Evolutionary History, Population Genetics and Quantitative Genetics of Thinhorn Sheep

by

Zijian Sim

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Systematics and Evolution

Department of Biological Sciences

University of Alberta

Abstract

The field of evolutionary biology is centered on the study of processes that shape the diversity of life on Earth. Processes that shape genetic variation can act on a wide-range of spatial and temporal scales. Large-scale events such as glaciations and mountain uplift have wide-ranging effects across a species range, while diversity may be modified at finer scales by phenomena such as dispersal between populations or by sexual selection in mating groups. Recent advances in genomic technologies offer researchers powerful new tools to interrogate wildlife genomes. While the shift to genomic techniques is occurring throughout the field, organisms that are closely related to well-studied domestic species are considered "genome-enabled" and uniquely able to take advantage of genomic tools developed for domestics. One such species is the charismatic thinhorn sheep of North America. *Ovis dalli* is an alpine and subalpine ungulate endemic to the mountains of northwestern North America. The thinhorn sheep range stretches from Alaska to the west, east to the Mackenzie Mountains of the Northwest Territories and south through Yukon to northern British Columbia. There are two recognized subspecies of thinhorn sheep, 1) the white coated Dall's sheep (*O. d. dalli*), thought to occupy Alaska, Northwest Territories, and western Yukon, and 2) the dark coloured Stone's sheep (*O. d. stonei*), which inhabits BC and southeast Yukon. In this thesis, I developed and applied genomic resources to study the processes that shape genetic variation in thinhorn sheep. First, I applied two single nucleotide polymorphism (SNP) arrays, the OvineHD and OvineSNP50 BeadChips, which were originally designed for the domestic sheep, to thinhorn sheep and showed the utility of these cross-species SNPs for phylogenetic and species discrimination analysis. Second, I used genome-wide SNP data gathered using the OvineHD BeadChip to study the phylogeographic history of thinhorn sheep. I found evidence that the evolution of Dall's and Stone's sheep was

mediated by isolation in separate refugia during the late Pleistocene. Bayesian analysis of admixture also indicates a potential zone of hybridization in southern Yukon where sheep from the two glacial refugia met as they recolonized North America following the recession of the glaciers. Third, I investigated the fine-scale population genetic structure of thinhorn sheep by applying markers developed from the data acquired in earlier chapters to genotype over 2800 rams. I describe the global population genetic structure of thinhorn sheep and revealed three previously unreported Stone's sheep genetic clusters in the Stikine/Skeena, Cassiar and Rocky Mountains of British Columbia. I also redefined the range for Stone's sheep and showed that is much more restricted than currently accepted subspecies maps indicate, and almost exclusively confined within BC. Finally, I used genome-wide SNP data and an "animal model" to perform the first estimates for heritability in three fitness-related traits (horn length, horn base circumference and horn volume) in thinhorn sheep without a pedigree. I also performed a genome-wide association analysis for associations between SNP effects and each of the three traits. I found horn length, horn base circumference and horn volume to be moderately heritable and identified two SNP loci with suggestive associations to horn length. Taken together, this thesis provided insights into the evolution of thinhorn sheep and developed genomic tools for other wildlife researchers. Results from this thesis can also be used to inform the conservation management of thinhorn sheep.

Preface

This thesis represents collaboration between the University of Alberta, British Columbia Ministry of Lands, Forests, Natural Resource Operations and Rural Development (Bill Jex), and Environment Yukon (Troy Hegel). While all analyses conducted are my own original work, I used "we" in the text of my data chapters to reflect the collaborative nature of my research.

Chapter 2 of this thesis is being prepared for submission to the journal Conservation Genetic Resources. I conceived of and designed the study, collected and analyzed the data, and drafted the manuscript. DWC provided advice, analytical guidance and feedback for the manuscript drafts. We received data from the International Sheep Genomics Consortium and L.K. Ernst Institute for Animal Husbandry. Not all authors have edited and approved the manuscript.

Chapter 3 of this thesis has been published as Sim Z, Hall JC, Jex B, Hegel TM, Coltman DW (2016) Genome-wide set of SNPs reveals evidence for two glacial refugia and admixture from postglacial recolonization in an alpine ungulate. Molecular Ecology, 25, 962-1083. JCH, DWC and I conceived of the study, performed the data analysis and/or interpreted the results . I performed the laboratory work and wrote the manuscript. BJ and TMH provided samples. All authors edited and approved the manuscript.

Chapter 4 of the thesis has been published as Sim Z, Davis CS, Jex B, Hegel TM, Coltman DW (2018) Management implications of highly resolved hierarchical population genetic structure in thinhorn sheep. Conservation Genetics. https://doi.org/10.1007/s10592-018-1123-2. All authors were involved in the conception and design of the study. BJ and TMH provided samples. I collected and analyzed the data, and drafted the manuscript. DWC and CSD provided analytical guidance and feedback for the manuscript drafts. All authors edited and approved of the manuscript.

Chapter 5 of the thesis is the result collaborative work by Sim Z, Davis C, Jex B, Hegel T, Coltman DW. I conceived of the study. DWC and I designed the study. I performed the data

analysis and wrote the manuscript. DWC provided help in results interpretation and provided substantial feedback on the manuscript. Not all authors have edited and approved the manuscript.

Acknowledgements

To my supervisory committee: Dave, thank you for taking me on as student after my failed Master's attempt at UofT. You gave me chance when many would not have and I am forever grateful for that. Thanks giving me the intellectual freedom to pursue my own research interest and for all the doors you opened for me. Corey, thanks for being my closest supervisor and a mentor for science and beyond. I am grateful for all your help and patient guidance with my lab work. It was a pleasure to be your TA for 392/592. Jocelyn, thank you for all your help with phylogenetic analyses and my phylogeography paper.

To my collaborators Bill and Troy, thank you for being wonderful partners and allowing me the opportunity to make a tangible contribution to conservation. Thanks also for going above and beyond with your support for my other projects. It was an honour to have worked with you two.

I want to thank my fellow Coltman Lab mates, and the wider Bio Sci and UofA community - Cathy, Michelle, Jess R, Jess H, Jamie, Anh, Rachel, Edward, Christianne and Ian. Special thanks to Josh, for all your help with my research and always being willing to read a draft or just have a good science conversation with. You embodied what it means to be the "senior grad student", the glue that held the lab together… with baked goods. Rene, thanks for all your help with bioinformatics and analyses. To Brandon, Murdoch, Kyle & Matt, thanks for the friendship - I miss the days when we can all still sit around a fire and chat about anything under the sun.

UBCO: Mike R, thank you for giving me my start in research. I can still remember how excited I was on the first day of the evolution class that you and Bob taught. The days I spent in your lab are some of my most cherished memories. Philippe, my original "supervisor". Thanks for taking me under your wing and teaching me your craft and mentoring me all while trying to complete a PhD and raising a kid. Thanks for always looking out for me.

Mike Turner, thanks for being a great friend. The music you introduced to me has been a constant companion in this academic journey. I look forward to many more adventures with you, musically or otherwise. May our friendship stay forever young.

To the extended hunting community. It has been a pleasure to have worked and interacted with many of you. Thanks for always making a rank outsider like me feel welcome.

To my colleagues in the Forensic Lab, thanks for being so supportive and never making me feel like doing my PhD was a burden on the lab. I am lucky to be part of such a great group.

I want to thank the ODAC family Mark, Zhong, LX, Jason, Seb, Weng, Francis, Nick, LJ, QY, YR, Eunice, May… for all the unspeakable things we went through ☺. Thanks to Mr Lim for all the life lessons, and always being willing to stretch and hold me to high standards. Mr Chin - you encouraged, fostered and grew my love for biology.

Thanks to my funding sources - University of Alberta, BC HCTF, BC MLFNRORD, Environment Yukon, Wild Sheep Foundation, Wild Sheep Society of BC, Alberta Innovates.

To my family: Dad, you were gone too soon. I'm sorry you never got to see me graduate and hold your own granddaughter. Know that you are in our heart always. Mum, thanks for always making sure I had what I needed to succeed. Zihui, thanks for being our family's rock and holding everyone together. Mama thanks for raising me and showing me what it means to love and care unconditionally. Gugu, thanks for loving me like a son. Thanks for showing me the qualities of tolerance, patience, hard work and a sense of adventure. There is no one in the world I respect more than you. Mum and dad (in Canada): thanks for allowing us to lean on you when we needed to most. Your enthusiasm for my research made the results all the more meaningful.

Tiff, thanks for being my greatest supporter and partner. Your love and support mean the world to me. Thanks for doing more than your fair share for the family and for your patience as I finished my thesis. This PhD is as much yours as it is mine. I was very lucky the day I decided to play flag football in Toronto. Lena, you fill daddy's heart with joy. Watching you grow up and discover the world has been magical. May your curious mind find you all the adventures and discoveries the world has to offer.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1 General introduction

The field of evolutionary biology is centered on the study of the processes that shape the diversity of life on Earth. Processes that shape variation can act across wide spatial and temporal scale. For instance, while large-scale events such as glaciations and mountain uplift have wide-ranging effects across a species range, diversity may be modified at finer scales by local phenomena such as dispersal between populations or by sexual selection in mating groups. Therefore, a comprehensive investigation of a species evolutionary history should examine variation at various spatial and temporal resolutions.

By virtue of its role as the hereditary material, DNA has long been of interest to evolutionary biologists. Waves of technological innovation (Maxam & Gilbert 1977; Mullis *et al.* 1986; Sanger & Coulson 1975; Southern 1975) have advanced our ability to examine and sequence the genomes of organisms to probe the effects of evolutionary and ecological processes, sparking studies into molecular evolution and ecology. Most recently, the development of massively parallel DNA sequencing and array technologies starting in the early 2000s have drastically reduced the unit cost of obtaining genomic data, thus ushering in the "next-generation sequencing era" (Levy & Myers 2016; Metzker 2010). For example, the first sequencing of the human genome was a multi-laboratory collaboration which took 10 years to the cost of US$300 million to complete (Collins *et al.* 1998). Over the 20 years since, the cost of genome sequencing has dropped precipitously and, as of 2016, approaches $1000/genome (Wetterstrand 2018).
For wild species, the ability to cost effectively gather genomic data has prompted a broad suite of evolutionary and ecological questions that were once restricted by the ability of small marker sets to probe (Davey *et al.* 2011; Helyar *et al.* 2011; Luikart *et al.* 2003). This has enabled new inquiries into genomic organization (Genome 2009; Yandell & Ence 2012), phylogenomics (Decker *et al.* 2009; McCormack *et al.* 2013), cryptic population structure (Steane *et al.* 2015), association mapping (Santure & Garant 2018) and heterozygosity fitness correlations (Chapman *et al.* 2009). Genomic data also has potential applications in conservation (Allendorf *et al.* 2010), such as in the delineation of biologically-sound conservation units for management (Funk *et al.*

2012), protection of threatened species (Steiner *et al.* 2013) and wildlife forensics in support of conservation law enforcement (Ogden 2011).

While the shift to genomic techniques is occurring throughout the field of evolutionary biology, organisms that are closely related to well-studied domestic species are considered "genome enabled" and uniquely able to take advantage of genomic tools developed for domestics (Kohn *et al.* 2006). One such species is the thinhorn sheep (*Ovis* dalli) of North America. *Ovis dalli* is perhaps most well known for massive horns found in adult males, and is of great cultural and ecological significance to the mountainous regions of northwestern North America. The thinhorn sheep is an alpine and subalpine ungulate whose range stretches from Alaska in the west, east to the Mackenzie Mountains of the Northwest Territories (NWT) and south through Yukon to northern British Columbia (BC) (Valdez & Krausman 1999). There are two recognized subspecies of thinhorn sheep: 1) the white-coated Dall's sheep (*O. d. dalli*), thought to occupy Alaska, NWT, and western Yukon, and 2) the dark coloured Stone's sheep (*O. d. stonei*), which inhabits BC and southeast Yukon (Demarchi & Hartwig 2004). The transition between Dall's and Stone's sheep occurs in a clinal fashion over a contact zone containing sheep of intermediate coat coloration (Sheldon 1911). This region extends from the Ogilvie Mountains in Yukon south to the northern end of the Cassiar Mountains in BC. Sheep occupying the transitional region are informally referred to as Fannin sheep (Feldhamer *et al.* 2003).

Repeated glaciations and movements of ice-sheets characterize the geological history of northwestern North America during the Pleistocene. Therefore, like many North American species, the evolutionary history of thinhorn sheep is inextricably linked to glacial movements (Shafer *et al.* 2010). Glacial movements in North America likely impacted sheep populations in several ways. In times of glacial advance, distinct populations may have been isolated in ice-free refugia, where they would have potentially been set on different evolutionary paths due to restricted gene flow and altered selective regimes (Hewitt 2000; Hewitt 1996). Alternatively, glacial advance is also associated with lower sea levels, thus enabling migration across over previously impassable water bodies (Elias *et al.* 1996). Glaciations can also affect the physical (e.g. U-shaped valleys) and biochemical (e.g. mineral deposits) state of landscapes (Anderson

2007).

Previous research has indicated that the ancestor for all North American mountain sheep is an argali-like (*Ovis ammon*) animal that crossed the Bering land bridge from Asia into North America during the early parts of the Pleistocene (Valdez & Krausman 1999). Phylogenetic studies using mitochondrial DNA (mtDNA) indicate that thinhorn sheep split from its sister species, the bighorn sheep (*O. canadensis*), approximately 1 to 1.4 million years ago during the middle Pleistocene (Bunch *et al.* 2006; Rezaei *et al.* 2010). A previous mtDNA study has also suggested that the two thinhorn sheep subspecies, Dall's and Stone's sheep, potentially evolved in separate glacial refugia during the Wisconsin glaciation of the Pleistocene (Loehr *et al.* 2006). At the population level, mountain range boundaries have been found to delineate population groups (Worley *et al.* 2004) while steep precipitous terrain has been associated with gene flow between population due to thinhorn sheeps preference for rocky habitat (Roffler *et al.* 2014). Other researchers have also found the distribution of some putatively ecologically significant genes to be associated with population structure (Loehr *et al.* 2008; Worley *et al.* 2006).

1.2 Thesis objectives and data chapters

For my doctoral studies, I researched the processes and mechanisms that shape the distribution of genetic variation in thinhorn sheep at various spatial and temporal scales. In this thesis, which is comprised of four data chapters, I 1) describe the development of genomic resources for thinhorn sheep, and used these resources to 2) investigate the range-wide phylogeography and glacial refugia of wild sheep in North America, 3) elucidate the fine-scale population genetic structure in thinhorn sheep, and 4) examine the genetic basis of horn size in thinhorn sheep.

In **Chapter 2,** I use a cross-species application of the domestic sheep OvineSNP50 and OvineHD BeadChips to develop genomic resources in a set of ungulate species. I examine the utility of cross-species single nucleotide polymorphism (SNP) data for the purposes of classification and species delimitation questions using phylogenetic and principal component analysis. I also discuss the potential application of this resource for wildlife forensics.

In **Chapter 3**, I apply the OvineHD BeadChip to examine the range-wide distribution of genetic variation and refugial origins of thinhorn sheep. I use phylogenetic inference to evaluate the hypothesis that thinhorn sheep survived in two refugia during the last glaciation. I also investigate the nature of postglacial recolonization using population genetic clustering and admixture analysis of relevant phylogenetic groups**.**

In **Chapter 4**, I analyze the fine-scale population genetic structure of thinhorn sheep by applying markers developed from the data acquired in **Chapters 2** and **3** to genotype over 2800 samples. I compare subspecies distributions derived from genetic data with the subspecies maps currently used for management. Finally, I assess the levels of genetic differentiation between currently used management units and compare the boundaries of current management units to SNP-derived genetic boundaries.

In **Chapter 5**, I use genome-wide SNP data from over 180 thinhorn rams acquired using the OvineHD Beadchip to examine the heritability and estimate variance components of horn size in thinhorn sheep. I also perform a genome-wide association study to explore links between SNP effects and horn size.

1.3 References

Allendorf FW, Hohenlohe PA, Luikart G (2010) Genomics and the future of conservation genetics. *Nat Rev Genet* **11**, 697-709.

Anderson SP (2007) Biogeochemistry of Glacial Landscape Systems. *Annual Review of Earth and Planetary Sciences* **35**, 375-399.

Bunch TD, Wu C, Zhang Y-P, Wang S (2006) Phylogenetic analysis of snow sheep (Ovis nivicola) and closely related taxa. *Journal of Heredity* **97**, 21-30.

Chapman JR, Nakagawa S, Coltman DW, Slate J, Sheldon BC (2009) A quantitative review of heterozygosity–fitness correlations in animal populations. *Molecular Ecology* **18**, 2746-2765.

Collins FS, Brooks LD, Chakravarti A (1998) A DNA polymorphism discovery resource for research on human genetic variation. *Genome Research* **8**, 1229-1231.

Davey JW, Hohenlohe PA, Etter PD*, et al.* (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics* **12**, 499-510.

Decker JE, Pires JC, Conant GC*, et al.* (2009) Resolving the evolution of extant and extinct ruminants with high-throughput phylogenomics. *Proceedings of the National Academy of Sciences* **106**, 18644-18649.

Demarchi R, Hartwig C (2004) Status of Thinhorn Sheep in British Columbia. *British Columbia Ministry of Water, Land and Air Protection, Biodiversity Branch - Wildlife Bulletin No. B-119*.

Elias SA, Short SK, Nelson CH, Birks HH (1996) Life and times of the Bering land bridge. *Nature* **382**, 60.

Feldhamer GA, Thompson BC, Chapman JA (2003) *Wild mammals of North America: biology, management, and conservation* JHU Press.

Funk WC, McKay JK, Hohenlohe PA, Allendorf FW (2012) Harnessing genomics for delineating conservation units. *Trends in Ecology & Evolution* **27**, 489-496.

Genome KCoS (2009) Genome 10K: A Proposal to Obtain Whole-Genome Sequence for 10 000 Vertebrate Species. *Journal of Heredity* **100**, 659-674.

Helyar SJ, Hemmer-Hansen J, Bekkevold D*, et al.* (2011) Application of SNPs for population genetics of nonmodel organisms: new opportunities and challenges. *Molecular Ecology Resources* **11**, 123-136.

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

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

Kohn MH, Murphy WJ, Ostrander EA, Wayne RK (2006) Genomics and conservation genetics. *Trends in Ecology & Evolution* **21**, 629-637.

Levy SE, Myers RM (2016) Advancements in Next-Generation Sequencing. *Annual Review of Genomics and Human Genetics* **17**, 95-115.

Loehr J, Worley K, Grapputo A*, et al.* (2006) Evidence for cryptic glacial refugia from North American mountain sheep mitochondrial DNA. *Journal of Evolutionary Biology* **19**, 419-430.

Loehr J, Worley K, Moe J, Carey J, Coltman DW (2008) MC1R variants correlate with thinhorn sheep colour cline but not individual colour. *Canadian Journal of Zoology* **86**, 147-150.

Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing. *Nat Rev Genet* **4**, 981-994.

Maxam AM, Gilbert W (1977) A new method for sequencing DNA. *Proceedings of the National Academy of Sciences of the United States of America* **74**, 560-564.

McCormack JE, Hird SM, Zellmer AJ, Carstens BC, Brumfield RT (2013) Applications of next-generation sequencing to phylogeography and phylogenetics. *Molecular Phylogenetics and Evolution* **66**, 526-538.

Metzker ML (2010) Sequencing technologies - the next generation. *Nat Rev Genet* **11**, 31-46.

Mullis K, Faloona F, Scharf S*, et al.* (1986) Specific Enzymatic Amplification of DNA In Vitro: The Polymerase Chain Reaction. *Cold Spring Harbor Symposia on Quantitative Biology* **51**, 263-273.

Ogden ROB (2011) Unlocking the potential of genomic technologies for wildlife forensics. *Molecular Ecology Resources* **11**, 109-116.

Rezaei HR, Naderi S, Chintauan-Marquier IC*, et al.* (2010) Evolution and taxonomy of the wild species of the genus Ovis (Mammalia, Artiodactyla, Bovidae). *Molecular Phylogenetics and Evolution* **54**, 315-326.

Roffler G, Talbot S, Luikart G*, et al.* (2014) Lack of sex-biased dispersal promotes fine-scale genetic structure in alpine ungulates. *Conservation Genetics*, 1-15.

Sanger F, Coulson AR (1975) A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of Molecular Biology* **94**, 441-448.

Santure AW, Garant D (2018) Wild GWAS—association mapping in natural populations. *Molecular Ecology Resources* **18**, 729-738.

Shafer ABA, Cullingham CI, Côté SD, Coltman DW (2010) Of glaciers and refugia: a decade of study sheds new light on the phylogeography of northwestern North America. *Molecular Ecology* **19**, 4589-4621.

Sheldon C (1911) *The wilderness of the Upper Yukon; a hunter's explorations for wild sheep in sub-Arctic Mountains* T. Fisher Unwin, New York.

Southern EM (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *Journal of Molecular Biology* **98**, 503-517.

Steane DA, Potts BM, McLean E*, et al.* (2015) Genome-wide scans reveal cryptic population structure in a dry-adapted eucalypt. *Tree Genetics & Genomes* **11**, 33.

Steiner CC, Putnam AS, Hoeck PEA, Ryder OA (2013) Conservation Genomics of Threatened Animal Species. *Annual Review of Animal Biosciences* **1**, 261-281.

Valdez R, Krausman PR (1999) *Mountain sheep of North America* University of Arizona Press.

Wetterstrand K (2018) DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP), National Human Genome Research Institute, NIH.

Worley K, Carey J, Veitch A, Coltman DW (2006) Detecting the signature of selection on immune genes in highly structured populations of wild sheep (Ovis dalli). *Molecular Ecology* **15**, 623-637.

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

Yandell M, Ence D (2012) A beginner's guide to eukaryotic genome annotation. *Nature Reviews Genetics* **13**, 329.

Chapter 2: Cross species application of ovine SNP arrays for phylogenetic inference and species identification

2.1 Abstract

Single nucleotide polymorphisms (SNPs) are increasingly used by wildlife researchers to investigate ecological, evolutionary and conservation related questions. While recent advances in genomic technologies have drastically decreased the cost of SNP genotyping, marker discovery still presents a challenge, particularly for non-model species. The cross species application of SNP genotyping arrays designed for domestic species to their wild relatives can potentially allow for rapid marker discovery and genotyping in a cost effective fashion. Here, we apply the OvineSNP50 BeadChip, which was originally designed for domestic sheep, to 13 wild ungulate species from the Antilocapridae and Bovidae families, and assessed the data for use in phylogenetic inference and species identification. We found that SNP data generated by cross-species amplification was able to recover a tree topology that matches the known phylogenetic relationships of the species tested while branch lengths corresponded with time since last common ancestor with domestic sheep. We found coding SNP data as genotypes instead of alleles was more successful in recovering the correct tree topology. We also found cross species data to be capable of species identification when combined with principal component analysis or an assignment test. We highlight a potential application of the OvineSNP50 BeadChip for wildlife forensic investigations.

2.2 Introduction

The use of single nucleotide polymorphisms (SNPs) has risen greatly in the past decade and SNPs are fast becoming the marker of choice for many molecular studies (Allendorf *et al.* 2010; Garvin *et al.* 2010; Slate *et al.* 2009). While the price of sequencing has dropped precipitously, marker discovery, protocol optimization and post genotyping data processing remain potential limiting factors (Davey *et al.* 2011). For some wild species, the use of commercial SNP resources developed for closely related domestic species (Kohn *et al.* 2006) mitigate many of the difficulties of "blind" marker discovery and genotyping by making available a dense genome-

wide panel of SNPs (Miller *et al.* 2012a; Miller *et al.* 2011; Pertoldi *et al.* 2010). The cross species use of domestic SNP resources also allows wildlife researchers to interrogate loci that are mapped to well-annotated genomes of their close domestic relatives. If specific loci require additional investigation post marker discovery, primers can also be easily designed by referencing the flanking sequences of the SNPs in question. Well-annotated genomes also allow for easy post genotyping data processing and eliminate the need for time-consuming sequence assembly. Since conversion rates are well documented (Haynes & Latch 2012; Miller *et al.* 2011; Sim *et al.* 2016) and generally known to be inversely correlated with time to last common ancestor (Haynes & Latch 2012; Miller *et al.* 2012a), researchers can reliably estimate the number of markers that will be obtained, greatly benefiting study planning.

Studies that can potentially utilize cross species applications of domestic SNP resources include phylogenetic and species identification investigations. While researchers have used domestic SNP arrays for phylogenetic inference in closely related wild species (Decker *et al.* 2009; Sim *et al.* 2016), none had yet compared the performance of different SNP coding schemes and ascertainment bias correction against well studied phylogenies as reference. As well, while specially designed SNP chips are routinely used for species identification in heavily monitored and managed taxa such as salmonids (Wenne *et al.* 2016), no study has attempted to perform species identification in a diverse selection of wild ungulates using cross-species application of a domestic SNP chip. The ability to quickly determine species with little to no development time can have many practical applications such as in the forensic investigation of wildlife trafficking incidents.

Here, we investigate the cross species application of SNPs genotyped using the OvineSNP50 BeadChip for phylogenetic inference on 13 wild ungulates from the Antilocapridae and Bovidae families, and compare the results of our analysis against published phylogenetic relationships. Since previous studies have found amplification rates and polymorphism to be inversely correlated with time to last common ancestor (TLCA) (Miller *et al.* 2012a), we expect branch lengths to show the same relationship i.e. species with shortest TLCA have the longest branches. We also assess the performance of ascertainment bias correction for invariant sites and different

SNP coding schemes on the results of our phylogenetic analysis. We attempt to demonstrate the utility of phylogenetic analysis for species delimitation. Additionally, we performed principal component analyses (PCA) to visualize the data and also to assess the use of PCAs for species delimitation.

2.3 Methods

2.3.1 Dataset

In total, we collected SNP data for 60 samples representing 13 wild ungulate taxa from Antilocapridae (*Antilocapra*) and Bovidae (*Bos, Ammotragus, Hemitragus, Ovibos, Ovis, Rupricapra*) (Table 1). Within Bovidae, samples were split into subfamily Bovinae (*Bos*) and Antilopinae (*Ammotragus, Hemitragus, Ovibos, Ovis, Rupricapra*). All taxa from the subfamily Bovinae in this study belong to the tribe Bovini, while the all taxa belonging to the subfamily Antiloponae were from the tribe Carpini *sensu lato* (Hassanin & Douzery 1999). Samples were genotyped using the ovineSNP50 BeadChip (Illumina) except for individuals from *O. canadensis* and *O. dalli,* which were genotyped using the OvineHD BeadChip (Illumina). We used the extract SNPs function in PLINK1.07 (--extract) (Purcell *et al.* 2007) to extract only SNPs that are present in the OvineSNP50 BeadChip from *O. canadensis* and *O. dalli* samples dataset. Development (Kijas *et al.* 2012; Kijas *et al.* 2014), genotyping and raw signal quality control (Deniskova *et al.* 2016; Miller *et al.* 2018; Miller *et al.* 2012a; Miller *et al.* 2011; Sim *et al.* 2016) of the SNP datasets has been described in previous studies (Table 2.1).

2.3.2 Quality control and summary statistics

All quality control and summary statistics were performed using PLINK1.07 (Purcell *et al.* 2007). First, we removed markers with locus specific call rates <0.9 within each species (--mind, --geno). Since call rates are known to correlate with TLCA, this quality control step prevents the inclusion of loci that perform poorly for one or few species but well for all or most other species. Next, we merged data from each species (--merge-list) and retained only loci with greater than >0.9 call rate in the merged 13 species dataset to minimize systematic missing data (i.e. loci with call rates >0.9 in only a few species). We also removed loci with minor allele frequency <0.01 (-

-maf). Since the *O. dalli* dataset only consists of males, we removed all loci on the sex chromosomes (--exclude).

2.3.3 Phylogenetic Inference

We performed phylogenetic analysis for the 13 ungulate species using maximum likelihood (ML) implemented in IQ-TREE (Trifinopoulos *et al.* 2016). To assess the robustness of our phylogenetic analysis, we explored two SNP coding schemes. 1) IUPAC: Each genotype was represented by its International Union of Pure and Applied Chemistry (IUPAC) ambiguity code and the input data were generated by concatenating all genotypes (e.g. for three genotypes AA, GG and GC, the resulting concatenation will be AGS), and 2) morphological: each genotype was coded by its unique state (e.g. AA=0, AT=1, TT=2, resulting in 012). Data conversion was performed using PLINK1.07 (--recodeA). We also assessed the use of the Mkv model for ascertainment bias correction (+ASC) (Lewis 2001), which is recommended for use in SNP datasets to account for the lack of invariant sites (Leaché *et al.* 2015). We allowed IQ-Tree to select the optimal substitution model using Bayesian Information Criterion (BIC) (Schwarz 1978), allowing free rate heterogeneity (-m TEST). For the IUPAC coded dataset, IQ-tree selected TIM+R2, which is a transition model with 1) variable base frequencies, 2) two transversion rates (AC=GT; AT=CG) and 3) free rate heterogeneity across sites, as the best model, while MK+FQ+G4, which is a Markov model with a discrete Gamma model to account rate heterogeneity, was selected for the morphological dataset. Pronghorn antelope (*Antilocapra americana*) samples were used as the outgroup in our phylogenetic analyses.

Tree topologies from our ML analysis and the currently accepted phylogenetic relationships derived from two mitochondrial (cytochrome *b* and 12S rRNA) and one nuclear (exon 4 κ-casein) genes (Ropiquet & Hassanin 2005) were visually inspected for differences in clades supported by >70% bootstrap (Mason-Gamer & Kellogg 1996; Seelanan *et al.* 1997). Previous studies have shown this approach to be useful in determining areas of incongruence between trees derived from different datasets (Merckx *et al.* 2013; Scheunert & Heubl 2014). We also compared branch lengths and TLCA. Estimates for TLCA were referenced from Miller *et al.* (2012a).

2.3.4 Species identification

We explored the use of SNP genotypes to differentiate between all 13 sampled species using PCA analyses (glPCA) implemented in the R package ADEGENET (Jombart 2008). PCAs were visualized using the s.class function in the R package ade4 (Dray & Dufour 2007). Since previous studies have shown that species specific polymorphism and missing data are inversely correlated with TLCA from the species for which the SNP array was discovered for (Miller *et al.* 2012a), we re-ran our PCA analysis without species within the *Ovis* genus (*O. orientalis, O. vignei*, *O. ammon*, *O. nivicola*, *O. canadensis*, and *O. dalli*) since the OvineSNP50 BeadChip was designed for domestic sheep, *O. aries*. We also tested the performance of an assignment test (Paetkau *et al.* 1995) to distinguish between all the tested species using GENODIVE (Meirmans & Van Tienderen 2004). We used a missing allele frequency of 0.002 (Paetkau *et al.* 2003), acceptance threshold of 0.01 and a Monte Carlo simulation of 10,000 individuals to generate the null distribution of likelihood values (Cornuet *et al.* 1999).

2.4 Results

2.4.1 Quality control and summary statistics

Of the initial 54,241 SNPs genotyped, 49,818 loci passed within species quality control for >0.9 call rate in at least one species. Once merged into one 13 species dataset, we removed 1,363 loci found on the sex chromosomes followed by 41,195 loci with call rates <0.9 and finally a further 2,015 loci with MAF <0.01. The final dataset consisted of 5,245 loci.

2.4.2 Phylogenetic inference

Overall, ML phylogenetic analysis with SNP loci found the family Bovidae to be monophyletic using both IUPAC and morphological coding schemes (Figure 2.1). Within Bovidae, the subfamilies Bovinae and Antilopinae (tribe Caprini *s.l.*) were found to be reciprocally monophyletic. Within subfamily Bovinae/tribe Bovini, we found strong support for reciprocal monophyly between *B. gaurus* and *B. mutus*. Within Antilopinae/tribe Caprini *s.l*, the morphological coding scheme revealed monophyletic relationships for all individuals belonging

to the same species while analysis using IUPAC coding found *O. vignei* to be paraphyletic with respect to *Ovis orientalis* (Figure 2.1). Branch lengths were very different for each species and were correlated with TLCA with *O. aries* (Figure 2.2). Tree topologies for ML analysis ran with and without ascertainment bias correction were identical.

2.4.3 Species delimitation

Principal component analysis revealed groupings of individuals of the same species and distinguished between individuals from different species although the degree of visual separation corresponded with TLCA (Figure 2.3). Other than *O. orientalis*, *O. vignei*, and *O. ammon*, individuals were very tightly clustered. The first principal component (PC1) separated all species within the genus *Ovis* while PC2 separated all other species analyzed. When all *Ovis* species were removed, clearer separation was observed among the remaining species. PC1 separated all members of the tribe Caprini *s.l.* while PC2 separated the rest of the species (Figure 2.4). In our assignment test, all samples were correctly assigned to their species ($p < 0.01$).

2.5 Discussion

Here, we found the OvineSNP50 BeadChip to be suitable for some forms of phylogenetic inference. We also found phylogenetic inference, assignment tests and PCAs were able to differentiate species identity. When used in phylogenetic inference, we found that the morphological coding scheme performed better than IUPAC coding in resolving known species relationships. The tree topology for the morphological coding scheme showed no hard conflicts with the current accepted phylogeny of ungulates (Hassanin & Ropiquet 2004; Rezaei *et al.* 2010; Ropiquet & Hassanin 2005), and all family, subfamily and tribe identities were reproduced with high confidence (bootstrap > 90%). Ascertainment bias correction did not change the tree topologies in our analysis, however since the uncorrected analysis reproduced the "correct" phylogeny, our dataset likely does not properly evaluate the effectiveness of ascertainment bias corrections. Branch lengths were found to be inversely proportionate with TLCA. This is unsurprising since previous studies have found polymorphism to be inversely correlated with time to last common ancestor (Miller *et al.* 2012a). Therefore, cross-species SNP arrays are likely not useful for any analysis where accurate branch lengths are required, such as in

molecular dating. For applications where only topologies are important, such as in species differentiation, cross species SNP chip applications will likely be useful.

Cross species application of the OvineSNP50 BeadChip have previously been shown to be useful in discriminating between two wild sheep species, *O. canadensis* and *O. dalli,* through PCA analysis (Miller *et al.* 2011). Here, we further show the utility of the Ovine50 BeadChip to differentiate between 13 ungulate species as all members of the same species cluster together. As expected, variation and thus the discriminatory ability of PCA analysis corresponded with TLCA. The principal separation of samples in our analysis was within the genus *Ovis* due to their close relation with *O. aries*. Within *Ovis,* the three species most closely related to *O. aries*, *O. ammon*, *O. orientalis*, and *O. vignei*, were also the most visually divergent. Our assignment test also correctly assigned all samples to their species using cross species SNPs.

The ability of researchers for some taxa to take advantage of genomic resources developed for closely related domestic animals such as sheep, horse and cow, have led them to be termed "genome-enabled" (Kohn *et al.* 2006). We have shown the OvineSNP50 BeadChip to be a valuable genome resource for SNP discovery and genotyping for studies that investigate phylogenetic relationships and species identification in wild ungulates. A potential practical application of species identification using the ovineSNP50 BeadChip may be in the field of wildlife forensics, which unlike human forensics, suffers from the lack of commercially optimized kits (Zhou *et al.* 2014). Additionally, tests for multiple species are also needed since unlike in human forensics (where non-human targets are excluded), wildlife forensic casework frequently encounter biological material with many possible species origins (e.g. ground meat consisting of mixtures of wild and domestic species). Wildlife regulations are frequently species specific (e.g. Wildlife Act of Alberta) so species identification is frequently the key question in an investigation. Matching or individualization analysis also requires the application of species-specific tests. Therefore, when the species in a sample is unknown, a time consuming 2-step protocol is required, (generally) first an mtDNA test for species identification, followed by a 8-15 microsatellite test for individual identification. The cross species use of the ovine50 BeadChip can potentially a enable a single protocol to perform marker discovery, species identification and

individualization, alleviating the need for lengthy development validation experiments and improving workflow efficiency in forensic casework.

.

Table 2.1 Taxonomic information and original reference for samples used in this study. *Samples could only be identified to the genus level

Family	Subfamily	Tribe	Species	Common name	N	SNP array	Original reference
Bovidae	Bovinae	Bovini	*Bos gaurus*	Gaur	2	OvineSNP50	Kijas et al. 2012
			Bos mutus	Yak	2	OvineSNP50	Kijas et al. 2012
	Antilopinae	Caprini *s.l.*	*Ammotragus lervia*	Babary sheep	4	OvineSNP50	Kijas et al. 2012
			*Hemitragus spp.**	Tahr	5	OvineSNP50	Kijas et al. 2012
			Ovibos moschatus	Muskox	6	OvineSNP50	Kijas et al. 2012
			Rupicapra spp. *	Chamois	4	OvineSNP50	Kijas et al. 2012
			Ovis ammon	Argali sheep	6	OvineSNP50	Kijas et al. 2012
			Ovis canadensis	Bighorn sheep	4	OvineHD	Miller et al. 2018
			Ovis dalli	Thinhorn sheep	4	OvineHD	Sim et al. 2016
			Ovis nivicola	Snow sheep	4	OvineSNP50	Deniskova et al. 2016
			Ovis orientalis	Mouflon sheep	8	OvineSNP50	Kijas et al. 2012
			Ovis vignei	Urial sheep	3	OvineSNP50	Kijas et al. 2012
Antilocaparidae			*Antilocapra americana*	Pronghorn	8	OvineSNP50	Kijas et al. 2012

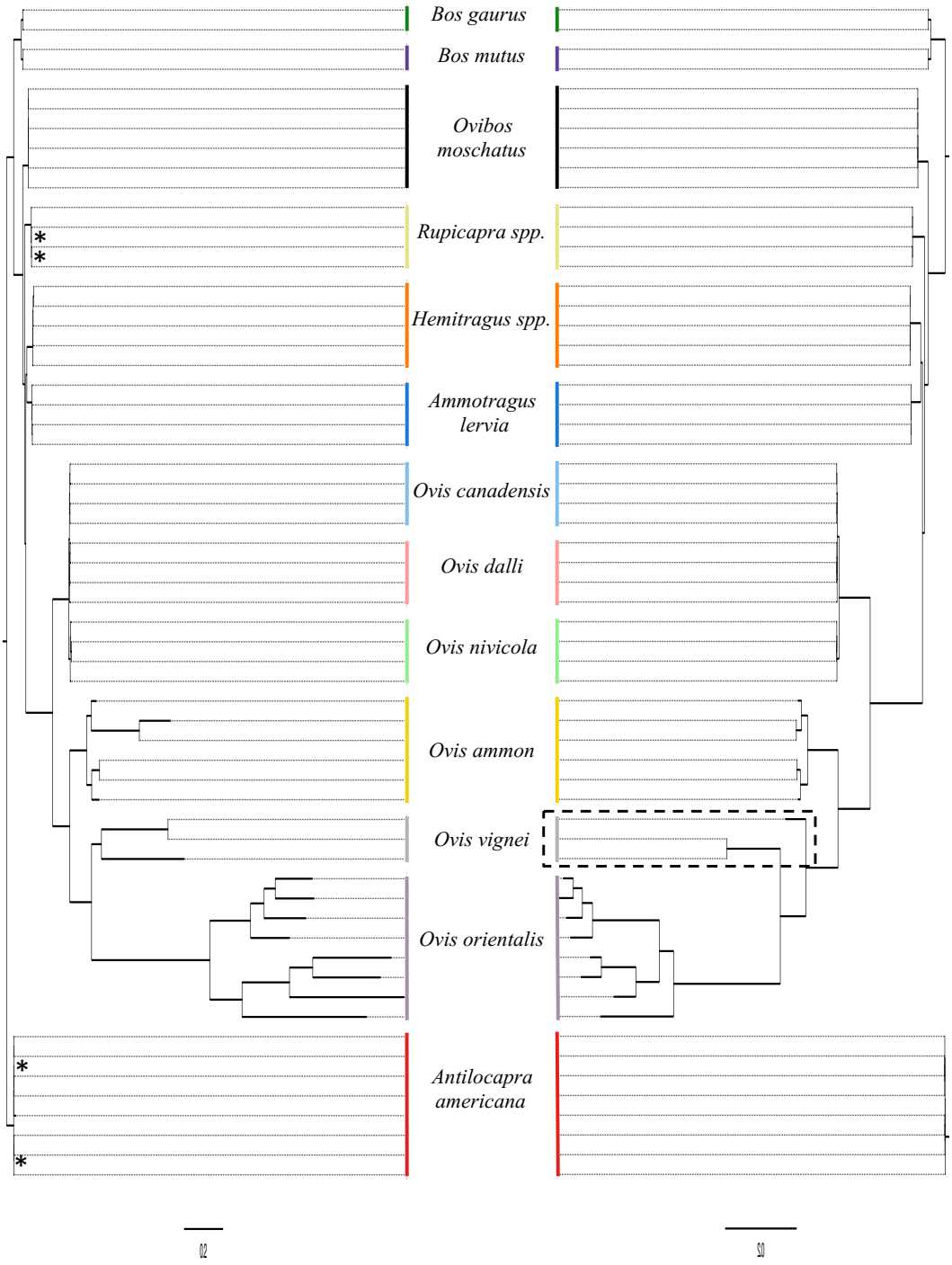

Figure 2.1 Maximum likelihood phylogenetic tree of 13 wild ungulate species rooted with Antilocapridae as outgroup. The tree on the left represent the tree generated using the morphological dataset and right represent that of the IUPAC dataset.* denotes branches that do not have >70% bootstrap support. Box drawn with dashed line represent paraphyly in *O vignei* in the tree generated using the IUPAC coding scheme. Coloured taxa labels are consistent throughout all figures

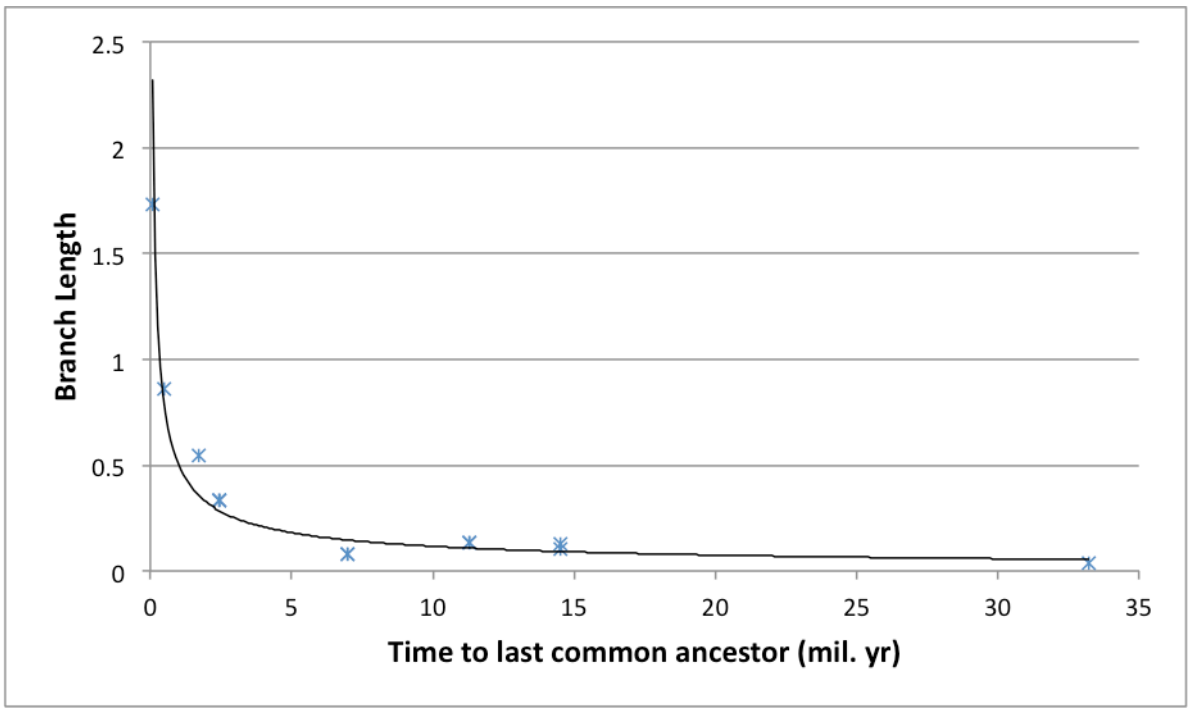

Figure 2.2 Branch length plotted as a function of time to last common ancestor. A non-linear weighted regression line is shown (formula for the line: $y = 0.5021x^{-0.635}$). Time to last common ancestor was referenced from Miller et al. 2012a.

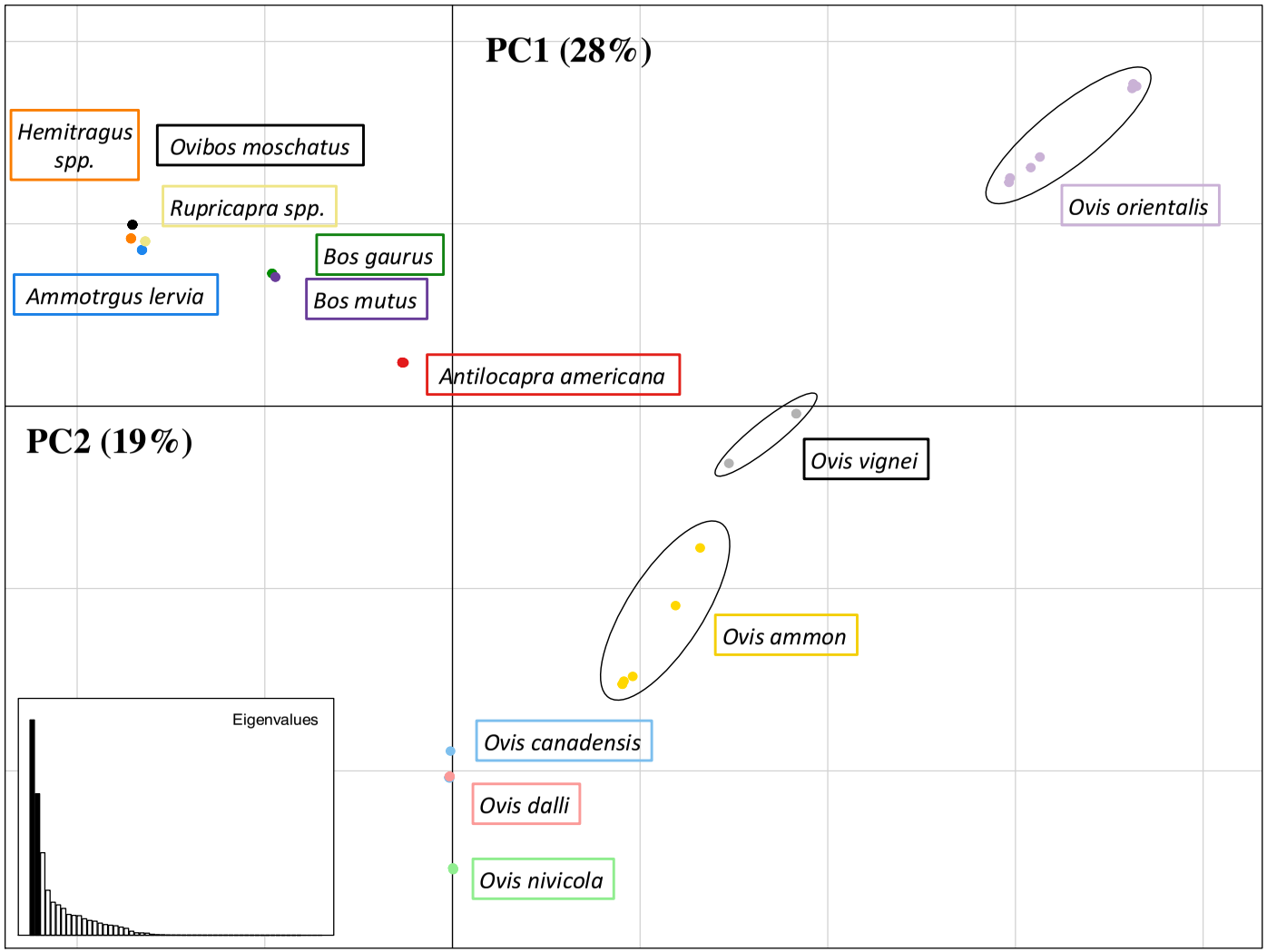

Figure 2.3 PCA plot of all 13 ungulate taxa for PC1 and PC2 (percent variance explained in parentheses). Coloured taxa labels are consistent throughout all figures.

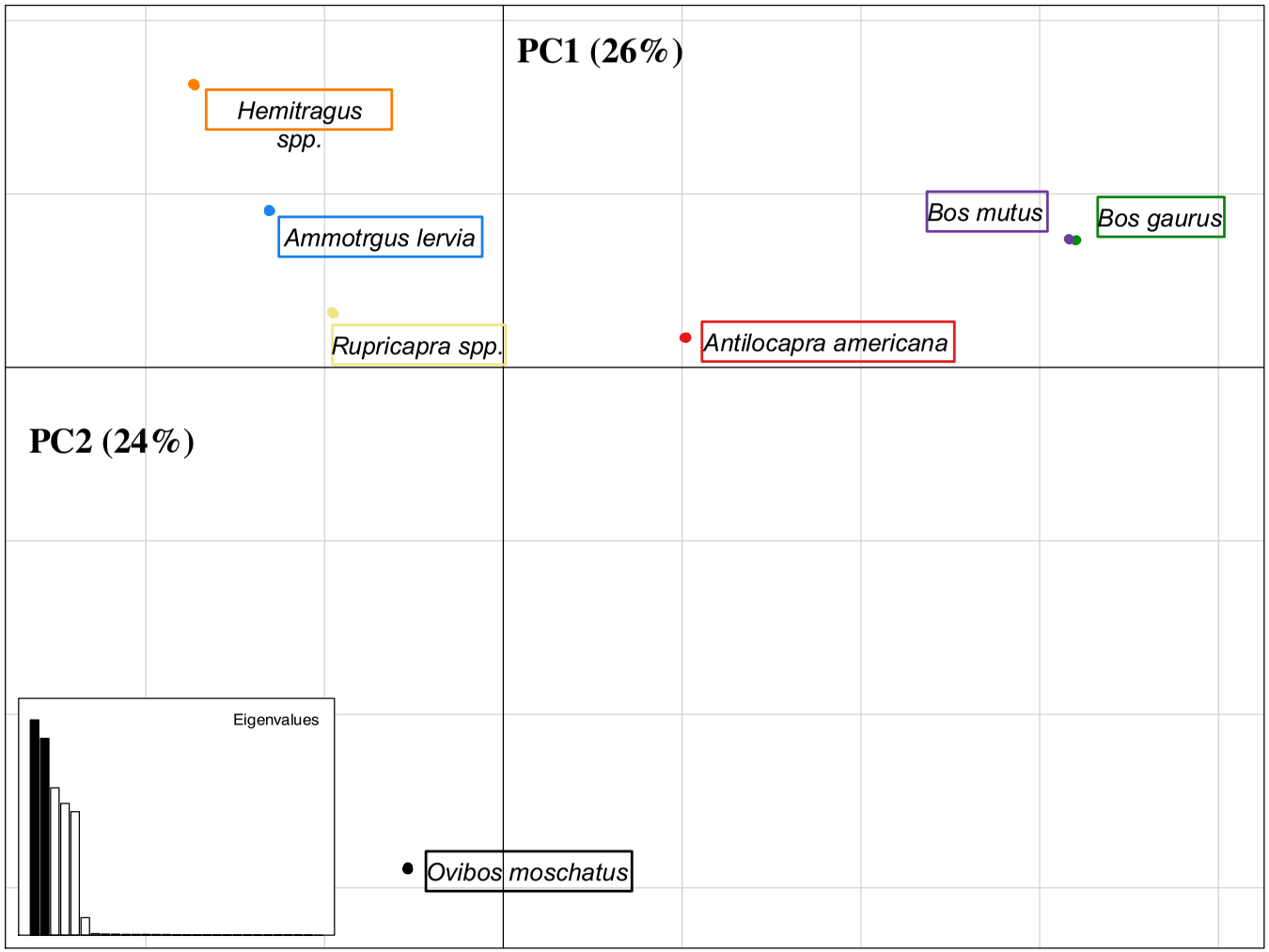

Figure 2.4 PCA without genus *Ovis* for PC1 and PC2 (percent variance explained in parentheses). Coloured taxa labels are consistent throughout all figures

2.6 References

Alberta WAo RSA 2000. *Alberta Queen's printer. Retrieved from* https://www.canlii.org/en/ab/laws/stat/rsa-2000-c-w-10/134378/rsa-2000-c-w-10.html.

Allendorf FW, Hohenlohe PA, Luikart G (2010) Genomics and the future of conservation genetics. *Nat Rev Genet* **11**, 697-709.

Cornuet J-M, Piry S, Luikart G, Estoup A, Solignac M (1999) New Methods Employing Multilocus Genotypes to Select or Exclude Populations as Origins of Individuals. *Genetics* **153**, 1989.

Davey JW, Hohenlohe PA, Etter PD*, et al.* (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics* **12**, 499-510.

Decker JE, Pires JC, Conant GC*, et al.* (2009) Resolving the evolution of extant and extinct ruminants with high-throughput phylogenomics. *Proceedings of the National Academy of Sciences* **106**, 18644-18649.

Deniskova TE, Okhlopkov IM, Sermyagin AA*, et al.* (2016) Whole genome SNP scanning of snow sheep (Ovis nivicola). *Doklady Biochemistry and Biophysics* **469**, 288-293.

Dray S, Dufour A-B (2007) The ade4 Package: Implementing the Duality Diagram for Ecologists. *2007* **22**, 20.

Garvin MR, Saitoh K, Gharrett AJ (2010) Application of single nucleotide polymorphisms to non-model species: a technical review. *Molecular Ecology Resources* **10**, 915-934.

Hassanin A, Douzery EJP (1999) The Tribal Radiation of the Family Bovidae (Artiodactyla) and the Evolution of the Mitochondrial Cytochrome b Gene. *Molecular Phylogenetics and Evolution* **13**, 227-243.

Hassanin A, Ropiquet A (2004) Molecular phylogeny of the tribe Bovini (Bovidae, Bovinae) and the taxonomic status of the Kouprey, Bos sauveli Urbain 1937. *Molecular Phylogenetics and Evolution* **33**, 896-907.

Haynes GD, Latch EK (2012) Identification of novel single nucleotide polymorphisms (SNPs) in deer (*Odocoileus* spp.) using the BovineSNP50 BeadChip. *Plos One* **7**, e36536.

Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**, 1403-1405.

Kijas JW, Lenstra JA, Hayes B*, et al.* (2012) Genome-Wide Analysis of the World's Sheep Breeds Reveals High Levels of Historic Mixture and Strong Recent Selection. *PLoS Biol* **10**, e1001258.

Kijas JW, Porto-Neto L, Dominik S*, et al.* (2014) Linkage disequilibrium over short physical distances measured in sheep using a high-density SNP chip. *Animal Genetics* **45**, 754-757.

Kohn MH, Murphy WJ, Ostrander EA, Wayne RK (2006) Genomics and conservation genetics. *Trends in Ecology & Evolution* **21**, 629-637.

Leaché AD, Banbury BL, Felsenstein J, Nieto-Montes de Oca A, Stamatakis A (2015) Short Tree, Long Tree, Right Tree, Wrong Tree: New Acquisition Bias Corrections for Inferring SNP Phylogenies. *Systematic Biology*.

Mason-Gamer RJ, Kellogg EA (1996) Testing for Phylogenetic Conflict Among Molecular Data Sets in the Tribe Triticeae (Gramineae). *Systematic Biology* **45**, 524-545.

Meirmans PG, Van Tienderen PH (2004) genotype and genodive: two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes* **4**, 792-794.

Merckx VSFT, Kissling J, Hentrich H*, et al.* (2013) Phylogenetic relationships of the mycoheterotrophic genus Voyria and the implications for the biogeographic history of Gentianaceae. *American Journal of Botany* **100**, 712-721.

Miller JM, Festa-Bianchet M, Coltman DW (2018) Genomic analysis of morphometric traits in bighorn sheep using the Ovine Infinium® HD SNP BeadChip. *PeerJ* **6**, e4364.

Miller JM, Kijas JW, Heaton MP, McEwan JC, Coltman DW (2012) Consistent divergence times and allele sharing measured from cross-species application of SNP chips developed for three domestic species. *Molecular Ecology Resources* **12**, 1145-1150.

Miller JM, Poissant J, Kijas JW, Coltman DW, Consortium ISG (2011) A genome-wide set of SNPs detects population substructure and long range linkage disequilibrium in wild sheep. *Molecular Ecology Resources* **11**, 314-322.

Paetkau D, Calvert W, Stirling I, Strobeck C (1995) Microsatellite analysis of population structure in Canadian polar bears. *Molecular Ecology* **4**, 347-354.

Paetkau D, Slade R, Burden M, Estoup A (2003) Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power. *Molecular Ecology* **13**, 55-65.

Pertoldi C, W√≥jcik J, Tokarska MÇ*, et al.* (2010) Genome variability in European and American bison detected using the BovineSNP50 BeadChip. *Conservation Genetics* **11**, 627-634.

Purcell S, Neale B, Todd-Brown K*, et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics* **81**, 559-575.

Rezaei HR, Naderi S, Chintauan-Marquier IC*, et al.* (2010) Evolution and taxonomy of the wild species of the genus Ovis (Mammalia, Artiodactyla, Bovidae). *Molecular Phylogenetics and Evolution* **54**, 315-326.

Ropiquet A, Hassanin A (2005) Molecular phylogeny of caprines (Bovidae, Antilopinae): the question of their origin and diversification during the Miocene. *Journal of Zoological Systematics and Evolutionary Research* **43**, 49-60.

Scheunert A, Heubl G (2014) Diversification of Scrophularia (Scrophulariaceae) in the Western Mediterranean and Macaronesia – Phylogenetic relationships, reticulate evolution and biogeographic patterns. *Molecular Phylogenetics and Evolution* **70**, 296-313.

Schwarz G (1978) Estimating the Dimension of a Model. *Ann. Statist.* **6**, 461-464.

Seelanan T, Schnabel A, Wendel JF (1997) Congruence and Consensus in the Cotton Tribe (Malvaceae). *Systematic Botany* **22**, 259-290.

Sim Z, Hall JC, Jex B, Hegel TM, Coltman DW (2016) Genome-wide set of SNPs reveals evidence for two glacial refugia and admixture from postglacial recolonization in an alpine ungulate. *Molecular Ecology* **25**, 962-1083.

Slate J, Gratten J, Beraldi D*, et al.* (2009) Gene mapping in the wild with SNPs: guidelines and future directions. *Genetica* **136**, 97-107.

Trifinopoulos J, Nguyen L-T, von Haeseler A, Minh BQ (2016) W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Research* **44**, W232-W235.

Wenne R, Drywa A, Kent M, Sundsaasen KK, Lien S (2016) SNP Arrays for Species Identification in Salmonids. In: *Marine Genomics: Methods and Protocols* (ed. Bourlat SJ), pp. 97-111. Springer New York, New York, NY.

Zhou H, Wu D, Chen R*, et al.* (2014) Developmental validation of a forensic rapid DNA-STR kit: Expressmarker 16. *Forensic Science International: Genetics* **11**, 31-38.

Chapter 3: Genome-wide set of SNPs reveals evidence for two glacial refugia and admixture from postglacial recolonization in an alpine ungulate

3.1 Abstract

Past glaciation events have played a major role in shaping the genetic diversity and distribution of wild sheep in North America. The advancement of glaciers can isolate populations in ice-free refugia, where they can survive until the recession of ice sheets. The major Beringian refugium is thought to have held thinhorn sheep (*Ovis dalli*) populations during times of glacial advance. While isolation in the major refugium can account for much of the genetic and morphological diversity seen in extant thinhorn sheep populations, mounting evidence suggest the persistence of populations in smaller minor refugia. We investigated the refugial origins of thinhorn sheep using ~10,000 SNPs obtained via a cross species application of the domestic sheep ovine HD BeadChip to genotype 52 thinhorn sheep and five bighorn sheep (*O. canadensis*) samples. Phylogenetic inference revealed a distinct lineage of thinhorn sheep inhabiting British Columbia, which is consistent with the survival of a group of thinhorn sheep in a minor refugium separate from the Beringian refugium. Isolation in separate glacial refugia likely mediated the evolution of the two thinhorn sheep subspecies, the white Dall's sheep (*O. d. dalli*), which persisted in Beringia, and the dark Stone's sheep (*O. d. stonei*), which utilized the minor refugium. We also found the first genetic evidence for admixture between sheep from different glacial refugia in south-central Yukon as a consequence of post-glacial expansion and recolonization. These results show that glaciation events can have a major role in the evolution of species inhabiting previously glaciated habitats and the need to look beyond established refugia when examining the evolutionary history of such species.

3.2 Introduction

The continent of North America was repeatedly glaciated during the Pleistocene age. The advance and retreat of the ice sheets have large-scale effects that dominate the evolutionary history of many North American species (Avise 2000). The advancement of glaciers can isolate populations in ice-free refugia, where they can embark on different evolutionary trajectories as a result of restricted gene flow and altered selective regimes (Hewitt 2000; Hewitt 1996). During

the Pleistocene glacial periods, two major refugia are thought to have existed in western North America: 1) the Beringian refugium, and 2) southern refugium, (generally south of the Canada-USA border) (Pielou 1991). While isolation in the major refugia can account for much of the genetic and morphological diversity seen in many extant North American plant and animal species such as the rockcress, *Boechara* spp. (Dobeš *et al.* 2004), American pika, *Ochotona princeps*, (Galbreath *et al.* 2009), and flying squirrel, *Glaucomys sabrinus*, (Bidlack & Cook 2001), mounting evidence suggests the persistence of populations in smaller, minor refugia (Shafer *et al.* 2010).

The pattern in which genetic variation is geographically distributed within a species is dependent upon the combined, and sometimes competing, effects of historical vicariance and contemporary dispersal (Zink *et al.* 2000). In vagile species, high levels of gene flow can mask the signatures of vicariance while low dispersal rates seen in more philopatric organisms can facilitate the retention of historical patterns of spatial genetic variation. Alpine ungulates are excellent candidates to study the effects of glacial vicariance since they have specific habitat requirements that limit dispersal, thus favouring the retention of signals from historical events. Previous work on the mountain goat (*Oreamnos americanus*) using both mitochondrial and microsatellite DNA supported two Pleistocene refugia, which contradicts fossil evidence that points to survival in only a single southern refugium (Shafer *et al.* 2011).

Despite its wide distribution in northwestern North America, the thinhorn sheep (*Ovis dalli*) has very specific habitat needs, thus restricting it to precipitous terrain in alpine and sub-alpine environments (Geist 1971; Valdez & Krausman 1999). Like other alpine species such the American pika and mountain goat (Henry *et al.* 2012; Shafer *et al.* 2011), thinhorn sheep avoid boreal habitat found at lower elevations (Valdez & Krausman 1999). Therefore, thinhorn sheep populations are often thought to be localized on "sky islands" (Shafer *et al.* 2011), which are naturally fragmented and result in very limited gene flow among populations (Worley *et al.* 2004). This philopatric nature makes the thinhorn sheep a good candidate to investigate the possible presence of alternative or minor refugia and patterns of post-glacial re-colonization.

The thinhorn sheep is an alpine ungulate discontinuously distributed throughout the mountains of northwestern North America (Figure 1), from Alaska, east to the Mackenzie Mountains of the Northwest Territories and south through the Yukon to northern British Columbia (BC) (Valdez & Krausman 1999). There are two traditionally recognized subspecies of thinhorn sheep based on pelage colour: 1) the white coated and more northerly distributed Dall's sheep (*O. d. dalli*), and 2) the dark coated and more southerly distributed Stone's sheep (*O. d. stonei*) (Valdez & Krausman 1999). The transition between Dall's and Stone's sheep occurs in a roughly clinal fashion over a contact zone containing sheep of intermediate coat colouration that extends from the Ogilvie Mountains in Yukon south to northern end of the Cassiar Mountains in BC (Demarchi & Hartwig 2004; Sheldon 1911). The validity of colour-based subspecies delineation is subject to some debate since no clean break exists between subspecies and degree of colouration is variable within populations (Loehr *et al.* 2006; Ramey 1993). Therefore, variation in coat colour might represent variation across a species range rather than being diagnostic of subspecific relationships. Genetic studies in other wild sheep species have also found morphology-based taxonomic groupings to be inconsistent with genetic evidence (Gutiérrez-Espeleta *et al.* 2000).

The thinhorn sheep has traditionally been thought to have survived only in the major Beringian refugium during the last glacial maximum, and subsequent to glacial retreat, re-colonized North America by moving south and east as the climate warmed and new habitats became available (Geist 1971). However, a recent mitochondrial DNA (mtDNA) based study (Loehr *et al.* 2006) showed evidence that a small group of thinhorn sheep might have also utilized a minor refugium located in northwestern BC, an area that remained ice-free during last glacial maximum due to the asynchronous advances of the Cordilleran and Laurentide ice sheets (Figure 1) (Catto *et al.* 1996).

By sequencing a 604 bp portion of the mtDNA control region, Loehr *et al.* (2007) found sheep that occupied BC possessed divergent haplotypes compared to sheep from other parts of the species range as well as a star-shaped haplotype network, which is consistent with a population bottleneck caused by isolation in a separate glacial refugium. Nuclear microsatellite markers also

show strong population structure between sheep from British Columbia and adjacent populations in southern Yukon (Worley *et al.* 2004). While Loehr *et al.* (2006) raised the possibility of the utilization of two Pleistocene refugia by thinhorn sheep, this hypothesis has yet to be tested using nuclear DNA or phylogenetic methods. Questions also remain about the manner of postglacial range expansion and potential admixture between individuals from different refugia.

In this study, we shed new light on the phylogeography of thinhorn sheep by using a genome-wide set of single nucleotide polymorphisms (SNPs). We used phylogenetic inference to evaluate the hypothesis that thinhorn sheep survived in two refugia during the last glaciation. We also investigated the nature of post-glacial recolonization using population genetic clustering and admixture analysis of relevant phylogenetic groups. If thinhorn sheep survived in two refugia, we expect a basal phylogenetic divergence between sheep that survived in the major Beringian refugium and sheep that persisted in the minor refugium. If, however, thinhorn sheep utilized only the Beringian refugium, we expect individuals sampled further away from the refugium to be nested within those that are more proximal (Brunsfeld *et al.* 2001). Given the smaller geographical extent of the minor refugium, we may expect to see lower levels of genetic diversity among sheep descended from individuals that occupied the minor refugium. Finally, we expect subspecies relationships in *O. dalli* to be reflective of refugial origins. In the single refugium scenario, *O. d. stonei* individuals should be nested within *O. d. dalli* individuals in the phylogenetic tree since the *O. d. dalli* range overlaps the historical boundary of the Beringian refugium while *O. d. stonei* are found further south in the species range. In the two refugia scenario, we expect a basally-positioned bifurcation that separates *O. d. dalli* and *O. d. stonei* individuals.

3.3 Methods and Materials

3.3.1 Sampling

We acquired 52 muscle (n=42) and blood samples (n=10) from six areas across the thinhorn sheep range, from northwest Alaska to northern BC, collected between 2001 and 2013 (Table 3.1; Figure 3.1). Spatial resolutions for samples are based on their jurisdictional origins. Samples from Alaska, BC, and Northwest Territories are geo-referenced while sample localities from

Yukon are resolved to the game management subzone level (Table 3.1; Figure 3.1). According to current subspecies maps (Demarchi & Hartwig 2004), samples from Alaska, Northwest Territories, northern Yukon, and southwest Yukon are from the Dall's sheep range while south-central Yukon and BC are from the Stone's sheep range. Muscle samples were collected during regulatory inspections of legally harvested rams. DNA from muscle tissue of five bighorn sheep (*Ovis canadensis*) originating from Ram Mountain, Alberta, Canada (n=4) and National Bison Range, Montana, USA (n=1) were used as an outgroup in subsequent analyses.

3.3.2 Molecular techniques and quality control

DNA extraction was performed using the Qiagen DNeasy Blood and Tissue Kit (Qiagen) following the manufacturer's protocol. Extracted DNA was quantified using the Qubit Fluorometer (Life Technologies) and normalized to 50ng/μL prior to genotyping. SNP genotyping was performed using the ovine HD BeadChip (Illumina) developed by the International Sheep Genomics Consortium (Kijas *et al.* 2014). Using the ovine HD BeadChip, samples were interrogated at 606,006 biallelic SNP loci identified from 41 breeds of domestic (*Ovis aries*) and wild sheep (including the bighorn and thinhorn sheep) using a combination of Sanger and next-generation sequencing. Raw signal intensities were converted into genotype calls using the GENOMESTUDIO software (Illumina).

Genotyping reliability was assessed by referencing GenTrain (GT) scores, a measure of genotype cluster quality, for each locus (Johnston *et al.* 2011). Loci with GT scores <0.6 were removed from the dataset. This criterion is consistent with previous studies that have applied domestic SNP arrays to wild species (Haynes & Latch 2012; Miller *et al.* 2012a; Miller *et al.* 2011). We also removed loci and individuals that had genotyping rates lower than 90%.

3.3.3 Phylogenetic inferences

Phylogenetic analyses were performed using maximum parsimony (MP) and Bayesian methods. MP inference was performed with unordered and equally weighted characters (Fitch 1971) using the software PAUP* 4.0b10 (Swofford 2003). A heuristic search was undertaken with the following search parameters: 100 random addition replicates, 10 trees held in each stepwise

addition, tree bisection and reconnection branch swapping and no tree discarded between replicates. Nodal support was evaluated using 1000 bootstrap replicates under the same parameters as the initial tree search (Felsenstein 1985).

Optimal nucleotide substitution model selection for Bayesian inference was determined using the Akaike Information Criterion (AIC) (Akaike 1974) as implemented in the software JMODELTEST2 (Darriba *et al.* 2012). The optimal model selected was GTR+Γ (Tavaré 1986), which allows for varying substitution rates and rate heterogeneity between sites. Two independent Bayesian runs with four Markov chain Monte Carlo (MCMC) chains per run, 10,000,000 generations, 25% (2,500,000) burn-in, uniform priors, and GTR+Γ substitution model were implemented in the program MRBAYES 3.2.2 (Ronquist & Huelsenbeck 2003). Substitution rates and rate heterogeneity were allowed to vary by chromosome**.** Stationarity was assessed using the following criteria: 1) average deviation of split frequencies < 0.01, 2) effective sampling size > 200 (Rambaut *et al.* 2014), 3) good trace plot mixing and flat trajectory, and 4) potential scale reduction factor (PSRF) approaching 1. Branch support was assessed by examining the posterior probabilities of each node in the post-burn-in consensus tree.
To ensure that phylogenetic signals are robust to SNP genotype coding, each phylogenetic inference was performed using each of two SNP coding schemes: 1) each nucleotide/allele in a SNP genotype is treated as a character and the dataset is generated by concatenating all SNP nucleotides (e.g. for three genotypes AA, GC, AT, the corresponding concatenation would be AAGCAT), and 2) each SNP genotype was coded by a unique character state (i.e. AA=0, AB=1, BB=2). The conversion of nucleotide SNP data to their respective coding schemes was performed using the program PLINK (Purcell *et al.* 2007). Hard incongruencies in tree topologies from the two coding schemes were assessed using the Templeton test, which implements a Wilcoxon signed-rank test to determine if one tree topology is significantly more parsimonious than another (Templeton 1983).

3.3.4 Genetic diversity

We calculated Nei's unbiased estimator of genetic diversity ($\hat{H}_S$) (Nei 1987) and effective number of alleles (A_E)(Nagylaki 1985) for each group identified by phylogenetic methods: 1)

Northern-West, 2) Northern-East, and 3) Southern Clades (see Results), using GENODIVE (Meirmans & Van Tienderen 2004).

3.3.5 Bayesian clustering and admixture analysis

We investigated the presence and degree of admixture between lineages that utilized separate glacial refugia with the Bayesian clustering software STRUCTURE 2.3.4 (Pritchard *et al.* 2000). We performed 20 independent runs for $K = 1$ to $K = 6$ (where K is the number of clusters) with 300,000 replicates and 100,000 burn-in replicates as recommended by Gilbert *et al.* (2012). We used no location priors and the correlated allele frequencies model. Runs were permuted and averaged using the LargeKGreedy algorithm (2000 random input order repeats) in CLUMPAK 1.1 (KOPELMAN *ET AL.* 2015). Support for K-values was assessed using the Evanno method as implemented in STRUCTURE HARVESTER 0.6.94 (Earl & vonHoldt 2012). Individual membership coefficients from STRUCTURE were graphically displayed using Microsoft Excel for Mac 2011 14.5.2 (Microsoft) and Adobe Illustrator CS 5 (Adobe).

3.4 Results

3.4.1 Genotyping efficiency and quality control

In total, 57 individuals (*O. dalli* = 52, *O. canadensis* =5) were genotyped at 606,006 loci with genotyping success of over 95% for all samples and loci. Of the 606,006 loci interrogated, 9,879 were found to be polymorphic and were retained for subsequent analysis. This level of polymorphism is consistent with previous studies that undertook cross-species amplifications of genomic SNP arrays developed for domestic species (Haynes & Latch 2012; Miller *et al.* 2012a; Miller *et al.* 2011).

3.4.2 Phylogenetic Inference

Both MP and Bayesian methods strongly support (with 100% bootstrap and posterior probabilities, respectively) a basal, reciprocally monophyletic relationship between northern and southern groups of thinhorn sheep, hereupon referred to as the Northern and Southern Clades, respectively (Figure 3.2). The monophyletic Southern Clade consists of all but one individual from the BC sample area and is generally south of the BC - Yukon border with the rest of the

individuals making up the Northern Clade. Bayesian phylogenetic inference also supports the subdivision of the Northern Clade into two groups, hereupon referred to as the Northern-West and the Northern-East clusters (Figure 3.2). The Northern-East cluster is paraphyletic with respect to the monophyletic Northern-West cluster (Figure 3.2). Sheep in the Southern Clade are also those geographically closest to the minor refugium proposed by Catto *et al.* (1996) in northeastern BC.

Templeton tests comparing hard incongruencies in tree topologies from different coding schemes were non-significant ($p>0.05$), thus indicating congruency in phylogenetic signals from the two SNP coding schemes.

3.4.3 Genetic diversity

We found genetic diversity to be lowest in the Southern Clade ($\hat{H}_S = 0.174$; $A_E = 1.277$), intermediate in the Northern-East cluster ($\hat{H}_S = 0.188$; $A_E = 1.296$), and highest in the Northern-West cluster ($\hat{H}_S = 0.198$; $A_E = 1.311$).

3.4.4 Bayesian clustering and admixture analysis.

The Evanno ΔK method indicated best support for K =2 (Figure S3.1, Supporting Information). At K = 2, cluster identities reflect the Northern Clade - Southern Clade division identified by phylogenetic inference (Figure 3.3). While the majority of individuals can be strongly assigned (>75%) to their respective clades, sheep in the Northern-East cluster, which belongs to the Northern Clade, were consistently admixed. Sheep from the Northern-East cluster shared 28.0% - 58.7% of their ancestry with sheep from the Southern Clade. The level of admixture in other individuals generally decreased with increasing geographical distance from the Northern-East cluster. In addition, there was also reasonable support for K = 3, which showed clustering of individuals from Noatak National Preserve and Gates of the Arctic National Preserve locations, both in the Brooks Range of northwestern Alaska (Figure S3.2,Information).

3.5 Discussion

3.5.1 Refugial origins

In this study, we examined the range-wide phylogeographic patterns of thinhorn sheep in North America by interrogating a genome-wide set of SNPs. Our results provide robust support for the existence of two Pleistocene glacial refugia for thinhorn sheep. Multiple lines of evidence support this double refugia scenario. Both MP and Bayesian phylogenetic methods provided strong support for the North-South split as the deepest division among individuals sampled from across the species distribution (Figure 3.2). The Southern Clade also consists entirely of one subspecies, Stone's sheep (*O. d. stonei*). Bayesian clustering also supports the North-South split, with cluster assignments that reflect the Northern and Southern Clades identified by phylogenetic methods (Figure 3.3). These results strongly suggest independent evolutionary histories between sheep belonging to the Southern Clade and the Northern Clade thinhorn sheep, which occupies the northern parts of the species distribution (Figure 3.2). We argue that this independence is likely a result of glacial-induced vicariance, with the Southern Clade confined to the minor refugium in northeast BC, away from the rest of the species, which occupied the Beringian refugium. The North-South split was also detected in a previous mtDNA based study on thinhorn sheep (Loehr *et al.* 2006), which found sheep that occupied the geographical range of the Southern Clade possess divergent haplotypes. Other similarly distributed mountain ungulates also exhibit similar North-South bifurcations that suggest two Pleistocene glacial refugia (Latch *et al.* 2009; Shafer *et al.* 2011).

Further evidence for the double refugia scenario may also be seen in the patterns of genetic diversity among the groups. Individuals from the Southern Clade were found to have the lowest levels of genetic diversity while those in the Northern-West cluster had the highest, as measured by both $\hat{H}_S$ and A_E. Worley *et al.* (2004) found similar patterns of genetic variation using 12 microsatellites. The reduced levels of genetic diversity observed in the Southern Clade may be due to the smaller geographical extent of the minor refugium, which may have led to a stronger bottleneck effect in these sheep compared to those in the relatively larger Beringian refugium. Similarly depressed genetic diversity can also seen in island bighorn sheep populations when compared to that of mainland populations (Gasca-Pineda *et al.* 2013). Contemporary

observations in other alpine species also tend to find that small isolated populations possess less genetic diversity (Henry *et al.* 2012).

While we did not perform molecular dating in this study, existing literature supports the placement of the North-South spilt in the late Pleistocene. First, the divergence between thinhorn and bighorn sheep was found to have occurred 0.97±0.37 MYA, during the Middle Pleistocene, using one mitochondrial and four nuclear genes (Gradstein & Ogg 2012; Rezaei *et al.* 2010). This date allows us to bracket any intraspecific divergence in the thinhorn sheep to be after that date. Molecular dating performed by Loehr *et al.* (2006) using a 604 bp portion of the mtDNA control region on thinhorn sheep from a similar range as this study dates the North-South split to 219,852 years before present. This divergence time predates the beginnings of the last glacial period, (~150,000 years ago) and is consistent with the scenario of refugia mediated divergence. Dating estimates from Loehr *et al.* (2006) also agree with estimates for mountain goats, which are thought to have also utilized the Catto *et al.* (1996) minor refugium (Shafer *et al.* 2011). Both the present study and Loehr *et al.* (2006) contradict the previous hypothesis that postulated the survival of thinhorn sheep in only the Beringian refugium (Geist 1971). If all thinhorn survived in a single refugium and post-glacial recolonization proceeded southwards from a single origin, we would expect individuals from sampling areas south of the Beringian refugium to be nested within populations from the north (Brunsfeld *et al.* 2001). Survival of thinhorn sheep in the major southern refugium is unlikely since the southern extent of the thinhorn sheep range does not extend beyond the proposed southern limits of the last glacial maximum and thus does not overlap the boundaries of the southern refugium (Tomasik & Cook 2005). Current thinhorn sheep distribution also does not include areas considered to have been part of the major southern refugium (Demarchi & Hartwig 2004; Dyke & Prest 1987).

3.5.2 Post-glacial re-colonization and admixture

Structure analysis revealed strong cluster assignments that correspond to phylogenetic clade identities for most individuals (Figure 3.3). The strong cluster assignment is likely a result of isolation in separate glacial refugia. A notable exception exists for sheep from the Northern-East cluster, which were found to be consistently admixed (Figure 3.3). The Northern-East cluster

also lies in the path of a major post-glacial recolonization route in northwestern North America (Shafer *et al.* 2010). Therefore, it is likely that the level of admixture seen in the Northern-East cluster is due to interbreeding of individuals with different refugial origins as a result of post-glacial range expansion from their respective refugia. This is the first study to reveal the nature of post-glacial recolonization in the thinhorn sheep and the first reported instance of an admixture zone between lineages of thinhorn sheep from different glacial refugium.

3.5.3 Morphology

Morphological distinction also lends support to the Northern-East cluster being a zone of hybridization. While sheep from the Northern-West cluster and the Southern Clade are consistently white (Figure 3.4a) and dark coloured (Figure 3.4b), respectively, sheep from the Northern-East are known to exhibit coat colour intergradation from dark coloured to almost completely white except for a dark tail (Figure 3.4c/d). We argue this morphological distinction is likely due to the hybrid origins of individuals from the Northern-East cluster.

3.5.4 Taxonomic implications

Under current subspecies designations, sheep from both the Southern Clade and Northern-East cluster belong to the Stone's sheep subspecies, *O. d. stonei* (Demarchi & Hartwig 2004). This designation is largely based on pelage colour; thinhorn sheep that are not completely white are classified as Stone's sheep while white sheep are classified as Dall's sheep, *O d. dalli* (Valdez & Krausman 1999). Our results show current taxonomic designations do not accurately reflect the evolutionary history of the species. Phylogenetic and clustering methods both reveal that sheep from the Northern-East cluster are more closely related to sheep from the Northern-West cluster (both in the Northern Clade) than that of the Southern Clade. Therefore, it is more appropriate to designate sheep from the Northern and Southern Clades each as their own subspecies i.e. Dall's and Stone's respectively. The view that current subspecific designations are inappropriate is also consistent with previous studies. Loehr *et al.* (2006) found patterns of mtDNA haplotype divergence to be unreflective of subspecific designations. Nuclear microsatellite evidence also shows the genetic differentiation between BC populations (Southern Clade in our study) and

populations from the rest of the species range to be the most substantial among all comparisons (Worley *et al.* 2004).

There are a number of possible explanations for this inconsistency. First, current delimitations may be based on inappropriate morphological characters. Subspecific designations of thinhorn sheep are largely based on pelage colour (Valdez & Krausman 1999). Stone's sheep are generally regarded to be of dark colouration and occupy the southerly reaches of the thinhorn sheep distribution in BC and north to south-central Yukon, while Dall's sheep are white and occupy Alaska, southwest and northern Yukon and the Mackenzie Mountains of the Northwest Territories. However, it is well known that thinhorn sheep coat colour consists of a complete colour intergradation between Stone's and Dall's sheep (Loehr *et al.* 2008; Sheldon 1911). Results from this study show that current criteria of color or geographical distribution do not consistently assign individuals to their respectively positions as described by phylogenetic inferences (Figure 3.2). Putative Stone's sheep individuals from south-central Yukon (Northern-east cluster) are also more closely related to Dall's sheep to the north and west, which means they are currently misclassified and should more accurately be designated as Dall's sheep. Admixture analysis also shows individuals from the Northern-East cluster consistently shared ancestry with sheep from the Southern Clade (Figure 3.3). This suggests that morphological commonalities between the Northern-East cluster and Southern Clade may be due to hybridization between the two groups.

3.6 Conclusions

Alpine species, with their low dispersal rates, tend to retain the legacy of historical vicariance, making them ideal study subjects to investigate the effects of past glaciation events on present day distribution of genetic diversity (Henry *et al.* 2012; Loehr *et al.* 2006; Worley *et al.* 2004). We have shown the effect of survival in two refugia on the phylogeographical patterns of the thinhorn sheep. We also show that admixture events following postglacial range expansion can have profound morphological effects, making subspecific assignment difficult. The results of this study highlight the intricate role glaciation events can have on the evolutionary history of a species and the need to look beyond established refugia.

Table 3.1 Sampling localities, jurisdictional origins (US state or Canadian Province/Territory), abbreviations used for each sample locality, the number of samples for each sample locality and Clade/cluster designation for each sample used in this study. NWT = Northwest Territories. BC = British Columbia

Sampling Locality	Jurisdiction	Abbrev.	N	Clade/Cluster
Gates of the Arctic National Preserve	Alaska	GAAR	3	Northern-West
Noatak National Preserve	Alaska	NNP	3	Northern-West
Central Alaskan Range	Alaska	CAR	2	Northern-West
Yukon-Charley Rivers National Preserve	Alaska	YUCH	2	Northern-West
Arctic Red River Basin	NWT	AR	1	Northern-West
Yukon game management zone 240	Yukon	YK240	1	Northern-West
Yukon game management zone 279	Yukon	YK279	1	Northern-West
Yukon game management zone 507	Yukon	YK507	1	Northern-West
Yukon game management zone 520	Yukon	YK520	1	Northern-West
Yukon game management zone 532	Yukon	YK532	1	Northern-West
Yukon game management zone 545	Yukon	YK545	1	Northern-West
Yukon game management zone 550	Yukon	YK550	2	Northern-West
Yukon game management zone 702	Yukon	YK702	1	Northern-West
Yukon game management zone 720	Yukon	YK720	1	Northern-West
Yukon game management zone 724	Yukon	YK724	1	Northern-West
Yukon game management zone 730	Yukon	YK730	2	Northern-West
Yukon game management zone 903	Yukon	YK903	1	Northern-West
Yukon game management zone 416	Yukon	YK416	1	Northern-East
Yukon game management zone 440	Yukon	YK440	1	Northern-East
Yukon game management zone 445	Yukon	YK445	1	Northern-East
Yukon game management zone 812	Yukon	YK812	3	Northern-East
Yukon game management zone 818	Yukon	YK818	1	Northern-East
Yukon game management zone 819	Yukon	YK819	2	Northern-East
Yukon game management zone 820	Yukon	YK820	2	Northern-East
Yukon game management zone 1001	Yukon	YK1001	1	Northern-East
Yukon game management zone 1005	Yukon	YK1005	1	Northern-East
Yukon game management zone 1015	Yukon	YK1015	1	Northern-East
Yukon game management zone 1021	Yukon	YK1021	1	Northern-East
BC game management zone 620	BC	BC620	2	Southern Clade
BC game management zone 623	BC	BC623	1	Southern Clade
BC game management zone 624	BC	BC624	1	Northern-West
BC game management zone 625	BC	BC625	1	Southern Clade
BC game management zone 742	BC	BC742	1	Southern Clade
BC game management zone 750	BC	BC750	1	Southern Clade

BC game management zone 751	BC	BC751	3	Southern Clade
BC game management zone 754	BC	BC754	1	Southern Clade
BC game management zone 757	BC	BC757	1	Southern Clade

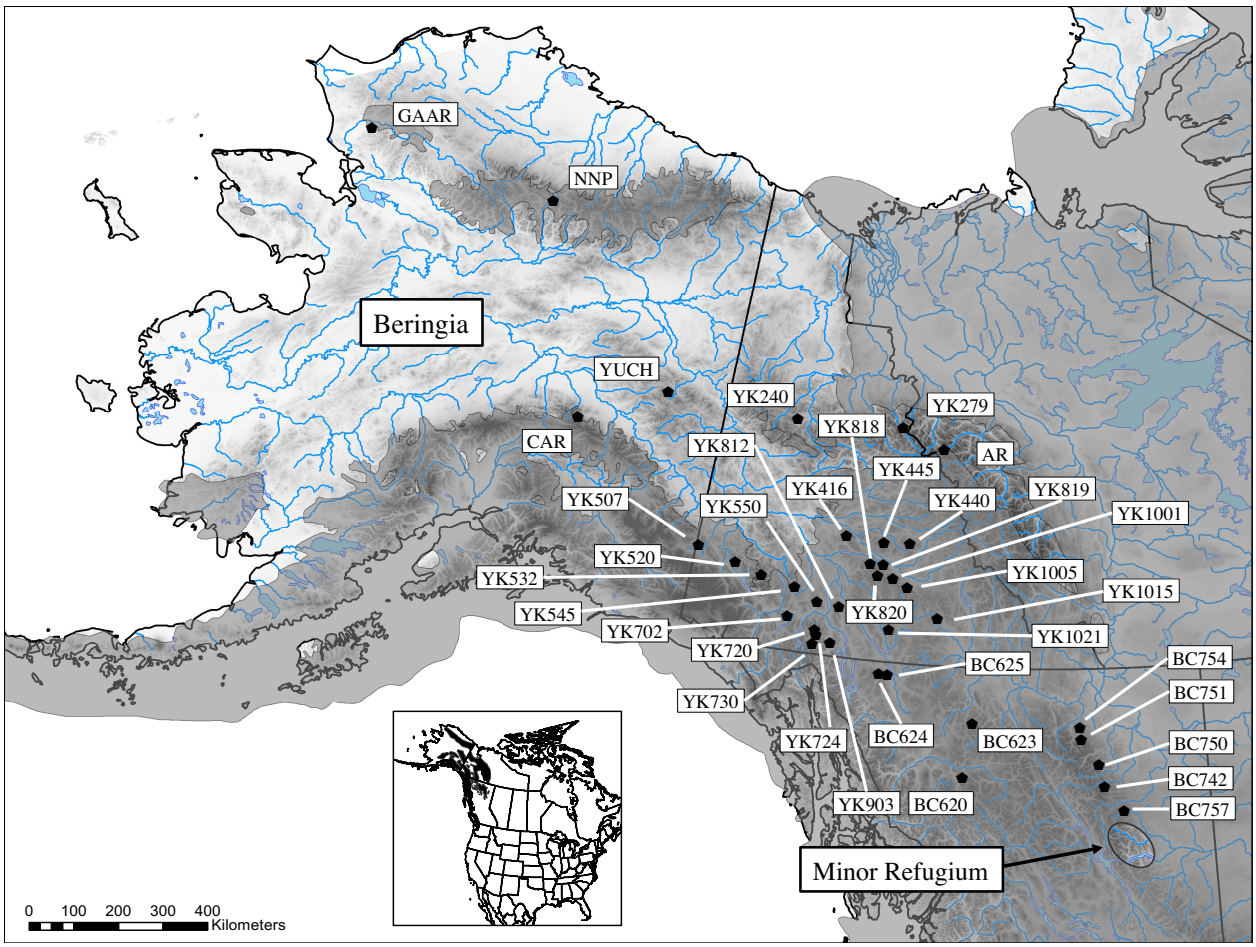

Figure 3.1 Map of range-wide sampling localities for thinhorn sheep. Sample locality abbreviations match that of Table 1. Inset map represents the worldwide distribution of *O. dalli.* Shaded area represents the glacial extent during the glacial maximum (~21,000 years ago) with the Beringian refugium (Dyke 2004) and the Catto *et al.* (1996) minor refugium labeled.

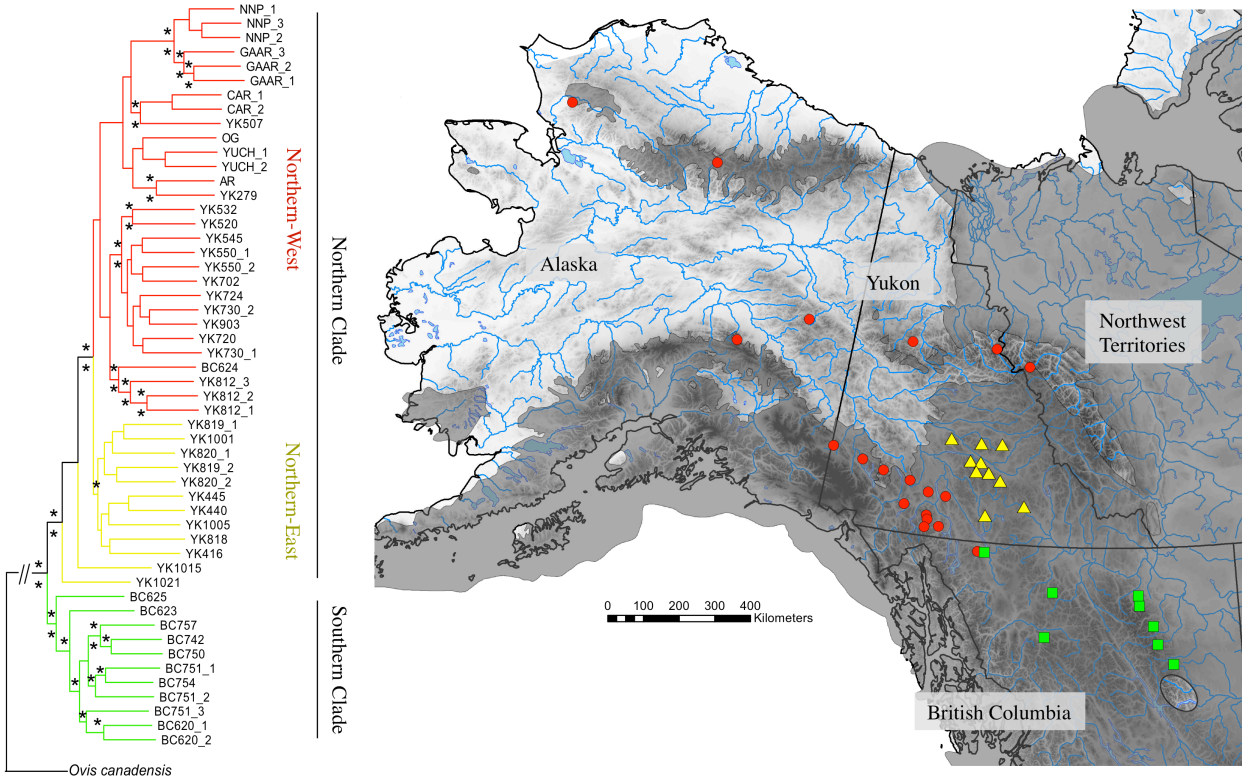

Figure 3.2 Bayesian tree of 52 thinhorn sheep individuals with individual clade/cluster assignments plotted on the adjacent map. * above and below branches denote 100% Bayesian posterior probabilities and maximum parsimony bootstrap support over 75%, respectively. The tree was rooted using 5 *Ovis canadensis* as the outgroup. Bars and labels to the right of the bars delineate the Northern and Southern Clades while coloured labels to the left of the Northern Clade bar distinguish cluster identities within the Northern Clade. Terminal taxa abbreviations match sampling localities in Table 1. Shaded area represents the glacial extent during the last glacial maximum (~21,000 years ago). (Red/● - Northern-West cluster; Yellow/▲ - Northern-East cluster; Green/■ - Southern Clade)

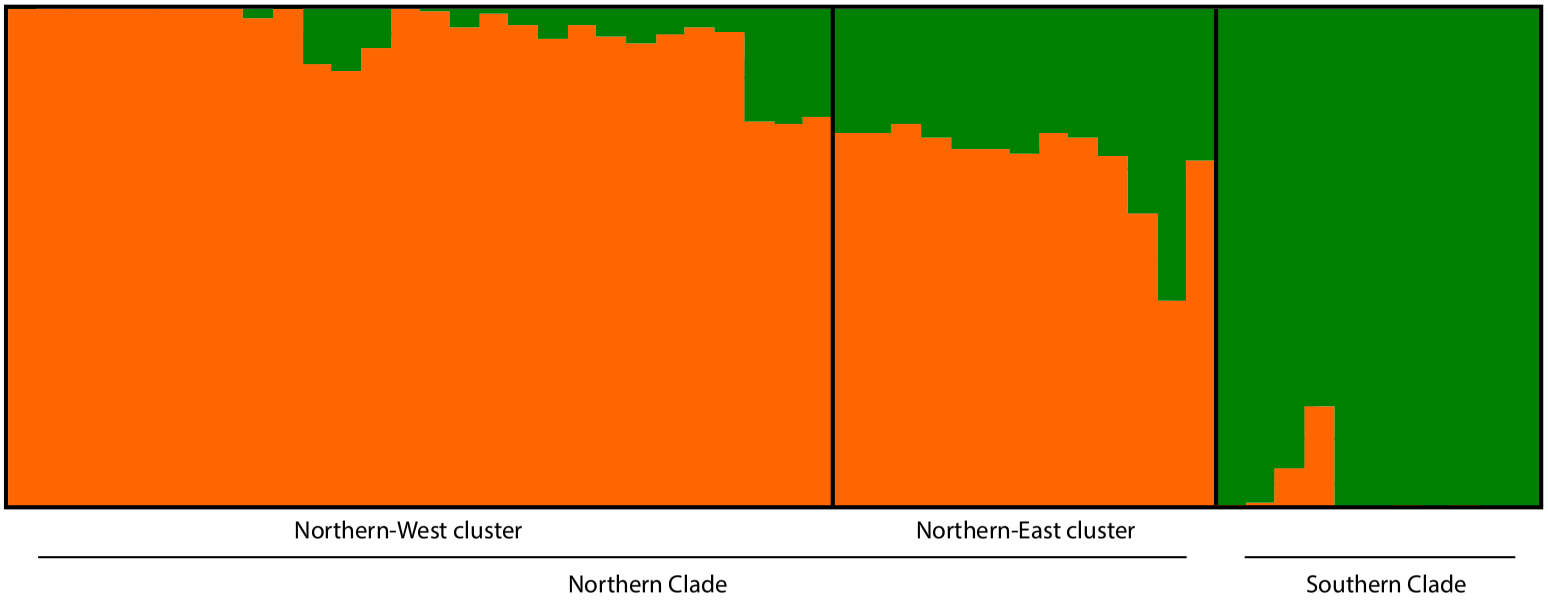

Figure 3.3 Admixture plot from Structure (K=2) showing the levels shared ancestry of sampled individuals. Each bar represents an individual with orange representing ancestry from the Northern Clade and green representing ancestry from the Southern Clade.

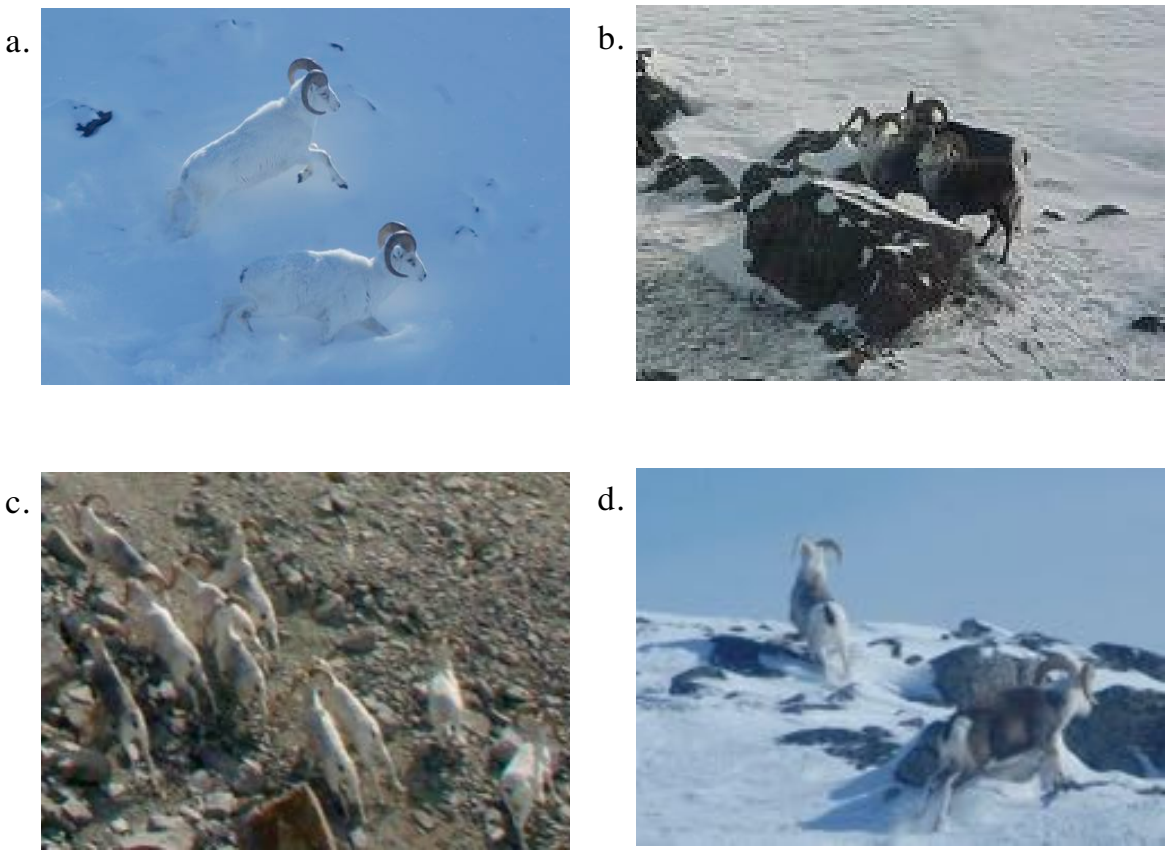

Figure 3.4 Pictures of sheep from different regions of the thinhorn sheep distribution showing representative pelage colourations: (a) Dall's sheep from Northern-West, (b) Stone's sheep from Southern Clade, (c, d) range of pelage colouration possibilities of sheep from Northern-East. Photo credit: Bill Jex

3.7 References

Akaike H (1974) A new look at the statistical model identification. *Automatic Control, IEEE Transactions on* **19**, 716-723.

Avise JC (2000) *Phylogeography: the history and formation of species* Harvard University Press.

Bidlack AL, Cook JA (2001) Reduced genetic variation in insular northern flying squirrels (Glaucomys sabrinus) along the North Pacific Coast. *Animal Conservation* **4**, 283-290.

Brunsfeld S, Sullivan J, Soltis D, Soltis P (2001) Comparative phylogeography of northwestern North America: A synthesis. In: *Integrating Ecology and Evolution in a Spatial Context* (eds. Silvertown J, Antonovics J), pp. 319-339. Blackwell Science, London.

Catto N, Liverman DGE, Bobrowsky PT, Rutter N (1996) Laurentide, cordilleran, and montane glaciation in the western Peace River ‚Äì Grande Prairie region, Alberta and British Columbia, Canada. *Quaternary International* **32**, 21-32.

Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* **9**, 772-772.

Demarchi R, Hartwig C (2004) Status of Thinhorn Sheep in British Columbia. BC Ministry of Water, Land and Air Protection, Biodiversity Branch. *Bull. No. B-119. 96pp*, 3.

Dobeš C, Mitchell-Olds T, Koch MA (2004) Intraspecific diversification in North American Boechera stricta (= Arabis drummondii), Boechera ×divaricarpa, and Boechera holboellii (Brassicaceae) inferred from nuclear and chloroplast molecular markers—an integrative approach. *American Journal of Botany* **91**, 2087-2101.

Dyke AS (2004) An outline of North American deglaciation with emphasis on central and northern Canada. *Quaternary glaciations: extent and chronology* **2**, 373-424.

Dyke AS, Prest VK (1987) Late Wisconsinan and Holocene history of the Laurentide ice sheet. *Géographie physique et Quaternaire* **41**, 237-263.

Earl D, vonHoldt B (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* **4**, 359-361.

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

Fitch WM (1971) Toward defining the course of evolution: minimum change for a specific tree topology. *Systematic Biology* **20**, 406-416.

Galbreath KE, Hafner DJ, Zamudio KR (2009) When cold is better: climate-driven elevation shifts yield complex patterns of diversification and demography in an apline speciatlist (American pika, *Ochotona princeps*). *Evolution* **63**, 2848-2863.

Gasca-Pineda J, Cassaigne I, Alonso RA, Eguiarte LE (2013) Effective Population Size, Genetic Variation, and Their Relevance for Conservation: The Bighorn Sheep in Tiburon Island and Comparisons with Managed Artiodactyls. *Plos One* **8**, e78120.

Geist V (1971) *Mountain sheep* Chicago Press.

Gilbert KJ, Andrew RL, Bock DG*, et al.* (2012) Recommendations for utilizing and reporting population genetic analyses: the reproducibility of genetic clustering using the program structure. *Molecular Ecology* **21**, 4925-4930.

Gradstein FM, Ogg JG (2012) Chapter 2 - The Chronostratigraphic Scale. In: *The Geologic Time Scale*, pp. 31-42. Elsevier, Boston.

Gutiérrez-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.

Haynes GD, Latch EK (2012) Identification of novel single nucleotide polymorphisms (SNPs) in deer (*Odocoileus* spp.) using the BovineSNP50 BeadChip. *Plos One* **7**, e36536.

Henry P, Sim Z, Russello MA (2012) Genetic evidence for restricted dispersal along continuous altitudinal gradients in a climate change-sensitive mammal: the American pika. *Plos One* **7**, e39077.

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

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

Johnston SE, McEwan JC, Pickering NK*, et al.* (2011) Genome-wide association mapping identifies the genetic basis of discrete and quantitative variation in sexual weaponry in a wild sheep population. *Molecular Ecology* **20**, 2555-2566.

Kijas JW, Porto-Neto L, Dominik S*, et al.* (2014) Linkage disequilibrium over short physical distances measured in sheep using a high-density SNP chip. *Animal Genetics* **45**, 754-757.

Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I (2015) Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources* **15**, 1179-1191.

Latch EK, Heffelfinger JR, Fike JA, Rhodes OE (2009) Species-wide phylogeography of North American mule deer (Odocoileus hemionus): cryptic glacial refugia and postglacial recolonization. *Molecular Ecology* **18**, 1730-1745.

Loehr J, Carey J, Hoefs M, Suhonen J, YlÖNen H (2007) Horn growth rate and longevity: implications for natural and artificial selection in thinhorn sheep (Ovis dalli). *Journal of Evolutionary Biology* **20**, 818-828.

Loehr J, Worley K, Grapputo A*, et al.* (2006) Evidence for cryptic glacial refugia from North American mountain sheep mitochondrial DNA. *Journal of Evolutionary Biology* **19**, 2071-2071.

Loehr J, Worley K, Moe J, Carey J, Coltman DW (2008) MC1R variants correlate with thinhorn sheep colour cline but not individual colour. *Canadian Journal of Zoology* **86**, 147-150.

Meirmans PG, Van Tienderen PH (2004) genotype and genodive: two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes* **4**, 792-794.

Miller JM, Kijas JW, Heaton MP, McEwan JC, Coltman DW (2012) Consistent divergence times and allele sharing measured from cross-species application of SNP chips developed for three domestic species. *Molecular Ecology Resources* **12**, 1145-1150.

Miller JM, Poissant J, Kijas JW, Coltman DW, Consortium ISG (2011) A genome-wide set of SNPs detects population substructure and long range linkage disequilibrium in wild sheep. *Molecular Ecology Resources* **11**, 314-322.

Nagylaki T (1985) Homozygosity, effective number of alleles, and interdeme differentiation in subdivided populations. *Proceedings of the National Academy of Sciences of the United States of America* **82**, 8611-8613.

Nei M (1987) *Molecular evolutionary genetics* Columbia university press.

Pielou E (1991) *After the Ice Age: The Return of Life to Glaciated North America* The University of Chicago Press, Chicago, IL.

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

Purcell S, Neale B, Todd-Brown K*, et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics* **81**, 559-575.

Rambaut A, Suchard M, Xie D, Drummond A (2014) Tracer v1. 6. *Available from Available at:* http://beast. *bio. ed. ac. uk/Tracer*.

Ramey RR, II (1993) *Evolutionary genetics and systematics of North American mountain sheep: Implications for conservation* Ph.D., Cornell University.

Rezaei HR, Naderi S, Chintauan-Marquier IC*, et al.* (2010) Evolution and taxonomy of the wild species of the genus Ovis (Mammalia, Artiodactyla, Bovidae). *Molecular Phylogenetics and Evolution* **54**, 315-326.

Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572-1574.

Shafer ABA, Cote SD, Coltman DW (2011) Hot spots of genetic diversity descended from multiple pleistocene refugia in an alpine ungulate. *Evolution* **65**, 125-138.

Shafer ABA, Cullingham CI, CÔTÉ SD, Coltman DW (2010) Of glaciers and refugia: a decade of study sheds new light on the phylogeography of northwestern North America. *Molecular Ecology* **19**, 4589-4621.

Sheldon C (1911) The wilderness of the Upper Yukon; a hunter's explorations for wild sheep in sub-Arctic Mountains. 354 pp, illus. *New York*.

Swofford DL (2003) (Swofford).

Tavaré S (1986) Some probabilistic and statistical problems in the analysis of DNA sequences. In: *American Mathematical Society: Lectures on Mathematics in the Life Sciences*, pp. 57-86. Amer Mathematical Society.

Templeton AR (1983) Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of humans and the apes. *Evolution* **37**, 221-244.

Tomasik E, Cook JA (2005) *Mitochondrial Phylogeography and Conservation Genetics of Wolverine (Gulo gulo) of Northwestern North America*.

Valdez R, Krausman PR (1999) *Mountain sheep of North America* University of Arizona Press.

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

Zink RM, Blackwell-Rago RC, Ronquist F (2000) The shifting roles of dispersal and vicariance in biogeography. *Proceedings of the Royal Society B-Biological Sciences* **267**, 497-503

Chapter 4: Management implications of highly resolved hierarchical population genetic structure in thinhorn sheep

4.1 Abstract

Patterns of genetic variation of a species can be shaped by events that occur at wide temporal and geographic scales. Geophysical processes, such as continental glaciations, can affect species vicariance at wide scales whereas processes that act at finer scales, such as gene flow between populations, can have more localized effects. Recent studies have shown that contemporary population structure should be interpreted within the context of historical events, such as ice-age vicariance, due to the hierarchical nature of genetic variation found in many species. The thinhorn sheep (*Ovis dalli*) is a mountain specialist found in northwestern North America, from the Brooks Range in Alaska, east through Yukon to the Mackenzie Mountains of Northwest Territories (NWT) and south to the Rocky and Coastal Mountains of British Columbia. In this study, we examined the population genetic structure of thinhorn sheep in light of genetic evidence regarding the glacier driven evolution of the two thinhorn sheep subspecies, *O. d. dalli* and *O. d. stonei,* using 153 biallelic single-nucleotide polymorphisms genotyped in over 2800 thinhorn rams. We found patterns of genetic variation to be generally consistent with genetic subspecies boundaries at the species-wide level, and mountain range and river valley boundaries at finer scales. By taking in account historical vicariance by conducting hierarchical analyses of population genetic structure, we revealed the presence of three previously unreported Stone's sheep genetic clusters in the Stikine/Skeena, Cassiar and Rocky Mountains and identified a new geographic range for Stone's sheep that is much more restricted than currently accepted subspecies maps indicate, and is almost exclusively confined within BC. Our results indicate that contemporary patterns of genetic variation can be influenced by events acting over a range of spatial and temporal scales, and thus the importance of interpreting the findings of present-day genetic structure in light of the phylogeographical history of the species in question.

4.2 Introduction

The distribution of genetic variation in a species is driven by a number of factors at different timescales and over different geographic scales that are typically the outcome of vicariant and

dispersal processes. Large-scale historical events, such as continental wide glaciation, can influence genetic differentiation at a species-wide level while the existence of barriers to gene flow at a landscape level can have more localized effects (Avise 2000; Shafer *et al.* 2010). Together, these factors shape the structure and genetic diversity of a species. Therefore, any attempts to study the contemporary genetic structure of a species should take into account its evolutionary history since the signals of historical events, such as glacial vicariance, and present day dispersal can be confounding (Zink *et al.* 2000). Alpine species that are widely distributed make ideal subjects to study the relative effects of historical vicariance and gene flow since their lack of dispersal can result in the retention of the signatures of historical events while such signals may be masked in more vagile species due to gene flow between historically isolated lineages.

The continent of North America underwent repeated glaciations during the Pleistocene epoch and vicariance as a result of ice-sheet advance has been shown to have driven much of broad-scale genetic differentiation among many of its native species (Shafer *et al.* 2010). One such species is the thinhorn sheep (*Ovis dalli*). The thinhorn sheep is an alpine ungulate endemic to the mountains of northwestern North America, from the Brooks Range in Alaska (AK), east through Yukon (YK) to the Mackenzie Mountains of Northwest Territories (NWT) and south to the Rocky and Coastal Mountains of British Columbia (BC) (Valdez & Krausman 1999). Currently, thinhorn sheep are classified into two subspecies based on pelage colour: 1) the white coated and northerly Dall's sheep (*O. d. dalli*) and 2) the dark coated and southerly Stone's sheep (*O. d. stonei*). This colour-based classification of subspecies is the source of some debate (Geist 1971; Loehr *et al.* 2006; Sim *et al.* 2016; Worley *et al.* 2004) since thinhorn sheep coat colour shows complete intergradation from pure white Dall's sheep types through shades of gray/brown to dark brown Stone's sheep types (Sheldon 1911). No clear break in coat colour exists between the subspecies and coat colours are often not uniform within a population (Demarchi & Hartwig 2004). Sheep with intermediate colouration, informally referred to as Fannin's sheep, are most commonly found in the Ogilvie Mountains (YK) as well as the Pelly (YK) and Casssiar Moutains (YK/BC) although grey saddled Dall's sheep have been known to exist even as far north as the Brooks Range in AK (Heimer pers. comm.). There have also been past attempts to

designate sheep inhabiting the Kenai Peninsula of AK as a subspecies, *O. d. kenaiensis*, based on skull morphology (Allen 1902).

Thinhorn sheep have substantial ecological, economic and cultural value, and they are a vital component of the natural heritage of North America (Jex *et al.* 2016), being highly valued by both sport and subsistence hunters, and by non-consumptive users alike. Wildlife agencies generally focus management toward protecting the long-term viability of the species and localized sub-populations, while providing opportunities for resource use. Effective conservation of thinhorn sheep populations relies on the establishment of biologically meaningful population units for management and protection of important seasonal habitats and movement corridors between those habitats. At present, management units can be based on subspecies designation, administrative boundaries, pelage colour, topology and biogeography. However, these measures may be imprecise and uninformative for determining sustainable harvest limits, or in determining informed risk or sensitivity ratings associated with resource development. The delineation of more precise population boundaries and data regarding the degree of connectivity between populations are necessary to inform resource development decision-making processes as well as habitat and landscape management, so that the viability of herds, local and broad scale ecology, and options for sustainable human uses are maintained and protected. The use of molecular markers to investigate the amount of gene flow and levels of genetic differentiation within and between population units can give managers information regarding the demographic independence of different thinhorn sheep herds, and the incorporation of spatially explicit analyses can inform the management unit delineation process by regulatory agencies (Moritz 1994). Information from a genetic analysis of population structure can also be used to designate groups that may be considered for special protection or management action under relevant conservation legislation in different jurisdictions (e.g Endangered or Threatened Species Listing - Species at Risk Act – Canada, British Columbia Wildlife Act). Precise knowledge of the boundaries of thinhorn populations and amount of gene flow between populations will help managers assess human encroachment and anthropogenic impacts, which has been identified by managers as one of the biggest future challenges in thinhorn sheep conservation management (Jex et. al. 2016).

Mitochondrial DNA (mtDNA) data show broad scale genetic variation to be strongly influenced by the Pleistocene glacial cycles and that the evolution of the two thinhorn sheep subspecies (*O. d. dalli* and *O. d. stonei*) was likely due to isolation in different glacial refugia (Loehr *et al.* 2006). A recent study using ~10,000 nuclear single nucleotide polymorphisms (SNPs) confirmed the role of glacial vicariance in *O. dalli* subspecies formation (Sim *et al.* 2016). This study found the most basal split in thinhorn sheep exists between sheep from BC and southern YK from the rest of the species. Using this delineation, many populations in southern YK that are currently identified as Stone's sheep based on coat colour could be more appropriately placed in the Dall's sheep clade. The same study also revealed admixture between *O. d. dalli* and *O. d. stonei* likely due to post-glacial range expansion and hybridization. The admixture zone is located in the Pelly (YK) and Cassiar Mountains (YK/BC), which coincide with a geographical region of mixed pelage colouration (Sim *et al.* 2016). Sheep from this area are informally named Fannin's sheep (pers obs).

As an alpine species, thinhorn sheep have specific habitat needs that are likely to influence the genetic structure of the species across its distribution. Thinhorn sheep require precipitous escape terrain that is adjacent to open meadows used for feeding, which generally restricts populations to alpine and subalpine habitats (Valdez & Krausman 1999). Boreal environments and river valleys found at lower elevations have been observed to form natural barriers to dispersal (Geist 1971). Recent microsatellite studies have found gene flow to be impeded by river valleys (Roffler *et al.* 2014) and genetic structure to be strongly influenced by mountain range boundaries (Worley *et al.* 2004).

While previous mtDNA (Loehr *et al.* 2006) and SNP (Sim *et al.* 2016) studies had strongly suggested that current subspecies boundaries are incongruent with genetic data, the lack of fine scale sampling has left the actual boundaries of the Dall's and Stone's subspecies unclear. The geographic boundaries of the admixture zone first reported by Sim *et al.* (2016) also need to be refined. In this study, we genotype over 2800 thinhorn sheep rams at 153 SNPs biallelic SNP markers to investigate the range-wide distribution of genetic variation in thinhorn sheep. We reanalyze the population genetic structure of thinhorn sheep in light of recent studies regarding

the glacial origins of the thinhorn subspecies, and investigate the relationship between spatial and genetic distance. We also compare subspecies distributions defined from genetic data with the current subspecies boundaries that are being used for management purposes. We assess the levels of genetic differentiation between currently used management units and compare the boundaries of these MUs to SNP derived genetic boundaries. We expect range wide genetic structure to be influenced by historical biogeography and mountain range boundaries to be more important at finer scales.

4.3 Methods

4.3.1 Sample Collection

We sampled 2820 thinhorn sheep rams from across the range of the species including all provinces, territories and states in North America where *O. dalli* exist (Table 4.1; Fig. 4.1). Samples were collected between 1996 and 2015 and comprise shavings produced when horns are drilled to insert an identification plug during compulsory inspection of harvested rams. Samples from BC came from 21 management units within the Skeena and Peace regions. YK individuals came from 11 game management zones, which can be broken down further into 175 subzones each of approximately 30-50km in diameter. NWT samples were collected from 8 outfitter management areas that are managed under the Dehcho and Sahtu regions. AK samples were collected from 13 game management units (Table 4.1; Fig. 4.1). Management units in each jurisdiction are defined by combination of political, traditional, subspecies and geographical boundaries.

For population-based analyses, we defined sampling localities by their respective jurisdictional management units (MUs): Management Units for BC, Game Management Zones for YK, Outfitter Management Zones for NWT and Game Management Units for AK. We excluded MUs with < 20 samples from population level analyses since empirical studies have found that when using ~100 SNPs, ~20 individuals are required for accurate estimates of heterozygosities and allele frequencies (Willing *et al.* 2012) as well as adequate statistical power for analyzing populations with intermediate levels of genetic differentiation (F_{ST} = 0.01) (Morin *et al.* 2009). We used all samples for individual-based analyses, but spatial analyses were only performed for

samples with fine-scale location data. BC and NWT samples were geo-referenced while samples from YK and AK were resolved only to the level of game management subunits.

4.3.2 Molecular Technique

Markers used in this study were sub-selected from those that successfully cross species amplified in 55 thinhorn rams (Sim *et al.* 2016) using the Ovine HD Beadchip (Kijas *et al.* 2014). Selection criteria included minor allele frequency (>0.01), even genomic spacing and 1 million bp distance from the closest coding region (as annotated in the domestic sheep genome). Locus specific amplification primers were designed for 180 SNP markers using Primer3 and the domestic sheep genome (Untergasser *et al.* 2012) (Table S1). Of the 180 SNP markers, 27 were eliminated due to non-specific amplification and/or formation of primer dimers yielding a final SNP typing panel of 153 markers (Table S1). All forward primers were 5' tailed with a CAG tag sequence (5'- CAGTCGGGCGTCATCA -3') (Faircloth *et al.* 2009) and all reverse primers were 5' tailed with the trP1 sequence (5'- CCTCTCTATGGGCAGTCGGTGAT -3') to facilitate incorporation of barcode sequences, sequencing primer binding sites and Ion Torrent emulsion PCR (emPCR) specific sequences.

We extracted DNA from horn core samples (0.01-0.04 g) using the Qiagen DNeasy Blood and Tissue Kit (Qiagen). We modified the manufacturer's protocol by increasing the volumes of Buffer ATL, proteinase K and Buffer AL by 50% to ensure that the horn core material was submerged during tissue lysis. We also increased the incubation time to 24 hrs.
Genotyping was performed by amplicon re-sequencing using an Ion Torrent personal genome machine (PGM) (Life Technologies). We generated templates for sequencing in eight 10 µL multiplex polymerase chain reactions (PCR) using an Eppendorf EP thermal cycler (Eppendorf) (Table S1). Each reaction contained 1.5 µL H_2O, 1x Qiagen Multiplex PCR Master Mix (Qiagen), 0.1 µM forward primer, 0.1 µM reverse primer, 2x Q solution (Qiagen) and 2 µL of template DNA. Cycling conditions were: 15 min at 95C, followed by 30 cycles of 30s at 95C, 90s at 58C and 60s at 72C, and a final extension for 30 min at 72C.

We pooled the PCR products from each of the 8 multiplex reactions per individual and then performed a 1:1000 dilution of the pooled product. We barcoded amplicons for each individual using a second 10 μL PCR consisting of 0.71 μL of H_2O, 1x Q5 reaction buffer (New England BioLabs, NEB), 0.128 μM of each dNTP, 3% DMSO, 0.16 uM of trP1 primer (Table S1), 0.1uM IonXpress Barcode primers, 0.3 U/μL of Q5 High-Fidelity polymerase (NEB) and 3 μL of the pooled and diluted template. Cycling conditions were: 30s at 98C, followed by 30s at 98C, 15s at 61C, 30s at 72C and a final extension for 2 min at 72C.

We pooled barcoded amplicons over sets of 96 individuals and gel extracted the pooled amplicons using the QIAquick Gel Extraction Kit (Qiagen). The size range selected was approximately 250-600 bp and gel excision was performed using a sterile scalpel aided by EtBr staining and UV illumination. Gel extracted product was further purified using the QIAquick PCR Purification Kit (Qiagen) in order to remove non-specific PCR products of small molecular weight that can preferentially amplify during emPCR and therefore dominate the sequencing space (Brown *et al.* 2017). We sequenced the purified PCR product using the Ion Torrent PGM sequencer and a 316 chip (Life Technologies) following the manufacturer's recommended protocol. We replicated the DNA extraction and genotyping steps for 10% of the individuals sampled for quality assessment. SNPs were called using the VARIANTCALLER v4.2 plug-in implemented in the Ion Torrent Server v4.2.1 (Life Technologies). VARIANTCALLER was also used to perform initial control for sequencing quality where reads with phred score < 20 for the SNP/variant base were discarded. During SNP calling, loci with read depth < 20 and heterozygote calls where the frequency of reads supporting the variant allele were < 0.15 were also discarded.

4.3.3 Quality control and summary statistics

We used PLINK 1.07 to calculate per locus and per individual missing data (Purcell *et al.* 2007). Individuals or loci with > 10% missing data were excluded from subsequent analysis. We tested for deviations from Hardy-Weinberg equilibrium at the management unit level using Nei's estimator of G_{IS} implemented in the software GENODIVE 2.0b27 (Meirmans & Van Tienderen 2004). We also performed pairwise linkage disequilibrium (LD) pruning in PLINK where we

discarded one of each pair of loci found with $r^2 > 0.5$ within a 5 locus sliding window using marker locations based on the domestic sheep genome.

We calculated expected (H_E) and observed (H_O) heterozygosities, minor allele frequencies (MAF), mean call rate and estimated F_{IS} using the heterozygosity-based G_{IS} (Nei 1987) for each management unit using GENODIVE 2.0b27 (Meirmans & Van Tienderen 2004). We also performed significance testing for deviations from Hardy-Weinberg equilibrium (HWE) using 999 permutation tests implemented in GENODIVE 2.0b27. To characterize genetic differentiation between managements units, we calculated pairwise F_{ST} from an Analysis of Molecular Variance (AMOVA) (Excoffier *et al.* 1992; Michalakis & Excoffier 1996) using GENODIVE 2.0b27 (Meirmans & Van Tienderen 2004). We used the R package ape (Paradis *et al.* 2003) to construct an unrooted neighbour joining population tree based on pairwise F_{ST} values.

4.3.4 Bayesian genetic clustering

We assessed population genetic structure using Bayesian clustering analysis implemented in the program STRUCTURE 2.3.4 (Pritchard *et al.* 2000). We performed 20 independent runs for $k = 1$ to $k = 20$ (where k is the number of clusters) using 500,000 MCMC replicates following a burn-in period of 100,000 replicates (Gilbert *et al.* 2012), using the admixture model with correlated allele frequencies and no location priors. We permuted and averaged STRUCTURE output for all runs at each k using CLUMPAK 1.1 (Kopelman *et al.* 2015). We also used CLUMPAK to detect the presence of multiple modes among STRUCTURE runs for each k (Janes *et al.* 2017). The ΔK (Evanno *et al.* 2005) and ln Pr(X|K) methods (Pritchard *et al.* 2000) were used to assess support for each k. Individuals were assigned to genetic clusters based on STRUCTURE membership coefficients ($q \geq 0.8$). Individuals that could not be assigned to any cluster ($q<0.8$) were considered admixed. Since mountain ungulates are known to exhibit hierarchical population structure (Shafer *et al.* 2011), we repeated our clustering analysis for each identified cluster until individuals could not be assigned to any cluster ($q < 0.8$) as defined by the optimal K from the ΔK method (Evanno *et al.* 2005) or if the ln Pr(X|K) method (Pritchard *et al.* 2000) supported $k = 1$ as the optimal K (Vähä *et al.* 2007). Admixed individuals were excluded from the next level of hierarchical analysis.

4.3.5 Short-term migration rates

We estimated the short-term migration rates between the genetic clusters identified by our Bayesian clustering analysis using BAYESASS 3.04 (Wilson & Rannala 2003). We performed a total of 10 runs with the following MCMC settings: burn-in=2,000,000, iterations= 15,000,000, sampling interval=1000. We adjusted the values of the mixing parameters (final values: deltaM=0.03, deltaA=0.15, deltaF=0.08) to ensure acceptance rates were between 20-60% for each parameter (Rannala 2015). We assessed run convergence by visually examining the stability of post burn-in log-probabilities in trace outputs for each run using TRACER 1.6 (Rambaut *et al.* 2014) and chose the best run using the Bayesian deviance method (Meirmans 2013).

4.3.6 Post-hoc AMOVA

We performed a hierarchical analysis of molecular variance (AMOVA) to explore the genetic variation explained by differences 1) within individuals, 2) between subspecies, and 3) between STRUCTURE determined genetic clusters nested within each subspecies. Individuals were assigned to clusters for which it has the highest membership coefficient. We did not test for significance in our AMOVAs due to the circularity of logic of testing for significance in pre-defined clusters based on the same dataset. All AMOVAs were performed using GENODIVE 2.0b27.

4.3.7 Isolation by distance

Isolation by distance (IBD) has been shown to confound tests for genetic clustering particularly when hierarchical partitioning of genetic variation is expected (Meirmans 2012), as is the case with many wide-spread North American ungulate species including the thinhorn sheep (Shafer *et al.* 2011; Worley *et al.* 2004). To investigate the spatial structure of genetic variation, we performed Mantel and partial Mantel tests to investigate correlations between genetic distances (Rousset 2000) and geographical distance for individuals with geo-referenced spatial data in the R package VEGAN 2.5-2 (Oksanen *et al.* 2018), using 9999 permutations for significance testing. Pairwise geographical and genetic distances between sampling points were calculated using

SPAGeDi 1.3a (Hardy & Vekemans 2002). To account for the effect of hierarchical genetic structure on the interpretations of the Mantel tests, we explored two population groupings for our Mantel tests as informed by biogeography and population structure: 1) all individuals (species level), and 2) individuals in their respective subspecies/glacial refugia (subspecies level) (Kuchta & Tan 2005). Since tests of IBD can be confounded by genetic structure, we performed two additional partial Mantel tests: 1) association between geographical and genetic distances while conditioning on a matrix denoting the cluster identity (i.e. 0 = belong to same genetic cluster, 1 = belong to different genetic cluster), and 2) association between genetic distance and genetic cluster while conditioning on a matrix denoting geographical distance between individuals (Meirmans 2012). We used VEGAN to compute a Mantel correlogram for each subspecies identified by the initial STRUCTURE analysis (Oksanen *et al.* 2018). We used 9999 permutations to test for significance and Sturges' formula to determine the number and breakpoints for distance classes.

4.4 Results

4.4.1 SNP genotyping and quality control

Of the 153 SNP markers genotyped, 19 had <90% genotyping success and 2 were monomorphic. Of the remaining 132 loci, no pairs of loci were found to be in significant LD. Two loci were found to be significantly out of HWE in 22 of the 53 sampling localities tested and were removed from subsequent analysis. In total, 130 SNP markers were retained. Of the 2820 individuals sampled, 1764 (BC=549, YK=610, NWT=353, AK=252) had > 90% genotyping success and were retained. Call rates for samples that passed quality control ranged from 0.927 to 1.000 with a mean call rate of 0.989.

4.4.2 Summary statistics and genetic differentiation

Expected heterozygosities (H_E) ranged from 0.108-0.295 while observed heterozygosities (H_O) ranged from 0.108-0.276 across all loci and sampling localities. Across a possible 6890 tests (number of loci x sampling localities), 414 (0.06) were found to be significantly out of HWE. Generally, diversity was highest in the Cassiar and Pelly Mountains (BC623, BC624, BC625,

BC626; YK8, YK10) and lowest in the AK and Mackenzie Mountain MUs (AK12, AK13, AK20, AK26; NWTS2, NWTS5, NWTG1) (Table 4.1).

4.4.3 Bayesian clustering

We detected a hierarchical pattern of genetic structure using Bayesian clustering analysis. Species wide, the ΔK method supported k=2 while the most supported number of clusters using the ln Pr(X|K) method could not be unambiguously determined since the ln Pr(X|K) curve transitioned gradually towards its asymptote between k=2 and k=3 (Figure S4.3; Figure S4.4). At k=2, the thinhorn sheep distribution is split roughly along a northeast and southwest axis approximately delineated by the Taku River and Teslin Lake watersheds with an admixture zone that stretches from Atlin Lake and the Pelly River to the north, and to the Stikine River in the south (Figure 4.3). This split has been identified by previous genetic studies (Loehr *et al.* 2006; Sim *et al.* 2016; Worley *et al.* 2004) with Dall's sheep (*O. d. dalli*) found in the northern cluster and Stone's sheep (*O. d. stonei*) in the southern cluster (Figure 3; Figure. 4). At k=3, the Dall's sheep cluster was split into two: one cluster found in the Brooks Range, Ogilvie Mountains and Mackenzie Mountains and the other in the Alaskan Range, Wrangell Range and northern Coast Mountains. Runs above k=3 showed signs of potentially spurious or non-convergent results. *k=4* splits Dall's sheep from the Alaskan and Wrangell Range, and northern Coast Mountain, while suggesting that sheep from Brooks Range are admixed with ancestry from a) Ogilvie Mountains and Mackenzie Mountains, and b) Alaskan Range, Wrangell Range and northern Coast Mountains. This is biologically implausible since the Brooks Range is geographically isolated (>500km away) from either of the apparent parent populations. All runs above k=5 have multiple modes with no k value possessing a common mode that was represented in >60% (12/20) of the runs. We performed the subsequent STRUCTURE analysis using the k=2 clustering as this grouping reflected sub-specific relationships.

When we performed the STRUCTURE analysis for the Dall's sheep cluster, the ΔK method supported k=2 while the ln Pr(X|K) method preferred k=5 (Figure S4.3; Figure S4.4). At k=2, the Dall's sheep cluster is split along a northwestern axis by the Tintina Trench with individuals from the Brooks Range, Ogilvie Mountains and Mackenzie Mountains (northern Dall's group)

forming one cluster, while individuals from the Central Alaskan Range, Wrangell Range and Coast Mountains (southern Dall's group) making up the other (Figure 4.4). This subdivision of the northern cluster into two is very similar to the *k*=3 split found in the species-wide STRUCTURE run. At *k*=5, we see clusters that correspond roughly to 1) Brooks Range, 2) Ogilvie Mountains, 3) Mackenzie Mountains, 4) Central Alaskan Range and 5) Wrangell Mountains/Coast Mountains. When we ran the STRUCTURE analysis separately for each of the clusters identified by *k*=2*,* the northern Dall's group split into three clusters; 1) Brooks Range, 2) Ogilvie Mountains, and 3) Mackenzie Mountains (Figure 4.4), while the southern Dall's group split into two clusters; 1) Central Alaskan Range and 2) Wrangell Mountains/Coast Mountains (Figure 4.4) for a total of 5 clusters nearly identical to those identified using the ln Pr(X|*K*) method performed on the entire Dall's cluster (Figure S4.5-S4.8). These 5 clusters, plus Stone's sheep, were also identified in a minor mode among *k*=6 (8/20 run), *k*=7 (3/20 runs) solutions for the species-wide STRUCTURE run.

When we performed the STRUCTURE analysis within the Stone's sheep cluster, the Δ*K* method supported *k*=2 while the ln Pr(X|*K*) method supported *k*=3 (Figure S4.9; Figure S4.10). At *k*=2, Stone's sheep are split roughly along the Kechika River. The eastern Stone's cluster consists of individuals occupying the Rocky Mountains while the western Stone's sheep cluster occupies the Coast and Skeena Mountains south of the Stikine River, as well as, Cassiar and Omineca Mountains (Figure 4.4). Minor modes of the species-wide *k=6* (2/20) and k=7 (5/20) STRUCTURE runs found a similar east-west split in Stone's sheep as well. At *k*=3, the cluster west of the Kechika River is split approximately north/south along the Stikine River (Figure 4.4). When we ran separate STRUCTURE analyses for the clusters identified at *k*=2, the western Stone's sheep cluster was roughly split along the Stikine River *(k*=2) according to both the Δ*K* and ln Pr(X|*K*) methods (Figure 4.4). We found poor support for breaking the eastern cluster any further since the ln Pr(X|*K*) method preferred *k*=1.

4.4.4 Post-hoc AMOVA

Our hierarchical AMOVA showed that 65.5% of genetic variation was within individuals, 13.8% was among individuals nested within each subspecies, and 20.7% was among subspecies at the

species level. Within Dall's sheep, 77.4% of the variation was found within individuals, 6.9% among individuals nested in genetic clusters and 15.7% among STRUCTURE determined genetic clusters. Within Stone's sheep, we found 85.7% of variation was within individuals, 5.4% among individuals nested in genetic clusters and 8.9% among genetic clusters.

4.4.5 Isolation by distance

At the species level, we found significant IBD (Mantel test: $r = 0.582$, $P < 0.0001$) that remained significant after accounting for genetic clustering (Partial Mantel test: $r = 0.292$, $P < 0.0001$). Correlation between genetic distance and genetic cluster was also significant after conditioning on geographical distance (Partial Mantel test*:* $r = 0.734$, $P < 0.0001$). At the subspecies level, we also found significant IBD (Dall's: $r = 0.488$, $P < 0.0001$; Stone's: $r = 0.246$, $P < 0.0001$) that remained significant after accounting for genetic clustering (Dall's: $r = 0.0634$, $P < 0.0001$; Stone's: $r = 0.0672$, $P < 0.001$) although the resultant correlations were marginal. Correlation between genetic distance and genetic cluster remained significant in both Dall's and Stone's sheep after conditioning on geographical distance (Dall's: $r = 0.478$, $P < 0.0001$; Stone's: $r = 0.411$, $P < 0.0001$).

Mantel correlograms illustrate that Mantel correlations generally decreased as geographical distance increased for both Dall's and Stone's sheep. Mantel correlations were positive for the first 4 (0 - 326.98km) and 3 (0 - 227.59km) distance classes in Dall's and Stone's sheep, respectively (Figure 4.5).

4.5 Discussion

4.5.1 Inference of population genetic structure

In this study, we investigated the population genetic structure of the world's thinhorn sheep herds by genotyping 2820 thinhorn sheep rams using 153 bialleic SNP markers. A previous genetic study had identified 8 genetic clusters using 12 microsatellites and 919 samples (Worley *et al.* 2004) but the authors noted difficulty in determining the optimal number of genetic clusters in their STRUCTURE analysis due to gradually asymptotic ln Pr(X|*K*) solutions. The authors also

raised the potential role of Pleistocene glaciations on the distribution of genetic variation in thinhorn sheep but did not explicitly account for its role in their analysis. Since then, studies using mtDNA (Loehr *et al.* 2006) and SNPs (Sim *et al.* 2016) have provided evidence for evolution of Dall's and Stone's sheep due to glacier mediated vicariance. Sim *et al.* (2016) also identified a potential admixture zone in southern YK and northern BC but had few samples and thus poor spatial resolution.

Bayesian analysis of genetic structure revealed a pattern of hierarchical genetic structure across the thinhorn sheep range delimited by subspecies and mountain range boundaries. At the broadest scale, genetic variation could be partitioned into two main genetic lineages (k=2) with spatial boundaries that agree with subspecies boundaries for the two thinhorn sheep subspecies, *O. d. dalli* and *O. d. stonei*, as determined by mtDNA (Loehr *et al.* 2006), SNP (Sim *et al.* 2016) and microsatellite markers (Loehr *et al.* 2006; Sim *et al.* 2016; Worley *et al.* 2004). We also found an extensive area of admixture in the contact zone between the Dall's and Stone's sheep clusters ranging from the Pelly Mountains in YK south to the Cassiar Mountains and Skeena Mountains of BC (Figure 4.3). The differentiation between Dall's and Stone's sheep has been shown by previous studies to be driven by vicariance due to isolation in different Pleistocene glacial refugia (Loehr *et al.* 2006; Sim *et al.* 2016). In this case, the broad scale partitioning of genetic variation is the legacy of historical vicariance, as the two genetic lineages are descendants of sheep that occupied different refugia. The admixture zone results from recent gene flow in the contact area between the two lineages following post-glacial recolonization (Sim *et al.* 2016). Loehr *et al.* (2006) similarly found, using a 604bp portion of the mtDNA control region, that sheep in the Peace region of BC, where the majority of *O. d. stonei* are found, possessed distinct haplotypes when compared to sheep elsewhere.

After accounting for differences between subspecies, genetic variation is largely shaped by mountain range boundaries and large-river watersheds. This is expected since the thinhorn sheep has an alpine and subalpine distribution, and requires steep slopes for escape terrain (Valdez & Krausman 1999). The use of escape terrain is the thinhorn sheep's primary defense against predation (Valdez & Krausman 1999). A recent microsatellite study found landscape features

that provide good escape terrain, such as steep slopes and open land cover, to be positively correlated with gene flow in a population of Dall's sheep in southcentral AK (Roffler *et al.* 2016). Hengeveldand Cubberley (2011) identified only occasional forays into low elevation habitats that were not linked to breeding activity or gene flow in a population of Stone's sheep in northern BC.

In Dall's sheep, genetic variation can be further broken down into two hierarchical levels according to our Bayesian clustering analysis. At the subspecies level, two clusters roughly north and south of the Tintina trench were identified. The Northern Dall's group, consisting of sheep found in the Brooks Range, Ogilvie Mountains and Mackenzie Mountains, and the Southern Dall's group, from the Central Alaskan Range and Coast Mountains (Figure 4.4), are separated by low elevation habitat, which is likely to impede movement. Further, mountain ranges occupied by the Northern and Southern groups also coincide with comparatively early glacial retreat and make up part of the post-glacial migration corridor in other North American mammals (Brunsfeld *et al.* 2001; Shafer *et al.* 2010). The Northern and Southern Dall's groups can be split into three: 1) Brooks Range, 2) Ogilvie Mountains, and 3) Mackenzie Mountains (Figure 4.4), and two: 1) Central Alaskan Range and 2) Wrangell Mountains/Coast Mountains, clusters respectively, for a total of 5 clusters (Figure 4.4). The Brooks Range is isolated in the northwest AK and surrounded by extensive lower elevation habitats unsuitable for thinhorn sheep movement. The Ogilvie and Mackenzie Mountains clusters are separated by the Arctic Red River Valley.

The Dall's sheep genetic clusters generally agree with Worley *et al.* (2004) with the exception of the Mackenzie Mountains of NWT. In their study, Worley *et al.* (2004) reported two clusters in the Mackenzie Mountains while we report only one. For our analysis, both ΔK method and ln Pr(X|K) method preferred $k = 3$ for the northern Dall's sheep group which does not split the Mackenzie Mountains. Worley et al. reported difficulty in unambiguously determining the optimal number of clusters for their STRUCTURE analysis with Mackenzie Mountains split occurring at $k = 8$ (of 8). More generally only small differences in ln Pr(X|K) for $k = 4$ and $k = 8$ were observed by Worley *et al.* (2004). F_{ST} values were also found to be lowest between

sampling localities in the Mackenzie Mountains in both Worley *et al.* (2004) and this study (Table 4.2). The Mackenzie Mountains east of the Arctic Red River likely represents continuous habitat for thinhorn sheep with no obvious geographical barriers. Results from our Mantel and partial Mantel tests also found significant IBD that was greatly reduced once genetic clustering was taken into account. Therefore, while we observed the split found in Worley *et al.* (2004) in the Mackenzie Mountains in our $k = 4$ Structure solution for the northern Dall's group, weak genetic differentiation, weak Mantel correlation after accounting for genetic clustering and the lack of obvious geographical barriers leads us to group the Mackenzie Mountains sheep in one cluster.

Worley *et al.* (2004) did not report genetic clusters within Stone's sheep since STRUCTURE analysis was not performed separately for Stone's sheep. Further, sheep from the Stikine/Skeena Mountains and Cassiar Mountains (Figure 4.4) were not well sampled. More recent recommendations (Evanno *et al.* 2005; Gilbert 2016; Gilbert *et al.* 2012; Meirmans 2012) and treatments (Latch *et al.* 2014; Shafer *et al.* 2011; Vähä *et al.* 2007) of STRUCTURE analysis allowed us to better delineate the spatial boundaries of genetic structure in thinhorn sheep. In this case, we accounted for differences in evolutionary history due to glacial vicariance by performing separate STRUCTURE analysis for Dall's and Stone's sheep (Loehr *et al.* 2006; Sim *et al.* 2016), which revealed three previously unreported genetic clusters of Stone's sheep along a generally east-west axis consisting of sheep found in: 1) Stikine and Skeena Mountains, 2) Cassiar Mountains and 3) Rocky Mountains, separated by the Stikine and Kechika Rivers respectively (Figure 4.4). Our *post-hoc* hierarchical AMOVA indicated that the greatest amount of genetic variation could be found among the two subspecies, *O. d. dalli* and *O. d. stonei*, which accounted for 24.3% of species-wide genetic variation. Differences found among STRUCTURE derived clusters accounted for 17% and 10.6% within *O. d. dalli* and *O. d. stonei* respectively. The relatively lower levels of differentiation between clusters in *O. d. stonei* may further explain why Worley *et al.* (2004) did not detect the three *O. d. stonei* clusters in their STRUCTURE analysis.

Since *O. d. dalli* and *O. d. stonei* are thought to have evolved due to isolation in different glacial refugia during the Pleistocene ice-age, our AMOVA and STRUCTURE results indicate that historical vicariance still plays a strong role in influencing present-day genetic structure (Loehr *et al.* 2006; Sim *et al.* 2016). Previous genetic studies have identified signatures of isolation in multiple glacial refugia in a variety of animal and plant taxa inhabiting western North America including rockcress (Dobeš *et al.* 2004) and mountain goat (Shafer *et al.* 2011). Therefore, it is likely that the same geological forces exerted on the thinhorn sheep during the Pleistocene that resulted in the hierarchical population structure seen today are also likely in play for other taxa inhabiting western North America. Our results indicate the need to interpret contemporary patterns of population genetic structure within the context of the phylogeograhical history of the species in question.

The use of Mantel and partial Mantel tests for spatial analyses in ecological studies has been the subject of some controversy in the literature, centered on statistical weaknesses of the test that can lead to false positives and the potentially inappropriate comparison of ecological variables using dissimilarity matrices (Guillot & Rousset 2013; Legendre & Fortin 2010; Legendre *et al.* 2015). However, it has also been suggested that the use of Mantel tests for analysis of isolation by distance can be useful especially when results are interpreted conservatively and supplemented with Mantel correlograms (Borcard & Legendre 2012; Diniz-Filho *et al.* 2013). Mantel correlogram analysis is used to compare the strength of Mantel correlations across different distance classes. In light of a significant Mantel test, performing a Mantel correlogram analysis can identify the distance class for which Mantel correlations are significant and thus the extent of isolation by distance. We found evidence for isolation by distance between genetic clusters identified by STRUCTURE but not for individuals within clusters. Mantel correlograms for Dall's and Stone's sheep indicated that positive Mantel correlations existed for distance classes of up to 326.98 km and 227.59 km respectively. These distances are similar in size to the spatial extent of the genetic clusters. Our results agree with the findings of Worley *et al.* (2004), which found significant IBD in species (Mantel $r = 0.747$, $P < 0.0001$) and subspecies level (Mantel $r = 0.452$, $P < 0.0001$) comparisons.

4.5.2 Subspecies designation

At the species-wide level, genetic variation was strongly partitioned into two genetic lineages that represent the two thinhorn sheep subspecies: *O. d. dalli* and *O. d. stonei*. This distribution of genetic diversity largely agrees with the findings of previous SNP (Sim *et al.* 2016), microsatellite (Worley *et al.* 2004), and mtDNA (Loehr *et al.* 2006) studies. With increased sampling across the species distribution, especially in the admixture zone between *O. d. dalli* and *O. d. stonei* around the Cassiar and Skeena Mountains, we were able to clarify the subspecies boundaries of *O. d. dalli* and *O. d. stonei*. We propose that the Dall's sheep range should include all of AK and NWT, all of YK except the Cassiar Mountains near the BC border, and the northwestern part of BC (Coast Range and Cassiar mountains, generally west of Atlin Lake). The Stone's sheep range should be restricted to the area east of Teslin Lake and south of Taku River (Figure 4.3; Figure 4.4). Individuals showing admixture between *O. d. dalli* and *O. d. stonei* (known traditionally as Fannin's sheep), are largely from the Pelly Mountains of YK south through the Cassiar Mountains and to the Stikine River of BC (Figure 4.3; Figure 4.4).

The subspecies boundaries presented here disagree with currently accepted subspecies boundaries based on coat colour (Demarchi & Hartwig 2004; Valdez & Krausman 1999). We generally found sheep of intermediate colouring, which current maps classify as Stone's sheep, to be either admixed or more strongly classified as Dall's sheep. As presently defined, sheep south of the Pelly Mountains and west of the Atlin-Teslin Lake complex are classified as Stone's sheep, despite only possessing intermediate coat colours. Prior to this study, no genetic data existed for sheep in the Pelly/Atlin-Teslin area. We were able to show that sheep in this region are mostly admixed with a majority of Dall's sheep ancestry (Figure 4.4). While our sampling only included rams and males are often found to be the more dispersive sex among mammals, a recent microsatellite study on fine-scale genetic structure of Dall's sheep in the Wrangell-St. Elias National Park and Preserve did not find evidence for sex-biased dispersal (Roffler *et al.* 2014). Therefore our findings of population structure will likely apply to both sexes. Furthermore, even if male-biased dispersal is occurring in other locations, our interpretations are likely to be conservative since the inclusion of the less dispersive sex may reveal more genetic structure.

4.5.3 Informing conservation management

Our findings indicate that Stone's sheep range is much more restricted than current subspecies maps indicate and is almost exclusively restricted within BC. As a result, managers may be incorrectly classifying thinhorn sheep in the Pelly and Cassair Mountains as Stone's sheep when performing population inventories, in effect over estimating the population of Stone's sheep by the number of sheep counted in those areas. These areas also encompass the entire distribution of Stone's sheep in Yukon, which have variously been estimated to contain between 21-27% of the global Stone's sheep population (Demarchi & Hartwig 2004) - only one individual from Yukon had a STRUCTURE membership coefficient of >0.8 for Stone's sheep. Our data show that Stone's sheep are rarer than currently thought and thus could be deserving of a higher level of consideration from wildlife and land managers. Stone's sheep may also be subject to higher hunting pressures in certain jurisdictions as a result of some proximity to human activity; localized hunting pressure for admixed or Fannin's sheep may also increase due to their hybrid status. Below the subspecies level, our results also indicate that the spatial boundaries Coast Mountains cluster of Dall's sheep and to a much less degree the Cassiar Mountains of Stone's sheep are shared between the Canadian province/territory of BC and YK. Therefore managers of each jurisdiction are encouraged to consider some degree of cross border co-management for these herds. Additionally, findings showing some degree of demographic correlation (Waples & Gaggiotti 2006) between the Mackenzie and Ogilevie/Wernecke clusters in Dall's sheep may also warrant some cross-jurisdictional management considerations between the Canadian territories of YK and NWT (Table 3.4). Similarly, potential demographic correlation of the Cassiar and Stikine clusters of Stone's sheep in BC may also suggest the need for some level of co-management between management units in BC (Table 3.4).

The use of molecular markers to assess the distribution of genetic variation in a managed game species can fill important information gaps for managers. Wildlife managers must consider many factors when prescribing management actions and regulatory frameworks within their respective jurisdictions, and land managers must weigh risk and benefit associated with human activities. Our study has identified genetic coherent population units that can form the basis of conservation and management unit delineation for those managers. Furthermore, knowledge of a hierarchical

distribution of genetic variation may inform also the regulatory framework of management authorities, since individual management units may be part of larger hierarchical genetic groupings, and efforts to influence conditions in one area may lead to unintended consequences somewhere else. Many of the processes (glacial vicariance, contemporary gene flow) responsible for the hierarchical distribution of genetic structure in thinhorn sheep may also apply to other wide ranging North American ungulates, such as mountain goat (Shafer *et al.* 2011), a sensitive mountain ungulate species that often requires a higher level of management focus.

Information from the fine-scale mapping of subspecies and genetic cluster boundaries found in this study can also directly inform managers in the designation of conservation status and species at risk listings. For example, the mapping of a zone of hybridization and core subspecies boundaries could be used by governments to update the Conservation Status Reports for Dall's and Stone's sheep in the various jurisdictions. This data can also potentially be used by non-government organizations to re-evaluate geographic sub-species boundaries and to update record lists.

4.6 Conclusion

We found genetic variation in thinhorn sheep to be distributed in a hierarchical fashion, reflecting the influence of processes across a wide temporal and spatial scale: historical vicariance due to continent-wide glaciation, post-glacial expansion out of ice-age refugia, admixture of lineages of different refugial origins as well as contemporary barriers to gene flow. Our results indicate the need for studies of present-day population structure to consider the effects of past and more broadly acting events such as glaciations and mountain uplifts. We show that signatures of historical events may still be readily observed in the distribution of neutral genetic diversity today, and knowledge regarding evolutionary history of a species should shape the interpretation of contemporary population genetic data.

Table 4.1 Sampling jurisdiction, sampling locality, abbreviations, number of samples and summary statistics.

Jurisdiction	Sampling locality	Abbrev.	N	Ho	Hs	Gis	out-of-HWE
Alaska	Game Mangement Unit 7	AK7	2	-	-	-	-
	Game Mangement Unit 11	AK11	6	-	-	-	-
	Game Mangement Unit 12	AK12	41	0.198	0.207	0.043	9
	Game Mangement Unit 13	AK13	27	0.157	0.18	0.129	9
	Game Mangement Unit 14	AK14	18	-	-	-	-
	Game Mangement Unit 15	AK15	2	-	-	-	-
	Game Mangement Unit 16	AK16	2	-	-	-	-
	Game Mangement Unit 19	AK19	17	-	-	-	-
	Game Mangement Unit 20	AK20	62	0.187	0.202	0.076	14
	Game Mangement Unit 24	AK24	6	-	-	-	-
	Game Mangement Unit 25	AK25	20	-	-	-	-
	Game Mangement Unit 26	AK26	48	0.174	0.181	0.04	8
British Columbia	Management Unit 618	BC618	1	-	-	-	-
	Management Unit 619	BC619	22	0.23	0.26	0.117	14
	Management Unit 620	BC620	30	0.224	0.237	0.054	5
	Management Unit 621	BC621	24	-	-	-	-
	Management Unit 622	BC622	1	-	-	-	-
	Management Unit 623	BC623	35	0.255	0.275	0.074	14
	Management Unit 624	BC624	46	0.257	0.276	0.07	14
	Management Unit 625	BC625	35	0.264	0.289	0.088	19
	Management Unit 626	BC626	57	0.275	0.292	0.058	17
	Management Unit 627	BC627	34	0.209	0.249	0.163	25
	Management Unit 628	BC628	9	-	-	-	-
	Management Unit 629	BC629	15	-	-	-	-
	Management Unit 736	BC736	7	-	-	-	-
	Management Unit 740	BC740	10	-	-	-	-
	Management Unit 742	BC742	38	0.248	0.258	0.038	10
	Management Unit 750	BC750	46	0.241	0.268	0.102	19
	Management Unit 751	BC751	50	0.253	0.269	0.058	15
	Management Unit 752	BC752	51	0.247	0.268	0.076	17
	Management Unit 754	BC754	27	0.248	0.27	0.082	8
	Management Unit 757	BC757	10	-	-	-	-
	Management Unit 758	BC758	1	-	-	-	-
Northwest Territories	Outfitter Management Area D/OT/01	NWTD1	48	0.212	0.224	0.054	9
	Outfitter Management Area D/OT/02	NWTD2	65	0.199	0.22	0.092	11
	Outfitter Management Area G/OT/01	NWTG1	30	0.212	0.217	0.025	7
	Outfitter Management Area S/OT/01	NWTS1	48	0.214	0.223	0.042	5
	Outfitter Management Area S/OT/02	NWTS2	44	0.204	0.214	0.047	4
	Outfitter Management Area S/OT/03	NWTS3	37	0.214	0.22	0.028	4
	Outfitter Management Area S/OT/04	NWTS4	35	0.213	0.222	0.041	4
	Outfitter Management Area S/OT/05	NWTS5	46	0.206	0.215	0.038	10
Yukon	Game Management Area 1	YK1	1	-	-	-	-
	Game Management Area 2	YK2	186	0.219	0.232	0.054	30
	Game Management Area 3	YK3	1	-	-	-	-
	Game Management Area 4	YK4	60	0.243	0.263	0.078	17
	Game Management Area 5	YK5	129	0.217	0.233	0.067	30
	Game Management Area 6	YK6	7	-	-	-	-
	Game Management Area 7	YK7	142	0.213	0.218	0.026	20
	Game Management Area 8	YK8	29	0.254	0.279	0.089	10
	Game Management Area 9	YK9	5	-	-	-	-
	Game Management Area 10	YK10	47	0.243	0.271	0.103	17
	Game Management Area 11	YK11	3	-	-	-	-

Table 4.2 Pairwise F_{ST} between samplings localities used in the study. Values above the diagonal represent pairwise F_{ST} values ($p > 0.05$ in bold) while values below the diagonal show p-values (asterisks represent $p < 0.05$).

	AK12	AK13	AK26	BC619	BC620	BC623	BC624	BC625	BC626	BC627	BC742	BC750	BC751	BC752	BC754	NWT D1	NWT D2	NWT G1	NWT S1	NWT S2	NWT S3	NWT S4	NWT S5	YK2	YK4	YK5	YK7	YK8	YK10
AK12	--	0.095	0.237	0.33	0.386	0.312	0.251	0.184	0.217	0.161	0.38	0.364	0.346	0.329	0.372	0.171	0.181	0.173	0.161	0.171	0.165	0.151	0.172	0.136	0.13	0.087	0.148	0.171	0.171
AK13	*	--	0.275	0.381	0.431	0.348	0.294	0.225	0.259	0.211	0.423	0.406	0.384	0.363	0.417	0.224	0.245	0.228	0.206	0.224	0.221	0.2	0.23	0.187	0.17	0.139	0.18	0.213	0.215
AK26	*	*	--	0.385	0.432	0.346	0.299	0.27	0.303	0.303	0.417	0.403	0.38	0.365	0.413	0.215	0.211	0.204	0.208	0.207	0.206	0.207	0.218	0.169	0.205	0.243	0.282	0.262	0.252
BC619	*	*	*	--	0.036	0.027	0.043	0.095	0.108	0.211	0.065	0.053	0.037	0.023	0.058	0.282	0.281	0.283	0.293	0.301	0.289	0.286	0.288	0.265	0.193	0.295	0.309	0.16	0.117
BC620	*	*	*	*	--	0.074	0.086	0.147	0.134	0.264	0.09	0.069	0.06	0.064	0.084	0.339	0.333	0.335	0.345	0.356	0.342	0.342	0.341	0.311	0.254	0.351	0.364	0.234	0.179
BC623	*	*	*	*	*	--	0.035	0.087	0.115	0.208	0.054	0.05	0.033	0.01	0.047	0.263	0.259	0.263	0.272	0.279	0.269	0.269	0.271	0.258	0.18	0.29	0.305	0.15	0.106
BC624	*	*	*	*	*	*	--	0.038	0.076	0.148	0.098	0.082	0.07	0.045	0.089	0.197	0.193	0.198	0.208	0.213	0.207	0.2	0.203	0.19	0.122	0.215	0.233	0.102	0.06
BC625	*	*	*	*	*	*	*	--	0.053	0.074	0.158	0.139	0.122	0.105	0.142	0.161	0.167	0.159	0.161	0.166	0.16	0.158	0.169	0.157	0.074	0.145	0.148	0.051	0.029
BC626	*	*	*	*	*	*	*	*	--	0.099	0.17	0.148	0.132	0.118	0.14	0.192	0.191	0.19	0.193	0.203	0.194	0.185	0.194	0.187	0.115	0.186	0.185	0.104	0.078
BC627	*	*	*	*	*	*	*	*	*	--	0.27	0.251	0.232	0.22	0.249	0.187	0.192	0.185	0.183	0.191	0.178	0.165	0.189	0.174	0.075	0.091	0.092	0.068	0.074
BC742	*	*	*	*	*	*	*	*	*	*	--	0.012	0.024	0.041	0.027	0.34	0.333	0.343	0.352	0.36	0.348	0.346	0.344	0.329	0.26	0.351	0.369	0.227	0.175
BC750	*	*	*	*	*	*	*	*	*	*	*	--	**0.003**	0.032	0.009	0.325	0.321	0.326	0.336	0.344	0.331	0.33	0.329	0.316	0.245	0.338	0.352	0.212	0.165
BC751	*	*	*	*	*	*	*	*	*	*	*	0.065	--	0.017	0.005	0.3	0.296	0.299	0.31	0.317	0.304	0.304	0.305	0.294	0.222	0.32	0.334	0.195	0.144
BC752	*	*	*	*	*	*	*	*	*	*	*	*		--	0.033	0.273	0.273	0.272	0.284	0.291	0.281	0.278	0.275	0.272	0.196	0.305	0.319	0.172	0.124
BC754	*	*	*	*	*	*	*	*	*	*	*	*	*	*	--	0.331	0.329	0.335	0.344	0.354	0.34	0.337	0.34	0.324	0.249	0.345	0.358	0.216	0.168
NWT D1	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	--	0.021	0.037	0.031	0.031	0.034	0.022	0.012	0.075	0.082	0.188	0.22	0.142	0.15
NWT D2	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	--	0.066	0.064	0.062	0.063	0.051	0.036	0.087	0.095	0.197	0.225	0.152	0.154
NWT G1	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	--	0.011	0.017	0.012	0.012	0.028	0.044	0.074	0.182	0.219	0.143	0.151
NWT S1	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	--	**0.001**	0.013	0.011	0.028	0.06	0.074	0.173	0.207	0.143	0.152
NWT S2	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	0.517	--	0.01	0.015	0.026	0.065	0.075	0.177	0.208	0.148	0.155
NWT S3	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	--	0.009	0.025	0.061	0.064	0.171	0.204	0.131	0.142
NWT S4	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	--	0.013	0.051	0.064	0.16	0.194	0.133	0.142
NWT S5	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	--	0.068	0.079	0.185	0.22	0.154	0.155
YK2	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	--	0.07	0.144	0.18	0.137	0.14
YK4	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	--	0.09	0.102	0.03	0.051
YK5	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	--	0.035	0.103	0.123
YK7	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	--	0.105	0.135
YK8	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	--	0.026
YK10	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	--

Table 4.3 Estimates of migration rates between genetic clusters identified by Bayesian clustering using BAYEASS, The standard deviation of the marginal posterior distribution for each estimate is in parenthesis. Shaded values represent migration rates between 0.01 and 0.1, shaded and bold values above 0.1. Diagonals represent the proportion of non-migrants in each cluster.

		Migration from:							
		Dall's sheep					Stone's sheep		
		Brooks	Ogilvie	Mackenzie	Car	Coast	Stikine	Cassiar	Rocky
Migration to:	Brooks	0.9431 (0.0137)	0.0328 (0.0109)	0.0040 (0.0039)	0.0040 (0.0040)	0.0040 (0.0040)	0.0040 (0.0040)	0.0040 (0.0040)	0.0040 (0.0039)
	Ogilvie	0.0024 (0.0022)	0.8848 (0.0143)	**0.1027 (0.0141)**	0.0019 (0.0018)	0.0031 (0.0024)	0.0017 (0.0017)	0.0017 (0.0017)	0.0017 (0.0017)
	Mackenzie	0.0018 (0.0018)	0.0066 (0.0043)	0.9512 (0.0090)	0.0018 (0.0018)	0.0332 (0.0074)	0.0018 (0.0018)	0.0018 (0.0017)	0.0018 (0.0017)
	Car	0.0018 (0.0018)	0.0125 (0.0048)	0.0018 (0.0018)	0.9738 (0.0068)	0.0047 (0.0031)	0.0018 (0.0018)	0.0018 (0.0018)	0.0018 (0.0018)
	Coast	0.0017 (0.0018)	0.0064 (0.0037)	0.0090 (0.0045)	0.0274 (0.0070)	0.9027 (0.0113)	0.0036 (0.0030)	0.0448 (0.0087)	0.0043 (0.0030)
	Stikine	0.0023 (0.0023)	0.0023 (0.0023)	0.0023 (0.0023)	0.0023 (0.0023)	0.0234 (0.0081)	0.7683 (0.0136)	**0.1932 (0.0141)**	0.0059 (0.0040)
	Cassiar	0.0018 (0.0018)	0.0019 (0.0020)	0.0019 (0.0019)	0.0018 (0.0018)	0.0624 (0.0107)	0.0146 (0.0069)	0.8409 (0.0149)	0.0748 (0.0126)
	Rocky	0.0018 (0.0018)	0.0018 (0.0018)	0.0018 (0.0017)	0.0018 (0.0018)	0.0051 (0.0031)	0.0131 (0.0048)	0.0684 (0.0110)	0.9062 (0.0119)

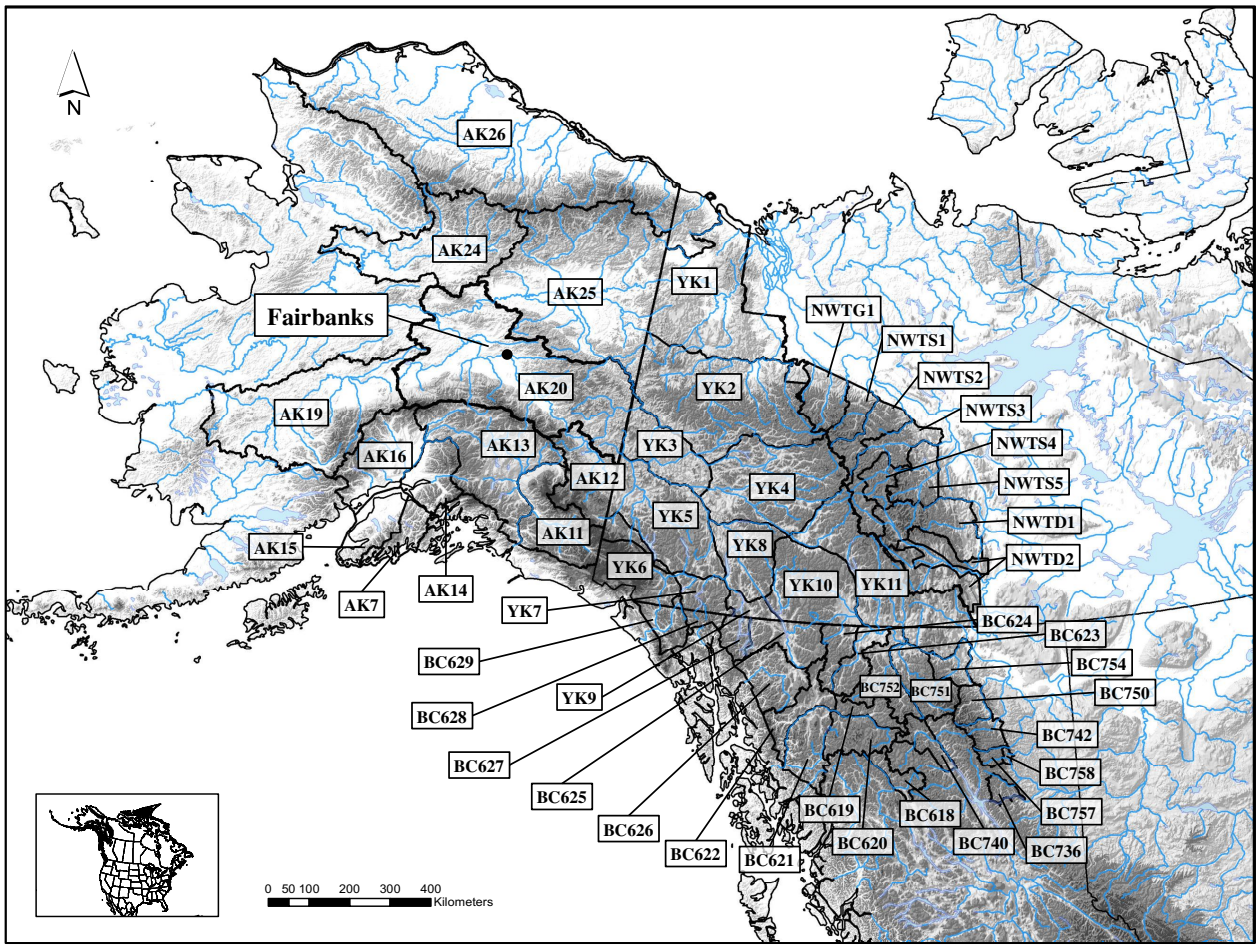

Figure 4.1 Map depicting sampling localities included in this study. Abbreviations for sampling localities are included in Table 1.

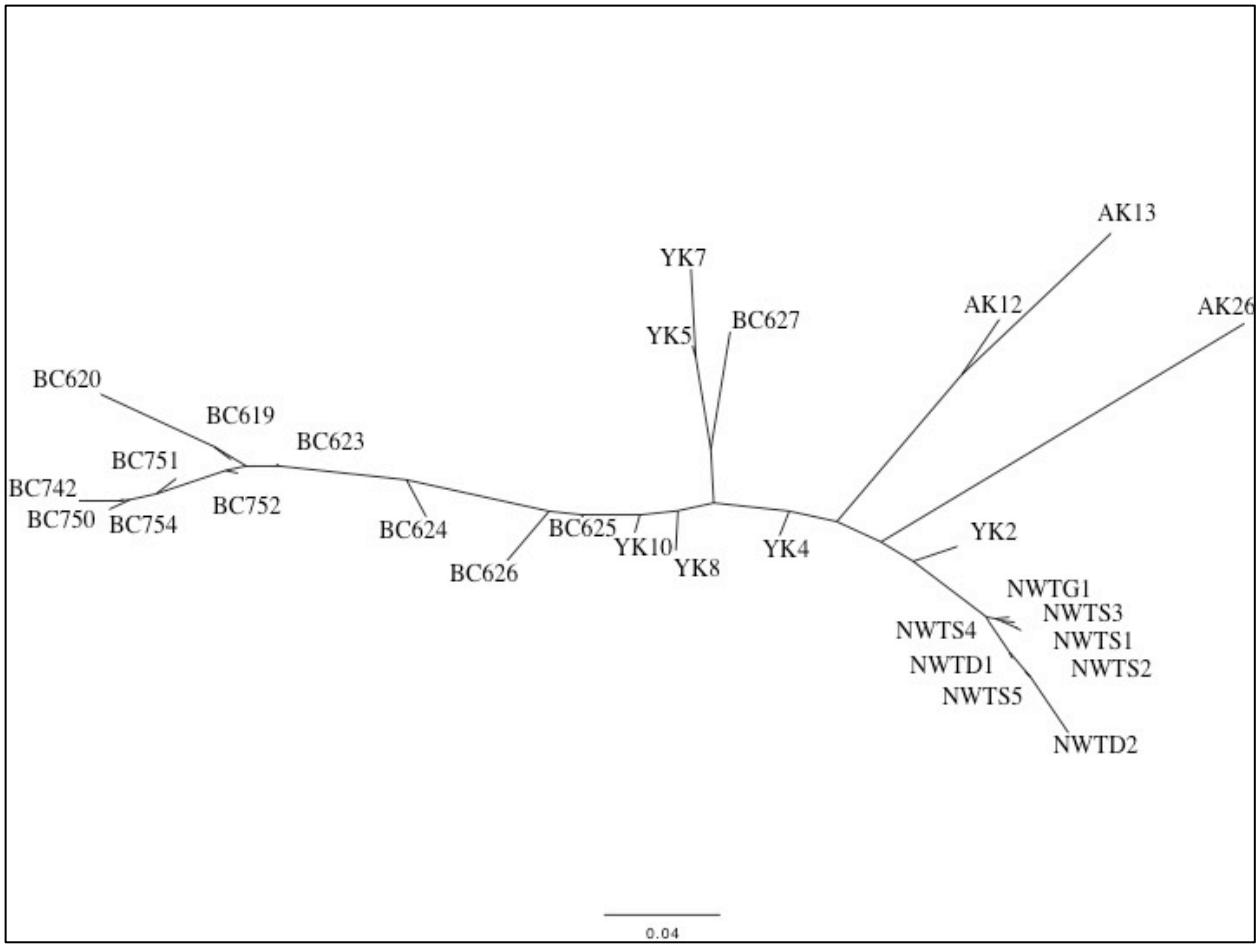

Figure 4.2 Unrooted neighbor joining population tree constructed using pair-wise F_{ST} values. Scale bar represents pairwise F_{ST}.

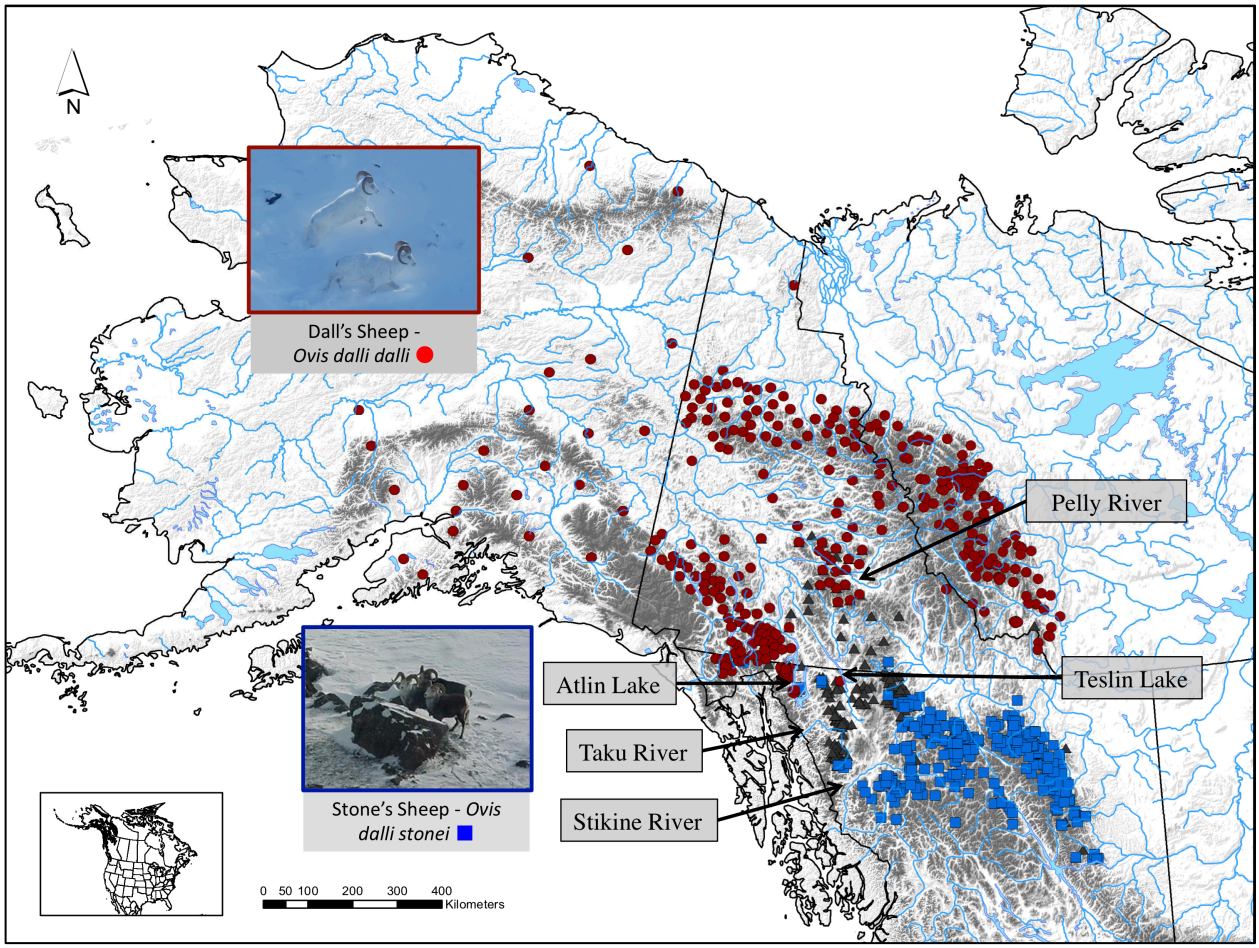

Figure 4.3 Map for species-wide STRUCTURE analysis at *k*=2 with samples plotted with subspecies identity (●=Dall's sheep cluster; ■=Stone's sheep cluster; ▲=Admixed). Representative pictures for mature Dall's and Stone's sheep rams are shown (photo credit: Bill Jex).

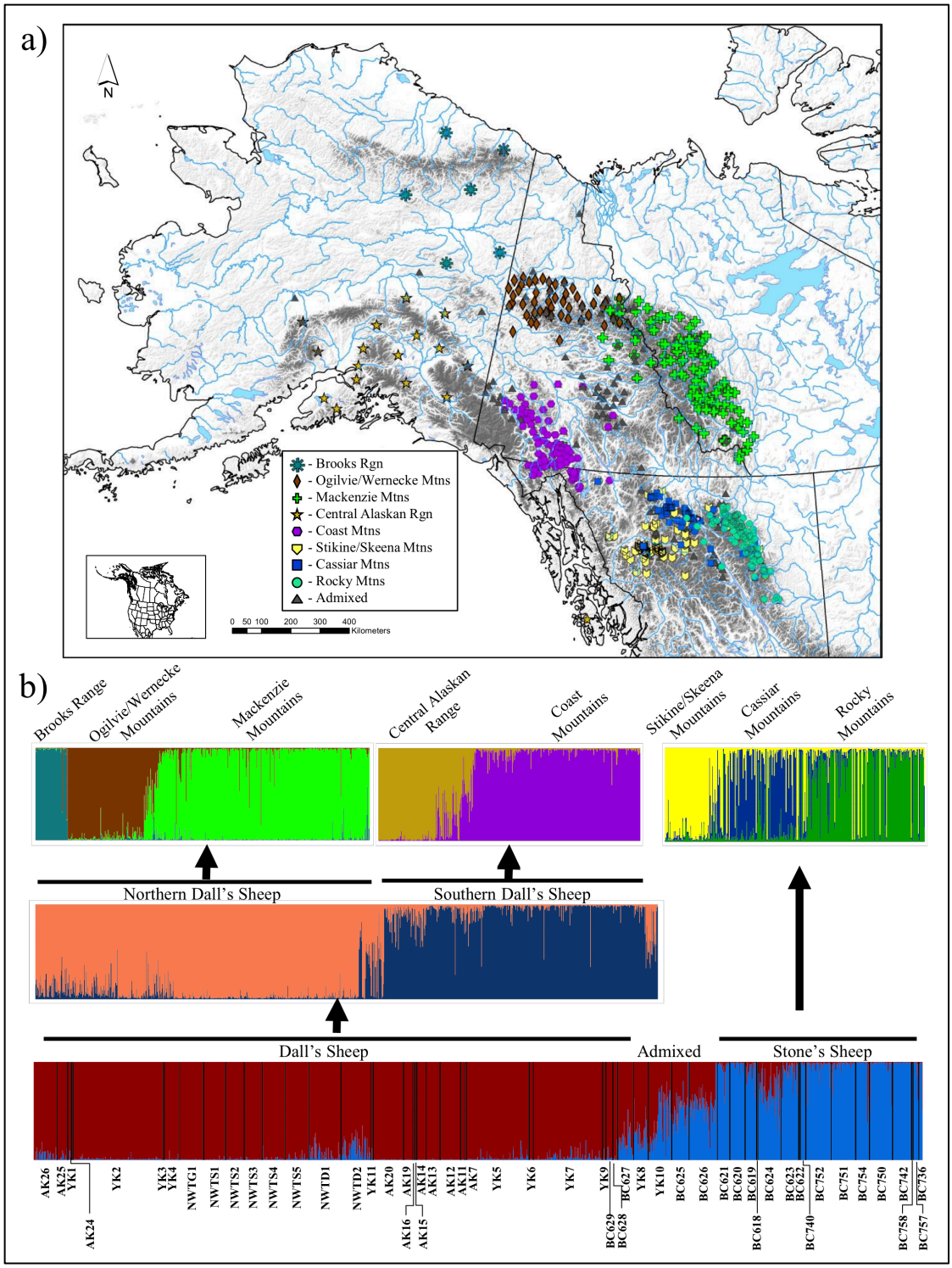

Figure 4.4 **a)** Map of sampled thinhorn sheep individuals with cluster identities assigned by STRUCTURE analysis. **b)** STRUCTURE barplot of each level of our hierarchical STRUCTURE analysis. The arrows indicate the progression from species-wide to within subspecies level population structure. Sample localities are as described in Table 1.

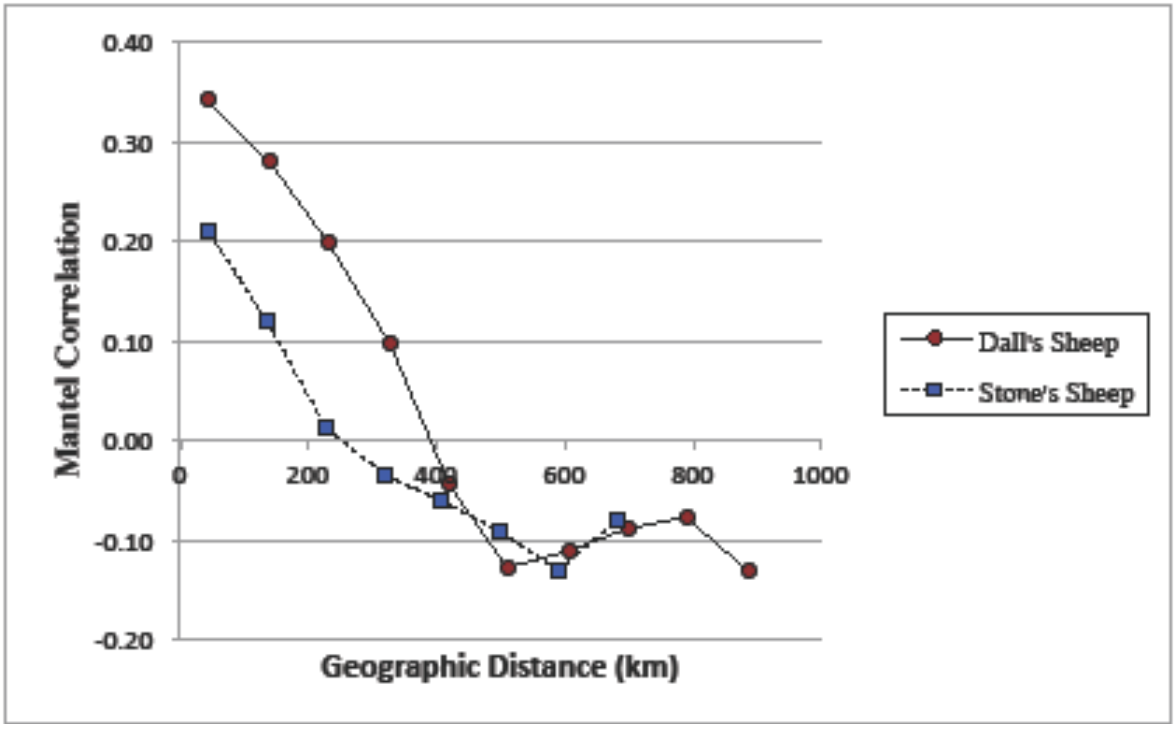

Figure 4.5 Mantel correlograms for spatial autocorrelation in Dall's and Stone's sheep

4.7 References

Allen JA (1902) A new sheep from the Kenai Peninsular. *Bulletin of the American Meseum of Natural History* **16**, 145-148.

Avise JC (2000) *Phylogeography: the history and formation of species* Harvard University Press, Cambrige, MA.

Borcard D, Legendre P (2012) Is the Mantel correlogram powerful enough to be useful in ecological analysis? A simulation study. *Ecology* **93**, 1473-1481.

Brown RR, Davis CS, Leys SP (2017) Clones or clans: the genetic structure of a deep-sea sponge, Aphrocallistes vastus, in unique sponge reefs of British Columbia, Canada. *Molecular Ecology* **26**, 1045-1059.

Brunsfeld S, Sullivan J, Soltis D, Soltis P (2001) Comparative phylogeography of northwestern North America: A synthesis. In: *Integrating Ecology and Evolution in a Spatial Context* (eds. Silvertown J, Antonovics J), pp. 319-339. Blackwell Science, London.

Demarchi R, Hartwig C (2004) Status of Thinhorn Sheep in British Columbia. *British Columbia Ministry of Water, Land and Air Protection, Biodiversity Branch - Wildlife Bulletin No. B-119*.

Diniz-Filho JAF, Soares TN, Lima JS*, et al.* (2013) Mantel test in population genetics. *Genetics and Molecular Biology* **36**, 475-485.

Dobeš C, Mitchell-Olds T, Koch MA (2004) Intraspecific diversification in North American Boechera stricta (= Arabis drummondii), Boechera ×divaricarpa, and Boechera holboellii (Brassicaceae) inferred from nuclear and chloroplast molecular markers—an integrative approach. *American Journal of Botany* **91**, 2087-2101.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14**, 2611-2620.

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**, 479-491.

Faircloth BC, Terhune TM, Schable NA*, et al.* (2009) Ten microsatellite loci from Northern Bobwhite (Colinus virginianus). *Conservation Genetics* **10**, 535-538.

Geist V (1971) *Mountain sheep* Chicago Press, Chicago, IL.

Gilbert KJ (2016) Identifying the number of population clusters with structure: problems and solutions. *Molecular Ecology Resources* **16**, 601-603.

Gilbert KJ, Andrew RL, Bock DG*, et al.* (2012) Recommendations for utilizing and reporting population genetic analyses: the reproducibility of genetic clustering using the program structure. *Molecular Ecology* **21**, 4925-4930.

Guillot G, Rousset F (2013) Dismantling the Mantel tests. *Methods in Ecology and Evolution* **4**, 336-344.

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.

Hengeveld PE, Cubberley JC (2011) Stone's sheep population dynamics and habitat use in the Sulphur / 8 Mile oil and gas pre-tenure plan area, northern British Columbia, 2005 – 2010. Synergy Applied Ecology, Mackenzie, British Columbia.

Janes JK, Miller JM, Dupuis JR*, et al.* (2017) The K = 2 conundrum. *Molecular Ecology* **26**, 3594-3602.

Jex B, Ayotte J, Bleich V*, et al.* (2016) Thinhorn sheep: conservation challenges and management strategies for the 21st Century. Wild Sheep Working Group, Western Association of Fish and Wildlife Agencies, Boise, Idaho, USA.

Kijas JW, Porto-Neto L, Dominik S*, et al.* (2014) Linkage disequilibrium over short physical distances measured in sheep using a high-density SNP chip. *Animal Genetics* **45**, 754-757.

Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I (2015) Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources* **15**, 1179-1191.

Kuchta SR, Tan A-M (2005) Isolation by distance and post-glacial range expansion in the rough-skinned newt, Taricha granulosa. *Molecular Ecology* **14**, 225-244.

Latch EK, Reding DM, Heffelfinger JR, Alcalá-Galván CH, Rhodes OE (2014) Range-wide analysis of genetic structure in a widespread, highly mobile species (Odocoileus hemionus) reveals the importance of historical biogeography. *Molecular Ecology* **23**, 3171-3190.

Legendre P, Fortin M-J (2010) Comparison of the Mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. *Molecular Ecology Resources* **10**, 831-844.

Legendre P, Fortin M-J, Borcard D (2015) Should the Mantel test be used in spatial analysis? *Methods in Ecology and Evolution* **6**, 1239-1247.

Loehr J, Worley K, Grapputo A*, et al.* (2006) Evidence for cryptic glacial refugia from North American mountain sheep mitochondrial DNA. *Journal of Evolutionary Biology* **19**, 419-430.

Meirmans PG (2012) The trouble with isolation by distance. *Molecular Ecology* **21**, 2839-2846.

Meirmans PG (2013) Nonconvergence in Bayesian estimation of migration rates. *Molecular Ecology Resources* **14**, 726-733.

Meirmans PG, Van Tienderen PH (2004) genotype and genodive: two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes* **4**, 792-794.

Michalakis Y, Excoffier L (1996) A Generic Estimation of Population Subdivision Using Distances Between Alleles With Special Reference for Microsatellite Loci. *Genetics* **142**, 1061-1064.

Morin PA, Martien KK, Taylor BL (2009) Assessing statistical power of SNPs for population structure and conservation studies. *Molecular Ecology Resources* **9**, 66-73.

Moritz C (1994) Defining "Evolutionarily Significant Units" for conservation. *Trends in Ecology & Evolution* **9**, 373-375.

Nei M (1987) *Molecular Evolutionary Genetics* Columbia University Press, New York, NY.

Oksanen J, Kindt R, Legendre P*, et al.* (2018) vegan: Community Ecology Package. **version 2.5-2**.

Paradis E, Strimmer K, Claude J, Noel Y, Bolker B (2003) Analyses of Phylogenetics and Evolution: the ape Package. *Bioinformatics* **20**, 289–290.

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

Purcell S, Neale B, Todd-Brown K*, et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics* **81**, 559-575.

Rambaut A, Suchard M, Xie D, Drummond A (2014) Tracer v1. 6. *Available from Available at:* http://beast. *bio. ed. ac. uk/Tracer*.

Rannala B (2015) BayeAss Edition 3.0 User's Manual.

Roffler G, Talbot S, Luikart G*, et al.* (2014) Lack of sex-biased dispersal promotes fine-scale genetic structure in alpine ungulates. *Conservation Genetics*, 1-15.

Roffler GH, Schwartz MK, Pilgrim KL*, et al.* (2016) Identification of landscape features influencing gene flow: How useful are habitat selection models? *Evolutionary Applications* **9**, 805-817.

Rousset (2000) Genetic differentiation between individuals. *Journal of Evolutionary Biology* **13**, 58-62.

Shafer ABA, Côté SD, Coltman DW (2011) Hot spots of genetic diversity descended from multiple pleistocene refugia in an alpine ungulate. *Evolution* **65**, 125-138.

Shafer ABA, Cullingham CI, Côté SD, Coltman DW (2010) Of glaciers and refugia: a decade of study sheds new light on the phylogeography of northwestern North America. *Molecular Ecology* **19**, 4589-4621.

Sheldon C (1911) *The wilderness of the Upper Yukon; a hunter's explorations for wild sheep in sub-Arctic Mountains* T. Fisher Unwin, New York.

Sim Z, Hall JC, Jex B, Hegel TM, Coltman DW (2016) Genome-wide set of SNPs reveals evidence for two glacial refugia and admixture from postglacial recolonization in an alpine ungulate. *Molecular Ecology* **25**, 962-1083.

Untergasser A, Cutcutache I, Koressaar T*, et al.* (2012) Primer3—new capabilities and interfaces. *Nucleic Acids Research* **40**, e115.

Vähä J-P, Erkinaro J, Niemelä E, Primmer CR (2007) Life-history and habitat features influence the within-river genetic structure of Atlantic salmon. *Molecular Ecology* **16**, 2638-2654.

Valdez R, Krausman PR (1999) *Mountain sheep of North America* University of Arizona Press.

Waples RS, Gaggiotti O (2006) INVITED REVIEW: What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular Ecology* **15**, 1419-1439.

Willing E-M, Dreyer C, van Oosterhout C (2012) Estimates of Genetic Differentiation Measured by FST Do Not Necessarily Require Large Sample Sizes When Using Many SNP Markers. *Plos One* **7**, e42649.

Wilson GA, Rannala B (2003) Bayesian Inference of Recent Migration Rates Using Multilocus Genotypes. *Genetics* **163**, 1177.

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

Zink RM, Blackwell-Rago RC, Ronquist F (2000) The shifting roles of dispersal and vicariance in biogeography. *Proceedings of the Royal Society B-Biological Sciences* **267**, 497-503.

Chapter 5: Heritability and genomic architecture of horn size in thinhorn sheep

5.1 Abstract

Understanding the genetic basis of fitness-related trait variation has long been of great interest to evolutionary biologists. Secondary sexual characteristics, such as horns in bovids, are particularly intriguing since they can be potentially affected by both natural and sexual selection. Until recently however, the study of fitness-related quantitative trait variation in wild species has been hampered by a lack of genomic resources, pedigree and/or phenotype data. Recent innovations in genomic technologies have enabled wildlife researchers to perform marker-based relatedness estimation and acquire adequate loci density, enabling both the "top-down" approach of quantitative genetics and the "bottom-up" approach of association studies to describe the genetic basis of fitness-related traits. Here we combine a cross species application of the OvineHD BeadChip, and horn measurements (horn length, base circumference and volume) from harvested thinhorn sheep to examine the heritability and to perform a genome-wide SNP association study of horn size in the species. Thinhorn sheep are a mountain ungulate that resides in mountainous regions of northwestern North America. Thinhorn sheep males grow massive horns that determine the social rank and mating success. We found horn length, base circumference and volume to be moderately heritable and two loci to be suggestively associated with horn length.

5.2 Introduction

The genetic basis of trait diversity is a fundamental area of inquiry in evolutionary biology. Knowledge regarding the generation, inheritance and maintenance of variation strikes at the core of our understanding of evolution. Genes that underlie fitness related traits are of particular interest since selection is thought to act most strongly on these relationships (Ellegren & Sheldon 2008). Key areas of inquiry include questions on the role of additive genetic variation (Lynch & Walsh 1998) and the elucidation of genomic architectures (Slate *et al.* 2009; Slate *et al.* 2010) in quantitative traits. Until recently, the study of fitness-related quantitative trait variation has been hampered by a lack of genomic resources, pedigree and/or phenotype data and has thus been restricted to either the laboratory or a select few wild species under long-term study (Kruuk &

Hill 2008; Slate *et al.* 2010).

Recent advances in genomic technologies have ushered in a new age of inquiry into the genetic basis of traits by dramatically lowering the per-unit cost of obtaining genetic data, particularly for non-model organisms (Davey *et al.* 2011; Helyar *et al.* 2011). This drop in price and ease of genetic data collection has advanced our ability to feasibly investigate a large enough number of loci to reasonably interrogate the genome for associations with phenotype (Garvin *et al.* 2010). Furthermore, new methods in relatedness estimation enabled by large genomic datasets have also allowed us to overcome the imprecision of those estimated using smaller marker sets and thus the need for difficult to obtain pedigree information (Coltman 2005; Csilléry *et al.* 2006; Gienapp *et al.* 2017). These key gains in our ability to perform marker-based relatedness estimation and acquire adequate loci density enabled by second and third generation DNA technologies opens the door for us to utilize both the "top-down" approach of quantitative genetics (Gienapp *et al.* 2017) and the "bottom-up" approach of association studies (Santure & Garant 2018) to describe the genetic basis of fitness-related traits.

Secondary sexual characteristics are intriguing targets of inquiry since they are potentially under both natural and sexual selection. While beneficial alleles may be expected to be driven to fixation by either process, the presence of both may yield counter-balancing selection pressures that maintain genetic variation (Kruuk *et al.* 2008). Horns in bovids, such as thinhorn sheep, are striking examples of a secondary sexual characteristic subject to sexual selection. Male mountain sheep grow massive horns that make up 8-12% of their body weight (Feldhamer *et al.* 2003). During the mating season, large horned males are more dominant and more likely to mate (Coltman *et al.* 2002; Geist 1971; Hogg 1984). The same relationship is not observed in females (Favre *et al.* 2008). From a management perspective, horn length (along with age) is one of two components that determine the legal status of a ram for harvest in most jurisdictions. As a practical effect, horn size based harvest restrictions may result in the preferential removal of large horned and/or fast growing males. If horn size is heritable, then these regulations can result in selection against large horned individuals, thus potentially effecting an evolutionary response in harvested thinhorn sheep populations. In the closely related bighorn sheep, similar harvest regulations have been found to result in both demographic (Festa-Bianchet *et al.* 2013) and

evolutionary (Coltman *et al.* 2003) changes in hunted populations.

Quantitative genetic studies of the closely related bighorn sheep have found horn size to be moderately to highly heritable (Coltman *et al.* 2007; Coltman *et al.* 2003; Miller *et al.* 2018; Poissant *et al.* 2008). Microsatellite-based QTL mapping has also found regions suggestively associated with horn morphology (Poissant *et al.* 2011). A region identified as a possible QTL for horn dimension in Poissant *et al.* (2011) was also found by a genome resequencing study to show signatures of a selective sweep in a separate population of bighorn sheep (Kardos *et al.* 2015). This region contains the gene coding for Relaxin-like receptor 2 (RXFP2), which has been shown to strongly influence horn development in domestic sheep and underwent strong positive selection due to artificial breeding for individuals lacking horns (Kijas *et al.* 2012). In a feral breed of domestic sheep, the Soay sheep of St. Kilda archipelago (Scotland), RXFP2 has been found to be strongly associated with discrete and quantitative variation in horn phenotype (Johnston *et al.* 2011). A more recent study found that variation in RXFP2 in Soay sheep is maintained by a life history trade-off - the allele conferring greater fecundity is associated with lower survival (Johnston *et al.* 2013). RXFP2 has been found in humans and mice to be positively correlated with testosterone levels in blood, while mutations in RXFP2 have been found to be associated with osteoporosis (Ferlin *et al.* 2008) and testicular descent (Feng *et al.* 2009) in mice and humans.

To date, no study has investigated the heritability or additive genetic variance of a fitness-related trait in thinhorn sheep. While some studies have sought to investigate candidate loci for association with pelage colour (Loehr *et al.* 2008) and signatures of selection (Worley *et al.* 2006), none has yet performed a genome-wide association study of a sexual selected trait with thousands of SNPs. In this study, we combine a cross-species application of a high density domestic SNP array, the OvineHD BeadChip, and horn measurements collected during regulatory inspections of harvested thinhorn sheep to examine the genetic architecture of horn size in thinhorn sheep. First, we estimate the heritability of three horn size metrics: 1) horn length, 2) base circumference, and 3) volume using an “animal model” (Kruuk 2004), a linear mixed effects model used in animal breeding. Then, we perform a genome-wide association analysis between SNP markers and each of the three horn size metrics. This is first study of its

kind for thinhorn sheep and one of the few for a wild species not under long-term study. Since harvest regulations for thinhorn sheep are in part based on minimum horn size, understanding the heritability and genomic architecture of horn size traits can also inform the conservation management of the species.

5.3 Methods

5.3.1 Sample origins and horn measurements

Horn measurements from 192 individuals were collected from 2013 - 2015 from hunter harvested Dall's sheep (*Ovis dalli dalli*) in game management units 5, 7 and 9 in Yukon, Canada (Figure 5.1). Dall's sheep are a northern subspecies of thinhorn sheep, which is one of two closely related mountain sheep species in North America (Valdez & Krausman 1999). Dall's sheep are notable in being the only white-coloured mountain sheep subspecies in North America. Populations of Dall's sheep occupy mountainous regions in Alaska, Yukon, western Northwest Territories and northwestern British Columbia (Feldhamer *et al.* 2003).

Hunting regulations in Yukon stipulate that only rams over the age of eight or possessing horns that "extend beyond a line running from the centre of the nostril to the lowermost edge of the eye" (also known as full curl) may be legally harvested (Yukon Hunting Regulations 2018-2019; Figure 5.2). Hunters are required to submit harvested Dall's sheep rams for registration during which measurements for 1) horn length, 2) horn base circumference, and 3) annuli length are taken. Horn length is measured from tip to base of the horn following the outside curvature of the horn using a flexible measuring tape. The longer of the right and left horn is reported. Horn base circumference is the circumference measurement of each annual growth segment. Horn growth in Dall's sheep occurs in between April to September according the seasonal patterns of North America (Bunnell 1978). The cessation of horn growth after the growing season creates annual growth rings, or annuli, which can be used to estimate the age of an individual (Geist 1966; Hemming 1969). Annuli length is the measurement between two growth segments. Annuli can be used to estimate the horn dimensions at the end of each preceding growing season using the proxy of annuli as representing the horn base for those years. We estimated the horn volume using the formula for a conical frustum: Volume = $1/3\pi H(r_1^2+r_1r_2+r_2^2)$. Where r_1 and r_2 represent

the base radii at either end of an annual growth segment and H is the length of the segment (Heimer & Smith III 1975).

5.3.2 Quality control

The horns of most rams show some degree of wear or less commonly breakage, which may make the first annulus difficult to distinguish. A missed first annulus will lead to the growth of year one (lamb tips) and year two growth to be recorded as a single growth increment, resulting in an overestimate of year one growth and underestimate for the subsequent years. Furthermore, age determination will also be in error. To mitigate the effect of mis-identified or missing first annulus, we removed individuals with biologically implausible growth increments i.e. year one growth > 160mm and sum of year one and two > 420mm (Bunnell 1978; Hik & Carey 2000). Horns with a visible first annulus but worn lamb tips may still result in an underestimate of year one growth. Therefore, we also 1) removed all measures of year one horn length and base circumference, 2) subtracted year one growth (tip to first annulus) for all horn length measurements, and 3) did not calculate the horn volume of the lamb tips.

5.3.3 SNP genotyping and quality control

We extracted DNA from 192 samples of Dall's sheep rams with the Qiagen DNeasy Blood and Tissue Kit (Qiagen) using standard protocol. We quantified the extracted DNA using the Qubit Fluorometer (Life Technologies) and normalized to 50 ng/uL in preparation for genotyping. We genotyped the sampled individuals using the OvineHD SNP BeadChip following the manufacturer's protocol (Qiagen). The OvineHD SNP BeadChip is a SNP array containing 606,006 markers originally designed for use in domestic sheep (*Ovis aries*) by the International Sheep Genomics Consortium (ISGC) (Kijas *et al.* 2014). Raw signals were converted into genotype calls using a custom cluster file provided by the ISGC, which was developed using a multibreed panel of 288 *O. aries* individuals (J. McEwan, unpublished), using the software GENOMESTUDIO (Illumina). We also used GENOMESTUDIO to cull low quality genotype calls using a GenCall (GC) score threshold of 0.8. The GC score of genotype call is an assessment of cluster quality based on how tightly clustered the raw signal of that genotype call is compared to other calls of identical genotype and can range from 0-1 (higher is better). Post GC score quality controlled genotype calls were exported in PLINK format using a custom plug-in (Illumina).

We used PLINK v1.07 (Purcell *et al.* 2007) to remove all individuals or loci with genotyping rate of <0.9, x-linked loci (based on assumed synteny with domestic sheep), and/or minor allele frequency of < 0.01. We also performed a check for Hardy-Weinberg equilibrium (HWE; $\alpha = 0.001$) but did not remove any loci due to HWE deviations since loci under selection, thus potentially associated with horn size, are expected to be out of HWE (Turner *et al.* 2011). We interrogated loci significantly associated with horn size for false positives due to deviations with HWE.

5.3.4 SNP-based quantitative genetics and genome-wide association

We used the R packages GENABEL (Aulchenko *et al.* 2007; Karssen *et al.* 2016) and its extension for repeated measures, REPEATABEL (Rönnegård *et al.* 2016), to perform a genome wide association study and estimate variance components for phenotypic variation.
Repeated measures data generally result from longitudinal studies, however, annualized pattern growth due to the cessation of horn growth each winter allows for the estimation of horn measurements in previous years. The use of repeated measures allows for better estimates of within-individual variation and has been found to increase the power of genome-wide association studies by providing year-to-year variation in a trait measurement (Rönnegård *et al.* 2016).

First, we used REPEATABEL to fit a linear mixed model to estimate variance component assuming no SNP effects. Our mixed model resembles an "animal model" where phenotypic variance (Vp) is partitioned into random effects of additive genetic (Va), permanent environment (Vpe), cohort (Vyb), year of measurement (Vym) and residual variation (Vr) after accounting for age as a fixed effect (Vp=Va+Vpe+Vyb+Vym+Vr). Permanent environmental effect was calculated from repeated measures of the same individual to account for variation associated with environment condition effects specific to that individual. Additive genetic variation was estimated using a genomic relationship matrix. Narrow sense heritability was calculated as the ratio of additive genetic variation and overall phenotypic variation (h^2=Va/Vp).

Second, we used the (co)variance matrix calculated in the first model fitting step to test for associations of individual SNPs with horn size measures using ordinary least squares. P-values

for SNP associations were calculated using Wald tests. We defined genome wide significant and suggestive significance of SNP association using 0.05/nSNPs and 1/nSNPs respectively (nSNPs - number of markers). In association analysis containing repeated measures of potentially related individuals there are concerns that significance may be inflated because 1) population stratification may overestimate SNP effects and 2) repeated measures of the same individual may be correlated. By using the (co)variance matrix constructed in the first model-fitting step, we can account for relatedness by using the GRM and within individual variance via the estimation of permanent environmental effects (Rönnegård *et al.* 2016). We calculated the genomic inflation factor (λ) (post genomic control) for each model by using GENABEL to perform a regression analysis of observed versus expected p-values (Aulchenko *et al.* 2007). We reviewed gene annotations of significantly associated loci in the *O. aries* genome (assembly 3.1, ISGC).

5.4 Results

5.4.1 Data acquisition and quality control

We acquired SNP genotypes and horn size measurement data for 192 Dall's sheep rams from southwestern Yukon, Canada (Figure 5.3, Table 5.1). Five individuals were excluded due to low call rates (<0.9) resulting in an overall per individual call rate of >0.989 for the remaining individuals. Subsequently, we excluded 131,836 loci due to poor cluster quality (GC score <0.8), 6,580 loci due to low locus-specific call rate (call rate <0.9), 154 loci for being x-linked and 460,801 for having MAF < 0.01, resulting in a SNP dataset of 6635 loci representing each chromosome. This degree of polymorphism is in line with other wild sheep studies employing domestic sheep based SNP chips for genotyping (Miller *et al.* 2011; Sim *et al.* 2016). A further seven individuals were culled for being biologically implausible and/or potential recording errors resulting in a final data set of 180 individuals genotyped at 6635 SNPs.

5.4.2 SNP-based quantitative genetics and GWAS

Marker based estimates of heritability for horn measures ranged from 0.33 - 0.36 (Table 5.1). Of the non-genetic effects, we found small but significant effects for year of birth and year of measurement (0.01 - 0.06) while permanent environmental effects were relatively larger (0.29 - 0.31). Manhattan plots for traits measured and associated Q-Q plots are shown in Figures 5.4 and

5.5, respectively. We found little to no evidence of genomic inflation post genomic control (all λ ~1; Figure 5,5).

No loci were found to be associated with any of the horn size traits at the genome-wide significance level. Two loci, 1) OAR2_43601714 and 2) OAR3_134140997 (both in HWE) showed suggestive association with horn length (Figure 4). OAR2_43601714 is located on chromosome 2 in the gene GFRA2 (GDNF family receptor alpha). In domestic sheep, GFRA2 codes for receptors that interact with glial cell line-derived neurotrophic factors (GDNF) (Jing *et al.* 1997). OAR3_134140997 is located on chromosome 3 in the putative gene FIGNL2 (Fidgetin-like Protein 2). In mammals, fidgetin like proteins belong to a family of ATPases associated with embryonic development (Cox *et al.* 2000; Frickey & Lupas 2004).

5.5 Discussion

In this study, we investigated the genetic basis of fitness-related traits in Dall's sheep by genotyping 192 animals using a cross-species application of a domestic sheep high density SNP array (over 600,000 loci). We achieve a typical conversion rate of about 1% (Miller *et al.* 2012a; Sim *et al.* 2016), which resulted in SNP panel of ~6000 loci, comparable to that of a medium density SNP array. By combining the genotype data with horn dimension measurements taken during compulsory inspections of harvested rams, we show that horn length, base circumference and volume are moderately heritable in thinhorn sheep. Point estimates of narrow sense heritability ranged from 0.33-0.36. This level of heritability is comparable to estimates of similar traits using pedigree data in the closely related bighorn sheep (*O. canadensis*). For example, Miller *et al.* (2018) found horn length and base circumference to be moderately heritable (0.15 and 0.23, respectively) in a population of bighorn in Ram Mountain, Alberta, Canada, while Poissant *et al.* (2008) found heritability of male horn volume to be 0.32 in the same bighorn sheep population. Studies of quantitative traits using genome-wide SNP data in place of pedigrees in other wild mammalian (Malenfant *et al.* 2018) and avian (Kardos *et al.* 2016; Lundregan *et al.* 2018) species has also yielded similar heritability values.

Our results also indicate the influence of permanent environmental effects on horn size to be moderate (0.29 - 0.34). This result is consistent with a previous study on Dall's sheep in Yukon,

which found that spring weather explained a large proportion of annual variation (0.18-0.46) in horn growth (Loehr *et al.* 2010). While Loehr *et al.* (2010) did not assess the heritability of horn size traits, it found only a small individual effect on horn size (0.026 - 0.079), which the authors argue indicate a small role for additive genetic on horn size. However, in the same paper, the authors noted that their inference regarding the small role of additive genetic effect on horn size did not agree with that of the moderate to high heritability estimates from bighorn sheep by Coltman *et al.* (2003), explaining that the differences could be due to the modelling choices and thus not necessarily inconsistent. Since we used an "animal model", we are unsurprised to find our inference of the relative role of additive genetic effect on horn size is similar to that of Coltman *et al.* (2003), which also used an "animal model", and thus not necessarily contradictory to that of Loehr *et al.* (2010).

We found two loci of suggestive association with horn length that mapped to genes GFRA2 and FIGNL2 in the domestic sheep genome. In humans, GFRA2 is a protein-coding gene for a co-receptor in the GDNF family of neurotrophic ligands (Jing *et al.* 1997). This family of ligands is involved in transmembrane signal transduction and plays an important role in the development of the central and peripheral nervous system (Lin *et al.* 1993). FIGNL2 is less well-studied, and a putative gene inferred from the human genome. In humans, FIGNL2 is thought to be a paralog of the gene Fidgetin (FIGN) which belongs to a superfamily of genes that code for ATPases associated with diverse cellular activities proteins (AAA proteins). This superfamily is made up of a wide range of molecular chaperones that facilitate cellular functions such as proteolysis and membrane fusion characterized by a common conserved ATP-binding domain of ~240 residues (Lupas & Martin 2002). While biologically plausible post-hoc rationales can be invoked regarding the mechanism underlying these associations we do not see an immediate connection. This and the suggestive nature of the statistical association indicate caution against over-interpretation. Ideally, these results should provide starting material for a validation study with finer scale genomic coverage using a separate population of Dall's sheep and more individuals.

Previous studies of Soay sheep, a free living feral breed of domestic sheep found on St. Klida archipelago (Scotland), have found a strong candidate for horn morphology in the gene RXFP2, located on chromosome 10 of the domestic sheep genome (Johnston *et al.* 2011). QTLs of

suggestive significance on chromosome 10 have also been identified in bighorn sheep that are co-localized in the region mapped to RXFP2 in domestic sheep (Poissant *et al.* 2011). Our post QC genotype data contained one locus (OAR10_29685536) in RXFP2 gene and this locus is not significantly associated with any horn size trait. In the 413,000bp (found by Miller *et al.* (2018) to be the half-length of LD in bighorn sheep) region up-and-downstream of RXFP2, all 162 loci were monomorphic. However, more recent research based on whole-genome resequencing of six pooled bighorn sheep populations indicate a selective sweep of the RXFP2 region consistent with positive selection (Kardos *et al.* 2015). Similar selective sweeps of RXFP2 have also been found in domestic sheep (Kijas *et al.* 2012). So while it may have been somewhat unexpected to find no association of between RXFP2 and horn size in thinhorn sheep, the presence of selective sweeps would prevent any signatures of association from being detected using our methods. We were unable to evaluate hypotheses regarding a selective sweep since a recent study has indicated that cross-species application of SNP chips are problematic for identifying runs of heterozygosities, a typical signature of selective sweeps (Shafer *et al.* 2016).

While quantitative genetic and association studies have long been performed for some wild animals, most are part of large longitudinal studies. These types of projects can be costly and logistically challenging to operate, thus putting them out of the reach of most wildlife researchers. By taking advantage of the "genome-enabled" (Kohn *et al.* 2006) status of thinhorn sheep, we gained access to a high density SNP array which gave us a final SNP panel equivalent to that of a medium density array (~6000 SNPs) with no cost towards marker discovery. Combining this SNP panel with recently developed methods for estimating genetic co-variance using a genomic related matrix allowed us to circumvent the need for a pedigree, which are available only in very few wild species at great cost and labour. Further, we utilized annuli produced by seasonal growth patterns as a form of repeated measure, thus giving us greater statistical power than would otherwise have been available for the given sample size. Our approach leverages 1) cross species application of domestic genomic resources, 2) existing governmental databases, and 3) annualized growth patterns to cost effectively perform a first quantitative genetic and genome-wide association study for thinhorn sheep and one of the first for a wild species outside of long-term active monitoring. Any of the three prongs in our analytical approach can be used by other studies to improve the cost effectiveness, speed and/or

power of their analysis. In conservation context, our finding that horn size is heritability can be used to inform management since our results show that horn size based harvest regulations can have the potential to induce an evolutionary response in hunted populations by favouring the survival of smaller horned rams.

Table 5.1 Number of individuals (N_{ind}), number of observations (N_{obs}), means, variances and estimated random effect sizes of three horn size traits in Dall's sheep.

Trait	N_{ind}	N_{obs}	Mean (SD)	h^2	V_{pe}	V_{yrmeas}	$V_{yrbirth}$	V_r
Horn Length	180	1647	627.7 (243.0)	0.33 (0.02)	0.35 (0.02)	0.02 (0.01)	0.01 (0.01)	0.29 (0.03)
Horn Base Circumference	180	1695	280.1 (69.5)	0.36 (0.03)	0.29 (0.03)	0.06 (0.02)	0.01 (0.01)	0.28 (0.04)
Horn Volume	177	1616	4525.9 (2953.9)	0.36 (0.02)	0.31 (0.02)	0.01 (0.01)	0.01 (0.01)	0.34 (0.03)

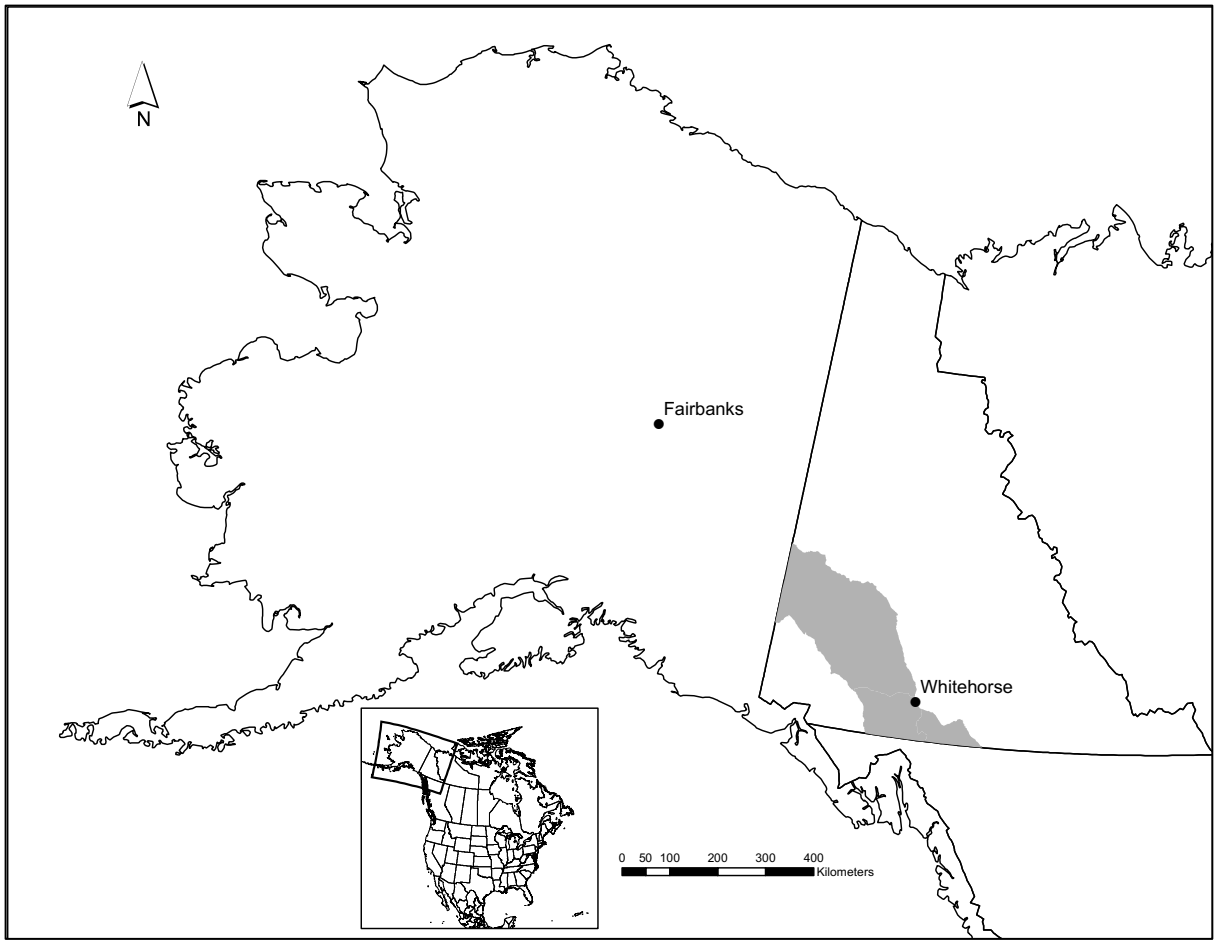

Figure 5.1 Map of study area. Shaded region represent region in southwest Yukon where samples originated.

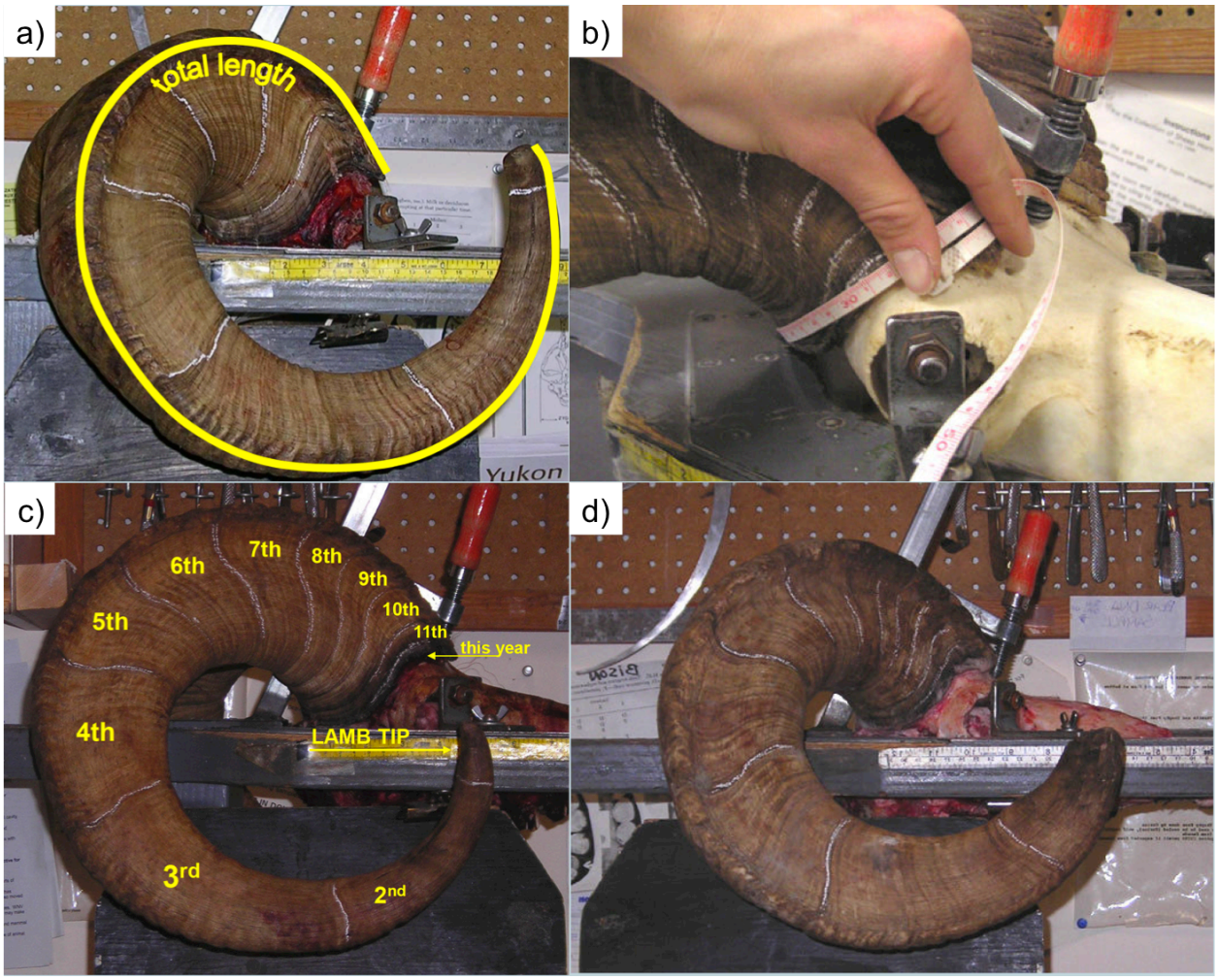

Figure 5.2 Pictures showing the horn measurement process: (a) Total horn length measurement, (b) horn base circumference, (c) marking of annuli to determine age, (d) example of horn with broken/worn lamb tip. Note that horns in (a), (b) and (c) are considered "full-curl". (Image courtesy of Yukon Ministry of Environment)

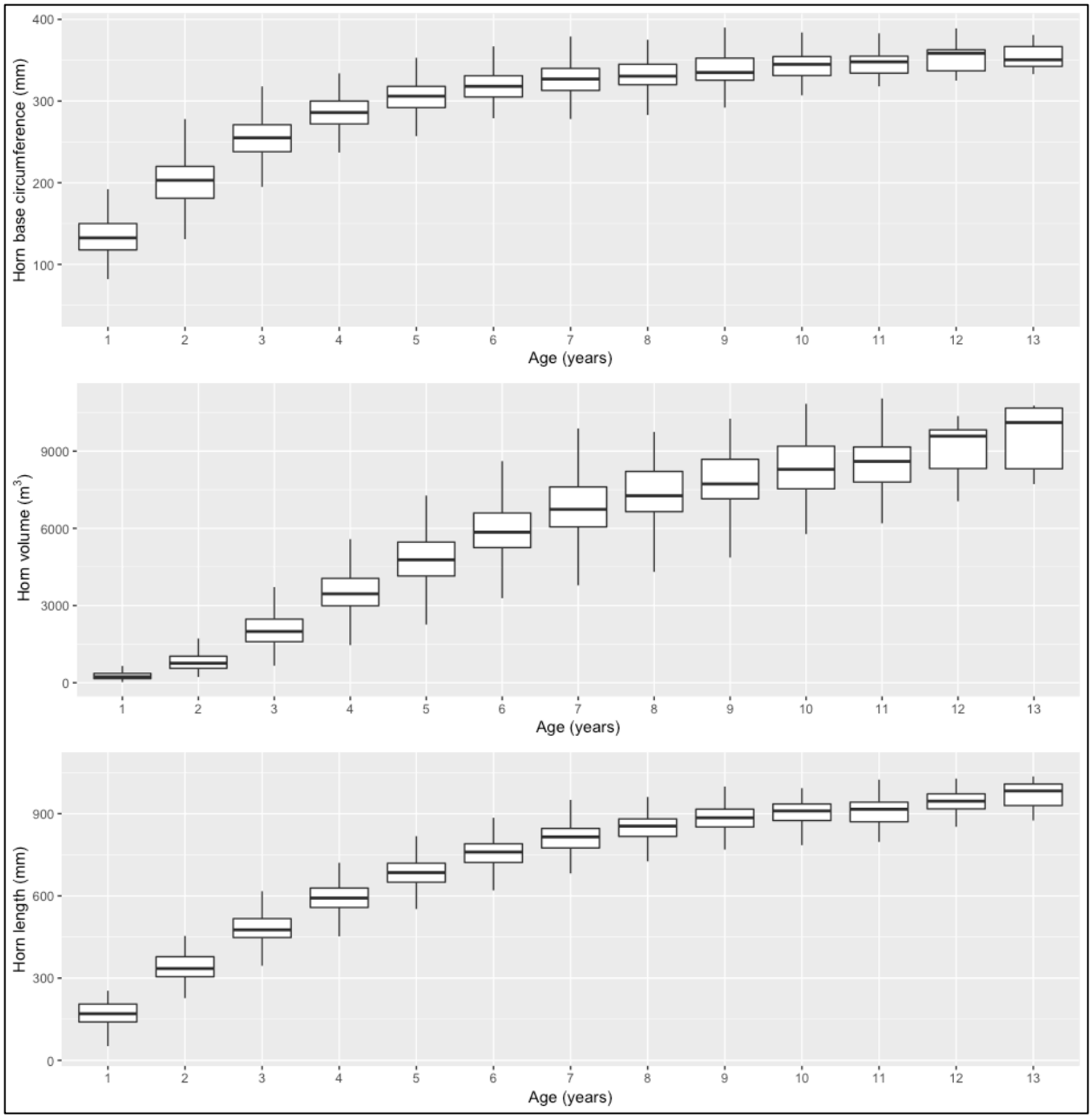

Figure 5.3 Plots of horn size traits as function of age class.

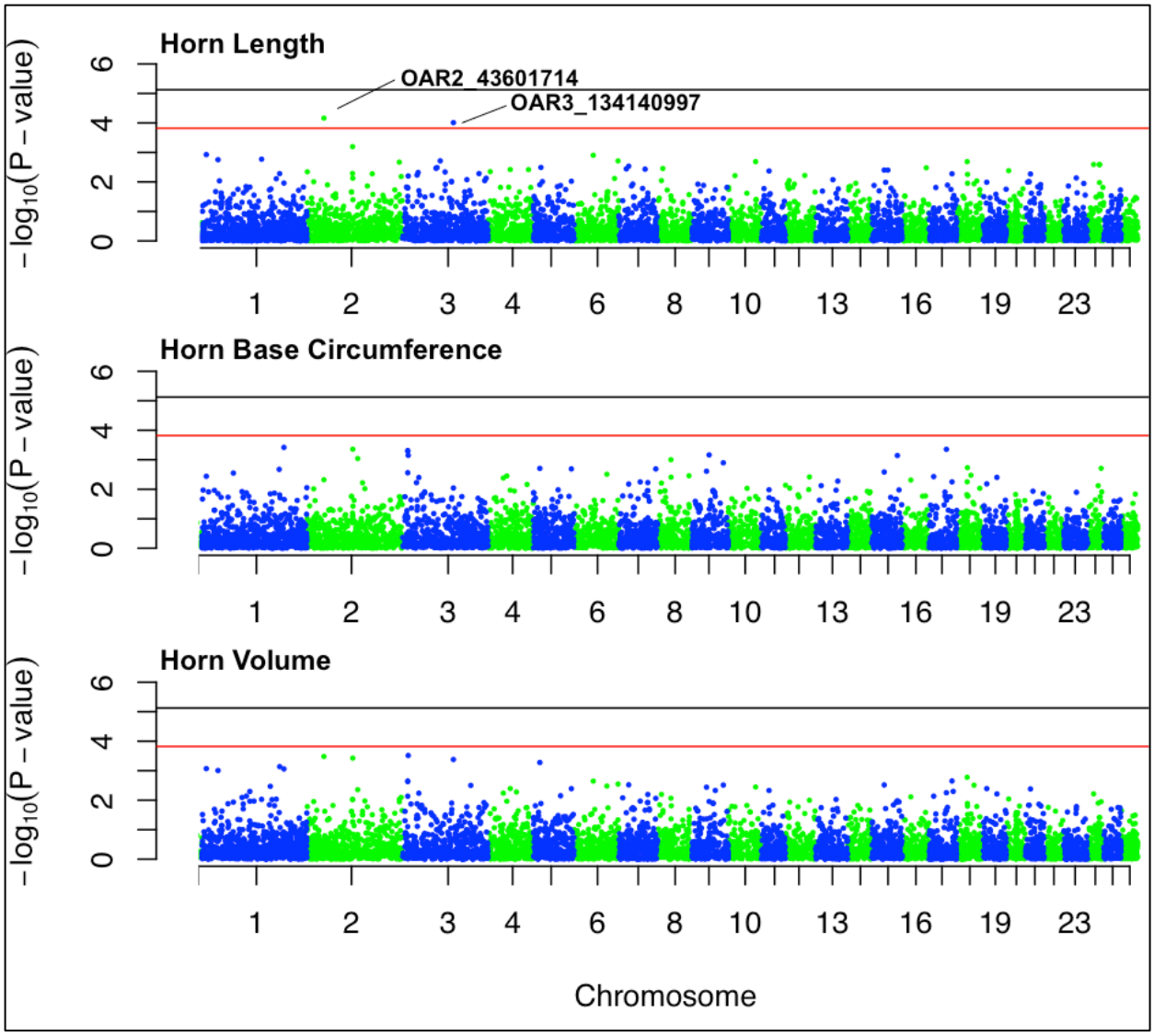

Figure 5.4 Manhattan plots for associations between SNPs and each of three horn size traits; horn length, horn base circumference, and horn volume. The black line represents the threshold for genome-wide significance and red line represents suggestive significance. The two labelled loci are of suggestive significance for horn length.

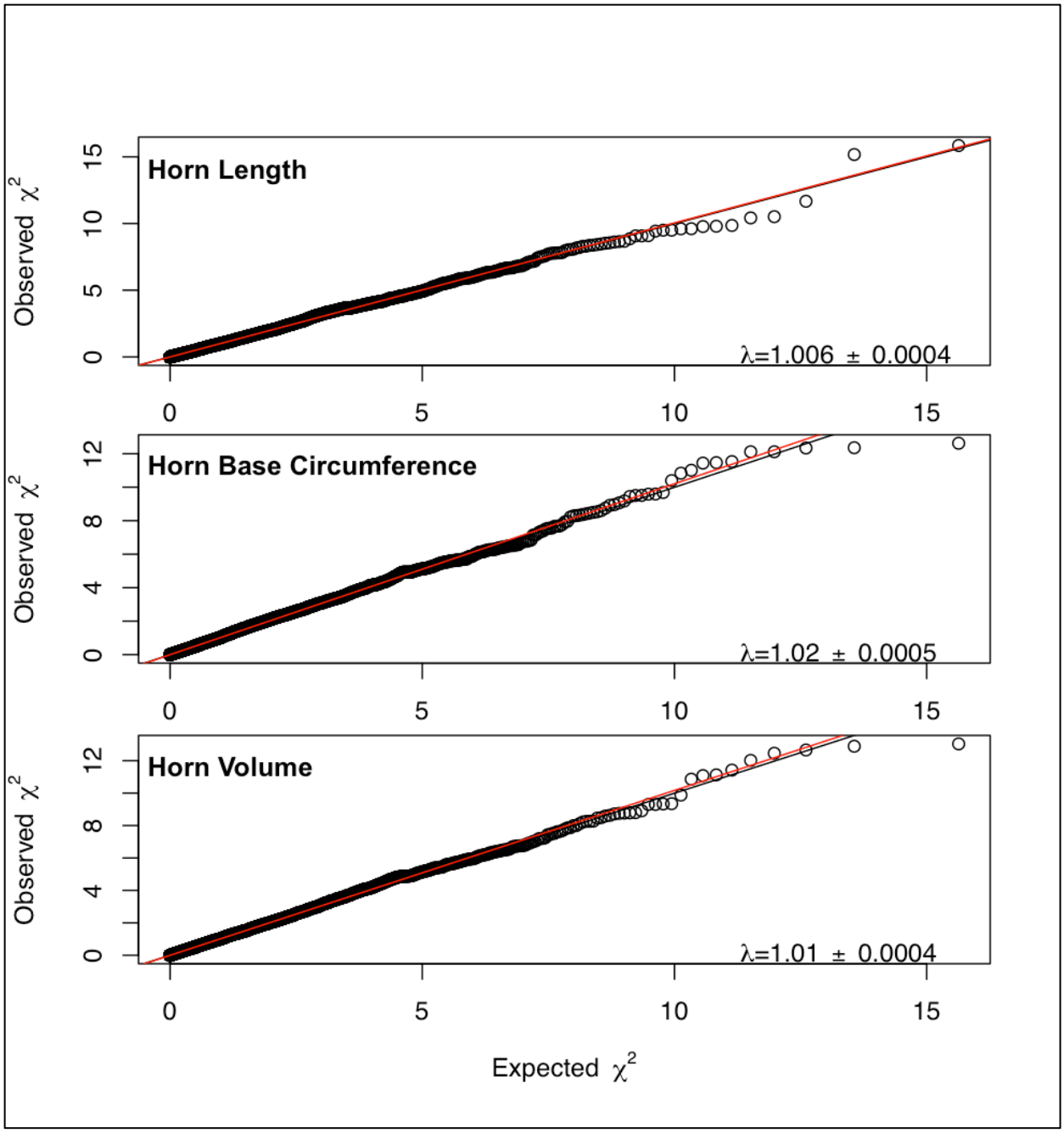

Figure 5.5 Q-Q plots for each horn size trait with the genome inflation factor and standard error printed on the bottom right corner of each plot. The black line represents a 1:1 correspondence while the red line is a regression through the observed data.

5.6 References

Aulchenko YS, Ripke S, Isaacs A, van Duijn CM (2007) GenABEL: an R library for genome-wide association analysis. *Bioinformatics* **23**, 1294-1296.

Bunnell FL (1978) Horn Growth and Population Quality in Dall Sheep. *The Journal of Wildlife Management* **42**, 764-775.

Coltman DW (2005) Testing marker-based estimates of heritability in the wild. *Molecular Ecology* **14**, 2593-2599.

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: Biological Sciences* **269**, 165.

Coltman DW, O'Donoghue P, Hogg JT, Festa-Bianchet M (2007) SELECTION AND GENETIC (CO)VARIANCE IN BIGHORN SHEEP. *Evolution* **59**, 1372-1382.

Coltman DW, O'Donoghue P, Jorgenson JT*, et al.* (2003) Undesirable evolutionary consequences of trophy hunting. *Nature* **426**, 655-658.

Cox GA, Mahaffey CL, Nystuen A, Letts VA, Frankel WN (2000) The mouse fidgetin gene defines a new role for AAA family proteins in mammalian development. *Nature genetics* **26**, 198.

Csilléry K, Johnson T, Beraldi D*, et al.* (2006) Performance of Marker-Based Relatedness Estimators in Natural Populations of Outbred Vertebrates. *Genetics* **173**, 2091.

Davey JW, Hohenlohe PA, Etter PD*, et al.* (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics* **12**, 499-510.

Ellegren H, Sheldon BC (2008) Genetic basis of fitness differences in natural populations. *Nature* **452**, 169-175.

Favre M, Martin JGA, Festa-Bianchet M (2008) Determinants and life-history consequences of social dominance in bighorn ewes. *Animal Behaviour* **76**, 1373-1380.

Feldhamer GA, Thompson BC, Chapman JA (2003) *Wild mammals of North America: biology, management, and conservation* JHU Press.

Feng S, Ferlin A, Truong A*, et al.* (2009) INSL3/RXFP2 Signaling in Testicular Descent: Mice and Men. *Annals of the New York Academy of Sciences* **1160**, 197-204.

Ferlin A, Pepe A, Gianesello L*, et al.* (2008) Mutations in the Insulin-Like Factor 3 Receptor Are Associated With Osteoporosis. *Journal of Bone and Mineral Research* **23**, 683-693.

Frickey T, Lupas AN (2004) Phylogenetic analysis of AAA proteins. *Journal of Structural Biology* **146**, 2-10.

Garvin MR, Saitoh K, Gharrett AJ (2010) Application of single nucleotide polymorphisms to non-model species: a technical review. *Molecular Ecology Resources* **10**, 915-934.

Geist V (1966) *Validity of Horn Segment Counts in Aging Bighorn Sheep*.

Geist V (1971) *Mountain sheep* Chicago Press, Chicago, IL.

Gienapp P, Fior S, Guillaume F*, et al.* (2017) Genomic Quantitative Genetics to Study Evolution in the Wild. *Trends in Ecology & Evolution* **32**, 897-908.

Heimer WE, Smith III AC (1975) Ram horn growth and population quality. *Wildlife Bulletin 5. Alaska Department of Fish and Game*.

Helyar SJ, Hemmer-Hansen J, Bekkevold D*, et al.* (2011) Application of SNPs for population genetics of nonmodel organisms: new opportunities and challenges. *Molecular Ecology Resources* **11**, 123-136.

Hemming JE (1969) Cemental Deposition, Tooth Succession, and Horn Development as Criteria of Age in Dall Sheep. *The Journal of Wildlife Management* **33**, 552-558.

Hik D, Carey J (2000) *Cohort variation in horn growth of Dall sheep rams in the southwest Yukon, 1969–99*.

Hogg JT (1984) Mating in bighorn sheep: multiple creative male strategies. *Science* **225**, 526.

Jing S, Yu Y, Fang M*, et al.* (1997) GFRα-2 and GFRα-3 Are Two New Receptors for Ligands of the GDNF Family. *Journal of Biological Chemistry* **272**, 33111-33117.

Johnston SE, Gratten J, Berenos C*, et al.* (2013) Life history trade-offs at a single locus maintain sexually selected genetic variation. *Nature* **502**, 93-95.

Johnston SE, McEwan JC, Pickering NK*, et al.* (2011) Genome-wide association mapping identifies the genetic basis of discrete and quantitative variation in sexual weaponry in a wild sheep population. *Molecular Ecology* **20**, 2555-2566.

Kardos M, Husby A, McFarlane SE, Qvarnström A, Ellegren H (2016) Whole-genome resequencing of extreme phenotypes in collared flycatchers highlights the difficulty of detecting quantitative trait loci in natural populations. *Molecular Ecology Resources* **16**, 727-741.

Kardos M, Luikart G, Bunch R*, et al.* (2015) Whole-genome resequencing uncovers molecular signatures of natural and sexual selection in wild bighorn sheep. *Molecular Ecology* **24**, 5616-5632.

Karssen LC, van Duijn CM, Aulchenko YS (2016) The GenABEL Project for statistical genomics. *F1000Research* **5**, 914.

Kijas JW, Lenstra JA, Hayes B*, et al.* (2012) Genome-Wide Analysis of the World's Sheep Breeds Reveals High Levels of Historic Mixture and Strong Recent Selection. *PLoS Biol* **10**, e1001258.

Kijas JW, Porto-Neto L, Dominik S*, et al.* (2014) Linkage disequilibrium over short physical distances measured in sheep using a high-density SNP chip. *Animal Genetics* **45**, 754-757.

Kohn MH, Murphy WJ, Ostrander EA, Wayne RK (2006) Genomics and conservation genetics. *Trends in Ecology & Evolution* **21**, 629-637.

Kruuk LEB (2004) Estimating genetic parameters in natural populations using the "animal model". *Philosophical Transactions of the Royal Society B: Biological Sciences* **359**, 873-890.

Kruuk LEB, Hill WG (2008) Introduction. Evolutionary dynamics of wild populations: the use of long-term pedigree data. *Proceedings of the Royal Society B: Biological Sciences* **275**, 593-596.

Kruuk LEB, Slate J, Wilson AJ (2008) New Answers for Old Questions: The Evolutionary Quantitative Genetics of Wild Animal Populations. *Annual Review of Ecology, Evolution, and Systematics* **39**, 525-548.

Lin LF, Doherty DH, Lile JD, Bektesh S, Collins F (1993) GDNF: a glial cell line-derived neurotrophic factor for midbrain dopaminergic neurons. *Science* **260**, 1130.

Loehr J, Worley K, Moe J, Carey J, Coltman DW (2008) MC1R variants correlate with thinhorn sheep colour cline but not individual colour. *Canadian Journal of Zoology* **86**, 147-150.

Lundregan SL, Hagen IJ, Gohli J*, et al.* (2018) Inferences of genetic architecture of bill morphology in house sparrow using a high-density SNP array point to a polygenic basis. *Molecular Ecology* **0**.

Lupas AN, Martin J (2002) AAA proteins. *Current Opinion in Structural Biology* **12**, 746-753.

Lynch M, Walsh B (1998) *Genetics and analysis of quantitative traits* Sinauer Sunderland, MA.

Malenfant RM, Davis CS, Richardson ES, Lunn NJ, Coltman DW (2018) Heritability of body size in the polar bears of Western Hudson Bay. *Molecular Ecology Resources* **18**, 854-866.

Miller JM, Festa-Bianchet M, Coltman DW (2018) Genomic analysis of morphometric traits in bighorn sheep using the Ovine Infinium® HD SNP BeadChip. *PeerJ* **6**, e4364.

Miller JM, Kijas JW, Heaton MP, McEwan JC, Coltman DW (2012) Consistent divergence times and allele sharing measured from cross-species application of SNP chips developed for three domestic species. *Molecular Ecology Resources* **12**, 1145-1150.

Miller JM, Poissant J, Kijas JW, Coltman DW, Consortium ISG (2011) A genome-wide set of SNPs detects population substructure and long range linkage disequilibrium in wild sheep. *Molecular Ecology Resources* **11**, 314-322.

Poissant J, Davis CS, Malenfant RM, Hogg JT, Coltman DW (2011) QTL mapping for sexually dimorphic fitness-related traits in wild bighorn sheep. *Heredity* **108**, 256.

Poissant J, Wilson AJ, Festa-Bianchet M, Hogg JT, Coltman DW (2008) Quantitative genetics and sex-specific selection on sexually dimorphic traits in bighorn sheep. *Proceedings of the Royal Society of London B: Biological Sciences* **275**, 623-628.

Purcell S, Neale B, Todd-Brown K*, et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics* **81**, 559-575.

Rönnegård L, McFarlane SE, Husby A*, et al.* (2016) Increasing the power of genome wide association studies in natural populations using repeated measures – evaluation and implementation. *Methods in Ecology and Evolution* **7**, 792-799.

Santure AW, Garant D (2018) Wild GWAS—association mapping in natural populations. *Molecular Ecology Resources* **18**, 729-738.

Shafer ABA, Miller JM, Kardos M (2016) Cross-Species Application of SNP Chips is Not Suitable for Identifying Runs of Homozygosity. *Journal of Heredity* **107**, 193-195.

Sim Z, Hall JC, Jex B, Hegel TM, Coltman DW (2016) Genome-wide set of SNPs reveals evidence for two glacial refugia and admixture from postglacial recolonization in an alpine ungulate. *Molecular Ecology* **25**, 962-1083.

Slate J, Gratten J, Beraldi D*, et al.* (2009) Gene mapping in the wild with SNPs: guidelines and future directions. *Genetica* **136**, 97-107.

Slate J, Santure AW, Feulner PGD*, et al.* (2010) Genome mapping in intensively studied wild vertebrate populations. *Trends in Genetics* **26**, 275-284.

Summary YHR (2018-2019) Yukon Hunting Regulations Summary (ed. Environment YMo), Yukon, Canada.

Turner S, Armstrong LL, Bradford Y*, et al.* (2011) Quality Control Procedures for Genome-Wide Association Studies. *Current Protocols in Human Genetics* **68**, 1.19.11-11.19.18.

Valdez R, Krausman PR (1999) *Mountain sheep of North America* University of Arizona Press.

Worley K, Carey J, Veitch A, Coltman DW (2006) Detecting the signature of selection on immune genes in highly structured populations of wild sheep (Ovis dalli). *Molecular Ecology* **15**, 623-637.

Chapter 6: Conclusion

6.1 General conclusion

My doctoral thesis was focused on the development and application of SNP resources to understand thinhorn sheep evolution. My research asked questions regarding the processes that govern the distribution of genetic variation at various spatial and temporal scales, ranging from historical continental-wide glaciations, to landscape level geographical barriers to gene flow down, down to the genetic basis of an individual's mating success.

In **Chapter 2**, I combined my own data and the data of Kijas *et al.* (2012), Miller *et al.* (2012a), Miller *et al.* (2018) and Deniskova *et al.* (2016) to test the cross-species application of a pair of SNP arrays originally developed domestic sheep for phylogenetic and principal component analysis. I evaluated the performance of two SNP data coding schemes for phylogenetic analysis and applied a new ascertainment bias correction method for SNP data. I show for the first time that coding SNP data as genotypes (i.e. each genotype was given a state were AA=0,AB=1,BB=3) was better than coding by alleles for phylogenetic analysis. I was not able to assess the effectiveness of the ascertainment bias correction method I used since the "uncorrected" analysis produced the "correct" phylogeny. I demonstrated the use of the OvineSNP50 BeadChip in species delimitation using both phylogenetic and principal component analysis. I also discussed the use of OvineSNP50 BeadChip in wildlife forensics as a cost and time saving resource.

In **Chapter 3**, I applied the OvineHD BeadChip to 55 thinhorn sheep from across the range of the species distribution to examined the role of ice-sheet movement on the evolution of the two thinhorn sheep subspecies, the white Dall's sheep and the dark Stone's sheep. Using phylogenetic inference, I found evidence that isolation in different glacial refugia likely mediated the evolution of Dall's and Stone's sheep. I found the first genetic evidence of an admixture zone containing hybrids of Dall's and Stone's sheep in the Pelly and Cassiar Mountains of Yukon and British Columbia (BC). This hybrid zone arised due to contact between lineages expanding out of their respective glacial refugia after glacial retreat. I also provided some initial clues using

broad resolution sampling and Bayesian clustering/admixture analysis that the geographical range of Stone's sheep may be more restricted than current subspecies maps (which are based on coat colour) indicate.

In **Chapter 4**, I used data generated by the **Chapters 2** and **3** to develop a SNP panel containing 153 biallelic SNPs to examine the global fine-scale population genetic structure of thinhorn sheep. I found the distribution genetic variation to be generally consistent with subspecies boundaries at the species-wide level, and mountain range and river valley boundaries at finer scales with some key exceptions. By performing hierarchical analyses of population genetic structure, I revealed the presence of three previously unreported Stone's sheep genetic clusters in the Stikine/Skeena, Cassiar and Rocky Mountains. I also identified a new geographic range for Stone's sheep that is much more restricted than currently accepted subspecies maps indicate, and almost exclusively located within BC.

In **Chapter 5,** I used genetic data acquired using the OvineHD BeadChip and phenotype data collected during compulsory inspections of harvested thinhorn rams to produce the first heritability estimates for a fitness-related trait in the species. I found horn length, horn base circumference and horn volume to be moderately heritable in thinhorn sheep. I also performed the first genome-wide association study between SNP effects and each of the three horn size traits for thinhorn sheep. I found 2 loci of suggestive significance for associations with horn length. While I did not see any immediate links between the molecular functions of these loci and horn size, this result should serve as impetus for future studies.

When I began my doctoral studies in 2013, the genomics era had only emerged in the recent past and the use of SNP data still relatively new. In this thesis, I was able to make methodological contributions to SNP data coding for phylogenetic inference at a time when the use of SNP data for phylogenetics was very new (Decker *et al.* 2009; Leaché & Oaks 2017). I was the first to discover genetic evidence for the hybrid nature of Fannin sheep. I also showed that the coat-colour based subspecies maps for thinhorn sheep are likely incorrect and used fine-scale genetic clustering analysis to re-draw the subspecies map for thinhorn sheep. In this new map, the Stone's sheep range is much restricted and almost entirely contained within BC. The SNP-based

subspecies and population boundaries I found are also being used to update management policies in relevant jurisdictions, most recently prompting discussions of co-management for populations that straddle the BC-Yukon border. I was also the first to produce heritability estimates and find suggestively significant SNP associations for a fitness related trait in thinhorn sheep and one of the first for a wild species not under long-term study.

A major weakness of this study can be attributed to the limitations of cross-species SNPs chips. The cross-species application of domestic SNP arrays on related wild species can be said to be a bit of a double-edged sword. On the one hand, it has enabled rapid and cost effective marker discovery and genotyping for many non-model species (Haynes & Latch 2012; Miller *et al.* 2012a). On the other hand, conversion rates remain low, meaning the vast majority of loci on the SNP array do not produce useful data. Furthermore, ascertainment bias caused by the evolutionary histories of loci cross species SNPs also restrict questions that may be answered. This is because for a locus to be found polymorphic in thinhorn sheep (or any non-target species), it must be necessarily be polymorphic in domestic sheep (if not it would not have been discovered in the first place). The most parsimonious explanation for why a loci is polymorphic in both domestic and thinhorn sheep now is that the site was polymorphic in the last common ancestor (LCA) of thinhorn and domestic sheep. The other explanation would be a site that was monomorphic in the LCA became variable independently in domestic and thinhorn, which is far less likely. For neutral loci, the retention of ancestral polymorphism is largely governed by drift and rate of fixation corresponds to time since LCA. This effect is seen in the correlation of polymorphism, call rates (Miller *et al.* 2012b; Miller *et al.* 2011), and branch length (Chapter 2) to time since LCA. This makes cross-species SNP data unsuited for questions where unbiased estimates monomorphism or homozygosity are important, such as in the detection of selective sweeps (Shafer *et al.* 2016) or determining time since divergence (Sim *et al.* 2016). Future studies can overcome the problem of ascertainment bias in cross species SNPs by discovering markers using the target species. More ambitiously, a universal wild sheep SNP chip could be developed using an ascertainment panel that included all 6 of the world's wild sheep species. Methodological improvements may also be able to correct for the ascertainment bias in cross SNP chip data if the fixation rate, and thus the tendency for monomorphism, can be precisely estimated and thus accounted for.

6.2 References

Decker JE, Pires JC, Conant GC*, et al.* (2009) Resolving the evolution of extant and extinct ruminants with high-throughput phylogenomics. *Proceedings of the National Academy of Sciences* **106**, 18644-18649.

Deniskova TE, Okhlopkov IM, Sermyagin AA*, et al.* (2016) Whole genome SNP scanning of snow sheep (Ovis nivicola). *Doklady Biochemistry and Biophysics* **469**, 288-293.

Haynes GD, Latch EK (2012) Identification of novel single nucleotide polymorphisms (SNPs) in deer (*Odocoileus* spp.) using the BovineSNP50 BeadChip. *Plos One* **7**, e36536.

Kijas JW, Lenstra JA, Hayes B*, et al.* (2012) Genome-Wide Analysis of the World's Sheep Breeds Reveals High Levels of Historic Mixture and Strong Recent Selection. *PLoS Biol* **10**, e1001258.

Leaché AD, Oaks JR (2017) The Utility of Single Nucleotide Polymorphism (SNP) Data in Phylogenetics. *Annual Review of Ecology, Evolution, and Systematics* **48**, 69-84.

Miller JM, Festa-Bianchet M, Coltman DW (2018) Genomic analysis of morphometric traits in bighorn sheep using the Ovine Infinium® HD SNP BeadChip. *PeerJ* **6**, e4364.

Miller JM, Kijas JW, Heaton MP, McEwan JC, Coltman DW (2012a) Consistent divergence times and allele sharing measured from cross-species application of SNP chips developed for three domestic species. *Molecular Ecology Resources* **12**, 1145-1150.

Miller JM, Poissant J, Hogg JT, Coltman DW (2012b) Genomic consequences of genetic rescue in an insular population of bighorn sheep (Ovis canadensis). *Molecular Ecology* **21**, 1583-1596.

Miller JM, Poissant J, Kijas JW, Coltman DW, Consortium ISG (2011) A genome-wide set of SNPs detects population substructure and long range linkage disequilibrium in wild sheep. *Molecular Ecology Resources* **11**, 314-322.

Shafer ABA, Miller JM, Kardos M (2016) Cross-Species Application of SNP Chips is Not Suitable for Identifying Runs of Homozygosity. *Journal of Heredity* **107**, 193-195.

Sim Z, Hall JC, Jex B, Hegel TM, Coltman DW (2016) Genome-wide set of SNPs reveals evidence for two glacial refugia and admixture from postglacial recolonization in an alpine ungulate. *Molecular Ecology* **25**, 962-1083.

Bibliography

Akaike H (1974) A new look at the statistical model identification. *Automatic Control, IEEE Transactions on* **19**, 716-723.

Alberta WAo RSA 2000. *Alberta Queen's printer. Retrieved from* https://www.canlii.org/en/ab/laws/stat/rsa-2000-c-w-10/134378/rsa-2000-c-w-10.html.

Allen JA (1902) A new sheep from the Kenai Peninsular. *Bulletin of the American Meseum of Natural History* **16**, 145-148.

Allendorf FW, Hohenlohe PA, Luikart G (2010) Genomics and the future of conservation genetics. *Nat Rev Genet* **11**, 697-709.

Anderson SP (2007) Biogeochemistry of Glacial Landscape Systems. *Annual Review of Earth and Planetary Sciences* **35**, 375-399.

Aulchenko YS, Ripke S, Isaacs A, van Duijn CM (2007) GenABEL: an R library for genome-wide association analysis. *Bioinformatics* **23**, 1294-1296.

Avise JC (2000) *Phylogeography: the history and formation of species* Harvard University Press, Cambrige, MA.

Bidlack AL, Cook JA (2001) Reduced genetic variation in insular northern flying squirrels (Glaucomys sabrinus) along the North Pacific Coast. *Animal Conservation* **4**, 283-290.

Borcard D, Legendre P (2012) Is the Mantel correlogram powerful enough to be useful in ecological analysis? A simulation study. *Ecology* **93**, 1473-1481.

Brown RR, Davis CS, Leys SP (2017) Clones or clans: the genetic structure of a deep-sea sponge, Aphrocallistes vastus, in unique sponge reefs of British Columbia, Canada. *Molecular Ecology* **26**, 1045-1059.

Brunsfeld S, Sullivan J, Soltis D, Soltis P (2001) Comparative phylogeography of northwestern North America: A synthesis. In: *Integrating Ecology and Evolution in a Spatial Context* (eds. Silvertown J, Antonovics J), pp. 319-339. Blackwell Science, London.

Bunch TD, Wu C, Zhang Y-P, Wang S (2006) Phylogenetic analysis of snow sheep (Ovis nivicola) and closely related taxa. *Journal of Heredity* **97**, 21-30.

Bunnell FL (1978) Horn Growth and Population Quality in Dall Sheep. *The Journal of Wildlife Management* **42**, 764-775.

Catto N, Liverman DGE, Bobrowsky PT, Rutter N (1996) Laurentide, cordilleran, and montane glaciation in the western Peace River ‚ÄìÄî Grande Prairie region, Alberta and British Columbia, Canada. *Quaternary International* **32**, 21-32.

Chapman JR, Nakagawa S, Coltman DW, Slate J, Sheldon BC (2009) A quantitative review of heterozygosity–fitness correlations in animal populations. *Molecular Ecology* **18**, 2746-2765.

Collins FS, Brooks LD, Chakravarti A (1998) A DNA polymorphism discovery resource for research on human genetic variation. *Genome Research* **8**, 1229-1231.

Coltman DW (2005) Testing marker-based estimates of heritability in the wild. *Molecular Ecology* **14**, 2593-2599.

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: Biological Sciences* **269**, 165.

Coltman DW, O'Donoghue P, Hogg JT, Festa-Bianchet M (2007) SELECTION AND GENETIC (CO)VARIANCE IN BIGHORN SHEEP. *Evolution* **59**, 1372-1382.

Coltman DW, O'Donoghue P, Jorgenson JT*, et al.* (2003) Undesirable evolutionary consequences of trophy hunting. *Nature* **426**, 655-658.

Cornuet J-M, Piry S, Luikart G, Estoup A, Solignac M (1999) New Methods Employing Multilocus Genotypes to Select or Exclude Populations as Origins of Individuals. *Genetics* **153**, 1989.

Cox GA, Mahaffey CL, Nystuen A, Letts VA, Frankel WN (2000) The mouse fidgetin gene defines a new role for AAA family proteins in mammalian development. *Nature genetics* **26**, 198.

Csilléry K, Johnson T, Beraldi D*, et al.* (2006) Performance of Marker-Based Relatedness Estimators in Natural Populations of Outbred Vertebrates. *Genetics* **173**, 2091.

Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* **9**, 772-772.

Davey JW, Hohenlohe PA, Etter PD*, et al.* (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics* **12**, 499-510.

Decker JE, Pires JC, Conant GC*, et al.* (2009) Resolving the evolution of extant and extinct ruminants with high-throughput phylogenomics. *Proceedings of the National Academy of Sciences* **106**, 18644-18649.

Demarchi R, Hartwig C (2004) Status of Thinhorn Sheep in British Columbia. *British Columbia Ministry of Water, Land and Air Protection, Biodiversity Branch - Wildlife Bulletin No. B-119*.

Deniskova TE, Okhlopkov IM, Sermyagin AA*, et al.* (2016) Whole genome SNP scanning of snow sheep (Ovis nivicola). *Doklady Biochemistry and Biophysics* **469**, 288-293.

Diniz-Filho JAF, Soares TN, Lima JS*, et al.* (2013) Mantel test in population genetics. *Genetics and Molecular Biology* **36**, 475-485.

Dobeš C, Mitchell-Olds T, Koch MA (2004) Intraspecific diversification in North American Boechera stricta (= Arabis drummondii), Boechera ×divaricarpa, and Boechera holboellii (Brassicaceae) inferred from nuclear and chloroplast molecular markers—an integrative approach. *American Journal of Botany* **91**, 2087-2101.

Dray S, Dufour A-B (2007) The ade4 Package: Implementing the Duality Diagram for Ecologists. *2007* **22**, 20.

Dyke AS (2004) An outline of North American deglaciation with emphasis on central and northern Canada. *Quaternary glaciations: extent and chronology* **2**, 373-424.

Dyke AS, Prest VK (1987) Late Wisconsinan and Holocene history of the Laurentide ice sheet. *Géographie physique et Quaternaire* **41**, 237-263.

Earl D, vonHoldt B (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* **4**, 359-361.

Elias SA, Short SK, Nelson CH, Birks HH (1996) Life and times of the Bering land bridge. *Nature* **382**, 60.

Ellegren H, Sheldon BC (2008) Genetic basis of fitness differences in natural populations. *Nature* **452**, 169-175.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14**, 2611-2620.

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**, 479-491.

Faircloth BC, Terhune TM, Schable NA*, et al.* (2009) Ten microsatellite loci from Northern Bobwhite (Colinus virginianus). *Conservation Genetics* **10**, 535-538.

Favre M, Martin JGA, Festa-Bianchet M (2008) Determinants and life-history consequences of social dominance in bighorn ewes. *Animal Behaviour* **76**, 1373-1380.

Feldhamer GA, Thompson BC, Chapman JA (2003) *Wild mammals of North America: biology, management, and conservation* JHU Press.

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

Feng S, Ferlin A, Truong A*, et al.* (2009) INSL3/RXFP2 Signaling in Testicular Descent: Mice and Men. *Annals of the New York Academy of Sciences* **1160**, 197-204.

Ferlin A, Pepe A, Gianesello L*, et al.* (2008) Mutations in the Insulin-Like Factor 3 Receptor Are Associated With Osteoporosis. *Journal of Bone and Mineral Research* **23**, 683-693.

Festa-Bianchet M, Pelletier F, Jorgenson JT, Feder C, Hubbs A (2013) Decrease in horn size and increase in age of trophy sheep in Alberta over 37 years. *The Journal of Wildlife Management*, n/a-n/a.

Fitch WM (1971) Toward defining the course of evolution: minimum change for a specific tree topology. *Systematic Biology* **20**, 406-416.

Frickey T, Lupas AN (2004) Phylogenetic analysis of AAA proteins. *Journal of Structural Biology* **146**, 2-10.

Funk WC, McKay JK, Hohenlohe PA, Allendorf FW (2012) Harnessing genomics for delineating conservation units. *Trends in Ecology & Evolution* **27**, 489-496.

Galbreath KE, Hafner DJ, Zamudio KR (2009) When cold is better: climate-driven elevation shifts yield complex patterns of diversification and demography in an apline speciatlist (American pika, *Ochotona princeps*). *Evolution* **63**, 2848-2863.

Garvin MR, Saitoh K, Gharrett AJ (2010) Application of single nucleotide polymorphisms to non-model species: a technical review. *Molecular Ecology Resources* **10**, 915-934.

Gasca-Pineda J, Cassaigne I, Alonso RA, Eguiarte LE (2013) Effective Population Size, Genetic Variation, and Their Relevance for Conservation: The Bighorn Sheep in Tiburon Island and Comparisons with Managed Artiodactyls. *Plos One* **8**, e78120.

Geist V (1966) *Validity of Horn Segment Counts in Aging Bighorn Sheep*.

Geist V (1971) *Mountain sheep* Chicago Press, Chicago, IL.

Genome KCoS (2009) Genome 10K: A Proposal to Obtain Whole-Genome Sequence for 10 000 Vertebrate Species. *Journal of Heredity* **100**, 659-674.

Gienapp P, Fior S, Guillaume F*, et al.* (2017) Genomic Quantitative Genetics to Study Evolution in the Wild. *Trends in Ecology & Evolution* **32**, 897-908.

Gilbert KJ (2016) Identifying the number of population clusters with structure: problems and solutions. *Molecular Ecology Resources* **16**, 601-603.

Gilbert KJ, Andrew RL, Bock DG*, et al.* (2012) Recommendations for utilizing and reporting population genetic analyses: the reproducibility of genetic clustering using the program structure. *Molecular Ecology* **21**, 4925-4930.

Gradstein FM, Ogg JG (2012) Chapter 2 - The Chronostratigraphic Scale. In: *The Geologic Time Scale*, pp. 31-42. Elsevier, Boston.

Guillot G, Rousset F (2013) Dismantling the Mantel tests. *Methods in Ecology and Evolution* **4**, 336-344.

Gutiérrez-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.

Hassanin A, Douzery EJP (1999) The Tribal Radiation of the Family Bovidae (Artiodactyla) and the Evolution of the Mitochondrial Cytochrome b Gene. *Molecular Phylogenetics and Evolution* **13**, 227-243.

Hassanin A, Ropiquet A (2004) Molecular phylogeny of the tribe Bovini (Bovidae, Bovinae) and the taxonomic status of the Kouprey, Bos sauveli Urbain 1937. *Molecular Phylogenetics and Evolution* **33**, 896-907.

Haynes GD, Latch EK (2012) Identification of novel single nucleotide polymorphisms (SNPs) in deer (*Odocoileus* spp.) using the BovineSNP50 BeadChip. *Plos One* **7**, e36536.

Heimer WE, Smith III AC (1975) Ram horn growth and population quality. *Wildlife Bulletin 5. Alaska Department of Fish and Game*.

Helyar SJ, Hemmer-Hansen J, Bekkevold D*, et al.* (2011) Application of SNPs for population genetics of nonmodel organisms: new opportunities and challenges. *Molecular Ecology Resources* **11**, 123-136.

Hemming JE (1969) Cemental Deposition, Tooth Succession, and Horn Development as Criteria of Age in Dall Sheep. *The Journal of Wildlife Management* **33**, 552-558.

Hengeveld PE, Cubberley JC (2011) Stone's sheep population dynamics and habitat use in the Sulphur / 8 Mile oil and gas pre-tenure plan area, northern British Columbia, 2005 – 2010. Synergy Applied Ecology, Mackenzie, British Columbia.

Henry P, Sim Z, Russello MA (2012) Genetic evidence for restricted dispersal along continuous altitudinal gradients in a climate change-sensitive mammal: the American pika. *Plos One* **7**, e39077.

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

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

Hik D, Carey J (2000) *Cohort variation in horn growth of Dall sheep rams in the southwest Yukon, 1969–99*.

Hogg JT (1984) Mating in bighorn sheep: multiple creative male strategies. *Science* **225**, 526.

Janes JK, Miller JM, Dupuis JR*, et al.* (2017) The K = 2 conundrum. *Molecular Ecology* **26**, 3594-3602.

Jex B, Ayotte J, Bleich V*, et al.* (2016) Thinhorn sheep: conservation challenges and management strategies for the 21st Century. Wild Sheep Working Group, Western Association of Fish and Wildlife Agencies, Boise, Idaho, USA.

Jing S, Yu Y, Fang M*, et al.* (1997) GFRα-2 and GFRα-3 Are Two New Receptors for Ligands of the GDNF Family. *Journal of Biological Chemistry* **272**, 33111-33117.

Johnston SE, Gratten J, Berenos C*, et al.* (2013) Life history trade-offs at a single locus maintain sexually selected genetic variation. *Nature* **502**, 93-95.

Johnston SE, McEwan JC, Pickering NK*, et al.* (2011) Genome-wide association mapping identifies the genetic basis of discrete and quantitative variation in sexual weaponry in a wild sheep population. *Molecular Ecology* **20**, 2555-2566.

Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**, 1403-1405.

Kardos M, Husby A, McFarlane SE, Qvarnström A, Ellegren H (2016) Whole-genome resequencing of extreme phenotypes in collared flycatchers highlights the difficulty of detecting quantitative trait loci in natural populations. *Molecular Ecology Resources* **16**, 727-741.

Kardos M, Luikart G, Bunch R*, et al.* (2015) Whole-genome resequencing uncovers molecular signatures of natural and sexual selection in wild bighorn sheep. *Molecular Ecology* **24**, 5616-5632.

Karssen LC, van Duijn CM, Aulchenko YS (2016) The GenABEL Project for statistical genomics. *F1000Research* **5**, 914.

Kijas JW, Lenstra JA, Hayes B*, et al.* (2012) Genome-Wide Analysis of the World's Sheep Breeds Reveals High Levels of Historic Mixture and Strong Recent Selection. *PLoS Biol* **10**, e1001258.

Kijas JW, Porto-Neto L, Dominik S*, et al.* (2014) Linkage disequilibrium over short physical distances measured in sheep using a high-density SNP chip. *Animal Genetics* **45**, 754-757.

Kohn MH, Murphy WJ, Ostrander EA, Wayne RK (2006) Genomics and conservation genetics. *Trends in Ecology & Evolution* **21**, 629-637.

Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I (2015) Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources* **15**, 1179-1191.

Kruuk LEB (2004) Estimating genetic parameters in natural populations using the "animal model". *Philosophical Transactions of the Royal Society B: Biological Sciences* **359**, 873-890.

Kruuk LEB, Hill WG (2008) Introduction. Evolutionary dynamics of wild populations: the use of long-term pedigree data. *Proceedings of the Royal Society B: Biological Sciences* **275**, 593-596.

Kruuk LEB, Slate J, Wilson AJ (2008) New Answers for Old Questions: The Evolutionary Quantitative Genetics of Wild Animal Populations. *Annual Review of Ecology, Evolution, and Systematics* **39**, 525-548.

Kuchta SR, Tan A-M (2005) Isolation by distance and post-glacial range expansion in the rough-skinned newt, Taricha granulosa. *Molecular Ecology* **14**, 225-244.

Latch EK, Heffelfinger JR, Fike JA, Rhodes OE (2009) Species-wide phylogeography of North American mule deer (Odocoileus hemionus): cryptic glacial refugia and postglacial recolonization. *Molecular Ecology* **18**, 1730-1745.

Latch EK, Reding DM, Heffelfinger JR, Alcalá-Galván CH, Rhodes OE (2014) Range-wide analysis of genetic structure in a widespread, highly mobile species (Odocoileus hemionus) reveals the importance of historical biogeography. *Molecular Ecology* **23**, 3171-3190.

Leaché AD, Banbury BL, Felsenstein J, Nieto-Montes de Oca A, Stamatakis A (2015) Short Tree, Long Tree, Right Tree, Wrong Tree: New Acquisition Bias Corrections for Inferring SNP Phylogenies. *Systematic Biology*.

Leaché AD, Oaks JR (2017) The Utility of Single Nucleotide Polymorphism (SNP) Data in Phylogenetics. *Annual Review of Ecology, Evolution, and Systematics* **48**, 69-84.

Legendre P, Fortin M-J (2010) Comparison of the Mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. *Molecular Ecology Resources* **10**, 831-844.

Legendre P, Fortin M-J, Borcard D (2015) Should the Mantel test be used in spatial analysis? *Methods in Ecology and Evolution* **6**, 1239-1247.

Levy SE, Myers RM (2016) Advancements in Next-Generation Sequencing. *Annual Review of Genomics and Human Genetics* **17**, 95-115.

Lin LF, Doherty DH, Lile JD, Bektesh S, Collins F (1993) GDNF: a glial cell line-derived neurotrophic factor for midbrain dopaminergic neurons. *Science* **260**, 1130.

Loehr J, Carey J, Hoefs M, Suhonen J, YlÖNen H (2007) Horn growth rate and longevity: implications for natural and artificial selection in thinhorn sheep (Ovis dalli). *Journal of Evolutionary Biology* **20**, 818-828.

Loehr J, Carey J, O'Hara RB, Hik DS (2010) The role of phenotypic plasticity in responses of hunted thinhorn sheep ram horn growth to changing climate conditions. *Journal of Evolutionary Biology* **23**, 783-790.

Loehr J, Worley K, Grapputo A*, et al.* (2006) Evidence for cryptic glacial refugia from North American mountain sheep mitochondrial DNA. *Journal of Evolutionary Biology* **19**, 419-430.

Loehr J, Worley K, Moe J, Carey J, Coltman DW (2008) MC1R variants correlate with thinhorn sheep colour cline but not individual colour. *Canadian Journal of Zoology* **86**, 147-150.

Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing. *Nat Rev Genet* **4**, 981-994.

Lundregan SL, Hagen IJ, Gohli J*, et al.* (2018) Inferences of genetic architecture of bill morphology in house sparrow using a high-density SNP array point to a polygenic basis. *Molecular Ecology* **0**.

Lupas AN, Martin J (2002) AAA proteins. *Current Opinion in Structural Biology* **12**, 746-753.

Lynch M, Walsh B (1998) *Genetics and analysis of quantitative traits* Sinauer Sunderland, MA.

Malenfant RM, Davis CS, Richardson ES, Lunn NJ, Coltman DW (2018) Heritability of body size in the polar bears of Western Hudson Bay. *Molecular Ecology Resources* **18**, 854-866.

Mason-Gamer RJ, Kellogg EA (1996) Testing for Phylogenetic Conflict Among Molecular Data Sets in the Tribe Triticeae (Gramineae). *Systematic Biology* **45**, 524-545.

Maxam AM, Gilbert W (1977) A new method for sequencing DNA. *Proceedings of the National Academy of Sciences of the United States of America* **74**, 560-564.

McCormack JE, Hird SM, Zellmer AJ, Carstens BC, Brumfield RT (2013) Applications of next-generation sequencing to phylogeography and phylogenetics. *Molecular Phylogenetics and Evolution* **66**, 526-538.

Meirmans PG (2012) The trouble with isolation by distance. *Molecular Ecology* **21**, 2839-2846.

Meirmans PG (2013) Nonconvergence in Bayesian estimation of migration rates. *Molecular Ecology Resources* **14**, 726-733.

Meirmans PG, Van Tienderen PH (2004) genotype and genodive: two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes* **4**, 792-794.

Merckx VSFT, Kissling J, Hentrich H*, et al.* (2013) Phylogenetic relationships of the mycoheterotrophic genus Voyria and the implications for the biogeographic history of Gentianaceae. *American Journal of Botany* **100**, 712-721.

Metzker ML (2010) Sequencing technologies - the next generation. *Nat Rev Genet* **11**, 31-46.

Michalakis Y, Excoffier L (1996) A Generic Estimation of Population Subdivision Using Distances Between Alleles With Special Reference for Microsatellite Loci. *Genetics* **142**, 1061-1064.

Miller JM, Festa-Bianchet M, Coltman DW (2018) Genomic analysis of morphometric traits in bighorn sheep using the Ovine Infinium® HD SNP BeadChip. *PeerJ* **6**, e4364.

Miller JM, Kijas JW, Heaton MP, McEwan JC, Coltman DW (2012a) Consistent divergence times and allele sharing measured from cross-species application of SNP chips developed for three domestic species. *Molecular Ecology Resources* **12**, 1145-1150.

Miller JM, Poissant J, Hogg JT, Coltman DW (2012b) Genomic consequences of genetic rescue in an insular population of bighorn sheep (Ovis canadensis). *Molecular Ecology* **21**, 1583-1596.

Miller JM, Poissant J, Kijas JW, Coltman DW, Consortium ISG (2011) A genome-wide set of SNPs detects population substructure and long range linkage disequilibrium in wild sheep. *Molecular Ecology Resources* **11**, 314-322.

Morin PA, Martien KK, Taylor BL (2009) Assessing statistical power of SNPs for population structure and conservation studies. *Molecular Ecology Resources* **9**, 66-73.

Moritz C (1994) Defining "Evolutionarily Significant Units" for conservation. *Trends in Ecology & Evolution* **9**, 373-375.

Mullis K, Faloona F, Scharf S*, et al.* (1986) Specific Enzymatic Amplification of DNA In Vitro: The Polymerase Chain Reaction. *Cold Spring Harbor Symposia on Quantitative Biology* **51**, 263-273.

Nagylaki T (1985) Homozygosity, effective number of alleles, and interdeme differentiation in subdivided populations. *Proceedings of the National Academy of Sciences of the United States of America* **82**, 8611-8613.

Nei M (1987) *Molecular Evolutionary Genetics* Columbia University Press, New York, NY.

Ogden ROB (2011) Unlocking the potential of genomic technologies for wildlife forensics. *Molecular Ecology Resources* **11**, 109-116.

Oksanen J, Kindt R, Legendre P*, et al.* (2018) vegan: Community Ecology Package. **version 2.5-2**.

Paetkau D, Calvert W, Stirling I, Strobeck C (1995) Microsatellite analysis of population structure in Canadian polar bears. *Molecular Ecology* **4**, 347-354.

Paetkau D, Slade R, Burden M, Estoup A (2003) Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power. *Molecular Ecology* **13**, 55-65.

Paradis E, Strimmer K, Claude J, Noel Y, Bolker B (2003) Analyses of Phylogenetics and Evolution: the ape Package. *Bioinformatics* **20**, 289–290.

Pertoldi C, W√≥jcik J, Tokarska MÇ*, et al.* (2010) Genome variability in European and American bison detected using the BovineSNP50 BeadChip. *Conservation Genetics* **11**, 627-634.

Pielou E (1991) *After the Ice Age: The Return of Life to Glaciated North America* The University of Chicago Press, Chicago, IL.

Poissant J, Davis CS, Malenfant RM, Hogg JT, Coltman DW (2011) QTL mapping for sexually dimorphic fitness-related traits in wild bighorn sheep. *Heredity* **108**, 256.

Poissant J, Wilson AJ, Festa-Bianchet M, Hogg JT, Coltman DW (2008) Quantitative genetics and sex-specific selection on sexually dimorphic traits in bighorn sheep. *Proceedings of the Royal Society of London B: Biological Sciences* **275**, 623-628.

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

Purcell S, Neale B, Todd-Brown K*, et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics* **81**, 559-575.

Rambaut A, Suchard M, Xie D, Drummond A (2014) Tracer v1. 6. *Available from Available at:* http://beast*. bio. ed. ac. uk/Tracer*.

Ramey RR, II (1993) *Evolutionary genetics and systematics of North American mountain sheep: Implications for conservation* Ph.D., Cornell University.

Rannala B (2015) BayeAss Edition 3.0 User's Manual.

Rezaei HR, Naderi S, Chintauan-Marquier IC*, et al.* (2010) Evolution and taxonomy of the wild species of the genus Ovis (Mammalia, Artiodactyla, Bovidae). *Molecular Phylogenetics and Evolution* **54**, 315-326.

Roffler G, Talbot S, Luikart G*, et al.* (2014) Lack of sex-biased dispersal promotes fine-scale genetic structure in alpine ungulates. *Conservation Genetics*, 1-15.

Roffler GH, Schwartz MK, Pilgrim KL*, et al.* (2016) Identification of landscape features influencing gene flow: How useful are habitat selection models? *Evolutionary Applications* **9**, 805-817.

Rönnegård L, McFarlane SE, Husby A*, et al.* (2016) Increasing the power of genome wide association studies in natural populations using repeated measures – evaluation and implementation. *Methods in Ecology and Evolution* **7**, 792-799.

Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572-1574.

Ropiquet A, Hassanin A (2005) Molecular phylogeny of caprines (Bovidae, Antilopinae): the question of their origin and diversification during the Miocene

Phylogénie moléculaire des caprinés (Bovidae, Antilopinae): origine et diversification au cours du Miocène. *Journal of Zoological Systematics and Evolutionary Research* **43**, 49-60.

Rousset (2000) Genetic differentiation between individuals. *Journal of Evolutionary Biology* **13**, 58-62.

Sanger F, Coulson AR (1975) A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of Molecular Biology* **94**, 441-448.

Santure AW, Garant D (2018) Wild GWAS—association mapping in natural populations. *Molecular Ecology Resources* **18**, 729-738.

Scheunert A, Heubl G (2014) Diversification of Scrophularia (Scrophulariaceae) in the Western Mediterranean and Macaronesia – Phylogenetic relationships, reticulate evolution and biogeographic patterns. *Molecular Phylogenetics and Evolution* **70**, 296-313.

Schwarz G (1978) Estimating the Dimension of a Model. *Ann. Statist.* **6**, 461-464.

Seelanan T, Schnabel A, Wendel JF (1997) Congruence and Consensus in the Cotton Tribe (Malvaceae). *Systematic Botany* **22**, 259-290.

Shafer ABA, Côté SD, Coltman DW (2011) Hot spots of genetic diversity descended from multiple pleistocene refugia in an alpine ungulate. *Evolution* **65**, 125-138.

Shafer ABA, Cullingham CI, Côté SD, Coltman DW (2010) Of glaciers and refugia: a decade of study sheds new light on the phylogeography of northwestern North America. *Molecular Ecology* **19**, 4589-4621.

Shafer ABA, Miller JM, Kardos M (2016) Cross-Species Application of SNP Chips is Not Suitable for Identifying Runs of Homozygosity. *Journal of Heredity* **107**, 193-195.

Sheldon C (1911) *The wilderness of the Upper Yukon; a hunter's explorations for wild sheep in sub-Arctic Mountains* T. Fisher Unwin, New York.

Sim Z, Hall JC, Jex B, Hegel TM, Coltman DW (2016) Genome-wide set of SNPs reveals evidence for two glacial refugia and admixture from postglacial recolonization in an alpine ungulate. *Molecular Ecology* **25**, 962-1083.

Slate J, Gratten J, Beraldi D*, et al.* (2009) Gene mapping in the wild with SNPs: guidelines and future directions. *Genetica* **136**, 97-107.

Slate J, Santure AW, Feulner PGD*, et al.* (2010) Genome mapping in intensively studied wild vertebrate populations. *Trends in Genetics* **26**, 275-284.

Southern EM (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *Journal of Molecular Biology* **98**, 503-517.

Steane DA, Potts BM, McLean E*, et al.* (2015) Genome-wide scans reveal cryptic population structure in a dry-adapted eucalypt. *Tree Genetics & Genomes* **11**, 33.

Steiner CC, Putnam AS, Hoeck PEA, Ryder OA (2013) Conservation Genomics of Threatened Animal Species. *Annual Review of Animal Biosciences* **1**, 261-281.

Summary YHR (2018-2019) Yukon Hunting Regulations Summary (ed. Environment YMo), Yukon, Canada.

Swofford DL (2003) *{PAUP*. Phylogenetic Analysis Using Parsimony (* and Other Methods). version 4.}* Sinauer Associates, Sunderland, MA.

Tavaré S (1986) Some probabilistic and statistical problems in the analysis of DNA sequences. In: *American Mathematical Society: Lectures on Mathematics in the Life Sciences*, pp. 57-86. American Mathematical Society, Providence, RI.

Templeton AR (1983) Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of humans and the apes. *Evolution* **37**, 221-244.

Tomasik E, Cook JA (2005) Mitochondrial Phylogeography and Conservation Genetics of Wolverine (Gulo gulo) of Northwestern North America. *Journal of Mammalogy* **86**, 386-396.

Trifinopoulos J, Nguyen L-T, von Haeseler A, Minh BQ (2016) W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Research* **44**, W232-W235.

Turner S, Armstrong LL, Bradford Y*, et al.* (2011) Quality Control Procedures for Genome-Wide Association Studies. *Current Protocols in Human Genetics* **68**, 1.19.11-11.19.18.

Untergasser A, Cutcutache I, Koressaar T*, et al.* (2012) Primer3—new capabilities and interfaces. *Nucleic Acids Research* **40**, e115.

Vähä J-P, Erkinaro J, Niemelä E, Primmer CR (2007) Life-history and habitat features influence the within-river genetic structure of Atlantic salmon. *Molecular Ecology* **16**, 2638-2654.

Valdez R, Krausman PR (1999) *Mountain sheep of North America* University of Arizona Press.

Waples RS, Gaggiotti O (2006) INVITED REVIEW: What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular Ecology* **15**, 1419-1439.

Wenne R, Drywa A, Kent M, Sundsaasen KK, Lien S (2016) SNP Arrays for Species Identification in Salmonids. In: *Marine Genomics: Methods and Protocols* (ed. Bourlat SJ), pp. 97-111. Springer New York, New York, NY.

Wetterstrand K (2018) DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP), National Human Genome Research Institute, NIH.

Willing E-M, Dreyer C, van Oosterhout C (2012) Estimates of Genetic Differentiation Measured by FST Do Not Necessarily Require Large Sample Sizes When Using Many SNP Markers. *Plos One* **7**, e42649.

Wilson GA, Rannala B (2003) Bayesian Inference of Recent Migration Rates Using Multilocus Genotypes. *Genetics* **163**, 1177.

Worley K, Carey J, Veitch A, Coltman DW (2006) Detecting the signature of selection on immune genes in highly structured populations of wild sheep (Ovis dalli). *Molecular Ecology* **15**, 623-637.

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

Yandell M, Ence D (2012) A beginner's guide to eukaryotic genome annotation. *Nature Reviews Genetics* **13**, 329.

Zhou H, Wu D, Chen R*, et al.* (2014) Developmental validation of a forensic rapid DNA-STR kit: Expressmarker 16. *Forensic Science International: Genetics* **11**, 31-38.

Zink RM, Blackwell-Rago RC, Ronquist F (2000) The shifting roles of dispersal and vicariance in biogeography. *Proceedings of the Royal Society B-Biological Sciences* **267**, 497-503.

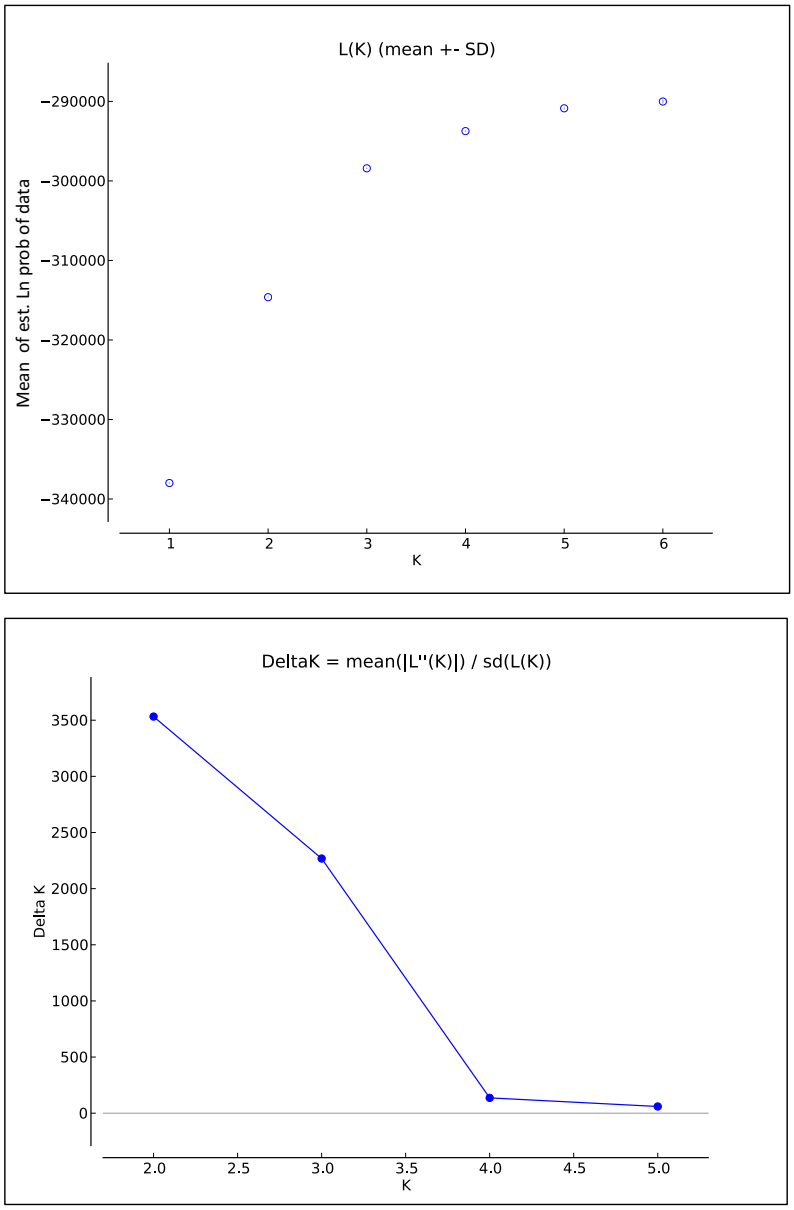

Figure S3.1 (a) Mean likelihood (±SD) calculated from 20 independent runs for K 1-5. (b) and Delta K plots calculated using the Evanno method in Structure Harvester.

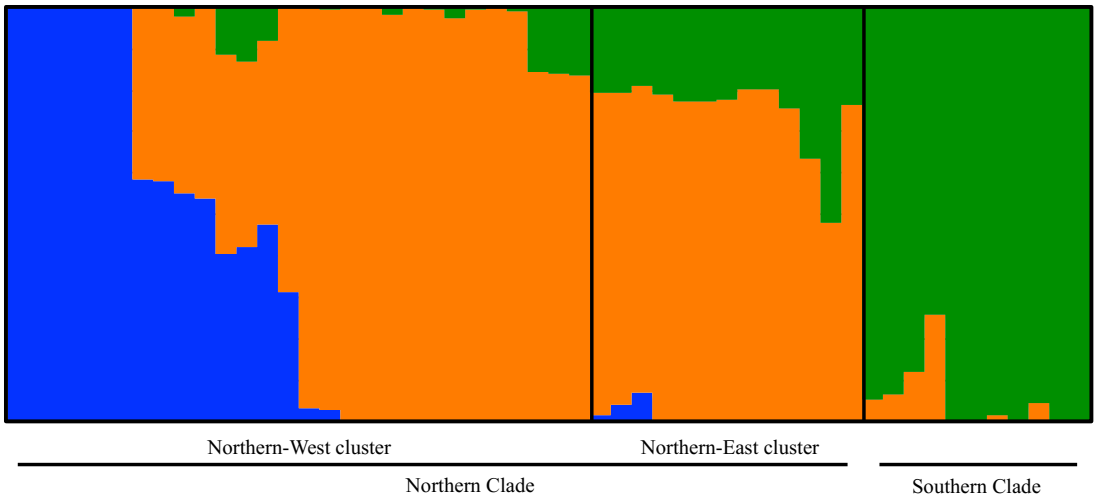

Figure S3.2 Structure admixture plot outputs for 20 independent runs for K=3.

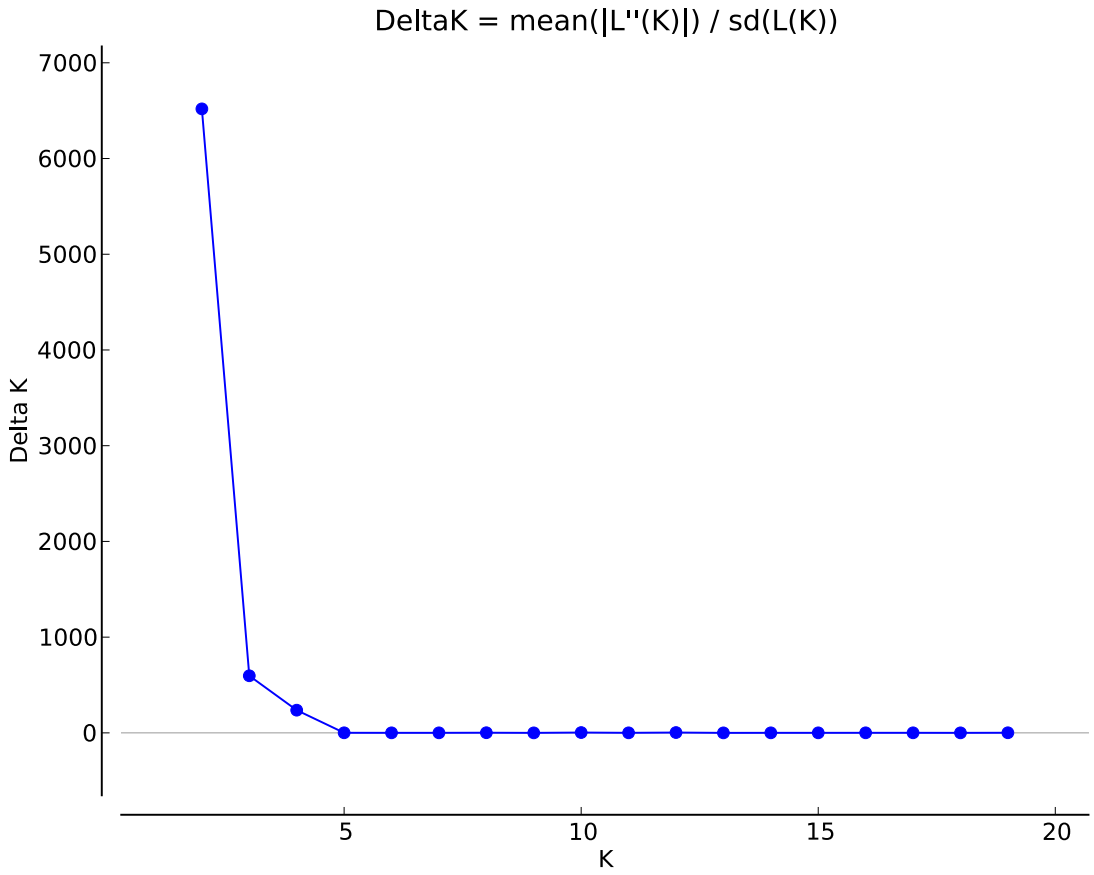

Figure S4.1 DeltaK plot calculated from 20 independent runs for K 1-20 for the species wide Structure analysis.

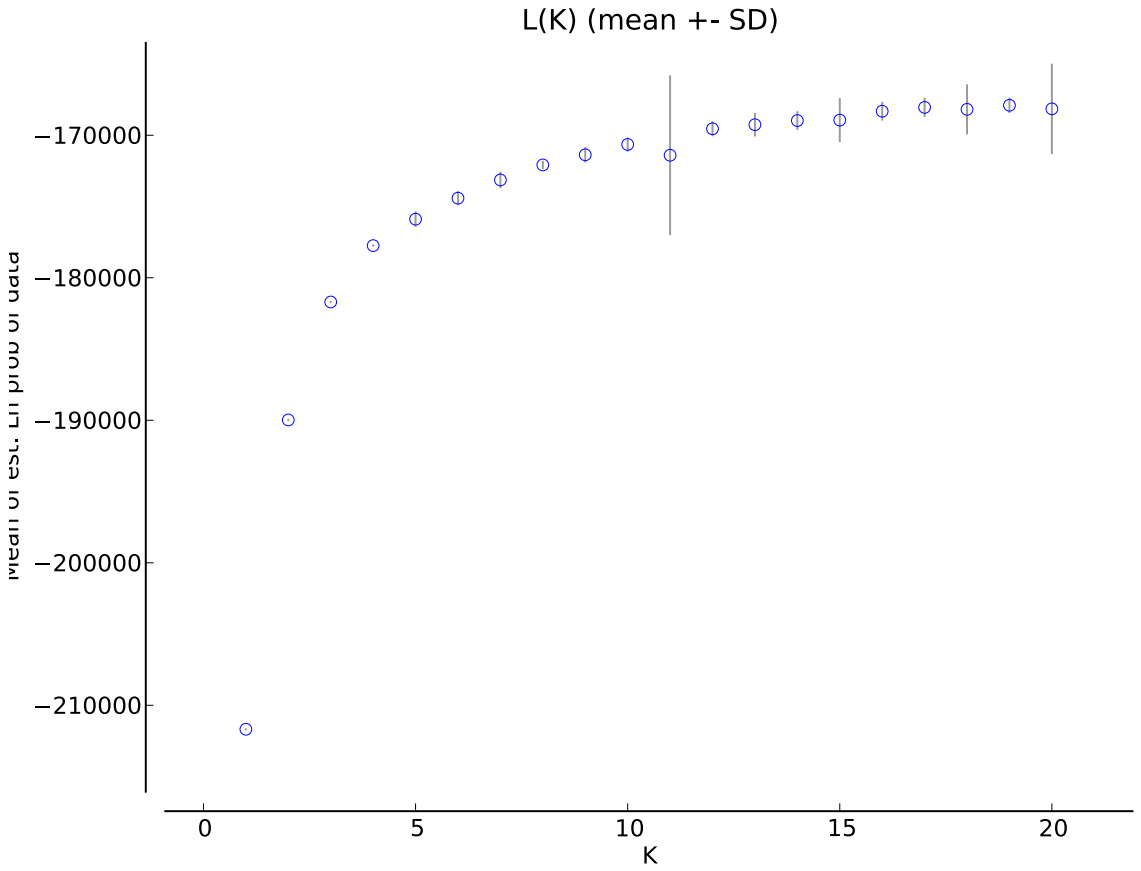

Figure S4.2 Mean likelihood (±SD) calculated from 20 independent runs for K 1-20 for the species wide Structure analysis.

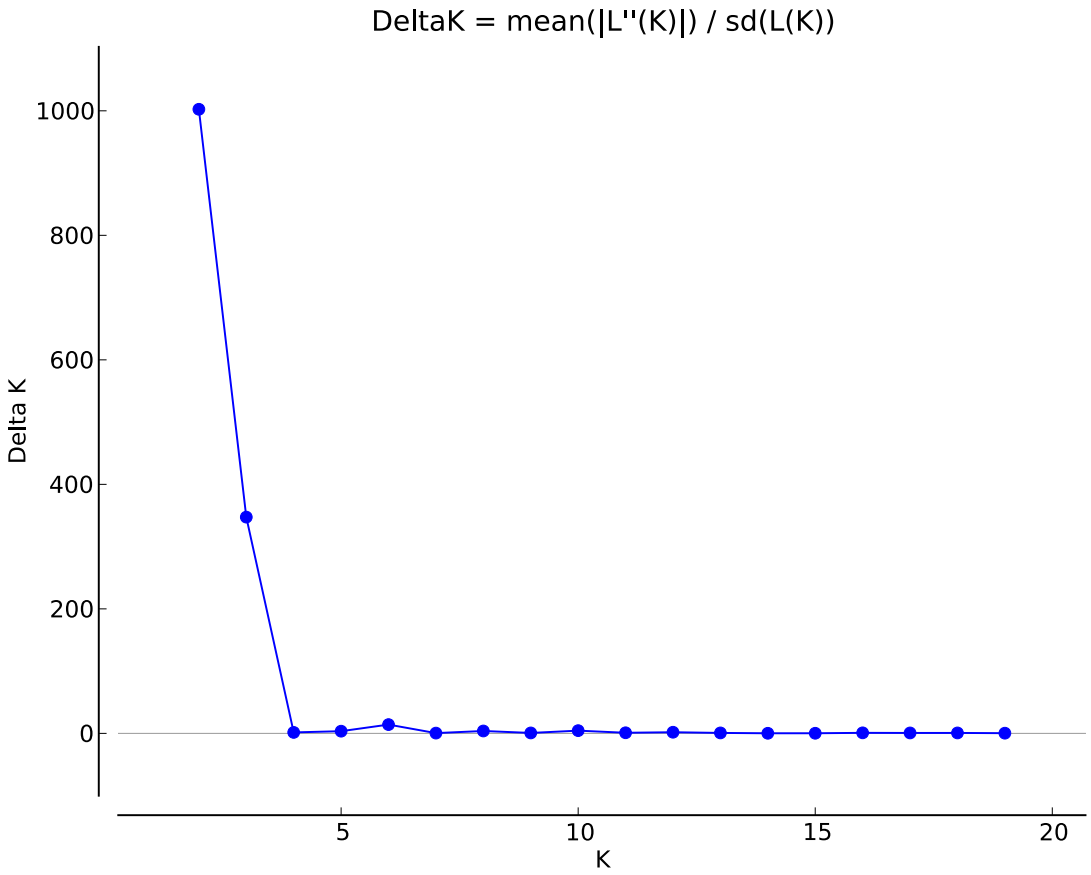

Figure S4.3 DeltaK plot calculated from 20 independent runs for K 1-20 for the Dall's sheep Structure analysis.

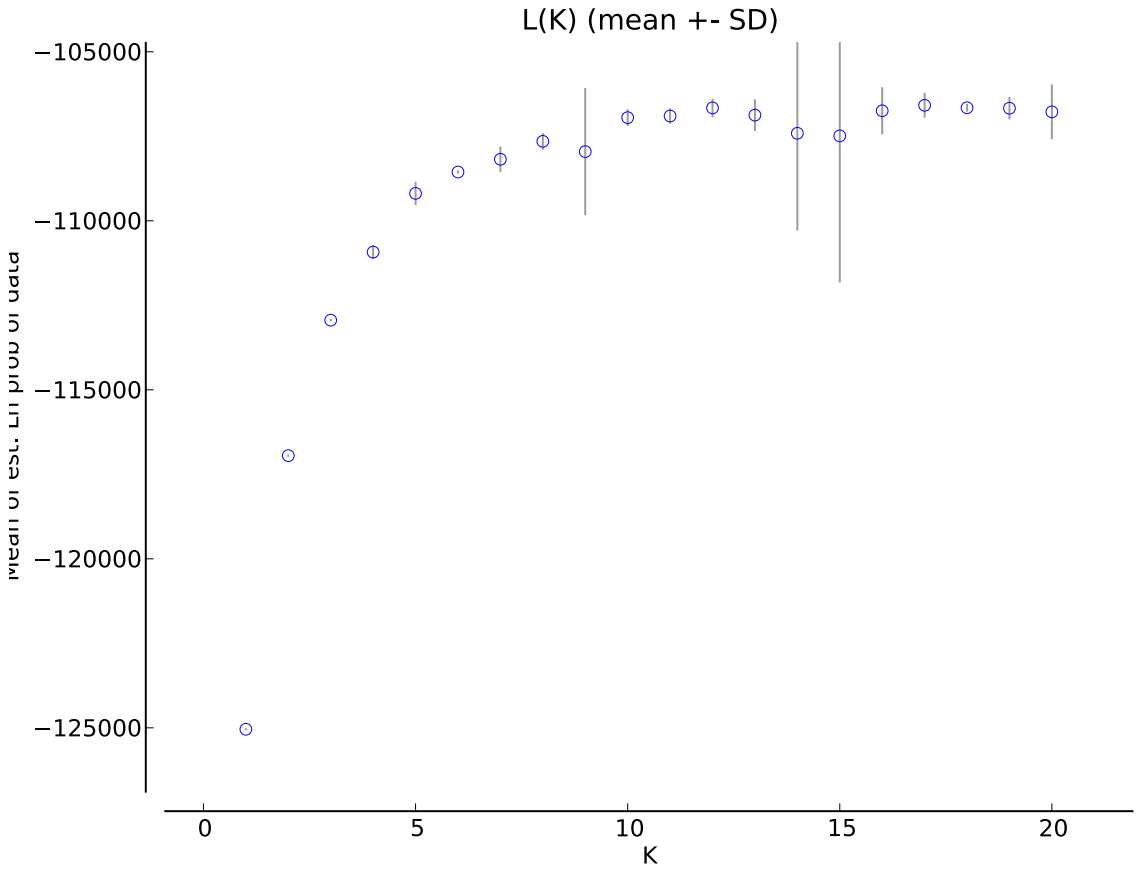

Figure S4.4 Mean likelihood (±SD) calculated from 20 independent runs for K 1-20 for the Dall's sheep Structure analysis.

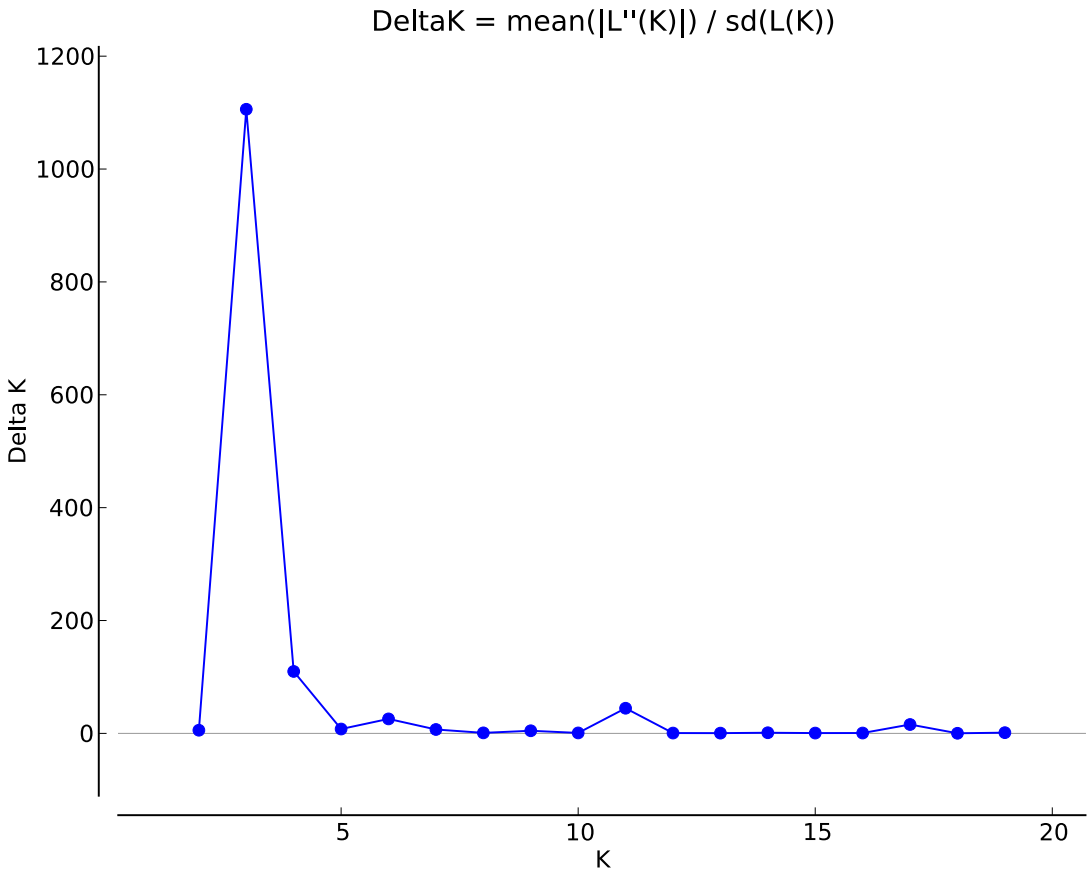

Figure S4.5 DeltaK plot calculated from 20 independent runs for K 1-20 for the Northern Dall's sheep Structure analysis.

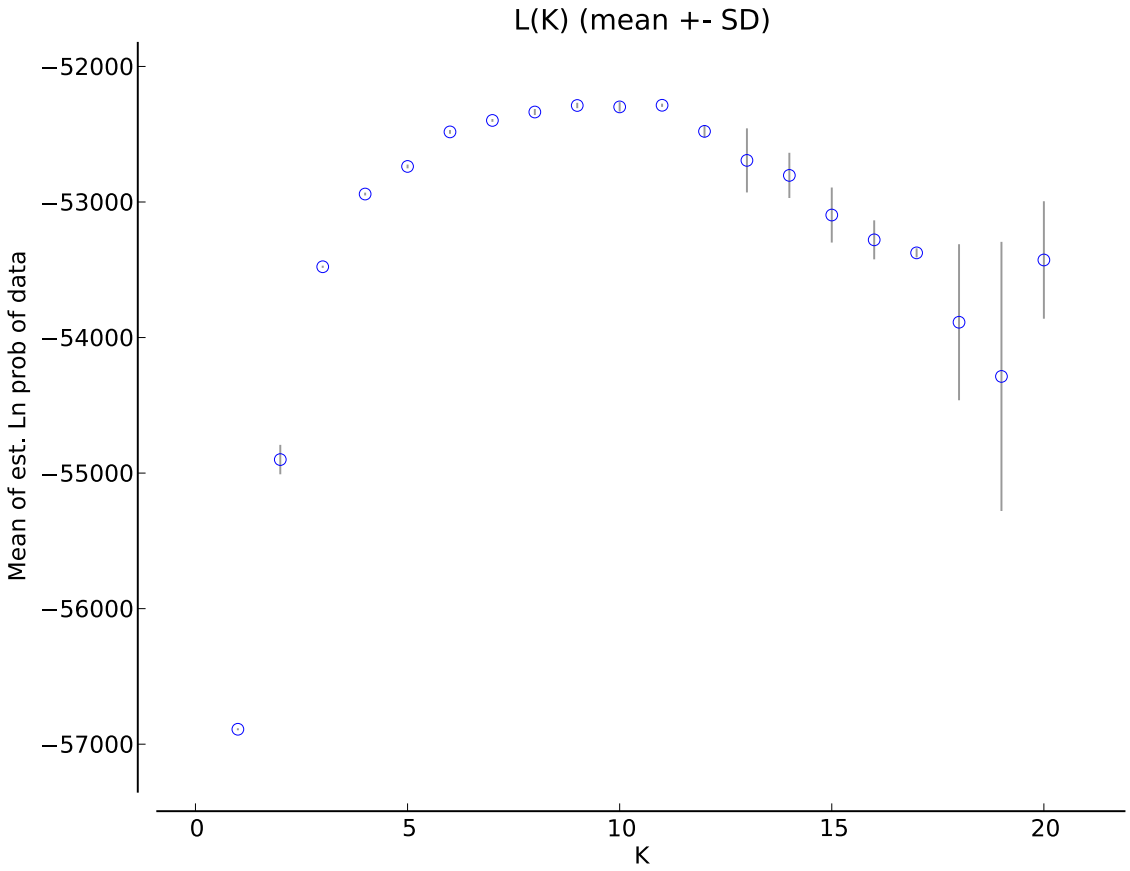

Figure S4.6 Mean likelihood (±SD) calculated from 20 independent runs for K 1-20 for the Northern Dall's sheep Structure analysis.

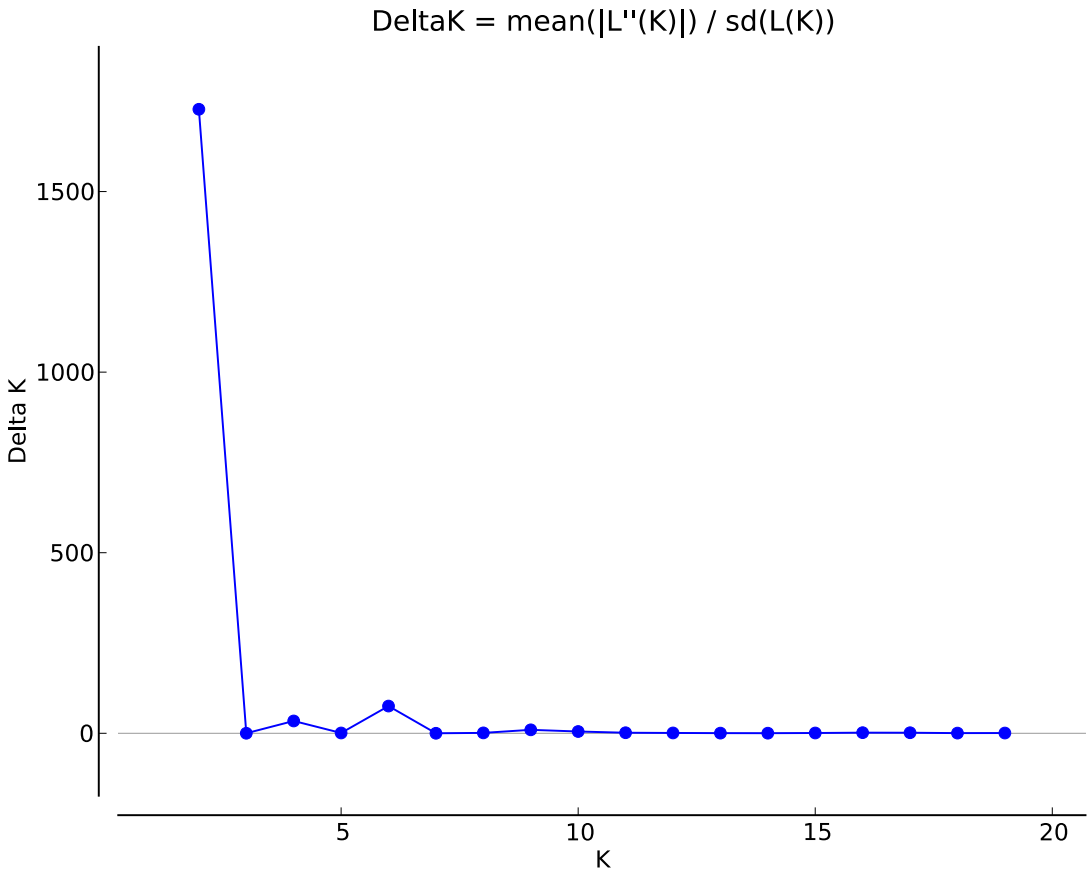

Figure S4.7 DeltaK plot calculated from 20 independent runs for K 1-20 for the Southern Dall's sheep Structure analysis.

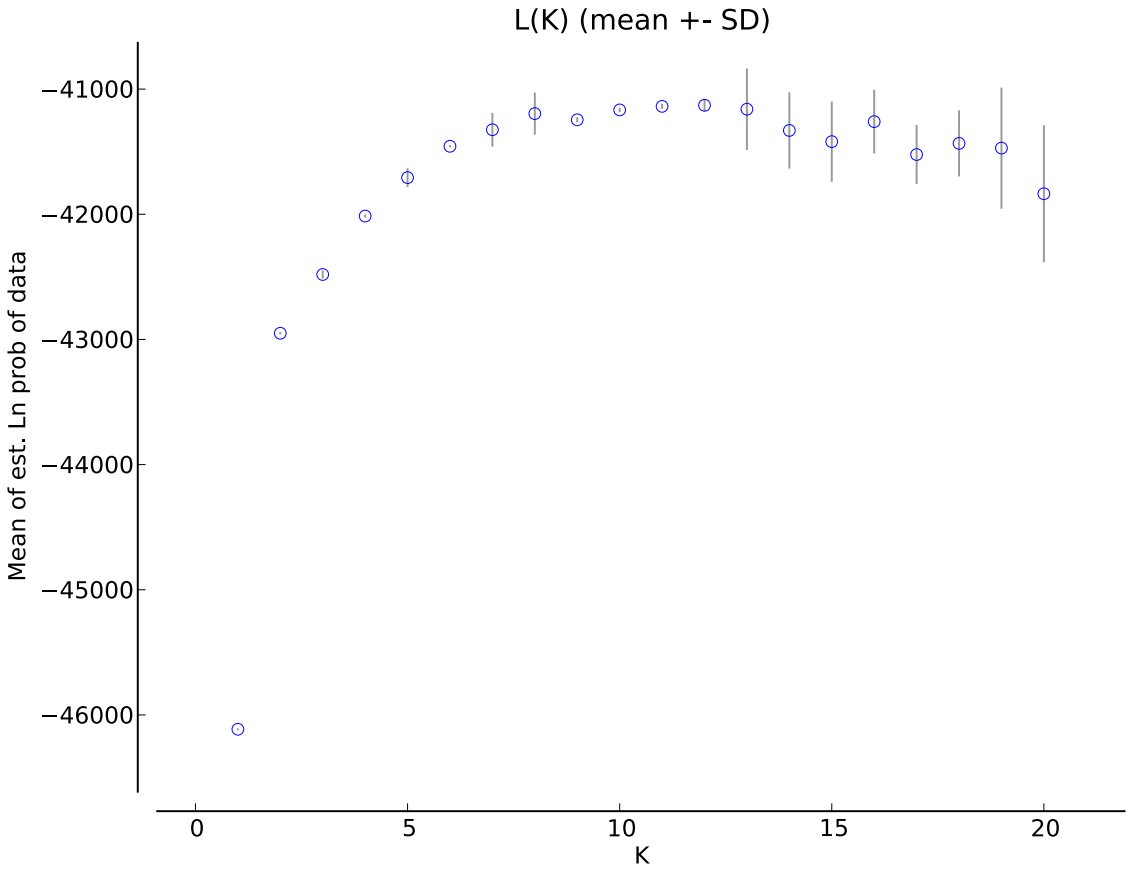

Figure S4.8 Mean likelihood (±SD) calculated from 20 independent runs for K 1-20 for the Southern Dall's sheep Structure analysis.

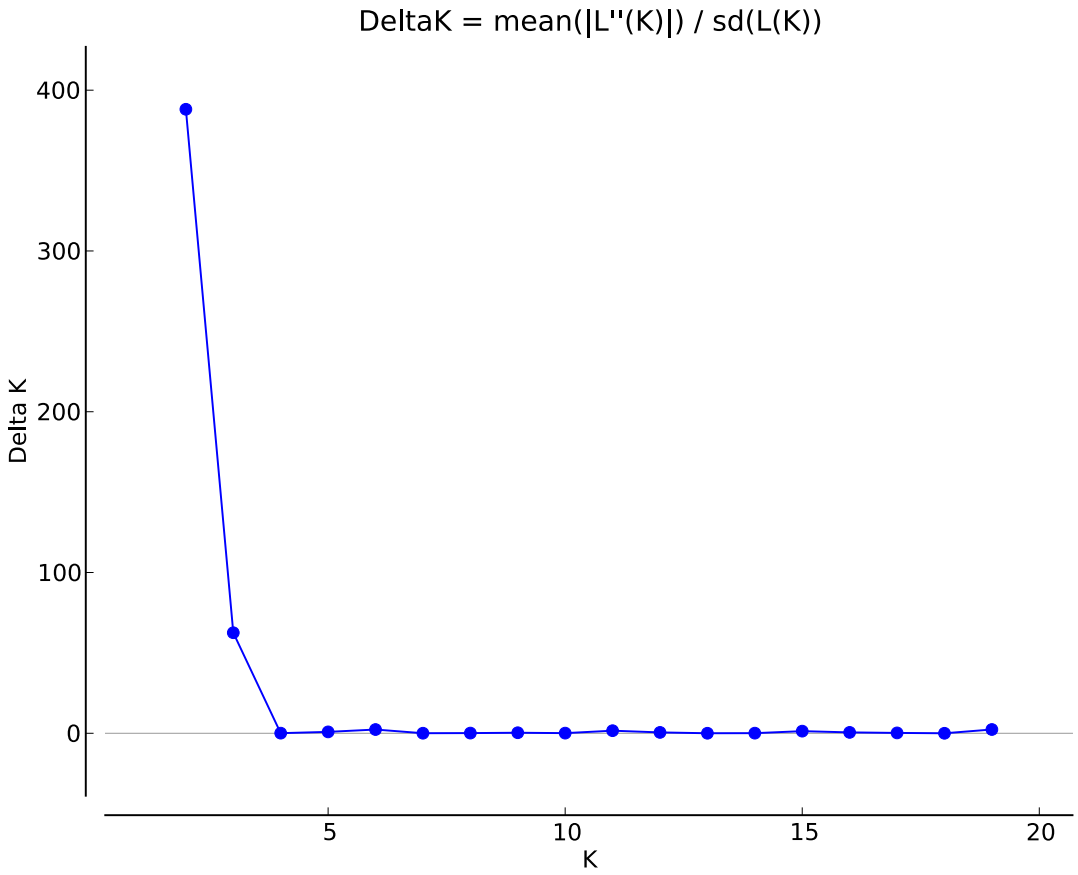

Figure S4.9 DeltaK plot calculated from 20 independent runs for K 1-20 for the Stone's sheep Structure analysis.

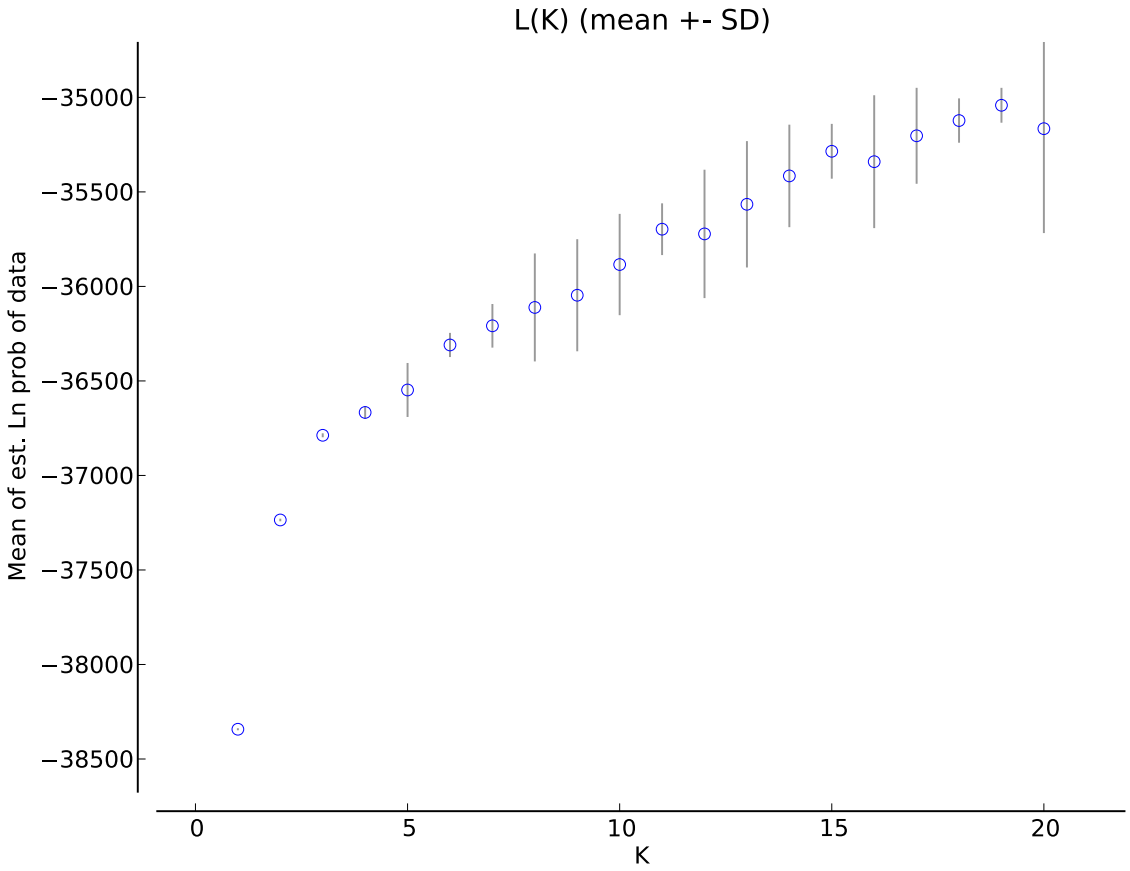

Figure S4.10 Mean likelihood (±SD) calculated from 20 independent runs for K 1-20 for the Southern Dall's sheep Structure analysis.