ABSTRACT

Comparing the Genetic Diversity of Late Pleistocene *Bison* with Modern *Bison bison* Using Ancient DNA Techniques and the Mitochondrial DNA Control Region

Kory C. Douglas

Mentors: Robert P. Adams, Ph.D., and Lori E. Baker, Ph.D.

The transition between the Pleistocene and Holocene Epochs brought about a mass extinction of many large mammals. The genetic consequences of such widespread extinctions have not been well studied. Using ancient DNA and phylogenetic techniques, the genetic diversity and phylogenetic relatedness of extinct Pleistocene *Bison* ranging from Siberia to mid-latitude North America (10,000 ybp to 50,000 ybp) were compared to extant *Bison bison*.

The mitochondrial DNA control region was sequenced from 10 *Bison priscus* skulls obtained from the Kolyma Region of Siberia, Russia. Control region sequences from other Pleistocene *Bison* species and *Bison bison* were obtained from Genbank. There is a measurable loss of genetic diversity in *Bison bison* compared to Pleistocene *Bison*. Furthermore, the Pleistocene *Bison* population was strongest in North America from a time period of 30,000 ybp to 10,000 ybp, and the genetic diversity present in this population is not represented in the *Bison bison* population.

Comparing the Genetic Diversity of Late Pleistocene *Bison* with Modern *Bison bison* Using Ancient DNA Techniques and the Mitochondrial DNA Control Region

by

Kory C. Douglas

A Thesis

Approved by the Department of Biology

Ann E. Rushing, Ph.D., Acting Chairperson

Submitted to the Graduate Faculty of
Baylor University in Partial Fulfillment of the
Requirements for the Degree
of
Master of Science

Approved by the Thesis Committee

Robert P. Adams, Ph.D., Co-Chairperson

Lori E. Baker, Ph.D., Co-Chairperson

Christopher M. Kearney, Ph.D.

Kieran P. McNulty, Ph.D.

Accepted by the Graduate School
August 2006

J. Larry Lyon, Ph.D., Dean

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to thank the Department of Biology at Baylor University for making this project possible. Through the Jack G. and Norma Jean Folmar Research Award and Bob Gardner Memorial Grant Award, I was afforded the opportunity to collect samples in Siberia, Russia, an experience I will never forget. I would like to thank Dr. Pavel A. Nikolskiy, Fedor K. Shidlovskiy, and the Ice Age Museum in Moscow, Russia for their generous donation of samples and also affording me the opportunity to collect samples in Siberia, Russia. I would also like to thank my mentors, Dr. Lori E. Baker and Dr. Robert P. Adams, for their guidance and mentorship both professionally and personally.

I would like to thank my other committee members, Dr. Christopher M. Kearney and Dr. Kieran P. McNulty, for their time, patience, and expertise. I especially want to thank Dr. Erich J. Baker for his help in constructing phylogenetic trees, and Dr. Diane Wycuff for her sound advice and friendship.

Lastly, I would like to thank my family for their continual emotional and financial support throughout my graduate career. Every positive quality I possess I owe to my parents who taught me the importance of sacrifice and hard work, and they continue to be my inspiration.

CHAPTER ONE

Introduction

Ancient DNA (aDNA) is a powerful tool, providing otherwise unobtainable genetic information from extinct species. Ancient DNA has been sequenced from animal and plant remains dated up to one hundred thousand years before present (ybp), and bacterial remains dated up to one million years before present (Goldenberg and others 1990; Soltis and others 1992; Lindahl 1997; Willerslev and others 2003). Although some of these studies have been criticized, the potential to analyze sequence diversity over hundreds of thousands of years can provide insights into the evolutionary origins of contemporary species. These origins have previously been inferred through the analysis of morphological data, genetic markers, and populational distribution patterns, all of which can only provide indirect evidence of the evolutionary processes leading to contemporary species (Willerslev and Cooper 2005). Analyzing genetic data, we are able to augment morphological and geological comparisons by examining the underlying sequence diversity of populations. Furthermore, with the availability of sequence data, phylogenetic relationships can be examined and population genetic indices can be compared providing further insight into evolutionary processes and population dynamics. When combined, morphological, geological and genetic data can present a more accurate and comprehensive portrait of the evolutionary processes leading to contemporary species, populational dynamics over time, and extinction events than if examined separately (Shapiro and others 2004; Willerslev and Cooper 2005; Poinar and others 2006).

Ancient DNA

With the advent of polymerase chain reaction (PCR), it became possible to amplify a single surviving DNA template, enabling the amplification of small amounts of aDNA (Mullis 1987; Pääbo 1989; Thomas and others 1989). Early studies were highly criticized, however, for a lack of suitable controls, facilities, and techniques to control for contamination, a major concern in aDNA research (Wayne and others 1999). For example, if a single double-stranded DNA molecule is amplified by PCR for 60 cycles, approximately 10^{18} amplified molecules can be produced in a volume of less than 50 micro-liters (μL) or about 5e16 molecules per micro-liter. This value increases if more than one original template molecule is present. Therefore, the act of opening a PCR tube may release microscopic aerosol droplets containing millions of copies of the template DNA per 0.005 μL (Cooper and Poinar 2000). If a sample becomes contaminated with amplified or extraneous DNA, the more robust contaminant DNA may 'out-compete' the target DNA for the primers leading to amplification of the contaminant DNA rather than the target DNA. Unfortunately, such contamination is difficult to differentiate from positive results until the DNA sequence data are obtained. Furthermore, in order to validate the obtained sequence, a sample must be independently replicated and the sequences compared (Cooper and Poinar 2000).

Because of the high likelihood of contamination, rigorous criteria have been developed to authenticate aDNA results. Physically isolated work areas, irradiation of tools and surfaces with ultra violet light, and negative amplification controls are necessary to decrease and evaluate contamination (Cooper and Poinar 2000). PCR robustness is approximately inversely correlative to product size (longer sequences yield

lower concentrations of amplicons). When analyzing aDNA, PCR products over 500 base pairs are rare because DNA degradation produces DNA molecules of approximately 200 base pairs or less (Pääbo 1989). Results should also be reproducible from the same, and different, DNA extracts of a single sample. Finally, the results should make phylogenetic sense (Cooper and Poinar 2000). If one is attempting to amplify ancient *Bison* DNA and the resulting DNA sequences are not similar to other *Bison* sequences, it is probable that the sample has been contaminated or a sequencing error has occurred.

The degradation and damage of DNA after cell death presents problems as well. Permanently cold conditions are most suitable to preserve DNA for long periods of time making the permafrost of Siberia and northern North America an excellent environment for preservation. However, even in the best conditions for DNA preservation, the enzymatic reactions of cells and the degradative actions of microorganisms can cause DNA damage and contamination (Pääbo and others 2004). Depurination and deamination of bases, oxidation, and the affects of background radiation are further concerns when studying aDNA. To address concerns raised by the occurrence of these effects, the same sample can be extracted and sequenced multiple times (Hofreiter and others 2001). In addition, the use of a high fidelity DNA polymerase with relatively lower error rates can further minimize sequencing errors (Hansen and others 2001).

Mitochondrial DNA

Due to DNA damage with time, preservation of amplifiable nuclear DNA is rare. However, extra-nuclear, multi-copy sources of DNA in the cell such as mitochondrial DNA (mtDNA) have proven to be more amenable to analysis (Hofreiter and others 2002). The mitochondrion is a maternally inherited organelle which is believed to have

resulted from an endosymbiotic relationship between nucleated cells and free living bacteria capable of utilizing oxygen (Margulis and Sagan 1986). Mitochondria carry out oxidative phosphorylation, generating much of the energy used in eukaryotic cells. Each cell has multiple mitochondria (approximately 500), and within each mitochondrion are multiple copies of the mitochondrial genome leading to estimates of 1,000 to 10,000 copies of the mitochondrial genome per cell (Bogenhagen and Clayton 1974). In addition, mitochondria and mitochondrial genomes have the capability to replicate, divide, and fuse independently from the somatic nuclear division cycle (Bereiter-Hahn and Voth 1994).

In contrast to the linear nature of the nuclear genome, the mitochondrial genome is circular, promoting stability over time and increasing its resistance to degradation. The mitochondrial genome consists of several coding regions, which code for proteins necessary for oxidative phosphorylation, and a non-coding region referred to as the control region (also referred to as the D-loop or hyper-variable region; Ingman and Gyllensten 2001). Mitochondrial DNA coding and non-coding regions have been targeted in aDNA studies of Woolly Mammoths (*Mammuthus primigenius*), Pleistocene Brown Bears (*Ursus arctos*), Pleistocene Cave Bears, (*Ursus spelaeus*), Adélie Penguins (*Pygoscelis adeliae*), as well as other species (Yang and others 1996; Barnes and others 2002; Hofreiter and others 2002; Lambert and others 2002).

Furthermore, DNA sequences of the control region of mtDNA have been extensively utilized in modern and ancient DNA studies of genetic diversity and phylogenetic relationships. The control region is a non-coding segment of the mitochondrial genome which serves as the origin of replication for the mtDNA and has

been identified as approximately 900-1100 base pairs in length in many eukaryotes (Mitomap 2005; Lai and others 2006). Studies of Pleistocene brown bears (*Ursus arctos)* (Leonard and others 2000; Orlando and others 2002), woolly mammoths (*Mammuthus primigenius*) (Joger and Gorrido 2001), and Pleistocene bison (Shapiro and others 2004) have reported the mtDNA control region to be a viable source of aDNA amplification and analysis.

Because it is non-coding, the control region is less conserved mutationally than regions which code for genes and has a much higher mutation rate with respect to the rest of the mitochondrial genome. The mutation rate of the entire mitochondrial genome for mammals is commonly accepted to be approximately 2% per million years (Brown and others 1979), whereas the mutation rate for the control region for mammals has been estimated at approximately 30% per million years (Bradley and others 1996; Shapiro and others 2004). Even within the control region, there are two regions which are more likely to show mutations. These regions are referred to as hyper-variable regions one and two (HVRI and HVRII) (Greenberg and others 1983; Gilbert and others 2005). Because of its uni-parental inheritance pattern and its high mutation rate, the control region mtDNA is a powerful tool with which to examine evolutionary relationships (Brown and others 1979; Wayne and others 1999).

Pleistocene Epoch

With the upper limit of animal DNA preservation being approximately 100,000 ybp, we are able to examine the genetic diversity of animals during the events of the Late Pleistocene. The Pleistocene Epoch [1.8 million ybp to 11,000 ybp] consisted of climatic changes between glacial and inter-glacial periods (Taylor and others 1993). The glacial

and inter-glacial periods are referred to in two ways: by common names and by marine isotopic stage (MIS) numbers. The four maximal glacial advances in North America were the Nebraskan (MIS 10), Kansan (MIS 8), Illinoian (MIS 6), and Wisconsinan (MIS 2-4). The respective inter-glacial periods of North America following the Nebraskan glaciation were the Aftonian (MIS 9), Yarmouth (MIS 7), and the Sangamon (MIS 5). The glacial and inter-glacial periods of North America correspond to comparable events in Asia and Europe (Kurtén and Anderson 1980) with alternative common names. See Table 1 for the glacial and inter-glacial cycles.

Table 1. Glacial Patterns of the Pleistocene Epoch: (1.8 Million ybp – Present) (Stoffer 2003)

Glacial/ Inter-Glacial	Approximate Time Span (in years before present)
Wisconsinan Glacial	100,000 – 11,000
Sangamon Inter-Glacial	250,000 – 100,000
Illinoian Glacial	400,000 – 250,000
Yarmouth Inter-Glacial	750,000 - 400,000
Kansan Glacial	900,000 - 750,000
Aftonian Inter-Glacial	1.65 million – 900,000
Nebraskan Glacial	1.8 million – 1.65 million

The last inter-glacial period of the Late Pleistocene, the Sangamon (MIS 5), preceded the Wisconsinan glaciation (MIS 2-4) which contained the Last Glacial Maximum (LGM) (Kurtén and Anderson 1980). During the Wisconsinan glaciation (at its maximum advancement circa 30,000 ybp), the Cordilleran Ice Sheet and Laurentide Ice Sheet were formed in North America. The Cordilleran Ice Sheet extended from the Aleutian Islands, across southern Alaska, down the entire Pacific coastline of Canada,

and east covering the Rocky Mountains. The Laurentide Ice Sheet extended from the Atlantic coast of Canada west to the Rocky Mountains and stretched as far south as present day Illinois, U.S.A. as depicted in Figure 1 (Nilsson 1983). The two ice sheets would have shared a boundary on the eastern side of the Rocky Mountains; however, this boundary has been a controversial topic.

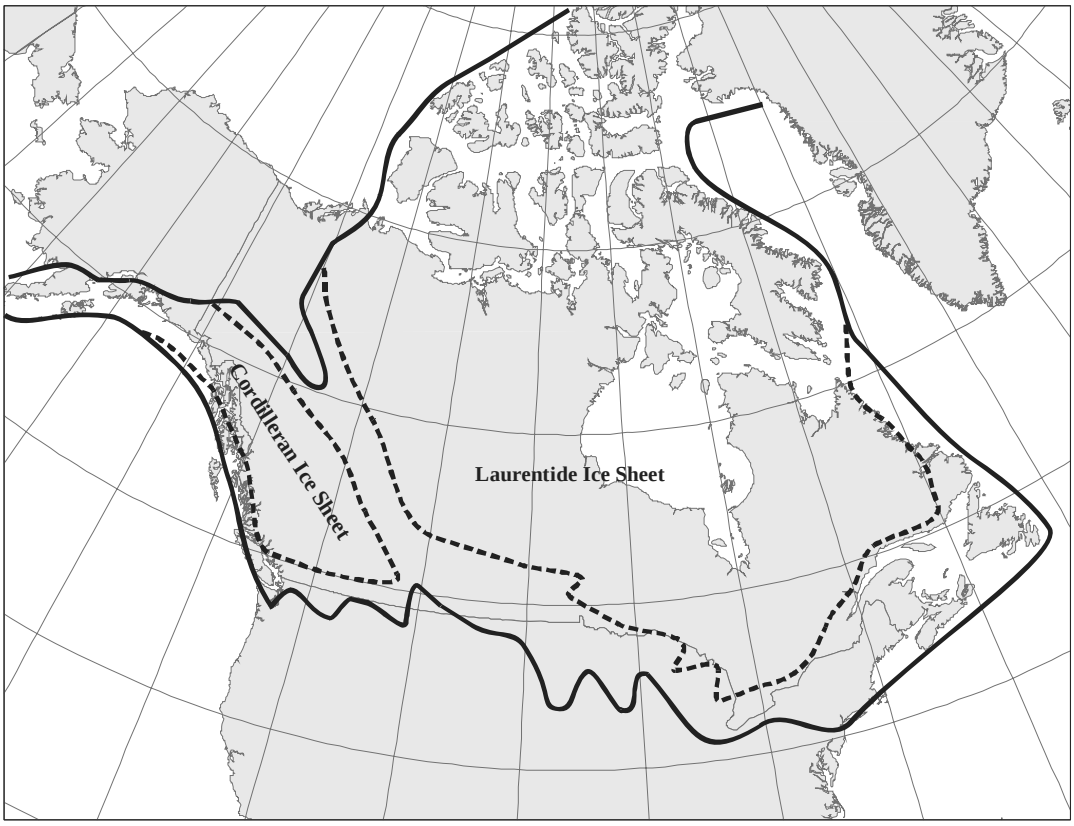

Figure 1. Cordilleran and Laurentide Ice Sheets and Ice Free Corridor During the Wisconsinan Glaciation (Solid Line indicates Glacial Position at Glacial Maximum circa 30,000 ybp. Dashed line indicates Glacial Positions circa 18,000 ybp) (adapted from Flint 1971).

During the glacial periods, the ice sheets were not stagnant but advanced and retreated iteratively (Stoffer 2003). Therefore, at times during the Wisconsinan Glaciation, the Laurentide and Cordilleran Ice Sheets did not merge completely, and an

ice-free corridor (IFC) connecting present day Alaska with mid-latitude North America existed (See Figure 1; Hopkins 1967; Flint 1971; Nilsson 1983; Burns 1996).

The mass glaciations trapped much of the world's water supply in expansive glaciers, lowering sea levels, resulting in uncovered landmasses previously submerged. Beringia, one of these landmasses, connected present-day Siberia with Alaska at the end of the Sangamon inter-glacial and throughout the Wisconsinan glaciation (Creager and McManus 1967), providing a new habitat and migration routes for both fauna and flora between Eurasia and North America. Furthermore, the IFC allowed for migration between Beringia and central North America. Carbon^{14} dates of recovered remains support the proposition that megafauna were present in Eurasia, Beringia, and North America throughout the Wisconsinan glaciation (Kurtén and Anderson 1980; Jorge and Gorrido 2001; Shapiro and others 2004). With an environment composed of stepped plains covered with grasses and grass-like plants, Beringia sustained a wide variety of fauna (Hopkins and others 1982; Guthrie 1990). Large mammals such as *Bison*, *Mammuthus* (Mammoth), and *Equus* (Horses) dominated the grazing lands of northeastern Asia, Beringia, and mid-latitude North America (Kurtén and Anderson 1980; Guthrie 1990). Well adapted for the colder, drier environment of the Late Pleistocene, these animals survived mainly on open tundra vegetation dominated by herbs and grasses (Edwards and others 2000).

Mass Extinctions

The mass extinction of megafauna at the Pleistocene-Holocene transition (circa 10,000 ybp) in North America was a unique event in Earth's history. Forty-three large mammal genera (weighing over 5 kg) became extinct and those which persisted, such as

Bison, markedly changed their geographical ranges (Barnosky 1989). By analyzing fossil remains, geological stratigraphy, and known history, hypotheses have been developed attempting in an effort to explain these mass extinctions. Two dominant hypotheses appear in the literature.

The first attributes the mass extinction event to over-hunting by man (McDonald 1981; Sutcliff 1985; Kurtén 1988). Mitochondrial control region and Y-chromosome analyses of North American Indians indicate that modern humans dispersed into North America approximately 30,000–34,000 ybp and migrated south into mid-latitude North America over the next 20,000 years (Underhill 1996; Starikovskaya 1998). Based on radiocarbon dated Clovis tools, it is generally thought that humans (*Homo sapiens sapiens*) were present in central North America at 11,200 ybp (Meltzer 1995). With the advancement of tools for hunting, such as Clovis and Folsom points, and the correlation between the arrival of modern humans in North America and the mass extinctions, it is argued that humans over-hunted the megafauna leading to their extinction (Kelly and Todd 1988; Kurtén 1988).

The second proposed explanation cites the dramatic climate change at the Pleistocene-Holocene transition as the cause of the mass extinctions (Guthrie 1984a). During maximum glacial coverage the westerly jet stream was split around the Laurentide Ice Sheet, and storm tracks were deflected further south than their current patterns (Kutzbach 1987). This effect would have reduced differences in seasonality forming a more harmonious climate than today resulting in expanded vegetational diversity. As the glaciers receded and weather patterns began to change, differences in seasonality increased and a reduction in vegetational diversity ensued (Barnosky 1989).

Fauna such as mammoth, mastodons, and bison, which required a wide variety of vegetation, became extinct or were extirpated across wide regions (Guthrie 1984b). Furthermore, studies of the pollen records suggest an influx of trees replaced the tundra and steppe plains in Beringia and Alaska approximately 10,500 ybp (Anderson and Brubaker 1994; Pisaric and others 2001). This elimination of suitable habitat led to a massive decline in population numbers, thus a loss of diversity, and ultimately extinction.

Bison

McDonald (1981), using morphological data, outlined the bison lineage as follows: the *Bison* and *Bos* genera split from *Leptobos* during the Pliocene with Bison speciating during the Pleistocene as shown in Figure 2.

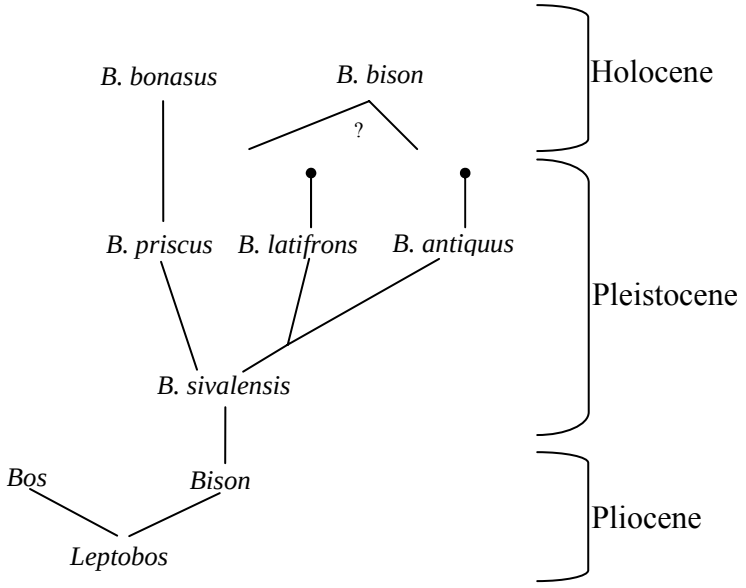

Figure 2. The Origins of *Bison*

The first *Bison*, *Bison sivalensis* or *B. sivalensis*, dispersed into Siberia during the Middle Pleistocene and gave rise to *B. priscus* (also called *B. alaskensis*), *B. latifrons*,

and *B. anitquus*. These three species coexisted during the Late Pleistocene with *B*ison *priscus* ranging from Siberia to central North America and *B. latifrons* and *B. antiquus* sharing habitat in North America. The Pleistocene forms of *Bison* were replaced by two Holocene forms *Bison bison* (North American bison) and *Bison bonasus* (Eurasian bison). *B. bonasus* arose from *B. priscus*; however, the exact predecessor of *B. bison* is unclear although *B. priscus* is its most likely predecessor.

B. priscus, the dominant species of bison in the Late Pleistocene, is of particular interest due to its abundant presence in the fossil record and modern descendents in North America and Europe, *B. bison* and *B. bonasus* respectively. *B. priscus* first appeared in Beringia during the Sangamon Inter-glacial period approximately 100,000 ybp (McDonald 1981) and quickly migrated throughout mid-latitude North America. Fossil remains of *B. priscus* from the Late Pleistocene range from Siberia to mid-latitude North America, and the corresponding carbon^{14} dates indicate its presence throughout the Late Pleistocene disappearing in the early Holocene (Guthrie 1970; Kurtén and Anderson 1980; McDonald 1981; Sutcliff 1985; Shapiro and others 2004). At this time, *B. bison*, a smaller form than *B. priscus*, arose and quickly established itself forming a large population on the central plains of North America. However, *B. bison* experienced a known bottleneck between the late 1860s into the early 1880's due to over-kill, primarily by Anglo-American hunters (McDonald 1981). The population was reduced to fewer than 300 individuals from whom the present *B. bison* population descends (Dary 1989).

Phylogenetics

Phylogenetics is the study of genetic relationships among different organisms. Through phylogenetics, the phylogeny or evolutionary history of an organism can be

determined and is usually depicted using a phylogenetic tree. Furthermore, molecular phylogenetics examines aligned amino acid or nucleotide sequences to make inferences about the ancestral relationships of species (Hartl 2000). There are two concerns in analyzing sequence data for phylogenetic inference. First, two or more independent mutations may occur at the same site (called multiple hits), meaning that two sites that differ may have undergone multiple mutations (Thornton and DeSalle 2000). Secondly, there is the possibility of homoplasy in which two sequences are identical not because of identity by descent from a common ancestor but due to certain mutational events. Parallel, convergent, and/or reverse mutations can lead to sequence similarity without giving an accurate portrayal of the mutational events leading to the sequences (Hartl 2000). Because of both multiple hits and homoplasy, the number of observed differences between two aligned sequences may underestimate the true number of mutations that have occurred.

Nucleotide sequences are more sensitive to multiple hits and homoplasy than amino acid sequences because there are only four nucleotides. Therefore, methods have been developed to address the problems presented by multiple hits and homoplasy. These methods attempt to simplify the assumptions made about the mutational processes. The simplest model, called the one-parameter, model was developed by Jukes and Cantor (1969) and assumes that any nucleotide is equally likely to mutate or be substituted by any other nucleotide. However, because transition (i.e. purine to purine) and transversion (i.e. purine to pyrimidine) rates vary, this assumption is often violated (Wakeley 1994). Correcting for these differing rates, Kimura (1980) developed a two-parameter model which assigns different rates to transitions and transversions. Tamura and Nei (1993)

developed a more extensive model, correcting for multiple hits by taking into account differences in substitution rates of nucleotides and the inequality of nucleotide frequencies. Further models have been developed with up to 12 parameters; however, these models quickly become computationally intense.

Utilizing these models, molecular phylogenetics has changed the approach to biological classification of organisms as nucleotide data are independent of morphological characteristics. Generally, however, the phylogenies determined by molecular phylogenetics support the phylogenies inferred from morphology (Patterson and others 1993). Phylogenetic trees are constructed from molecular data in three main ways: distance-based methods, such as neighbor-joining and minimum evolution; parsimony methods; and the maximum likelihood method.

Distance-based methods use the pairwise differences between sequences to construct trees. For example, the neighbor-joining method sequentially groups the most closely related sequences (Saitou and Nei 1987), whereas the minimum evolution method examines every possible tree and selects the one that minimizes the total branch lengths (Rzhetsky and Nei 1992). Distance methods also assume a constant rate of evolution and perform poorly when this assumption is violated, as is often the case (Nei and Kumar 2000). Parsimony methods search among all possible trees to identify the tree with the minimum number of mutational steps or most parsimonious tree (Hartl 2000). However, when evolutionary rates vary widely among sequences, parsimony methods can exhibit long-branch attraction (Page and Holmes 1998). For example, there are four sequences with the known phylogeny of ((A, B), (C, D)) as seen in Figure 3. A and B share a

common ancestor as do C and D. The long branches of A and C have experienced more substitutions than the other branches.

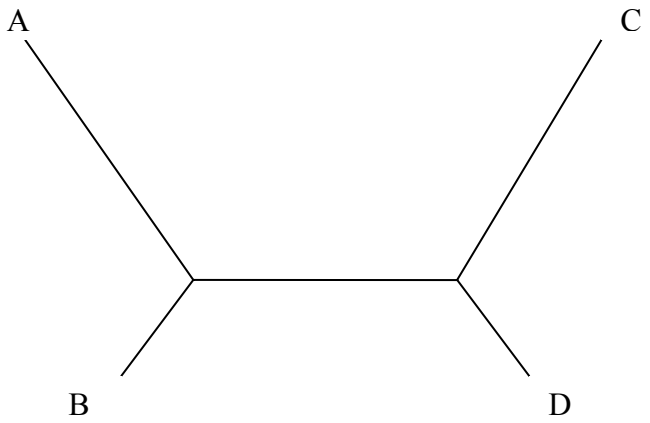

Figure 3. Long Branch Attraction Using Maximum Parsimony

If the number of substitutions is too great, the parsimony methodology will support the tree ((A, C), (B, D)) rather than the correct tree. This tendency toward long-branch attraction yields incorrect trees and is hard to detect. Lastly, the maximum likelihood method assumes a model of nucleotide substitution and identifies the tree that maximizes the probability of obtaining the observed sequence data. This method is tolerant of violations of its assumptions and performs well even when substitution rates differ in different branches. Intuitively, maximum likelihood gives the most accurate tree; however, this method requires intensive computations (Hartl 2000).

Research Hypotheses

Because of their wide distribution, sustained presence throughout the Pleistocene, and presence of modern descendents, *Bison* make an excellent model to investigate the events affecting megafauna at the Pleistocene-Holocene transition. By examining the

genetic diversity and phylogenetic relationships of *Bison* with respect to the mitochondrial DNA control region, several questions can be addressed:

1. How does the genetic diversity of Pleistocene *Bison* compare with the genetic diversity of *Bison bison*?

   Null Hypothesis: The amount of genetic diversity in the modern *Bison bison* population is the same as the amount of genetic diversity present in the Late Pleistocene *Bison* population.

2. Was the Pleistocene *Bison* population panmictic or was this population subdivided?

   Null Hypothesis: The Pleistocene *Bison* population was subdivided.

3. Is the genetic diversity present in Pleistocene *Bison* represented in the *Bison bison* population?

   Null hypothesis: The genetic diversity present in Pleistocene *Bison* is represented in the modern *Bison bison* population.

CHAPTER TWO

Materials and Methods

Sampling

Bone samples from 33 ancient bison skulls discovered in the Kolyma Region of Siberia were obtained. The skulls were identified morphologically as Pleistocene *Bison* by Dr. Andrei Sher and Dr. Pavel Nickolsky of the Ice Age Museum in Moscow and Geological Institute of the Russian Academy of Sciences. Two samples were donated from private collections, 29 samples were acquired from the Ice Age Museum in Moscow, Russia, and two samples were collected personally in the field at the Duvanny Yar site on the Kolyma River (Lat. 68°38'12" N, Long. 159°09'46" E). A Dremel® rotary tool was used to remove 4.0-5.0 grams of bone from each skull. After each sample was removed from the skull, the Dremel® was thoroughly cleaned with a sodium hypochlorite (bleach) solution followed by distilled, de-ionized water (ddH_20) and ethanol to prevent cross contamination of samples. Furthermore, separate blades were used for each sample and then discarded. The samples were placed in separate Ziploc® bags and inventoried by name, number, and photograph. Samples were labeled BP (*Bison priscus*) followed by a number beginning with 100 and transported to Baylor University by the author.

Contamination Prevention

Ancient DNA extraction, amplification, and sequencing require meticulous attention to detail to prevent contamination. Because of the low concentration of DNA in

ancient samples, modern DNA contamination can easily mask the aDNA at any stage of the analysis. Extraction, PCR, electrophoresis, and sequencing procedures were carried out in four separate rooms with separate ventilation systems establishing a gradient between the rooms. Samples were processed up-gradient from extraction room to PCR room to post-PCR rooms. For example, no PCR product was present in the extraction room at any time.

Before every procedure (extraction, PCR, etc.), all surfaces and tools were cleaned with a 10% dilution of store purchased sodium hypochlorite (bleach), followed by ddH_20, ethanol, and finally irradiated with ultra violet (UV) light for varying amounts of time. Prior to extractions, surfaces and tools were UV irradiated approximately 20 hours, and before PCR, reaction assembly surfaces and tools were UV irradiated for three hours. Sterile microcentrifuge tubes were used and UV irradiated as well. To prevent reagents and extraneous DNA from contaminating the pipettes, sterile, filtered pipette tips were used. Extractions and PCRs were carried out in isolated hoods that were sterilized as abovementioned between procedures. To control for the possibility of human contamination, PCRs were performed using the author's extracted genomic and mitochondrial DNA as a template with the designed bison primers yielding no amplicons. Furthermore, aprons covering the arms, surgical masks, and surgical caps were worn to prevent potential contamination. A double layer of powder free latex gloves was worn. The first pair was put on outside of the hood, and the second pair was put on from a glove box opened inside the hood. Additionally, mock extractions were conducted in the absence of any bone, using all buffers made in the lab to ensure that the buffers were free

from contamination. Extraction blanks and PCR blanks were also run with each sample to test for contamination.

Extraction

A silica/guanidine thiocyanate extraction method was used to extract DNA from the samples (Baker and others 2001). This method uses DNA-binding silica and is preferable to phenol/chloroform precipitation methods under low DNA concentration conditions. In addition to the silica extraction method, sample BP100 was also independently extracted using the Geneclean® for Ancient DNA kit (Catalogue # 1002-200) to confirm the DNA sequence of that sample. Both methods extract both nuclear and mitochondrial DNA.

Sample Preparation

Using a Dremel® rotary tool, 280-400 milligrams of bone from each sample was removed for extraction. Samples were cut in an external flow hood with tools and surfaces sterilized between samplings. Each bone sample was washed with a 10% bleach solution to remove surface contaminants, rinsed with ddH_20 to remove the bleach, and UV irradiated for 10 minutes. The samples were vacuum packaged, frozen in liquid nitrogen for eight minutes, wrapped in an absorbent liner, and crushed using a hammer. The crushed bone was placed in a sterile 1.5 mL tube. After each sample was processed, all tools used were washed with a 10% bleach solution and rinsed with ddH_20 and ethanol to remove contaminants.

Silica/Guanidine Thiocyanate Method (Baker and others 2001)

Extraction buffer (10 M guanidine thiocyanate (GuSCN) (Fisher Scientific®), 0.1 M Tris-HCl (pH= 6.4), 0.2 M EDTA (pH= 8.0), 1.3% Triton X-100) was prepared according to Boom and others (1990). Twelve grams of GuSCN, 10 mL of the Tris-HCl buffer, and 2.2 mL of the EDTA buffer were mixed in a 50 mL vial. The solution was placed in warm water to help dissolve the GuSCN. Two hundred fifty micro-liters of Triton X-100 (Acros® Organics) and 1.5 grams of DNA binding silica (Sigma®) were then added. The solution was vortexed and centrifuged twice for 10 minutes at 4,000 rpm to precipitate the silica from the extraction buffer. The supernatant was aliquoted into 1000 µL volumes in brown 1.5 mL tubes. The extraction buffer was stored in the dark and used for up to three weeks.

To extract the DNA, 1000 µL of extraction buffer was added to the 1.5 mL tubes containing the crushed bone samples and incubated at 60 degrees Celsius for at least 18 hours. The samples were removed from the incubator, centrifuged for five minutes at 13,000 rpm, and 250 µL of the supernatant was transferred to a new, sterile 1.5 mL tube. Five hundred micro-liters of sodium iodide (NaI) and five micro-liters of DNA glassmilk from Geneclean® III Kit (Catalogue # 1001-600) were added. The samples were vortexed then incubated at 55 degrees Celsius for 15 minutes. It is important to mix the extract well to promote the binding of the DNA to the silica in the glassmilk. The samples were centrifuged for five minutes at 13,000 rpm, and the supernatant was removed. To remove residual salts, 500 µL of New Wash prepared by manufacturer recommendation was added to the sample. The glassmilk pellet was re-suspended, the sample was centrifuged for five minutes at 13,000 rpm, and the supernatant removed.

This step was repeated. The pellet was re-suspended in 50 µL of ultra-pure, molecular bio-grade, nuclease free water purchased from USB Corporation® (Product # 71786). Samples were incubated at 56 degrees Celsius for 10 minutes and centrifuged for five minutes at 13,000 rpm. The supernatant was removed, placed in a sterile 0.6 mL tube, and stored at negative 20 degrees Celsius.

Geneclean® for Ancient DNA Kit

Crushed bone from the sample was placed in 1000 µL of DeHybernation Solution A and incubated at 55 degrees Celsius overnight. The extract was removed from the incubator, centrifuged for five minutes at 13,000 rpm, and the supernatant was transferred to a sterile 1.5 mL microcentrifuge tube. Three-hundred micro-liters of Ancient DNA Glassmilk were added, and the sample was incubated at room temperature for 30 minutes while mixing. The extract was transferred to the spin filter provided and centrifuged for one minute at 14,000 rcf. The catch tube was emptied, 500 µL of Salton Wash #1 was added, and centrifuged as above. After emptying the catch tube, 500 µL of Salton Wash #2 was added, and centrifuged as above. The catch tube was emptied, 500 µL of Ancient DNA Alcohol Wash was added, and centrifuged as above. The last step was repeated, and the sample was then centrifuged for two minutes at 14,000 rcf to remove any of the remaining alcohol wash. The pellet was re-suspended in 50 µL of ultra-pure, molecular bio-grade, nuclease free water, and the sample centrifuged for one minute at 14,000 rcf. The extract remaining in the catch tube was transferred to a 0.6 mL tube and stored at negative 20 degrees Celsius.

Extractions were performed in a Labconco® Purifier Class II Biosafety Cabinet in a room dedicated to extraction only. Extreme caution was exercised throughout the

extraction procedure to control for contamination. Furthermore, when multiple samples were processed concurrently, spatial separation within the hood was strictly maintained. No more than eight samples were processed at one time.

Primer Design

Primers were obtained from Bio-Synthesis, Inc. (Lewisville, Texas). Primers were designed using modern mitochondrial DNA (mtDNA) control region sequences of *Bison bison* (North American bison), *Bison bonasus* (Eurasian bison), and their respective cattle crossbreeds (See Appendix A). These sequences were obtained from Dr. James Derr of the Department of Veterinary Pathobiology at Texas A&M University. The sequences were aligned, and the primers were designed from the most conserved regions of the mtDNA control region between the respective sequences except for primer F15755 (See Table 2). Primer F15755 was developed using a previously published primer by Shapiro and others (2004) and a *Bos taurus* (cow) mtDNA genome sequence containing the control region (Genbank NC_006853) because the mitochondrial sequence needed to create F15755 was outside of the sequence provided by Dr. Derr.

To standardize the numbering system for the ancient bison mtDNA control region, the first forward primer (F15755) was located on the *Bos taurus* mtDNA sequence (NC_006853). The 5' base of F15755 corresponds to the base numbered 15755 in this *Bos taurus* mtDNA sequence. Furthermore, two *Bison priscus* sequences obtained from Genbank (AY748794 and AY748796) were compared to the *Bos taurus* sequence to find a base number for their first base with respect to the *Bos taurus* sequence. Once the base number at the beginning of the *Bison priscus* sequence was determined, the remaining sequence was numbered accordingly. Except for F15755, the primers were

numbered according to their 3' base position in the *Bison priscus* Numbered Reference Sequence in Appendix A. Primer F15755 begins outside of this reference sequence.

Table 2. Paired Primers

Paired Primers 5'-3' (F=Forward, R=Reverse)	Annealing Temperature (degrees Celsius)
F15755 ACCCCCAAAGCTGAAGTTCT R15925 GGTGAGTGACGTTAGCTAGTG	59
F15869 AAAAATTCCAATAACTCAACA R16014 GGGCATATAATTTAATGTAC	47
F16004 CATATTATGTATATAGTACATTAAA R16156 CGGCATGGTAGTTAAGCTCG	53
F16146 CACTAGATCACGAGCTTAAT R16264 GAAGAAAGAACCAGATGTCT	51
F16254 GAACTTTATCAGACATCTGG R16376 TACCAAATGAATGACAGCAC	51
F16366 ACACATAACTGTGCTGTCAT R16512 TTATGTCCTGTAACCATTGA	51

The primers are between 20 and 25 base pairs in length with annealing temperatures between 47 and 59 degrees Celsius. The annealing temperatures were not allowed to exceed the elongation temperature required for the sequencing reaction (60 degrees Celsius). Annealing temperatures were determined using the formula [(A+T) * 2 + (G + C) * 4] – 5]. This is a simplified formula to determine approximate annealing temperatures. The annealing temperature of a primer can also be estimated as 5-10 degrees Celsius lower than the melting temperature (T_m) (Erlich 1989). As indicated by the formula, the hydrogen bonding patterns between adenine (A)/thymine (T) with two hydrogen bonds and the guanine (G)/cytosine (C) with three hydrogen bonds affect the annealing temperature. A primer designed in an A/T rich region requires more bases to

reach an acceptable annealing temperature as opposed to a primer designed in a G/C rich region.

Six over-lapping primer sets were created to ensure continuous sequence reads. Primer pairs produced amplicons between 118 bp and 170 bp resulting in a 528 bp sequence of the mitochondrial DNA control region. Furthermore, Netprimer (2005) was used to assess primer dimerization, hairpin formations, and to calculate melting temperatures all of which affect the optimization of PCR (Griffin 1994). The primer concentrations were adjusted using molecular bio-grade, nuclease-free water to make 20 micro-molar (μM) stocks of each primer. Subsequently, aliquots of 25 μL of each individual primer were pipette to prevent contamination of the stock and reduce the number of freeze/thaw activities.

Polymerase Chain Reaction (PCR)

All reactions were prepared in a Labconco® PCR Enclosure in a room dedicated to PCR. Standard PCR mixtures consisted of 2.5 μL of 10X PCR Gold Buffer (Applied Biosystems®), 1.7 μL of 25 mM $MgCl_2$ (Applied Biosystems®), 2.0 μL of 10 mM mixture of the four deoxynucleotide triphosphates (dNTPs) (Applied Biosystems®), 0.5 μL of each 20 μM stock primer (forward and reverse), 0.2 μL of AmpliTaq Gold™ (Applied Biosystems®), 1.0 μL of DNA template, and molecular bio-grade water for 25 μL total volume. Reaction mixtures were subjected to the PCR thermo-cycles of denaturation, annealing, and elongation as shown in Table 3.

However, some templates were difficult to amplify due to apparent co-extracted inhibitors. It was sometimes necessary to dilute DNA templates to dilute inhibitors or to add additional template to obtain amplification. In such cases, serial dilutions of 1:10,

1:100, and 1:1000 were made using the original extract and molecular bio-grade water. Furthermore, the amount of diluted extract added to the PCR mixture was adjusted between 1 µL, 2 µL, and 5 µL depending on the sample.

Table 3. Cycles of the Polymerase Chain Reaction

	Initial Denaturation	Denaturation	Annealing	Elongation	# of Cycles	Final Elongation
Temp.	95°C	95°C	Varies	70°C	60	70°C
Time	4 min.	30 sec.	30 sec.	30 sec.		7 min

All reactions were amplified on either a PTC-100™ or a DNA Engine Opticon™ 2 both made by MJ Research®. Unlike the PTC-100™, the DNA Engine Opticon™ 2 has a gradient temperature block allowing for several annealing temperatures, and therefore several primer pair reactions, to be utilized concurrently. All primer sets were attempted simultaneously with sample BP115 with successful results. However, due to the unpredictable and inconsistent results produced by the DNA Engine Opticon, all other samples were processed using just one primer set at a time and running duplicate or triplicate reactions.

DNA Visualization and Purification

The PCR products were electrophoresed in a 12% acrylamide gel, stained with ethidium bromide (EtBr), and visualized on a UV transilluminator to estimate quantity and quality. Electrophoresis is the process of separating DNA strands of different lengths by applying an electric current to the gel which contains the DNA. Strands of DNA are negatively charged, and when an electric field is applied, the DNA will migrate or run in

the direction of the positive charge. Using a medium such as a polyacrylamide gel, strands of different lengths will migrate at different speeds through the gel creating bands of DNA (Chrambach and Rodbard 1971). The different bands can then be visualized using EtBr, which binds to DNA and fluoresces when placed in ultra violet light (Figure 4).

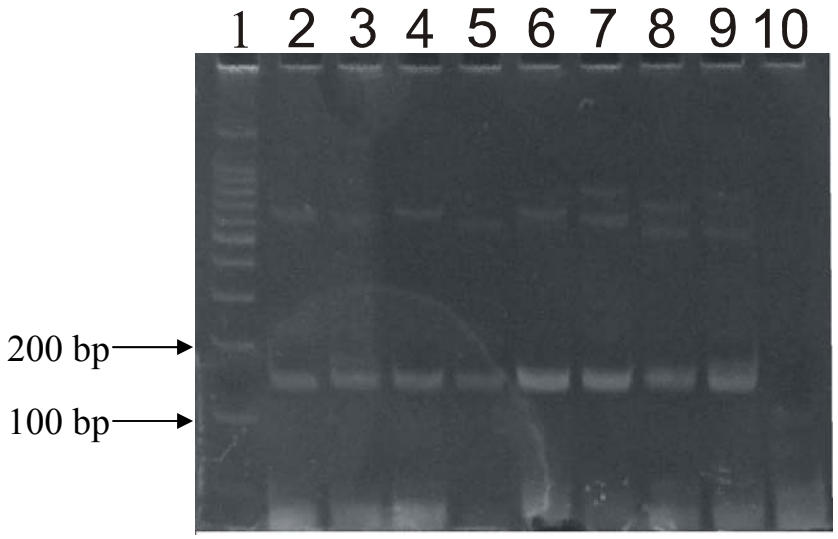

Figure 4. Polyacrylamide Gel of PCR Products from Eight Samples of Ancient *Bison* Mitochondrial DNA using Primers F16146 and R16264. Lane 1- 100 base pair DNA Ladder, Lanes 2-9- Samples, Lane 10- PCR Blank.

All apparatuses used to make polyacrylamide gels and execute polyacrylamide gel electrophoresis (PAGE) were purchased from Bio-Rad® (Mini Protean® 3 Series). Twelve percent polyacrylamide gels were prepared in vertical plates at 0.75 mm thickness with combs forming 10 wells capable of holding up to 25 µL. To make a 12% polyacrylamide gel, 3.85 mL of acrylamide, 2.0 mL of bis-acrylamide, 3.0 mL ddH_20, 1.0 mL of 10X TAE, 210 µL of 3% ammonium persulfate, and 5.0 µL of TEMED (N, N, N, N–Tetramethyl Ethylenediamine) were mixed in a 50 mL beaker and poured into the assembled vertical plates. Four to six gels were prepared at a time. Gels were left

standing for 30 minutes to form. For storage, the ends of the gel plates were covered by damp Kimwipes®, wrapped in a plastic wrap, and stored at a four degrees Celsius. The gels could be stored up to three months. For PAGE, gels were removed from storage, unwrapped, and the comb was removed. The gel was then loaded onto the vertical PAGE apparatus. Careful attention was given to firmly locking the gel in position as leaking of the running buffer can occur when using vertical gels, and 1X TAE was added as the running buffer.

To identify the target band by amplicon size, a 100 bp DNA Ladder by Promega® (Catalogue # G2101) was used. The ladder consisted of 11 double-stranded DNA fragments with sizes increasing by 100 bp increments from 100 bp to 1,000 bp and a 1,500 bp fragment. Five micro-liters (650 ng of DNA) of the ladder, which contains approximately 150 ng of the 500 bp DNA fragment and 50 ng of each of the other ten DNA fragments, was used per gel. Samples were loaded using the Blue/Orange loading dye that is provided with the above DNA ladder by Promega® (Catalogue # G190A). PAGE was run at 150 volts for 75 minutes using a FB300 power source (Fisher Scientific®). After electrophoresis, gels were removed from the plates and placed in an EtBr bath and visualized by a Foto/UV® 21 ultra violet light box (Fotodyne® Incorporated). The gel image was captured using a Kodak® EDAS 290 and processed using the Kodak® 1D 3.6.1 imaging system.

Target DNA was then gel purified using the bandstab technique (Wilton and others 1997). A bandstab is a gel purification method in which a pipette tip is used to stab the target DNA band displacing a small portion of the gel into the pipette tip. Using a transfer pipet, the displaced gel is expelled from the pipette tip into a previously

prepared PCR reaction mixture to serve as the template (Wilton and others 1997). The previously prepared PCR mixture contained the same constituents as described earlier in the PCR section. The thermo-cycling conditions of the PCR were also the same but 30 cycles were run instead of 60 cycles.

DNA Quantification

The bandstab product was purified using the QIAquick® PCR Purification Kit from Qiagen® (Catalogue # 28106). The protocol provided by the manufacturer was followed except for the final elution step. The DNA was eluted from the filter with 30 µL of molecular bio-grade water for sequencing purposes as opposed to the elution buffer provided in the kit. To assess the concentration of DNA in the purified sample, five micro-liters of sample DNA were run on a 12% polyacrylamide gel at 150 volts for 75 minutes using the 100 bp DNA Ladder by Promega®. Figure 5 shows the resultant gel after purification of the PCR product.

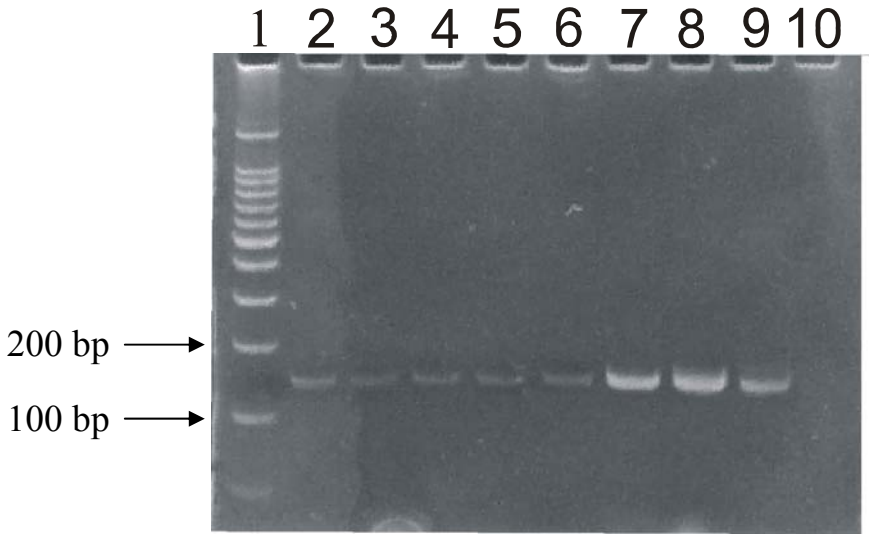

Figure 5. Polyacrylamide Gel After PCR Purification. Lane 1- 100 base pair DNA Ladder, Lanes 2-9- PCR Purified Samples.

Since the lengths and masses of the bands of the DNA ladder are known (see previous section), the Kodak® 1D 3.6.1 imaging system software could be utilized to calculate the band masses of the target bands by comparing them to the known band masses of the DNA ladder (Kodak® 1D 3.6.1 Users Manual). The concentration of DNA can then be determined by dividing the calculated band masses by the sample volume electrophoresed (five micro-liters). Purified sample concentrations ranged from 6 ng/μL to 60 ng/μL.

Sequencing

The CEQ™ DTCS Quick Start Kit was used for sequencing. The provided DTCS Master Mix consists of the deoxynucleotide triphosphates (dNTPs) needed for DNA strand replication, individual dye conjugated dideoxynucleotide triphosphates (ddNTPs), and all other PCR constituents except template DNA, ddH_20, and primers (CEQ™ 8000 Customer Training Guide). ddNTPs lack the hydroxyl group attached to the 3' carbon of the deoxyribose sugar. Therefore, with the addition of a ddNTP, elongation of the DNA strand is prevented terminating the sequence at a specific nucleotide. The resulting product is a mixture of DNA strands of different lengths representing the entire template strand with each terminal 3' end corresponding to the specific ddNTP that was incorporated (Sanger and others 1977). In addition, each ddNTP is conjugated to a unique fluorescent dye, emitting a specific wavelength of light when fluoresced.

Ten micro-liter reactions consisting of four micro-liters of DTCS Master Mix, 1.0 μL of 5.0 μM of a single primer, template DNA, and ddH_20 were created. Both forward and reverse primers were used for sequencing, but only one primer (forward or reverse) was used per reaction, yielding single stranded DNA products for sequencing. The

volume of template DNA added varied with the concentration of the purified sample. Double stranded DNA (dsDNA) concentrations should range between 50 and 100 femtamolar according to manufacturer specifications. To meet this criterion for dsDNA 200 bp in length, approximately 15 ng of dsDNA was needed. Previously determined template DNA concentrations were used to determine the volume of each purified sample that was added to the reaction mixture, and the amount of ddH_20 was adjusted to bring the final reaction volume to 10 μL.

For optimal results sequencing reactions were prepared on ice to prevent DNA polymerase activity. The samples were subjected to 30 thermal cycles of 96 degrees Celsius for 20 seconds (denaturation), the annealing temperature of the respective primer used for 20 seconds, and 60 degrees Celsius for two minutes (elongation). The products were then purified using Ethanol Plate Precipitation in a CEQ™ Sample Plate as instructed by the manufacturer. A 50 μL stop solution consisting of 20 μL of 3 M sodium acetate (NaOAc) (pH= 5.2), 20 μL of 100 mM EDTA (pH= 8.0), and 10 μL of glycogen was made, and 5.0 μL of this premix was added to the 10 μL sequencing reaction product. The above recipe results in sufficient stop solution for one column (eight samples). The samples were transferred to a CEQ™ sample plate, and 60 μL of cold 95% ethanol was added to each well. The sample plate was covered with aluminum foil to prevent sample loss. The samples were mixed and centrifuged at 4,000 rpm for 15 minutes at four degrees Celsius. After centrifugation, the aluminum foil was removed, the plate was inverted on eight folds of paper towels cut to fit the centrifuge plate holder, and centrifuged at 300 rpm for 20 seconds to remove the ethanol. The resulting DNA pellets were rinsed with 200 μL of 70% cold ethanol and centrifuged again at 4,000 rpm

for 15 minutes at four degrees Celsius. After centrifugation, the plate was inverted again onto eight folds of paper towels cut to fit the centrifuge plate holder and centrifuged at 300 rpm for 20 seconds. The 70% cold ethanol rinse was repeated to remove any remaining residual salts, and the pellets were vacuum dried for 10 minutes. The DNA pellets were re-suspended in 40 µL of Sample Loading Solution (SLS) provided by the manufacturer, and a drop of mineral oil was added to each well to prevent evaporation. The sample plate was covered by aluminum foil and stored at four degrees Celsius.

Samples were sequenced using the Beckman- Coulter CEQ™ 8000 Genetic Analysis System. The system utilizes capillary electrophoresis to separate DNA fragments of different sizes. By applying an electric field to the capillary tubes, ions will migrate in a medium such as polyacrylamide gel depending on their charge (Baker 1995). As the negatively charged DNA strands migrate through the gel, the dye-labeled ddNTPs are fluoresced by a laser, and a sequence is rendered. The LFR-1 separation method on the CEQ™ system with a 4.2 kV current for an 85 minute duration was chosen because of the fragment sizes (<200 bp) of the single stranded DNA being sequenced. Sequence data were originally analyzed using the default PCR sequence analysis parameters. To get a more accurate analysis, raw sequence data were re-analyzed indicating a PCR product <200 bp in length and defining a start and stop time for the analysis. The start time (typically around 20 minutes from the beginning of the run) and stop time (typically around 40 minutes from the beginning of the run) was inferred by reviewing the raw data and determining when the sequence first provided output and when the output ended. Analyzed data, as shown in Figure 6, were printed and read using the *Bison priscus* Numbered Reference Sequence (BPNS) in Appendix A as a guide. The International

Union of Biochemistry (IUB) ambiguity codes were used to represent heterozygous bases (See Table 4). Base changes were recorded only if both forward and reverse sequences showed the mutation.

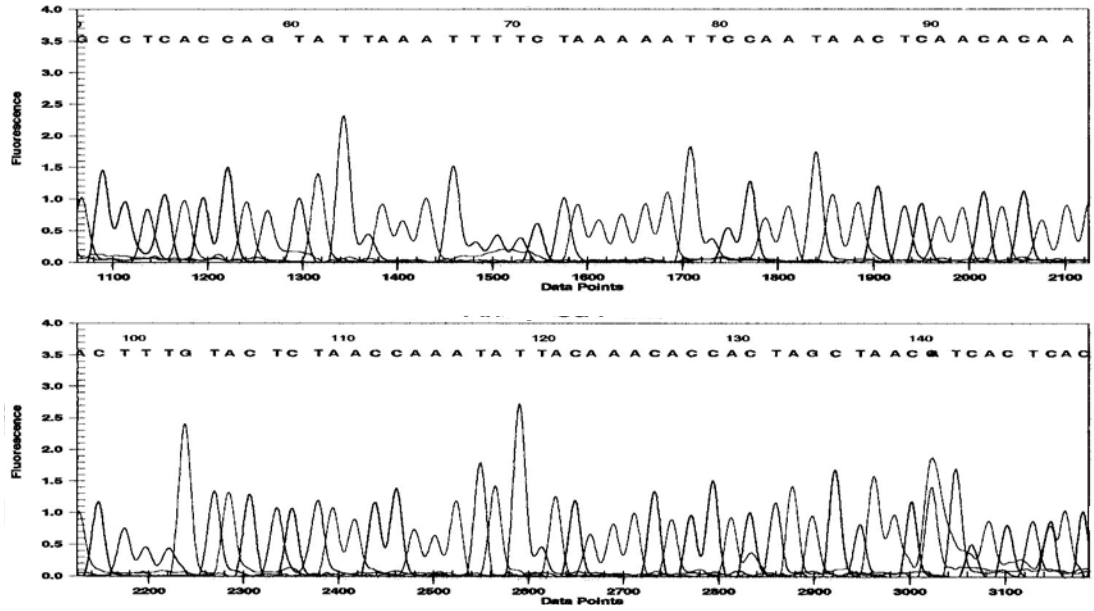

Figure 6. Portion of Analyzed Data from Sample BP100 Produced by CEQ™ 8000 Genetic Analysis System. Sequence Begins at Position 15825 on the BPNS.

Table 4. IUB Code Table (Liébecq 1992)

IUB (Degenerate Bases) Code Table											
IUB Code	N	V	B	H	D	K	S	W	M	Y	R
Bases	A,C,G,T	G,A,C	G,T,C	A,T,C	G,A,T	G,T	G,C	A,T	A,C	C,T	A,G

Sequence Alignment

The mtDNA control region sequences were aligned using Multiple Sequence Alignment by ClustalW® version 1.83 a freeware program (Thompson and others 1994). ClustalW® was used to align the sequences in a three step procedure. All the sequences are compared to each other to form pairwise alignments and then a dendogram is

constructed grouping sequences by similarity. The final multiple alignment is performed using the dendogram as a guide tree (Thompson and others 1994). For the pairwise alignments, the slow/accurate option was chosen with a gap penalty of 15 and gap extension penalty of 6.66. The program uses these parameters to score each pairwise alignment and subsequently to form the guide tree. Using the guide tree, the two gap penalties, and the pairwise scores, the most similar sequences are aligned first and the least similar sequences aligned last (Thompson and others 2003). Alignment results were carefully examined to ensure proper alignment. Manual adjustments were made to correct obvious mistakes. A guanine rich region at 15946 in the BRNS in Appendix A was eliminated from the analysis because of expected DNA polymerase slippage causing erroneous sequences (Derr 2006).

Analysis Software

After alignment, Arlequin® version 3.01 was used to analyze the standard molecular diversity indices (Excoffier and Schneider 2005). Furthermore, PHYLIP® DNAML version 3.5c (Felsentein 2004) using the Redhat Linux® version 3.23 operating system and MEGA® version 3.1 (Kumar and others 2004) using the Microsoft® Windows XP operating system were used to examine phylogenetic relationships. To obtain bootstrap values, the SeqBoot® and Consense® applications available in the PHYLIP® 3.65 package were used.

CHAPTER THREE

Results

Sequence Results

In comparing the two extraction methods, the silica/guanidine thiocyanate method produced more consistent results than the Geneclean® for Ancient DNA kit method. When using the Bio 101 Geneclean® for Ancient DNA kit, extracts had to be diluted before PCR in order to produce the target amplicons. With the silica/guanidine thiocyanate method, samples generally amplified successfully without having to dilute the original extract. Although mitochondrial control region sequence data were obtained for 12 of the 33 ancient bison samples, only 10 samples yielded enough target sequence data to include in the analysis, seven full target sequences (528 bp) and three partial target sequences (415 bp). These 10 samples comprise the BP population referred to below. The polymorphic loci of these 10 sequenced samples with respect to the *Bison priscus* Numbered Reference Sequence (BNRS) are presented in Appendix B. Two samples, BP100 and BP101, were carbon^{14} dated by the Arizona AMS Laboratory and found to be >39,900 and 32,100 +/- 1,300 years before present (ybp) respectively. BP100 yielded sequence data; however, BP101 did not. Carbon^{14} dating on the remaining samples which yielded target sequence data is ongoing.

For analysis, modern *Bison bison* mtDNA control region sequences (n=24) were obtained from Genbank (Accession Numbers AY428859, AF083357-64, U12935-6, U12944-48, U12955-59, U12864) and from Texas A&M University (*Bison bison* 1 and *Bison bison* 2 in Appendix A). These sequences represent several *Bison bison* herds

present in the United States and Canada, as well as *Bison bison* located at several zoos worldwide (Ward and others 1999; Wilson and Strobeck 1999). In the diversity analysis, these sequences comprise the modern bison population. In the phylogenetic analysis, the Genbank sequences are referred to by their respective accession numbers, and the Texas A&M sequences are identified as B.bis1 and B.bis2 respectively.

Ancient *Bison* mtDNA control region sequences were also obtained from Genbank. These sequences were published by Shapiro and others (2004), and geographical and carbon^{14} dating information can be found in the supporting online materials of the above publication available at *Science* Online. Only samples with full mtDNA control region sequences were used. These samples are indicated by SBS (Siberian Bison Sample) followed by a number or NABS (North American Bison Sample) followed by a number. Studies of the mtDNA control region sequences of *Bison bison* and various species of *Bos* have determined that interbreeding between the two has occurred over the last 10,000 years most likely due to low *Bison* population numbers (Ward and others 1999; Ritz and others 2000; Verkaar and others 2004; Derr 2006). Artifacts of the *Bos* control region are still present in some *Bison bison* control region sequences. To limit these *Bison* and *Bos* hybrid control region sequences, only ancient samples with carbon^{14} dates over 10,000 ybp and validated modern *Bison bison* samples were used for analysis.

Standard Diversity Indices

Standard genetic diversity indices provide information on the diversity present in sequence data. Using Arlequin® version 3.01, the number of polymorphic sites, mean number of pairwise differences, and nucleotide diversity were examined to assess

differences in diversity (Excoffier and Schneider 2005). A 5% level of missing data was allowed, and a rate of one was set for the transition, transversion, and insertion/deletion rates.

As previously mentioned, the ancient bison sequences from Genbank were divided into two geographic regions: North American (Alaska, Canada, and the United States), and Siberian (Siberia, the Ural Mountains, and Mongolia/Northern China). These divisions were labeled NABS (North American Bison Samples) (n=111) and SBS (Siberian Bison Samples) (n=77). To form an inclusive Siberian population, the BP (n=10) sequence data were included with the ancient Siberian bison sequence data forming SBS+BP (n=87). The BP sequences are a representative sample of the ancient bison population from the Kolyma Region of Siberia. Furthermore, the NABS population was divided into samples carbon^{14} dated between 10,000 and 30,000 ybp referred to as NABS^{1} (n=59) and samples carbon^{14} dated greater than 30,000 ybp referred to as NABS^{2} (n=52) to test for differences in diversity with respect to the timing of the Last Glacial Maximum. The dates of the samples in NABS^{1} place them during the Last Glacial Maximum up to the Pleistocene-Holocene transition whereas the dates of the samples of NABS^{2} precede the Last Glacial Maximum.

Since the diversity indices being examined are heavily dependent on population size, 24 randomly selected sequences from the SBS+BP, SBS, NABS, NABS^{1}, and NABS^{2} populations were used for analysis to match the number of modern *Bison bison* sequences available. The BP sequences generated by the author were also separately analyzed from the SBS sequences. The diversity indices presented give measurements of the genetic diversity within each population.

Number of Polymorphic Sites

The number of polymorphic sites is a measure of the number of nucleotide positions where a polymorphism exists in any one of the sequences analyzed. Because the number of usable loci varies between populations, the percentage of polymorphic sites per number of usable loci is given for better comparison. Table 5 gives the number of polymorphic sites observed in each population, the number of usable loci, and the percentage of polymorphic sites per usable loci.

Table 5. Number of Polymorphic Sites (n=24 for all populations except BP)

	SBS + BP	BP (n=10)	SBS	NABS	NABS[1] (<30,000)	NABS[2] (>30,000)	Modern
Number Polymorphic Sites	48	35	41	69	63	75	20
Number of Usable Loci (<5 % missing data)	409	409	522	500	500	500	516
Polymorphic Sites per Usable Loci	0.1174	0.0856	0.0785	0.1380	0.1260	0.1500	0.0388

NABS, NABS[1], and NABS[2] have the highest percentage of polymorphic sites of all the populations at 0.1380, 0.1260, and 0.1500 respectively, and the modern population has the lowest at 0.0388. NABS[2] has a slightly higher percentage of polymorphic sites than do the NABS[1] and the NABS populations. Interestingly, SBS+BP at 0.1174 has a higher percentage than both of its constituents. This can most likely be attributed to sampling error.

Mean Number of Pairwise Differences

The mean number of pairwise differences is the mean number of differences between all the haplotypes in the sample when compared to every other sequence individually (Tajima 1983; Tajima 1993). Table 6 gives the mean number pairwise differences for the populations examined with one standard deviation.

Table 6. Mean Number of Pairwise Differences (n=24 for all populations except BP)

	SBS+BP	BP (n=10)	SBS	NABS	NABS[1] (<30,000)	NABS[2] (>30,000)	Modern
Mean Number of Pairwise Differences	8.370 +/- 4.015	9.733 +/- 4.875	8.844 +/- 4.226	13.866 +/- 6.450	14.322 +/- 6.652	13.591 +/- 6.328	3.949 +/- 2.049

Figure 7 shows the mean number of pairwise differences with lines indicating two standard deviations or a 95% confidence limit.

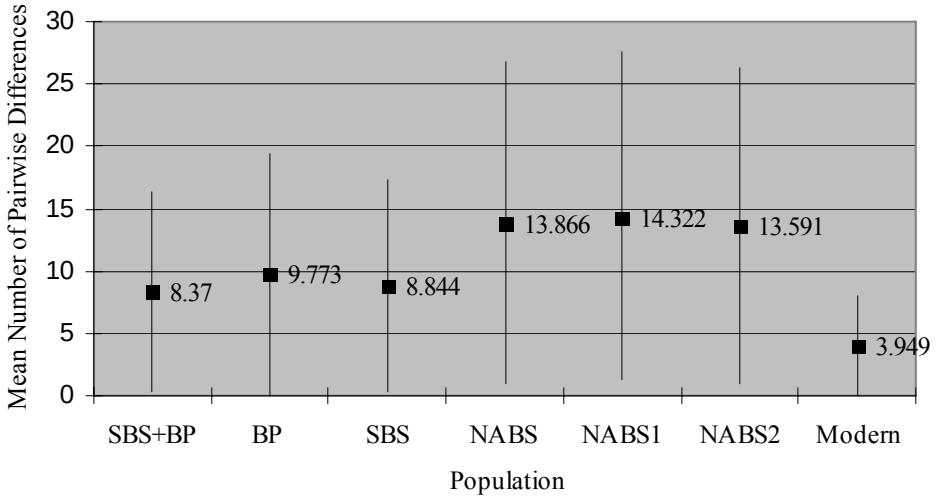

Figure 7. Mean Pairwise Differences +/- Two Standard Deviations

Standard deviation is the square root of the variance, and variance is an indicator of diversity. Similar sequences yield a lower variance, thus a lower diversity. The standard deviation of the modern *Bison bison* population is +/- 2.049, whereas standard deviations of the SBS+BP and NABS populations are +/- 4.015 and +/- 6.450 respectively. Furthermore, the populations containing the Siberian bison samples show slightly less variance than the populations containing the North American bison samples. However, the variances of the SBS+BP, SBS, and BP populations are similar, and the variances of the NABS, NABS[1], and NABS[2] are similar.

Nucleotide Diversity

Nucleotide diversity measures the probability of nucleotide differences per site between any two randomly selected sequences (Nei and Li 1979) and can also be used as a measure of diversity. Table 7 gives the nucleotide diversities for the populations examined with one standard deviation.

Table 7. Nucleotide Diversity (n=24 for all populations except BP)

	SBS+BP	BP (n=10)	SBS	NABS	NABS[1] (<30,000)	NABS[2] (>30,000)	Modern
Nucleotide Diversity (average over loci)	0.0204 +/- 0.010	0.0238 +/- 0.014	0.0169 +/- 0.009	0.0277 +/- 0.014	0.0286 +/- 0.015	0.0271 +/- 0.014	0.008 +/- 0.004

Figure 8 shows the calculated nucleotide diversity for each population with lines indicating two standard deviations. The variance of the nucleotide diversity of the modern *Bison bison* population is 0.004, whereas the variance of the SBS+BP and NABS populations are +/- 0.010 and +/- 0.014 respectively. These results agree with the mean

number of pairwise differences measurements. The variances of NABS, NABS[1], and NABS[2] are similar; however, the variances of SBS+BP, SBS, and BP are not as uniform. The BP population shows a variance more similar to the NABS, NABS[1], and NABS[2] populations than to SBS+BP and SBS. Population size may have affected these measurements.

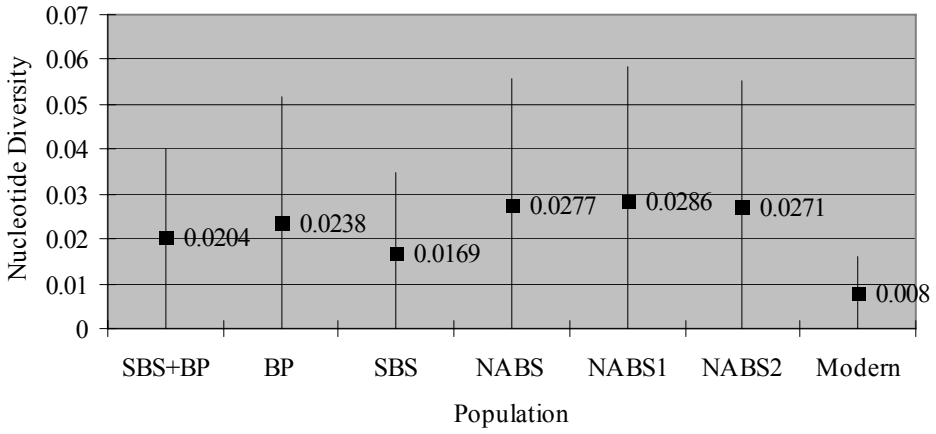

Figure 8. Nucleotide Diversity (over all loci) +/- Two Standard Deviations

Phylogenetic Analysis

Phylogenetic trees were constructed using DNAML from the PHYLIP® package, a maximum-likelihood program (Felsenstein 2004). The maximum-likelihood method identifies the tree that maximizes the probability of obtaining the observed sequences (Hartl 2000). The transition (i.e. base changes between purines) to transversion (i.e. base changes from purine to pyrimidine) ratio was set to 2.000. Two mtDNA control region sequences from the related genus *Bos* were obtained from Genbank and used as an out-group (Accession Numbers NC_006853 and NC_005971).

Figure 9 shows the maximum-likelihood tree created with the modern bison population and the BP population. Branch lengths that are significantly different from

zero at P<0.01 are marked with an asterisk above the branch, and bootstrap values over fifty percent are indicated below branches. Bootstrap values were obtained using 1000 replicates. Note that the modern bison population and BP population form two distinct clades. Moreover, the modern population does not branch from the BP population.

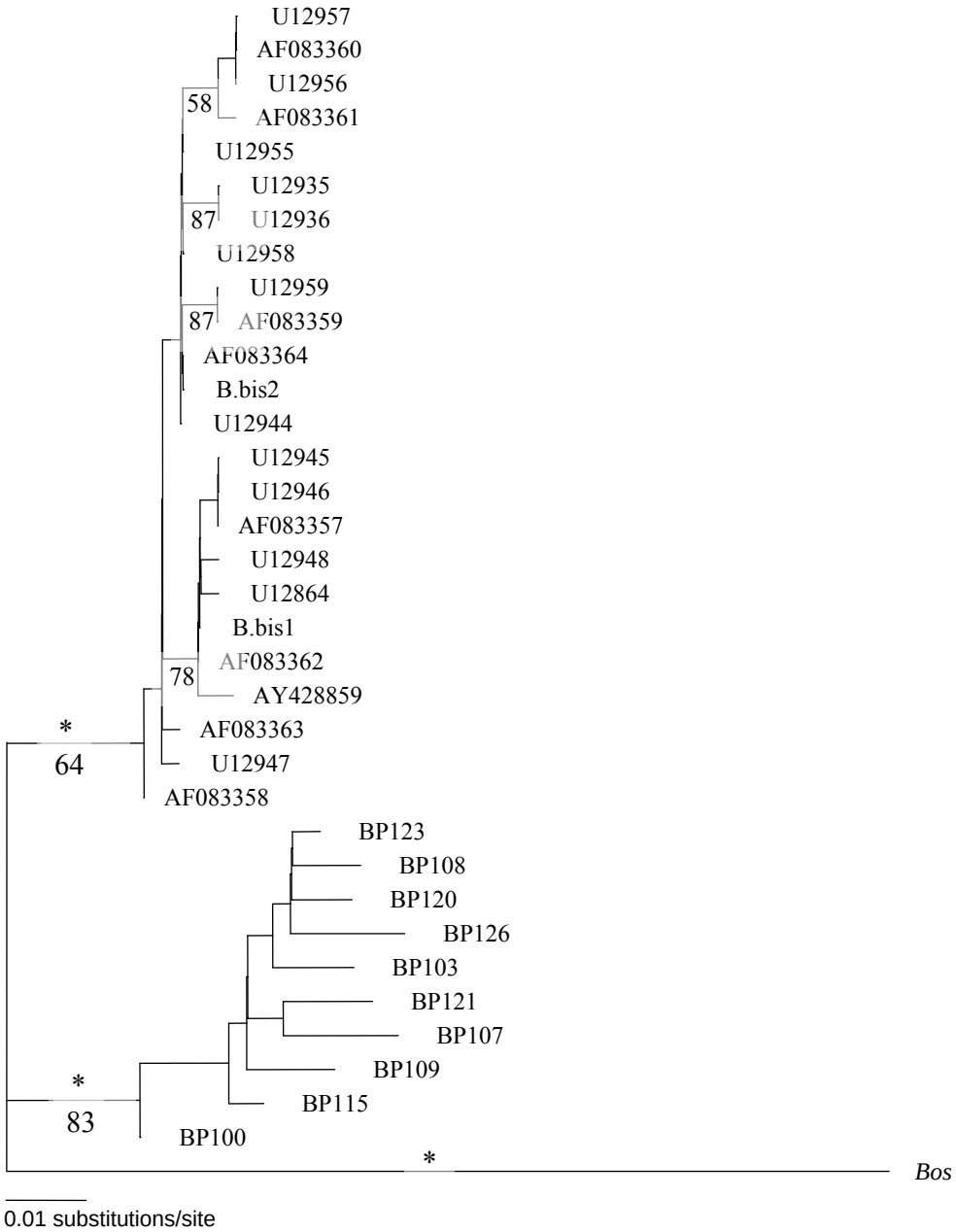

Figure 9. Maximum-Likelihood Tree of *Bison bison* and *Bison priscus*

For further examination, all ancient bison sequences (NABS, SBS, and BP) and modern bison sequences were analyzed to create a maximum-likelihood phylogenetic tree. The resulting maximum-likelihood tree is shown in Appendix C. Due to the overall size of the resulting tree, a simplified tree depicted in Figure 10 was constructed to show relevant branches within the phylogenetic tree. Appendix C is required in order to fully understand Figure 10.

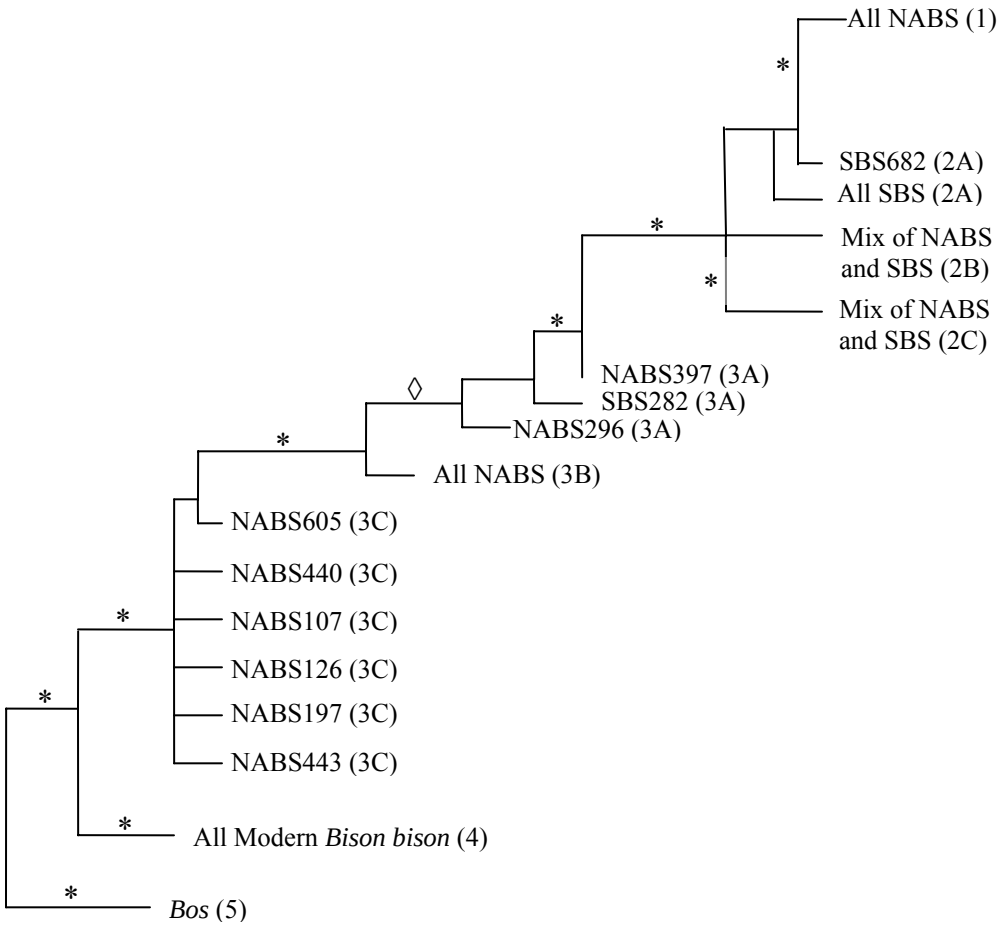

Figure 10. Overall Maximum-Likelihood Tree Simplified from Tree in Appendix C. An asterisk indicates branch lengths which are significantly different from zero at $P< 0.01$. A diamond indicates branch lengths which are significantly different from zero at $P< 0.05$. The number and letter in the parenthesis indicate Clade and Sub-Clade as indicated in Appendix C.

For simplification, five major clades are identified with Clades 2 and 3 having three sub-clades each (A, B, and C). Clade 1 consists of sequences from NABS only. As expected, the BP samples sequenced by the author are found in Clades 2 and 3 which consist of sequences from SBS and some sequences from NABS. The modern *Bison bison* population forms Clade 4 and branches separately from the NABS, SBS, and BP populations. Finally, Clade 5 consists of the two *Bos* sequences used as the out-group. The intensive computations required to generate bootstrap values for this tree are in process.

CHAPTER FOUR

Discussion and Conclusions

In comparison of modern *Bison bison* (*B. bison*) and Late Pleistocene North American *Bison* (also referred to as ancient North American *Bison*) with respect to the number of polymorphic loci, the ancient North American *Bison* had a 3.6 fold greater percentage of polymorphic loci than the modern *B. bison*. The mean number of pairwise differences and nucleotide diversity of ancient North American *Bison* show a 3.2 and 3.5 fold higher variance, respectively, than the modern *B. bison*. In addition, comparing the modern *B. bison* population to Late Pleistocene Siberian *Bison* (also referred to as ancient Siberian *Bison*) reveals a 3.0 fold greater percentage of polymorphic loci in ancient Siberian *Bison* than the modern population. Figure 11 shows the differences in the percentage of polymorphic sites per usable loci for the three populations.

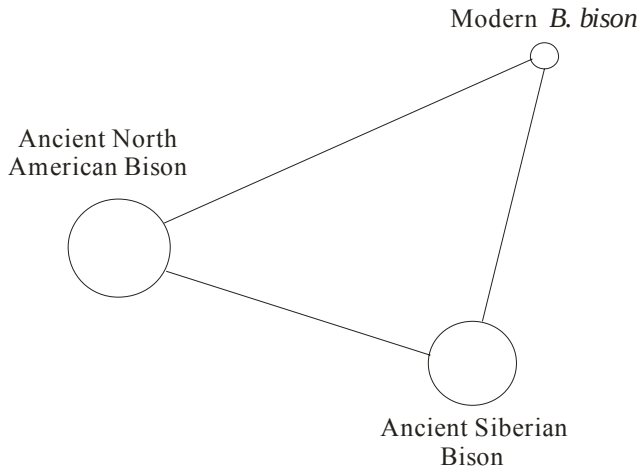

Figure 11. Differences in Percentage of Polymorphic Sites per Usable Loci

Furthermore, the mean number of pairwise differences and nucleotide diversity of the ancient Siberian *Bison* show a 2.0 and 2.5 fold higher variance, respectively, than the modern *B. bison* population. According to these data, the populations of ancient Siberian and ancient North American *Bison* exhibit considerably more diversity than the modern *B. bison* population.

Ancient North American *Bison* show a 1.7 and 1.4 fold increase in variance with respect to the mean number of pairwise differences and nucleotide diversity compared to the ancient Siberian *Bison* population. This indicates that slightly more genetic diversity may have been present in the ancient *Bison* population present in North America as opposed to that in Siberia. *Bison* and *Bos* split from *Leptobos* in Eurasia during the Pliocene, and *Bison* migrated to North America during the Pleistocene (McDonald 1981). According to this analysis, during the Late Pleistocene the *Bison* population in North America had more genetic diversity thus a stronger population even though the genus originally evolved in Eurasia.

Comparing the ancient North American *Bison* samples which were radiocarbon dated to the time period during the Last Glacial Maximum (LGM) and the ancient North American *Bison* samples which were radiocarbon dated to the time period before the LGM, the ancient North American *Bison* population during the LGM contains slightly more genetic diversity than the ancient North American *Bison* population before the LGM with respect to mean number of pairwise differences and nucleotide diversity. This would suggest that ancient North American *Bison* were increasing in genetic diversity up to the LGM (circa 30,000 ybp), and the genetic diversity of the Late Pleistocene *Bison*

population in North America slightly increased throughout the LGM to the Pleistocene-Holocene transition (circa 10,000 ybp).

Examining the phylogenetic tree shown in Figure 10, modern *B. bison* form a clade separate from the Late Pleistocene *Bison* sampled. The most closely related clades to *B. bison* (Clades 3B and 3C) contain samples discovered in North America indicating that the modern population is most closely related to a small group of ancient North American *Bison*. Furthermore, Shapiro and others (2004) found that the genetic diversity present in Late Pleistocene *Bison* is not represented in modern *B. bison*. This phylogenetic analysis supports their conclusion. The genetic diversity present in modern *B. bison* did not originate from the Late Pleistocene *Bison* population analyzed. One possible explanation is that the modern population originated from a bison population south of the ice sheet before the LGM that were not sampled in this study. Another possible explanation is that the population bottleneck of the late 1800s eliminated the late Pleistocene *Bison* mitochondrial DNA haplotypes from the *Bison* population.

In addition, Clade 3C contains three samples (NABS440, 443, 605) discovered south of the ice sheets formed during the LGM, and three samples (NABS107, 126, 197) discovered north of the ice sheets formed during the LGM. These six samples have radiocarbon dates ranging from 60,000 ybp to 20,000 ybp. Their grouping in the same clade suggests that similar haplotypes were present both north and south of the ice sheets formed during the LGM and indicates a panmictic population from mid-latitude North America to Alaska. These data may indicate that *Bison* migrated between Alaska and mid-latitude North America during the LGM when large ice sheets should have prevented this migration and supports postulation of the existence of a migrational route, if not an

ice-free corridor (IFC), between Alaska and mid-latitude North America throughout the LGM even though the IFC is not believed to have been viable as a migration route before 18,000 ybp (Pielou 1991).

To further analyze the IFC, radiocarbon dates of all the ancient North American *Bison* samples were compiled and mapped. Figure 12 and Table 8 show the locations of the Late Pleistocene North American *Bison* samples and respective radiocarbon dates with emphasis placed on the samples located in the IFC and mid-latitude North America.

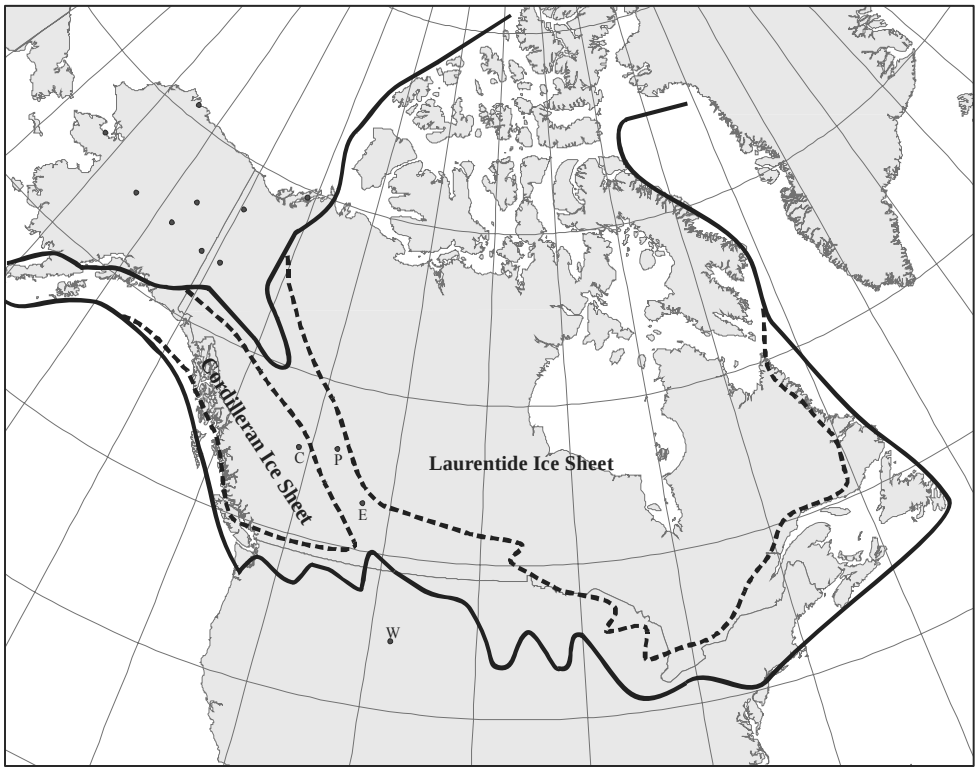

Figure 12. Ice Free Corridor. Solid Line Approximately Indicates LGM circa 30,000 ybp and Dashed Line Indicates Glacial Positions around 18,000 ybp. (C- Chetwynd, BC; P- Peace River, BC; E- Edmonton, AB; W- Natural Trap Cave, WY).

The samples discovered around Edmonton, Alberta, are very interesting with dates ranging from 60,400 +/- 2900 ybp to 34,050 +/- 450 ybp. According to the proposed

glacial boundaries during this part of the Wisconsinan glaciation, this area was covered by ice that prevented migration between Beringia and mid-latitude North America (Flint 1971; Pielou 1991). However, these samples indicate that *Bison* were present at this location during the Wisconsinan glaciation.

Table 8. Locations and Associated Radiocarbon Dates of Late Pleistocene North American *Bison* Samples.

Location	$Carbon^{14}$ Dates (ybp)
Alaska and Yukon Territory (Black Dots)	Dates from 1700 to >60,000
Peace River, British Columbia, Canada (P)	10,230 +/- 55 (NABS 254) 10,340 +/- 40 (NABS 342) 10,460 +/- 65 (NABS 202) 10,505 +/- 45 (NABS 348)
Chetwynd, British Columbia (C)	11,240 +/- 70 (NABS 237)
Edmonton, Alberta, Canada (E)	60,400 +/- 2900 (NABS 440) >59,400 (NABS 455) 56,300 +/- 3100 (NABS 473) >55,200 (NABS 452) >52,600 (NABS 457) 34,050 +/- 450 (NABS 443)
Natural Trap Cave, Wyoming, USA (W)	20,020 +/- 150 (NABS 359) 20,380 +/-90 (NABS 605)

In Figure 10, Sub-Clades 2B, 2C, and 3A are composed of a mix of ancient Siberian and ancient North American *Bison* samples. The mixing of these two populations indicates that migration existed both between Siberia and North America and between North America and Siberia. These data further support the hypothesis of a panmictic *Bison* population present in Beringia. Sub-Clade 2A, composed of ancient Siberian *Bison*, gives rise to Clade 1 which is composed of ancient North American *Bison* samples only. This suggests a dispersal of ancient *Bison* from Siberia into North America, forming a strong population which inhabited Eastern Beringia. However, these

ancient North American *Bison* samples are the most distantly related of all samples analyzed from the modern *B. bison* samples. This would indicate that the genetic diversity of the ancient North American *Bison* population comprising Clade 1 was not passed to the modern *B. bison* population.

In conclusion, the genetic diversity present in modern *Bison bison* is less than the genetic diversity that was present in Late Pleistocene *Bison*, and the genetic diversity of Late Pleistocene *Bison* is not represented in the modern population. Furthermore, although a small amount of subdivision is evident according to these data, the Late Pleistocene *Bison* population appears to be panmictic.

Future Research

The viability of ancient DNA research has been demonstrated with the successful sequencing of Late Pleistocene *Bison* mitochondrial DNA control region; however, further work is indicated. Other regions of ancient *Bison* DNA, both nuclear and mitochondrial, should be sequenced and similarly analyzed to substantiate the resultant phylogeny. In addition, *Bison* samples from Alaska to mid-latitude North America which span the Holocene Epoch (from 10,000 ybp to present) should be collected and sequenced, exercising great care to control for *Bison/Bos* hybrid sequences. Incorporating these data with those presented here from Late Pleistocene *Bison* samples and modern *Bison bison* samples will give further insight into the evolution and phylogeny of *Bison*. Furthermore, more Late Pleistocene species should be similarly studied to ascertain the population dynamics of other species at this time and compare these dynamics to those of *Bison*.

This study presented techniques by which ancient DNA can be extracted, amplified, sequenced, authenticated, and analyzed. By adhering to strict protocols, *Bison* mitochondrial DNA control region sequences were obtained from ancient bone samples, one of which was radiocarbon dated over 39,900 years before present. Through the analysis of these sequences, the genetic diversity and phylogeny of Late Pleistocene *Bison* was compared to that of an extant form of bison, *Bison bison*. Future ancient DNA studies should build and improve upon these techniques to examine other ancient populations.

APPENDICES

APPENDIX A

Bison Reference Sequences

Bison bison 1
AAAATTCCAATAACTCAACACAAACTTTGTACTCTAACCAAATACTGCAAACACCACTAGCTA
ACGTCACTCACCCCCAAAATGCATTACCCAAACGGGGGGAAATATACATAACATTAATGTAAT
AAAAACATATTATGTATATAGTACATTAAATTATATGCCCCATGCATATAAGCAAGTACTTAT
CCTCTATTGACAGTACATAGTACATAAAGTTATTAATTGTACATAGCACATTATGTCAAATCTA
CCCTTGGCAACATGCATACCCCTTCCATTAGATCACGAGCTTAATTACCATGCCGCGTGAAAC
CAGCAACCCGCTAGGCAGAGGATCCCTCTTCTCGCTCCGGGCCCATGAACCGTGGGGGTCGCT
ATTTAATGAACTTTATCAGACATCTGGTTCTTTCTTCAGGGCCATCTCACCTAAAATCGCCCAT
TCTTTCCTCTTAAATAAGACATCTCGATGGACTAATGGCTAATCAGCCCATGCTCACACATAA
CTGTGCTGTCATACATTTGGTATTTTTTTATTTTGGGGGATGCTTGGACTCAGCTATGGCCGTC
AAAGGCCCTGACCCGGAGCATCTATTGTAGCTGGACTTAACTGCACCTTGAGCACCAGCATAA
TGGTAAGCATGCACATATAGTCAATGGTTACAGGACATAAACTGTATTATATAT

Bison bison 2
AAAATTCCAATAACTCAACACAAACTTTGTACTCTAACCAAATATTGCAAACACCACTAGCTA
ACGTCACTCACCCCCAAAATGCATTACCCAAACGGGGGAATATACATAACATTAATGTAATAA
AAACATATTATGTATATAGTACATTAAATTATATGCCCCATGCATATAAGCAAGTACTTATCCT
CTATTGACAGTACATAGTACATAAAGTTATTAATTGTACATAGCACATTATGTCAAATCTACC
CTTGGCAACATGCATATCCCTTCCATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCA
GCAACCCGCTAGGCAGAGGATCCCTCTTCTCGCTCCGGGCCCATGAACCGTGGGGGTCGCTAT
TTAATGAACTTTATCAGACATCTGGTTCTTTCTTCAGGGCCATCTCACCTAGAATCGCCCATTC
TTTCCTCTTAAATAAGACATCTCGATGGACTAATGGCTAATCAGCCCATGCTCACACATAACT
GTGCTGTCATACATTTGGTATTTTTTTATTTTGGGGGATGCTTGGACTCAGCTATGGCCGTCAA
AGGCCCTGACCCGGAGCATCTATTGTAGCTGGACTTAACTGCACCTTGAGCACCAGCATAATG
GTAGGCATGCACATATAGTCAATGGTTACAGGACATAAACTGTATTATAT

Bison bonasus 1
AAATTTCCAATATCCAACACAGACTTTGTACCCCAACCAAATATTACAAACACCACTAGCTAA
CAGCATACACCCCATACACATATCACAGAACGCGCTGCCCAAGCAGGGTATATACATAATATT
AATGTATTAAAGACATAATATGTATATAGTACATTAAATTATATGCCCCATGCATATAAGCAA
GTACATGATGACTATTAATAGTACATAATACATACAATTATTAATTGTACATAACATATTATGT
CAAGTCCATTCTTGGTAACATACGTACCCCTTCCATTAGATCACGAGCTTAATTACCATGCCGC
GTGAAACCAGCAACCCGCTAGGCAAAGGACTTCTCTTCTCGCTCCGGGCCCATGAACCGTGGG
GGTCGCTATTTAATGAACTTTATCAGACATCTGGTTCTTTCTTCAGGGCCATCTCATCTAAAAC
CGTCCATTCTTTCCTCTTAAATAAGACATCTCGATGGACTAATGGCTAATCAGCCCATGCTCAC
ACATAACTGTGCTGTCATACATTTGGTATTTTTTTATTTTGGGGGATGCTTGGACTCAGCTATG
GCCGTCAAAGGCCCCGACCCGGAGCATATATTGTAGCTGGACTTAACTGCATCTTGAGCACCA
GCATAATGGTAAGCGTGGACATTACAGTCAATGGTTACAGGACATAATTATATTATATAT

Bison bonasus 2
AAAATTTCCAATATCCAACACAGACTTTGTACCCCAACCAAATATTACAAACACCACTAGCTA
ACAGCATACACCCCATACACATATCACAGAACGCGCTGCCCAAGCAGGGTATATACATAATA
TTAATGTATTAAAGACATAATATGTATATAGTACATTAAATTATATGCCCCATGCATATAAGC
AAGTACATGATGACTATTAATAGTACATAATACATACAATTATTAATTGTACATAACATATTA
TGTCAAGTCCATTCTTGGTAACATACGTACCCCTTCCATTAGATCACGAGCTTAATTACCATGC
CGCGTGAAACCAGCAACCCGCTAGGCAAAGGACTTCTCTTCTCGCTCCGGGCCCATGAACCGT
GGGGGTCGCTATTTAATGAACTTTATCAGACATCTGGTTCTTTCTTCAGGGCCATCTCATCTAA
AACCGTCCATTCTTTCCTCTTAAATAAGACATCTCGATGGACTAATGGCTAATCAGCCCATGCT
CACACATAACTGTGCTGTCATACATTTGGTATTTTTTTATTTTGGGGGATGCTTGGACTCAGCT
ATGGCCGTCAAAGGCCCCGACCCGGAGCATATATTGTAGCTGGACTTAACTGCATCTTGAGCA
CCAGCATAATGGTAAGCGTGGACATTACAGTCAATGGTTACAGGACATAATTATATTATATAT

Bison bison DC
AAAAAtCCCAATAACTCAACACAGAATTTGCACCCTAACCAAATATTACAAACACCACTAGCT
AACATAACACGCCCATACACAGACCACAGAATGAATTACCTACGCAAGGGGTAATGTACATA
ACATTAATGTAATAAAGACATAATATGTATATAGTACATTAAATTATATGCCCCATGCATATA
AGCAAGTACATGACCTCTATAGCAGTACATAATACATATAATTATTGACTGTACATAgTACATT
ATGTCAAATTCATTCTTGATAGTATATCTATTATATATTCCTTACCATTAGATCACGAGCTTAA
TTACCATGCCGCGTGAAACCAGCAACCCGCTAGGCAGGGATCCCTCTTCTCGCTCCGGGCCCA
TAAACCGTGGGGGTCGCTATCCAATGAATTTTACCAGGCATCTGGTTCTTTCTTCAGGGCCATC
TCATCTAAAACGGTCCATTCTTTCCTCTTAAATAAGACATCTCGATGGACTAATGGCTAATCAG
CCCATGCTCACACATAACTGTGCTGTCATACATTTGGTATTTTTTTATTTTGGGGGATGCTTGG
ACTCAGCTATGGCCGTCAAAGGCCCTGACCCGGAGCATCTATTGTAGCTGGACTTAACTGCAT
CTTGAGCACCAGCATAATGATAAGCGTGGACATTACAGTCAATGGTCACAGGACAT

Bison bonasus DC
AAAAAATCCCAATAACTCAACACAGAATTTGCACCCTAACCAAATATTACAAACACCACTAG
CTAACATAACACGCCCATACACAGACCACAGAATGAATTACCTACGCAAGGGGTAATGTACA
TAACATTAATGTAATAAAGACATAATATGTATATAGTACATTAAATTATATGCCCCATGCATA
TAAGCAAGTACATGACCTCTATAGTAGTACATAATACATATAATTATTGACTGTACATAGTAC
ATTATGTCAAATTCATTCTTGATAGTATATTTATTATATATTCCTTACCATTAGATCACGAGCTT
AATTACCATGCCGCGTGAAACCAGCAACCCGCTAGGCAGGGATCCCTCTTCTCGCTCCGGGCC
CATAAACCGTGGGGGTCGCTATCCAATGAATTTTACCAGGCATCTGGTTCTTTCTTCAGGGCC
ATCTCATCTAAAACGGTCCATTCTTTCCTCTTAAATAAGACATCTCGATGGACTAATGGCTAAT
CAGCCCATGCTCACACATAACTGTGCTGTCATACATTTGGTATTTTTTTATTTTGGGGGATGCT
TGGACTCAGCTATGGCCGTCAAAGGCCCCGACCCGGAGCATCTATTGTAGCTGGACTTAACTG
CATCTTGAGCACCAGCATAATGATAAGCGTGGACATTACAGTCAATGG

Bison priscus Numbered Reference Sequence
15760 AAAAGCTGAAGTTCTATTTAAACTATTCCCTGAACGCTATTAATATAGTTCCATAAATGC
15820 AAAGAGCCTCACCAGTATTAAATTTACTAAAAATTCCAATAACTCAACACAAACTTTGT
15880 ACTCTAACCAAATACTGCAAACACCACTAGCTAACGTCACTCACCCCCAAAATGCATTAC
15940 CCAAACGGGGGGAAATATACATAACATTAATGTAATAAAAACATATTATGTATATAGTAC
16000 ATTAAATTATATGCCCCATGCATATAAGCAAGTACTTATCCTCTATTGACAGTACATAGT
16060 ACATAAAGTTATTAATTGTACATAGCACATTATGTCAAATCTACCCTTGGCAACATGCAT
16120 ACCCCTTCCATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTAG
16180 GCAGAGGATCCCTCTTCTCGCTCCGGGCCCATGAACCGTGGGGGTCGCTATTTAATGAAC
16240 TTTATCAGACATCTGGTTCTTTCTTCAGGGCCATCTCACCTAAAATCGCCCATTCTTTCC
16300 TCTTAAATAAGACATCTCGATGGACTAATGGCTAATCAGCCCATGCTCACACATAACTGT
16360 GCTGTCATACATTTGGTATTTTTTTATTTTGGGGGATGCTTGGACTCAGCTATGGCCGTC
16420 AAAGGCCCTGACCCGGAGCATCTATTGTAGCTGGACTTAACTGCACCTTGAGCACCAGCA
16480 TAATGGTAAGCATGCACATATAGTCAATGGTTACAGGACATAAACTGTATTATATAT

APPENDIX B

Table of Polymorphisms

Sample	15894	15895	15896	15931	15940	15955
BPNS	C	T	G	-	C	T
BP100	T	-	A	-	-	C
BP103	T	-	A	-	-	C
BP107	T	-	A	-	T	C
BP108	T	-	A	A	-	C
BP109	T	-	A	-	-	C
BP115	T	-	A	-	-	C
BP120	T	-	A	A	-	C
BP121	T	-	A	-	-	C
BP123	T	K	A	-	-	C
BP126	T	-	A	A	-	C

Sample	15956	16038	16039	16040	16041	16048
BPNS	A	T	C	C	T	A
BP100	G	A	-	T	C	G
BP103	G	A	-	-	C	-
BP107	G	A	T	T	C	-
BP108	G	A	T	-	C	-
BP109	G	A	T	T	C	-
BP115	G	A	T	T	C	-
BP120	G	A	A	-	C	-
BP121	G	A	T	T	C	-
BP123	G	A	T	-	-	-
BP126	G	A	T	-	C	-

Sample	16049	16053	16057	16061	16067	16076
BPNS	C	A	A	C	G	T
BP100	T	-	-	-	-	-
BP103	T	-	-	-	-	C
BP107	T	G	G	-	-	C
BP108	T	-	-	-	-	C
BP109	T	-	-	T	A	C
BP115	T	-	-	-	-	C
BP120	T	-	-	-	-	C
BP121	T	-	G	-	A	C
BP123	T	-	-	-	-	C
BP126	T	-	-	-	-	C

Sample	16087	16103	16104	16109	16121	16130
BPNS	C	C	C	G	C	T
BP100	-	-	-	A	T	-
BP103	-	-	-	A	T	C
BP107	-	-	T	A	T	C
BP108	-	-	T	A	T	C
BP109	-	-	-	A	T	C
BP115	-	-	-	A	T	C
BP120	-	Y	T	A	T	C
BP121	T	-	-	A	T	C
BP123	-	-	T	A	T	C
BP126	-	-	T	A	-	C

Sample	16147	16183	16188	16189	16190	16205
BPNS	T	G	T	C	C	G
BP100	-	-	-	-	-	-
BP103	-	-	C	-	-	-
BP107	-	-	C	-	T	-
BP108	-	A	C	-	-	-
BP109	-	-	C	T	-	-
BP115	-	-	C	-	-	-
BP120	C	-	C	-	-	A
BP121	-	-	C	-	T	-
BP123	-	-	C	-	-	-
BP126	-	-	-	-	-	-

Sample	16213	16215	16216	16220	16239	16278
BPNS	A	C	C	G	C	C
BP100	G	T	T	-	-	T
BP103	G	T	T	A	T	T
BP107	G	T	T	-	T	?
BP108	G	T	T	-	T	?
BP109	G	T	T	-	-	-
BP115	G	T	T	-	T	-
BP120	G	T	T	-	-	T
BP121	-	T	T	-	-	T
BP123	G	T	T	-	-	T
BP126	-	T	T	-	-	?

Sample	16285	16286	16288	16368
BPNS	T	C	C	A
BP100	C	T	T	-
BP103	C	-	T	T
BP107	?	?	?	?
BP108	?	?	?	?
BP109	-	T	T	T
BP115	C	T	T	-
BP120	C	T	T	T
BP121	C	T	T	T
BP123	C	-	T	T
BP126	?	?	?	?

BPNS- *Bison priscus* Numbered Reference Sequence
BP- *Bison priscus*

APPENDIX C

Maximum Likelihood Tree

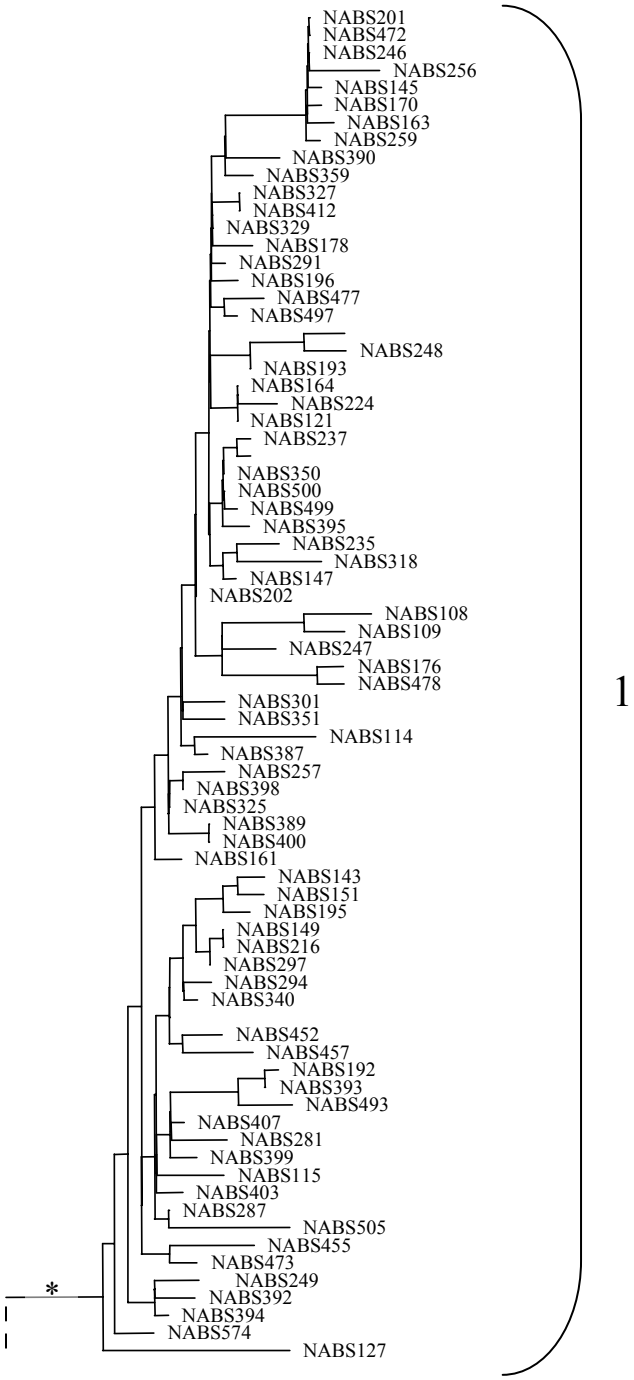

Figure C.1. Clade 1: Ancient North American Bison

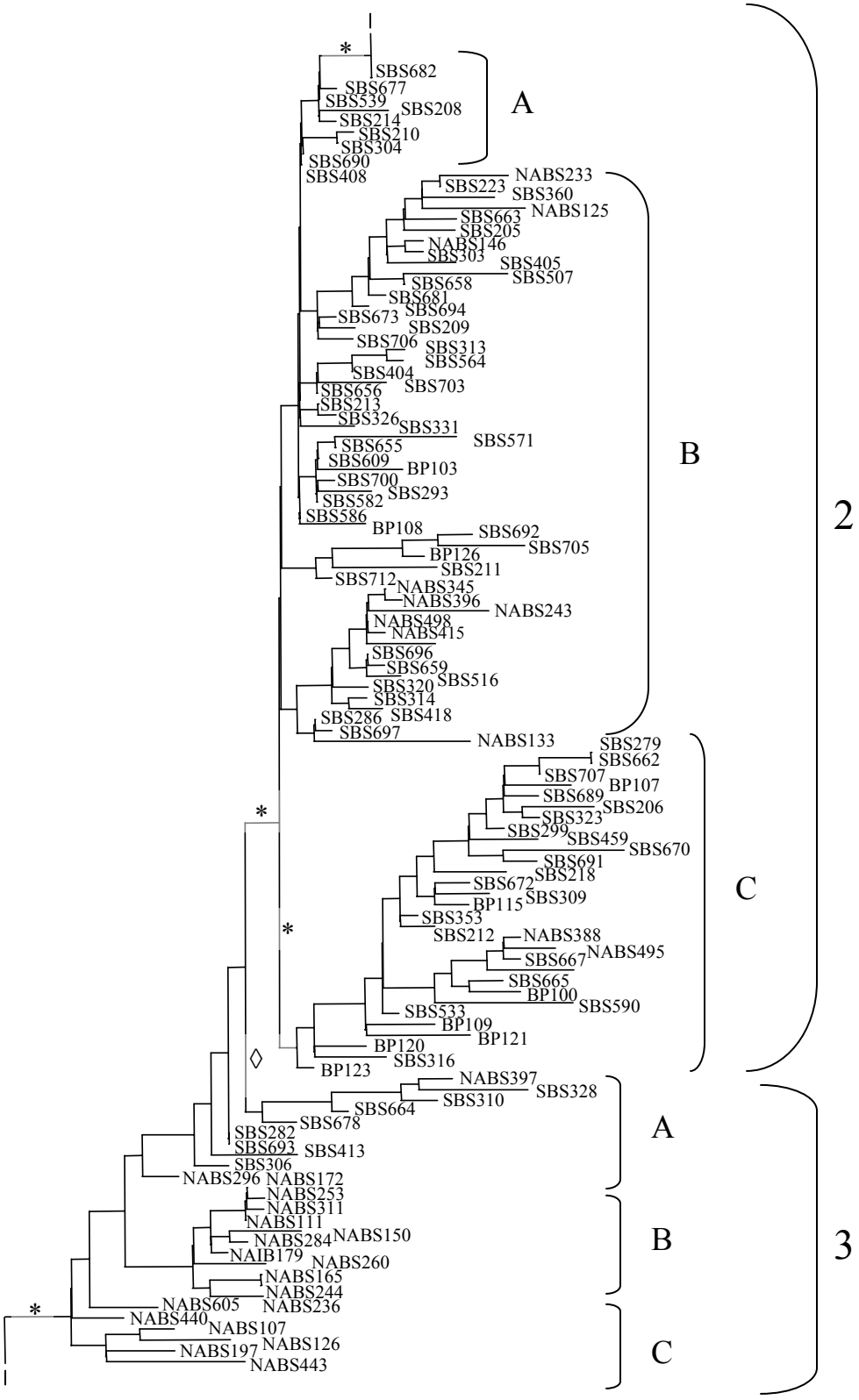

Figure C.2. Clades 2 and 3: Ancient Siberian and Ancient North American Bison

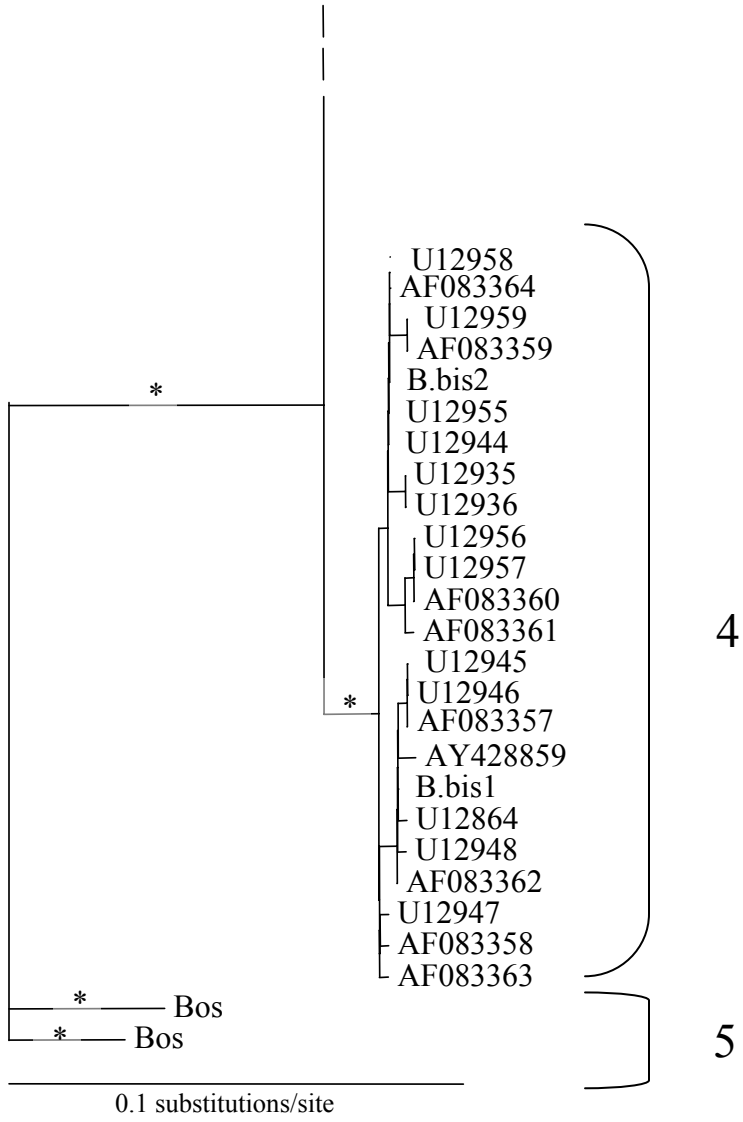

Figure C.3. Clades 4 and 5: Modern *Bison bison* and *Bos* (Outgroup)

Branch lengths significantly different from zero at P<0.01 are indicated by an asterisk whereas a diamond indicates branch lengths significantly different from zero at P<0.05 for Figures C.1 through C.3.

LITERATURE CITED

Anderson PM, Brubaker LB. 1994. Vegetation history of north central Alaska: a mapped summary of late-Quaternary pollen data. Quaternary Science Reviews 13: 71–92.

Baker D. 1995. Capillary Electrophoresis. New York: Wiley.

Baker LE, McCormick WF, Matteson KJ. 2001. A silica-based mitochondrial DNA extraction method applied to forensic hair shafts and teeth. J Forensic Science 46 (1): 126-130.

Barnes I, Matheus P, Shapiro B, Jensen D, Cooper A. 2002. Dynamics of Pleistocene population extinctions in Beringian brown bears. Science 295: 2267–2270.

Barnosky AD. 1989. The late Pleistocene event as a paradigm for widespread mammal extinction. In mass extinctions processes and evidence, SK Donovan editor. New York: Columbia University Press.

Boom R, Sol CJA, Salimans MMM, Jansen CL, Wertheim-van Dillen PME, Van der Noordaa J. 1990. Rapid and simple method for purification of nucleic acid. Journal of Clinical Microbiology 28: 495–503.

Bradley DG, MacHugh DE, Cunningham P, Loftus RT. 1996. Mitochondrial diversity and the origins of African and European cattle. Anthropology 93: 5131-5135.

Bereiter-Hahn J, Voth M. 1994. Dynamics of mitochondria in living cells: shape changes, dislocations, fusion, and fission of mitochondria. Microscopy Research and Technique 27: 198-219.

Bogenhagen D, Clayton DA. 1974. The number of mitochondrial deoxyribonucleic acid genomes in mouse L and human HeLa cells. J Biol Chem 249: 7991–7995.

Brown WM, George M, Wilson AC. 1979. Rapid evolution of animal mitochondrial DNA. Proc. Natl. Acad. Sci. USA 76: 1967-1971.

Burns, JA. 1996. Vertebrate paleontology and the alleged ice-free corridor: the meat of the matter. Quaternary International 32: 107-112.

Cooper A, Poinar H. 2000. Ancient DNA: do it right or not at all. Science 289: 530-531.

Chrambach A, Rodbard D. 1971. Polyacrylamide gel electrophoresis. Science 72: 440-451

Creager JS, McManus DA. 1967. Geology of the floor of the Bering and Chukchi seas-american studies." In The Bering land bridge. DM Hopkins editor. Stanford: Stanford University Press.

Dary DA. 1989. The buffalo book. Chicago: The Swallow Press, Inc.

Derr J. 2006. Discussion of analysis of bison mtDNA control region.

Edwards ME, Anderson PM, Brubaker LB, Ager TA, Andreev AA, Bigelow NH, Cwynar LC, Eisner WR, Harrison SP, Hu FS, Jolly D, Lozhkin AV, MacDonald GM, Mock CJ, Ritchie JC, Sher AV, Spear RW, Williams JW, Yu G. 2000. Pollen-based biomes for Beringia 18,000, 6000 and 0 14C yr bp. Journal of Biogeography 27: 521-554.

Erlich HA. 1989. PCR technology. New York: Stockton Press.

Excoffier LG, Schneider S. 2005. Arlequin ver 3.0: an integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online 1:47-50.

Felsentein J. 2004. Phylip phylogeny inference package. Avail from: http://evolution.genetics.washington.edu/phylip.html.

Flint RF. 1971. Glacial and Quaternary Geology. New York: John Wiley and Sons, Inc.

Gilbert MTP, Shapiro B, Drummond A, Cooper A. 2005. Post-mortem DNA damage hotspots in bison (*Bison bison*) provide evidence for both damage and mutational hotspots in human mitochondrial DNA. Journal of Archaeological Science 32: 1053-1060.

Goldenberg E, Giannassi D, Clegg M, Smiley C, Durbin M, Henderson D, Zurawski G. 1990. Chloroplast DNA from a Miocene *Magnolia* species. Nature 344: 656-658.

Greenberg BD, Newbold JE, Sugino A. 1983. Intraspecific nucleotide sequence variability surrounding the origin of replication in human mitochondrial. Gene 21: 33-49.

Griffin HG, Griffin AM. 1994. PCR technology current innovations. Boca Raton: CRC Press.

Guthrie RD. 1970. Bison evolution and zoogeography in North America during the Pleistocene. The Quarterly Review of Biology 45: 1-15.

Guthrie RD. 1984a. The evidence for middle-Wisconsin peopling of Beringia: an evaluation. Quaternary Research 22: 231–241.

Guthrie RD. 1984b. Mosaics, allelochemics, and nutrients: An ecological theory of late Pleistocene magafaunal extinctions. In Quaternary Extinctions. PS Martin and RG Klein editors. Tucson: University of Arizona Press.

Guthrie RD. 1990. Frozen fauna of the mammoth steppe. Chicago: The University of Chicago Press.

Hansen AJ, Willerslev E, Wiuf C, Mourier T, Arctander P. 2001. Statistical evidence for miscoding lesions in ancient DNA templates. Molecular Biology and Evolution 18: 262-265.

Hartl D. 2000. A primer of population genetics. Sunderland, MA: Sinauer Associated, Inc.

Hofreiter M, Serre D, Poinar H, Kuch M, Pääbo S. 2001. Ancient DNA. Nature Reviews Genetics 2: 353-359.

Hofreiter M, Capelli C, Krings M, Waits L, Conard N, Munzel S, Rabeder G, Nagel D, Paunovic M, Jambresic G, Meyer S, Weiss G, Pääbo S. 2002. Ancient DNA analyses reveal high mitochondrial DNA sequence diversity and parallel morphological evolution of late Pleistocene cave bears. Molecular Biology and Evolution 19: 1244-1250.

Hopkins DM. 1967. The Bering land bridge. Stanford: Stanford University Press.

Hopkins DM, Mattews JV, Schweger CE, Young SE, editors. 1982. Paleoecology of Beringia. New York: Academic Press.

Ingman M, Gyllensten U. 2001. Analysis of the complete human mtDNA genome: methodology and inferences for human evolution. The Journal of Heredity 92: 454-461.

Joger U, Gorrido G. 2001. Phylogenetic position of *Elephas*, *Loxodonta*, and *Mammuthus* based on molecular evidence. The World of Elephants International Congress, Rome, Italy: 544-547.

Jukes TH, Cantor CR. 1969. Evolution of protein molecules. In mammalian protein metabolism. HN Munro editor. New York: Academic Press.

Kelly RL, Todd LC. 1988. Coming into the country: Early paleoindian hunting and mobility. American Anitquity 53: 231-244.

Kimura M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. Journal of Molecular Evolution 16: 111-120.

Kumar S, Tamura K, Nei M. 2004. MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment briefings. Bioinformatics 5: 150-163.

Kurtén B. 1988. Before the Indians. New York: Columbia University Press.

Kurtén B, Anderson E. 1980. Pleistocene mammals of North America. New York: Columbia University Press.

Kutzbach JE. 1987. Model simulations of the climatic patterns during the deglaciation of North America. In North America and adjacent oceans during the last deglaciation: the geology of North America K-3. WF Ruddiman and HE Wright editors. Geological Society of America, Boulder, CO: 425-446.

Lai SJ, Liu YP, Liu YX, Li XW, Yao YG. 2006. Genetic diversity and origin of Chinese cattle revealed by mtDNA D-loop sequence variation. Mol. Phylogenet. Evol. 38: 146-154.

Lambert DM, Ritchie PA, Millar CD, Holland B, Drummond AJ, Baroni C. 2002. Rates of evolution in ancient DNA from Adélie penguins. Science 295: 2270-2273.

Leonard JA, Wayne RK, Cooper A. 2000. Population genetics of ice age brown bears. Proc Natl Acad Sci USA 97: 1651-1654.

Liébecq C, editor. 1992. Biochemical nomenclature and related documents. 2nd edition. London: Portland Press.

Lindahl T. 1997. Facts and artifacts of ancient DNA. Cell 90: 1-3.

Margulis L, Sagan D. 1986. Origins of Sex: Three Billion Years of Genetic Recombination. New Haven: Yale University Press.

McDonald J N. 1981. North American bison: their classification and evolution. Berkeley, CA: University of California Press.

Meltzer DJ. 1995. Clocking the first Americans. Annual Review of Anthropology 24: 21-45.

MITOMAP: A human mitochondrial genome database. 2005. http://www.mitomap.org.

Mullis KB, Faloona FA. 1987. Specific synthesis of DNA in vitro via a polymerase catalyzed chain reaction. Meth. Enzymol. 155: 335-350.

Nei M, and Li W. 1979. Mathematical model for studying genetic variance in terms of restriction endonucleases. Proc. Nati. Acad. Sci. USA 76: 5269-5273.

Nei, M, Kumar S. 2000. Molecular evolution and phylogenetics. Oxford University Press, New York.

Netprimer. Premier biosoft international. 2005. <www.premierbiosoft.com/netprimer/netprlaunch/netprlaunch.html>

Nilsson T. 1983. The Pleistocene: geology and life in the quaternary ice age. Boston: D. Reidel Publishing Company.

Orlando L, Bonjean D, Bocherens H, Thenot A, Argant A, Otte M, Hanni C. 2002. Ancient DNA and the population genetics of cave bears (*Ursus spelaeus*) through space and time. Molecular Biology and Evolution 19: 1920-1933.

Pääbo S. 1989. Ancient DNA: Extraction, characterization, molecular cloning, and enzymatic amplification. Proc. Natl. Acad. Sci. 86: 1939-1943.

Pääbo S, Poinar H, Serre D, Jaenicke-Després V, Hebler J, Rohland J, Kuch M, Krause J, Vigilant L , Hofreiter M. 2004. Genetic analysis from ancient DNA. Annu. Rev. Genet. 38: 645-679.

Page RDM, Holmes CE. 1998. Molecular evolution: A phylogenetic approach. Malden, MA: Blackwell Science Ltd.

Patterson C, Williams DC, Humphries CJ. 1993. Congruence between molecular and morphological phylogenies. Annu. Rev. Ecol. System 24: 153-188.

Pielou EC. 1991. After the ice age: the return of life to glaciated North America. Chicago and London: The University of Chicago Press.

Pisaric MFJ, MacDonald GM, Velichkob AA, Cwynar LC. 2001. The lateglacial and postglacial vegetation history of the northwestern limits of Beringia, based on pollen, stomate and tree stump evidence. Quaternary Science Reviews 20: 235-245.

Poinar HN, Schwarz C, Qi J, Shapiro B, MacPhee RDE, Buigues B, Tikhonov A, Huson D, Tomsho LP, Auch A, Rampp M, Miller W, Schuster SC. 2006. Metagenomics to paleogenomics: large-scale sequencing of mammoth DNA. Science 311: 392-394.

Ritz LR, Glowatzki-Mullis ML, MacHugh DE, Gaillard L. 2000. Phylogenetic analysis of the tribe Bovini using microsatellites. Animal Genetics 31: 178.

Rzhetsky A, Nei M. 1992. A simple method for estimating and testing minimum evolution trees. Molecular Biology and Evolution 9: 945-967.

Saitou N, Nei M. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Molecular Biology and Evolution 4: 406-425.

Sanger F, Nicklen S, Coulson AR. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74: 5463-5467.

Shapiro B, Drummond AJ, Rambaut A, Wilson MC, Matheus PE, Sher AV, Pybus OG, Gilbert MTP, Barnes I, Binladen J, Willerslev E, Hansen AJ, Baryshnikov GF, Burns JA, Davydov S, Driver JC, Froese DG, Harington CR, Keddie G, Kosintsev P, Kunz ML, Martin LD, Stephenson RO, Storer J, Tedford R, Zimov S, Cooper A. 2004. Rise and fall of Beringian steppe bison. Science 306: 1561-1565.

Soltis P, Soltis D, Smiley C. 1992. An *rbc*L sequence from a Miocene *Taxodium* (bald cypress). Proc. Natl. Acad. Sci. USA 89: 449-551.

Starikovskaya YB, Sukernik RI, Schurr TG, Kogelnik AM, Wallace DC. 1998. mtDNA Diversity in Chukchi and Siberian eskimos: Implications for the genetic history of ancient Beringia and the peopling of the new world. American Journal of Human Genetics 63: 1473-1491.

Stoffer, PW. 2003. Geology of Badlands National Park: A preliminary report: U.S. geological survey open-file report 03-35.

Sutcliffe AJ. 1985. On the track of ice age mammals. Cambridge: Harvard University Press.

Tajima F. 1983. Evolutionary relationships of DNA sequences in finite populations. Genetics 105: 437-460.

Tajima F. 1993. Measurement of DNA polymorphism. In mechanisms of molecular evolution. Introduction to molecular paleopopulation biology. N Takahata and AG Clark editors. Tokoyo, Sunderland, MA: Japan Scientific Societies Press.

Tamura K, Nei M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Molecular Biology and Evolution 10: 512-526.

Taylor KC, Lamorey GW, Doyle GA , Alley RB, Grootes PM, Mayewski PA, White J, Barlow LK. 1993. The 'flickering switch' of late Pleistocene climate change. Nature 361: 432-436.

Thomas RH, Schaffner W, Wilson AC, Pääbo S. 1989. DNA phylogeny of the extinct marsupial wolf. Nature 340: 465-467.

Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTALW version 1.83: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22: 4673–4680.

Thompson JD, Higgins DG, Gibson TJ. 2003. Online help for Clustal W version 1.83. Avail from: http://align.genome.jp/clustalw/clustalw_help.html.

Thornton JW, DeSalle R. 2000. Gene family evolution and homology: genomics meets phylogenetics. Annual Review of Genomics and Human Genetics 1: 41-73.

Underhill PA, Jin L, Zemans R, Oefner PJ, Cavalli-Sforza LL. 1996. A pre-Columbian Y chromosome-specific transition and its implications for human evolutionary history. Genetics 93: 196-200.

Verkaar ELC, Nijman IJ, Beeke M, Hanekamp E, Lenstra JA. 2004. Maternal and paternal lineages in cross-breeding bovine species. Has wisent a hybrid origin? Mol. Bio. and Evol. 21: 1165-1170.

Wakely J. 1994. Substitution-rate variation among sites and the estimation of transition bias. Mol. Biol. Evol. 11: 436-442.

Ward TJ, Bielawski JP, Davis SK, Templeton JW, Derr JN. 1999. Identification of domestic cattle hybrids in wild cattle and bison species: A general approach using mtDNA markers and the parametric bootstrap. Animal Conservation 2: 51-57.

Wayne RK, Leonard JA, Cooper A. 1999. Full of sound and fury: The recent history of ancient DNA. Annual Review of Ecology and Systematics 30: 457-477.

Willerslev E, Hansen AJ, Binladen J, Brand TB, Gilbert MTP, Shapiro B, Bunce M, Wiuf C, Gilichinske DA, Cooper A. 2003. Diverse plant and animal genetic records from Holocene and Pleistocene sediments. Science 300: 791-795.

Willerslev E, Cooper A. 2005. Ancient DNA. Proceedings of the Royal Society B 272: 3-16.

Wilson GA, and Strobeck C. 1999. Genetic variation within and relatedness among wood and plains bison populations. Genome 42: 483-496.

Wilton SD, Lim L, Dye D, Laing N. 1997. Bandstab: A PCR-based alternative to cloning PCR products. BioTechniques 22: 642-645.

Yang H, Golenberg EM, Shoshani J. 1996. Phylogenetic resolution within the *Elephantidae* using fossil DNA sequence from the American mastodon (*Mammut americanum*) as an outgroup. Proc. Natl. Acad. Sci. USA 93: 1190-1194.