAGTAATGGTGCTCTGGTGAAC
O.dalli*2     ....................................................
O.dalli*3     ............C.......................................
O.canadensis  ............C.......................................

O.dalli*1     TGGAAGAAGGGCTCTGAGCAACCTCGCTCTGCACAGCCTTGAGCAACTCAGACTA
O.dalli*2     ....................................................
O.dalli*3     .........A..........................................
O.canadensis  ....................................................

O.dalli*1     GTGTTTCTTCATCTACAACACGCAGATACTAGCACAGAGGTCACAGGGCTGTGGG
O.dalli*2     ....................................................
O.dalli*3     ....................................................
O.canadensis  ....................................................

O.dalli*1     GAGGATCCAGTGCGATAACACTGGTCAAGCCCGGGGCACAGTGCTGGAACAAAGT
O.dalli*2     ....................................................
O.dalli*3     ....................................................
O.canadensis  ....................................................

O.dalli*1     GCTAGGATATCAGCCATTAATAATTGCTTAAGTTAGTTTTGGTTTGGGGTGGTTT
O.dalli*2     ....................................................
O.dalli*3     ....................................................
O.canadensis  ....................................................
```

```
O.dalli*1     TTGTTGTTGTTGTTCAGCCCCTTTGCTCTCACCCTAGGGTTGCGGAAGCTGGAAG
O.dalli*2     C.....................................................
O.dalli*3     .....................................................
O.canadensis  .....................................................

O.dalli*1     CCTTTTTTGGATTTCTTATTACCATAATGGCTTTGACCTTCGGC
O.dalli*2     ...........................................
O.dalli*3     ...........................................
O.canadensis  ...........................................
```

Figure 5.8. Thinhorn NRAMP sequences from exon 5 through to exon 7. Putative exons are shown highlighted and in order from exon 5.

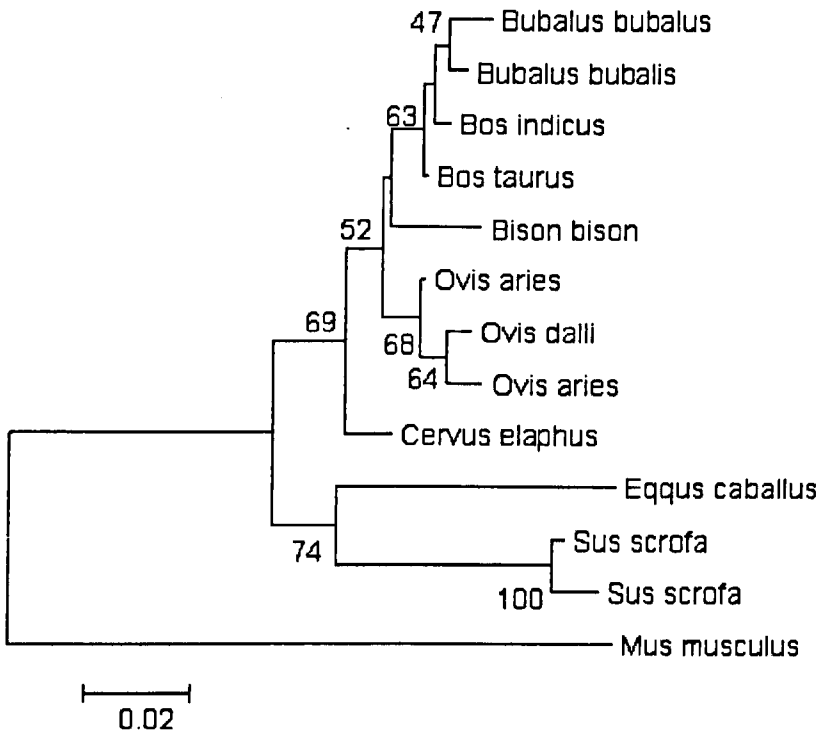


Figure 5.9. Neighbour joining tree of NRAMP coding sequences from exons 5 to 7 including all available GenBank ungulate species and rooted with a mouse sequence. Genetic distance was calculated using Jukes Cantor method and bootstrap support was provided by 2000 replications.

5.4. DISCUSSION

5.4.1. Variation within the major histocompatibility complex

Many previous studies to investigate variation in ungulate MHC genes involve species that have undergone population bottlenecks, such as moose and bison (Ellegren *et al.*, 1996; Mikko & Andersson, 1995; Mikko *et al.*, 1997). Bighorn sheep populations have also undergone bottlenecks due to pathogen related die-offs (Gutierrez-Espeleta *et al.*, 2001). The only study to be carried in sheep concerns the small population of the Soays' of St. Kilda (Paterson, 1996). Here characterisation of the MHC gene DRB3 was made from samples originating from one population. In contrast this chapter represents a study of sequences obtained from multiple populations of a wild sheep species with no history of demographic bottleneck or population collapse due to the introduction of novel disease.

Seventeen DRB3 alleles were found from 22 thinhorn sheep. This level of allelic diversity exceeds that of moose and bison, although is of a similar magnitude to both domestic (Schwaiger *et al.*, 1994) and bighorn sheep (Gutierrez-Espeleta *et al.*, 2001). Similarities in both allelic and amino acid diversity between sheep species are in contrast to their differing disease and population demographic histories. Theory predicts the presence of many alleles and high levels of heterozygosity within genes of the MHC region, a result of balancing selection (Hughes & Nei, 1988). Indeed these patterns are present in most species investigated. In contrast, population bottlenecks should reduce both the number of alleles and levels of heterozygosity at loci through the stochastic loss of genetic variation that occurs in small populations. Within the MHC expected patterns in diversity do not always correlate with historical demography. Several species that have undergone bottlenecks still exhibit extensive MHC variation (Aguilar *et al.*, 2004; Gutierrez-Espeleta *et al.*, 2001), while some species that have shown expansive population growth have low level variation within the MHC region (Mikko *et al.*, 1999). The lack of differential MHC

variability between sheep species suggests that extensive variation in the ovine MHC class II region may not confer resistance to current disease.

Despite the apparent lack of association between disease status and MHC diversity there are several lines of evidence suggesting that balancing selection is acting in both thinhorn and bighorn DRB3 sequences. The majority of variable sites within the thinhorn DRB3 exon are found in the first and second codon positions. This is reflected in significantly greater levels of nonsynonymous than synonymous substitutions. An increased rate of substitution at replacement sites is a key consequence of selection favouring new variants (Hughes & Nei, 1988), such as is expected in regions involved in pathogen response. Moreover, there is an approximate three-fold increase in the number of variable sites within the putative protein binding region (PBR) reflecting the functional significance of this region. Changes in amino acid in the PBR are more likely to produce differential responses to invading pathogens. The pattern of increased nonsynonymous substitutions, especially at regions involved in protein binding, has been widely documented in a number of species (Gutierrez-Espeleta *et al.*, 2001; Paterson, 1998; Pfau *et al.*, 1999; Hedrick *et al.*, 2001; Garrigan & Hedrick, 2001). It implies functionality rather than neutrality causes the accumulation and maintenance of variation at MHC genes. The lack of evidence of selection in non-PBR regions is evident from a non significant d_N/d_S ratio revealing that codons not involved in antigen recognition have undergone homogenisation through recombination or gene conversion.

I observed extensive sharing of variable residues in amino acid sequences across sheep species. However, five novel residues were present in thinhorn that were not observed in either bighorn or domestic sheep. Of these residues, four are located at codons within the PBR. The distribution of variable amino acids differs from the other two sheep species, where most novel residues are located in codons not involved with protein binding. It is unclear whether differences in amino acid polymorphism between species are due to differences in selective pressure or if

mechanisms generating polymorphisms vary in regard to site between species (Swarbrick *et al.*, 1995).

The phylogeny of ungulate MHC-DRB3 sequences also provides evidence of selection. Balancing selection acts to maintain allelic variants over long time periods resulting in trans-species polymorphisms (the presence of one allele in two or more different species). Several of the thinhorn alleles characterised here can also be found in domestic and bighorn sheep. Trans-species polymorphism results from deep allelic coalescence, where allelic lineages appear older than those of the species from which they originate. Modern thinhorn and bighorn sheep are relatively new species, having evolved from an ancestral species which colonised North America via Beringia during the last ice age. Mitochondrial DNA estimates a divergence time of between 70,000 and 340,000 years ago between thinhorn and bighorn populations (Loehr *et al.*, 2005). The presence of shared DRB3 alleles indicates a more recent time of divergence not reflected in neutral gene regions, a pattern expected from genes under balancing selection. The maintenance of DRB3 allelic lineages is most pronounced in a division of sheep that predates even the divergence of cattle and sheep species (Figure 5.5). Such strong evidence of ancient MHC lineages can arise from both convergent evolution and the retention of ancestral alleles (Yeager & Hughes, 1999). Problems associated with using a region under the influence of selection to reconstruct species phylogeny are clearly seen in Figure 5.3., where tree topologies differ with the type of mutation within the gene. An excess of non-synonymous sites tends to increase branch lengths, inflating the genetic distance between species.

The DRB region constitutes a family of genes of which only DRB3 is considered here. Although only one copy of this gene is presumed to be expressed in cattle, two copies are expressed in sheep (Dutia *et al.*, 1994). The deepest divergence of MHC-DRB3 alleles (Figure 5.5) separates sequences into two groups (A and B) divided by a cluster of cattle alleles. This relates to an estimated divergence time between groups A and B of approximately 19 million years ago. It may be that the larger group of alleles (A) represent one expressed gene copy, while the more genetically

distant group (even to cattle sequences) represents the second. On no occasion were more than two alleles found upon sequencing any one sample. If two gene copies were expressed, I could expect to recover three or four different alleles from some individuals. However, the amplification of alleles from several different gene copies has been described in other sheep species (Jugo & Vicario, 2000).

Evidence of balancing selection present in DRB3 sequences may not reflect current selection but be the result of historical processes. It could be that presently the neutral processes of drift and mutation exceed the strength of selection at thinhorn MHC genes. Evidence of population structure within the DRB3 locus supports the influence of neutral processes at the gene. It is however possible that negative frequency dependent selection (a form of balancing selection) could also result in population structure. In this case the fitness associated with a genotype varies with the changing genetic composition of a population; decreasing as the allele becomes more common. This type of selection is associated with genetic variation in heterogeneous environments. It can therefore be hypothesised that different allelic compositions are favoured between isolated populations, thus resulting in a degree of population structure between them. Levels of linkage disequilibrium are expected to be raised immediately after a selective event (Przeworski, 2002). The low coefficients of linkage disequilibrium between microsatellites surrounding the DRB3 gene provide evidence against recent balancing selection, a result supported by the relatively weak test statistic of Tajima's *D*. As both these tests of selection are relatively weak these conclusions should be interpreted with caution. Moreover, balancing selection could be acting on other MHC genes not considered here, such as those located in the class I region.

The limited evidence I was able to obtain correlating sequence and microsatellite alleles appears to show there to be more information within sequences than within linked neutral markers. In at least two cases two sequence alleles were correlated to the same microsatellite genotype. This is in contrast to the pattern reported in Soay sheep, where there were more microsatellite alleles than correlated sequences

(Paterson, 1998). A more detailed analysis of correlations between marker types was impossible given the presence of null alleles in the microsatellite locus OLADRB.

5.4.2. Variation within interferon gamma sequences

Thinhorn IFNG has lower nucleotide diversity than both MHC and NRAMP. Non-coding regions are characterised in few organisms, so species comparisons of diversity with both thinhorn IFNG exonic and intronic regions is impossible. However, cross species alignments of exons 2 through 4 are possible. These reveal the presence of similarly low levels of variation in other ungulates. For example, both *Capra hircus* and *Equus caballus* have nucleotide polymorphism equal to that of thinhorn. Across the entire IFNG coding region of *O. aries* only five polymorphic bases are present.

Both thinhorn IFNG coding sequences encode the same translated amino acid sequence. This is expected as the amino acid sequences of all three domestic sheep alleles are also fixed. Indeed most characterised species show a greater number of nucleotide than amino acid sequences corresponding to a higher rate of synonymous than nonsynonymous substitutions in IFNG. It appears that any selection at IFNG acts to conserve amino acid residues by eliminating novel nonsynonymous mutations. This is in contrast to expectations of balancing selection. Of the three characterised domestic sheep nucleotide sequences, one has been correlated with disease resistance. The common and primitive thinhorn allele is not a disease resistant form.

An examination of nucleotide variation points to there being no balancing selection acting on the thinhorn IFNG gene. This conclusion is supported by an IFNG gene tree. Unlike the MHC based tree, tree topology follows species phylogeny, with strong bootstrap support at most nodes (Figure 5.7). Genes under the influence of neutral drift are expected to show evidence of population structure. The restriction of allele *O. dalli*1* to northern Alaska provides some evidence of this, although the lack

of variation results in a lower magnitude of structuring than was present in the MHC region. Caution should be taken when interpreting these results as showing a lack of selection as no robust tests of selection can be carried out on these data.

Selective sweeps result in a reduction of genetic variation at linked loci due to the fixation of alleles linked to advantageous mutations. The resultant localised increase in linkage disequilibrium (Kim & Nielsen, 2004) has been used as a method of identifying gene regions that deviate from neutrality. There is no evidence of allelic correlation between the two IFNG markers, the gene sequence and the linked microsatellite KP6. A lack of linkage disequilibrium between closely linked markers can not however be used as evidence against selection, as linkage disequilibrium can be eroded in very few generations by recombination.

5.4.3. Variation within NRAMP sequences

This study represents the first work to examine evidence of selection at the NRAMP gene in any species. Homologous intronic bovid sequences contain 13 polymorphic sites (Ables *et al.*, 2002) compared to the four present within thinhorn. The greatest number of polymorphisms within both species is found in intron 5. This is also the case in other mammals such as the Rhesus macaque where intronic polymorphisms are correlated to TB resistance (Deinard *et al.*, 2002). A relationship between intronic variation and disease raises a potentially interesting discussion point in thinhorn. I report increased intronic diversity in the south of the species range. It is in these regions that sheep habitat is most influenced by anthropogenic factors, indeed I previously hypothesised that a reduction in neutral microsatellite diversity in the southern limits of the species range is a result of human habitation (Worley *et al.*, 2004). However, any correlation between intronic variation in sheep NRAMP and pathogen resistance remains untested. Interestingly the increased sequence variability in the southern species range is in contrast to variability in linked microsatellites, where there are significantly fewer heterozygotes at OVINRAO1 in the southern range than there are in northern populations.

A comparison between levels of polymorphism within both exonic and intronic regions of the NRAMP gene of thinhorn and other ungulates cannot be made, due to the unavailability of intronic sequences from other species. However comparisons of exonic sequences reveal there to be few nucleotide differences separating the sequences of available species, for example exonic sequences of thinhorn and domestic sheep vary by one synonymous and one nonsynonymous fixed difference. Overall sequence diversity is of a similar magnitude to that observed in IFNG.

The association between the microsatellite OVINRAO1 and NRAMP intronic sequence is imperfect, indicating that recombination has occurred between the 1769 base pairs separating them. There is no information in the literature regarding the distance between IFNG coding regions and the linked microsatellite KP6. An unsuccessful attempt was made to infer this distance by BLASTing KP6 microsatellite primer sequences against a cattle genome scaffold. Given the linkage evident between NRAMP sequence and microsatellite I hypothesise that the distance between IFNG exons and KP6 comprises a greater number of nucleotides.

5.4.4. Prion Protein

An analysis of the group Caprinae reveals evidence of balancing selection in the PrP gene dating to before domestication (J. Slate, in press). However, of the many PrP sequence variants observed in domestic sheep, only two were present in thinhorn. As there are reliable associations between sequence polymorphism and resistance to scrapie, the evidence here suggests an absence of selective pressure in response to prion diseases in wild sheep. Alternative (although less likely) conclusions are that the resistance alleles in domestic sheep do not correlate to those in wild sheep or that the thinhorn variant is resistant to a prion agent other than scrapie.

5.4.5. Conclusions

I have successfully characterised coding regions of four immune related genes in thinhorn sheep. Patterns of variation within an MHC gene support the hypothesis of balancing selection in this region. Despite known associations with disease resistance, the remaining three genes contain no evidence of balancing selection. This is however not surprising, as the approaches used have low statistical power to detect selection from such small numbers of unique sequence alleles. Although linked microsatellites also yielded no evidence of selection at IFNG or NRAMP, this is not a reflection of the accuracy of microsatellite markers in identifying the absence of selection. There remains a discrepancy between the degree of balancing selection identified by sequence data and microsatellite loci at the MHC. A possible explanation for this discrepancy is the different fingerprints left by contemporary and historical selection. Balancing selection at the MHC appears ancient in origin. As the mutation rate of sequences is much lower than those present at microsatellites, the signatures of ancient selection are not expected to remain within microsatellites for more than a few short generations. It therefore appears that there is no statistical evidence of contemporary selective pressures in response to disease in genes related to the thinhorn immune response. Alternatively, selection could be acting on all characterised genes, but be too weak to be detected by the few microsatellite loci used here. The magnitude of genetic variation present within thinhorn immune genes (with the exception of PrP) is similar to that present within the coding sequences of other ungulate species, irrespective of disease and demographic histories. This adds support to a conclusion of weak or no current day selection within immune genes.

CHAPTER 6. GENERAL DISCUSSION

This thesis details an analysis of the genetic variation present within several different types of genetic marker in thinhorn sheep. It represents a comprehensive study of the predominant processes regulating genetic variation within an unmanaged species. Unifying conclusions, significant contributions and future directions for this field are discussed below.

6.1. Processes regulating genetic variation in thinhorn sheep

Identifying the processes that regulate genetic variation in natural populations is an important research area in evolutionary biology. The results detailed in previous chapters quantify the magnitude of variation present within different genetic markers. The combined results can be used to describe the relative impact of neutral and selective processes acting on thinhorn sheep.

As hypothesised, significant population structure is present throughout the thinhorn species range. Detailed empirical evidence of both substructure due to ewe groups within sample areas and genetic differentiation between mountain ranges is presented in Chapter 2. Results reflect thinhorn natal philopatry and the restriction of gene flow through unsuitable habitat. Population structure is also evident from mtDNA (Chapter 3). Patterns of population structure present between mtDNA and microsatellite markers are extremely similar. Both show relatively lower genetic differentiation between neighbouring mountain ranges than that present between distant regions. At similar geographic scales, the degree of microsatellite population structure is more pronounced than from mtDNA due to the higher mutation rates and bi-parental inheritance in the former loci.

More surprisingly, large scale population structure is also evident within coding sequences of genes involved in the immune response (Chapter 5). MHC sequences present in more than one sample region are likely to come from adjacent mountain

ranges, mirroring the results of mtDNA and microsatellites. Widespread evidence of population structure across coding and non-coding loci can be attributed to the predominance of neutral processes over natural selection in regulating levels of genetic variation in thinhorn sheep. The prevalence of neutral processes is also reflected in immune linked microsatellites (Chapter 4). Evidence of isolation-by-distance relationships mirror those produced by neutral drift in populations with limited dispersal.

Mutations that result in a change in amino acid residue are likely to be eliminated by selection due to the likely decrease in fitness they confer. Selection against deleterious mutations is therefore one process that results in a reduction of the genetic variation in coding regions relative to non-coding DNA. However, the role played by differing mutation rates is perhaps more significant. The power to detect population structure is associated with the degree of allelic diversity within loci. More variable loci are more useful for elucidating population structure. This distinction between loci accounts for the relatively low level population structure revealed by MHC alleles relative to those from non-coding loci. MHC loci are not ideal markers for interpretation of population structure due to the known signature of balancing selection they contain. The inability to identify the effects of selective processes in a thinhorn MHC locus suggests that selection is very weak in this species. IFNG and NRAMP sequences also show some limited evidence of regional genetic differentiation. However, this mostly reflects a difference in the magnitude of sequence diversity between mountain ranges.

There are conflicts between the magnitudes of genetic diversity observed at different loci within the same sample area. Genetic diversity within non-linked microsatellite loci is greater within sample areas of the Pelly and Mackenzie mountain ranges than within Alaskan and BC areas. This may be attributed to the relative isolation of the Alaskan ranges, anthropogenic effects in BC in addition to both regions being on the fringes of the species range. Genetic variation in immune linked microsatellites varies between loci, but patterns of diversity approximate those of non-linked

microsatellites. In contrast, mtDNA haplotype diversity in northern and central Alaska is higher than that within central Yukon sample regions. BC diversity remains relatively low. Disparity in levels of diversity between loci can reveal evidence of differential processes regulating genetic variation. However, both microsatellites and mtDNA are evolutionarily neutral markers, and as such the levels of variation they contain are dependent on neutral genetic drift and gene flow. Differences in genetic diversity between neutral loci therefore reflect their differential mutation rates and mode of inheritance. The disparity between levels of genetic variation in mtDNA and microsatellites within Alaskan populations is likely to reflect historical processes. MtDNA is able to reveal evidence of historical phylogeography not possible from microsatellite loci. Contemporary Alaskan thinhorn populations shared the relatively large refugia of Beringia during periods of glaciation. It can be hypothesised that the size of the Alaskan population at this time was therefore larger than those in other smaller refugia. As the effects of drift are not as pronounced in large populations a higher level of variation may have been maintained within Alaska during periods of glaciation. This is reflected in the relatively high mtDNA diversity in extant Alaskan populations.

As gene coding regions may also be under the influence of natural selection, the genetic variation within different genes can vary as a result of differential selective processes. Although the associated sample sizes are small, patterns of genetic diversity within thinhorn coding sequences appear to differ substantially from those of neutral markers. Diversity within the MHC gene DRB3 appears lower in the Mackenzie and Pelly ranges than that present in Alaska and BC. This pattern is mirrored by patterns the genetic diversity within IFNG alleles. The NRAMP coding sequence is fixed across much of the species range, only Pelly and BC sample regions show variation. Although differences in the distributions of genetic diversity between coding regions can be evidence of natural selection, there is no contemporary evidence of strong selection acting on genes associated with the immune system of thinhorn sheep. Evidence of a disparity between levels of genetic variation within populations from functional and neutral loci is also present in moose

(Wilson *et al.*, 2003). In this case the functional locus was also an MHC gene, indicating that differences between levels of variation present between locus types in both sheep and moose are largely reflective of historical versus contemporary processes.

6.2. Delineation of mountain sheep species and subspecies

Taxonomic divisions within North American mountain sheep are unresolved from morphometric and previous genetic analyses (Ramey, 1993; Cowan, 1940). Microsatellites appear to show a slight genetic distinction between samples areas where the majority of herds comprise different subspecies in excess of that expected by geography alone. This verification of subspecies is not supported by mtDNA, where analysis is on an individual basis rather than the grouping approach implemented for microsatellite analysis. None of the immune related coding sequence alleles characterised is confined to one subspecies. Overall evidence presented in this thesis suggests the lack of clear genetic distinction between thinhorn subspecies.

MtDNA also provides evidence against the genetic division of thinhorn and bighorn sheep into separate species. There is some evidence of allele sharing between thinhorn and bighorn immune genes. MHC alleles can not be used for taxonomic inference due to the effects of balancing selection. Alleles of IFNG sequences are shared between thinhorn, bighorn and domestic sheep species, reflecting the low level diversity within this region. However, some genetic distinction between thinhorn and bighorn is present at the NRAMP locus. The most prevalent bighorn allele appears to be lacking in thinhorn. In addition to results presented in this thesis, I have observed significant differences between thinhorn and bighorn microsatellite alleles. Although many microsatellites alleles are found between species, they are present in very different allele distributions. Combined evidence relating to the species recognition of bighorn and thinhorn sheep appears to show that there is no current gene flow between species, revealed by differing microsatellite profiles and

the lack of overlap between their ranges. Sequence similarities reflect the fact that North American mountain sheep are evolutionarily young species, only showing diversification in the relatively recent past. In conclusion, I suggest thinhorn and bighorn sheep remain considered as separate species.

6.3. Major contributions

This study makes significant contributions within each study area. Chapter 2 represents one of the first studies to describe the extent of population structure across the entire native species range of a mountain dwelling large mammal, detailing empirical evidence relating to the effects of habitat limitations. Chapter 3 provides evidence for the existence of cryptic glacial refugia in western Canada throughout the Pleistocene glaciations, previously only suggested from geological and palynological evidence (Szeicz & MacDonald, 2001; Mandryk *et al.*, 2001). Both chapters are published in peer reviewed journals (Worley *et al.*, 2004; Loehr *et al.*, 2005).

A major unifying contribution of this work is as an empirical study aiming to identify natural selection in a highly structured population, and of the problems associated with this issue. Prior to this work there was theoretical evidence outlining the problems expected from locating the fingerprint of selection in such populations (Muirhead, 2001), but no empirical examples are present in the literature. This study has raised a number of issues relating to the subject, including the identification of a bias to report balancing selection, even in populations with little evidence of structure. It acts as an example for future research into the identification of selection in natural populations; showing the importance of the ability to discount population structure as a cause of deviation from neutral expectations by comparing variation in putative selected loci with neutral genetic variation within study populations.

The patterns of genetic variation present at linked microsatellite markers and nucleotide sequence data in thinhorn sheep reveal that caution must be taken before

making conclusions about selection from only one line of evidence. The data presented here reveal that this is a rather unsupported method, even in the MHC where linkage is known to be extremely tight between loci.

6.4. Implication of results to species management

The magnitude of genetic diversity present across the thinhorn species range is of importance to game managers. Licences for hunting thinhorn sheep generate large revenue in northern Canada, and hence any significant fall in their numbers may impact the economy of these regions. This thesis shows there are high levels of diversity over all the species range. More relevant to sheep management is the identification of several genetically distinct thinhorn groups between mountain ranges. Northern Alaskan populations are especially distinct through their geographic isolation from other regions, as revealed by both microsatellites and DNA sequences.

Genetic diversity across the thinhorn range reflects gradual changes caused by an isolation-by-distance relationship rather than discrete distinctions between sample areas. Distinctions between current management zones, delineated by lines drawn on the range map, are therefore not good representations of the genetic distinctions between differentiated sheep populations. It would be useful to consider larger areas with boundaries based on the limits of suitable habitat when examining population numbers in the future.

Similar levels of MHC diversity are present between thinhorn and bighorn sheep (Gutierrez-Espeleta *et al.*, 2001). In addition, there is no known association between diversity at the MHC and resistance to novel disease in the latter species. This suggests that if disease appears in the thinhorn population, sheep will have no resistance and the consequences will be similar to those observed in bighorn populations. The lack of strong evidence of selection at other immune related genes also suggests susceptibility to any correlated disease. If the association between PrP genotype and scrapie resistance present in domestic sheep is applicable to other prion

diseases, we may also predict no innate resistance of thinhorn to diseases such as CWD.

6.5. Future directions and questions

Perhaps the most intriguing questions concerning North American wild sheep populations are those relating to disease in bighorn. There is no strong evidence of current selective pressures acting on the immune response of thinhorn sheep, a species with no known history of widespread disease. It may be interesting to apply the methods utilised in this study to populations of bighorn both before and after a disease outbreak to examine evidence for changing allele frequencies at loci involved in the immune response. Subtle effects of selection as revealed by differences in allele frequencies may be best observed using microsatellite loci. Although this method was unsuccessful in identifying selection in thinhorn, I suggest that the use of many more markers within a population with little population structure will overcome some of the difficulties encountered. Analysis of thinhorn MHC sequences was only able to identify evidence of historical balancing selection, obscuring any current non-neutral processes. However, the comparison of sequence diversity within one population before and after a disease related bottleneck may be more revealing of current selection.

Several issues have arisen during the analysis of data included in this thesis. Firstly, the significance of linkage disequilibrium differs when homozygotes individuals are dropped from analysis. There are no theoretical discussions of this aspect of calculating linkage disequilibrium; thinhorn data show these effects empirically. Secondly, any attempt to identify natural selection via the use of linked microsatellites has low power. As a result many loci are required to locate regions under the influence of selection. Any future attempts aiming to locate selection in natural populations should consider this fact, and possibly aim to analyse variation within one region more concisely, including many more loci than were considered here.

Although thinhorn microsatellites have the power to correctly assign individuals to the correct mountain ranges or origin, they are unable to identify the origin of sheep on a smaller scale. A much finer scale of assignment is required by game managers when investigating forensics cases concerning illegal hunting. Although the genotypes completed in this thesis have been deposited with the Alberta Department of Wildlife and Fisheries forensics laboratory, individuals must be genotyped over additional microsatellite loci to increase the power and use for such agencies.

REFERENCES

Abbot, R. J., Smith, L. C., Milne, R. I., *et al.* (2000) Molecular analysis of plant migration and refugia in the Arctic. *Science*, **289,** 1343-1346.

Ables, G. P., Nishibori, M., Kanemaki, M., *et al.* (2002) Sequence analysis of the NRAMP1 genes from different bovine and buffalo breeds. *Journal of Veterinary Medical Science*, **64,** 1081-1083.

Aguilar, A., Roemer, G., Debenham, S., *et al.* (2004) High MHC diversity maintained by balancing selection in an otherwise genetically monomorphic mammal. *Proceedings of the National Academy of Science, USA*, **101,** 3490-3494.

Arranz, J. J., Bayon, Y. & San Primitivo, F. (1998) Genetic relationships among Spanish sheep using microsatellites. *Animal Genetics*, **29,** 435-440.

Austyn, J. M. & Wood, K. J. (1993) *Principles of cellular and molecular immunology*. Oxford University Press.

Avise, J. C. & Walker, D. (1998) Pleistocene phylogeographic effects on avian populations and the speciation process. *Proceedings of the Royal Society of London B*, **265,** 457-463.

Avraham, A., Band, M., Yoffe, O., *et al.* (1993) Bovine dinucleotide repeat polymorphism at the ARO28 locus. *Animal Genetics*, **24,** 147.

Barichello, N. J., Carey, J. & Hoefs, M. (1989) Mountain sheep status and harvest in the Yukon: A summary of distribution, abundance and the registered harvest, by game management zone. Whitehorse, Yukon: Yukon department of renewable resources.

Barthel, R., Feng, J., Piedrahita, J. A., *et al.* (2001) Stable transfection of the bovine *NRAMP1* gene into murine RAW264.7 cells: Effect on *Brucella abortus* survival. *Infection and Immunity*, **69,** 3110-3119.

Barton, N. H. (1995) Linkage and the limits to natural selection. *Genetics*, **140,** 821-841.

Barton, N. H. (2000) Genetic hitchhiking. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, **355,** 1553-1562.

Baylis, M. & Goldmann, W. (2004) The genetics of scrapie in sheep and goats. *Current Molecular Medicine*, **4,** 385-396.

Beaumont, M. A. & Balding, D. J. (2004) Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology*, **13,** 969-980.

Beaumont, M. A. & Nichols, R. A. (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London Series B-Biological Sciences*, **263,** 1619-1626.

Belkhir. (1996) GENETIX; available at http://www.university-montp2.fr/~genetix/genetix.htm.

Benjamini, E., Sunshine, G. & Leskowitz, S. (1996) *Immunology: A short course.* New York: Wiley.

Bernatchez, L. & Landry, C. (2003) MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? *Journal of Evolutionary Biology*, **16,** 363-377.

Bishop, M. D., Kappes, S. M., Keele, J. W., *et al.* (1994) A genetic-linkage map for cattle. *Genetics*, **136,** 619-639.

Blackwell, J. M. & Searle, S. (1999) Genetic regulation of macrophage activation: understanding the function of *Nramp1* (*=Ity/Lsh/Bcg*). *Immunology Letters*, **65,** 73-80.

Blattman, A. N. & Beh, K. J. (1992) Dinucleotide repeat polymorphism within the ovine major histocompatibility complex. *Animal Genetics*, **23,** 392-392.

Boehm, U., Klamp, T., Groot, M., *et al.* (1997) Cellular responses to interferon-gamma. *Annual Review of Immunology*, **15,** 749-795.

Bowcock, A. M., Kidd, J. R., Mountain, J. L., *et al.* (1991) Drift, admixture, and selection in human evolution - A study with DNA polymorphisms. *Proceedings of the National Academy of Sciences of the United States of America*, **88,** 839-843.

Boyce, W. M., Hedrick, P. W., Muggli-Cockett, N. E., *et al.* (1997) Genetic variation of major histocompatibilty complex and microsatellite loci: A comparison in bighorn sheep. *Genetics*, **145,** 421-433.

Boyce, W. M., Ramey, R. R., II, Rodwell, T. C., *et al.* (1999) Population subdivision among desert bighorn sheep (*Ovis canadensis*) ewes revealed by mitochondrial DNA analysis. *Molecular Ecology*, **8,** 99-106.

Broders, H. G., Mahoney, S. P., Montevecchi, W. A., *et al.* (1999) Population genetic structure and the effect of founder events on the genetic variability of moose, *Alces alces*, in Canada. *Molecular Ecology*, **8,** 1309-1315.

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

Brunhoff, C., Galbreath, K. E., Federov, V. B., *et al.* (2003) Holarctic phylogeography of the root vole (*Microtus oeconomus*): implications for late Quaternary biogeography of high latitudes. *Molecular Ecology*, **12,** 957-968.

Brunner, P. C., Douglas, M. R., Osinov, A., *et al.* (2001) Holarctic phylogeography of arctic charr (*Salvelinus alpinus* L.) inferred from mitochondrial DNA sequences. *Evolution*, **55,** 573-586.

Buchanan, F. C. & Crawford, A. (1992) Ovine dinucleotide repeat polymorphism at the MAF209 locus. *Animal Genetics*, **23,** 183.

Buchanan, F. C. & Crawford, A. (1993) Ovine microsatellites at the OarFCB11, OarFCB128, OarFCB193, OarFCB266 and OarFCB304 loci. *Animal Genetics*, **24,** 145.

Bueler, H., Fischer, M., Lang, Y., *et al.* (1992) Normal development and behaviour of mice lacking the neuronal cell-surface PrP protein. *Nature*, **356,** 577-582.

Burbrink, F. T., Lawson, R. & Slowinski, J. B. (2000) Mitochondrial DNA phylogeography of the polytypic North American rat snake (*Elaphe obsoleta*): A critique of the subspecies concept. *Evolution*, **54,** 2107-2118.

Bussmann, V., Lantier, I., Pitel, F., *et al.* (1998) cDNA cloning, structural organization, and expression of the sheep NRAMP1 gene. *Mammalian Genome*, **9,** 1027-1031.

Bustamante, C. D., Nielsen, R. & Hartl, D. L. (2003) Maximum likelihood and Bayesian methods for estimating the distribution of selective effects among

classes of mutations using DNA polymorphism data. *Theoretical Population Biology*, **63,** 91-103.

Canonne-Hergaux, F., Gruenheid, S., Govoni, G., *et al.* (1999) The Nramp1 protein and its role in resistance to infection and macrophage function. *Proceedings of the Association of American Physicians*, **111,** 283-289.

Caplazi, P. A., O'Rourke, K. I. & Baszler, T. V. (2004) Resistance to scrapie in PrP ARR/ARQ heterozygous sheep is not caused by preferential allelic use. *Journal of Clinical Pathology*, **57,** 647-650.

Carmichael, L. E., Nagy, J. A., Larter, N. C., *et al.* (2001) Prey specialization may influence patterns of gene flow in wolves of the Canadian Northwest. *Molecular Ecology*, **10,** 2787-2798.

Casgrain, P. & Legendre, P. (2001) *The R package for multivariate and spatial analysis*, Version 4.0 - *Users manual.* Department des Sciences biologiques, Universite de Montreal. http://www.fas.umontreal.ca/BIOL/legendre.

Catto, N., Liverman, D. G. E., Bobrowsky, P. T., *et al.* (1996) Laurentide, Cordilleran, and Montane glaciation in the western Peace River Grande Prairie Region, Alberta and British Columbia, Canada. *Quaternary International*, **32,** 21-32.

Catto, N. R. (1996) Richardson Mountains, Yukon-Northwest Territories: The northern portal of the postulated 'ice-free corridor'. *Quaternary International*, **32,** 3-19.

Clark, T. E., Levin, D. B., Kavanaugh, D. H., *et al.* (2001) Rapid evolution in the *Nebria gregaria* group (Coleoptera: Carabidae) and the paleogeography of the Queen Charlotte Islands. *Evolution*, **55,** 1408-1418.

Clarke, B. C. & Kirby, D. R. S. (1966) Maintenance of histocompatibility polymorphism. *Nature*, **211,** 999-1000.

Coltman, D. W., Festa-Bianchet, M., Jorgenson, J. T., *et al.* (2002) Age-dependent sexual selection in bighorn rams. *Proceedings of the Royal Society of London Series B-Biological Sciences*, **269,** 165-172.

Coltman, D. W., O'Donoghue, P., Jorgenson, J. T., *et al.* (2003a) Undesirable evolutionary consequences of trophy hunting. *Nature*, **426,** 655-658.

Coltman, D. W., Pilkington, J. G. & Pemberton, J. M. (2003b) Fine-scale genetic structure in a free-living ungulate population. *Molecular Ecology*, **12,** 733-742.

Coltman, D. W., Pilkington, J. G., Smith, J. A., *et al.* (1999) Parasite-mediated selection against inbred Soay sheep in a free-living, island population. *Evolution*, **53,** 1259-1267.

Coltman, D. W., Wilson, K., Pilkington, J. G., *et al.* (2001) A microsatellite polymorphism in the gamma interferon gene is associated with resistance to gastrointestinal nematodes in a naturally-parasitized population of Soay sheep. *Parasitology*, **122,** 571-582.

Cornuet, J. M., Piry, S., Luikart, G., *et al.* (1999) New methods employing multilocus genotypes to select or exclude populations as origins of individuals. *Genetics*, **153,** 1989-2000.

Cote, S. D., Dallas, F., Marshall, F., *et al.* (2002) Microsatellite DNA evidence for genetic drift and philopatry in Svalbard reindeer. *Molecular Ecology*, **11,** 1923-1930.

Cowan, I. M. (1940) Distribution and variation in the native sheep of North America. *The American Midland Naturalist*, **24,** 505-580.

Crandall, K. A., Harris, D. J. & Fetzner, J. W. (2000) The monophyletic origin of freshwater crayfish estimated from nuclear and mitochondrial DNA sequences. *Proceedings of the Royal Society of London Series B-Biological Sciences*, **267,** 1679-1686.

Davis, E. G., Eichenberger, M. R., Grant, B. S., *et al.* (2000) Microsatellite marker of interferon-gamma receptor 1 gene correlates with infection following major trauma. *Surgery*, **128,** 301-305.

de Gortari, M. J., Freking, B. A., Kappes, S. M., *et al.* (1997) Extensive genomic conservation of cattle microsatellite heterozygosity in sheep. *Animal Genetics*, **28,** 274-290.

Deinard, A. S., Lerche, N. W. & Smith, D. G. (2002) Polymorphism in the Rhesus macaque (*Macaca mulatta*) NRAMP1 gene: lack of an allelic association to tuberculosis susceptibility. *Journal of Medical Primatology*, **31,** 8-16.

Ditchkoff, S. S., Lochmiller, R. L., Masters, R. E., *et al.* (2001) Major histocompatibility complex associated variation in secondary sexual traits of white-tailed deer (*Odocoileus virginianus*): Evidence for good genes advertisement. *Evolution*, **55,** 616-625.

Dobeš, C. H., Mitchell-Olds, T. & Koch, M. A. (2004) Extensive chloroplast haplotype variation indicates Pleistocene hybridisation and radiation of North American *Arabis drummondii*, *A.* x *divaricarpa*, and *A. holboellii* (*Brassicaceae*). *Molecular Ecology*, **13,** 349-370.

Duk-Rodkin, A. & Hughes, O. L. (1991) Age relationships of Laurentide and montane glaciations, Mackenzie Mountains, Northwest Territories. *Geographie Physique et Quaternaire*, **41,** 237-263.

Dutia, B. M., McConnell, I., Ballingall, K. T., *et al.* (1994) Evidence for the expression of two distinct MHC class II DRB like molecules in the sheep. *Animal Genetics*, **25,** 235-241.

Dyke, A. S., Andrews, J. T., Clark, P. U., *et al.* (2002) The Laurentide and Innuitian ice sheets during the last glacial maximum. *Quaternary Science Reviews*, **21,** 9-31.

Dyke, A. S. & Prest, V. K. (1987) Late Wisconsin and Holocene history of the Laurentide ice sheet. *Geographie Physique et Quaternaire*, **41,** 237-263.

Ede, A. J., Pierson, C. A. & Crawford, A. (1995) Ovine microsatellites at the OarCP9, OarCP16, OarCP20, OarCP21, OarCP23 and OarCP26 loci. *Animal Genetics*, **26,** 129-130.

Ehrich, D., Fedorov, V. B., Stenseth, N. C., *et al.* (2000) Phylogeography and mitochondrial DNA (mtDNA) diversity in North American collared lemmings (*Dicrostonyx groenlandicus*). *Molecular Ecology*, **9,** 329-337.

Ellegren, H., Mikko, S., Wallin, K., *et al.* (1996) Limited polymorphism at major histocompatibility complex (MHC) loci in the Swedish moose (*A. alces*). *Molecular Ecology*, **5,** 3-9.

Ellegren, H., Moore, S., Robinson, N., *et al.* (1997) Microsatellite evolution - A reciprocal study of repeat lengths at homologous loci in cattle and sheep. *Molecular Biology and Evolution*, **14,** 854-860.

Ellegren, H., Smith, N. G. C. & Webster, M. T. (2003) Mutation rate variation in the mammalian genome. *Current Opinion in Genetics & Development*, **13**, 562-568.

Ewens, W. J. (1972) Sampling theory of selectively neutral alleles. *Theoretical Population Biology*, **3**, 87-&.

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

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

Fay, J. C. & Wu, C. I. (2000) Hitchhiking under positive Darwinian selection. *Genetics*, **155**, 1405-1413.

Federov, V. B. & Stenseth, N. C. (2002) Multiple glacial refugia in the North American arctic: inference from phylogeography of the collared lemming (*Dicrostonyx groenlandicus*). *Proceedings of the Royal Society of London B*, **269**, 2071-2077.

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

Felsenstein, J. (1993) Phylogeny Inference Package (PHYLIP), Version 3.5c. Seattle: University of Washington.

Festa-Bianchet, M. (1991) The social system of bighorn sheep - grouping patterns, kinships and female dominance rank. *Animal Behaviour*, **42**, 71-82.

Fitzsimmons, N. N., Buskirk, S. W. & Smith, M. H. (1997) Genetic changes in reintroduced Rocky Mountain bighorn sheep populations. *Journal of Wildlife Management*, **61**, 863-872.

Flagsted, Ø. & Røed, K. H. (2003) Refugial origins of reindeer (*Rangifer tarandus* L.) inferred from mitochondrial DNA sequences. *Evolution*, **57**, 658-670.

Flajnik, M. F. & Kasahara, M. (2001) Comparative genomics of the MHC: glimpses into the evolution of the adaptive immune system. *Immunity*, **15**, 351-362.

Fleming, M. D. & Andrews, N. C. (1998) Mammalian iron transport: An unexpected link between metal homeostasis and host defence. *Journal of Laboratory and Clinical Medicine*, **132,** 464-468.

Forbes, S. H. & Hogg, J. T. (1999) Assessing population structure at high levels of differentiation: microsatellite comparisons of bighorn sheep and large carnivores. *Animal conservation*, **2,** 223-233.

Forbes, S. H., Hogg, J. T., Buchanan, F. C., *et al.* (1995) Microsatellite evolution in congeric mammals: domestic and bighorn sheep. *Molecular Biology and Evolution*, **12,** 1106-1113.

Foreyt, W. J. (1989) Fatal *Pasteurella haemolytica* pneumonia in bighorn sheep after direct contact with clinically normal domestic sheep. *American Journal of Veterinary Research*, **50,** 341-344.

Foreyt, W. J., Silflow, R. M. & Lagerquist, J. E. (1996) Susceptibility of Dall's sheep (*Ovis dalli dalli*) to pneumonia caused by *Pasteurella haemolytica. Journal of Wildlife Diseases*, **32,** 586-593.

Fu, Y. X. (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, **147,** 915-925.

Gabrielsen, T. M., Bachmann, K., Jakobsen, K. S., *et al.* (1997) Glacial survival does not matter: RAPD phylogeography of Nordic *Saxifraga oppositifolia. Molecular Ecology*, **6,** 831-842.

Galbreath, K. E. & Cook, J. A. (2004) Genetic consequences of Pleistocene glaciations for the tundra vole (*Microtus oeconomus*) in Beringia. *Molecular Ecology*, **13,** 135-148.

Garrigan, D. & Hedrick, P. N. (2001) Class I MHC polymorphism and evolution in endangered California Chinook and other Pacific salmon. *Immunogenetics*, **53,** 483-489.

Gaudieri, S., Dawkins, R. L., Habara, K., *et al.* (2000) SNP profile within the human major histocompatibility complex reveals an extreme and interrupted level of nucleotide diversity. *Genome Research*, **10,** 1579-1586.

Geist, V. (1971) *Mountain sheep: A Study in Behavior and Evolution.* Chicago: University of Chicago Press.

Geist, V. (1999) Adaptive strategies in Mountain sheep. In: *Mountain sheep of North America* (Ed. by Valdez, R. & Krausman, P.), pp. 192-208. Tucson: The Univerisity of Arizone Press.

Georges, M. & Massey, J. (1992) Polymorphic DNA markers in Bovidae. In: *WO Publ. No. 92/13120*. Geneva: World Intellectual Property Organisation.

Gilad, Y., Rosenberg, S., Przeworski, M., *et al.* (2002) Evidence for positive selection and population structure at the human MAO-A gene. *Proceedings of the National Academy of Sciences of the United States of America*, **99,** 862-867.

Glinka, S., Ometto, L., Mousset, S., *et al.* (2003) Demography and natural selection have shaped genetic variation in *Drosophila melanogaster*: A multi-locus approach. *Genetics*, **165,** 1269-1278.

Golden, J. L. & Bain, J. F. (2000) Phylogeographic patterns and high levels of chloroplast DNA diversity in four *Packera* (Asteraceae) species in southwestern Alberta. *Evolution*, **54,** 1566-1579.

Goldmann, W., Chong, A., Foster, J., *et al.* (1998) The shortest known prion protein gene allele occurs in goats, has only three octapeptide repeats and is non-pathogenic. *Journal of General Virology*, **79,** 3173-3176.

Goldstein, D. B., Linares, A. R., Cavallisforza, L. L., *et al.* (1995) An evaluation of genetic distances for use with microsatellite loci. *Genetics*, **139,** 463-471.

Goldstein, E. J., Millspaugh, J. J., Washburn, B. E., *et al.* (2005) Relationships among fecal lungworm loads, fecal glucocorticoid metabolites, and lamb recruitment in free-ranging Rocky Mountain bighorn sheep. *Journal of Wildlife Diseases*, **41,** 416-425.

Greenwood, P. J. (1980) Mating systems, philopatry and dispersal in birds and mammals. *Animal Behaviour*, **28,** 1140-1162.

Groth, D. M. & Wetherall, J. D. (1994) Dinucleotide repeat polymorphism within the ovine major histocompatibility complex class I region. *Animal Genetics*, **25,** 61-61.

Gruenheid, S., Pinner, E., Desjardins, M., *et al.* (1997) Natural resistance to infection with intracellular pathogens: The Nramp1 protein is recruited to the

membrane of the phagosome. *Journal of Experimental Medicine*, **185,** 717-730.

Gulland, F. M. D., Albon, S. D., Pemberton, J. M., *et al.* (1993) Parasite associated polymorphism in a cyclic ungulate population. *Proceedings of the Royal Society of London Series B-Biological Sciences*, **254,** 7-13.

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

Gutierrez-Espeleta, G. A. (1999) Neutral and adaptive genetic variation in Desert bighorn sheep (*Ovis canadensis*): Implications for conservation. In: *PhD Thesis*: PhD thesis, Arizona State University.

Gutierrez-Espeleta, G. A., Hedrick, P. W., Kalinowski, S. T., *et al.* (2001) Is the decline of desert bighorn sheep from infectious disease the result of low MHC variation? *Heredity*, **86,** 439-450.

Gutierrez-Espeleta, G. A., Kalinowski, S. T., Boyce, W. M., *et al.* (2000) Genetic variation and population structure in desert bighorn sheep: implications for conservation. *Conservation Genetics*, **1,** 3-15.

Hambuch, T. M. & Lacey, E. A. (2002) Enhanced selection for MHC diversity in social tuco-tucos. *Evolution*, **56,** 841-845.

Harding, C. V. & Unanue, E. R. (2000) Cellular mechanisms of antigen processing and the function of class I and II major histocompatibility complex molecules. *Cell Regulation*, **1,** 499-509.

Hardy, O. J. & Vekemans, X. (2002) SPAGEDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes*, **2,** 618-620.

Hedrick, P. W. (1999) Balancing selection and MHC. *Genetica*, **104,** 207-214.

Hedrick, P. W. (2002) Pathogen resistance and genetic variation at MHC loci. *Evolution*, **56,** 1902-1908.

Hedrick, P. W., Parker, K. M. & Lee, R. N. (2001) Using microsatellite and MHC variation to identify species, ESUs, and MUs in the endangered Sonoran topminnow. *Molecular Ecology*, **10,** 1399-1412.

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

Hewitt, G. M. (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*, **58,** 247-276.

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

Hiendleder, S., Kaupe, B., Wassmuth, R., *et al.* (2002) Molecular analysis of wild and domestic sheep questions current nomenclature and provides evidence for domestication from two different subspecies`. *Proceedings of the Royal Society of London B*, **269,** 893-904.

Hillis, D. M., Moritz, C. & Mable, B. K. (1996) *Molecular systematics*. Sunderland: Sinauer Associates.

Hofreiter, M., Serre, D., Rohland, N., *et al.* (2004) Lack of phylogeography in European mammals before the last glaciation. *Proceedings of the National Academy of Sciences of the United States of America*, **101,** 12963-12968.

Hogg, J. T. (2000) Mating systems and conservation at large spatial scales. In: *Vertebrate Mating Systems* (Ed. by Apollonio, M., Festa-Bianchet, M. & Mainardi, D.), pp. 332. Singapore: World Scientific Publishing.

Holder, K., Montgomerie, R. & Friesen, V. L. (1999) A test of the glacial refugium hypothesis using patterns of mitochondrial and nuclear DNA sequence variation in rock ptarmigan (*Lagopus mutus*). *Evolution*, **53,** 1936-1950.

Holder, K., Montgomerie, R. & Friesen, V. L. (2000) Glacial vicariance and historical biogeography of rock ptarmigan (*Lagopus mutus*) in the Bering region. *Molecular Ecology*, **9,** 1265-1278.

Horin, P., Smola, J., Matiasovic, J., *et al.* (2004) Polymorphisms in equine immune response genes and their associations with infections. *Mammalian Genome*, **15,** 843-850.

Huang, S.-W. & Yu, H.-T. (2003) Genetic variation of microsatellite loci in the major histocompatibility complex (MHC) region in the southeast Asian house mouse (*Mus musculus castaneus*). *Genetica*, **119,** 201-218.

Hudson, R. R., Kreitman, M. & Aguade, M. (1987) A test of neutral molecular evolution based on nucleotide data. *Genetics*, **116,** 153-159.

Hughes, A. L. & Nei, M. (1988) Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature*, **335,** 167-170.

Hughes, A. L. & Nei, M. (1989) Nucleotide substitution at major histocompatibility complex class II loci: Evidence for overdominant selection. *Proceedings of the National Academy of Science, USA*, **86,** 958-962.

Hutchison, D. W. & Templeton, A. R. (1999) Correlation of pairwise genetic and geographic distance measures: Inferring the relative influences of gene flow and drift on the distribution of genetic variability. *Evolution*, **53,** 1898-1914.

Huttley, G. A., Smith, M. W., Carrington, M., *et al.* (1999) A scan for linkage disequilibrium across the human genome. *Genetics*, **152,** 1711-1722.

Jarvi, S. I., Tarr, C. L., McIntosh, C. E., *et al.* (2004) Natural selection of the major histocompatibility complex (MHC) in Hawaiian honeycreepers (Drepanidinae). *Molecular Ecology*, **13,** 2157-2168.

Jeffery, K. J. M. & Bangham, C. R. M. (2000) Do infectious diseases drive MHC diversity? *Microbes and Infection*, **2,** 1335-1341.

Jenkins, N. P., Brooks, N. H. & Hutchinson, I. V. (2000) An interferon-gamma microsatellite polymorphism is associated with coronary artery disease severity. *European Heart Journal*, **21,** 248-248.

Jepsen, B. I., Siegismund, H. R. & Fredholm, M. (2002) Population genetics of the native caribou (*Rangifer tarandus groenlandicus*) and the semi-domestic reindeer (*Rangifer tarandus tarandus*) in Southwestern Greenland: Evidence of introgression. *Conservation Genetics*, **3,** 401-409.

Jugo, B. M. & Vicario, A. (2000) Single-strand conformational polymorphism and sequence polymorphism of Mhc-DRB in Latxa and Karrantzar sheep: implications for Caprinae phylogeny. *Immunogenetics*, **51,** 887-897.

Kauer, M. O., Dieringer, D. & Schlotterer, C. (2003) A microsatellite variability screen for positive selection associated with the "Out of Africa" habitat expansion of *Drosophila melanogaster*. *Genetics*, **165,** 1137-1148.

Kim, Y. & Nielsen, R. (2004) Linkage disequilibrium as a signature of selective sweeps. *Genetics*, **167,** 1513-1524.

Kimura, M. (1955) Random genetic drift in multi-allelic loci. *Evolution*, **9,** 419-435.

Kimura, M. (1968) Evolutionary rate at the molecular level. *Nature*, **217,** 624-626.

Kimura, M. (1981) Estimation of evolutionary distances between homologous nucleotide sequences. *Proceedings of the National Academy of Science, USA*, **78,** 454-458.

Kimura, M., Maruyama, T. & Crow, J. F. (1963) Mutation load in small populations. *Genetics*, **48,** 1303-&.

Kimura, M. & Ohta, T. (1978) Stepwise mutation model and distribution of allelic frequencies in a finite population. *Proceedings of the National Academy of Science, USA*, **75,** 2868-2872.

Knowles, L. L. (2004) The burgeoning field of statistical phylogeography. *Journal of Evolutionary Biology*, **17,** 1-10.

Knowles, L. L. & Madison, W. P. (2002) Statistical phylogeography. *Molecular Ecology*, **11,** 2623-2635.

Kohn, M. H., Pelz, H. J. & Wayne, R. K. (2000) Natural selection mapping of the warfarin-resistance gene. *Proceedings of the National Academy of Sciences of the United States of America*, **97,** 7911-7915.

Kostia, S., Kantanen, J., Kolkkala, M., *et al.* (1998) Applicability of SSCP analysis for MHC genotyping: fingerprinting of Ovar-DRB1 exon 2 alleles from Finnish and Russian breeds. *Animal Genetics*, **29,** 453-455.

Kreitman, M. (2000) Methods to detect selection in populations with applications to the human. *Annual Review of Genetics*, **1,** 539-559.

Kreitman, M. & Akashi, H. (1995) Molecular evidence for natural selection. *Annual Review of Ecology and Systematics*, **26,** 403-422.

Kruglyak, S., Durrett, R., Schug, M. D., *et al.* (2000) Distribution and abundance of microsatellites in the yeast genome can be explained by a balance between slippage events and point mutations. *Molecular Biology and Evolution*, **17,** 1210-1219.

Kumar, G. S. N., Govindiah, M. G., Nagaraj, C. S., *et al.* (1999) Does the natural resistance associated macrophage protein (Nramp) gene confer resistance/susceptibility in bovines against mycobacterial infection? *Current Science*, **77**, 809-812.

Kumar, S., Tamura, K., Jakobsen, I. B., *et al.* (2001) MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics*, **17**, 1244-1245.

Kutz, S. J., Veitch, A. M., Hoberg, E. P., *et al.* (2001) New host and geographic records for two protostrongylids in Dall's sheep. *Journal of Wildlife Diseases*, **37**, 761-774.

Kyle, C. J. & Strobeck, C. (2002) Connectivity of peripheral and core populations of North American wolverines. *Journal of Mammalogy*, **83**, 1141-1150.

Landry, C. & Bernatchez, L. (2001) Comparative analysis of population structure across environments and geographical scales at major histocompatibility complex and microsatellite loci in Atlantic salmon (*Salmo salar*). *Molecular Ecology*, **10**, 2525-2539.

Langefors, A., Lohm, J., Grahn, M., *et al.* (2001) Association between major histocompatibility complex class IIB alleles and resistance to *Aeromonas salmonicida* in Atlantic salmon. *Proceedings of the Royal Society of London B*, **268**, 479-485.

Levson, V. M. & Rutter, N. W. (1996) Evidence of Cordilleran Late Wisconsinan glaciers in the 'ice- free corridor'. *Quaternary International*, **32**, 33-51.

Lewontin, R. C. & Krakauer, J. (1973) Distribution of gene frequency as a test of theory of selective neutrality of polymorphisms. *Genetics*, **74**, 175-195.

Loehr, J., Worley, K., Grapputo, A., *et al.* (2005) North American mountain sheep phylogeography: Evidence for multiple refugia and gene flow along the 'ice-free corridor'. *Journal of Evolutionary Biology*, in press.

Luikart, G. & Allendorf, F. W. (1996) Mitochondrial DNA variation and genetic population structure in Rocky Mountain bighorn sheep (*Ovis canadensis condenses*). *Journal of Mammalogy*, **77**, 109-123.

Mackintosh, C. G., Qureshi, T., Waldrup, K., *et al.* (2000) Genetic resistance to experimental infection with *Mycobacterium bovis* in red deer (*Cervus elaphus*). *Infection and Immunity*, **68,** 1620-1625.

Maddox, J. F., Davies, K. P., Crawford, A. M., *et al.* (2001) An enhanced linkage map of the sheep genome comprising more than 1000 loci. *Genome Research*, **11,** 1275-1289.

Mandryk, C. A. S., Josenhans, H., Fedje, D. W., *et al.* (2001) Late Quaterary paleoenvironments of Northwestern North America: implications for inland versus coastal migration routes. *Quaternary Science Reviews*, **20,** 301-314.

Manel, S., Berthier, P. & Luikart, G. (2002) Detecting wildlife poaching: identifying the origin of individuals with Bayesian assignment tests and multilocus genotypes. *Conservation Biology*, **16,** 650-659.

Matthews, G. D. & Crawford, A. M. (1998) Cloning, sequencing and linkage mapping of the NRAMP1 gene of sheep and deer. *Animal Genetics*, **29,** 1-6.

Maudet, C., Miller, C., Bassano, B., *et al.* (2002) Microsatellite DNA and recent statistical methods in wildlife conservation management: applications in Alpine ibex [*Capra ibex (ibex)*]. *Molecular Ecology*, **11,** 421-436.

Maynard Smith, J. & Haigh, J. (1974) The hitch-hiking effect of a favourable gene. *Genetical Research*, **23,** 23-35.

Mikko, S. & Andersson, L. (1995) Low major histocompatibilty complex class II diversity in European and North American moose. *Proceedings of the National Academy of Sciences of the United States of America*, **92,** 4259-4263.

Mikko, S., Roed, K., Schmutz, S., *et al.* (1999) Monomorphism and polymorphism at Mhc DRB loci in domestic and wild ruminants. *Immunological Reviews*, **167,** 169-178.

Mikko, S., Spencer, M., Morris, B., *et al.* (1997) A comparative analysis of Mhc DRB3 polymorphism in the American bison (*Bison bison*). *Journal of Heredity*, **88,** 499-503.

Monello, R. J., Murray, D. L. & Cassirer, E. F. (2001) Ecological correlates of pneumonia epizootics in bighorn sheep herds. *Canadian Journal of Zoology-Revue Canadienne De Zoologie*, **79,** 1423-1432.

Moritz, C. (1994) Defining 'evolutionary significant units' for conservation. *Trends in Ecology & Evolution*, **9,** 373-375.

Muirhead, C. A. (2001) Consequences of population structure on genes under balancing selection. *Evolution*, **55,** 1532-1541.

Nei, M. (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, **89,** 538-590.

Nei, M. & Gojobori, T. (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution*, **3,** 418-426.

Nei, M. & Maruyama, T. (1975) Lewontin-Krakauer test for neutral genes - Comment. *Genetics*, **80,** 395-395.

Nielsen, R. (2001) Statistical tests of selective neutrality in the age of genomics. *Heredity*, **86,** 641-647.

Nielsen, R. & Wakely, J. (2001) Distinguishing migration from isolation: a Markov chain Monte Carlo approach. *Genetics*, **158,** 885-896.

Nikolajczyk, B., Nelsen, B. & Sen, R. (1996) Precise alignment of sites required for mu enhancer activation in B cells. *Molecular Cell Biology*, **16,** 4544-4554.

Noon, T. H., Wesche, S. L., Cagle, D., *et al.* (2002) Hemorrhagic disease in bighorn sheep in Arizona. *Journal of Wildlife Diseases*, **38,** 172-176.

Obexer-Ruff, G., Sattler, U., Martinez, D., *et al.* (2003) Association studies using random and "candidate" microsatellite loci in two infectious goat diseases. *Genetics Selection Evolution*, **35,** S113-S119.

Otto, S. P. (2000) Detecting the form of selection from DNA sequence data. *Trends in Genetics*, **16,** 526-529.

Paetkau, D., Amstrup, S. C., Born, E. W., *et al.* (1999) Genetic structure of the world's polar bear populations. *Molecular Ecology*, **8,** 1571-1584.

Paetkau, D., Waits, L. P., Clarkson, P. L., *et al.* (1997) An empirical evaluation of genetic distance statistics using microsatellite data from bear (Ursidae) populations. *Genetics*, **147,** 1943-1957.

Page, R. D. M. (1996) TREEVIEW: An application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences*, **12,** 357-358.

Paterson, K. A. & Crawford, A. M. (2000) Ovine microsatellite OarKP6 isolated from a BAC containing the ovine interferon gamma gene. *Animal Genetics*, **31,** 343-343.

Paterson, S. (1996) Major histocompatibility complex variation in Soay sheep. In: *Department of Genetics*, pp. 127. Cambridge: University of Cambridge.

Paterson, S. (1998) Evidence for balancing selection at the major histocompatibility complex in a free-living ruminant. *Journal of Heredity*, **89,** 289-294.

Paterson, S., Wilson, K. & Pemberton, J. M. (1998) Major histocompatibility complex variation associated with juvenile survival and parasite resistance in a large unmanaged ungulate population (*Ovis aries L.*). *Proceedings of the National Academy of Sciences of the United States of America*, **95,** 3714-3719.

Payseur, B. A., Cutter, A. D. & Nachman, M. W. (2002) Searching for evidence of positive selection in the human genome using patterns of microsatellite variability. *Molecular Biology and Evolution*, **19,** 1143-1153.

Penty, J. M., Henry, H. M., Ede, A. J., *et al.* (1993) Ovine microsatellites at the OarAE16, OarAE54, OarAE57, OarAE119 and OarAE129 loci. *Animal Genetics*, **24,** 219.

Perez, T., Albornoz, J. & Dominguez, A. (2002) Phylogeography of chamois (*Rupicapra spp.*) inferred from microsatellites. *Molecular Phylogenetics and Evolution*, **25,** 524-534.

Pfau, R. S., Van Den Bussche, R. A., McBee, K., *et al.* (1999) Allelic diversity at the Mhc-DQA locus in cotton rats (*Sigmodon hispidus*) and a comparison of DQA sequences within the family Muridae (Mammalia : Rodentia). *Immunogenetics*, **49,** 886-893.

Pielou, E. C. (1991) *After the ice age: return of life to glaciated North America.* Chicago: University of Chicago Press.

Pitel, F., Lantier, I., Gellin, J., *et al.* (1996) Two polymorphic microsatellite markers close to the ovine NRAMP gene. *Animal Genetics*, **27,** 60-61.

Pokorny, V., McLean, L., McQueen, F., *et al.* (2001) Interferon-gamma microsatellite and rheumatoid arthritis. *Lancet*, **358,** 122-123.

Polziehn, R. O., Beech, R., Sheraton, J., *et al.* (1996) Genetic relationships among North American bison populations. *Canadian Journal of Zoology-Revue Canadienne De Zoologie*, **74,** 738-749.

Polziehn, R. O., Hamr, J., Mallory, F. F., *et al.* (2000) Microsatellite analysis of North American wapiti (Cervus elaphus) populations. *Molecular Ecology*, **9,** 1561-1576.

Posada, D. & Crandall, K. A. (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics*, **14,** 817-818.

Primmer, C. R., Ellegren, H., Saino, N., *et al.* (1996) Directional evolution in germline microsatellite mutations. *Nature Genetics*, **13,** 391-393.

Primmer, C. R., Koskinen, M. T. & Piironen, J. (2000) The one that did not get away: individual assignment using microsatellite data detects a case of fishing competition fraud. *Proceedings of the Royal Society of London Series B-Biological Sciences*, **267,** 1699-1704.

Primmer, C. R. & Matthews, M. E. (1993) Ovine dinucleotide repeat polymorphism at the Vias-s5 locus. *Animal Genetics*, **24,** 146.

Pritchard, J. K., Stephens, M. & Donnelly, P. (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155,** 945-959.

Pritchard, J. K. & Wen, W. (2002) Documentation for *structure* software: Version 2.0. University of Chicago.

Prusiner, S. B. (1982) Novel proteinaceous infectious particles cause scrapie. *Science*, **216,** 136-144.

Przeworski, M. (2002) The signature of positive selection at randomly chosen loci. *Genetics*, **160,** 1179-1189.

Puurtinen, M., Knott, K. E., Suonpaa, S., *et al.* (2004) Genetic variability and drift load in populations of an aquatic snail. *Evolution*, **58,** 749-756.

Ramey, R. R. (1993) Evolutionary genetics and systematics of North American mountain sheep: implications for conservation. Cornell University.

Ramey, R. R. (1995) Mitochondrial DNA variation, population structure, and evolution of mountain sheep in the south-western United-States and Mexico. *Molecular Ecology*, **4,** 429-439.

Rannala, B. & Mountain, J. L. (1997) Detecting immigration using multilocus genotypes. *Proceedings of the National Academy of Science, USA*, **94,** 9197-9201.

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

Reiter, Z. (1993) Interferon - a major regulator of natural-killer cell-mediated cytotoxicity. *Journal of Interferon Research*, **13,** 247-257.

Rice, W. R. (1989) Analysing tables of statistical tests. *Evolution*, **43,** 223-225.

Robertson, A. (1975) Lewontin-Krakauer test for neutral genes - Comment. *Genetics*, **80,** 396-396.

Rogers, A. R. & Harpending, H. (1992) Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution*, **9,** 552-569.

Rosenblum, E. B., Hoekstra, H. E. & Nachman, M. W. (2004) Adaptive reptile color variation and the evolution of the Mc1r gene. *Evolution*, **58,** 1794-1808.

Rozas, J. & Rozas, R. (1999) DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics*, **15,** 174-175.

Rozen, S. & Skaletsky, H. J. (2000) Primer3 on the WWW for general users and for biologist programmers. In: *Bioinformatics Methods and Protocols: Methods in Molecular Biology* (Ed. by Krawetz, S. & Misener, S.), pp. pp365-386. Totowa, NJ: Humana Press.

Rudolph, K. M., Hosch, T. L. & Heimer, W. E. (2001) NRAMP final project report. Caldwell: Wildlife health laboratory, Idaho department of fish and game.

Rudolph, K. M., Hunter, D. L., Foreyt, W. J., *et al.* (2003) Sharing of Pasteurella spp. between free-ranging bighorn sheep and feral goats. *Journal of Wildlife Diseases*, **39,** 897-903.

Sage, R. D. & Wolff, J. O. (1986) Pleistocene glaciations, fluctuating ranges and low genetic variability in a large mammal (*Ovis dalli*). *Evolution*, **40,** 1092-1095.

Salamon, H., Klitz, W., Easteal, S., *et al.* (1999) Evolution of HLA class II molecules: Allelic and amino acid site variability across populations. *Genetics*, **152,** 393-400.

Saltercid, L. & Flajnik, M. F. (1995) Evolution and developmental regulation of the major histocompatibility complex. *Critical Reviews in Immunology*, **15,** 31-75.

Sambrook, J., Fritsch, E. & Maniatis, T. (1989) *Molecular Cloning: a Laboratory Manual.* New York, USA: Cold Spring Harbor Laboratory Press.

Sander, P., Hamann, H., Pfeiffer, I., *et al.* (2004) Analysis of sequence variability of the bovine prion protein gene (PRNP) in German cattle breeds. *Neurogenetics*, **5,** 19-25.

Satta, Y., Kupfermann, H., Li, Y. J., *et al.* (1999) Molecular clock and recombination in primate MHC genes. *Immunological Reviews*, **167,** 367-379.

Satta, Y., O'hUigin, C., Takahata, N., *et al.* (1994) Intensity of natural selection at the major histocompatibility complex loci. *Proceedings of the National Academy of Science, USA*, **91,** 7184-7188.

Sauermann, U., Nurnberg, P., Bercovitch, F. B., *et al.* (2001) Increased reproductive success of MHC class II heterozygous males among free-ranging rhesus macaques. *Human Genetics*, **108,** 249-254.

Schierup, M. H., Mikkelsen, A. M. & Hein, J. (2001) Recombination, balancing selection and phylogenies in MHC and self-incompatibility genes. *Genetics*, **159,** 1833-1844.

Schierup, M. H., Vekemans, X. & Charlesworth, D. (2000) The effect of subdivision on variation at multi-allelic loci under balancing selection. *Genetical Research*, **76,** 51-62.

Schlotterer, C. (2002) A microsatellite-based multilocus screen for the identification of local selective sweeps. *Genetics*, **160,** 753-763.

Schmidt, P., Ludt, C., Kuhn, C., *et al.* (1996) A diallelic tetranucleotide repeat, (GT(3))(5 or 6), within intron 1 of the ovine interferon-gamma gene. *Animal Genetics*, **27,** 437-438.

Schneider, S. & Excoffier, L. (1999) Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: Application to human mitochondrial DNA. *Genetics*, **152,** 1079-1089.

Schneider, S., Roessli, D. & Excoffier, L. (2000) Arlequin: A software for population genetics data analysis.: Genetics and Biometry laboratory, Department of Anthropology, University of Geneva.

Schwaiger, F. W., Buitkamp, J., Weyers, E., *et al.* (1993) Typing of MHC-DRB genes with the help of intronic simple repeated DNA sequences. *Molecular Ecology*, **2,** 260-272.

Schwaiger, F. W., Weyers, E., Buitkamp, J., *et al.* (1994) Interdependent MHC-DRB exon plus intron evolution in Artiodactyls. *Molecular Biology and Evolution*, **11,** 239-249.

Seabury, C. M., Honeycutt, R. L., Rooney, A. P., *et al.* (2004) Prion protein gene (PRNP) variants and evidence for strong purifying selection in functionally important regions of bovine exon 3. *Proceedings of the National Academy of Sciences of the United States of America*, **101,** 15142-15147.

Sheldon, C. (1911) *The Wilderness of the Upper Yukon.* New York: Charles Scribner's Sons.

Sibly, R. M., Whittaker, J. C. & Talbot, M. (2001) A maximum-likelihood approach to fitting equilibrium models of microsatellite evolution. *Molecular Biology and Evolution*, **18,** 413-417.

Sigurdardottir, S., Borsch, C., Gustafsson, K., *et al.* (1991) Cloning and sequence analysis of 14 DRB alleles of the bovine major histocompatibility complex by using the polymerase chain reaction. *Animal Genetics*, **22,** 199-209.

Simmons, N. M. (1982) Seasonal ranges of Dall's sheep, Mackenzie Mountains, Northwest Territories. *Arctic*, **35,** 512-518.

Slate, J., Coltman, D. W., Goodman, S. J., *et al.* (1998) Bovine microsatellite loci are highly conserved in red deer (*Cervus elaphus*), sika deer (*Cervus nippon*) and Soay sheep (*Ovis aries*). *Animal Genetics*, **29,** 307-315.

Slatkin, M. (1987) Gene flow and the geographic structure of natural populations. *Science*, **236,** 787-792.

Slatkin, M. (1993) Isolation by distance in equilibrium and nonequilibrium populations. *Evolution*, **47,** 264-279.

Slatkin, M. (1995a) Hitchhiking and associative overdominance at a microsatellite locus. *Molecular Biology and Evolution*, **12,** 473-480.

Slatkin, M. (1995b) A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, **139,** 1463-1463.

Slatkin, M. & Hudson, R. R. (1991) Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics*, **129,** 555-562.

Slatkin, M. & Wiehe, T. (1998) Genetic hitch-hiking in a subdivided population. *Genetical Research*, **71,** 155-160.

Stajich, J. E. & Hahn, M. W. (2005) Disentangling the effects of demography and selection in human history. *Molecular Biology and Evolution*, **22,** 63-73.

Stone, R. T., Pulido, J. C. & Duyk, G. M. (1995) A small insert genomic library highly enriched for microsatellite repeat sequences. *Mammalian Genome*, **6,** 714-724.

Storz, J. F. (1999) Genetic consequences of mammalian social structure. *Journal of Mammalogy*, **80,** 553-569.

Storz, J. F. & Dubach, J. M. (2004) Natural selection drives altitudinal divergence at the albumin locus in deer mice, *Peromyscus maniculatus*. *Evolution*, **58,** 1342-1352.

Storz, J. F., Payseur, B. A. & Nachman, M. W. (2004) Genome scans of DNA variability in humans reveal evidence for selective sweeps outside of Africa. *Molecular Biology and Evolution*, **21,** 1800-1811.

Sunyaev, S., Kondrashov, F. A., Bork, P., *et al.* (2003) Impact of selection, mutation rate and genetic drift on human genetic variation. *Human Molecular Genetics*, **12,** 3325-3330.

Swarbrick, P. A., Buchanan, F. C. & Crawford, A. (1991) Ovine dinucleotide repeat polymorphism at the MAF36 locus. *Animal Genetics*, **22,** 377-378.

Swarbrick, P. A., Schwaiger, F. W., Epplen, J. T., *et al.* (1995) Cloning and sequencing of expressed DRB genes of the red deer (*Cervus elaphus*) MHC. *Immunogenetics*, **42,** 1-9.

Swofford, D. L. (2002) PAUP*: phylogenetic analysis using parsimony (*and other methods). Sunderland: Sinauer Associates.

Szeicz, J. M. & MacDonald, G. M. (2001) Montane climate and vegetation dynamics in easternmost Beringia during the Late Quaternary. *Quaternary Science Reviews*, **20,** 247-257.

Tajima, D. (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123,** 585-595.

Talbot, S. L. & Shields, G. F. (1996) Phylogeography of brown bears (*Ursus arctos*) of Alaska and paraphyly within the Ursidae. *Molecular Phylogenetics and Evolution*, **5,** 477-494.

Thursz, M. R., Thomas, H. C., Greenwood, B. M., *et al.* (1997) Heterozygote advantage for HLA class-II type in hepatitis B virus infection. *Nature Genetics*, **17,** 11-12.

Tranulis, M. A. (2002) Influence of the prion protein gene, *Prnp*, on scrapie susceptibility in sheep. *APMIS*, **110,** 33-43.

Tremblay, N. O. & Schoen, D. J. (1999) Molecular phylogeography of *Dryas intergrifolia*: glacial refugia and postglacial recolonisation. *Molecular Ecology*, **8,** 1187-1198.

Troy, C. S., MacHugh, D. E., Bailey, J. F., *et al.* (2001) Genetic evidence for Near-Eastern origins of European cattle. *Nature*, **410,** 1088-1091.

Valdez, R. & Krausman, P. R. (1999) *Mountain sheep of North America.* Tucson: The University of Arizona Press.

van Deventer, S. J. H. (2000) Cytokine and cytokine receptor polymorphisms in infectious disease. *Intensive Care Medicine*, **26,** S98-S102.

van Haeringen, W. A., Gwakisa, P. S., Mikko, S., *et al.* (1999) Heterozygosity excess at the cattle DRB locus revealed by large scale genotyping of two closely linked microsatellites. *Animal Genetics*, **30,** 169-176.

van Hooft, W. F., Groen, A. F. & Prins, H. H. T. (2003) Genetic structure of African buffalo herds based on variation at the mitochondrial D-loop and autosomal microsatellite loci: Evidence for male-biased gene flow. *Conservation Genetics*, **4,** 467-477.

van Oosterhout, C., van Heuven, M. K. & Brakefield, P. M. (2004) On the neutrality of molecular genetic markers: pedigree analysis of genetic variation in fragmented populations. *Molecular Ecology*, **13,** 1025-1034.

Vandendries, E. R., Johnson, D. & Reinke, R. (1996) Orthodenticle is required for photoreceptor cell development in the Drosophila eye. *Developmental Biology*, **173,** 243-255.

Veitch, A. & Simmons, E. (1999) Mackenzie Mountain non-resident and non-resident alien hunter harvest summary 1999. Norman Wells, NWT, Canada: Department of Resources, Wildlife and Economic Development.

Vidal, S., Gros, P. & Skamene, E. (1995) Natural resistance to infection with intracellular parasites - molecular genetics identifies Nramp1 as the Bcg/Ity/Lsh locus. *Journal of Leukocyte Biology*, **58,** 382-390.

Vigouroux, Y., McMullen, M., Hittinger, C. T., *et al.* (2002) Identifying genes of agronomic importance in maize by screening microsatellites for evidence of selection during domestication. *Proceedings of the National Academy of Sciences of the United States of America*, **99,** 9650-9655.

Vitalis, R., Dawson, K. & Boursot, P. (2001) Interpretation of variation across marker loci as evidence of selection. *Genetics*, **158,** 1811-1823.

Wall, J. D. (1999) Recombination and the power of statistical tests of neutrality. *Genetical Research*, **74,** 65-79.

Wall, J. D., Andolfatto, P. & Przeworski, M. (2002) Testing models of selection and demography in *Drosophila simulans*. *Genetics*, **162,** 203-216.

Waser, P. M. & Strobeck, C. (1998) Genetic signatures of interpopulation dispersal. *Trends in Ecology and Evolution*, **13,** 43-44.

Watson, P. J. & Davies, R. L. (2002) Outbreak of *Pasteurella multocida* septicaemia in neonatal lambs. *Veterinary Record*, **151,** 420-422.

Watterson, G. A. (1978) The homozygosity test of neutrality. *Genetics*, **88,** 405-417.

Wegner, K. M., Reusch, T. B. H. & Kalbe, M. (2003) Multiple parasites are driving major histocompatibility complex polymorphism in the wild. *Journal of Evolutionary Biology*, **16,** 224-232.

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

Wenink, P. W., Groen, A. F., Roelke-Parker, M. E., *et al.* (1998) African buffalo maintain high genetic diversity in the major histocompatibility complex in spite of historically known population bottlenecks. *Molecular Ecology*, **7,** 1315-1322.

Wickström, L. M., Haukisalmi, V., Varis, S., *et al.* (2003) Phylogeography of the circumpolar *Paranoplocephala arctica* species complex (Cestoda: Anoplocephalidae) parasitising collared lemmings (*Dicrostonyx* spp.). *Molecular Ecology*, **12,** 3359-3371.

Wiehe, T. (1998) The effect of selective sweeps on the variance of the allele distribution of a linked multiallele locus: Hitchhiking of microsatellites. *Theoretical Population Biology*, **53,** 272-283.

Wiehe, T. & Slatkin, M. (1998) Epistatic selection in a multi-locus Levene model and implications for linkage disequilibrium. *Theoretical Population Biology*, **53,** 75-84.

Wiener, P., Burton, D., Ajmone-Marsan, P., *et al.* (2003) Signatures of selection? Patterns of microsatellite diversity on a chromosome containing a selected locus. *Heredity*, **90,** 350-358.

Willis, K. J. & Whittaker, R. J. (2000) The refugial debate. *Science*, **287,** 1406-1407.

Wilson, G. A. & Strobeck, C. (1999) Genetic variation within and relatedness among wood and plains bison populations. *Genome*, **42,** 483-496.

Wilson, G. A., Strobeck, C., Wu, L., *et al.* (1997) Characterization of microsatellite loci in caribou *Rangifer tarandus*, and their use in other artiodactyls. *Molecular Ecology*, **6,** 697-699.

Wilson, P., Grewal, S., Rodgers, A., *et al.* (2003) Genetic variation and population structure of moose (*Alces alces*) at neutral and functional DNA loci. *Canadian Journal of Zoology*, **81,** 670-683.

Worley, K., Strobeck, C., Arthur, S., *et al.* (2004) Population genetic structure of North American thinhorn sheep (*Ovis dalli*). *Molecular Ecology*, **13,** 2545-2556.

Wright, S. (1978) *Evolution and the Genetics of Populations. Volume IV: Variability Within and Among Natural Populations*. Chicago: University of Chicago Press.

Xu, H. Y., Chakraborty, R. & Fu, Y. X. (2005) Mutation rate variation at human dinucleotide microsatellites. *Genetics*, **170,** 305-312.

Yamada, N. A., Smith, G. A., Castro, A., *et al.* (2002) Relative rates of insertion and deletion mutations in dinucleotide repeats of various lengths in mismatch repair proficient mouse and mismatch repair in human cells. *Mutation Research*, **499,** 213-225.

Yang, Z. H. & Nielsen, R. (2000) Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Molecular Biology and Evolution*, **17,** 32-43.

Yeager, M. & Hughes, A. L. (1999) Evolution of the mammalian MHC: natural selection, recombination, and convergent evolution. *Immunological Reviews*, **167,** 45-58.

Zhu, Y., Queller, D. C. & Strassmann, J. E. (2000) A phylogenetic perspective on sequence evolution in microsatellite loci. *Journal of Molecular Evolution*, **50,** 324-338.

APPENDIX A. MOLECULAR METHODS

A.1. Modified DNA extraction from horn samples using Qiagen tissue kit

300-500µl volume of loosely packed horn cuttings were placed in a labelled 1.5µl tube ensuring that large pieces broken up. Where possible the inner regions of the core were chosen in preference to horn from the outer regions, which was often thicker and paler in colour. The inner horn is more likely to yield a greater final DNA concentration. 450-550µl of Buffer ATL was added to the horn ensuring the cuttings were covered. 50-60µl Proteinase K were added to the horn before vortexing. Samples were left to digest at 55°C overnight. After digestion the liquid volume was transferred to a new labelled 1.5µl tube and the volume was noted. An equal volume of Buffer AL was then added before vortexing and incubation at 70°C for 10 minutes. DNA was then precipitated by the addition of 420µl 100% ethanol. Samples were vortexed before addition of half of each mix to a Qiagen spin column. Liquid was spun through the spin columns by centrifuging at 13,000 rpm for 1 minute. Liquid was poured from the collection tube and spin columns replaced. The remaining half of the mixture was added and centrifugation repeated. Collection tubes were then discarded and spin columns placed into clean tubes. 500µl Buffer AW1 was added and each sample was centrifuged for 1 minute at 13,000 rpm. The liquid was discarded and 500µl Buffer AW2 added to each column. Columns were centrifuged for 1 minute and the eluted liquid was discarded. An additional 2 minute centrifuge was conducted to discard any remaining alcohol from the spin columns. Each spin column was placed in a clean 1.5µl labelled tube and 200µl Buffer AE added. Samples were incubated at 70°C for 5 minutes prior to final elution by 1 minute centrifugation at 13,000 rpm.

A. 2. Ligation protocol of DRB PCR products

A. LIGATIONS

a. Briefly centrifuge the pGEM-T vector and Control Insert DNA tubes to collect contents at the bottom of tubes.

b. Set up ligation reactions as below in 0.5ml tubes.

c. Vortex the 2X Rapid Ligation Buffer vigorously before each use.

	Standard Reaction	Positive Control	Background Control
2X Rapid Ligation Buffer, T4 DNA Ligase	5µl	5µl	5µl
pGEM-T Vector (50ng)	1µl	1µl	1µl
PCR product	Xµl*	-	-
Control insert DNA	-	2µl	-
T4 DNA Ligase (3 Weiss units/µl)	1µl	1µl	1µl
ddH_2O to volume of	10µl	10µl	10µl

* see below for optimising Insert:Vector molar ratios

d. Mix reactions by pipetting. Incubate reactions overnight at 4°C for maximum transformations.

B. TRANSFORMATIONS

a. Prepare 2 LB/ampicillin/IPTG/X-Gal plates for each ligation, plus two plates for determining transformation efficiency. Equilibrate the plates to room temperature before use.

b. Centrifuge ligation reactions to collect contents at the bottom of the tubes. Add 5µl of each ligation reaction to a sterile 1.5ml tube on ice. Set up another tube with 0.1ng uncut plasmid.

c. Remove tube of frozen JM109 cells from –70°C and place in an ice bath until just thawed (~5min). Mix by gently flicking.

d. Transfer 100µl cells into each tube from step b.

e. Gently flick tubes to mix and place on ice for 45mins.

f. Heat-shock cells for 90s in a water bath at exactly 42°C (do not shake).
g. Return cells to ice for 2min.
h. Add 600µl LB broth to tubes containing transformed cells and 900µl to the tube with uncut plasmid.
i. Incubate for 1 hour at 37°C with shaking (~150rpm).
j. Plate 100µl each culture onto plates. Dilute the transformation control (1:10) with LB broth before plating.
k. Incubate the plates overnight (16-24 hours) at 37°C. White colonies generally contain inserts.

C. OPTIMISING INSERT: VECTOR MOLAR RATIOS

a. The concentration of PCR product should be estimated.
b. pGEM-T vector is ~3kb and 50ng/µl
c. Estimate the amount of PCR product using the formula below.

$$\frac{\text{ng of vector} \times \text{kb size of insert}}{\text{kb size of vector}} \times \text{insert:vector molar ratio} = \text{ng of insert}$$

so if 0.3kb PCR product is used with 50ng of 3kb vector, with ratio between 3:1 and 1:3, then;

$$\frac{50\text{ng vector} \times 0.3\text{kb insert}}{3\text{kb vector}} \times \frac{3}{1} \text{ or } \frac{1}{3} = 1^2/_3 \text{ and } 15\text{ng insert.}$$

APPENDIX B: REFERENCE SEQUENCES

1. GenBank accession numbers for all MHC DRB sequences used in this thesis

Alces alces
X83278
X83279
X83283
X83284
X83282
X83285
AF043191
AF043193
AF043190
AF043192
AF043194

Bison spp (*B. bonasus)
X98649
X98653
X98651
X98652
X98650
X98648
X98647
U82824
U82825
AY805099
AY805102
AY805105
AY805106
AY805107
AY805108
AF022373*
AF022374*

Bos indicus
X75035
Z30649
Z30652
Z36543
Z36542
Z48224
X87646
X87653
X87663
X87665
X87666
X87667
X87668
X87669
X87670
Z82023
Z82024
Z82028
Z82029
Z82031
Z82032
Z82033
Z82034
AF272867
AJ001999
AJ002000
AJ002002
AJ002004
AJ277153
AJ277154
AJ277155
AJ277156
AF261953
AF261954
AF272862
AF272863
AF272864
AF272866
AF272867
AF272868
AF272869
AF272871
AF272872
AF272873
AF272874
AF272875
AF272876
AF272878
AF272879
AF272881
AF272882
AF272883
Z82996
Z82997
Z82998
Z82999
Z83000
AF209710

Bos taurus
D45358
AB048811
AB048815
X92409
AB053167
AF376811
U64462
U64464
M99050
M99051
M99052
M30012
U77067
U78548
U90921
AJ505018
AY125892
AF008234
AF008235
M94924
M94926
AF515452
AF010395
AF010396
U00123
U00125

U00128
U00130
U00131
U00133
U00134
U00135
U00137
U00138
U00139
U00141
U00142
Y18308
Z30650
Z30651
L36154
L36155
AB033386
AB033396
AB033387
Z36536
Z36539
AF144543
AF144544
AF144545
AF144546
AF144547
AF144548
AF149790
X87641
X87644
X87647
X87649
X87650
X87651
X87652
X87654
X87657
X87660
X87662
X87664

Bubalus bubalis
AF261955
AF261956
AF270653
AF270654
AF270655
AF270656
AF270657
AF270658
AF270659
AF270661
AF270663
AF270664
AF270666
AF270667
AF270668
AF270670
AF270672
AF270674
AF385473
AF385474
AF385475
AF385476
AF385477
AF385478
AF385479
AY496063

Capra aegagrus
U00183
U00184
U00185
U00186
U00187
U00188
U00189
U00190
U00191
U00192
U00193
U00194
U00195
U00196
U00197
U00198
U00199
U00200
U00201
U00202
U00203
Z92710
Z92711
Z92712
Z92713
Z92714
Z92715
Z92716
Z92717
Z92718
Z92719
Z92720
Z92721
Z92723
Z92724
Z92725

Capra hircus
AB008347
AB008348
AB008350
AB008351
AB008352
AB008353
AB008354
AB008355
AB008358
AB008359
AB008360
AB008361
AB008362
AF376809
AY496061
AY496062
AY496935

Capra pyrenaica
AF461692
AF461693
AF461694
AF461695
AF461696
AY351788

Cervus elaphus
U11101
U11102
U11105

U11106
U11107
U11108
U11109
U11110
U11111
U11112
U11113
U11114
U11115
U11116
U11117
U11118
U11119
U11120
U11121
U11122
U11123
U11223
U11224
U11226
U11227
U11228
U11229
U11230
U11231
U11232
U11233
U11234
U11210
U11211
U11212
U11213
U11214
U11215
U11216
U11217
U11218
U11219
U11220
U11221

Damaliscus pygargus
AJ302736
AJ302738
AJ302739
AJ302740
AJ302741
AJ302742
AJ302743
AJ302744
AJ302745
AJ302746
AJ302747
AJ302748
AJ302749
AJ302750
AJ302751
AJ302752
AJ302753
AJ302754
AJ302755
AJ302756
AJ302757
AJ302758
AJ302759
AJ302760
AJ302761
AJ302762

Ovibos moschatus
AF012715

Ovis aries
AB061317
AB061318
AB061319
AB061320
AB061321
AB061322
AB061323
AB061372
AF376812
M73984
Y09961
Y10245
Y10246
Y10248
Y10249
Z92726
Z92727
Z92728
Z92730
Z92731
Z92732
Z92733
Z92734
AY230000
AY248695
AY307083
AY227049
U00204
U00205
U00206
U00207
U00208
U00209
U00210
U00213
U00215
U00216
U00217
U00218
U00219
U00221
U00222
U00223
U00224
U00225
U00226
U00227
U00228
U00229
U00232
U00233
U00234
U00235
U00236
U00237
AY496059
AY496060
AF036558
AF036559
AF036560
AF036561
AF036562
AF036564
AB017205

AB017206
AB017207
AB017209
AB017211
AB017213
AB017214
AB017215
AB017216
AB017217
AB017218
AB017219
AB017220
AB017221
AB017222
AB017223
AB017224
AB017225
AB017226
AB017227
AB017228
AB017229
X76905
AY691948
AF126433
AF126434
AF126436
AF126437
AF126438
AF126439
AF126441

Ovis canadensis
AF324840
AF324841
AF324842
AF324843
AF324844
AF324845
AF324846
AF324847
AF324848
AF324849
AF324850
AF324851
AF324852
AF324853
AF324854
AF324855
AF324856
AF324857
AF324858
AF324859
AF324860
AF324861

Ovis dalli
AJ920396
AJ920397
AJ920398
AJ920399
AJ920400
AJ920401
AJ920402
AJ920403
AJ920404
AJ920405
AJ920406
AJ920407
AJ920408
AJ920409
AJ920410
AJ920411
AJ920412

Rupicapra pyrenaica
AY212149
AY212150
AY212151
AY212152
AY212153
AY212154
AY212155
AY212156
AY212157
AY898752
AY898753
AY898754
AY898755

Rupicapra rupicapra
AF336340
AY368437
AY368438
AY368439
AY368441
AY368442
AY368443
AY368444
AY368446
AY368447
AY368448
AY368449
AY368451
AY368453
AY368454

Syncerus caffer
AF059233
AF059234
AF059235
AF059236
AF059237
AF059238
AF059239
AF059240
AF059241
AF385481

2. GenBank accession numbers for IFNG sequences used in this thesis

Bos taurus
Z54144
NM 174086
AF533686

Bubalus bubalis
AF484688

Ovis aries
AY575608
X52640
AY575609

Sus scrofa
X53085
AY188089

Camelus bactrianus
AB107657

Lama glama
AB107652

Cervus elaphus
X63079
L07502

Capra hircus
U34232
AY304501

Equus asinus
AF055341

Equus caballus
U04050
D28520

Homo sapien
X13274

3. GenBank accession numbers for NRAMP sequences used in this thesis

Ovis aries
Z34916
U70255

Bison bison
U39614

Bos taurus
U12862

Bubalus bubalis
U27105
AY860618

Bos indicus
AY338470

Sus scrofa
NM 213821
U55068

Equus caballus
AF354445

Cervus elaphus
AF005379

Mus musculus
NM 013612

4. Exon 2 and 3 and surrounding introns of *O. dalli* Interferon Gamma– all fixed, with exons highlighted.

GCCAATTTTTTCTTGTGTTGTTTTATTTTCCAG**AATGCAAGTAACCCAGATGTAGCTAAGGGTGGGCC**
TCTTTTCTCAGAAATTTTGAAGAATTGGAAAGAGGTAAGCTGAACATTTCCATTTGGCTGATTTTCCT
GTTGCTTTTTTTCTGATGGATAAATTCACATCAACCTCTCTTTGTGCTCTTTTCTCCCAAG**GAGAGTG**
ACAAAAAGATTATTCAGAGCCAAATTGTCTCCTTCTACTTCAAACTCTTTGAAAACCTCAAAGATAAC
CAGGTCATTCAAAGGAGCATGGATATCATCAAGCAAGACATGTTTCAGAAGTTCTTGAACGGCAGCTC
TGAGAAACTGGAGGACTTCAAAAGGCTGATTCAAATTCCGGTGAGAGGATCTTAATGCTTTCTTTGGT
ATCATTGCAGAGACTCTTACAAAGCACTTCATTCCCAGAAAATAGAAATTAGATACCAGGTATAAAGC
TCTGAAGTGAAGTTGTAGTCAAAATGCCTCTTTGCTAGTTCTCTTTTGCCTTGGGTGACTTTGCAGAG
TCAGTTATTGGAGGCACTTAAG

APPENDIX C: MANUSCRIPTS BASED ON DATA INCLUDED IN THIS THESIS

C.1. Worley K, Strobeck C, Arthur S, *et al.* (2004) Population genetic structure of North American thinhorn sheep (*Ovis dalli*). *Molecular Ecology* **13**, 2545-2556.

C. 2. Loehr J, Worley K, Grapputo A, *et al.* (2005) Evidence for cryptic glacial refugia from North American mountain sheep mitochondrial DNA. *Journal of Evolutionary Biology, in press*

Molecular Ecology (2004) 13, 2545–2556 doi: 10.1111/j.1365-294X.2004.02248.x

Population genetic structure of North American thinhorn sheep (*Ovis dalli*)

K. WORLEY,* C. STROBECK,† S. ARTHUR,‡ J. CAREY,§ H. SCHWANTJE,¶ A. VEITCH** and D. W. COLTMAN*

*Department of Animal and Plant Sciences, University of Sheffield, Western Bank, Sheffield, UK, S10 2TN, †Department of Biological Sciences, University of Alberta, Edmonton, Alberta, AB, Canada, T6G 2E9, ‡Alaska Department of Fish and Game, 1300 College Road, Fairbanks, AK, USA, 99701–1599, §Department of Environment, Government of Yukon, Whitehorse, Yukon, Canada, Y1A 2C6, ¶Biodiversity Branch, Ministry of Water, Land and Air Protection, PO Box 9338, 2975 Jutland Road, Victoria, BC, Canada, V8W 9M1, **Department of Resources, Wildlife and Economic Development, Sahtu Region, Government of the Northwest Territories, PO Box 130, Norman Wells, NWT, Canada, X0E 0V0*

Abstract

The thinhorn sheep (*Ovis dalli* ssp.) provides a rare example of a North American large mammal that occupies most of its native range and maintains close to ancestral population size. There are currently two recognized subspecies, Dall's sheep (*O. d. dalli*) and Stone's sheep (*O. d. stonei*), the validity of which remains uncertain. We investigated the spatial genetic structure of thinhorn sheep populations representing both subspecies by genotyping individuals ($n = 919$) from across the species range at 12 variable microsatellite loci. We found high levels of genetic diversity within ($H_E = 0.722$) and significant genetic structure among the 24 sampled areas ($F_{ST} = 0.160$). Genetic distance measures and Bayesian clustering analyses revealed the presence of at least eight subpopulations that are delineated by mountain range topology. A strong overall pattern of isolation-by-distance is evident across the sampling range ($r = 0.75$, $P < 0.001$) suggesting limited dispersal and extensive philopatry. Partial Mantel tests of this relationship showed mountain range distinctions represent significant barriers to gene flow ($P = 0.0001$), supporting the Bayesian analyses. Genetic structure was more strongly pronounced in southern Yukon and Alaska than elsewhere. We also show evidence for genetic differences between the two currently recognized thinhorn subspecies.

Keywords: assignment test, F_{ST}, isolation-by-distance, microsatellites, *Ovis*, ungulate

Received 24 February 3004; revision received 30 April 2004; accepted 30 April 2004

Introduction

Many species show geographical stratification of genetic diversity. At selectively neutral loci, population genetic structure is primarily determined by the interplay between genetic drift and the rate of gene flow between geographically separated subpopulations (Slatkin 1987). In theory, relatively little migration is required for the homogenization of allele frequencies across populations at equilibrium under the island model (Wright 1978). However, as most species have a dispersal range considerably less than that of the species range, population structure often reflects a pattern of isolation-by-distance (Slatkin 1993). Quantifying levels of genetic structure can therefore give an insight into the colonization history of a species (Perez *et al.* 2002), rates of dispersal (Waser & Strobeck 1998) and the effects of environmental barriers on gene flow (Paetkau *et al.* 1999; Carmichael *et al.* 2001; Kyle & Strobeck 2002). In trophy game animals, such as thinhorn sheep (*Ovis dalli* spp.), knowledge of genetic structure may also have practical implications for harvest management and for forensic analyses of suspected illegal hunting (Primmer *et al.* 2000).

Due to the variety of factors that affect population structure, there are differences in scale and pattern of structure across different taxa. Many signatures of structure can be predicted based on life history traits. Large carnivores, such as wolves and bears, generally have larger potential dispersal distances and therefore decreased levels of structure than mountain ungulates (Forbes & Hogg 1999). Forest- or

Correspondence: K. Worley. Fax: + 44 (0) 114 2220002; E-mail: k.worley@sheffield.ac.uk

plains-dwelling ungulates show low to moderate levels of genetic structure (Broders *et al.* 1999; Wilson & Strobeck 1999; Polziehn *et al.* 2000) as they are able to utilize larger areas of continuous suitable habitat and may conduct long-distance migration. Animals with more stringent habitat preference criteria, such as mountain sheep, may be expected to exhibit more pronounced population structure.

Wild sheep are dependent upon habitat with steep rugged cliffs for use as escape terrain with nearby open grazing areas for feeding (Geist 1971; Valdez & Krausman 1999). For most populations, the overall annual range consists of relatively unlimited summer habitat, restricted only by accessibility to escape terrain and bounded by natural barriers. Topographical features such as forested valleys and rivers present natural boundaries to dispersal. The range shrinks in winter to smaller areas with limited snow accumulation, open grassland and available escape terrain, resulting in higher sheep densities and setting the upper limit to herd size. Mountain sheep are highly philopatric, with individuals utilizing the same seasonal home ranges each year (Geist 1971; Festa-Bianchet 1991). As is the case with most mammals (Greenwood 1980), dispersal is largely male-biased (Hogg 2000) and ewes rarely leave the natal home range (Festa-Bianchet 1991; Valdez & Krausman 1999). The combination of limited gene flow and small population size should cause genetic differences to accumulate rapidly between geographically separated populations.

Perhaps unsurprisingly, previous studies on mountain sheep have demonstrated considerable genetic structure. Mitochondria DNA analyses in desert bighorn (*Ovis canadensis nelsoni*) ewes showed significant differentiation at the level of the home range group (Ramey 1995; Boyce *et al.* 1999). Studies using nuclear markers have found significant genetic distances between Rocky Mountain bighorn (*O. c. canadensis*) populations sampled across the species range (Forbes & Hogg 1999), and on a much finer scale between recently reintroduced populations (Fitzsimmons *et al.* 1997).

Human activity is also likely to have influenced the genetic structure of bighorn sheep and other mountain ungulates. Translocations, population declines due to unregulated hunting and habitat loss and the transmission of novel pathogens from domestic sheep have severely impacted bighorn sheep population structure (Luikart & Allendorf 1996). These factors make it difficult to sample representatively the native range of the bighorn sheep. In contrast, thinhorn occupy most of their historic range in approximately ancestral numbers. Their range encompasses much of mountainous northwestern North America, from Alaska to the Mackenzie River in the east, and to the Pine Pass of the Rocky Mountains of northeastern British Columbia in the southeast. There are an estimated 130 000 individuals present (Barichello *et al.* 1989; Valdez & Krausman 1999; Veitch & Simmons 1999). Cases of domestic livestock introduction in these mostly remote areas are few, having little impact on wild sheep numbers. In addition there have been limited recorded translocations either into or out of populations. Thinhorn sheep are therefore an ideal model in which to study mountain ungulate population structure as it evolved prior to anthropogenic influence.

Thinhorn sheep are classified in two subspecies on the basis of coat colour. The more abundant white Dall's sheep (*O. dalli dalli*) are found across much of the species range, except the far south. The darker Stone's sheep (*O. d. stonei*) are less numerous and are found in the Yukon and northern British Columbia only. The taxonomic validity of the subspecies status in thinhorn and other wild sheep species is debatable. Genetic analyses of desert bighorn have shown that currently recognized subspecies do not provide a full explanation for phylogenetic boundaries (Ramey 1993; Gutierrez-Espeleta 1999). Rather, it could be that differences between individuals are representative of variation in one species over a geographical range. Genetic study may aid our understanding of the unresolved taxonomy of thinhorn.

In this study we aimed to quantify levels of genetic variation in thinhorn sheep populations across the species range. A previous thinhorn study using allozyme markers found little variation (Sage & Wolff 1986). Here we used ungulate-derived microsatellites (de Gortari *et al.* 1997; Slate *et al.* 1998) to investigate population structure and differentiation between thinhorn subspecies. We hypothesized that thinhorn sheep are highly genetically structured, with genetic variation partitioned among contiguous mountain ranges and subspecies.

Materials and methods

Sample locations and collection

We sampled 919 thinhorn sheep from 24 sampling areas across the species range, from northwestern Alaska to British Columbia (Fig. 1). Samples for DNA extraction were collected between 1994 and 2002 and comprised flakes of horn produced by drilling horn material to insert metal identification plugs (horn 'corings') of hunted rams taken at kill registration or compulsory inspection. Alaskan samples comprised whole blood taken from both rams and ewes during the period 1999–2002. Sample area boundaries were defined from game management zones or units in the Northwest Territories (NWT) and northern British Columbia (BC) and by groups of adjacent management subzones from the Yukon Territory (YK). Alaskan sample areas comprised individuals collected from the same mountain range, with maximum distance separating samples of between 20 km (in the Central Alaskan Range: CAR) and 140 km (in the Gates of the Arctic Reserve: GAAR). The 24 sample areas represent seven contiguous mountain blocks defined in Table 1. These mountain blocks consist of continuous

Table 1 Genetic variability estimates from populations included in the study (mean number of alleles per locus (A), observed (H_O) and expected heterozygosities (H_E), and mean estimates of F_{IS} (Weir & Cockerham 1984)). Also included are the number of loci showing significant deviation from Hardy–Weinberg Equilibrium at each population (at $P < 0.05$). *Global departure across all loci from HWE; significance of F_{IS} is indicated *$P < 0.05$, **$P < 0.01$

Region	Sample area	Abbr.	Mountain group	N	% Dall's in sample	H_O	H_E	Loci not in HWE	A	F_{IS}
Alaska	Central Alaskan Range	CAR	Alaska range	25	100	0.58	0.58	1	5.17	0.020
	Gates of the Arctic National Park Preserve	GAAR	Brookes range	34	100	0.55	0.60	3*	5.17	0.107**
	Noatak National Preserve	NK	Brookes range	25	100	0.63	0.60	0	3.97	−0.028
	Yukon-Charley Rivers National Preserve	YUCH	White Hills	36	100	0.57	0.59	4*	5.25	0.050*
Northwest Territories	Game management zone D/OT/01	D/OT/01	Mackenzie mountains	40	100	0.62	0.65	0	6.92	0.055**
	Game management zone D/OT/02	D/OT/02	Mackenzie mountains	42	100	0.54	0.57	5*	6.92	0.070**
	Game management zone G/OT/01	G/OT/01	Mackenzie mountains	40	100	0.60	0.62	1	6.92	0.047*
	Game management zone S/OT/01	S/OT/01	Mackenzie mountains	40	100	0.62	0.63	2	7.50	0.040
	Game management zone S/OT/02	S/OT/02	Mackenzie mountains	40	100	0.64	0.65	3	7.33	0.024
	Game management zone S/OT/03	S/OT/03	Mackenzie mountains	40	100	0.60	0.65	1	7.08	0.096**
	Game management zone S/OT/04	S/OT/04	Mackenzie mountains	40	100	0.59	0.66	1	7.25	0.119**
	Game management zone S/OT/05	S/OT/05	Mackenzie mountains	39	100	0.64	0.67	2	7.42	0.061*
Yukon Territories	Game management zone 2 (Ogilvie mountains)	Y2W	Mackenzie mountains	40	93	0.62	0.67	4	6.58	0.087**
	Game management zone 2 (Wernecke mountains)	Y2E	Mackenzie mountains	40	100	0.63	0.66	4	6.58	0.057*
	Game management zone 4	Y4	Pelly mtns	40	36	0.59	0.66	4*	7.17	0.116**
	Game management zone 5	Y5	Southwest range	40	100	0.56	0.58	3	5.67	0.051*
	Game management zone 7	Y7	Southwest range	40	100	0.55	0.59	5	6.17	0.064**
	Game management zone 8	Y8	Pelly mtns	40	38	0.57	0.61	2	6.00	0.070**
	Game management zone 10	Y10	Pelly mtns	40	32	0.60	0.64	0	6.33	0.070**
British Columbia	Game zone 742	BC742	Northern Rockies	40	0	0.48	0.49	2	4.42	0.042
	Game zone 750	BC750	Northern Rockies	40	0	0.51	0.54	2	5.17	0.071**
	Game zone 751	BC751	Northern Rockies	40	0	0.47	0.52	2	5.58	0.104**
	Game zone 752	BC752	Northern Rockies	38	0	0.47	0.48	1	5.00	0.043
	Game zone 754	BC754	Northern Rockies	40	0	0.46	0.51	2	4.83	0.101**

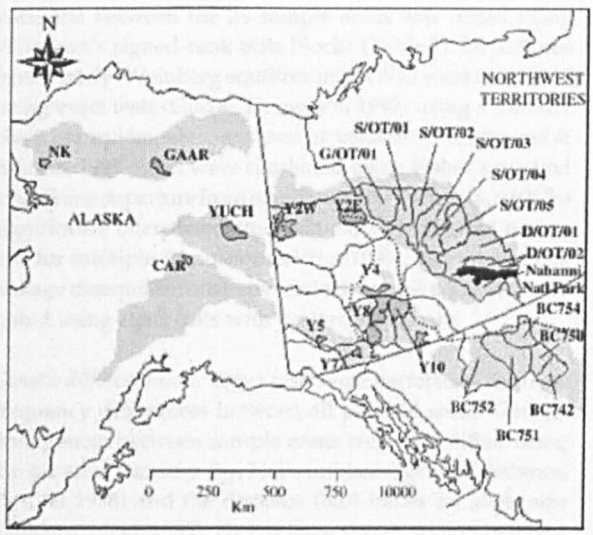


Fig. 1 North American *Ovis dalli* population locations included in this study. Sampling locations in Yukon and Alaska were within the black shaded areas. Populations in the Northwest Territories and British Columbia cover all the area within the labelled region. Abbreviations for sample area names are described in Table 1. Lightly shaded regions represent the full extent of mountain ranges. The dotted line represents the northern limits of Stone's sheep range.

upland regions separated by unsuitable sheep habitat such as major river valleys or lowland forest. Between 25 and 42 individuals were sampled per area (mean of 38). Horn corings were stored dry in sealed paper envelopes, while whole blood samples were stored at –20 °C in EDTA.

Populations also varied with regard to the subspecies of the sheep sampled. NWT and Alaskan populations comprised all Dall's sheep, while Yukon populations contained herds of both Dall's and Stone's sheep (see Table 1). British Columbia individuals were exclusively Stone's.

Molecular techniques

Genomic DNA was extracted from approximately 0.5 mL of horn material per sample using a tissue extraction kit (Qiagen, Crawley, West Sussex, UK). Blood samples were extracted using a phenol–chloroform technique (Sambrook *et al.* 1989) from 200 μL of blood. Twelve dinucleotide microsatellite loci developed in domestic sheep and cattle (Table 2) were amplified for each individual. Loci were chosen based on levels of variability and previous successful use in bighorn sheep (Coltman *et al.* 2002; Coltman *et al.* 2003a).

Each polymerase chain reaction (PCR) was carried out in 10 μL reactions. Reactions contained 2 μL of DNA template, 80 μmol of each primer, 0.16 mM dNTPs, 2 mM $MgCl_2$ and 0.5 units *Taq* polymerase (Sigma, Gillingham, Dorset, or Bioline, London, UK). Additionally, the quality of some products was improved by the addition of 25 mg/mL bovine serum albumin (BSA). The PCR profile consisted of 35 cycles of 30 s each at 94 °C and 54 °C followed by 40 s at 72 °C. Cycles were preceded by 5 min at 94 °C and terminated with 10 min at 72 °C. PCR products were genotyped using an ABI 377 sequencer and analysed using the software GENESCAN and GENOTYPER (Applied Biosystems, Foster City, CA, USA).

Data analyses

Genetic diversity. We quantified genetic variability within each sample area by the number of alleles per locus (A) and expected heterozygosity (H_E) as calculated by GENETIX version 4.01 (Belkhir 1996). GENETIX was also used to determine these values for each locus. Homogeneity of genetic

Table 2 Comparison of the genetic diversity found at the 12 microsatellite markers used (number of alleles over all populations (A), size range of alleles and mean observed (H_O) and expected heterozygosities (H_E) of all sample areas). Also included are the number of populations out of 24 that show significant deviation from HWE ($P < 0.05$). *Global departure from HWE

Locus	A	Allele size range (bp)	H_O	H_E	Populations deviating from HWE	Reference
AE16	9	82–98	0.6250	0.6628	6*	(Penty *et al.* 1993)
BM1225	13	225–265	0.6284	0.6813	6*	(Bishop *et al.* 1994)
BM4505	11	239–275	0.4775	0.5472	9*	(Bishop *et al.* 1994)
BM4513	16	131–161	0.7122	0.7316	1	(Bishop *et al.* 1994)
BM848	22	201–243	0.7525	0.8305	10*	(Bishop *et al.* 1994)
BMC1222	4	288–294	0.1937	0.2254	1	(de Gortari *et al.* 1997)
CP26	19	123–165	0.6939	0.7168	3	(Ede *et al.* 1995)
FCB266	13	82–106	0.6262	0.6498	4	(Buchanan & Crawford 1993)
MAF209	5	109–119	0.2020	0.2067	1	(Buchanan & Crawford 1992)
MAF36	13	84–116	0.7227	0.7353	5	(Swarbrick *et al.* 1991)
TGLA126	11	112–144	0.5379	0.5361	5	(Georges & Massey 1992)
TGLA387	12	129–153	0.6771	0.7031	3	(Georges & Massey 1992)

variation between the 24 sample areas was tested using Wilcoxon's signed-rank tests blocks (Table 1). Departures from Hardy–Weinberg equilibrium (HWE) were examined using exact tests (Guo & Thompson 1992) using a Markov chain as implemented by GENEPOP version 3.3 (Raymond & Rousset 1995). Loci were combined using Fisher's method to examine departure from equilibrium for each area, with the significance interpreted after sequential Bonferroni correction for multiple comparisons (Rice 1989). The presence of linkage disequilibrium between loci over all areas was also tested using exact tests with GENEPOP software.

Genetic differentiation. Exact tests were performed for allele frequency differences between all pairs of areas. Genetic divergences between sample areas were quantified using the distance statistics F_{ST}, Nei's unbiased genetic distance, D_s (Nei 1978) and the distance $(\delta\mu)^2$ based on allele size (Goldstein *et al.* 1995). D_S has been shown to fare better at fine-scale population differentiation, while $(\delta\mu)^2$ has proved more useful for examining relationships between more distinct populations and for estimating evolutionary times (Paetkau *et al.* 1997). F_{ST} is insensitive when migration rates are low. We therefore used D_S rather than F_{ST} when constructing isolation-by-distance plots and trees. F_{ST} was calculated with GENEPOP, other distance measures were calculated using the software SPAGEDI version 1.0 (Hardy & Vekemans 2002). Significance of F_{ST} was tested with 10 000 permutations using GENETIX. PHYLIP 3.5 (Felsenstein 1993) was used together with the program TREEVIEW 1.5 (Page 1996) to construct an unrooted tree from the D_S distance matrix.

The Bayesian methodology of STRUCTURE version 2.0 (Pritchard *et al.* 2000) was implemented to determine the level of genetic substructure in the data set independently of sampling areas. To estimate the number of subpopulations (K), five independent runs of $K = 1$–20 were carried out at 100 000 Markov chain Monte Carlo (MCMC) repetitions. The most probable number of populations was taken using the log-likelihood of K. Individuals were then assigned to each subpopulation, based on the highest percentage membership (q). Sample locations of these individuals were then plotted on a map of the species range to examine the relative geographical position of subpopulations.

Isolation-by-distance. The relationship between genetic and geographical distances was examined to assess isolation-by-distance (Slatkin 1993). Geographic distances between pairs of areas were calculated from linear distances between mean latitude and longitude positions of samples from each area and plotted against the genetic distances D_S and $(\delta\mu)^2$. We also investigated the effects of mountain block and colour polymorphism on genetic distance controlling for the effects of geographical distance by partial Mantel tests (10 000 permutations) as calculated by the R-PACKAGE version 4.0 (Casgrain & Legendre 2001). To assess the effects of subspecies and mountain block on genetic distance independently of geographical distance, we used a general linear model of the residuals from linear regressions of genetic on geographical distances.

Results

Genetic diversity and tests of disequilibrium

Twelve locus microsatellite profiles were recorded for 919 thinhorn sheep. To assess the accuracy of profiles obtained from horn material, a subset of 50 individuals were retyped. In all cases identical genotypes were returned (total of 600 genotypes from 12 loci), showing the profiles to be repeatable. Overall, our rate of missing data was 1.9%.

Between four (*BMC1222*) and 22 (*BM848*) alleles were found at the 12 loci (mean 12.33, SE 0.43). Expected heterozygosities per locus ranged from 0.229 (*MAF209*) to 0.922 (*BM848*) (Table 2). At less diverse loci several alleles were present in Mackenzie and southern YK populations, while one allele was fixed in Alaska and BC. Populations had a mean number of alleles per locus (A) between 3.97 and 7.50 (Table 1). Measures of genetic variation did not differ between areas (A: $Z = 0.1065$, $P = 0.9152$; H_O: $Z = -0.0152$, $P = 0.9879$), although the highest measures of genetic diversity were observed in areas from the Mackenzie and Yukon (Pelly and southwest) mountain ranges. F_{IS} for all but one area was positive, ranging from −0.028 to 0.119. Several significant departures from HWE due to heterozygote deficit were found, with every locus showing deviation in at least one area (see Tables 1 and 2 for distribution). The repeatability of our microsatellite genotypes, low rate of missing data and consistent levels of HWE disequilibrium across loci suggest that heterozygote deficiencies did not arise from PCR artefacts.

Tests for genotypic disequilibrium revealed no significant linkage after correcting for multiple comparisons. When uncorrected, linkage was suggested between *BM1225* and *FCB266* ($\chi^2 = 71.35$, d.f. = 48, $P = 0.016$). Physical linkage between these markers is unlikely as they are located on separate chromosomes in domestic sheep (Maddox *et al.* 2001) (16 and 25, respectively).

Genetic differentiation

Allele frequencies between pairs of sample areas differed significantly at up to all 12 loci (Table 3). Significant genic differentiation (all with $P < 0.0001$) was recorded between all but two pairs of areas (BC751 and BC754; $\chi^2 = 20.867$, d.f. = 24, and G/OT/01 and S/OT/03: $\chi^2 = 35.288$, d.f. = 24, $P = 0.0642$) when tests were combined across loci. Global F_{ST} was 0.160 with pairwise F_{ST} ranging from −0.0034 to 0.3575 (Table 3). The greatest levels of differentiation were

Table 3 Genetic distance matrix of pairwise F_{ST} (Weir & Cockerham 1984) as calculated by GENEPOP (Raymond & Rousset 1995) below the diagonal. All F_{ST} values are significant at < 0.01 except those marked *. Values above the diagonal are the number of significantly differentiated loci (from the 12 used) between pairwise sample areas

	CAR	GAAR	NK	YUCH	D/OT/01	D/OT/02	G/OT/01	S/OT/01	S/OT/02	S/OT/03	S/OT/04	S/OT/05	Y2W	Y2E	Y4	Y5	Y7	Y8	Y10	BC742	BC750	BC751	BC752	BC754
CAR		11	11	11	10	11	11	11	12	10	10	12	10	9	11	11	12	11	11	11	11	11	11	11
GAAR	0.140		8	11	12	12	12	12	12	12	12	12	11	11	12	12	11	11	11	12	12	12	12	12
NK	0.201	0.099		11	12	12	12	12	12	12	12	12	12	11	12	12	12	12	12	12	12	12	12	12
YUCH	0.112	0.148	0.191		12	12	12	11	11	12	12	12	12	10	11	12	12	11	11	12	12	12	12	12
D/OT/01	0.157	0.138	0.216	0.141		10	8	9	9	7	6	7	11	11	10	11	12	11	11	11	11	11	11	11
D/OT/02	0.164	0.167	0.240	0.150	0.065		11	11	10	10	10	11	12	11	10	11	12	12	12	11	10	10	10	11
G/OT/01	0.150	0.160	0.215	0.115	0.036	0.083		4	5	3	7	10	12	10	11	12	12	11	11	12	11	11	11	12
S/OT/01	0.140	0.130	0.207	0.124	0.037	0.066	0.023		3	4	5	11	12	9	12	12	12	12	12	12	12	12	12	12
S/OT/02	0.135	0.127	0.193	0.124	0.043	0.048	0.030	0.013		5	5	8	11	11	11	12	12	12	12	12	12	12	12	12
S/OT/03	0.144	0.145	0.210	0.121	0.018	0.064	0.003*	0.023	0.025		4	8	11	10	11	12	12	11	11	12	11	11	11	12
S/OT/04	0.128	0.119	0.188	0.115	0.017	0.056	0.026	0.008	0.019	0.013		8	11	9	11	12	12	11	11	12	12	11	11	12
S/OT/05	0.114	0.138	0.198	0.134	0.036	0.058	0.061	0.044	0.026	0.043	0.029		12	12	9	12	12	11	11	12	11	11	11	12
Y2W	0.119	0.155	0.186	0.108	0.096	0.113	0.081	0.066	0.052	0.083	0.069	0.082		10	12	12	12	12	12	12	12	12	12	12
Y2E	0.104	0.108	0.166	0.068	0.057	0.095	0.031	0.035	0.043	0.033	0.033	0.068	0.059		12	11	12	12	12	11	11	11	11	11
Y4	0.125	0.152	0.206	0.137	0.104	0.114	0.117	0.094	0.089	0.107	0.083	0.094	0.106	0.104		11	11	10	8	11	11	11	11	11
Y5	0.168	0.196	0.234	0.180	0.210	0.253	0.217	0.201	0.203	0.207	0.182	0.204	0.178	0.173	0.101		10	11	12	11	10	11	11	11
Y7	0.197	0.203	0.245	0.243	0.220	0.259	0.239	0.225	0.207	0.215	0.203	0.200	0.199	0.219	0.138	0.123		9	10	12	12	12	11	12
Y8	0.182	0.167	0.232	0.160	0.132	0.187	0.155	0.129	0.144	0.138	0.120	0.136	0.156	0.133	0.086	0.107	0.153		8	12	12	11	11	12
Y10	0.179	0.183	0.220	0.169	0.132	0.165	0.152	0.136	0.150	0.136	0.116	0.137	0.147	0.134	0.054	0.121	0.192	0.082		12	11	11	11	12
BC742	0.289	0.281	0.342	0.276	0.200	0.251	0.231	0.222	0.244	0.209	0.202	0.219	0.250	0.217	0.215	0.283	0.301	0.170	0.172		6	8	7	10
BC750	0.266	0.261	0.324	0.261	0.171	0.225	0.205	0.196	0.214	0.181	0.179	0.195	0.219	0.194	0.182	0.259	0.272	0.154	0.146	0.025		5	5	4
BC751	0.264	0.266	0.338	0.259	0.176	0.221	0.211	0.204	0.219	0.190	0.182	0.197	0.230	0.205	0.192	0.263	0.278	0.156	0.147	0.047	0.022		6	0
BC752	0.298	0.297	0.358	0.283	0.210	0.254	0.242	0.227	0.246	0.217	0.210	0.224	0.250	0.229	0.206	0.282	0.310	0.152	0.151	0.044	0.029	0.026		5
BC754	0.265	0.273	0.345	0.260	0.182	0.229	0.215	0.207	0.225	0.196	0.185	0.198	0.235	0.209	0.196	0.266	0.291	0.162	0.152	0.047	0.020	−0.003*	0.034	

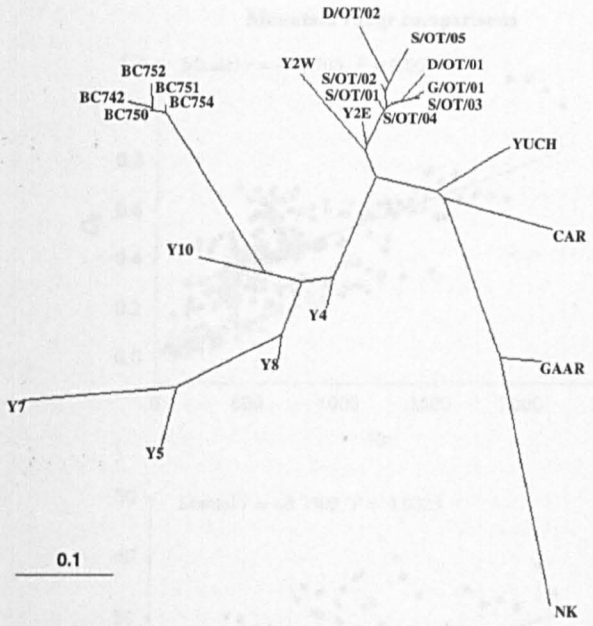


Fig. 2 Neighbour-joining tree of genetic distances (D_S) between sample areas of thinhorn sheep (*O. dalli*) constructed using TREEVIEW (Page 1996).

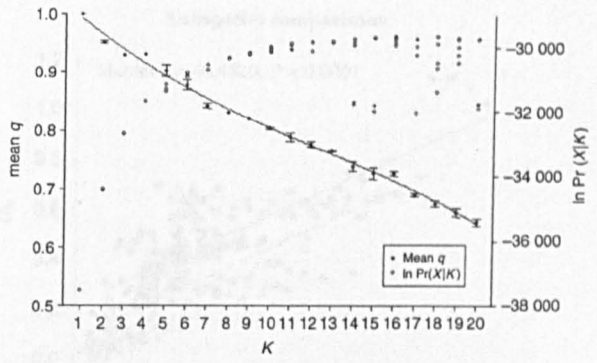


Fig. 3 Likelihood plot of structure results (MCMC with 100 000 repetitions) (Pritchard *et al.* 2000). Mean q is the mean confidence assignment of all individuals to their most probable cluster, indicating robustness of assignment (shown with standard error). ln Pr(X | K) is the log likelihood for each value of K, the number of simulated clusters. The most likely K is that where ln Pr(X | K) is maximized. The plot illustrates the difficulty in deciding on the most likely number of subpopulations in the data set due to isolation-by-distance.

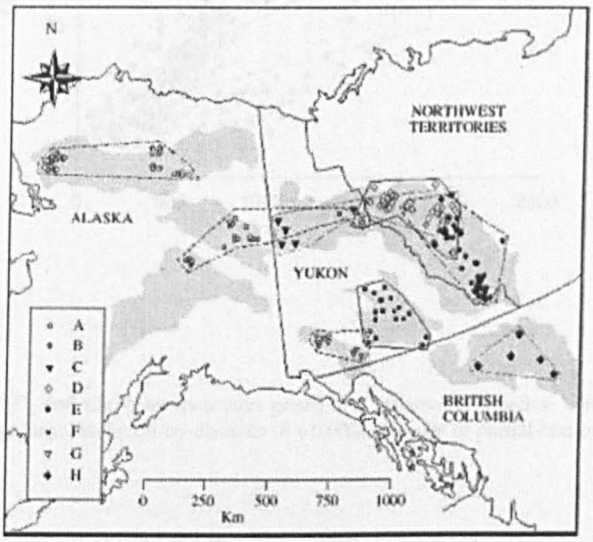


Fig. 4 Geographic locations of genetically similar clusters individuals as assigned by STRUCTURE at $K = 8$ (individuals with $q > 0.9$ only). Boundaries of each cluster are defined by lines surrounding all individuals within each. Due to the close sampling sites of some animals, one point on the map does not equal one sheep, but a variable number and is intended only as a guide to the geographical limitations of the cluster. Mountain range limitations are included in lighter shading to indicate obvious barriers to gene flow.

observed between Brooks Range and BC populations, whereas comparisons within the BC range and Mackenzie Mountains showed little genetic structure (Table 3). Patterns in the statistics D_S and $(\delta\mu)^2$ paralleled those found in F_{ST}. Correlation between D_S and F_{ST} was high ($r = 0.953$), whereas that of D_S and $(\delta\mu)^2$ ($r = 0.693$) and F_{ST} and $(\delta\mu)^2$ ($r = 0.741$) were moderate. An unrooted neighbour-joining tree constructed based on D_S revealed that sample areas on the same mountain range clustered together (Fig. 2).

The Bayesian structure analysis did not return an unequivocal number of genetic clusters as ln Pr(X | K) appears to gradually reach an asymptote beyond $K = 8$ (Fig. 3). We therefore examined patterns of cluster composition and confidence of cluster assignment at various values of K to determine the most probable number of subpopulations. Up to high values of K the cluster of BC individuals (which first occurred at $K = 2$) remained robust, indicating that this group is the most genetically differentiated within the data set. Other clear clusters were concordant with mountain range boundaries. The clustering at $K = 4$ consisted of Alaskan, Mackenzie, southern Yukon and BC groups. The division of Alaska into a Brooks range and mid-Alaskan cluster occurred at $K = 6$, along with a split of the Mackenzie range into two groups. After $K = 8$, the confidence of assignments fell dramatically, with some clusters having no individuals assigned at $q > 0.9$. Also, after $K = 8$, most new clusters were formed by further divisions within the Mackenzie range. We therefore decided to assign individuals to eight subpopulation clusters (Fig. 4). The split at $K = 2$ also showed evidence of subspecies differentiation. Individuals cross-assigning into the BC cluster were from sample areas consisting of mixed subspecies (Y4, 8 and 10). Stone's sheep from these mixed sample areas cross-assigned to the purely Stone's BC areas significantly more than Dall's ($\chi^2 = 4.81$,

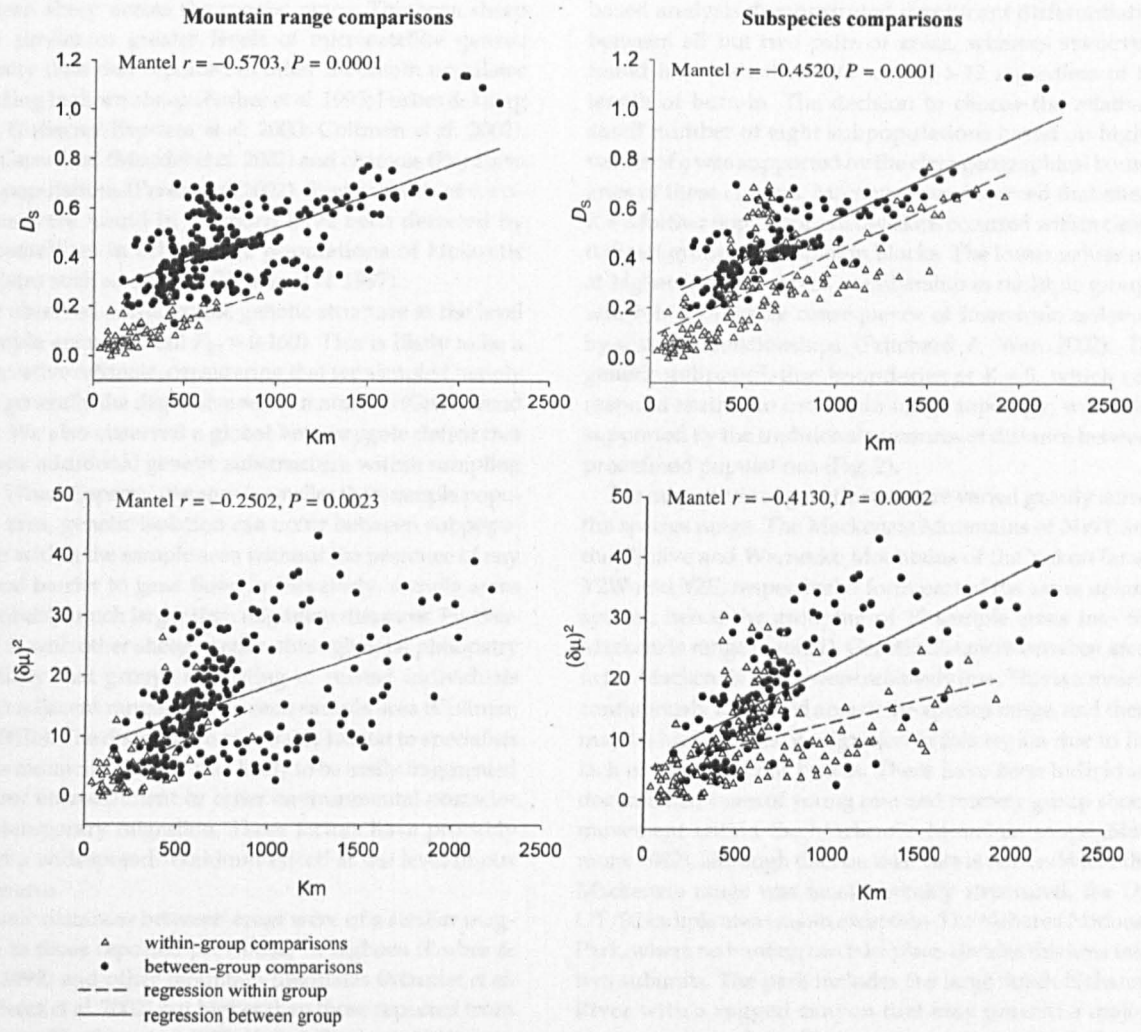


Fig. 5 Isolation-by-distance relationships between population pairs (D_S and $(\delta\mu)^2$) by mountain group and putative subspecies. Two measures of genetic distance are plotted (D_S and $(\delta\mu)^2$). Both show significant isolation-by-distance ($P = 0.0001$). Results of partial Mantel tests for the effect of mountain range and subspecies are shown.

d.f. = 1, $P < 0.05$). Cross-assigned sheep had lower confidences of assignment, q, than sheep originating from BC.

Isolation-by-distance

A significant isolation-by-distance (IBD) relationship was detected with both genetic distance measures (D_S: Mantel $r = 0.747$, $P = 0.0001$; $(\delta\mu)^2$: Mantel $r = 0.593$, $P = 0.0001$), although the measure based on allele identity fitted the data with slightly less scatter. Areas were split into those comprising mostly Dall's (between 90 and 100% of individuals) and those comprising mostly Stone's (between 62 and 100% of individuals) to examine patterns of IBD between and within subspecies (Fig. 5). Partial Mantel tests showed a highly significant effect of subspecies status on D_S (Mantel $r = -0.452$, $P = 0.0001$). When IBD was considered among populations between and within mountain ranges (as defined in Table 1) partial Mantel tests showed a significant within-range effect (Mantel $r = -0.570$, $P = 0.0001$). Analysis of residuals from a regression of genetic on geographical distance showed that D_S was affected by both mountain range ($t = 6.376$, $P < 0.0001$) and subspecies ($t = 4.511$, $P < 0.0001$). However $(\delta\mu)^2$ was affected only by subspecies ($t = 6.179$, $P < 0.0001$) and not by mountain range ($t = 0.418$, $P = 0.676$).

Discussion

This study has revealed for the first time the presence of extensive genetic variation and population substructure in

thinhorn sheep across the species range. Thinhorn sheep show similar or greater levels of microsatellite genetic diversity than that reported in other mountain ungulates including bighorn sheep (Forbes *et al.* 1995; Forbes & Hogg 1999; Gutierrez-Espeleta *et al.* 2000; Coltman *et al.* 2002), ibex (*Capra ibex*) (Maudet *et al.* 2002) and chamois (*Rupicapra* spp.) populations (Perez *et al.* 2002). Similar levels of variation as were found in thinhorn have been detected by microsatellites in other large populations of Holarctic ungulates such as caribou (Wilson *et al.* 1997).

We observed considerable genetic structure at the level of sample areas (overall $F_{ST} = 0.160$). This is likely to be a conservative estimate, considering that we sampled mainly rams, generally the dispersive sex in mammals (Greenwood 1980). We also observed a global heterozygote deficit that suggests additional genetic substructure within sampling areas. When dispersal distance is smaller than sample population area, genetic isolation can occur between subpopulations within the sample area without the presence of any physical barrier to gene flow. In this study, sample areas are probably much larger than migration distances. Furthermore, as with other sheep species, through natal philopatry it is likely that groups consisting of related individuals inhabit adjacent ranges within each sample area (Coltman *et al.* 2003b). The distribution of suitable habitat to specialists such as mountain sheep is also likely to be easily fragmented by forest encroachment or other environmental obstacles to contemporary migration. These factors have probably created a widespread 'Wahlund effect' at the level of our sample area.

Genetic distances between areas were of a similar magnitude to those reported previously in bighorn (Forbes & Hogg 1999) and other mountain ungulates (Maudet *et al.* 2002; Perez *et al.* 2002) but higher than those reported from carnivores (Paetkau *et al.* 1999; Kyle & Strobeck 2002) over similar geographical distances. This is expected due to the lower levels of migration and smaller home ranges characteristic of wild sheep populations. We found genetic differentiation between thinhorn sheep populations sampled less than 40 km apart show, at a similar scale for differentiation as that reported in desert bighorn sheep using mtDNA (Boyce *et al.* 1999) and microsatellites (Gutierrez-Espeleta *et al.* 2000).

Although we have shown fairly strong genetic differentiation and a robust pattern of isolation-by-distance among areas, the delineation of our sampling areas may be partly an artefact of wildlife management and political boundaries. The Bayesian method (Pritchard *et al.* 2000) is robust to bias introduced by sample area boundaries, as populations are not defined a priori. However, the results of the STRUCTURE analyses were not clear to interpret. This may be partly because the algorithm implemented by STRUCTURE is not well-suited for situations where there is isolation-by-distance (Pritchard & Wen 2002). The traditional area-based analysis demonstrated significant differentiation between all but two pairs of areas, whereas STRUCTURE found inconsistent results with $K > 13$ regardless of the length of burn-in. The decision to choose the relatively small number of eight subpopulations based on higher values of q was supported by the clear geographical boundaries of these clusters. Moreover, we observed that after a $K = 8$ further population subdivisions occurred within clearly defined groups of mountain blocks. The lower values of q at higher K reflect mixed membership in multiple groups, which is a probable consequence of finer-scale isolation-by-distance relationships (Pritchard & Wen 2002). The genetic subpopulation boundaries at $K = 8$, which correspond mainly to mountain range topology, were also supported by the traditional measures of distance between predefined populations (Fig. 2).

The magnitude of genetic structure varied greatly across the species range. The Mackenzie Mountains of NWT and the Ogilive and Wernecke Mountains of the Yukon (areas Y2W and Y2E, respectively) form part of the same upland system, hence the grouping of 10 sample areas into the Mackenzie range (Table 1). Genetic distances between areas in the Mackenzie range were relatively low. This is a mainly continuously inhabited area of the species range, and there may be higher rates of migration in this region due to the lack of major habitat breaks. There have been individual documented cases of young ram and nursery group sheep movement within the Mackenzie Mountain range (Simmons 1982), although data on their fate is scarce. While the Mackenzie range was mostly weakly structured, the D/OT/02 sample area was an exception. The Nahanni National Park, where no hunting can take place, divides this area into two subunits. The park includes the large South Nahanni River with a rugged canyon that may present a major barrier to gene flow. This may explain relatively large genetic distances between this area and the rest of the Mackenzie range.

Thinhorn sheep population structure will also be influenced by patterns of range expansion following the recession of the Laurentide and Cordilleran ice sheets after the last Pleistocene glaciation (Sage & Wolff 1986). At the last glacial maximum the southern Yukon and BC were under ice, whereas the Mackenzie Mountains, west-central Yukon and much of Alaska were largely unglaciated (Dyke *et al.* 2002). The current BC populations may have therefore been founded by colonization of a small number of Stone's sheep migrating southward from refugia in south-central Yukon. Alternately, BC thinhorn populations could be relicts from small ice-free refugia that existed in the northern Rocky Mountains in the seam between the Laurentide and Cordilleran ice sheets (Catto *et al.* 1996). Either scenario is consistent with the lower levels of observed genetic variation (Table 1) and genetic homogeneity (Fig. 2) of BC thinhorn populations.

Higher levels of genetic variation and structure were evident in the southern Yukon sample areas. If these areas were mainly ice-free during the last period of glaciation, then higher levels of variation would be consistent with the relative age of these populations. Pronounced population structure and long branch lengths separating these populations (Fig. 2) may therefore reflect bottleneck-like effects of recolonization from multiple refugia in this region. However, it may also be a consequence of reduced contemporary gene flow due to post-Pleistocene forest encroachment and more recent anthropogenic habitat fragmentation.

Alaskan areas were highly genetically differentiated. Genetic distance measures between Alaskan areas were greater than expected from geographical distance alone, and we found very high q-values for individuals in the two clusters based in the region (Fig. 4). This is consistent with topographical features of the region that may limit gene flow. Distances through suitable sheep habitat in Alaska are larger than the linear distance between sampling locations due to large areas of unsuitable habitat, which create semi-insular populations. The Alaskan interior lowlands and the Yukon River valley separate the Central Alaska and Brooks ranges, leading to large genetic distances between the GAAR/NK and YUCH sample areas in Alaska (Table 2, Fig. 2) and discrimination of these groups by STRUCTURE (Fig. 4). The range covered by YUCH is relatively continuous with that of the West Ogilvie Mountains (Y2W), however, providing easier route of gene flow between YUCH and Y2W. This is also reflected in the STRUCTURE analyses, with spread of the mid-Alaskan-based cluster into northern Yukon. We also note that genetic distances may be upwards biased between Alaskan populations, and genetic diversity decreased as a result of the sampling procedure in these regions. Ewes collected close together are likely to be more related than the hunted ram samples from elsewhere. However, we suggest that even with this factor, genetic distances between Alaskan ranges are higher than those between other mountain groups.

After accounting for subspecies and mountain range effects, thinhorn sheep population structure is well described by an isolation-by-distance pattern (Fig. 3) which reflects hypothetical regional equilibrium between gene flow and drift (Hutchison & Templeton 1999). The gradient of the overall isolation-by-distance plot is greater than many carnivore populations, but of a similar magnitude to bighorn sheep (Forbes & Hogg 1999). This positive relationship is caused typically by limited dispersal. Wild sheep are highly philopatric with little migration from the natal region, and a strong association with winter range. Low rates of gene flow among thinhorn sheep populations are suggested by high F_{ST} values, large degree of population differentiation and the relatively steep gradient of the isolation-by-distance plot.

Differences in pelage colour and skull measurements delineate the classification of *O. d. dalli* and *O d. stonei* subspecies (Ramey 1993). Our data also provide evidence of genetic differentiation between subspecies (Fig. 5). Genetic distances between areas consisting of the same subspecies were smaller than distances between areas of different subspecies after correcting for geographical distance. However, genetic distances are quite variable and there is considerable overlap between the within- and between-subspecies comparisons, a trend also reported in bighorn sheep (Forbes & Hogg 1999). Additional evidence of genetically differentiated subspecies comes from STRUCTURE analyses. Stone's sheep from mixed sample areas clustered more with BC samples than with areas comprising pure Dall's. In these mixed areas sheep had lower assignment confidences at $K = 2$ than in areas of one subspecies, reflecting their mixed ancestry. This provides evidence of a zone of introgression whereby 'Stone's' alleles have been introduced from the South. Genetic evidence is supported by pelage patterns with many sheep from southern Yukon showing an intermediate colour. These so-called fannin sheep are presumed to be the result of cross-breeding between the two subspecies. In thinhorn therefore genetic evidence of subspecies status is concordant with pelage colour. This is not often seen in wild sheep.

Figure 5 shows a greater effect of subspecies than mountain range on $(\delta\mu)^2$ than on D_S. As measures of allele size variance may better reveal distinctions between more deeply divergent populations (Forbes *et al.* 1995; Paetkau *et al.* 1997), this suggests that the genetic difference between subspecies is more ancient than the division resulting from mountain block.

In summary, microsatellite analyses of thinhorn sheep populations have demonstrated considerable genetic structure across the species range. The pattern of genetic differentiation is broadly consistent with isolation-by-distance. However, subspecies differentiation, colonization from multiple glacial refugia and the effects of mountain range topology complicate this pattern. Our microsatellite data provide some support for the subspecies status of Dall's and Stone's sheep.

Acknowledgements

This research was supported by the Department of Environment (Yukon), the Department of Resources, Wildlife and Economic Development, Sahtu Region (NWT), the Natural Sciences and Engineering Research Council (Canada), the Natural Environment Research Council (UK) and the University of Sheffield (UK). We thank Richard Popko (NWT), Philip Merchant (YK), Jim Lawlor, Chris Kleckner and Layne Adams (AK) and all the hard-working provincial, territorial and state wildlife officers, technicians and biologists who have collected and catalogued the samples used for this study.

References

Barichello NJ, Carey J, Hoefs M (1989) *Mountain Sheep Status and Harvest in the Yukon: a Summary of Distribution, Abundance and the*

Registered Harvest, by Game Management Zone. Yukon Department of Renewable Resources, Whitehorse, Yukon.

Belkhir (1996) GENETIX. Available at: http://www.University-montp2.fr/~genetix/genetix.htm.

Bishop MD, Kappes SM, Keele JW *et al.* (1994) A genetic-linkage map for cattle. *Genetics*, **136**, 619–639.

Boyce WM, Ramey RRII, Rodwell TC, Rubin ES, Singer RS (1999) Population subdivision among desert bighorn sheep (*Ovis canadensis*) ewes revealed by mitochondrial DNA analysis. *Molecular Ecology*, **8**, 99–106.

Broders HG, Mahoney SP, Montevecchi WA, Davidson WS (1999) Population genetic structure and the effect of founder events on the genetic variability of moose, *Alces alces*, in Canada. *Molecular Ecology*, **8**, 1309–1315.

Buchanan FC, Crawford A (1992) Ovine dinucleotide repeat polymorphism at the MAF209 locus. *Animal Genetics*, **23**, 183.

Buchanan FC, Crawford A (1993) Ovine microsatellites at the OarFCB11, OarFCB128, OarFCB193, OarFCB266 and OarFCB304 loci. *Animal Genetics*, **24**, 145.

Carmichael LE, Nagy JA, Larter NC, Strobeck C (2001) Prey specialization may influence patterns of gene flow in wolves of the Canadian Northwest. *Molecular Ecology*, **10**, 2787–2798.

Casgrain P, Legendre P (2001) *The R Package for Multivariate and Spatial Analysis*, Version 4.0. Users' Manual. Department des Sciences biologiques, Université de Montreal. Available at: http://www.fas.umontreal.ca/BIOL/legendre.

Catto N, Liverman DGE, Bobrowsky PT, Rutter N (1996) Laurentide, Cordilleran, and Montane glaciation in the western Peace River Grande Prairie Region, Alberta and British Columbia, Canada. *Quaternary International*, **32**, 21–32.

Coltman DW, Festa-Bianchet M, Jorgenson JT, Strobeck C (2002) Age-dependent sexual selection in bighorn rams. *Proceedings of the Royal Society of London Series B: Biology Sciences*, **269**, 165–172.

Coltman DW, O'Donoghue P, Jorgenson JT *et al.* (2003a) Undesirable evolutionary consequences of trophy hunting. *Nature*, **426**, 655–658.

Coltman DW, Pilkington JG, Pemberton JM (2003b) Fine-scale genetic structure in a free-living ungulate population. *Molecular Ecology*, **12**, 733–742.

Dyke AS, Andrews JT, Clark PU *et al.* (2002) The Laurentide and Innuitian ice sheets during the last glacial maximum. *Quaternary Science Reviews*, **21**, 9–31.

Ede AJ, Pierson CA, Crawford A (1995) Ovine microsatellites at the OarCP9, OarCP16, OarCP20, OarCP21, OarCP23 and OarCP26 loci. *Animal Genetics*, **26**, 129–130.

Felsenstein J (1993) *Phylogeny Inference Package (PHYLIP)*, Version 3.5c. University of Washington, Seattle.

Festa-Bianchet M (1991) The social system of bighorn sheep — grouping patterns, kinship and female dominance rank. *Animal Behaviour*, **42**, 71–82.

Fitzsimmons NN, Buskirk SW, Smith MH (1997) Genetic changes in reintroduced Rocky Mountain bighorn sheep populations. *Journal of Wildlife Management*, **61**, 863–872.

Forbes SH, Hogg JT (1999) Assessing population structure at high levels of differentiation: microsatellite comparisons of bighorn sheep and large carnivores. *Animal Conservation*, **2**, 223–233.

Forbes SH, Hogg JT, Buchanan FC, Crawford AM, Allendorf FW (1995) Microsatellite evolution in congeneric mammals — domestic and bighorn sheep. *Molecular Biology and Evolution*, **12**, 1106–1113.

Geist V (1971) *Mountain Sheep: a Study in Behavior and Evolution*. University of Chicago Press, Chicago.

Georges M, Massey J (1992) Polymorphic DNA markers in Bovidae. In: *WO Publication No. 92/13120*. World Intellectual Property Organisation, Geneva.

Goldstein DB, Linares AR, Cavallisforza LL, Feldman MW (1995) An evaluation of genetic distances for use with microsatellite loci. *Genetics*, **139**, 463–471.

de Gortari MJ, Freking BA, Kappes SM *et al.* (1997) Extensive genomic conservation of cattle microsatellite heterozygosity in sheep. *Animal Genetics*, **28**, 274–290.

Greenwood PJ (1980) Mating systems, philopatry and dispersal in birds and mammals. *Animal Behaviour*, **28**, 1140–1162.

Guo SW, Thompson EA (1992) Performing the exact test of Hardy–Weinberg proportion for multiple alleles. *Biometrics*, **48**, 361–372.

Gutierrez-Espeleta GA (1999) *Neutral and adaptive genetic variation in Desert bighorn sheep* (Ovis canadensis): *implications for conservation*. PhD Thesis, Arizona State University.

Gutierrez-Espeleta GA, Kalinowski ST, Boyce WM, Hedrick PW (2000) Genetic variation and population structure in desert bighorn sheep: implications for conservation. *Conservation Genetics*, **1**, 3–15.

Hardy OJ, Vekemans X (2002) SPAGEDI: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes*, **2**, 618–620.

Hogg JT (2000) Mating systems and conservation at large spatial scales. In: *Vertebrate Mating Systems* (eds Apollonio M, Festa-Bianchet M, Mainardi D), p. 332. World Scientific Publishing, Singapore.

Hutchison DW, Templeton AR (1999) Correlation of pairwise genetic and geographic distance measures: inferring the relative influences of gene flow and drift on the distribution of genetic variability. *Evolution*, **53**, 1898–1914.

Kyle CJ, Strobeck C (2002) Connectivity of peripheral and core populations of North American wolverines. *Journal of Mammalogy*, **83**, 1141–1150.

Luikart G, Allendorf FW (1996) Mitochondrial DNA variation and genetic population structure in Rocky Mountain bighorn sheep (*Ovis canadensis canadensis*). *Journal of Mammology*, **77**, 109–123.

Maddox JF, Davies KP, Crawford AM *et al.* (2001) An enhanced linkage map of the sheep genome comprising more than 1000 loci. *Genome Research*, **11**, 1275–1289.

Maudet C, Miller C, Bassano B *et al.* (2002) Microsatellite DNA and recent statistical methods in wildlife conservation management: applications in Alpine ibex *Capra ibex* (ibex). *Molecular Ecology*, **11**, 421–436.

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, **89**, 538–590.

Paetkau D, Amstrup SC, Born EW *et al.* (1999) Genetic structure of the world's polar bear populations. *Molecular Ecology*, **8**, 1571–1584.

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

Page RDM (1996) TREEVIEW: an application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences*, **12**, 357–358.

Penty JM, Henry HM, Ede AJ, Crawford A (1993) Ovine microsatellites at the OarAE16, OarAE54, OarAE57, OarAE119 and OarAE129 loci. *Animal Genetics*, **24**, 219.

Perez T, Albornoz J, Dominguez A (2002) Phylogeography of chamois (*Rupicapra* spp.) inferred from microsatellites. *Molecular Phylogenetics and Evolution*, **25**, 524–534.

Polziehn RO, Hamr J, Mallory FF, Strobeck C (2000) Microsatellite analysis of North American wapiti (*Cervus elaphus*) populations. *Molecular Ecology*, **9**, 1561–1576.

Primmer CR, Koskinen MT, Piironen J (2000) The one that did not get away: individual assignment using microsatellite data detects a case of fishing competition fraud. *Proceedings of the Royal Society of London Series B: Biology Sciences*, **267**, 1699–1704.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Pritchard JK, Wen W (2002) *Documentation for Structure Software*, Version 2.0. University of Chicago, Chicago.

Ramey RR (1993) *Evolutionary genetics and systematics of North American mountain sheep: implications for conservation*. PhD Thesis, Cornell University.

Ramey RR (1995) Mitochondrial DNA variation, population structure, and evolution of mountain sheep in the southwestern United States and Mexico. *Molecular Ecology*, **4**, 429–439.

Raymond M, Rousset F (1995) GENEPOP, version 1.2: population genetics software for exact test and ecumenicism. *Journal of Heredity*, **86**, 248–249.

Rice WR (1989) Analysing tables of statistical tests. *Evolution*, **43**, 223–225.

Sage RD, Wolff JO (1986) Pleistocene glaciations, fluctuating ranges, and low genetic variability in a large mammal (*Ovis dalli*). *Evolution*, **40**, 1092–1095.

Sambrook J, Fritsch E, Maniatis T (1989) *Molecular Cloning: a Laboratory Manual*, 2nd edn. Cold Spring Harbor Laboratory Press, New York.

Simmons NM (1982) Seasonal ranges of Dall's sheep, Mackenzie Mountains, Northwest Territories. *Arctic*, **35**, 512–518.

Slate J, Coltman DW, Goodman SJ *et al.* (1998) Bovine microsatellite loci are highly conserved in red deer (*Cervus elaphus*), sika deer (*Cervus nippon*) and Soay sheep (*Ovis aries*). *Animal Genetics*, **29**, 307–315.

Slatkin M (1987) Gene flow and the geographic structure of natural populations. *Science*, **236**, 787–792.

Slatkin M (1993) Isolation by distance in equilibrium and nonequilibrium populations. *Evolution*, **47**, 264–279.

Swarbrick PA, Buchanan FC, Crawford A (1991) Ovine dinucleotide repeat polymorphism at the MAF36 locus. *Animal Genetics*, **22**, 377–378.

Valdez R, Krausman PR (1999) *Mountain Sheep of North America*, p. 353. University of Arizona Press, Tucson.

Veitch A, Simmons E (1999) *Mackenzie Mountain Non-Resident and Non-Resident Alien Hunter Harvest Summary 1999*. Department of Resources, Wildlife, and Economic Development, Norman Wells, NWT.

Waser PM, Strobeck C (1998) Genetic signatures of interpopulation dispersal. *Trends in Ecology and Evolution*, **13**, 43–44.

Weir BS, Cockerham CC (1984) Estimating *F*-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.

Wilson GA, Strobeck C (1999) Genetic variation within and relatedness among wood and plains bison populations. *Genome*, **42**, 483–496.

Wilson GA, Strobeck C, Wu L, Coffin JW (1997) Characterization of microsatellite loci in caribou *Rangifer tarandus*, and their use in other artiodactyls. *Molecular Ecology*, **6**, 697–699.

Wright S (1978) *Evolution and the Genetics of Populations*, Vol. IV. *Variability Within and Among Natural Populations*. University of Chicago Press, Chicago.

This study forms part of Kirsty Worley's PhD thesis, supervised by Dave Coltman, on evolutionary genetics of thinhorn sheep. Dave Coltman is now an associate professor at the University of Alberta who is interested in the evolutionary genetics of mountain ungulates and other handsome critters. This project began in Curtis Strobeck's wildlife genetics laboratory. Jean Carey is the Yukon's sheep and goat management biologist. Helen Schwantje is the Wildlife Veterinarian for British Columbia's Ministry of Water, Land and Air Protection and has a special interest in wild sheep and their health. Alasdair Veitch is a senior wildlife biologist with the Government of the Northwest Territories and is responsible for research and management on a wide range of species — primarily ungulates, carnivores and birds. Steve Arthur is an ungulate specialist with the Alaskan Department of Fish and Game.

---

doi:10.1111/j.1420-9101.2005.01027.x

Evidence for cryptic glacial refugia from North American mountain sheep mitochondrial DNA

J. LOEHR,* K. WORLEY,† A. GRAPPUTO,* J. CAREY,‡ A. VEITCH§ & D. W. COLTMAN¶

*University of Jyväskylä, Faculty of Biological and Environmental Sciences, Jyväskylä, Finland
†Department of Animal and Plant Sciences, University of Sheffield, Western Bank, Sheffield, UK
‡Yukon Department of Environment, Whitehorse, YT, Canada
§Northwest Territories Department of Environment and Natural Resources, Sahtu Region, Norman Wells, NT, Canada
¶Department of Biological Sciences, University of Alberta, Edmonton, AB, Canada

Keywords:

control region;
glaciation;
hybridization;
ice-free corridor;
mitochondrial DNA;
Ovis;
refugia.

Abstract

The separation of populations by ice sheets into large refugia can account for much of the genetic diversity found in present day populations. The evolutionary implications of small glacial refugia have not been as thoroughly explored. To examine refugial origins of North American mountain sheep *Ovis* spp., we analyzed a 604 bp portion of the mitochondrial DNA (mtDNA) control region from 223 *O. dalli* and *O. canadensis*. Major refugia were identified in eastern Beringia and southern North America, and we found evidence for two smaller refugia situated between the Laurentide and Cordilleran glaciers. Our results are the first to demonstrate support for survival of any organism in the latter two refugia. These refugia also appear to have conserved a genetic signal that confirms past hybridization of *O. dalli* and *O. canadensis*.

Introduction

Ice sheets can serve to fragment and isolate populations in differing refugia (Hewitt, 1996,2000), and this process of vicariance has been attributed as one of the major catalysts for the development of biodiversity in present temperate and arctic environments (Pielou, 1991; Willis & Whittaker, 2000). While isolation in major glacial refugia has been found to account for much genetic and morphological diversity (e.g. Ehrich *et al.*, 2000; Holder *et al.*, 2000; Brunhoff *et al.*, 2003; Flagsted & Røed, 2003; Dobeš *et al.*, 2004; Galbreath & Cook, 2004), quite understandably very little evidence is available to assess the evolutionary importance of smaller refugia. Smaller refugia are often difficult to identify due to a lack of fossil evidence or sufficient detail from studies of glacial limits.

Although examples are rare, genetic evidence has helped verify these 'cryptic' refugial sites and reconstruct their paleoecology. Plants of the *Packera* genus probably survived in southwestern Alberta, Canada in small ice-free areas between the Laurentide and Cordilleran ice sheets or in nunataks (small ice-free 'islands' extending above the ice sheets) in the same area (Golden & Bain, 2000). Mitochondrial DNA (mtDNA) evidence supports the survival of the endemic Norwegian lemming *Lemmus lemmus* in Scandinavian refugia (Fedorov & Stenseth, 2001). There is strong evidence for the existence of ice-free areas in the northwest of the Canadian arctic archipelago; there is congruence among molecular evidence for refugial populations of rock ptarmigan *Lagopus mutus* (Holder *et al.*, 1999) and the arctic plants *Dryas integrifolia* (Tremblay & Schoen, 1999) and *Saxifraga oppositifolia* (Abbott *et al.*, 2000). The collared lemming *Dicrostonyx groenlandicus* also survived in this region (Fedorov & Stenseth, 2002). This finding has been independently verified by a lemming parasite *Paranoplocephla arctica* (Wickström *et al.*, 2003).

Small populations may have their genetic signal of refugial origin erased if post-glacial migration from other refugial populations is great enough. This may explain the lack of signal from *Rangifer tarandus*, which was hypothesized to have survived glaciation in the Canadian arctic archipelago (Flagsted & Røed, 2003), and in *S. oppositifolia*, which may have survived in northern Norway (Gabrielsen *et al.*, 1997). Thus the signal of past refugial separation may be stronger in

Correspondence: John Loehr, University of Jyväskylä, Faculty of Biological and Environmental Sciences, P.O. Box 35, 40014, Jyväskylä, Finland.
Tel: 358 14 260 1211; fax: 358 14 260 2321;
e-mail: johloeh@cc.jyu.fi

species with limited dispersal abilities (e.g. Fedorov & Stenseth, 2002).

Here we investigate the refugial origins of two species of mountain sheep (*Ovis* spp.) in North America. Mountain sheep are highly philopatric and are ideal study subjects to probe for the existence of cryptic montane refugia because they are superbly adapted to glacial conditions and currently inhabit areas within sight of large ice sheets (Geist, 1971). The treeless landscape of montane glacial environments affords sheep ample foraging opportunity on grasses and sedges and an abundance of suitable escape terrain for predator avoidance.

The two mountain sheep species are recognized based on morphological evidence (Cowan, 1940; Ramey, 1993) and their divergence is thought to be a result of long-term separation by glacial activity. *O. dalli* (thinhorn sheep) occupy mountain ranges in northwestern North America, and *O. canadensis* (bighorn sheep) inhabit areas in southwestern Canada and the United States (Fig. 1). Cowan (1940), Pielou (1991) and Geist (1999) hypothesized that the two North American mountain sheep species were isolated from one another by the ice sheets that separated Beringia and southern North America (Fig. 2). Two subspecies of thinhorn sheep *O. d. dalli* and *O. d. stonei* are also of particular interest in this study. These subspecies are based on regional coat colour differences, however there is complete intergradation between light and dark morphs.

Sheep currently inhabit four regions known to have been ice-free during the Wisconsinan glaciations (approximately 70 000–10 000 years ago). The two major refugia of Beringia and southern North America are well documented. However, portions of the Mackenzie Mountains of Canada's Northwest Territories (also known as 'Easternmost Beringia'; Dyke & Prest, 1987; Duk-Rodkin & Hughes, 1991), and a region further south in northeastern British Columbia (Catto *et al.*, 1996) also remained ice-free. The glacial history of these regions is complex. Easternmost Beringia was influenced by Laurentide and montane glaciation, and the ice-free area appears to have been effectively cut off from other parts of Beringia during much of the last ice age (Catto, 1996). Northeastern British Columbia was a crossroads for Laurentide and Cordilleran ice sheets and montane glaciers. From investigations near Fort St. John, British Columbia (Fig. 2) Catto *et al.* (1996) found that coalescence of ice sheets did not occur during the last ice age. Since the ice sheets did not advance at the same time in this region a temporally and geographically shifting ice-free zone existed. Numerous isolated foothills of the Rocky Mountains also remained ice-free during glacial maximum.

Both of these areas are within the disputed Ice-Free Corridor, a proposed migration route between Beringia and southern North America (e.g. see Pielou, 1991; Mandryk *et al.*, 2001 for a discussion of this region). Most

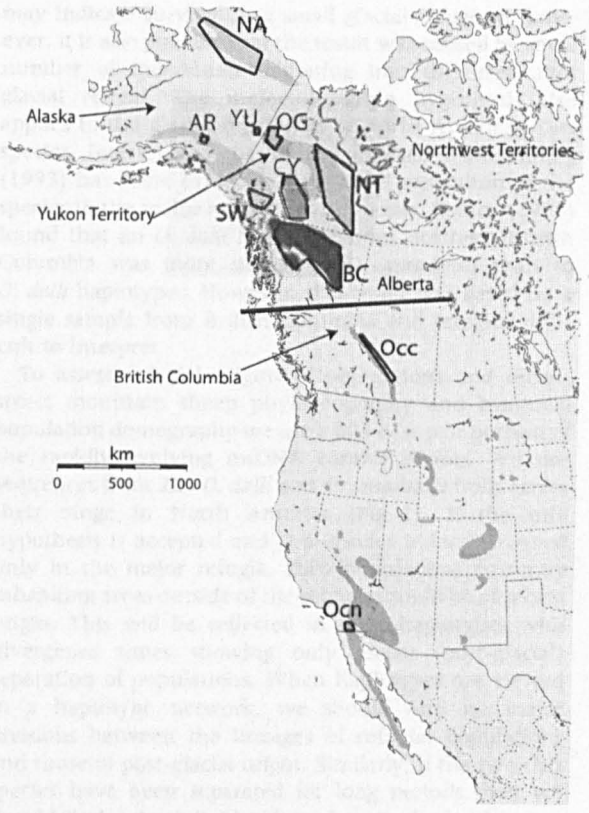


Fig. 1 Distribution of mountain sheep (*Ovis* spp.) in North America and regions sampled for mtDNA and relevant territory, province and state names. Shaded areas represent approximate boundaries of sheep habitation. The solid line divides putative species: *O. dalli* to the north and *O. canadensis* to the south. For *O. dalli* darkness of shaded areas represent relative darkness of coat pelage. All *O. dalli* populations except OG, CY and BC are almost exclusively comprised of white coloured sheep (see Materials and methods for site names and further description of colour variation and subspecies of *O. dalli*).

now agree that the Ice-Free Corridor was not open throughout the last ice age since coalescence of ice sheets has been verified at its southern end in central Alberta (Levson & Rutter, 1996) but little is known about the importance of areas in this region as glacial refugia. Other than geologic evidence to support ice-free conditions, we are not aware of any studies that confirm habitation by any organism of the two postulated refugia throughout glaciation. Palynological data, however, confirms a herb and shrub dominated vegetation (containing plants suitable for consumption by sheep) in the Mackenzie Mountains (Szeicz & MacDonald, 2001) during the Lateglacial (ca. 12 000–10 000 C year BP) and Catto *et al.* (1996) postulated that the region in northeastern British Columbia could have supported plant and animal life.

Based on evidence from glacial limits it appears that sheep could have survived glaciation in smaller glacial

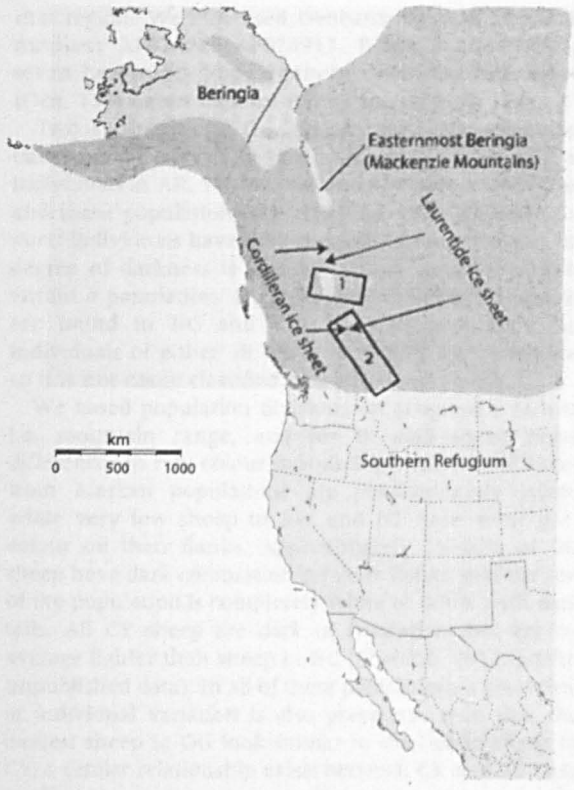


Fig. 2 Glacial coverage of North America during glacial maximum (approximately 21 000 years ago). Glacial limit approximations are based on Duk-Rodkin (1999), Mandryk *et al.* (2001) and Catto *et al.* (1996). North America was glaciated by both the Cordilleran and Laurentide ice sheets. Arrows show general direction of ice sheet advance. Two major refugia existed in Beringia and areas of southern North America. A small ice-free area in the Mackenzie Mountains has also been verified. Box 1 represents the study area of Catto *et al.* (1996), who found that the independently advancing and retreating ice sheets did not coalesce in this region. This region should be viewed as a temporally and geographically fluctuating ice free zone. For example between 14 000 and 10 000 years ago the Cordilleran glacier advanced further east, but by this time the Laurentide glacier had retreated. Little or no evidence is available from the region between Easternmost Beringia and box 1. We have represented this area as covered in ice; however, the exact history of this region is uncertain. The glacial history of the area in box 2 differs from box 1 since coalescence did occur between ice sheets at glacial maximum (e.g. Levson & Rutter, 1996), presumably making survival of sheep in this region impossible.

refugia in the Northwest Territories and British Columbia (Fig. 2). In the context of this paper, we wish to test the null hypothesis that sheep populations survived only in the major refugia as proposed by Cowan (1940), Pielou (1991) and Geist (1999). Recent analysis of nuclear DNA found that populations of *O. dalli* in British Columbia have lower genetic diversity than populations that inhabited major glacial refugia (Worley *et al.*, 2004). This may indicate survival in a small glacial refugium; however, it is also possible that the result was caused by small number of individuals migrating into this area after glacial retreat. The major refugium hypothesis also applies to the glacial separation of North American *Ovis* species. Interestingly, in his unpublished work Ramey (1993) has some evidence that glacial separation of the species in the major refugia did not occur. Ramey (1993) found that an *O. dalli* haplotype from northern British Columbia was more similar to *O. canadensis* than to *O. dalli* haplotypes. However, this result was based on a single sample from British Columbia and remains difficult to interpret.

To assess refugial origins of populations and reconstruct mountain sheep phylogeography and historical population demography we use a 604 base pair portion of the rapidly evolving mtDNA control region. We use sequences from 223 *O. dalli* and *O. canadensis* from across their range in North America (Fig. 1). If the null hypothesis is accepted and *Ovis* species indeed survived only in the major refugia, then populations currently inhabiting areas outside of the refugia should be of recent origin. This will be reflected in their haplotypes with divergence times showing only recent (post-glacial) separation of populations. When haplotypes are viewed in a haplotype network, we should find no major divisions between the lineages of refugial populations and those of post-glacial origin. Similarly, if the two *Ovis* species have been separated for long periods then we should find a deep division between species haplotypes. Here we also wish to test the validity of *O. dalli* subspecies. Since these subspecies intergrade across their range monophyly is not expected. However, if subspecies groups are supported we should find that analysis of mtDNA molecular variance shows a significant difference between subspecies. Finally, we wish to test to what degree glaciation has affected mountain sheep populations. Since much of current sheep range was covered by glaciers during the ice age, we expect to find evidence for post-glacial rapid expansion in areas covered by ice.

Materials and methods

We sampled sheep from eight areas across the range of *O. dalli* (Fig. 1). Three locations were sampled in Alaska (in brackets: abbreviation for population and number of samples from location): Alaska Range (AR, 3), Yukon Charley Rivers National Preserve (YU, 4) and Northern Alaska (NA, 4). Five geographic regions were sampled in Canada: Ogilvie Mountains (OG, 9); Mackenzie Mountains, Northwest Territories (NT, 16); Pelly Mountains, Central Yukon (CY, 18); Southwest Yukon (SW, 12); and British Columbia (BC, 22).

We sampled Rocky Mountain bighorn *O. c. canadensis* from the Canadian Rocky Mountains (Occ, 11) and incorporated one sequence from Genbank (accession number AY091486, Hiendleder *et al.*, 2002) also from

that region. We also used Genbank samples (accession numbers AF076911–AF076917, Boyce *et al.*, 1999) of seven haplotypes from southern California *O. c. nelsoni* (Ocn, 124) desert bighorn in our analysis.

Two subspecies of *O. dalli* are currently recognized based on differences in colour morphology. Nearly all individuals in AR, YU, NA, NT and SW have a white coat and these populations are classified as *O. d. dalli*. *O. d. stonei* individuals have dark coat coloration, however the degree of darkness is variable and is never consistent within a population. Sheep representing this subspecies are found in BC and CY. The OG population has individuals of either all white or darker coloration and so it is not easily classified into either subspecies.

We based population divisions on geographic factors, i.e. mountain range, and for *O. dalli* sheep major differences in ram colour morphology. All *O. dalli* sheep from Alaskan populations are predominantly white, while very few sheep in SW and NT have some grey colour on their flanks. Approximately 35–40% of OG sheep have dark colouration on their flanks, and the rest of the population is completely white or white with dark tails. All CY sheep are dark in coloration, but are on average lighter than sheep in BC (Sheldon, 1911; Loehr, unpublished data). In all of these populations a great deal of individual variation is also prevalent, such that the darkest sheep in OG look similar to the lighter sheep in CY; a similar relationship exists between CY and BC (also see Fig. 1).

Most samples of *O. dalli* were taken from horn core samples of hunter-killed rams in Yukon, British Columbia and Northwest Territories from 1994 to 2000. Alaskan samples comprised whole blood samples of both rams and ewes collected between 1999 and 2002. Rocky mountain *O. canadensis* samples also comprised horn core samples. Horn corings were stored dry in sealed paper envelopes, while blood samples were stored at −20 °C in EDTA. Genomic DNA was extracted from approximately 0.5 mL of horn material per sample using a tissue extraction kit (Qiagen, Crawley, West Sussex, UK). Blood samples were extracted using a phenol chloroform technique (Sambrook *et al.*, 1989) from 200 μL of blood.

The human primers L15999 (5′-ACC ATC AAC ACC CAA AGC TGA-3′) and H16498 (5′-CCT GAA GTA GGA ACC AGA TG-3′) were used to amplify the control region of mtDNA in the 5′ to 3′ direction. Polymerase chain reactions (PCR) were carried out in 50 μL reactions containing 5 μL DNA template, 0.4 μM of each primer, 100 μM dNTP's, 1.5 mM $MgCl_2$ and 1 unit *Taq* polymerase (Bioline, London). PCR profiles comprised 32 cycles of 94 °C denaturing for 30 s, 50 °C annealing for 30 s, and extension at 72 °C for 40 s. Cycles were preceded by 3 min at 94 °C and terminated with 5 min at 72 °C. PCR products were gel purified (Qiagen) and quantified before sequencing. Purified PCR products were sequenced using primers as above and run on an ABI 3730. Editing and alignment was carried out using the software SeqScape (Applied Biosystems, Foster City, USA).

A Siberian snow sheep *O. nivicola* sample was used as an outgroup for analysis. It was chosen because *O. nivicola* are the closest interspecific relative of *O. dalli* and *O. canadensis*. A minimum spanning network was constructed using the pair-wise number of differences among haplotypes as implemented in Arlequin ver. 2.001 (Schneider *et al.*, 2000).

To determine population structure based on species/subspecies groups we used Analysis of Molecular Variance (AMOVA) as implemented in Arlequin (Excoffier *et al.*, 1992). *O. canadensis* subspecies comparisons were not made since sampling from *O. canadensis* sheep was less extensive. A detailed mtDNA analysis of *O. canadensis* subspecies has been done by Ramey (1993).

Haplotype divergence times were calculated with Mdiv (Nielsen & Wakely, 2001). This program employs a Bayesian framework using the Hasegawa–Kishino–Yano nucleotide substitution model (HKY 1985) to correct for multiple hits, and produces a joint estimation of multiple demographic parameters (θ, nonequilibrium-based migration rate, time to most recent common ancestor (TMRCA) and time of divergence for any pair of populations). These two estimates differ because haplotypes can start to diverge before a single population diverges into two or more populations. Using both estimates it is possible to estimate the time that haplotypes began to diverge and when populations separated or experienced secondary contact.

To calculate the time since the divergence of haplotypes we used a molecular clock of 24% per locus per million years, which has been estimated for domestic and wild sheep from the complete length of the mtDNA control region (Hiendleder *et al.*, 2002). Since the control region evolves too rapidly for comparison based on fossil evidence of separate species, an indirect method was used. Hiendleder *et al.* (2002) first estimated the protein-coding sequence distance between the most divergent lineages of sheep (*O. canadensis* and *O. aries*). This was then referenced against molecular divergence of the complete mtDNA genome for cow (*Bos* spp.) and sheep, and then further referenced using fossil evidence of a divergence of the *Ruminantia* and *Cetacea* 60-million-year ago. To be conservative we also consider a much faster mutation rate estimate of 38% per locus per million years, which has been estimated using cattle mtDNA from a 240-bp portion of the same portion of the control region that we sequenced (Troy *et al.*, 2001).

We calculated indicators of within population genetic diversity (haplotype diversity and nucleotide diversity) using Arlequin. To probe for evidence of population expansion we calculated Fu's F_S test (Fu, 1997), which is considered a sensitive indicator of population expansion. It compares the number of haplotypes observed with the number of haplotypes expected in a random sample. This test assumes that recombination does not

occur and uses an infinite-sites model. Further analysis of population expansion was done using mismatch distribution analysis (MDA). MDA assesses the shape of the distribution of the number of observed differences between pairs of DNA sequences. A population that has experienced rapid growth exhibits a unimodal wave when results from MDA are graphed. MDA produces an age expansion parameter (τ), which is a relative measure of the time (in generations) since population expansion and is useful to date the initiation of rapid population growth (see Rogers & Harpending, 1992; Schneider & Excoffier, 1999). MDA also produces an estimate of effective population size at the start (θ_0) and at the end (θ_1) of expansion.

Results

There were 64 unique haplotypes defined by 90 polymorphic sites, including 84 transitions and 6 transversions in the 223 *O. dalli* and *O. canadensis* sequences we analyzed. The mean number of pair-wise differences between sequences was 17.31 (SD ± 7.72) and nucleotide diversity was 0.035 (SD ± 0.017). Population specific results are presented in Table 1.

We found extensive paraphyly between *O. dalli* and *O. canadensis*. The minimum spanning network (Fig. 3) showed that *O. dalli* haplotypes from BC and NT and *O. canadensis* haplotypes from southern Canada were more closely related to each other than to haplotypes from their putative species. A large percentage of *O. dalli* genetic variation was explained by dividing populations into their respective mountain ranges, showing that molecular variance was geographically structured (Table 2). Divisions based on species or *O. dalli* subspecies resulted in no significant difference (Table 3).

The shortest divergence times were estimated at 7000 years for OG and YU, and 11 000 years for CY and SW (Table 4). We found a divergence time of about 68 000 years for the BC and southern Canada *O. canadensis* haplotypes. The divergence time between these two populations is of particular interest. To be conservative we also calculated the divergence time using the much faster mutation rate calculated for cattle (Troy *et al.*, 2001). The faster mutation rate resulted in a divergence time of 43 305 years before present. Again using the mutation rate of 24% calculated for sheep we estimated the TMRCA. The TMRCA for BC/Occ was about 431 000 (Table 4). A similar TMRCA was found for most populations surveyed. However, we found a relatively short TMRCA for populations at the opposite ends of mountain sheep distribution. *O. dalli* sheep from northern Alaska and *O. c. nelsoni* sheep from the southern US shared a common ancestor around 250 000 years ago. In the haplotype network, *O. c. nelsoni* haplotypes had the least number of steps to Siberian snow sheep *O. nivicola*.

To test for evidence of population growth we used Fu's F_S test. This test is considered significant at $\alpha = 0.05$ when $P = 0.02$ (Fu, 1997). Population growth was found for NA and Occ and bordered on significance for NT (Table 5). For these populations MDA also indicated population growth since the null hypothesis of population expansion was not rejected. Results for Fu's F_S and P (SSD) were also consistent in the case of Ocn, since both tests rejected the possibility of growth. The OG population also appears to not have experienced growth since the P (SSD) result neared significance and θ_0 and θ_1 were also nearly identical (Table 5).

There were inconsistent test results for the CY, SW and AR populations. Fu's F_S test detected no growth for these populations, but rapid population expansion could not be ruled out by the MDA test. This was also the case for BC. However, the star shaped pattern of haplotypes from BC suggests a population bottleneck making the population expansion scenario more likely (Fig. 3). Twenty of the 22 BC samples were found in the star shaped haplogroup (see Fig. 3, haplotypes 1–10). The remaining two haplotypes (22 and 23) were similar to an NT haplogroup and are a probable result of contact between NT and BC about 48 000 years ago. We then tested the BC haplotypes 1–10 again for evidence of population expansion. In this case Fu's test was significant ($F_S = -4.77$, $P = 0.003$) demonstrating strong support for population expansion. Population growth was also supported by MDA analysis (P (SSD) = 0.31). The value of τ was estimated at 1.83 (95% CI = 0.42–2.79) which was similar to that estimated for all BC haplotypes.

Table 1 Genetic diversity for *O. dalli* and *O. canadensis* populations based on mtDNA control region. Number of samples (N), number of unique haplotypes (Hn), haplotype diversity (h), nucleotide diversity (π), number of segregating sites (S) (the * indicates that a transversion was observed in the population) and mean number of pair-wise differences (d) between haplotypes.

	OG	NT	CY	SW	AR	YU	NA	BC	Occ	Ocn
N	9	16	18	12	3	4	4	22	11	124
Hn	3	14	12	6	3	1	4	12	8	7
h (±95%CI)	0.722 (0.097)	0.975 (0.035)	0.948 (0.033)	0.894 (0.054)	1.000 (0.272)	0	1.000 (0.177)	0.896 (0.045)	0.946 (0.054)	0.773 (0.025)
π (±95%CI)	0.020 (0.011)	0.024 (0.013)	0.027 (0.014)	0.011 (0.006)	0.017 (0.014)	0	0.004 (0.003)	0.013 (0.007)	0.008 (0.005)	0.013 (0.007)
S	22*	47*	35*	14	11*	0	3	39	11*	20*
d (SD)	10.89 (5.47)	11.24 (5.39)	12.69 (6.00)	5.14 (2.68)	7.33 (4.73)	0	1.67 (1.22)	6.97 (3.40)	3.02 (1.70)	6.56 (3.12)

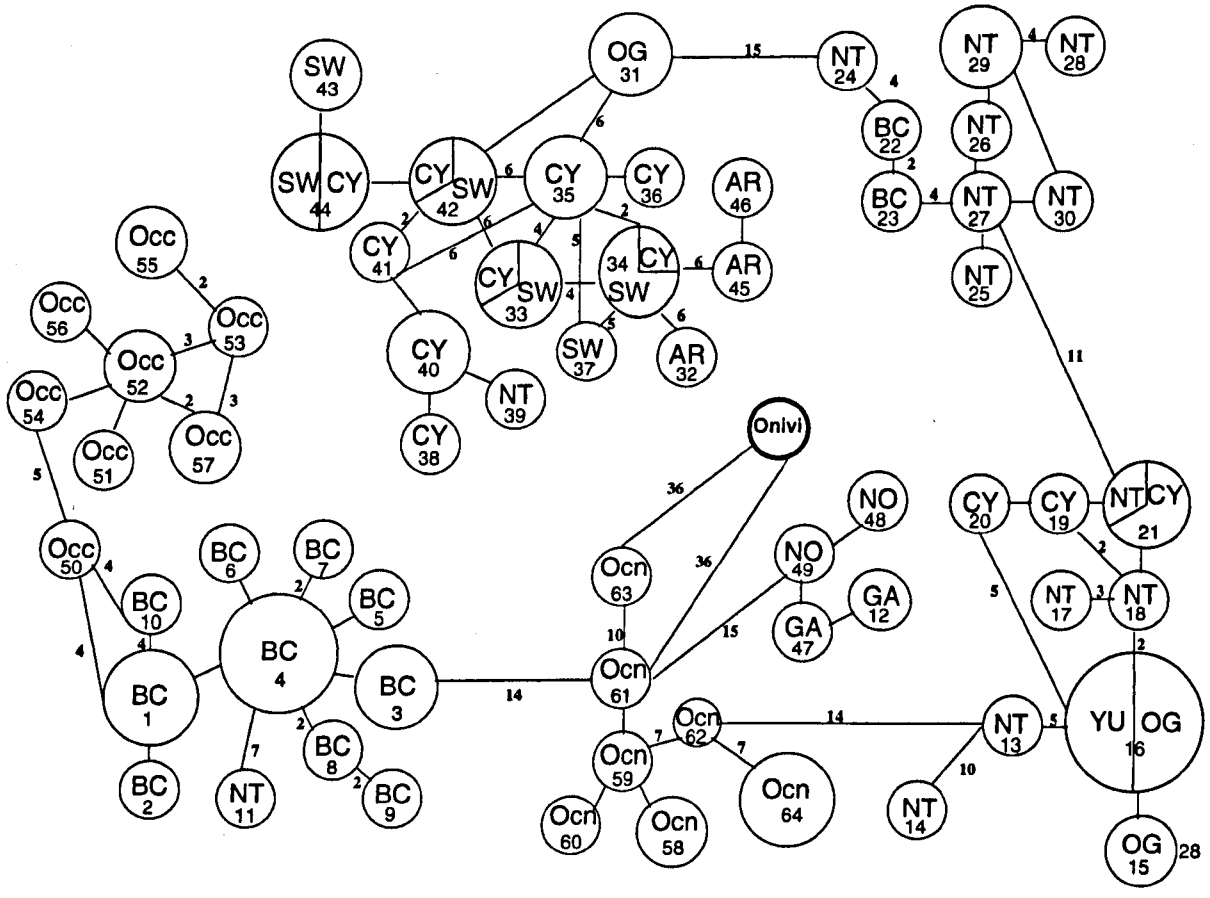


Fig. 3 Minimum spanning network of haplotypes, rooted with *Ovis nivicola*. Occ and Ocn populations are *O. canadensis*, all others are *O. dalli* (see methods and Fig. 1). *O. d. stonei* populations are CY and BC. Intergradation between subspecies occurs in OG. All remaining *O. dalli* populations represent the *O. d. dalli* subspecies. Circle size is proportional to the number of times the haplotype was observed. An exception is the Ocn haplotype group. Since the number of samples in this population was very high, haplotypes are represented in proportion to number of samples from the same population. Numbers inside circles refer to haplotype identification number. Numbers next to lines refer to number of steps between haplotypes. If no number is present then one step separates haplotypes.

Table 2 Analysis of molecular variance (AMOVA) of *O. dalli* and *O. canadensis* populations based on mountain range and species.

	Molecular variance of *O. dalli* and *O. canadensis*			
Criteria for division	Components	% of total variation	Fixation indices	*P*
Mountain range	Among populations (i.e. mountain ranges)	67.5	Φ_{st} 0.67	<0.001
	Within populations	32.5		
Species	Among species groups	9.4	Φ_{ct} 0.09	0.18
	Among populations within species groups	58.9	Φ_{sc} 0.65	<0.001
	Within populations	31.7	Φ_{st} 0.63	<0.001

Discussion

Multiple refugia and gene flow between species

Our results support the possibility of four glacial refugia. As expected, two major refugia were verified: one northwest of the major ice sheets in Beringia, and a second in southern North America. Support for long-term isolation in these large refugia is found in population divergence times. Divergence times of 167–307 000 years for *O. c. nelsoni* from southern North America and *O. dalli* populations from Beringia suggest

Table 3 AMOVA of *O. dalli* based on mountain range and subspecies.

	Molecular variance of *O. dalli* only			
Criteria for division	Components	% of total variation	Fixation indices	*P*
Mountain range	Among populations (i.e. mountain ranges)	57.8	Φ_{st} 0.58	<0.001
	Within populations	42.1		
Subspecies	Among subspecies groups	−11.1	Φ_{ct} −0.11	1.0
	Among populations within subspecies groups	69.8	Φ_{sc} 0.63	<0.001
	Within populations	41.3	Φ_{st} 0.53	<0.001

that contact has not occurred since the mid Pleistocene. During the last ice age the mass of ice separating these two refugia were thought to have completely divided *O. dalli* from *O. canadensis* (Pielou, 1991; Geist, 1999). Our evidence does not support this hypothesis, instead the existence of two additional refugia in the Mackenzie Mountains and northeast British Columbia is probable. This region appears to have allowed historic gene flow between the two species, and harboured populations in refugia during the last glacial maximum.

There is evidence for contact between NT, CY, OG and BC that predates glacial maximum, suggesting a refugium in the Mackenzie Mountains existed (Table 4). A divergent haplogroup is also dominated by NT haplotypes (see Fig. 3 haplotypes 22–30). Two BC haplotypes exist in this haplogroup, however based on divergence estimates this is a result of contact between these populations prior to glacial maximum. Since divergence estimates can only be approximate there is still a possibility that gene flow also occurred after glacial maximum. Although this refugium may not have been completely separated from Beringia, it appears evident that within this region there has been considerable differentiation of haplotypes, supporting a hypothesis of continual long-term habitation of the Mackenzie Mountains.

The BC *O. dalli* population also had divergence times with adjacent populations that predated glacial maximum. Interestingly, this population appears to be a product of hybridization between species since haplotypes similar to both *O. dalli* and *O. canadensis* were present. However, most haplotypes in this *O. dalli* population were more similar to *O. canadensis* haplotypes than to other *O. dalli*. This finding contradicts Geist's (1971) hypothesis of *O. dalli* range expansion from a Beringian refugium into British Columbia. A second explanation may be that the BC population was colonized by *O. canadensis* haplotypes after ice sheets retreated. Evidence against this comes from an estimated divergence time between BC *O. dalli* and *O. c. canadensis* that predates glacial maximum by about 50 000 years.

Colour morphology also suggests that a post-glacial colonization of BC by *O. canadensis* is unlikely. *O. dalli* in BC share similar coat pelage with sheep further north in CY and both of these populations differ greatly from the coat colour of *O. canadensis*. If BC were initially colonized by *O. canadensis* and sheep from CY then migrated south and interbred with *O. canadensis* (resulting in a change in morphology to resemble CY *O. dalli*) then CY haplotypes would have been found in BC. However, post-Wisconsinan glaciation gene flow between these populations is not supported: no CY haplotypes were found in BC. Nuclear DNA microsatellite evidence also showed a high degree of population structure between BC and CY suggestive of limited gene flow (Worley *et al.*, 2004). However, it is still possible that a small number of migrants introduced the present coat colour to the BC population, and a strong selective advantage lead to it spreading throughout the population. Unfortunately, horn morphology and body size do not give a clear indication of refugial origin since BC sheep are intermediate in size between *O. dalli* and Occ sheep (Geist, 1971; Ramey, 1993).

We propose a third explanation, which is most congruent with the evidence available. Although parts of the range occupied by BC *O. dalli* remained ice-free (Catto *et al.*, 1996), no large ice-free areas have been identified in the southern Canadian Rocky Mountains currently inhabited by *O. canadensis* (e.g. Levson & Rutter, 1996). The mtDNA from *O. canadensis* supports a scenario in which the southern Canadian Rockies were recolonized after ice retreated (Luikart & Allendorf, 1996). Thus we propose that contact occurred between *O. canadensis* and *O. dalli* prior to glacial maximum. As ice advanced the BC *O. dalli* population survived glaciation in small ice-free areas described by Catto *et al.* (1996), while *O. canadensis* populations in the Canadian Rockies were overcome by ice and went extinct. The star shaped pattern in the haplotype network for BC (Fig. 3) indicates a possible population bottleneck (Slatkin & Hudson, 1991) consistent with this scenario. Rapid expansion for BC was estimated at 7300 years ago. This relatively recent onset of population growth agrees well with melting of ice in this region, which has been dated to about 10 000 years ago (Catto *et al.*, 1996). Indicators of nuclear genetic diversity were relatively low for BC populations, also lending some support to a bottleneck hypothesis (Worley *et al.*, 2004). Results should be interpreted with some caution since extinction events,

Table 4 Below diagonal: Mdiv estimates of divergence times for population pairs. Numbers in parentheses are 95% credibility intervals of the estimate. Divergence times before present are presented below credibility intervals. Above diagonal: Mdiv estimate and number of years to most recent common ancestor. The times in years (in bold) were obtained based on a mutation rate for the control region (24% per million years) and an average generation time of 3 years.

	OG	NT	CY	SW	BC	YU	AR	NA	Occ	Ocn
OG	–	2.34 **297 694**	4.40 **326 070**	4.93 **296 750**	3.97 **402 040**	8.83 **253 029**	4.54 **283 750**	5.45 **319 336**	5.1 **348 633**	3.25 **341 504**
NT	0.30 (0.17–1.94) **38 644**	–	2.74 **309 580**	2.40 **301 687**	2.64 **399 011**	2.23 **311 152**	1.96 **309 823**	1.97 **348 341**	2.54 **355 865**	1.55 **341 283**
CY	0.95 (0.45–4.87) **70 274**	1.39 (0.43–4.73) **156 803**	–	5.77 **355 468**	3.75 **445 649**	4.57 **308 237**	3.37 **333 638**	3.81 **402 332**	3.97 **411 991**	3.05 **427 358**
SW	1.72 (0.69–4.88) **103 729**	1.32 (0.48–4.86) **165 998**	0.18 (0.00–1.86) **11 336**	–	4.39 **463 687**	5.46 **277 266**	2.09 **144 504**	6.43 **426 155**	5.74 **434 237**	3.90 **418 438**
BC	1.74 (0.56–4.85) **235 059**	0.32 (0.28–4.86) **48 434**	1.19 (0.58–4.81) **141 501**	2.10 (0.79–4.84) **221 891**	–	3.87 **349 711**	3.66 **438 914**	3.97 **407 855**	4.99 **457 417**	2.78 **378 268**
YU	0.23 (0.06–1.95) **6 531**	0.28 (0.19–4.86) **39 156**	3.86 (0.58–11.7) **260 349**	4.16 (0.84–9.74) **211 250**	1.06 (0.46–4.86) **95 786**	–	4.51 **237 245**	9.50 **239 974**	5.6 **272 708**	2.77 **282 771**
AR	0.97 (0.35–4.88) **60 625**	1.16 (0.28–4.87) **183 365**	0.81 (0.38–4.85) **77 836**	1.32 (0.33–4.85) **91 266**	1.87 (0.55–4.84) **224 254**	4.24 (0.76–9.74) **223 042**	–	4.88 **376 167**	4.90 **430 026**	2.72 **423 583**
NA	4.16 (0.76–9.74) **243 750**	1.21 (0.34–4.86) **213 956**	2.08 (0.52–4.86) **219 646**	5.72 (1.12–9.76) **379 099**	2.14 (0.64–9.70) **219 852**	9.59 (2.10–34.2) **242 247**	4.18 (0.64–9.74) **322 208**	–	4.86 **294 258**	2.27 **253 306**
Occ	4.08 (0.43–7.78) **278 906**	1.93 (0.22–2.93) **270 401**	3.05 (0.79–5.82) **316 517**	4.48 (0.62–7.81) **338 917**	0.75 (0.36–1.94) **68 567**	4.7 (0.86–7.82) **228 880**	3.87 (0.90–7.79) **339 633**	4.30 (0.35–5.87) **260 352**	–	1.86 **256 477**
Ocn.	2.12 (0.82–5.86) **222,766**	1.10 (0.62–2.60) **242,201**	1.57 (0.84–9.68) **219,964**	2.86 (1.10–9.70) **306,854**	1.34 (0.56–4.18) **182,331**	7.84 (4.1–19.24) **177,625**	1.94 (0.20–4.92) **302,115**	1.5 (0.54–9.72) **167,383**	1.03 (0.22–4.86) **142,027**	–

Table 5 Indicators of historical population demography. Fu's F_S is an indicator of population growth. Other demographic parameters (95% CI) were calculated using MDA: τ (time estimate for onset of expansion), θ_0, (population size estimate at onset of growth) and θ_1 (estimate after growth). P (SSD_{obs}), measures the probability of deviation from a hypothesis of rapid population growth according to the mismatch distribution (see Fig. 4). Years since expansion initiation are estimated from τ. No statistics were calculated for YU, since all haplotypes were the same in this population.

	OG	NT	CY	SW	BC	AR	NA	Occ	Ocn
Fu's F_S (P value)	7.75 (0.99)	−3.85 (0.03)	0.07 (0.48)	1.08 (0.72)	−0.82 (0.40)	0.81 (0.41)	−2.18 (0.01)	−2.97 (0.02)	11.42 (0.99)
τ	7.00 (2.53–11.25)	11.83 (4.91–26.83)	3.00 (0.98–25.58)	7.55 (3.82–12.38)	2.00 (0.00–5.07)	11.72 (6.49–30.72)	1.84 (0.00–3.80)	3.21 (1.27–4.24)	12.27 (5.02–23.26)
θ_0	2.10 (0.00–2.84)	4.86 (0.00–11.50)	12.32 (0.00–29.2)	0.00 (0.00–2.16)	1.46 (0.00–7.70)	0.00 (0.00–17.00)	0.00 (0.00–1.35)	0.00 (0.00–2.61)	0.00 (0.00–10.48)
θ_1	2.10 (0.22–59.13)	21.60 (12.41–191.19)	14.61 (3.31–310.61)	13.22 (9.04–3301.97)	38.71 (4.80–6773.3)	841.09 (24.06–7013.6)	3776.3 (3761.7–6946.3)	5010.60 (32.70–9225.63)	21.23 (16.49–156.40)
P (SSD_{obs})	0.07	0.59	0.19	0.41	0.11	0.29	0.44	0.16	<0.001
Years since expansion initiation	25 584	43 238	10 964	27 595	7309	42 825	6724	11 732	44 846

genetic drift, or a selective sweep may have altered the composition of haplotypes in populations since the end of the ice ages and could confound our results.

Historic population demography

During the last ice age, overall habitat available to sheep was limited by the extent of ice sheets. With the retreat of ice, new territory was made available, allowing some populations to expand. Rapid population growth was most obvious for the Occ, NA and NT populations (Table 5, Fig. 4). Contradictory results were initially found for BC, but strong evidence for population growth was found when the haplogroup distinct to BC was analyzed. All populations with signs of rapid expansion now occupy regions that had extensive glacial coverage (Figs 1 and 2). Initiation of population expansion was roughly dated to the period during and following glacial maximum (22 000–10 000 years ago). The exception is NT, which appears to have experienced expansion prior to glacial maximum (Table 5). This may hint that glacial maximum occurred there earlier than elsewhere. Temporal estimation of expansion events based on a single locus must, however, be interpreted with caution (Hillis *et al.*, 1996). The Ocn and OG populations showed no sign of population growth (Table 5, Fig. 4), which appears consistent with the reduced presence or complete absence of glaciers in the regions these populations now occupy (Fig. 2).

Historical demography for the CY, SW and AR populations remains less clear. Much of the current habitat occupied by these populations was covered by ice (Fig. 2). Therefore, it might be expected that population growth would be easily detected. However, we found contradictory results for population expansion (see Results and Table 5). The discrepancies found between F_S and P (SSD) may be a result of a genetic bottleneck occurring in a population after the initial expansion. Such a pattern has potential to erase the signal of earlier expansion events leading to a lower probability of significant F_S tests (Excoffier & Schneider, 1999). However, it is also possible that these populations have not expanded greatly in size. For example, during glacial maximum there could have been large sheep populations in areas of Beringia near glaciers, which then shifted into newly available territory as the glaciers retreated. At the same time their original territory may have become unsuitable due to growth of forests as the climate moderated.

Taxonomy

Paraphyly was found between *O. dalli* and *O. canadensis* and was also evident in *O. dalli* subspecies (Fig. 3). AMOVA at the species level produced nonsignificant results (Table 2), and for *O. dalli* the variance based on subspecies was negative, indicating that some individuals were more closely related to members of the other subspecies than they were to their own (Table 3). This confirms earlier unpublished work by Ramey (1993), also reporting paraphyly. Species classification is currently based on analysis of skull morphology (Cowan, 1940) but fewer significant differences between species were found upon reanalysis of the same data (Ramey, 1993). At the molecular level *O. dalli* and *O. canadensis* do not appear to fulfil species criteria since their mtDNA lineages are not reciprocally monophyletic (e.g. Moritz, 1994). However, basing species on mtDNA monophyly may be overly stringent (Crandall *et al.*, 2000). Instead species status may be better based on criteria of historic and current genetic and ecological exchangeability. Crandall *et al.* (2000) suggested that evidence for inexchangeability can support species division, while evidence for exchangeability supports synonymization. Our results indicate that historic gene flow has occurred, thus the species have demonstrated historic exchangeability. Current gene flow between putative species can be assessed with future research involving nuclear DNA markers. Some evidence for ecological inexchangeability may be found in differences in skull, horn, body size and colour morphology. On the other hand at their closest geographic point of contact the habitat and life history traits of these species are very similar indicating exchangeability (Geist, 1971).

Current taxonomy groups *O. dalli* populations into *O. d. stonei* or *O. d. dalli*. There is some support for this classification in nuclear DNA. Worley *et al.* (2004) found significant subspecies differentiation by grouping BC and CY together as Stone's sheep and all others as Dall's sheep, However, mtDNA and the clinal nature of colour morphology do not support this grouping. In addition OG differs in colour morphology from Dall's sheep populations, yet it is geographically and genetically isolated from other dark sheep populations. In sum our evidence confirms that the OG, CY and BC populations have been launched on independent evolutionary paths: isolation has occurred geographically and genetically for long time periods, and colour morphology has diverged. The current *O. d. stonei* subspecies designation does not appear useful to describe the evolutionary history of these populations, and we recommend that they be managed as separate entities.

Conclusions

Pleistocene glaciations are a major catalyst of the evolution of diversity (e.g. Pielou, 1991; Avise & Walker, 1998; Holder *et al.*, 2000), and our research demonstrates the importance of looking beyond well-known major glacial refugia to smaller ice-free areas to establish their evolutionary importance. We have shown that for North American mountain sheep such areas may have preserved a record of these species' evolutionary history. It appears that these small refugia have allowed for

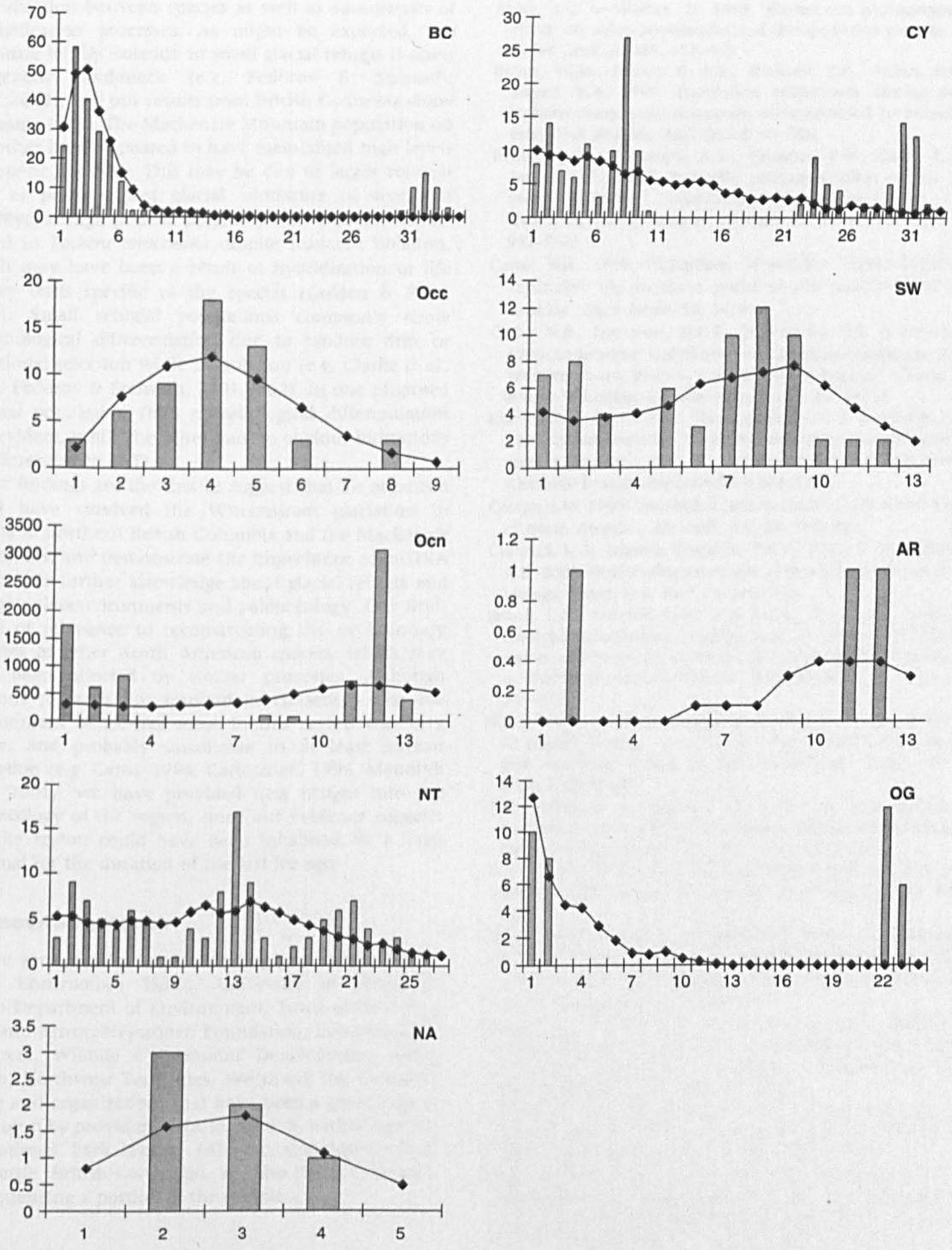


Fig. 4 Mismatch distribution for sequences from *Ovis dalli* and *O. canadensis* populations. Bars represent observed values and lines expected values for a model of sudden population expansion. *Y* axis: number of pairs, *X* axis: number of pair-wise differences. When graphed a population that has undergone rapid population expansion will exhibit a unimodal wave in the distribution of pair-wise genetic differences. In our data this trend is most evident in Occ and least evident in Ocn.

hybridization between species as well as stimulation of diversification processes. As might be expected, the signature left by isolation in small glacial refugia is often a genetic bottleneck (e.g. Fedorov & Stenseth, 2001,2002), and our results from British Columbia show this same trend. The Mackenzie Mountain population on the other hand appeared to have maintained high levels of genetic diversity. This may be due to larger refugial size, or possibly post glacial admixture of divergent haplotypes. High genetic polymorphism was also maintained in *Packera contermina* despite nunatak isolation, which may have been a result of hybridization or life history traits specific to the species (Golden & Bain, 2000). Small refugial populations commonly show morphological differentiation due to random drift or directional selection while in isolation (e.g. Clarke *et al.*, 2001; Fedorov & Stenseth, 2001,2002). In one proposed refugial population (BC) morphological differentiation was evident, while the other had no obvious indications of differentiation (NT).

Our findings are the first to suggest that an organism could have survived the Wisconsinan glaciation in refugia in northern British Columbia and the Mackenzie Mountains, and demonstrate the importance of mtDNA as a tool to further knowledge about glacial refugia and ice age paleoenvironments and paleoecology. Our finding is of relevance to reconstructing the evolutionary histories of other North American species, which may have been affected by similar processes. Although previous research has resulted in consensus that the environment of ice-free areas in this region was very severe, and probably unsuitable to at least human habitation (e.g. Catto, 1996, Catto *et al.*, 1996, Mandryk *et al.*, 2001), we have provided new insight into the paleoecology of the region, since our evidence suggests that the region could have been inhabited by a large mammal for the duration of the last ice age.

Acknowledgments

We are very grateful for the funding provided by Ella ja Georg Ehrnroothin Säätiö, University of Jyväskylä, Yukon Department of Environment, Town of Faro, and Ellen and Artturi Nyyssönen Foundation, Department of Resources, Wildlife & Economic Development, Sahtu Region, Northwest Territories. We thank the following people and organizations that have been a great help in this project by providing DNA samples: S. Arthur and the US National Park service (Alaskan samples) and H. Schwantje (British Columbia). We also thank J. Gratten for sequencing a portion of the samples.

References

Abbott, R.J., Smith, L.C., Milne, R.I., Crawford, R.M.M., Wolff, K. & Balfour, J. 2000. Molecular analysis of plant migration and refugia in the arctic. *Science* **289**: 1343–1346.

Avise, J.C. & Walker, D. 1998. Pleistocene phylogeographic effects on avian populations and the speciation process. *Proc. R. Soc. Lond. B.* **265**: 457–463.

Boyce, W.M., Ramey II, R.R., Rodwell, T.C., Rubin, E.S. & Singer, R.S. 1999. Population subdivision among desert bighorn sheep (*Ovis canadensis*) ewes revealed by mitochondrial DNA analysis. *Mol. Ecol.* **8**: 99–106.

Brunhoff, C., Galbreath, K.E., Fedorov, V.B., Cook, J.A. & Jaarola, M. 2003. Holarctic phylogeography of the root vole (*Microtus oeconomus*): implications for late Quaternary biogeography of high latitudes. *Mol. Ecol.* **12**: 957–968.

Catto, N.R. 1996. Richardson Mountains, Yukon-Northwest Territories: the northern portal of the postulated 'ice-free corridor'. *Quat. Intern.* **32**: 3–19.

Catto, N.R., Liverman, D.G.E., Bobrowsky, P.T. & Rutter, N. 1996. Laurentide, Cordilleran, and Montane Glaciation in the Western Peace River – Grande Prairie Region, Alberta and British Columbia, Canada. *Quat. Intern.* **32**: 21–32.

Clarke, T.E., Levin, D.B., Kavanaugh, D.H. & Reimchen, T.E. 2001. Rapid evolution in the Nebria gregaria group (Coleoptera: Carabidae) and the paleogeography of the Queen Charlotte Islands. *Evolution* **55**: 1408–1418.

Cowan, I.M. 1940. Distribution and variation in the native sheep of North America. *Am. Midl. Nat.* **24**: 505–580.

Crandall, K.A., Bininda-Emonds, O.R.P., Mace, G.M. & Wayne, R.K. 2000. Considering evolutionary processes in conservation biology. *Trends Ecol. Evol.* **15**: 290–295.

Dobeš, C.H., Mitchell-Olds, T. & Koch, M.A. 2004. Extensive chloroplast haplotype variation indicates Pleistocene hybridization and radiation of North American *Arabis drummondii*, *A. divaricarpa*, and *A. holboellii* (Brassicaceae). *Mol. Ecol.* **13**: 349–370.

Duk-Rodkin, A. 1999. Glacial limits map of Yukon Territory: Geological Survey of Canada, Open File 3694, Indian and Northern Affairs Canada Geoscience Map, 1999-2 Scale 1:1 000 000.

Duk-Rodkin, A. & Hughes, O.L. 1991. Age relationships of Laurentide and montane glaciations, Mackenzie Mountains, Northwest Territories. *Geog. Phys. Quat.* **45**: 79–91.

Dyke, A.S. & Prest, V.K. 1987. Late Wisconsinan and Holocene history of the Laurentide ice sheet. *Geog. Phys. Quat.* **41**: 237–263.

Ehrich, D., Fedorov, V.B., Stenseth, N.C., Krebs, C.J. & Kenney, A. 2000. Phylogeography and mitochondrial DNA (mtDNA) diversity in North American collared lemmings (*Dicrostonyx groenlandicus*). *Mol. Ecol.* **9**: 329–337.

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

Excoffier, L., & Schneider, S. 1999. Why hunter gatherer societies do not show signs of Pleistocene demographic expansions. *Proc. Natl. Acad. Sci.* **96**: 10597–10602.

Fedorov, V.B. & Stenseth, N.C. 2001. Glacial survival of the Norwegian Lemming (*Lemmus lemmus*) in Scandanavia: inference from mitochondrial DNA variation. *Proc. R. Soc. Lond. B.* **268**: 809–814.

Fedorov, V.B. & Stenseth, N.C. 2002. Multiple glacial refugia in the North American arctic: inference form phylogeography of the collared lemming (*Dicrostonyx groenlandicus*). *Proc. R. Soc. Lond. B.* **269**: 2071–2077.

Flagsted, Ø., & Røed, K.H. 2003. Refugial origins of reindeer (*Rangifer tarandus* L.) inferred from mitochondrial DNA sequences. *Evolution* **57**: 658–670.

Fu, Y.-X. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147**: 915–925.

Gabrielsen, T.M., Bachmann, K., Jakobsen, K.S. & Brochmann, C. 1997. Glacial survival does not matter: RAPD phylogeography of Nordic *Saxifraga oppositifolia*. *Mol. Ecol.* **6**: 831–842.

Galbreath, K.E. & Cook, J.A. 2004. Genetic consequences of Pleistocene glaciations for the tundra vole (*Microtus oeconomus*) in Beringia. *Mol. Ecol.* **13**: 135–148.

Geist, V. 1971. *Mountain Sheep: A Study in Behavior and Evolution*. University of Chicago Press, Chicago.

Geist, V. 1999. Adaptive strategies in mountain sheep. In: *Mountain Sheep of North America* (R. Valdez & P. Krausman, eds), pp. 192–208. The University of Arizona Press, Tucson, USA.

Golden, J.L. & Bain, J.F. 2000. Phylogeographic patterns and high levels of chloroplast DNA diversity in four *Packera* (Asteraceae) species in southwestern Alberta. *Evolution* **54**: 1566–1579.

Hewitt, G.M. 1996. Some genetic consequence of ice ages, and their role in divergence and speciation. *Biol. J. Linn. Soc.* **58**: 247–276.

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

Hiendleder, S., Kaupe, B., Wassmuth, R. & Janke, A. 2002. Molecular analysis of wild and domestic sheep questions current nomenclature and provides evidence for domestication from two different subspecies. *Proc. R. Soc. Lond. B.* **269**: 893–904.

Hillis, D.M., Moritz, C. & Mable, B.K. 1996. *Molecular Systematics*. Sinauer Associates, Sunderland.

Holder, K., Montgomerie, R. & Friesen, V.L. 1999. A test of the glacial refugium hypothesis using patterns of mitochondrial and nuclear DNA sequence variation in rock ptarmigan (*Lagopus mutus*). *Evolution* **53**: 1936–1950.

Holder, K., Montgomerie, R. & Friesen, V.L. 2000. Glacial vicariance and historical biogeography of rock ptarmigan (*Lagopus mutus*) in the Bering region. *Mol. Ecol.* **9**: 1265–1278.

Levson, V.M. & Rutter, N.W. 1996. Evidence of Cordilleran late Wisconsinan glaciers in the "ice-free corridor". *Quat. Intern.* **32**: 33–51.

Luikart, G. & Allendorf F.W. 1996. Mitochondrial DNA variation and genetic population structure in Rocky Mountain bighorn sheep. *J. Mammal.* **77**: 123–131.

Mandryk, C.A.S., Josenhans, H., Fedje, D.W. & Mathewes, R.W. 2001. Late Quaternary paleoenvironments of Northwestern North America: implications for inland versus coastal migration routes. *Quat. Sci. Rev.* **20**: 301–314.

Moritz, C. 1994. Defining 'evolutionary significant units' for conservation. *Trends Ecol. Evol.* **9**: 373–375.

Nielsen, R., Wakely, J. 2001. Distinguishing migration from isolation: a Markov chain Monte Carlo approach. *Genetics* **158**: 885–896.

Pielou, E.C. 1991. *After the Ice Age: Return of Life to Glaciated North America*. University of Chicago Press, Chicago.

Ramey, R.R. 1993. *Evolutionary Genetics and Systematics of North American Mountain Sheep: Implications for Conservation*, PhD thesis, Cornell University. p. 250.

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

Sambrook, J., Fritsch, E.F. & Maniatis, T. 1989. *Molecular Cloning: A Laboratory Manual*, 2nd edn. Cold Spring Harbour Laboratory Press, New York.

Schneider, S., & Excoffier, L. 1999. Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: application to human mitochondrial DNA. *Genetics* **152**: 1079–1089.

Schneider, S., Roessli, D. & Excoffier, L. 2000. *Arlequin: A software for population genetics data analysis*. Ver *2001*. Genetics and Biometry Lab, Deptartment of Anthropology, University of Geneva.

Sheldon, C. 1911. *The Wilderness of the Upper Yukon*, 1st edn. Charles Scribner's Sons, New York.

Slatkin, M., & Hudson, R.R. 1991. Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics* **129**: 555–562.

Szeicz, J.M. & MacDonald, G.M. 2001. Montane climate and vegetation dynamics in easternmost Beringia during the late Quaternary. *Quat. Sci. Rev.* **20**: 247–257.

Tremblay, N.O. & Schoen, D.J. 1999. Molecular phylogeography of *Dryas integrifolia*: glacial refugia and postglacial recolonization. *Mol. Ecol.* **8**: 1187–1198.

Troy, C.S., MacHugh, D.E., Bailey, J.F., Magee, D.A., Loftus, R.T., Cunningham, P., Chamberlain, A.T., Sykes, B.C. & Bradley, D.G. 2001. Genetic evidence for Near-Eastern origins of European cattle. *Nature* **410**: 1088–1091.

Wickström, L.M., Haukisalmi, V., Varis, S., Hantula, J., Fedorov, V.B. & Henttonen, H. 2003. Phylogeography of the circumpolar *Paranoplocephala arctica* species complex (Cestoda: Anoplocephalidae) parasitizing collared lemmings (*Dicrostonyx* spp.). *Mol. Ecol.* **12**: 3359–3371.

Willis, K.J. & Whittaker, R.J. 2000. The refugial debate. *Science* **287**: 1406–1407.

Worley, K., Strobeck, C., Arthur, S., Carey, J., Schwantje, H., Veitch, A. & Coltman, D.W. 2004. Population genetic structure of North American thinhorn sheep *Ovis dalli*. *Mol. Ecol.* **13**: 2545–2556.

Received 5 May 2005; revised 2 August 2005; accepted 9 August 2005