Spring 1-1-2014

Population Genetics of Ovis and Phylogenetics of Caprinae: A Comparison of Different Genetic Markers for Evaluation of Diversity at Multiple Taxonomic Levels

Catherine Colleen Driscoll
University of Colorado Boulder, catherine.driscoll@colorado.edu

Follow this and additional works at: https://scholar.colorado.edu/ebio_gradetds
Part of the Genetics Commons, and the Population Biology Commons

Recommended Citation
https://scholar.colorado.edu/ebio_gradetds/49

This Dissertation is brought to you for free and open access by Ecology & Evolutionary Biology at CU Scholar. It has been accepted for inclusion in Ecology & Evolutionary Biology Graduate Theses & Dissertations by an authorized administrator of CU Scholar. For more information, please contact cuscholaradmin@colorado.edu.
POPULATION GENETICS OF *OVIS* AND PHYLOGENETICS OF *CAPRINAE*: A COMPARISON OF DIFFERENT GENETIC MARKERS FOR EVALUATION OF DIVERSITY AT MULTIPLE TAXONOMIC LEVELS

by

CATHERINE C. DRISCOLL

B.A., University of California Berkeley 2000

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirement for the degree of

Doctor of Philosophy

Department of Ecology and Evolutionary Biology

2014
This thesis entitled:
Population Genetics of *Ovis* and Phylogenetics of Caprinae: A Comparison of Different Genetic
Markers For Evaluation of Diversity at Multiple Taxonomic Levels
Written by Catherine C. Driscoll
Has been approved for the Department of Ecology and Evolutionary Biology

_______________________________________  Jeffry B. Mitton
_______________________________________  Robert P. Guralnick
_______________________________________  Rebecca J. Safran
_______________________________________  David W. Stock
_______________________________________  John D. Wehausen

Date_____________________

The final copy of this thesis has been examined by the signatories, and we find that both the
content and the form meet acceptable presentation standards of scholarly work in the above
mentioned discipline.
Driscoll, Catherine Colleen (Ph.D., Ecology and Evolutionary Biology)

Population Genetics of Ovis and Phylogenetics of Caprinae: A Comparison of Different Genetic Markers For Evaluation of Diversity at Multiple Taxonomic Levels

Thesis directed by Professor Jeffry B. Mitton

This research was initiated due to concern that one of the four Rocky Mountain National Park (RMNP) bighorn sheep (Ovis canadensis) herds (the Mummy) could be suffering from inbreeding depression following a documented pneumonia outbreak and die-off in 1994. Failure of this herd to recover led to a comprehensive population genetic analysis of five bighorn sheep herds (Big Thompson, Continental Divide, Mummy, Never Summer and St. Vrain) to address four main questions:

1) Using data from 13 microsatellite loci: How much genetic variation exists in non-coding DNA in each herd?

2) Using sequence data from two distinct regions of the maternally inherited mitochondrial genome (Cytochrome Oxidase I gene (COI) and D-Loop control region (DL)): To what extent are ewes a source of gene flow between herds?

3) Using sequence data from two regions of the Y-chromosome gene SRY (promoter (PRO) and complete gene coding region (ORF)): to what extent are rams a source of gene flow between herds?

4) Do these data support evidence of a recent and severe genetic bottleneck in any of these five herds such that inbreeding depression could be a significant factor in long-term viability?
Our results support that all five herds have high levels of heterozygosity, low population substructure maintained by on-going gene flow, and no evidence of a recent genetic bottleneck. All data clearly refute the hypothesis that decreased genetic variation is a significant factor in the Mummy herd’s failure to thrive and instead suggest that environmental factors should be investigated.

Unfortunately, our analyses of Y-chromosome data showed the SRY sequences to be invariant among RMNP bighorn herds; however we realized these data would still be useful in a broader context and we therefore expanded analyses to include all genera within the subfamily Caprinae (Artiodactyla, Bovidae) and incorporated comparative data from the mitochondrial gene ND5. These data provide comprehensive and complementary phylogenetic analyses for which the more conserved Y-chromosome sequences generally showed better resolution of the most basal nodes, while the more variable mitochondrial sequences were usually better able to resolve relationships among recently diverged genera as well as between species within individual genera.
I dedicate this document to my father, James Glynn Driscoll III, without whom this research would not have been possible, and this dissertation may never have been completed. And I am certain that had you not been there to guide me, I would still be wandering the Never Summer Mountains, trying to find my way home.
ACKNOWLEDGMENTS

This dissertation is truly the product of both time and energy contributed by so many people and I am eternally grateful to each and every one of them. I am indebted to my advisor, Dr. Jeff Mitton, and my CU committee: Drs. Rob Guralnick, Rebecca Safran and David Stock, who worked tirelessly to help me reach this goal; providing support, feedback and countless edits along the way. I would also like to especially thank my outside committee member, Dr. John Wehausen, who not only taught me many of the laboratory and analytical techniques necessary for this research but who also provided endless support throughout this process. Thank you John, our long discussions have been invaluable.

Dr. David Armstrong, you are one of the most influential professors that I have ever encountered and I aspire to be the educator you are. Thank you for setting such an impressive example, I will endeavor to follow in your path. Dr. John Basey, thank you for your tireless dedication to undergraduate education; I have admired your ability to inspire both teaching assistants and students alike since I came to CU. I hope you keep up your wonderful work. Dr. Erin Bissell, thank you for the many semesters of teaching we accomplished together, we weathered the education equivalent of “Apocalypse Now” and survived, for several years. I learned so much along the way.

Graduate school has been described as a war of attrition, which requires unwavering resolve in the face of seemingly insurmountable obstacles; even when you cannot find the courage to carry on, even when you believe you cannot succeed. That resolve was only possible for me because I had the support of such wonderful people, and I cannot express the depth of my gratitude for the love and support of so many family and friends.
I want to thank my father, James Driscoll III, who contributed to this process in too many ways to name. Beginning with our field work: I will never forget the cold, early mornings searching for bighorn, and inevitably finding them on the absolute top of the ridge of the highest peak around and steadfastly making the arduous climb straight up the side of some incredibly steep mountainside. Day after day we labored to find and collect bighorn sheep samples for these genetic analyses found herein. We saw beauty that is unrivaled by any experience in my life. I will never forget those weeks in the backcountry of Rocky Mountain National Park.

I want to thank my mother, Patricia Martin, and my step father, Dan Martin, for unending, unwavering love. There is no way to express the value of your support. Dan, you have always been proud of me and I hope you know how much that has, and still does, mean. Thank you. Mom, we weathered lean times while I was growing up and your incredible optimism made my childhood rich in too many ways to name. I want to thank you especially for coming to Colorado for my defense and sitting with me, waiting, for one of the most frightening moments of my life. You are a rock, mom, and I love you.

Of course there are so many friends and colleagues to thank that I just have to simply list them and hope they know how much they mattered. Dr. Robert Baker, we met the first day of graduate school, and you have been a friend ever since. I feel so lucky to have gotten to know you so much better in the last year. Tim Singleton, you have been a wonderful friend and roommate almost from the moment I arrived in Colorado, thank you so much. Dr. Mike Robeson, you have been a great friend and often advisor on the finer points of phylogenetics and I hope we continue our video chats for years to come. Dr. Se Jin Song, I can’t imagine anyone putting up with sharing a tiny office with me for so many years and managing to like me anyway, I treasure your friendship. Mathew Arnold, thank you for being my intrepid explorer, I
will always remember how much fun we had weathering the rainstorms during field collection and I am so proud of you. Dr. Sarah Wise, thank you for, quite simply, being you. Your star shines so brightly. Joey Hubbard, you are a great friend and roommate and I am so glad to know you. Dr. Loren Sackett, I will always remember your smiling face. Ryan Lynch, thank you for pushing me when I needed it and never letting me back down, you helped me find the strength to see this through. Dr. Rob Roy Ramey III, thank you for unwavering support and for helping me find my voice. Corey Hazekamp, you were the best undergraduate I ever had the pleasure of working with, I can’t wait to see what great things you accomplish.

I would especially like to thank the EBIO staff, specifically Jill Skarstad and Linda Bowden, to whom I went for innumerable questions and who always had the answers. Thank you for your time and help and seemingly unending knowledge.

Finally, I would like to thank the National Park Service and the National Science Foundation for providing the majority of the funding for my research. This work would not have been possible without these resources, including the contributions of Judy Visty, Sherri Huwer, and Janet George.
## CONTENTS

### CHAPTER I
- Introduction and Background.................................................................1
- Development of Molecular Techniques, Markers, and Their Uses........1
- Population Genetics.................................................................................3
- Phylogenetics.........................................................................................6

### CHAPTER II
- Population Genetics of Rocky Mountain Bighorn Sheep Estimated From Microsatellite and Mitochondrial Data ..................................................10
  - Introduction..........................................................................................10
  - Study Area..........................................................................................13
  - Materials and Methods.................................................................18
  - Results...............................................................................................24
  - Discussion..........................................................................................41

### CHAPTER III
- Evaluating the Relative Resolving Power of Y-Chromosome and Mitochondrial Markers for Inferring Phylogenetic relationships among genera within the Subfamily Caprinae (Artiodactyla, Bovidae)..............................................47
  - Introduction..........................................................................................47
  - Materials and Methods.................................................................49
  - Results...............................................................................................54
  - Discussion..........................................................................................61
CHAPTER IV

General Conclusions.................................................................70

Population Genetics of Rocky Mountain Bighorn Sheep............70

Phylogenetics of Caprinae Genera...........................................73

LITERATURE CITED........................................................................76
TABLES

TABLE

1. Historical data for sheep translocations involving the five RMNP herds (BT, CD, MM, NS and SV) movements of sheep into, between, and out of these native herds ................................................................. 17

2. Microsatellite loci general information by locus including: location in genome (Chr.), range of allele sizes (in base pairs) identified in this study (Range), total number of different alleles detected (Alleles), PCR annealing temperature in degrees Celsius (Anneal), fluorescent label for each primer pair (Tag), source reference for locus information (Cited source) and NCBI accession number (Accession) ................................................................. 20

3. Data collection summary by marker type and individual herd. Number of sample collection sites (N), total microsatellite genotypes (Microsatellite), total mitochondrial Cytochrome Oxidase I amplifications (mtDNA COI) and total mitochondrial D-Loop control region amplifications (mtDNA DL) for each of the five bighorn sheep herds (BT: Big Thompson, CD: Continental Divide, MM: Mummy, NS: Never Summer, SV: St. Vrain) in north-central Colorado, USA .................................................. 25

4. Genetic variation estimates of microsatellite genotype data, including range and average number of alleles per locus (Na), Heterozygosity excess (He), Heterozygosity expected at equilibrium (Heq), probability of heterozygosity excess or deficiency (Two-tailed Wilcoxon Test, P(w)) and probability of heterozygosity excess under TPM (Sign Test, P(s)) ................. 26

5. Pairwise fixation indices (Fst; left side) and effective migration rates (Nm; right side) estimated from microsatellite genotypes (above diagonal) and mitochondrial haplotypes (below diagonal) among all five bighorn sheep herd pairs in north-central Colorado, USA ................................................. 26
6. Coefficient of inbreeding ($F_{is}$) estimates for three microsatellite data subsets including: genotypes for only the four potentially “non-neutral” loci: OLA, KERA, SOMA and TCRBV (Four loci), genotypes for the nine putatively “neutral” loci: ADC, AE16, FCB266, FCB304, HH62, MAF33, MAF36, MAF48 and MAF209 (Nine loci) and genotypes from the complete data set of all 13 Loci (All loci) for five bighorn sheep herds in north-central Colorado, USA.................................................................30

7. Pairwise isolation-by-distance (IBD) estimates from mitochondrial and microsatellite data for among all five bighorn sheep herd.........................32

8. Distribution and abundance of each of the twelve mitochondrial haplotypes identified in this study among all five herds of bighorn sheep in north-central Colorado, USA.................................................................37

9. Estimates of genetic diversity for the mitochondrial DNA haplotypes comprising 1,937 bp of Cytochrome Oxidase I (COI) and D-Loop (DL) sequences for each of the five bighorn sheep herds from north-central Colorado, USA including total sample size (N), segregating sites (S), number of haplotypes (h), haplotype diversity ($H_d$), average number of nucleotide differences (K), nucleotide diversity ($\pi$)........................................38

10. Sample information by individual Caprinae genus for Y-chromosome (SRY ORF and SRY PRO) and mitochondrial (ND5) data sets. .................50

11. SRY and ND5 primer sequences used for all sequence amplifications........51

12. Summary DNA sequence statistics by data set including sequence length, number of haplotypes (h), nucleotide diversity ($\pi$), nucleotide diversity with Jukes-Cantor correction ($\pi^{JC}$), number of segregating sites (S), average number of nucleotide differences (K), theta per nucleotide site as estimated from S ($\Theta_W$), and theta per sequence as estimated from S ($\Theta$)........53

13. Support of individual nodes for each data set as estimated with posterior probabilities (PP), ML Bootstrap percentages (BPML) and MP bootstrap percentages (BPMP). Nodes not found or with less than 50% bootstrap support or 0.5 posterior probability are indicated with an asterisk (*). Nodes for which support could not be evaluated because not all genera were available are indicated with a dash (-)..................................................................59
FIGURES

FIGURE

1. Relief map of north-central Colorado, USA showing 26 DNA collection sampling sites for the five bighorn sheep (*Ovis canadensis*) herds in and around Rocky Mountain National Park included in this study. Polygons each represent collection sites for a single herd: BT (circles), CD (diamonds), MM (squares), NS (hexagons) and SV (triangles)........................15

2. Map of the study area showing general geographic ranges for each of the five bighorn sheep herds. Overall fixation index (F<sub>st</sub>) estimates for both microsatellite and mitochondrial are indicated in the top left box. Pairwise fixation indices are indicated in boxes between herd pairs the highest and lowest values for each marker type in bold. Herd abbreviations: BT = Big Thompson, CD = Continental Divide, MM = Mummy, NS = Never Summer, SV = St. Vrain................................................28

3. Map of study area showing general geographic ranges for each of the five bighorn sheep herds. Overall effective migration (Nm) estimates for both microsatellite and mitochondrial data are indicated in top left box. Pairwise effective migration estimates are indicated in boxes between herd pairs with the highest and lowest values for each marker type in bold. Herd abbreviations: BT = Big Thompson, CD = Continental Divide, MM = Mummy, NS = Never Summer, SV = St. Vrain................................................29

4. Isolation-by-distance and reduced major axis regression analysis of microsatellite genotype data. Plot of log genetic similarity (M) vs. log geographic distance (km) for all ten bighorn sheep population pairs. Herd abbreviations: BT = Big Thompson, CD = Continental Divide, MM = Mummy, NS = Never Summer and SV = St. Vrain. Line of best fit for these data follows the equation y = -1.28x + 2.55(R<sup>2</sup> = 0.167, p = 0.13).............33

5. Scatter plots showing individual microsatellite deviations from expected heterozygosity across seven different mutation models of microsatellite evolution where IAM = Infinite Allele Model, SMM = Strict Stepwise Mutation Model and TPM = Two-Phase Mutation Model as implemented in the program BOTTLENECK.................................................................35
6. BOTTLENECK bar chart of allele distributions by frequency classes from least common (1) to most common (10) (left); Scatter plot distributions of individual microsatellite deviations from expected heterozygosity at equilibrium (right). Herd abbreviations: Big Thompson (BT), Continental Divide (CD), Mummy (MM), Never Summer (NS), St. Vrain (SV).

7. Isolation-by-distance and reduced major axis regression analysis of mitochondrial sequence data as implemented in Isolation by Distance Web Service. Plot of log genetic similarity (M) vs. log geographic distance (km) for all ten bighorn sheep population pairs. Herd abbreviations: BT = Big Thompson, CD = Continental Divide, MM = Mummy, NS = Never Summer and SV = St. Vrain. Line of best fit for these data follows the equation $y = -1.54x + 0.56$ ($R^2 = 0.69$, $p = 0.02$).

8. Comparative cladograms of Caprinae genera based on two regions of the Y-chromosome gene SRY: the SRY promoter (PRO, left) and the SRY open reading frame (ORF, right). Node supports are given as posterior probabilities (PP, Bayesian, above branch) and bootstrap values (ML/MP, below branch). Nodes with support of less than 50 were collapsed.

9. Comparative cladograms of the subfamily Caprinae based on ND5 gene sequence (left) and combined SRY ORF/PRO gene sequence (right). Node supports are shown on branches with posterior probabilities (Bayesian) above the line and bootstrap values (ML/MP) below the line. Nodes with branch supports all below 50 were collapsed and are not shown.

10. Cladogram of Caprinae genera based from combined ND5/PRO/ORF sequence data. Node supports are shown on branches with posterior probabilities (Bayesian) above the line and bootstrap values (ML/MP) below the line. Branches with supports all below 50 were collapsed and are not shown.
CHAPTER I

The study of evolution has been a central focus of biology and was greatly advanced by the work of both Charles Darwin (1899-1882) and Alfred Russel Wallace (1823-1913) who each independently articulated the theory of natural selection as a mechanism by which the composition of heritable characteristics in a given population changed through time (Darwin 1859, Wallace 1870). At that time, however, the hereditary mechanism of such characteristics was still unknown. Advancements in the field of genetics, many of which stemmed from the work of Gregor Mendel (1822-1884), ultimately provided a theory of heredity to explain the genetic mechanism of inheritance and its relation to the theory of natural selection. The integration of Mendelian genetics and evolution by natural selection was one of the central results of the Modern Synthesis (Huxley 1942) which brought together ideas and data from multiple fields of study.

DEVELOPMENT OF MOLECULAR TECHNIQUES, MARKERS, AND THEIR USES

The development of molecular techniques for quantifying genetic variation has revolutionized evolutionary studies (Lewontin and Hubby 1966). Before the development of these methods the primary focus of scientific analyses included morphological, physiological, and behavioral data to investigate evolutionary processes. Following the discovery of the structure of DNA (Watson and Crick 1953), protein electrophoresis provided some of the first data describing allozyme variation in populations based on differences in the number and composition of amino acids (Manos 1987, Leberg 1992). However, this method had limited ability to identify variation as it only discerned differences that affected amino-acid identity
within protein-coding regions of the genome. This was a significant drawback because not only is a very small percentage of genomic DNA expressed as proteins, but these regions are often subject to different selective pressures relative to neutral genetic markers. Subsequent developments allowing analyses of DNA polymorphisms throughout the genome, including single nucleotide polymorphisms (SNP), restriction fragment length polymorphisms (RFLP), variable number of tandem repeats (VNTR), minisatellites and microsatellites, improved both the breadth and depth of genomic variation detection. The invention of polymerase chain reaction (PCR) by Kary Mullis in 1983 greatly increased the efficiency of genetic data collection. Concurrent advances in DNA sequencing, from Sanger sequencing (Sanger et al. 1977), to fully automated sequencing (Applied Biosystems, 1987), to more recent advances in next generation sequencing, have resulted in virtually unlimited availability of information for diverse genetic markers.

This diversity of potential genetic data requires a careful evaluation of which genetic marker or markers are appropriate for an individual study. The variation among genetic markers includes differences in ploidy, inheritance pattern, variation, and evolutionary rate. Mitochondrial markers are among the more common employed, because they are haploid, and thus easier to amplify and sequence, they are generally highly variable, and they exhibit maternal inheritance. Mitochondrial DNA analyses include: use as a molecular clock for dating origins and divergences among taxa (Hasegawa et al. 1985, Cann et al. 1987), for tracking female-mediated gene flow (Richards et al. 2003, Meadows et al. 2006), and for tracing female lineages and introgression within and between groups (Avise and Ellis 1986, Lehman et al. 1991). Nuclear autosomal markers exhibit a vast array of variation. Microsatellites are one of the more popular markers, as they are both highly variable and relatively easy to amplify and score.
Because microsatellites are non-coding, they are generally considered neutral and can be used to accurately evaluate extant genetic variation uninfluenced by selection. Microsatellites have been used in a wide variety of analyses including evaluation of variation within populations for analyses of inbreeding depression (Taylor et al. 1994, Luikart et al. 1998), population substructure and connectivity within metapopulations (Johnson et al. 2003, Palo et al. 2004), medical evaluation of disease risk (Hearne et al. 1992, Duyao et al. 1993), DNA fingerprinting for identification of kinship (Queller et al. 1993, Blouin 2003) and forensic investigations (Norrgard 2008, Katsanis et al. 2013). As a result, Y-chromosome represents a unique genetic marker in that it is haploid, is generally highly conserved, is paternally inherited, and only exists in males. Thus, Y-chromosome genetic data can be useful for tracking male-mediated gene flow and introgression (Kikkawa et al. 2003, Rootsi et al. 2004) and for tracing patrilineal lineages and origins of species (Hammer et al. 1995, Crusiani 2011). Moreover, precisely because the Y-chromosome is unique, Y-chromosome genetic data can be used to compliment or to contrast with analyses of other markers (Tosi et al. 2000, Wood et al. 2005, Pidancier et al. 2006).

POPULATION GENETICS

One of the fields concerned with evolutionary processes is population genetics, which has more recently emerged and is mainly a product of the work of three scientists: John Haldane, Ronald Fisher and Sewall Wright in the 1920s and 1930s. Haldane contributed to the quantification of population genetic approaches, to the understanding of the effects of mutation and migration on natural selection, and was the first to define the concept of mutational load to explain costs associated with natural selection (Haldane 1932). R. A. Fisher was instrumental in
advancing statistical approaches for quantifying genetics, including analysis of variance and maximum likelihood, and his work was the basis for the field of quantitative genetics. He was also the first to define heterozygote advantage as a result of his work which involved appreciating that many natural polymorphisms are not selectively neutral and thus selection was often a significant force in natural populations (Fisher 1930). Wright is credited with defining the mathematics of genetic drift to explain stochastic processes causing genetic change in populations across generations as well as F-statistics to explain genetic variation within and among individuals and populations of the same species.

Thanks in part to advances in molecular genetics, F-statistics are more commonly used today as a method of comparing estimates of heterozygosity (H) at different levels within a population in order to evaluate parameters such as inbreeding and the degree of structure within a metapopulation, defined as a group of populations of the same species that are spatially separated but interact to some degree, in other words, “A population of populations” (Levins 1969). The term heterozygosity refers to the existence of two different alleles at a particular genetic locus. Heterozygosity can be used to measure the total proportion of heterozygous loci within a single individual, and can be extended for use as a population parameter, to measure the proportion of individuals who are heterozygous for a particular locus in a population. \( F_{is} \), more commonly referred to as the inbreeding coefficient, is the F-statistic concerned with comparing the heterozygosity of a single individual to the heterozygosity of its population. \( F_{is} \) ranges from negative one (all individuals are heterozygous, no observed homogygotes) to positive one (all individuals are homozygous, no observed heterozygotes). \( F_{is} \) values near zero are typically accepted as indicating a reasonably variable, “healthy” population while an \( F_{is} \) approaching positive one is deemed increasingly inbred and potentially at risk for the effects of inbreeding
depression. Inbreeding measures the extent to which two alleles at a particular locus among individuals are identical by descent. When the number of alleles or the number of individuals in a population decrease, the degree of homozygosity and inbreeding generally increase. This can have a significant effect on fitness by increasing the union of recessive deleterious alleles and thus the occurrence of phenotypic traits less well suited for a particular environment. Several factors can dramatically impact this relationship, however, including: population size, mutation rate and levels of current genetic variation, mating structure and departures from non-random mating (and resultantly effective population size), the type and extent of selection, and stochastic forces such as genetic drift. Although real world population genetic analyses are complicated by these factors, increased heterozygosity has been repeatedly linked to increased fecundity (Cothran et al. 1983, McAlpine 1993, Gemmell and Slate 2006), resistance to infection and disease (Allison 1954, Gulland et al.1993), life-span (Mitton and Jeffers 1989, Hartl et al. 1992) and population persistence (Soulé 1976, Brook et al. 2002, Frankham 2005, O’Grady et al. 2006). Thus, the inbreeding coefficient is a commonly estimated population genetic parameter when evaluating genetic variation in populations.

The F-statistic concerned with the degree of subdivision within a population or metapopulation, $F_{st}$, compares the heterozygosity of a single population to that of the larger population. $F_{st}$ ranges from zero (a panmictic population with identical identities and frequencies of all alleles between all subpopulations) to one (complete genetic division of subpopulations with no shared alleles). The degree of substructure, or isolation of subpopulations, within a metapopulation can affect overall heterozygosity, because uneven distributions of alleles will influence the incidence of different heterozygous combinations and may also cause some alleles to exist only in certain subpopulations and not in others (sometime referred to as “private
alleles”). The result of increasing substructure within a population is that certain heterozygous combinations never actually occur, decreasing overall heterozygosity, known as the Wahlund effect, even when the population is in Hardy-Weinberg equilibrium. This can be problematic for evaluation of genetic variation in natural populations, because, with increased population substructure (As \( F_s \) gets closer to 1), the absolute number and frequencies of alleles in the whole population will belie the actual heterozygosities of each subpopulation which are, in reality, much lower than expected by estimates made from data for the whole population.

**PHYLOGENETICS**

Another field concerned with quantifying the effects of evolutionary processes through time is phylogenetics. The field of phylogenetics is primarily focused on resolving evolutionary relationships between groups, or clades, which may comprise any taxonomic unit from a single species through to entire domains of life. This often involves using cladistics to group organisms based on their synapomorphies, or shared derived characteristics, in order to infer evolutionary history from a most recent common ancestor.

One of the earliest phylogenetic theories was proposed by Ernst Haeckle which is succinctly summarized as “ontogeny recapitulates phylogeny”. This idea centers around the thought that the way in which organisms develop during their lifetime reflects the evolutionary history of their species, and thus shared ontogenic traits can be used to infer phylogenetic relationships among taxa. This theory has since been generally disproved but reflects one of the main goals of phylogenetics which is to identify diagnostic data that can accurately explain the evolutionary relationships among groups of organisms at all taxonomic levels.
The ability to statistically evaluate support for observed phylogenetic relationships has been provided by three main methods: Maximum Parsimony (MP), Maximum Likelihood (ML) and Bayesian Inference (Bayes). MP analyses, as first described by Walter Fitch (1971), identifies the “most parsimonious” tree which is the tree requiring the fewest state changes to explain the evolutionary relationships among taxa given observed data. This method of tree estimation was the first to be widely implemented, mainly due to its relative simplicity and lower computational intensity, though this does vary depending on the search method (Camin and Sokal 1965, Fitch 1971). As computational capabilities allowed for more intensive analyses, the popularity of both ML and Bayesian methods increased. Maximum Likelihood is a method of tree estimation that seeks to identify the “most likely” phylogenetic tree (or evolutionary hypothesis) given the observed data using a uniform probability distribution and various substitution models. Popularized by R. A. Fisher in the early 20th century, ML analyses have been used to quantify evolutionary relationships for many decades (Edwards and Cavalli-Sforza 1963, Neyman 1974). Bayesian tree estimation methods are based on the theorems developed by reverend Thomas Bayes (1701-1761) and rediscovered posthumously in 1812 by Pierre-Simon Laplace (Truscott and Emory 1902). Bayesian inference is closely related to ML methods with the additional ability to set a priori assumptions to modify probability distributions from the simple uniform distribution implemented in ML and, like ML, identify of the most likely tree given the data, though Bayesian analyses also allow assignment of a prior.

Accurately identifying phylogenetic relationships can be complicated by the limited ability to accurately place clades in the context of larger phylogenies and can be exacerbated by a number of factors. Time since divergence of related taxa can be an important consideration when selecting diagnostic data because less variable markers can make it difficult to discern
relationships among recently diverged taxa while more variable data can obscure relationships among taxa that diverged in the more distant past. Horizontal gene transfer, the transmission of genetic information between unrelated organisms within a single generation via interspecific hybridization, can confuse evolutionary signals and result in underestimation of time since divergence between taxa. Although horizontal gene transfer is less common in eukaryotes, it has been repeatedly documented, particularly between mitochondrial DNA and nuclear DNA (Lopez et al. 1994, Turner et al. 2003). Loss of information from incomplete fossil records, also known as taxon sampling error, can also make phylogenetic inferences difficult, and even when the fossil record is complete, there is often a bias for characters that preserve well, such as teeth or bones, over those that don’t, such as soft tissues and ontological traits (Kidwell and Holland 2002, Crepet et al. 2004). Consideration of the effects of these potential complications is important, but, despite this, phylogenetics has provided important data regarding past and present evolutionary patterns at all taxonomic levels.

Obfuscation of evolutionary signals can also arise due to homoplasy, wherein the same phenotypic character is independently derived in two separate lineages, often due to convergent evolution, and not as a result of a derived character from recent common ancestor (Chen et al. 1997, Emery and Clayton 2004). Thus, homoplastic traits do not accurately reflect a shared evolutionary history and can confuse taxonomies, especially those based on limited data, such as analyses of just a single character. Discordant phylogenetic signals can also result from analyses of different marker types, such as morphological data, behavioral data, and genetic data, reflecting differences in the evolutionary histories of these data. More recently, genetic data has been predominantly used in phylogenetic studies, and the disparity between analyses of diverse genetic markers is often due to differences in variation, mutation rate, and inheritance pattern.
Phylogenetics is further complicated in that interpretations of analyses are relative to the scope of inquiry. For example, comparisons between two taxa may appear monophyletic, paraphyletic, or polyphyletic depending on scope of analyses and the taxa included. Consideration of the effects of these potential complications is important, but, despite this, there have been important advances in the field of phylogenetics regarding past and present evolutionary patterns at all taxonomic levels.
CHAPTER II

INTRODUCTION

Caughley (1994) distilled conservation biology into two topics of which one, the small population paradigm, concerns the extinction vulnerability of small populations. Among the potential factors involved in this vulnerability are loss of genetic diversity from genetic drift and the potential for resulting inbreeding depression (Soulé 1980). The ability of immigration to counteract the effects of genetic drift in largely isolated populations has long been recognized (Mills and Allendorf 1996) and has led in part to the important inclusion of landscape-level metapopulation considerations in conservation biology (Hanski and Gilpin 1997). Potential negative effects of inbreeding, however, remain a conservation challenge for animal species living in populations that are both small and more isolated than populations had been historically (IUCN: http://www.iucn.org, Baillie et al. 2004, Reed 2005, Vilas et al. 2006). Inbreeding problems associated with decreased heterozygosity can reflect increased expression of deleterious recessive alleles, exacerbating the causes of inbreeding depression attributed to both the overdominance and dominance hypotheses (Hedrick and Kallinowski 2000, Reed et al. 2002). Decreased heterozygosity is one factor which has been empirically demonstrated to have negative effects on individual fitness and long-term persistence of natural populations (Koehn et al. 1973, Raikkonen et al. 2006, Luikart et al. 2008a). While inbreeding depression is probably most common and acute in captive breeding populations, such as those in zoos (Ralls and Ballou 1986, Ralls et al. 1988, Frankham 2005, Gilligan et al. 2003, Boaks et al. 2007), it can also occur in parks and wildlife areas (Keller and Waller 2002, Crnokrak and Barrett 2002, O’Grady et al.
If detected, genetic rescue through the influx of genes from nearby populations may be a possible management strategy to alleviate the costs associated with inbreeding (Hedrick 2001, Hogg et al. 2006, Bergl et al. 2008).

It has been repeatedly shown that levels of fitness and inbreeding depression vary between laboratory and wild populations of the same species (Dobzhansky et al. 1963). Thus, it is important to evaluate the factors relevant to maintenance of genetic variation in natural settings when considering the risks associate with inbreeding depression. The degree to which gene flow ameliorates the risk of inbreeding depression in natural populations of plants and animals is dependent on several demographic factors, including: Life history and size of individual populations (Barrett et al. 1991, Manel et al. 2003, Angeloni et al. 2011), the distance between populations, the degree and pattern of individual dispersal among populations, individual life span, mating and reproductive structure (Hanksi 1998, Saccheri et al. 1998, Manel et al. 2003). Thus, species that are more continuously distributed generally have higher rates of gene flow and are less likely to experience decreased heterozygosity.

Bighorn sheep (Ovis canadensis) are a North American species that, because of narrow habitat requirements, occur in a naturally fragmented pattern of often small, disjunct populations. Where small, fragmented populations are geographically isolated, genetic drift and inbreeding will steadily erode genetic diversity (Schwartz et al.1986). Reduction of a population’s genetic diversity has been documented to result in problems that affect individual and population fitness (Hogg et al. 2006, Johnson et al. 2011, Rioux-Paquette et al. 2011, Miller et al. 2012). Gene flow can alleviate some of these problems, however, and bighorn sheep fit a metapopulation model well, with male biased interdemic movements commonly documented by radio telemetry (Bleich et al. 1990, 1996). Resulting gene flow among subpopulations has been measured
genetically (Epps et al. 2005, 2006, 2007) and is recognized as a critical biological process for this species that counteracts the effects of genetic drift in functional metapopulations (Schwartz et al. 1986, Bleich et al. 1990, 1996).

Bighorn sheep in and around Rocky Mountain National Park (RMNP) potentially represent a similar metapopulation structure as those documented in previous studies. This study was initiated, in part, as a request for genetic data to test the hypothesis that one of the four RMNP herds might be suffering from inbreeding depression. The Mummy herd declined in size following a pneumonia outbreak in the mid 1990s with continued poor recruitment thereafter. The failure of this herd to rebound to historic numbers could be due to a number of genetic or environmental causes, of which the genetic hypothesis was most testable.

Testing the hypothesis that the Mummy herd is suffering from inbreeding depression involves measuring appropriate genetic variables in the larger metapopulation that includes RMNP and the surrounding area. A complete description of a metapopulation involves genetic descriptions of the populations and estimates of gene migration among them. Nearby herds within this metapopulation provide convenient comparisons for population estimates of heterozygosity (He) and the inbreeding coefficient ($F_{st}$) to determine if the population of interest is genetically depauperate. Estimates of migration among populations will allow managers to assess whether particular herds are virtually isolated from the others or are regularly receiving genes from nearby populations.

To test the hypothesis that inbreeding depression was the cause of dwindling herd size and continued low recruitment in the Mummy herd, we examined data from both mitochondrial and microsatellite markers to estimate several population genetic parameters including: effective migration (Nm), heterozygosity, isolation-by-distance (IBD), population substructure ($F_{st}$), and
inbreeding ($F_{is}$). These analyses provide a comprehensive evaluation of both within and among population genetic variation, the extent to which these individual herds are part of a larger metapopulation structure, and whether the Mummy herd (or any of the other four herds) is suffering from the effects of a recent and severe genetic bottleneck.

Like many hypotheses in population biology, an asymmetrical experimental design exists in this research. A finding that genetic diversity is not compromised, gene flow is common, and inbreeding coefficients are not remarkable will cleanly disprove the inbreeding hypothesis. However, an alternative finding consistent with inbreeding does not necessarily imply that inbreeding is the primary factor driving population dynamics, given numerous other potential factors. Under such situations, heterozygosity-fitness correlations and genetic rescue experiments would be needed to further evaluate the role of genetics in population dynamics.

STUDY AREA

Our study was conducted in the Rocky Mountains of north-central Colorado in an area including RMNP, Indian Peaks Wilderness, Roosevelt National Forest and Comanche Peak Wilderness (Fig. 1). Five Rocky Mountain bighorn sheep (Ovis canadensis) populations were included in our analyses based on their presence in and relative proximity to RMNP and assessment of potential for gene flow within the larger metapopulation. Four of these herds (Continental Divide (CD), Mummy (MM), Never Summer (NS) and St. Vrain (SV)) maintain some or all of their annual range within the 415 square miles (1,075 square kilometers) of RMNP while the fifth (Big Thompson (BT)) maintains an annual range east of RMNP within Roosevelt National Forest. Historic data identify a single East Side (ES) herd as the combination of two
distinct herds, the Mummy and the Cow Creek (CC); however, current management recognizes a single herd (MM) in this range. In addition, the BT and SV herds were collectively referred to as the BT herd by wildlife managers prior to 1998 after which the two populations were recognized as separate herds. Most recent population estimates for the RMNP herds (CD: 120, MM: 64-78, NS: 180, and SV: 50-100) and the BT herd (85) were based on McClintock and White (2007), the Colorado Division of Wildlife (CDOW) (George et al. 2009) and RMNP data (Judy Visty and Cherie Yost, personal communication).
Figure 1. Relief map of north-central Colorado, USA showing 26 DNA collection sampling sites for the five bighorn sheep (*Ovis canadensis*) herds in and around Rocky Mountain National Park included in this study. Polygons each represent collection sites for a single herd: BT (circles), CD (diamonds), MM (squares), NS (hexagons) and SV (triangles).
CDOW and Park data show that these five herds have served as either the source or destination for sheep transplantations 11 times in the past (Table 1). These include six transplantations from study herds to herds outside our study area (NS: 1979, 1982, and 1984; CC: 1983, 1988, and 1989), three transplantations between study herds (MM to BT: 1983 and 1985; CC to BT: 1987) and two transplantations into study herds from outside sources (Tarryall to CC: 1977; Georgetown to BT: 2000). The three transplants of East Side (CC & MM) sheep to the BT herd represent major induced “migration” events within this metapopulation. The transplantations of outside sheep into study herds represented about 21% the subsequent CC herd census (126 sheep) and 31% of the subsequent BT herd census (77 sheep) and thus were substantial introductions of potentially novel genetic variation into this metapopulation. Based solely on this historic transplantation data, the MM and BT herds should share a higher percentage of genetic variation relative to the other three herds. This shared variation could be predominantly due to mitochondrial haplotypes rather than microsatellite genotypes given that most of the transplanted sheep were female, bighorn sheep exhibit a highly polygynous mating structure, and bighorn ewes tend to remain philopatric to natal herds.
Table 1. Historical data for sheep translocations involving the five RMNP herds (BT, CD, MM, NS and SV) movements of sheep into, between, and out of these native herds.

<table>
<thead>
<tr>
<th>Source Herd</th>
<th>Census</th>
<th>Destination</th>
<th>Census</th>
<th>Year</th>
<th>Translocated sheep</th>
<th>Ram</th>
<th>Ewe</th>
<th>Yrlng</th>
<th>Lamb</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sheep movement out of study area</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ES (CC)</td>
<td>-</td>
<td>Bristol Head</td>
<td>-</td>
<td>1983</td>
<td>3 11 0 5 19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ES (CC)</td>
<td>200(ES)</td>
<td>Glenwood Canyon</td>
<td>-</td>
<td>1988</td>
<td>- - - - 27</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ES (CC)</td>
<td>200(ES)</td>
<td>Lower Poudre</td>
<td>50</td>
<td>1989</td>
<td>2 9 2 5 18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NS</td>
<td>-</td>
<td>AZ &amp; NV</td>
<td>-</td>
<td>1979</td>
<td>1 11 0 8 20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NS</td>
<td>-</td>
<td>Purgatoire Canyon</td>
<td>-</td>
<td>1982</td>
<td>2 10 0 5 17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NS</td>
<td>-</td>
<td>Dinosaur South</td>
<td>-</td>
<td>1984</td>
<td>1 13 0 5 19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sheep movement within study area</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ES (MM)</td>
<td>43</td>
<td>BT</td>
<td>-</td>
<td>1983</td>
<td>2 9 0 8 19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ES (MM)</td>
<td>70</td>
<td>BT</td>
<td>100</td>
<td>1985</td>
<td>5 14 0 7 26</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ES (CC)</td>
<td>200(ES)</td>
<td>BT</td>
<td>100</td>
<td>1987</td>
<td>0 9 8 9 26</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sheep movement into study area</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tarryall</td>
<td>-</td>
<td>CC</td>
<td>-</td>
<td>1977</td>
<td>2 14 0 4 20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Georgetown</td>
<td>450</td>
<td>BT</td>
<td>50</td>
<td>2000</td>
<td>5 13 0 4 22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Previously identified factors influencing bighorn sheep migration and gene flow between these populations include both natural features (distance, landscape topography, forage and water availability) as well as artificial features (roads, buildings, cultivated areas and fences). Three distinct mountain ranges are included within this area: the Mummy Range crossing the northeastern corner of RMNP; the Continental Divide transecting the Park from North to South;
and the Never Summer Range along the western border separating RMNP and Routt National Forest. In addition, three U.S. highways (7, 34 and 36) and numerous hiking trails transect annual sheep migration routes at multiple locations throughout the study area. This area also includes the town of Estes Park, several small communities, and many farms and ranches.

MATERIALS AND METHODS

Bighorn sheep DNA samples were obtained non-invasively by repeated sampling across 36 locations within the study area between 2005 and 2010 (Fig. 1). Fresh fecal pellets were collected from a single, distinct pile and allowed to dry immediately. The outermost layer of each pellet was carefully scraped in an attempt to obtain intestinal cells, and total genomic DNA was extracted from that material with the QIAamp DNA Stool Minikit™ (Qiagen) incorporating modifications following Wehausen et al. (2004). We found this DNA collection approach, described in studies of bighorn sheep and other ruminants (Maudet et al. 2004, Luikart et al. 2008b, Poole et al. 2011, Godinho et al 2012), to be an efficient and effective non-invasive method for isolating usable DNA.

Microsatellite data development - Thirteen known and unlinked dinucleotide microsatellite loci were selected to develop data on nuclear genetic population structure, from which gene flow and other population genetic measures could be estimated (Table 2). Four of these are considered candidate adaptive loci due to locations within known gene regions (ADC, KERA, OLA and SOMA) of which two (ADC and OLA) are within genes with known disease-related function (Schwaiger et al. 1993, Wood and Phua 1994, Crawford et al 1995, Lucy et al.
1998; Table 2). The other nine loci are not associated with genes and have consistently met expectations for neutral markers (Crawford et al. 2006, Luikart and Cornuet 2008b). All loci were amplified using one of three kits: AmpliTaq Gold® PCR Kit (Applied Biosystems), Type-it Microsatellite PCR Kit (Qiagen) or Multiplex PCR Kit (Qiagen). Amplifications were performed with fluorescently labeled forward primers (FAM, HEX or TET) in a final volume of 25µl containing 50ng of template DNA, 2 mM MgCl₂, 0.05 mM dNTPs, 0.4µM of each primer and 0.25 - 2 units of polymerase. General PCR cycling steps were: initial denature at 93°C for 7:30’ followed by 30 cycles of: 95°C for 30”, 55 to 63°C (depending on locus) for 40”, 72°C for 30” and a final extension at 72°C for 2’. Annealing temperatures were optimized for individual microsatellite loci to maximize primer binding specificity and peak signal strength. PCR reactions were multiplexed where possible.
Table 2. Microsatellite loci general information by locus including: location in genome (Chr.), range of allele sizes (in base pairs) identified in this study (Range), total number of different alleles detected (Alleles), PCR annealing temperature in degrees Celsius (Anneal), fluorescent label for each primer pair (Tag), source reference for locus information (Cited source) and NCBI accession number (Accession).

<table>
<thead>
<tr>
<th>Locus</th>
<th>Chr.</th>
<th>Range</th>
<th>Alleles</th>
<th>Anneal</th>
<th>Tag</th>
<th>Cited source</th>
<th>Accession</th>
</tr>
</thead>
<tbody>
<tr>
<td>AE16</td>
<td>13</td>
<td>82-106</td>
<td>11</td>
<td>62</td>
<td>HEX</td>
<td>Penty <em>et al.</em> 1993</td>
<td>L11047</td>
</tr>
<tr>
<td>ADC***</td>
<td>-</td>
<td>85-113</td>
<td>8</td>
<td>62</td>
<td>FAM</td>
<td>Crawford <em>et al.</em> 1995</td>
<td>N/A</td>
</tr>
<tr>
<td>FCB266</td>
<td>25</td>
<td>88-102</td>
<td>6</td>
<td>62</td>
<td>FAM</td>
<td>Buchanan and Crawford, 1993</td>
<td>L01535</td>
</tr>
<tr>
<td>MAF36</td>
<td>22</td>
<td>88-104</td>
<td>8</td>
<td>56</td>
<td>TET</td>
<td>Swarbrick <em>et al.</em> 1991</td>
<td>M80519</td>
</tr>
<tr>
<td>SOMA**</td>
<td>16</td>
<td>94-118</td>
<td>7</td>
<td>59</td>
<td>TET</td>
<td>Lucy <em>et al.</em> 1998</td>
<td>U15731</td>
</tr>
<tr>
<td>MAF209</td>
<td>17</td>
<td>107-119</td>
<td>6</td>
<td>62</td>
<td>TET</td>
<td>Buchanan and Crawford, 1992a</td>
<td>M80358</td>
</tr>
<tr>
<td>HH62</td>
<td>16</td>
<td>108-122</td>
<td>9</td>
<td>56</td>
<td>HEX</td>
<td>Ede <em>et al.</em> 1994</td>
<td>N/A</td>
</tr>
<tr>
<td>MAF33</td>
<td>9</td>
<td>123-131</td>
<td>6</td>
<td>56</td>
<td>HEX</td>
<td>Buchanan and Crawford, 1992b</td>
<td>M77200</td>
</tr>
<tr>
<td>MAF48*</td>
<td>-</td>
<td>123-133</td>
<td>7</td>
<td>62</td>
<td>FAM</td>
<td>Buchanan <em>et al.</em> 1991</td>
<td>M62645</td>
</tr>
<tr>
<td>FCB304</td>
<td>19</td>
<td>134-150</td>
<td>9</td>
<td>62</td>
<td>FAM</td>
<td>Buchanan and Crawford, 1993</td>
<td>L01535</td>
</tr>
<tr>
<td>TCRBV62**</td>
<td>4</td>
<td>167-175</td>
<td>4</td>
<td>62</td>
<td>FAM</td>
<td>Buitkamp <em>et al.</em> 1993</td>
<td>L18957</td>
</tr>
<tr>
<td>KERA**</td>
<td>-</td>
<td>175-179</td>
<td>3</td>
<td>62</td>
<td>HEX</td>
<td>Tasheva <em>et al.</em> 1998</td>
<td>AF036962</td>
</tr>
<tr>
<td>OLADRBps***</td>
<td>-</td>
<td>278-302</td>
<td>8</td>
<td>62</td>
<td>HEX</td>
<td>Beraldi <em>et al.</em> 2006</td>
<td>AH003856</td>
</tr>
</tbody>
</table>

*Locus location unknown; **Loci within a gene region; ***Loci within disease-related gene region.
PCR products were electrophoresed with TAMRA size standards (350 or 500) on an ABI Prism 377 automated sequencer and chromatograms were analyzed using GENESCAN software (Applied Biosystems). Samples were genotyped independently two and five (average four) times to minimize errors due to allelic drop-out and/or false alleles. Duplicate samples with the same genotypes were identified and removed from analyses.

**Microsatellite data analyses** - Microsatellite data were used to estimate heterozygosity (He) and deviations from Hardy-Weinberg equilibrium as implemented in GENEPOP Web version 4.2 (Raymond and Rousset 1995, Rousset 2008). Independent analyses were performed on three different data sets: A) all 13 microsatellites, B) the nine neutral loci, and C) the four candidate adaptive loci, in order to evaluate whether these microsatellites are all selectively neutral or if there is evidence of selection acting on any individual locus. These data were also used to estimate overall and pairwise genetic distances ($F_{st}$) and overall and per locus inbreeding coefficients ($F_{is}$). Both overall and pairwise effective migration rates ($Nm$) were estimated using the private alleles method (those alleles present in just a single herd) following Barton and Slatkin (1986).

The relationship between genetic similarity and geographic distance was analyzed in the program Isolation by Distance Web Service (IBDWS) version 3.23 (Jensen et al. 2005). Approximate geographic coordinates (latitude, longitude) for all sample collection sites were determined using Google Earth. Pairwise distances among herds were measured from the estimated center of each herd’s annual range using the spherical distance (in kilometers) between coordinate pairs. Genetic relatedness was estimated following Slatkin’s measure of genetic similarity, defined as $M = (1/F_{st} - 1)/4$ (Slatkin 1993). The correlation between geographic distance and genetic distance was evaluated statistically with a Partial Mantel test using 10,000
iterations as implemented in IBDWS. The significance of this relationship was evaluated by regressing the $\log_{10}$ transformed pairwise geographic distances against the corresponding $\log_{10}$ transformed genetic distances using the Reduced Major Axis (RMA) regression.

Genetic structure was also examined using a model-based Bayesian clustering method as implemented in STRUCTURE 2.2 (Pritchard et al. 2000). This method clusters samples into K populations based on their microsatellite genotypes and the allele frequencies at each locus in a way that minimizes both Hardy-Weinberg disequilibrium and linkage disequilibrium (LD). K was set to 1-8 and for each K the program was run with 750,000 MCMC steps after a burn-in of 250,000 iterations using the admixture model with correlated allele frequencies and no prior sample information. The Evanno method was used to correct for overestimation of K and the most likely number of populations was estimated from the rate of change in log probabilities between each K evaluated (Evanno et al. 2005). Our a priori expectation was significant support for K=5 herds if genetic differentiation ($F_{st}$) was high and effective migration (Nm) was low among herds.

We evaluated whether a recent ($2N_e - 4N_e$ generations) and severe reduction in effective population size ($N_e$) had occurred which would provide evidence for a genetic bottleneck as implemented in the program BOTTLENECK (Cornuet and Luikart 1996, Luikart and Cornuet 1998, Luikart et al. 1998). This analysis operates on the principle that allele number ($Na$) is reduced faster than observed heterozygosity immediately after a bottleneck. Evidence of a bottleneck is evaluated by comparison of observed heterozygosity ($H_e$), calculated from allele frequencies, with heterozygosity expected at mutation-drift equilibrium ($H_{eq}$), calculated from allele number. Analyses of data were run under three mutation models: Infinite allele model (IAM), the stepwise mutation model (SMM), and the two phase model (TPM) that combines
IAM and SMM. We evaluated data under seven different model parameter combinations: TPM with 5% IAM and variance of 12, TPM with 5% IAM and variance of 30, TPM with 10% IAM and variance of 12, TPM with 10% IAM and variance of 30, TPM with 20% IAM and variance of 12, TPM with 20% IAM and variance of 30 and a strict SMM model. These parameter combinations were selected based on author recommendations (Piry et al. 1999) and previous work in bighorn sheep (Ramey et al. 2000, Luikart et al. 2011) and all were run with 10,000 iterations. BOTTLENECK also evaluates allele frequency distributions graphically via a qualitative “mode-shift” indicator of bottlenecked populations identified by a disproportionate loss of rare alleles. The Wilcoxon test was used to identify significant deviations from expected heterozygositites at mutation-drift equilibrium.

Mitochondrial DNA data development - In order to increase the likelihood of obtaining data with optimal variation for high population genetic resolution, we selected two regions of the mtDNA genome for inclusion in these analyses. Both Cytochrome Oxidase I (COI) and the D-Loop control region (DL) had been previously analyzed in multiple Ovis species, providing potentially useful comparative data for future analyses. Sequence amplification included 1,053 bp of COI and 963 bp of the DL using the AmpliTaq Gold® PCR Kit with individual reaction composition as previously described for microsatellite PCR. COI was amplified with primers COIF (5’-GCAGAGTTTGAAAGCTGCT-3’) and COIR (5’-AGCTGACGTGAAGTAAGC-3’) (Hiendleder et al. 1998) using a PCR profile which included an initial denature at 94°C for 7.5min followed by 35 cycles of: 94°C for 1min, 60°C for 70sec, 72°C for 90sec and a final extension at 72°C for 5min. DL was amplified with primers P1F (5’-CAACACCCAAAAGCTGAAGTTTC-3’) and P4R (5’-CTAGGCATTTCAGTGCCTTG-3’) (Zhao et al. 2001) using a PCR profile which included an initial denature at 94° for 7:30’
followed by 30 cycles of: 94°C for 1min, 62°C for 70sec, 72°C for 70sec and a final extension at
72°C for 5min. All sequences were aligned and analyzed in Sequencher™ version 4.6 (Gene
Codes Corp.). COI and DL sequences were each analyzed separately and then concatenated into
a single haplotype per individual for all subsequent analyses.

Mitochondrial DNA data analyses - Molecular diversity for each population was
evaluated for several parameters including: the number of haplotypes (h), haplotype diversity
(Hd), nucleotide diversity (π), effective migration (Nm) and population differentiation (Fst) in the
program DnaSP (Librado and Rozas 2009). The correlation between log genetic similarity (M)
and log geographic distance (km) was analyzed in the program Isolation by Distance Web
Service (IBDWS) version 3.23 (Jensen et al. 2005) as described for microsatellites above with
the addition that mtDNA data was analyzed using both traditional Fst as well as Φst to allow
incorporation of sequence distance information.

RESULTS

Microsatellite data analyses - Microsatellite data yielded genotypes for 137 individuals
after removal of poor quality amplifications and duplicate samples (Table 3). Observed
heterozygosity, averaged across all thirteen loci, ranged from 0.59 to 0.72 and no estimates
deviated significantly from expected under Hardy-Weinberg equilibrium (Table 4). Overall and
pairwise population subdivision, as estimated by Fst, was low and comparable among all three
data sets (13, nine, and four loci respectively), and we therefore present results for analyses of
the complete 13 loci genotypes. Overall and pairwise Fst estimates were all low (0.02-0.09 for
pairwise; Table 5, Fig. 2). Likewise, pairwise and overall F\textsubscript{st} estimates from individual loci were all near zero (F\textsubscript{st} = -0.021 to 0.158 & F\textsubscript{st} = 0.023 to 0.070, respectively).

Table 3. Data collection summary by marker type and individual herd. Number of sample collection sites (N), total microsatellite genotypes (Microsatellite), total mitochondrial Cytochrome Oxidase I amplifications (mtDNA COI) and total mitochondrial D-Loop control region amplifications (mtDNA DL) for each of the five bighorn sheep herds (BT: Big Thompson, CD: Continental Divide, MM: Mummy, NS: Never Summer, SV: St. Vrain) in north-central Colorado, USA.

<table>
<thead>
<tr>
<th>Herd</th>
<th>N</th>
<th>Microsatellite</th>
<th>mtDNA COI</th>
<th>mtDNA DL</th>
</tr>
</thead>
<tbody>
<tr>
<td>BT</td>
<td>3</td>
<td>15</td>
<td>22</td>
<td>22</td>
</tr>
<tr>
<td>CD</td>
<td>12</td>
<td>48</td>
<td>70</td>
<td>76</td>
</tr>
<tr>
<td>MM</td>
<td>10</td>
<td>28</td>
<td>36</td>
<td>42</td>
</tr>
<tr>
<td>NS</td>
<td>6</td>
<td>29</td>
<td>50</td>
<td>52</td>
</tr>
<tr>
<td>SV</td>
<td>5</td>
<td>17</td>
<td>31</td>
<td>34</td>
</tr>
<tr>
<td>Total</td>
<td>36</td>
<td>137</td>
<td>209</td>
<td>226</td>
</tr>
</tbody>
</table>
Table 4. Genetic variation estimates of microsatellite genotype data, including range and average number of alleles per locus (Na), Heterozygosity excess (He), Heterozygosity expected at equilibrium (Heq), probability of heterozygosity excess (One-tailed Wilcoxon test, P(e)), probability of heterozygosity excess or deficiency (Two-tailed Wilcoxon Test, P(w)) and probability of heterozygosity excess under TPM (Sign Test, P(s)).

<table>
<thead>
<tr>
<th>Herd</th>
<th>Na</th>
<th>He</th>
<th>Heq</th>
<th>F_is</th>
<th>P(e)</th>
<th>P(w)</th>
<th>P(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BT</td>
<td>2-7 (4.5)</td>
<td>0.66</td>
<td>0.61</td>
<td>0.09</td>
<td>0.08</td>
<td>0.17</td>
<td>0.55</td>
</tr>
<tr>
<td>CD</td>
<td>3-14 (7.4)</td>
<td>0.72</td>
<td>0.70</td>
<td>0.09</td>
<td>0.17</td>
<td>0.34</td>
<td>0.33</td>
</tr>
<tr>
<td>MM</td>
<td>4-7 (5)</td>
<td>0.59</td>
<td>0.60</td>
<td>-0.01</td>
<td>0.68</td>
<td>0.68</td>
<td>0.48</td>
</tr>
<tr>
<td>NS</td>
<td>3-8 (5)</td>
<td>0.66</td>
<td>0.63</td>
<td>-0.01</td>
<td>0.37</td>
<td>0.73</td>
<td>0.54</td>
</tr>
<tr>
<td>SV</td>
<td>3-7 (4.23)</td>
<td>0.63</td>
<td>0.59</td>
<td>0.04</td>
<td>0.17</td>
<td>0.34</td>
<td>0.55</td>
</tr>
</tbody>
</table>

Table 5. Pairwise fixation indices ($F_{st}$; left side) and effective migration rates (Nm; right side) estimated from microsatellite genotypes (above diagonal) and mitochondrial haplotypes (below diagonal) among all five bighorn sheep herd pairs in north-central Colorado, USA.

<table>
<thead>
<tr>
<th>$F_{st}$</th>
<th>BT</th>
<th>CD</th>
<th>MM</th>
<th>NS</th>
<th>SV</th>
<th>Nm</th>
<th>BT</th>
<th>CD</th>
<th>MM</th>
<th>NS</th>
<th>SV</th>
</tr>
</thead>
<tbody>
<tr>
<td>BT</td>
<td>–</td>
<td>0.02</td>
<td>0.03</td>
<td>0.09</td>
<td>0.02</td>
<td>BT</td>
<td>1.91</td>
<td>2.66</td>
<td>1.11</td>
<td>1.67</td>
<td></td>
</tr>
<tr>
<td>CD</td>
<td>0.26</td>
<td>–</td>
<td>0.03</td>
<td>0.03</td>
<td>0.04</td>
<td>CD</td>
<td>2.25</td>
<td>–</td>
<td>3.40</td>
<td>4.28</td>
<td>1.70</td>
</tr>
<tr>
<td>MM</td>
<td>0.25</td>
<td>0.02</td>
<td>–</td>
<td>0.09</td>
<td>0.05</td>
<td>MM</td>
<td>2.01</td>
<td>2.01</td>
<td>–</td>
<td>2.16</td>
<td>1.15</td>
</tr>
<tr>
<td>NS</td>
<td>0.44</td>
<td>0.10</td>
<td>0.18</td>
<td>–</td>
<td>0.09</td>
<td>NS</td>
<td>1.74</td>
<td>13.14</td>
<td>1.80</td>
<td>–</td>
<td>1.04</td>
</tr>
<tr>
<td>SV</td>
<td>0.25</td>
<td>0.16</td>
<td>0.19</td>
<td>0.19</td>
<td>–</td>
<td>SV</td>
<td>2.57</td>
<td>2.62</td>
<td>1.27</td>
<td>3.09</td>
<td>–</td>
</tr>
</tbody>
</table>

Estimates of both overall and pairwise effective migration (Nm) were comparable for analyses of the four, nine and 13 loci data sets (Nm = 8.96, range: 0.79 to 6.06, Nm = 8.82, range: 0.96 to 4.41 and Nm = 8.82, range: 1.04 to 4.28 respectively), and we therefore present
data for all 13 loci. Effective migration was moderate to high for both overall and pairwise herd comparisons (Table 5, Fig. 3). These estimates suggest that gene flow is likely contributing to low population subdivision among the five herds.
Figure 2. Map of the study area showing general geographic ranges for each of the five bighorn sheep herds. Overall fixation index ($F_{st}$) estimates for both microsatellite and mitochondrial are indicated in the top left box. Pairwise fixation indices are indicated in boxes between herd pairs the highest and lowest values for each marker type in bold. Herd abbreviations: BT = Big Thompson, CD = Continental Divide, MM = Mummy, NS = Never Summer, SV = St. Vrain.
Figure 3. Map of study area showing general geographic ranges for each of the five bighorn sheep herds. Overall effective migration (Nm) estimates for both microsatellite and mitochondrial data are indicated in top left box. Pairwise effective migration estimates are indicated in boxes between herd pairs with the highest and lowest values for each marker type in bold. Herd abbreviations: BT = Big Thompson, CD = Continental Divide, MM = Mummy, NS = Never Summer, SV = St. Vrain.
Inbreeding, as measured with the inbreeding coefficient ($F_{is}$), was estimated for each herd using three data sets consisting of four, nine and 13 loci (Table 6). Results between analyses were comparable for each herd and no large deviations were observed between these three estimates. All $F_{is}$ values were near zero as expected for non-bottlenecked herds with high gene flow; these estimates show no evidence of inbreeding.

Table 6. Coefficient of inbreeding ($F_{is}$) estimates for three microsatellite data subsets including: genotypes for only the four potentially “non-neutral” loci: OLA, KERA, SOMA and TCRBV (Four loci), genotypes for the nine putatively “neutral” loci: ADC, AE16, FCB266, FCB304, HH62, MAF33, MAF36, MAF48 and MAF209 (Nine loci) and genotypes from the complete data set of all 13 Loci (All loci) for five bighorn sheep herds in north-central Colorado, USA.

<table>
<thead>
<tr>
<th>Herd</th>
<th>Four loci</th>
<th>Nine loci</th>
<th>All loci</th>
</tr>
</thead>
<tbody>
<tr>
<td>Big Thompson</td>
<td>−0.104</td>
<td>0.164</td>
<td>0.090</td>
</tr>
<tr>
<td>Continental Divide</td>
<td>0.117</td>
<td>0.067</td>
<td>0.090</td>
</tr>
<tr>
<td>Mummy</td>
<td>−0.068</td>
<td>0.019</td>
<td>−0.010</td>
</tr>
<tr>
<td>Never Summer</td>
<td>−0.047</td>
<td>0.019</td>
<td>−0.010</td>
</tr>
<tr>
<td>St. Vrain</td>
<td>0.118</td>
<td>0.005</td>
<td>0.040</td>
</tr>
</tbody>
</table>

Isolation-by-distance analyses of the 10 herd pairs indicated a trend towards a negative relationship between log genetic relatedness ($M$) and log geographic distance ($R^2 = 0.167$, $p = 0.130$, $y = -1.28x + 2.55$), but no statistically significant relationship was revealed (Table 7). We find that the herd pair with the lowest $M$ is also the farthest geographically (NS/BT; 47.54 km),
while the pair with the highest M is the nearest geographically (CD/NS; 11.58 km); however, this trend is not consistent across all herd pairs (Fig. 4).
Table 7. Pairwise isolation-by-distance (IBD) estimates from mitochondrial and microsatellite data for among all five bighorn sheep herds.

<table>
<thead>
<tr>
<th>Herd pair</th>
<th>Geographic distance (km)</th>
<th>Log geographic distance (km)</th>
<th>Genetic similarity (M)</th>
<th>Log genetic similarity (M)</th>
<th>Genetic similarity (M)</th>
<th>Log genetic similarity (M)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BT/CD</td>
<td>37.33</td>
<td>1.57</td>
<td>0.47</td>
<td>-0.33</td>
<td>10.4</td>
<td>1.017</td>
</tr>
<tr>
<td>BT/MM</td>
<td>25.26</td>
<td>1.40</td>
<td>2.08</td>
<td>0.32</td>
<td>8.46</td>
<td>0.927</td>
</tr>
<tr>
<td>BT/NS</td>
<td>47.54</td>
<td>1.70</td>
<td>0.39</td>
<td>-0.42</td>
<td>2.58</td>
<td>0.412</td>
</tr>
<tr>
<td>BT/SV</td>
<td>24.37</td>
<td>1.39</td>
<td>0.75</td>
<td>-0.13</td>
<td>11.63</td>
<td>1.066</td>
</tr>
<tr>
<td>CD/MM</td>
<td>13.835</td>
<td>1.141</td>
<td>0.746</td>
<td>-0.127</td>
<td>8.1</td>
<td>0.908</td>
</tr>
<tr>
<td>CD/NS</td>
<td>11.577</td>
<td>1.064</td>
<td>4.151</td>
<td>0.618</td>
<td>6.92</td>
<td>0.840</td>
</tr>
<tr>
<td>CD/SV</td>
<td>31.828</td>
<td>1.503</td>
<td>0.834</td>
<td>-0.079</td>
<td>6.92</td>
<td>0.840</td>
</tr>
<tr>
<td>MM/NS</td>
<td>22.501</td>
<td>1.352</td>
<td>0.67</td>
<td>-0.174</td>
<td>2.56</td>
<td>0.669</td>
</tr>
<tr>
<td>MM/SV</td>
<td>28.568</td>
<td>1.456</td>
<td>0.714</td>
<td>-0.146</td>
<td>4.67</td>
<td>0.669</td>
</tr>
<tr>
<td>NS/SV</td>
<td>43.318</td>
<td>1.637</td>
<td>0.691</td>
<td>-0.161</td>
<td>2.66</td>
<td>0.425</td>
</tr>
</tbody>
</table>
Figure 4. Isolation-by-distance and reduced major axis regression analysis of microsatellite genotype data. Plot of log genetic similarity (M) vs. log geographic distance (km) for all ten bighorn sheep population pairs. Herd abbreviations: BT = Big Thompson, CD = Continental Divide, MM = Mummy, NS = Never Summer and SV = St. Vrain. Line of best fit for these data follows the equation $y = -1.28x + 2.55 (R^2 = 0.167, p = 0.13)$. 
Qualitative evaluation of overall metapopulation substructure (from Q plots), as evaluated in STRUCTURE, shows no clear genetic division between herds for any of the three microsatellite data sets (four, nine and 13 loci). Maximum likelihood (ML) values for predicted herd number (K) were not significantly different between K = four, five or six herds, indicating low confidence in assignment of individuals to specific herds based on microsatellite genotypes. These data further support $F_{st}$ and $Nm$ estimates, suggesting low metapopulation substructure among these five herds.

Of the 455 microsatellite heterozygosity (He) estimates obtained from BOTTLENECK across all herd/locus/model combinations, only 63 deviated significantly from expected heterozygosity at mutation-drift equilibrium (Fig. 5). Within these deviations we did not find a consistent pattern for any specific locus among herds or across multiple loci for any single herd. The largest number of deviations was found in the Mummy herd (19 of 91) and the fewest in the Big Thompson herd (7 of 91). For each individual herd the scatter plots patterns were consistent across all seven microsatellite mutation models (Fig. 6). We identified a clear positive correlation between increasing percentage of stepwise mutations (SMM) in the two-phase mutation model (TPM) and number of significant deviations from expected heterozygosity. This is consistent with the general understanding that the strict SMM is a poor model for describing microsatellite evolution. In addition, the Sign Test revealed no significant heterozygosity excess, and the Wilcoxon test found no significant deviations from expected heterozygosity for any of the five herds under any of the seven mutation models tested (Table 4). The “mode-shift” bar graphs showed normal, L-shaped distributions of microsatellite variation by frequency class across all five herds, with the possible exception of the Big Thompson herd, but this is most likely due to small sample size.
Figure 5. Scatter plots showing individual microsatellite deviations from expected heterozygosity across seven different mutation models of microsatellite evolution where IAM = Infinite Allele Model, SMM = Strict Stepwise Mutation Model and TPM = Two-Phase Mutation Model as implemented in the program BOTTLENECK.
Figure 6. BOTTLENECK bar chart of allele distributions by frequency classes from least common (1) to most common (10) (left); Scatter plot distributions of individual microsatellite deviations from excepted heterozygosity at equilibrium (right). Herd abbreviations: Big Thompson (BT), Continental Divide (CD), Mummy (MM), Never Summer (NS), St. Vrain (SV).
**Mitochondrial sequence variation and genetic diversity** - High quality sequence data for both COI and DL were obtained from 211 DNA samples and, after removal of duplicates, our final mitochondrial data set included complete haplotypes for 190 individuals from all five populations (Table 3). The full length 1,937 bp sequence contained 105 polymorphic sites, five within COI and 100 within DL. A single invariant 79 bp indel (insertion or deletion) was identified within the DL sequence and found to be polymorphic in the Big Thompson, Mummy and St. Vrain herds while only the deletion was detected in the Continental Divide and Never Summer herds. A total of twelve unique haplotypes were identified from among these herds (Table 8).

Table 8. Distribution and abundance of each of the twelve mitochondrial haplotypes identified in this study among all five herds of bighorn sheep in north-central Colorado, USA.

<table>
<thead>
<tr>
<th>Haplotype</th>
<th>Big Thompson</th>
<th>Continental Divide</th>
<th>Mummy</th>
<th>Never Summer</th>
<th>St. Vrain</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>13</td>
<td>2</td>
<td>21</td>
<td>0</td>
<td>1</td>
<td>37</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>16</td>
<td>23</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>11</td>
<td>7</td>
<td>10</td>
<td>0</td>
<td>28</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>36</td>
<td>0</td>
<td>13</td>
<td>0</td>
<td>49</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>9</td>
<td>1</td>
<td>16</td>
<td>9</td>
<td>35</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>20</td>
<td>70</td>
<td>30</td>
<td>41</td>
<td>29</td>
<td>190</td>
</tr>
</tbody>
</table>
No single haplotype was shared among all five herds. Six of the 12 total haplotypes were found in only a single herd. Overall haplotype diversity ($H_d$) and nucleotide diversity ($\pi$) were 0.826 and 0.129 respectively. The highest haplotype and nucleotide diversities were found in the Never Summer herd while the lowest were found in the Big Thompson herd (Table 9).

Table 9. Estimates of genetic diversity for the mitochondrial DNA haplotypes comprising 1,937 bp of Cytochrome Oxidase I (COI) and D-Loop (DL) sequences for each of the five bighorn sheep herds from north-central Colorado, USA including total sample size (N), segregating sites (S), number of haplotypes (h), haplotype diversity ($H_d$), average number of nucleotide differences (K), nucleotide diversity ($\pi$).

<table>
<thead>
<tr>
<th>Herd</th>
<th>N</th>
<th>S</th>
<th>h</th>
<th>$H_d$</th>
<th>K</th>
<th>$\pi$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Big Thompson</td>
<td>20</td>
<td>3</td>
<td>2</td>
<td>0.479</td>
<td>1.44</td>
<td>0.014</td>
</tr>
<tr>
<td>Continental Divide</td>
<td>70</td>
<td>42</td>
<td>7</td>
<td>0.682</td>
<td>12.07</td>
<td>0.115</td>
</tr>
<tr>
<td>Mummy</td>
<td>30</td>
<td>38</td>
<td>4</td>
<td>0.469</td>
<td>13.18</td>
<td>0.130</td>
</tr>
<tr>
<td>Never Summer</td>
<td>41</td>
<td>40</td>
<td>5</td>
<td>0.703</td>
<td>16.56</td>
<td>0.159</td>
</tr>
<tr>
<td>St. Vrain</td>
<td>29</td>
<td>23</td>
<td>6</td>
<td>0.633</td>
<td>9.88</td>
<td>0.093</td>
</tr>
<tr>
<td>Total</td>
<td>190</td>
<td>45</td>
<td>12</td>
<td>0.826</td>
<td>13.03</td>
<td>0.129</td>
</tr>
</tbody>
</table>

Overall population subdivision was much higher for mitochondrial haplotypes ($F_{st} = 0.25$) than for microsatellite genotypes. The range of pairwise $F_{st}$ estimates was likewise larger ($F_{st} = 0.02$ to 0.26) with a comparable low value but a much larger high value than for microsatellite data. The lowest pairwise estimate was between CD and MM and the largest was between BT and NS (Table 5, Fig. 2). The overall effective migration among herds estimated from mitochondrial sequences ($N_m = 8.82$) was lower than from microsatellite genotypes (Fig. 3). The range of pairwise effective migration estimates between herd pairs was likewise larger than for microsatellite data with a similar low but a much larger high value (Table 5, Fig 3).
Isolation-by-distance analyses of the 10 herd pairs indicated a significant negative correlation between genetic relatedness (M) and log geographic distance for all four analysis combinations of geographic distance or log geographic distance vs. genetic relatedness or log genetic relatedness (Table 7). These data were also consistent for analyses using both F_{st} and Φ_{st}. Consistent with our microsatellite data, we find that the herd pair with the lowest M is also the farthest geographically (NS/BT) while the pair with the highest M is the closest geographically (CD/NS), however this trend is not consistent across all herd pairs and these results may reflect influences of translocations to BT and MM from outside herds (Fig. 7).
Figure 7. Isolation-by-distance and reduced major axis regression analysis of mitochondrial sequence data as implemented in Isolation by Distance Web Service. Plot of log genetic similarity (M) vs. log geographic distance (km) for all ten bighorn sheep population pairs. Herd abbreviations: BT = Big Thompson, CD = Continental Divide, MM = Mummy, NS = Never Summer and SV = St. Vrain. Line of best fit for these data follows the equation $y = -1.54x + 0.56$ ($R^2 = 0.69$, $p = 0.02$).
DISCUSSION

The distribution of bighorn sheep is typically naturally fragmented due to specific habitat requirements, and population sizes are typically small, making bighorn sheep susceptible to genetic drift, which can result in loss of allelic variation and decreased heterozygosity (Schwartz et al. 1986, Bleich et al. 1990, Gutiérrez-Espeleta et al. 2000, Berger 1990). However, migration and resulting gene flow among neighboring herds can ameliorate the effects of genetic drift by introducing new alleles and contributing to the maintenance of genetic variation in individual herds. We found no evidence of a recent genetic bottleneck or severe reduction in genetic diversity for any of these five Rocky Mountain bighorn sheep herds in and near RMNP. Our data do not support the hypothesis that inbreeding depression was a factor in the failure of the Mummy herd to recover to historic numbers following the pneumonia outbreak in the mid 1990s. Our data also suggest that heterozygosity within these five herds is maintained by gene flow within this metapopulation.

Heterozygosity is an important component of population genetic analyses, because it is often positively correlated with increased fitness, resistance to disease and long-term persistence of natural populations (Keller and Waller 2002, Jannikke et al. 2006, Xu et al. 2007, O’Grady et al. 2006). Positive correlations between heterozygosity and both fitness (FitzSimmons et al. 1995) and resistance to disease have also been documented in bighorn sheep (Coltman et al. 1999, Luikart et al. 2008a). Disease is often a significant factor in long-term persistence, of bighorn populations because they are particularly susceptible to pneumonia outbreaks which can result in large die-offs or local extinctions (Bailey 1986, Shackleton et al. 1999, Cassirer and Sinclair, 2007, Wehasausen et al. 2011). We found expected heterozygosities in RMNP.
0.59-0.72) to be comparable to the highest values reported in other studies of native bighorn sheep populations (Boyce et al. 1997, Forbes and Hogg 1999, Epps et al. 2010, Johnson et al 2011, Luikart et al. 2011) and higher than those reported for restored bighorn sheep populations (Whitaker et al. 2004).

Our inbreeding coefficient estimates support these heterozygosity analyses in that no significant loss of heterozygosity, as indicated by a positive F_is, was detected. All F_is estimates were near zero (-0.01 – 0.09) suggesting that no herd has experienced a recent genetic bottleneck which would be indicated by a disproportionate loss of rare alleles. It is therefore unlikely that any of these herds are suffering from the effects of inbreeding depression. Factors other than genetic diversity are therefore likely responsible for the Mummy herd’s failure to thrive.

Metapopulation analyses suggest that high heterozygosity is being maintained by gene flow among these five herds. Estimates of population substructure from microsatellite data were low (F_st = 0.02 - 0.09, Fig. 2), effective migration rates were reasonably high (Nm = 1.04 to 4.28, Fig. 3), and no significant correlation was found between allelic distribution and geographic distance (Fig. 4). In contrast, mitochondrial data showed higher genetic isolation among herds (F_st = 0.02 to 0.44, Fig. 2) and a significant negative correlation between haplotype distribution and geographic distance (Fig. 6). The ranges of both effective migration rates (Nm = 1.04 to 13.14) and population substructure were larger for mitochondrial data than for microsatellite data, suggesting that ewes preferentially move between certain herds more than between others. Additionally, the pattern of population substructure is asymmetrically distributed across this metapopulation. The two lowest F_st estimates involved the CD herd (CD/MM = 0.02, CD/NS = 0.10) which could be attributed increased migration between these herds due to shared lambing areas where herds overlap, such as at Specimen Mountain on the west side of RMNP, which
could also explain the extremely high effective migration estimate between the NS and CD herds (Nm = 13.14). The four highest Fst estimates were between the BT herd and each of the Park herds (BT/NS = 0.44, BT/CD = 0.26, BT/MM = 0.25, BT/SV = 0.25) which could indicate that ewe-specific movement is inhibited to some extent by the town of Estes Park (population 6,017), which lies directly between RMNP and the BT herd.

Recent research has shown that both slope and ruggedness of terrain are significant factors in bighorn habitat selection (Sappington et al. 2007) and that highways block gene flow in this species (Epps et al. 2005). This could help explain the limited dispersal of ewes across the relatively flat landscape between RMNP and the BT range which includes highways 36, 14 and 7. In contrast, our microsatellite data do not show this same disparity between movement among Park herds and movement between RMNP herds and the BT herd, suggesting that ram dispersal across populated areas is not inhibited to the same extent as ewe dispersal or that motivation to migrate is higher in rams because of the potential for increased mating success in non-natal herds.

The disparity between Fst estimates from microsatellite and mitochondrial data could also be explained to some extent by differences in the evolution and inheritance patterns of these two markers. Genes located on the mitochondrial chromosome are vital to metabolic processes involved in aerobic respiration and are thus more conserved than microsatellite alleles which are neutral and rapidly mutate via replication slippage and are thus generally highly variable in populations. Mitochondria are also haploid and maternally inherited while microsatellites are diploid and biparentally inherited. Thus, when the sex ratio is equal, mitochondrial markers have an effective population one quarter that of microsatellite markers (Birky et al. 1983). However, this disparity is offset, to some extent, by the highly polygynous mating structure of bighorn
sheep in which most adult ewes reproduce each year while generally only the most dominant rams successfully mate (Geist 1966, Geist 1971, Coltman et al. 2005).

A closer look at mitochondrial haplotype distributions also reveals an unexpected pattern among these herds in that haplotypes one and two are the only haplotypes detected in the BT herd while the NS herd, at the opposite end of this metapopulation, has five haplotypes, but is lacking haplotypes one and two (Table 8). Thus, our data show no haplotypes in common between the BT and NS herds, which suggests that females may not have moved between these populations and bred successfully in recent history. The mtDNA differentiation can be seen more clearly if haplotypes one and two are pooled into "haplotype E" and all other haplotypes are pooled into "haplotype W", so that the BT herd is fixed (frequency is 100%) for the E haplotype, and the NS herd is fixed for the W haplotype. If we use Sewall Wright's calculation of $F_{st}$ to compare these populations for the pooled data, then $F_{st} = 1.0$. But when we use Nei's formulation (Nei et al. 1975) for multiallelic data, the value is $F_{st} = 0.25$. This is because Nei's method, used most commonly for multiallelic data, considers only the average decrement in heterozygosity within sub populations in comparison to heterozygosity in the full population, but it does not measure which alleles are present in the various subpopulations, or the degree of allelic differentiation among sub populations. This complete mitochondrial differentiation between the NS and BT herds raises questions concerning whether the intervening territory marks an historic boundary to ewe-mediated gene flow or whether this differentiation results from the genetic signature of novel mitochondrial haplotypes introduced via sheep translocation from the Georgetown herd and that gene flow has yet to occur to any great extent since then.

Interpretation of these metapopulation substructure data must be tempered, however, by the history of sheep translocations involving study herds (Table 1). Starting in 1977, 20 sheep
from the Tarryall population (South Platte River Canyon, CO) were introduced into the MM herd. During the next decade, a total of 71 sheep were translocated between the MM herd to the BT herd on three separate occasions; in 1983 (19 sheep), in 1985 (26 sheep) and in 1987 (26 sheep). The most recent translocation occurred in 2000 when 22 sheep from the Georgetown population (Georgetown, CO) were introduced into the BT herd. These translocations were primarily composed of ewes, between 59 and 99 of the total 113 sheep moved, depending on the unknown sexes of both yearlings and lambs (George et al. 2009).

This, and the highly polygynous mating structure of bighorn sheep, leads to two predictions. First The MM and BT herds should each have a higher proportion of specific mitochondrial haplotypes and microsatellite alleles compared with other herds due to the induced migration from the Tarryall and Georgetown herds. If incorporated into local gene pools, this induced migration should result in higher pairwise $F_{st}$ estimates between each of these two herds and the other three (CD, NS and SV) herds. Second, due to the repeated translocation of sheep (primarily ewes) from the MM herd to the BT herd, these herds should exhibit a higher proportion of shared genetic variation, primarily mitochondrial haplotypes, resulting in lower $F_{st}$ estimates relative to the other herds. Our data, however, do not match either of these predictions. Although BT mitochondrial pairwise $F_{st}$ estimates are the highest in this metapopulation, the BT/MM estimate is also high. The corresponding BT microsatellite pairwise $F_{st}$ estimates are all low and comparable with other pairwise values (Figure 3). MM mitochondrial pairwise $F_{st}$ estimates are variable, ranging from 0.02 to 0.26 and the Mm microsatellite pairwise $F_{st}$ estimates are also low and comparable with others in this metapopulation. This could be attributed to low transplantation success rates which have been repeatedly documented in bighorn, to the loss of transplanted sheep subsequent loss during the pneumonia induced die-off.
Taken together, these data indicate that the degree to which translocations of sheep have influenced our results is unclear.

Quantification of population genetics for natural populations is hindered by a number of factors, including incomplete historical data, uneven sampling, potentially unrealistic assumptions for parameter estimations, and potential species-specific complications. Thus, interpretation of population genetic data can be complex and, while comprehensive evaluation of different genetic markers can help elucidate important patterns, data limitations need to be acknowledged. Although the transplantation history makes some interpretations difficult, our analyses of both mitochondrial and microsatellite data nevertheless found that the five Rocky Mountain bighorn sheep herds have high heterozygosity, show no evidence of inbreeding, and exhibit low metapopulation substructure due to on-going gene flow, a finding consistent with other analyses of bighorn sheep metapopulations with unimpaired migration. While our data are clearly not consistent, they do clearly disprove the hypothesis that decreased genetic variation is responsible for the Mummy herd’s failure to thrive and suggest that other factors should be considered.
CHAPTER III

INTRODUCTION

The subfamily Caprinae (Artiodactyla, Bovidae) arose around 14.1 MYA in Asia with subsequent radiation around the world (Hernández-Fernández and Vrba 2005). The Caprinae includes sheep, goats and closely related horned species that are often characterized by adaptations to northern latitudes and high altitudes. Unlike other, more homogeneous taxonomic groups, genera of Caprinae are characterized by a diverse set of morphological traits and do not share any unambiguous synapomorphies (Gentry 1992), which may explain why many taxonomies have been proposed for this group and a consensus has not been reached. In the last fifty years many different genera have been proposed within this group. Of these, 10 to 12 are currently generally accepted: Ammotragus (aoudad), Budorcas (takin), Capra (goat, ibex, markhor, tur), Capricornis (serow), Naemorhedus (goral), Oreamnos (mountain goat), Ovibos (musk ox), Ovis (true sheep), Pseudois (bharal, blue sheep) and Rupicapra (chamois). The additional genus Hemitragus (tahr) was recently found to be polyphyletic and two additional genera, Nilgiritragus (Nilgiri tahr) and Arabitragus (Arabian tahr), have been added as a result (Ropiquet and Hassanin 2005b). Both of these newer genera have been subsequently supported by additional data (Shafer and Hall 2010, Bibi et al. 2012). Inclusion of two additional genera, Saiga (saiga) and Pantholops (chiru), has been historically debated (Simpson 1945, Nowak 1991, Gentry 1992, Groves and Shields 1996, Gatesy et al. 1997, Hassanin et al. 1997, Hassanin and Douzery 1999), yet more recent molecular analyses have placed Saiga with Gazella in the subfamily Antilopinae (Ropiquet and Hassanin 2005a), while Pantholops has been supported as

Not only are the absolute number of genera of the Caprinae debated, but also the evolutionary relationships among them, for which many different taxonomies have been proposed based on a diverse set of morphological, behavioral, ecological, and genetic data. These analyses have led to taxonomies which range from including all genera within a single tribe (Caprini) (Hassanin and Douzery 1999) to dividing genera among as many as four different tribes (Caprina, Ovibovini, Panthalopina, and/or Rupicaprin) (Simpson 1945, Groves and Shields 1996, Gatesy 1997, Hassanin et al. 2009, Arif et al. 2012, Bibi et al. 2012). One reason for the incongruence of findings among studies is that very few have included most or all of the genera in this subfamily, instead focusing on the placement of only one or a few taxa. In addition, most genetic analyses to date have focused solely on mitochondrial markers, with the majority of these focused on a single gene (Groves and Shields 1996, Gatesy et al. 1997, Hassanin et al. 1998, Hassanin and Douzery 1999, Hassanin et al. 2009, An et al. 2010, Bibi et al. 2012), while only a few included multiple mitochondrial markers (Hassanin and Douzery, 2003, Ropiquet and Hassanin 2005a) or whole mitogenome analyses (Hassanin et al. 2009, Arif et al. 2012). At the same time, relatively few nuclear markers have been developed for these genera (Matthee and Davis 2001, Ropiquet and Hassanin 2003, Ropiquet and Hassain 2005b, Ropiquet and Hassanin, 2006, Pérez et al. 2011), and studies comparing nuclear and mitochondrial phylogenies report that results are often not concordant between data sets (Ropiquet and Hassanin 2006, Schaefer and Hall 2010).

To address the previous lack of comprehensive phylogenetic analyses of Caprinae, we provide data in this study which include most or all Caprinae genera. Our phylogenetic analyses
include twelve of the thirteen extant genera of Caprinae (excepting *Nilgritragus*). To address the prior paucity of comparative analyses incorporating diverse markers, our data include three distinct sequences from two different markers; the mitochondrial gene ND5 and two regions of the nuclear gene sex-determining region Y (SRY). Our data provide the first SRY open reading frame (ORF) sequence data for 15 species in nine genera and the first SRY promoter (PRO) sequence data for 13 species in 11 genera. With these data we compare the relative resolving power of the highly variable ND5 gene with the highly conserved SRY gene to determine if and when each marker is better at clarifying the evolutionary relationships among Caprinae genera.

**MATERIALS AND METHODS**

*Data sets* - Three distinct data sets were developed for this study; 1) the full-length SRY ORF sequence (ORF); 2) the partial SRY promoter sequence (PRO); and the partial ND5 sequencer (ND5). We combined all available non-redundant published sequences (ORF: 2, PRO: 11, ND5: 5) with our own sequence data (ORF: 31, PRO: 26, ND5: 27) to ultimately include data for 12 genera and 24 species within the subfamily Caprinae. We analyzed each of the three data sets independently as well as a fourth combined Y-chromosome data set (PRO/ORF) and a fifth complete dataset (ND5/PRO/ORF). For combined analyses we only included samples with both PRO and ORF data and only samples with at least two of the three markers were included in the complete ND5/PRO/ORF data set (Table 10).
<table>
<thead>
<tr>
<th>Genus</th>
<th>Species</th>
<th>SRY ORF samples</th>
<th>GenBank Accession no.</th>
<th>Reference</th>
<th>SRY PRO samples</th>
<th>GenBank Accession no.</th>
<th>Reference</th>
<th>ND5 samples</th>
<th>GenBank Accession no.</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ammotragus</td>
<td>lavia</td>
<td>1</td>
<td>EU938019</td>
<td>Meadows &amp; Kijas, 2009</td>
<td>1</td>
<td>present study</td>
<td></td>
<td></td>
<td>EU938020</td>
<td>Meadows &amp; Kijas, 2009</td>
</tr>
<tr>
<td>Arabitragus</td>
<td>jayakai</td>
<td>1</td>
<td>FJ207523</td>
<td>Hassaanin, 2009</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Budorcas</td>
<td>taxicolor</td>
<td>1</td>
<td>D82963</td>
<td></td>
<td>1</td>
<td>AF53341</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Capra</td>
<td>aegagrus</td>
<td>1</td>
<td>JN992655</td>
<td>present study</td>
<td>1</td>
<td>D82963</td>
<td>JN992656</td>
<td></td>
<td>JF680994</td>
<td>present study</td>
</tr>
<tr>
<td></td>
<td>ibex</td>
<td>1</td>
<td>JN992664, JN992665</td>
<td>present study</td>
<td>2</td>
<td>JN992658, JN992655</td>
<td>present study</td>
<td></td>
<td>JN992660</td>
<td>present study</td>
</tr>
<tr>
<td></td>
<td>falconeri</td>
<td></td>
<td>JN992660</td>
<td>present study</td>
<td>1</td>
<td>JN992669</td>
<td>present study</td>
<td></td>
<td>JN992669</td>
<td>present study</td>
</tr>
<tr>
<td>Capricornis</td>
<td>umbratilis</td>
<td>1</td>
<td>JN992668</td>
<td>present study</td>
<td>1</td>
<td>JN992668</td>
<td>present study</td>
<td></td>
<td>JN992668</td>
<td>present study</td>
</tr>
<tr>
<td>Hemitragus</td>
<td>jomonensis</td>
<td>1</td>
<td>JN992670</td>
<td>present study</td>
<td>1</td>
<td>JN992679</td>
<td>present study</td>
<td></td>
<td>JN992679</td>
<td>present study</td>
</tr>
<tr>
<td>Oreamnos</td>
<td>americanus</td>
<td>1</td>
<td>JN992671</td>
<td>present study</td>
<td>1</td>
<td>JN992682</td>
<td>present study</td>
<td></td>
<td>JN992682</td>
<td>present study</td>
</tr>
<tr>
<td>Ovis</td>
<td>ammon</td>
<td>1</td>
<td>Z38265</td>
<td>Payen &amp; Cotinot, 1994</td>
<td>1</td>
<td>EU938022</td>
<td>Meadows &amp; Kijas, 2009</td>
<td>1</td>
<td>present study</td>
<td></td>
</tr>
<tr>
<td></td>
<td>canadensis</td>
<td>many IN992662, IN992669</td>
<td>JN400247 - JN400247</td>
<td>present study</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>JF680994</td>
<td>present study</td>
</tr>
<tr>
<td></td>
<td>dalli dalli</td>
<td>1</td>
<td>IN992668</td>
<td>present study</td>
<td>1</td>
<td>JN992680</td>
<td>present study</td>
<td></td>
<td>JN992680</td>
<td>present study</td>
</tr>
<tr>
<td></td>
<td>dalli stonei</td>
<td>1</td>
<td>JN992670</td>
<td>present study</td>
<td>1</td>
<td>JN992679</td>
<td>present study</td>
<td></td>
<td>JN992679</td>
<td>present study</td>
</tr>
<tr>
<td></td>
<td>nivicola</td>
<td>1</td>
<td>JN992671</td>
<td>present study</td>
<td>1</td>
<td>JN992682</td>
<td>present study</td>
<td></td>
<td>JN992682</td>
<td>present study</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>JN992683</td>
<td>present study</td>
<td></td>
<td>JN992683</td>
<td>present study</td>
</tr>
<tr>
<td>Saiga</td>
<td>tartarica</td>
<td>1</td>
<td>Z38265</td>
<td>Payen &amp; Cotinot, 1994</td>
<td>1</td>
<td>EU938022</td>
<td>Meadows &amp; Kijas, 2009</td>
<td>1</td>
<td>present study</td>
<td></td>
</tr>
<tr>
<td>Pantholops</td>
<td>hodgsonii</td>
<td>1</td>
<td>JN992672</td>
<td>present study</td>
<td>1</td>
<td>JN992682</td>
<td>present study</td>
<td></td>
<td>JN992682</td>
<td>present study</td>
</tr>
<tr>
<td>Pseudois</td>
<td>nayaur</td>
<td>1</td>
<td>JN992673</td>
<td>present study</td>
<td>1</td>
<td>JN992683</td>
<td>present study</td>
<td></td>
<td>JN992683</td>
<td>present study</td>
</tr>
<tr>
<td></td>
<td>scaepferi</td>
<td>1</td>
<td>JZ38265</td>
<td>Payen &amp; Cotinot, 1994</td>
<td>1</td>
<td>EU938023</td>
<td>Meadows &amp; Kijas, 2009</td>
<td>1</td>
<td>present study</td>
<td></td>
</tr>
<tr>
<td>Rupicapra</td>
<td>rupicapra</td>
<td>1</td>
<td>JN547786</td>
<td>Perez et al, 2011</td>
<td>1</td>
<td>FJ207539</td>
<td>Hassaanin et al, 2009</td>
<td>1</td>
<td>present study</td>
<td></td>
</tr>
<tr>
<td></td>
<td>pyrenaica</td>
<td>1</td>
<td>JN547785</td>
<td>Perez et al, 2011</td>
<td>1</td>
<td>FJ207537</td>
<td>Hassaanin et al, 2009</td>
<td>1</td>
<td>present study</td>
<td></td>
</tr>
</tbody>
</table>
Sequence analyses - The complete 727 bp SRY ORF was amplified with primers previously designed from mouflon sheep (*O. o. musimon*) genomic sequence (Payen and Cotinot 1994). A 481 bp segment of the SRY promoter located 1,238 bp upstream from the ORF start codon was amplified with primers designed for this study (Table 11). All Y-chromosome amplifications were conducted in 25µl reactions containing 50ng of template DNA, 2mM MgCl₂, 0.05mM dNTPs, 0.4µM of each primer and 0.25 units of HotStart-IT Taq DNA Polymerase (Affymetrix Inc. Santa Clara, CA). The PCR profile included: One cycle of 5min at 94°C followed by 40 cycles of: 1min at 94°C, 1min at 61°C (PRO) or 52°C (ORF), 1min at 72°C and one final cycle of 5min at 72°C. Sequencing was conducted by Functional Biosciences, Inc. (Madison, WI). All alignments, including both sequences from single samples as well as complete alignments including all Caprinae samples, were performed in Sequencher™ version 4.6 (Gene Codes Corporation; Ann Arbor, MI). All sequences were checked by hand for accuracy and polymorphisms were confirmed manually.

<table>
<thead>
<tr>
<th>Locus</th>
<th>Primer name</th>
<th>Primer sequence (5’-3’)</th>
<th>bp</th>
<th>Tm (°C)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORF</td>
<td>SRY1-F</td>
<td>TGCCAGGAGGTATTGAGGGG</td>
<td>723</td>
<td>52</td>
<td>Payen and Cotinot, 1994</td>
</tr>
<tr>
<td></td>
<td>SRY1-R</td>
<td>CAGAGGAGCAGTTATTTTG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PRO</td>
<td>SRY3A-F</td>
<td>AACAATTGCATGTAGCTCCAGA</td>
<td>481</td>
<td>61</td>
<td>present study</td>
</tr>
<tr>
<td></td>
<td>SRY3A-R</td>
<td>CAATGACAGTAGAAAACCTCAAAGG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ND5</td>
<td>ND5A-F</td>
<td>AGTAGTTATCCATTCGCTTTAGG</td>
<td>665</td>
<td>58</td>
<td>(Wehausen, unpublished data)</td>
</tr>
<tr>
<td></td>
<td>ND5A-R</td>
<td>AAGATTTGGTGGAGATCTCAGGTG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ND5</td>
<td>ND5B-F</td>
<td>CCAACACAGCAGCCTACAAGC</td>
<td>690</td>
<td>61</td>
<td>(Wehausen, unpublished data)</td>
</tr>
<tr>
<td></td>
<td>ND5B-R</td>
<td>GGTCCTTTGGGAGTGAATCCCGGTG</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
We sequenced 1,110 bp of the mtDNA gene ND5, the largest and one of the faster evolving mitochondrial protein coding genes (Lopez-Ramirez and Saavedra-Molina, 1997) via two overlapping segments amplified with separate primer pairs and annealing temperatures (Table 11). For 22uL reactions, PCRs included 1X PCR buffer, 1.9mM MgCl₂, 0.16mM each dNTP, 5X BSA (New England Biolabs), 0.4uM each primer, and 8.8 units Amplitaq Gold (Applied Biosystems). The PCR profile included one cycle of 7 min at 94°C followed by 35 cycles of: 1 min at 94°C, 70 sec at 58/61°C, and 1.5 min at 72°C. After examining 5uL of PCR products on an agarose gel, cleanup was: 12uL PCR product + 2.4uL shrimp alkaline phosphatase (SAP) + 2.4uL exonuclease I (Exo1) at 37°C for 30min then 80°C for 10min. Cycle sequencing reactions combined 7.5uL of cleaned PCR product, 3uL BigDye3.1 (Applied Biosystems), 0.255uL of 10uM primer, and 4.295uL water, and used the following temperature profile: 25 cycles of 96°C 10sec, 50°C 5sec, and 60°C 4min. This was cleaned with DyeEx columns (Qiagen) and dried down. Rehydration used 0.4uL dextran blue dye and 5uL deionized formamide. All samples were sequenced in both directions and electrophoresed on an ABI 373 or 377 automated DNA sequencer using associated Sequence Analysis software (ABI). Forward and reverse sequences were aligned and edited using Sequencher software. Calculations are summarized in Table 12, including: haplotype number (h), number of segregating sites (S), genetic distance (K), and estimates of nucleotide diversity (π, o) were performed in DnaSP version 4.0 (Rozas et al. 2003).
Table 12: Summary DNA sequence statistics by data set including sequence length, number of haplotypes (h), nucleotide diversity (π), nucleotide diversity with Jukes-Cantor correction (πJC), number of segregating sites (S), average number of nucleotide differences (K), theta per nucleotide site as estimated from S (ӨW), and theta per sequence as estimated from S (Ө).

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Length (bp)</th>
<th>h</th>
<th>π</th>
<th>πJC</th>
<th>S</th>
<th>K</th>
<th>ӨW</th>
<th>Ө</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORF</td>
<td>727</td>
<td>21</td>
<td>0.024</td>
<td>0.025</td>
<td>116</td>
<td>17.059</td>
<td>0.038</td>
<td>26.667</td>
</tr>
<tr>
<td>PRO</td>
<td>481</td>
<td>21</td>
<td>0.026</td>
<td>0.026</td>
<td>61</td>
<td>11.909</td>
<td>0.032</td>
<td>15.030</td>
</tr>
<tr>
<td>ND5</td>
<td>1110</td>
<td>39</td>
<td>0.108</td>
<td>0.118</td>
<td>471</td>
<td>119.520</td>
<td>0.101</td>
<td>111.403</td>
</tr>
<tr>
<td>PRO/ORF</td>
<td>1208</td>
<td>18</td>
<td>0.025</td>
<td>0.024</td>
<td>142</td>
<td>28.100</td>
<td>0.034</td>
<td>40.661</td>
</tr>
<tr>
<td>ND5/PRO/ORF</td>
<td>2318</td>
<td>27</td>
<td>0.072</td>
<td>0.121</td>
<td>442</td>
<td>122.248</td>
<td>0.104</td>
<td>114.674</td>
</tr>
</tbody>
</table>

Phylogenetic analyses - Each of the three individual data sets (ORF, PRO and ND5) were analyzed separately as well as together in a concatenated fourth (PRO/ORF) and fifth (ND5/PRO/ORF) data set to maximize molecular information and total number of genera included. All PRO sequences were assigned a simple nucleotide model while the codon model with partitioned base sites was used for all ORF and ND5 sequences. The combined Y-chromosome data set was partitioned as PRO (1-481 bp), ORF (482-1,208 bp) and the full concatenated data set was partitioned as: ND5 (1-1,110 bp), PRO (1,111-1,591 bp), ORF (1,592-2,318 bp). Maximum parsimony (MP), maximum likelihood (ML) and Bayesian methods were used to reconstruct phylogenies for each data set. All trees were rooted with the outgroup Saiga tartarica and were analyzed and edited in FigTree 1.3.1 (Rambaut 2006).

Appropriate models of nucleotide substitution for ML and Bayesian analyses were identified using both the –lnL and the Akaike Information Criterion (AIC) scores (Akaike 1974,
Hurvich and Tsai 1989) as implemented in jModelTest 0.1.1 (Posada and Crandall 1998; Posada, 2008). Selected models were: GTR + G for ORF, PRO and PRO/ORF data sets and GTR+ I + G for ND5 and ND5/PRO/ORF data sets.

MP analyses were performed in PAUP 4.0 (Swofford 2003). Unweighted MP trees (Hasegawa 1985) were estimated using heuristic searches with indels designated as a 5th state change and no limit on the number of trees retained. Branch support was assessed with 10,000 bootstrap replicates using tree bisection-reconnection (TBR) branch swapping. ML analyses were performed in RAxML 1.3 (Silvestro and Michalak 2011). Branch supports were calculated with 10,000 thorough bootstrap replicates on the extended majority rule (MRE) consensus tree for each data set. Bayesian analyses were performed in MrBayes version 3.2.1 (Huelsenbeck and Ronquist 2001) using the GTR + G model for ORF, PRO and PRO/ORF data sets and the GTR + I + G for the ND5 and ND5/PRO/ORF data sets as identified in JModelTest from both – Lnl and Aikakie Information Criteria (AIC) scores. All analyses included 1,000,000 runs sampling every 100 trees with a burn-in of 25% (2,500 trees).

RESULTS

We sequenced 727 bp of the Y-chromosome SRY ORF which included the entire gene coding region. Our final data set included 44 unique sequences representing eight Caprinae genera (not including the outgroup S. tartarica). The SRY ORF contained 116 variable sites and 55 of these were parsimony informative. Two indels were detected, both of which occurred in C. sumatrensis: a three base insertion of GGT for at position 72 and a single base insertion of A for at position 679 which results in a frameshift and a protein that is 15 amino acids shorter than that
of the other genera examined. Whether these indels are characteristic of *C. sumatrensis* remains to be independently confirmed, as our sequence was determined from two reads of a single individual and no other SRY ORF sequence was not publicly available we were unable to confirm these indels with independent data. The 483 bp SRY PRO data set contained 21 unique sequences representing 10 genera for which 61 sites were polymorphic, and 33 were parsimony informative. The 1,110 bp ND5 data set contained 26 unique sequences representing 13 genera for which 471 sites were polymorphic, and 361 were parsimony informative. The 1,208 bp PRO/ORF data set contained 18 unique sequences representing seven genera for which 136 sites were polymorphic, and 46 were parsimony informative. The 2,318 bp ND5/PRO/ORF data set contained 27 unique sequences representing 10 genera for which 442 sites were polymorphic, and 324 were parsimony informative.

**Phylogenetic Analyses** – DNA polymorphism and statistics are summarized by data set in Table 12. Three cladograms of Caprinae genera based on different combinations of data are presented in Figures 7, 8 and 9. Support of specific nodes for each data set, as estimated with posterior probabilities (PP), maximum likelihood (BPML), and maximum parsimony (BPMP) are shown in Table 13.
Figure 8. Comparative cladograms of Caprinae genera based on two regions of the Y-chromosome gene SRY: the SRY promoter (PRO, left) and the SRY open reading frame (ORF, right). Node supports are given as posterior probabilities (PP, Bayesian, above branch) and bootstrap values (ML/MP, below branch). Nodes with support of less than 50 were collapsed.
Figure 9. Comparative cladograms of the subfamily Caprinae based on ND5 gene sequence (left) and combined SRY ORF/PRO gene sequence (right). Node supports are shown on branches with posterior probabilities (Bayesian) above the line and bootstrap values (ML/MP) below the line. Nodes with branch supports all below 50 were collapsed and are not shown.
Figure 10. Cladogram of Caprinae genera based from combined ND5/PRO/ORF sequence data. Node supports are shown on branches with posterior probabilities (Bayesian) above the line and bootstrap values (ML/MP) below the line. Branches with supports all below 50 were collapsed and are not shown.
Table 13. Support of individual nodes for each data set as estimated with posterior probabilities (PP), ML Bootstrap percentages (BPML) and MP bootstrap percentages (BPMP). Nodes not found or with less than 50% bootstrap support or 0.5 posterior probability are indicated with an asterisk (*). Nodes for which support could not be evaluated because not all genera were available are indicated with a dash (-).

<table>
<thead>
<tr>
<th>Species included in node</th>
<th>PRO</th>
<th>ORF</th>
<th>PRO / ORF</th>
<th>ND5</th>
<th>ND5 / PRO / ORF</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PP</td>
<td>BPML</td>
<td>BPMP</td>
<td>PP</td>
<td>BPML</td>
</tr>
<tr>
<td>Capra</td>
<td>1.00</td>
<td>90</td>
<td>86</td>
<td>1.00</td>
<td>98</td>
</tr>
<tr>
<td>Ovis</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>1.00</td>
<td>99</td>
</tr>
<tr>
<td>Ovis - New World</td>
<td>0.96</td>
<td>85</td>
<td>64</td>
<td>*</td>
<td>61</td>
</tr>
<tr>
<td>Ovis - Old World</td>
<td>*</td>
<td>82</td>
<td>*</td>
<td>1.00</td>
<td>99</td>
</tr>
<tr>
<td>Ovis / Oreamnos</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>0.99</td>
<td>92</td>
</tr>
<tr>
<td>Oreamnos / Budorcas</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Capra / Hemitragus / Psuedoi</td>
<td>1.00</td>
<td>73</td>
<td>97</td>
<td>1.00</td>
<td>99</td>
</tr>
<tr>
<td>Capra / Hemitragus</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>0.96</td>
<td>86</td>
</tr>
<tr>
<td>Pseudois</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Ovibos / Capricornis</td>
<td>1.00</td>
<td>90</td>
<td>82</td>
<td>1.00</td>
<td>97</td>
</tr>
<tr>
<td>Ovibos / Capricornis / Naemorhedus</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Capricornis / Naemorhedus</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Rupicapra</td>
<td>1.00</td>
<td>100</td>
<td>100</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Budorcas / Pseudois / Hemitragus / Capra / Oreamnos</td>
<td>1.00</td>
<td>85</td>
<td>79</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Budorcas / Pseudois / Hemitragus / Capra</td>
<td>0.92</td>
<td>59</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Budorcas / Capricornis / Ovibos</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>0.52</td>
<td>92</td>
</tr>
<tr>
<td>Budorcas / Capricornis / Ovibos / Capra / Hemitragus / Pseudois / Oreamnos</td>
<td>1.00</td>
<td>84</td>
<td>79</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Budorcas / Capricornis / Ovibos / Capra / Hemitragus / Pseudois</td>
<td>0.92</td>
<td>70</td>
<td>53</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Budorcas</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Budorcas / Capra</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Ammotragus / Arabitragus</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Pantholops</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
MP analyses consistently provided lower support and less resolution relative to both ML and Bayesian methods for individual nodes in each of the four trees. Bayesian posterior probabilities tended to be higher than either MP or ML bootstrapping at individual nodes. In general the Y-chromosome data provided better resolution, as judged by posterior probabilities and bootstrap values, than the mitochondrial data for the earliest divergences within Caprinae (Table 13). Within the Y-chromosome data, SRY PRO was more variable than SRY ORF and generally provided better resolution, though these data sets were individually less informative than the combined PRO/ORF data set (Fig. 9) and exhibit some topological incongruities, with several nodes showing low to no support across analyses. These incongruent or poorly supported nodes include the placement of Budorcas and Oreamnos as well as the organization of species within both Capra and Ovis, which likely reflects low variation among these relatively recent divergences within these clades (Fig. 8).

The combined SRY data provided better resolution at more basal nodes than did ND5, which was unable to resolve relationships among five (PP and ML) or seven (MP) of the major Caprinae clades, resulting in a large basal polytomy. However, SRY variation was insufficient to resolve relationships among species within more recently diverged groups, while ND5, in contrast, provided better resolution at these more distal nodes across data sets and tree estimation methods. Table 13 summarizes support for individual nodes across all three analyses for each of the five data sets.

Mitochondrial data strongly supported the placement of Pantholops as sister to all other genera of Caprinae as well as grouping of Ammotragus and Arabitragus (Fig. 9). These data also support the division of tahrs (originally the single genus Hemitragus) into the Himalayan tahr (genus Hemitragus) and the Arabian tahr (genus Arabitragus). Four of the five data sets show
high support for *Hemitragus* as sister to *Capra* while the fifth, PRO, lacked sufficient resolution and instead included it in a polytomy with *Capra* and *Pseudois*. All ND5 analyses group the Arabian tahr with the aoudad (*Ammotragus lervia*) but were unable to resolve placement of this clade within the Caprinae, instead including it in the large basal polytomy.

**DISCUSSION**

*Clades supported by both Y-chromosome and mitochondrial phylogenies* - Several nodes within our Caprinae phylogenetic trees are supported independently by all available data, including *Capra, Capra/Hemitragus/Pseudois, Ovibos/Capricornis*, and *Rupicapra*. Additional nodes that are supported by at least two of three markers or four of five data sets included *Ovis* (excluding PRO), *Ovis* New World clade (Excluding ORF), *Ovis* Old World clade, and *Capra/Hemitragus, Ovibos/Capricornis* (Table 13).

The monophyly of true sheep (genus *Ovis*) was strongly supported for four of the five data sets (Table 13) and three of these, ND5, PRO/ORF and ND5/PRO/ORF, also supported the separation of *Ovis* into two distinct clades: Old World species (*O. ammon, O. orientalis* and *O. vignei*) and New World species (*O. dalli, O. canadensis* and *O. nivicola*) (Figs. 8 and 9) while the fourth, ORF, also supported the old world clade (PP = 1.0, ML = 99, MP = 98) but found low (ML = 61) to no (MP and Bayes) support for the new world clade (Fig. 8). In contrast, PRO suggested paraphyly of *Ovis* with strong support for the grouping of new world sheep (PP = 1, ML = 94, MP = 98) with all other Caprinae genera separate from the other true sheep (Old World) species (Fig. 8). It is likely that low Y-chromosome variation is contributing to poor resolution within *Ovis* due to the relatively recent divergence and radiation of this genus (Ramey
1993, Bunch et al. 2006, Loehr et al. 2006, Rezaei et al. 2010, Meadows et al. 2011), and that this is responsible for the observed discordance between our topologies. Our data thus best supports the hypothesis that *Ovis* is a monophyletic group in accordance with the majority of our data and previous findings (Hassannin and Douzery 1999, Hernandez-Fernandez and Vrba 2005, Ropiquet and Hassanin 2005a, 2005b).

Relationships among species within each of the *Ovis* clades also differed among data sets. Both ND5 (PP = 1, ML = 99, MP = 100) and the combined ND5/PRO/ORF (PP = 1, ML = 100, MP = 100) data sets show high support for the grouping of *O. o. musimon* with *O. o. orientalis*. SRY PRO also shows moderate support for this association (PP = .62, ML = 68, MP < 50), and all of these data agree with previous findings that mouflon sheep are the wild progenitor of domestic sheep (Rezaei et al. 2010, Meadows et al. 2011). However, the two Y-chromosome data sets showed low (ORF/PRO; PP = .75) to no (ORF) support for this grouping.

The monophyly of goats and related species (genus *Capra*) was well supported across all analyses and data sets. Likewise, all analyses supported a *Capra/Pseudois/Hemitragus* clade except ND5 MP which excluded *Pseudois*. For four of the five data sets *Hemitragus* was sister to *Capra* while *Pseudois* was more distantly related to both genera. The fifth data set, PRO, was unable to resolve these relationships, showing a polytomy at this node. Resolution of species within *Capra* differed between data sets. ORF showed division of all four species with high support for the earliest divergence of the ibex (*C. ibex*), followed by the markhor (*C. falconeri*) with wild goat (*C. aegagrus*) and domestic goat (*C. hircus*) most recently diverged (Fig. 8). The PRO data also supported the early divergence of ibex but was unable to resolve the relationships among the other three species (Fig. 8). The combined PRO/ORF data set showed high support for all divisions within *Capra* in agreement with the ORF data set (Fig. 9). In contrast, our ND5
data found wild goat and markhor to be most closely related and were unable to further resolve the *Capra* relationships, showing a three-way polytomy at this node (Fig 8). The combined ND5/PRO/ORF data set gave high support to the early divergence of ibex, followed by divergence of markhor but showed low confidence and conflicting placements for wild goat and domestic goat among tree estimation methods (Fig. 10).

While several studies support the monophyly of *Capra* and its association with *Pseudois* and *Hemitragus* (Groves and Shields 1996, Hassanin et al. 1998, Shafer and Hall 2010), many others have yielded conflicting results (Pidancier et al. 2006, Ropiquet and Hassanin 2006, Hassanin et al. 2009, Bibi et al. 2012). Both Pidancier et al. (2006) and Ropiquet and Hassanin (2006) found strong discordance for relationships among *Capra* species between mtDNA and Y-chromosome markers which they attributed to lineage sorting of ancestral polymorphisms, introgressive hybridization of mtDNA haplotypes, and differing selection pressures acting on each marker type. Hassanin et al. (2009) found that evolutionary signals differed among genes across the mtDNA genome and that results for ND5 were particularly discordant with the other mitochondrial genes, placing *C. hircus* completely outside the “goat-like” clade (*Pseudois/Hemitragus/Capra*). They did not support the paraphyly of *Capra* however, instead attributing this discordance to either exogenous (via interspecies hybridization) or endogenous (via amplification of nuclear copies of mtDNA, or Numts) contamination. Additional analyses of all available *C. hircus* mtDNA data found many problems in the goat reference genome including detection of several Numt sequences, one of which covered most of ND5 and another which was in *Hemitragus* and *Pseudois*, suggesting a nuclear integration event occurred in the common ancestor to this clade. Bibi et al. (2012) used both mtDNA and morphological data to place an extinct *Capra* species (*C. wodaramoya*) and found several areas of conflict between
phylogenies. Both placement of the genus, as either sister to or polyphyletic with Hemitragus, and relationships among Capra species differed between datasets. For example, morphological data supports monophyly of ibex species (C. ibex, C. nubiana, C. pyrenaica) while mtDNA shows a paraphyletic grouping of these species. A common problematic area was placement of the basal relationships among Caprinae genera which they suggest is due to rapid adaptive radiation of these genera in the Miocene.

The monophyly of the Chamois (R. rupicapra and R. pyrenaica) was also supported by both ND5 and ND5/PRO/ORF data sets. The placement of this genus within Caprinae did vary among data sets, however, with our PRO data placing it sister to the Capra/Hemitragus/Pseudois clade (PP: 1, ML BP: 95, MP BP: 93) while ND5 was unable to resolve this placement and ND5/PRO/ORF analyses were incongruent in that Bayesian (PP: 1) and ML (BP: 100) analyses agree with the PRO analyses in placing Rupicapra sister to Capra/Hemitragus/Pseudois while MP (BP: 100) analyses group Rupicapra with Ammotragus all with high support. Perez et al. (2011) also found phylogenetic discordance between Y-chromosome and mtDNA markers. Their SRY data supported a weak association between Rupicapra, Capra and Ammotragus with all Rupicapra belonging to a single unique clade while mtDNA data formed three, well differentiated, clades within Rupicapra. They also find that the distance between species pair for mtDNA was about two to three times that of SRY and suggest that the phylogenetic discordance between mtDNA and Y-chromosome data is due to disparity in time of divergence of haplotypes for each and that all modern Rupicapra are descended from a single very young male lineage.

The Ovibos/Capricornis clade (which also included Naemorhedus where data were available) was well supported across all five data sets; placement of this clade within Caprinae differed among data sets, however. ORF lacked adequate resolution to place this clade,
including it in a basal polytomy, while both PRO and PRO/ORF data sets placed this clade with good statistical support as the second divergence within the larger Capra/Hemitragus/Pseudois/Budorcas/(Rupicapra/Ammotragus) group after the divergence of Oreamnos (Figs. 8 and 9). ND5 ML and Bayesian analyses both highly supported placement of this clade as the earliest divergence within the Caprinae while MP was unable to resolve this branch and instead includes it in a large basal polytomy (Fig. 9). ND5/PRO/ORF analyses are also incongruent in that there was low support for grouping with Budorcas (Bayesian), while both ML and MP were unable to resolve the placement of Ovibos/Capricornis and instead included it in a three-way polytomy within the larger clade including all genera except Ovis (Fig. 10).

**Clades with discordance between Y-chromosome and mitochondrial phylogenies** - We found conflicting results for the placement of the takin (genus Budorcas). Our ND5 data shows an association of Budorcas with Oreamnos (PP: .96, ML BP: 73, MP BP: 52), while our ND5/PRO/ORF data shows weak (PP: .52, ML BP: 54) to no (MP) association with the Capricornis/Ovibos clade, and both the ND5 and ORF data sets lack sufficient resolution to reliably place these clades within the larger Caprinae phylogeny. While our PRO/ORF analyses place Budorcas sister to the Capra/Hemitragus/Pseudois clade (PP: 1, ML BP: 100, MP BP: 98) in agreement with Hassanin et al., (2009), our PRO data place it sister to the Capra/Hemitragus/Pseudois/Rupicapra clade but with low (ML BP: 53) to no (PP, MP) support.

Placement of the mountain goat (genus Oreamnos) was inconsistent among data sets and analyses. Three of our data sets (PRO, PRO/ORF & ND5/PRO/ORF) place Oreamnos sister to a large clade including all Caprinae genera except Ovis and Ammotragus. However, these individual data sets differ in the relative relationship of Oreamnos to specific genera within this
large clade. The PRO and PRO/ORF data both show a closer association with the *Capricornis/Ovibos* clade, while the full ND5/PRO/ORF data set associates *Oreamnos* with a *Budorcas/Capricornis/Ovibos* clade. In contrast, the ORF data include *Oreamnos* in a clade with *Ovis*, while the mitochondrial data group *Oreamnos* with *Budorcas* as part of a large basal polytomy.

The *Oreamnos/Budorcas* clade was strongly supported for ND5 ML and Bayesian (but not MP) analyses and this relationship is in agreement with previous findings (Ramey 2000) based on half the ND5 data. This is not consistent, however, with the findings of Hassanin et al. (2009), which supports our ORF data in associating *Oreamnos* with *Ovis* based on analyses of multiple mitochondrial markers, though interestingly their ND5 data, like ours, did not support this association. Other studies have grouped *Oreamnos* with *Rupicapra* (Fernandez and Vrba 2005), with a *Capricornis/Ovibos/Neamorhedus* clade using cytochrome *b* mitochondrial data (Hassanin et al. 1998), or have found conflicting results among data sets or analyses (Groves and Shields 1996, Shafer and Hall 2010, Bibi et al. 2012). The genus *Oreamnos* has been labeled a “rogue taxon” because its placement within the Caprinae is sensitive to both the types of markers used and the analytical methods employed. A total evidence approach employed by Shafer and Hall (2010) found the most support for a relatively basal position for *Oreamnos* sister to all Caprinae genera except *Naemorhedus* and *Ovibos*. This is similar to findings by Ropiquet and Hassanin (2005), but a definitive consensus has yet to be reached. It is possible that placement of *Oreamnos* is hindered by the lack of variation or speciation within the genus. This could be due to their adaptation to extreme environments, slow growth rates, and long generation times, which may have decreased the evolutionary rate in mountain goats.
Comparison of taxonomic resolution for ND5 and SRY - Several evolutionary and population dynamic factors contribute to diversity among specific genetic markers including evolutionary rate, selection, inheritance pattern, ploidy, and recombination. Population dynamic factors include generation time, effective population size, sex-biased dispersal patterns, and mating structure. Both Y-chromosome and mitochondrial markers are haploid and uniparentally inherited, which predict an effective population (Ne) one-quarter that of diploid nuclear markers. However, differing mutation rates and species-specific population dynamics also contribute to nucleotide diversity. Most Caprinae genera exhibit dispersal patterns which would predict higher Y-chromosome diversity. However, this is potentially offset, to some extent, by highly polygynous mating structures where a small number of males contribute a disproportionately large amount of genetic variation to subsequent generations, ultimately decreasing the Y-chromosome Ne in Caprinae genera.

While available data are limited, both theoretical and empirical data suggest that Y-chromosome markers could provide a unique and informative phylogenetic perspective for understanding Caprinae taxonomy. For example, it has been shown that increased germ-cell divisions in spermatogenesis is responsible for higher male mutation rates in humans (Haldane 1947), and this has been subsequently supported in ungulates with data from five Caprinae genera where the observed mutation rate was three to four times higher in males than females (Lawson and Hewitt 2002). In contrast, observed Y-chromosome nucleotide diversity ($\pi_y = 0.90 \times 10^{-4}$) in Ovis was approximately 10% of expected ($\pi_y = 8.36 \times 10^{-4}$), even after correction for a lower male Ne and a higher male-to-female mutation rate. This Ovis Y-chromosome nucleotide diversity estimate corresponded to about 1.4% of the mitochondrial observed nucleotide diversity ($\pi_{mt} = 6.15 \pm 0.87 \times 10^{-3}$) (Meadows et al. 2004). This can be at least partly explained by
purifying selection acting to conserve the SRY coding sequence which is necessary for normal development of male anatomy. Thus, there is reason to expect that Caprinae Y-chromosome markers will be more conserved than mitochondrial markers, which would prove more useful for identifying basal divergences within this subfamily due to a lack of sequence saturation over long periods of evolutionary time (Hartl and Clark 1997, Barbosa and Carranza 2010, Perez et al. 2011). Our data supports this hypothesis in that the substitutions per site (K) for ND5 was seven (ORF) to ten (PRO) times higher than SRY (Table 12). Furthermore, all tree estimation methods find a large basal polytomy for ND5 (Fig. 9), while all tree estimation methods find better resolution for all three SRY data sets with strong support for basal divergences (Figs. 8 and 9).

One reason that has been proposed for the observed discordance in evolutionary patterns between Y-chromosome and mitochondrial markers in the Caprinae data is that all 13 mammalian mitochondrial genes are involved in the production of energy used to make ATP and maintain body temperature and are therefore especially important for adaptations necessary for living in colder environments at higher latitudes and altitudes. Because many genera of the Caprinae utilize these environments, it is hypothesized that selection acting on mitochondrial genes has contributed to the evolution of this group. For example, positive selection for 75 bp D-loop repeats as been documented and attributed to a response to the more severe oxidative stress and higher metabolic rates resulting from high-altitude life (Hassanin et al. 2009). In addition, it has been suggested that phylogenetic placement of problematic genera such as Capra and Oreamnos has been difficult because of past interspecific hybridization and subsequent selection for mitochondrial genotypes more fit for these environments, ultimately leading to
discordant evolutionary signals between nuclear and mitochondrial markers (Ropiquet and Hassanin 2006).

**Caprinae phylogenetics** - The taxonomic structure of the subfamily Caprinae has been problematic since Simpson’s original classification (1945) which recognized 11 genera divided into four tribes: 1) Caprini (goats, sheep, and related species), 2) Rupicaprini (chamois, goral, and serow), 3) Ovibovini (muskox and takin) and 4) Saigini (antelope and chiru). Subsequent molecular studies have found all four of these tribes to be either poly- or paraphyletic (Gentry 1992, Gatesy 1997, Hassanin et al. 1998, Ropiquet and Hassanin 2005, Hassanin et al. 2009, Shafer and Hall 2010, Arif et al. 2012) leading to the proposal of many additional taxonomies. This taxonomic discordance has been attributed to a number of factors, including the lack of any diagnostic synapomorphies uniting this group. Rather, the subfamily Caprinae includes genera with diverse adaptations which have contributed to rapid radiation throughout high altitude habitats across Africa, Asia, Europe, and North America. Additionally, individual genera vary both in times since divergence as well as diversity and species composition within clades. Additional factors, such as interspecific hybridization, variable evolutionary rates between markers, and differing life history and demographic variables, all contribute to conflicting results among the many phylogenetic analyses including genera within this subfamily. This discordance is exacerbated by a paucity of comprehensive comparative analyses incorporating diverse markers, because the overwhelming majority of available data is from mitochondrial markers. Thus continued development of additional nuclear markers, including more diverse Y-chromosome markers, should complement existing mitochondrial data and prove useful for increasing resolution of the evolutionary relationships within this subfamily.
CHAPTER IV

POPULATION GENETICS OF ROCKY MOUNTAIN BIGHORN SHEEP

Population genetic analyses have proved incredibly useful for evaluation of genetic variation, gene flow, metapopulation structure, inbreeding and ultimately informed management of populations. Comparative analyses of different markers can significantly improve the value of genetic data, however correct interpretation requires consideration of the differences inherent in the markers used. Additionally, interpretation of these data is often context sensitive in that individual populations or species differ in myriad ways. Quantification of population genetics for natural populations, specifically, is often hindered by a number of factors, including incomplete historical data, uneven sampling, unrealistic assumptions for parameter estimations, and complications due to species-specific demography and dynamics. Thus, interpretation of population genetic data can be complex and comprehensive evaluation of different genetic markers may help discern patterns but conflicts and discrepancies should also be acknowledged.

Bighorn sheep often exist in small, isolated populations due to specific habitat requirements and naturally isolated ranges that are increasingly fragmented due to anthropogenic factors. Thus, bighorn herds are often susceptible to inbreeding depression due to factors including disease and genetic drift, which can result in loss of allelic variation and decreased heterozygosity. Migration and gene flow among neighboring herds can ameliorate the effects of these factors by introducing new alleles that contribute to the maintenance of genetic variation in individual herds. However, bighorn herds are often subject to the pressures associated with habitat fragmentation, recent research has shown that both slope and ruggedness of terrain are
significant factors in bighorn habitat selection (Sappington et al. 2007) and that highways block
gene flow in bighorn (Epps et al. 2005), which can limit dispersal of sheep among adjacent
herds.

Five Rocky Mountain bighorn sheep (*Ovis canadensis*) populations were included in our
analyses based on their presence in and relative proximity to RMNP and assessment of potential
for gene flow within the larger metapopulation. Four of these herds (Continental Divide (CD),
Mummy (MM), Never Summer (NS) and St. Vrain (SV)) maintain some or all of their annual
range within the 415 square miles of RMNP while the fifth (Big Thompson (BT)) maintains an
annual range east of RMNP within Roosevelt National Forest. This dissertation research was
initiated due to concern regarding long-term persistence of the Mummy herd following a
pneumonia outbreak in the mid 1990s with a significant die-off and continued poor recruitment
thereafter. This could be due to a number of genetic or environmental causes, and we therefore
gathered data from both mitochondrial haplotypes and microsatellite genotypes to determine if
decreased genetic variation could explain the Mummy herd’s failure to rebound to historic
numbers.

We find no evidence of a severe reduction in genetic diversity or recent genetic
bottleneck for any of these five Rocky Mountain bighorn sheep herds. Observed
heterozygosities were comparable to (Forbes and Hogg 1999, Epps et al. 2010, Johnson et al
2011, Luikart et al. 2011) or higher than (Boyce et al. 1997) those reported for other native
bighorn populations and higher than those reported for restored bighorn populations (Whitaker et
al. 2004). Metapopulation analyses suggest that high heterozygosity is being maintained, to some
extent, by gene flow among these five herds. Estimates of population subdivision from
microsatellite data were low and, likewise, effective migration rates were reasonably high, and
no significant correlation was found between allelic distribution and geographic distance. In contrast, mitochondrial data show higher genetic isolation among herds and a significant negative correlation between haplotype distribution and geographic distance. The ranges of both effective migration rates and population substructure were larger for mitochondrial data than for microsatellite data, suggesting that ewes preferentially move between some herds more than between others.

The disparity between metapopulation structure as estimated from microsatellite and mitochondrial data could also be explained to some extent by differences in the evolution and inheritance patterns of these two markers. Genes located on the mitochondrial chromosome are vital to metabolic processes involved in aerobic respiration and are thus more conserved than microsatellite alleles which are neutral and rapidly mutate via replication slippage and are thus generally highly variable in populations (Ballard and Whitlock, 2004). Mitochondria are also haploid and maternally inherited while microsatellites are diploid and bparentally inherited. Thus, when the sex ratio is equal, mitochondrial markers have an effective population one quarter that of microsatellite markers (Birky et al. 1983). However, this disparity is offset, to some extent, by the highly polygynous mating structure of bighorn sheep in which most adult ewes reproduce each year while generally only the most dominant rams successfully mate (Geist 1966, Geist 1971, Coltman et al. 2005).

Interpretation of these metapopulation substructure data must be tempered, however, by the history of sheep movement both into and between study herds. These translocations were primarily composed of ewes (George et al. 2009) and this, together with the highly polygynous mating structure of bighorn sheep, suggests that the MM and BT herds should each have a higher proportion of private mitochondrial haplotypes and microsatellite alleles and higher pairwise }$F_{st}$}. 

72
estimates between each of these two herds and the other three (CD, NS and SV) herds, and the BT and MM herds should exhibit a higher proportion of shared genetic variation, primarily mitochondrial haplotypes, resulting in lower subdivision relative to the other herds. Our data do not match either of these predictions, however, and suggest that these transplantations are not appreciably altering interpretation of the metapopulation structure. This could be explained by low transplantation success which has been documented in bighorn (Gross et al. 2000, Singer et al. 2000) which may have resulted in few transplanted ewes surviving and reproducing in their transplanted herds.

Despite mitigating factors inherent in evaluation of genetic variation in natural populations our data clearly show that none of these five Rocky Mountain bighorn herds are suffering from a genetic bottleneck and that this is likely due to continued gene flow within this metapopulation. Our results also reiterate the utility of comparative mitochondrial and microsatellite analyses which have proved useful for discerning relationships both within and among many ungulate species (Sarno et al. 2004, Randi et al. 2005, Shafer et al. 2010) and highlight the advantages of evaluating population genetic parameters from complementary markers for other species (Brunner et al. 1998, Shaw et al. 1999).

PHYLOGENETICS OF CAPRINAE GENERA

Phylogenetics is a discipline that has greatly benefitted from relatively recent advances in the ability to quantify individual variation at the molecular level, augmenting or complementing previous analyses centered on morphological and behavioral data. This is especially true in
genera and species with low evolutionary rates or complicated evolutionary and demographic histories such that divergence is difficult to assess and multiple lines of evidence are required (Avise et al. 1992, Martin et al. 1992, Kuhner and Felsenstein 1994, Yang 1996, Wendel and Doyle 1998). Taxonomic organization of genera within the subfamily Caprinae is particularly difficult because no unique synapomorphies unite this diverse group and phylogenetic analyses to date have focused primarily on mitochondrial data. This has resulted in continued disagreement regarding resolution of Caprinae phylogenetics, and recent papers have repeatedly called for analyses which incorporate more diverse markers to order to aid elucidation of the relationships within this group and related ungulate genera (Zachos et al. 2003, Hernandez-Fernandez and Vrba 2005, Lorenzen et al. 2008).

To address the previous lack of comprehensive phylogenetic analyses of Caprinae, we provide comparative analyses of two different markers; the mitochondrial gene ND5 and two regions of the nuclear gene sex-determining region Y (SRY) for twelve of the thirteen extant Caprinae genera. Our data provide the first SRY open reading frame (ORF) sequence data for 15 species in nine genera and the first SRY promoter (PRO) sequence data for 13 species in 11 genera. With these data we compare the relative resolving power of the highly variable ND5 gene with the highly conserved SRY gene to determine if and when each marker is better at clarifying the evolutionary relationships among Caprinae genera. Our data highlight the complex evolutionary history of Caprinae genera and the importance of incorporating data from diverse markers with disparate evolutionary signals by highlighting inherent incongruities between them, resulting from differences including evolutionary rate, variation, inheritance pattern and effective population size. These analyses suggest that both ND5 and SRY have complementary phylogenetic value, in accordance with results from other species (Sasazaki et al. 2005).
2007, Schiffer et al. 2007, Yoon et al. 2008, Nussberger et al. 2013). Our more variable ND5 data were more useful for discerning recent divergences, which agrees with analyses in other species (Horai et al. 1995, Questiau et al. 1998, Tserenbataa et al. 2004, Palsboll et al. 2007, Smith et al. 2008). In contrast, our more conserved SRY data proved more useful for discerning past divergences, which has been shown in other species (Jobling et al. 2003, Geraldes et al. 2005, Rangel-Villalobos et al. 2009, Lippold et al. 2011), although the low variation in the SRY markers impeded our analyses of population genetics in bighorn sheep, in congruence with other findings from other sheep species (Meadows and Kijas 2009, Oner et al. 2011, Shafter et al. 2011).


Avise, J. C., and D. Ellis. 1986. Mitochondrial DNA and evolutionary genetics of higher animals [and discussion]. Philosophical Transactions of the Royal Society B: Biological Sciences 312: 325-342.


Buitkamp, J., F.-W. Schwaiger, and J.-T. Epplen. 1993. Ovis aries T cell receptor gene V-region, exons 1 (3' end) and 2 (5' end). NCBI.


Questiau et al. 1998


