UNCOVERING THE GENOMIC SIGNATURES OF POPULATION STRUCTURE AND INTROGRESSION IN AMERICAN BISON

A Dissertation

by

DAVID FORGACS

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	James N. Derr
Committee Members,	Terje Raudsepp
	Penny K. Riggs
	David W. Threadgill
Intercollegiate Faculty Chair,	David W. Threadgill

December 2019

Major Subject: Genetics

ABSTRACT

Much of wildlife research is plagued by outdated molecular methods and bison studies are no exception. The aim of this work is to push the boundaries of population genetics in bison and to update and build upon previously existing techniques to provide a more accurate assessment of population health, genetic diversity, and the level of domestic cattle introgression in the bison genome. I also aimed to develop protocols and analyses that would swiftly translate to other wildlife species. To that effect, I looked at several different genomic markers, including the whole mitochondrial genome, microsatellites, as well as genome-wide SNPs.

A great deal of my research revolved around the single most important US bison herd at Yellowstone National Park. Using whole mitochondrial sequences from 25 Yellowstone bison, the levels of genetic diversity and population subdivision were established, and constructing a large dataset of 65 bison from all across North America helped reveal the signatures of the most recent historic bottleneck and how it has affected the herd.

In order to conduct future experiments with a less constrained and more random way of sampling bison, the quality of DNA from fecal samples was also assessed by directly comparing them to paired blood samples from the same 50 Yellowstone bison. In addition to establishing a set of 15 markers that performed reliably with bison fecal samples, we have also shown that there is a statistically significant reduction in heterozygosity estimates from fecal samples if the paired sample method is not

implemented during marker selection. This discrepancy is often overlooked by previous researchers, causing them to present data that underestimate the genetic diversity in various wildlife species.

Indubitably, the biggest threat facing bison genomic integrity today is domestic cattle introgression that has left the majority of bison today riddled with domestic cattle chromosomal segments. I set out to create a SNP-based in-depth genotyping assay to differentiate cattle- and bison-like segments with high resolution. To that effect, I analyzed 40 bison from all across North America and found that the least introgressed of the herds appears to be the one at Santa Catalina Island, a herd with known high levels of mitochondrial albeit low levels of nuclear cattle introgression. Thus, no herd could be identified to date that did not have a recent cattle influence, including herds that were previously thought to be non-introgressed, such as the herds at Yellowstone National Park, Wind Cave National Park, and the plains bison of Canada. The likely reason for introgression in many of these herds can be traced back to three male bison that were introduced into Yellowstone National Park in 1902. In addition, multiple historical hybridization events between plains and wood bison were uncovered, some of which occurred in the United States.

DEDICATION

First and foremost, I would like to dedicate my dissertation to my mom, Judit Langermann, who is absolutely the strongest woman I know. I have learned unspeakable volumes from her perseverance and grit. I would also like to dedicate it to my wife, Courtney Caster, who has stood by my side and always encouraged me to work harder and kept my spirits up all throughout my Ph.D. work. Thank you to the rest of my family: my sister, Noémi Forgács, who always kept me in check, my dad, György Forgács who got a PhD at a time when dissertations had to be written by typewriter, my grandparents, Édi & István and Ági & Tamás, and my great aunt Jutka, who have always supported me throughout my studies. I would also like to thank my second family, Teresa and Stephen, and all my friends for their unceasing support. I am truly lucky to know you all. I owe you all EVERYTHING. Oh, and of course Ellie, Quinn, and Eowyn for being the most therapeutic cat royalties that have ever meowed.

ACKNOWLEDGEMENTS

I owe enormous gratitude to my committee chair, Dr. Derr, who has taken me on as a graduate student and supported my development for all these years. I would also like to thank the rest of my committee, Dr. Raudsepp, Dr. Riggs, and Dr. Threadgill for their guidance, mentorship, and support throughout the course of my Ph.D., as well as for taking an interest in my research. I owe enormous gratitude to Rick Wallen and Brian Davis, two of the best collaborators I could ever ask for.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a dissertation committee consisting of James N. Derr of the Department of Veterinary Pathobiology, Terje Raudsepp of the Department of Veterinary Integrative Biosciences, Penny K. Riggs of the Department of Animal Science, and David W. Threadgill of the Department of Molecular & Cellular Medicine.

The samples analyzed for Chapter II were provided by Rick L. Wallen and P.J. White of the Yellowstone National Park Service, and Min X. Cai and Lauren K. Dobson of the Department of Veterinary Pathobiology at Texas A&M University. The telemetry data was also collected by Rick Wallen. The samples analyzed in Chapter III were provided by Rick Wallen and P.J. White, and some of the data collection was supervised by Amy Boedeker of the Department of Veterinary Pathobiology at Texas A&M University. The samples analyzed for Chapter IV were provided by the Department of the Interior, National Park Service, the Canadian Museum of Nature, and the Turner Foundation. A considerable number of the scripts were written or edited by Brian Davis of the Department of Veterinary Integrative Biosciences at Texas A&M University, and run on Texas A&M's High Performance Research Computing platform Ada. The previously unpublished samples were prepared, sequenced, and curated by Genome Quebec with contributions by Paul Stothard and Tianfu Yang from the University of Alberta. Sam Stroupe has been instrumental in the historical research and primary literature analysis. Caitlin Curry helped with the graphic design on Figure IV.9. The

samples analyzed in Appendix E were provided by Wind Cave National Park and The Nature Conservancy. All other work conducted for this dissertation was completed by me independently.

Funding Sources

This research was made possible in part by the Department of the Interior, National Park Service under Grant Number P12AC71337 (formerly P12AT51121). Additional funding came from The Nature Conservancy and Turner Enterprises, Inc. The library preparation and sequencing effort was funded by the Alberta Agriculture and Forestry Strategic Research and Development Grant, the Agriculture and Agri-Food Canada: Canadian Agricultural Adaption Program, Mitacs Accelerate, and the CanSeq150 Program. Graduate study was supported in part by the Texas A&M Genetics Fellow Scholarship.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION*

American bison (*Bison bison*) is one of the few remnants of the megafauna that was abundant until the last Ice Age around 15 thousand years ago (Cooper et al. 2015). It is the largest mammal of the Americas, measuring up to 6.5 feet at the shoulder and weighing up to 2,000 lbs. The common ancestor of bison and the domestic cattle (*Bos taurus*) dates back to approximately 1 million years ago (Douglas et al. 2011), when the lineages started to diverge in Eurasia, most likely in modern-day India and China (Halbert 2003, Marsolier-Kergoat 2015). During the middle Pleistocene, some bison migrated west to occupy Europe and formed two species: the woodland bison (*Bison schoetensacki*) and the steppe bison (*B. priscus*) which later gave rise to modern European bison, also known as the wisent (*B. bonasus*) (Marsolier-Kergoat 2015). Some of these steppe bison also migrated east and crossed the Bering strait to reach the Americas approximately 300-130 thousand years ago (Marsolier-Kergoat 2015) and occupied the plains of North America as far south as Mexico. Today, there are two recognized North American subspecies, the plains bison (*B. bison bison*) which is currently restricted to protected federal lands and private ranches in the United States

* Parts of this chapter were reprinted with permission from "Mitochondrial genome analysis reveals historical lineages in Yellowstone bison" by David Forgacs, Rick L. Wallen, Lauren K. Dobson, and James N. Derr, 2016. PLoS ONE 11(11):e0166081 and "Evaluation of fecal samples as a valid source of DNA by comparing paired blood and fecal samples from American bison (*Bison bison*)" by David Forgacs, Rick L. Wallen, Amy L. Boedeker, and James N. Derr, 2019. BMC Genet 20(22): s12863-019-0722-3.

and Canada, as well as the wood bison (*B. bison athabascae*), which occur in a few national parks and sanctuaries across Canada.

The number of bison plummeted drastically shortly after European settlement in North America due to hunting and habitat destruction. In the 16th century, the number of bison was estimated to be around 25-30 million which declined to fewer than a hundred bison by the early 1900s (Department of the Interior 1902, Taylor 2011). Coder (1975) gives a detailed history of the few remarkable individuals who set out to save bison from certain extinction at the end of the 19th century. James McKay and William Alloway started capturing bison calves and having them raised by domestic cattle heifers in Canada's Manitoba province. Charles Goodnight and his wife Mary in Texas were solely responsible for saving the southern plains bison, a putative subspecies that is not recognized anymore due to high levels of gene flow throughout the 20th century. Walking Coyote of the Pend d'Oreille tribe in Montana initially captured some bison calves as a peace offering to his tribe after he was excommunicated due to his practices of polygamy and intertribal marriage but eventually ended up keeping most of them. Michel Pablo and Charles Allard later bought his animals and further grew the herd on a reservation where the herd was growing steadily, largely unmanaged. Frederick Dupree was a French-Canadian fur trapper who captured bison in Montana and around the Yellowstone region and raised them on a ranch in South Dakota. In Kansas, Charles "Buffalo" Jones, hunter-turned-rancher, assembled the largest bison herd at the time, comprised of 56 animals (Coder 1975).

However, the efforts of these individuals to rescue bison were marred by forcible bison-cattle hybridization efforts in an attempt to create larger, sturdier and more cold-tolerant cattle (Coder 1975). Though several contemporary zoologists, including William T. Hornaday spoke out against the practice, unfortunately, the McKay-Alloway herd, the Goodnight herd, the Dupree herd, and the Jones herd were all experimenting with crossing bison with domestic cattle, producing several hybrids (Hornaday 1908, Coder 1975). Goodnight even published a manuscript in the early days of the Journal of Heredity about the ins and outs of bison-cattle hybridization (Goodnight 1914). The only ranchers at the time who were mindful of hybridization were Pablo and Allard who never attempted hybridization on their own, and even kept hybrids that they acquired from Charles Jones' herd separated to avoid unintentional crossbreeding (Department of the Interior 1902, Coder 1975).

Hybridization between bison and domestic cattle, which results in cattle genetic introgression into the bison lineage, is an anthropogenically mediated process that has never been reported to occur in the wild (Steklenev and Yasinetskaya 1982). Cattle ranchers of the late 19th century pioneered a technique, generally involving a male bison calf separated from its mother at an early age, raised by a cattle heifer. Such a calf would readily mate with cattle heifers once it has reached maturity (Jones 1907, Goodnight 1914). The reciprocal cross, breeding female bison with domestic cattle bulls was rarely successful (Goodnight 1914, Steklenev and Yasintskaya 1982). The vast majority of male first- or second-generation hybrids were born sterile consistent with Haldane's rule which states that the heterogametic sex is often inviable during hybridization (Haldane

1922). Female first-generation hybrids, however, readily reproduced and produced fertile offspring of their own. Human-mediated hybridization has been largely abandoned since the early 20th century, but due to the extremely small population size of bison at the time, most of the 500,000 bison today are the descendants of cattle-bison hybrids. While the amount of cattle genetic material is significantly reduced with each generation of backcrossing to bison, the mitochondrial genome which is uniparentally inherited on the maternal lineage does not experience a similar dilution effect and stays purely cattle-like even today. Nuclear-mitochondrial incompatibility – the condition when the cooperation and interaction between the nuclear and the mitochondrial genomes within the same cell is stymied because they were evolved in different species – causes reduced fertility and a decreased efficiency in energy conversion in a host of recently introgressed species (Johnson 2010, Montooth et al. 2010, Ma et al. 2016). Similarly, in bison-cattle hybrids, though the hybrids look indistinguishable from bison after as few as 4-5 generations of back-crossing, bison-cattle hybrids tend to grow up to be smaller and weigh less than non-introgressed members of their species (Baynes 1909, Derr et al. 2012). In an experiment where bison, cattle, and hybrid cell cultures were tested where the cells were deprived of necessary nutrients, the cultures with cattle mitochondria and bison nuclear genome performed significantly worse than either non-introgressed culture (Derr, unpublished). This led to the hypothesis that nuclear-mitochondrial incompatibility causes a lower efficiency of energy conversion, due to the many ways the nucleus and the mitochondria in a cell collaborate to perform cellular respiration and essential metabolic functions.

In order to assess the proportion of bison whose DNA contain cattle genomic regions, Ward et al. (1999) implemented a mitochondrial genotyping test amplifying a highly conserved region in the 16S rRNA gene that is identical in bison and cattle, and a second, informative marker in the D-loop region that is present in cattle but absent in bison (TPW) (Ward et al. 1999). Testing over 10,000 bison from 150 different US herds in the early 2000s, 3.7% of bison tested positive for cattle mitochondria. Nearly all private herds had signs of cattle introgression, and the only federal herds that tested negative for cattle mitochondria were Yellowstone National Park in Wyoming, Wind Cave National Park in South Dakota, Wichita Mountains Wildlife Refuge in Oklahoma, and Fort Niobrara National Wildlife Refuge in Nebraska. Canadian wood and plains bison herds on federal lands were also void of any signs of mitochondrial introgression. Nuclear microsatellite testing was developed by Schnabel (2001) and Halbert (2003) for the purposes of the detection of nuclear cattle introgression and parentage testing. While the 14 microsatellite loci that had specific bison and cattle alleles in order to assess a bison's nuclear introgression status was rudimentary due to the low number of microsatellites covering tiny fragments across fewer than half of the chromosomes, they provided a deeper snapshot into herd status than before. Based on the results of the nuclear microsatellite testing, even more animals and herds were shown to be introgressed than previously thought, including the Fort Niobrara herd.

However, some herds remained pristine and free from any measurable cattle introgression so far. The most important of these is without a doubt the Yellowstone National Park herd, which is the largest US herd, comprising over 5,000 bison at times.

Though the Yellowstone herd started from only 22 bison in the beginning of the 20th century, strict protection of the animals and the land helped make it a true conservation success story (Coder 1975). The terrain of Yellowstone is mountainous and poor in vegetation compared to the tallgrass prairie which is the natural landscape for bison. While bison have continuously lived in the Yellowstone area since before European settlement, recently they have been migrating outside the park to the north into Montana and to the west into Idaho. With the increased population and human-bison conflict, bison were not welcome outside of Yellowstone National Park any longer (Montana v. US, Case No. CV95-6-H-CCL). This is due to the perceived threat of damage bison can cause to property and people, as well as the risk of disease transmission, especially of brucellosis from bison to the many cattle ranches vital to the region's economy. *Brucella abortus* is a Gram-negative coccobacillus and the zoonotic disease agent of brucellosis. Brucellosis causes fever, chills, aches, and spontaneous abortions. They spread through contact with an aborted fetus, as well as other bodily fluids. Humans can also contract brucellosis by drinking unpasteurized milk. Vaccines have been developed for cattle, but for reasons unknown, they are inefficient and unreliable in bison. While no known case of bison-to-cattle transmission of brucellosis has ever been documented, livestock owners are nonetheless worried as more than two thirds of adult female Yellowstone bison carry the disease (Blanton et al. 2015). Quarantines have proven inefficient in the past as even bison that have tested consistently negative through serological and blood testing for over 6 months have been known to latently express and spread the disease (Hobbs et al. 2015). Thus, the Interagency Bison Management Plan (IBMP) was

formulated, which through hazing and population control/culling keeps the bison from leaving Yellowstone National Park. As a result, on the northern boundary of the park, close to 1,000 bison are killed annually. Despite the culling and loss of calves to brucellosis, the Yellowstone bison population is thriving and the number of births outnumber the mortality rate.

A second genetically important herd can be found in Wind Cave National Park in South Dakota. Just like the Yellowstone herd where some of their founders originated from, these bison have never tested positive for cattle nuclear or mitochondrial alleles using the current genetic tests, and they are free of brucellosis. The National Park Service has partnered with The Nature Conservancy to establish several satellite herds across multiple midwestern states in order to increase the number of Wind Cave lineage bison without the restriction of the limited carrying capacity of Wind Cave National Park of around 350-400 bison (Appendix E) (Halbert and Derr, 2008).

This dissertation is split into 3 main chapters. Chapter II focuses on the Yellowstone bison herd with the aim of informing genetic data-based management decisions in the face of intense population management. As the annual reduction of the Yellowstone bison population is a continued necessity, understanding more about the genetic background and population structure of these animals is crucial to avoid the unnecessary loss of genetic diversity. The analysis of the mitochondrial genomes of 25 Yellowstone bison has revealed high haplotype diversity and no evidence for population subdivision. The population reached its nadir in 1902, when a survey found only 22 indigenous bison remaining in the central region of Yellowstone National Park. In order

to save the herd from extinction, managers reintroduced bison to the Lamar Valley in the northern region of the park, including 3 bulls from the Goodnight herd in Texas and 18 females from the Pablo-Allard herd in northern Montana (Coder 1975, Ogilvie 1979). The indigenous and introduced herds were united in 1915 and the two herds were allowed to comingle ever since. Comparing these 25 Yellowstone bison to the mitochondrial sequences of bison from across North America revealed two distinct Yellowstone lineages that can be traced back to 1902 when the introduced herd was brought into the park to supplement the last few remaining indigenous bison. This relatively large data set opens up many lines of inquiry, and we provide evidence supporting hybridization between the two North American bison subspecies. Based on Douglas et al. (2011)'s published sequences, Pringle (2011) proposed that two non-synonymous point mutations in bison mitochondrial DNA cause significant impairment of mitochondrial oxidative phosphorylation (IMOP). However, his results were solely based on in silico analysis and no phenotype has ever been associated with the mutations to substantiate the detrimental effect of the IMOP mutations. We set out to i) validate the Pringle hypothesis based on historical data, ii) map out the mitochondrial haplotypes in Yellowstone, in relation to other North American bison populations, and iii) estimate genetic diversity and infer population structure from the genetic data.

Chapter III focuses on the direct comparison of fecal and blood samples from the same individuals to determine any discrepancies in the efficiencies of the markers. The collection and analysis of fecal DNA is a common practice, especially when dealing with wildlife species that are difficult to track or capture. While fecal DNA is known to

be lower quality than traditional sources of DNA, such as blood or other tissues, few investigations have verified fecal samples as a valid source of DNA by directly comparing the results to high quality DNA samples from the same individuals. Our goal was to compare DNA from fecal and blood samples from the same 50 American plains bison (*Bison bison*) from Yellowstone National Park, analyze 35 short tandem repeat (STR) loci for genotyping efficiency, and compare heterozygosity estimates. We discovered that some of the fecal-derived genotypes obtained were significantly different from the blood-derived genotypes from the same bison. We also found that fecal-derived DNA samples often underestimated heterozygosity values, in some cases by over 20%. These findings highlight a potential shortcoming inherent in previous wildlife studies that relied solely on a multi-tube approach, using exclusively low-quality fecal DNA samples with no quality control to account for false alleles and allelic dropout. Herein, we present a rigorous marker selection protocol that is applicable for a wide range of species and report a set of 15 STR markers for use in future bison studies that yielded consistent results from both fecal and blood-derived DNA (Forgacs et al. 2019).

Chapter IV focuses on an examination of the bison genome based on a genome-wide SNP analysis. The use of microsatellites in parentage and introgression testing has been overtaken by SNP genotyping and genotyping by sequencing in recent years. Currently, the only technology available for genetic testing in bison is through fluorescent microsatellite genotyping. The bison genotyping test that is currently offered as a service only uses a handful of nuclear microsatellites and a single mitochondrial marker. This test is not sufficiently robust to provide an in-depth analysis of

introgression status in bison-cattle hybrids that have been backcrossing to bison for the last century or even longer. Following the whole genome sequencing of 40 plains and wood bison and 21 domestic cattle from select breeds, we have investigated cattle introgression in 6 major North American bison herds. Based on results from phylogenetic trees, ABBA-BABA tests, structure analyses, and PCAs, we have identified a number of early generation hybrids that show extremely high levels of cattle introgression, but surprisingly, the herd with the least cattle introgression appeared to be the Santa Catalina Island bison herd, a herd where a history of cattle introgression is well documented. Thus, it seems like all the bison herds tested – including those without any previous molecular evidence for introgression, such as Yellowstone and Wind Cave National Parks – have experienced some level of recent domestic cattle introgression. Without any herds that we have analyzed so far that lack cattle introgression, the creation of the genetic test for cattle introgression is unfeasible and largely academic in nature.

However, other insights were also gained from this dataset. It documents the pervasiveness of the Yellowstone lineage across all modern plains and wood bison herds, due to a high level of hybridization between the subspecies, and in fact gives some credence to early reports by numerous zoologists who claimed that the "mountain bison" described along the Rocky Mountains and even in Yellowstone National Park are in fact wood bison. Thus, the Yellowstone herd today could be made up of the indigenous (wood) bison and the introduced (plains) bison, forming the two mitochondrial clades still visible today and described in Chapter II.

Early historical records are sparse, incomplete, hard to come by, and often inaccurate. In my research, I only used statements that are backed by multiple independently verifiable sources. While some of the explanations herein may not stand the test of time, they are based on a vast amount of molecular genetic data that was interrogated using numerous different independent tests.

CHAPTER II

MITOCHONDRIAL GENOME ANALYSIS REVEALS HISTORICAL LINEAGES IN YELLOWSTONE BISON*

Introduction

One of the most iconic species living in Yellowstone National Park is the American plains bison (*Bison bison bison*). American bison survived multiple historic and recent population bottlenecks due to habitat reduction, commercial hunting, and diseases from imported domestic livestock (Isenberg 2000). Populations undergoing major reductions in size with constrained areas of distribution are vulnerable to the effects of inbreeding and the loss of genetic diversity through genetic drift (Hitchings and Beebee 1998, Furlan et al. 2012).

Yellowstone bison have existed on the same landscape for hundreds of years and there is no evidence of domestic cattle introgression (Cahalane 1944, Meagher 1973, Halbert and Derr 2007). The population reached its nadir in 1902, with as few as 22 indigenous animals remaining in the central area of the Park. As a result, managers reintroduced bison to the Lamar Valley in the northern region of the National Park, including 18 females from the Pablo-Allard herd in northern Montana, three bison bulls from the Goodnight herd in Texas, and three calves from the indigenous bison in central

* Parts of this chapter were reprinted with permission from "Mitochondrial genome analysis reveals historical lineages in Yellowstone bison" by David Forgacs, Rick L. Wallen, Lauren K. Dobson, and James N. Derr, 2016. PLoS ONE 11(11):e0166081.

Yellowstone (Meagher 1973, Coder 1975). The indigenous and introduced herds began commingling in 1915 and have intermixed seasonally to some extent ever since (Meagher 1973).

Today, the Yellowstone bison population occupies approximately 1 million acres of suitable habitat near the headwaters of the Yellowstone and Madison River watersheds (Plumb et al. 2009). The core habitat available to Yellowstone bison is protected by the management boundaries and conservation policies of the National Park Service. Additional suitable habitat for bison extends outside the park into Montana, but only constitutes less than 10% of the total conservation area (Blanton et al. 2015). Bison coexist with a full suite of native ungulates and predators, exposing them to competition for food, predation, and survival under substantial environmental extremes. Thus, Yellowstone bison have likely retained adaptive capabilities that may be diminished in other bison populations across North America that are managed like livestock (White and Wallen 2012).

Halbert et al. (2012) evaluated 46 nuclear microsatellite loci from Yellowstone bison and found evidence of moderately high level of genetic diversity (0.626) and gene patterns indicating the existence of at least two subpopulations (the northern and the central herds) with limited gene flow between them. However, not much work has been conducted to describe genetic diversity based on the mitochondrial genome. Mitochondrial DNA (mtDNA) analyses provide insight into how historical events shaped and influenced population genetic diversity without the complicating issues of diploidy and recombination, inherent with the nuclear genome. Mitochondrial haplotype

diversity is a valuable indicator of population health because mtDNA codes for genes that play a crucial part in ribosomal activity, cellular respiration, and energy production. The mitochondrial genome contains 13 protein-coding genes, as well as genes that code for the small and large rRNA subunits (12S and 16S respectively), and tRNAs. The mitochondrial genome is haploid and inherited only through the maternal lineage, making it easier to track populations without having to account for heterozygotes. In addition, mtDNA is more sensitive to inbreeding, loss of diversity, and genetic drift because only one parent plays a role in its transmission.

Ward et al. (1999) analyzed 53 bison from across North America that had no evidence of cattle mtDNA and described eight unique bison haplotypes based on partial D-loop sequences from the mitochondrial genome. Analyzing DNA sequences from this highly variable 600 base pair region, the authors reported two haplotypes (which they named haplotypes 6 and 8) from the five Yellowstone bison analyzed. Gardipee (2007) collected DNA samples from 153 Yellowstone bison and developed a method to distinguish between the two haplotypes previously described by Ward et al. (1999) by sequencing a 470 base pair section of the D-loop control region.

Analyzing the complete mitochondrial genome, Douglas et al. (2011) found 17 unique mtDNA haplotypes during a broad survey of plains bison (*B. b. bison*) and wood bison (*B. b. athabascae*). Wood bison are phenotypically distinct from plains bison and historically limited to Canada and the State of Alaska. Most of the 17 haplotypes came from animals in private herds, which have largely undocumented histories and cannot be traced back to a particular lineage. Notable exceptions are bHap2 which includes a bison

at the National Bison Range in Montana, bHap10 includes a Fort Niobrara National Wildlife Refuge bison from Nebraska, bHap17 is from a Yellowstone National Park bison, bHap13 and 16 from the Caprock Canyon State Park in Texas, and wHap14 and 15 from wood bison at Elk Island National Park in Canada.

Based on the published sequences from Douglas et al. (2011), Pringle (2011) proposed that two non-synonymous point mutations in bison mitochondrial DNA cause significant impairment of mitochondrial oxidative phosphorylation (IMOP). One of these mutations causes an isoleucine to asparagine amino acid change in the ATP6 gene while the other is a valine to alanine change in the cytochrome b gene (Appendix A). His conclusions were deduced solely from comparative *in silico* analysis of homologous sequences in other mammals such as dogs and humans where similar mutations are known to cause a mitochondrial disease (Li et al. 2006). To our knowledge, no phenotype has ever been described to substantiate the detrimental effect of the IMOP mutations in bison.

The objective of our research was to better characterize and understand haplotype frequencies in Yellowstone bison. Previous attempts to delineate mitochondrial haplotype diversity in bison took a much broader approach, analyzing only a few bison for a single location across the United States, which likely resulted in significant local diversity going undetected. We evaluated the amount of genetic diversity in mtDNA in Yellowstone bison and developed a molecular method to test for differentiation between the two primary breeding herds (the northern and central herds). In addition, we assessed the overall genetic health of Yellowstone bison and analyzed the allegedly detrimental

IMOP mutations to identify potential selective differences between bison that express IMOP mutations and bison that are wild type.

Results

Ten different haplotypes were found in the 25 modern samples from Yellowstone bison (Table II.1, Figure II.1). Seven bison belonged in YNPHap1 and ten to YNPHap2. The rest of the haplotypes were unique to only a single animal sequenced in this study.

Haplotype	Northern herd	Central herd	Unknown
YNPHap1	4	3	
YNPHap2	4	4	2
YNPHap3	0	1	
YNPHap4	1	0	
YNPHap5	1	0	
YNPHap6	0	1	
YNPHap7	1	0	
YNPHap8	0	1	
bHap17	0	1	
Templeton	0	0	1
Total	11	11	3

Table II.1: Mitochondrial haplotype distribution in 25 bison associated with the northern or central herds in Yellowstone National Park. Three samples were collected from bison near the north boundary at a capture facility, but their movement histories are unknown. Reprinted with permission from Forgacs et al. 2016.

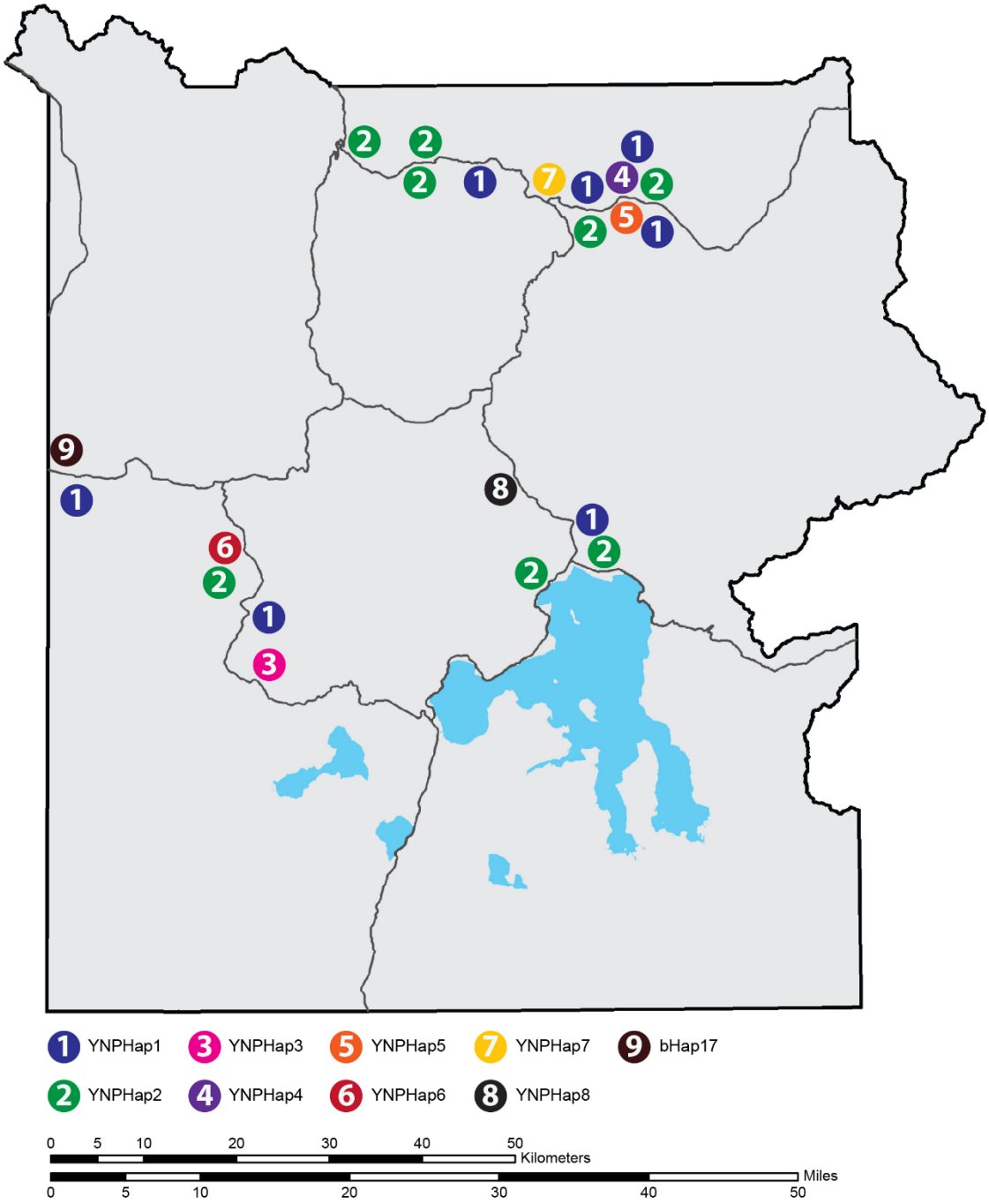


Figure II.1: Mitochondrial haplotype distribution in Yellowstone National Park. The sampling location and haplotype identity of each Yellowstone bison in this study based on their association with either the northern or central herds. Three additional samples (Templeton and two YNPHap2 bison) were collected from bison near the north boundary at a capture facility, but their movement histories are unknown, and they were omitted from this figure. Reprinted with permission from Forgacs et al. 2016.

Haplotype diversity among all 25 modern Yellowstone bison was calculated as 0.7800 (+/- 0.0649) with a mean difference between the haplotypes of 0.00103. The AMOVA test for population subdivision between the northern and central herds yielded an F_{ST} value of -0.06 ($p = 0.76$). Arlequin is known to produce slightly negative F_{ST} values in cases where variation within the population is larger than variation between the groups that comprise the population (Excoffier and Lischer 2010). In such cases, F_{ST} should be treated as zero (Emelianov et al. 2004, Elshibli and Korpelainen 2008, Yuan et al. 2010). Three of the 25 Yellowstone bison (Templeton, YNP5885, and YNP5899) were sampled after they were removed from the population at the northern boundary of the Park, but they were not part of the telemetry study and were, therefore, excluded from the population subdivision analysis. bHap17 was sampled from the west boundary capture operation which is a migration path used only by the central herd.

While bison from the northern breeding group tend to remain in the northern area for their entire lives, bison born in central Yellowstone will either be year-round residents of the central range or migrate to the northern range to spend the winter. Observations over recent years indicate many bison from the central herd have emigrated to become residents in northern Yellowstone year-round (White and Wallen 2012, Geremia et al. 2014) (Table II.2).

Haplotypes	Year-round in central Yellowstone National Park	Winter migrant to northern from central Yellowstone National Park	Year-round in northern Yellowstone National Park	Emigrant from central to northern Yellowstone National Park	Total
YNPHap1	1	2	4	0	7
YNPHap2	3	1	1	3	8
YNPHap3	1	0	0	0	1
YNPHap4	0	0	0	1	1
YNPHap5	0	0	0	1	1
YNPHap6	0	1	0	0	1
YNPHap7	0	0	1	0	1
Total	5	4	6	5	20

Table II.2: Mitochondrial haplotypes based on different life history strategies. Only the migration status of the 20 animals in the radio telemetry study were known and included. Reprinted with permission from Forgacs et al. 2016.

The analyses of all 65 plains and wood bison from across North America based on the combined data from Douglas et al. (2011) and this study revealed 30 different haplotypes with a total of 120 polymorphic sites. Forty-six of these sites were synonymous, 37 were non-synonymous, and 37 were polymorphic sites in tRNA- and rRNA-coding regions (Appendix A). The only genes with no nucleotide differences between the 65 bison were the ND4L protein-coding gene, and the tRNA-Val, tRNA-Gln, tRNA-Met, tRNA-Tyr, tRNA-His, tRNA-Ser, tRNA-Glu, and the tRNA-Pro genes. The Templeton-Crandall-Singh (TCS) network tree (Figure II.2, Figure II.3) identified two distinct clades of haplotypes separated by 10 polymorphisms unique to every member of each clade. Due to the large constraint on the mitochondrial genome, non-synonymous mutations in protein-coding genes that cause a change in the amino acid

sequence are deemed especially important. In an attempt to reduce the noise in the tree in Figure II.2, and to see if the clear divide between the two clades is still supported, a second tree only considering the 37 non-synonymous mutations was created (Figure II.3). This tree disregards all nucleotide differences in tRNA and rRNA genes, as well as all synonymous changes in the 13 protein-coding genes in the mitochondrial genome.

Some bison that formerly represented unique haplotypes collapsed into clusters that are compilations of haplotypes that were unique when all polymorphisms were taken into consideration, but lack any non-synonymous mutations to differentiate between them. Cluster 1 includes haplotypes CCSP1, bHap13, bHap16, and wHap15; Cluster 2 includes YNPHap1, YNPHap4, and EINP1; Cluster 3 includes EINP2, and wHap14; and Cluster 4 includes YNPHap2, bHap2, bHap10, bHap11, and bHap17. This TCS network further supports that the separation between the two clades is not just an artifact in the dataset (Figure II.3).

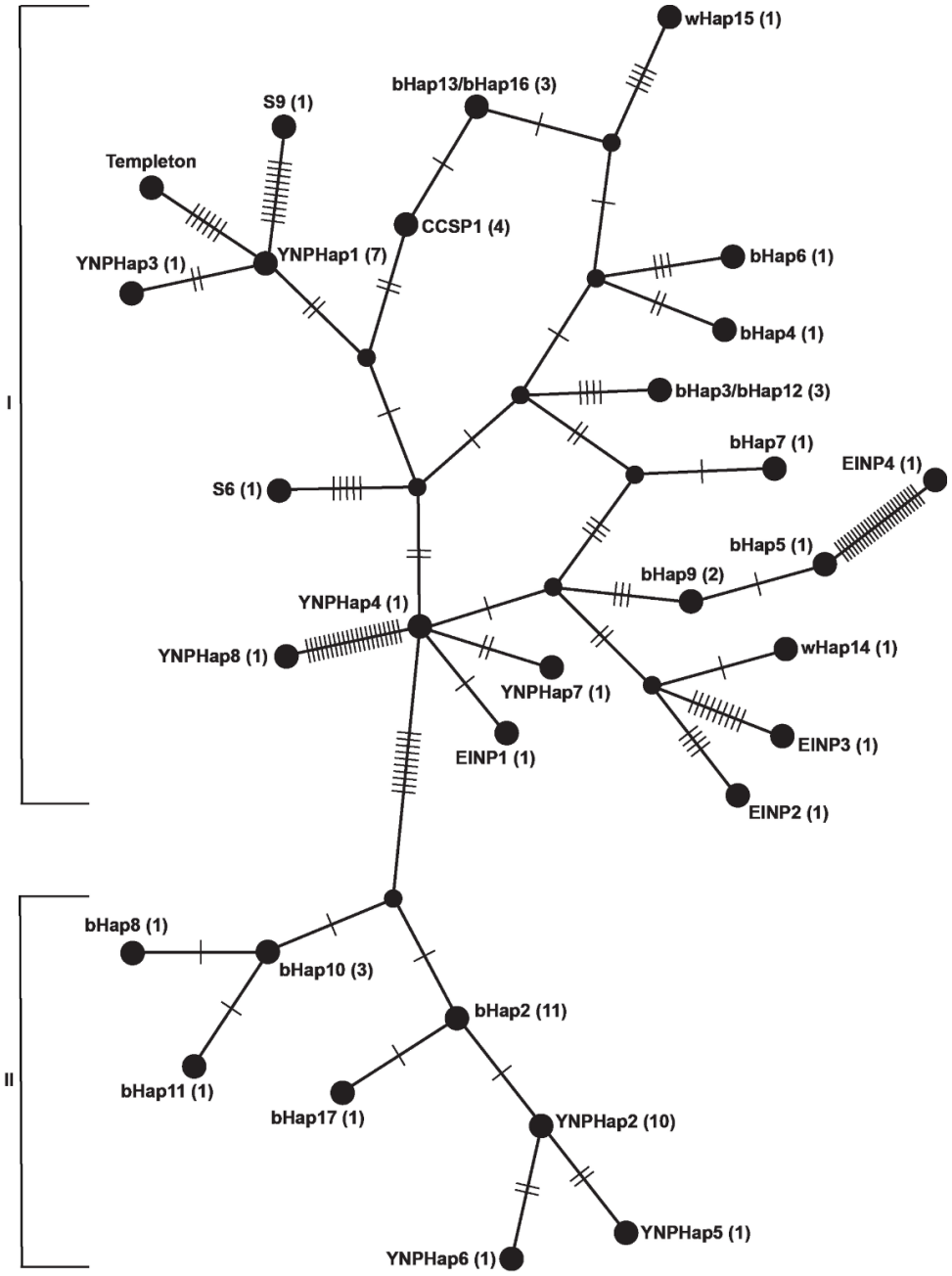


Figure II.2: TCS network tree of North American bison based on all polymorphic sites. Each dash represents one single nucleotide difference between two neighboring haplotypes. The numbers in the parentheses after the name of each haplotype denote the number of bison belonging to each haplotype. Roman numerals I and II represent the two clades in this analysis. Animals with a wHap and EINP prefix denote wood bison haplotypes, while YNPHap, Templeton, and bHap17 denote Yellowstone bison. S6 and S9 are two historical bison sampled in or near the modern day Yellowstone National Park. Reprinted with permission from Forgacs et al. 2016.

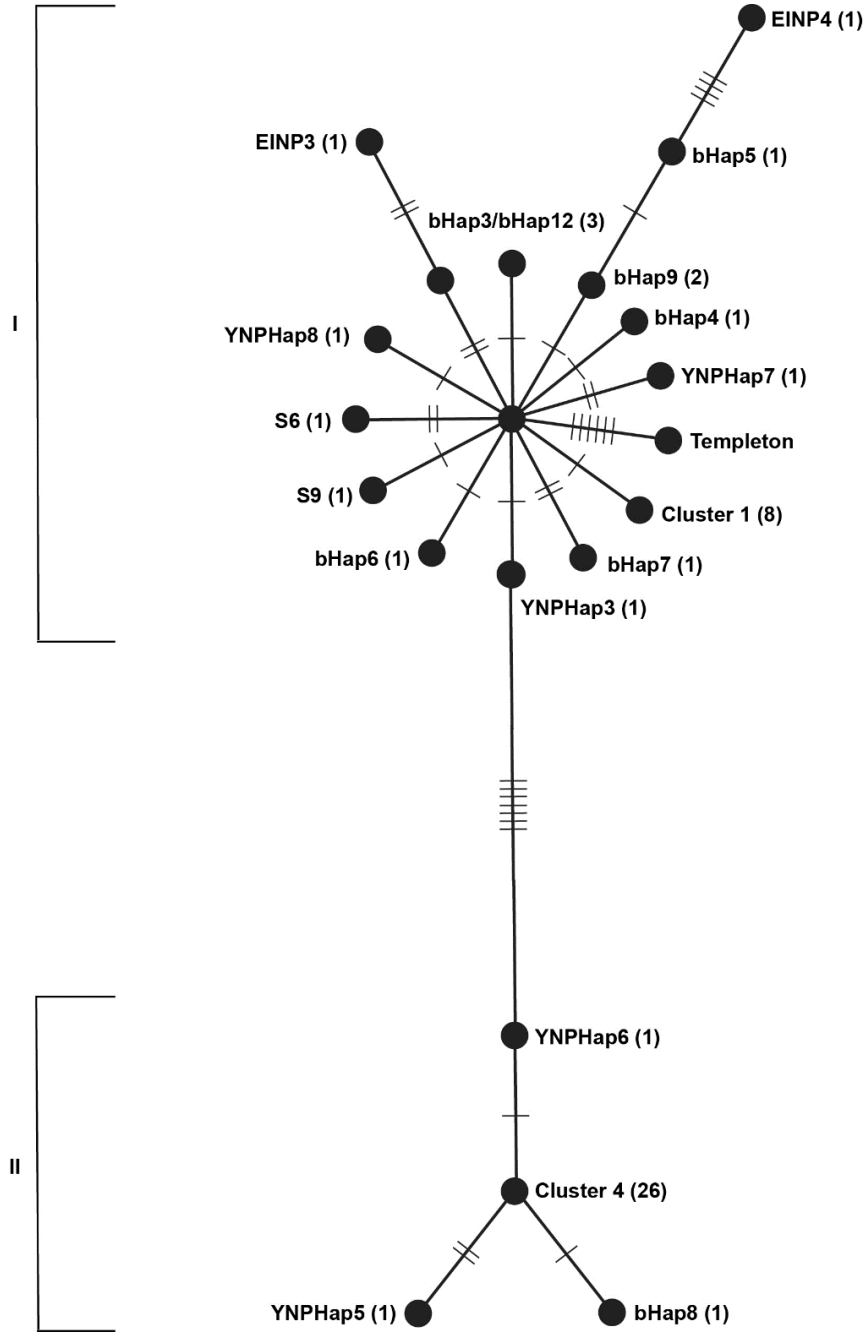


Figure II.3: TCS network tree of North American bison based on non-synonymous sites. All polymorphic sites that are synonymous or not in protein-coding genes were disregarded for this tree. Each dash represents one single nucleotide difference between two neighboring haplotypes. The numbers in the parentheses after the name of each haplotype denote the number of bison belonging to each haplotype. Roman numerals I and II represent the two clades formed in this analysis. Reprinted with permission from Forgacs et al. 2016.

Thirteen of the 25 Yellowstone bison belonging to YNPHap2, YNPHap5, YNPHap6, and bHap17 exhibited both alleged detrimental IMOP mutations, while the other 12 bison belonging to YNPHap1, YNPHap3, YNPHap4, YNPHap7, YNPHap8, and Templeton did not have the mutations. All animals exhibited either both or neither point mutations. Clades I and II align in such a way that all animals in Clade I that includes the indigenous Yellowstone bison are wild type, while all animals in Clade II that include the bison introduced in 1902 have both IMOP mutations (Figure II.2).

Discussion

The analysis of the mitochondrial genome of 25 Yellowstone bison yielded ten unique haplotypes, demonstrating high haplotype diversity in this population 0.78 (+/- 0.06). These haplotypes had little differentiation between them with the overall mean difference of 0.00103, which demonstrates the historic bottleneck and the subsequent process of increasing diversity due to the population boom and good management practices. High diversity is associated with greater population health and higher fitness in animals (Lacy 1987, Mattila et al. 2012). Similar results were found in a woodrat population in Texas where haplotype diversity was high (0.974 overall) but nucleotide diversity was low (0.008), suggestive of a period of low effective population size followed by rapid expansion (Mendez-Harclerode et al. 2007).

Using the 22 Yellowstone bison sampled for this study where sampling locations were well documented, population subdivision was tested on a geographic scale, using the northern and central herds as the two hypothetical populations. This study did not

find any evidence of population subdivision between the two herds based on mitochondrial DNA. Halbert et al. (2012) found evidence for population subdivision and the existence of at least two breeding herds within the Yellowstone bison population based on STRUCTURE analysis (Pritchard et al. 2000) using 43 nuclear microsatellites, but reported similar F_{ST} values to ours (0.0321). The reason for the difference in the findings could be due to differences in the structure and function of the genomic regions analyzed, the differences in mutation rates, and the sensitivities of the statistical tests used.

When the 10 Yellowstone haplotypes were compared with the 20 other bison haplotypes from across the United States and Canada, there was a clear division between Clade I and II. The comparatively large genetic distance between the two clades – 10 single nucleotide polymorphisms in contrast to the smaller distance between internal nodes within each clade – shows the signature of a major historic event that explains the separation between the two. Clade I contains haplotypes that are more similar to the two historical bison (S6 and S9) that lived in or near modern day Yellowstone National Park, while Clade II consists of bison that are genetically more dissimilar to those. Clade II also includes animals that are known to have originated from the Pablo-Allard herd, such as the samples from the National Bison Range, Fort Niobrara National Wildlife Refuge, as well as the bison introduced to northern Yellowstone National Park in 1902. Thus, we conclude that Clade I contains haplotypes that are more closely associated with the indigenous bison that lived in the area for hundreds of years, while Clade II has

haplotypes that resemble the bison introduced to the park from northern Montana in 1902.

Our findings indicate that no bison in Clade I have the mutations implicated in IMOPs, while both mutations were present in all bison in Clade II. Therefore, all the introduced bison from the Pablo-Allard herd and their descendants have these mutations. Comparing the frequency of IMOPs in 1902 with current frequencies enables a comparison of Yellowstone bison mtDNA haplotype frequencies over a span of 110 years. If these mutations truly cause a detrimental phenotype, as hypothesized by Pringle (Pringle 2011), a substantial reduction in the frequency of haplotypes with IMOPs would be expected due to strong negative selection. In 1902, there were a total of 10 indigenous females and 18 introduced females, while the 25 modern-day bison sampled are split with 13 animals with IMOPs and 12 that are wild type (Table II.3). Because the descendants of the indigenous bison lacked the IMOP mutations (Clade I), and the descendants of the introduced animals possessed the mutations (Clade II), we can compare the frequency of 1902 introduced bison and the current bison that were found to carry the IMOP mutations, as well as the frequency of the 1902 indigenous bison and the current frequency of wild type bison. Due to the statistically non-significant change in haplotype frequencies in the Yellowstone population ($p = 0.412$) based on Fisher's exact test and the lack of any kind of reported lesion or disease that affect a large proportion of Yellowstone bison, we did not find evidence to support Pringle's hypothesis.

	IMOPs	Wild type
Number of bison in 1902	18 (introduced females)	10 (indigenous females)
Frequency in 1902	0.643	0.357
Bison in current study	13	12
Current frequency	0.520	0.480

Table II.3: Historical and present frequencies of IMOP mutations in Yellowstone bison. Reprinted with permission from Forgacs et al. 2016.

The present study has contributed eight new haplotypes to a large dataset with over a million base pairs of mitochondrial DNA sequence analyzed from bison across North America. These data have significant implications beyond the Yellowstone bison population. For example, the bison network (Figure II.3) indicates that wood bison (*Bison bison athabascae*) are not a single monophyletic group. Approximately 6,600 plains bison from the Conrad herd were used to supplement the local population in Wood Bison National Park in the 1920s (Coder 1975). Hybridization between plains and wood bison is well documented (Peden and Kraay 1979, Wilson and Strobeck 1999), which certainly confuses their current taxonomical status. Elk Island National Park received 23 animals from Wood Bison National Park in 1965. While Cluster 3 (Figure II.3) represents a monophyletic group that is markedly different from the closest plains bison haplotype, others, such as wHap15, represent animals that may phenotypically look like wood bison but their mitochondrial DNA is closely related to plains bison. The closest haplotypes to wHap15 are bHap13/bHap16. These two haplotypes are from the descendants of the historical bison herd of Charles Goodnight which now constitutes the Caprock Canyon State Park herd in Texas. Historical documents show that Conrad purchased a bison heifer from Goodnight before he sold plains bison to the Canadian

government that they used to supplement the bison herd in Wood Buffalo National Park (Coder 1975).

Mitochondrial DNA analysis can be used as an indicator for population loss because it is four times more sensitive to any reductions in the effective population size compared to the nuclear genome (Bolfikova and Hulva 2012). Sequencing and analyzing a large dataset, such as whole mitochondrial DNA from many animals in a wildlife population, can reveal various facets that are telling about the history and genetic health of that population. Therefore, a vast amount of new information can be used to improve the conservation and management of species at risk of population reduction or extinction. Advances in sequencing technologies have allowed us to sample more animals and sequence whole mitochondrial genomes to discover eight new haplotypes and provide a finer resolution than previous studies. Due to the high number of haplotypes identified relative to the number of individuals sampled in this study, a higher proportion of the Yellowstone population should be sampled to capture a more comprehensive array of haplotype diversity that exists in Yellowstone bison. Our analysis suggests that Yellowstone bison represent nearly half – 10 of 22 modern plains bison haplotypes – of all the known haplotypes in plains bison from recently sampled individuals.

These data could be used in the future to develop a mitochondrial DNA-based assay to screen Yellowstone bison to determine lineage. Based on our findings, sequencing a 1223 bp region using 18F and 18R primers from Douglas et al. (2011) is sufficient to determine if a particular bison belongs to Clade I or Clade II. This sequence

contains two of the 10 mutations separating the two clades which is enough to delineate whether the bison is a descendant of the indigenous or the introduced herd. We plan on sequencing more Yellowstone bison to develop a SNP-based assay to gain a more accurate picture of the genetic differentiation and migration patterns.

The status of the Yellowstone bison population based on our findings of high haplotype diversity and lack of population subdivision appears to be genetically healthy, especially for a population with a history of intensive management that included periods of extreme reductions in size. In recent years, as the number of bison has grown exponentially and more bison leave the park during the winter, culling of animals to control their abundance and distribution has become necessary. Our finding that there is no subdivision based on mtDNA support that Yellowstone bison can be managed – for mitochondrial haplotype diversity – as a single population with multiple breeding segments. Before new management standards and policies are defined for the Yellowstone bison population, additional studies involving population structure and genetic diversity based on both mtDNA and nuclear genetic diversity assessments need to be conducted.

Methods

Approximately 30 adult female bison are fitted with radio telemetry collars each year in Yellowstone National Park to obtain demographic and movement information (Geremia et al. 2014, Hobbs et al. 2015). During November 2011 to January 2012, tail hairs and blood was collected from 20 of these bison (Table II.4). The bison were

chemically immobilized with Carfentanil and Xylazine following standard capture procedures periodically revised, reviewed, and endorsed by supervising veterinarians within the National Park Service and approved by park management. The National Park Service conducts reviews by an agency endorsed Institutional Animal Care and Use Committee which has reviewed anesthesia procedures conducted by the Yellowstone Bison Ecology and Management Program and has approved it as a monitoring and surveillance project. The IACUC reviews capture procedures every 3 years with the latest review completed in October of 2015. The blood samples were spotted on Whatman FTA cards and stored at room temperature. Bison were selected to represent four distinct survival strategies and movement patterns (Table II.2), including year-round residents in northern Yellowstone (n = 6), year-round residents in central Yellowstone (n = 5), bison that emigrated from central to northern Yellowstone and remained for breeding (n = 5), and bison that migrated from central to northern Yellowstone during winter, but returned to central Yellowstone for breeding (n = 4).

Yellowstone animal ID	TAMU ID	Haplotype	Sex	Age	Life history strategy (based on telemetry study)	Sampling location
YNP1586/99-088	13657	bHap17	F	6	Central year-round	Madison Valley
Ear Tag #80	2011002044	Templeton	M	5	Unknown (not included on map)	Gardiner Basin
3225	2012002859	YNPHap1	F	13+	Central year-round	Hayden Valley
Yell-047	2012002869	YNPHap1	F	10+	Northern year-round	Lamar Valley
Yell-056	2012002871	YNPHap1	F	9+	Northern year-round	Lamar Valley
Yell-358	2012002877	YNPHap1	F	5	Central winter migrant	Madison Valley
Yell-360	2012002878	YNPHap1	F	5	Northern year-round	Blacktail Deer Plateau
Yell-370	2012002884	YNPHap1	F	5	Central winter migrant	Old Faithful Basin
Yell-371	2012002885	YNPHap1	F	5	Northern year-round	Lamar Valley
WG091210-004	2009005885	YNPHap2	F	adult	Unknown (not included on map)	Gardiner Basin
WG091210-018	2009005899	YNPHap2	F	calf	Unknown (not included on map)	Gardiner Basin
Yell-017	2012002860	YNPHap2	F	13+	Central year-round	Pelican Creek
Yell-020	2012002861	YNPHap2	F	9	Northern emigrant from Central	Blacktail Deer Plateau
Yell-033	2012002864	YNPHap2	F	10	Northern emigrant from Central	Lamar Valley
Yell-036	2012002865	YNPHap2	F	12+	Central year-round	Old Faithful Basin
Yell-050	2012002870	YNPHap2	F	9	Northern year-round	Lamar Valley
Yell-066	2012002873	YNPHap2	F	4	Central year-round	Hayden Valley
Yell-067	2012002874	YNPHap2	F	2	Central migrant	Gardiner Basin

Table II.4: List of samples and relating haplotypes for the 25 modern Yellowstone bison from the study. Reprinted with permission from Forgacs et al. 2016.

Yellowstone animal ID	TAMU ID	Haplotype	Sex	Age	Life history strategy (based on telemetry study)	Sampling location
Yell-301	2012002876	YNPHap2	F	10	Northern emigrant from Central	Blacktail Deer Plateau
Yell-070	2012002875	YNPHap3	F	3	Central year-round	Old Faithful Basin
Yell-029	2012002862	YNPHap4	F	9	Northern emigrant from Central	Lamar Valley
Yell-031	2012002863	YNPHap5	F	13+	Northern emigrant from Central	Lamar Valley
Yell-040	2012002867	YNPHap6	F	7	Central winter migrant	Old Faithful Basin
Yell-362	2012002880	YNPHap7	F	5	Northern year-round	Lamar Valley
YNP1861	13932	YNPHap8	F	calf	Central year-round	Hayden Valley

Table II.4: Continued.

This study focuses only on the protein-coding, tRNA and rRNA genes of the mitochondrial genome and excludes most of the D-loop control region due to difficulties in sequencing long mononucleotide runs and lack of known adaptive evolutionary function. mtDNA was extracted from these Yellowstone samples (haplotype ID: YNPHap n = 20) and sequenced with a 3130 Genetic Analyzer (Life Technologies). The mitochondrial genomes were amplified based on the PCR protocol described in Douglas et al. (2011), and assembled using NCBI Reference Sequence NC_012346.1 (Achilli et al. 2008). The sequences were compared to previously published full mitochondrial haplotypes from Douglas et al. (2011) (haplotype IDs: bHap for plains bison and wHap

for wood bison, n = 31), with the exception of bPub1 which is a zoo animal of unknown lineage that was excluded from further analysis. The previously unpublished sequences can be found in the NCBI database (Appendix B).

Illumina whole genome sequencing was conducted for 14 additional bison. Four of these animals were from Caprock Canyon State Park in Texas (CCSP, 5x coverage); four were wood bison from Elk Island National Park in Canada (EINP, 5x coverage); four were additional Yellowstone animals that were collected from bison during winter removal operations (YNP1861, YNP5885, YNP5899, 10x coverage, and Templeton, 75x coverage); and two were historical museum specimens (S9 and S6, 10x coverage). The coverage values represent the average depth across the entire genome, mitochondrial reads are much more abundant due to the presence of hundreds of copies of the mitochondrial genome for each cell (Botero-Castro et al. 2013). For modern samples used in this study, the mitochondrial coverage ranges from 86x-603x. S9 was collected from the Lamar Valley in northern Wyoming in 1856, and S6 was collected from southern Montana in 1886, near the northern boundary of Yellowstone National Park (Dobson 2015) and the mitochondrial coverage for these historical samples was approximately 58x. There is limited information about the breeding ranges or movements of these 14 additional bison. Sequences were trimmed using FASTQ-MCF requiring a nucleotide quality score of more than 20 for each base and retaining only those reads that had a sequence length of more than 70 bases (Aronesty 2011). These filtered paired-end sequences were aligned to a previously published complete mitochondrial genome of a bison (YNP1586, GenBank ID: GU947004.1) using the

default settings in the Burrows-Wheeler Alignment 0.6.2 software package (BWA-MEM) (Li 2013). The resulting alignment files were sorted and indexed using SAMtools 0.1.18 (Li 2009). Read group information was added using the AddOrRelpaceReadGroups option of PicardTools 1.7.1 (https://github.com/broadinstitute/picard/releases/tag/1.128). The Genome Analysis Toolkit 3.1.1 (GATK) (McKenna et al. 2010) option RealignerTargetCreator was used to realign the mapped reads to account for INDEL shifted coordinates. Genetic variants, SNVs, and INDELs were identified against the mitochondrial bison sequence for each aligned sample and were filtered according to the GATK Best Practices recommendations (DePristo et al. 2011, Van der Auwera et al. 2013). These variants were placed into variant call formatted (VCF) files in order to make a consensus sequence for each sample using VCFtools package (Danecek et al. 2011).

The older historical bison, S9 from the year 1856, shows increased frequency of C to T nucleotide substitutions, potentially due to de-amination well documented in historical samples (Hofreiter et al. 2001) (Appendix A). However, all of these were unique to S9 only, meaning they only affect the length terminal node and not where the branch is located. Thus, these potentially de-aminated sites were presented in Figure II.2 as they were sequenced, the only effect being the likely over-estimated genetic difference between S9 and the basal haplotype YNPHap1.

The notion that the sequences analyzed in this study were truly mitochondrial in origin and not NUMTs – inactive copies of pieces of mtDNA copied into the nuclear genome – is supported by the lack of frameshift and nonsense mutations (Calvignac et

al. 2011) and a minimum coverage of 4x for any given base, without any sign of heteroplasmy. The mean length of numts are 100-300 bp (Calvignac et al. 2011), therefore primer sets amplifying 573-1223 bp fragments were used to avoid amplifying short nuclear copies during Sanger sequencing. Parts of mitochondrial sequences that had several SNPs close together in one, or a small subset of bison, have been checked using BLAST and yielded no suspicious nuclear sequences with a high identity score (Yao et al. 2008).

All the data from the 65 animals were compiled and aligned using ClustalW in MEGA version 6.0 (Tamura et al. 2013). The alignments were trimmed to the same length, 15,548 bp. Each polymorphic site was characterized based on whether it represented a synonymous or non-synonymous mutation or was in a non-coding region for further analysis (Appendix A).

Primers 18F (CTTTACCGCCATRGAACTAATCTT) and 18R (GTTCCTAAGACCAACGGATRA) were implemented from Douglas et al. (2011) to amplify a region of the mitochondrial genome containing two SNPs that separate Clade I and Clade II. The two SNPs are both G/A transitions that cause non-synonymous changes in the ND4 gene, Ile10678Met and Ala11108Thr. Sequencing the 1223 bp region flanked by these primers can serve as a diagnostic test to determine clade identity in lieu of sequencing the entire mitochondrial genome.

Haplotype diversity was calculated for the Yellowstone bison as $\widehat{H}e=\frac{n}{n-1}(1-\sum_{i=1}^{n}\hat{p}i^{2})$, where n is the number of bison and p is the frequency of each haplotype. Overall mean difference was determined by averaging the number of base

substitutions per site over all sequences using the Maximum Composite Likelihood model in MEGA version 6.0 (Tamura et al. 2004). Population subdivision was calculated using Arlequin's AMOVA feature to acquire the F_{ST} values based on the presence or absence of panmixia between the northern and central breeding herds. Phylogenetic networks were created using alignments imported in PopART v. 1.7 (http://popart.otago.ac.nz) and drawn as a TCS network using statistical parsimony (Figure II.2, Figure II.3) (Templeton et al. 1992, Clement et al. 2000). A maximum likelihood tree with 500 bootstraps, under the model of gamma-distributed rate heterogeneity amongst sites and a proportion of invariant sites (G+I) was also created using MEGA version 6.0, using water buffalo (*Bubalus bubalis*) (GenBank ID: AY488491.1), yak (*Bos grunniens*) (AY684273.2), domestic cattle (*Bos taurus*) (GU947021.1), and European bison (*Bison bonasus*) (HQ223450.1), as the outgroups.

CHAPTER III

EVALUATION OF FECAL SAMPLES AS A VALID SOURCE OF DNA BY COMPARING PAIRED BLOOD AND FECAL SAMPLES FROM AMERICAN BISON*

Introduction

Genetic variation is the basis for evaluating biodiversity within and between populations; without genetic variation, populations could not evolve or adapt to changing environmental conditions (Geffen et al. 2007). Modern wildlife conservation is becoming increasingly more dependent on the use of genetic tools due to their efficiency and reliability in identifying species and relationships within and between populations, and even resolving taxonomic conflicts. Collecting fecal material as a non-invasive sampling strategy has been widely used in situations where capturing wildlife is dangerous, time consuming, or unfeasible (King et al. 2008, Costa et al. 2017). Some researchers have warned against using fecal DNA without first conducting a pilot study to assess genotyping error rates (Maudet et al. 2004, Taberlet et al. 1999); most studies, however, adapt or develop markers for use with fecal DNA without validating the markers by using paired DNA from a high quality source (e.g. blood, soft tissue). Instead, they opt for a multi-tube approach that involves repeat testing of the same fecal

* Parts of this chapter were reprinted with permission from "Evaluation of fecal samples as a valid source of DNA by comparing paired blood and fecal samples from American bison (*Bison bison*)" by David Forgacs, Rick L. Wallen, Amy L. Boedeker, and James N. Derr, 2019. BMC Genet 20(22): s12863-019-0722-3.

samples to estimate genotyping errors rather than direct comparison of low versus high quality DNA isolated from the same individuals (Bellemain et al. 2005, Brazeal et al. 2017, Broquet et al. 2004, De Barba et al. 2017, Morin et al. 2018), with few exceptions (Ernest et al. 2000, Fernando et al. 2003, Kohn et al. 1999, Parsons 2001, Zhang et al. 2018). Thus, without a direct validation of the fecal-derived alleles, uncertainty arises about how allelic dropout and false allele rates could bias the reported results. Some studies compared results from genotyping fecal samples under different conditions (Murphy et al. 2002, Panasci et al. 2011, Zhu et al. 2017) and different methods for DNA extraction, such as rehydrating and carefully removing only the mucosal coat of the scat (Ball et al. 2007) or using a magnetic bead protocol (Flagstad et al. 1999), but to our knowledge, a direct and comprehensive comparison between a large number of paired samples at numerous loci has not been performed in any wildlife species.

The two most common issues with using low quality samples identified by previous STR studies are allelic dropout and false allele rates (Johnson et al. 2007). Allelic dropout occurs when heterozygotes are mistakenly reported as homozygotes due to the stochastic loss of one of the two alleles during amplification. Allelic dropout can lead to underestimating heterozygosity and genetic diversity, as well as misjudging the level and direction of natural selection (Taberlet et al. 1999, Johnson et al. 2007). In wildlife studies using fecal DNA markers, allelic dropout was often high, determined as 8% in a study on mountain lions (Ernest et al. 2000), 11.1% in wolves (Creel et al. 2003), and up to 49% in chimpanzees (Morin et al. 2001). However, all of these studies inferred dropout rates by re-amplifying DNA from the same fecal samples several times

and comparing those results (the multi-tube approach); none of them used matching high-quality DNA samples for comparison.

False allele rate (misprinting) describes a scenario where neither of the two alleles present in an individual amplifies, and instead, completely different alleles are detected. Misprinting can be due to PCR error, DNA degradation, or contamination. False alleles can also bias population metrics by providing inaccurate genetic diversity, heterozygosity, and population structure estimates, as well as overestimating census size because different samples from the same animal can appear to be unique. False allele rates in previous studies have ranged from negligible in coyotes (Kohn et al. 1999) and dolphins (Parsons 2001) to 5.6% in wolves (Creel et al. 2003), and 18.3% in barbary macaques (Lathuilliere et al. 2001).

The use of short tandem repeats (STRs), commonly known as microsatellites, is a well-established approach to analyzing non-coding repetitive regions for genetic analyses (Jarne and Lagoda 1996, Morgante and Olivieri 1993). STRs are typically considered to be neutral markers that are not generally subject to strong selection (Queller et al. 1993), therefore, they are especially useful for population genetic and forensic analyses due to the high variability in the number of repeats between individuals (Algee-Hewitt et al. 2016). Thus, by using just a handful of STR loci, low probability of identity (P_{ID}) and low likelihood of odds (LOD) scores can be achieved (Morton 1955). For population-level analyses, that means the probability of two animals possessing the exact same genotype across all markers by mere chance is low.

In order to assess the efficiency of STR markers with fecal DNA, we analyzed paired fecal and blood samples from American plains bison (*Bison bison bison*) from Yellowstone National Park. Yellowstone bison have been at the forefront of numerous conflicts for the past 150 years with regards to population management and genetic diversity (Bidwell 2010, Halbert et al. 2012, White and Wallen 2012, Wallen et al. 2015). In order to better understand the complete range of genetic diversity in Yellowstone bison, it is necessary to establish a truly random sampling practice. Previous approaches to estimate diversity indices have been heavily reliant on opportunistic sampling: when a portion of the population moves beyond the park boundary, bison are subject to capture and removal, thus animals that are sampled are ultimately the ones eliminated from the population (Forgacs et al. 2016, Halbert 2003, Halbert and Derr 2008). While these studies provide valuable preliminary data on the genetic status of Yellowstone bison, the use of non-invasive methods would provide a unique opportunity to sample a larger and more representative proportion of the population. Previous analysis using fecal DNA from Yellowstone bison has been conducted, but the lack of validation using paired high-quality DNA samples and the low number of loci tested necessitated further research (Gardipee 2007). If a set of validated markers was established for use with bison fecal DNA, then the opportunity to further explore the genetic diversity of the Yellowstone population and reevaluate information that was previously collected could be used to improve our understanding of this unique bison population. It could also prove useful with other free-ranging bison

herds, such as those at Elk Island National Park and Wood Buffalo National Park in Canada, and the Henry Mountains bison herd in Utah.

The objective of our research was to verify fecal samples as a reliable source of DNA by comparing 50 fecal samples from Yellowstone bison with blood samples from the same animals. We looked at the rates of allelic dropout and false alleles, as well as the overall feasibility of using fecal matter as the source of DNA for future genetic analyses in this population. In addition, most wildlife research currently lacks information regarding the validity of the STR markers used with fecal samples due to a lack of comparison to high quality DNA samples from the same animals. Here, we present a case study to show the importance and necessity of the meticulous design and validation of markers.

Methods

Fifty bison from the Yellowstone National Park bison herd that ventured outside the park were captured during January and February of 2015. The 41 females and 9 males ranged 1-4 years in age. Fecal samples were obtained directly from the rectum and stored at -20°C in 95% ethanol. Simultaneously, blood was drawn from the same fifty animals and spotted on Whatman FTA (Flinders Technology Associates) cards (GE Health Care, USA) (Pairs 1-50, Table III.1). Fecal DNA extractions were performed with the QIAamp Fast DNA Stool Kit (Qiagen, Valencia, California, USA), following the manufacturer's protocols.

Matched pair ID	Animal ID	Fecal sample ID	FTA sample ID	Collection date	Sex
Pair 1	81HL-6203	2015004267	2015004213	2/17/2015	female
Pair 2	81HL-6207	2015004268	2015004214	2/17/2015	female
Pair 3	81HL-6208	2015004269	2015004215	2/17/2015	female
Pair 4	81HL-6210	2015004270	2015004216	2/17/2015	female
Pair 5	81HL-6213	2015004271	2015004217	2/17/2015	female
Pair 6	81HL-6216	2015004272	2015004218	2/17/2015	female
Pair 7	81HL-6224	2015004273	2015004219	2/17/2015	female
Pair 8	81HL-6230	2015004274	2015004220	2/17/2015	female
Pair 9	81HL-6231	2015004275	2015004221	2/17/2015	female
Pair 10	81HL-6233	2015004276	2015004222	2/17/2015	female
Pair 11	81HL-6276	2015004277	2015004223	2/18/2015	male
Pair 12	81HL-6287	2015004278	2015004224	2/18/2015	female
Pair 13	81HL-6288	2015004279	2015004225	2/18/2015	female
Pair 14	81HL-6290	2015004280	2015004226	2/18/2015	male
Pair 15	81HL-6291	2015004281	2015004227	2/18/2015	female
Pair 16	81HL-6297	2015004282	2015004228	2/18/2015	male
Pair 17	81HL-6317	2015004283	2015004229	2/19/2015	female
Pair 18	81HL-6326	2015004284	2015004230	2/19/2015	female
Pair 19	81HL-6336	2015004285	2015004231	2/19/2015	female
Pair 20	81HL-6337	2015004286	2015004232	2/19/2015	female
Pair 21	81HL-6346	2015004287	2015004233	2/19/2015	female
Pair 22	81HL-6657	2015004288	2015004234	1/22/2015	female
Pair 23	81HL-6659	2015004289	2015004235	1/22/2015	female
Pair 24	81HL-6664	2015004290	2015004236	1/22/2015	female
Pair 25	81HL-6669	2015004291	2015004237	1/22/2015	female
Pair 26	81HL-6670	2015004292	2015004238	1/22/2015	female
Pair 27	81HL-6723	2015004293	2015004240	1/26/2015	male
Pair 28	81HL-6735	2015004294	2015004241	1/26/2015	male
Pair 29	81HL-6736	2015004295	2015004242	1/26/2015	male
Pair 30	81HL-6740	2015004296	2015004243	1/26/2015	male
Pair 31	81HL-6742	2015004297	2015004244	1/26/2015	male
Pair 32	81HL-6745	2015004298	2015004245	1/26/2015	male
Pair 33	81HL-6790	2015004299	2015004246	2/10/2015	female
Pair 34	81HL-6793	2015004300	2015004247	2/10/2015	female
Pair 35	81HL-6795	2015004301	2015004248	2/10/2015	female
Pair 36	81HL-6796	2015004302	2015004249	2/10/2015	female
Pair 37	81HL-6801	2015004303	2015004250	2/10/2015	female
Pair 38	81HL-6802	2015004304	2015004251	2/10/2015	female
Pair 39	81HL-6810	2015004305	2015004252	2/10/2015	female
Pair 40	81HL-6812	2015004306	2015004253	2/10/2015	female
Pair 41	81HL-6883	2015004309	2015004255	1/20/2015	female
Pair 42	81HL-6895	2015004310	2015004256	1/20/2015	female
Pair 43	81HL-6896	2015004311	2015004257	1/20/2015	female
Pair 44	81HL-6906	2015004312	2015004259	1/20/2015	female
Pair 45	81HL-6952	2015004313	2015004260	1/21/2015	female
Pair 46	81HL-6955	2015004314	2015004261	1/21/2015	female
Pair 47	81HL-6960	2015004315	2015004262	1/21/2015	female
Pair 48	81HL-6964	2015004316	2015004263	1/21/2015	female
Pair 49	81HL-6965	2015004317	2015004264	1/21/2015	female
Pair 50	81HL-6980	2015004318	2015004265	unknown	female

Table III.1: List of the 50 pairs of fecal and blood samples from Yellowstone bison. Reprinted with permission from Forgacs et al. 2019.

Blood samples were extracted from the FTA cards by soaking the individual 1.20 mm punches in 200 µL of 20 mM NaOH followed by incubation at 50°C for 30 min with periodic inversion. After aspirating the NaOH, 200 µL of 10% TE buffer was added to the punches and then removed after 2 minutes at room temperature. The punches were left uncovered to dry overnight and were used directly in the PCR reaction.

Both fecal and blood samples went through the same PCR protocol, using 35 STR markers adapted from Schnabel, (2001) and Halbert (2003) (Table III.2). PCR was run with 3.175 µL of fluorescently labeled stock primer mix (concentrations listed in Table III.2), 0.5 µL of 10X MasterAmp (Epicentre Technologies, Madison, Wisconsin, USA), 0.25 µL of 10 mM dNTPs, and 0.075 µL of Promega GoTaq (Promega, Madison, Wisconsin, USA) which was added to 1 µL of fecal DNA extract or FTA punch. During PCR, the samples were heated for 3 min at 96°C, followed by 4 cycles of 20 s at 96°C, 30 s at 58°C, and 90 s at 65°C, decreasing the initial temperatures by 1°C per cycle. Then the samples were subjected to 26 cycles of 20 s at 96°C, 30 s at 54°C, and 90 s at 65°C. The reaction was concluded with 1 min at 96°C, 1 min at 54°C, and 20 min at 65°C. For all fecal samples, a second PCR was set up with the same parameters, using 1 µL of amplified DNA from the previous PCR reaction as the template for the second round in order to obtain the sufficient amount of copies for genotyping.

Multiplex	Locus	Primer name	Final concentration (µM)	Label	Sequence 5' -> 3'	Allele size range (bp)
80	BM1862	BM1862F	5	6-FAM	AAGCAAAAAGGCTGATGGC	199-215
		BM1862R	5	none	TTGCAGATACTGGCAAGTGG	
	BMS1001	BMS1001F	5	NED	GAGCCAATTCCTACAATTCTCTT	107-115
		BMS1001R	5	none	AGACATGGCTGAAATGACTGA	
	BMS1074	BMS1074F	4	NED	CAGTAGCCAAGATATGGAAGCA	152-160
		BMS1074R	4	none	AGCTCCTTGCTGCTACAAATG	
	BMS1675	BMS1675F	10	6-FAM	CATTAGAAAGCTGATTGGAGGG	85-91
		BMS1675R	10	none	TAATAATCAGTGCCGCTCCC	
	BMS1716	BMS1716F	20	HEX	GTGGGTTGGAGAGGTACAAG	185-197
		BMS1716R	20	none	AGAAATGGCCTTGAGAAAGAG	
	HUJ246	HUJ246F	5	NED	ACTCCAGTTTTCTTTCCTGGG	242-264
		HUJ246R	5	none	TGCCATGTAGTAGCTGTGTGC	
82	BM4311	BM4311F	5	6-FAM	TCCACTTCTTCCCTCATCTCC	90-104
		BM4311R	5	none	GAAGTATATGTGTGCCTGGCC	
	BM6017	BM6017F	5	HEX	TCTTCTGTTTTCCTCCATCCC	104-122
		BM6017R	5	none	GGAAACTAGCTTATGCTGTGGG	
	BM711	BM711F	5	6-FAM	CAGCATCAGCAACTAACATAGG	161-177
		BM711R	5	none	TGGACCATGAGGGAAGTCTC	
	INRA189	INRA189F	5	NED	TACACGCATGTCCTTGTTTCGG	96-100
		INRA189R	5	none	CTCTGCATCTGTCCTGGACTGG	
	TGLA122	TGLA122F	10	NED	CCCTCCTCCAGGTAAATCAGC	136-150
		TGLA122R	10	none	AATCACATGGCAAATAAGTACATAC	
85	BL1036	BL1036F	5	NED	TAGCTTATGCCATTGTTTTTGC	175-193
		BL1036R	5	none	ATCTGATGTGGGTTTCTGACTG	
	BM4107	BM4107F	8	HEX	ATAGGCTTTGCATTGTTCAGG	159-185
		BM4107R	8	none	AGCCCCTGCTATTGTGTGAG	
	BM47	BM47F	10	6-FAM	ACAGGAAGGAGAAGGGGAAG	103-119
		BM47R	10	none	CCGGGGTCACATGACTCTG	
	BMS1315	BMS1315F	5	HEX	AAGCCATTGATTGTAGATTGGG	135-149
		BMS1315R	5	none	GAGTTTCCTTTTTCCCCCAC	
	BMS1857	BMS1857F	5	6-FAM	TAATTGATCACAAAGAGGAGCC	142-170
		BMS1857R	5	none	GATCCCAGAGAATCACTCACC	
	ILSTS102	ILSTS102F	5	NED	CAGGACTGAGTAACTAAGGC	113-156
		ILSTS102R	5	none	AGGAGACAGCTACAAACCCC	

Table III.2: List of the 35 STR markers used in the study. Reprinted with permission from Forgacs et al. 2019.

Multiplex	Locus	Primer name	Final concentration (µM)	Label	Sequence 5' -> 3'	Allele size range (bp)
PRTG 1	BM17132	BM17132F	9	6-FAM	ATCTGCCAGTATCACATCAACA	81-103
		BM17132R	9	none	GTTACTTTTCCAGGCATGAAGC	
	BMS410	BMS410F	8	NED	GGCTGAAAAGCTGTGGTGTT	81-103
		BMS410R	8	none	TTGCCACATTTACCTTCTTTCA	
	BMS510	BMS510F	5	VIC	TGCTGCATGATTCTCATTCC	88-98
		BMS510R	5	none	AGCCTTCCTGTTCTCTGCTG	
	BMS527	BMS527F	10	6-FAM	TCAGTGAAAGCAAGAGAAATATCC	159-183
		BMS527R	10	none	TTCCATTCCCTTTGAATATCCC	
	RM372	RM372F	15	VIC	TTCAACCCAACATCCACTTG	114-138
		RM372R	15	none	ACTTAGATTTCCAAGCCCAGG	
PRTG 2	BM1225	BM1225F	15	NED	TTTCTCAACAGAGGTGTCCAC	233-273
		BM1225R	15	none	ACCCCTATCACCATGCTCTG	
	BM1706	BM1706F	6	6-FAM	ACAGGACGGTTTCTCCTTATG	230-256
		BM1706R	6	none	CTTGCAGTTTCCCATACAAGG	
	BM1905	BM1905F	8	NED	GTCCATGGGTTCACAAAGAG	170-196
		BM1905R	8	none	ACGCCTGCTGATGCTGTAG	
	BM2113	BM2113F	6	6-FAM	GCTGCCTTCTACCAAATACCC	127-153
		BM2113R	6	none	CTTCCTGAGAGAAGCAACACC	
	BM4440	BM4440F	5	NED	CCCTGGCATTCAACAAGTGT	121-147
		BM4440R	5	none	CACCCTGTTAGGAATCACTGG	
	BM720	BM720F	8	VIC	ACATCTCATTCTTGTGTCATGG	203-239
		BM720R	8	none	GAAATTGACTTTAGGGTTCCCC	
PRTY A	BM4513	BM4513F	5	NED	GCGCAAGTTTCCTCATGC	132-134
		BM4513R	5	none	TCAGCAATTCAGTACATCACCC	
	BM7145	BM7145F	12	NED	ATTATGTTCCAGATTCCATTCCA	108-112
		BM7145R	12	none	CAGCACTGTTTCATAAACTATGGG	
	BMS2270	BMS2270F	2	6-FAM	(GTTTCTT)CTGCGTTAACACCCCACC	66-70
		BMS2270R	2	none	(GTGTCTT)GCAGGAAGGCTGATGCAC	
	SPS113	SPS113F	3	VIC	CCTCCACACAGGCTTCTCTGACTT	128-132
		SPS113R	3	none	CCTTAACTTGCTTGAGTTATTGCCC	

Table III.2: Continued.

Multiplex	Locus	Primer name	Final concentration (µM)	Label	Sequence 5' -> 3'	Allele size range (bp)
PRTY BC	AGLA293	AGLA293F	10	VIC	GAAACTCAACCCAAGACAACTCAAG	218-220
		AGLA293R	10	none	ATGACTTTATTCTCCACCTAGCAGA	
	BM4307	BM4307F	20	6-FAM	ATAACACAAAAAGTGGAAAACACTC	183-187
		BM4307R	20	none	ATTTTATCTCAGGTCCCTTTTTATC	
	CSSM42	CSSM42F	12	NED	GGGAAGGTCCTAACTATGGTTGAG	167-171
		CSSM42R	12	none	ACCCTCACTTCTAACTGCATTGGA	

Table III.2: Continued.

After the amplification of the STR markers, the samples were genotyped using the AB Genetic Analyzer 3130xl. The results were analyzed using GeneMapper v. 3.7 software (Applied Biosystems, Carlsbad, CA) and the genotyping calls were made separately for fecal and blood samples to avoid any bias. When the fecal/blood pairs from the same animals were compared for each STR locus, they were noted as matching (same genotype for both), not matching (different genotype) or unreadable (either blood or fecal sample has yielded no intelligible data). All mismatches were verified using the multi-tube approach by a minimum of two independent amplification and genotyping runs. In the case of the Y chromosome marker INRA189, even a single instance of amplification of the allele was scored as a male.

A final STR panel was chosen based on the ability to genotype fecal DNA with the highest reliability (>95% matches). These loci were mapped to the *Bos taurus* (UMD 3.1) reference genome and the genetic distance was determined based on the 1.23-1.25 cM/Mb recombination frequency reported in domestic cattle (Arias et al. 2009). The expected frequency of each allele was determined based on 10,000-17,000 plains bison from across North America previously genotyped at each locus (Derr, unpublished). The probability of identity (P_{ID}) was calculated by assuming the most conservative estimate that a bison was homozygous for the most common allele at each locus, following the directions in Butler (2005) (Appendix C). All the linked loci were less than 50 map units apart, so only the loci with the lowest major allele frequencies were included in the probability of identity calculation from each chromosome, providing the most conservative estimate.

Regression analyses were performed to evaluate the relationship between the efficiency of each marker and several physical and chemical properties of the STR loci such as the location on the chromosome, length of the motif, C/G content, primer attributes and secondary structure.

Heterozygosity (H_O) was calculated both by animal and by locus. Heterozygosity estimates were calculated using fecal samples and blood samples separately, and the differences in heterozygosity were calculated based on samples that produced readable data for both. Overall statistical significance between all fecal and blood samples and pairwise statistical significance for each animal and marker was tested using a paired Student's t-test.

Results

DNA from the 50 paired fecal and blood samples were analyzed at 35 STR loci. On average, 82.80% of the samples matched (range: 62.86% - 100%) at the markers tested, 13.31% yielded unreadable data (range: 0% - 37.14%), and the fecal and blood samples did not match at 3.89% of the markers (range: 0% - 22.86%) (Figure III.1).

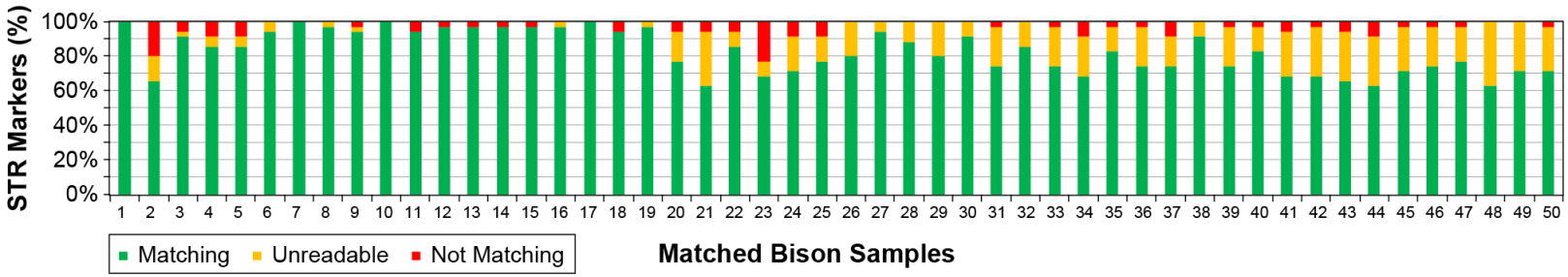


Figure III.1: Efficiency of each pair of bison fecal and blood samples at 35 STR loci. The percentage of markers that matched for each pair, the percentage of unreadable results due to no amplification or unclear genotyping results, and the percentage of non-matching pairs where the blood and fecal samples from the same bison yielded different genotyping results are shown. Each fecal-blood pair from the same bison is referred to by number 1-50 (Table III.1). Reprinted with permission from Forgacs et al. 2019.

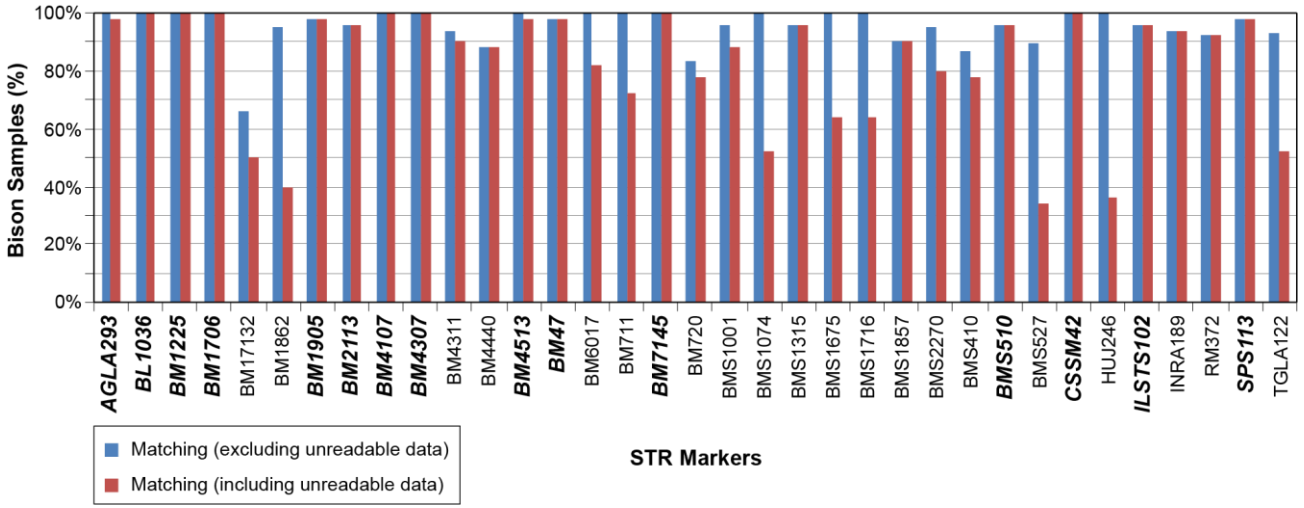


Figure III.2: Efficiency of each STR marker for the 50 paired bison fecal and blood samples. The bars represent the percentage of samples that matched for each of the markers. The STR markers chosen for the final panel are shown in bold and italics. Reprinted with permission from Forgacs et al. 2019.

The percent of matching pairs for each STR marker was also determined by two methods (Figure III.2). The efficiency was calculated as the proportion of samples for each marker where the blood and fecal samples matched (Percent matching (including unreadable data)), as well as the proportion of the samples that produced readable data divided by the sum of those that either matched or did not match (Percent matching (excluding unreadable data)). The difference between the results from the two methods is the percent of unreadable samples.

The allelic dropout rate, calculated as the instances where only one of the two alleles in a blood sample were present in the matching fecal sample, was 0.023 overall, which means allelic dropout accounted for 60.3% of all mismatches between fecal and blood samples observed in this study. In order to assess potential contamination, cases were also examined where neither allele matched at a certain marker and the false allele rate was determined to be 0.0046, accounting for only 11.8% of all mismatches. The remainder of the mismatches were attributed to other types of discrepancies, such as when an individual appeared to be a heterozygote based on fecal genotypes, whereas the blood-based genotype was homozygous for one of the alleles. Therefore, in this example they still share an allele but the second allele from the fecal-derived DNA can be assumed to be an artifact.

Fifteen of the 35 STRs were chosen for a final panel of markers that performed at an exceptionally high efficiency with over 95% of the paired fecal and blood samples matching (Table III.3). The 15 STR markers reside on 10 chromosomes, and the closest STRs on the same chromosome are located 14.76 Mb (18.15 cM) apart on chromosome

2. The probability of identity (P_{ID}), after accounting for linkage, was calculated as 2.318×10^{-6}.

Marker	Percent matching (excluding reruns)	Percent matching (including reruns)	Chromosome	Major allele frequency (x_i)	Major genotypic frequency (x_i^2)
BM7145	100%	100%	1	0.800	0.640
BM4307	100%	100%	1	0.782	0.612
BM2113	96%	96%	2	0.360	0.129
CSSM42	100%	100%	2	0.607	0.368
AGLA293	100%	98%	5	0.958	0.917
SPS113	98%	98%	10	0.613	0.376
BL1036	100%	100%	14	0.313	0.098
BM4513	100%	98%	14	0.872	0.761
BM1706	100%	100%	16	0.660	0.436
BM1225	100%	100%	20	0.433	0.187
BM4107	100%	100%	20	0.363	0.132
BM1905	98%	98%	23	0.495	0.245
BM47	98%	98%	23	0.729	0.531
ILSTS102	96%	96%	25	0.531	0.282
BMS510	96%	96%	28	0.466	0.217

Table III.3: List of 15 STR loci chosen due to their high fidelity and efficiency. The major allele frequency for each STR was determined based on an extensive library of 10-17,000 plains bison, and used to estimate the probability of identity (2.318×10^{-6}). Reprinted with permission from Forgacs et al. 2019.

We have investigated a number of different parameters to identify the factors responsible for the discrepancy in the efficiency of different markers when tested on fecal material. However, no relationship was observed between the efficiency of the STR markers and primer length ($R^2 = 0.0016$, $p = 0.82$), average primer G/C content ($R^2 = 0.0087$, $p = 0.60$), amplicon length ($R^2 = 4 \times 10^{-5}$, $p = 0.97$), amplicon G/C content ($R^2 = 0.0051$, $p = 0.69$), primer melting temperature ($R^2 = 0.018$, $p = 0.27$), number of

hairpin bases (R^2 = 0.0154, p = 0.31), homodimers (R^2 = 0.0041, p = 0.60), or heterodimers (R^2 = 0.0052, p = 0.55) (Figure III.3). These calculations were based on the *Bos taurus* genome (UMD 3.1), and amplicon length and G/C content may differ slightly in bison. Because all bovid autosomal chromosomes are acrocentric, and the cattle reference assemblies were built starting from the centromere, we could assess if the distance of the STR regions from the centromere had any effect on the efficiency of the STR marker. However, no relationship was detected (R^2 = 0.0465, p = 0.22) (Figure III.3).

Heterozygosity estimates were compared between fecal and blood samples and calculated in two ways: (i) what percent of loci were heterozygous in each bison (heterozygosity by animal, Figure III.4), and (ii) what percent of animals were heterozygous for each locus (heterozygosity by locus, Figure III.5). The hemizygous Y chromosome marker (INRA189) was excluded from this analysis.

For heterozygosity by animal, overall, the estimates from the fecal samples showed a significant 2.7% reduction in heterozygosity (p < 0.0001***) compared to the estimates from blood samples (Figure III.4). Nearly half of the animals showed no changes in heterozygosity, while others experienced a reduction of up to 20.7% (p < 0.05*). There were 3 samples where the heterozygosity estimates were actually higher in fecal samples, but none of those differences were significant (Pairs 11, 47, and 50).

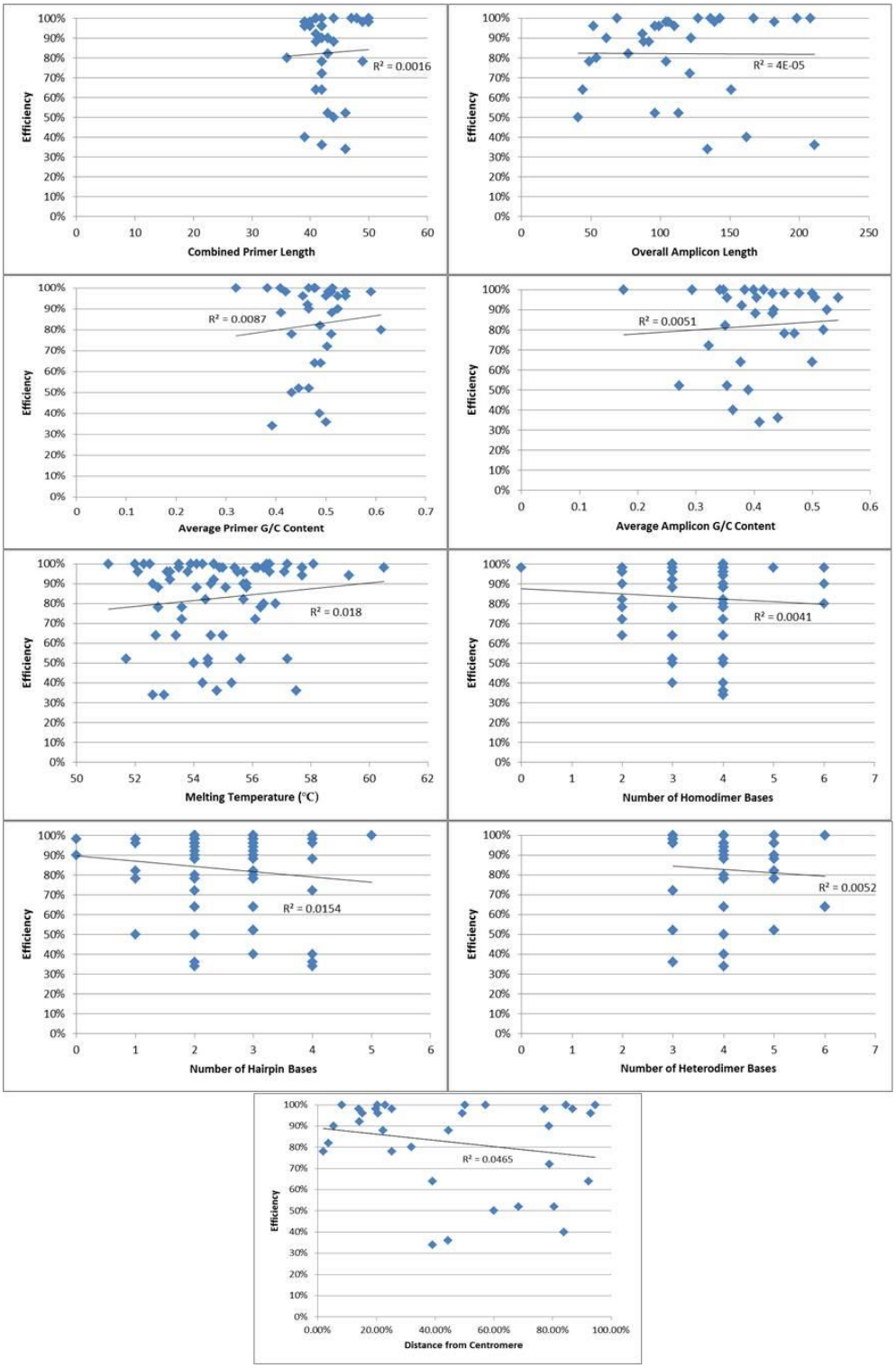


Figure III.3: Efficiency of STR markers based on various parameters tested. Reprinted with permission from Forgacs et al. 2019.

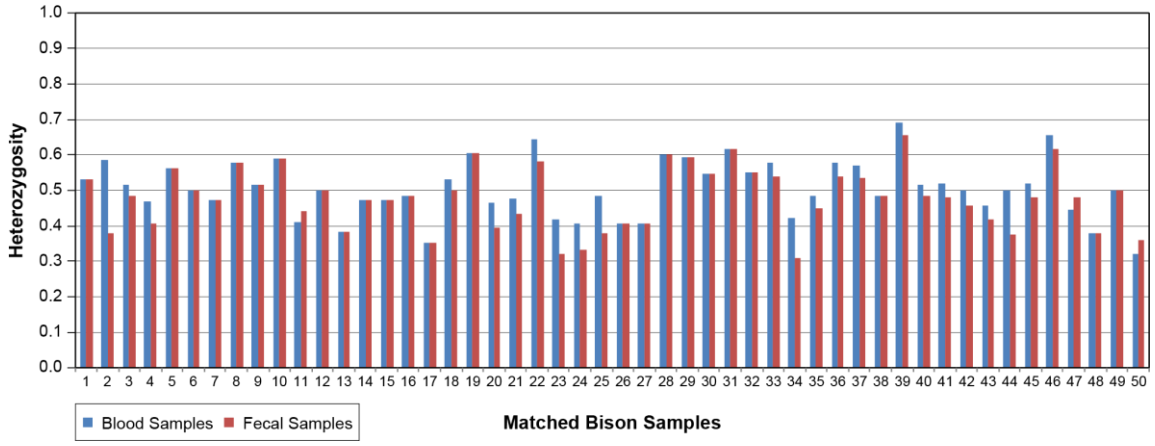


Figure III.4: Heterozygosity estimates from blood and fecal samples by animal. A significant overall reduction in heterozygosity is seen when fecal samples are used ($p < 0.0001$). Each fecal-blood pair from the same bison is referred to by number 1-50 (Table III.1). Reprinted with permission from Forgacs et al. 2019.

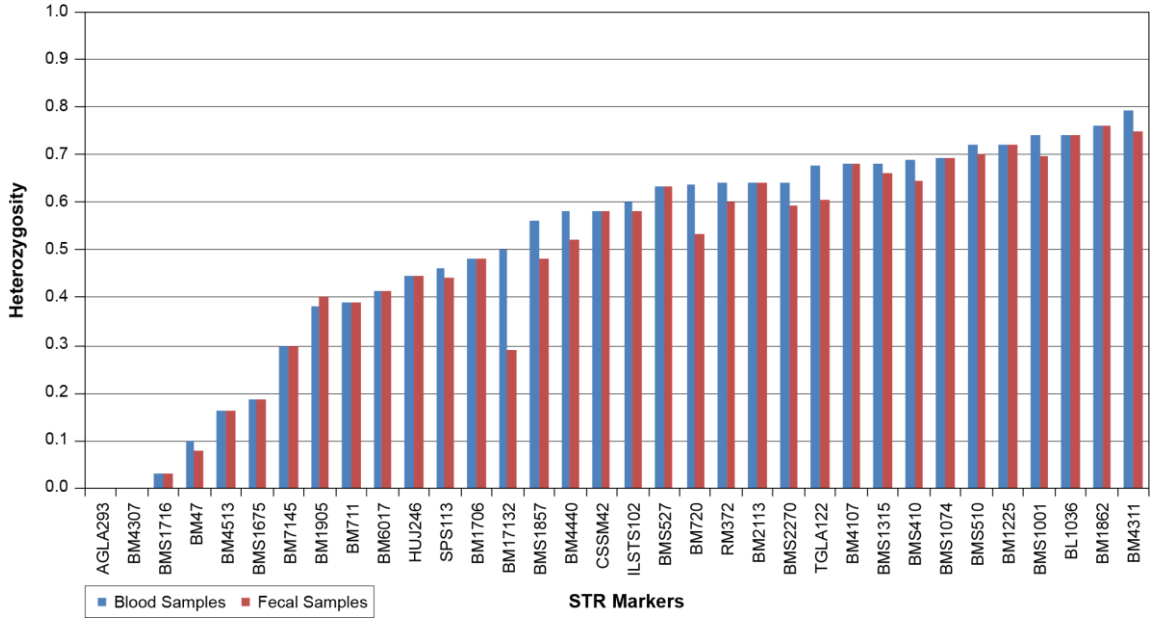


Figure III.5: Heterozygosity estimates from blood and fecal samples by STR marker. A significant reduction is seen when fecal samples are used ($p < 0.005$). Reprinted with permission from Forgacs et al. 2019.

Similarly, there was a significant overall reduction in fecal-derived heterozygosity by locus ($p < 0.005$**) (Figure III.5). Over half of the loci showed no change in heterozygosity between the blood and the fecal samples, while others showed a significant reduction as high as 21.1% ($p < 0.05$*). Only a single STR locus showed an increase in heterozygosity from fecal samples (BM1905) but it was not significant.

Discussion

We evaluated 35 STR markers for use with fecal DNA by comparing genotypes generated from paired blood and fecal samples from 50 bison and investigated how heterozygosity is affected by using different types of samples. We concluded that the majority of mismatches between fecal and blood DNA were caused by allelic dropout, which has been previously identified as a major drawback to using fecal DNA (Taberlet et al. 1999, Johnson et al. 2007). Yet, fecal marker validation beyond simply re-analyzing the poor quality fecal samples has remained rare.

During DNA extraction, bison hairs were found in many fecal samples. The origin of these hair samples is unknown; they could be from the same individual or from other bison. The source of these hairs could either be hair ripped out during sample collection or ingested during grazing or grooming. While extreme care was taken to avoid hairs during DNA extraction, some shed hair follicles may be a source of DNA during amplification. Cases where neither allele matched were rare (0.46% of cases) and explained only 12.12% of mismatched genotypes between blood and fecal paired samples. Due to the rarity of this event, and the fact that it was not a consistent mismatch

across multiple loci in the same fecal-blood pairs, we can conclude that it is highly unlikely that foreign DNA has been amplified. This supports the notion that either the presence of hair follicle DNA did not affect our analysis, or that the hair in the fecal samples came from that individual.

A single Y chromosome marker was tested in the study (INRA189), however, we discourage using fecal DNA with any presence-absence markers where the absence of signal is a genotype in and of itself. A lack of signal could be misinterpreted as a null allele (in this case, a female bison), while it may have been the result of a failure in the amplification of the allele from fecal DNA. Thus, sex can only be determined with some level of certainty in cases where a heterozygous X chromosome marker and no amplification on the Y can be used to definitively identify some females, while Y chromosome amplification and only one X chromosome allele amplifying (hemizygosity) is strong evidence for a male. It is important to note that neither our X (BM6017) nor our Y chromosome (INRA189) markers performed with high enough reliability to be included in our final bison panel.

After careful comparison with paired blood samples, 15 of the 35 STRs were identified as highly reliable markers for bison genotyping using fecal material. However, verifying fecal markers by merely repeating the extraction or the amplification step can lead to erroneous results. Given that the number of bison in North America is currently around 500,000, these 15 markers yield a sufficiently low probability of identity for use in population genetics studies (Department of the Interior 2014).

We compared heterozygosity based on fecal- and blood-derived DNA and conclude that there is a highly significant ($p < 0.005$) decrease in heterozygosity estimates from fecal samples. Our results show that without a careful assessment of the markers, significantly skewed population metrics may be reported. A significant proportion of conservation genetics studies use fecal samples in animal species that are rare or endangered because in many of those cases, invasive sampling is not feasible (Kurose et al. 2005, Silva et al. 2015, Taberlet et al. 1997). Fecal DNA can lead to an artificially decreased heterozygosity estimate, which may negatively skew estimates of the genetic diversity of the species. This discrepancy in population metrics emphasizes the need for better quality control in the form of paired high-quality DNA samples to verify the reliability of each marker. We advise researchers to first test their markers by using a direct comparison between the low- and high-quality sources of DNA in a subset of their samples to validate those markers for use in low quality samples. Once verified, they can use those markers with a high level of confidence in the rest of their samples without the need to collect DNA from other sources.

In conclusion, careful consideration needs to be taken when designing fecal DNA studies. While numerous methods are used to evaluate genotyping error rates, we contend that a paired sample-based genotyping method such as the one described here will provide the most accurate validation of fecal DNA markers. This method will also yield the best sampling protocol and the most precise results. We present a set of 15 reliable STR markers for use with bison fecal DNA. Based on these verified markers, a population-wide study can be conducted using fecal DNA to evaluate genetic diversity

parameters and population metrics in Yellowstone bison whose capture is unfeasible inside the national park. It would also aid with the evaluation of other herds that are under protection or lack the infrastructure to capture bison and acquire high quality DNA samples for analysis. While amplification was possible from DNA extracted from every fecal sample, some markers amplified more reliably than others. Due to the lack of consensus on what factors influence marker performance, the rigorous assessment of markers used in all fecal DNA studies is necessary to provide valid results. By comparing fecal samples to blood from the same animals we show a significant reduction (in some cases in excess of 20%) in heterozygosity, a common population genetics metric often used in fecal DNA studies. Our case study accentuates the need in fecal DNA studies concerning all animals, not just bison, to verify individual markers and fecal samples with a matching high-quality DNA sample to validate each marker.

CHAPTER IV

BISON POPULATION PHYLOGENOMICS, SUBSPECIES HYBRIDIZATION, AND CATTLE INTROGRESSION BASED ON SNP ANALYSIS

Introduction

The recovery of few species has been documented and overseen as carefully as that of the American bison. Its demographic history has varied from the time it first crossed the Bering strait to occupy the American continent and spread south all the way to Central America to the end of the 19th century when it was hunted to near-extinction. Since then, thanks to a few conservation-minded entrepreneurs who captured the last few remaining wild bison and raised them on their ranches, they have managed to rebound. However, such an extreme bottleneck leaves a genomic signature that is easily recognizable, even after a hundred years of population expansion. Many of those same ranchers also dabbled in the business of bison-cattle introgression, creating a number of viable and fertile hybrids in the second half of the 19th century, though there are multiple independent records of human-mediated bison-cattle hybridization from naturalists like Pehr Kalm and Albert Gallatin dating back to as early as the beginning of the 18th century (Barsness 1985). These cattle ranchers did not operate in a vacuum – they often exchanged animals between herds, creating a homogenizing effect amongst bison across the United States and Canada. To this day, no large-scale genome-level inquiries have been conducted in order to gain a more complete and accurate picture of how cattle

introgression, bison relocation, and management resulted in the current genetic constitution of bison.

Two subspecies of the American bison (*Bison bison*) are recognized. The plains bison (*Bison bison bison*) has historically ranged from southern Canada to Mexico. Their preferred habitat is the open prairie where they migrate in order to find food during the resource-scarce winter months. The other subspecies is the wood bison (*Bison bison athabascae*) that occupies Alaska and Canada. Larger, darker, and more aloof than their plains counterparts, they prefer spending their lives in the boreal forests. They spend the winters, when resources are scarce, in valleys and plains, but retreat to the forests in the mountains for the warmer months (Christman 1971). Anecdotally, bison that lived in the Rocky Mountains were consistently described as different from plains bison. Observers – contemporary trappers, hunters, and outdoorsmen – described their appearance much like that of wood bison (Hornaday 1889, Christman 1971, Meagher 1973). As a matter of fact, mammalogists of the day even referred to them as "mountain bison" or *Bison athabascae*, the Latin name for the Canadian wood bison today. Christman (1971) goes as far as to state that the only surviving mountain bison herd is the one in Yellowstone National Park based on identified specimens, historical records, bison bones found in the area. He claims that mountain bison have hybridized with plains bison in the region. However, today's post-bottleneck plains bison populations are phenotypically very distinct from any wood bison populations, so little credence has been given to the notion that wood bison once roamed the lower 48 states of the United States.

In order to reconstruct the genetic architecture of contemporary bison, we have sequenced the genomes of plains and wood bison from important lineages in North America, some with no expected introgression levels, others with high introgression levels based on mitochondrial and nuclear microsatellite testing. Some of the herds that had no prior evidence of cattle introgression include the Yellowstone National Park (WY) and the Wind Cave National Park (SD) herds, as well as Canadian plains bison, while those of high known introgression levels include the Santa Catalina Island (CA), and the Caprock Canyon State Park (TX) bison herds. Other herds from which animals were sequenced include ones with unknown levels of cattle introgression, such as the Canadian wood bison populations. We have included a number of historical plains and wood bison that pre-date the last major bottleneck, and potentially some of the introgression events as well.

A whole genome approach can reveal much more information about the population structure, ancestry, hybridization, and introgression of species compared to previous approaches, which were generally limited to analyzing small fragments of the genome, such as microsatellites, single-gene sequencing, or RAD-Seq. In the case of bison, mitochondrial inheritance can be tracked with very high rigor by using a single marker in the D-loop region due to its uniparental inheritance and no dilution effect. On the other hand, our microsatellite test only interrogates a couple dozen nuclear markers, permitting no more than a rudimentary snapshot into the bison genome. Bison have 29 autosomes, and using such a low number of introgression markers have led to entire chromosomes being overlooked. With the advance of genomics and high throughput

short-read sequencing, we can now take an in-depth look at the entire bison genome. However, as with many non-model species, an adequate chromosome-level assembly of bison does not yet exist. Fortunately, domestic cattle (*Bos taurus*) is a closely related species with very similar genomic architecture and excellent genomic resources, which is why it was used as a reference throughout this study without creating substantial bias.

The history of most bison herds in North America are interconnected, dating back to the most recent population bottleneck in the latter half of the 19th century when bison were hunted to near-extinction. A handful of cattle ranchers and entrepreneurs set out to save the species by catching the last few remaining wild bison and keeping them on ranches. Thanks to their efforts, bison have been saved from certain extinction, but the interference created a new issue. Their primary goal was not bison conservation, but livestock husbandry, and most of them experimented with the hybridization of bison and cattle (Coder 1975). They traded and sold animals amongst themselves and eventually these were the bison used in restocking today's national parks and other bison herds. The only place in the United States where bison have existed in the wild continuously since pre-Columbian times is the Yellowstone area, and all other animals were reintroduced from private herds.

The ancestry of most modern herds can be traced back to Yellowstone bison. But Yellowstone itself is not a homogeneous population. As discussed in Chapter II, two different mitochondrial haplogroups (clades) are present in Yellowstone today, the descendants of the indigenous bison that occupied the area for hundreds of years, and the

descendants of the introduced bison that were brought to the park from the Pablo-Allard and Goodnight herds to boost the numbers of the dwindling population.

Some of the animals in the Pablo-Allard herd came from a series of previous herds that practiced bison-cattle hybridization: the McKay-Alloway, the Bedson, and the Buffalo Jones herds (Colpitts 2002). Charles Goodnight was a pioneer in hybridization, even publishing his findings in the Journal of Heredity (Goodnight 1914). Considering that the Pablo-Allard herd sent 18 females and the Goodnight herd sent 3 males to Yellowstone, it is surprising that no cattle mitochondrial DNA has been detected in any Yellowstone bison tested. All the Goodnight bison were all males, which would explain the complete lack of mitochondrial cattle influence while allowing for nuclear introgression. However, Goodnight always claimed that he kept the hybrids physically separated from his non-introgressed herd (Coder 1975), though others disagreed (Department of the Interior 1902). Further analysis of contemporary reports, on the other hand, suggests that the 3 males were in fact picked by Charles "Buffalo" Jones, a prominent bison producer and stout advocate of bison-cattle hybridization. He and his brother N.C. Jones owned the particular bison sent to Yellowstone when they ran Goodnight's hybridization experimental program, and in 1902, as the first game warden of Yellowstone National Park, Buffalo Jones was tasked to handpick the males to be brought to Yellowstone. In several correspondences, he advocated for bison-cattle hybridization experiments, going so far as to advise that the vanished bison herds should be replaced by hybrids all over the United States (Will save the buffalo 1901). Jones also intended to keep up his hybridization experiments in Yellowstone against the wishes of

the Department of the Interior (as seen from his correspondence with his good friend and President of the United States, Theodore Roosevelt), and eventually resigned as game warden in order to commit himself fully to the creation of hybrids (Jones 1905).

Bison from the Pablo-Allard herd, in addition to supplementing the Yellowstone herd, were also later sent up to Canada in large numbers (6,600 bison) that have been sympatric with wood bison thereafter, producing hybrids between the subspecies (Coder 1975, Forgacs et al. 2016). Yellowstone animals have also been a major source of bison for other public herds, such as Wind Cave National Park (another herd with no prior cattle introgression detected), and Catalina Island (founded with approximately equal number of Yellowstone and Goodnight-lineage bison).

The Goodnight herd formed the basis of the Texas State bison herd at Caprock Canyons State Park, but due to their high rates of inbreeding and a very low number of founders, Yellowstone-lineage bison were introduced from the Castle Rock herd owned by Turner Enterprises, Inc (Halbert 2004). Therefore, in some way, virtually all US and Canadian plains bison herds share ancestry with the Yellowstone bison herd.

In order to take an in-depth look at recent genomic changes due to hybridization and the differentiation across ecotypes, a study of killer whales used a combination of maximum likelihood trees, PCA plots, TreeMix directional gene flow analyses, and ABBA-BABA tests. Based on the congruency of the results from those tests, they were able to determine which ecotypes were the closest and which ones were the most distant genetically, and they could assign directionality to the hybridization (Foote et al. 2016). A similar approach proved useful to determine introgression and genetic divergence in

Andean cloud forest birds (Winger 2017), domestic cattle progenitors (Park et al. 2015), and silverleaf whiteflies (Elfekih et al. 2018). Following a similar strategy, in this study, we analyzed the genomes of 40 North American bison from 6 distinct herds in order to assess their levels of cattle introgression, subspecies hybridization, and relatedness to learn more about the demographic history of the American bison and develop a SNP genotyping test for the detection of cattle nuclear and mitochondrial markers.

Methods

Hair, tissue, or bone samples from modern and historical bison from across North America were collected from 25 bison and DNA was extracted at the DNA Technologies Core Lab at Texas A&M University according to the protocols published by Curry (2019) and Curry and Derr (2019). In addition, 15 more bison previously sequenced by Dobson (2015), 21 cattle from four breeds (Angus, Charolais, Hereford, and Holstein) and a water buffalo (*Bubalus bubalis*) were selected for analysis from the NCBI database (Table IV.1). The cattle breeds were chosen strategically to cover breeds that were known to be used in bison-cattle hybridization experiments. However, other breeds that were known to be subject to introgression, such as Galloway cattle and Longhorns could not be included due to a lack of available good quality genome sequences.

Sample	Source	Location	Collection date	Subspecies
10345	Irvine, Colonel A.	unknown	1885	Closer to plains
AL704	CMNMA	Alberta	1910	Wood
Angus_01	NCBI	SRR1425124	-	*Bos taurus*
Angus_02	NCBI	SRR1365144	-	*Bos taurus*
Angus_04	NCBI	SRR1343172	-	*Bos taurus*
Angus_05	NCBI	SRR1365129	-	*Bos taurus*
Angus_06	NCBI	SRR2016763	-	*Bos taurus*
Angus_07	NCBI	SRR4477870	-	*Bos taurus*
Angus_08	NCBI	SRR4280084	-	*Bos taurus*
Angus_10	NCBI	SRR4279959	-	*Bos taurus*
Bubalus bubalis	NCBI	SRR7284794	-	*Bubalus bubalis*
CCSP48	Dobson 2015	Caprock Canyons State Park (TX)	2000	Plains
CCSP50	Dobson 2015	Caprock Canyons State Park (TX)	2001	Plains
CCSP61	Dobson 2015	Caprock Canyons State Park (TX)	2003	Plains
CCSP662	DNA Technologies Lab	Caprock Canyons State Park (TX)	2018	Plains
CCSP667	DNA Technologies Lab	Caprock Canyons State Park (TX)	2018	Plains
CCSP68	Dobson 2015	Caprock Canyons State Park (TX)	2004	Plains
Charolais_03	NCBI	SRR1355258	-	*Bos taurus*
Charolais_04	NCBI	SRR1365145	-	*Bos taurus*
Charolais_06	NCBI	SRR1365122	-	*Bos taurus*
Charolais_07	NCBI	SRR1348578	-	*Bos taurus*
CMNMA10420	CMNMA	unknown	July 1919	Closer to wood
CMNMA55398	CMNMA	Cypress Hills, Saskatchewan	1904	Plains
CMNMA55399	CMNMA	Swift Current, Saskatchewan	1871	Plains
CMNMA7	CMNMA, T. C. Weston	unknown	Prior to 1912	Closer to plains
CPB1803	CMNMA, Charles M. Sternberg	Drumheller, Alberta	August 1912	Plains
CPB5591	CMNMA, T. C. Weston	Junction of Bow and Belly Rivers, Alberta	1883	Plains
CPB8833	CMNMA, Charley Alloway	Prince Alberta, Saskatchewan	1909	Plains
EIP828	DNA Technologies Lab	Elk Island National Park, Alberta	2008	Plains
EIP872	DNA Technologies Lab	Elk Island National Park, Alberta	early-mid 2000s	Plains
EIW151	Dobson 2015	Elk Island National Park, Alberta	early 2000s	Wood
EIW233	Dobson 2015	Elk Island National Park, Alberta	early 2000s	Wood
EIW26	Dobson 2015	Elk Island National Park, Alberta	early 2000s	Wood
EIW788	Parks Canada	Elk Island National Park, Alberta	January 2018	Wood
EIW789	Parks Canada	Elk Island National Park, Alberta	January 2018	Wood

Table IV.1: List and origin of samples.

Sample	Source	Location	Collection date	Subspecies
EIW95	Dobson 2015	Elk Island National Park, Alberta	early 2000s	Wood
Hereford_02	NCBI	SRR1365142	-	*Bos taurus*
Hereford_03	NCBI	SRR1355261	-	*Bos taurus*
Hereford_04	NCBI	SRR1365126	-	*Bos taurus*
Hereford_05	NCBI	SRR1365131	-	*Bos taurus*
Hereford_06	NCBI	SRR2226524	-	*Bos taurus*
Hereford_09	NCBI	SRR4280185	-	*Bos taurus*
Holstein_03	NCBI	SRR1365147	-	*Bos taurus*
Holstein_06	NCBI	SRR934409	-	*Bos taurus*
Holstein_08	NCBI	SRR1425131	-	*Bos taurus*
RRYNP2	DNA Technologies Lab	Red Rock Turner Ranch (MT)	2010s	Plains
S6	Dobson 2015	Dawson County (MT)	1886	Plains
S9	Dobson 2015	Yellowstone National Park (WY)	August 1856	Plains
SCI612	DNA Technologies Lab	Santa Catalina Island (CA)	1990	Plains
SCI662	DNA Technologies Lab	Santa Catalina Island (CA)	1992	Plains
Templeton	Dobson 2015	Yellowstone National Park (WY)	mid 2000s	Plains
WB1896	CMNMA	unknown	1913	Wood
WB299	CMNMA, Warburton Pike	Canada	1892	Wood
WB4538	CMNMA, Parks Canada	Wood Buffalo National Park	1921	Wood
WB625	CMNMA, F. White	Lake Athabasca	unknown	Wood
WBNP700	CMNMA	Wood Buffalo National Park	November 1937	Wood
WC525	DNA Technologies Lab	Wind Cave National Park (SD)	2016	Plains
WC530	DNA Technologies Lab	Wind Cave National Park (SD)	2016	Plains
YNP1856	Dobson 2015	Yellowstone National Park (WY)	early-mid 1990s	Plains
YNP1861	Dobson 2015	Yellowstone National Park (WY)	early-mid 1990s	Plains
YNP5885	Dobson 2015	Yellowstone National Park (WY)	early 2000s	Plains
YNP5899	Dobson 2015	Yellowstone National Park (WY)	late 2000s	Plains

Table IV.1: Continued.

Genomic libraries were prepared and sequenced by Delta Genomics (Edmonton, Alberta, Canada). 100 ng to 1 µg of genomic DNA was sheared using the Covaris S2 focused sonicator (Covaris Inc.) to achieve a fragment size ranging between 300-400 bp. Sheared DNA fragments were used for library preparation according to respective library preparation protocols that were compatible with the Illumina next generation

sequencing platform. Libraries were prepared according to the NEXTflex^{TM} DNA Sequencing Kit protocol (Bio-O Scientific, Austin, TX) and indexed using NEXTflex^{TM} DNA Barcodes (Bio-O Scientific). Size selection of end repaired product during library preparation followed the gel-free size selection clean up processes using Agencourt AMPure XP magnetic beads (Beckman Coulter, Mississauga, Ontario). Finally, 2-10 cycles of PCR amplification were performed on libraries depending on the input DNA concentration. Library quality was analyzed with the DNA1000 chip on the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA) and library quantification was carried out using the Qubit HS assay (Life Technologies). All libraries passing quality control were sequenced through the paired-end sequencing workflow at 2 x 150 cycles on the HiSeqX at McGill University and Génome Québec Innovation Centre. The average genome coverage and maximum depth are listed in Table IV.2.

Sample	Coverage	Maximum depth
10345	11.3679	28.41975
AL704	10.8424	27.106
Angus_01	19.1267	47.81675
Angus_02	18.5725	46.43125
Angus_04	14.8074	37.0185
Angus_05	17.4386	43.5965
Angus_06	11.2209	28.05225
Angus_07	16.1648	40.412
Angus_08	18.2662	45.6655
Angus_10	13.7282	34.3205
Bubalus bubalis	18.8526	47.1315
CCSP48	2.09453	5.236325
CCSP50	2.16277	5.406925
CCSP61	2.14178	5.35445
CCSP662	36.0636	90.159

Table IV.2: Average genome coverage and maximum depth of each sample.

Sample	Coverage	Maximum depth
CCSP667	36.4572	91.143
CCSP68	1.97277	4.931925
Charolais_03	18.8956	47.239
Charolais_04	18.6464	46.616
Charolais_06	16.4929	41.23225
Charolais_07	14.7596	36.899
CMNMA10420	18.2153	45.53825
CMNMA55398	7.59779	18.994475
CMNMA55399	10.8234	27.0585
CMNMA7	15.9824	39.956
CPB1803	9.98444	24.9611
CPB5591	5.70318	14.25795
CPB8833	32.5486	81.3715
EIP828	36.954	92.385
EIP872	41.0589	102.64725
EIW151	1.43945	3.598625
EIW233	2.28936	5.7234
EIW26	2.01788	5.0447
EIW788	46.6065	116.51625
EIW789	45.6908	114.227
EIW95	1.47304	3.6826
Hereford_02	19.2686	48.1715
Hereford_03	18.4451	46.11275
Hereford_04	17.7301	44.32525
Hereford_05	17.329	43.3225
Hereford_06	11.1229	27.80725
Hereford_09	10.2669	25.66725
Holstein_03	19.325	48.3125
Holstein_06	17.9286	44.8215
Holstein_08	17.7626	44.4065
RRYNP2	44.4567	111.14175
S6	11.4163	28.54075
S9	23.2322	58.0805
SCI612	36.7173	91.79325
SCI662	56.9796	142.449
Templeton	27.9834	69.9585
WB1896	8.58449	21.461225
WB299	46.3513	115.87825
WB4538	24.3917	60.97925

Table IV.2: Continued.

Sample	Coverage	Maximum depth
WB625	5.57323	13.933075
WBNP700	30.5641	76.41025
WC525	35.9989	89.99725
WC530	38.7848	96.962
YNP1856	11.9979	29.99475
YNP1861	10.3843	25.96075
YNP5885	7.16443	17.911075
YNP5899	3.20402	8.01005

Table IV.2: Continued.

The fastq files were aligned to the *Bos taurus* genome build ARS-UCD1.2 using the Burrows-Wheeler Aligner (BWA-MEM 0.7.15) (Li and Durbin 2009) and sorted and indexed by SAMtools 1.3.1 (Li et al. 2009). Some alignments were created using SpeedSeq (Chiang et al. 2015). In cases where samples were run on multiple lanes, they were merged using the SAMtools merge option. Next, haplotypes were called for all the aligned bam files using GATK HaplotypeCaller (Van der Auwera et al. 2013) chromosome by chromosome, then merged using the GenotypeGVCFs, CatVariants and CombineGVCFs options. The dataset was pruned using VCFtools 0.1.16 (Danecek et al. 2011), keeping only biallelic SNPs (minimum and maximum number of alleles were set to 2), and removing indels and outliers in the depth of coverage (only keeping variants within the 10x-100x range). The resulting dataset of 96,903,559 variants served as the basis of all downstream analysis and further data pruning. All additional reductions and manipulations of this dataset were performed using VCFtools. In some analyses the population assignment was given based on the sampling location and whether they were historical (1856-1937) or modern (1990-present) samples (Table IV.3).

.Sample	Population	Population abbreviation
CCSP48	Caprock Canyons State Park (TX)	CCSP
CCSP50	Caprock Canyons State Park (TX)	CCSP
CCSP61	Caprock Canyons State Park (TX)	CCSP
CCSP662	Caprock Canyons State Park (TX)	CCSP
CCSP667	Caprock Canyons State Park (TX)	CCSP
CCSP68	Caprock Canyons State Park (TX)	CCSP
10345	historical Canadian plains bison	hist_cpb or CP
CMNMA55398	historical Canadian plains bison	hist_cpb or CP
CMNMA55399	historical Canadian plains bison	hist_cpb or CP
CMNMA7	historical Canadian plains bison	hist_cpb or CP
CPB1803	historical Canadian plains bison	hist_cpb or CP
CPB5591	historical Canadian plains bison	hist_cpb or CP
CPB8833	historical Canadian plains bison	hist_cpb or CP
EIP828	modern Canadian plains bison	modern_cpb or CP
EIP872	modern Canadian plains bison	modern_cpb or CP
AL704	historical Canadian wood bison	hist_wb or WB
CMNMA10420	historical Canadian wood bison	hist_wb or WB
WB1896	historical Canadian wood bison	hist_wb or WB
WB299	historical Canadian wood bison	hist_wb or WB
WB4538	historical Canadian wood bison	hist_wb or WB
WB625	historical Canadian wood bison	hist_wb or WB
WBNP700	historical Canadian wood bison	hist_wb or WB
EIW151	modern Canadian wood bison	modern_wb or WB
EIW233	modern Canadian wood bison	modern_wb or WB
EIW26	modern Canadian wood bison	modern_wb or WB
EIW788	modern Canadian wood bison	modern_wb or WB
EIW789	modern Canadian wood bison	modern_wb or WB
EIW95	modern Canadian wood bison	modern_wb or WB
S6	historical Greater Yellowstone Area (MT)	`hist_YNP or YNP
S9	historical Greater Yellowstone Area (WY)	hist_YNP or YNP
RRYNP2	modern Greater Yellowstone Area (MT)	modern_YNP or YNP
Templeton	modern Greater Yellowstone Area (MT)	modern_YNP or YNP
YNP1856	modern Greater Yellowstone Area (WY, MT)	modern_YNP or YNP
YNP1861	modern Greater Yellowstone Area (WY, MT)	modern_YNP or YNP
YNP5885	modern Greater Yellowstone Area (WY, MT)	modern_YNP or YNP
YNP5899	modern Greater Yellowstone Area (WY, MT)	modern_YNP or YNP
SCI612	Santa Catalina Island (CA)	SCI
SCI662	Santa Catalina Island (CA)	SCI
WC525	Wind Cave National Park (SD)	WC
WC530	Wind Cave National Park (SD)	WC

Table IV.3: Population assignment of the 40 bison in this study.

A series of principal component analyses (PCAs) were performed on a reduced dataset in order to establish patterns in the data between individuals and populations. In

the first iteration, no animals were removed, but all SNPs with missing data for more than 30% of the individuals were discarded and the dataset was further thinned to one variant every 1 kb resulting in 2.3 million variants. In the second iteration, *B. bubalis* was removed and the dataset was reduced to variants with missing data for less than 20% of individuals, resulting in 13.5 million variants. PCA plots were created using the SNPRelate package in RStudio (Zheng et al. 2012). 3D PCA plots were created using the rgl package in RStudio (Adler and Murdoch 2017). Historical and modern Canadian wood bison (CW) and plains bison (CP) were clustered together.

The number of population-specific SNPs were determined for each population by reducing the dataset by all individuals from each population and calculating the difference between the total number of variants between all bison (68,730,679) and the remaining number of variants. For the Canadian plains bison, wood bison, and Yellowstone populations where both historical and modern animals were available, the number of population-specific variants was calculated for the historical, modern, and the combined populations.

Faststructure 1.0 (Raj et al. 2014) was used to determine population structure. The unthinned dataset was transformed into Plink format (Purcell et al. 2007) and the bed file was used to test the number of distinct populations K = 2 through K = 12, which was visualized using Distruct.

A rooted and an unrooted maximum likelihood tree was created using RaxML (Stamatakis 2014) on the Cipres Science Gateway server (Miller et al. 2010) using 1000 bootstraps and the GTR + gamma rate heterogeneity model. The datasets were thinned to

6-8 million variants in order to accommodate the maximum file size allowed on the server. The rooted version of the tree was created using all 40 bison and a Hereford cattle as the outgroup, while the unrooted version only included the 40 bison. The results were visualized using FigTree 1.4.4 (Rambaut 2009).

TreeMix 1.13 (Pickrell and Pritchard 2012) was used to determine the directionality of transfer of genes between populations. First, a representative individual from each of the populations was tested. Templeton (the bison reference genome from Yellowstone), a single animal from each of the four cattle breeds, and the outgroup *Bubalus bubalis* were also included in this analysis. The data was analyzed in blocks of 500 bp, incrementally with up to 5 migration edges allowed. In an independent run, all bison and cattle were assigned to their respective populations, analyzed in blocks of 1000 bp, with up to 8 migration edges allowed.

D-statistics were calculated and analyzed by running the ABBA-BABA test using ANGSD (Soraggi et al. 2018). The final Z scores and D-statistics are based on all variants across all bam files. The test was run on an individual-by-individual basis as well as population-by-population. The results were ranked by Z scores and the non-significant results were discarded ($p > 0.05$). A negative D-statistic suggests that H3 is closer to H1 than H2, while a positive D-statistic points to H3 being closer to H2 than H1. nABBA values reveal how many blocks of variants suggested H3 was closer to H2, while nBABA values show how many times H3 was closer to H1. For the individual-by-individual results, there were simply too many lines to independently show, so for each H3 individual, the number of times each other individual or population (H1 or H2) was

significantly closer was counted and ranked from largest (closer than any other individual or population) to smallest (farther than any other individual or population). The results were compared between the regular run and one that takes ancient transitions into consideration, as cytosine to thymine de-aminations have been documented in S9 mitochondrial DNA, but there was no significant difference between the two runs.

Results

The phylogenetic and population structure analyses performed on the 40 bison, 21 cattle, and a water buffalo yielded some unforeseen and unexpected results. The principal component analysis showed that the introgression status of the animals did not match the a priori assumptions that SCI and CCSP were highly introgressed, YNP, WC, and Canadian plains bison were putatively non-introgressed, and the other populations were potentially introgressed (Figure IV.1). The bison were spread out along PC2, and introgressed and non-introgressed populations were not clearly separated and defined. Though individuals from the same populations sometimes clustered together (like CCSP and WC), that was not always the case (like YNP and CW), indicating members of those populations had varying levels of introgression (Figure IV.2). A 3-dimensional PCA excluding *B. bubalis* showed a clear division between cattle and bison along PC3, but not between plains and wood bison along PC2 (Figure IV.3). The PCA results indicated that while bison and cattle are significantly different from each other, the two subspecies, and the introgressed and non-introgressed herds are not.

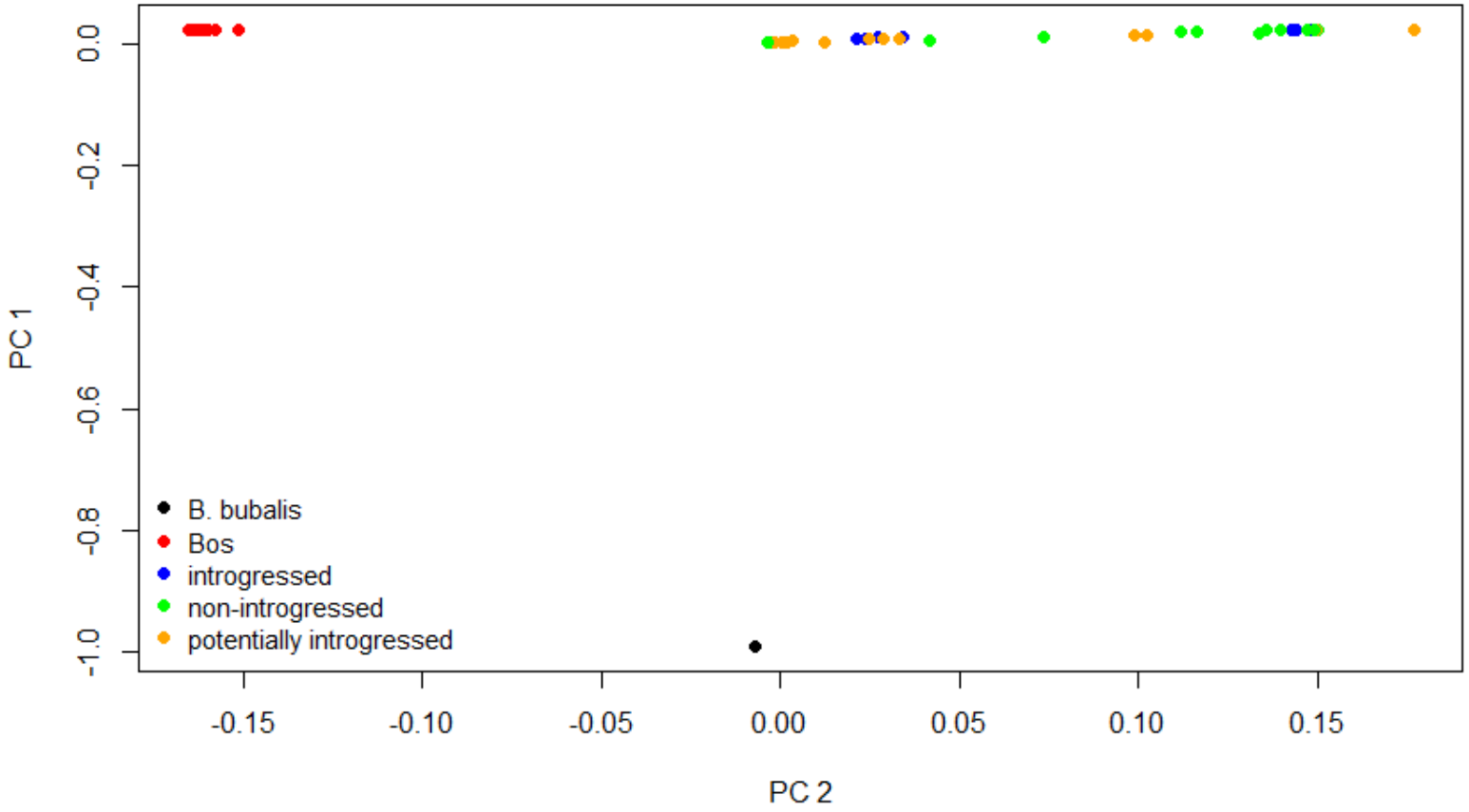


Figure IV.1: PCA based on a priori introgression status. PC1 = 58.9%, PC2 = 13.2%.

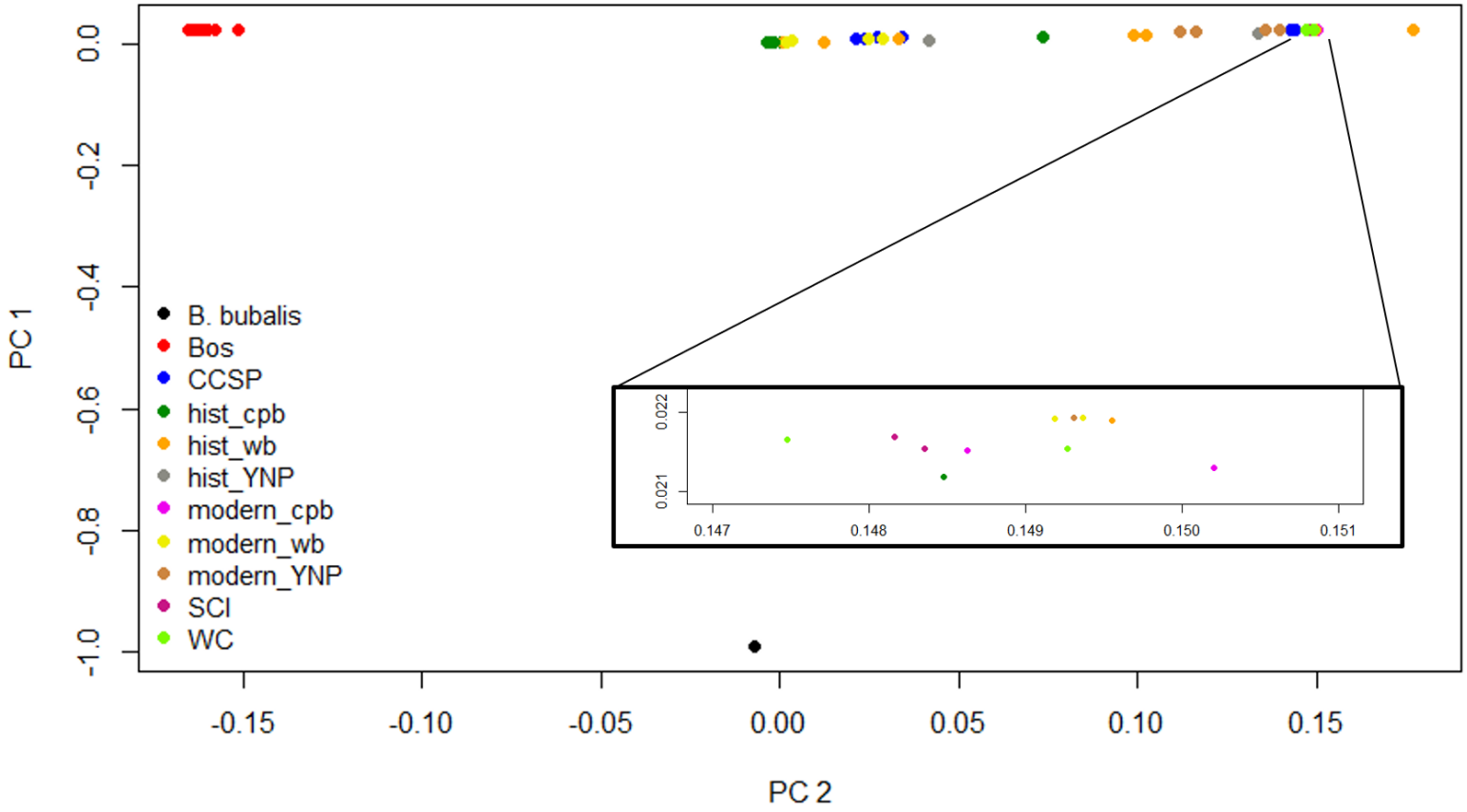


Figure IV.2: PCA based on bison population assignment. PC1 = 58.9%, PC2 = 13.2%.

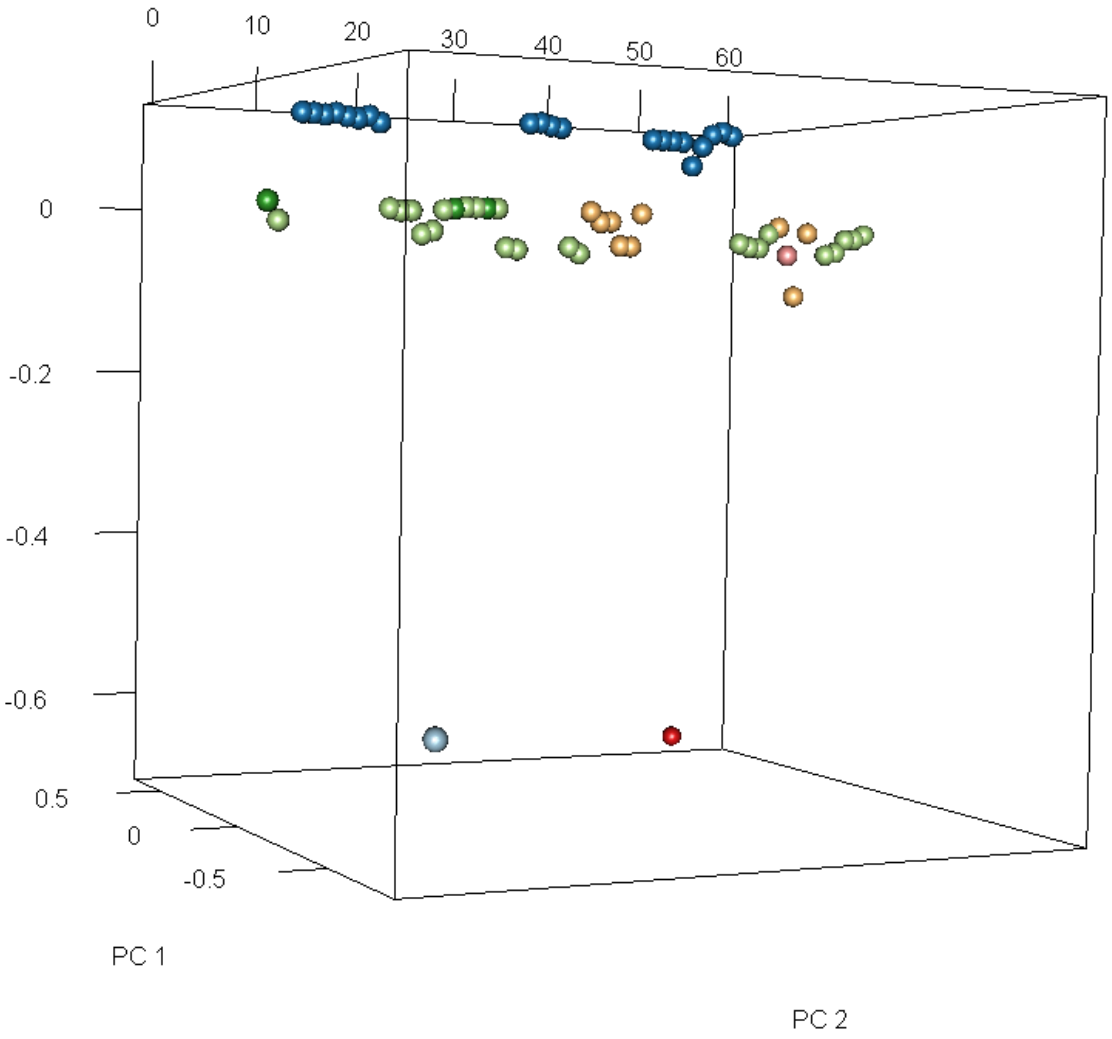


Figure IV.3: 3-D PCA shows separation between cattle and bison, but not between plains and wood bison. Blue represents *Bos taurus* cattle, plains bison are green, wood bison are yellow, and red and silver are two Canadian wood bison, AL704 and WBNP700 respectively. PC1 = 13.3%, PC2 = 12.2%, PC3 = 9.2%.

The number of population-specific SNPs demonstrated that wood bison do not have more unique variants compared to all the plains bison groups (Table IV.4). This is likely due to extensive hybridization between plains and wood bison. Populations with less unique variation are more similar to other populations tested, and therefore have a more recent shared demographic history. The populations with the largest numbers are the most unique.

Population	Population-specific SNPs
historical Canadian plains bison	22,143
all Canadian plains bison	80,484
historical Yellowstone area bison	196,100
Caprock Canyons State Park bison	199,581
all Yellowstone bison	285,986
modern Yellowstone bison	1,518,141
historical wood bison	3,049,047
Santa Catalina Island bison	7,513,537
modern wood bison	8,049,883
Wind Cave National Park bison	9,425,484
all wood bison	9,906,227
modern Canadian plains bison	10,635,748

Table IV.4: Number of population-specific SNPs. The total number of SNPs amongst the 40 bison tested was 68,730,679.

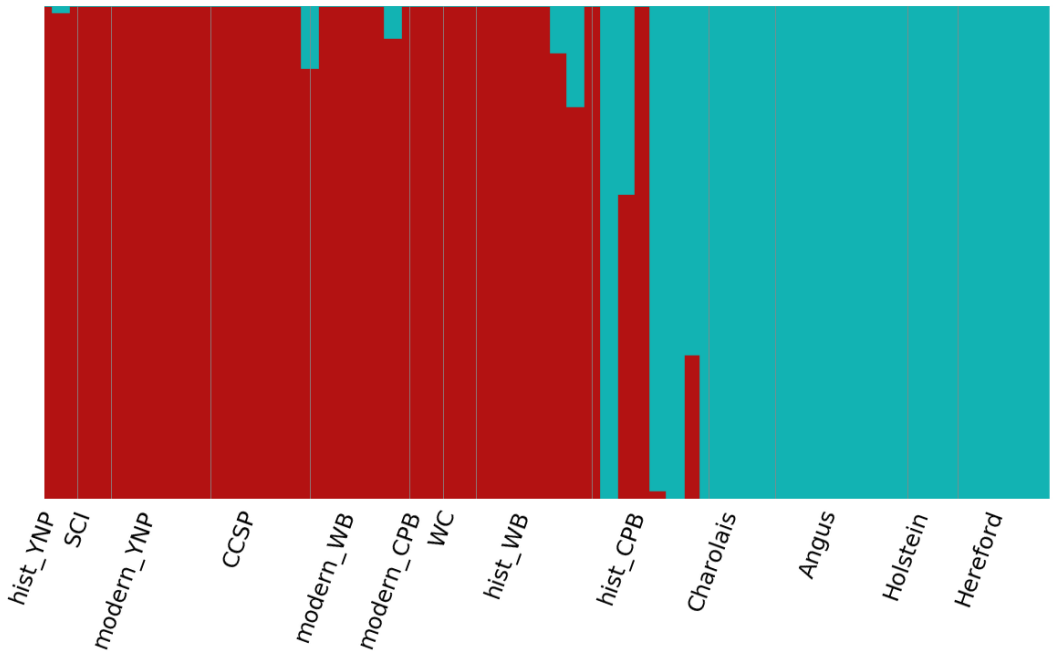


Figure IV.4: Population structure of bison and cattle at k = 2.

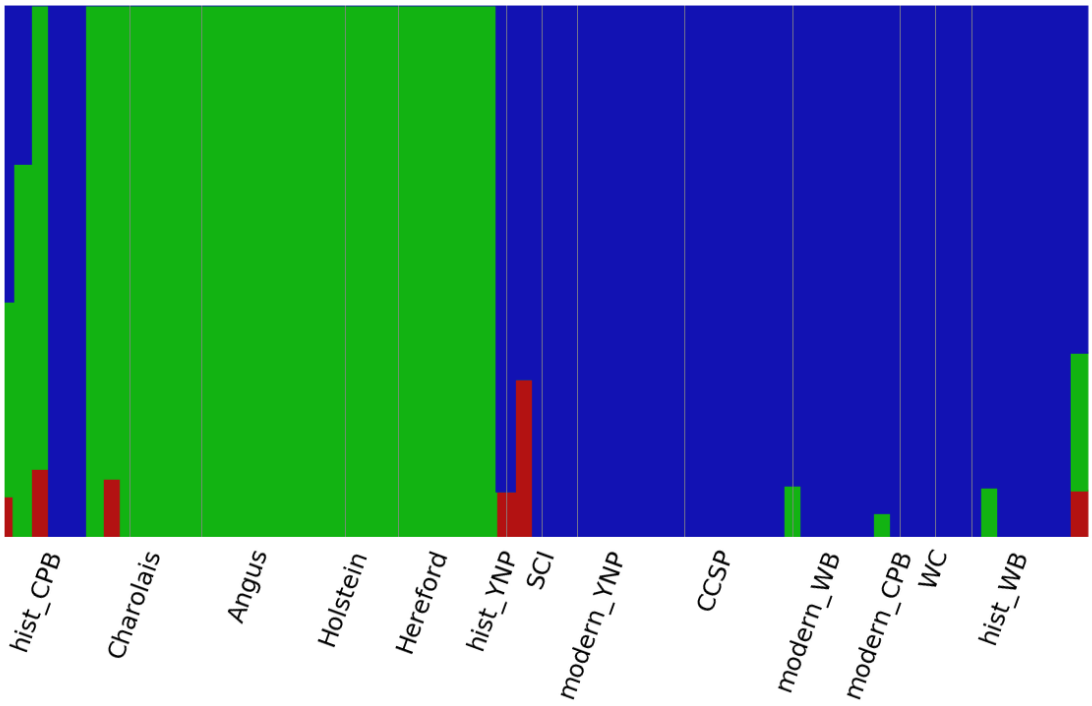


Figure IV.5: Population structure of bison and cattle at k = 3.

Between the 61 cattle and bison in this study, the number of likely distinct populations was determined as 2 or 3 by fastSTRUCTURE. At k = 2, one of the populations incorporated all cattle and some highly introgressed bison from historical and modern wood bison, and historical Canadian plains bison; while the other clade included the remainder of the bison (Figure IV.4). At k = 3, the third cluster represented genetic material shared by all historical bison populations (plains, wood, and Yellowstone bison) (Figure IV.5).

The maximum likelihood phylogeny identified a number of historical bison that were more closely related to cattle than any of the bison from this study, indicating high levels of cattle introgression (Appendix D). In addition, we identified two monophyletic clades with limited bootstrap support: plains bison and wood bison. Within the plains bison clade, modern Yellowstone bison, modern Canadian plains bison, and Caprock Canyon State Park bison each form a monophyletic clade, while Santa Catalina Island and Wind Cave National Park bison form two closely related sister groups. The unrooted maximum likelihood tree still assorts the highly introgressed bison and the plains bison in a similar manner, but in this version, the wood bison are split into two groups: one that is significantly divergent from plains bison (bootstrap = 96) and one where the separation from the plains bison clade is poorly supported (bootstrap = 31) (Figure IV.6, Figure IV.7).

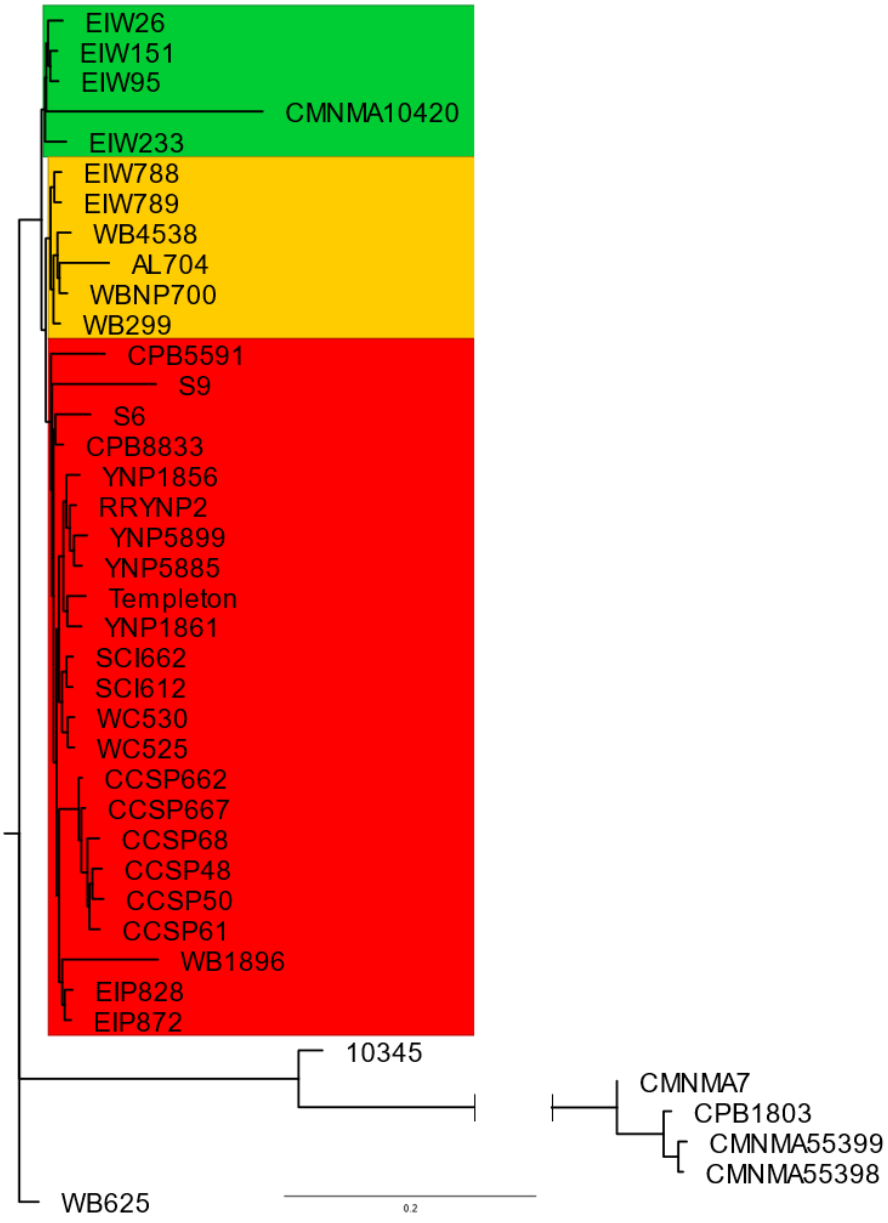


Figure IV.6: Unrooted maximum likelihood phylogram. The wood bison clade with strong bootstrap support (96) is highlighted in green, the wood bison clade with weak bootstrap support (31) is highlighted yellow, and the plains bison clade is highlighted red.

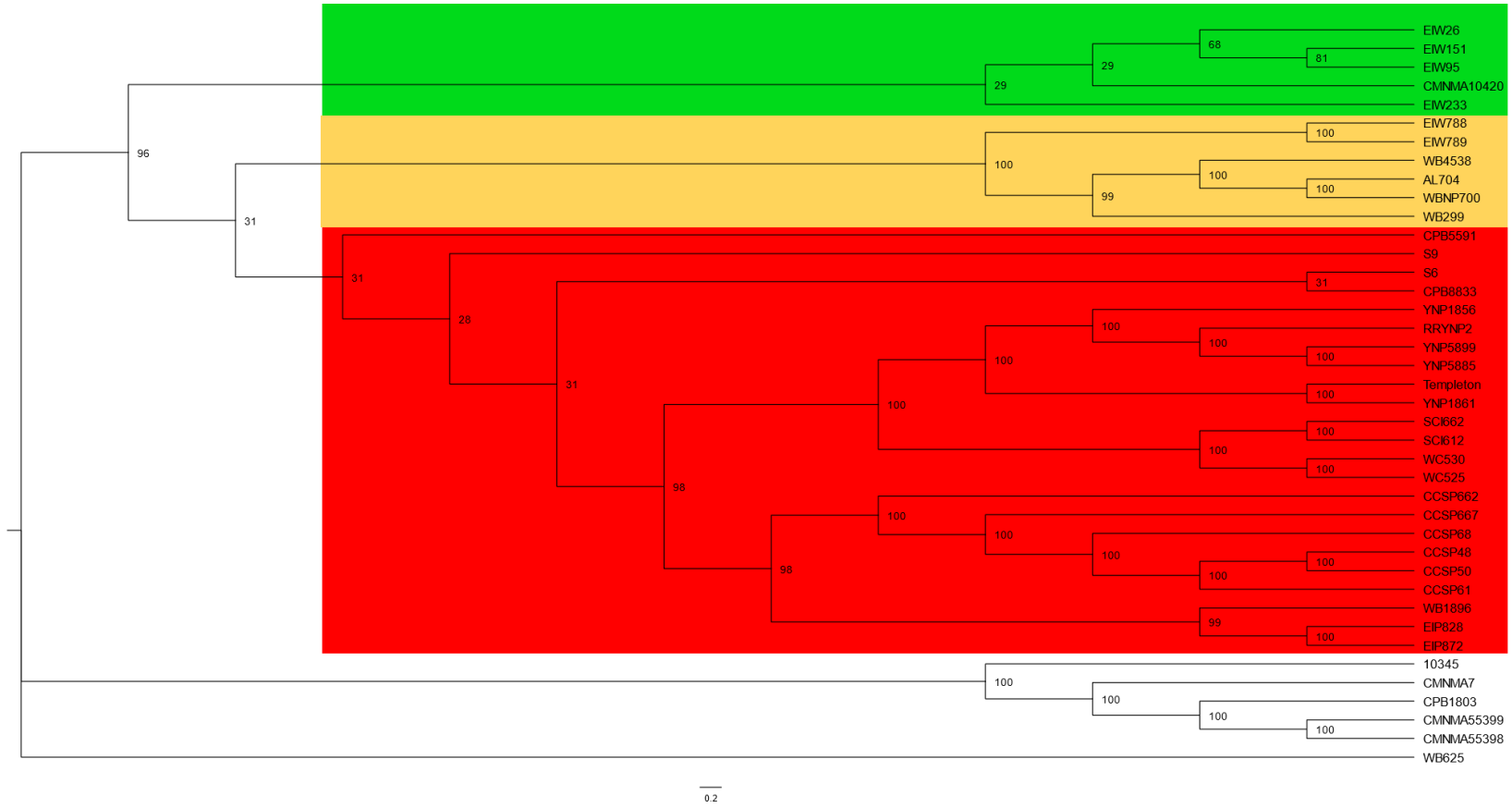


Figure IV.7: Unrooted maximum likelihood cladogram. The wood bison clade with strong bootstrap support (96) is highlighted in green, the wood bison clade with weak bootstrap support (31) is highlighted yellow, and the plains bison clade is highlighted red. Bootstrap values are displayed for each node.

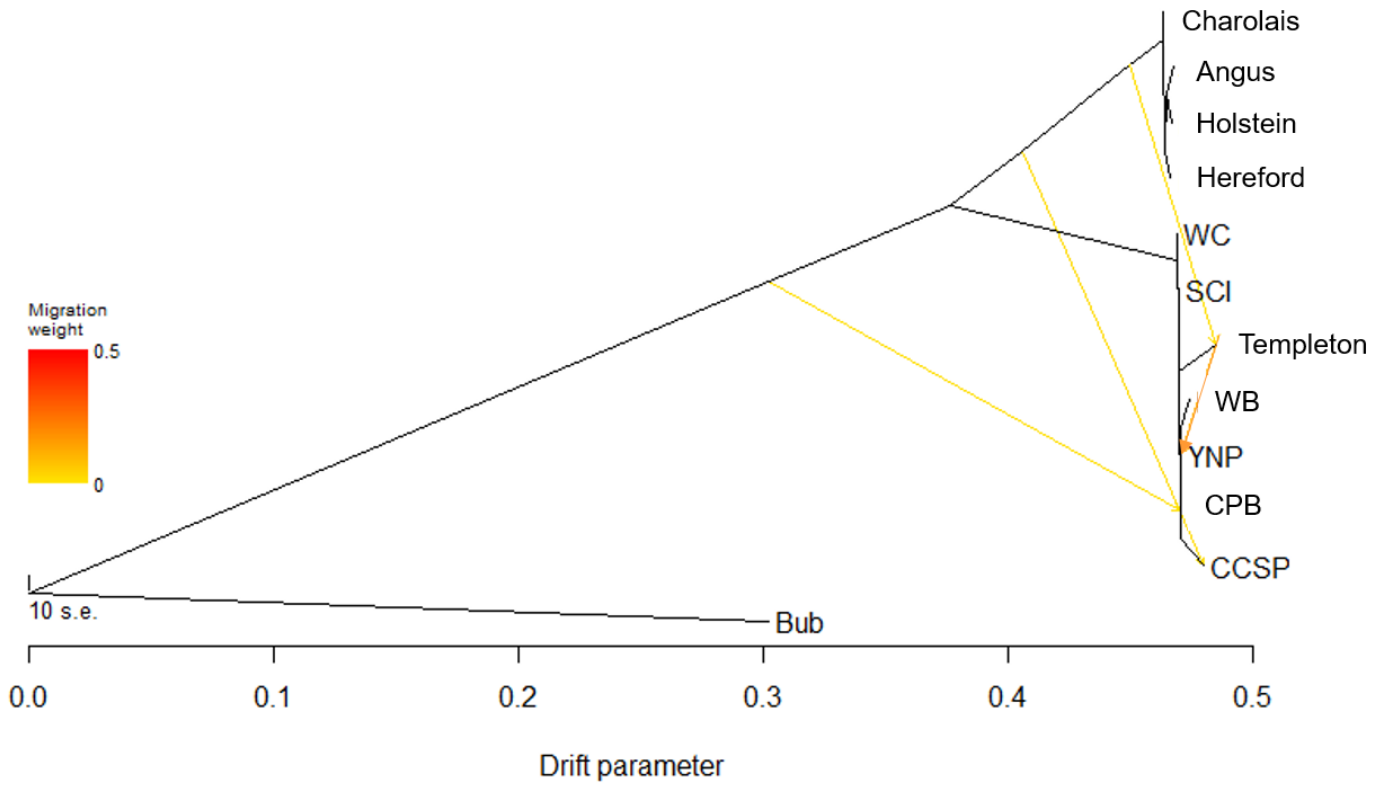


Figure IV.8: Admixture analysis based on a representative member from each population. Templeton, the bison reference genome, was also included in the analysis.

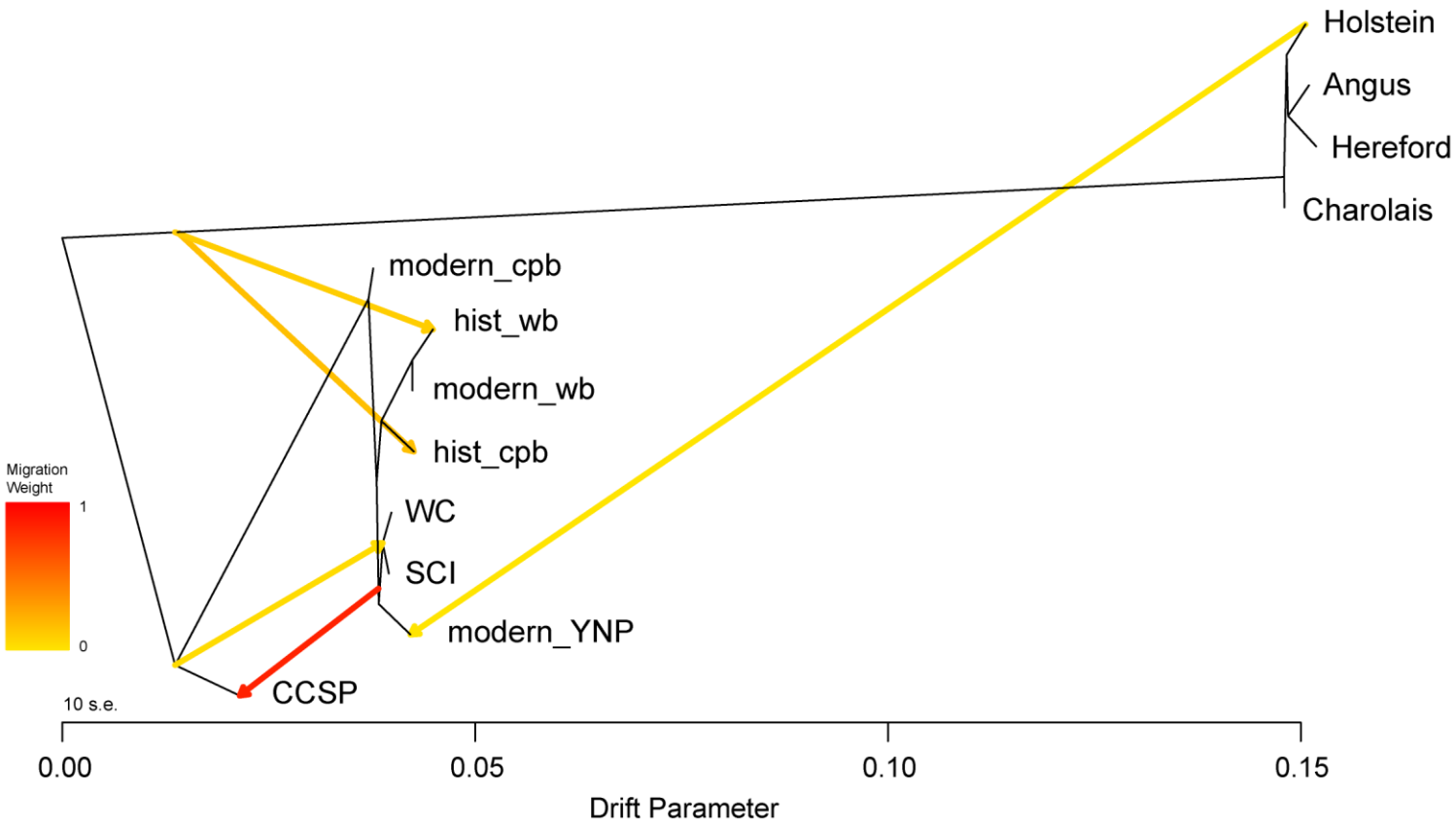


Figure IV.9: Admixture analysis based on all members of each population.

The inference of the patterns of gene flow and admixture revealed the populations with the highest levels of cattle introgression. Whether only one representative member of each population was used (Figure IV.8) or all 40 bison were included (Figure IV.9), the emerging pattern suggested that the most highly introgressed herds (measured by the strongest gene flow from cattle) supported by both TreeMix analyses are Canadian plains bison and modern Yellowstone bison. There is further evidence identifying cattle introgression in historical wood bison, and Caprock Canyons bison based on one but not both of the admixture analyses.

The subspecies of three of the historical Canadian bison were not known. We used population-level ABBA-BABA to compare each to historical wood and plains bison in order to determine which one was the closer match. They were all significantly closer to one of the two subspecies, CMNMA7 and 10345 to plains bison and CMNMA10420 to wood bison (Table IV.5).

D	Z	nABBA	nBABA	H1	H2	H3
0.0672	41.255	77709	67929	hist_wb	**hist_cpb**	10345
0.504	163.540	203665	67078	hist_wb	**hist_cpb**	CMNMA7
-0.074	-87.104	105416	122318	**hist_wb**	hist_cpb	CMNMA10420

Table IV.5: Subspecies determination of unknown bison using the ABBA-BABA test. The group that is closer to the test group H3 is bolded in each row. $p < 1 \times 10^{-6}$ for each row.

There is also an abundance of evidence for the hybridization of plains and wood bison (Table IV.6). Historical plains bison are more similar to historical wood bison than to

many of the modern plains bison herds. The Canadian hybridization event is recent as there is no evidence of hybridization between historical plains and wood bison, but there is evidence of a separate hybridization event between wood bison and Yellowstone bison (Table IV.6).

D	Z	nABBA	nBABA	H1	H2	H3
-0.133	-39.535	1065900	1392904	**modern_wb**	CCSP	hist_cpb
-0.107	-66.442	1065901	1322524	**modern_wb**	modern_cpb	hist_cpb
-0.104	-55.030	1070661	1319240	**modern_wb**	SCI	hist_cpb
-0.108	-50.337	1051954	1307690	**modern_wb**	WC	hist_cpb
-0.032	-20.017	1026566	1093440	**modern_wb**	YNP	hist_cpb
0.065	54.161	1555096	1366491	hist_wb	**modern_cpb**	hist_cpb
0.066	42.37	1574125	1378279	hist_wb	**SCI**	hist_cpb
0.065	34.0367	1550135	1361435	hist_wb	**WC**	hist_cpb
0.137	91.427	1561300	1185690	hist_wb	**YNP**	hist_cpb
-0.131	-73.423	1130944	1472835	**modern_wb**	modern_cpb	YNP

Table IV.6: ABBA-BABA results showing introgression between Canadian plains and wood bison. The introgression appears to be recent because it is affecting modern, but not historical plains bison, while the hybridization of Yellowstone bison and wood bison seems to be a separate event. The group that is closer to the test group H3 is bolded in each row. $p < 1 \times 10^{-6}$ for each row.

Both the individual-by-individual and a population-level ABBA-BABA tests demonstrated that the two Santa Catalina Island bison comprised the least introgressed herd tested (Table IV.7 and Table IV.8). The closest populations to Angus, Charolais, and Holstein cattle were the other three cattle breeds followed by the most heavily introgressed bison. However, Hereford cattle were closer to three of the heavily introgressed historical Canadian plains bison (CMNMA55398, CMNMA55399, and CPB1803) than to any of the other cattle breeds tested.

Angus	Proximity	Charolais	Proximity	Hereford	Proximity	Holstein	Proximity
Holstein	42	Angus_01	42	CMNMA55398	42	Angus_01	42
Hereford	41	Holstein_03	41	CMNMA55399	41	Hereford_04	41
Charolais	40	Hereford_04	40	CPB1803	40	Charolais_03	40
CMNMA55398	39	CMNMA55398	39	Angus_01	39	CMNMA55398	39
CMNMA55399	38	CMNMA55399	38	Holstein_03	38	CMNMA55399	38
CPB1803	37	CPB1803	37	Charolais_03	37	CPB1803	37
CMNMA7	36	CMNMA7	36	CMNMA7	36	CMNMA7	36
10345	35	10345	35	10345	35	10345	35
WB1896	34	WB1896	34	WB1896	34	WB1896	34
CPB5591	33	CPB5591	33	CPB5591	33	CPB5591	33
WB625	31	WB625	31	WB625	31	CCSP68	31
WBNP700	31	WBNP700	31	WBNP700	31	WB625	31
CCSP68	30	CCSP68	30	CCSP68	30	WBNP700	30
CCSP61	29	CCSP61	29	CCSP61	29	CCSP61	29
CCSP48	27	CCSP48	27	EIW233	28	CCSP48	27
EIW233	27	EIW233	27	AL704	26	EIW233	27
AL704	26	AL704	26	CCSP48	26	AL704	26
CCSP50	25	CCSP50	25	CMNMA10420	26	CCSP50	25
CMNMA10420	25	CMNMA10420	25	CCSP50	24	CMNMA10420	25
EIW26	23	EIW26	23	EIW26	23	EIW26	23
CCSP667	21	CCSP667	21	CCSP667	21	CCSP667	21
YNP5885	21	YNP5885	21	YNP5885	21	YNP5885	21
YNP5899	20	YNP5899	20	YNP5899	20	YNP5899	20
YNP1861	18	YNP1856	19	YNP1856	18	YNP1861	18
YNP1856	17	YNP1861	18	S9	17	CCSP662	17
CCSP662	16	S9	15	YNP1861	17	YNP1856	17
Templeton	16	Templeton	15	Templeton	15	S9	15
S9	15	WB299	15	WB299	15	Templeton	15
EIW95	14	CCSP662	14	CCSP662	14	WB299	15
WB299	14	EIW95	14	EIW95	14	EIW95	14
EIW789	13	EIW789	13	EIW789	13	EIW789	13
EIW151	11	EIW151	11	EIW151	12	RRYNP2	11
RRYNP2	11	EIW788	11	S6	11	EIW151	10
EIW788	10	RRYNP2	10	EIW788	9	EIW788	10
S6	9	S6	9	RRYNP2	9	S6	9
WC530	7	WC530	7	WB4538	7	WC530	7
WB4538	6	WB4538	6	WC530	6	WB4538	6
EIP828	5	EIP828	5	EIP828	5	EIP828	5
SCI612	4	SCI612	4	SCI612	4	SCI612	4
CPB8833	3	CPB8833	3	CPB8833	3	CPB8833	3
SCI662	2	SCI662	2	SCI662	2	SCI662	2
WC525	1	WC525	1	WC525	1	WC525	1
EIP872	0	EIP872	0	EIP872	0	EIP872	0

Table IV.7: Distance of each individual from the four cattle herds tested. Proximity is a measure of how many times the individual tested was closer to each cattle herd than another animal tested.

D	Z	nABBA	nBABA	H1	H2	H3
-0.395	-140.917	681819	1573282	**hist_cpb**	SCI	*Bos*
-0.231	-100.152	633764	1013839	**hist_wb**	SCI	*Bos*
-0.066	-25.563	535324	611135	**modern_wb**	SCI	*Bos*
0.067	35.991	941393	823796	SCI	**hist_YNP**	*Bos*
0.015	5.321	652624	633027	SCI	**WC**	*Bos*
0.074	29.711	622420	536419	SCI	**YNP**	*Bos*

Table IV.8: ABBA-BABA results demonstrating the distance of Santa Catalina Island bison is greater from cattle than a number of herds previously thought to be non-introgressed. $p < 1 \times 10^{-6}$ for each row.

YNP5885	YNP5899	YNP1856	YNP1861	Templeton	RRYNP2	Proximity
YNP5899	YNP5885	YNP5885	YNP5899	YNP1861	YNP5899	42
RRYNP2	RRYNP2	YNP5899	YNP1856	RRYNP2	YNP5885	41
YNP1856	YNP1861	YNP1861	YNP5885	YNP5899	YNP1856	40
YNP1861	YNP1856	RRYNP2	RRYNP2	YNP1856	YNP1861	39
Templeton	Templeton	Templeton	Templeton	YNP5885	EIW788	38
EIW788	EIW788	EIW788	EIW788	EIW789	EIW789	37
EIW789	EIW789	EIW789	EIW789	EIW788	Templeton	36
WB299	WB299	WB299	WB299	WB299	WB299	35
WC530	SCI612	WC530	WC530	WC530	WC530	34
SCI612	WC530	SCI612	SCI612	SCI612	SCI612	33
EIP828	WC525	WC525	WC525	SCI662	WC525	32
WC525	EIP828	EIP828	SCI662	WC525	SCI662	31
SCI662	SCI662	SCI662	EIP828	EIP828	EIP828	30
EIW151	EIW151	EIW151	EIW151	EIW151	EIP872	29
EIP872	EIP872	EIP872	EIP872	EIW95	CCSP662	28
CCSP662	EIW95	CCSP662	CCSP662	EIP872	CCSP667	27
EIW95	CCSP667	EIW95	EIW95	CCSP667	EIW151	26
CCSP667	CCSP662	CCSP667	CCSP667	CCSP662	EIW95	25
CPB8833	CPB8833	CPB8833	CPB8833	CPB8833	CPB8833	24
CCSP50	CCSP50	CCSP50	CCSP50	CCSP50	CCSP50	23
EIW26	EIW26	EIW26	EIW26	EIW26	EIW26	22
CCSP48	CCSP48	CCSP48	CCSP48	CCSP48	CCSP48	21
CCSP61	CCSP61	CCSP61	CCSP61	CCSP61	CCSP61	20
CCSP68	CCSP68	CCSP68	CCSP68	CCSP68	CCSP68	19
EIW233	EIW233	EIW233	EIW233	EIW233	EIW233	18
S6	S6	S6	S6	S6	AL704	17
WBNP700	WBNP700	AL704	AL704	WBNP700	WB4538	16
WB4538	AL704	WB4538	WB4538	AL704	S6	15
AL704	WB4538	WBNP700	WBNP700	WB4538	WBNP700	14
WB625	WB625	WB625	WB625	WB625	WB625	13
CPB5591	CPB5591	CPB5591	CPB5591	CPB5591	CPB5591	12
WB1896	WB1896	WB1896	WB1896	WB1896	WB1896	11
S9	S9	S9	S9	S9	S9	10
10345	10345	10345	10345	10345	10345	9
Hereford_04	Hereford_04	Hereford_04	Hereford_04	Hereford_04	Hereford_04	7
CMNMA7	CMNMA7	CMNMA7	CMNMA7	CMNMA7	CMNMA7	7
Holstein_03	Holstein_03	Holstein_03	Holstein_03	Holstein_03	Holstein_03	6
Charolais_03	Charolais_03	Charolais_03	Charolais_03	Charolais_03	Charolais_03	5
Angus_01	Angus_01	Angus_01	Angus_01	Angus_01	Angus_01	4
CMNMA10420	CMNMA10420	CMNMA10420	CMNMA10420	CMNMA10420	CMNMA10420	4
CPB1803	CPB1803	CPB1803	CPB1803	CPB1803	CPB1803	2
CMNMA55398	CMNMA55398	CMNMA55398	CMNMA55398	CMNMA55398	CMNMA55398	1
CMNMA55399	CMNMA55399	CMNMA55399	CMNMA55399	CMNMA55399	CMNMA55399	0

Table IV.9: Distance of each individual from the six contemporary Yellowstone bison tested. The proximity is a measure of how many times the individual tested was closer to each Yellowstone bison than another animal tested.

A strong connection exists between Yellowstone- and Yellowstone-derived herds and wood bison (Table IV.9). All six modern Yellowstone bison are closer to a handful of modern wood bison than to any of the other plains bison herds. The wood bison that were adjacent to Yellowstone bison in the individual ABBA-BABA test were the same as the ones that did not form a well-supported clade separate from plains bison based on the unrooted maximum likelihood tree (Figure IV.7).

Discussion

The initial scope of this research project was to develop a SNP-based genotyping assay for the detection of cattle introgression in bison. However, the results suggested that all of the bison herds tested in this study – including Yellowstone National Park, Wind Cave National Park, and Canadian plains bison herds previously believed to be free of any signs of introgression – have some level of hybridization with cattle. This is supported by the PCA results that showed anticipated introgressed and non-introgressed herds cluster together, both individual and population-level ABBA-BABA tests that revealed that a bison herd with known high levels of introgression is the least introgressed herd tested, the TreeMix results that demonstrated directional gene flow from cattle to several bison herds, and several independent historical sources tracing the origin of many founding populations to a number of bison-cattle hybrids.

The Yellowstone National Park bison herd is key to understanding the impact of historical events on modern bison genetics. Considerable information is known about the origin of the herd, and there were only a limited number of founding events. Before

1902, only a small herd of indigenous bison at Yellowstone survived the intense hunting effort of the preceding century. In 1902, bison were introduced to the Park from two sources. The largest influx of bison was 18 females from the Pablo-Allard herd, a herd that was made up of both bison captured from the wild and bison acquired from known introgressed herds. While the historical records state that the hybrids and non-hybrids were kept separate on the Pablo-Allard ranch, the most convincing argument the 18 females sent to the Park could not have been hybrids is the lack of mitochondrial introgression in Yellowstone bison.

The nuclear markers currently used for the detection of bison-cattle introgression are few and cover a very small area of the genome – the 14 microsatellite markers span fewer than half the chromosomes – and they were selected based on the assumption that the Yellowstone and Canadian plains bison are non-introgressed (Schnabel 2001, Halbert 2003). As a result, any nuclear introgression in those two herds would go undetected by the test, whereas some markers may not have the ability to scrutinize between bison and cattle, but instead between those Yellowstone/Canadian plains bison herds and markers present in the rest of bison and cattle. Due to the design of that introgression test, any cattle-like markers in those two herds would be concealed from the introgression test. The current mitochondrial DNA test, on the other hand, is based on a fixed difference between a variety of cattle breeds and a wide range of bison, and mtDNA does not get diluted over generations of back-crossing due to the unilaterally maternal inheritance of the mitochondrial genome. Hence, the current mitochondrial test would have been able to detect any signs of cattle introgression in the maternal lineage.

The chance that some of those 18 Pablo-Allard bison were hybrids, but none of them contributed cattle mitochondria to Yellowstone, is insignificant.

The other source of bison brought into Yellowstone was Charles Goodnight's bison. Though he was also keen on hybridization experiments, historical records show he also kept his hybrids separate from the ones he caught from the wild. However, the 3 bulls did not come from his wild herd: they were picked by Buffalo Jones who Goodnight put in charge of his introgressed herd. Jones was a big proponent of bison-cattle hybridization, advocating for the culture of cross-breeding and repopulating the federal herds with introgressed bison. As the first game warden of Yellowstone National Park, he had the opportunity to pick the three bulls from his introgressed herd and introduced them to Yellowstone as the only breeding males in the introduced herd (Will save the buffalo 1901, Jones 1905, Coder 1975). This explains why there has never been any evidence of mitochondrial introgression in Yellowstone bison. Though it was only 3 bulls, the effect that those Goodnight bison had on the herd is undeniable. Bison from Caprock Canyons State Park, the herd made up mostly of the descendants of the Goodnight bison, are significantly closer to modern Yellowstone bison than to historical (pre-1902) Yellowstone area bison ($D = 0.364$, $p < 1 \times 10^{-6}$).

Yellowstone served as a source of bison in the past century to nearly all other major herds in the United States and Canada. Yellowstone-lineage animals were sent to Wind Cave National Park (which also only shows nuclear cattle introgression), Caprock Canyons State Park, and Santa Catalina Island. Some of the Canadian bison in this study are early generation hybrids, showing a lot more cattle-like genetic variants than all

other bison. Such plains bison include CMNMA55398, CMNMA55399, CPB1803, CMNMA7, 10345, and the wood bison CMNMA10420. The historical evidence for where the cattle introgression came from is harder to pinpoint, but we know that the McKay-Alloway herd was actively involved in hybridization, Goodnight bison were sent to Banff National Park, and 6,600 bison from the Conrad herd (the successor of the Pablo-Allard herd) were sent to Elk Island National Park (Coder 1975, COSEWIC 2013).

Most historical records agree that bison in the Yellowstone area prior to the 1902 introduction of non-native bison were markedly different from plains bison. They refer to them as mountain bison, using the same Latin nomenclature as to the wood bison subspecies. They describe them as darker, stouter, and habitually more in favor of heavily wooded mountainous areas. All of these point to the fact that the indigenous Yellowstone bison may have been wood bison that occupied the slopes of the Rocky Mountains all the way down to New Mexico (Hornaday 1889, Christman 1971, Meagher 1973). Others, however, disagree and state no difference between plains bison and mountain bison existed (Garretson 1938).

A very strong connection between wood bison and not just Yellowstone, but also Yellowstone-derived herds was revealed using the ABBA-BABA test, suggesting some modern Yellowstone bison are closer to historical wood bison than to other bison from Yellowstone (Table IV.9). In the unrooted maximum likelihood phylogeny, there are two wood bison groups. One of the clades is separate from the plains bison, potentially showing wood bison with little or no hybridization with plains bison. The other clade has

low bootstrap support, potentially displaying more recently introgressed plains and wood/mountain bison. These bison are also the ones that appear to be closer to Yellowstone bison than any other plains bison herd based on the ABBA-BABA test, giving considerable support to the notion that the some of the descendants of the Yellowstone bison may have been mountain bison. On the other hand, the rooted maximum likelihood tree shows all (not heavily introgressed) wood bison as a single, albeit not strongly supported monophyletic clade (Appendix D), and the TreeMix analysis reveals no strong gene flow between wood bison and Yellowstone bison. Yet, the only way the Santa Catalina Island bison in California could be genetically close to wood bison is if wood bison were present in Yellowstone at the time of the foundation of the SCI herd in 1924. This strongly suggests that the genetic proximity between wood and plains bison is not just the result of Pablo-Allard herd animals being sent to both Yellowstone and Canada to interbreed with local bison. In that case, the proximity of Catalina bison to Canadian plains bison and historical Yellowstone bison would have been greater than their proximity to wood bison. We also found that modern Yellowstone bison are significantly closer to wood bison than to modern Canadian plains bison, suggesting widespread hybridization between the subspecies prior to their Canadian introduction (Table IV.6, Table IV.9). Each of the tests performed have their strengths and weaknesses: while the maximum likelihood trees are only based on a subset of SNPs, the ABBA-BABA results that are based on all variants may be misleading when there has not been enough divergence between the populations. In cases of very recent divergence, the two lineages have simply not had enough time to

accumulate enough unique genetic variation to be detectable using the ABBA-BABA test.

Population structure analysis of bison very distinctly reveals the two major forces that shaped all modern bison, and unfortunately both are human-mediated and irreversible. The first and biggest is bison-cattle introgression that has left bison with notable genomic changes (Figure IV.4) that appear to permeate all modern bison. The second is the enormous effect that the most recent population bottleneck had on bison, as most differences between bison that are not heavily introgressed is variation that was present in all historical but no modern herds. This means that any two modern bison herds tested – including plains and wood bison herds – share more of their genetic make-up with each other than with historical bison (Figure IV.5). The enormous loss of bison diversity can also be seen by the vast distance between modern and historical Yellowstone bison, suggesting the historical bison in this analysis were from lineages that are now extinct (Figure IV.9).

Further analysis needs to be performed on a larger number of bison genomes, including herds that have no history of post-1902 Yellowstone-lineage bison in secluded Canadian herds, such as the Mackenzie Bison Sanctuary in the Northwest Territories of Canada. In addition, including a larger set of cattle breeds may provide extra information, such as in the case of Hereford cattle, which are closer to CMNMA55398, CMNMA55399, and CPB1803 (three highly introgressed bison) than to any of the other cattle breeds. Therefore, it is clear that the introgression in those bison came from Hereford cattle or a very closely related breed. Moreover, analyzing a sliding window

phylogeny could pinpoint the exact regions of cattle introgression in bison, and the function of nearby regions and genes could be investigated and linked to existing phenotypes.

In conclusion, this study of 40 bison genomes revealed the high degree of relatedness between modern bison herds due to a great deal of human-mediated gene flow between the populations causing a homogenizing effect, and the multiple recent hybridization events between the plains and the wood bison subspecies. At this point, a definitive bison-cattle introgression test is not feasible since non-introgressed bison weren't actually included in the analysis. Even if a small group of non-introgressed bison were found, it could only be an appropriate test for the amount of introgression, and not the presence/absence of it, as it seems no bison herd in the United States is completely devoid of cattle introgression. This is a cautionary tale for conservation geneticists in the future never to repeat the mistakes of the past by artificially hybridizing, and subsequently releasing introgressed organisms into the wild.

CHAPTER V

CONCLUSIONS

In 1902, a national survey revealed that the number of bison left in the United States was 1,143 – most of which were in private herds, many of those hybrids. Only 72 bison were counted in the wild: 50 in Colorado, and 22 in Yellowstone National Park. This survey set in motion the introduction of 18 Pablo-Allard lineage females and 3 Goodnight lineage males to the indigenous Yellowstone bison. Canadian bison did not fare much better. In the same year, 669 Canadian bison were counted, 600 of which were wild, the remaining 69 were in private hands, some of the latter were also introgressed (Department of the Interior 1902). A manic scramble ensued, establishing federal herds restocked with privately raised bison. As most of the private herds consisted of already introgressed bison and their introgression status was rarely known with any degree of certainty, a number of introgressed animals were used to establish the bison herds of today. This is especially poignant in the case of the Yellowstone bison herd, where only 3 introgressed bulls were the likely culprits of introducing cattle genetics to a previously non-introgressed herd. Due to Yellowstone bison being sent to a number of bison herds that were formerly believed to be non-introgressed, such as Wind Cave National Park, and Fort Niobrara National Wildlife Refuge, the cattle nuclear genes spread throughout United States and Canada undetected.

Nevertheless, bison are thriving today, as shown by the detailed study of the Yellowstone herd in Chapter II. Though the winters are harsh and resources are scarce,

and the landscape is far from ideal for bison, regardless of the high prevalence of brucellosis, and yearly culling, the Yellowstone bison herd is thriving both in numbers and in terms of genetic diversity.

Their high genetic diversity is most likely due to the fact that in the past 20 years, as the number of Yellowstone bison have greatly increased, so did their movement in the Park. Based on radio telemetry data, we know that about 50% of Yellowstone bison from the central herd migrate to the northern area of the park, either annually or permanently (Table II.2). Increased migration is associated with an increase in gene flow as well, leading to a homogenization effect that would leave the bison as a single panmixed population. Halbert (2003) hypothesized that there were three genetically distinct subpopulations, a trend that seemed to have changed over the course of the next decade when they reported only two (Halbert et al. 2012). The analysis of the entire mitochondrial genome in Chapter II showed significant differences between two clades of Yellowstone bison: one that we hypothesized goes back to the indigenous bison lineage, and the other made up of descendants of bison introduced in 1902 to Yellowstone National Park.

Though Halbert et al. (2012) found clear evidence of two genetically different genetic clusters in Yellowstone, those clusters were not entirely based on geography. Similar to the telemetry migration data, their data also showed the central herd as being genetically unique, while the northern herd was divided between 51.5% central-like animals (bison which are genetically indistinguishable from central herd bison) and 40.7% that formed a separate genetically distinct cluster. The study also found a number

of bison in each herd that were the result of hybridization between the two clusters, possessing loci that appear to be the results of recombination (<10% of bison). In order to see how the Yellowstone bison have changed since Halbert's study, we compared it with data from the 50 Yellowstone bison from Chapter III, all collected at the Gardiner winter cull location. These 50 bison are a skewed sample because they were all captured at the northern boundary of the park during the same time of year. When population structure was investigated using only those 50 animals with the Structure software package (Pritchard et al. 2000), the results were inconclusive due to the small sample size, while assigning them along with the 661 bison from Halbert et al. (2012) resulted in evidence for two populations in Yellowstone. However, including bison sampled 10-20 years ago introduces a temporal bias to the analysis that causes the results to be difficult to interpret. We found the animals to be in complete panmixia with regards to geographic location based on whole mitochondrial genome analysis. Halbert et al. (2012) argued that based on F_{ST} values, most of the gene flow between the northern and the central herds is perpetrated by males, while we found evidence that the females are also responsible for the gene flow, due to the even distribution of mtDNA haplotypes between the hypothesized herds.

Annual removal of animals during the late winter months is a court-mandated necessity to keep animals from leaving Yellowstone National Park. Most of the animals are taken either from the northern boundary (Gardiner area) or West Yellowstone. If there truly is unilateral gene flow from the central to the northern herd, then animals taken in the Gardiner basin will be a mixture of both populations, while those taken at

West Yellowstone will be exclusively central animals and their (potentially mixed-ancestry) offspring. Having a clear picture of the subdivision of Yellowstone bison would better inform population management practices by avoiding uneven culling in case there is evidence for multiple subpopulations in Yellowstone National Park.

We have identified two major mitochondrial clades after analyzing whole mitogenomes of 65 bison. We concluded that Clade I includes the descendants of the indigenous Yellowstone bison that were present in the Park for hundreds of years, and Clade II includes the descendants of the bison introduced in 1902 from the Pablo-Allard herd. Besides Yellowstone bison, Clade II also includes other herds from the Pablo-Allard lineage, while Clade I also includes historical Yellowstone area bison. What was confounding at first is that while all wood bison in our analysis were in Clade I, they did not form a monophyletic cluster, but instead appeared to be intertwined with plains bison haplotypes. At the time of the publication of the research, we reported this as simply a consequence of the hybridization of plains and wood bison in Canada after a large number of plains bison were sent north in the 1930s, but in light of new information it seems like it is more likely due to a separate event. That is, the indigenous Yellowstone bison were at least in part wood bison. 19th century naturalists called them mountain bison, and described the stark morphological and behavioral differences between them and plains bison. Even the Latin name they gave them was the same as that of the wood bison subspecies. Our nuclear genome analysis also suggests the close relationship of Yellowstone bison and wood bison, supported by principal component analyses, maximum likelihood trees, and ABBA-BABA tests. Thus, the most likely explanation to

the intermingled plains and wood bison haplotypes is also the crossing of mountain/ wood bison and plains bison in Yellowstone National Park.

Besides genetic diversity, population structure, and subspecies differentiation, we have also analyzed two point mutations in the mitochondrial genome that were allegedly detrimental to bison health (Pringle 2011). We have compared the frequency of the mutation over the last century and found no change. Thus, there is no strong evolutionary force acting on either of those mutations and the claim that they would cause an impairment of cellular respiration is unsubstantiated.

In the future we could compare a large number of recent samples from the entire area of the park (ideally 25% northern boundary winter cull, 25% West Yellowstone winter cull, 25% summer northern herd and 25% summer central herd animals). For a robust analysis, we propose to collect samples from 200 bison, extract their DNA using the appropriate methods, amplify and genotype them at all 46 loci Halbert's 2012 study used, as well as with the SNP markers being developed by Stroupe et al. (in preparation). Computationally assigning those animals to k number of hypothetical populations ($k = 1$ through $k = 5$) and interpreting the most likely number of populations and how they relate to geographic sampling location and life history would afford us the opportunity to unequivocally provide a definitive picture of the current status of Yellowstone bison subpopulations.

However, the collection of samples from bison is difficult. On many federal lands, such as Yellowstone National Park, there are no round-ups, and the animals are not run through chutes. Currently, all our Yellowstone samples are taken along the

northern and western boundaries of the Park, where animals are captured during the winter to be culled. Sample collection inside the Park is very expensive and permits and licenses are difficult to obtain if chemical immobilization of the bison is necessary. Therefore, we carried out a study to compare DNA genotyping efficiency and accuracy of paired blood and fecal samples from the same 50 Yellowstone bison. Using fecal samples would allow the collection of biological materials from bison in an easier and less intrusive manner, while also permitting the truly random sampling of bison.

Through bison fecal DNA analysis, we have identified that there is a vast difference in accuracy between STR markers. While the mechanism for this is unknown, the allelic dropout and false allele rates at some markers were so different between the blood and fecal samples that the discrepancy was statistically significant. Most previous studies using fecal DNA have not validated their markers using a handful of paired samples to aid marker choice, and we have found that caused a significant underestimation of genetic diversity and overestimation of inbreeding depression. As most fecal studies are conducted with wildlife species that are too rare and elusive to capture, many researchers have unwittingly inflated the direness of the genetic health of their study populations. Using the paired approach, we have validated 15 STR markers that provide accurate results with fecal samples in bison.

While the STR markers are a cheap and easy way of catching a glimpse of the underlying genetic data, a SNP-based analysis would provide a much more holistic genome-wide approach to the analysis of bison genetics. Based on over 96 million SNPs

between water buffalo, cattle, and 40 bison, we could take a more in-depth look at introgression, subspecies differentiation, and population structure of bison.

The first major finding of the study was that none of the herds tested were void of cattle introgression. According to our findings, the least introgressed herd was the Santa Catalina Island bison herd, a herd of known high levels of introgression. As a matter of fact, while neither of the two SCI bison sequenced tested positive for any of the nuclear introgression STR markers, one of them (SCI612) has cattle mitochondrial DNA. Based on a nuclear STR and mitochondrial variant analysis of 391 SCI bison, 46% had cattle mitochondrial DNA, while only 7.4% of them tested positive for at least one nuclear introgression marker.

We also uncovered evidence that there were multiple hybridization events between plains and wood bison, not only occurring in Elk Island National Park where the two subspecies cohabitate, but also in Yellowstone National Park where wood bison followed south along the Rocky Mountains all the way down to New Mexico (Christman 1971). This is evident based on the close genetic proximity of Yellowstone bison and the paraphyletic group of wood bison in the unrooted maximum likelihood tree (Figure IV.7), and based on the close genetic relationship of Yellowstone-lineage herds that could have never had a direct contact with Canadian wood bison, such as the Santa Catalina Island and the Wind Cave National Park herds.

Even though more data is needed in order to develop a SNP-based genotyping assay for bison-cattle introgression, a SNP-based parentage test is currently being developed (Stroupe et al. in preparation). Mitochondrial introgression is a lot more

straightforward and 11 species-defining and 9 clade-differentiating SNPs were developed to aid future introgression and lineage analyses (Table V.1, Table V.2). The clade-differentiating SNPs distinguish between bison that fall into Clade I and Clade II based on the mitochondrial study.

Identifier	Adapter regions and SNP
Bisonbosmt287	TTTGACTAAGTTATATTAAT[C/T]AGGGTTGGTAAATCTCGTGC
Bisonbosmt655	AACCCCGATAAACCTCACCA[G/A]TTCTTGCTAATACAGTCTAT
Bisonbosmt2071	AAGGACTTGTATGAATGGCC[A/G]CACGAGGGTTTTACTGTCTC
Bisonbosmt2141	CGTGAAGAGGCGGGAATGCA[T/C]AAATAAGACGAGAAGACCCT
Bisonbosmt2514	GATCTGAGTTCAGACCGGAG[C/T]AATCCAGGTCGGTTTCTATC
Bisonbosmt3391	AACGTAGAATATGCAGCAGG[G/A]CCATTTGCCCTCTTCTTCAT
Bisonbosmt5308	GATTTACAGTCTAATGCTTT[A/G]CTCAGCCATTTTACCCATGT
Bisonbosmt6857	TTTGAAGAACCCACCTATGT[C/T]AACCTAAAATAAGAAAGGAA
Bisonbosmt9353	TGACATTTCGTAGACGTAGT[T/C]TGACTTTTCCTCTATRTTTC
Bisonbosmt11191	GAACGAATCCACAGCCGAAC[T/C]ATAATTCTAGCTCGAGGCCT
Bisonbosmt14107	ACATGGAATCTAACCATGAC[C/T]AATGATATGAAAAACCATCG

Table V.1: Species determining mitochondrial variants with adapter regions. [bison/cattle] The numbers after the identifier are based on the position of the SNP in the mitochondrial sequences published in Forgacs et al. 2016.

Identifier	Adapter regions and SNP
Clademt5366	TTCTCAACCAACCATAAAGA[T/C]ATCGGTACCCTTTATCTACT
Clademt5852	AACATAAAGCCCCCCGCAAT[A/G]TCACAGTACCAAMCCCCTCT
Clademt8108	ATCAAAACAAATAATGAGTA[T/A]CCACAACCCCAAAGGACAAA
Clademt8504	CACTACAACAGCTCTAATTA[C/T]ATTCATCGTTCTAATCCTAC
Clademt10678	CTGATTTATATCCAAAACAT[A/G]GTAGGATCCCTAAATTTCCT
Clademt11108	GAAGTTACATAGGAGCAACC[G/A]CTCTCATGATTGCCCATGGCC
Clademt12107	ATTAACAAATTCTTCAAATA[C/T]CTACTCCTATTCCTCATCAC
Clademt12222	TTCTACTCATCGGATGATGA[T/C]ACGGACGAGCAGACGCAAAT
Clademt14445	TATCTGCTTATATATGCACG[T/C]AGGACGAGGCCTATATTACG

Table V.2: Clade differentiating mitochondrial variants with adapter regions. [Clade I/Clade II] The numbers after the identifier are based on the position of the SNP in the mitochondrial sequences published in Forgacs et al. 2016.

Humans have had a dichotomous effect on bison. While the human-mediated transfer of animals caused the spread of cattle genes across America, it was also likely a saving grace for bison today. Although gene flow created a homogenizing effect making all bison genetically similar to one another, it has also maximized the genetic diversity within each herd. Most recently, this was showcased by the Caprock Canyons State Park herd, which was on its way to extinction due to inbreeding depression, until a number of Yellowstone-lineage bison from the Castle Rock herd were introduced (Halbert et al. 2004). Since then, the signs of inbreeding depression have ceased and the Caprock Canyons herd is thriving.

Bison are survivors and there are no obstacles that have proven impossible for them to overcome. My hope is that the studies in this dissertation have paved the way to further inquiries into the bison genome. I hope it will also inform conservation genetic studies beyond bison, for the same concepts apply for a variety of vertebrate species. Lastly, I hope that their protection will not stop at purely ceremonial acts (such as officially naming bison the U.S. National Mammal), but also herald in legitimate conservation action so that the bottlenecks and the mistakes of the past will never be repeated.

REFERENCES

Achilli A, Olivieri A, Pellecchia M, Uboldi C, Colli L, Al-Zahery N, et al. Mitochondrial genome of extinct Aurochs survive in domestic cattle. Curr Biol. 2008 Feb 26; 18(4): R157-8.

Adler D, Murdoch D. 3D visualization using OpenGL. https://r-forge.r-project.org/projects/rgl/.

Algee-Hewitt BF, Edge MD, Kim J, Li JZ, Rosenberg NA. Individual identifiability predicts population identifiability in forensic microsatellite markers. Curr Biol. 2016; 26(7): 935-42.

Arias JA, Keehan M, Fisher P, Coppieters W, Spelman R. A high density linkage map of the bovine genome. BMC Genet. 2009; 10: 18.

Aronesty E. ea-utils: Command-line tools for processing biological sequencing data. Durham, NC: Expression Analysis. 2011.

Babicki S, Arndt D, Marcu A, Liang Y, Grant JR, Maciejewski A, et al. Heatmapper: web-enabled heat mapping for all. Nucleic Acids Res. 2016 Jul 8; 44(W1): W147-53.

Ball MC, Pither R, Manseau M, Clark J, Petersen SD, Kingston S, et al. Characterization of target nuclear DNA from faeces reduces technical issues associated with the assumptions of low-quality and quantity template. Cons Genet. 2007; 8(3): 577-86.

Barsness L. Heads, hides, & horns. Fort Worth (TX): Texas Christian University Press; 1985.

Baynes EH. Report of the secretary Ernest Harold Baynes. Annual Report of the American Bison Society. 1909; 43-64.

Bellemain E, Swenson JE, Tallmon D, Brunberg S, Taberlet P. Estimating population size of elusive animals with DNA from hunter-collected feces: four methods for brown bears. Cons Biol. 2005; 19(1): 150-61.

Bidwell D. Bison, boundaries, and brucellosis: risk perception and political ecology at Yellowstone. Soc Nat Resour. 2010; 23(1): 14-30.

Blanton DW, White PJ, Wallen RL, Auttelet KL, Stewart AJ, Bramblett AM. The population – attributes, behavior, distribution, resource use, and trends. In: White PJ, Wallen RL, Hallac DE. Yellowstone bison – conserving an American icon. Yellowstone National Park: The Yellowstone Association; 2015; pp. 1-17.

Bolfíková B, Hulva P. Microevolution of sympatry: landscape genetics of hedgehogs *Erinaceus europaeus* and *E. roumanicus* in Central Europe. Heredity. 2012 Mar; 108(3): 248-55.

Botero-Castro F, Tilak M, Justy F, Catzeflis F, Delsuc F, Douzery EJP. Next-generation sequencing and phylogenetic signal of complete mitochondrial genomes for resolving the evolutionary history of leaf-nosed bats (Phyllostomidae). Mol Phylogenet Evol. 2013 Dec; 69(3): 728-39.

Brazeal JL, Weist T, Sacks BN. Noninvasive genetic spatial capture-recapture for estimating deer population abundance. J Wildl Manage. 2017; 81(4): 629-40.

Broquet T, Petit E. Quantifying genotyping errors in noninvasive population genetics. Mol Ecol. 2004; 13(11): 3601-8.

Butler JM. Forensic DNA typing: biology, technology, and genetics of STR markers. 2nd ed. Burlington, MA: Elsevier Academic Press; 2005.

Cahalane VH. Restoration of wild bison. Transactions of the North American Wildlife Conference. 1944; 9: 135-143.

Calvignac S, Konecny L, Malard F, Douady CJ. Preventing the pollution of mitochondrial datasets with nuclear mitochondrial paralogs (numts). Mitochondrion. 2011 Mar; 11(2): 246-54.

Chiang C, Layer RM, Faust GG, Lindberg MR, Rose DB, Garrison EP, et al. SpeedSeq: ultra-fast personal genome analysis and interpretation. Nat Meth. 2015: doi:10.1038/nmeth.3505.

Christman GM. The Mountain Bison. The American West. 1971 May; 8(3): 44-7.

Clement M, Posada D, Crandall KA. TCS: a computer program to estimate gene genealogies. Mol Ecol. 2000 Oct; 9(10): 1657-9.

Coder GD. The national movement to preserve the American buffalo in the United States and Canada between 1880 and 1920. Ph.D. Dissertation, Columbus (OH): The Ohio State University; 1975.

Colpitts G. Game in the Garden: A human history of wildlife in western Canada to 1940. Vancouver, British Columbia, Canada: UBC Press; 2002.

Cooper A, Turney C, Hughen KA, Brook BW, McDonald HG, Bradshaw CJA. Abrupt warming events drove Late Pleistocene Holarctic megafaunal turnover. Science. 2015 Aug 7; 349(6248): 602-6.

COSEWIC. COSEWIC assessment and status report on the plains bison *Bison bison bison* and the wood bison *Bison bison athabascae* in Canada. Committee on the Status of Endangered Wildlife in Canada. 2013; 1-109.

Costa V, Rosenbom S, Monteiro R, O'Rourke SM, Beja-Pereira A. Improving DNA quality extracted from fecal samples – a method to improve DNA yield. Eur J Wildl Res. 2017; 63: 3.

Creel S, Spong G, Sands JL, Rotella J, Zeigle J, Joe L, et al. Population size estimation in Yellowstone wolves with error-prone noninvasive microsatellite genotypes. Mol Ecol. 2003; 12(7): 2003-9.

Curry CJ. A century of conservation genetics of the lion (*Panthera leo*). Ph.D. Dissertation, College Station (TX): Texas A&M University; 2019.

Curry CJ, Derr JN. Development of lion miniSTRs for use with modern and historical DNA samples. African J of Wildlife Research. 2019; 49(1): 64-74.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call formar and VCFtools. Bioinformatics. 2011 Aug 1; 27(15): 2156-8.

De Barba M, Miquel C, Lobreaux S, Quenette PY, Swenson JE, Taberlet P. High-throughput microsatellite genotyping in ecology: improved accuracy, efficiency, standardization and success with low-quantity and degraded DNA. Mol Ecol Res. 2017; 17(3): 492-507.

Department of the Interior. The American bison in the United States and Canada. Senate document, 57th Congress. 1902; 1(445): 1-43.

Department of the Interior. DOI bison report: Looking forward. Natural Resource Report NPS/NRSS/BRMD/NRR. 2014; 821.

DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011 May; 43(5): 491-8.

Derr JN, Hedrick PW, Halbert ND, Plough L, Dobson LK, King J, et al. Phenotypic effect of cattle mitochondrial DNA in American bison. Cons Biol. 2012 Aug 2; 26(6): 1130-6.

Dobson LK. Sequencing the genome of the North American bison. Ph.D. Dissertation: College Station (TX): Texas A&M University; 2015.

Douglas KC, Halbert ND, Kolenda C, Childer C, Hunter DL, Derr JN. Complete mitochondrial DNA sequence analysis of *Bison bison* and bison-cattle hybrids: function and phylogeny. Mitochondrion. 2011 Jan; 11(1): 166-75.

Elshibli S, Korpelainen H. Microsatellite markers reveal high genetic diversity in date palm (*Phoenix dactylifera* L.) gerplasm from Sudan. Genetica. 2008 Oct; 134(2): 251-60.

Emelianov I, Marec F, Mallet J. Genomic evidence for divergence with gene flow in host races of the larch budmoth. Proc Biol Sci. 2004 Jan 7; 271(1534): 97-105.

Ernest HB, Penedo MCT, May BP, Syvanen M, Boyce WM. Molecular tracking of mountain lions in the Yosemite Valley region in California: genetic analysis using microsatellites and faecal DNA. Mol Ecol. 2000; 9(4): 433-41.

Excoffier L, Lischer HE. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol. 2010 May; 10(3): 564-7.

Fernando P, Vidya TNC, Rajapakse C, Dangolla A, Melnick DJ. Reliable noninvasive genotyping: fantasy or reality? J Hered. 2003; 94(2): 115-23.

Flagstad O, Røed K, Stacy JE, Jakobsen KS. Reliable noninvasive genotyping based on excremental PCR of nuclear DNA purified with a magnetic bead protocol. Mol Ecol. 1999; 8(5): 879-83.

Foote AD, Vijay N, Ávila-Arcos MC, Baird RW, Durban JW, Fumagalli M, et al. Genome-culture coevolution promotes rapid divergence of killer whale ecotypes. Nature Communications. 2016 May 31; 7: 11693.

Forgacs D, Wallen RL, Dobson LK, Derr JN. Mitochondrial genome analysis reveals historical lineages in Yellowstone bison. PLoS ONE. 2016; 11(11): e016608.

Forgacs D, Wallen RL, Boedeker AL, Derr JN. Evaluating the use of fecal DNA as a non-invasive sampling technique by comparing matched blood and fecal samples from American bison (*Bison bison*). BMC Genet. 2019; 20(22). doi:10.1186/s12863-019-0722-3.

Furlan E, Stoklosa J, Weeks A, Griffiths J, Gust N, Ellis R, et al. A small population size and extremely low levels of genetic diversity in island populations of the platypus, *Ornithorhynchus anatinus*. Ecol Evol. 2012 Apr; 2(4): 844-57.

Gardipee FM. Development of fecal DNA sampling methods to assess genetic population structure of Greater Yellowstone bison. M.Sc. Thesis, Missoula (MT): University of Montana; 2007.

Garretson MS. The American bison. New York (NY): New York Zoological Society; 1938.

Geffen E, Luikart G, Waples R. Impacts of modern molecular genetic techniques on conservation biology. Publications, Agencies and Staff of the U.S. Department of Commerce. 2007; 461.

Geremia C, White PJ, Hoeting JA, Wallen RL, Watson FG, Blanton D, et al. Integrating population- and individual-level information in a movement model of Yellowstone bison. Ecol Appl. 2014 Mar; 24(2): 346-62.

Goodnight C. My experience with bison hybrids. J Hered. 1914; 5(5): 197-9.

Halbert ND. The utilization of genetic markers to resolve modern management issues in historic bison populations: implications for species conservation. Ph.D. Dissertation, College Station (TX): Texas A&M University; 2003.

Halbert ND, Raudsepp T, Chowdhary BP, Derr JN. Conservation genetics analysis of the Texas State bison herd. J Mammal. 2004 Oct 20; 85(5): 924-31.

Halbert ND, Derr JN. A comprehensive evaluation of cattle introgression into US federal bison herds. J Hered. 2007 Jan-Feb; 98(1): 1-12.

Halbert ND, Derr JN. Patterns of genetic variation in UD federal bison herds. Mol Ecol. 2008; 17(23): 4963-77.

Halbert ND, Gogan PJ, Hedrick PW, Wahl JM, Derr JN. Genetic population substructure in bison at Yellowstone National Park. J Hered. 2012 May-June; 103(3): 360-70.

Haldane JBS. Sex ratio and unisexual sterility in hybrid animals. J Genet. 1922; 12(2): 101-9.

Hitchings SP, Beebee TJ. Loss of genetic diversity and fitness in common toad (*Bufo bufo*) populations isolated by inimical habitat. J Evol Biol. 1998 May; 11(3): 269-83.

Hobbs NT, Geremia C, Treanor J, Wallen RL, White PJ, Hooten MB, et al. State-space modeling to support management of brucellosis in the Yellowstone bison population. Ecol Monogr. 2015 Nov; 85(4): 525-56.

Hofreiter M, Jaenicke V, Serre D, von Haeseler A, Pääbo S. DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. Nucleic Acids Res. 2001 Dec 1; 29(23): 4793-9.

Hornaday WT. The extermination of the American bison. Report of the National Museum. 1889; IV: 29-57.

Hornaday WT. Census of living American bison on January 1st, 1908. Annual Report of the American Bison Society. 1908; 70-8.

Isenberg AC. The destruction of the bison: An environmental history, 1750-1920. Cambridge (UK): Cambridge University Press; 2000.

Jarne P, Lagoda PJ. Microsatellites, from molecules to populations and back. Trends Ecol Evol. 1996; 11(10): 424-9.

Johnson NA. Hybrid incompatibility genes: remnants of a genomic battlefield? Trend in Genet. 2010; 26(7): 317-25.

Johnson PCD, Haydon DT. Maximum-likelihood estimation of allelic dropout and false allele error rates from microsatellite genotypes in the absence of reference data. Genetics. 2007; 175(2): 827-42.

Jones CJ. N.A.R.S., RG 78, Records of the National Park Service, Letters Relating to Yellowstone. 1905; Letter No. 176. National Archives Identifier: 7268838.

Jones CJ. Breeding cattelo. Annu Report of the Am Breeders' Assoc. 1907; 3: 161-5.

King RA, Read DS, Traugott M, Symondson WO. Molecular analysis of predation: a review of best practice for DNA-based approaches. Mol Ecol. 2008; 17(4): 947-63.

Kohn MH, York EC, Kamradt DA, Haught G, Sauvajot RM, Wayne RK. Estimating population size by genotyping faeces. Proc Biol Sci. 1999; 266(1420): 657-63.

Kurose N, Masuda R, Tatara M. Fecal DNA analysis for identifying species and sex of sympatric carnivores: a noninvasive method for conservation on the Tsushima Islands, Japan. J Hered. 2005; 96(6): 688-97.

Lacy, RC. Loss of genetic diversity from managed populations: interacting effects of drift, mutation, immigration, selection, and population subdivision. Cons Biol. 1987 Aug; 1(2): 143-58.

Lathuilliere M, Menard N, Gautier-Hion A, Crouau-Roy B. Testing the reliability of noninvasive genetic sampling by comparing analyses of blood and fecal samples in Barbary macaques (*Macaca sylvanus*). Am J Primatol. 2001; 55(3): 151-8.

Lehmkuhl JF. Determining size and dispersion of minimum viable populations for land management planning and species conservation. Environ Manag. 1984; 8: 167.

Li FY, Cuddon PA, Song J, Wood SL, Patterson JS, Shelton GD, et al. Canine spongiform leukoencephalomyelopathy is associated with a missense mutation in cytochrome b. Neurobiol Dis. 2006 Jan; 21(1): 35-42.

Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009; 25: 1754-60.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. Bioinformatics. 2009 Aug 15; 25(16): 2078-9.

Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint arXiv: 2013 May 26; 1303.3997.

Ma H, Gutierrez NM, Morey R, Van Dyken C, Kang E, Hayama T, et al. Incompatibility between nuclear and mitochondrial genomes contributes to an interspecies reproductive barrier. Cell Metabol. 2016; 24(2): 283-94.

Marsolier-Kergoat MC, Palacio P, Berthonaud V, Maksud F, Stafford T, Bégouën R, et al. Hunting the extinct steppe bison (*Bison priscus*) mitochondrial genome in the Trois-Frères paleolithic painted cave. PLoS ONE. 2015; 10(6): e0128267.

Mattila HR, Rios D, Walker-Sperling VE, Roeselers G, Newton ILG. Characterization of the active microbiotas associated with honey bees reveals healthier and broader communities when colonies are genetically diverse. PLoS ONE. 2012 Mar 12; 7(3): e32962.

Maudet C, Luikart G, Dubray D, Von Hardenberg A, Taberlet P. Low genotyping error rates in wild ungulate faeces sampled in winter. Mol Ecol Res. 2004; 4(4): 772–5.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010 Sep; 20(9): 1297-303.

Meagher MM. The bison of Yellowstone National Park. Washington DC: Scientific Monograph Series, National Park Service; 1973.

Mendez-Harclerode FM, Strauss RE, Fulhorst CF, Milazzo ML, Ruthven DC, Bradley RD. Molecular evidence for high levels of intrapopulation genetic diversity in woodrats (*Neotoma micropus*). J Mammal. 2007 Apr 1; 88(2): 360-70.

Miller, MA, Pfeiffer W, Schwartz T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: Proceedings of the Gateway Computing Environments Workshop (GCE). 2010 Nov 14; pp 1-8.

Montooth KL, Meiklejohn CD, Abt DN, Rand DM. Mitochondrial-nuclear epistasis affects fitness within species but does not contribute to fixed incompatibilities between species of *Drosophila*. Evolution. 2010; 64(12): 3364-79.

Morgante M, Olivieri AM. PCR-amplified microsatellites as markers in plant genetics. Plant J. 1993; 3(1): 175-82.

Morin DJ, Waits LP, McNitt DC, Kelly MJ. Efficient single-survey estimation of carnivore density using fecal DNA and spatial capture-recapture: a bobcat case study. Popul Ecol. 2018; 60: 197.

Morin PA, Chambers KE, Boesch C, Vigilant L. Quantitative polymerase chain reaction analysis of DNA from noninvasive samples for accurate microsatellite genotyping of wild chimpanzees (*Pan troglodytes verus*). Mol Ecol. 2001; 10(7): 1835-44.

Morton NE. Sequential tests for the detection of linkage. Am J Hum Genet. 1955; 7(3): 277-318.

Murphy MA, Waits LP, Kendall KC, Wasser SK, Higbee JA, Bogden R. An evaluation of long-term preservation methods for brown bear (*Ursus arctos*) faecal DNA samples. Cons Genet. 2002; 3(4): 435-40.

Ogilvie SC. The Park Buffalo. Calgary, Alberta, Canada: National and Provincial Parks Association of Canada; 1979.

Panasci M, Ballard WB, Breck S, Rodriguez D, Densmore III LD, Wester DB, et al. Evaluation of fecal DNA preservation techniques and effects of sample age and diet on genotyping success. J Wildl Manage. 2011; 75(7): 1616-24.

Parsons KM. Reliable microsatellite genotyping of dolphin DNA from faeces. Mol Ecol Res. 2001; 1(4): 341-4.

Peakall RO, Smouse PE. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Molec Ecol Notes. 2006 Mar; 6(1): 288-95.

Peakall RO, and Smouse PE. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinformatics. 2012; (28): 2537-9.

Peden DG, Kraay GJ. Comparison of blood characteristics in plains bison, wood bison, and their hybrids. Can J Zool. 1979 Sep; 57(9): 1778-84.

Pickrell JK, Pritchard JK. Inference of population splits and mixtures from genome-wide allele frequency data. PLoS Genet. 2012; 8(11): e1002967.

Plumb GE, White PJ, Coughenour MB, Wallen RL. Carrying capacity, migration, and dispersal in Yellowstone bison. Biol Cons. 2009 Nov; 142(11): 2377-87.

Pringle TH. Widespread mitochondrial disease in North American bison. Nature Precedings. 2011 Feb; hdl:10101/npre.2011.5645.1.

Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000 Jun; 155(2): 945-59.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a toolset for whole-genome association and population-based linkage analysis. Am J of Human Genet. 2007: 81.

Queller DC, Strassmann JE, Hughes CR. Microsatellites and kinship. Trends Ecol Evol. 1993; 8(8): 285-8.

Raj A, Stephens M, Pritchard JK. fastSTRUCTURE: Variational Inference of Population Structure in Large SNP Data Sets. Genetics. 2014 Jun; 197: 573-89.

Rambaut A. Computer program distributed by the author, website: http://tree.bio.ed.ac.uk/software/figtree/ [accessed July 29, 2019].

Schnabel RD. Developing DNA-based technologies in North American bison: parentage testing, linkage mapping and QTL scans. Ph.D. Dissertation. College Station (TX): Texas A&M University; 2001.

Silva TL, Godinho R, Castro D, Abaigar T, Brito JC, Alves PC. Genetic identification of endangered North African ungulates using noninvasive sampling. Mol Ecol Res. 2015; 15(3): 652-61.

Soraggi S, Wiuf C, Albrechtsen A. Powerful inference with the D-statistic on low-coverage whole-genome data. G3: Genes, Genomes, Genetics. 2018 Feb 1; 8(2): 551-66.

Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014 May; 30(9): 1312–3.

Steklenev EP, Yasinetskaya NI. Results of crossing of the bison (*Bison bison bison* L.) with the domestic cow (*Bos (bos) taurus typicus*) and characteristics of the chromosome complexes of the hybrid progeny. Cyto and Genet. 1982; 16: 28-33.

Taberlet P, Camarra JJ, Griffin S, Uhres E, Hanotte O, Waits LP, et al. Noninvasive genetic tracking of the endangered Pyrenean brown bear population. Mol Ecol. 1997; 6(9): 869-76.

Taberlet P, Waits L, Luikart G. Noninvasive genetic sampling: look before you leap. Trends Ecol Evol. 1999; 14(8): 323-7.

Tamura K, Nei M, Kumar S. Prospects for inferring very large phylogenies by using the neighbor-joining method. Proc Natl Acad Sci USA. 2004 Jul 27; 101: 11030-35.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular evolutionary genetics analysis version 6.0. Mol Biol Evol. 2013 Dec; 30(12): 2725-9.

Taylor MS. Buffalo hunt: International trade and the virtual extinction of the North American bison. The Amer Econ Rev. 2011; 101(7): 3162-95.

Templeton AR, Crandall KA, Sing CF. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. Genetics. 1992 Oct; 132(2): 619-33.

Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, et al. From FastQ data to high-confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinformatics. 2013 Oct; 43: 11.10.1-33.

Wallen RL, White PJ, Geremia C. Historical Perspecive. In: White PJ, Wallen RL, Hallac DE, editors. Yellowstone bison—Conserving an American icon in modern society. Yellowstone National Park: The Yellowstone Association; 2015. p. 45-66.

Ward TJ, Bielawski JP, Davis SK, Templeton JW, Derr JN. Identification of domestic cattle hybrids in wild cattle and bison species: a general approach using mtDNA markers and the parametric bootstrap. Anim Cons. 1999 Feb; 2(1): 51-7.

Weng Z, Saatchi M, Schnabel RD, Taylor JF, Garrick DJ. Recombination locations and rates in beef cattle assessed from parent-offspring pairs. Genet Sel Evol. 2014; doi: 10.1186/1297-9686-46-34.

White PJ, Wallen RL. Yellowstone bison – should we preserve artificial population substructure or rely on ecological processes? J Hered. 2012 Sep-Oct; 103(5): 751-3.

Will save the buffalo. The Topeka Daily Capital. 1901 Dec 24; 5.

Wilson GA, Strobeck C. Genetic variation within and relatedness among wood and plains bison populations. Genome. 1999 Jun; 42(3): 483-96.

Yao YG, Kong QP, Salas A, Bandelt HJ. Pseudomitochondrial genome haunts disease studies. J Med Genet. 2008 Dec; 45(12): 769-72.

Yuan M, Wei D, Zhang K, Gao Y, Liu Y, Wang B, et al. Genetic diversity and population substructure of *Panonychus citri* (Acari: Tetranychidae) in China based on mitochondrial COI gene sequences. J Econ Entomol. 2010 Dec; 103(6): 2204-13.

Zhang X, Xu Y, Liu H, Jiang G. Risks involved in fecal DNA-based genotyping of microsatellite loci in the Amur tiger *Panthera tigris altaica*: a pilot study. J For Res. 2018; 29: 525.

Zheng X, Levine D, Shen J, Gogarten S, Laurie C, Weir B. A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics. 2012; 28(24): 3326-28.

Zhu Y, Liu H, Yang H, Li Y, Zhang H. Factors affecting genotyping success in giant panda fecal samples. PeerJ. 2017; 5: e3358.

APPENDIX A

LIST OF POLYMORPHIC SITES IN THE BISON MITOCHONDRIAL GENOME

Base position	Gene	Consensus codon/base	Amino acid	Alternate codon/base	Alternate amino acid	Haplotypes with alternate codon/base
18	tRNA-Phe	C		T		S9
54	tRNA-Phe	C		T		S9
93	12S rRNA	T		C		wHap15
130	12S rRNA	C		T		S9
520	12S rRNA	A		T		EINP2
533	12S rRNA	T		G		EINP2
534	12S rRNA	A		T		EINP2
851	12S rRNA	C		T		EINP4
940	12S rRNA	C		T		bHap3/bHap12
1157	16S rRNA	C		T		S9
1161	16S rRNA	A		G		wHap15
1606	16S rRNA	C		T		S9
1609	16S rRNA	A		C		S9
1624	16S rRNA	G		A		S9
1715	16S rRNA	G		A		EINP1, EINP3
2206	16S rRNA	T		C		S6
2226	16S rRNA	C		T		wHap15
2688	tRNA-	G		A		bHap11
2939	ND1	GCC	ala	ACC	thr	EINP3
3019	ND1	CCT	pro	CCG	pro	YNPHap1, YNPHap3, CCSP1, S6, S9, Templeton, bHap3/bHap12, bHap4, bHap6, bHap13/bHap16, wHap15
3019	ND1	CCT	pro	CCA	pro	bHap7
3543	ND1	TCA	ser	TTA	leu	EINP2, EINP3, wHap14
3708	tRNA-Ile	A		C		YNPHap8
4042	ND2	AAC	asn	ACC	thr	Templeton

Non-synonymous mutations are in red, synonymous mutations are in green and non-coding tRNA or rRNA regions are in black. The two IMOP mutations are highlighted yellow. Reprinted with permission from Forgacs et al. 2016.

Base position	Gene	Consensus codon/base	Amino acid	Alternate codon/base	Alternate amino acid	Haplotypes with alternate codon/base
4048	ND2	AAC	asn	ACC	thr	Templeton
4053	ND2	CGA	arg	GGA	gly	Templeton
4994	tRNA-Trp	C		T		YNPHap8
5012	tRNA-Trp	C		A		YNPHap8
5032	tRNA-Ala	C		T		YNPHap8, EINP4
5099	tRNA-	C		T		EINP4
5242	tRNA-	C		T		YNPHap8
5366	COX1	GAT	asp	GAC	asp	YNPHap2, YNPHap5, YNPHap6, bHap2, bHap8, bHap10, bHap11, bHap17
5852	COX1	ATA	ile	ATG	met	YNPHap2, YNPHap5, YNPHap6, bHap2, bHap8, bHap10, bHap11, bHap17
5865	COX1	ACC	thr	CCC	pro	Templeton
5909	COX1	TTA	leu	TTC	phe	Templeton
6692	COX1	ATG	met	ATA	ile	EINP4
6701	COX1	TTC	phe	TTT	phe	EINP4
6704	COX1	ATT	ile	ATC	ile	EINP4
6722	COX1	ATG	met	ATA	ile	EINP4
6728	COX1	ATA	ile	ATG	met	YNPHap8
6737	COX1	ATC	ile	ATT	ile	YNPHap8, EINP4
6752	COX1	TTT	phe	TTC	phe	EINP4
6922	tRNA-Ser	T		A		YNPHap8
6935	tRNA-Ser	C		nothing		bHap3/bHap12
6936	tRNA-Ser	T		nothing		bHap3/bHap12
7008	tRNA-	C		T		YNPHap8
7040	COX2	TTC	phe	TTT	phe	YNPHap8
7049	COX2	GCT	ala	GCA	ala	YNPHap8, EINP3, EINP4
7109	COX2	TTC	phe	TTT	phe	EINP3
7139	COX2	ATC	ile	ATT	ile	YNPHap8
7161	COX2	CTG	leu	TTA	leu	YNPHap8
7214	COX2	ATT	ile	ATC	ile	YNPHap1, YNPHap3, CCSP1, S9, Templeton, bHap4, bHap6, bHap13/bHap16
7355	COX2	GAG	glu	GAA	glu	YNPHap8
7376	COX2	TAC	tyr	TAT	tyr	YNPHap8
7637	COX2	CCC	pro	CCT	pro	EINP4
7661	COX2	CTG	leu	CTA	leu	EINP4
7685	COX2	GCG	ala	GCA	ala	EINP3, EINP4

Continued.

Base position	Gene	Consensus codon/base	Amino acid	Alternate codon/base	Alternate amino acid	Haplotypes with alternate codon/base
7750	tRNA-Lys	A		T		EINP4
7768	tRNA-Lys	C		T		EINP4
7876	ATP8	AAT	asn	AAC	asn	YNPHap1, YNPHap3, CCSP1, S6, S9, Templeton, bHap3/bHap12, bHap4, bHap6, bHap7,
7880	ATP8	GAA	glu	AAA	lys	bHap3/bHap12
8108	ATP6	ATC	ile	AAC	asn	YNPHap2, YNPHap5, YNPHap6, bHap2, bHap8, bHap10, bHap11, bHap17
8430	ATP6	GGA	gly	GGG	gly	bHap3/bHap12
8458	ATP6	GCT	ala	ACT	thr	CCSP1, bHap13/bHap16, wHap15
8504	ATP6	ACA	thr	ATA	ile	YNPHap2, YNPHap5, YNPHap6, bHap2, bHap8, bHap10, bHap11, bHap17
8624	COX3	ACT	thr	ACC	thr	YNPHap8
8935	COX3	ACT	thr	ATT	ile	S9
9001	COX3	CTG	leu	CAG	gln	YNPHap5, YNPHap7
9002	COX3	CTG	leu	CTA	leu	YNPHap1, YNPHap3, S9, Templeton
9015	COX3	GTC	val	CTC	leu	YNPHap5
9015	COX3	GTC	val	ATC	ile	YNPHap7
9114	COX3	TTA	leu	CTA	leu	bHap4
9176	COX3	TCC	ser	TCA	ser	EINP2
9224	COX3	GGC	gly	GGG	gly	bHap17
9252	COX3	TTA	leu	CTA	leu	EINP4
9266	COX3	TTT	phe	TTC	phe	every haplotype in Douglas et al. 2011
9290	COX3	TTC	phe	TTT	phe	EINP4
9434	tRNA-Gly	C		T		EINP3
9448	tRNA-Gly	G		A		EINP3
9582	ND3	GGA	gly	GGG	gly	S6
9627	ND3	TTC	phe	TTT	phe	S9
9648	ND3	TTC	phe	TTT	phe	S9
9723	ND3	ATG	met	ATA	ile	S6
9795	ND3	CTG	leu	CTA	leu	EINP4
9827	tRNA-	T		C		EINP4
9842	tRNA-	C		T		EINP4
9856	tRNA-	G		A		EINP4
10431	ND4	AAC	asn	ATC	ile	bHap4
10453	ND4	TTC	phe	TTT	phe	YNPHap8
10470	ND4	TTA	leu	TCA	ser	bHap6
10525	ND4	TAT	tyr	TAC	tyr	bHap3/bHap12

Continued.

Base position	Gene	Consensus codon/base	Amino acid	Alternate codon/base	Alternate amino acid	Haplotypes with alternate codon/base
10678	ND4	ATA	ile	ATG	met	YNPHap2, YNPHap5, YNPHap6, bHap2, bHap8, bHap10, bHap11, bHap17
10921	ND4	AAC	asn	AAT	asn	YNPHap2, YNPHap5, YNPHap6, bHap2, bHap17
11108	ND4	GCT	ala	ACT	thr	YNPHap2, YNPHap5, YNPHap6, bHap2, bHap8, bHap10, bHap11, bHap17
11359	ND4	GGA	gly	GGG	gly	S6
11406	ND4	ATA	ile	ACA	thr	EINP4, bHap5
11483	ND4	CTG	leu	TTG	leu	YNPHap2, YNPHap5, YNPHap6, bHap2, bHap17
11492	ND4	TTA	leu	ATA	ile	EINP4, bHap5, bHap9
11507	ND4	TTA	leu	CTA	leu	bHap6
11718	tRNA-	G		A		YNPHap8
12107	ND5	TAC	tyr	TAT	tyr	YNPHap2, YNPHap5, YNPHap6, bHap2, bHap8, bHap10, bHap11, bHap17
12159	ND5	TTC	phe	CTC	leu	bHap8
12222	ND5	TAC	tyr	CAC	his	YNPHap2, YNPHap5, YNPHap6, bHap2, bHap8, bHap10, bHap11, bHap17
12840	ND5	GCC	ala	ACC	thr	bHap7
13292	ND5	TAC	tyr	TAT	tyr	EINP2, EINP3, wHap14
13573	ND6	ATA	ile	ATG	met	EINP4
13594	ND6	TTT	phe	TTA	leu	EINP4
13624	ND6	TGA	trp	TGG	trp	EINP4
13682	ND6	TCC	ser	TTC	phe	S6
13699	ND6	ATT	ile	ATC	ile	YNPHap3, YNPHap6
13707	ND6	TGG	trp	CGG	arg	YNPHap3, YNPHap6
13810	ND6	TTA	leu	TTT	phe	Templeton
14026	ND6	GTA	val	GTG	val	EINP4, bHap5, bHap9
14160	CYTB	AAC	asn	AGC	ser	bHap7
14445	CYTB	GTA	val	GCA	ala	YNPHap2, YNPHap5, YNPHap6, bHap8, bHap10, bHap11, bHap17
14519	CYTB	GTA	val	ATA	ile	EINP2, EINP3, wHap14
14527	CYTB	GCC	ala	GCT	ala	CCSP1, bHap13/bHap16
14914	CYTB	GAT	asp	GAC	asp	EINP4, bHap5, bHap9
14935	CYTB	CCA	pro	CCG	pro	bHap6
15100	CYTB	ATT	ile	ATC	ile	EINP3
15195	CYTB	ATC	ile	ACC	thr	EINP3
15344	tRNA-Thr	A		G		YNPHap3, S9, Templeton

Continued.

APPENDIX B

GENBANK ACCESSION NUMBERS FOR ALL PREVIOUSLY UNPUBLISHED MITOCHONDRIAL HAPLOTYPES

Haplotype	GenBank accession number
YNPHap1	KX451352
YNPHap2	KX451353
YNPHap3	KX451354
YNPHap4	KX451355
YNPHap5	KX451356
YNPHap6	KX451357
YNPHap7	KX451358
YNPHap8	KX451359
CCSP	KX451360
EINP1	KX451361
EINP2	KX451362
EINP3	KX451363
EINP4	KX451364
S6	KX451365
S9	KX451366
Templeton	KX451367

Reprinted with permission from Forgacs et al. 2016.

APPENDIX C

STATISTICAL ANALYSIS OF FECAL DNA GENOTYPING DATA

Panel	Marker	Total # of samples	Match	Unreadable	No match	Percent matching (excluding unreadable data)	Percent matching (including unreadable data)
Mix 80	BM1862	50	20	29	1	95.24%	40.00%
Mix 80	BMS1001	50	44	4	2	95.65%	88.00%
Mix 80	BMS1074	50	26	24	0	100.00%	52.00%
Mix 80	BMS1675	50	32	18	0	100.00%	64.00%
Mix 80	BMS1716	50	32	18	0	100.00%	64.00%
Mix 80	HUJ246	50	18	32	0	100.00%	36.00%
Mix 82	BM4311	50	45	2	3	93.75%	90.00%
Mix 82	BM6017	50	41	9	0	100.00%	82.00%
Mix 82	BM711	50	36	14	0	100.00%	72.00%
Mix 82	INRA189	50	47	0	3	94.00%	94.00%
Mix 82	TGLA122	50	26	22	2	92.86%	52.00%
Mix 85	*BL1036	50	50	0	0	100.00%	100.00%
Mix 85	*BM4107	50	50	0	0	100.00%	100.00%
Mix 85	*BM47	50	49	0	1	98.00%	98.00%
Mix 85	*ILSTS102	50	48	0	2	96.00%	96.00%
Mix 85	BMS1315	50	48	0	2	96.00%	96.00%

Paired fecal and blood sample matching percentages. Reprinted with permission from Forgacs et al. 2019.

Panel	Marker	Total # of samples	Match	Unreadable	No match	Percent matching (excluding unreadable data)	Percent matching (including unreadable data)
Mix 85	BMS1857	50	45	0	5	90.00%	90.00%
PRTG1	*BMS510	50	48	0	2	96.00%	96.00%
PRTG1	BM17132	50	25	12	13	65.79%	50.00%
PRTG1	BMS410	50	39	5	6	86.67%	78.00%
PRTG1	BMS527	50	17	31	2	89.47%	34.00%
PRTG1	RM372	50	46	0	4	92.00%	92.00%
PRTG2	*BM1225	50	50	0	0	100.00%	100.00%
PRTG2	*BM1706	50	50	0	0	100.00%	100.00%
PRTG2	*BM1905	50	49	0	1	98.00%	98.00%
PRTG2	*BM2113	50	48	0	2	96.00%	96.00%
PRTG2	BM4440	50	44	0	6	88.00%	88.00%
PRTG2	BM720	50	39	3	8	82.98%	78.00%
PRTY A	*BM4513	50	49	1	0	100.00%	98.00%
PRTY A	*BM7145	50	50	0	0	100.00%	100.00%
PRTY A	*SPS113	50	49	0	1	98.00%	98.00%
PRTY A	BMS2270	50	40	8	2	95.24%	80.00%
PRTY BC	*AGLA293	50	49	1	0	100.00%	98.00%
PRTY BC	*BM4307	50	50	0	0	100.00%	100.00%
PRTY BC	*CSSM42	50	50	0	0	100.00%	100.00%

Continued.

Panel	Marker	Allelic dropout	Neither allele matches	Blood heterozygosity	Fecal heterozygosity	Difference	Number of alleles	Allelic richness
Mix 80	BM1862	0	0	0.762	0.762	0.000	10	0.500
Mix 80	BMS1001	2	0	0.739	0.696	0.043	5	0.800
Mix 80	BMS1074	0	0	0.692	0.692	0.000	5	0.8
Mix 80	BMS1675	0	0	0.188	0.188	0.000	4	0.75
Mix 80	BMS1716	0	0	0.031	0.031	0.000	6	0.5
Mix 80	HUJ246	0	0	0.444	0.444	0.000	7	0.429
Mix 82	BM4311	2	0	0.792	0.750	0.042	7	0.857
Mix 82	BM6017	0	0	0.415	0.415	0.000	6	0.833
Mix 82	BM711	0	0	0.389	0.389	0.000	7	0.571
Mix 82	INRA189	2	1	n/a	n/a	n/a	3	0.667
Mix 82	TGLA122	2	0	0.679	0.607	0.071	7	0.714
Mix 85	*BL1036	0	0	0.740	0.740	0.000	8	0.500
Mix 85	*BM4107	0	0	0.680	0.680	0.000	9	0.556
Mix 85	*BM47	1	0	0.100	0.080	0.020	6	0.333
Mix 85	*ILSTS102	1	0	0.680	0.660	0.020	8	0.375
Mix 85	BMS1315	1	0	0.560	0.480	0.080	7	0.571
Mix 85	BMS1857	1	4	0.600	0.580	0.020	14	0.357
PRTG1	*BMS510	1	1	0.500	0.289	0.211	6	0.667
PRTG1	BM17132	11	2	0.689	0.644	0.044	10	0.500
PRTG1	BMS410	2	0	0.720	0.700	0.020	11	0.364
PRTG1	BMS527	0	0	0.632	0.632	0.000	12	0.500

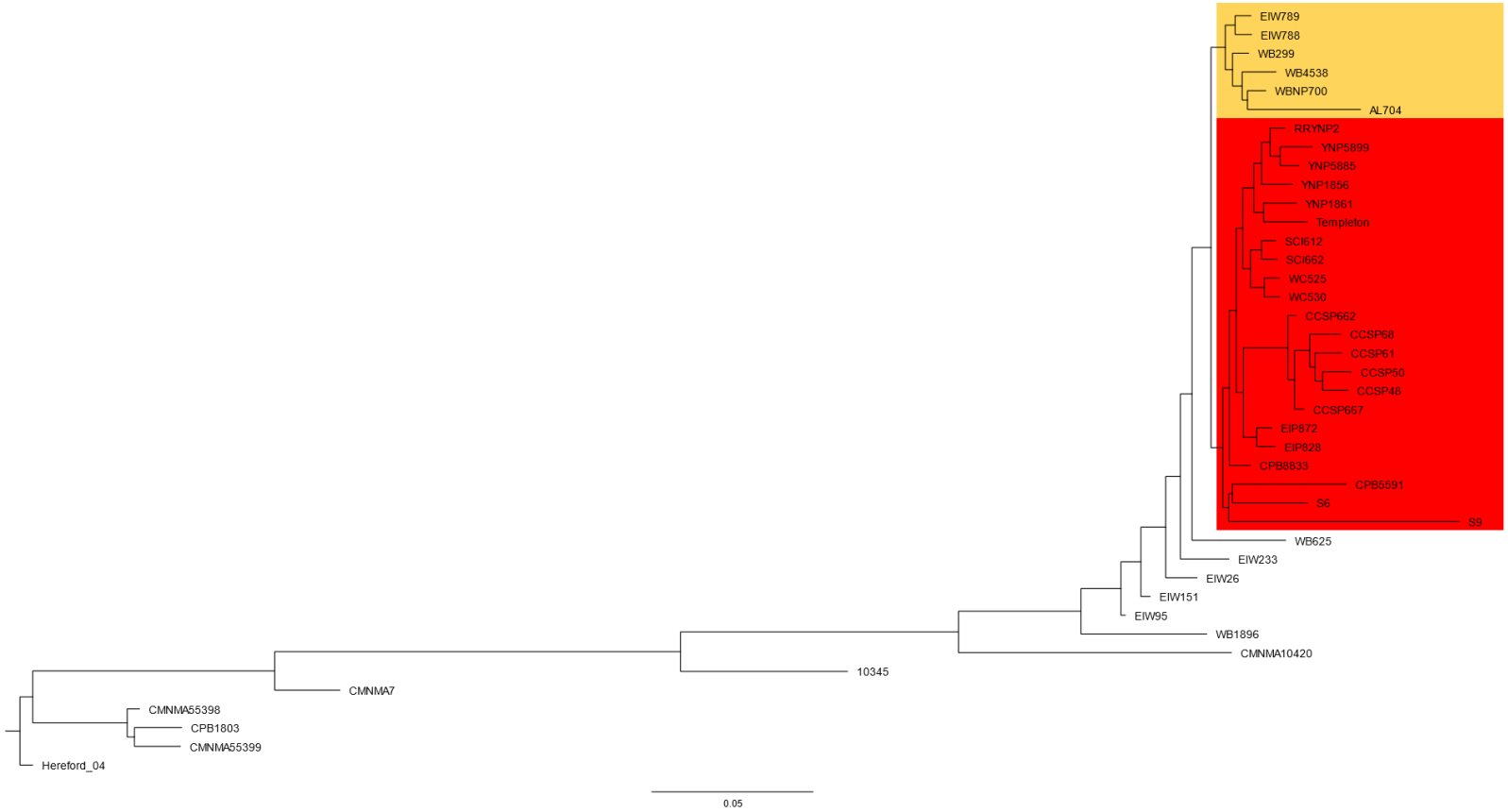
Discrepancies between paired blood and fecal samples in different population genetic metrics. The markers with a star were included in the final STR panel for bison fecal DNA. Reprinted with permission from Forgacs et al. 2019.

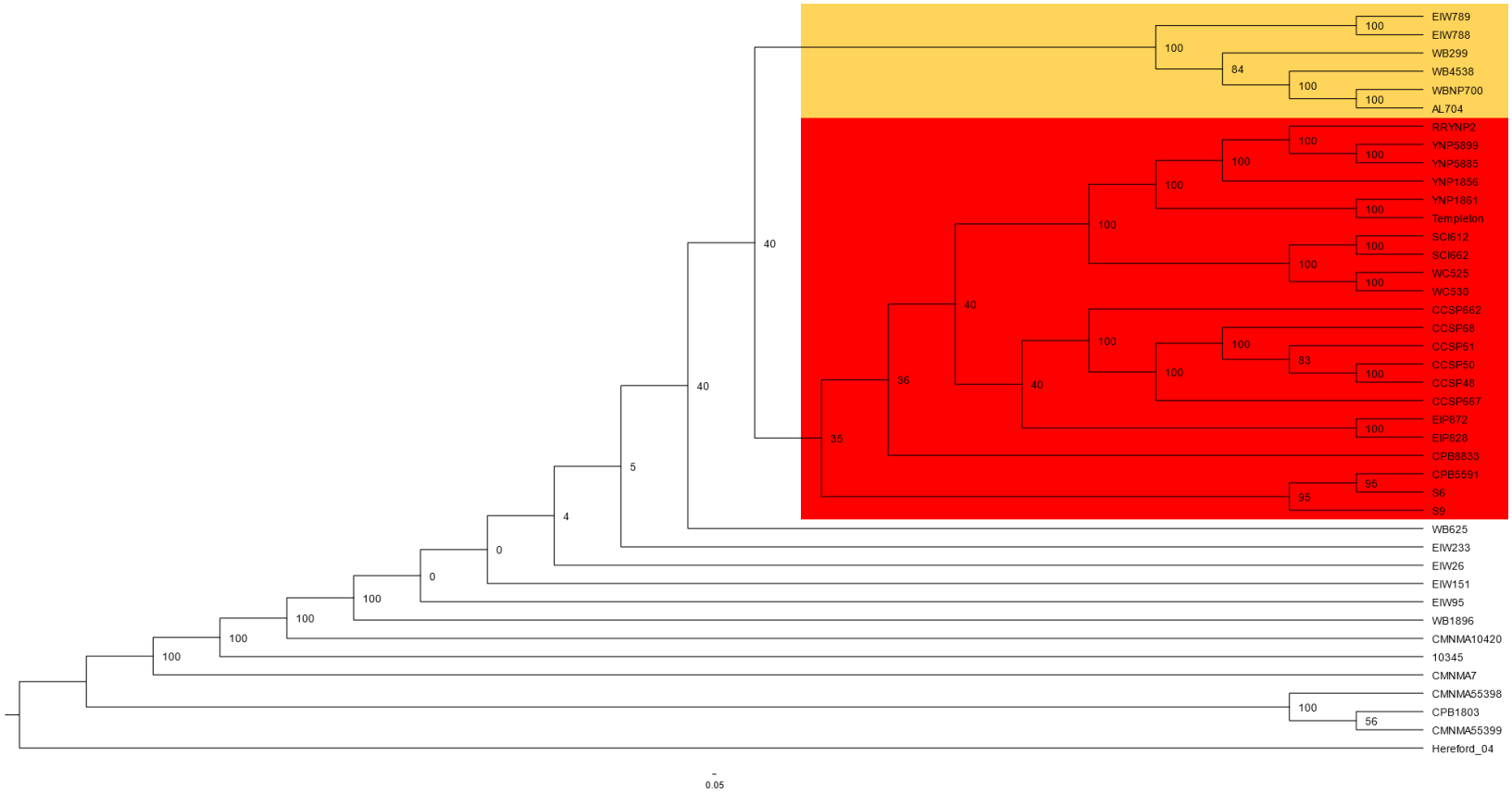
Panel	Marker	Allelic dropout	Neither allele matches	Blood heterozygosity	Fecal heterozygosity	Difference	Number of alleles	Allelic richness
PRTG1	RM372	1	0	0.640	0.600	0.040	12	0.333
PRTG2	*BM1225	0	0	0.720	0.720	0.000	13	0.385
PRTG2	*BM1706	0	0	0.480	0.480	0.000	14	0.286
PRTG2	*BM1905	0	0	0.380	0.400	-0.020	10	0.300
PRTG2	*BM2113	1	0	0.640	0.640	0.000	13	0.308
PRTG2	BM4440	4	0	0.580	0.520	0.060	11	0.455
PRTG2	BM720	5	0	0.638	0.532	0.106	13	0.462
PRTY A	*BM4513	0	0	0.163	0.163	0.000	2	1.000
PRTY A	*BM7145	0	0	0.300	0.300	0.000	3	0.667
PRTY A	*SPS113	1	0	0.643	0.595	0.048	3	0.667
PRTY A	BMS2270	2	0	0.460	0.440	0.020	3	0.667
PRTY BC	*AGLA293	0	0	0.000	0.000	0.000	2	0.500
PRTY BC	*BM4307	0	0	0.000	0.000	0.000	3	0.333
PRTY BC	*CSSM42	0	0	0.580	0.580	0.000	3	1.000

Continued.

APPENDIX D

ROOTED MAXIMUM LIKELIHOOD TREES BASED ON GENOMIC DATA





Maximum likelihood phylogram and cladogram rooted with a representative Hereford cattle. The plains bison clade is highlighted red, the wood bison clade is highlighted green. Bootstrap values are displayed for each node.

APPENDIX E

THE NATURE CONSERVANCY SATELLITE HERD REPORT

The aim of this project was to genetically test bison from Wind Cave National Park and six satellite herds managed by The Nature Conservancy, and provide information on the population health, determine genetic subdivision, and advise the potential transfer of animals.

We tested 712 total bison at 34 nuclear microsatellite markers established by Halbert, 2003. The bison came from Broken Kettle Grasslands (BKG; 153), Dunn Ranch (DR; 135), Kankakee Sands (KS; 47), Nachusa Grasslands (NG; 96), Smoky Valley Ranch (SVR; 190), Tallgrass Prairie National Preserve (TPNP; 20), and Wind Cave National Park (WC; 71). Our analysis has revealed that the Smoky Valley Ranch herd has a number of non-Wind Cave lineage bison, some of which show evidence of cattle introgression based on the nuclear microsatellite analysis, and as such in its current state cannot be considered as a satellite herd. DNA was extracted and amplified from each of the 712 tail hair samples, and subsequently genotyped on an ABI 3130 Genetic Analyzer. The genotyping calls were made using GeneMapper and STRand, and the analysis was done using Arlequin (Excoffier and Lischer 2010), GenAlEx (Peakall and Smouse 2006, 2012), Heatmapper (Babicki et al. 2016) and STRUCTURE (Pritchard et al. 2000).

The genotyping failure rate was determined to be 1.47% overall (BKG: 0.4%, DR: 3.1%, KS: 0.6%, NG: 1.6%, SVR: 1.8%, TPNP: 0.7%, WC: 0.5%), meaning that

even after repeated testing, we could not verify 355 of the 24,208 data points and those were omitted from all further analysis. All the alleles were also cross-checked with alleles that were previously documented in Wind Cave National Park, which revealed 8 SVR bison (367, 372, 373, 688, 698, 706, 714, and 715) had at least one allele previously only found in cattle and bison-cattle hybrids. The comparison also revealed unique alleles never before seen in WC in 60 of the SVR herd (all from the production side), as well as one allele in two very likely related KS bison that is unique, most likely due to random mutation. None of the bison from any of the herds tested positive for cattle mitochondrial DNA.

The total heterozygosity (a measure of genetic diversity) ranged from 0.563 in BKG to 0.604 in TPNP which is slightly lower than what Halbert and Derr, 2008 found in WC (0.65) but they used a larger number of bison and markers in their analysis. The small range of heterozygosity values indicates that all satellite herds have similar diversity indices. The total number of alleles (another measure of genetic diversity) ranged from 98 in TPNP to 116 in SVR. This is likely due to TPNP having the fewest number of bison tested, and SVR having the second most, with a number of unique and introgressed alleles in the production herd. As many diversity indices for SVR are skewed, hereafter we have also calculated the metrics using only the 47 SVR bison that are intended to be kept after Fall 2019 when all non-WC lineage bison will be removed from the herd. Each satellite herd encompassed 37-44% of the total number of alleles found in plains bison (allelic richness). WC and SVR have the highest number of alleles and allelic richness of all of the non-introgressed satellite herds (103, 39.2%).

Herd Name	Heterozygosity	Total Number of Alleles	Allelic Richness
Broken Kettle Grasslands	0.563	102	0.388
Dunn Ranch	0.579	100	0.380
Kankakee Sands	0.597	99	0.376
Nachusa Grasslands	0.572	101	0.384
Smoky Valley Ranch	0.581	116	0.441
Smoky Valley Ranch (keepers)	0.589	103	0.392
Tallgrass Prairie National Preserve	0.604	98	0.373
Wind Cave National Park	0.580	103	0.392
Yellowstone National Park	--------	133	0.506

Population genetics metrics for each satellite herd and select federal herds for comparison.

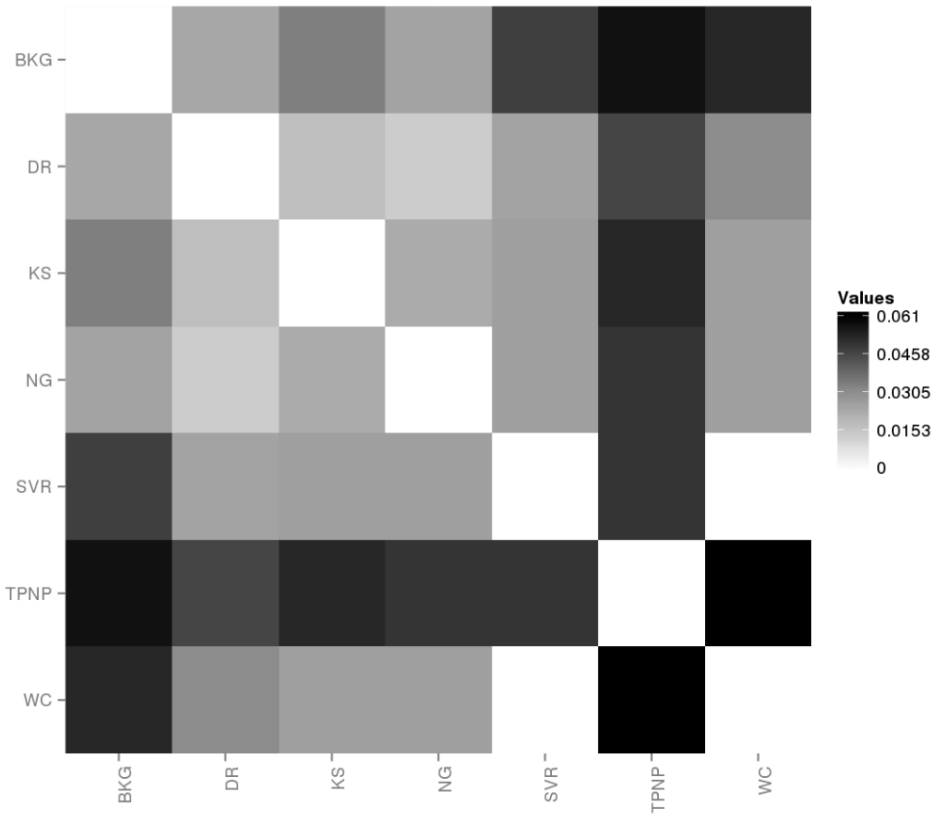
Most loci for each herd were in Hardy-Weinberg equilibrium, meaning that the allele and genotypic frequencies were close to being unchanged over generations. Only a very few of the 34 loci we tested violated H-W E (BKG: 5, DR: 3, KS: 1, NG: 3, SVR: 4, TPNP: 1, WC: 3). The violation is most likely not due to limited herd size, as the number of bison tested from each herd is inversely proportional to the number of loci in violation.

Analyses of population structure revealed that all satellite herd animals cluster closer to the Wind Cave herd than to any other foundation herd. None of the 712 animals form a separate cluster, which is surprising given that some of the SVR bison are hybrids with no known Wind Cave ancestry. Further analyses revealed that the 712 bison genetically belong to the same indistinguishable metapopulation, meaning the satellite herds have so far done a great job at preserving Wind Cave genetic diversity without creating genetic differentiation. The FST value based on an AMOVA test is negligible

(0.014) showing further evidence that there is no genetic population subdivision between the satellite herds. Pairwise structure analysis revealed that the two herds that are genetically the closest are SVR and WC, and the genetically most distant are TPNP and BKG.

	BKG	DR	KS	NG	SVR	TPNP	WC
BKG	0.00000						
DR	0.01166	0.00000					
KS	0.02429	0.01276	0.00000				
NG	0.01317	0.00966	0.01643	0.00000			
SVR	0.01453	0.00730	0.00690	0.01018	0.00000		
TPNP	0.03031	0.02165	0.02475	0.02257	0.01236	0.00000	
WC	0.01976	0.01185	0.00999	0.00976	0.00004	0.02306	0.00000

Table representing the pairwise differences between the satellite herds.



Heatmap representing the pairwise differences between the satellite herds.

This information should be taken into consideration when moving animals between satellite herds, as moving between herds that are more genetically distant (larger number in the pairwise table) will have a lot stronger positive effect. The distance between the entire SVR herd and all other herds would have been inflated by the non-WC animals, so only the Wind Cave lineage SVR bison were included in this analysis.

Due to uneven sex ratios in every satellite herd, the non-Fisherian model for calculating effective population size was used. Below is the comparison of the actual and the effective population sizes for each herd.

Herd Name	**Actual Breeding Age Population Size (F+M)**	**Effective Population Size**
Broken Kettle Grasslands	122 (93+29)	88.43
Dunn Ranch	102 (76+26)	77.49
Kankakee Sands	42 (27+15)	38.57
Nachusa Grasslands	76 (50+26)	68.42
Smoky Valley Ranch (w/ production animals)	125 (117+8)	29.95
Smoky Valley Ranch (keepers)	45 (36+9)	28.80
Tallgrass Prairie National Preserve	43 (26+17)	41.12
TOTAL for all satellite herds (only bison w/ WC origin)	430 (308+122)	349.54
Wind Cave National Park (2000) (Halbert and Derr, 2008)	368 (228+140)	346.96
TOTAL WC lineage bison	798 (536+262)	703.92

Comparison of the actual and the effective population size for each satellite herd.

Generally, a minimum effective population size of 50 is necessary in order to reduce inbreeding in the short-term (Lehmkuhl 1984). Naturally, moving animals in and out of those herds may inflate the effective population size and reduce the risk of inbreeding depression. When adding animals to herds, the ones with the lowest effective population size should be prioritized, the top 3 being KS, TPNP, and SVR once the production bison have been removed. To demonstrate the importance of breeding males for the genetic integrity of small herds, KS could achieve an effective population size of greater than 50 by adding 9 males or 50 females; TPNP could achieve it by adding 8 males or 25 females; and SVR could achieve it by adding 11 males but could not achieve it by adding virtually any number of females without also increasing the number of males. However, this method of calculating effective population does not take genetics or breeding success into consideration, only the proportion of males and females, where the closer the sex ratio is to 1:1, the closer the effective population size gets to the census size.

The carrying capacity of each satellite herd also need to be taken into consideration. Based on the number of samples submitted from each of the satellite herds, none of them have reached their maximum carrying capacity yet. It is especially crucial to boost the number of individuals in those herds where there is room to spare and the effective population size is currently fewer than 50 bison. The goal of the satellite herds is to reach an effective population of 1,000 but given the current status of the herds even if all the carrying capacities were reached, the census size would only reach 896.

Herd Name	Carrying Capacity
Broken Kettle Grasslands	250
Dunn Ranch	201
Kankakee Sands	75
Nachusa Grasslands	120
Smoky Valley Ranch	150
Tallgrass Prairie National Preserve	100

Carrying capacity of each of the satellite herds.

The genetic analysis and parentage testing of 40 additional SVR calves born in 2018 revealed that only 29 can be confidently matched to a mother and father which we have samples from. All with the exception of one of these calves was sired by a single bull (982000423576569). Nine out of the remaining 11 calves had a significant match to a mother but not to a father, suggesting there was at least another breeding bull in the herd besides the 9 Wind Cave lineage bulls we were provided samples from.

Of the 29 calves where parentage could be identified with high confidence, only 2 of the calves had Wind Cave lineage mothers, all the others were born to production herd mothers. As a near even ratio of production to Wind Cave lineage females were present, it shows a clear preference of the bull for non-Wind Cave lineage bison. This could be potentially due to production females being tamer and less resistant to mating, or due to negative assortative mating, where mates less like your genetics are deemed most desirable.

Summary of recommendations:

1. The first priority should be boosting the number of animals (especially the number of males) in herds with an effective population size of fewer than 50 bison. This could be best achieved by adding Wind Cave bison, or transferring bison from another (larger) satellite herd.

2. Transfers between herds with the greatest pairwise distance between them should be practiced often, as they will provide the most significant genetic boost.

3. In order to achieve a larger effective population size, the uneven sex ratios should be addressed by aiming to more closely reflect a 1:1 sex ratio in each satellite herd. However, effective population size is just a metric that is meant to keep inbreeding low and genetic diversity high in populations. To that effect, even more important than the equal sex ratio is that a variety of males are mating with the females, which is not always the case as shown with the 2018 Smoky Valley Ranch calves. Annual parentage testing of calves will help identify harem bulls which would need to have their access to females restricted to achieve greater genetic diversity. A more accurate measure of effective population size based on genetic data and the change of allele frequencies over time would require a longitudinal study of multiple non-overlapping generations of bison which would take too long and immediate action is recommended.

5. The Smoky Valley Ranch herd in its current state is not a proper satellite herd. Still a lot of the bison are from non-Wind Cave lineages, and the some of those animals even used to have detectable cattle alleles or alleles novel to Wind Cave. The removal of production animals and calves who were born to a Wind Cave bull and a production

bison cow is necessary in the shortest possible timeframe based on continued rigorous parentage testing of calves, as well as competent record keeping for the herd. Once only confirmed Wind Cave-lineage bison remain, a large influx of Wind Cave bison is warranted due to the very low actual and effective population sizes for this herd.