Kreutzer, McKensie, Allele Characterization of Ten Short Tandem Repeat Loci of North American Bears (Ursids) Using Next-Generation Sequencing. Master of Science (Forensic Genetics), May, 2014, 58 pp., 6 tables, 13 figures, 47 references.

An identification method that provides higher genetic resolution than capillary electrophoresis (CE) is needed for isolated bear populations that possess low genetic diversity. Amplification conditions were optimized for ten bear STR loci. Amplicons were used to develop a library for next-generation sequencing (NGS) on the Ion Torrent™ PGM™ Sequencer. Through ligation of DNA barcode adaptors, seven black bear (*Ursus americanus*) samples were sequenced together. Sequencing reads were aligned to a virtual ladder and analyzed in NextGENe® software. Allele concordance was shown between CE and NGS. Variants within alleles (SNPs and INDELs) showed that NGS provided higher genetic resolution. These results have implications for improving individual identification and population assignment in wildlife forensics and conservation for populations with low genetic diversity.

ALLELE CHARACTERIZATION OF TEN SHORT TANDEM REPEAT LOCI OF NORTH AMERICAN BEARS (URSIDS) USING NEXT-GENERATION SEQUENCING

THESIS

Presented to the Graduate Council of the

Graduate School of Biomedical Sciences

University of North Texas

Health Science Center at Fort Worth

In Partial Fulfillment of the Requirements

For the Degree of

MASTER OF SCIENCE

By

McKensie Kreutzer, B.S.

Fort Worth, Texas

May, 2014

ACKNOWLEDGEMENTS

I would first like to thank the genetics unit from the US Fish and Wildlife Service Forensic Laboratory for the opportunity to conduct this research for them. Special thanks go to Dr. Mary Curtis for graciously giving her time to help me throughout the process. I would next like to thank my major professor, Dr. John Planz, for his guidance and support throughout this project, and for always challenging me to think about things critically. I would like to thank Dr. Michael Allen for input and guidance on project design. I would like to acknowledge David Visi and Shantanu Shewale for providing assistance throughout the sequencing process. I would especially like to thank my lab mate, Taylor Howard, for going through this entire project alongside me and providing me with countless hours of entertainment. Finally, I would like to acknowledge my husband, Andy, and my parents, Ron and Maureen Smith, for extending their unwavering encouragement and support throughout my entire Master's program.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Bears are mammals that belong to the Family Ursidae. North American populations are represented in the Genus *Ursus*, and are divided into three species: brown bears (*Ursus arctos*), black bears (*Ursus americanus*), and polar bears (*Ursus maritimus*). These species are recognized by discrete morphological differences, such as body size and coat color, as well as ecological differences in the various habitats and ranges each occupy. Genetic differences, such as varying allele frequencies at short tandem repeat (STR) loci, also play an important role in separating these species (Waits *et al.* 1999) and characterizing geographic subpopulations within the same species (Paetkau *et al.* 1995; Waits *et al.* 1999; Waits *et al.* 2000).

Throughout the Holarctic, several subpopulations of brown bears with different subspecific taxonomic designations inhabit various ranges (Taberlet *et al.* 1997). Subspecies include Eurasian brown bears (*Ursus arctos arctos*), Alaskan brown bears (*U. a. alascensis*), and Kamchatka brown bears (*U. a. beringianus*), among others. Brown bears in inland regions of North America are more commonly referred to as grizzly bears (*U. a. horribilis*) (Miller & Waits 2003). South of the Canadian border, grizzly bears exist in less than one percent of their historic range (Paetkau *et al.* 1998), maintaining populations only in the Yellowstone and Northern Continental Divide ecosystems (Miller & Waits 2003). The Archipelago Islands are home to another brown bear subspecies, Kodiak bears (*U. a. middendorffi*), that diverged from continental populations approximately 10,000 years ago (Paetkau *et al.* 1998) and are known to

be the largest land-based predator in the world. There are an estimated 50,000 to 65,000 brown bears dispersed throughout North America (Figure 1). However, a few subpopulations, such as Kodiak bears and Yellowstone bears, are physically isolated from other subpopulations and consequently experience limited gene flow with increased inbreeding (Waits *et al.* 2000) leading to decreased genetic diversity (Paetkau *et al.* 1998; Miller & Waits 2003). Biologists maintain that conservation efforts are essential to preserve these isolated populations, especially on the genetic level (Skrbinsek *et al.* 2012).

Figure 1: Current range map for brown bears in North America (Jonkel 2005)

Black bears are the most widely distributed species of bear on the North American continent (Figure 2) and inhabit a wide range of coastal and mountainous regions (Brown *et al.* 2009). There are an estimated 735,000 to 941,000 individuals (Williamson 2002) and as such, are classified as least concern on the International Union for Conservation of Nature and Natural

Resources (IUCN) Red List (IUCN 2013). As with brown bears, several subspecies exist across various habitats, with a few regarded as isolated and/or threatened. British Columbia is home to one such rare black bear subspecies, Kermode bears (*Ursus americanus kermodei*), that are all white due to a recessive allele in a pigmentation gene (Hedrick & Ritland 2012); only 100 to 500 individuals are estimated to remain. Florida black bear (*U. a. floridanus*) populations are heavily fragmented and listed as threatened by the Florida Game and Freshwater Fish Commission (Dixon *et al.* 2007). Similarly, Louisiana black bears (*U. a. luteolus*) are considered threatened in east Texas, southern Louisiana, and southwestern Mississippi (IUCN 2013), where population declines are primarily attributed to human exploitation and habitat loss due to agriculture (Lowe 2011). Subpopulations of black bears in California also experience isolation due to geographic barriers and distance (Brown *et al.* 2009). Physical isolation of subpopulations decreases gene flow and has adverse effects on genetic diversity (Waits *et al.* 2000); therefore, conservation and management plans that conserve genetic diversity are important to maintain rare, threatened, and fragmented populations (Dixon *et al.* 2007).

Unfortunately, subpopulation isolation due to habitat fragmentation is not the only threat to black bears. Poachers hunt for meat, paws, hides, claws, teeth, skulls, and gallbladders to fulfill demands from various markets (Williamson 2002). Asiatic black bear (*Ursus thibetanus*) gallbladders are commonly used in Traditional Chinese Medicines (TCMs), but with diminishing Asiatic black bear populations, demands must be met from elsewhere. It is becoming increasingly common for poachers to target North American species as substitutes for declining Old World species in the TCM trade. Interestingly, DNA analysis has shown that American black bears are the closest relation to Asiatic black bears (Krause *et al.* 2008), making them prime targets for poachers (Williamson 2002).

Figure 2: Current range map for black bears in North America (Manitoba Conservation and Water Stewardship 2013)

Polar bears are largely philopatric and are geographically distributed with limited overlap throughout the circumpolar Arctic (Figure 3) (Paetkau *et al.* 1995). Polar bears are commonly considered to be a marine mammal (Gunderson 2009), as most of their lives are spent on ice-covered water (Paetkau *et al.* 1999). Biologists estimate that only 20,000 to 25,000 individuals remain worldwide. Reports state that eight of nineteen recognized subpopulations are in decline due to various factors that include habitat loss (IUCN Polar Bear Specialist Group 2009) and reduced access to available prey, as well as reduced prey abundance (Paetkau *et al.* 1999). To compound the problem of population decline, polar bear populations have characteristically low genetic diversity. Due to these concerns, polar bears are currently classified as vulnerable on the IUCN Red List (IUCN 2013) and threatened on the United States Fish and Wildlife Service

Endangered Species List (U.S. Fish and Wildlife Service 2013). Therefore, conservation and management plans which conserve genetic diversity are critical to maintain polar bear existence.

Figure 3: Current range map for polar bears in North America (Manitoba Conservation and Water Stewardship 2013)

As previously discussed, all bear populations in North America, regardless of species or subspecies, are at risk of population fragmentation and human-caused mortality. Population fragmentation can be the direct result of habitat loss and/or destruction (Kocijan *et al.* 2011) due to human population growth and expansion into bear territory in the form of anthropogenic barriers such as cities and highways (Williamson 2002; Karamanlidis *et al.* 2012). Natural barriers can also play a role in further isolating populations (Paetkau *et al.* 1998) if those barriers lead to decreased gene flow (Bull *et al.* 2011) and the creation of distinct subpopulations (Dixon *et al.* 2007; Brown *et al.* 2009). Often, subpopulations become physically isolated (Taberlet *et al.*

1997; Paetkau *et al.* 1998b; Miller & Waits 2003), which leads to increased inbreeding and decreased heterozygosity and allelic diversity (Taberlet *et al.* 1997; Paetkau *et al.* 1998). Studies have shown that heterozygosity is correlated with population fitness (Lacy 1997; Reed & Frankham 2003; Dlugosch & Parker 2008) and that decreased effective population size and inbreeding depression reduce population viability (Lacy 1997). These factors lead to diminished adaptive ability in changing environments, thereby increasing the extinction risk of the subpopulation (Ohnishi *et al.* 2007). Population fragmentation is further compounded by human-caused mortality, such as poaching for bear parts markets (Williamson 2002). Genetic bottlenecks that arise from severe population decline can produce the same decrease in genetic diversity that is caused by inbreeding (Karamanlidis *et al.* 2012). Isolation and inbreeding can be natural events that occur over long periods of time and tend to equilibrate themselves (Paetkau *et al.* 1998); however, population fragmentation and human-caused mortality can occur more rapidly and pose a more immediate threat to bears (Miller & Waits 2003). The IUCN recognizes genetic diversity as one of three biodiversity measures deserving conservation (Reed & Frankham 2003; IUCN 2013); henceforth, management plans need to incorporate strategies to preserve genetic diversity in order to protect bear populations.

The field of conservation genetics utilizes techniques and analytical methods to assist conservation managers in combating issues of population fragmentation and decline. Genetic monitoring is used by managers to evaluate the genetic diversity of a subpopulation and determine if intervention is needed to expand the gene pool (De Barba *et al.* 2010). Population genetic monitoring is critical to determine if the genetic diversity of the subpopulation has been positively altered to increase adaptive ability and improve chances of long term subpopulation

persistence. Paternity analysis via genetic testing can also be performed to estimate male reproductive success (Craighead *et al.* 1995) in monitored populations.

The same analyses that are used by conservation managers are used by wildlife forensic analysts in wildlife crime cases to determine species and individual identification (Andreassen *et al.* 2012), parentage analysis, and assignment of a sample to a geographic origin (Ogden 2011). Various methods of genetic testing have been used in the past (e.g. allozymes and mitochondrial DNA) that did not provide enough variation to resolve individuals in populations of species with low genetic variability (Paetkau & Strobeck 1994). The discovery of DNA fingerprints, known as variable number tandem repeats (VNTRs) or minisatellites (Jeffreys *et al.* 1985), occurred in the 1980s and led to many successful prosecutions in human and wildlife crime investigations (Ogden 2011). Soon after, DNA profiling transitioned to short tandem repeats (STRs), otherwise called microsatellites. STR analysis investigates variation in very short repetitive sequence motifs found throughout the non-coding regions of the genome (Paetkau & Strobeck 1994). Currently, STR loci are amplified using polymerase chain reaction (PCR) and then sorted by size via gel or capillary electrophoresis to generate a unique multilocus genotype, referred to as a profile. Bear DNA profiling commonly utilizes a suite of ten dinucleotide STR loci that vary in length from approximately 80 to 280 base pairs (Table 1) (Waits *et al.* 2000; Skrbinsek *et al.* 2012) and were originally characterized in the 1990s (Paetkau & Strobeck 1994; Paetkau *et al.* 1995; Taberlet *et al.* 1997). These STR markers provide high levels of genetic variation with mean expected heterozygosity of greater than 60% (Paetkau *et al.* 1995; Bull *et al.* 2011; Skrbinsek *et al.* 2012) and average allelic richness of 3.1-9.2 (Brown *et al.* 2009) in most bear populations. However, their utility is reduced with lower levels of diversity found in isolated populations with limited gene flow (Paetkau *et al.* 1998b; Dixon *et al.* 2007; Brown *et al.* 2009).

Therefore, to improve the ability to perform individual identification and geographic assignment, analytical methods that provide better genetic resolution are needed.

Table 1: List of ten bear loci commonly used. Repeat motif is based on the GenBank® reference sequence. Base pair range is from previously reported values (Paetkau *et al.* 1998; Waits *et al.* 2000; Skrbinsek *et al.* 2012).

Locus	Reference Repeat Motif	Reported Base Pair Range
G1A	$(TG)_{19.5}$	180-202
G1D	$(GT)_{17.5}$	168-190
G10B	$(TG)_{21}$	130-166
G10C	$(TG)_{21.5}$	93-117
G10H	$(TG)_{25.5}(ATTT)_{6.2}$	244-279
G10J	$(TG)_{19.5}$	78-110
G10L	$(TG)T(TG)_{33.5}$	135-171
G10M	$(GT)_{21}$	196-222
G10P	$(GT)_{21}TT(GT)_{6}$	139-175
G10X	$(TG)_{20.5}TTC(TG)_{7}$	125-163

The United States Fish and Wildlife Service National Fish and Wildlife Forensics Laboratory (USFWS-NFWFL) is the only federal forensic laboratory devoted to analyzing evidence in wildlife crime cases (U.S. Fish and Wildlife Service 2011). The laboratory consists of several units including Criminalistics, Chemistry, Morphology, Pathology, and Genetics. Between 400 and 800 cases are processed annually, of which the Genetics unit receives between 80 and 100 (Curtis 2013). Approximately ten of these cases are bear-related. Each case typically contains one to three items of evidence, although it is not uncommon to receive as many as twenty-seven items. Most bear cases deal with American black bear; however, in the last five years, the laboratory has also processed three grizzly bear cases, one polar bear case, and one Asiatic black bear case. Since 2008, the laboratory has processed nine individual matching cases and nine species identifications on evidence such as tissue, meat, bone, blood, claws, and pelts.

Cases typically deal with situations such as out-of-season hunting and illegal poaching. In addition, the laboratory has processed twenty-seven gallbladder identifications in cases dealing with TCMs that are traded on the black market. TCMs utilize the bile and/or bile salts from bear gallbladders for medicinal purposes, but legality issues arise when the gallbladders come from protected bear populations as regulated by the Convention on International Trade in Endangered Species (CITES) (Coghlan *et al.* 2012). To combat black market trade, law enforcement officers borrow gallbladders from the USFWS-NFWFL reference collection to use in covert operations to catch suspects (Curtis 2013). Given the immensity of poaching and illegal trade, a genetic testing method is needed that can provide better accuracy of assigning a bear to a subspecies and geographic location to improve evidentiary weight in court.

High resolution sequencing methods, such as next-generation sequencing technologies, allow internal variants such as single nucleotide polymorphisms (SNPs) and insertions and deletions (INDELs) within STR regions to be detected. Thus, next-generation sequencing may provide better genetic resolution for individual identification and determining geographic provenance (Guichoux *et al.* 2011). This is especially important in wildlife forensics for identifications from populations exhibiting decreased genetic variation or for compromised evidence samples that conventional STR typing cannot resolve. The Ion Torrent™ Personal Genome Machine™ (PGM™) Sequencer by Life Technologies™ is a next-generation technology that utilizes semi-conductor sequencing. Sequencing takes place on a proprietary complementary metal oxide semiconductor (CMOS) chip overlaid with a microfluidic system containing millions of micro-scale wells (Life Technologies 2011). DNA fragments attach to and replicate on small beads that are then deposited into the wells. The chip is sequentially flooded with a different nucleotide every fifteen seconds. As the nucleotides flow across the chip, some

are incorporated into the strand. Hydrogen protons are released that change the pH of the solution. This change in pH is measured and converted to voltage by an ion sensitive layer beneath each well. Voltage is recorded, signaling the software to make a base call. Ultimately, all high resolution sequencing methods have been shown to provide near-perfect consensus accuracy (Life Technologies 2011; Liu *et al.* 2012; Quail *et al.* 2012).

In this study, it was hypothesized that when compared to capillary electrophoresis, next-generation sequencing of STR loci would attain better genetic resolution among bear populations, and would therefore improve the accuracy of individual identification and assignment of an individual to its population of origin. The main objective of this study was to use an existing capillary electrophoresis assay to develop a targeted next-generation sequencing library, and then show allele call concordance between the two methods. A secondary objective of this study was to investigate internal variants, such as SNPs and INDELs, within the loci.

CHAPTER II

MATERIALS AND METHODS

Specimens Examined:

Tissue samples from six American black bears (Table 2) were donated from the USFWS-NFWFL reference collection in Ashland, OR, and were extracted using an organic phenol-chloroform-isoamyl alcohol protocol (Appendix A) with Microcon® YM-100 (EMD Millipore® Corp., Billerica, MA) concentration in order to retrieve quality DNA for downstream analysis. One previously extracted sample (Table 2) was retrieved from the University of North Texas Health Science Center (UNTHSC) Evolutionary Genetics Laboratory. Extracts were quantified using a Qubit® 2.0 Fluorometer (Life Technologies™, Carlsbad, CA) (Appendix B) to determine the concentration of DNA present in the extracts. Normalization dilutions (5 ng/µL) were then performed to prepare for singleplex PCR.

Table 2: American black bear tissue samples used in this study. Samples were obtained from [a]USFWS-NFWFL and [b]UNTHSC Evolutionary Genetics Laboratory.

Sample Number	State
H30863[a]	AK
G30314[a]	CO
H30431[a]	FL
D21381[a]	CA
G41360[a]	PA
G40770[a]	ME
Bear JB 8/9[b]	NY

Primer Preparation:

Existing primers were analyzed for melting temperature (T_m), hairpin structure, and primer-dimer formation using AutoDimer (STRbase). A standard nucleotide Basic Local Alignment Search Tool (BLAST) (Altschul *et al.* 1990), available from the National Center for Biotechnology Information (NCBI), was used to perform primer redesign as required. Primers (Table 3) without fluorophores were ordered from Life Technologies™. Primers were reconstituted using TE^{-4} buffer (10mM Tris, 0.1mM EDTA, pH 8.0) to 300µM and quantified on the NanoDrop 2000 (Thermo Fisher Scientific Inc., Waltham, MA) to verify stock concentrations. Primers were each diluted to a 10µM working concentration.

Table 3: PCR primers for the ten bear loci used in this study. Previously reported primers are from [a](Paetkau *et al*. 1995) and [b](Taberlet *et al.* 1997). *Denotes primers that were modified from those published. Primers for G1D were newly designed for this study.

Locus	Forward $(GT)_n$ strand primer	Reverse $(CA)_n$ strand primer
G1A[a]	GACCCTGCATACTCTCCTCTGATG	GCACTGTCCTTGCGTAGAAGTGAC
G1D	GGAAAAGTTTTGGACATTTTCTTC	CAAGAACACTTGATACCTAGCAC
G10B[a]	*GTTCTGTTGAATTTGGTTTGC	*GCAGGACAAATCACAGAAAC
G10C[a]	AAAGCAGAAGGCCTTGATTTCCTG	GGGGACATAAACACCGAGACAGC
G10H[b]	CAACAAGAAGACCACTGTAA	AGAGACCACCAAGTAGGATA
G10J[b]	*GATCAGATATTTTCAGCTTTTGTGTGTG	*ATAACCCCTCACACTCCACTTC
G10L[a]	GTACTGATTTAATTCACATTTCCC	*CCTACCCATGCGATAAAATTG
G10M[a]	*CCTCATCGTAGGTTGTATTTTCCTCC	*CCAAATAATTTAAATGCATCCCAGGG
G10P[a]	AGGAGGAAGAAAGATGGAAAAC	TCATGTGGGGAAATACTCTGAA
G10X[a]	CCCTGGTAACCACAAATCTCT	*CTCAGTTATCTGTGAAATCAAAAC

Singleplex Reactions:

Singleplex amplification conditions for each locus were optimized (Appendix C). All singleplex reactions had a 25 µL volume and were carried out on one sample designated as a reference (H30431). For G1A, G10C, G10J, and G10M, the reaction contained 2.5 µL of 10X

PCR Buffer II, 3.0 µL of 25mM $MgCl_2$, 1.0 µL of 10mM dNTP mix, 2.5 µL of Bovine Serum Albumin (1.6 µg/µL), 0.5 µL of AmpliTaq Gold® (5 U/µL)(Life Technologies™), 2.5 µL of 5M Betaine (Sigma-Aldrich®, St. Louis, MO), 1.2 µL of 10µM forward primer, 1.2 µL of 10µM reverse primer, 2.0 µL of template DNA (5 ng/µL), and 8.6 µL of Ultrapure distilled water. For G1D, G10B, G10H, G10L, G10P, and G10X, the reaction contained 2.5 µL of 10X PCR Buffer II, 2.0 µL of 25mM $MgCl_2$, 1.0 µL of 10mM dNTP mix, 2.5 µL of Bovine Serum Albumin (1.6 µg/µL), 0.5 µL of AmpliTaq Gold® (5 U/µL)(Life Technologies™), 2.5 µL of 5M Betaine (Sigma-Aldrich®), 1.2 µL of 10µM forward primer, 1.2 µL of 10µM reverse primer, 2.0 µL of template DNA (5 ng/µL), and 9.6 µL of Ultrapure distilled water. Thermal cycling was carried out on the Eppendorf MasterCycler® pro S (Eppendorf International, Hamburg, Germany). Conditions for G1A, G10C, G10J, and G10M were 95°C for 11 minutes, 35 cycles of 95°C for 10 seconds, 60°C for 30 seconds, and 72°C for 30 seconds, and finally, 70°C for 10 minutes. Conditions for G1D, G10B, G10H, G10L, G10P, and G10X were 95°C for 11 minutes, 35 cycles of 95°C for 10 seconds, 52°C for 30 seconds, and 72°C for 1 minute, and finally, 70°C for 10 minutes. Locus amplification was confirmed by analyzing each reaction with the Agilent DNA 1000 Kit (Agilent Technologies Inc., Santa Clara, CA) on the Agilent 2100 Bioanalyzer (Agilent Technologies Inc.) (Appendix D) for amplicon length and concentration.

Following singleplex optimization with sample H30431, singleplex reactions were carried out on a second sample, G40770, to confirm optimization. All other samples were amplified. The ten singleplex reactions for each sample were combined and purified using the QIAquick® PCR Purification Kit (Qiagen® Inc., Valencia, CA) (Appendix E). Following purification, product quality and concentration were verified with the Agilent DNA 1000 Kit (Agilent Technologies Inc.) on the Agilent 2100 Bioanalyzer (Agilent Technologies Inc.).

Library Preparation and Sequencing:

The purified amplicon pools from the seven samples were used to prepare the library for next-generation sequencing on the Ion Torrent™ Personal Genome Machine™ (PGM™) Sequencer (Life Technologies™). Library preparation was carried out using the NEBNext® Fast DNA Library Prep Set for Ion Torrent™ (New England BioLabs® Inc., Ipswich, MA) (Appendix F). First, end repair was performed to convert the DNA fragments to blunt-ended DNA and phosphorylate the 5' ends to enable subsequent adaptor ligation. Following end repair, NEXTflex™ DNA Barcodes (Bioo Scientific™ Corp., Austin, TX) were ligated. Barcode adaptors are unique sequences of approximately 50 base pairs that allow multiple samples to be pooled together and sequenced in one run. Adaptor ligated fragments were purified using Microcon® YM-100 (EMD Millipore® Corp.) concentrators.

Each amplicon library was quantified using the Agilent High Sensitivity DNA Kit (Agilent Technologies Inc.) on the Agilent 2100 Bioanalyzer (Agilent Technologies Inc.) to confirm adaptor ligation. Amplicon libraries underwent an additional amplification to enrich the quantity of adaptor ligated amplicons in the library. Following enrichment, amplicon libraries were purified with Agencourt® AMPure® XP Beads (Beckman Coulter, Indianapolis, IN) to remove excess adaptors and primers. Amplicon libraries were quantified on the Qubit® 2.0 Fluorometer (Life Technologies™) and dilutions were made to 26pM per manufacturer's recommendation. All libraries were pooled together and underwent emulsion PCR on the Ion OneTouch™ System (Life Technologies™) using the Ion PGM™ Template OT2 400 Kit (Life Technologies™) (Appendix G). Following emulsion PCR, the library was quantified on the Qubit® 2.0 Fluorometer (Life Technologies™) using the Ion Sphere™ Quality Control Kit (Life Technologies™), to detect the percent of Ion Sphere™ Particles (ISPs) bound to DNA. The

library was enriched on the Ion OneTouch™ ES (Life Technologies™) and was injected onto the Ion PGM™ 316™ Chip v2 (Life Technologies™) for sequencing (Appendix H).

Data Analysis:

Sequencing reads were aligned to a virtual ladder that was created from limited sequence data reported in GenBank® (Table 4), a repository available from NCBI. To develop the virtual ladder, the largest allele previously reported for each locus was concatenated into one FASTA file that incorporated forward and reverse sequences. BAM data files of all PGM™ sequence reads for each barcoded library were converted into FASTA files and uploaded into NextGENe® (SoftGenetics®, State College, PA) for sequence analysis. Rigorous alignment was conducted to align reads to their respective locus via primer sequence and allow gaps in the reads when the repeat sequence was shorter than the reference. Rigorous alignment uses an algorithm to determine the read alignment with the least amount of mismatches to the reference. Allele calls were made for each locus to determine a STR profile for each individual. This was done by counting the number of reads that possessed the same number of repeats at a locus. The repeat number associated with the greatest number of reads was called as the true allele. Repeat numbers with lesser numbers of reads were called as stutter. For heterozygotes, two repeat lengths shared approximately equal numbers of reads, while all other repeat lengths at that locus had a smaller number of reads and were called as stutter. A consensus sequence was developed to characterize each locus by analyzing and recording single nucleotide polymorphisms (SNPs) and insertions and deletions (INDELs), including the STR repeat, in hopes of obtaining better genetic resolution.

The sequencing data was compared to previous fragment sizing data from the capillary electrophoresis assay used by the USFWS-NFWFL. Concordance between assays was evaluated. The level of resolution between assays was also evaluated to determine if next-generation sequencing provided better genetic resolution and thus, more accurate identification and assignment of an individual to its population of origin.

Table 4: Reference information from GenBank® for each locus. Sequence data was obtained by cloning isolates from *Ursus americanus* (black bear, ear tag #203) in Banff National Park, Canada. Sequences were all entered in April, 1995.

Locus	Accession
G1A	U22095
G1D	U22094
G10B	U22084
G10C	U22085
G10H	U22086
G10J	U22087
G10L	U22088
G10M	U22089
G10P	U22091
G10X	U22093

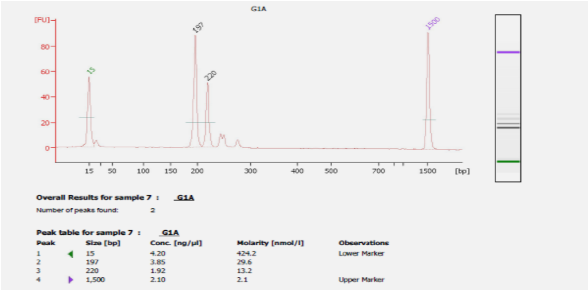
CHAPTER III

RESULTS

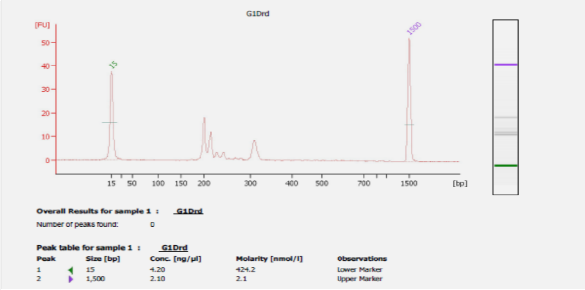
Singleplex Reactions:

Singlepex amplification conditions were successfully optimized. Electropherograms generated of the ten singleplex reactions for samples H30431 (Figure 4) and G40770 (Appendix I) from the Agilent 2100 Bioanalyzer (Agilent Technologies Inc.), confirmed that each primer pair amplified its respective targeted STR locus. Stutter was minimal in the electropherograms, and was generally contained within a single peak that was larger than the target peak. Non-specific products were greatly reduced or eliminated by optimizing the amounts of the reaction components, annealing temperature, and number of cycles in PCR, and by performing primer redesign. A key addition to the reaction mix was Betaine, which reduces base composition dependence of DNA strand melting during PCR. Electropherograms generated of the pooled singleplexes for all samples (Figure 5) confirmed amplicon presence, except for sample G40770, which did not contain product. However, since all singleplex reactions for G40770 contained product, the combined sample was assumed to have product. Absence of DNA in the electropherogram was likely due to pipetting error or a blocked well on the Agilent DNA 1000 Kit chip.

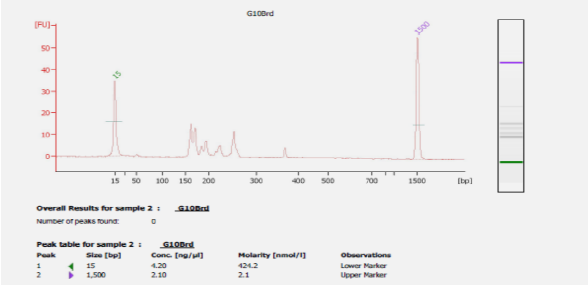
H30431/G1A



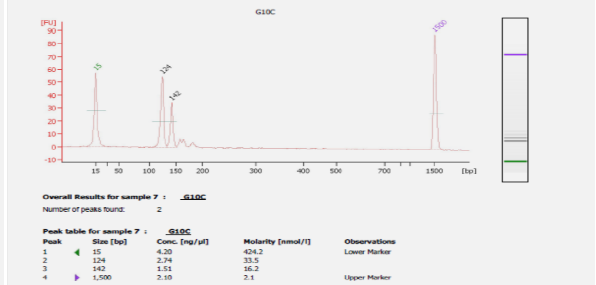
H30431/G1D



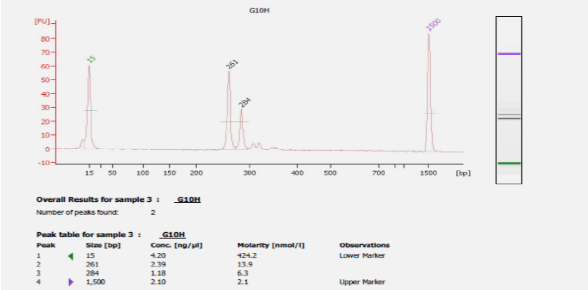
H30431/G10B



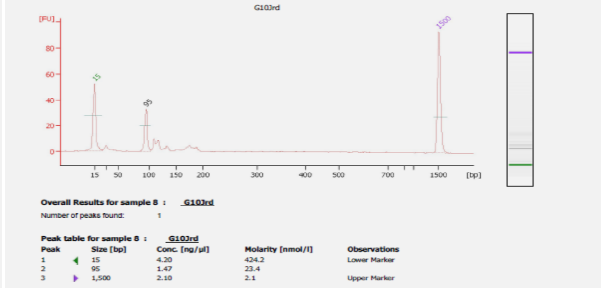
H30431/G10C



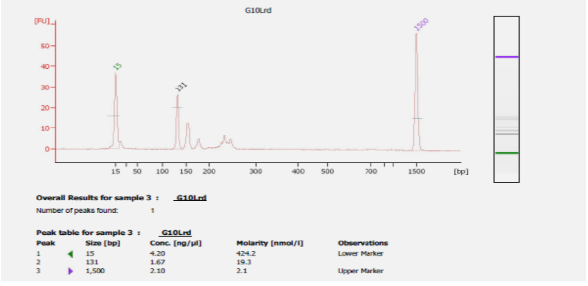
H30431/G10H



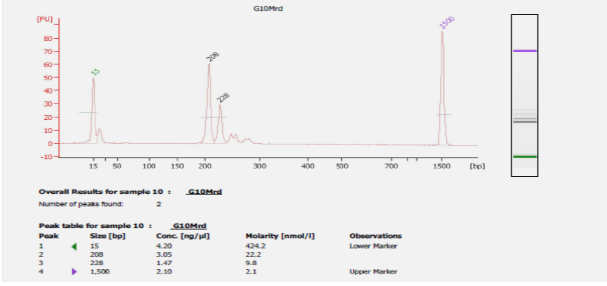
H30431/G10J



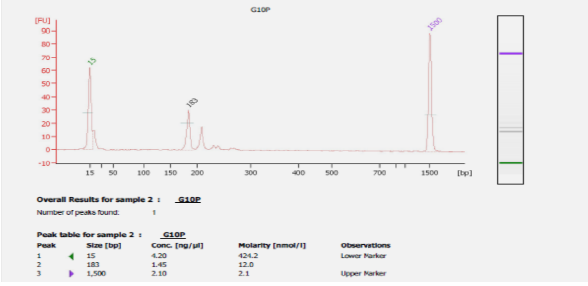
H30431/G10L



H30431/G10M



H30431/G10P



H30431/G10X

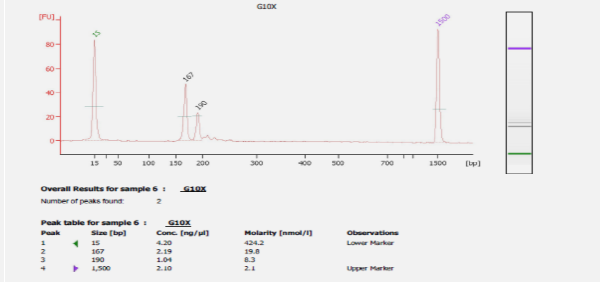


Figure 4: Singleplex reaction electropherograms of sample H30431, generated using the Agilent DNA 1000 Kit on the Agilent 2100 Bioanalyzer (Agilent Technologies Inc.).

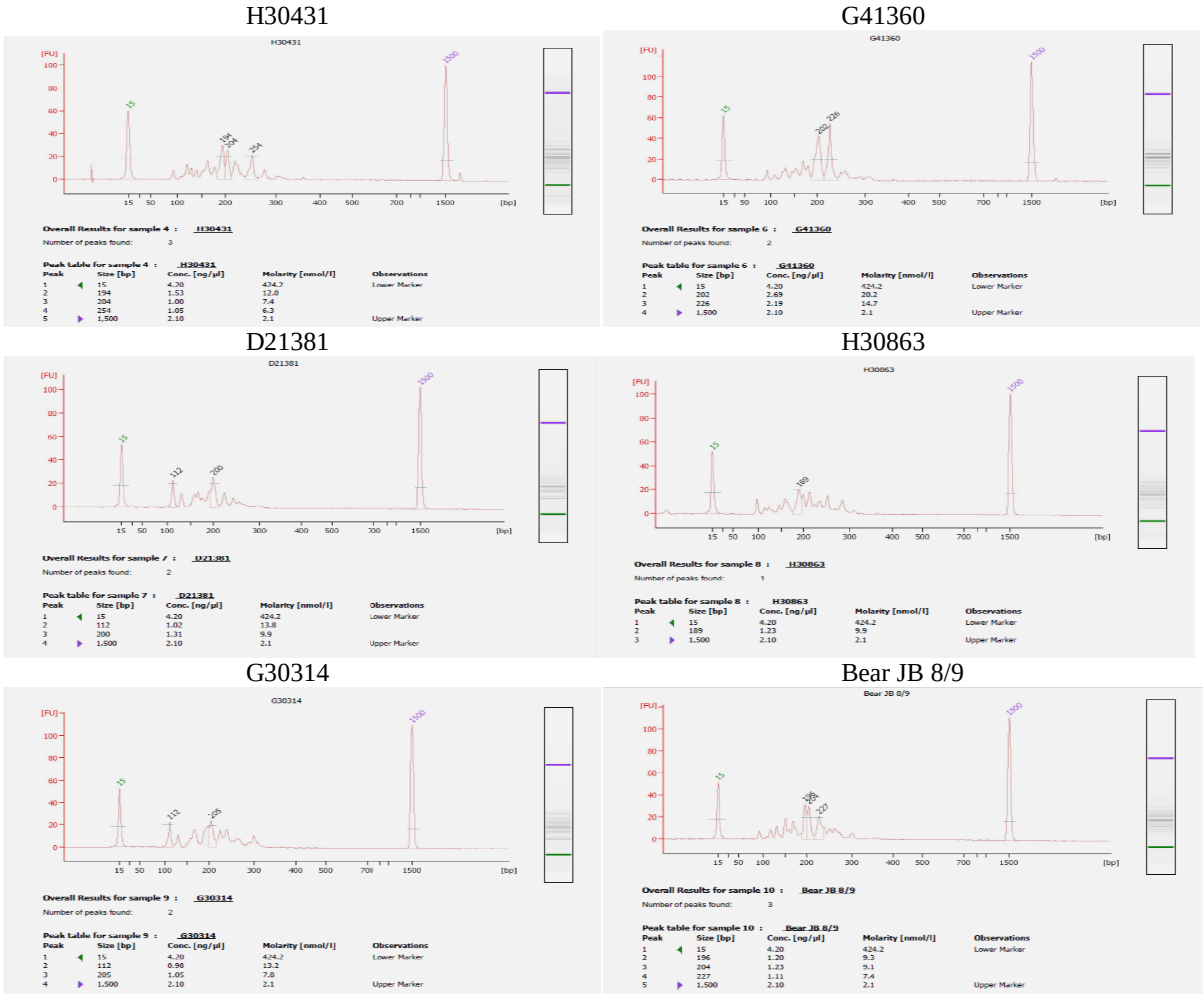


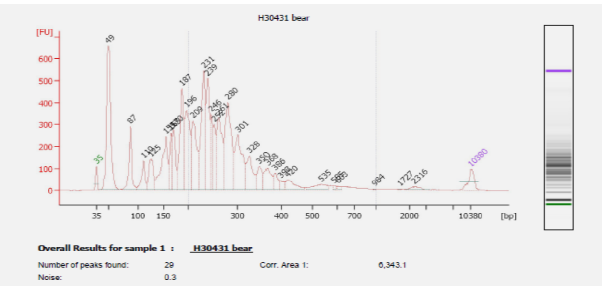
Figure 5: Electropherograms of samples after singleplex pooling, generated using the Agilent DNA 1000 Kit on the Agilent 2100 Bioanalyzer (Agilent Technologies Inc.). G40770 not shown.

Library Preparation:

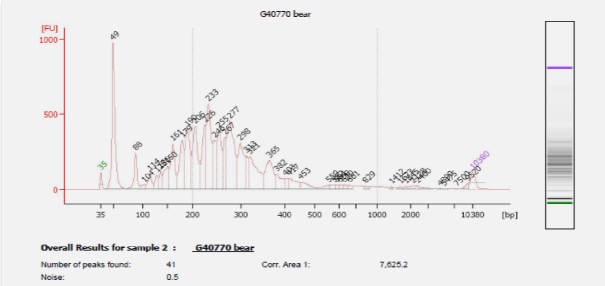
Library preparation was carried out on the purified amplicon pools. Following end repair, adaptor ligation, and purification, each sample was assessed. Electropherograms from the Agilent 2100 Bioanalyzer (Agilent Technologies Inc.) indicated that all samples experienced successful ligation (Figure 6), based on the median base pair size being approximately 40 bp larger post-ligation. Samples underwent additional amplification and purification to maximize adaptor ligated DNA and minimize excess primers and adaptors. Following emulsion PCR, the

library was quantified and had 75% of ISPs bound to DNA. The ideal range is 10-30% to ensure polyclonal ISPs are minimal, and although much higher than the ideal amount, it was decided that enrichment and sequencing proceed as planned on the library.

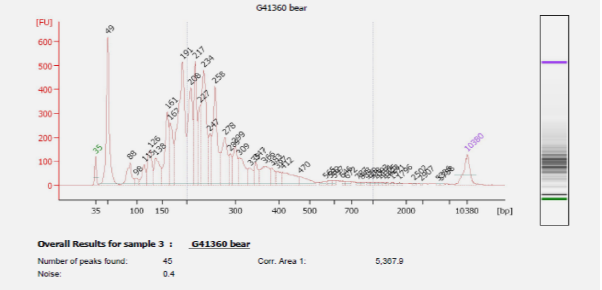
H30431



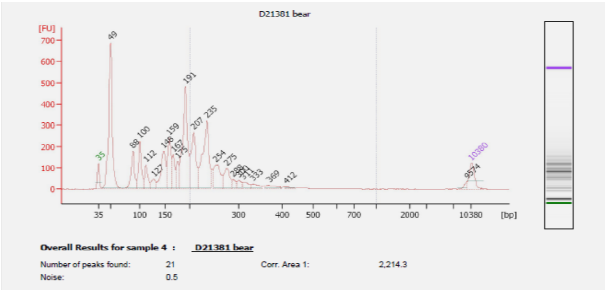
G40770



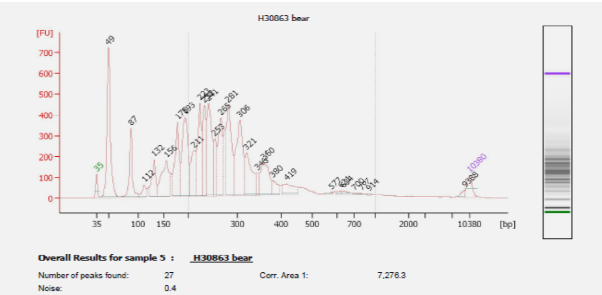
G41360



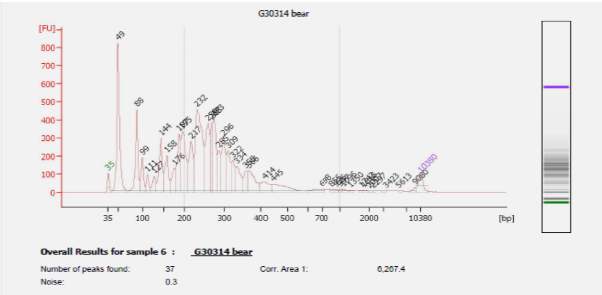
D21381



H30863



G30314



Bear JB 8/9

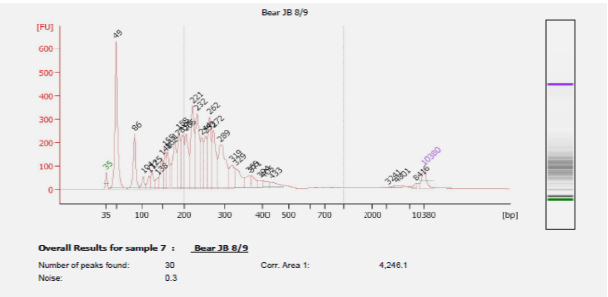


Figure 6: Electropherograms of samples after library preparation, generated using the Agilent High Sensitivity DNA Kit on the Agilent 2100 Bioanalyzer (Agilent Technologies Inc.).

Sequencing and Data Analysis:

The sequencing run was analyzed following completion. The chip had 66% loading efficiency, with 87% enrichment efficiency from the Ion OneTouch™ ES (Life Technologies™). Clonal beads made up 82% of beads present, while polyclonal beads made up only 18% - much less than expected based on the pre-enrichment quantification value. The final library was 47% of the loaded sample and contained 38% usable reads. The BAM data files were uploaded into the NextGENe® software to be analyzed. Table 5 lists the total reads, aligned reads, mean read length, and coverage level per locus for each sample. More than 94% of the reference sequence obtained 5X coverage or greater in all samples. Combined, the library generated 342,745 total reads, of which 218,683 were aligned to the reference. It was discovered that the sequencing chemistry struggled with repetitive motifs, as a large number of reads failed part of the way through the repeat region. Due to this, the average read length captured by the software for each locus of each sample was shorter than the anticipated average length of approximately 150 base pairs. Each sample had an average coverage per locus of at least 888X, and at least 15% of reads extended across the entire locus. Therefore, an adequate number of complete reads were available for analysis (Figure 7).

Table 5: Sequencing data obtained for each sample from the NextGENe® software.

Sample	Total Reads	Aligned Reads	Mean Read Length	Average Coverage	Percent with ≥5X Coverage
H30431	39,211	24,302	121	888	98.27
G40770	42,873	27,387	120	968	100
G41360	54,692	34,687	122	1291	99.855
D21381	64,220	39,701	122	1483	96.370
H30863	52,869	34,905	121	1329	94.967
G30314	40,019	25,755	121	1049	100
Bear JB 8/9	48,861	31,946	123	1175	100

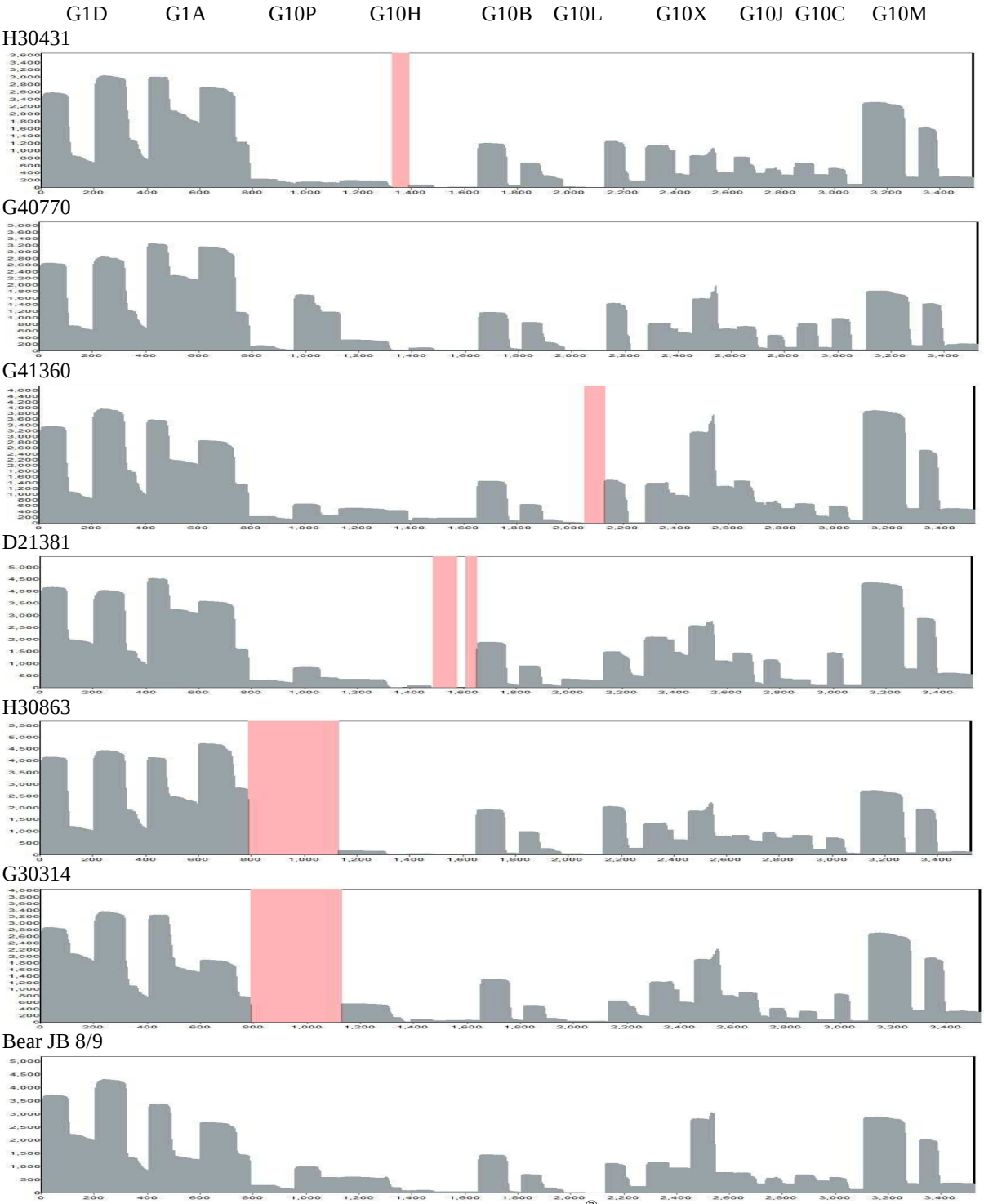


Figure 7: Coverage curves obtained from the NextGENe® software for each sample. Each peak represents a locus (both forward and reverse directions were sequenced). Pink bands indicate where no coverage was obtained. The large pink bands in samples H30863 and G30314 correspond to locus G10P, where the entire locus dropped out.

Aligned reads were analyzed for internal variants such as SNPs (transitions and transversions) and INDELs (Table 6). Each sample contained an average of five SNPs and five INDELs. Transversions made up 60% of the total number of detected SNPs, and insertions made up 68% of the total number of detected INDELs. Approximately half of variants were found in the flanking region preceding the repeats, while the other half were found in the flanking region following the repeats. Only two variants were detected in the repeat region. Two transitions (C → T in G10P and T → C in G10L), two transversions (A → T in G10B (Figure 8) and A → T in G10M), and three insertions (C in G1D (Figure 8) and A and C in G1A) were detected in all samples. G1D had a T deletion observed in all but three samples (H30431, H30863, and Bear JB 8/9) (Figure 8). Another deletion (T in G10X) was observed in four samples (G40770, H30863, G30314, and Bear JB 8/9). One transversion (G → T in G10L) was noted in two samples (G40770 and G41360), and one deletion (T in G10L) was also observed in only two samples (D21381 and G30314). All other variants were unique to one sample. One interesting variant, a C → A in the G1A locus, was observed in only one sample (H30863); however, in sample G40770, half of reads possessed a C and the other half possessed an A, indicating heterozygosity at that SNP for that individual. All other samples were homozygous for C at that location. The G10H locus is also noteworthy, as it appeared to be the most variable of the loci sequenced. Samples H30431, G41360, and Bear JB 8/9 all possessed a GG insertion, sample D21381 had a G → T transversion, sample H30863 had a C → G transversion, and sample G30314 had a heterozygous G/T SNP (Figure 9).

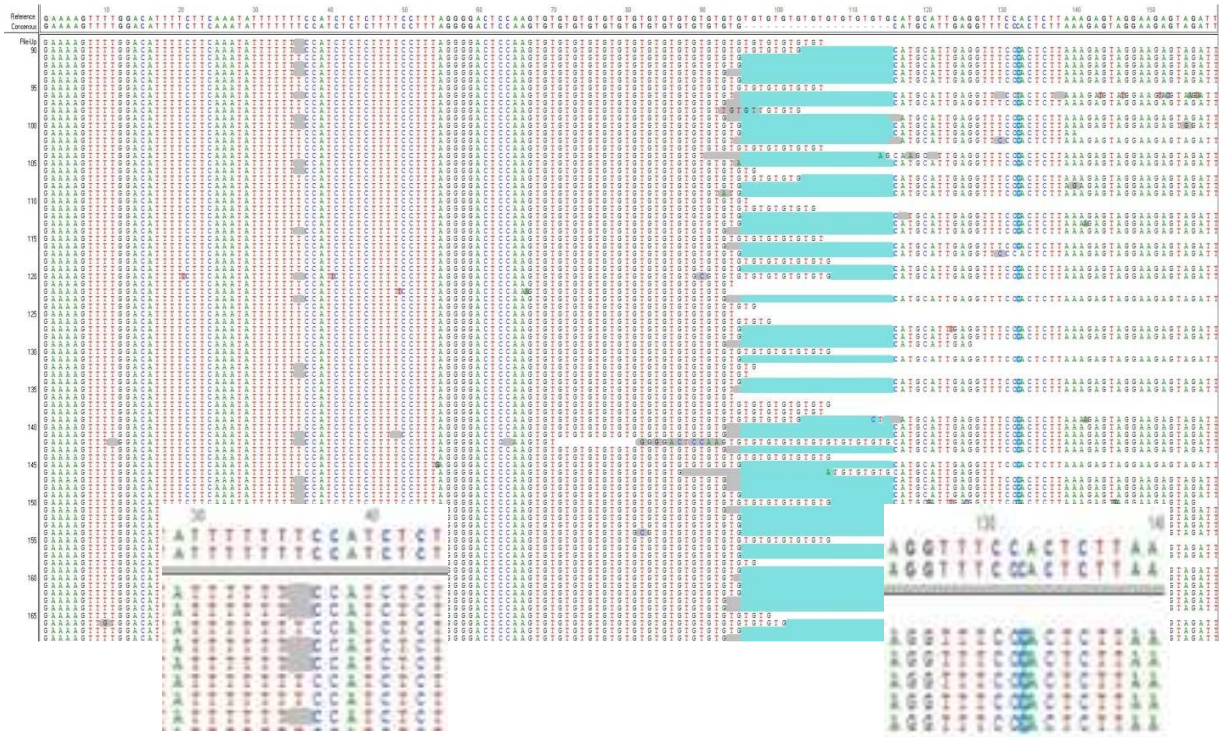
Allele repeat number for each locus was determined using the NextGENe® data when interpretation could be definitively made. This was accomplished by counting the number of reads that possessed the same number of repeats. The repeat number associated with the greatest

number of reads was called as the true allele. Repeat numbers with lesser reads were called as stutter. In the case of heterozygotes, two repeat lengths shared approximately equal numbers of reads, while all other repeat lengths at that locus had a smaller number of reads and were called as stutter. Due to the nature of the virtual ladder, not all sequence reads aligned in a manner that allowed the true allele(s) to be identified. Many of the GenBank® reference sequences used for the ladder contained ambiguous bases, which caused some reads to align in a way such that the allele was not interpretable. Figure 10 depicts read alignment generated using the NextGENe® software for sample H30863 at locus G10J, where the true allele could be determined, and sample G40770 at locus G10X, where the true allele could not be determined definitively. Allele size in bases pairs was calculated from the number of repeats determined from samples that were complete, e.g. the read spanned the repeat region and both flanking regions. Allele repeat number and size are listed in Table 6. Allele size was converted to its reported size by accounting for primer redesign in this study (sizes not shown). The allele calls determined by sequencing were compared to previous CE sizing results obtained from USFWS when such data was available. Allele calls/sizes were concordant within two bp. Loci were also evaluated for discriminatory ability. The G10C locus contained seven true alleles in six individuals, making it the most discriminatory locus in this limited dataset. The least discriminatory locus was G10M, with only two true alleles detected in four individuals. The average number of alleles (as repeats) per locus was 2.7, while the average number of alleles (as size) per locus was 2.8. INDELs not observed in the reference sequences may explain this minor difference. Loci may have the same number of repeats, but one locus may have an INDEL while others do not. For example, in G1D, three samples possessed the same repeat number but one sample also contained an insertion not related to the repeat motif, and therefore possessed an allele size that was one bp larger than the others.

Table 6: SNPs and INDELs for the seven bear samples, and alleles (repeats) and sizes (bp) for interpreted samples. Heterozygotes have two alleles and sizes listed. No data (ND) was available for G10P as it dropped out in samples H30863 and G30314, likely due to alignment issues.

		Sample						
Locus		H30431	G40770	G41360	D21381	H30863	G30314	Bear JB 8/9
G1D	SNPs	0	0	0	0	0	0	0
	INDELs	1	2	2	2	1	2	1
	Allele			14.5		14.5, 17.5	14.5, 18.5	17.5
	Size			181		182, 188	181, 189	188
G1A	SNPs	0	1	0	0	1	0	0
	INDELs	2	2	2	2	2	2	2
	Allele	18			18.5, 20.5	16		18, 20.5
	Size	181			182, 186	177		181, 186
G10P	SNPs	1	1	1	1	ND	ND	1
	INDELs	0	0	0	0	ND	ND	0
	Allele			29	27.5	ND	ND	
	Size			163	160	ND	ND	
G10H	SNPs	0	0	0	1	1	1	0
	INDELs	1	0	2	0	0	0	2
	Allele			20.5		32.5		
	Size			230		250		
G10B	SNPs	3	1	1	1	1	1	1
	INDELs	1	1	1	1	1	1	1
	Allele	19	19	24		19	22.5	24
	Size	145	145	155		145	152	155
G10L	SNPs	1	2	2	1	1	1	1
	INDELs	0	0	0	1	0	1	0
	Allele			26				
	Size			137				
G10X	SNPs	0	0	0	0	0	0	0
	INDELs	0	1	0	0	1	1	1
	Allele							
	Size							
G10J	SNPs	0	0	0	0	0	0	0
	INDELs	0	0	0	0	0	0	0
	Allele	12.5		12.5		14.5	19.5	12.5
	Size	84		84		88	98	84
G10C	SNPs	0	0	0	0	0	0	0
	INDELs	0	0	0	0	1	1	0
	Allele	17.5	20.5		13.5	17.5, 21.5	14.5	16, 22
	Size	105	111		97	105, 112	98	102, 114
G10M	SNPs	1	1	1	1	1	1	1
	INDELs	0	0	0	0	0	0	0
	Allele		20.5	20.5	20		20	
	Size		193	193	192		192	

G30314 G1D



G10B

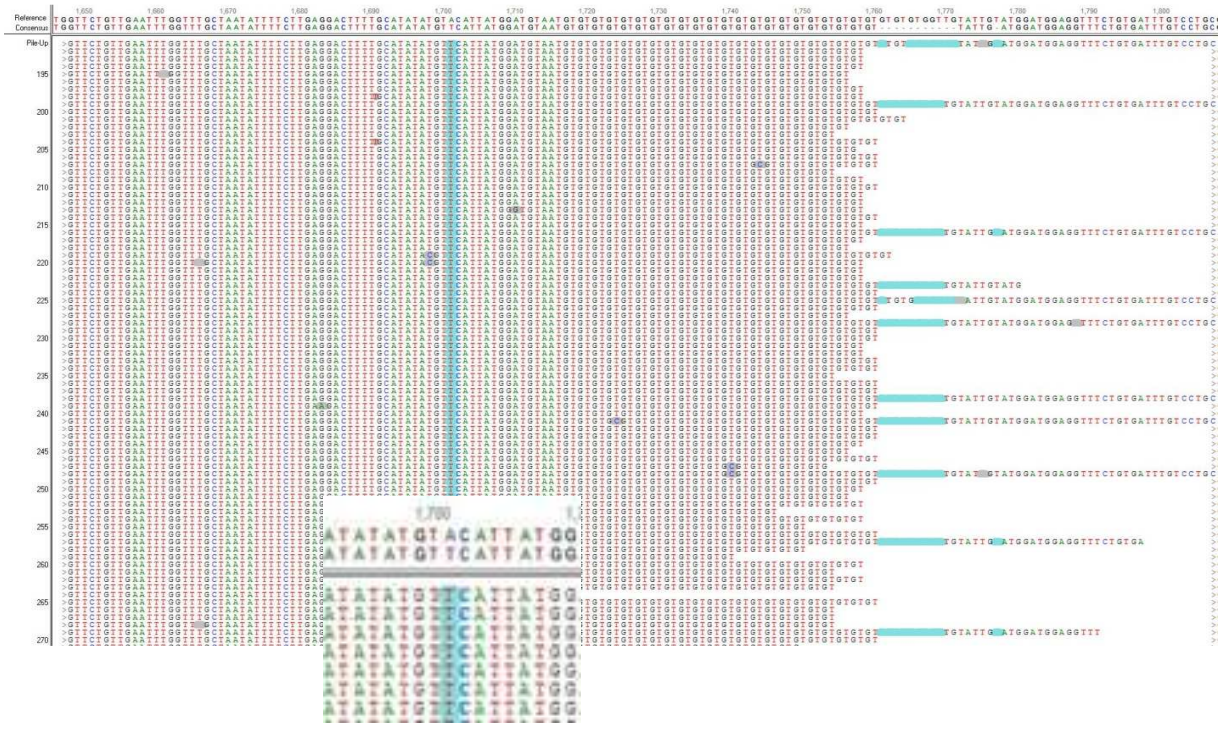
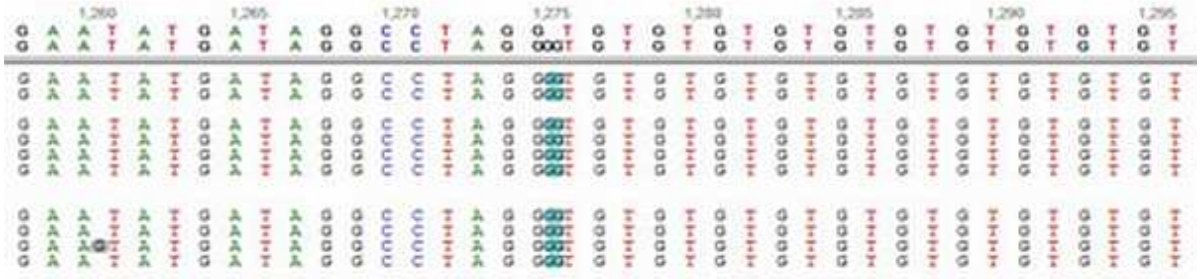
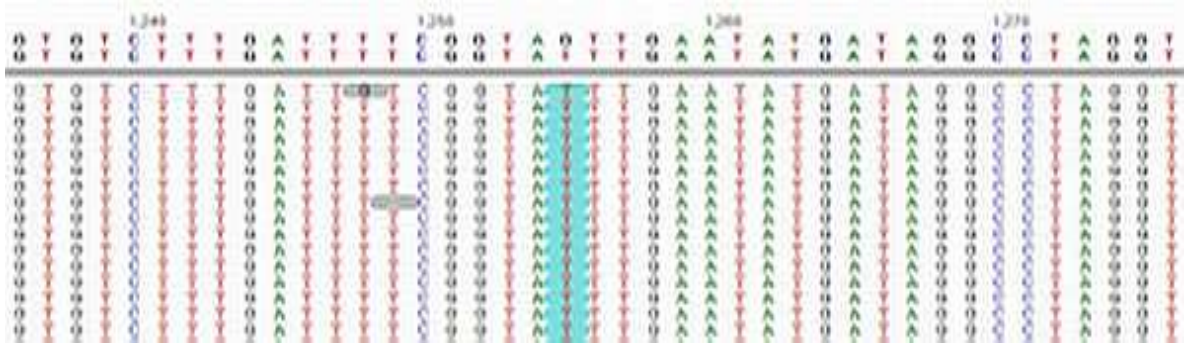


Figure 8: Examples of variants observed during analysis. G1D contains a T deletion observed in all samples except H30431, H30863, and Bear JB 8/9; G1D also has a C insertion observed in all samples (detail in insets). G10B possesses an A→T SNP observed in all samples (detail in inset).

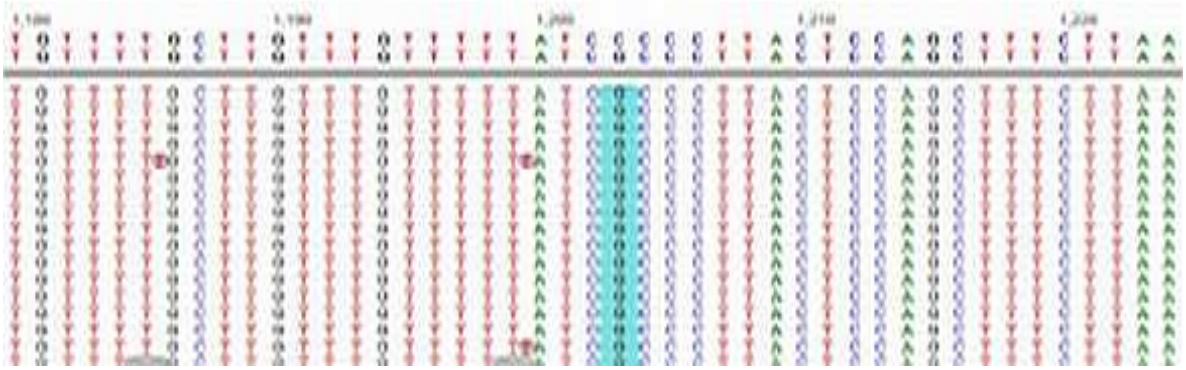
G10H GG insertion



G10H G→T SNP



G10H C→G SNP



G10H G/T heterozygote

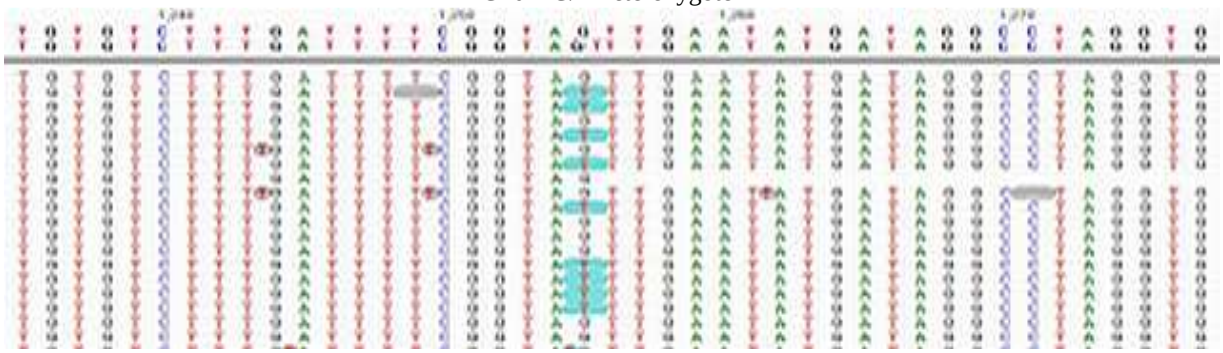
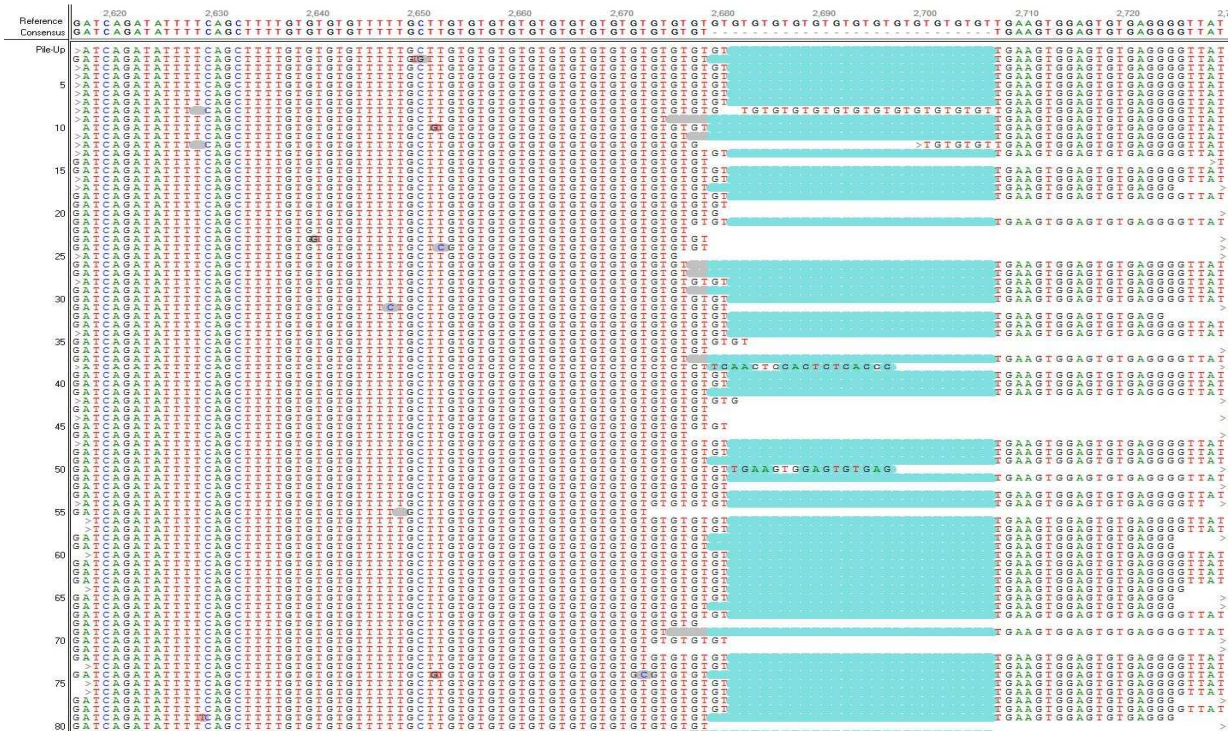


Figure 9: Variants at the G10H locus. Samples H30431, G41360, and Bear JB 8/9 all possessed a GG insertion, sample D21381 had a G → T transversion, sample H30863 had a C → G transversion, and sample G30314 had a heterozygous G/T SNP.

H30863 G10J



G40770 G10X

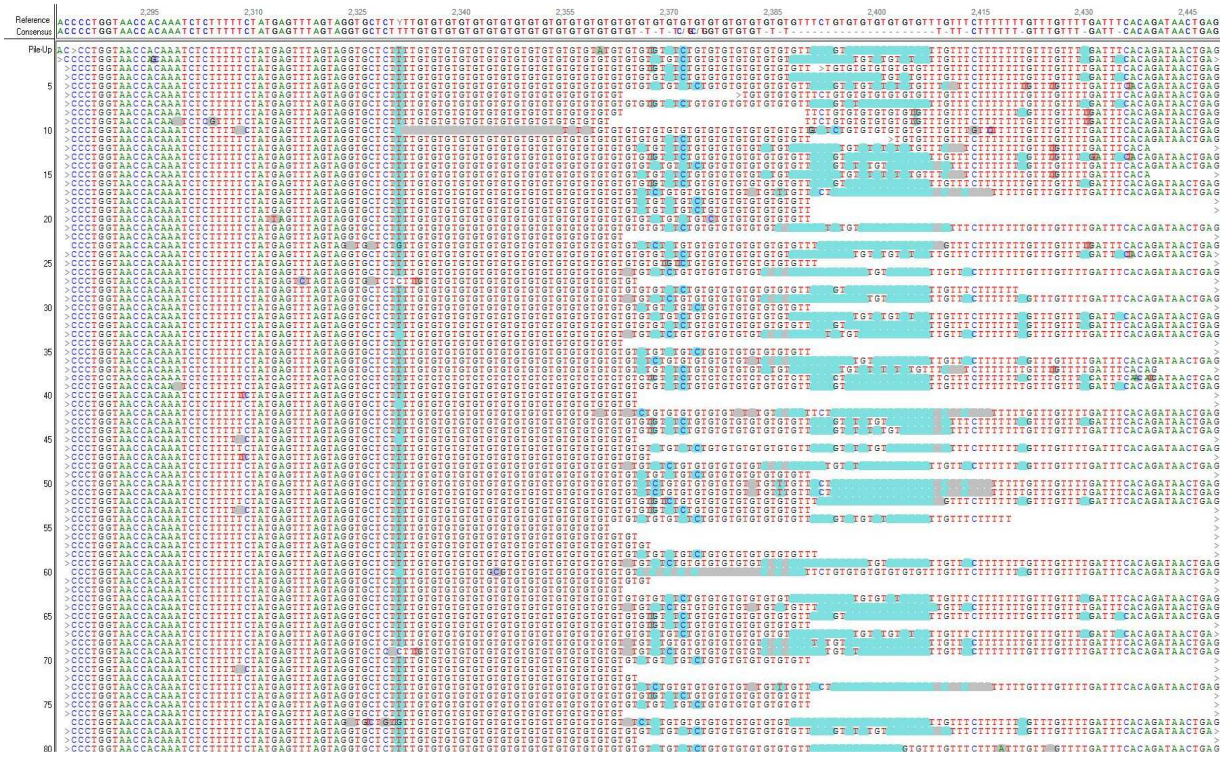


Figure 10: Example of sequence data where true allele could be determined (H30863 G10J) since gaps are all the same size, and could not be determined definitively (G40770 G10X) since gaps are uninterpretable.

Prior to this study, only one reference sequence was cataloged for each locus in GenBank®. All references were sequenced in the mid-1990s by use of a PCR cycle sequencing kit with subsequent product separation on a standard acrylamide sequencing gel and visualization via autoradiography (Paetkau & Strobeck 1994; Paetkau *et al*. 1995). As a result of the limitations of the technology at that time, numerous ambiguous bases existed in some of the references. The G10H locus is one example. In the GenBank® reference, 17 ambiguous bases (IUPAC codes: M, V, S, K, Y, W, R, D, H, and B) (Appendix J) were included. Part of the aim of this study was to update these sequences and definitively call all of the ambiguous bases. As a result of this study, ambiguous bases have been unambiguously defined (Figure 11) and sequences for each locus have been confirmed using sequencing data from both forward and reverse strands. The reference ladder was subsequently updated with the new information and a second analysis was performed in NextGENe®. Samples successfully aligned to the updated reference, and thus confirmed the base calls. These confirmed sequences for each locus will be submitted to GenBank®, allowing us to achieve our study aim. An example of the second analysis with the revised reference sequence is given in Figure 12. The updated reference sequences for the bear STR loci based on the specimens examined are listed in Appendix J.

H30431 G10M

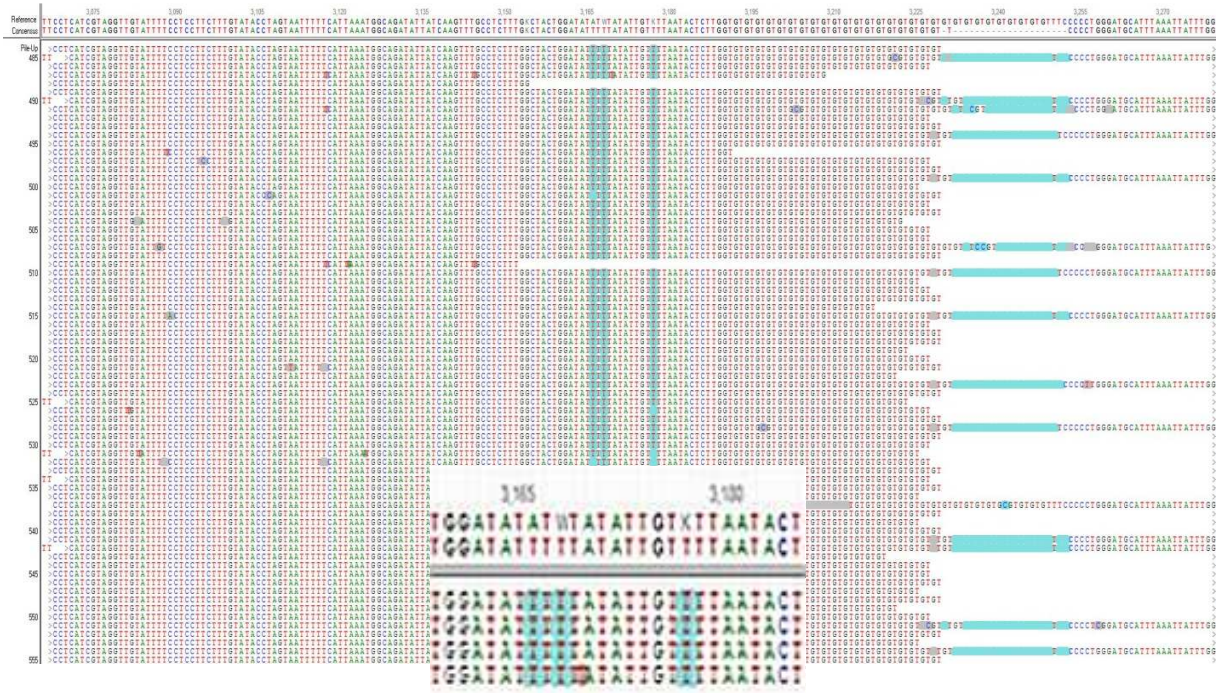


Figure 11: Example of ambiguous bases (W and K) in the reference sequence for the G10M locus that were called by sequencing data (detail in inset).

G41360 G10H

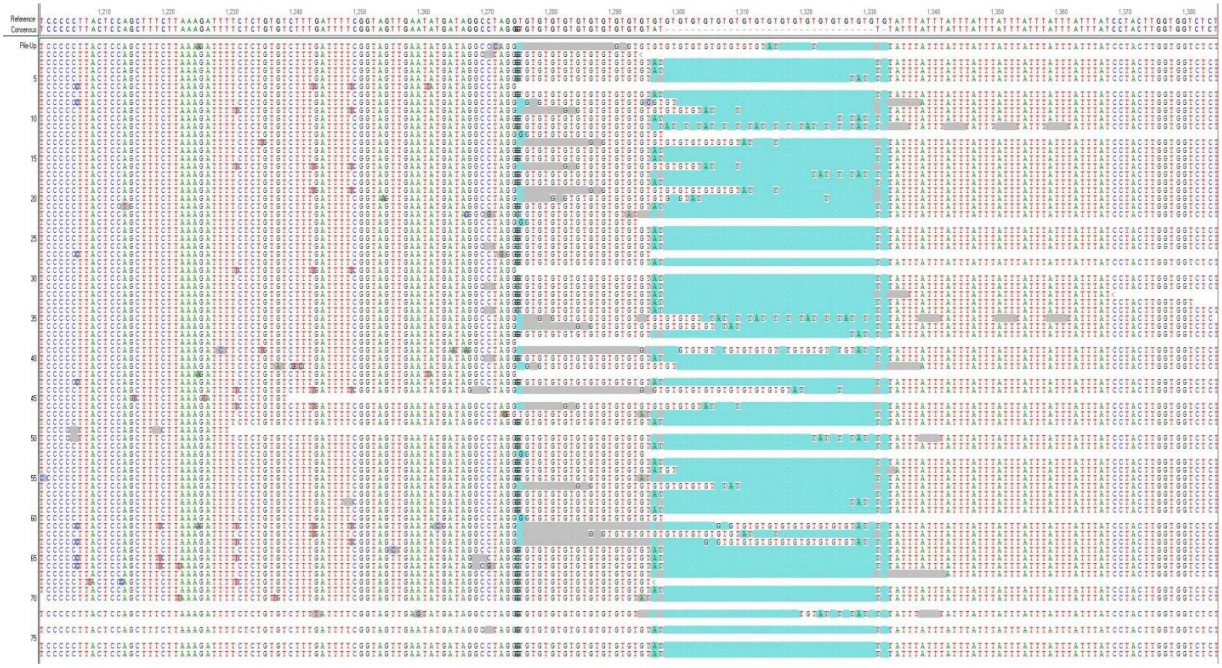


Figure 12: Example of second analysis performed with revised reference sequence. In the previous reference sequence for G10H, 17 ambiguous bases were included. Following next-generation sequencing, all bases have been called and the sample aligned successfully.

CHAPTER IV

DISCUSSION

Prior to this study, the method for obtaining a profile for an individual consisted of performing three multiplex PCR reactions and then sorting the loci by size via capillary electrophoresis. Figure 13 contains an example of a profile obtained for an individual by the USFWS-NWFWL. Dinucleotide repeats have a tendency to produce high amounts of stutter that are one repeat more or less than the true allele (Walsh *et al.* 1996). However, stutter can also be generated that is two and three repeats greater or less than the true allele. These stutter products are present in Figure 13. Non-specific PCR products are also observed in Figure 13, which could be a result of weak primer design, low annealing temperature, too many cycles in PCR, long extension time, or incorrect $MgCl_2$ concentration. Primers used by the USFWS-NFWFL were originally designed from the reference sequences available in GenBank®. As previously mentioned, these references were sequenced with old technologies and contained many ambiguous bases, making proper primer design difficult for some loci. In this study, primers were redesigned in order to increase product specificity and singleplex reactions were optimized to maximize the presence of specific DNA product for each locus and minimize the presence of stutter and non-specific products. This was accomplished by performing several rounds of singlepex PCR, each with varying amounts of the reaction components, PCR cycles, and annealing temperatures, until an optimized protocol was determined. However, dinucleotide behavior is not ideal for accurate STR analysis, so future studies should focus on locating better STR loci for identification purposes. Currently, the UNTHSC Evolutionary Genetics Laboratory

is preparing to conduct a study to mine Ursid tetranucleotides that may provide better resolution for forensic and population analysis of bears.

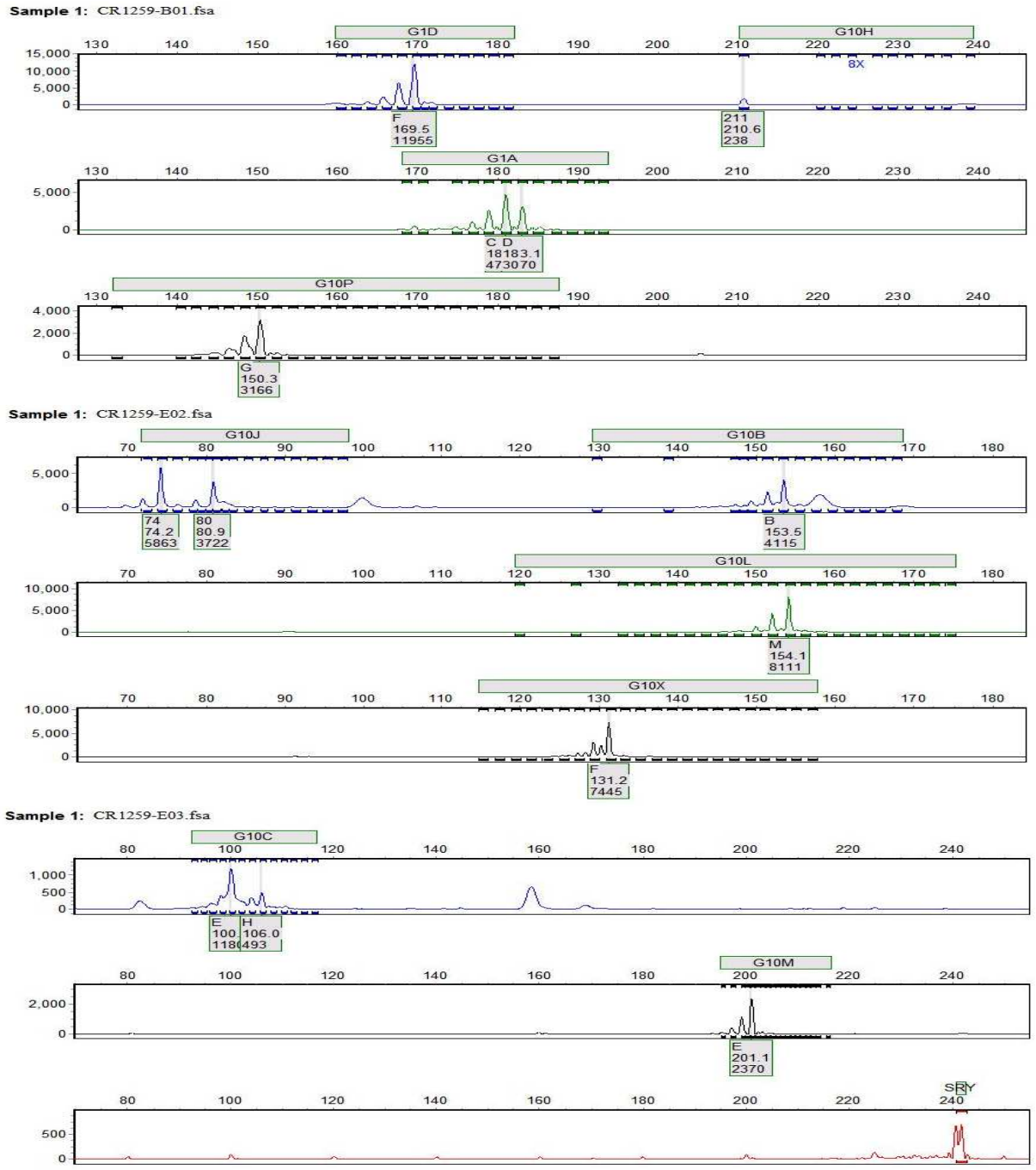


Figure 13: CE electropherograms from a sample run by the USFWS-NFWFL. Each locus has extensive stutter. Non-specific products are also seen.

This study confirmed that the sequencing chemistry of the Ion Torrent™ PGM™ Sequencer is not ideally suited for sequencing STR loci. Each locus had high average coverage (>888X), and >15% of reads extended across the entire locus. However, a large portion of sequencing reads prematurely terminated in the repeat region. This could be a result of the polymerase destabilizing, therefore causing the chemistry to get "hung-up" on the repeats and fail to finish sequencing the fragment. Future studies should examine other chemistries that are better suited to handle repeat regions and homopolymeric stretches.

This study also found that the NextGENe® software is not currently capable of organizing and sorting reads so that true alleles may be easily gleaned from the data. Allele calls were able to be made, but in instances of a heterozygous locus with stutter products, true alleles were difficult to discern since there were sometimes as many as eight repeat lengths (two true alleles, each with three stutter products, for example) found throughout the reads. There are other software tools available for the analysis of STR regions that rely on a preloaded human genome version as a reference (Gymrek *et al.* 2012; Langmead *et al.* 2009; Warshauer *et al.* 2013). However, when a genome is not available for reference, as is the case with many wildlife species, a reference ladder must be created *in silico* and uploaded for use. Unfortunately, not all available software tools allow other references to be entered, and for those that do, it is essential that the references be as accurate as possible. For bears, only one published reference sequence for each locus existed in GenBank®. These reference sequences were sequenced using a PCR cycle sequencing kit with subsequent product separation on a standard acrylamide sequencing gel and visualization via autoradiography. Cycle sequencing combined dideoxy sequencing with thermal cycling, but had a tendency to superimpose sequences, causing the flanking regions surrounding the repeat region to appear out of phase, thus resulting in ambiguous calls. These

inaccuracies and ambiguous bases present in the reference ladder most likely caused the software's inability to properly align the data. As part of this study, the ambiguous bases reported in the reference sequences were clarified, as a result of successfully determining the true base sequence. These sequences were used to update the reference sequences. Future bioinformatics studies should focus on developing a software pipeline that allows a reference ladder to be uploaded and then allows for sorting and organization of STR data, such that incomplete reads and stutter products could be sorted from true alleles and make data interpretation easier and more accurate.

As a result of this study, previously unknown information was revealed about ten Ursid STR loci. Information about SNPs and INDELs in these STR loci was discovered for North American black bears. Each individual possessed an average of five SNPs and five INDELs that were contained within the flanking regions of the loci. Four of these SNPs and three of these INDELs were present in all individuals. All samples used in this study were from black bear, and all reference data came from a black bear; therefore, these variants were likely due to poor sequencing of the reference data in the mid-1990s. As previously mentioned, the sequencing method used for this reference data was known to superimpose sequences and cause flanking regions to appear out of phase. Another possible explanation is that the reference data and samples did not actually come from the same species, and therefore, variants were due to species differences. Future sequencing studies are needed to compare SNP and INDEL differences between black bears, brown bears, and polar bears, to investigate if these variants are found in all black bears, but not other species. The reference sequences confirmed by this study will be used to update GenBank® and expand the available reference data for bears.

SNPs and INDELs that were polymorphic may potentially be used to assist in individual identification or geographic assignment. Of specific interest are the variants that were unique to a single individual, as was the case with the C → G SNP in G10H in sample H30863 (Figure 9), two heterozygous SNPs (C/T and A/G) in G10B in sample H30431, a C deletion in G10C in sample H30863, and an A deletion in G10C in sample G30314. These SNPs and INDELs may be a mutation unique to that individual, thus drastically increasing the discriminatory power of the locus, or may be population specific. Another interesting discovery was that the three individuals that possessed the GG insertion in G10H were all in geographic proximity to one another. The individuals were recorded to have come from Pennsylvania, New York, and Florida (Florida black bear are a threatened population and relocations from other areas are common, therefore, an individual from the northeast may have been relocated to Florida). This discovery lends credence to the prospect of geographic assignment using SNPs and INDELs. As all black bear samples used in this study originated from different geographic areas, population studies need to be conducted to determine if more bears from the same area contain the same variant.

Allele calls that could be made were concordant with previous CE data. Thus, this study has demonstrated that next-generation sequencing applications are a viable option for DNA profiling. However, for this approach to be streamlined, sequencing chemistries and bioinformatics pipelines more suited to non-human STR loci need to be developed. Upon analyzing the data, several loci appeared to have limited polymorphism in the samples provided. For a profile to be useful for individual identification, STR loci must have high discriminatory power. Loci that provide a greater amount of polymorphism would increase discriminatory power and therefore be helpful for forensic purposes. As already discussed, the UNTHSC Evolutionary Genetics Laboratory is preparing to mine for new loci that may provide more

polymorphism among bears. In the meantime, SNP and INDEL data from this study may be used in combination with allele calls to increase the discriminatory power of a profile.

In conclusion, this study uncovered data that can help identify individual black bears. With future population studies, these data may also help to determine geographic origin. Allele calls were shown to be concordant with CE data. SNP and INDEL information may be used in conjunction with STR alleles to provide better genetic resolution and more discriminatory power. Population studies from various geographic areas need to be conducted to determine if these SNPs and INDELs are population specific or are polymorphisms unique to the individual. Implications of this study include better genetic resolution that may assist conservation officers and wildlife forensic analysts in individual identification, while also providing higher discriminatory power for wildlife crime cases that go to court.

APPENDIX

APPENDIX A

Protocol for Organic DNA Extraction: Blood, Tissue, and Other Biological Material
(Adapted from UNT Center for Human Identification Procedure Manual)

Stain Extraction Buffer:

1. Combine 10 ml of Tris (pH 8.0), 20 ml of 5M NaCl, 20 ml of 0.5M EDTA, 200 ml of 10% SDS, and 600 ml of diH_2O. Titrate to pH 8.0 with concentrated HCl.

Cell Lysis and Protein Digestion:

1. Dissect 1-2 cm^2 of tissue and transfer to a labeled, sterile 2 ml microcentrifuge tube.
2. Add 500 µl of Stain Extraction Buffer, 10 µl of 1M dithiothreitol (DTT), and 5 µl of 20 mg/ml Proteinase K to each sample.
3. Briefly vortex samples and pulse spin.
4. Incubate samples at 56°C (±1°C) for 2-24 hours.

Phenol-Chloroform-IAA Extraction:

1. After digestion, briefly pulse spin samples to force condensation build up during incubation into bottom of tube.
2. Add 515 µl of phenol-chloroform-isoamyl alcohol (PCIA, 25:24:1) to each sample.
3. Vortex samples for 15-30 seconds to attain a milky emulsion, then centrifuge for 3 minutes at 16,000-19,000 x g.
4. Carefully remove the aqueous phase (top layer) from each sample. Avoid drawing any of the protein interface or organic solvent into the pipette tip. Transfer to a Microcon filtration device as described below.
5. The DNA can now be concentrated by Microcon 100 concentration.

Microcon 100 Concentration:

1. Assemble and label a Microcon® YM-100 Centrifugal Filter Device for each sample. Pre-wet the concentrator by adding 100 µl of sterile water to the filter side.
2. Carefully transfer the aqueous phase to the appropriately labeled concentrator, being careful to avoid transfer of the protein interface of organic solvents. Sample volume transferred should not exceed 400 µl.
3. Centrifuge the Microcon® YM-100 at 500 x g for 30 min. Discard flow-through.
4. Add 400 µl of sterile water to each concentrator.
5. Centrifuge at 500 x g for 55 min.
6. Discard filtrate tube and add 50 µl of hot (80-90°C) sterile water to the filter side of each concentrator. Place an appropriately labeled retentate tube on top of each concentrator.
7. Briefly vortex the concentrator with the retentate tube facing up, then invert the assembly and centrifuge at 1000 x g for 3 min. Discard the concentrators.
8. Record volume of DNA. Store at 4°C for short term storage or -10°C or less for long term storage.

APPENDIX B

Protocol for Qubit® 2.0 Fluorometer DNA Quantification (Adapted from Qubit® 2.0 Fluorometer User Manual)

Running New Standards for Calibration:

1. On the Home Screen, choose **DNA**. Standards Screen is automatically displayed.
2. Press **Yes** to read new standards. A prompt to insert Standard #1 appears on the screen.
3. Insert Standard #1 into the Sample Chamber and press **Read**. Ensure that you are using the Standard #1 appropriate for the DNA assay. Reading takes approximately 3 seconds.
4. Insert Standard #2 and press **Read**. Ensure that you are using the Standard #2 appropriate for the DNA assay.
5. The new standards graph with data points for standards connected by a line appears on the screen.

Reading Samples:

1. Choose **Sample** to go to the Sample Screen.
2. Insert a sample into the Sample Chamber and press **Read**. Measurement takes approximately 3 seconds. Upon completion of the measurement, the result is displayed on the screen. The number displayed is the concentration of DNA in the assay tube.
3. To read the next sample, remove the sample from the Sample Chamber, insert the next sample, and press **Read Next Sample**.
4. Repeat sample readings until all samples have been read.

Dilution Calculator:

1. To calculate the concentration of your original sample, press **Calculate Stock Conc**.
2. Using the volume roller wheel, select the volume of your original sample that you have added to the assay tube. When you stop scrolling, the original sample concentration is calculated based on the measured assay concentration.
3. To change the units in which the original sample concentration is displayed, press **ng/mL**. A pop-up window showing the current unit selection opens.
4. Select the unit for your original sample concentration by touching the desired unit in the unit selection pop-up window. To close the unit selection pop-up window, touch anywhere on the screen outside the pop-up. The Qubit® 2.0 Fluorometer automatically converts the units to your selection when the unit selection pop-up window is closed.

APPENDIX C

Protocol for the Singleplex Amplification of Ten Bear Loci

Primers:

Locus	Forward $(GT)_n$ strand primer	Reverse $(CA)_n$ strand primer
G1A	GACCCTGCATACTCTCCTCTGATG	GCACTGTCCTTGCGTAGAAGTGAC
G1D	GGAAAAGTTTTGGACATTTTCTTC	CAAGAACACTTGATACCTAGCAC
G10B	GTTCTGTTGAATTTGGTTTGC	GCAGGACAAATCACAGAAAC
G10C	AAAGCAGAAGGCCTTGATTTCCTG	GGGGACATAAACACCGAGACAGC
G10H	CAACAAGAAGACCACTGTAA	AGAGACCACCAAGTAGGATA
G10J	GATCAGATATTTTCAGCTTTTGTGTGTG	ATAACCCCTCACACTCCACTTC
G10L	GTACTGATTTAATTCACATTTCCC	CCTACCCATGCGATAAAATTG
G10M	CCTCATCGTAGGTTGTATTTTCCTCC	CCAAATAATTTAAATGCATCCCAGGG
G10P	AGGAGGAAGAAAGATGGAAAAC	TCATGTGGGGAAATACTCTGAA
G10X	CCCTGGTAACCACAAATCTCT	CTCAGTTATCTGTGAAATCAAAAC

Singleplexes for G1A, G10C, G10J, and G10M:

1. Add the following to a PCR tube for each reaction (total 25 µL):

10X PCR Buffer II	2.5 µl
25mM $MgCl_2$	3.0 µl
10mM dNTP mix	1.0 µl
Bovine Serum Albumin (1.6 µg/µL)	2.5 µl
AmpliTaq Gold® (5 U/µL)(Life Technologies™)	0.5 µl
5M Betaine (Sigma-Aldrich®)	2.5 µl
10µM forward primer	1.2 µl
10µM reverse primer	1.2 µl
Template DNA (5 ng/µL)	2.0 µl
Ultrapure distilled water	8.6 µl

2. Run on a thermal cycler as follows:

1 cycle	95°C	11 minutes
35 cycles	95°C	10 seconds
	60°C	30 seconds
	72°C	30 seconds
1 cycle	70°C	10 minutes
Hold	4°C	forever

Singleplexes for G1D, G10B, G10H, G10L, G10P, and G10X:

1. Add the following to a PCR tube for each reaction (total 25 µL):

10X PCR Buffer II	2.5 µl

25mM $MgCl_2$	2.0 µl
10mM dNTP mix	1.0 µl
Bovine Serum Albumin (1.6 µg/µL)	2.5 µl
AmpliTaq Gold® (5 U/µL)(Life Technologies™)	0.5 µl
5M Betaine (Sigma-Aldrich®)	2.5 µl
10µM forward primer	1.2 µl
10µM reverse primer	1.2 µl
Template DNA (5 ng/µL)	2.0 µl
Ultrapure distilled water	9.6 µl

2. Run on a thermal cycler as follows:

1 cycle	95°C	11 minutes
35 cycles	95°C	10 seconds
	52°C	30 seconds
	72°C	1 minute
1 cycle	70°C	10 minutes
Hold	4°C	forever

APPENDIX D

Protocol for Post-Amplification Quantification using Agilent 2100 Bioanalyzer and Agilent DNA 1000 Kit (Taken from manufacturer's protocol)

Preparing the Gel-Dye Mix:

1. Allow DNA dye concentrate and DNA gel matrix to equilibrate to room temperature for 30 minutes.
2. Vortex DNA dye concentrate and add 25 µl of the dye to a DNA gel matrix vial.
3. Vortex solution well and spin down. Transfer to spin filter.
4. Centrifuge the solution at 2240 g ± 20% for 15 minutes. Protect solution from light. Store at 4°C.

Loading the Gel-Dye Mix:

1. Allow the gel-dye mix to equilibrate to room temperature for 30 minutes before use.
2. Put a new DNA chip on the chip priming station.
3. Pipette 9.0µl of gel-dye mix in the well marked with the shaded (G).
4. Make sure that the plunger is positioned at 1 ml and then close the chip priming station.
5. Press plunger until it is held by the clip.
6. Wait exactly 60 seconds, and then release the clip.
7. Wait for 5 seconds. Slowly pull back plunger to 1 ml position.
8. Open the chip priming station and pipette 9,0 µl of gel-dye matrix in the two wells marked with unshaded (G)s.

Loading the Markers:

1. Pipette 5 µl of marker in all 12 sample wells and ladder well. Do not leave any wells empty.

Loading the Ladder and the Samples:

1. Pipette 1 µl of DNA ladder in the well marked with the ladder symbol.
2. In each of the 12 sample wells pipette 1 µl of sample (used wells) or 1 µl of de-ionized water (unused wells).
3. Put the chip horizontally in the adapter and vortex for 1 min at the indicated setting (2400 rpm).
4. Run the chip in the Agilent 2100 bioanalyzer within 5 minutes.

APPENDIX E

QIAquick® PCR Purification Kit Protocol (Adapted from manufacturer's protocol)

Before Starting:

1. This protocol is for the purification of up to 10 µg PCR products (100 bp to 10 kb in size).
2. Add 24 ml of 100% ethanol to Buffer PE before use.
3. All centrifugation steps are carried out at 13,000-15,000 x g in a conventional table-top microcentrifuge at room temperature.
4. Add 120 ml of pH indicator I to Buffer PB. The yellow color of Buffer PB with pH indicator I indicates a pH of ≤ 7.5. The adsorption of DNA to the membrane is only efficient at pH ≤ 7.5.

Procedure:

1. Add 5 volumes of Buffer PB to 1 volume of the PCR reaction and mix. If the color of the mixture is orange or violet, add 10 µl 3M sodium acetate, pH 5.0, and mix. The color of the mixture will turn yellow.
2. Place a QIAquick® column in a provided 2 ml collection tube.
3. To bind DNA, apply the sample to the QIAquick® column and centrifuge for 1 minute. Discard flow-through and place the QIAquick® column back in the same tube.
4. To wash, add 750 µl of Buffer PE to the QIAquick® column and centrifuge for 1 minute. Discard flow-through and place the QIAquick® column back in the same tube.
5. Centrifuge the QIAquick® column once more in the provided 2 ml collection tube for 1 minute to remove residual wash buffer.
6. Place each QIAquick® column in a clean 1.5 ml microcentrifuge tube.
7. To elute DNA, add 30 µl of water (pH 7.0-8.5) to the center of the QIAquick® membrane, let the column stand for 1 minute, and then centrifuge for 1 minute.

APPENDIX F

Protocol for the NEBNext® Fast DNA Library Prep Set for Ion Torrent™ (Taken from manufacturer's protocol)

End Repair:

1. Mix the following components in a sterile microfuge tube on ice:

Fragmented DNA (10 ng-100 ng)	20 µl
NEBNext® End Repair Reaction Buffer	6 µl
NEBNext® End Repair Enzyme mix	3 µl
Sterile water	31 µl

2. Incubate in a thermal cycler for 20 min at 25°C, followed by 10 min at 70°C, hold at 4°C.
3. Pulse spin the microfuge tube and return to ice.

Preparation of Adaptor Ligated DNA:

1. Add the following to the previous microfuge tube for a total volume of 100 µL, then mix by pipetting up and down several times:

Sterile water	18 µl
T4 DNA Ligase Buffer for Ion Torrent™	10 µl
NEXTflex™ Barcode Adaptors for Ion Torrent™	2.5 µl
NEXTflex™ DNA P1 Adaptor for Ion Torrent™	2.5 µl
Bst WarmStart DNA Polymerase	1 µl
T4 DNA Ligase	6 µl

2. Incubate in a thermal cycler for 15 min at 25°C, followed by 5 min at 65°C, hold at 4°C.

Cleanup of Adaptor Ligated DNA Using Microcon® YM-100:

1. Assemble and label a Microcon® YM-100 Centrifugal Filter Device for each sample. Pre-wet the concentrator by adding 100 µl of sterile water to the filter side of each concentrator.
2. Carefully transfer the adapter ligated DNA to the appropriately labeled concentrator.
3. Centrifuge the Microcon® YM-100 at 500 x g for 30 min. Discard flow-through.
4. Add 400 µl of sterile water to each concentrator.
5. Centrifuge at 500 x g for 55 min.
6. Discard filtrate tube and add 50 µl of hot (80-90°C) sterile water to the filter side of each concentrator. Place an appropriately labeled retentate tube on top of each concentrator.
7. Briefly vortex the concentrator with the retentate tube facing up, then invert the assembly and centrifuge at 1000 x g for 3 min. Discard the concentrators.

PCR Amplification of Adaptor Ligated DNA:

1. Mix the following components in a sterile microfuge tube:

Adaptor Ligated DNA	20 µl

Primers	4 µl
Sterile water	6 µl
NEBNext High-Fidelity 2X PCR Master Mix	50 µl

2. Run on a thermal cycler as follows:

Initial Denaturation	98°C	30 seconds
8 cycles	98°C	10 seconds
	58°C	30 seconds
	72°C	30 seconds
1 cycle	72°C	5 minutes
Hold	4°C	forever

Clean Up of Amplified Library:

1. Add 100 µl (1.0X volume) of AMPure XP Reagent to the sample and mix by pipetting up and down.
2. Incubate for 5 minutes at room temperature.
3. Pulse-spin and place in a magnetic rack for 3 minutes until the beads have collected to the wall of the tube and the solution is clear.
4. Carefully remove and set aside the supernatant without disturbing the beads.
5. Keep the tube on the magnet and add 500 µl freshly prepared 80% ethanol. Incubate 30 seconds, then carefully remove and discard the supernatant.
6. Repeat step 5.
7. Pulse-spin, return to the magnet and remove any residual ethanol with a pipet.
8. Keeping the tube in the magnetic rack with the cap open, air dry the beads for 5 minutes at room temperature.
9. Resuspend the beads in 35 µl of 0.1 x TE.
10. Pulse-spin, return to the magnet until the beads have collected to the wall of the tube and the solution is clear.
11. Transfer approximately 30 µl of supernatant to a fresh tube. Be careful not to transfer any beads.
12. Quantify the library on the Qubit® 2.0 Fluorometer.

APPENDIX G

Protocol for the Ion PGM™ Template OT2 400 Kit for use with Ion OneTouch™ 2 System
(Taken from the manufacturer's protocol)

Set up the Ion OneTouch™ 2 Instrument:

1. Install the 2 Ion OneTouch™ Recovery Tubes and the Ion OneTouch™ Recovery Router, then close the centrifuge lid.
2. Remove the used cleaning adaptor, insert the plate, and pull the handle to close the heat block. Thread the disposable tubing through the catch and pinch valve.
3. Insert the disposable injector, then confirm automatic placement of the disposable injector above the router by briefly pressing then releasing the spring-loaded top of the Injector Hub. You should hear a click.
4. Install the Ion OneTouch™ Oil on the left front port. Invert the Ion OneTouch™ Oil bottle 3 times, then fill the Reagent Tube half-full with Oil. Install the Reagent Tube. Minimize bubbles.
5. Install the Ion PGM™ OT2 Recovery Solution on the right front port. Invert the bottle of Recovery Solution 3 times, then fill the Reagent Tube a quarter-full with Recovery Solution. Install the Reagent Tube. Minimize bubbles.
6. Empty the waste container.

Prepare the Amplification Solution:

1. Allow the Ion PGM™ Template OT2 400 Reagent Mix to come to room temperature. Vortex for 30 seconds, then centrifuge for 2 seconds.
2. Vortex the Ion PGM™ Template OT2 400 Reagent B for 1 minute, then centrifuge for 2 seconds.
3. Centrifuge the Ion PGM™ Template OT2 400 Enzyme Mix for 2 seconds. Place on ice.
4. Centrifuge the Ion PGM™ Template OT2 400 Reagent X for 2 seconds. Place on ice.
5. Place the Ion PGM™ Template OT2 400 Ion Sphere™ Particles at room temperature.
6. Dilute the Amplicon Library to 26pM. Vortex the diluted library for 5 seconds, then centrifuge for 2 seconds. Place on ice.
7. In a 1.5 ml Eppendorf LoBind® Tube at room temperature, add the following:

Ion PGM™ Template OT2 400 Reagent Mix	500 µl
Ion PGM™ Template OT2 400 Reagent B	285 µl
Ion PGM™ Template OT2 400 Enzyme Mix	50 µl
Ion PGM™ Template OT2 400 Reagent X	40 µl
Diluted library	25 µl

8. Vortex the solution for 5 seconds, then centrifuge for 2 seconds.
9. Vortex the Ion PGM™ Template OT2 400 Ion Sphere™ Particles (ISPs) for 1 minute, centrifuge 2 seconds, pipet up and down, then add 100 µl to the amplification solution to obtain a total volume of 1000 µl.
10. Vortex the complete amplification solution for 5 seconds.

Fill and Install the Ion PGM™ OneTouch Plus Reaction Filter Assembly:

1. Pipet 1000 µl of the amplification solution through the sample port.
2. Pipet 1000 µl of the Ion OneTouch™ Reaction Oil through the sample port.
3. Pipet 500 µl of the Ion OneTouch™ Reaction Oil through the sample port.
4. Invert then install the filled Ion PGM™ OneTouch Plus Reaction Filter Assembly into the 3 holes on the top of the stage of the Ion OneTouch™ 2 Instrument.

Run the Ion OneTouch™ 2 Instrument:

1. Close the lid, touch **Run**, then select **PGM: Ion PGM™ Template OT2 400 Kit**, touch **Next**, touch **Assisted** or **Expert**. To cancel, touch **Abort**, then touch **Yes**. Remove the samples ≤16 hours after starting the run.

Recover the Template-Positive Ion PGM™ Template OT2 400 Ion Sphere™ Particles:

1. Follow the screen prompts to centrifuge the sample. If you removed the Reaction Tubes at the end of the run before the Ion OneTouch™ 2 Instrument had spun the sample or have not processed the sample within 15 minutes, centrifuge the sample on the instrument. On the home screen, touch **Options**, then touch **Final Spin**, then follow any screen prompts to centrifuge the sample.
2. Discard the Recovery Router and remove the Recovery Tubes.
3. Remove all but 100 µL of solution from each Recovery Tube.
4. With a new tip and using the same tip for both tubes, resuspend the Ion PGM™ Template OT2 400 Ion Sphere™ Particles (ISPs) in the remaining Ion PGM™ OT2 Recovery Solution. Pipet the pellet up and down until each pellet disperses in the solution.
5. Remove 2 µL from one of the tubes and place in a 0.2 mL PCR tube. Perform a Qubit® 2.0 Fluorometer quantification using the necessary reagents and the Ion program.
6. Add 500 µl of Ion OneTouch™ Wash Solution to each Recovery Tube and pipet up and down to disperse the ISPs.
7. Transfer the suspensions from both tubes to a new labeled 1.5 ml Eppendorf LoBind® Tube. You can now store at 2°-8°C for up to 3 days.
8. Before enrichment, incubate the ISPs at 50°C for 2 minutes, then centrifuge at 15,500 x g for 2.5 minutes.
9. Remove all but 100 µl of supernatant. With a new tip, pipet the solution up and down 10 times to resuspend the ISPs.
10. Transfer the entire volume into Well 1 of the 8-well strip.

Enrich the Template-Positive Ion PGM™ Template OT2 400 Ion Sphere™ Particles with the Ion OneTouch™ ES:

1. Prepare Melt-Off solution by combining 280 µl of Tween® Solution and 40 µl of 1 M NaOH.
2. Vortex the tube of Dynabeads® MyOne™ Streptavidin C1 Beads for 30 seconds, then centrifuge for 2 seconds. Pipet up and down until the dark pellet of beads disperses. Transfer 13 µl to a 1.5 ml Eppendorf LoBind® Tube. Place the tube on a magnet for 2

minutes, then carefully remove and discard the supernatant without disturbing the pellet of Dynabeads® MyOne™ Streptavidin C1 Beads. Add 130 µl of MyOne™ Beads Wash Solution. Remove from the magnet, vortex for 30 seconds, then centrifuge for 2 seconds.

3. Fill the wells of the 8-well strip as follows:
   - Well 1 100 µl of template-positive ISPs
   - Well 2 130 µl of resuspended Dynabeads® MyOne™ Streptavidin C1 Beads
   - Well 3 300 µl of Ion OneTouch™ Wash Solution
   - Well 4 300 µl of Ion OneTouch™ Wash Solution
   - Well 5 300 µl of Ion OneTouch™ Wash Solution
   - Well 6 Empty
   - Well 7 300 µl of Melt-Off solution
   - Well 8 Empty
4. Confirm that the square-shaped tab is on the left, then insert the filled 8-well strip so that it is pushed all the way to the right end of the slot of the Tray.
5. Prepare the Ion OneTouch™ ES by loading a new tip in the Tip Arm. Ensure that the bottom of the Tip Arm is not resting on top of the thumb screw.
6. Add 10 µL of Neutralization Solution to a new 0.2 ml PCR tube and insert the opened tube into the hole in the base of the Tip Loader.
7. Turn ON the Ion OneTouch™ ES and wait for the instrument to initialize. The screen displays “rdy”.
8. Press **Start/Stop**. The screen displays “run” during the run. The run takes ~35 minutes.
9. At the end of the run, the instrument displays “End” and beeps every 60 seconds. Press the **Start/Stop** button to silence this alarm and reset the instrument for the next run.
10. Securely close and remove the PCR tube containing the enriched ISPs. Mix the contents of the PCR tube by gently inverting the tube 5 times.
11. Remove the used tip and 8-well strip.
12. You can proceed on to sequencing.

APPENDIX H

Protocol for Sequencing with the Ion 316™ Chip v2 on the Ion PGM™ Sequencer (Taken from manufacturer's protocol)

Prepare Enriched, Template-Positive ISPs:

1. Vortex the Control Ion Sphere™ Particles and centrifuge for 2 seconds.
2. Add 5 µl of Control ISPs directly to the entire volume of enriched, template-positive ISPs in a 0.2 ml non-polystyrene PCR tube.

Anneal the Sequencing Primer:

1. Mix the contents of the tube by pipetting up and down. Centrifuge for 2 minutes at 15,500 x g.
2. Carefully remove the supernatant without disturbing the pellet, leaving 15 µl in the tube.
3. Add 12 µl of the Sequencing Primer and confirm that the total volume is 27 µl.
4. Pipet up and down to disrupt the pellet.
5. Run on a thermal cycler at 95°C for 2 minutes, then 37°C for 2 minutes, using the heated lid option.

Chip Check:

1. Remove a new chip from its packaging and label it. Save the package.
2. Place the chip on the Ion PGM™ Sequencer grounding plate.
3. Press **Run** on the main menu and follow the touchscreen prompts to prepare to test a new Ion PGM™ Chip.
4. When prompted, ground yourself by touching the grounding pad next to the chip clamp on the instrument and replace the old chip in the chip socket with the new one. Close the chip clamp.
5. When prompted, use the barcode scanner to scan the barcode on the chip package.
6. Pres **Chip Check** on the touchscreen.
7. During the initial part of Chip Check, visually inspect the chip in the clamp for leaks. If there is a leak, press **Abort** to stop the flow to the chip.
8. When chip Check is complete, press **Next** if the chip passes.
9. Following a successful Chip Check, remove the new chip and place it on the grounding plate. Insert a used chip in the socket and close the clamp.
10. Completely empty the waste bottle as instructed in the touchscreen.
11. Proceed immediately to load the chip.

Bind Sequencing Polymerase to the ISPs:

1. After annealing the Sequencing Primer, remove the ISPs from the thermal cycler and add 3 µl of Ion PGM™ Sequencing 400 Polymerase to the ISPs.
2. Pipet the sample up and down to mix, and incubate at room temperature for 5 minutes.

Load the Chip:

1. Tilt the chip 45 degrees so that the loading port is the lower port.
2. Insert the pipette tip firmly into the loading port and remove as much liquid as possible from the loading port. Discard the liquid.
3. Place the chip upside-down in the centrifuge adaptor bucket and transfer the bucket to the Minifuge with the chip tab pointing in.
4. Centrifuge for 5 seconds to completely empty the chip. Remove the chip from the bucket and wipe off any liquid on the bucket.
5. Place the chip back in the centrifuge adaptor bucket and place the bucket on a flat, stable surface.
6. Following polymerase incubation, collect the entire sample (~30 µl) into a pipette tip and insert the tip firmly into the loading port of the chip.
7. Dial down the pipette to gently deposit the ISPs at a rate of ~1 µL per second. To avoid introducing bubbles, leave a small amount of sample in the pipette tip (~0.5 µl).
8. Remove and discard any displaced liquid from the other port of the chip.
9. Transfer the chip to the Minifuge with the chip tab pointing in and centrifuge for 30 seconds, then remove from the centrifuge bucket.
10. Mix the sample in the chip by setting a pipette to 25 µl, tilting the chip 45 degrees so that the loading port is the lower port, inserting the pipette tip into the loading port, and without removing the tip, slowly pipetting the sample in and out of the chip 3 times.
11. Centrifuge for 30 seconds with the chip tab pointing out.
12. Repeat the chip mixing step in step 10, then centrifuge for 30 seconds with the chip tab pointing in.
13. Repeat the chip mixing step in step 10, this time pipetting the sample in and out 5 times.
14. Tilt the chip at a 45-degree angle and slowly remove as much liquid as possible from the loading port by dialing the pipette. Discard the liquid.
15. If some liquid remains in the chip, perform a 5 second quick spin with the chip tab pointing out and remove and discard any additional liquid.
16. If some liquid remains in the chip after the quick spin, lightly and rapidly tap the point of the chip tab against the benchtop a few times, and remove and discard any collected liquid. Do not flush the chip.
17. When chip loading is complete, press **Next** on the touchscreen.

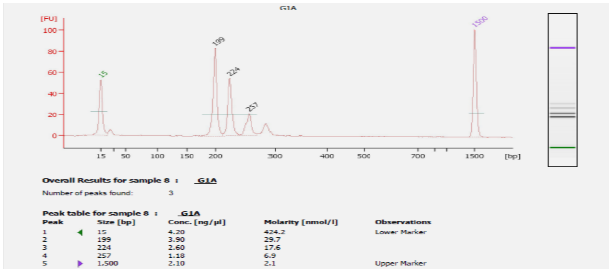
Select the Planned Run and Perform the Run:

1. Press the **Browse** button and select the name of the plan you created, then press **Next**.
2. Confirm the settings are correct and make any changes using the buttons and dropdown lists if necessary. Then press **Next**.
3. When prompted, load and clamp the chip, then press **Next**.
4. Visually inspect the ship for leaks before closing the cover.
5. When calibration is complete, the touchscreen will indicate whether it was successful. If it passes, press **Next** to proceed with the sequencing run.
6. After 90 seconds, the run will automatically begin.
7. When the run is complete, the touchscreen will return to the Main Menu. You can then proceed with another fun or perform cleaning if required.

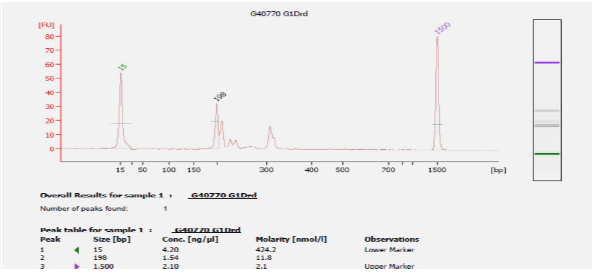
APPENDIX I

Singleplex reaction electropherograms of sample G40770, generated using the Agilent DNA 1000 Kit on the Agilent 2100 Bioanalyzer (Agilent Technologies Inc.).

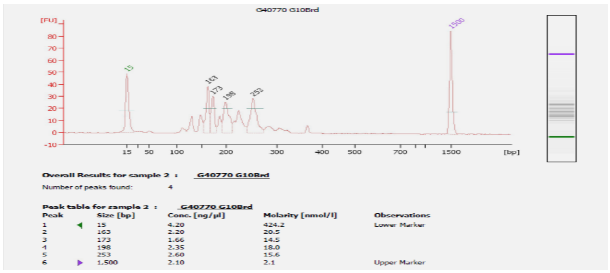
G40770/G1A



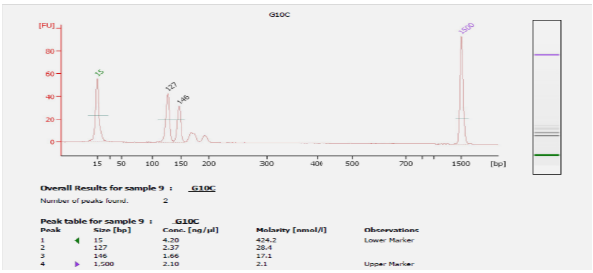
G40770/G1D



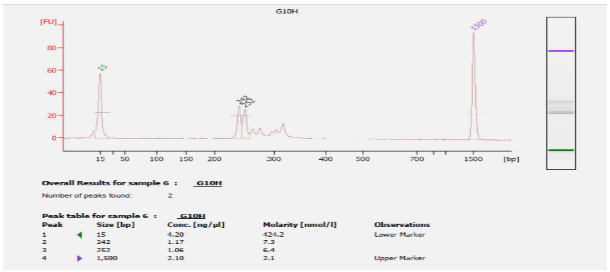
G40770/G10B



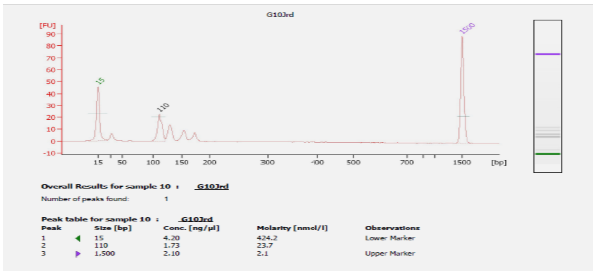
G40770/G10C



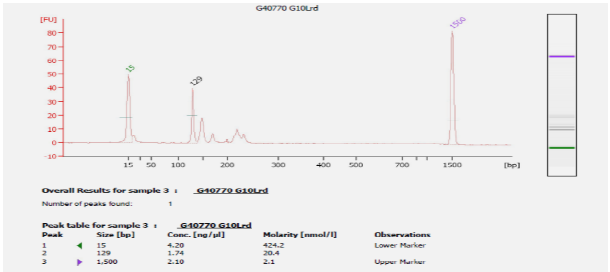
G40770/G10H



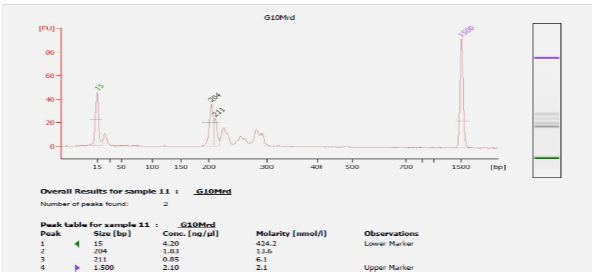
G40770/G10J



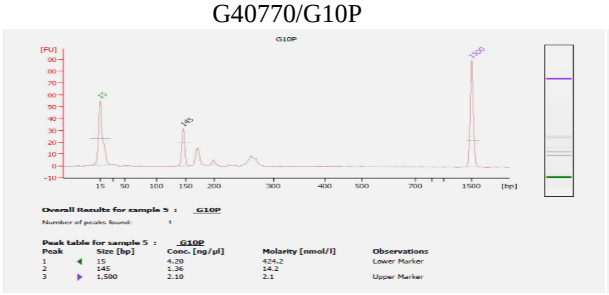
G40770/G10L



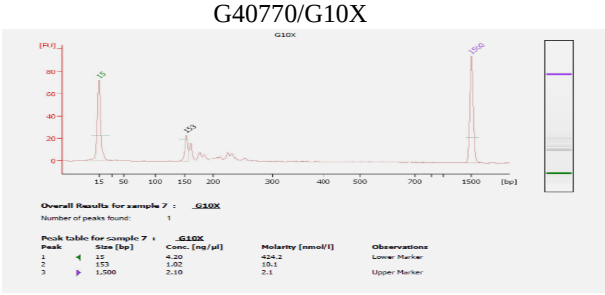
G40770/G10M



G40770/G10P



G40770/G10X



APPENDIX J

Reference Sequences for Ten Ursid STR Loci

G1D:

NCBI: size 187, $(GT)_{17.5}$
GGAAAAGTTTTGGACATTTTCTTCAAATATTTTTTTCCATCTCTCTTTTCCTTTAGGGGACTCCAAG
TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGCATGCATTGAGGTTTCCACTCTTAAAGAGTAG
GAAGAGTAGATTATGTATTATACCTTGCTGGGTGCTAGGTATCAAGTGTTCTTG
New: size 187, $(GT)_{17}$
GGAAAAGTTTTGGACATTTTCTTCAAATATTTTTTTCCATCTCTCTTTTCCTTTAGGGGACTCCAAG
TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTCATGCATTGAGGTTTCCCACTCTTAAAGAGTAG
GAAGAGTAGATTATGTATTATACCTTGCTGGGTGCTAGGTATCAAGTGTTCTTG

G1A:

NCBI: size 182, $(TG)_{19.5}$
GACCCTGCATACTCTCCTCTGATGGGCCAGGCCGGGCCAGGCTGTGTGTGTGTGTGTGTGTGTGT
GTGTGTGTGTGTGTGTATGAGTGATTGAAGTCCTGCTTCTTTTTCACTAATCTCGTCCAGAGGAGA
CTCTATTGGCTTCTATTTCTTCTTCCTGTCACTTCTACGCAAGGACAGTGC
New: size 180, $(TG)_{17.5}$
GACCCTGCATACTCTCCTCTGATGGGCCAGGCCGGGCCAGGCTGTGTGTGTGTGTGTGTGTGTGT
GTGTGTGTGTGTATGAGTGATTGAAGTCCTGCTTCTTTTTCACTAAATCTCGTCCCAGAGGAGACT
CTATTGGCTTCTATTTCTTCTTCCTGTCACTTCTACGCAAGGACAGTGC

G10P:

NCBI: size 161, $(GT)_{21}TT(GT)_6$
AGGAGGAAGAAAGATGGAAAACGYTAAGAAATGWWTGTACTTCACTTGCTCTTTAGACWYCGT
GTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTTTGTGTGTGTGTGTCTGACGATGTAG
CTTAGGCCTTTTCAGAGTATTTCCCCACATGA
New: size 155 $(GT)_{18}TT(GT)_6$
AGGAGGAAGAAAGATGGAAAACGTTAAGAAATGAATGTACTTCACTTGCTTTTTAGACTTCGTGT
GTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTTTGTGTGTGTGTGTCTGACGATGTAGCTTAGGCC
TTTTCAGAGTATTTCCCCACATGA

G10H:

NCBI: size 245, $(TG)_{25.5}(ATTT)_{6.2}$
CAACAAGAAGACCACTGTAATTCTTATCCTTGTTCTYTCTAGTAAGRGRGTGTGTTGTGTTGCGTG
TGTGTGTDTATCCCCCCCCACTCCAGCTCTCTYAAAGAGAKWCTCTCTSTCTCTGAGWYWCGGKA
GATGTATATRARAGASCCMGVTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT
GTGTGTATTTATTTATTTATTTATTTATTTATCCTACTTGGTGGTCTCT
New: size 235, $(GT)_{21}(ATTT)_{6.2}$
CAACAAGAAGACCACTGTAATTCTTATCCTTGTTCTTTCTAGTAAGGTTTTGTTTTGTTTTGCTTGT
TTGTTTTTATCCCCCTTACTCCAGCTTTCTTAAAGATTTTCTCTGTGTCTTTGATTTTCGGTAGTTGA
ATATGATAGGCCTAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTATTTATTTA
TTTATTTATTTATTTATCCTACTTGGTGGTCTCT

G10B:

NCBI: size 150, $(TG)_{21}$
GTTCTGTTGAATTTGGTTTGCTAATATTTTCTTGAGGACTTTTGCATATATGTACATTATGGATGTA
ATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGGTTGTATTGTATGGATGGAGGTT
TCTGTGATTTGTCCTGC
New: size 151, $(TG)_{22}$
GTTCTGTTGAATTTGGTTTGCTAATATTTTCTTGAGGACTTTTGCATATATGTTCATTATGGATGTA
ATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTTGTATTGTATGGATGGAGGT
TTCTGTGATTTGTCCTGC

G10L:

NCBI: size 152, $(TG)T(TG)_{33.5}$
GTACTGATTTAATTCACATTTCCCTAGTTTAATTTGTACTCATTTGTTGTGTGTGTGTGTGTGTGTG
TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTTCGTACTTAGTTCTATACA
ATTTTATCGCATGGGTAGG
New: size 157, $(TG)_{37.5}$
GTACTGATTTAATTCACATTTCCCCAGTTTAATTTGTACTCATTTGTGTGTGTGTGTGTGTGTGTGT
GTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTTCGTACTTAGTTCTA
TACAATTTTATCGCATGGGTAGG

G10X:

NCBI: size 150, $(TG)_{20.5}TTC(TG)_7$
CCCTGGTAACCACAAATCTCTTTTTCTATGAGTTTAGTAGGTGCTCTYTTGTGTGTGTGTGTGTGT
GTGTGTGTGTGTGTGTGTGTGTGTTTCTGTGTGTGTGTGTGTTTGTTTCTTTTTTTGTTTGTTTTGAT
TTCACAGATAACTGAG
New: size 145, $(TG)_{18.5}TTC(TG)_7$
CCCTGGTAACCACAAATCTCTTTTTCTATGAGTTTAGTAGGTGCTCTTTTGTGTGTGTGTGTGTGTG
TGTGTGTGTGTGTGTGTGTTTCTGTGTGTGTGTGTGTTTGTTTCTTTTTTGTTTGTTTTGATTTCACA
GATAACTGAG

G10J:

NCBI: size 98, $(TG)_{19.5}$
GATCAGATATTTTCAGCTTTTGTGTGTGTTTTTGCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT
GTGTGTGTTGAAGTGGAGTGTGAGGGGTTAT
New:
Unchanged from original reference

G10C:

NCBI: size 113, $(TG)_{21.5}$
AAAGCAGAAGGCCTTGATTTCCTGAAATAGGTCTATATTGTGTGTGTGTGTGTGTGTGTGTGTGTG
TGTGTGTGTGTGTGTAAAACACCTGCTGTCTCGGTGTTTATGTCCCC
New:
Unchanged from original reference

G10M:

NCBI: size 194, $(GT)_{21}$
CCTCATCGTAGGTTGTATTTTCCTCCTTCTTTGTATACCTAGTAATTTTTCATTAAATGGCAGATAT
TATCAAGTTTGCCTCTTTGKCTACTGGATATATWTATATTGTKTTAATACTCTTGGTGTGTGTGTG
TGTGTGTGTGTGTGTGTGTGTGTGTGTGTTTCCCCCTGGGATGCATTTAAATTATTTGG
New: size 194, $(GT)_{21}$
CCTCATCGTAGGTTGTATTTTCCTCCTTCTTTGTATACCTAGTAATTTTTCATTAAATGGCAGATAT
TATCAAGTTTGCCTCTTTGGCTACTGGATATTTTTATATTGTTTTAATACTCTTGGTGTGTGTGTGT
GTGTGTGTGTGTGTGTGTGTGTGTGTGTGTTTCCCCCTGGGATGCATTTAAATTATTTGG

IUPAC codes:

M	A or C
V	A or C or G
S	G or C
K	G or T
Y	C or T
W	A or T
R	A or G
D	A or G or T
H	A or C or T
B	C or G or T

REFERENCES

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *Journal of Molecular Biology,* **215(3)**, 403-410.

Andreassen R, Schregel J, Kopatz A, et al (2012) A forensic DNA profiling system for Northern European brown bears (Ursus arctos). *Forensic Science International-Genetics,* **6**, 798-809.

Brown SK, Hull JM, Updike DR, Fain SR, Ernest HB (2009) Black bear population genetics in California: signatures of population structure, competitive release, and historical translocation. *Journal of Mammalogy,* **90**, 1066-1074.

Bull RAS, Cushman SA, Mace R, et al (2011) Why replication is important in landscape genetics: American black bear in the Rocky Mountains. *Molecular Ecology,* **20**, 1092-1107.

Coghlan ML, Haile J, Houston J, et al (2012) Deep sequencing of plant and animal DNA contained within Traditional Chinese Medicines reveals legality issues and health safety concerns. *Plos Genetics,* **8**, 436-446.

Craighead L, Paetkau D, Reynolds HV, Vyse ER, Strobeck C (1995) Microsatellite analysis of paternity and reproduction in Arctic grizzly bears. *Journal of Heredity,* **86**, 255-261.

Curtis M (2013) Bear cases. **USFWS genetics lab case information**.

De Barba M, Waits LP, Garton EO, et al (2010) The power of genetic monitoring for studying demography, ecology and genetics of a reintroduced brown bear population. *Molecular Ecology,* **19**, 3938-3951.

Dixon JD, Oli MK, Wooten MC, Eason TH, McCown JW, Cunningham MW (2007) Genetic consequences of habitat fragmentation and loss: the case of the Florida black bear (Ursus americanus floridanus). *Conservation Genetics*, **8**, 455-464.

Dlugosch KM, Parker IM (2008) Founding events in species invasions: genetic variation, adaptive evolution, and the role of multiple introductions. *Molecular Ecology,* **17**, 431-449.

Guichoux E, Lagache L, Wagner S, et al (2011) Current trends in microsatellite genotyping. *Molecular Ecology Resources,* **11**, 591-611.

Gunderson A (2009) University of Michigan Animal Diversity Web: Ursus maritimus, http://animaldiversity.ummz.umich.edu/accounts/Ursus_maritimus. **2013**.

Gymrek M, Golan D, Rosset S, Erlich Y (2012) lobSTR: A short tandem repeat profiler for personal genomes. *Genome Research,* **22**, 1154-1162.

Hedrick PW, Ritland K (2012) Population genetics of the white-phased "spirit" black bear of British Columbia. *Evolution,* **66**, 305-313.

IUCN (2013) Red list of threatened species, http://www.iucnredlist.org/. **2013**.

IUCN Polar Bear Specialist Group (2009) 15th meeting of PBSG in Copenhagen, Denmark 2009: Press Release.

Jeffreys AJ, Wilson V, Thein SL (1985) Hypervariable minisatellite regions in human DNA. *Nature,* **314(7)**, 67-73.

Jonkel C (2005) Internet Center for Wildlife Damage Management: Grizzly/Brown Bears, http://icwdm.org/handbook/carnivor/GrizzlyBears.asp. **2013**.

Karamanlidis AA, Straka M, Drosopoulou E, et al (2012) Genetic diversity, structure, and size of an endangered brown bear population threatened by highway construction in the Pindos Mountains, Greece. *European Journal of Wildlife Research,* **58**, 511-522.

Kocijan I, Galov A, Cetkovic H, Kusak J, Gomercic T, Huber D (2011) Genetic diversity of Dinaric brown bears (*Ursus arctos*) in Croatia with implications for bear conservation in Europe. *Mammalian Biology,* **76**, 615-621.

Krause J, Unger T, Nocon A, et al (2008) Mitochondrial genomes reveal an explosive radiation of extinct and extant bears near the Miocene-Pliocene boundary. *Bmc Evolutionary Biology,* **8**, 220.

Lacy RC (1997) Importance of genetic variation to the viability of mammalian populations. *Journal of Mammalogy,* **78**, 320-335.

Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology,* **10**, R25.

Life Technologies (2011) Ion Torrent: Amplicon Sequencing Application Note.

Liu L, Li Y, Li S, et al (2012) Comparison of next-generation sequencing systems. *Journal of Biomedicine and Biotechnology,* 251364.

Lowe CL (2011) *Estimating populations parameters of the Louisiana black bear in the upper Atchafalaya river basin.* Master of Science, The University of Tennessee, Knoxville.

Manitoba Conservation and Water Stewardship (2013) Species At Risk: Polar Bear Fact Sheet, http://www.gov.mb.ca/conservation/wildlife/sar/fs/pbear.html. **2013**.

Manitoba Conservation and Water Stewardship (2013) Wild Animals of Manitoba: Black Bear Fact Sheet, http://www.gov.mb.ca/conservation/wildlife/mbsp/fs/blbear.html. **2013**.

Miller CR, Waits LP (2003) The history of effective population size and genetic diversity in the Yellowstone grizzly (*Ursus arctos*): implications for conservation. *Proceedings of the National Academy of Sciences,* **100**, 4334-4339.

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

Ohnishi N, Saitoh T, Ishibashi Y, Oi T (2007) Low genetic diversities in isolated populations of the Asian black bear (*Ursus thibetanus*) in Japan, in comparison with large stable populations. *Conservation Genetics*, **8**, 1331-1337.

Paetkau D, Amstrup SC, Born EW, et al (1999) Genetic structure of the world's polar bear populations. *Molecular Ecology*, **8**, 1571-1584.

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

Paetkau D, Shields GF, Strobeck C (1998b) Gene flow between insular, coastal and interior populations of brown bears in Alaska. *Molecular Ecology*, **7**, 1283-1292.

Paetkau D, Strobeck C (1994) Microsatellite analysis of genetic variation in black bear populations. *Molecular Ecology*, **3**, 489-495.

Paetkau D, Waits LP, Clarkson PL, et al (1998) Variation in genetic diversity across the range of North American brown bears. *Conservation Biology*, **12**, 418-429.

Quail MA, Smith M, Coupland P, et al (2012) A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *Bmc Genomics*, **13**, 341.

Reed DH, Frankham R (2003) Correlation between fitness and genetic diversity. *Conservation Biology*, **17**, 230-237.

Skrbinsek T, Jelencic M, Waits LP, Potocnik H, Kos I, Trontelj P (2012) Using a reference population yardstick to calibrate and compare genetic diversity reported in different studies: an example from the brown bear. *Heredity*, **109**, 299-305.

Taberlet P, Camarra JJ, Griffin S, et al (1997) Noninvasive genetic tracking of the endangered Pyrenean brown bear population. *Molecular Ecology*, **6**, 869-876.

U.S. Fish and Wildlife Service (2013) Endangered Species list, http://www.fws.gov/endangered. **2013**.

U.S. Fish and Wildlife Service (2011) Forensics Laboratory, http://www.fws.gov/lab. **2013**.

Waits L, Paetkau D, Strobeck C (1999) Chapter 3: Genetics of the Bears of the World. In: *Status Survey and Conservation Action Plan: Bears* (eds. Servheen C, Herrero S, Peyton B) pp. 25-32 IUCN, Cambridge, U.K.

Waits L, Taberlet P, Swenson JE, Sandegren F, Franzen R (2000) Nuclear DNA microsatellite analysis of genetic diversity and gene flow in the Scandinavian brown bear (*Ursus arctos*). *Molecular Ecology*, **9**, 421-431.

Walsh PS, Fildes NJ, Reynolds R (1996) Sequence analysis and characterization of stutter products at the tetranucleotide repeat locus vWA. *Nucleic Acids Research*, **24**, 2807-2812.

Warshauer DH, Lin D, Hari K, et al (2013) STRait Razor: a length-based forensic STR allele-calling tool for use with second generation sequencing data. *Forensic Science International-Genetics*, **7**, 409-417.

Williamson DF (2002) *In the Black: Status, Management, and Trade of the American Black Bear (Ursus americanus) in North America* TRAFFIC North America, World Wildlife Fund, Washington D.C.