Population Genetic Analyses of Rocky Mountain Bighorn Sheep (*Ovis canadensis*) Herds in and Around Estes Park, Colorado

Draft report submitted to:

Rocky Mountain National Park

Estes Park, Colorado

CATHERINE DRISCOLL, *Ph.D. Candidate, University of Colorado, Department of Ecology and Evolutionary Biology, N235 Ramaley, Campus Box 224, Boulder, CO 80309*

COREY HAZEKAMP, *Research Technician, University of Colorado, Department of Ecology and Evolutionary Biology, N235 Ramaley, Campus Box 224, Boulder, CO 80309*

JEFFRY B. MITTON, *Primary Investigator, University of Colorado, Department of Ecology and Evolutionary Biology, N235 Ramaley, Campus Box 224, Boulder, CO 80309*

EXECUTIVE SUMMARY

The primary objectives of this study included estimation of several population genetic parameters for the five herds of bighorn sheep (*Ovis canadensis*) in and around Rocky Mountain National Park (RMNP) in Estes Park, Colorado. These herds include all four RMNP herds: Continental Divide, Mummy, Never Summer and St. Vrain as well as the Big Thompson herd located east of the RMNP along Hwy 34 (Map 1). The impetus of this work was to determine if the struggling Mummy herd's low recruitment and failure to thrive could be attributed to inbreeding depression. We included four distinct genetic markers in these analyses: 1) genotypes at 13 microsatellite loci; 2) sequences of the Y chromosome gene SRY; 3) sequences of two mitochondrial DNA regions (D-Loop and COI) and 4) sequencing of two exons and two introns in the Major Histocompatibility Complex (MHC) gene DRB1. These markers collectively include both coding and non-coding regions, both haploid and diploid regions and both neutral and selective regions. Together the data collected from these markers provide a comprehensive estimation of genetic variation both within and among these five bighorn sheep herds.

Patterns of genetic differentiation were similar for microsatellite and mitochondrial markers and they indicated that no single herd is isolated from the others within this metapopulation. Specifically, there was no significant difference in microsatellite or mitochondrial genetic variation detected between the Mummy herd and any of the other four herds. Levels of variation at individual microsatellite loci were comparable among the five herds and no outliers were detected. Similarly, observed heterozygosity did not differ significantly from expected heterozygosity for any one herd. Based on overall and individual F_{is} estimates we did not find evidence of inbreeding depression in any herd, including the Mummy herd.

Genetic differentiation between the five herds, measured with F_{st}, was substantially higher when estimated with mtDNA markers than with microsatellites. Because mtDNA is inherited only from mothers, while microsatellites are inherited from both parents, the discrepancy between estimates of F_{st} suggests differential movement of males and females among herds. Higher levels of differentiation among herds measured with mtDNA indicate that females move among herds much less than males. The highest mtDNA differentiations are between the Big Thompson herd and each of the Park herds, indicating that ewe-mediated gene flow is higher within RMNP than either into or out of the Park. Limited movement by ewes is most convincingly demonstrated by the fact that the two most geographically separated herds, the Big Thompson and the Never Summer, have no mtDNA haplotypes in common. Additionally, the most common haplotype in the St. Vrain herd is completely absent in both the Never Summer and Mummy herds (and is very rare in the Continental Divide herd) further supporting more limited and asymmetrical migration of ewes then rams and consequently more limited and asymmetrical maternal then paternal gene flow. In contrast, the differentiation estimated with microsatellite genotypes is lower and more uniform among herds with no obvious difference between RMNP herds and the Big Thompson herd. Together, these data

suggest that: 1) all five herds have reasonable and comparable levels of genetic variation; 2) that there is low herd substructure within this metapopulation; 3) that this is most likely a result of geneflow due primarily to migration of rams among herds.

Both ORF and Promoter Y-chromosome sequences were identical for all bighorn sheep in the five herds examined here. Thus we find that there is insufficient variation within these markers for estimation of ram-mediated gene flow among herds. However, we expanded this data set to include all species within the *Ovis* genus in order to produce the most extensive systematic study of sheep to date. These data include the first ORF sequences for all six wild sheep species: bighorn (*O. canadensis*), snow (*O. nivicola*), dall's (*O. d. dalli* and *O. d. stonei*), argali (*O. ammon*), Urial (*O. vignei*) and mouflon (*O. orientalis*) as well as the first snow sheep promoter sequences. This work increases taxonomic resolution and will be useful for elucidating both temporal and geographic patterns of speciation within the *Ovis* genus.

Sequence analyses from two regions of the MHC DRB1 gene revealed that gene duplication has occurred within the *Ovis* genome. We identify from three to 20 potential alleles within a single individual suggesting between two and 10 total loci. Because sequence heterogeneity might be attributable to errors in DNA amplification, bacterial cell replication or sequencing. Therefore we conducted several additional experiments to verify that many of these alleles are, in fact, real and not artifacts generated during data collection. Although MHC variation has been identified in other organisms we are the first to document DRB1 gene duplication in bighorn sheep. These novel data will be useful for future MHC research and further study is warranted.

Map 1: Geographic ranges and population estimates for the Big Thompson (BT), Continental Divide (CD), Mummy (MM), Never Summer (NS), and St. Vrain (SV) herds.

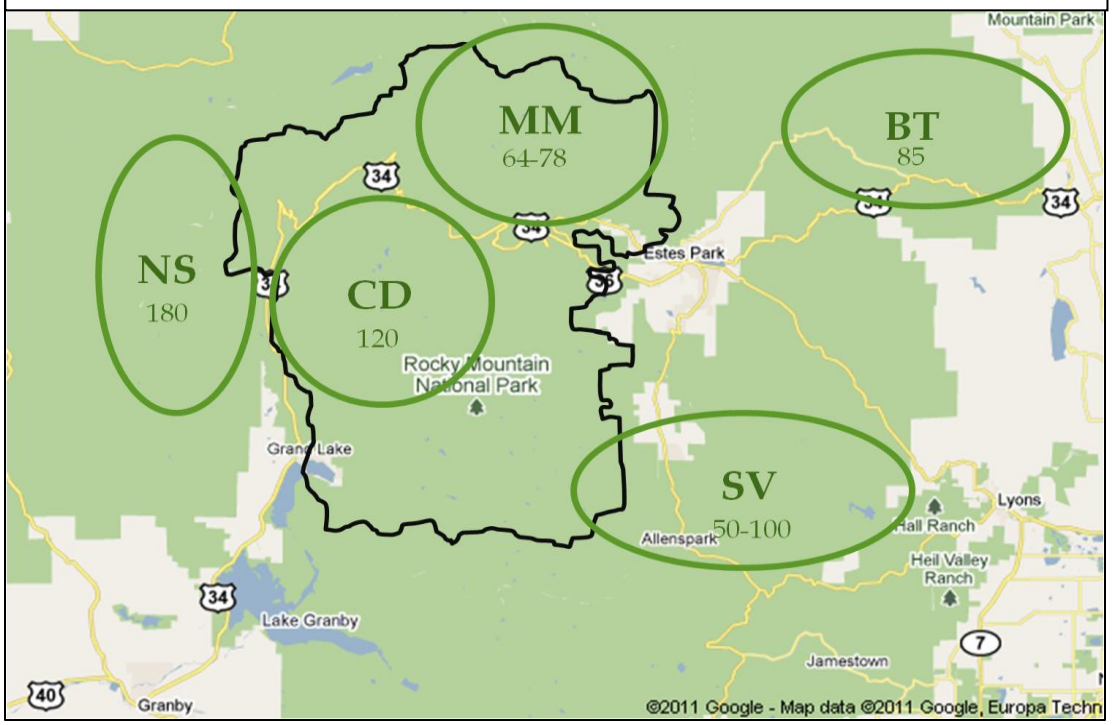


TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION AND BACKGROUND

I. *Rationale*

Rocky Mountain Bighorn Sheep (*Ovis canadensis*) are the emblematic animal of Rocky Mountain National Park (RMNP), appearing as a logo on park brochures and entrance stickers almost from its founding in 1915. Four bighorn herds maintain annual ranges that include land within Park boundaries, two on the east side of the Park (Mummy & St. Vrain) and two on the west side (Never Summer & Continental Divide). One additional herd, the Big Thompson, maintains a range nearby, but outside Park boundaries, along Hwy 34 between Estes Park and Ft. Collins. Because of the proximity of the Big Thompson herd to the four RMNP herds and the potential for gene flow, we determined that inclusion of the Big Thompson herd in this study was appropriate (Map 1).

While all four RMNP herds are potentially accessible to visitors, the Mummy herd is generally considered the most impacted by human interactions because access to a natural mineral lick at Sheep Lakes requires this herd to cross Fall River Road. As a result, the Mummy herd is often observed along this road on either side the of the Fall River Road Park entrance. Because of the Mummy herd's frequent interaction with Park visitors, there is concern that stress may be adversely affecting the Mummy herd. Anecdotal evidence indicates that a *pneumonia* outbreak caused a significant die-off of Mummy herd sheep in the 1990s. Additional evidence of declining populations prompted an intensive study of the Park herds' distributions and abundances in 2003 and 2004 (McClintock, 2007). This project used radio-collared ewes, ground surveys and aerial surveys to estimate the number of sheep in each herd. Results indicated there were approximately 60 sheep in the Mummy Range herd, 120 in the Continental Divide herd, and 180 in the Never Summer herd. The St. Vrain and Big Thompson herds were not included in the McClintock study. Estimations of population sizes for both the St. Vrain herd (50) and the Big Thompson herd (100) were provided by Colorado Parks and Wildlife. According to the McClintock study, if the reproductive and survivorship conditions recorded in 2003 and 2004 remain constant, the Mummy Range population is projected to die out (<1 individual in each age class) within 10 years, and the two West-side herds (Continental Divide and Never Summer) are projected to die out within 150 years. These projections lend some urgency to investigating this issue in multiple ways, including genetic analyses.

Describing the genetics of Park bighorn herds is an essential step that will provide direction for future monitoring or management actions. Accurate population estimates using radio collars and aerial surveys are expensive and risky for both humans and animals. As the cost of DNA identification of individuals has come down, the feasibility of estimating animal populations using non-invasive methods has become a reality. Methods for extracting high quality genetic material from fecal pellets of Rocky Mountain bighorn sheep have been successfully used in two laboratories, USDA Rocky Mountain Research Station in Missoula, MT (Luikart, 2006), and the lab of Dr. Jeff Mitton at CU Boulder (Driscoll, 2007). By providing a stock of DNA available for future analysis as well as the genotypes of individuals at 13 different loci (a DNA "fingerprint"), the present study allows the Park to move toward a monitoring method based primarily on DNA analyses.

II. *Objectives*

Data from four types of genetic markers were collected for this project in order to address four main questions:

1) Using data from 13 microsatellite loci: How much genetic variation exists in non-coding (evolutionarily neutral) DNA in these herds? This part of the project compared observed heterozygosity (H_o) vs. expected heterozygosity (H_e) as well as inbreeding coefficients (F_{is}), fixation indices (F_{st}) and migration (Nm) for all five herds. Low genetic variation may be a factor in the decline of these herds. Both overall and pairwise F_{st} values were used to estimate metapopulation substructure between herds in order to provide a comprehensive picture of genetic structure or pattern of differentiation among Park bighorn herds.

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? Measures of population connectivity via geneflow were estimated with both pairwise F_{st} values and effective migration rates (Nm).

3) Using sequence data from introns one and two and exons two and three of the MHC gene DRB1, known to be involved in the immune response system: Is there evidence that these herds are experiencing or have recently been subject to strong selection pressure that might ultimately contribute to the decline or extinction of any of these herds?

4) Using sequence data from two regions of the Y chromosome gene SRY (promoter (Pro) and complete gene coding region (ORF)): are males a source of gene flow between herds? If so, to what extent?

The answers to the above four questions provide estimates of relative genetic variation and rates of gene flow between all five herds using multiple markers. Analyses of several population genetic parameters such as: 1) degree of metapopulation substructure (F_{st}); 2) extent of inbreeding (F_{is}) within each herd and 3) migration rates between herds (Nm) allow Park managers to better determine whether genetic or environmental drivers are causing the Mummy herd decline and whether herd supplementation is a strategy that should be considered as a herd preservation tool.

III. *First Goal: Estimate Neutral Genetic Variation and Evidence for Inbreeding Depression*

Analyses of microsatellite data have been used to estimate neutral genetic variation in many species (Maudet *et al.* 2001; Maudet *et al.* 2004; Vial *et al*. 2003; Tapio *et al.* 2005; Kasapidis and Magoulas, 2008; Kongrit *et al*. 2008; Walser and Heckel, 2008) including sheep. From these prior studies, we selected 14 variable loci to characterize neutral genetic variation in bighorn sheep (Table 1). We genotyped each sheep in order to determine: A) total number of alleles at each locus for each herd, B) total haplotypes shared between herds, C) haplotypes unique to a single herd and D)

pairwise frequencies of shared haplotypes among all herds. These data were used to estimate degree of subdivision (F_{st}) between herds within the metapopulation. Microsatellite data were also used to determine if the Mummy (or any) herd was significantly less variable than the others. This information should assist future Park herd management decisions.

Historically, the extent of inbreeding in a given population has been estimated either through analyses of pedigrees in captive populations or by inference from heterozygosity in natural populations (Wright, 1931; Hartl and Clark, 2007). The inbreeding coefficient, F_{is}, can be calculated from microsatellite data and compared between herds in order to estimate relative levels of inbreeding. A population bottleneck is believed to result in the loss of low frequency (rare) alleles, but there are no strict standards established for determining whether "inbreeding" exists in a population based on high or low microsatellite variation within or across species. In this study, comparisons of F_{is} among herds in close proximity were used to measure the extent of inbreeding within each herd.

Of the fourteen microsatellite loci included in this study, thirteen were shown to be highly variable in bighorn sheep (AE16, MAF209, MAF33, MAF36, MAF 48, FCB266, SOMA, HH62, FCB304, TCRBV, ADC, KERA and OLA-DRB). Four of these are located within introns of genes (ADC, KERA, OLADRB and SOMA). Two of these are within genes involved in disease-related processes (ADC and OLADRB), while the remaining nine loci are unassociated with known gene regions. Variables such as F_{st} and F_{is} estimated with these non-coding diploid microsatellite markers were compared to the same variables estimated from mtDNA markers

IV. *Second Goal: Estimate gene flow from COI and DL mtDNA Sequence Analyses*

Several characteristics of mitochondrial DNA make it useful for population genetic analyses. The mtDNA genome is circular, has a higher mutation rate than genomic DNA, is maternally inherited (haploid) and thus does not undergo recombination. Because of these characteristics, mitochondrial DNA has been employed in studies of phylogeography and systematics for myriad species (Avise, 2000) including several within the genus *Ovis* (Luikart *et al.* 1996A; Hiendleder *et al.* 1998; Meadows *et al.* 2005; Loehr *et al.* 2006). Two mtDNA regions are included in this study. The gene cytochrome oxidase I (COI) encodes the last enzyme in the electron transport chain of cellular respiration and is therefore vital to the survival of the cell. Because of its role in energy production, COI variation is generally low. In contrast, the D-Loop region (DL) is the non-coding, highly variable region involved in replication and transcription of genes. Variation at these two regions spans the spectrum of variation of mtDNA markers and can therefore provide an interesting contrast for analysis of ewe-mediated gene flow.

V. *Third Goal: Compare Variation between DRB1 Exons Two and Three & Introns One and Two*

The MHC is a group of genes known to be involved in immune function. As a result, these genes are often under diversifying and/or balancing selection to increase genetic variation. MHC genes are generally considered the most variable genes in the genome and diversity of MHC alleles improves immune function through an increased ability to identify and fight infection and disease (Accola 1995; Lukas *et al*. 2004; Wigley, 2004). Thus, MHC genes have been extensively studied in many organisms (Hedrick, 1993; Markow *et al*. 1993; Gaudieri, 2000; Horton *et al.* 2008; Landis *et al.* 2008). MHC molecules can be divided into either class I or class II on the basis of both structure and function. Class II molecules, such as DRB1, present peptide fragments from foreign invaders to helper T-cells which ultimately triggers cell-mediated immune response. Because diversity in these proteins confers an increased ability to recognize foreign antigens and mount a successful defense, selection on MHC genes may act to increase heterozygosity in natural populations.

The *Ovis aries* DRB1 gene is 11.7kb long with five distinct exons (Ballingal *et al.* 2008A). Exon two encodes the antigen-binding region of the DRB1 protein and is thus directly involved in identification of foreign antigens (Dukkipati *et al*. 2006). Several studies have provided evidence supporting diversifying selection acting on this exon (Sayers *et al.* 2005, Worley *et al.* 2006; Hughes and Nei, 1988; Schaschl *et al.* 2006; Mona *et al*. 2008). Exon two is well characterized in many organisms, including sheep (Hedrick, 1993; Belov *et al.* 2007; Ballingall, 2008B; Horton *et al.* 2008). In contrast, exon three is of similar length but has no involvement in antigen presentation. It could therefore provide a useful comparison for relative signatures of selection. One way of calculating this is with a dN/dS ratio. This compares the number of sequence mutations that result in a change of the amino acid sequence (dN) to the number of mutations that have no impact on the amino acid sequence (dS); ratios of above 1 are often used as evidence for diversifying selection (Schaschl *et al*. 2006). Indeed, our previous analysis of the dN/dS ratios for 12 full-length mammalian DRB1 gene sequences show an average dN/dS ratio of 2.72 for the antigen-binding region of exon two (Driscoll and Mitton, unpublished data). In contrast, no other DRB1 exon showed dN/dS ratios departing from neutral expectations, lending support to the theory that diversifying selection is acting to increase variation for exon two specifically (Driscoll *et al.* unpublished data).

While MHC genes have been extensively studied in domestic sheep (*Ovis aries*), few studies have included wild sheep species. Of these, only two have analyzed DRB1 exon two in bighorn sheep (Gutierrez-Espeleta *et al.* 2001 & Worley, 2006) and no studies to date have analyzed exon three, intron one or intron two. We have designed two sets of novel primers that specifically amplify two regions of the bighorn sheep DRB1 genomic sequence. Primer set In1Ex2A amplifies an 826bp region including all 270bp of exon two and 556bp of adjacent intron one. Primer set In2Ex3A amplifies an 879bp region including all 281bp of exon three and 598bp of adjacent intron two. Together, these data will be valuable for a better understanding of the genomic variation within MHC gene DRB1.

VI. *Fourth Goal: Estimate Ram-Mediated Gene Flow Using the Y Chromosome Gene SRY*

The Y chromosome is a potentially valuable genetic marker because of its male-specific region (MSY) and unique male-to-male transmission pattern. The MSY accounts for 95% of the Y chromosome and contains a suite of genes involved in male sexual development. This region is non-homologous to the X chromosome and therefore cannot recombine during meiosis, making markers in this region potentially interesting for population genetic studies, especially those involving species with male-dominant mating systems and asymmetric migration patterns, and also for phylogenetic studies (Meadows, 2006; Heindleder *et al.* 2002).

The Sex-Determining Region Y (SRY), within the MSY, encodes a testes-determining factor involved in mammalian sex determination. SRY is well characterized in several species (Ali and Hasnain, 2002; Meadows *et al*. 2006; Waters *et al*. 2007; Marchal *et al*. 2008) making it one of the most useful Y chromosome markers to date (Hammer *et al*. 1998; Deng *et al*. 2004; Geraldes *et al*. 2005; Rosser *et al*. 2000; Luo *et al*. 2007). The 723bp intronless gene encodes a High Mobility Group (HMG-box) motif that is highly conserved among a wide variety of DNA-binding proteins, enabling SRY to act as a transcription factor activating a developmental cascade ultimately contributing to the male phenotype.

While SRY is one of the most extensively studied Y chromosome genes to date, data are still lacking for many genera, including the *Ovis* genus. The *Ovis* genus can be divided into three major taxonomic groups: Pachyceriforms (snow: *O. nivicola*, thinhorn: *O. dalli* and bighorn: *O. canadensis*); Argaliforms (argali: *O. ammon*); and Moufloniforms (mouflon: *O. orientalis* and urial: *O. vignei*). Ecological data add support to a division between 1) the stockier Pachyceriforms inhabiting precipitous mountainous habitats across Russia and North America and 2) the more agile Argaliforms/Moufloniforms inhabiting open, rolling grassland habitats across Europe, the Middle East and Asia (Geist, 1976).

Fossil evidence suggests that the *Ovis* genus evolved from a sheep-like ancestor between 2.42 and 3.19 MYA with the subsequent split of Pachyceriforms around 1.4 to 1.6 MYA (Rezaei *et al.* 2010; Bunch, 2006). Speciation within the Pachyceriforms began with the migration of snow sheep across the Bering Strait from Eurasia during the last ice age. Approximately 1 MYA, isolation in separate North American refugia during the Wisconsinan glaciations promoted the evolution of thinhorn sheep to the north and bighorn sheep to the south (Pielou, 1991; Loehr *et al.* 2006; Cowan, 2012). Divergence of Argaliforms and Moufloniforms occurred between 1-1.7 MYA with their spread across Eurasia. The most recent split of Moufloniforms into mouflon sheep and urial sheep occurred less than 1 MYA (Bunch, 2006, but see Meadows *et al.* 2011; Rezaei *et al.* 2010). Further taxonomic resolution within these subgenera remains difficult, however, and many competing classifications have been proposed based on a diverse set of criteria: morphological (pelage color, body type , horn size and shape), ecological (habitat type, geographic distribution) and genetic

(DNA polymorphism, chromosome number, ability to hybridize) (Tsalkin, 1951 reviewed in Reading *et al.* 2001; Haltenorth 1963 reviewed Shakleton, 1991; Valdez and Krausman 1982; Valdez and Batten, 1982; Geist, 1991; Shakleton 1997; Reale and Festa-Bianchet, 2000; Tserenbataa *et al.* 2004; Wilson and Reeder, 2005; Bunch *et al*. 2006). More recent genetic data support a clear monophyletic structure for species within both the Pachyceriforms and the Argaliforms; however phylogenetic structure of the Moufloniform clade remains unclear. Comprehensive genetic studies including samples from all *Ovis* species would likely resolve these issues and yield molecular markers needed for protection and management of wild populations, yet only two such studies have been published to date (Bunch *et al.* 2006; Rezaei *et al.* 2010).

While the domestic sheep (*Ovis aries*) genome has been well-characterized, data for markers on the Y-chromosome remains scarce for wild sheep species. This is especially true for bighorn sheep; only two partial (520bp) SRY promoter sequences from a single Canadian population have been published to date (Meadows, 2004; Meadows, 2006). For this study we sequenced two regions of the SRY gene: a 520bp section of the promoter region located 1,218 bases upstream of the Open Reading Frame (ORF) and a 954bp fragment including the complete 723bp ORF (Payen and Cotinot, 1994). These data were collected in order to isolate and estimate gene flow among herds due to ram migration and to provide a contrast to the gene flow estimates obtained from mtDNA sequence analyses.

MATERIALS AND METHODS

I. *Field Sample Collection and DNA Extraction*

Non-invasive field collection of bighorn sheep DNA was achieved via fecal pellet extraction. This was done in order to limit stress and risk to wild populations as well as decrease costs associated with previous invasive collection methods. Most samples were obtained in back country areas of RMNP on an opportunistic basis. Animals were located with spotting scopes and observed from afar in order to minimize trauma to sheep. After the sheep left the area, fresh fecal pellets, identified as bright green and moist, were collected. Older samples were identified as dark brown and dry and were avoided. Observation of sheep usually allowed sex and general age class identification and field notes were recorded at each site. Individual samples generally consisted of 15-20 pellets collected from a single distinct pile. These were subsequently stored in paper bags and allowed to dry immediately and thoroughly to avoid mold contamination. Epithelial cells were then scraped from the outmost layer of pellets with razor blades until 120-180mg were obtained. DNA extraction was performed with the QIAamp DNA Stool Minikit™ (Qiagen Corp, 2009) incorporating modifications following Wehausen *et al.* 2004. Selection of this procedure was based on our initial evaluation of alternate storage and extraction protocols which indicated that the above method produced more consistently high quality DNA than previous methods involving ethanol storage or whole-pellet extraction.

II. *Microsatellite Genotyping*

Data Collection

Amplification of microsatellite loci was achieved using one of the following kits: AmpliTaq Gold® PCR Kit (Applied Biosystems, 2007), Type-it Microsatellite PCR Kit (Qiagen Corp, 2009) or Multiplex PCR Kit (Qiagen Corp, 2009). 25µl PCR reactions were carried out in either an Eppendorf Mastercycler gradient thermocycler (Eppendorf AG, 2005) or a PCT-100 Peltier Thermal Cycler (Bio-Rad Laboratories Inc. 2005) in 96 well polycarbonate plates with silicone sealing mats (Corning Corp.). General PCR cycling steps were: initial denature at 93^{o} for 7:30’ followed by 30 cycles of: 95^{o} for 30”, 55-63^{o} for 40”, 72^{o} for 30” and a final extension at 72^{o} for 2’. Annealing temperatures were optimized for each microsatellite locus in order to maximize primer binding specificity and obtain clear, strong peaks. Whenever possible, PCR reactions were multiplexed such that multiple primer pairs with compatible annealing temperatures were combined into a single reaction. This allowed amplification of multiple microsatellite loci simultaneously which increased efficiency and decreased cost. All forward primers were labeled with one of three fluorescent dyes (Hex, Fam or Tet) for detection of amplified products.

Each sample was genotyped at 13 previously identified variable microsatellite loci (Table 1). Amplifications and scoring of peaks were both independently replicated in two to five (average four) separate reactions in order to decrease errors due to either allelic drop-out or poor sample quality. When replicate amplifications gave inconsistent results the sample was considered unreliable and removed from further analyses. All successful microsatellite amplifications were denatured with Hi-Di Formamide (1.5µl/sample, Applied Biosystems, 2003) and run in 36cm acrylamide gels on an ABI Prism 377 automated sequencer (Applied Biosystems, 2000) with Blue Dextran loading solution (0.4µl/sample, Promega Corporation, 2012). Microsatellite peak sizes were standardized across gel lanes using TAMRA size standards (350 or 500, Applied Biosystems). After run completion, gel lanes were tracked using GeneScan (Applied Biosystems, 2000), and microsatellite genotypes were manually scored to ensure accuracy and consistency. When two DNA samples had identical genotypes across all 13 loci, we removed the duplicates from all data sets in order to avoid incorrect population parameter estimations due to redundant sampling.

Table 1: Microsatellite loci general information and source references

Locus	Size Range (bp)	No. of Alleles	Annealing Temp (^{o}C)	Fluorescent Tag	Multiplex Group – PCR, Gel	Literature Reference	Genbank Accession
AE16	82-106	11	62	HEX	A, 1	Penty *et al.* 1993	L11047
ADC	85-113	8	62	FAM	A, 1	Crawford *et al.* 1995	N/A
FCB266	88-102	6	62	FAM	B, 2	Buchanan and Crawford, 1993	L01535
MAF36	88-104	8	56	TET	C, 3	Swarbrick *et al.* 1991	M80519
SOMA**	94-118	7	59	TET	N/A, 4	Lucy *et al.* 1998	U15731
MAF209	107-119	6	62	TET	B, 2	Buchanan and Crawford, 1992a	M80358
HH62	108-122	9	56	HEX	N/A, 4	Ede *et al.* 1994	N/A
MAF33	123-131	6	56	HEX	C, 3	Buchanan and Crawford, 1992b	M77200
MAF48	123-133	7	62	FAM	A, 3	Buchanan *et al*. 1991	M62645
FCB304	134-150	9	62	FAM	B, 2	Buchanan and Crawford, 1993	L01535
TCRBV62**	167-175	4	62	FAM	A, 3	Buitkamp et al. 1993	L18957
KERA**	175-179	3	62	HEX	B, 2	Tasheva *et al.* 1998	AF036962
OLADRBps**	278-302	8	62	HEX	A, 3	Beraldi *et al*. 2006	AH003856
MMP9*	184-196	N/A	N/A	TET	B, 4	Maddox, 2001	AJ133770

*Removed from final analyses due to inconsistent amplification and poor peak resolution

**Loci known to be within gene regions

Data Analyses

We used three programs to estimate population genetic parameters from microsatellite data and to perform all statistical tests. The program Genepop (Raymond and Rousset, 1995; Rousset, 2008) estimated overall and pairwise metapopulation substructure between herds (F_{st}), overall and per locus inbreeding coefficients within each herd (F_{is}) and overall and pairwise effective migration rates between herds (Nm).

F_{st} is a standardized measure of allelic frequency differentiation between two herds or among several herds, and it varies from 0.0, indicating no difference between groups, to 1.0, indicating no sharing of alleles between groups. F_{is} is a measure of the departure of genotypic frequencies from expected distributions under random mating. This value varies from -1.0, indicating an extreme excess of heterozygotes, to +1.0, indicating a complete absence of heterozygotes (Wright, 1931). Thus, inbreeding is indicated by large positive values of F_{is}.

We also calculated a multilocus estimate of the effective number of migrants (Nm), where N is the effective population size and m is the migrations rate. This estimate was based on the relationship between Nm and the frequency of private alleles observed (those alleles occurring only in a single herd) (Barton and Slatkin, 1986).

The program STRUCTURE (Pritchard *et al.* 2000) provided qualitative and graphical characterization of metapopulation subdivision. This was accomplished by first estimating the "optimal" number of total herds within the metapopulation as determine solely based on genotypic data. We then created "Q plots" to show the program's relative confidence in assignment of each individual sheep to a herd based solely on the genotype of that individual. Thus, we provide a qualitative comparison of how well the genetic data mirror the geographical divisions between herds.

The program BOTTLENECK (Cornuet and Luikart, 1997) was used to identify recent ($2N_e$-$4N_e$ generations) and severe reductions in effective population size (N_e) as a reliable indicator of genetic bottlenecks. This procedure compares observed heterozygosity (H_e, calculated from allele frequencies) to heterozygosity at equilibrium (H_{eq}, calculated from allele number). Because allele number (Na) is reduced faster than observed heterozygosity immediately after a bottleneck event, excess heterozygosity ($H_e > H_{eq}$, distinct from an excess of heterozygotes $H_o > H_e$) indicates that a recent and severe genetic bottleneck has occurred. Tests were run under three mutation models: The Infinite Allele Model (IAM), the Stepwise Mutation Model (SMM) and the Two Phase Model (TPM) (80-95% SMM and 5-20% IAM as recommended by the authors) with 5,000 replications per run. The Two Phase Model is generally considered the most accurate for microsatellite analyses because it combines the Stepwise Mutation Model which predicts mutations as single repeat increases or decreases (i.e. 175 to 173 or 177) with the Infinite Allele Model which predicts mutations as changes of any size and in either direction (i.e. 175 to 171 or 193). Previous research shows that the majority of microsatellite variation is created via replication slippage where a single repeat is either added or removed during normal cell division, thus, following recommendations by the program authors, we included 80-95% SMM and 5-20% IAM in our TPM to reflect the most accurate method for estimating microsatellite evolution.

BOTTLENECK was also used to graph allele frequency distributions as a qualitative "mode-shift" indicator of bottlenecked populations identified by a disproportionate lack of rare alleles. Because

rare alleles are continually created via mutation or introduced via gene flow, they usually comprise the largest group of allelic variation in a normal, healthy herd. However, because these alleles occur in low frequency, they are the first to be lost from a population during a bottleneck. This can be identified graphically by a truncation of the rarest allele classes. In contrast, the allele frequency distribution of a non-bottlenecked population has an L-shaped graphical distribution such that the rarest allele classes are the most frequent.

III. *Mitochondrial DNA Sequencing*

Data Collection

Cytochrome Oxidase I (COI) and D-Loop control region (DL) amplifications were performed with the AmpliTaq Gold® PCR Kit. Primers COIF (5' GCAGAGTTTGAAGCTGCT 3') and COIR (5' AGCTGACGTGAAGTAAGC 3') were used to amplify a 1053bp region of COI. Thermocycling included an initial denature at 94° for 7:30' followed by 35 cycles of: 94° for 60", 60° for 70", 72° for 90" and a final extension at 72° for 5'. Primers P1F (5'-CAACACCCAAAGCTGAAGTTC-3') and P4R (5'-CTAGGCATTTTCAGTGCCTTGC-3') were used to amplify a 961 bp (±75bp indel) fragment of the DL region. The amplification profile included an initial denature at 94° for 7:30' followed by 30 cycles of: 94° for 60", 62° for 70", 72° for 70" and a final extension at 72° for 5'.

Data Analyses

All mtDNA sequences were aligned and analyzed in Sequencher™ versions 4.2 and 4.6 (Gene Codes Corp.). COI and DL sequence sets were each analyzed separately and then concatenated into a single haplotype per individual for subsequent analyses. The program DnaSP (Librado and Rozas, 2009) was used to perform all calculations, including estimates of F_{st}, F_{is}, Nm, K, A, Hd.

IV. *SRY Gene Sequencing*

Data Collection

Both the SRY Open Reading Frame (ORF) and Promoter (Pro) amplifications were conducted in 25ul reactions using HotStart-IT Taq DNA Polymerase (Affymetrix Inc. Santa Clara, CA) with an initial denature at 94° for 5' followed by 40 cycles of: 94° for 60", 61° for 60" (Pro) or 52° for 60" (ORF), 72° for 60" with a final extension at 72° for 5'. From published *Ovis aries* SRY gene sequence we designed *Ovis*-specific primers SRY3A-F (5'-AACAATTGCATGTAGCTCCAGA-3') and SRY3A-R (5'-CAATGACAGTAGAAACTTCAAAGG-3') to amplify a 481bp segment of the SRY promoter region located 1238 bases upstream from the SRY ORF start codon. The complete

723bp coding region for SRY was amplified with previously published primers (Payen and Cotinot, 1993). All sequencing was conducted by Functional Biosciences (Madison, WI).

Data Analyses

All SRY sequences were aligned in Sequencher (Versions 4.2 and 4.6, Gene Codes Corp; Ann Arbor, MI). Phylogenetic analyses were performed on three separate data sets: A) ORF sequences; B) Pro sequences; C) concatenated ORF and Pro sequences together. Rooted and unrooted trees were estimated for each data set both with and without a designated outgroup (*Capra aegagrus*; wild goat). Each data set was analyzed using four different methods: Maximum Parsimony, Neighbor-Joining, Maximum Likelihood and Bayesian search algorithms. All sequences generated for this project were submitted to NCBI: SRY ORF accessions JN992654-JN992678 (Table 11) and SRY Pro accessions JN00247-JN00253 & JN992679-JN992695 (Table 12).

PAUP 4.0 (Swofford, 2000) was used to estimate trees using Maximum Parsimony, Neighbor-Joining and Maximum Likelihood methods. We estimated unweighted Maximum Parsimony trees (Hasegawa, 1985) using heuristic searches with indels designated as a 5th state change. We estimated Maximum Likelihood trees using the GTR + I model of DNA substitution with TBR branch swapping and bootstrapping (1,000 replicates). jModelTest 0.1.1 (Posada and Crandall, 1998; Posada, 2008) was used to find the optimal DNA substitution model for all Maximum Likelihood analyses. Bayesian analyses were performed in MrBayes version 3.1 (Huelsenbeck and Ronquist. 2001) using a codon model for the ORF (bases 1-723) and a simple nucleotide model for the Promoter (bases 724-1204). All resulting phylograms were analyzed and edited in FigTree 1.3.1 (Rambaut, 2006).

IV. ***MHC DRB1 Gene Sequencing***

Data Collection

For this project we initially used previously published primers (268F/275R; Ballingall *et al.* 2008A) for preliminary data collection and then designed two novel sets of primers based on these data that amplify two distinct regions of the DRB1 gene (Figure 16). Primers In1Ex2A-F (5'-TGGAGGGGACATGTCAGTTT-3') and In1Ex2A-R (5'- CACACACTGCTCCACACTGG) amplify 831bp of DRB1 which includes all 270bp of exon two as well as 516bp of the adjacent intron one. Primers In2Ex3A-F (5'-TTTCCCCCAATTTATCCACA-3') and In2Ex3A-R (5'-GGTGGGTGAGGGACTTATGA-3') amplify all 279bp of exon three as well as 563bp of the adjacent intron two. All amplifications were conducted in 25µl reactions using HotStart-IT Taq DNA Polymerase (Affymetrix Inc. Santa Clara, CA) with an initial denature at 94^{o} for 5' followed by 30 cycles of: 94^{o} for 30", 55-62^{o} for 30" and 72^{o} for 60" with a final extension at 72^{o} for 10'. Successful PCR was confirmed by agarose gel electrophoresis. These amplifications were then cloned using the TOPO TA Cloning Kit for Sequencing with One Shot MAX Efficiency DH5α-T1 *Escherichia coli* chemically competent cells (Invitrogen Corp. 2004). Bacterial colonies were grown

at 37^{o}c for 12-16 hours on LB/Agar with the antibiotic Kanamycin as a selective agent. Isolated colonies were picked and grown in an LB/Kanamycin solution at 37^{o}c for 10-12 hours with continuous agitation. Plasmids were isolated from bacterial cultures with QIAprep Spin Miniprep Kits (Qiagen corp. 2006). Sequencing of all purified plasmid clones was performed by Functional Biosciences (Madison, WI).

Our initial goals included cloning and sequencing a minimum of 20% of the individuals from each herd. This design required sequencing of six distinct clones from each individual in order to confirm homozygosity or heterozygosity at each locus. This also reduced the chance for misidentification of heterozygotes as homozygotes due to errors in PCR, cloning or sequencing.

Data Analyses

All MHC sequences were aligned and analyzed in Sequencher^{TM} (versions 4.2 and 4.6 Gene Codes Corporation; Ann Arbor, MI). Chromatograms were also manually checked for errors and all single nucleotide polymorphisms were individually confirmed.

RESULTS

I. *Field Sample Collection of DNA from Fecal Pellets*

A total of 36 sample sets were collected from locations throughout RMNP and the surrounding area, including Indian Peaks Wilderness Area, Roosevelt National Forest, Comanche Peak Wilderness, and between Estes Park and Ft. Collins along highway 7. From the 365 total samples collected, 224 ultimately provided high quality DNA. A final sample set of 193 was used for subsequent analyses after removal of all duplicate samples that were identified through microsatellite genotyping. All sample sizes collected per herd are summarized by marker type in Table 2.

Table 2: Summary of all data sample sets by marker type

Herd	Microsat	mtDNA	SRY - Pro	SRY - ORF	MHC
Big Thompson (BT)	15	20	4	0	2
Continental Divide (CD)	48	70	30	1	2
Mummy (MM)	28	30	14	1	0
Never Summer (NS)	29	41	37	0	0
St. Vrain (SV)	17	29	17	2	2
Total	137	190	102	4	6

II. *Microsatellite Genotype Analyses*

Microsatellite data were collected for 14 variable loci (Table 1); however one locus, MMP9, was ultimately excluded due to poor amplification and unreliable genotyping. Thus all analyses and parameter estimations presented here are derived from the remaining 13 loci. 137 samples were included in this data set. Approximately 22,000 data points were generated for use in the following analyses.

First, we found no significant deviation between observed heterozygosity (Given as H_e by BOTTLENECK but more commonly denoted H_o) and heterozygosity expected at equilibrium (Given as H_{eq} by BOTTLENECK but more commonly denoted H_e) (Tables 3A-E). This is shown most effectively by the "Prob" calculation in the TPM section of these tables. A TPM "Prob" value of ≤ 0.05 indicates that deviations between H_e and H_{eq} are statistically significant. While a few "Prob" values do fall below the 0.05 threshold at a few individual loci (BT: FCB266 (0.02); CD: MAF36 (0.04), FCB266 (0.05); MM: SOMA (0.05); SV: MAF209 (0.04), MAF33 (0.02), KERA (0.03)), the vast majority of loci give values well above this threshold across all herds. Thus, we found no strong or consistent signal of significant deviation from expected heterozygosity for any individual microsatellite locus or for any one of these five herds.

Second, we found that allele frequency class bar graphs show normal, L-shaped distributions of microsatellite variation across all five herds (Figure 1). Thus, the rarest alleles do comprise the majority of observed allelic variation within each herd. This is a hallmark of healthy, genetically variable populations and therefore no evidence of a disproportionate loss of rare alleles (frequency classes 1, 2, or 3) is seen here.

Finally, scatter plots of individual microsatellite deviations from expected heterozygosity (H_e - H_{eq}) show no evidence of heterozygosity excess for any herd (Figure 1). Excess heterozygosity would be indicated by a disproportionately high distribution of data points above the line Y= 0. However, we find that the distributions of deviations across this line are relatively even for each herd. Although the BT distribution is slightly skewed toward excess heterozygosity, this is likely due to this herd's smaller sample size relative to the others. Overall we find that observed microsatellite variation supports that these are healthy herds with no indication of a recent or severe genetic bottleneck.

Tables 3A-E: BOTTLENECK summary statistics by microsatellite locus for each herd under three different mutation models: Two-Phase Model (TPM), Infinite Allele Model (IAM) and Stepwise Mutation Model (SMM).

3A. Big Thompson Herd				TPM				IAM				SMM			
Locus	N	Na	He	Heq	S.D.	DH/sd	Prob	Heq	S.D.	DH/sd	Prob	Heq	S.D.	DH/sd	Prob
FCB266	30	3	0.68	0.47	0.14	1.45	0.02	0.41	0.16	1.71	0.01	0.51	0.13	1.35	0.03
MAF209	30	5	0.68	0.68	0.09	0.00	0.41	0.62	0.12	0.48	0.40	0.71	0.08	-0.46	0.25
FCB304	30	3	0.62	0.47	0.15	1.01	0.16	0.42	0.16	1.26	0.10	0.51	0.13	0.86	0.22
KERA*	30	2	0.52	0.28	0.16	1.44	0.07	0.24	0.16	1.74	0.05	0.28	0.16	1.45	0.06
AE16	30	7	0.81	0.78	0.06	0.39	0.43	0.74	0.09	0.81	0.22	0.81	0.04	-0.06	0.39
ADC	30	5	0.69	0.68	0.09	0.15	0.47	0.62	0.12	0.60	0.33	0.72	0.07	-0.29	0.31
MAF48	30	4	0.69	0.59	0.11	0.82	0.23	0.54	0.14	1.09	0.13	0.63	0.09	0.58	0.34
TCRBV*	29	3	0.52	0.48	0.14	0.32	0.48	0.41	0.16	0.65	0.35	0.51	0.13	0.09	0.44
OLA*	30	7	0.75	0.78	0.06	-0.54	0.23	0.74	0.08	0.14	0.47	0.81	0.04	-1.33	0.10
MAF33	30	5	0.74	0.68	0.09	0.70	0.26	0.62	0.12	1.02	0.13	0.71	0.08	0.39	0.43
MAF36	30	5	0.65	0.68	0.09	-0.26	0.33	0.61	0.12	0.32	0.45	0.72	0.07	-0.88	0.16
SOMA*	30	5	0.6	0.68	0.09	-0.88	0.16	0.62	0.12	-0.20	0.34	0.71	0.07	-1.55	0.07
HH62	28	5	0.67	0.69	0.09	-0.21	0.34	0.62	0.12	0.37	0.44	0.72	0.07	-0.77	0.19
Means	29.77	4.5	0.66	0.61	0.11	0.34	0.28	0.55	0.13	0.77	0.26	0.64	0.09	-0.05	0.23

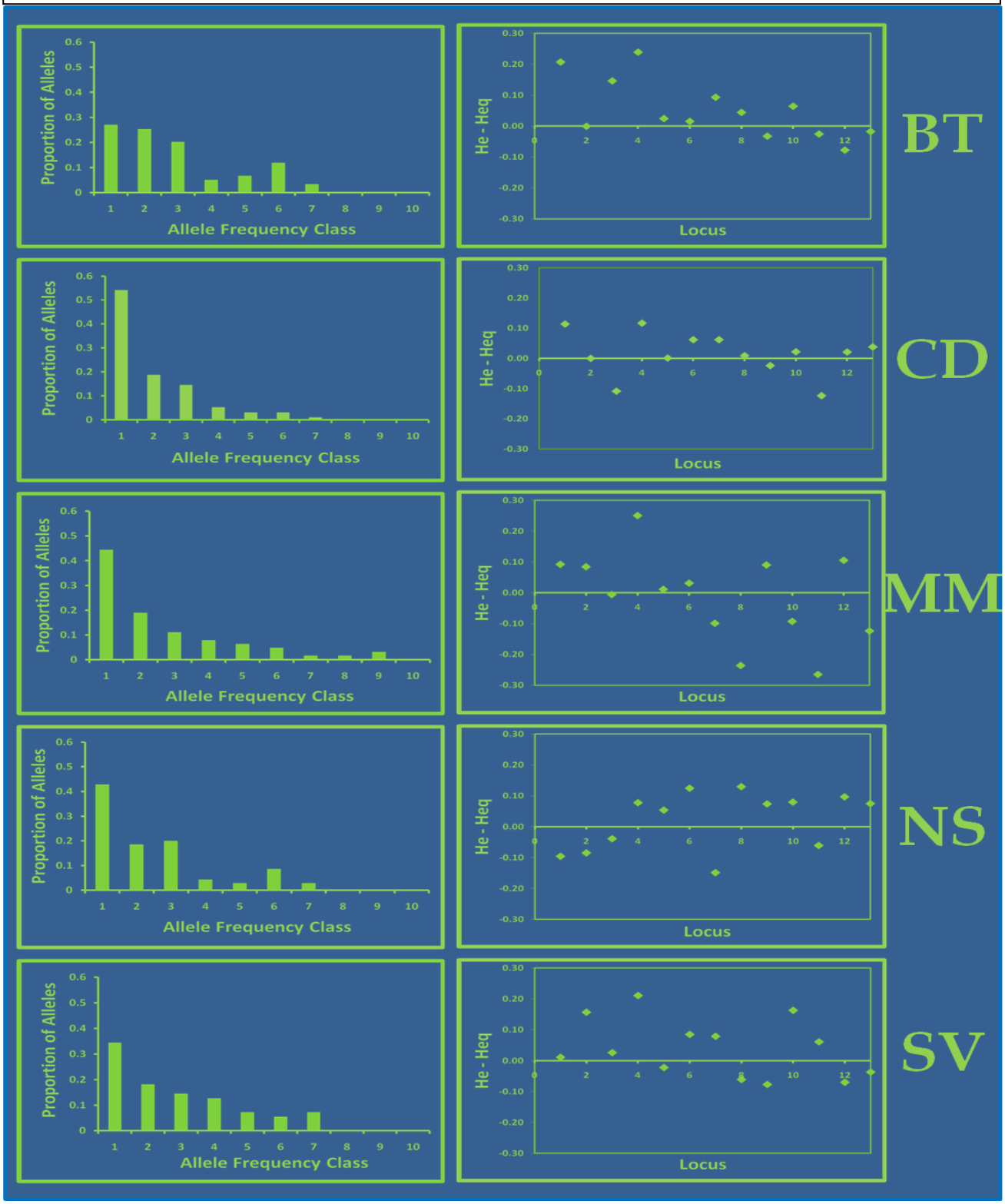
3B. Continental Divide Herd				TPM				IAM				SMM			
Locus	N	Na	He	Heq	S.D.	DH/sd	Prob	Heq	S.D.	DH/sd	Prob	Heq	S.D.	DH/sd	Prob
FCB266	91	6	0.79	0.68	0.09	1.21	0.05	0.58	0.14	1.44	0.02	0.73	0.07	0.94	0.15
MAF209	91	8	0.76	0.76	0.07	-0.04	0.40	0.68	0.11	0.66	0.29	0.80	0.04	-1.06	0.13
FCB304	95	9	0.68	0.78	0.06	-1.77	0.06	0.71	0.10	-0.28	0.31	0.83	0.04	-4.06	0.00
KERA*	95	3	0.53	0.40	0.16	0.76	0.24	0.33	0.18	1.10	0.16	0.45	0.14	0.53	0.34
AE16	90	14	0.87	0.87	0.03	0.01	0.42	0.82	0.06	0.81	0.18	0.90	0.02	-1.06	0.12
ADC	90	7	0.78	0.72	0.08	0.75	0.22	0.63	0.13	1.18	0.07	0.77	0.05	0.19	0.52
MAF48	96	5	0.68	0.62	0.11	0.53	0.36	0.51	0.16	1.00	0.15	0.67	0.08	0.09	0.45
TCRBV*	96	4	0.54	0.52	0.14	0.10	0.46	0.43	0.18	0.61	0.34	0.58	0.11	-0.41	0.28
OLA*	83	8	0.74	0.76	0.07	-0.32	0.31	0.68	0.11	0.47	0.38	0.80	0.04	-1.52	0.08
MAF33	95	7	0.74	0.72	0.08	0.23	0.49	0.63	0.13	0.85	0.20	0.77	0.05	-0.59	0.22
MAF36	92	9	0.66	0.79	0.06	-2.16	0.04	0.71	0.11	-0.45	0.25	0.83	0.04	-4.39	0.00
SOMA*	90	9	0.81	0.79	0.06	0.36	0.44	0.71	0.10	0.92	0.15	0.83	0.04	-0.47	0.26
HH62	91	7	0.76	0.72	0.08	0.48	0.38	0.64	0.13	0.94	0.15	0.77	0.06	-0.24	0.33
Means	29.77	4.5	0.66	0.70	0.08	0.01	0.30	0.62	0.13	0.71	0.20	0.75	0.06	-0.93	0.22

3C. Mummy Herd				TPM				IAM				SMM			
Locus	N	Na	He	Heq	S.D.	DH/sd	Prob	Heq	S.D.	DH/sd	Prob	Heq	S.D.	DH/sd	Prob
FCB266	48	4	0.66	0.56	0.13	0.74	0.26	0.49	0.16	1.07	0.15	0.61	0.11	0.45	0.40
MAF209	56	5	0.72	0.64	0.11	0.81	0.21	0.55	0.15	1.15	0.09	0.69	0.08	0.42	0.40
FCB304	56	3	0.42	0.44	0.16	-0.09	0.39	0.37	0.17	0.31	0.45	0.47	0.14	-0.36	0.30
KERA*	56	2	0.49	0.24	0.17	1.53	0.09	0.21	0.17	1.70	0.09	0.25	0.17	1.48	0.10
AE16	49	6	0.72	0.71	0.09	0.15	0.47	0.64	0.13	0.66	0.31	0.75	0.06	-0.46	0.27
ADC	48	5	0.68	0.65	0.10	0.32	0.47	0.57	0.14	0.76	0.27	0.69	0.08	-0.18	0.34
MAF48	55	7	0.65	0.74	0.07	-1.36	0.09	0.68	0.11	-0.29	0.31	0.79	0.05	-2.71	0.02
TCRBV*	55	4	0.32	0.55	0.13	-1.77	0.07	0.48	0.17	-0.94	0.20	0.61	0.10	-2.86	0.02
OLA*	55	5	0.73	0.64	0.11	0.87	0.19	0.56	0.15	1.17	0.09	0.69	0.08	0.54	0.35
MAF33	56	6	0.61	0.70	0.09	-1.10	0.13	0.62	0.13	-0.12	0.37	0.74	0.06	-2.12	0.04
MAF36	56	4	0.29	0.55	0.13	-1.99	0.06	0.48	0.16	-1.16	0.18	0.60	0.11	-2.85	0.02
SOMA*	56	6	0.81	0.70	0.09	1.23	0.05	0.62	0.13	1.44	0.01	0.74	0.06	1.01	0.11
HH62	56	6	0.57	0.70	0.09	-1.40	0.10	0.63	0.13	-0.42	0.27	0.74	0.06	-2.78	0.02
Means	29.77	4.5	0.66	0.60	0.11	-0.16	0.20	0.53	0.14	0.41	0.21	0.64	0.09	-0.80	0.18

3D. Never Summer Herd				TPM				IAM				SMM			
Locus	N	Na	He	Heq	S.D.	DH/sd	Prob	Heq	S.D.	DH/sd	Prob	Heq	S.D.	DH/sd	Prob
FCB266	54	6	0.61	0.70	0.09	-1.11	0.13	0.63	0.12	-0.21	0.33	0.74	0.06	-2.15	0.04
MAF209	54	6	0.61	0.70	0.09	-1.02	0.13	0.62	0.13	-0.08	0.39	0.75	0.06	-2.14	0.04
FCB304	58	6	0.66	0.70	0.09	-0.44	0.26	0.62	0.13	0.30	0.48	0.74	0.06	-1.27	0.10
KERA*	58	3	0.51	0.43	0.16	0.47	0.41	0.36	0.17	0.86	0.26	0.48	0.14	0.23	0.49
AE16	58	8	0.83	0.78	0.06	0.90	0.16	0.72	0.10	1.18	0.05	0.81	0.04	0.46	0.38
ADC	56	3	0.55	0.44	0.15	0.78	0.24	0.36	0.18	1.09	0.16	0.48	0.13	0.56	0.33
MAF48	58	5	0.49	0.63	0.11	-1.35	0.11	0.56	0.15	-0.46	0.28	0.68	0.08	-2.48	0.03
TCRBV*	57	3	0.56	0.43	0.16	0.84	0.21	0.37	0.17	1.14	0.13	0.48	0.13	0.64	0.29
OLA*	53	7	0.82	0.75	0.07	1.01	0.12	0.68	0.11	1.29	0.04	0.79	0.05	0.70	0.26
MAF33	56	4	0.63	0.56	0.13	0.61	0.33	0.47	0.16	0.99	0.17	0.60	0.10	0.30	0.47
MAF36	58	7	0.68	0.74	0.07	-0.87	0.15	0.67	0.12	0.04	0.42	0.78	0.05	-2.16	0.04
SOMA*	54	6	0.80	0.70	0.08	1.17	0.07	0.63	0.13	1.35	0.03	0.74	0.06	0.88	0.17
HH62	58	6	0.77	0.70	0.09	0.83	0.19	0.62	0.13	1.19	0.07	0.74	0.06	0.47	0.37
Means	29.77	4.5	0.66	0.64	0.10	0.14	0.19	0.56	0.14	0.67	0.22	0.68	0.08	-0.46	0.23

3E. St. Vrain Herd				TPM				IAM				SMM			
Locus	N	Na	He	Heq	S.D.	DH/sd	Prob	Heq	S.D.	DH/sd	Prob	Heq	S.D.	DH/sd	Prob
FCB266	29	4	0.61	0.59	0.12	0.14	0.48	0.54	0.14	0.52	0.37	0.64	0.10	-0.26	0.33
MAF209	34	4	0.74	0.58	0.12	1.30	0.04	0.53	0.14	1.51	0.02	0.63	0.10	1.19	0.06
FCB304	34	4	0.61	0.59	0.12	0.20	0.50	0.52	0.15	0.64	0.31	0.63	0.10	-0.14	0.36
KERA*	34	3	0.68	0.47	0.15	1.42	0.03	0.40	0.17	1.66	0.01	0.51	0.13	1.35	0.03
AE16	34	7	0.75	0.77	0.07	-0.33	0.29	0.73	0.09	0.29	0.47	0.80	0.05	-1.11	0.13
ADC	32	3	0.56	0.46	0.15	0.65	0.30	0.41	0.16	0.89	0.22	0.50	0.13	0.44	0.38
MAF48	32	4	0.67	0.59	0.11	0.71	0.27	0.52	0.15	1.01	0.16	0.63	0.10	0.46	0.40
TCRBV*	32	4	0.53	0.59	0.12	-0.53	0.24	0.52	0.14	0.04	0.44	0.63	0.10	-1.03	0.14
OLA*	34	4	0.51	0.59	0.12	-0.67	0.22	0.52	0.15	-0.08	0.40	0.63	0.10	-1.19	0.12
MAF33	33	4	0.75	0.59	0.12	1.38	0.02	0.53	0.15	1.55	0.01	0.63	0.10	1.25	0.04
MAF36	32	3	0.53	0.46	0.15	0.43	0.42	0.41	0.16	0.73	0.30	0.50	0.13	0.19	0.49
SOMA*	32	7	0.71	0.78	0.06	-1.09	0.13	0.73	0.09	-0.27	0.30	0.81	0.04	-2.19	0.03
HH62	34	4	0.55	0.58	0.12	-0.25	0.31	0.52	0.15	0.23	0.49	0.63	0.10	-0.78	0.18
Means	29.77	4.5	0.66	0.59	0.12	0.26	0.25	0.53	0.14	0.67	0.27	0.63	0.10	-0.14	0.21

Figure 1: 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).



Genepop analyses of microsatellite variation yielded low estimates of population differentiation both among all herds (F_{st} = 0.047) as well as between pairs of herds (F_{st} = 0.02 – 0.09) (Table 5, Figure 5). These analyses reveal little genetic differentiation or herd subdivision within this metapopulation. The number of sheep moving among these five herds was estimated to be Nm = 8.82 per generation with a range of Nm = 1.04 to 4.28 for pairwise comparisons of herds (Table 7, Figure 7). This value uses the frequency of private alleles (those alleles present in just a single herd) to estimate the number of sheep moving between herds each generation (Barton and Slatkin, 1986). These data show a general isolation-by distance pattern of movement such that gene flow estimates are higher between closer herds and generally decrease with increasing geographic distance.

Values of F_{is}, the inbreeding coefficient, varied from F_{is} = -0.01 to 0.09. Values for all individual herds were close to F_{is} = 0.0 as expected under random mating and thus these data yield no evidence of inbreeding (Table 6, Figure 6).

Additional Nm and F_{is} calculations were also estimated from two subsets of the microsatellite data: A) the 9 "neutral" loci only (Nm=8.82; F_{is} = 0.01 to 0.16) and B) the 4 other loci located within gene regions (Nm=8.97; F_{is} = -0.10 to 0.12) (Tables 6 & 7). We performed these additional calculations to determine whether selection might skew Nm and F_{is} estimates obtained from microsatellites located near or in gene regions. However, estimates from all three data sets are in close agreement, indicating no skew in estimates associated with potential selection and lending support to overall conclusions.

Table 4: Genepop F_{st} summary statistics: Pairwise F_{st} values per microsatellite locus (A-M) and average pairwise F_{st} values across all 13 loci (N). Herd abbreviations: Big Thompson (BT), Continental Divide (CD), Mummy (MM), Never Summer (NS), St. Vrain (SV).

A

FCB 266	BT	CD	MM	NS
CD	0.0308			
MM	0.0193	0.0568		
NS	0.0668	0.0599	0.0475	
SV	0.0723	0.1049	-0.008	0.1163

B

MAF209	BT	CD	MM	NS
CD	0.0404			
MM	0.0247	0.0026		
NS	0.2437	0.1113	0.1323	
SV	0.0216	0.009	0.0196	0.0983

C

FCB304	BT	CD	MM	NS
CD	-0.0009			
MM	0.0303	0.0472		
NS	0.0826	0.0643	0.225	
SV	0.0169	0.0296	0.1586	0.0029

D

KERA	BT	CD	MM	NS
CD	-0.0261			
MM	-0.0204	0.0043		
NS	0.0091	-0.0043	0.0509	
SV	0.0519	0.056	0.0894	0.0572

E

AE16	BT	CD	MM	NS
CD	0.0143			
MM	0.0381	0.0397		
NS	0.0793	0.011	0.0628	
SV	0.0233	0.0281	-0.01	0.0618

F

ADC	BT	CD	MM	NS
CD	0.0204			
MM	0.0323	0.0178		
NS	0.0388	0.0759	0.1216	
SV	0.0206	0.0637	0.0206	0.0723

G

MAF48	BT	CD	MM	NS
CD	0.0169			
MM	0.0031	0.005		
NS	0.0865	0.0313	0.034	
SV	-0.0084	0.0439	0.0617	0.1519

H

OLA	BT	CD	MM	NS
CD	-0.0157			
MM	-0.0083	-0.002		
NS	0.0647	0.0479	0.063	
SV	0.0869	0.0738	0.1382	0.1633

I

HH62	BT	CD	MM	NS
CD	0.0645			
MM	-0.0032	0.0659		
NS	0.1281	0.0102	0.1332	
SV	-0.0051	0.0731	-0.012	0.1405

J

MAF33	BT	CD	MM	NS
CD	-0.0014			
MM	0.0192	0.0349		
NS	0.0593	0.0127	0.047	
SV	-0.0134	0.0038	0.0625	0.0744

K

TCRBV	BT	CD	MM	NS
CD	0.008			
MM	0.0272	0.0429		
NS	0.0391	0.0082	0.1216	
SV	0.0098	-0.021	0.0245	0.0093

L

MAF36	BT	CD	MM	NS
CD	0.0002			
MM	0.1135	0.0686		
NS	0.0003	-0.0083	0.1119	
SV	-0.0203	0.0087	0.0451	0.0279

M

SOMA	BT	CD	MM	NS
CD	0.1119			
MM	0.094	0.0133		
NS	0.158	0.0091	0.0087	
SV	0.0087	0.0393	0.0544	0.0892

N

HERD	BT	CD	MM	NS
CD	0.0231			
MM	0.0288	0.0301		
NS	0.0876	0.0344	0.0899	
SV	0.021	0.0409	0.0524	0.0848

Table 5: Genepop estimates of: F_{is}, F_{st} and F_{it} per locus across all herds (Table A) and F_{is} per locus using both the Cockerham & Weir method (F_{is} W&C) and the Robertson & Hill (F_{is} R&H) methods for each herd (B-D).

A

All Herds			
Locus	Fis	Fst	Fit
FCB266	0.151	0.059	0.201
MAF209	-0.036	0.069	0.035
FCB304	0.116	0.070	0.178
KERA*	0.070	0.023	0.188
AE16	0.037	0.034	0.070
ADC	-0.017	0.053	0.037
MAF48	0.044	0.036	0.078
TCRBV*	0.080	0.028	0.106
OLA*	-0.083	0.053	-0.025
MAF33	0.020	0.028	0.047
MAF36	0.142	0.033	0.171
SOMA*	-0.034	0.049	0.016
HH62	0.015	0.065	0.079
Average	0.041	0.047	0.079

B

Mummy		
Locus	F_{is}(W&C)	F_{is}(R&H)
FCB266	-0.065	-0.039
MAF209	0.111	0.034
FCB304	-0.187	-0.143
KERA	-0.089	-0.090
AE16	0.117	0.024
ADC	-0.217	-0.160
MAF48	0.152	0.082
TCRBV	0.432	0.183
OLA	-0.281	-0.187
MAF33	-0.243	-0.099
MAF36	-0.110	-0.047
SOMA	-0.066	-0.069
HH62	0.257	0.059
Average	-0.015	

C

Continental Divide		
Locus	F_{is}(W&C)	F_{is}(R&H)
FCB266	0.187	0.134
MAF209	-0.148	-0.045
FCB304	0.162	0.175
KERA	0.413	0.217
AE16	-0.094	-0.042
ADC	0.189	0.249
MAF48	-0.009	0.125
TCRBV	0.166	0.067
OLA	0.022	-0.014
MAF33	0.150	0.102
MAF36	0.172	0.120
SOMA	-0.044	-0.012
HH62	0.011	0.080
Average	0.090	

D

Never Summer		
Locus	F_{is}(W&C)	F_{is}(R&H)
FCB266	0.192	0.030
MAF209	-0.027	-0.057
FCB304	0.059	0.004
KERA	-0.228	-0.119
AE16	0.132	0.086
ADC	-0.297	-0.251
MAF48	0.008	0.014
TCRBV	-0.197	-0.155
OLA	-0.025	-0.030
MAF33	0.099	0.217
MAF36	0.139	0.154
SOMA	0.177	0.129
HH62	-0.167	-0.098
Average	-0.010	

E

St. Vrain		
Locus	F_{is}(W&C)	F_{is}(R&H)
FCB266	0.222	0.125
MAF209	-0.197	-0.198
FCB304	-0.057	-0.057
KERA	0.398	0.394
AE16	-0.017	-0.020
ADC	-0.128	-0.082
MAF48	0.122	0.029
TCRBV	0.030	-0.047
OLA	0.082	-0.044
MAF33	0.093	0.073
MAF36	0.173	0.274
SOMA	-0.091	-0.063
HH62	-0.070	-0.094
Average	0.043	

F

Big Thompson		
Locus	F_{is}(W&C)	F_{is}(R&H)
FCB266	0.222	0.243
MAF209	0.220	0.079
FCB304	0.577	0.574
KERA	0.229	0.239
AE16	0.179	0.327
ADC	0.041	0.019
MAF48	0.031	0.082
TCRBV	0.062	0.014
OLA	-0.252	-0.110
MAF33	-0.174	-0.109
MAF36	0.292	0.190
SOMA	-0.349	-0.137
HH62	0.148	0.125
Average	0.90	

Table 6: Genepop estimates of inbreeding coefficient (F_{is}) by herd for three microsatellite data subsets: A) 4 "non-neutral" loci: OLA, KERA, SOMA & TCRBV; B) 9 "neutral" loci: ADC, AE16, FCB266, FCB304, HH62, MAF33, MAF36, MAF48 & MAF209; and C) All 13 Loci together.

Herd	A) F_{is} (4 Loci)	B) F_{is} (9 Loci)	C) F_{is} (All Loci)
Mummy	**-0.068**	**0.019**	**-0.01**
Continental Divide	**0.117**	**0.067**	**0.09**
Never Summer	**-0.047**	**0.019**	**-0.01**
St. Vrain	**0.118**	**0.005**	**0.04**
Big Thompson	**-0.104**	**0.164**	**0.09**

Table 7: Nm rates for: A) All 13 loci; B) 9 "neutral" loci and C) 4 "non-neutral" loci.

A		MM	CD	NS	SV
	CD	3.410			
	NS	2.163	4.284		
	SV	1.151	1.732	1.037	
	BT	2.658	1.914	1.109	1.666

Overall Nm: 8.825

B		MM	CD	NS	SV
	CD	3.181			
	NS	2.230	4.413		
	SV	2.286	2.156	1.281	
	BT	2.286	1.819	1.051	0.964

Overall Nm: 8.815

C		MM	CD	NS	SV
	CD	5.630			
	NS	1.849	3.588		
	SV	6.062	0.785	0.714	
	BT	6.062	2.548	1.628	2.605

Overall Nm: 8.965

Structure analyses show low overall metapopulation substructure which supports conclusions from F_{st} estimates. Maximum likelihood (ML) values for predicted herd number (K) were not significantly different between K = 4, 5 or 6. This broad ML curve (Figure 2B) indicates that Structure was unable to clearly assign a definitive herd number to this metapopulation due to low genetic differentiation between populations.

This trend is further supported by Q plots (Figure 2A) showing the proportion of each individual's genotype that Structure attributed to each potential population. We also include additional Q plots representing various subsets of these data: A) K= five herds with only the nine neutral loci included; B) K = 4 herds where data for the BT herd was excluded; C) K = 4 herds using only the nine neutral loci and excluding the BT herd; and D) A sample Q plot showing strong population substructure for comparison purposes (Figure 3).

We also present these data as pie charts which show the proportion of sheep that were assigned by Structure to one of the K = 5 herds based on these microsatellite data (Figure 4). Perfect correlation between an individual sheep's actual herd identity and Structure's correct assignment to that herd would be represented by five different solid, single color pies analogous to $F_{st} = 1$. No correlation between an individual sheep's actual herd identity and Structure's correct assignment to that herd would be represented by five identical, equally proportioned pies analogous to $F_{st} = 0$. Our results are much more indicative of the second scenario of an overall F_{st} much closer to 0 than 1, further supporting low population substructure within this metapopulation.

Figure 2A: STRUCTURE Q plots for K = 4, 5, and 6 potential herds

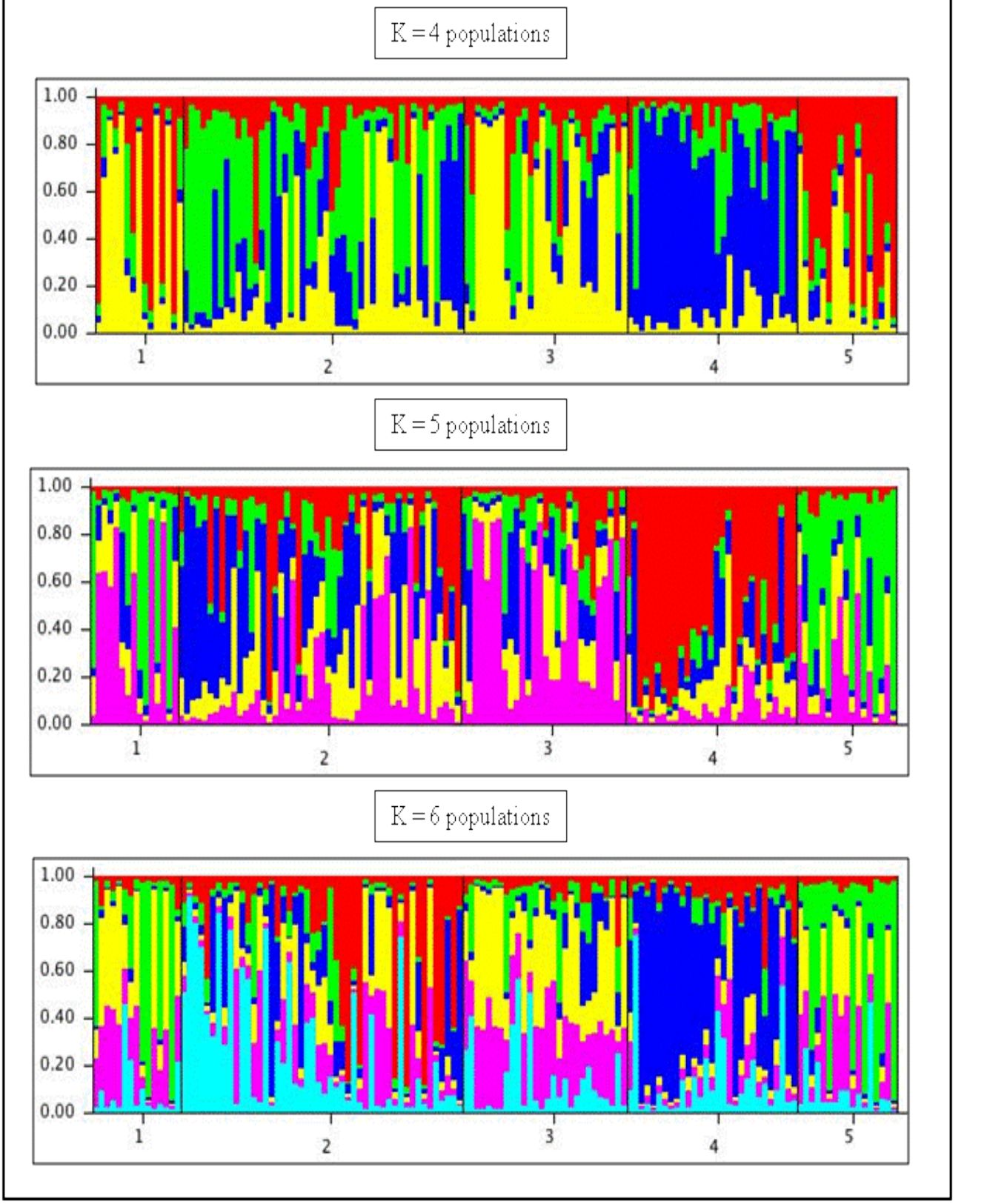


Figure 2B: Structure analysis of Maximum Likelihood values for each possible number of herds (K) = 1 – 8.

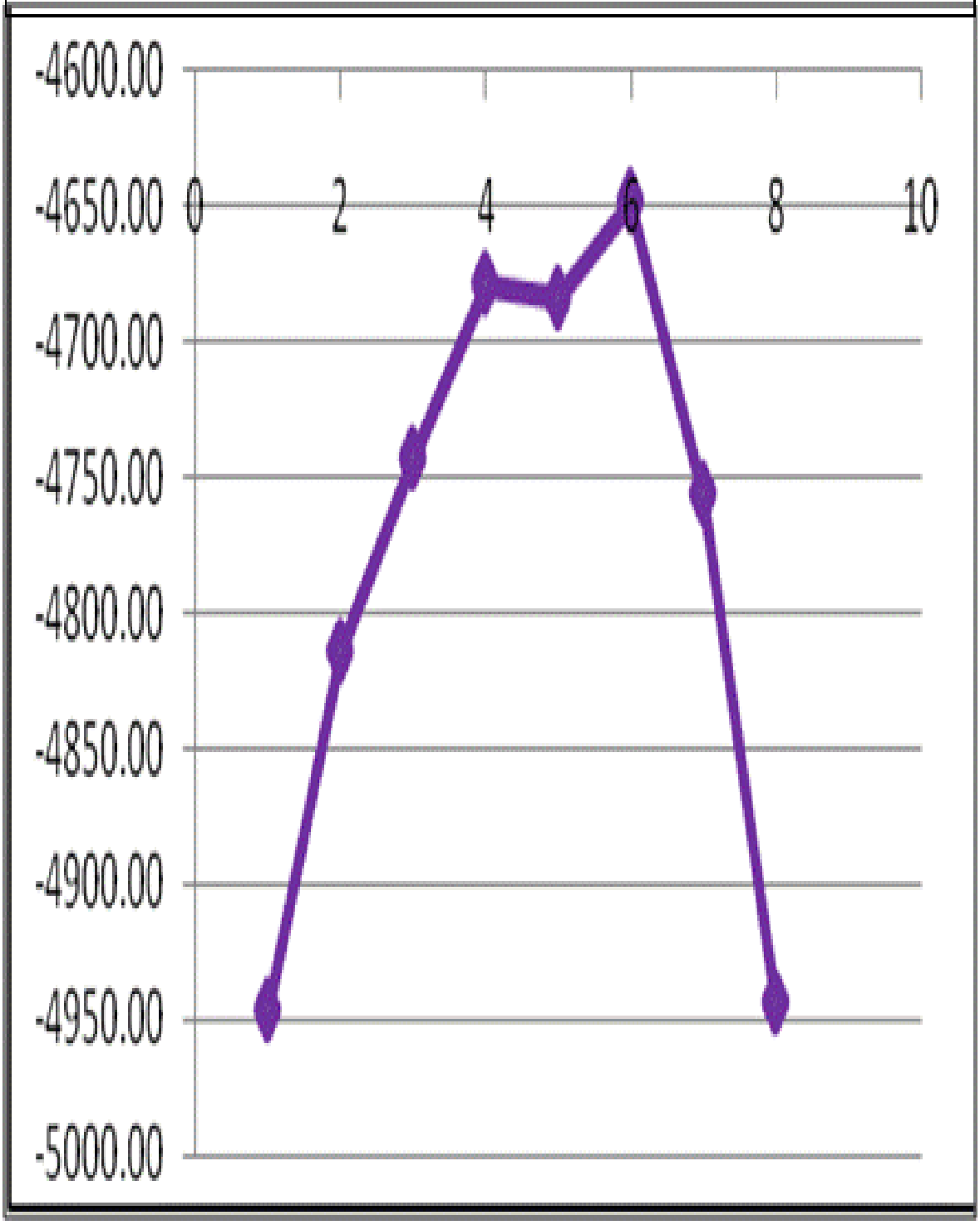


Figure 3: STRUCTURE Q plots for various subsets of microsatellite data

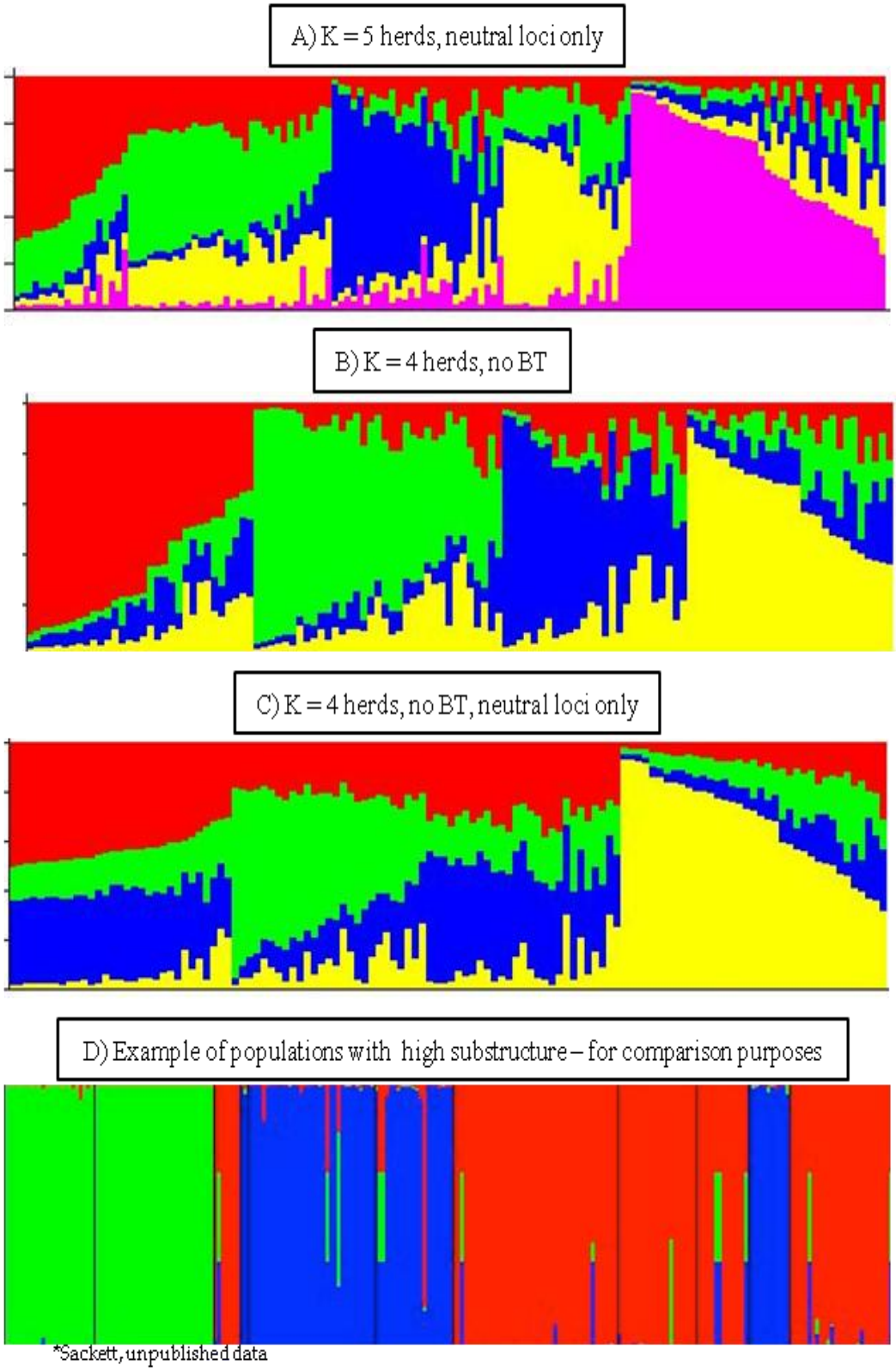


Figure 4: Overall fixation index (F_{st}) estimated from microsatellite genotypes by the program Genepop. Pie charts show proportion of sheep from each herd that were assigned to the five populations designated by the program Structure.

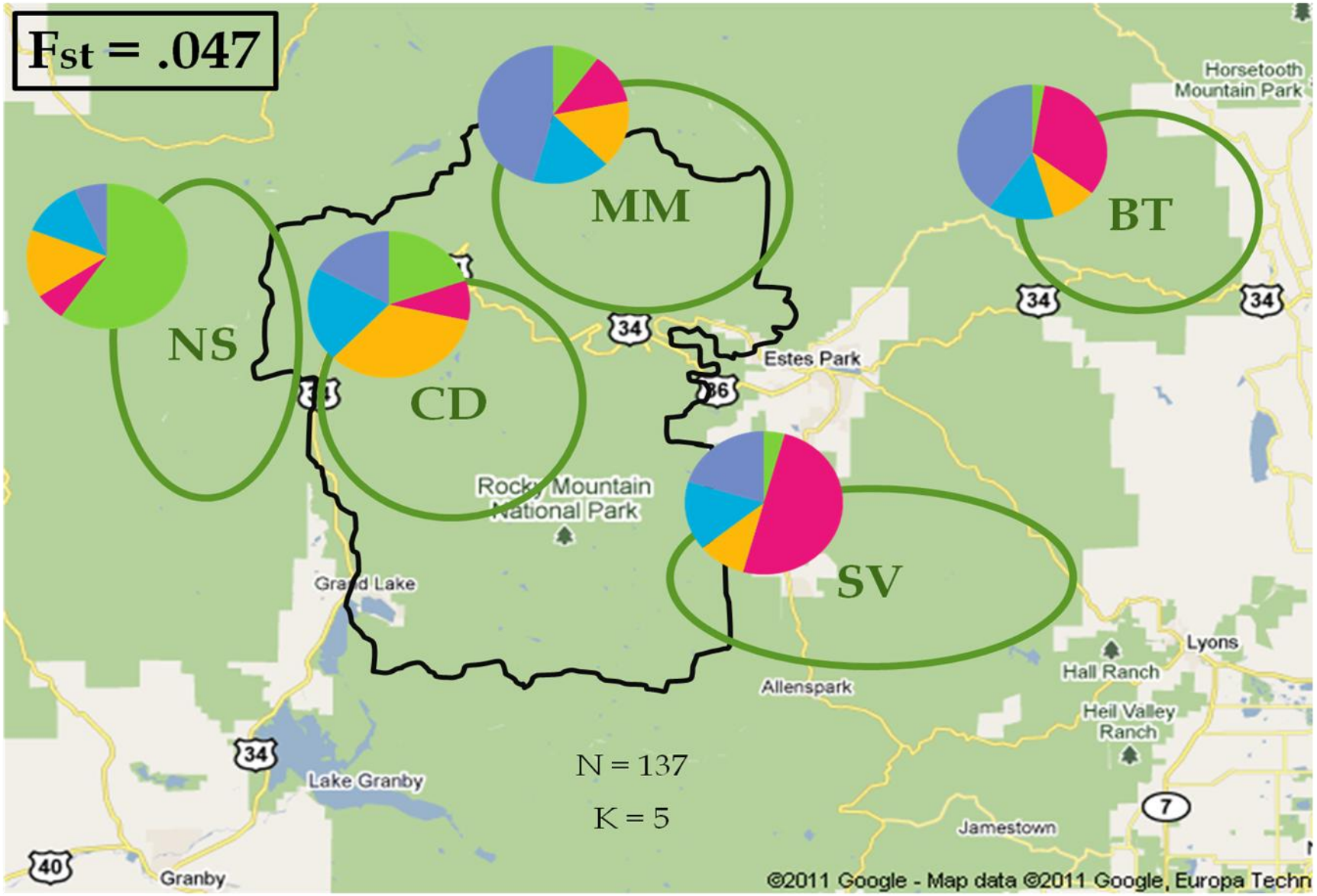


Figure 5: Overall and pairwise fixation indices (F_{st}) estimated from microsatellite genotypes by the program Genepop. Arrows indicate pairwise F_{st} values between all five herds in this metapopulation.

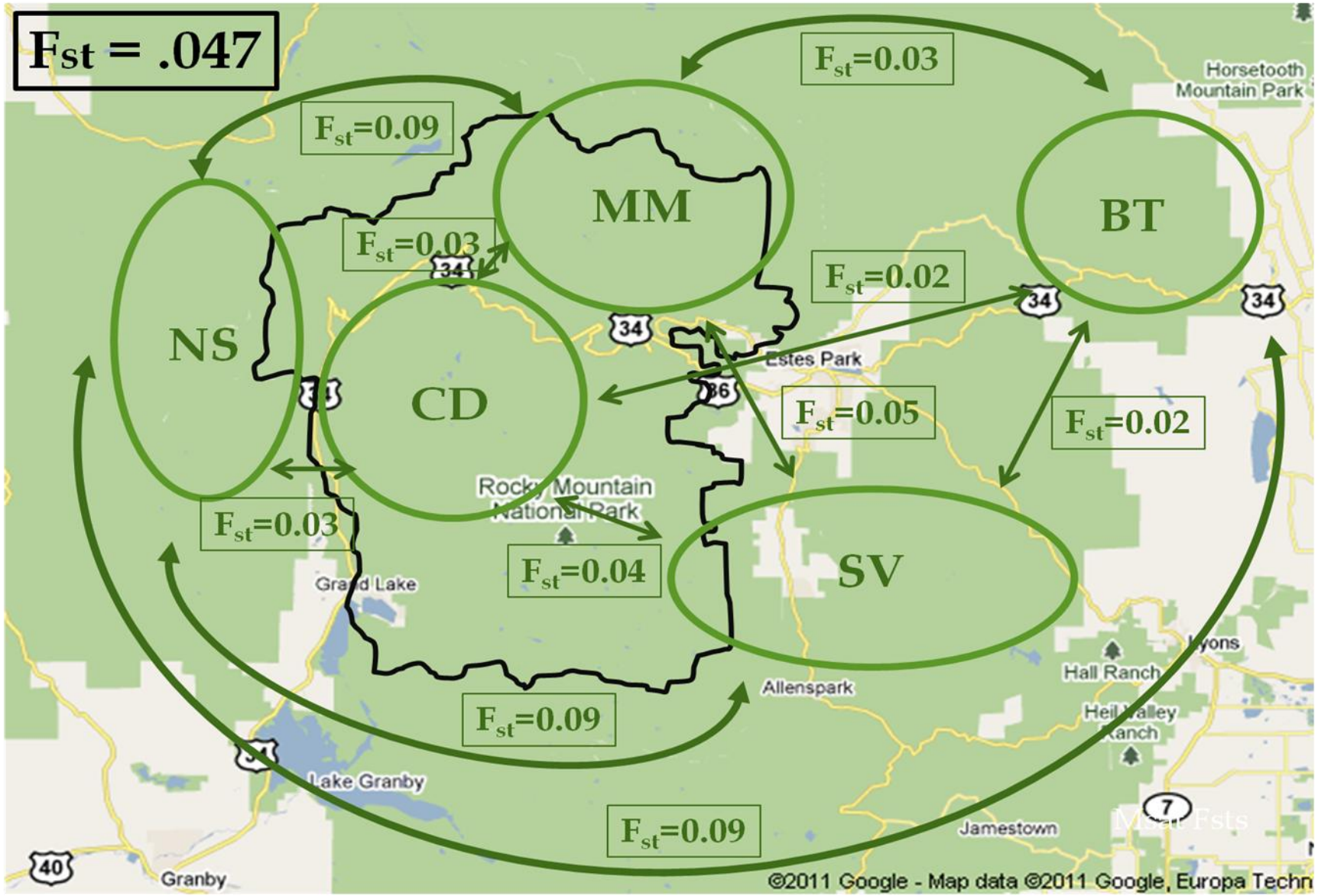


Figure 6: Overall and individual herd inbreeding coefficients (F_{is}) estimated from microsatellite genotypes from all 13 loci by the program Genepop.

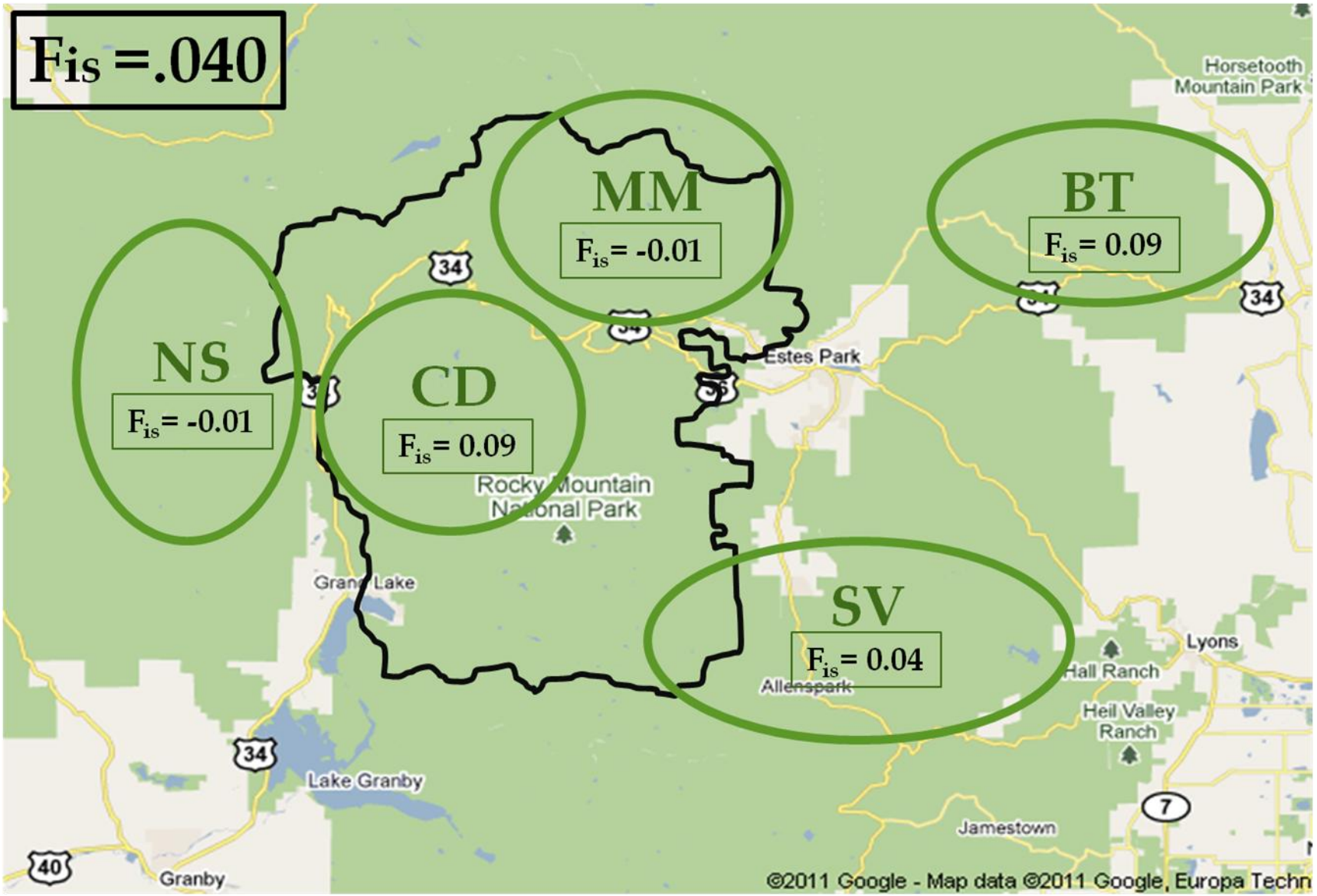
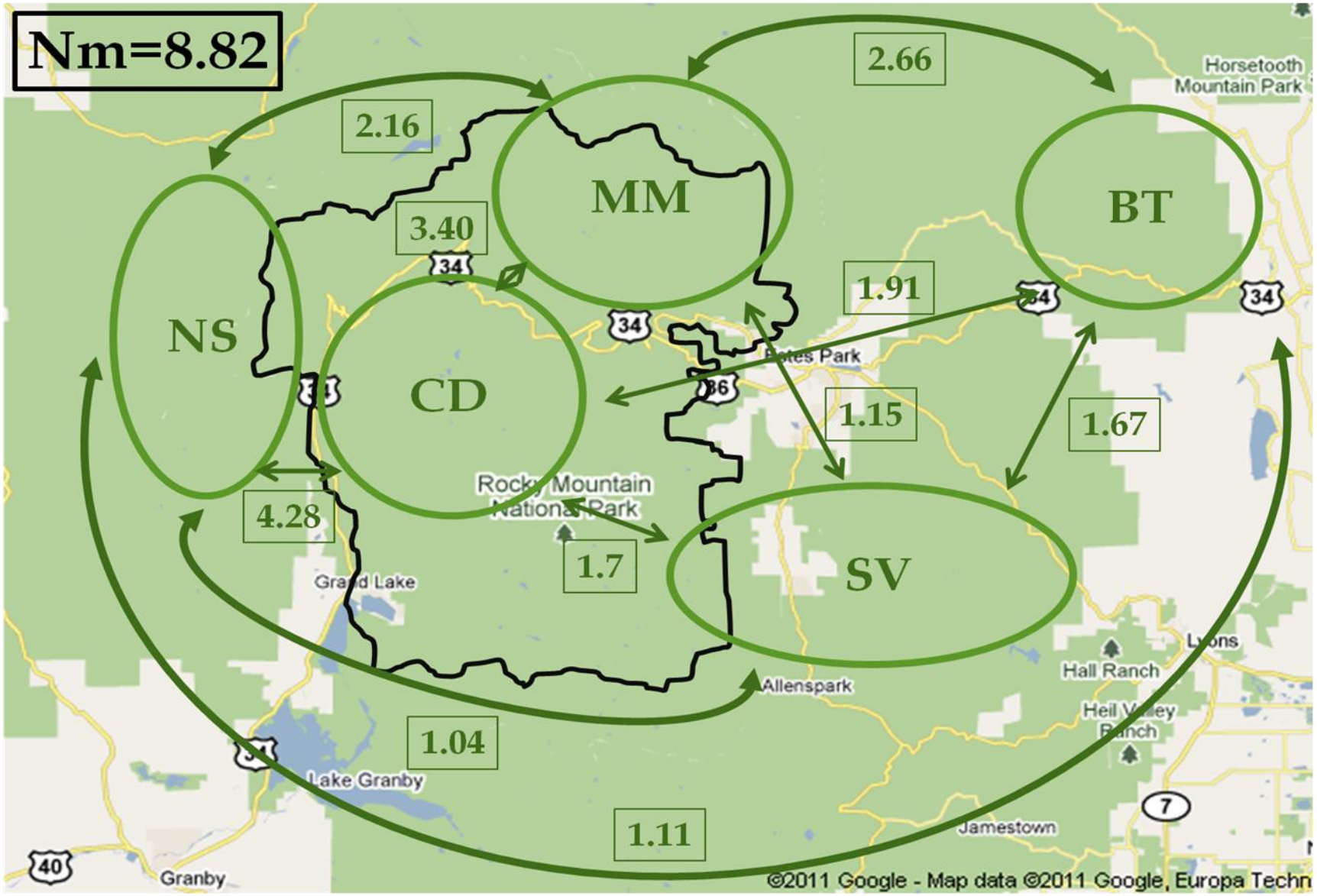


Figure 7: Overall and pairwise effective migration (Nm) rates estimated from microsatellite genotypes by the program Genepop. Arrows indicate pairwise Nm values between all five herds in this metapopulation.



III. ***Mitochondrial DNA Sequence Analyses***

Summary mtDNA analyses by herd are show in Table 8 and all pairwise comparisons for these data are shown in Table 9. Twelve unique haplotypes were identified from 190 total individuals (Table 10). We found a total of five polymorphic sites in COI and 101 polymorphic sites in D-Loop (excluding the 79 base indel). The sequence within this 79bp DL indel was found to be invariant (i.e. was either present or absent) for all sheep sequenced and so was represented as a single SNP in all analyses in order to avoid errors in parameter estimations. Our data are graphically represented by pie charts showing relative distributions of these 12 haplotypes among all herds (Figure 8). We also show pairwise fixation indices (F_{st}) between all herds (Figure 9). Finally, we show pairwise effective migration (Nm) rates between herds (Figure 10). Together, these data demonstrate higher estimates of population substructure from mtDNA sequences than estimates obtained from microsatellite genotypes.

Table 8: DnaSP mtDNA estimates by herd. Total sample size (N), Total segregating sites (S), total haplotypes (h), Haplotype diversity (Hd), average number of nucleotide differences (K), nucleotide diversity (Pi), and nucleotide diversity with Jukes and Cantor correction (Pi(JC).

Herd	N	S	h	Hd	K	Pi	Pi (JC)
Big Thompson	20	3	2	0.479	1.437	0.0140	0.014
Continental Divide	70	42	7	0.682	12.065	0.117	0.144
Mummy	30	38	4	0.469	13.184	0.127	0.160
Never Summer	41	40	5	0.703	16.561	0.161	0.197
St. Vrain	29	23	6	0.633	9.877	0.959	0.109

Table 9: DnaSP Pairwise mtDNA summary data. Overall Nm = 1.52.

Herd 1	Herd 2	Hs	Ks	Kxy	G_{st}	D_{st}	G_{st}	Nst	Nm	F_{st}	Dxy	Da
BT	CD	0.640	9.372	8.779	0.182	0.008	0.081	0.262	2.25	0.255	0.087	0.022
BT	MM	0.473	8.485	9.750	0.068	0.013	0.136	0.262	2.01	0.250	0.097	0.024
BT	NS	0.633	11.292	15.635	0.224	0.031	0.221	0.441	1.74	0.439	0.155	0.068
BT	SV	0.560	6.170	7.271	0.171	0.010	0.142	0.245	2.57	0.252	0.072	0.018
CD	MM	0.620	12.103	12.613	0.199	0.002	0.018	0.010	2.01	0.016	0.125	0.002
CD	NS	0.690	13.287	15.453	0.037	0.009	0.062	0.096	13.14	0.102	0.153	0.016
CD	SV	0.663	10.993	12.514	0.164	0.009	0.079	0.158	2.62	0.158	0.124	0.020
MM	NS	0.606	14.868	16.583	0.217	0.011	0.074	0.103	1.80	0.117	0.164	0.019
MM	SV	0.540	11.340	14.015	0.282	0.015	0.124	0.196	1.27	0.193	0.139	0.027
NS	SV	0.667	13.338	15.832	0.139	0.016	0.114	0.208	3.09	0.194	0.157	0.030

Table 10: Mitochondrial haplotypes and distributions among herds

Haplotype	Big Thompson	Continental Divide	Mummy	Never Summer	St. Vrain	Total
1	13	2	21	0	1	37
2	7	0	0	0	16	23
3	0	11	7	10	0	28
4	0	36	0	13	0	49
5	0	9	1	16	9	35
6	0	10	0	0	0	10
7	0	1	0	0	0	1
8	0	1	0	0	2	3
9	0	0	1	0	0	1
10	0	0	0	1	0	1
11	0	0	0	1	0	1
12	0	0	0	0	1	1
total	20	70	30	41	29	190

Figure 8: Overall fixation index (F_{st}) estimated from mtDNA sequence data by DnaSP. Pie charts indicate proportion of sheep in each herd with each of the 12 color-coded haplotypes.

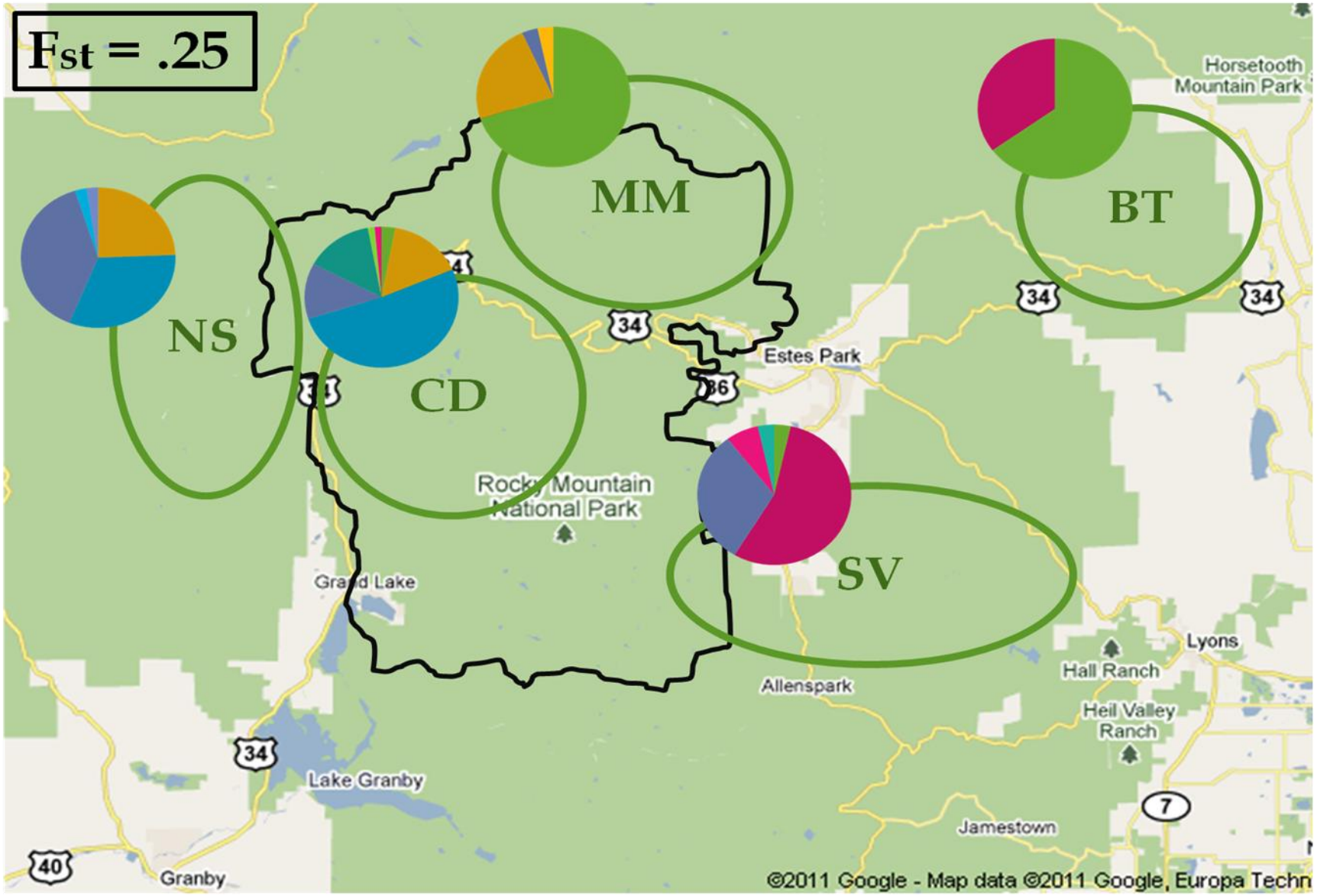


Figure 9: Overall and pairwise fixation indices (F_{st}) estimated from mtDNA sequences by DnaSP. Arrows indicate pairwise F_{st} values between all five herds in this metapopulation.

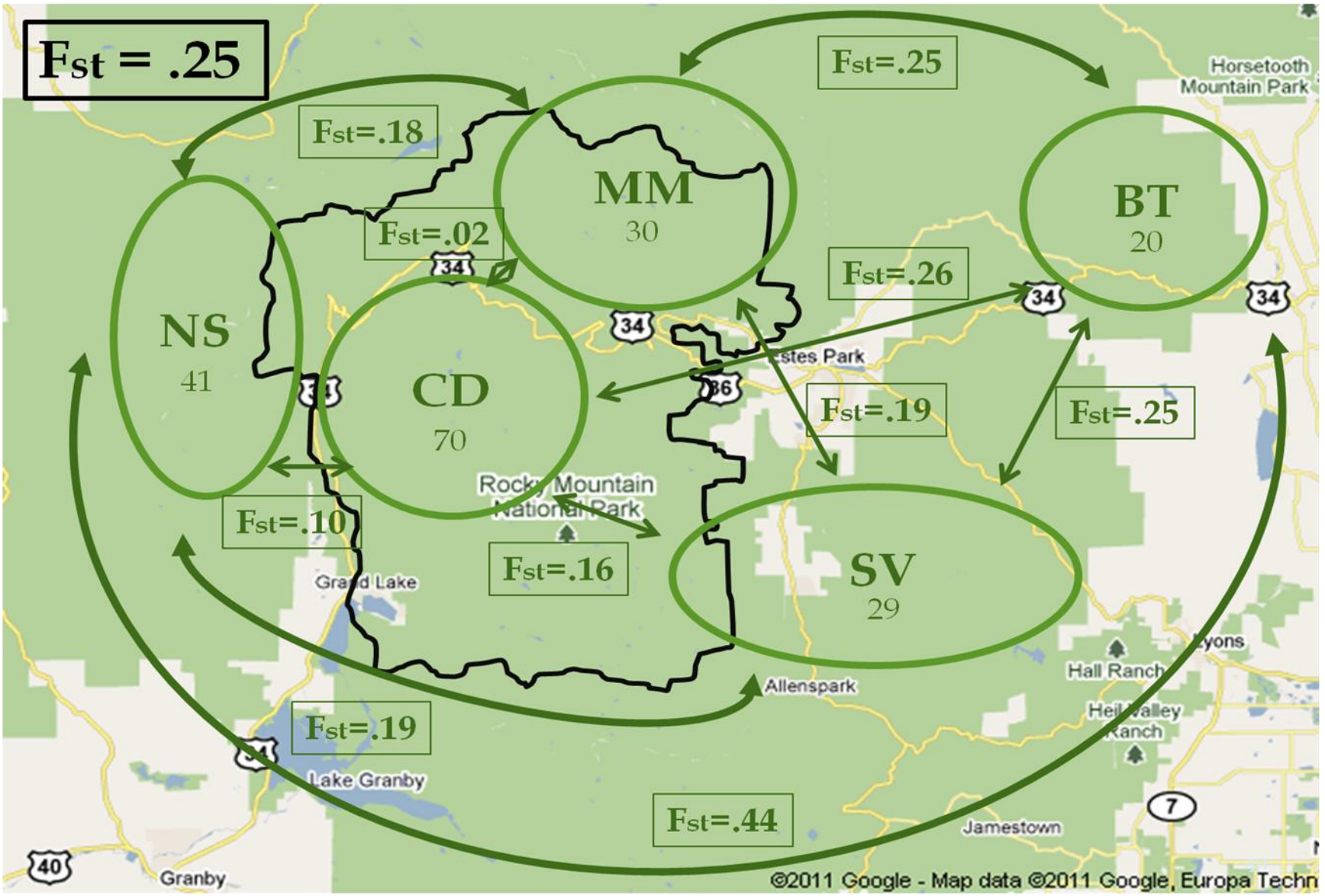
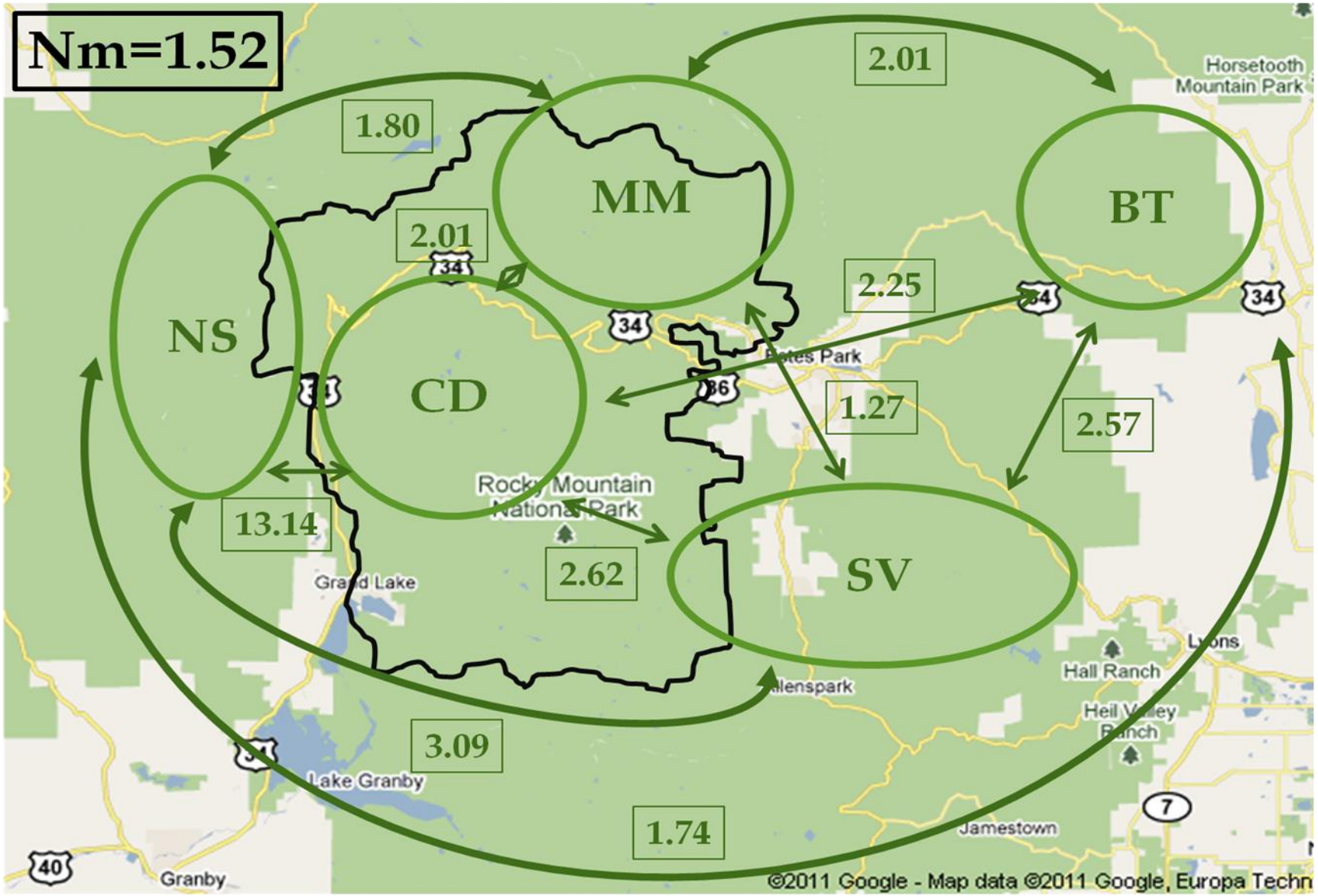


Figure 10: Overall and pairwise effective migration (Nm) rates estimated from mtDNA sequences by DnaSP. Arrows indicate pairwise Nm values between all five herds in this metapopulation.



IV. *SRY Gene Sequencing Analyses*

Both of our initial SRY Open Reading Frame (ORF) and Promoter (Pro) sequence data sets were invariant within and among RMNP bighorn sheep herds (Table 2). That is, we found absolutely no genetic variation at these two markers among the 102 rams whose DNA we analyzed for this part of the study. Due to this absence of genetic variation, SRY markers could not be used for population-level analyses. Thus, we were unable to track ram-mediated gene flow using these SRY data.

In order to make some productive use of these data, we expanded each data set to include samples from all six wild sheep species within the *Ovis* genus: bighorn sheep (*Ovis c. canadensis, Ovis c. nelsoni*), snow sheep (*Ovis nivicola*), dall's sheep (*Ovis d. dalli, Ovis d. stonei*), argali sheep (*Ovis ammon*), urial sheep (*Ovis vignei*) and mouflon sheep (*Ovis orientalis*). Our final data sets included DNA sequences from three sources: 1) amplifications performed in the Mitton lab on samples collected from RMNP and the surrounding area; 2) amplifications performed in the Mitton lab on samples provided by collaborators (Dr. John Wehausen, Dr. Rob Ramey III, The Colorado National Monument, and James Driscoll); 3) previously published sequences obtained from NCBI GenBank.

These expanded data sets included 20 additional ORF sequences and 19 additional Pro sequences generated in the Mitton lab as well as all available published sequences. Detailed species IDs, NCBI accession numbers, sample collection information and haplotypes are given in Table 11 (SRY ORF) and Table 12 (SRY Pro).

The SRY promoter has been more extensively sequenced than the SRY ORF and therefore our sample sizes varied by data set due to differences in availability of published sequences for these two regions. For these analyses we generated three distinct data subsets by combining all available published sequences with those we generated specifically for this study: 1) all ORF sequences (26 total, representing 23 different populations); 2) all promoter sequences (160 total, representing 31 populations) and 3) concatenated ORF and promoter sequences (18 total, representing 16 populations) (Tables 11 & 12, all individuals included in the concatenated data set are indicated with an asterisk). All three data sets include at least two individuals from each of the six wild sheep species with the exception of the concatenated data set for which we were only able to amplify both ORF and Pro sequences from the same individual for a single *O. musimon* DNA sample (ARM). For visual simplicity we represent all RMNP sequences with a single sample in each phylogenetic tree presented here.

From this expanded data set we identified seven ORF Single Nucleotide Polymorphisms (SNPs) at six polymorphic sites (OY1-OY6). Of these, five sites were parsimony informative which corresponded to four distinct haplotypes (Table 13). Four of these seven SNPs were synonymous substitutions (C/T at 144bp, C/T at 268bp, A/G at 417bp and A/G at 591bp) while three were non-synonymous substitutions (A/C at 16bp: Lys ->Gln, codon 6; C/G at 144bp: Asn->Lys, codon 48; C/T at 151bp: Arg->Cys, codon 51). None of the non-synonymous SNPs were within the SRY

HMG-box motif of the DNA-binding region and are therefore unlikely to have a significant phenotypic effect.

We also identified nine Promoter SNPs at eight polymorphic sites (OY1-OY8) which corresponded to 7 distinct haplotypes (Table 14). From these data we found intraspecies variation in: *O. aries* (N = 8), *O. vignei* (N = 7), *O. canadensis* (N = 114), *O. dalli* (N = 6) and *O. d. stonei* (N = 3) (Table 14).

For all three data sets we generated both rooted and unrooted phylogenetic trees with and without a designated outgroup: 1) ORF only (Figures 11 & 13), 2) Promoter only (Figures 12 & 13) and 3) Concatenated ORF and Promoter sequences (Figures 14 & 15). For all trees we represent the number of base changes between clades as branch lengths. Each data set was analyzed using four different tree estimation methods: Maximum Parsimony, Maximum Likelihood, Bayesian and Neighbor-Joining search algorithms. Tree topology was consistent across all methods, lending support to this phylogenetic structure of the *Ovis* genus. For simplicity only Maximum Parsimony analyses are shown here (Figures 11-15).

Tree topology from all three data sets show a clear division between "new world" sheep (Pachyceriforms: bighorn, thinhorn and snow) and "old world" sheep" which are divided into Moufloniforms (mouflon and urial) and Argaliforms (argali). These data support and extend previous findings based on more limited analyses and inferences of evolutionary relationships based on these data will prove useful for comparisons with other genetic markers.

Our data include the first snow sheep SRY promoter sequences as well as the first SRY ORF sequences for all six wild sheep species. Our expanded data set also provides the first evidence of intraspecies SRY variation within the *Ovis* genus. With these data we performed the first phylogenetic analyses to include all six wild sheep species using Y chromosome markers, clarifying taxonomic relationships within this genus. While these data are not useful for tracking ram-mediated gene flow within RMNP, they do contribute to increased taxonomic resolution between wild sheep species as well as providing valuable genetic data for future studies.

Table 11: SRY ORF sample IDs, collection information, NCBI accessions and haplotype IDs. Samples indicated with an asterisk were also included in the concatenated data set.

Species	Common Name	Location/Source	Accession	ID	n	H1	H2	H3	H4
O. c. canadensis	**RM Bighorn**	**CD herd, RMNP**	JN992663	**CR9F**	**1**	**1**			
O. c. canadensis	**RM Bighorn**	**NS Herd, RMNP**	JN992662	**BC23***	**1**	**1**			
O. c. canadensis	**RM Bighorn**	**SV Herd, RMNP**	JN992666, JN992667	**SV1D,11D***	**2**	**2**			
O. c. nelsoni	**Desert Bighorn**	**Unknown**	JN992660	**NV08**	**1**	**1**			
O. c. nelsoni	**Desert Bighorn**	**Durango, CO**	JN992658, JN992659	**DB64*, DB78***	**2**	**2**			
O. c. sierrae	**SN Bighorn**	**Bishop, CA**	JN992661	**SN07***	**1**	**1**			
O. canadensis	**RM Bighorn**	**Pikes Peak, CO**	JN992664, JN992665	**PP01,PP11***	**2**	**2**			
O. nivicola	**Snow**	**Dzhugdzhur Mtns, Russia**	JN992671	**AB344**	**1**	**1**			
O. nivicola	**Snow**	**Unknown**	JN992672	**AB345***	**1**	**1**			
O. nivicola	**Snow**	**Unknown**	JN992673	**CK1STA***	**1**	**1**			
O. d. dalli	**Thinhorn**	**Alaska Range, AK**	JN992668	**88T***	**1**		**1**		
O. d. dalli	**Thinhorn**	**Unknown**	JN992669	**DRIT***	**1**		**1**		
O. d. stonei	**Stone's**	**Atlin, B.C.**	JN992670	**JS1***	**1**		**1**		
O. ammon	**Argali**	**Mongolia**	JN992656	**GA392***	**1**			**1**	
O. ammon	**Argali**	**Central Asia**	JN992657	**SU492***	**1**			**1**	
O. ammon	**Argali**	**Kargash Mtns, Kazikhstan**	JN992654	**AMIT**	**1**			**1**	
O. ammon	**Argali**	**Unknown**	JN992655	**CK3T***	**1**			**1**	
O. aries	**Domestic**	**unknown**	**Z30265**	**Z30265**	**1**			**1**	
O. orientalis	**Mouflon**	**Armenia**	JN992674	**ARM***	**1**			**1**	
O. orientalis	**Mouflon**	**Armenia**	JN992675	**OR672**	**1**			**1**	
O. orientalis	**Mouflon**	**Armenia**	JN992676	**WAP08**	**1**			**1**	
O. vignei	**Urial**	**Turkmenistan**	JN992677	**AU125***	**1**			**1**	
O. vignei	**Urial**	**Unknown**	JN992678	**CK2T***	**1**				**1**
Totals					**26**	**13**	**3**	**9**	**1**

Table 12: SRY Pro sequence IDs, collection information, NCBI accessions and haplotype IDs. Samples indicated with an asterisk were also included in the concatenated data set.

Species	Sample Location	Accession Numbers	Sample ID	n	Haplotype	H1	H2	H3	H4	H5	H6	H7
O. ammon	Unknown	EU938024-EU938027	N/A	4	ACTTGGGA(4)				4			
O. ammon	Mongolia	JN992692	GA392*	1	ACTTGGGA(4)				1			
O. ammon	Central Asia	JN992691	SU492*	1	ACTTGGGA(4)				1			
O. ammon	Unknown	JN992693	CK3T*	1	ACTTGGGA(4)				1			
O. aries	Various	EU938038-EU938045	N/A	8	ACTTGGAA(6) GCTTGGAA(7)						6	2
O. c. canadensis	SV herd, RMNP	JN400249	SV*	17	ATCGAGTG(1)	17						
O. c. canadensis	MM herd, RMNP	JN400247	MM,NF,SP	14	ATCGAGTG(1)	14						
O. c. canadensis	NS Herd, RMNP	JN400248	BC*,RR,SG	37	ATCGAGTG(1)	37						
O. c. canadensis	CD Herd, RMNP	JN400250	CR,MI,RC	30	ATCGAGTG(1)	30						
O. c. canadensis	BT Herd, Drake	JN400251	BT	4	ATCGAGTG(1)	4						
O. c. canadensis	Pikes Peak, CO	JN400253	PP*	3	ATCGAGTG(1)	3						
O. c. canadensis	Aldrich Mtn, OR	JN400252	AM	6	ATCGAGTG(1)	6						
O. c. canadensis	Alberta Canada	JN992690	Cad8	1	ATCGAGTG(1)	1						
O. c. canadensis	Alberta, Canada	EU938032,EU938033	N/A	2	ATTTAGTA(3)			2				
O. c. nelsoni	Durango, CO	JF680994	DB*	2	ATCGAGTG(1)	2						
O. c. sierrae	Bishop, CA	JF680995	SN*	8	ATCGAGTG(1)	8						
O. d. dalli	Alaska Mtns, AK	JN992688	88T*	1	ATTGAGTG(2)		1					
O. d. dalli	Hula Hula AK	JN992689	ANWR12	1	ATTGAGTG(2)		1					
O. d. dalli	Brooks Range, AK	JN992681	ODD3	1	ATCGAGTG(1)	1						
O. d. dalli	Northwest AK	EU938034,EU938036	N/A	2	ATTTAGTA(3)			2				
O. d. dalli	Unknown	JN992680	DRIT*	1	ATTGAGTG(2)		1					
O. d. stonei	Atlin, B.C.	JN992679	JS1*	1	ATTGAGTG(2)		1					
O. d. stonei	Yukon B.C.	EU938035, EU938037	N/A	2	ATTTAGTA(3)			2				
O. orientalis	Ukraine	EU938022,EU938023	N/A	2	ACTTGGAA(6)						2	
O. orientalis	Armenia	JN992684	ARM*	1	ACTTGGAA(6)						1	
O. nivicola	Unknown	JN992682	AB345*	1	ATTGAGTG(2)		1					
O. nivicola	Unknown	JN992683	CK1STA*	1	ATTGAGTG(2)		1					
O. vignei	Urial	JN992687	CK2T*	1	ACTTATGA(5)					1		
O. vignei	Turkmenistan	JN992685	AU125*	1	ACTTGGGA(4)				1			
O. vignei	Urial	JN992686	AfUrial	1	ACTTGGGA(4)				1			
O. vignei	Unknown	EU938028-EU93831	N/A	4	ACTTATGA(5)					4		
Total				160		123	6	6	9	5	9	2

Table 13: SRY ORF haplotypes by species: Distributions and SNP IDs

Species	n	Haplotype	OY1	OY2	OY3	OY4	OY5	OY6
O. c. canadensis	6	1	C	T	T	T	A	G
O. c. sierrae	1	1	C	T	T	T	A	G
O. c. nelsoni	3	1	C	T	T	T	A	G
O. nivicola	3	1	C	T	T	T	A	G
O. d. dalli	2	2	C	T	T	T	A	A
O. d. stonei	1	2	C	T	T	T	A	A
O. ammon	4	3	C	C	C	C	G	G
O. orientalis	3	3	C	C	C	C	G	G
O. vignei	1	3	C	C	C	C	G	G
O. vignei	1	4	A	G	C	C	G	G
O. aries	1	3	C	C	C	C	G	G

Table 14: SRY Promoter haplotypes by species: Distributions and SNP IDs

Species	n	Haplotype	OY1	OY2	OY3	OY4	OY5	OY6	OY7	OY8
O. c. canadensis	112	1	A	T	C	G	A	G	T	G
O. c. canadensis	2	3	A	T	T	T	A	G	T	A
O. c. nelsoni	2	1	A	T	C	G	A	G	T	G
O. c. sierrae	8	1	A	T	C	G	A	G	T	G
O. d. dalli	1	1	A	T	C	G	A	G	T	G
O. d. dalli	3	2	A	T	T	G	A	G	T	G
O. d. dalli	2	3	A	T	T	T	A	G	T	A
O. d. stonei	1	2	A	T	T	G	A	G	T	G
O. d. stonei	2	3	A	T	T	T	A	G	T	A
O. nivicola	2	2	A	T	T	G	A	G	T	G
O. ammon	7	4	A	C	T	T	G	G	G	A
O. vignei	2	4	A	C	T	T	G	G	G	A
O. vignei	5	5	A	C	T	T	A	T	G	A
O. orientalis	3	6	A	C	T	T	G	G	A	A
O. aries	6	6	A	C	T	T	G	G	A	A
O. aries	2	7	G	C	T	T	G	G	A	A

Figure 11: SRY ORF maximum parsimony tree. Colors indicate different *Ovis* species; branch lengths indicate DNA base changes between clades

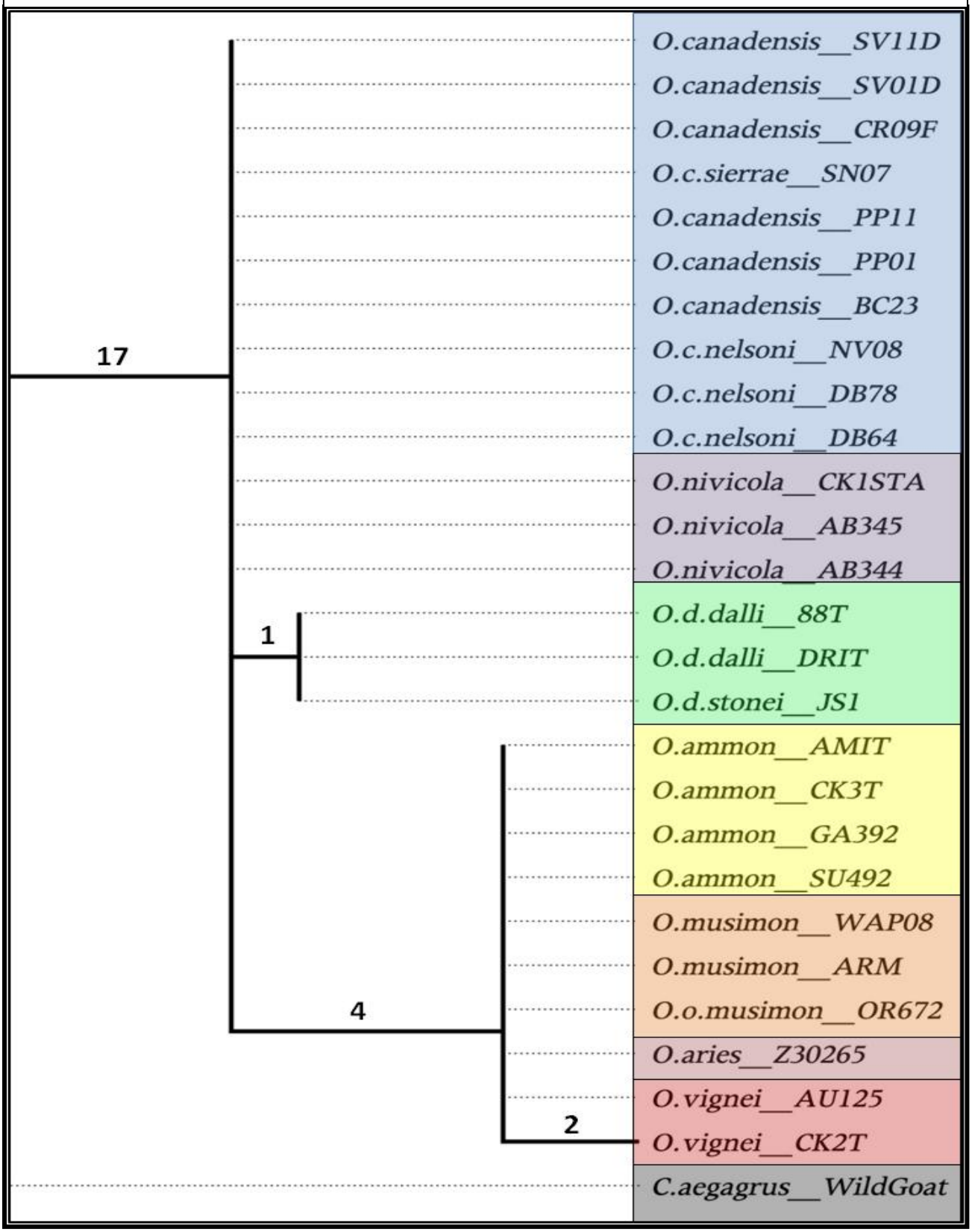


Figure 12: SRY Promoter maximum parsimony tree. Colors indicate different *Ovis* species. Branch lengths indicate DNA base changes between clades

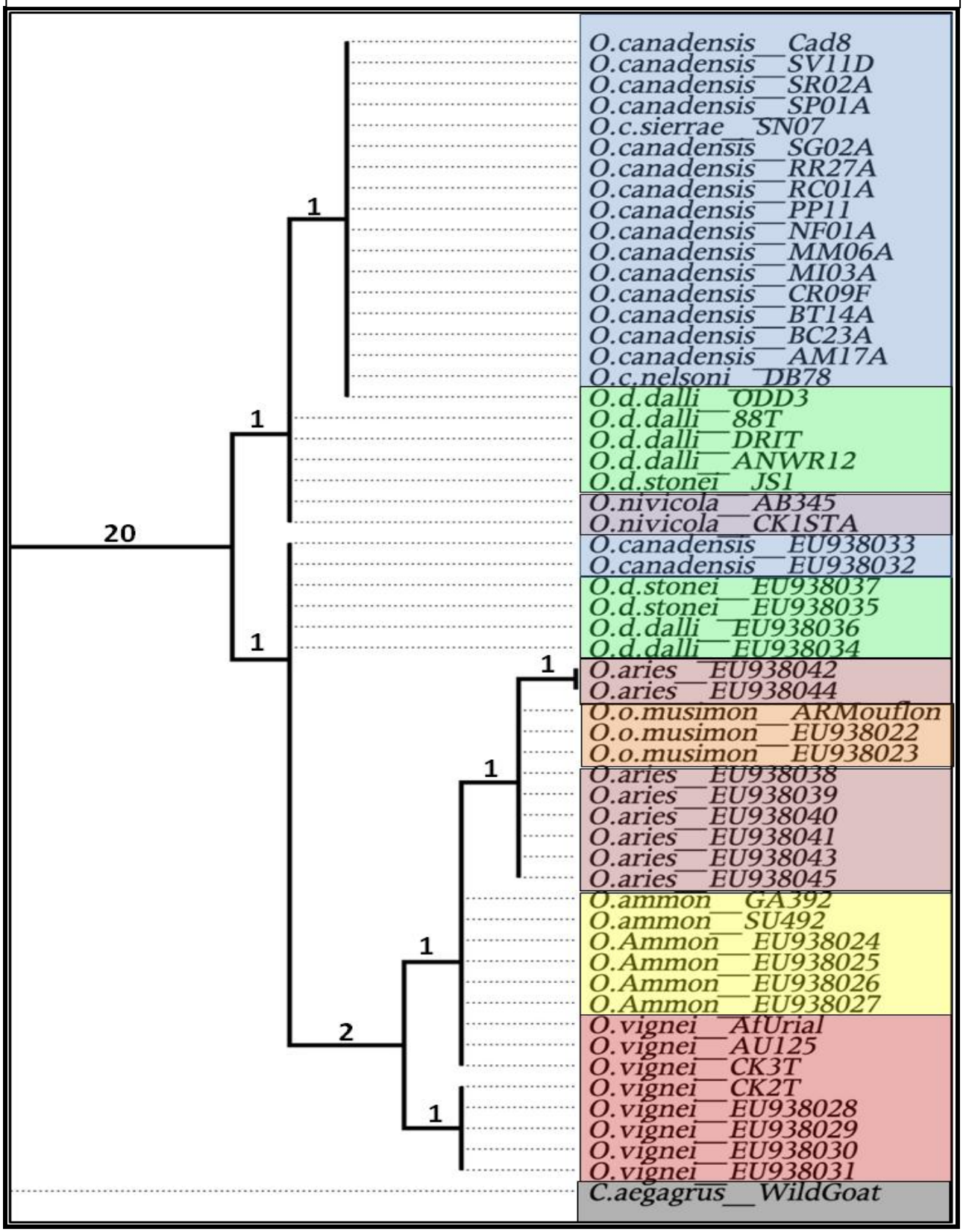


Figure 13: Comparison of ORF and Promoter maximum parsimony trees. Branch lengths indicate DNA base changes between clades

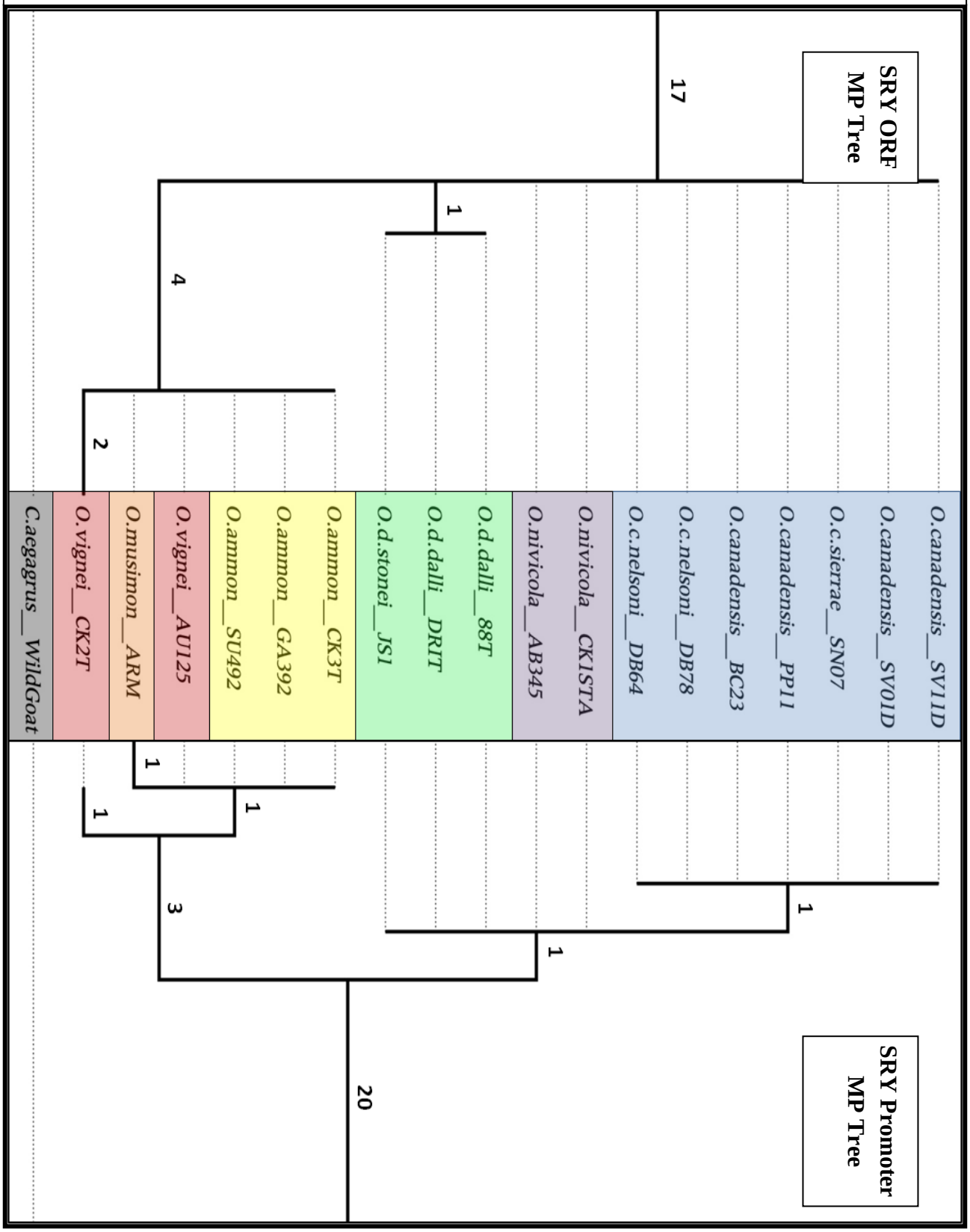


Figure 14: SRY ORF and promoter sequence data set: maximum parsimony unrooted phylogram.

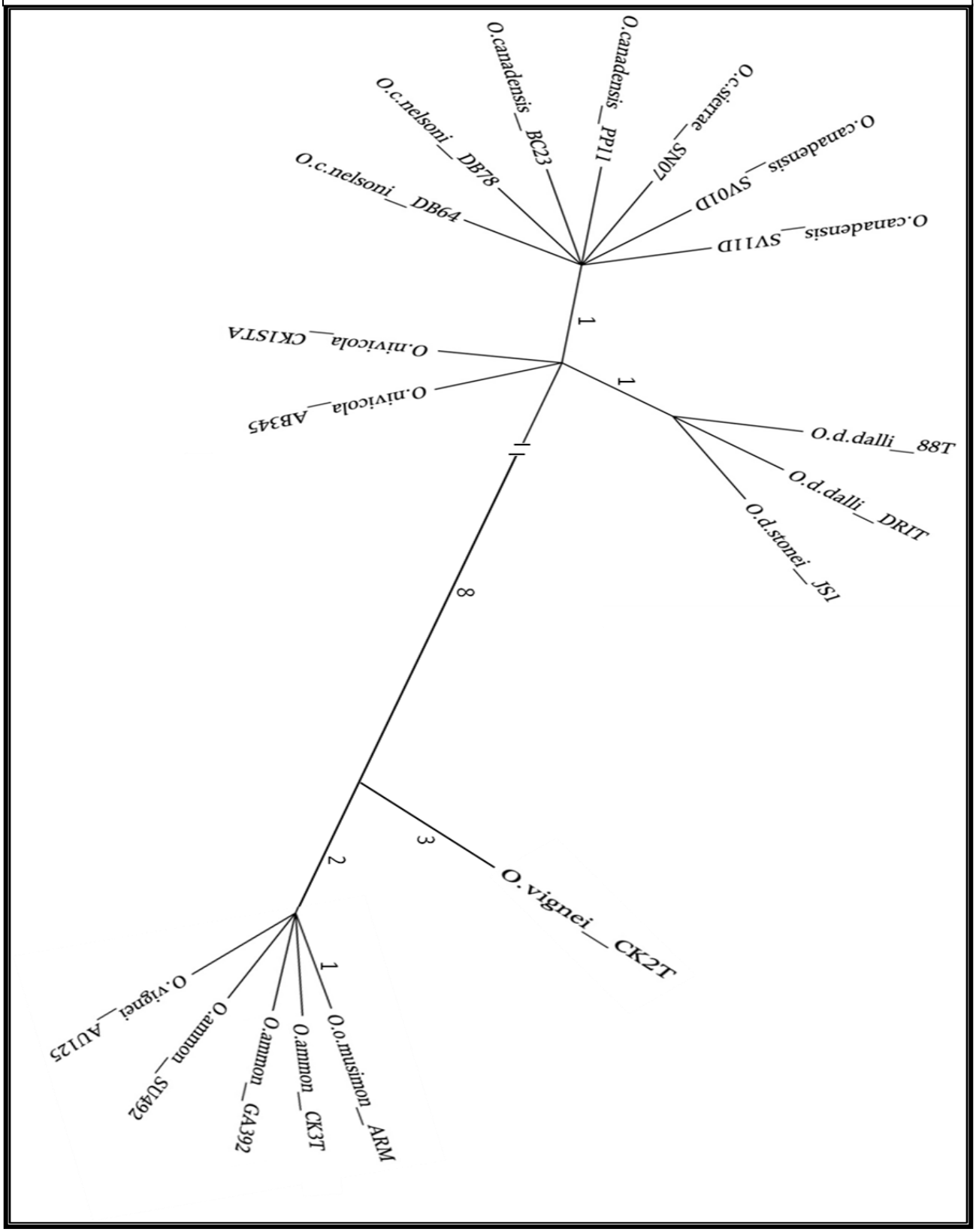
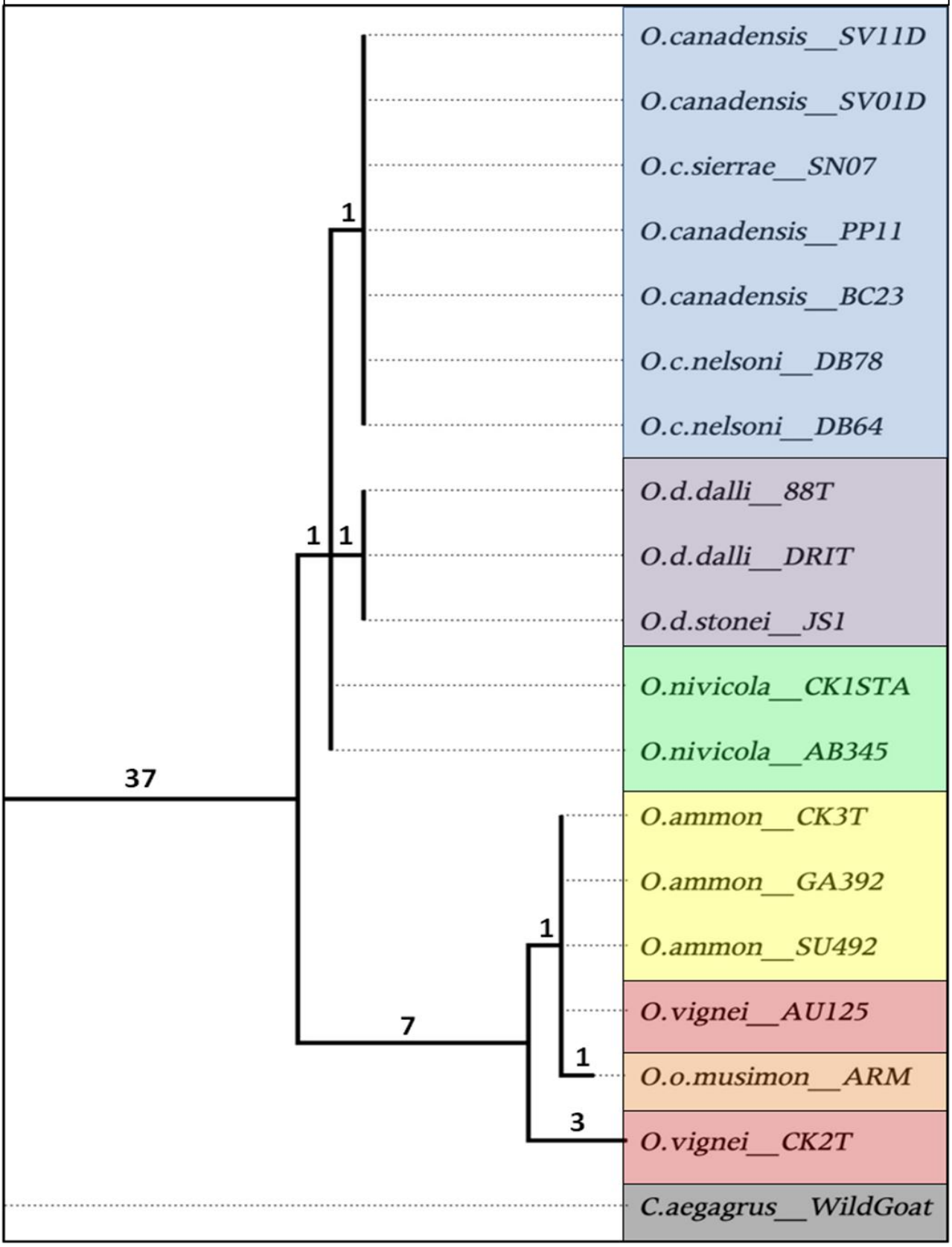


Figure 15: SRY concatenated ORF/Pro data set: maximum parsimony phylogenetic tree.



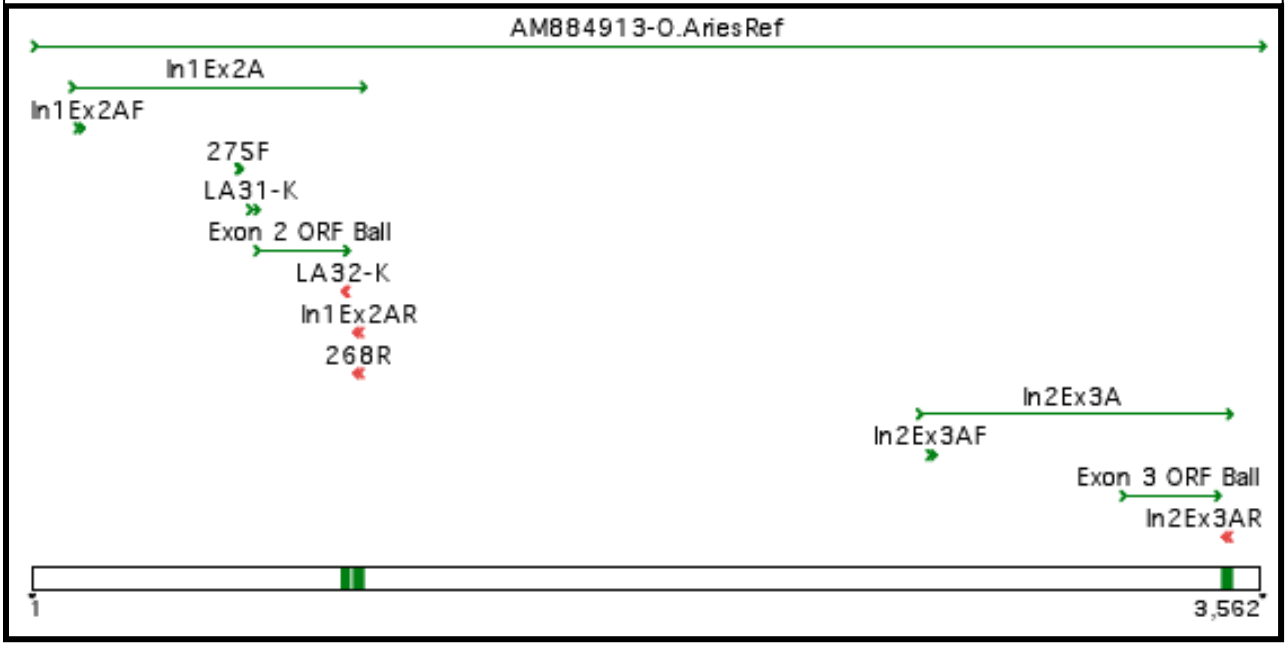
V. *MHC DRB1 Gene Sequence Analyses*

From our initial data collection a total of 27 DRB1 sequences were successfully amplified from ten different individuals. These sequences were then cloned and transformed, and 342 individual colonies were grown. From these, plasmids were extracted and DRB1 inserted DNA was sequenced. Ultimately, 298 high quality DRB1 sequences were included in these analyses and results are summarized by gene region in Table 15. These included amplifications from three separate primer sets covering two distinct regions (Figure 16). We obtained between two and 63 distinct sequences per individual.

Table 15: MHC DRB1 sequence information: sample IDs and distributions by region

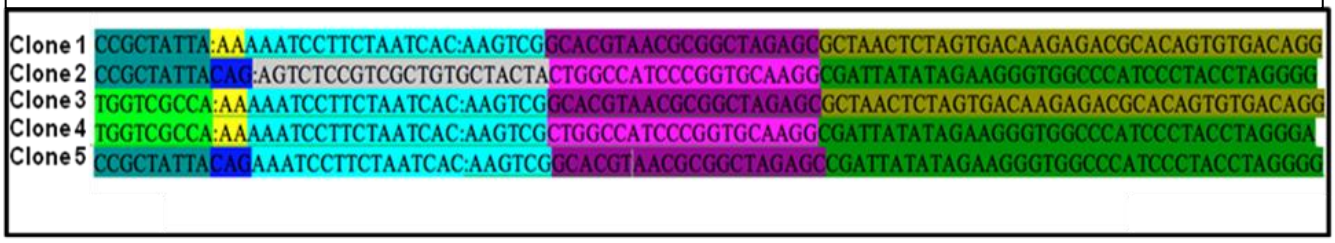
Primers	**268/275**			**In1Ex2A-F/R**			**In2Ex3A-F/R**			**Totals**
Sample	**PCR**	**Clones**	**Seqs**	**PCR**	**Clones**	**Seqs**	**PCR**	**Clones**	**Seqs**	
PP2	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**9**	**14**	**24**
PP9	**0**	**0**	**0**	**2**	**30**	**14**	**1**	**0**	**0**	**46**
BT3	**1**	**16**	**0**	**1**	**16**	**4**	**0**	**0**	**0**	**38**
BT10	**1**	**12**	**1**	**3**	**47**	**63**	**1**	**8**	**16**	**152**
CR6A	**1**	**16**	**0**	**1**	**16**	**9**	**0**	**0**	**0**	**43**
MI1A	**1**	**0**	**2**	**1**	**7**	**14**	**1**	**5**	**10**	**41**
SV7A	**1**	**0**	**2**	**1**	**16**	**8**	**0**	**0**	**0**	**28**
SV3D	**1**	**16**	**3**	**2**	**30**	**34**	**1**	**7**	**14**	**105**
MM7A	**1**	**20**	**20**	**1**	**25**	**37**	**1**	**23**	**0**	**128**
MM7C	**1**	**30**	**18**	**1**	**30**	**18**	**1**	**30**	**0**	**129**
Total	**8**	**110**	**43**	**13**	**217**	**201**	**6**	**82**	**54**	**734**

Figure 16: Annotated DRB1 gene sequence alignment including reference *Ovis aries* sequence (NCBI: AM884913) with all primers indicating relative location and length of regions amplified.



From these data we observed more alleles than could be explained by a single autosomal locus in a diploid individual. Figure 17 illustrates just one example of this phenomenon wherein five distinct clones were identified from a single sample (BT10). Sequences shown here contain all 92 variable bases within the complete 831bp In1/Ex2 amplification. Individual colors define distinct sequences within this region for ease of comparison.

Figure 17: MHC DRB1 In1/Ex2A: Five distinct cloned haplotypes obtained from single individual



A reasonable concern for this project was that sequences could be altered due to potential errors at three difference steps in the data collection process. Introduced mutations and heteroduplexes can lead to an overestimation of allele number and these artifacts must be ruled out as potential causes of observed allelic diversity (L'abbe *et al.* 1992; Longeri *et al.* 2002). Mutations can be initially introduced due to polymerase errors during the PCR process. To address this we used a polymerase with proofreading capability to reduce amplification artifacts. The second potential source of error is due to mistakes made by bacterial cells when replicating transformed plasmids. We therefore cloned and sequenced both the COI and SRY genes for several of the same individuals and

compared them to the independently collected data from these projects. This allowed us to show that mutations were not introduced into these sequences due to the cloning process because the cloned and non-cloned sequences matched for individuals. Finally, sequencing errors can cause misidentification of chromatograms leading to changes in correct base assignments. To avoid this all sequences were read in both directions and chromatograms were double checked manually to confirm that all peaks were called correctly. Because we designed primers based on bighorn sheep sequence data it is likely that we were able to document additional polymorphism that was not previously identified due to selective amplification of less optimal primers as well as inclusion of shorter amplifications which did not contain the full-length exon one or any of exon three, intron one or intron two sequence.

Together these preliminary In1/Ex2 results show a potential for up to twenty alleles (ten loci) per individual. We performed several experiments in order to confirm that this variation was indeed due to amplification of multiple loci as a result of DRB1 gene duplication and not due to mutations introduced via the data collection process. These data showed that completion of data collection was beyond the scope of this project and therefore we did not gather additional sequence data for this or the In2/Ex3 region.

CONCLUSIONS

I. *Microsatellite Genotype Analyses*

BOTTLENECK data analyses indicate no evidence for a recent genetic bottleneck in any of the five bighorn sheep herds included in this study. We evaluated microsatellite alleles with two qualitative analyses; by histograms of allele frequency classes and as scatter plots of individual loci deviations from expected heterozygosity. Each showed normal microsatellite allele distributions for all five herds. Quantitative analyses of observed and expected microsatellite heterozygosity also show no significant or consistent trend in deviations under the TPM model of microsatellite evolution.

Genepop data analyses give relatively low overall and pairwise fixation indices (F_{st} = 0.047 and 0.021 - 0.090, respectively), indicating very little subdivision between bighorn herds in this metapopulation. Pairwise F_{st} values show a general pattern of increasing isolation-by-distance wherein the closest herds tend to be the least genetically differentiated (lowest pairwise F_{st}) while the farthest are the most genetically differentiated (highest pairwise F_{st}). Most microsatellite variation found in this metapopulation is well represented in each herd with very few private alleles existing within a single population. Moreover, there is an interesting pattern in the data related to the Big Thompson herd. The BT herd has the three lowest pairwise F_{st} values between it and the Mummy, Continental Divide and St. Vrain herds (F_{st} = 0.03, 0.02 & 0.02 respectively). This may indicate that BT sheep are, in fact, regularly migrating into these three other herds and that unique genetic variation (private alleles) within the BT herd may be benefiting RMNP herds via maintenance of increased genetic variation through continuing gene flow.

A second interesting pattern in the data is that the three highest pairwise F_{st} values are between the Never Summer herd and the Mummy, Big Thompson and St. Vrain herds (F_{st} = 0.09, 0.088 & 0.085 respectively). This could indicate that the geography of the Never Summer mountain range is creating a barrier to sheep migration and thus gene flow within the Park. It is also possible that sheep from outside the Park are migrating into the Never Summer herd and contributing unique microsatellite alleles, causing a higher degree of genetic isolation for this herd. That the Never Summer herd could be the most isolated of the five herds is supported by STRUCTURE's qualitative metapopulation analyses which were able to "correctly" assign the largest proportion of NS sheep to their own herd (Figure 4). In contrast, STRUCTURE was least able to "correctly" assign the Continental Divide sheep to a single herd, indicated by the more evenly proportioned colors in the CD pie chart. This is likely explained by the CD herd's geographically central location within this metapopulation allowing the opportunity for higher rates of migration and thus geneflow with all other herds.

This pattern of low metapopulation substructure is further supported by relatively high overall and pairwise migration estimates (Nm = 8.82 and 1.04 - 4.28 respectively). These data also suggest that migration (and thus geneflow) between herds is likely maintaining a high proportion of among vs. within population genetic variation. Like the fixation indices, these data exhibit a general pattern of increasing isolation-by-distance such that the farthest herds have the lowest pairwise migration rates (Nm = 1.11 (NS/BT) and Nm = 1.04 (NS/SV)) while the closest herds have the highest pairwise migration rates (Nm = 4.28 (NS/CD) and Nm = 3.4 (CD/MM)). A notable exception to this pattern is the low MM/SV effective migration rate (Nm = 1.15). Although these herds are relatively close to one another geographically it is likely that Estes Park and high human population density in the area separating these herds is limiting migration between these herds to some extent.

Because we find high migration rates and low fixation indices across all herds it stands to reason that inbreeding coefficients should be near zero, indicating healthy, non-bottlenecked herds. And indeed, we find that all inbreeding coefficients are also very close to zero (F_{is} = -0.01 - 0.09), and that individual heterozygosity across microsatellite loci is consistently high. Thus the chance that decreased genetic variation could lead to inbreeding depression is low for all herds. Therefore, we find no indication that individual heterozygosity is decreased due to non-random mating within any of these five herds. Due to concern for the Mummy herd specifically, it is worth noting that this herd's inbreeding coefficient was estimated as F_{is} = -.01.

In addition to analyses performed on the full 13 microsatellite loci, we calculated both Nm and F_{is} for two subsets of these data because four of the microsatellite loci used in this study are known to be in gene regions. Thus, there was potential for selection skewing our results. We therefore estimated these parameters using only the 4 gene loci, as well as for the 9 remaining "neutral" loci, separately. Effective migration rates estimated from the 4 gene loci (Nm = 8.97), from the 9 "neutral" loci (Nm = 8.82) and from all 13 loci (Nm = 8.83) are virtually identical and therefore both "neutral" and "non-neutral" markers supported similar rates of migration estimated from microsatellite data.

We also find that the range of inbreeding coefficient estimates for the 4 loci (F_{is} = -0.068 – 0.118) and the 9 loci (F_{is} = 0.005 – 0.164) are both reasonably close to estimates for all 13 loci (F_{is} = -0.01 – 0.09). Although there is variation between the three F_{is} data sets and among the five herds, no directional pattern or cause for concern for any individual herd was detected. In general, these microsatellite data indicate high levels of gene flow and low herd substructure within this metapopulation which is preventing genetic isolation and inbreeding depression in any of these five herds. We also find high individual heterozygosity within all herds and therefore low risk for inbreeding depression in the foreseeable future.

II. ***Mitochondrial Sequence Analyses***

The F_{st} values estimated from microsatellite and mtDNA data (F_{st} = 0.04 and 0.25, respectively) indicate that nuclear DNA markers are much more evenly distributed among herds than are mitochondrial DNA markers. The most parsimonious explanation for this disparity is that rams move more frequently between herds than do ewes. This is because mtDNA is maternally inherited and therefore mtDNA gene flow is due solely to ewe migration while nuclear DNA geneflow is due to both ram and ewe migration.

A closer look at these data reveals additional interesting patterns in haplotype distributions among these herds. The pie diagrams in Figure 8 show that haplotypes 1 and 2 are the only haplotypes detected in the Big Thompson herd. In contrast, the Never Summer range has five haplotypes, but is lacking haplotypes 1 and 2 (Table 10). In other words, these herds have no haplotypes in common, indicating that females have not moved between these populations and bred successfully in recent history. The mtDNA differentiation can be seen more clearly if haplotypes 1 and 2 are pooled into "haplotype E" and all other haplotypes are pooled into "haplotype W", so that the Big Thompson herd is fixed (frequency is 100%) for the E haplotype, and the Never Summer herd is fixed for the W haplotype. If we use Sewall Wright's calculation of F_{st} to compare these populations for these pooled data, then F_{st} = 1.0. But when we use Nei's formulation (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. When the data are pooled into only two alleles, it is also evident that the Never Summer herd and the St. Vrain herds are likewise strongly differentiated.

The complete differentiation of the Never Summer and Big Thompson herds raises the question concerning whether the intervening territory marks an historic boundary to ewe-mediated gene flow or whether the Big Thompson herd was introduced from another geographic area with different mitochondrial haplotypes and gene flow has yet to occur to any great extent.

A closer look at the mtDNA pairwise F_{st} estimates between individual herds (Figure 9) shows a consistent pattern of increasing isolation-by-distance where the most distant herds (BT/NS; F_{st} = 0.44) are the most differentiated and the closest herds (MM/CD; F_{st} = 0.02 and MM/NS; F_{st} = 0.10) are the least differentiated. These data suggest that the Continental Divide mountain range does not provide a significant geographic barrier to gene flow between RMNP bighorn sheep herds. It is also clear from the four pairwise BT comparisons (F_{st} = 0.25-0.44) that the Big Thompson herd is more genetically differentiated from each of the four RMNP herds than any of those herds are to each other (F_{st} = 0.02-0.19). This pattern is not reflected in the microsatellite data (Figure 5) however, where BT pairwise F_{st} values are much lower (F_{st} = 0.02-0.09) and are comparable to those between RMNP herds (F_{st} = 0.03-0.09). This suggests that gene flow between the Big Thompson herd and

the RMNP herds is primarily ram- mediated and that ewes, for the most part, are not migrating between the two regions.

Likewise, mitochondrial data gives a much lower overall effective migration estimate than microsatellite data (Nm = 1.52 and 8.82, respectively). These estimates, like the F_{st} data, indicate that geneflow is likely occurring predominantly due to ram migration while ewes are remaining philopatric to their natal herds. Mitochondrial DNA is maternally inherited and therefore gene flow of mitochondrial markers will occur only if ewes migrate into a new herd and successfully mate, producing lambs that carry their mitochondrial chromosome. This also means that rates of genetic drift for mitochondrial markers will be higher because the effective population (n_e) of these markers is ¼ that of nuclear (autosomal) markers such as microsatellites. This is because for diploid autosomal markers an offspring will receive one of two alleles from each parent with the potential of four possible alleles and up to four possible genotypic combinations. In contrast each offspring will inherit only one mitochondrial marker from its mother and will be haploid for all associated genetic markers. This has important population genetic implications for species such as bighorn sheep where the dispersal of individuals is heavily sex-biased. Therefore sex ratios of introduced sheep should be taken into account for any future transplantation into RMNP herds.

A closer look at the pairwise migration estimates from both mitochondrial and microsatellite markers in Figures 10 and 7 (Nm = 1.27 - 13.14 and 1.04 - 4.28 respectively) shows that rates between individual pairs of herds are actually more similar than the global Nm estimates might indicate. For example: BT/CD: Nm = 2.25 (mtDNA) and 1.91 (Msat), MM/SV: Nm = 1.27 (mtDNA) and 1.15 (Msat) and MM/NS: Nm = 1.80 (mtDNA) and 2.16 (Msat). However, several estimates are quite different, most notably the contrast between estimates for CD/NS: Nm = 13.14 (mtDNA) and 4.28 (Msat). This could be due to shared lambing ground at "the crater" next to Specimen Mountain where herd ranges overlap and ewes from both herds interact each spring and early summer. Because ewes from both herds are more likely to come into contact here it is reasonable that a higher rate of ewe migration (and thus mitochondrial gene flow) is occurring between the Never Summer and Continental Divide herds. This pattern is also reflected in the Msat migration estimate, because although it is a much lower absolute value, it is still the highest pairwise Nm value from that analysis.

III. *SRY Gene Sequence Analyses*

The initial goals of this study included using SRY gene sequence to isolate and estimate ram-mediated gene flow between bighorn sheep herds within this metapopulation. Unfortunately, after extensive sequencing of these populations, we found no genetic variation within either the SRY gene promoter or ORF for Rocky Mountain bighorn sheep. Thus, we conclude that this region is not useful as a population genetic marker and is therefore not suitable for tracking ram-mediated migration in RMNP bighorn sheep herds. This absence of variation is likely due to two primary

factors: 1) evolutionary constraint on the SRY gene because of its vital role in normal male development and 2) because the SRY gene occurs only in males, decreasing its effective population (N_e), and is haploid so there is no recombination during meiosis. Thus, the single copy of this gene in males and absence of the gene in females gives it an effective population size of only ¼ that of normal diploid nuclear genes. In addition, the inability for this gene cluster to recombine further limits its variability.

We therefore expanded the scope of this part of the project to include SRY gene sequence data for all other species in the *Ovis* genus which resulted in the most extensive phylogenetic analyses for this genus to date. We provide the first published SRY ORF sequence data for any of the six wild sheep species as well as the first published promoter sequence for snow sheep. We also identify the first intraspecies SRY gene variation for wild sheep species. These data are potentially useful for increased clarification of phylogenetic relationships between all sheep species.

Phylogenetic analyses of both the SRY ORF and Promoter data sets independently support division of the *Ovis* genus into two distinct clades: Pachyceriforms (bighorn, thinhorn and snow) and Moufloniforms/Argaliforms (musimon, urial, argali and domestic). We find that genetic data clearly support division between these taxonomic groups but are insufficient to resolve division of the Argaliform/Moufloniform groups. This may be because of ecological factors imposing similar evolutionary constraints. While genetic variation within these clades is generally low, our concatenated ORF/Pro data set does provide clear separation of bighorn, thinhorn and snow sheep species within the Pachyceriform clade. However more data are necessary before clear resolution of subspecies within this clade is possible. It is likely that additional data from more variable markers will aid in this endeavor. Each of our data sets separately show incidences of closer grouping between species than within species. Several factors may be contributing to this pattern including gene flow between these species, a geographic component to selection at this gene, or failure of accurate species identification.

In addition to providing support for previous phylogenetic relationships, we provide data for novel intraspecies polymorphism within bighorn, thinhorn and urial sheep. We find no evidence for variation within either argali or mouflon sheep, which is interesting because mouflon sheep have been proposed as the progenitor of all extant sheep species and thus are most likely to exhibit increased genetic variation due to diversification. It is likely that both low SRY gene variation and small sample size are limiting more accurate estimates of genetic diversity for all species. Generally, the lack of resolution within the Moufloniform/Argaliform clade suggests that time since diversification is shorter and that population isolation has remained lower, preventing clear speciation within this clade.

IV. *MHC Gene Sequence Analyses*

It is expected that for a single gene in a diploid individual there should be at most two different alleles. However, our preliminary DRB1 sequencing results gathered from ten different Rocky Mountain bighorn sheep show between three and twenty unique In1/Ex2 sequences within a single individual. We conducted several additional experiments in order to eliminate factors that could potentially contribute to sequence variability from errors during PCR, bacterial cell division, or sequencing. Although we do find evidence of false alleles due to a single or a few point mutations, these additional experiments confirm that our sequencing results do show several distinct DRB1 sequences and therefore provide evidence of DRB1 gene duplication. Our data are supported by two additional lines of evidence. First, past studies have focused exclusively on just the 270bp exon two coding region and would therefore have missed much of the variation we find within the intronic region. Second, previous studies designed their primers based on sequences from cow (*Bos taurus)* or domestic sheep (*Ovis aries)* and thus selective amplification of only a subset of all existing alleles was likely occurring due to SNPs at primer binding sites which prevented identification of certain alleles in bighorn sheep.

Therefore we find that these multiple alleles are in fact the result of past DRB1 gene duplication events and are not artifacts due to errors during the data collection process. Thus, these data suggest that gene duplication has occurred resulting in between two and forty copies of the DRB1 gene within the *Ovis* genome. While gene duplication has been identified in the MHC previously, this is the first evidence of MHC gene duplication in bighorn sheep. This is a novel finding which deserves further study. Unfortunately, because this phenomenon had yet to be reported in the literature we initially designed the project to include less extensive data collection than is clearly necessary based on our preliminary data. Thus we were unable to perform population-level analyses of diversifying selection on the MHC gene DRB1 in these bighorn sheep herds. Data from this project will be available for future work characterizing the genomic structure of this region.

ACKNOWLEDGMENTS

We would like to thank all of the incredible and dedicated people of Rocky Mountain National Park for the support, information and expertise provided during this project. The participation of myriad RMNP employees and volunteers (including the Bighorn Brigade) was vital for successful DNA sample collection. Funding for this project was provided by the National Park Service, Rocky Mountain National Park (J1526075090/UCOB-33), the Rocky Mountain Nature Association, the National Science Foundation (Graduate Research Fellowship 100039091), the Department of Ecology and Evolutionary Biology and the University of Colorado, Boulder. In particular we would like to acknowledge the contributions of Judy Visty, Mary Kay Watry, Janet George, Dr. Rob Roy Ramey III, Dr. John Wehausen, Dr. Robert Baker, Tim Singleton, Mathew Arnold and Jim Driscoll for invaluable information, resources, logistical support, laboratory assistance and field assistance over the past several years. Thank you so much!

LITERATURE CITED

1. Accolla, R.S., Adorini, L., Sartoris, S., Sinigaglia, F. & Guardiola, J. MHC: orchestrating the immune response. *Immunology Today* **16**, 8-11 (1995).

2. Applied Biosystems Genescan Software. (2000).

3. Applied Biosystems Hi-Di ^{TM} Formamide. (2003).

4. Applied Biosystems Automated DNA Sequencing. (2000).

5. Applied Biosystems TAMRA Size Standards.

6. Applied Biosystems Amplitaq Gold PCR Kit. (2007).

7. Avise, J. *Phylogeography: the history of formation of species*. (Harvard University Press: Cambridge, 2000).

8. Ballingall, K.T., Fardoe, K. & McKeever. D. J. Genomic organisation and allelic diversity within coding and non-coding regions of the Ovar-DRB1 locus. *Immunogenetics* **60**, 95-103 (2008).

9. Ballingall, K.T. *et al.* Genetic and proteomic analysis of the MHC class I repertoire from four ovine haplotypes. *Immunogenetics* **60**, 177-84 (2008).

10. Barton, N.H. & Slatkin, M. A quasi-equilibrium theory of the distribution of rare alleles in a subdivided population. *Heredity* **56**, 409-15 (1986).

11. Belov, K. *et al.* Characterization of the opossum immune genome provides insights into the evolution of the mammalian immune system. *Genome research* **17**, 982-91 (2007).

12. Beraldi, D. *et al.* Development of a linkage map and mapping of phenotypic polymorphisms in a free-living population of Soay sheep (Ovis aries). *Genetics* **173**, 1521-37 (2006).

13. Bio-Rad Laboratories Inc. PTC-100 ® Thermal Cycler, 1-88 (2005).

14. Buchanan, F.C. & Crawford, A.M. Ovine microsatelittes at the OarFCB11, OarFCB128, OarFCB193, OarFCB266 and OarFCB304 loci. *Animal Genetics1* **24**, 145 (1993).

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

16. Buchanan, F.C. & Crawford, A.. Ovine dinucleotide repeat polymophism at the MAF33 locus. *Animal Genetics1* **23**, 186 (1992).

17. Buchanan, F.C., Swarbrick, P.A. & Crawford, A.M. Ovine dinucleotide repeat at the MAF48 locus. *Animal Genetics1* **22**, 379-380 (1991).

18. Buitkamp, J., Schwaiger, F.-W. & Epplen, J.-T. Ovis aries T cell receptor gene V-region, exons 1 (3' end) and 2 (5; end). NCBI Accession L18957 (1993).

19. Bunch, T.D., C, W., Zhang, Y.-P. & Wang, S. Phylogenetic analysis of snow sheep (Ovis nivicola) and closely related taxa. *The Journal of heredity* **97**, 21-30 (2006).

20. Corning Incorporated Corning ® Thermowell ® Sealing Mats.

21. Cornuet, J.M. & Luikart, G. Description and power analysis of two tests for detecting recent populatin bottlenecks from allele frequency data. *Genetics* **144**, 2001-2014 (1997).

22. Cowan, I.M. Distribution and Variation in the Native Sheep of North America. *American Midland Naturalist* **24**, 505-580 (2012).

23. Crawford, A.M. *et al.* Genetic Linkage Map of the Sheep Genome. *Genetics* **140**, 703-724 (1995).

24. Deng, W. *et al.* Evolution and migration history of the Chinese population inferred from Chinese Y-chromosome evidence. *Journal of human genetics* **49**, 339-48 (2004).

25. Dukkipati, V.S.R., Blair, H.T., Garrick, D.J. & Murray, A. "OVAR-Mhc" - ovine major histocompatibility complex: structure and gene polymorphisms. *Genetics and Molecular Research* **5**, 581-608 (2006).

26. Ede, A.J., Peirson, H., Henry, H. & Crawford, A.M. Ovine microsatellites at the OarAE64, OarHH22, OarHH56, OarHH62 and OarVH4 loci. *Animal Genetics* **25**, 51-52 (1994).

27. Gaudieri, S., Dawkins, R.L., Habara, K., Kulski, J.K. & Gojobori, T. SNP Profile within the Human Major Histocompatibility Complex Reveals an Extreme and Interrupted Level of Nucleotide Diversity SNP Profile within the Human Major Histocompatibility Complex Reveals an Extreme and Interrupted Level of Nucleotide Diversity. *Genome Research* **10**, 1579-1586 (2000).

28. Geist, V. On the taxonomy of giant sheep (Ovis ammon Linnaeus, 1766). *Canadian Journal of Zoology* **69**, 706-723 (1991).

29. Geist, V. *Mountain Sheep: A Study in Behavior and Evolution*. (University of Chicago Press: Chicago, IL, 1976).

30. Geraldes, a, Rogel-Gaillard, C. & Ferrand, N. High levels of nucleotide diversity in the European rabbit (Oryctolagus cuniculus) SRY gene. *Animal genetics* **36**, 349-51 (2005).

31. Gutierrez-Espeleta, G. a, Hedrick, P.W., Kalinowski, S.T., Garrigan, D. & Boyce, W.M. Is the decline of desert bighorn sheep from infectious disease the result of low MHC variation? *Heredity* **86**, 439-50 (2001).

32. Haltenorth, T. Klassifikation der Saugetiere: Artiodactyla. *Handb. Zoo.* **8**, 1-167 (1963).

33. Hammer, M.F. *et al.* Out of Africa and back again: nested cladistic analysis of human Y chromosome variation. *Molecular biology and evolution* **15**, 427-41 (1998).

34. Hartl, D.L. & Clark, A.G. *Principles of population genetics*. (Sinauer Associates, Inc.: Sunderland, 2007).at <http://tocs.ulb.tu-darmstadt.de/183030028.pdf>

35. Hartyl, D. & Clark, A. *Principles of Population Genetics*. 1-565 (Sinauer Associates, Inc.: Sunderland, MA (2007).

36. Hasegawa, M. Dating of the Human-Ape Splitting by a Molecular Clock of Mitochondrial DNA. *Journal of Molecular Evolution* **22**, 160-174 (1985).

37. Hedrick, P.W. Evolutionary Genetics of the Major Histocompatibility Complex. *The American Naturalist* **143**, 945-964 (1994).

38. Hiendleder, S., Mainz, K., Plante, Y. & Lewalski, H. Analysis of Mitochondrial DNA Indicates That Domestic Sheep Are Derived From Two Different Ancestral Maternal Sources: No Evidence for Contributions From Urial and Argali Sheep. *Journal of Heredity* **89**, 113-120 (1998).

39. Horton, R. *et al.* Variation analysis and gene annotation of eight MHC haplotypes: the MHC Haplotype Project. *Immunogenetics* **60**, 1-18 (2008).

40. Huelsenbeck, J.P. & Ronquist, F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754-755 (2001).

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

42. Invitrogen Corporation One Shot ® MAX Efficiency ® DH5α ™ -T1 Competent Cells. (2004).

43. Kasapidis P & magoulas A Development and application of microsatellite markers to address the population of the horse mackerel Trachurus trachurus. *Fisheries Research* **89**, 132-135 (2008).

44. Kongrit, C. *et al.* Isolation and characterization of dinucleotide microsatellite loci in the Asian elephant (Elaphus maximus). *Molecular Ecology* **8**, 175-177 (2008).

45. L'Abbe, D.L., Belmaaza, A., Francine, D. & Chartrand, P. Elimination of heteroduplex artifacts when sequencing HL:A genes amplified by polymerase chain reacion (PCR). *Immunogenetics* **35**, 395-397 (1992).

46. Landis, E.D., Purcell, M.K., Thorgaard, G.H., Wheeler, P.A. & Hansen, J.D. Transcription profiling MHC class I genes in rainbow trout with infectious hematopoietic necrosis virus. *Molecular Immunology* **45**, 1646-1657 (2008).

47. Librado, P. & Rozas, J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451-1452 (2009).

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

49. Longeri, M., Zanotti, M. & Damiani, G. Recombinant DRB sequences produced by mismatch repair of heteroduplexes during cloning in Escherichia coli. *European journal of immunogenetics: official journal of the British Society for Histocompatibility and Immunogenetics* **29**, 517-523 (2002).

50. Lucy, M.C., Johnson, G.S., Shibuya, H., Boyd, C.K. & Herring, W.O. Rapid communication: polymorphic (GT)n microsatellite in the bovine somatotropin receptor gene promoter T. *Journal of Animal Science* **76**, 2209-2210 (1998).

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

52. Lukas, D. *et al.* Major histocompatibility complex and microsatellite variation in two populations of wild gorillas. *Molecular Ecology* **13**, 3389-3402 (2004).

53. Luo, S.-J. *et al.* Development of Y chromosome intraspecific polymorphic markers in the Felidae. *The Journal of heredity* **98**, 400-13 (2007).

54. Maddox, J.F. Mapping the matrix metalloproteinase 9 (MMP9) gene and the BL1071 microsatellite to ovine chromosome 13 (Oar13). *Animal Genetics* **32**, 316-331 (2001).

55. Marchal, J. a, Acosta, M.J., Bullejos, M., Díaz de la Guardia, R. & Sánchez, A. Origin and spread of the SRY gene on the X and Y chromosomes of the rodent Microtus cabrerae: role of L1 elements. *Genomics* **91**, 142-51 (2008).

56. Markow, T. *et al.* HLA polymorphism in the Havasupai: evidence for balancing selection. *American journal of human genetics* **53**, 943-52 (1993).

57. Maudet, C., Luikart, G. & Taberlet, P. Developpment of microsatellite multiplexes for wild goats using primers designed from domestic Bovidae. *Genetics Selection Evolution* **33**, S193-S203 (2001).

58. Maudet, C. *et al.* A standard set of polymorphic microsatellites for threatened mountain ungulates (Caprini, Artiodactyla). *Moleculary Ecology Notes* **4**, 49-55 (2004).

59. McClintock, B.T. & White, G.C. Bighorn Sheep Abundance Following a Suspected Pneumonia Epidemic in Rocky Mountain National Park. *The Journal of Wildlife Management* **71**, 183-189 (2007).

60. Meadows, J.R.S. *et al.* Globally dispersed Y chromosomal haplotypes in wild and domestic sheep. *Animal Genetics* **37**, 444-453 (2006).

61. Meadows, J.R.S., Hawken, R.J. & Kijas, J.W. Nucleotide diversity on the ovine Y chromosome. *Animal Genetics* **35**, 379-385 (2004).

62. Meadows, J.R.S., Hiendleder, S. & Kijas, J.W. Haplogroup relationships between domestic and wild sheep resolved using a mitogenome panel. *Heredity* **106**, 700-706 (2011).

63. Meadows, J.R.S. *et al.* Mitochondrial sequence reveals high levels of gene flow between breeds of domestic sheep from Asia and Europe. *Journal of Heredity* **96**, 494-501 (2005).

64. Mona, S. *et al.* Disentangling the effects of recombination, selection, and demography on the genetic variation at a major histocompatibility complex class II gene in the alpine chamois. *Molecular Ecology* **17**, 4053-4067 (2008).

65. Payen, E.J. & Cotinot, C.Y. Sequence evolution of SRY gene within Bovidae family. *Mammalian Genome: Official Journal of the International Mammalian Genome Society* **5**, 723-725 (1994).

66. Payen, E.J. & Cotinot, C.Y. Comparitive HMG-box sequences of the SRY gene between sheep, cattle and goats. *Mammalian Genome: Official Journal of the International Mammalian Genome Society* **21**, 2772 (1993).

67. Penty, J.M., Henry, H.M., Ede, A.J. & Crawford, A.M. Ovis aries nucleotide repeat polymorphism. NCBI Accession L11047 (1993).

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

69. Posada, D. jModelTest: phylogenetic model averaging. *Molecular biology and evolution* **25**, 1253-1256 (2008).

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

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

72. Promega Corporation Blue Dextran Loading Solution DV435. (2012).

73. Qiagen Corp. Type-it ® Microsatellite PCR Handbook. (2009).

74. Qiagen Corporation QIAprep ® Miniprep QIAamp DNA Stool Minikit. (2009).

75. Rambaut, A. FigTree v. 1.3.1. (2006).at <http://tree.bio.ed.ac.uk/software/figtree/ >

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

77. Reading, R.P., Amgalanbaatar, S. & Wingard, G. The Open Country. *The Open Country* 25-32 (2001).

78. Réale, D. & Festa-Bianchet, M. Quantitative genetics of life-history traits in a long-lived wild mammal. *Heredity* **85**, 593-603 (2000).

79. Rezaei, H.R. *et al.* Evolution and taxonomy of the wild species of the genus Ovis (Mammalia, Artiodactyla, Bovidae). *Molecular phylogenetics and evolution* **54**, 315-26 (2010).

80. Rosser, Z.H. *et al.* Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. *American journal of human genetics* **67**, 1526-43 (2000).

81. Rousset, F. GENEPOP '007: A complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources* **8**, 103-106 (2008).

82. Sayers, G. *et al.* Major Histocompatibility Complex DRB1 gene: its role in nematode resistance in Suffolk and Texel sheep breeds. *Parasitology* **131**, 403-409 (2005).

83. Schaschl, H., Wandeler, P., Suchentrunk, F., Obexer-Ruff, G. & Goodman, S.J. Selection and recombination drive the evolution of MHC class II DRB diversity in ungulates. *Heredity* **97**, 427-37 (2006).

84. Schaschl, H., Goodman, S.J. & Suchentrunk, F. Sequence analysis of the MHC class II DRB alleles in Alpine chamois (Rupicapra r. rupicapra). *Developmental & Comparative Immunology* **28**, 265-277 (2004).

85. Shakleton, DM Lovari, S. Classification Adopted for the Caprinae Survey. *Sheep and Goats and their Relatives: Status survey and conservation action plan for Caprinae* 9-14 (1997).

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

87. Swofford, D. *PAUP* Phylogenetic Analysis Using Parsimony (*and Other Methods)*. Sinauer Associates: Sunderland, MA (2000).

88. Taberlet, P., Coissac, E., Pansu, J. & Pompanon, F. Conservation genetics of cattle, sheep, and goats. *Comptes Rendus Biologies* **334**, 247-254 (2011).

89. Tapio, I. *et al.* Unfolding of population structure in Baltic sheep breeds using microsatellite analysis. *Heredity* **94**, 448-456 (2005).

90. Tsalkin, V.I. *Wild Mountain Sheep of Europe and Asia*. Moscow Society of Naturalists: Moscow, Russia (1951).

91. Tserenbataa, T., Ramey, R.R., Ryder, O.A., Quinn, T.W. & Reading, R.P. A population genetic comparison of argali sheep (Ovis ammon) in Mongolia using the ND5 gene of mitochondrial DNA; implications for conservation. *Molecular Ecology* **13**, 1333-1339 (2004).

92. Valdez, R. & Batten, J. *the Wild Sheep of the World*. Wild Sheep and Goat International: Mesilla, NM 1-186 (1982).

93. Valdez, R. & Krausman, P.R. Mountain Sheep of North America. *Mountain Sheep of North American* 1-3 (1982).

94. Vial, L., Maudet, C. & Luikart, G. Thirty-four polymorphic microsatellites for European roe deer. *Molecular Ecology* **3**, 523-527 (2003).

95. Walser, B. & Heckel, G. Microsatellite markers for the common vole (Microtus arvalis) and their cross-species utility. *Conservation Genetics* **9**, 479-481 (2008).

96. Waters, P.D., Wallis, M.C. & Marshall Graves, J. a Mammalian sex--Origin and evolution of the Y chromosome and SRY. *Seminars in Cell & Developmental Biology* **18**, 389-400 (2007).

97. Wehausen, J.D., Ramey, R.R. & Epps, C.W. Experiments in DNA extraction and PCR amplification from bighorn sheep feces: the importance of DNA extraction method. *The Journal of Heredity* **95**, 503-509 (2004).

98. Wigley, P. Genetic resistance to Salmonella infection in domestic animals. *Research in Veterinary Science* **76**, 165-169 (2004).

99. Wilson, D.E. & Reeder, D.M. *Mammal Species of the World: A Taxonomic and Geographic Reference*. The Johns Hopkins University Press: Baltimore, MD (2005).

100. Worley, K., Carey, J., Veitch, a & Coltman, D.W. Detecting the signature of selection on immune genes in highly structured populations of wild sheep (Ovis dalli). *Molecular Ecology* **15**, 623-37 (2006).

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

102. Wright, S. Evolution in mendelian populations. *Genetics* **16**, 97-159 (1931).