Graham, Michelle L., Discovery and Characterization of Tetranucleotide Short Tandem Repeats in North American Bears (Ursids). Master of Science (Forensic Genetics), April 2015, 82 pg., 8 tables, 11 figures, 54 references.

Accurate individual identification is essential to wildlife crime investigations and conservation genetics. Current methodology utilizes dinucleotide short tandem repeats (STRs) that can be difficult to type accurately and have high mutation rates; however, tetranucleotide STRs have greater stability and allele diversity. The main objective of this study was to identify potential tetranucleotide STR loci and internal variants for the American black bear, brown bear, and polar bear. Barcoded genome libraries were prepared for each species from extracted and enzymatically fragmented DNA, size selected, quantified, enriched using biotinylated RNA baits to capture twelve common mammalian sequence motifs, and massively parallel sequenced. One potential locus was identified using the NextGENe® software and six potential loci were identified using algorithm mining.

DISCOVERY AND CHARACTERIZATION OF TETRANUCLEOTIDE SHORT TANDEM REPEATS IN NORTH AMERICAN BEARS (URSIDS)

Michelle Graham, B.S.

APPROVED:

_____

Major Professor

_____

Committee Member

_____

Committee Member

_____

University Member

_____

Chair, Department of Molecular and Medical Genetics

_____

Dean, Graduate School of Biomedical Science

DISCOVERY AND CHARACTERIZATION OF TETRANUCLEOTIDE SHORT TANDEM REPEATS IN NORTH AMERICAN BEARS (URSIDS)

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

Michelle Graham, B.S.

Fort Worth, TX

April 2015

ACKNOWLEDGEMENTS

I would like to thank my major professor, Dr. John Planz, for his guidance and patience throughout this entire process. Without his insight, I would never have worked through the hectic details of this project. I would like to thank Dr. Michael Allen for his expertise on next-generation sequencing and his input when things started getting messy. I would like to thank the US Fish and Wildlife Forensic Laboratory for contributing the samples used for this study. Next, I would like to thank Shantanu Shewale for assisting with the library preparation optimization and for keep things in the lab interesting. To my FGEN girls, thank you for making the last two years truly amazing; I don't think I could have made it through without your constant support. Finally, to my friends and family: thank you for always believing in my dreams, even when they seemed impossible.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

Chapter I:

Introduction

Short tandem repeats, also known as STRs or microsatellites, are repetitive DNA sequences containing unique core repeat units of 2-6 nucleotides [1, 2]. The use of STRs is well established in human forensic casework, with tetranucleotide repeat motifs being the preferred markers in sample analysis. Tetranucleotide repeat regions in DNA are characterized by high variability and co-dominance with the occurrence of stutter and heterozygosity balance decreasing with increased motif size [1]. The length variation of these repeat motifs allows for a number of alleles for each locus, producing a discrete genotype or "DNA fingerprint" that can be used for the genetic identification of individuals.

Similar approaches have been applied to wildlife genetic studies. The resulting information can be used for a wide range of purposes from studying the evolutionary history of a species to mapping out the genetic predisposition of a disease phenotype in an individual [3, 4]. Areas such as population data [5], gene flow [6], migration patterns [7, 8], and wildlife crime investigations [9] relating to conservation genetics and wildlife forensics are studied using the most recent technology. STR identification is specifically used for conservation efforts such as translocation [8], monitoring population size and variation [5, 6, 8, 10-12], and investigations into illegal hunting and other wildlife crime investigations [9, 13, 14]. Gene flow between populations [6], level of species inbreeding [5], and the identification of lineage sorting [15] are all studied to ensure the survival of wildlife species, including those in the family Ursidae, or bears.

The mammalian family Ursidae consists of eight members, distributed across the globe. Over the past few decades, numerous studies including allozyme markers [16], mitochondrial DNA (mtDNA) sequence analysis and restriction enzyme analysis [17-19], and dinucleotide DNA fingerprinting [10-12] have been performed to determine the phylogenetic relationship between each ursine family member. Three members of Ursidae reside in North America: the American black bear (*Ursus americanus*), the brown (grizzly) bear (*Ursus arctos*), and the polar bear (*Ursus maritimus*). Each species possesses distinguishable differences in morphology such as body size and coat color as well as ecological differences in habitat and range.

North American brown bears occupy the Holarctic, the range from northern Canada to as far south as Mexico. Several subspecies of brown bears exist within this region: brown bears of the Kodiak Archipelago, the Kodiak bear (*U. a. middendorffi*) [6]; subpopulations along coastal Alaska and Canada, the Alaskan brown bear (*U. a. alascensis*) [20]; and brown bears of the interior regions of North America, the grizzly bear (*U. a. horribilis*) [20]. Figure 1 shows the current habitat range for brown bears in North America.

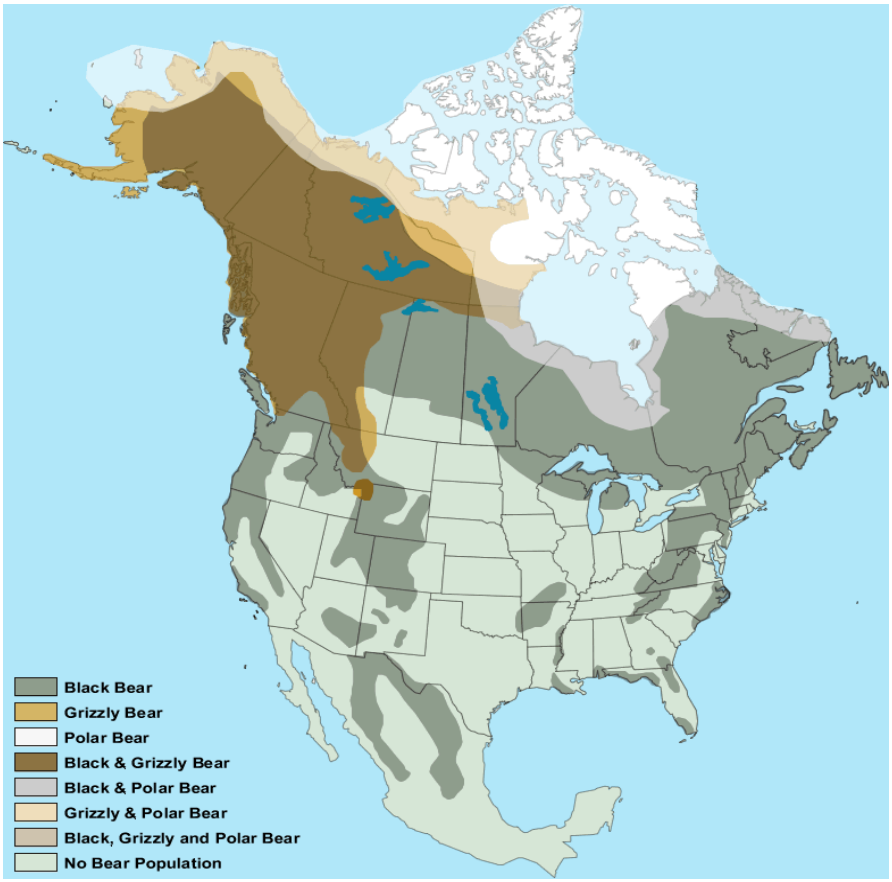


Figure 1: Current ranges for the brown (grizzly) bear, American black bear, and polar bears in North America [21]

The number and distribution of the brown bear has declined worldwide by >50% since the mid-1800s, with the grizzly bear inhabiting less than one percent of their historic habitat south of the Canadian border [6, 22]. Grizzly bears are currently located within the Yellowstone and Northern Continental Divide ecosystems [20]. A few brown bear subpopulations, notably those of the Kodiak and Yellowstone bears, are physically isolated from other subpopulations. This isolation, along with habitat destruction and overexploitation by humans, results in a loss of genetic diversity caused by inbreeding depression due to reduced gene flow [6, 20, 22]. In order to reduce the extinction probability of brown bears in North America, conservation efforts are essential to maintaining the isolated populations [20, 22].

The American black bear is the most widespread bear species in North America, spanning from the northernmost regions of Canada to southern United States and Mexico and inhabiting 41 states and 11 Canadian provinces (see Figure 1) [21, 23]. Like the brown bear, several subspecies of American black bears exist across their habitat: the Kermode bear in British Columbia (*U. a. kermodei*), the Cinnamon black bear of the western United States (*U. a. cinnamomum*), the Louisiana black bear (*U. a. luteolus*), and the Florida black bear (*U. a. floridanus*), among many others [23, 24]. Although the American black bear as a species are listed as "Least Concern," there are subspecies listed as "Threatened" [25]. The Louisiana black bear is listed as "Threatened" for east Texas, Louisiana, and Mississippi. Kermode bears is a rare genetic variant that results in paler coat, skin, and eye color [23, 24]. These subspecies and other threatened subpopulations exhibit heavy fragmentation, and have been subjected to human exploitation, and physical isolation. In addition, black bears are also threatened by poaching for brain, spinal cord, gallbladder, paws, claws, meat, and bone [23]. Due to declining Asiatic black bear populations, demand for their parts is being substituted through the illegal poaching of North American black bears [23]. All of these factors combined result in a similar loss of genetic diversity as seen in brown bears, requiring conservation efforts for isolated population and improved investigational techniques for wildlife crimes.

Polar bears exist in the ice-covered waters of the circumpolar Arctic (see Figure 1), remaining predominantly philopatric to particular areas due to the distribution of their primary prey of seals [11, 26]. Nineteen recognized subpopulations exist worldwide with current polar bear numbers are estimated to be between 20,000-25,000 individuals [26, 27]. These low numbers in combination with low reproductive rates and poaching leave polar bears particularly vulnerable to overharvesting and the effects of climate and human-based changes to their habitat [26]. Analysis

of genetic data show that as a result of high mobility of individuals and gene flow, there is little to no genetic subdivision between adjacent subpopulations and moderate subdivision across larger regions of the Arctic [28]. Because of these factors, polar bears are classified as "Vulnerable" by both the International Union for Conservation of Nature and Natural Resources and the World Wildlife Foundation [27, 29] and "Threatened by United States Fish and Wildlife Service [30], making the need for conservation and management plans critical for maintenance of this species.

The limited amount of the genetic data that exists for bear species have been generated from dinucleotide STR genotyping [5-7, 12] and mitochondrial DNA sequencing [31, 32]. Dinucleotide STRs have short amplicon sizes of two nucleotides that allow for their use on samples such as hair and feces [5-7, 12, 14] and have undergone validation and database building, making them the currently used method for individual and species identification. Although tetranucleotide STRs are the preferred markers for forensic casework, dinucleotide STRs are the ones widely used in wildlife forensics. For wildlife DNA profiling to update to the practices currently used in human forensics, new STRs based on tetranucleotide repeats must be identified.

STR typing relies upon differences in repeated DNA sequences present within eukaryotic genomes at numerous locations or loci [38]. STR markers are characterized by the type of repeat unit present and the number of repeats the locus contains. Simple repeat motifs are repeat units of identical length and sequence, compound repeat motifs are comprised of two or more simple repeat units, and complex repeat motifs contain several repeat blocks of variable unit length and structure with variable intervening sequences [39]. Allele designations are made for the three repeat motifs by identifying the number of repeat units at a given locus [38, 39]. Variation in the number of repeat units in STR markers occurs by slippage of the DNA polymerase enzyme during DNA replication [38]. Non-consensus alleles, those containing variations in the repeat motif for that

locus, are referred to as microvariants and are named for the number of full repeat units and the portion of the partial repeat unit. An example of this is the 9.3 microvariant at the human TH01 locus, which possesses nine full tetranucleotide repeats and an incomplete repeat unit of three nucleotides [38].

The current practices for STR typing are multi-locus PCR amplification of extracted DNA followed by the size-separation method of capillary electrophoresis (CE) [2]. However, the main limitation to CE analysis of STRs is that only allele length can be determined; this limits the quantitative value of STR, as CE is unable to detect variation within an STR repeat motif [33]. In recent years, however, improvements in DNA sequencing have advanced dramatically to the point that such technologies can be used not only for the study of human genetics but also for forensic casework such that these individual base variations can be identified [1, 34-40].

One of the more common deep-sequencing techniques used is inter-simple sequence repeat PCR (ISSR-PCR) [34]. ISSR-PCR uses PCR primers containing simple repeat probes that bind to regions thought to be abundant in microsatellites [1, 34]. The PCR product is subsequently amplified and sequenced to identify microsatellite sequences [34]. More recent methods use probe adaptor or RNA bait hybridization targeted for flanking regions of microsatellite sequences [1, 35-37]. During hybridization, fragmented DNA undergoes ligation of adaptors or RNA baits, PCR amplification, and sequencing [35-37]. Small DNA fragment size (100-1000 bp) is essential to ensure optimal enrichment product for sequencing [35-37].

In post-sequencing data analysis, one common tool used is Bowtie [40], a short read aligner that uses an *in silico* reference genome and offers high performance with a relatively low computational cost. However, the limitation to this software is that Bowtie cannot characterize novel alleles or allele variants and has no insertion/ deletion (indel) tolerance, requiring further

characterization using additional sequencing software [40]. Another software, lobSTR [41], conducts STR profiling by performing algorithm-based STR allele identification following alignment by software such as Bowtie. While lobSTR is an accurate method for detecting STR alleles, it requires a reference sequence to perform allele characterization and can only identify a simple repeat motif using the default settings; the identification of complex repeat motifs requires some additional manipulation of the software, making lobSTR a less applicable approach for locus discovery and characterization [41]. STRait Razor [42], a Perl-based software tool that identifies STR alleles based on allele length, combines a conceptually similar approach to CE with the ability to identify any changes to the repeat motifs of forensically relevant STR loci [42]. However, in order to use any of the aforementioned software for STR locus discovery, a reference genome is required.

The NextGENe® software (SoftGenetics®, State College, PA) uses another approach to post-sequencing data analysis. The NextGENe® platform uses a BLAST-Like Alignment Tool (BLAT) method and can be used to identify single nucleotide polymorphisms (SNPs), indels, and the de novo assembly and discovery of short tandem repeat loci [43]. NextGENe® analyzes data from different sequencing platforms such as the Ion Torrent (Life Technologies™, Carlsbad, CA) Personal Genome Machine® (PGM™) and the MiSeq sequencer (Illumina, Inc., San Diego, CA) with similar accuracy [44].

In this study, it was hypothesized that tetranucleotide short tandem repeat (STR) loci can be discovered and characterized in bears using next-generation sequencing and that these tetranucleotide STRs can be used for species and individual identification of the three primary bears species in North America: the American black bear, the brown (Grizzly) bear, and the polar bear. The main objective of this study was to perform a preliminary identification of STR loci for

each bear species. Further characterization of individual locus alleles within and between species as well as allelic frequencies will be conducted in future studies.

Chapter II

Materials and Methods

Specimens Examined:

Tissue samples from four brown (Grizzly) bears (Table 1) and four polar bears (Table 1) were donated by the United States Fish and Wildlife Forensic Laboratory (Ashland, OR); these samples were extracted using an organic phenol chloroform-isoamyl alcohol (PCIA) protocol (Appendix A) with ethanol precipitation. Each species was extracted separately to prevent species cross-contamination. Four previously extracted American black bear samples (Table 1) [45] were included in downstream analyses. Extracts were quantified using a Qubit® 2.0 Fluorometer (Life Technologies™, Carlsbad, CA; Appendix B) to determine DNA concentration in sample extracts.

Table 1: Brown bear, American black bear, and polar bear tissue samples used in this study

Species	**Sample Number**			
Brown (Grizzly) Bear	B20917	9010hH	D40120	H30201
American Black Bear	G40770	H30863	H30413	Bear 8/9
Polar Bear	M302_15	M304_41	M30304	M302_57

Library Preparation:

DNA extracts from the bears were used to prepare samples for next-generation sequencing on the Ion Torrent™ Personal Genome Machine® (PGM™) Sequencer® (Life Technologies™, Carlsbad, CA). To optimize library preparation, only brown bear samples were utilized; polar bear and American black bear samples will be interrogated in future studies. Library preparation was performed using the NEBNext® Fast DNA Fragmentation and Library Prep Set for Ion Torrent™ (New England Biolabs®, Inc., Ipswich, MA; for protocol see Appendix D). Enzymatic fragmentation was optimized to produce fragments of 300-400 base pairs in length for downstream analysis on the PGM™. Each fragmentation reaction included the following reagent volumes: variable volumes of extracted DNA for a total input of 1 μg, NEBNext® DNA Fragmentation Reaction Buffer (2 μL), NEBNext® DNA Fragmentation Master Mix (1.5 μL), and a variable volume of molecular grade water for a total reaction volume of 20 μL. Tests were performed to determine optimal thermal cycler parameters for fragmentation. Manufacturer recommended thermal cycling parameters called for 20 minutes at 25°C for enzymatic activation and 10 minutes at 70°C for enzymatic deactivation. For 300-400 base pairs, several incubation times were tested: 3 minutes, 4.5 minutes, 6 minutes, 8 minutes, 10 minutes, and 12.5 minutes. Optimal parameters were determined to be 12.5 minutes at 25°C followed by 10 minutes at 70°C (Appendix D). End repair of the fragmented DNA was performed concurrently by enzymes in the fragmentation reaction mix. Sizing quality of fragmented DNA was assessed with the Agilent High Sensitivity DNA Kit (Agilent Technologies, Inc. Santa Clara, CA) on the Agilent 2100 Bioanalyzer (Agilent Technologies, Inc.; see Appendix C for protocol).

Following fragmentation, NEXTflex™ DNA barcode adaptors (Bioo Scientific™ Corp., Austin TX), approximately 50 base pairs in length, were ligated to the fragmented DNA. The

ligation reaction was performed using reagents from the NEBNext® Fast DNA Fragmentation and Library Prep Set for Ion Torrent™ kit. The following reagents and volumes were added to the fragmentation reaction for a total ligation reaction volume of 50 µL: molecular grade water (3 µL), T4 DNA Ligase Buffer (10 µL), NEXTflex™ DNA Barcode Adaptor (5 µL; 18-29, one per sample), NEXTflex™ DNA P1 Adaptor (5 µL), *Bst* 2.0 WarmStart DNA Polymerase (1 µL), and T4 DNA Ligase (6 µL). Sequences for all DNA barcode adaptors and P1 DNA adaptor are listed in Table 2. The ligation reaction was incubated on a thermal cycler for 15 minutes at 25°C followed by 5 minutes at 65°C (Appendix D).

Table 2: DNA barcode adaptor and DNA P1 adaptor sequences used in this study

NEXTflex™	**Sequence[a]**
P1 Adapter	5' CCACTACGCCTCCGCTTTCCTCTCTATGGGCAGTCGGTGAT
	5' ATCACCGACTGCCCATAGAGAGGAAAGCGGAGGCGTAGTGGTT
A18 Adapter	**5' CCATCTCATCCCTGCGTGTCTCCGACTCAGAGGCAATTGCGAT**
	5' ATCGCAATTGCCTCTGAGTCGGAGACACGC
A19 Adapter	**5' CCATCTCATCCCTGCGTGTCTCCGACTCAGTTAGTCGGACGAT**
	5' ATCGTCCGACTAACTGAGTCGGAGACACGC
A20 Adapter	**5' CCATCTCATCCCTGCGTGTCTCCGACTCAGCAGATCCATCGAT**
	5' ATCGATGGATCTGCTGAGTCGGAGACACGC
A21 Adapter	**5' CCATCTCATCCCTGCGTGTCTCCGACTCAGTCGCAATTACGAT**
	5' ATCGTAATTGCGACTGAGTCGGAGACACGC
A22 Adapter	**5' CCATCTCATCCCTGCGTGTCTCCGACTCAGTTCGAGACGCGAT**
	5' ATCGCGTCTCGAACTGAGTCGGAGACACGC
A23 Adapter	**5' CCATCTCATCCCTGCGTGTCTCCGACTCAGTGCCACGAACGAT**
	5' ATCGTTCGTGGCACTGAGTCGGAGACACGC
A24 Adapter	**5' CCATCTCATCCCTGCGTGTCTCCGACTCAGAACCTCATTCGAT**
	5' ATCGAATGAGGTTCTGAGTCGGAGACACGC
A25 Adapter	**5' CCATCTCATCCCTGCGTGTCTCCGACTCAGCCTGAGATACGAT**
	5' ATCGTATCTCAGGCTGAGTCGGAGACACGC
A26 Adapter	**5' CCATCTCATCCCTGCGTGTCTCCGACTCAGTTACAACCTCGAT**
	5' ATCGAGGTTGTAACTGAGTCGGAGACACGC
A27 Adapter	**5' CCATCTCATCCCTGCGTGTCTCCGACTCAGAACCATCCGCGAT**
	5' ATCGCGGATGGTTCTGAGTCGGAGACACGC
A28 Adapter	**5' CCATCTCATCCCTGCGTGTCTCCGACTCAGATCCGGAATCGAT**
	5' ATCGATTCCGGATCTGAGTCGGAGACACGC
A29 Adapter	**5' CCATCTCATCCCTGCGTGTCTCCGACTCAGTCGACCACTCGAT**
	5' ATCGAGTGGTCGACTGAGTCGGAGACACGC

[a]: Underlined portion of barcode sequence is unique to that individual barcode

Size Selection:

Selecting fragment sizes of 300-400 bp while maximizing DNA retention is essential for library enrichment. To this purpose, two size selection methods were tested: a bead-based dual bead size selection and silica-based column double size selection. The modified column size selection showed the highest DNA retention while selecting for the desired fragment size and was therefore used for subsequent species preparation.

Proper fragment size distribution was assessed using the Agilent High Sensitivity DNA kit on the Agilent 2100 Bioanalyzer. Size selected DNA was amplified using the manufacturer recommended conditions for five cycles (Appendix D). Ligated DNA was quantified using the Qubit® 2.0 Fluorometer; dilutions of the prepared libraries were made to normalize samples to a range of 29.4 ng/ μL to 145 ng/ μL if necessary. Each species was pooled in equimolar concentration for library enrichment for a target input range of 100-500 ng.

Library Enrichment and Sequencing

Library enrichment was performed using a kit provided by MYcroarray (Ann Arbor, MI). For enrichment of the pooled species, three master mixes were prepared: a Library master mix, a Hybridization master mix, and a Capture Baits master mix (Appendix F). Sequences for the biotinylated RNA baits are included in Table 3.

Table 3: Sequences of the twelve biotinylated RNA baits used in this study

RNA Bait	Sequence
>TNRSTR-1	TGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTC TGTCTGTCTGTCTGTCTGTCTGTC
>TNRSTR-2	TCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTA TCTATCTATCTATCTATCTATCTA
>TNRSTR-3	TATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATC TATCTATCTATCTATCTATCTATC
>TNRSTR-4	CATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATT CATTCATTCATTCATTCATTCATT
>TNRSTR-5	TCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTG TCTGTCTGTCTGTCTGTCTGTCTG
>TNRSTR-6	TCCATCCATCCATCCATCCATCCATCCATCCATCCATCCATCCATCC ATCCATCCATCCATCCATCCATCCA
>TNRSTR-7	ATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCT ATCTATCTATCTATCTATCTATCT
>TNRSTR-8	TTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCT TCCTTCCTTCCTTCCTTCCTTCC
>TNRSTR-9	TTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT CTTTCTTTCTTTCTTTCTTTCT
>TNRSTR-10	CCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTC CTTCCTTCCTTCCTTCCTTCCTT
>TNRSTR-11	CTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT TTCTTTCTTTCTTTCTTTCTTT
>TNRSTR-12	TGTATGTATGTATGTATGTATGTATGTATGTATGTATGTATGTATGT ATGTATGTATGTATGTATGTATGTA

Each master mix was incubated as followed: Library master mix at 95°C for five minutes; Hybridization master mix at 65°C for three minutes, while keeping Library master mix in the thermalcycler; and Capture Baits master mix at 65°C for two minutes, while all additional master mixes in the thermalcycler. Library master mix (7 μL) and Hybridization master mix (13 μL) was transferred to the Capture Baits master mix and mixed via pipetting. The pooled Capture Solution was incubated at 65°C for 36 hours. The entire hybridization process is outlined in Figure 2.

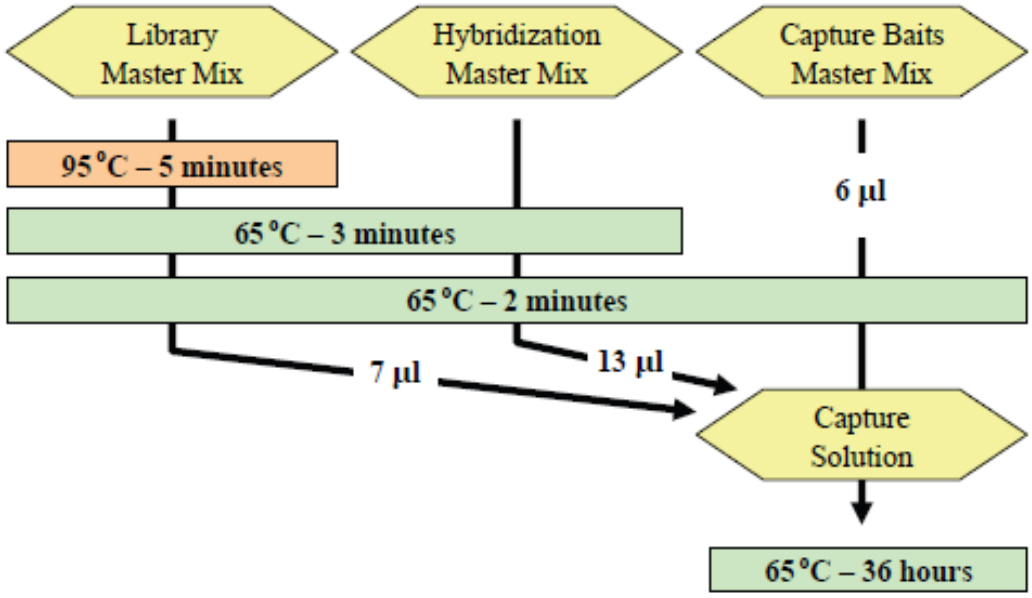


Figure 2: Hybridization process of library enrichment using the MYcroarray kit [46].

Following hybridization, DNA bound to RNA baits was recovered by means of streptavidin-coated magnetic beads and eluted using a 100 mM NaOH solution (Appendix F), followed by cleanup using the QIAquick PCR Purification kit (Qiagen; Appendix G), a post-capture amplification (Appendix F), and an additional QIAquick purification to remove excess primers and reagents. Enriched library quality was assessed via Agilent 2100 Bioanalyzer (Agilent Technologies; Appendix C) and quantified using a Qubit® 2.0 Fluorometer (Life Technologies™; Appendix B) prior to sequencing on the Ion Torrent™ Personal Genome Machine® (PGM™).

Pooled, enriched libraries were diluted to 26 pM per manufacturer's recommendation and underwent emulsion PCR on the Ion OneTouch™ System (Life Technologies™; Appendix H) using Ion Sphere™ Particles (ISPs). The library was further enriched on the Ion OneTouch™ ES (Life Technologies™) and injected onto the Ion Torrent™ Personal Genome Machine® (PGM™) System 316™ Chip v2 (Life Technologies™) for sequencing (Appendix I).

Data Analysis

BAM data files for all PGM™ sequencing reads were uploaded into the NextGENe® software v2.4.0.1 (SoftGenetics®, State College, PA) and converted to FASTA files. Due to the lack of reference sequence for the ursid species and smaller fragment sizes resulting from RNA baiting enrichment, a *de novo* assembly was performed using a Floton assembly method. This method was chosen to reduce the number of homopolymer errors to produce accurate assembly results. STR repeat motifs were characterized by searching the sequence data for the twelve RNA bait core repeat motifs with at least three repeats in each sequence. Allele calls for each individual sample within a species was determined by counting the number of reads possessing the same number of repeats at a given locus. Heterozygous genotypes, those with two alleles per locus, were characterized as alleles with approximately the same number of reads per repeat motif. A consensus sequence was developed for each STR locus by analyzing and recording single nucleotide polymorphisms (SNPs), insertions and deletions (INDELs), and any additional repeat blocks of variable unit length and structure with variable intervening sequences, a characteristic of complex repeat motifs.

In addition to the NexGENe® analysis, a computer algorithm was developed to mine for the reverse complement of the RNA bait sequences for three, seven, nine, and twelve repeats from the total pool of sequencing reads. The number of lines or "hits" were counted for each motif length at each baited location within an individual bear genome. Each repeat motif was sorted to remove options with insufficient flanking region for PCR primer binding and those without 40-60% G/C content. The remaining repeat motifs for all brown bear individuals were compared to identify matching flanking regions for shared STR loci using the MEGA5.2: Molecular Evolutionary Genetic Analysis v. 5.2 software. Similar flanking regions were imported into the

MEGA 5.2 software and aligned in the Alignment Explorer using default settings; sequences were identified by individual and location within the genome as specified within the FASTA files. Matching bases between sequences are indicated by a "*" above those bases.

Chapter III:

Results

Post-Library Preparation Size Selection

For the post-library preparation, size selection, and all downstream processes, only brown bear samples were utilized; polar bear and American black bear samples will undergo library preparation, library enrichment using biotinylated RNA baits, and sequencing in future studies. Selecting fragment sizes of 300-400 bp while maximizing DNA retention is essential for library enrichment. To this purpose, two size selection methods were tested. First, the recommended approach of a bead-based dual bead size selection using AMPure® XP Beads (Beckman Coulter, Inc., Brea, CA) was performed. Several ratios were used to select for ~300 bp with differing volumes of adaptor-ligated samples added to each size selection reaction. Table 4 lists the Bead: DNA ratios used as well as DNA input volume.

Table 4: Bead: DNA ratios tested during bead-based dual bead size selection of library prepared samples

DNA Volume (µL)	First Bead Selection	Second Bead Selection
20	0.8X	0.7X
50	0.8X	0.7X
100	0.8X	0.7X
50	0.85X	0.75X
100	0.6X	0.15X
100	0.5X	0.15X

Samples size selected using the 0.8X/ 0.7X and 0.85X/ 0.75X Bead: DNA ratio selected for fragment sizes smaller than that required for downstream processing, ~200 bp, as shown in Figure 3.

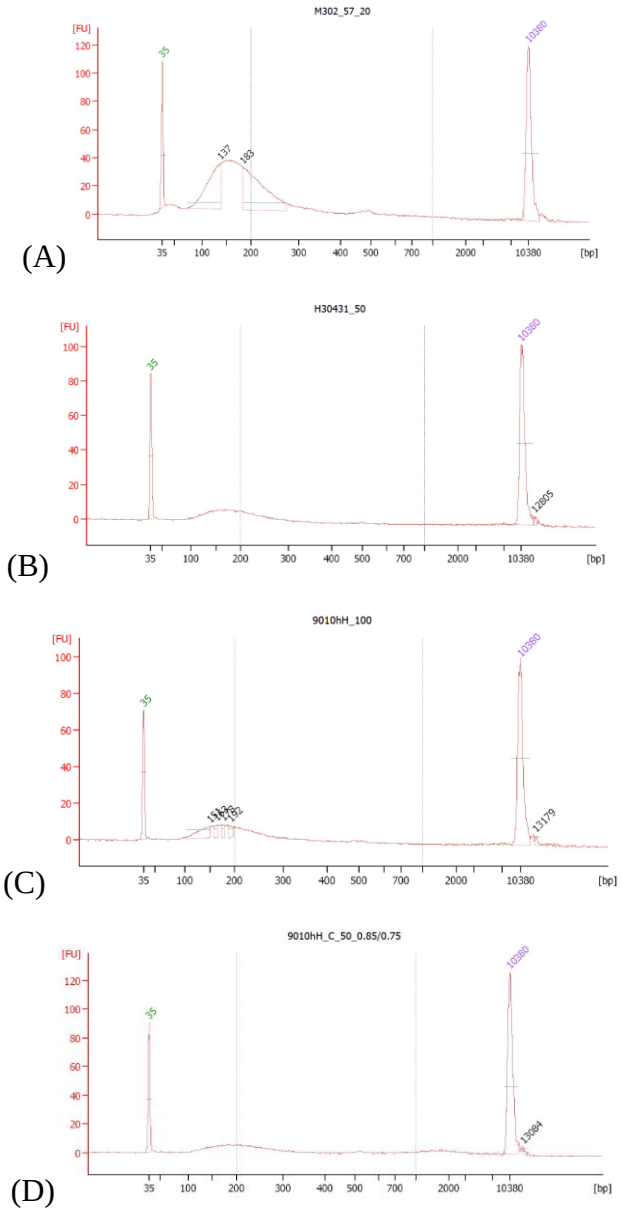


Figure 3: Bead-based dual size selection results using a 0.8X/ 0.7X or 0.85X/ 0.75X Bead: DNA ratio. (A) 0.8X/ 0.7X ratio with 20 µL input volume (M302_57, polar bear), (B) 0.8X/ 0.7X ratio with 50 µL input volume (H30431, American black bear), (C) 0.8X/ 0.7X ratio with 100 µL input volume (9010hH, brown bear), and (D) 0.85X/ 0.75X ratio with 50 µL input volume (9010hH, brown bear). All samples exhibited fragment sizes of ~200 bp as opposed to the desired fragment sizes of ~300 bp..

Samples prepared using either a 0.6X/ 0.15X or 0.5X/ 0.15X, while containing appropriately sized fragments of 300-400 bp, also exhibited an unexpected secondary selected range around 2000 bp and greater as shown in Figure 4.

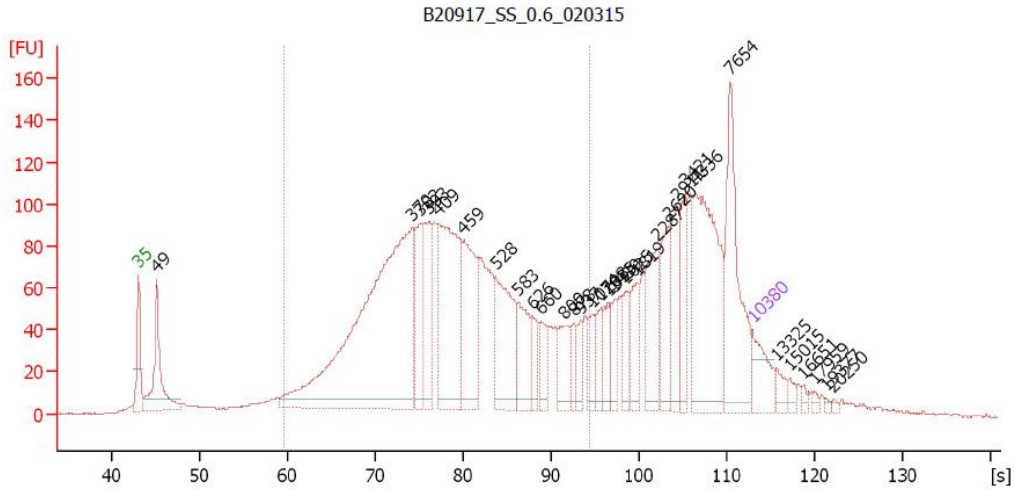


Figure 4: Example of secondary selection range for bead-based dual bead size selection of library prepared samples. As shown above for individual B20917 (brown bear), all samples size selected using a bead: DNA ratio of 0.6 X/ 0.15X or 0.5 X/ 0.15 X contained fragments of 300-400 bp as desired as well as an unexpected secondary range of ~2000 bp and greater.

A column-based size selection method was explored to determine if fragment size quality and DNA quantity could be improved in library prepared samples. The ligated DNA fragments were size selected using the Select-a-Size DNA Clean and Concentrator™ (Zymo Research, Irvine, CA; Appendix E), which includes a silica-based matrix. A double size selection was conducted to remove fragments of ≥700 bp and ≤200 bp. While samples undergoing column-based size selection did not exhibit narrow peaks as seen with bead-based size selection, increased DNA quantity was observed. Following fragmentation/end repair, barcode adaptor ligation, size selection, amplification, and purification, each samples was assessed on the Agilent 2100 Bioanalyzer (Agilent Technologies Inc.); electropherograms indicated that all brown bear samples were successfully ligated with sufficient DNA quantity for library enrichment as shown in Figure 5.

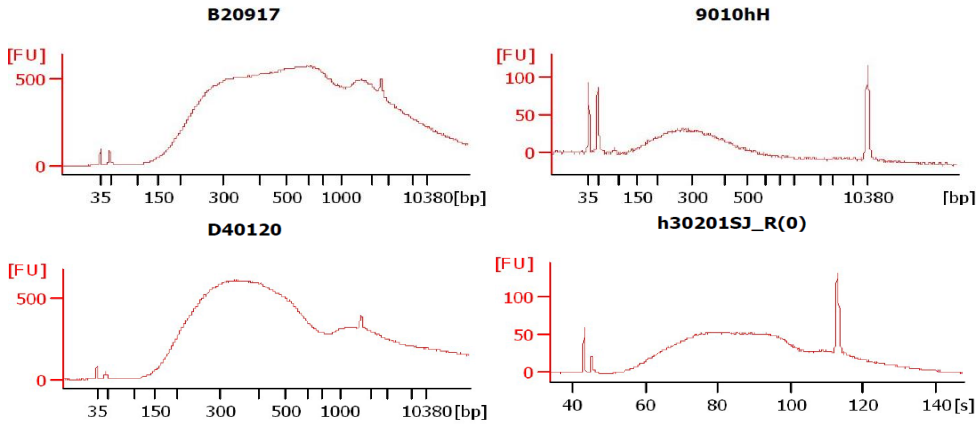


Figure 5: Electropherograms from the Agilent 2100 Bioanalyzer (Agilent Technologies Inc.) of all four brown bear samples (B20917-top left, 9010hH-top right, D40120-bottom left, and H30201-bottom right) after library preparation and size selection. All samples exhibit fragment sizes peaking at ~300 bp as required for library enrichment and sequencing.

Due to the higher quantity of DNA yielded from this method, a column-based double size selection was chosen to size select library prepared DNA for fragments of 300-400 bp. This will ensure sufficient DNA quantity for library enrichment using biotinylated RNA baits and optimal fragment size required for sequencing via the Ion Torrent™ PGM™ (Life Technologies™).

Library Enrichment and Sequencing

Library enrichment yielded no visible DNA in the resulting product after 14 amplification cycles; however, observable PCR product was seen after 36 amplification cycles. Sequencing data was analyzed following successful completion. The chip had 68% loading efficiency with 100% enrichment efficiency from the Ion OneTouch™ ES (Life Technologies™). Beads present within the sequencing run consisted of 35% clonal beads. The final library was 85% of the pooled library and contained 30% usable reads. All run summary results are shown in Figure 6.

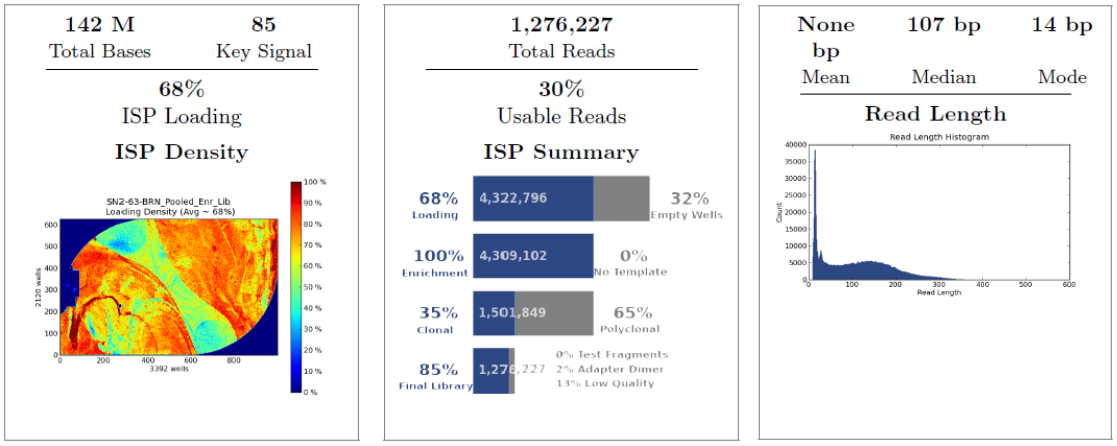


Figure 6: Ion Torrent™ PGM™ System (Life Technologies™, Carlsbad, CA) run summary results. The 316™ v2 chip had 68% loading efficiency with 100% enrichment efficiency. Beads presented with 35% clonal beads with 85% of the pooled library sequenced and containing 30% usable reads.

BAM files generated from the sequencing run were uploaded into the NextGENe® software, converted to FASTA files, and analyzed using a Floton *de novo* assembly method.

NextGENe® Data Analysis

De novo assembly of sequencing data from the Ion Torrent™ PGM™ revealed that out of twelve repeat motifs used to mine for tetranucleotide STR regions within the library prepared bear genomes, five repeat motifs presented with at least one identifiable tetranucleotide repeat motif of at least three repeats for all four tested individuals. The number of detected baited repeat motif regions for each individual is summarized in Table 5.

Table 5: Number of detected tetranucleotide repeat motifs for twelve RNA baits from preliminary analysis of sequencing data using NextGENe® software

	Sample			
Bait	**B20917**	**9010hH**	**D40120**	**H30201**
TNRSTR-1	1	0	5	5
TNRSTR-2	0	0	3	3
TNRSTR-3	0	0	4	4
TNRSTR-4	0	0	3	3
TNRSTR-5	1	0	5	5
TNRSTR-6	**8**	**1**	**15**	**15**
TNRSTR-7	0	0	2	2
TNRSTR-8	**3**	**3**	**7**	**7**
TNRSTR-9	**3**	**2**	**2**	**2**
TNRSTR-10	**3**	**3**	**6**	**6**
TNRSTR-11	**2**	**2**	**4**	**4**
TNRSTR-12	0	0	1	1
Total	21	11	57	57

[a] Baits in **bold** are detected tetranucleotide repeat motifs that have at least three repeats and are present in all four brown bear individuals

Out of the five tetranucleotide repeat motifs recovered in all four individuals using NextGENe® (those shown in **bold** in Table 5), four out of five repeat motifs are present within dense homopolymeric A regions within the individual genomes: TNRSTR-8, TNRSTR-9, TNRSTR-10, and TNRSTR-11. An example of an observed homopolymeric region flanking a detected tetranucleotide repeat motif of at least three repeats is shown in Figure 7.

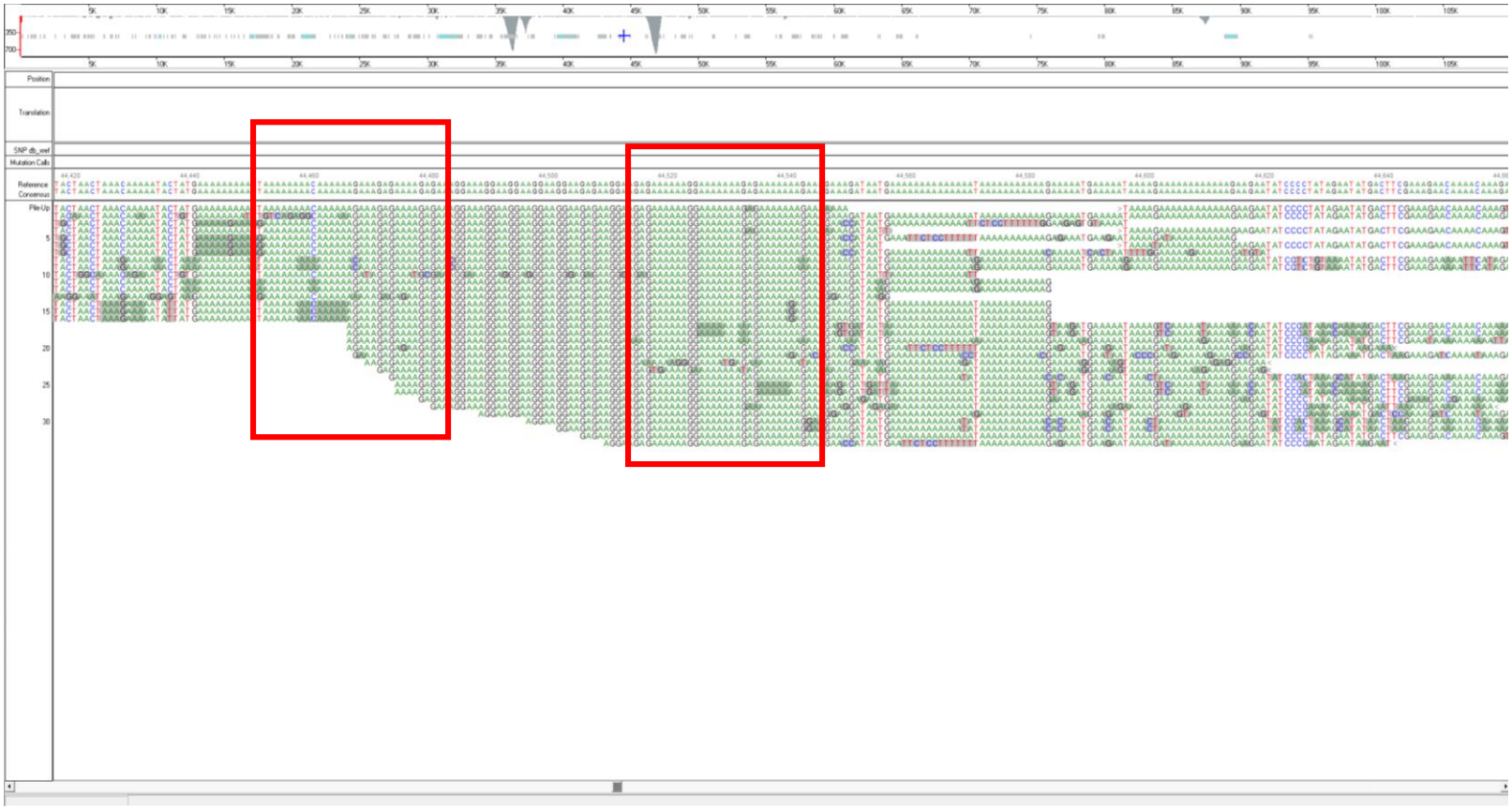


Figure 7: Example of homopolymeric flanking region observed in TNRSTR-8, -9, -10, and -11 from preliminary analysis of sequencing data using NextGENe® software. Example is from individual D40120 baited with TNRSTR-8 at repeat start location 44490 within the genome. Homopolymeric flanking regions are identified by red box with the repeat motif region centered between the two boxes.

Figure 8: Example of flanking region for region baited with bait TNRSTR-6 from preliminary analysis of sequencing data using NextGENe® software. Example is from individual H30201 at repeat start location 19048 within the genome. Flanking regions are identified by red box with the repeat motif region centered between the two boxes.

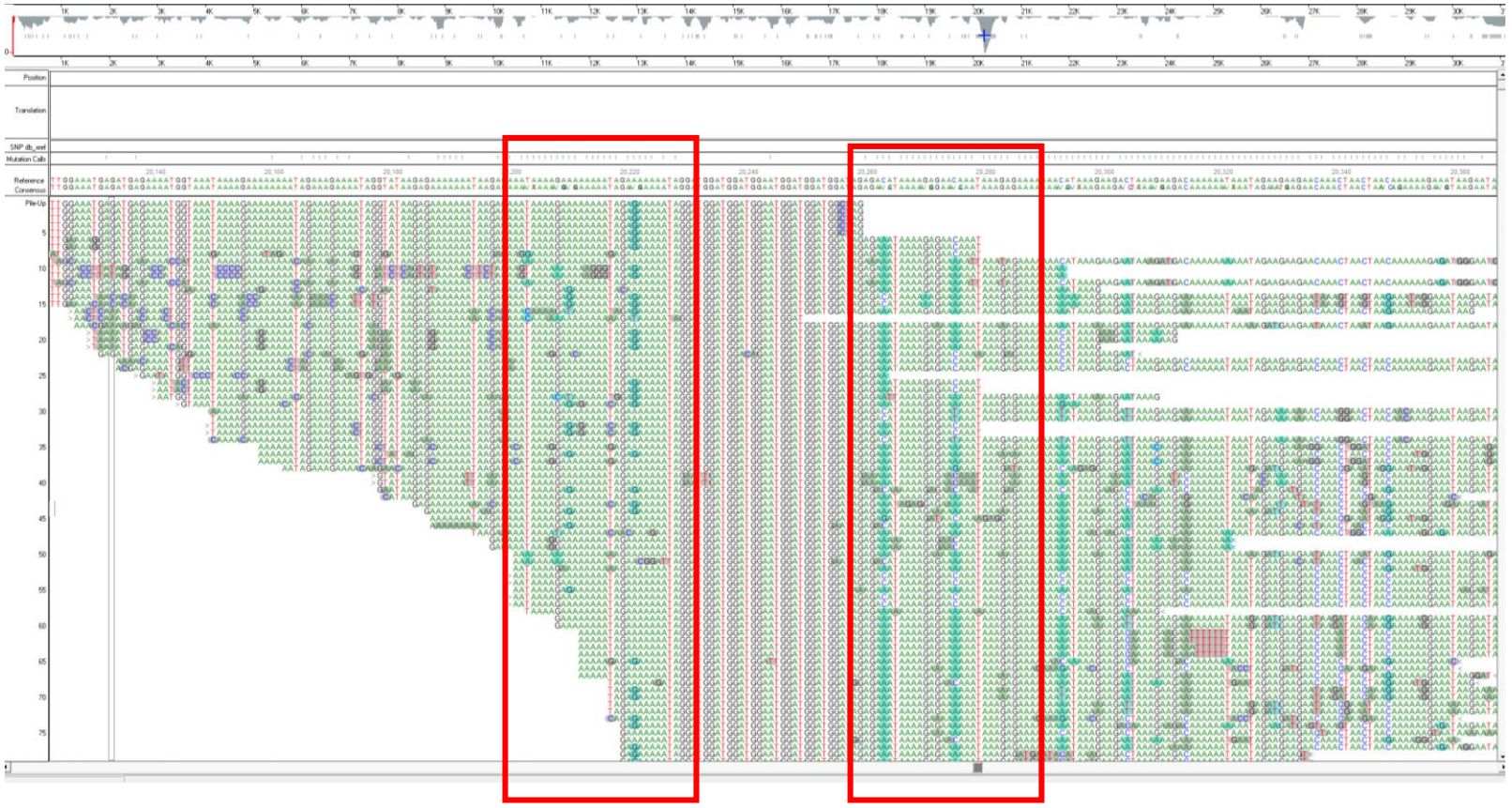


Figure 9: Homopolymeric A flanking region in individual 9010hH at repeat start location 20233. Flanking regions are identified by red box.

Table 6: Results of algorithm mining for twelve biotinylated RNA baits for regions containing three, seven, nine, and twelve repeats

B20917					9010hH				
Bait	Algorithm-3 Repeats	Algorithm-7 Repeats	Algorithm-9 Repeats	Algorithm-12 Repeats	Bait	Algorithm-3 Repeats	Algorithm-7 Repeats	Algorithm-9 Repeats	Algorithm-12 Repeats
TNRSTR-1	1433	27	0	0	TNRSTR-1	757	7	0	0
TNRSTR-2	3832	1600	837	109	TNRSTR-2	1336	438	184	17
TNRSTR-3	3862	1619	844	124	TNRSTR-3	1411	466	202	18
TNRSTR-4	1476	443	174	7	TNRSTR-4	374	55	9	3
TNRSTR-5	1265	12	0	0	TNRSTR-5	713	0	0	0
TNRSTR-6	6633	1936	841	46	TNRSTR-6	5618	1001	346	13
TNRSTR-7	3802	1575	812	108	TNRSTR-7	1337	428	179	14
TNRSTR-8	19593	10299	6856	2581	TNRSTR-8	6420	2219	1232	374
TNRSTR-9	853	54	16	7	TNRSTR-9	552	16	3	0
TNRSTR-10	19273	10117	6658	2461	TNRSTR-10	129	2141	1181	345
TNRSTR-11	831	49	15	7	TNRSTR-11	543	18	3	0
TNRSTR-12	343	117	42	4	TNRSTR-12	191	25	6	0
Total	63196	27848	17095	5454	Total	19381	6814	3345	784

Note: All numbers listed in Table 6 refer to the number of lines of data generated via algorithm mining from the FASTA files generated for the sequencing data and not to the number of regions within the genome containing baited repeat motifs.

Table 6 (Continued):

D40120					H30201				
Bait	Algorithm-3 Repeats	Algorithm-7 Repeats	Algorithm-9 Repeats	Algorithm-12 Repeats	Bait	Algorithm-3 Repeats	Algorithm-7 Repeats	Algorithm-9 Repeats	Algorithm-12 Repeats
TNRSTR-1	5557	78	3	0	TNRSTR-1	1779	46	1	0
TNRSTR-2	23383	8837	4093	348	TNRSTR-2	9966	3671	1649	112
TNRSTR-3	23837	9048	4238	377	TNRSTR-3	10162	3769	1752	137
TNRSTR-4	5891	1333	334	22	TNRSTR-4	0	415	116	1
TNRSTR-5	4989	44	0	0	TNRSTR-5	1625	23	0	0
TNRSTR-6	49306	12757	4669	202	TNRSTR-6	13337	3510	1310	45
TNRSTR-7	23313	8648	3953	323	TNRSTR-7	9885	3579	1583	106
TNRSTR-8	67722	33323	21245	7323	TNRSTR-8	25402	12491	7927	2787
TNRSTR-9	3617	215	60	12	TNRSTR-9	1406	68	17	0
TNRSTR-10	67739	32760	20747	6973	TNRSTR-10	25434	12348	7798	2711
TNRSTR-11	3496	191	52	8	TNRSTR-11	1307	59	15	0
TNRSTR-12	2262	525	129	4	TNRSTR-12	1165	311	117	3
Total	281112	107759	59523	15592	Total	101468	40290	22285	5902

See **Note** above

Only regions baited with TNRSTR-6 contains flanking regions of sufficient diversity for potential use as primer binding sites; an example of this is shown in Figure 8. Despite the increased diversity in flanking regions for samples containing repeat motifs for TNRSTR-6, individual 9010hH possessed a homopolymeric A flanking region for this repeat motif at repeat start location 20233, weakening its use as a potential individualizing STR locus. This repeat motif region is shown in Figure 9. Due to lower coverage of detected tetranucleotide repeat motif regions (< 10X) or repeat motif regions with higher coverage located within homopolymeric flanking regions, no allele designations were made for any brown bear individual.

Algorithm Data Analysis

Algorithm searches on individual FASTA files for baited repeat motifs containing three, seven, nine, and twelve repeats revealed a much greater diversity of potential STR repeat motifs as shown in Table 6.

Manual searching of the number and variety of repeats for each repeat motif was used as a preliminary indicator of repeat motif viable. Search criteria included analyzing repeat locations for repeat motifs that meet or exceed nine repeats for each bait. For an STR locus to be highly discriminating, that marker must have sufficient allelic variation, larger repeat numbers suggests higher genetic variation for a given locus. In addition, flanking regions for potential primer binding sites were used to ultimately determine the viability of repeat motifs. Primer binding sites to be used for amplification must contain between 25-30 bp as well as 40-60% GC content.

Hundreds of thousands of reads were generated by the Ion Torrent™ PGM™ for each individual: B20917-123,285, 9010hH-150,488, D40120-622,589, and H30201-211,396. By using a computer algorithm to search for locations within the genome containing the reverse complement

of each RNA bait sequence, these reads were reduced to more specific locations within each individual genome. While the location data for each repeat motif was more specific than the original whole-genome data, lines of data generated from the algorithm search numbered in the millions. Due to the vast number of strands for each repeat motif, only a sampling of repeat motifs (12 per individual at the nine and twelve repeat number level) were analyzed. Each repeat motif was designated as one of the following: non-viable due to no results for nine or twelve alleles or insufficient flanking region, potential with further mining of additional locations for a given repeat motif within the genome, or impractical without more concentrated mining of the repeat motifs using both more focused algorithm mining and more specific biotinylated RNA baits. The results of the manual searching of repeat motifs via algorithm mining and their respective designations are presented in Table 7.

Table 7: Results of manual searching of twelve RNA baits repeat motifs via algorithm mining and their respective designations

Bait	Non-Viable	Potential	Impractical
TNRSTR-1	X		
TNRSTR-2		X	
TNRSTR-3		X	
TNRSTR-4		X	
TNRSTR-5	X		
TNRSTR-6		X	
TNRSTR-7		X	
TNRSTR-8			X
TNRSTR-9	X		
TNRSTR-10			X
TNRSTR-11	X		
TNRSTR-12		X	

Four out of twelve repeat motifs were determined to be non-viable. These repeat motifs did not contain nine or more repeats for a given motif or did not possess flanking regions that met the

criteria for use as primer binding sites. Of the remaining eight repeat motifs, six were found to have potential for STR repeat motifs and two were considered impractical for use as STR repeat motifs without additional narrowing of DNA mining motifs to isolate more specific locations within the DNA as well as adequate flanking regions.

For all twelve repeat motifs sampled, only one matching flanking region was discovered: TNRSTR-12 for individuals D40120 and H30201 at locations 01855:01624 and 00140:01192, respectively; this matching flanking region was confirmed using MEGA 5.2 software as shown in Figure 10.

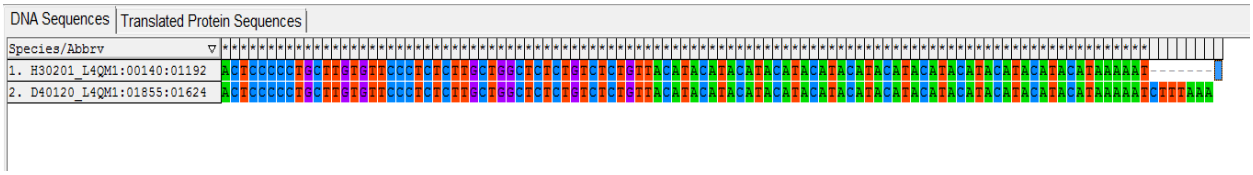


Figure 10: Matching flanking region and repeat unit length for individuals H30201 and D40120 for repeat motif TNRSTR-12. Matching nucleotides between sequences are indicated by a "*" above those bases.

It should be noted that the analysis of the six potential STR repeat motifs characterized in the algorithm mining consisted of only a sampling of locations for each individual sample. Future studies conducted must include a more in-depth mining process to further narrow the number of potential STR repeat motifs within the genome as well as identify viable flanking regions for primer binding sites.

Chapter IV:

Discussion

This study presents several viable repeat motifs for short tandem repeat loci to be used for species identification and individualization of individual bear samples, ultimately indicating that STRs can be enriched and identified in non-model species for which reference genomes are not available using RNA baits. By enriching library prepared genomic DNA for ursid species with biotinylated RNA baits, the number of possible STR repeat motifs within the genome is reduced from hundreds of thousands of possibilities for each repeat motif to thousands, each with identifiable locations within the genome that can be further characterized through different assembly methods or algorithm mining techniques. These thousands of repeat motifs can be further narrowed through specific criteria such as repeat number and flanking region usefulness. In-depth analysis of each repeat motif is somewhat simplified by mapping the location of a given repeat motif within the genome; however, the vast number of each repeat motif makes the task of narrowing the options for potential loci for STR typing a long, detailed process.

The reasoning for the failure to prepared libraries for all polar bear and American black bear samples is unknown. For all eight samples, significant peaks were present in all electropherograms analyzed using the Agilent 2100 High Sensitivity DNA Kit (Agilent Technologies Inc.) on the Agilent 2100 Bioanalyzer (Agilent Technologies Inc.), as shown in Figure 11. Each electropherogram shows a single overblown peak present at 45-50 bp, indicating that adaptor ligation was unsuccessful.

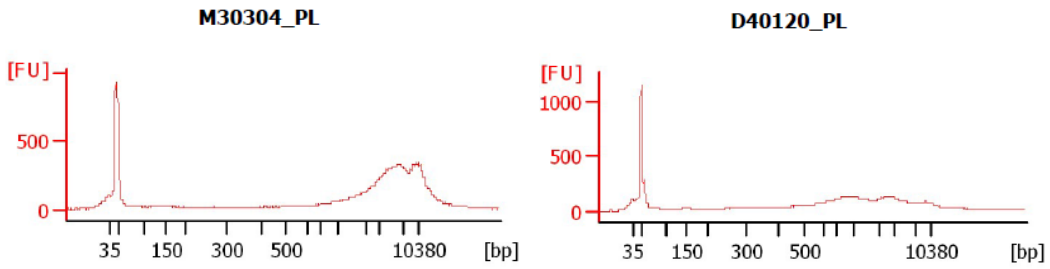


Figure 11: Example electropherograms of polar bear (left) and American black bear (right) samples from failed ligation reactions. X-axis units, base pairs; y-axis, fluorescence units (FU)

Several avenues were explored to determine the cause. Initially, the cause for ligation failure was believed to be related to the T4 ligase used during adaptor ligation; however, two separate experiments using a new T4 ligase as well as a new ligase and different barcode adaptor yielded results consistent to those of the initial failed samples. A new fragmentation and ligation kit was used to prepare each species with results similar to those seen in Figure 8. These results are highly unusual considering the close relatedness of polar and American black bears to brown bears [3, 15]. Further experiments will be conducted to explore the nature of this failure, including options such as re-extraction of bear samples and purification of sample extracts.

Sequencing data from the Ion Torrent™ PGM™ analyzed via the NextGENe® software showed that this method is not ideal for *de novo* assembly of sequence data for the purpose of identifying potential short tandem repeat loci. Less than half of the baited repeat motifs were detected in all individuals and those that were identified using the software were located within poly-Adenine regions, weakening their use as individualizing loci. The lack of reference sequence reduces the software's ability to properly align the data and improve the quality of information derived from sequencing data. Additionally, flanking regions and repeats may be reduced or eliminated due to the low average read length of 107 bp. However, all repeat motifs identified in each individual sample show that longer read lengths are present within the sequence data,

meaning that the entirety of a repeat motif region as well as associated flanking regions are characterizable from the sequencing data. A targeted read length of 300-400 bp was desired to reduce the occurrence of incomplete repeat motifs and flanking regions that may exist within the ursid genomes. Future studies should explore alternate sequencing platforms that are better suited to handle homopolymeric regions and repeat regions.

Algorithm mining presents better quality data regarding the identification of useful STR motifs. Over half of the baited loci were shown to contain viable options for individualizing loci. Due to the vast number of possible options for STR loci, only a small subsampling of options were characterized. A representative portion of potential STR loci with viable flanking regions is presented in Table 8. The complete sequence for each potential STR locus is listed in Appendix J and aligned sequences for these repeat motifs are included in Appendix K. The algorithm used to identify potential STR repeat motifs is included in Appendix L.

Table 8: Flanking regions of a representative sampling of potential STR loci derived from sequencing data using algorithm mining

Bait[a]	Sample	Allele	Location	Start Flanking Region	End Flanking Region
2	D40120	12	00067:00331	ACTCTGGGATGCCAGGAGATAGATAGATGA	NA
2	B20917	13	00345:00976	ATGGTAGGTTGTGGGTGCTTGGTAGGTACA	NA
2	9010hH	12	01275:02242	CAGGCACCAGGCACTGGGACAAACACATGA	NA
2	D40120	13	00111:01688	CATGTACCCTGAAAGTTAGGCTTGAACATA	TACATGCACACACACACACACACACACAC
2	B20917	13	00473:02343	CCCCATCAAGGTAGAGCACTGGGAGATAGA	NA
2	9010hH	12	00220:02501	CTCCCTGTGCCCCTCCCCTACTCTCTAAAA	NA
2	B20917	12	00414:01962	GAAAGTAAGTTGGCTGCTAGTTGGATATGA	NA
2	9010hH	12	00994:01697	GCACCCAGGCACTGGGGACAAAACACATGA	NA
2	D40120	12	00132:01455	GCTATCTGGATGGATGGATGGATGGATGGA	NA
2	D40120	12	00103:00269	GGATGGATGGATGGATAGATAGATAGATGA	NA
2	9010hH	12	00720:00152	GTACGTGTAAAGCGTAATCAAGATGATGGA	NA
2	9010hH	12	01498:02010	GTACGTGTAAAGCGTAATCAAGATGATGGA	NA
2	9010hH	12	01515:02388	GTACGTGTAAAGCGTAATCAAGATGATGGA	NA
2	9010hH	12	00705:01286	NA	TGATGGCTGATAGATGGATGGATGGATGGAT
2	9010hH	12	01433:02260	NA	ATTTTTTTCTCTAAGACTCTCTTGGCCCACC
2	D40120	12	00056:00530	NA	TATACACGCATACACACACACACAACTTTGG
2	H30201	12	00174:00856	NA	GATGATAGAAAGAAAGATGGATGGAGGGATG
2	H30201	12	00200:00431	NA	TGATAGATAGATGATAGATAGAGCGAGCCAG
2	B20917	12	00292:01351	TAATAATCTCTCCCTCCCTCTTTCTCTGCT	AGGAAGGAAGGAAGGAAGGAAGGAAGGAAG G
2	9010hH	12	01118:01382	TAGGGTGAAGAGACAGATGTAGATAGATGA	NA
2	B20917	12	00363:00761	TGACTCTTCCCATGACCTATCAGACTCACA	NA
3	9010hH	13	01560:03045	AATGATAGCTAGCTAGCTAGCTAGCTAGAT	NA
3	B20917	12	00363:00761	ACTCTTCCCATGACCTATCAGACTCACATA	NA
3	9010hH	12	00994:01697	AGGCACCCAGGCACTGGGGACAAAACACAT	NA
3	9010hH	12	00220:02501	CCCTGTGCCCCTCCCCTACTCTCTAAAATA	NA
3	B20917	13	00473:02343	GAGGTGCCCCATCAAGGTAGAGCACTGGGA	NA

Table 8 (Continued):					
3	D40120	12	00067:00381	GGACTCTGGGATGCCAGGAGATAGATAGAT	NA
3	9010hH	13	01118:01382	GGTAGGGTGAAGAGACAGATGTAGATAGAT	NA
3	B20917	12	00345:00976	GGTAGGTTGTGGGTGCTTGGTAGGTACATA	NA
3	9010hH	12	01433:02260	NA	GAATTTTTTTCTCTAAGACTCTCTTGGCCCA
3	B20917	12	00044:00693	NA	ATAGATAGATCTCAGTGGCTGACGACCACTC
3	D40120	12	00056:00530	NA	TACACGCATACACACACACACAACTTTGGGA
3	D40120	12	00135:01817	NA	TGGGTCTGGTTATGTCAACTGCTCATCAGCT
3	H30201	12	00174:00856	NA	GATGATAGAAAGAAAGATGGATGGAGGGATG
3	B20917	13	00414:01962	TGGAAAGTAAGTTGGCTGCTAGTTGGATAT	NA
3	D40120	12	00132:01455	TGGCTATCTGGATGGATGGATGGATGGATG	NA
3	9010hH	12	01275:02242	TTCAGGCACCAGGCACTGGGACAAACACAT	NA
3	B20917	12	00316:00564	TTGGGCTCCACACCCAACATTAAATAGACA	NA
4	9010hH	12	00538:02095	NA	AATCGATGGCAGAGAGAGAGGAAAAGAGGGA
4	B20917	13	00492:00964	CCTAAGCAGGAGGGAACCAAGTATTGGCTG	ACGCCTGCAAAGAGAAGCATTTCCA
4	B20917	13	00756:00465	AGCAGGAGGGAACCAAGTATTGGCCGAATG	ACGCCTGCAAAGAGAAGCATTTCCAGTTCAG
4	B20917	13	00713:01683	CCTAAGCAGGAGGGAACCAAGTATTGGCTG	ACGCCTGCAAAGAGAAGCATTTCCAGTTCAG
4	D40120	12	01382:00868	NA	AGATTTAGAGTACACACTCAACTCATTCCTG
4	9010hH	13	01581:02150	CAGTGCTGTAGCATCCAGTCAACGAATGTG	ATGGCAGAGAGAGAGGAAAAGAGGGAGAAGG
4	B20918	12	00373:01725	ATGTCCCTGATCCCTAGAACAGTGCCTGGC	NA
4	D40120	12	01611:02792	GAAGATAGAAAGTGCCAAGGGTACAAGATG	NA
4	H30201	12	00590:02025	GAGGACATGCAGTAGCTGTTTGCAG	NA
4	D40120	12	01405:00731	GTGCATAAACGCTTATGGAACGAATGGATG	NA
4	D40120	14	00330:00439	TCCTGTCAGACCCCAGGAGCCTTCTTTATG	NA
4	D40120	14	00453:01820	TCCTGTCAGACCCCAGGAGCCTTCTTTATG	NA
4	D40120	14	01020:01996	TCCTGTCAGACCCCAGGAGCCTTCTTTATG	NA
4	D40120	13	01851:02429	TCCTGTCAGACCCCAGGAGCCTTCTTTATG	NA
6	B20917	12	00123:01798	CCTTTGAGACATACAAGTGGAGATACTCAG	AAATGGATAAGGCTGGGTTTGAGAGAAGTGA
6	H30201	12	00424:01228	ACAGGGTAGGTACTCACTAGGGCTTTGTGA	AGAAAAATAAAGGGACCTCCTGTGACCTCTA

Table 8 (Continued):					
6	B20917	12	00271:01307	NA	AGATGACTCTACCCTAAATGGACCAACATGA
6	B20917	12	00590:01541	GAAAGGAGAATGGATAGATGGGAGAGAGGA	AGGAAGGAAGGAAGGATGGATGAGTGGATGA
6	H30201	12	00209:00355	AGTGCTATGACACCATCCAGGGCCAAGGAT	AGGATGGACGAGTGGATGGATGGATG
6	D40120	12	00072:00038	GATCTGGAGAATGTTTGAGGAGTGAAGAGA	CAGATGGACGGATGATGGATGGTGAGTGGAT
6	9010hH	12	00915:01341	NA	CGGTTGAATATCTTTACGTGCCTCCCATTGA
6	9010hH	12	01257:00847	NA	CGGTTGAATATCTTTACGTGCCTCCCATTGA
6	H30201	12	00467:01607	GGAAATTTTTTGGATGGATGGGTGGATGGG	CTGACAGGTAGGTGAATGGATGGGTGGATGG
6	H30201	12	00721:00428	GGAAATTTTTTGGATGGATGGGTGGATGGG	CTGACAGGTAGGTGAATGGGATGGGTGGATG
6	9010hH	12	01075:01620	GACAGATGGATGGACAGTTGGATGGGTGGG	GGGCGAGCAGACATCCTAATGAACCGATAGC
6	B20917	13	00609:01670	CAACCAATGTTGGCTGAATCCAGAAAGGGA	NA
6	B20917	12	00441:00074	CTAGATAGATGCTACCAAAAGGAAGGTACT	NA
6	9010hH	12	00535:00439	CTGGCACTTGTCAGATAATGGATGGAGAGA	NA
6	9010hH	12	00278:02401	GAGTATCTCAGTGTCTTCTGAGTAAAAGGG	NA
6	D40120	12	00087:00590	GATGCCGTGAAGACTGCTGAAATGATAACA	NA
6	B20917	12	00322:01269	GATGGAAGGATGCATGGAAAGACAATAGGA	NA
6	9010hH	12	00380:02349	GATGGAGGGATGGATGGAAGGAAGGATAGA	NA
6	H30201	12	00305:02286	GCCCTGCACACTGTAGATGCTCCCAAATAC	NA
6	H30201	12	00477:01292	GGTGAATGGATAAGTGGGTTGATGAATGTG	NA
6	B20917	12	00404:02338	GTTAGATAGGTAGGAGGAAGGATGGAAGGA	NA
6	9010hH	12	01141:01505	GTTGGATGGATGGAGAGATGGATGGAAGAA	NA
6	H30201	12	00297:02142	GTTTGATGAGTGGATGCACCTATGAACAAC	NA
6	H30201	12	00062:00492	TTTCTGTGTGCTCTTCTAACCACTTGACTT	NA
6	B20917	12	00455:01402	TTTGTCCTGGACACACACAGTGCTTAGTAC	NA
6	9010hH	12	00898:02606	NA	TAAATGGATGGATGGATCAATGGATGGTCGA
6	9010hH	12	00512:01591	NA	TAAGTAGATGGGATGATGGATGGAAGGATGG
6	9010hH	12	01556:02582	ATGGGTGGATGAATGGATAGATGGAAGGAA	TAGAAGGGTGGATGGATAAAACTCAAACTCA
6	B20917	12	00145:01851	AGTGGCTGGCTGGCTGGCTAAGTGGTTGGT	TAGATGGATGGGTGGATGGGTGGATGATTGG
6	H30201	12	00021:01613	TACCACACTACCCATAGATGGATGGATGCA	TGAGTGGATGGGTGGATGGATAAGCAAATGA
6	9010hH	12	00618:01236	NA	TGGGGACAGCTGCACAGCTAGGGGA

Table 8 (Continued):					
6	9010hH	12	00707:02717	GCTAGTGATACTCAACAGGTATTCACTGGT	TGGTTAGATGGATGGATAGATGGATGGATGG
6	D40120	13	00116:01009	CTTCATTGCACCCCCAGTGTACCCAAGGGA	TTGATGGATGGATGGATGATGTGCAAAGATG
7	H30201	12	00259:00735	NA	AACAGACAGAGACAGAGAGACATGGTGTTGG
7	D40120	13	00111:01688	NA	ACATGCACACACACACACACACACACACACA
7	D40120	12	00056:00530	NA	ATACACGCATACACACACACACAACTTTGGG
7	H30201	12	00174:00856	NA	GATGATAGAAAGAAAGATGGATGGAGGGATG
7	9010hH	12	00807:02415	NA	GATATAGTGATAGAGATGGATGAATGGATGG
7	9010hH	12	00705:01286	NA	GATGGCTGATAGATGGATGGATGGATGGATG
7	9010hH	12	01118:01382	AGGGTGAAGAGACAGATGTAGATAGATGAT	NA
7	9010hH	12	02058:02175	AGGGTGAAGAGACAGATGTAGATAGATGAT	NA
7	D40120	12	00136:00688	AGTGTTCACAGATGGGTGAATGGGTAAAGA	NA
7	D40120	12	00132:01455	CTATCTGGATGGATGGATGGATGGATGGAT	NA
7	D40120	12	00067:00331	CTCTGGGATGCCAGGAGATAGATAGATGAT	NA
7	B20917	12	00363:00761	GACTCTTCCCATGACCTATCAGACTCACAT	NA
7	9010hH	13	01560:03045	GATAGCTAGCTAGCTAGCTAGCTAGATGAT	NA
7	9010hH	12	01753:01282	GATAGCTAGCTAGCTAGCTAGCTAGATGAT	NA
7	B20917	12	00316:00564	GTTGGGCTCCACACCCAACATTAAATAGAC	NA
7	D40120	12	00106:01578	TAGGGCTTTCCTAAGAAGAGAACCAATATG	NA
7	9010hH	12	00220:02501	TCCCTGTGCCCCTCCCCTACTCTCTAAAAT	NA
7	B20917	13	00473:02343	TGAGGTGCCCCATCAAGGTAGAGCACTGGG	NA
7	B20917	13	00345:00976	TGGTAGGTTGTGGGTGCTTGGTAGGTACAT	NA
12	D40120	12	00835:00580	AGGGTGCACACACACACTCTCTCTCTCAAA	NA
12	D40120	12	01315:02002	AGGGTGCACACACACACTCTCTCTCTCAAA	NA
12	H30201	12	01426:01321	CTGGGTTCATCGTTCCTGTCCCTGGGAAAA	TAAAGTCACGCAGGCCTTCTGCTTGCCAGAA
12	H30201	12	01664:03211	CTGGGTTCATCGTTCCTGTCCCTGGGAAAA	NA
12	*D40120*	*12*	*01855:01624*	*TTCCCTCTCTTGCTGGCTCTCTGTCTCTGT*	*NA*
12	*H30201*	*12*	*00140:01192*	*TTCCCTCTCTTGCTGGCTCTCTGTCTCTGT*	*NA*

[a]: baits in *italics* possess matching flanking regions from two or more individuals

For the potential STR loci listed in Table 8, tens of thousands of lines of data for these combined baited sequences were generated. In order to identify more optimal regions within the ursid genome that could be used for STR typing, a more streamlined pipeline of algorithm mining of sequencing data must be developed.

TNRSTR-2, TNRSTR-3, and TNRSTR-7 all exhibit similar repeat patterns, TCTA, TATC, and ATCT, respectively, with many of the flanking regions possessing multiple repeats of one of the baits used in this study. This information suggests the possibility of compound or complex repeat motifs for these given baits. The similarity in repeat motif design was intentional to allow for the RNA baits to interrogate across compound and complex repeat structure. A closer analysis of these repeat motifs is needed to determine if these motifs are indeed compound or complex motifs in the individual ursid genomes as well as the repeat structure and flanking regions for these repeat motifs. A modified algorithm pipeline of these repeat motifs is recommended for this purpose with search criteria identifying two or more repeat motifs located within the same region of the genome; the distance between repeat motifs will be specified such that the motifs are considered part of the same repeat structure as opposed to being considered in linkage disequilibrium. TNRSTR-4 and TNRSTR-6 present with the most varied flanking regions of these RNA baits used in this study. These repeat motifs do not contain as many repeat motifs within the flanking regions, suggesting that these motifs have a simple repeat structure; however, TNRSTR-4, as shown in Table 10, has a higher instance of repeats greater than twelve (i.e. thirteen or fourteen).

The TNRSTR-12 repeat motif shows the most promise as a potential STR locus as it is the only baited region that presented a matched flanking region between individuals. The presence of a shared flanking region between individuals shows promise for discovering additional sites for

use as primer binding sites for future STR identification, allelic designation, and database development. Future bioinformatics studies will need to focus on additional mining to better sort and organize STR loci as well as identifying viable flanking regions for primer binding sites.

Choosing flanking regions that results in strong primer design is important for selecting STR loci. Non-specific PCR product can result from weak primer design as well as low annealing temperatures, too many amplification cycles, incorrect $MgCl_2$ concentrations, or extension times that are too long. Stronger primer binding sites consist of 40-60% GC content and are 25-30 bp in length. Additionally, these primer binding sites should not include repeat motifs of any type, including dinucleotide repeat motifs or alternate tetranucleotide repeat motifs, as these may result of weak primer binding and ambiguous PCR results. Following manual analysis of the algorithm mined data generated from this study, further efforts should be made to narrow the flanking regions found to have sufficient length and GC content to remove flanking regions found to have additional repeat motifs from the pool of potential STR loci.

The Ion Torrent™ PGM™ Sequencher (Life Technologies™) sequencing chemistry and other current second-generation deep sequencing platforms are not ideal for sequencing STR repeat motifs. Ion Torrent™ exhibit INDEL errors when sequencing homopolymeric stretches in the genome. Unfortunately, such homopolymeric regions are abundant for many types of repeats [41]. Improvements of the Ion Torrent™ PGM™ sequencing chemistry in conjunction with appropriate alignment methods will improve future STR calls. The Floton *de novo* assembly method of the NextGENe® software is designed to help reduced the errors associated with these homopolymer stretches using a "flow-based" approach. Despite correcting homopolymer errors in the sequencing data, this does not change the fact that the repeat motifs used in this study are located within these regions.

This study confirmed that the NextGENe® software is not currently capable of organizing and sorting sequencing data to assemble sequencing data for the purposes of identifying potential STR loci without the use of a reference sequence. For bears, the only published reference sequences that exist in GenBank® are those pertaining to dinucleotide repeat loci. Although these reference sequences may be used to align sequencing data, these sequences do contain inaccuracies and ambiguous bases that would affect the overall alignment of the data. Other software tools are available for the analysis of potential STR loci that use a preloaded human reference genome [41, 42]. Future bioinformatics studies should focus on developing a software pipeline that allows a reference sequence to be uploaded and allow the identification and organization of STR repeat motifs. This approach would allow true alleles to be sorted from stutter products and incomplete reads and make data interpretation more accurate and easier for analysts.

Algorithm mining for the repeat motifs used in this study proved more successful in identifying viable options for STR loci. By using algorithms to search the sequencing data generated from the Ion Torrent™ PGM™ for the reverse complement of the RNA bait sequences, hundreds of thousands of potential STR loci were narrowed down to tens of thousands. While only a sampling of the repeat motifs manually analyzed are included in this study, there are still thousands more to sort and characterize to determine their potential as STR loci. By analyzing these repeat motifs, more matching flanking regions between individuals can be identified and explored in future studies. Overall, these results show that next-generation sequencing applications are a viable option for characterizing STR loci for use in DNA profiling when used in conjunction with algorithm mining of repeat motifs. Future bioinformatics studies should focus on designing algorithms targeted for detection of compound and complex repeat structures. The manual search criteria for both repeat motifs and flanking region outlined in this study should be used for these

repeat structures. The results of these bioinformatics studies can then be used to design PCR primers for STR loci amplification and potential STR loci can be further characterized regarding their linkage between loci as well as discriminatory power.

In conclusion, this study identified several preliminary options for individualizing STR loci. Additional characterization will need to be performed on the sequencing data to illuminate more specific options for individualizing STR loci. Fine-tuning the library preparation will be required to allow for sample preparation across bear species while achieving optimal DNA retention and desired fragment length for downstream analyses. Once additional loci are identified within the bear genomes, additional individuals can be analyzed in order to determine allele frequency of species and subpopulations. Much like the tetranucleotide STR loci used in human forensic analyses, similar tetranucleotide STR loci in ursids will provide greater discriminating power to individual STR profiles generated in wildlife crime investigations and conservation analyses by providing more individualizing information and thus adding significance or "weight" to evidence provided in wildlife forensic cases.

APPENDIX

APPENDIX A [47]

Protocol for Organic DNA Extraction: Blood, Tissue, and Other Biological Material

(Adapted from the UNT Center for Human Identification Procedure Manual)

Stain Extraction Buffer (SEB):

1. Combine SDS (10 % v/v, 10 mL), EDTA (0.5 M, 1 mL), NaCl (5 M, 1 mL), Tris-HCl (pH 8.0, 0.5 mL), and Nuclease-free H_2O (30 mL). Confirm that solution is pH 8.0 and adjust accordingly.

SEB Working Solution:

1. Combine dithiothreitol (DTT, ~0.06 g) and SEB (10 mL).

TE^{-4} Buffer:

1. Combine Tris-HCl (pH 8.0, 1 M, 500 µL), EDTA (0.5 M, 100 µL), and Nuclease-free H_2O (49.4 mL).

Cell Lysis and Protein Digestion

1. Dissect 1-2 cm^2 of tissue and transfer to a labelled, sterile 0.1 mL microcentrifuge tube.
2. Add SEB Working Solution (500 µL) and Proteinase K (20 mg/ mL, 5 µL).
3. Briefly vortex and pulse spin samples.
4. Incubate samples at 56°C (±1°C) for 2-24 hours.

Phenol-Chloroform Isoamyl Alcohol (PCIA) Extraction:

1. After digestion, briefly pulse spin samples to force condensation built up during incubation to bottom of tube.
2. Add phenol-chloroform isoamyl alcohol (PCIA, 25:24:1, 500 µL) to each sample.
3. Vortex samples for 15-30 seconds to attain a milky emulsion, then centrifuge for three (3) minutes at 14,000-18,000 x g.
4. Carefully remove the aqueous layer (top layer) from each sample. Avoid drawing any of the protein interface or organic solvent into the pipette tip. Transfer to a new appropriately labelled, sterile tube.

Ethanol Precipitation

1. Add cold absolute ethanol (1.0 mL) to the aqueous extract. Vortex gently and incubate at -20°C or less for 30 minutes.
2. Centrifuge for 20 minutes at 14,000-18,000 x g and decant ethanol when finished.
3. Add ethanol (70 %, 1.0 mL) and centrifuge at 14,000-18,000 x g for 10 minutes. Carefully remove the majority of ethanol (70 %) solution using a transfer pipette; avoid disturbing the pelleted DNA. Evaporate remaining alcohol by drying sample on a heat block at 56°C (±1°C). Avoid over drying sample.
4. Add 50-100 µL of TE^{-4} buffer and let DNA resolubilize at 56°C (±1°C) for two (2) hours.
5. Store extract at 4°C for short term storage and -10°C or less for long term storage.

APPENDIX B [48]

Protocol for Qubit® 2.0 Fluorometer DNA Quantification

(Taken from the Qubit® 2.0 Fluorometer User Manual)

Performing the Qubit® dsDNA BR Assay:

1. Make the Qubit® working solution by diluting the Qubit® dsDNA BR reagent 1:200 in Qubit® dsDNA BR buffer. Use a clean plastic tube each time you make the Qubit® working solution. Do not mix the working solution in a glass container.
2. The final volume in each assay tube must be 200 µL. Each standard tube requires 190 µL of Qubit® working solution and each sample tube requires anywhere from 180-199 µL. Prepare sufficient Qubit® working solution to accommodate all standards and samples.
3. Load 190 µL of Qubit® working solution into each of the tubes used for standards.
4. Add 10 µL of each Qubit® standard to the appropriate tubes, then mix by vortexing 2-3 seconds. Be careful not to create bubbles.
5. Load the Qubit® working solution into individual assay tubes so that the final volume in each tube after adding sample is 200 µL.
6. Add each of your samples to assay tubes containing the correct volume of Qubit® working solution, then mix by vortexing 2-3 seconds. The final volume in each tube should be 200 µL.
7. Allow all tubes to incubate at room temperature for two (2) minutes.

Running New Standards for Calibration:

1. On the Home Screen of the Qubit® 2.0 Fluorometer, press **DNA**, then select **dsDNA Broad Range** as the assay type. The Standards Screen is displayed. If you have already performed a calibration for the selected assay, the Qubit® 2.0 Fluorometer prompts you to choose between reading new standards and using the previous calibration.
2. On the Standards Screen, select to run a new calibration or to use the last calibration
3. Press **Yes** to run a new calibration, then:
4. Insert the tube containing Standard #1 in the Qubit® 2.0 Fluorometer, close the lid, then press **Read**. The reading takes approximately three (3) seconds. Remove Standard #1.
5. Insert the tube containing Standard #2 in the Qubit® 2.0 Fluorometer, close the lid, then press **Read**. The reading takes approximately three (3) seconds. Remove Standard #2.
6. After the measurement is completed, the result is displayed on the screen.

Reading Samples:

1. Choose **Sample** to go to the Sample Screen.
2. Insert a sample assay tube in the Qubit® 2.0 Fluorometer and press **Read.** The reading takes approximately three (3) seconds. After the measurement is completed, the result is displayed on the screen. The number displayed is the concentration of DNA in the sample assay tube.
3. To read the next sample, remove the sample from the Qubit® 2.0 Fluorometer, insert the next sample, and then press **Read Next Sample**.
4. Repeat sample readings until all samples have been read.

Dilution Calculator:

1. After the sample measurement is completed, press **Calculate Stock Conc.** The Dilution Calculator Screen containing the volume roller wheel is displayed.
2. Using the volume roller wheel, select the volume of your original sample that you added to the assay tube. When you stop scrolling, the Qubit® 2.0 Fluorometer calculates the original sample concentration based on the measured assay concentration.
3. To change the units in which the original sample concentration is displayed, press **µg/ mL**. A pop-up window opens, showing the current unit selection.
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 once the unit selection pop-up window is closed.

APPENDIX C [49]

Protocol for Sizing and Analysis of DNA Fragments using the Agilent 2100 Bioanalyzer and Agilent High Sensitivity DNA Reagents

(Adapted from Manufacturer's Protocol)

Preparing the Gel-Dye Matrix:

1. Allow High Sensitivity DNA dye concentrate and High Sensitivity DNA gel matrix to equilibrate to room temperature for 30 min.
2. Add 25 μL of High Sensitivity DNA dye concentrate to a High Sensitivity DNA gel matrix vial.
3. Vortex solution well and spin down. Transfer to spin filter.
4. Centrifuge at 2240 x g ± 20% for 10 min. Protect solution from light. Store at 4°C. Use prepared gel-dye matrix within six (6) weeks of preparation.

Loading the Gel-Dye Matrix:

1. Allow the gel-dye mix to equilibrate to room temperature for 30 min before used.
2. Put a new High Sensitivity DNA chip on the chip priming station.
3. Pipette 9 μL of gel-dye matrix in the well marked (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 for exactly 60 s then release clip.
7. Wait for 5 s, then slowly pull back the plunger to the 1 mL position.
8. Open the chip priming station and pipette 9 μL of gel-dye mix in the wells marked [G].

Loading the Markers:

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

Loading the Ladder and the Samples:

1. Pipette 1 μL of High Sensitivity DNA ladder in the well marked with the ladder symbol.
2. In each of the 11 sample wells, pipette 1 μL of sample (used wells) or 1 μL of marker (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 instrument within 5 min.

APPENDIX D [50]

Protocol for NEBNext® Fast DNA Fragmentation and Library Prep Set for Ion Torrent

(Adapted from Manufacturer's Protocol)

Fragmentation and End Repair of DNA Protocol:

1. Mix the following components in a sterile microfuge tube on ice for a total volume of 18.5 µL:
    a. DNA (10 ng-1 µg): 1-15.5 µL
    b. NEBNext® DNA Fragmentation Reaction Buffer: 2 µL
    c. Sterile H_2O: variable
2. Vortex for three (3) seconds, pulse spin, and place on ice.
3. Vortex the vial of NEBNext® DNA Fragmentation Master Mix for three (3) seconds and pulse spin to collect liquid from the sides of the tube.
4. Add 1.5 µL of NEBNext® DNA Fragmentation Master Mix to the microfuge tube, vortex for three (3) seconds, and pulse spin.
5. Incubate in a thermal cycler for 12.5 min at 25°C, 10 min at 72°C, and hold at 4°C.
6. Pulse spin the microfuge tube and place on ice.

Preparation of Adapter Ligated DNA:

1. Add the following reagents to the microfuge tube for a total volume of 30 µL:
    a. Sterile H_2O: 3 µL
    b. NEBNext® T4 DNA Ligase Buffer for Ion Torrent: 10 µL
    c. NEXTflex™ Forward Barcode Adaptor*: 5 µL
    d. NEXTflex™ DNA P1 Adaptor: 5 µL
    e. NEBNext® *Bst* 2.0 WarmStart DNA Polymerase: 1 µL
    f. NEBNext® T4 DNA Ligase: 6 µL

    *A different barcode adaptor should be used for each sample
2. The total volume in the microfuge tube should be 50 µL. Mix the contents by pipetting up and down several times.
3. Incubate in a thermal cycler for 15 minutes at 25, 5 minutes at 65°C, and hold at 4°C.
4. Add Stop Buffer (5 µL), vortex, and pulse-spin.

Amplification of Size Selected DNA:

1. Mix the following components in a sterile microfuge tube for a total volume of 100 µL:
    a. Adapter ligated DNA: 1-40 µL
    b. Primers: 10 µL
    c. Sterile H_2O: variable
    d. NEBNext® High-Fidelity 2X PCR Master Mix: 50 µL
2. Place the tubes in a thermalcycler and run the following program:

Step	Temperature (°C)	Time (seconds)
Initial Denaturation	98	30
5 Cycles	98	10
	58	30
	72	30
Final Extension	72	300
Hold	4	∞

3. Purify the PCR product using QIAquick PCR purification kit following the manufacturer's instructions. Use 30 µL of buffer EB or molecular grade water for the final elution step.
4. Measure the DNA concentration.

APPENDIX E [51]

Protocol for Select-a-Size DNA Clean and Concentration™: Double Size Selection Protocol

(Adapted from Manufacturer's Protocol)

Deplete the Tope Fragments (≥700 bp):

1. Bring DNA sample up to 100 µL with DNA Elution Buffer.
2. Add Select-a-Size DNA Binding Buffer (500 µL). Mix thoroughly by pipetting up and down 5 times.
3. Transfer the mixture to a Zymo-Spin™ IIC Column in a 1.5 mL microcentrifuge tube. Centrifuge at ≥10,000 x g for 30 seconds. **Save the flow-through.**

Deplete the Bottom Fragments (300 bp or Below):

1. Choose the desired cutoff from the table below. Add the indicated volume of 95% ethanol to the flow-through and mix thoroughly.

DNA Fragments Retained	Additional Volume of 95% Ethanol
≥300 bp	0 µL
≥200 bp	30 µL
≥150 bp	70 µL
≥100 bp	150 µL
≥50 bp	300 µL

2. Transfer the mixture from step 4 to a Zymo-Spin™ IC-S Column in a collection tube and centrifuge at ≥10,000 x g for 30 seconds. Discard the flow-through.
3. Add DNA Wash Buffer (700 µL) to the column. Centrifuge at ≥10,000 x g for 30 seconds. Discard the flow-through.
4. Add DNA Wash Buffer (200 µL) to the column. Centrifuge at ≥10,000 x g for 60 seconds. Discard the collection tube.
5. Transfer the column to a 1.5 mL microcentrifuge tube; add DNA Elution Buffer (≥10 µL) directly to the column matrix and incubate fir a minimum of 1 minute at room temperature. Centrifuge at ≥10,000 x g for 30 seconds to elute the DNA.

APPENDIX F [46]

Protocol for Sequence Enrichment for Targeted Sequencing

(Adapted from Manufacturer's Protocol)

Hybridization:

1. Set the following program on a thermalcycler:

Step	Temperature (°C)	Time (Min)
#1	95	5
#2	65	3
#3	65	2
#4	65	∞

2. Prepare Library Master Mic in a nuclease-free tube and mix by vortexing. Set aside until Step #5.

Component	Amount (µL)
Block #1	2.5
Block #2	2.5
Block #3	0.6
Sequencing Library (100-500 ng)	3.4

3. Prepare Hybridization Master Mic in a nuclease-free tube and mix by vortexing. Set aside until Step #6.

Component	Amount (µL)
Hyb #1	20
Hyb #2	0.8
Hyb #3	8
Hyb #4	8

4. Prepare Capture Baits Master Mix in a nuclease-free tube and mic by pipetting. Set aside until Step #7.

Component	Amount (µL)
Capture Probe (baits)	5
RNase Block	1

5. Transfer the tube containing the Library Master Mix to the thermalcycler and start the program in Step #1. This will denature the DNA library for 5 minutes at 95°C.
6. Once the thermalcycler program reaches Step #2 temperature (65°C), transfer the tube containing the Hybridization Master Mix to the thermalcycler. Leave the Library Master Mix in the thermalcycler. This will pre-warm the Hybridization Master Mic for 3 minutes at 65°C.
7. Once the thermalcycler program reaches Step #3 temperature (65°C), transfer the tube containing the Capture Baits Master Mix to the thermalcycler. Leave all other tubes in the thermalcycler. This will pre-warm the Capture Baits Master Mix for 2 minutes at 65°C.

8. While keeping tubes at 65°C, transfer 7 µL of Library Master Mix and 13 µL of Hybridization Master Mix to Capture Baits Master Mix and mix by pipetting.
9. Hybridize solution at 65°C for 36 hours. Depending the application, hybridization time may need some optimization between 24 and 48 hours.

Recovery of Captured Targets:

1. Transfer 50 µL of MyOne™ Streptavidin C1 magnetic beads to a new 1.5 mL tube.
2. Pellet beads using a magnetic particle stand and discard the supernatant.
3. Add 200 µL Binding Buffer to beads to wash. Vortex tube for 5-10 seconds, place on magnetic particle stand for two minutes to pellet the beads, and remove and discard supernatant.
4. Repeat Step #3 twice for a total of three washes.
5. Resuspend the beads in 200 µL Binding Buffer.
6. Transfer the hybridization solution to the Binding Buffer/ Beads and incubate 30 minutes at room temperature on a rotator. Pellet beads with magnetic particle stand for two minutes and remove supernatant.
7. Add 500 µL Wash Buffer 1 to the beads and briefly vortex to resuspend. Incubate 15 minutes at room temperature. Pellet beads with magnetic particle stand for two minutes and remove supernatant.
8. Add 500 µL 65°C Wash Buffer 2 to the beads and briefly vortex to mix. Incubate for 10 minutes at 65°C. Pellet beads with magnetic particle stand for two minutes and remove supernatant.
9. Repeat Step #8 twice for a total of three 65°C washes. After third wash, make sure all additional buffer is removed.

Elution of Enriched Library:

1. Add 50 µL freshly prepared Elution Buffer to beads from Step #9 of Section II.
2. Vortex for 5-10 seconds to mix.
3. Incubate 10 minutes at room temperature.
4. Pellet the beads and transfer supernatant to a tube containing 70 µL Neutralization Buffer.

Enriched Library Cleanup:

1. Concentrate and desalt the solution using a QIAquick PCR Purification column following manufacturer's manual. The binding buffer should contain the pH indicator and pH should be adjusted if necessary. Elute with 30-µL buffer EB or molecular grade water.

Post-Capture Amplification:

1. Prepare PCR Master Mix on ice in a nuclease-free tube and mix by pipetting.

Component	**Amount (µL)**
Nuclease-free water	13
Primers	2
High-Fidelity 2X PCR Master Mix	25
Captured library	10

2. Place the tubes in a thermalcycler and run the following program:

Step	Temperature (°C)	Time (seconds)
#1	98	30
#2	98	20
#3	58	30
#4	72	30
#5	Repeat Step #2-#4 for 14 cycles	
#6	72	300
#7	4	∞

3. Purify the PCR product using QIAquick PCR purification kit following the manufacturer's instructions. Use 30 µL of buffer EB or molecular grade water for the final elution step.
4. Measure the DNA concentration.

APPENDIX G [52]

Protocol for QIAquick PCR Purification Kit

(Adapted from Manufacturer's Protocol)

1. Add 5 volumes Buffer PB to 1 volume of the PCR reaction and mix. If the color of the mixture is orange or violet, add 10 μL 3 M 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 30-60 seconds at 17,900 x g (13,000 rpm) at room temperature. Discard flow-through and place the QIAquick column back in the same tube.
4. To wash, add 750 μL Buffer PE to the QIAquick column. Centrifuge at 17,900 x g (13,000 rpm) for 30-60 seconds at room temperature. Discard flow-through and place the QIAquick column back in the same tube.
5. Centrifuge the QIAquick column once more in the provided 2 L 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 50 μL Buffer EB (10 mM Tris-Cl, pH 8.5) or water (pH 7.0-8.5) to the center of the QIAquick membrane and centrifuge the column for 1 minute. For increased DNA concentration, add 30 μL elution buffer to the center of the QIAquick membrane, let the column stand for 1 minute, and then centrifuge.
8. If the purified DNA is to be analyzed on a gel, add 1 volume of Loading Dye to 5 volumes of purified DNA. Mix the solution by pipetting up and down before loading the gel.

APPENDIX H [53]

Protocol for Ion PGM™ Template OT2 400 Kit

(Adapted from Manufacturer's Protocol)

Prepare the Amplification Solution:

1. Prepare the reagents as follows:

Reagents	Preparation
Ion PGM™ Template OT2 400 Reagent Mix	1. Allow the reagent mix to come to room temperature before use. 2. Vortex the solution for 30 seconds, then centrifuge the solution for 2 seconds 3. Keep the reagent mix at room temperature during use.
Ion PGM™ Template OT2 400 PCR Reagent B	1. Vortex the reagent for 1 minute, then centrifuge the solution for 2 seconds. 2. Inspect the reagent.
Ion PGM™ Template OT2 400 Enzyme Mix	1. Centrifuge the Enzyme Mix for 2 seconds. 2. Place on ice.
Ion PGM™ Template OT2 400 Reagent X	1. Centrifuge Reagent X for 2 seconds. 2. Place on ice.
Ion PGM™ Template OT2 400 Ion Sphere™ Particles	Place the suspension at room temperature.

2. Depending on the library type and concentration, dilute the library as shown in the table below. Use the library dilution within 48 hours of preparation.

	Ion AmpliSeq™ DNA Library	**Ion AmpliSeq™ RNA Library**	**gDNA Fragment or Amplicon Library**	**Total RNASeq Library**
Library Concentration	100 pM	100 pM	26 pM	20 pM
Volume of Library	2 µL	4 µL	25 µL	25 µL
Volume of Nuclease-Free Water	23 µL	21 µL	0 µL	0 µL
Total Volume of Diluted Library to Add to the Amplification Solution	25 µL	25 µL	25 µL	25 µL

   a. Vortex the diluted library for 5 seconds, then centrifuge for 2 seconds.

b. Place the diluted library on ice.

3. In a 1.5 mL Eppendorf LoBind® Tube at 15° to 30°C, add the following components. Add each component, then pipet the amplification solution up and down to mix.

Order	Reagent	Volume
1	Ion PGM™ Template OT2 400 Reagent Mix	500 µL
2	Ion PGM™ Template OT2 400 Reagent B	285 µL
3	Ion PGM™ Template OT2 400 Enzyme Mix	50 µL
4	Ion PGM™ Template OT2 400 Reagent X	40 µL
5	Diluted Library	25 µL
	Total	**900 µL**

4. Vortex the solution prepared in step 3 at maximum speed for 5 seconds, then centrifuge the solution for 2 seconds.
5. Vortex the Ion PGM™ Template OT2 400 Ion Sphere™ Particles (ISPs) for 1 minute, centrifuge 2 seconds, pipet up and down.
6. Add the Ion PGM™ OT2 400 Ion Sphere™ Particles to the amplification solution.

Order	Reagent	Volume
1	Amplification solution	900 µL
2	Ion PGM™ Template OT2 400 Ion Sphere™ Particles	100 µL
	Total	**1000 µL**

7. Vortex the complete amplification solution prepared in step 6 at maximum speed for 5 seconds.

Fill and Install the Ion PGM™ OneTouch Plus Reaction Filer Assembly on the Ion OneTouch™2 Instrument

1. Pipet 1000 µL of the amplification solution through the sample port.
2. Pipet 1000 µL of Ion OneTouch™ Reaction Oil (27 mL size) through the sample port.
3. Pipet 500 µL of Ion OneTouch™ Reaction Oil through the sample port.
4. Invert then install the filled Ion PGM™ OneTouch™ Plus Reaction Filter Assembly into the three holes on the top 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 a run, 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 has spun the sample or have not processed the sample within 15 minutes, centrifuge the sample of 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 Ion PGM™ OT2 Recovery 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. Process the ISPs:
    a. Add 500 µL Ion OneTouch™ Wash Solution to each Recovery Tube.
    b. Pipet the ISPs up and down to disperse the ISPs, then transfer each suspension of ISPs to a new labeled 1.5-mL Eppendorf LoBind® Tube.
    c. Heat the ISPs at 50°C for 2 minutes, then centrifuge the combined suspensions for 2.5 minutes at 15,500 x g.
    d. Remove all but 100 µL of the Wash Solution from the tube.

Enrich the Template-Positive Ion PGM™ Template OT2 400 Ion Sphere ™ Particles with Ion Torrent One Touch™ ES:

Prepare Reagents and Fill the 8-Well Strip:

1. .Prepare fresh Melt-Off Solution by combining in this order:

Order	**Component**	**Volume**
1	Tween® Solution	280 µL
2	1 M NaOH	40 µL
	Total	**320 µL**

Wash and Resuspend the Dynabeads® MyOne™ Streptavidin C1 Beads:

1. Vortex the tube for 30 seconds to thoroughly resuspend the beads, then centrifuge the tube of Dynabeads® MyOne™ Streptavidin C1 Beads for 2 seconds.
2. Open the tube, then use a new tip to pipet up and down the dark pellet of beads until the pellet disperses. *Immediately* proceed to the next step.
3. Transfer 13 µL of Dynabeads® MyOne™ Streptavidin C1 Beads to a new 1.5-mL Eppendorf LoBind® Tube.
4. Place the tube on a magnet such as a DynaMag™-2 magnet for 2 minutes, then *carefully* remove and discard the supernatant without disturbing the pellet of Dynabeads® MyOne™ Streptavidin C1 Beads.
5. Add 130 µL of MyOne™ Bead Wash Solution to the Dynabeads® MyOne™ Streptavidin C1 Beads.
6. Remove the tube from the magnet, vortex the tube for 30 seconds, and centrifuge the tube for 2 seconds.

Fill the 8-Well Strip:

1. Ensure that the template-positive ISPs from the Ion OneTouch™ 2 Instrument are in 100 µL of Ion OGM™ OT2 Wash Solution and are in Well 1 of the 8-sell strip. Well 1 with the ISPs is on the *left*.
2. If you have not done so already, assess the quality of the unenriched, template-positive Ion PGM™ Template OT2 400 Ion Sphere™ Particles using the Qubit® 2.0 Fluorometer, Guava® easyCyte™ Flow Cytometer, or the (demonstrated protocol) Applied Biosystems® Attune® Acoustic Focusing Cytometer.
3. Fill in the remaining wells as follows, then immediately proceed to step 4:

Well Number	Reagent to Dispense in Well
Well 1 (Well closest to the square-shaped tab)	*Entire* template-positive ISP sample (100 µL; prepared in step 1 of procedure)
Well 2	130 µL of Dynabeads® MyOne™ Streptavidin C1 Beads resuspended in MyOne™ Beads Wash Solution
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 freshly-prepared Melt-Off solution
Well 8	Empty

4. Confirm that the square-shaped tab is on the left, then insert the filled 8-well strip with the 8-well pushed all the way to the right end of the slot of the Tray

Prepare the Ion OneTouch™ ES:

1. Load a new tip in the Tip Arm.
2. Ensure that the back/ bottom end of the Tip Arm is not resting on top of the thumb screw, causing the Tip Arm to tilt forward.
3. Add 10 µL of Neutralization Solution to a new 0.2-mL PCR tube.
4. *Insert the opened 0.2-mL PCR tube with the Neutralization Solution into the hole in the base of the Tip Loader.*

Perform the Run:

1. Confirm that a new tip and opened 0.2-mL PCR tube with the Neutralization Solution have been loaded and that the 8-well strip is correctly loaded. Ensure that Well 1 (ISP sample) is the left-most well and that the 8-well strip is pushed to the far-right position within the slot.
2. Pipet the contents of Well 2 up and down to resuspend the beads before starting the run. Do not introduce bubbles into the solution.
3. If necessary, turn ON the Ion OneTouch™ ES and wait for the instrument to initialize: The screen displays "rdy". The Tip Arm performs a series of initialization movements and returns to the home position (~5 seconds).
4. Press **Start/ Stop**. The screen displays "run" during the run. The run takes ~35 minutes.
5. At the end of the run, the instrument displays "End" and deeps every 60 seconds. Press the **Start/ Stop** button to silence this alarm and reset the Ion OneTouch™ ES for the next run. The instrument can be left on between runs.
6. *Immediately after the run*, securely close and remove the PCR tube containing the enriched ISPs.
7. Mix the contents of the PCR tube by gently inverting the tube 5 ties.
8. Remove the used tip and the 8-well strip.

APPENDIX I [54]

Protocol for Sequencing on Ion 316™ Chip v2

(Taken from Manufacturer's Protocol)

Prepare Enriched, Template-Positive ISPs:

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

Annealing the Sequencing Primer:

1. Mix the contents of the tube by thoroughly 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 (visually compare to 15 µL liquid in a separate tube).
3. Add 12 µL of the Sequencing Primer and confirm that the total volume is 27 µL (add Annealing Primer if necessary).
4. Pipet the sample up and down thoroughly to disrupt the pellet.
5. Program a thermal cycler for 95°C for 2 minutes and then 37°C for 2 minutes, using the heated lid option.
6. Place the tube in the thermal cycler and run the program. After cycling, the reaction can remain in the cycler at room temperature while you proceed with Chip Check.

Chip Check:

1. Remove a new chip from its packaging and label it to identify the experiment. Save the chip package to scan the barcode later.
2. Place the chip on the Ion PGM™ Sequencer grounding plate or in the Ion centrifuge adaptor/ rotor bucket.
3. Press **Run** on the main menu to insert and cleaning chip, you can use the same used chip that was used for initialization.
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 for the experiment. **Do not wear gloves when transferring the chips on and off the instrument.** Close the chip clamp.
5. When prompted, use the barcode scanner to scan the barcode located on the chip package, or press **Change** to enter the barcode manually.
6. Press **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 the **Abort** button immediately to stop the flow to the chip.
8. When Chip Check is complete:
    a. If the chip passes, press **Next**
    b. If the chip fails, open the chip clamp, re-seat the chip in the socket, close the clamp, and press **Calibrate** to repeat the procedure. If the chip passes, press **Next**. If the chip still fails, press **Main Menu** and restart the experiment with a new chip.

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 through the following steps 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, for a total final volume of 30 µL.
2. Pipet the sample up and down to mix, and incubate at room temperature for 5 minutes.

Load the Chip:

Remove Liquid from the Chip:

1. Tilt the chip 45 degrees so that the loading port us the lower port.
2. Insert the pipette 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 tap pointing in** (toward the center of the MiniFuge).
4. Centrifuge for 5 seconds to completely empty the chip. Remove the chip from the bucket and wipe off any liquid on the bucket.

Load the Sample on the Chip:

1. Place the Ion PGN™ Chip back in the centrifuge adaptor bucket and place the bucket on a flat, stable surface such as the benchtop.
2. Following polymerase incubation, collect the entire sample (~30 µL) into a Rainin® SR-L200F pipette tip and insert the tip firmly into the loading port of the chip.
3. Dial down the pipette to gently and slowly deposit the ISPs at a rate of ~1 µL per second. To avoid introducing bubbles into the chip, leave a small amount of sample in the pipette tip (~0.5 µL).
4. Remove and discard any displaced liquid from the other port of the chip.
5. Transfer the chip to the MiniFuge **with the chip tab pointing in** (toward the center of the MiniFuge).
6. Centrifuge for 30 seconds, then remove the chip from the centrifuge bucket.
7. Mix the sample in the chip:
    a. Set the pipette volume to 25 µL.
    b. Tilt the chip 45 degrees so that the loading port is the lower port, and insert the pipette tip into the loading port.
    c. Without removing the tip, slowly pipet the sample in and out of the chip three times. **Pipet slowly to avoid creating bubbles.**
8. Centrifuge the chip for 30 seconds **with the chip tab pointing out** (away from the center of the MiniFuge).
9. Repeat the chip mixing in step #7 one more time, then spin for 30 seconds **with the chip tab pointing in** (toward the center of the MiniFuge).
10. Repeat the chip mixing in step #7, this time pipetting the sample in and out of the chip *five times*.

11. 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.
12. 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. **Do not spin the chip upside down**.
13. 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.
14. When chip loading is complete, press **Next** on the touchscreen.

Select the Planned Run and Perform the Run:

1. Press the **Browse** button next to the **Planned Run** field and select the name of the plan you created, then press **Next**.
2. The run settings will be automatically populated based on the Planned Run. Confirm that these settings are correct. Make any changes using the buttons and dropdown lists if necessary.

APPENDIX J

FASTA Sequences of Potential STR Repeat Motifs

[a]: Locations refer to the location of the sequence within the FASTA file

[b]: Repeat motifs are indicated in ***bold italics***. FASTA sequences with matching flanking regions have repeat motifs indicated in ***underlined bold italics***

Bait	Sample	Allele	Location[a]	FASTA Sequences[b]
2	D40120	12	00067:00331	GCTTAACAACTGCACCACCAGGTGCCCCTGGAGATTTTTTAATATTGTAAATTGTC ATCTGGACTCTGGGATGCCAGGAGATAGATAGATGA***TAGATAGATAGATAGATAGA TAGATAGATAGATAGATAGATAGATAGA***TGATAGATCACTTATTTAGAG
2	B20917	13	00345:00976	TGCCAATAAGCTTTTGTTTTGAATGGTAGGTTGTGGGTGCTTGGTAGGTACA ***TAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGA*** TCGACCCACCAACCAGGCCAGGCATGCCACTAATGATGAGTGTCAGGTA
2	9010hH	12	01275:02242	ACAAATATGACTGATGTTCAGGCACCAGGCACTGGGACAAACACATGA***TA GATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGA***CA
2	D40120	13	00111:01688	ATTATAGTTTGTAGGTAGCCCTATAGTTAGCCATGTACCCTGAAAGTTAGGCTTGA ACATA***TAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGA***TAGA TACATGCACACACACACACACACACACACACACA
2	B20917	13	00473:02343	CCAGGATATTCAGCTTGAACTTCCGGTGTGAGATAGACTGCATTTCCCTGA AAGAGCTACAGAGGAGTTTCTAGTCTCATGTTCTTCCAGAGCCTTGAGGGG CCTTGAGGTGCCCCATCAAGGTAGAGCACTGGGAGA***TAGATAGATAGATAG ATAGATAGATAGATAGATAGATAGATAGATAGATAGA***TAGACAGACA
2	9010hH	12	00220:02501	CTCTTGATTGATTTTTGGCTCAGGTCATGATCTCTCTCTCCCTGTGCCCCTCC CCTACTCTCTAAAA***TAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATA GATAGATAGA***TAAACA
2	B20917	12	00414:01962	ATTCTTTAGAATAGTAATAGAGAGGATTATCTATATATACTTTGTTTTTCTG TTATTCTTACATGCAACTAGGATAAGGCAATCACTTGCCAAAAAAACTACA GAATGGAAAGTAAGTTGGCTGCTAGTTGGATATGA***TAGATAGATAGATAGATAGA TAGATAGATAGATAGATAGATAGATAGATAGATAGA***TATAGAT

Bait	Sample	Allele	Location	FASTA Sequence
2	9010hH	12	00994:01697	ACAAATATGACTGATGTTCAGGCACCCAGGCACTGGGGACAAAACACATG A***TAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGA***CAG ACAGAGGTAGAT
2	D40120	12	00132:01455	CGCTTGATAACGGATTGACTGTAGAACAAGTAGATGTCAAGGATTGTTATAGGTTT CTGGATAGATAGTGGCTATCTGGATGGATGGATGGATGGATGGA***TAGATAGATAGA TAGATAGATAGATAGATAGATAGATAGATAGATAGA***T
2	D40120	12	00103:00269	CTATGATAACAAGGATGGATGGATGGATGGATGGATGGATGGATAGATAGA TAGATGA***TAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGA***CA G
2	9010hH	12	00720:00152	GTTTGGAAAAGGGCAAACTTCTTACAAGCTCTGGGATCAGATTGTACGTGT AAAGCGTAATCAAGATGATGGA***TAGATAGATAGATAGATAGATAGATAGATAGAT AGATAGATAGATAGATAGA***TTCTGTAGGAATTTAAAGCAATTAAGGTATTTG CTAT
2	9010hH	12	01498:02010	GTTTGGAAAAGGGCAAACTTCTTACAAGCTCTGGGATCAGATTGTACGTGT AAAGCGTAATCAAGATGATGGA***TAGATAGATAGATAGATAGATAGATAGATAGAT AGATAGATAGATAGATAGA***TTCTGTAGGAATTTAAA
2	9010hH	12	01515:02388	GTTTGGAAAAGGGCAAACTTCTTACAAGCTCTGGGATCAGATTGTACGTGT AAAGCGTAATCAAGATGATGGA***TAGATAGATAGATAGATAGATAGATAGATAGAT AGATAGATAGATAGATAGA***TTCTGTAGGAATTTAAAGCAATTAAGGTATTTG CTAT
2	9010hH	12	00705:01286	ATTATAGTTAGATAAA***TAGATAGATAGATAGATAGATAGATAGATAGATAGATAGA TAGATAGATAGA***TGATGGCTGATAGATGGATGGATGGATGGATGGATGGAT GGTTGGATGGATGGTTAGGTGGGTGGGTGGGTGGATGGATGGATGAACAG ACGGATAGATGGATGGATGGATAGGT
2	9010hH	12	01433:02260	GTAATGCTGAATGTTAAAAAATATATAGAGATAGATGA***TAGATAGATAGAT AGATAGATAGATAGATAGATAGATAGATAGATAGA***ATTTTTTTCTCTAAGAC TCTCTTGGCCCACCTATGAAAACAGATTAAAATTTTTTTTCTCAACCGAGAA TATTCGCAGTGAAACCTGGGTGGCTCAGTCGGTTAAGCATCTGCCTTCG
2	D40120	12	00056:00530	GCTATAGAAAGAAACAAAACCTTAGATATA***TAGATAGATAGATAGATAGATAGATA GATAGATAGATAGATAGATAGA***TATACACGCATACACACACACACAACTTTGGGAG AT

Bait	Sample	Allele	Location	FASTA Sequence
2	H30201	12	00174:00856	GTTGATTAGTTAATTAAATCTAGAGTGGCGTTAAGTTTAATCTATGAGATAT AGATAGATGATAGATAGATAGATGATAGATAGATGA***TAGATAGATAGATAG ATAGATAGATAGATAGATAGATAGATAGATAGA***TGATAGAAAGAAAGATGG ATGGAGGGATGGATGGATGGATGGATGGATGGATGGATGGATGGGTGGAT GGATGGATGGATGGTTGGTTGGATGGATTGGAGAGATAGATGATAGATAG GTAGATAGAT
2	H30201	12	00200:00431	CTTCCAAGTGGTGCTGCCTGCTGGCCCAAGGACCACACTTCGTGTTGTATTG GTATGGAAATAAATAAGTAGGTAGGTAGGTAGAT***TAGATAGATAGATAGATAGAT AGATAGATAGATAGATAGATAGATAGATAGA***TGATAGATAGATGATAGATA GAGCGAGCCAGCCGTGAGCAGGGTAGAAGCAGGCATTCCTGTGCATCTCG TTTCCCAGCTCAGTGTGTGTGCTTC
2	B20917	12	00292:01351	GACTTGTCAGCCTCCATAAGTGGGTGAGCCAACCTATTTGTTGTAATAATCT CTCCCTCCCTCTTTCTCTGCT***TAGATAGATAGATAGATAGATAGATAGATAGATAGA TAGATAGATAGATAGA***AGGAAGGAAGGAAGGAAGGAAGGAAGGAAGG
2	9010hH	12	01118:01382	GGTGGTTGGTAGGGTGAAGAGACAGATGTAGATAGATGA***TAGATAGATAG ATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGA***TATAGATAGATAGAT AGATAGATAGATAGAT
2	B20917	12	00363:00761	AGTCCTTTCTGCTTTAGAAACAAAGCCAGTAAAGAATTTAATGTCTTGGGC AATATGCAGGAACTTGGAGTTAACTTATCTGAATGACTCTTCCCATGACCT ATCAGACTCACA***TAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGA TAGATAGA***TATGAT
3	9010hH	13	01560:03045	ATATAGATAGATGATAGATAATGATAGCTAGCTAGCTAGCTAGCTAGAT***GA TAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATA***GA TGATAGATAGATAGATACAGATAGATGATAAGAAAATGATAGATGATAGA TAATGATTGAT
3	B20917	12	00363:00761	AGTCCTTTCTGCTTTAGAAACAAAGCCAGTAAAGAATTTAATGTCTTGGGC AATATGCAGGAACTTGGAGTTAACTTATCTGAATGACTCTTCCCATGACCT ATCAGACTCACATA***GATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGA TAGATAGATA***TGAT

Bait	Sample	Allele	Location	FASTA Sequence
3	9010hH	12	00994:01697	ACAAATATGACTGATGTTCAGGCACCCAGGCACTGGGGACAAAACACAT***GATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATA***GACAGACAGAGGTAGAT
3	9010hH	12	00220:02501	CTCTTGATTGATTTTTGGCTCAGGTCATGATCTCTCTCTCCCTGTGCCCCTCCCCTACTCTCTAAAATA***GATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATA***AACA
3	B20917	13	00473:02343	CCAGGATATTCAGCTTGAACTTCCGGTGTGAGATAGACTGCATTTCCCTGAAAGAGCTACAGAGGAGTTTCTAGTCTCATGTTCTTCCAGAGCCTTGAGGGGCCTTGAGGTGCCCCATCAAGGTAGAGCACTGGGA***GATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATA***GACAGACA
3	D40120	12	00067:00381	GCTTAACAACTGCACCACCAGGTGCCCCTGGAGATTTTTTAATATTGTAAATTGTCATCTGGACTCTGGGATGCCAGGAGATAGATAGAT***GATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATA***GATGATAGATCACTTATTTAGAG
3	9010hH	13	01118:01382	GGTGGTTGGTAGGGTGAAGAGACAGATGTAGATAGAT***GATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATA***TAGATAGATAGATAGATAGATAGATAGAT
3	B20917	12	00345:00976	TGCCAATAAGCTTTTGTTTTGAATGGTAGGTTGTGGGTGCTTGGTAGGTACATA***GATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATA***GATCGACCCACCAACCAGGCCAGGCATGCCACTAATGATGAGTGTCAGGTA
3	9010hH	12	01433:02260	GTAATGCTGAATGTTAAAAAATATATAGAGATAGAT***GATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATA***GAATTTTTTTCTCTAAGACTCTCTTGGCCCACCTATGAAAACAGATTAAAATTTTTTTTCTCAACCGAGAATATTCGCAGTGAAACCTGGGTGGCTCAGTCGGTTAAGCATCTGCCTTCG
3	B20917	12	00044:00693	AAAGAATTCTTACAACTTAATATTAGAAAAACAATAACAAAATGACTTTATATAATAAAAGCTACTATTAAGATAAAATACTATACACACACAGAGATCAAGAGATAGACAGATGATAGATAGGTAGATGATAGATGATAGATAGGTAGAT***GATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATA***ATAGATAGATCTCAGTGGCTGACGACCACTCACATGAAGGAGATGAAATCTGCT

Bait	Sample	Allele	Location	FASTA Sequence
3	D40120	12	00056:00530	GCTATAGAAAGAAACAAAACCTTAGATATATA***GATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATA*** TACACGCATACACACACACACAACTTTGGGAGAT
3	D40120	12	00135:01817	TTATAGATTAGACAGATAATTGATAGATGATAGATAGAT***GATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATA*** TGGGTCTGGTTATGTCAACTGCTCATCAGCTAGCCTCCAAGACC
3	H30201	12	00174:00856	GTTGATTAGTTAATTAAATCTAGAGTGGCGTTAAGTTTAATCTATGAGATATAGATAGATGATAGATAGATAGATGATAGATAGAT***GATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATA*** GATGATAGAAAGAAAGATGGATGGAGGGATGGATGGATGGATGGATGGATGGATGGATGGATGGATGGGTGGATGGATGGATGGATGGTTGGTTGGATGGATTGGAGAGATAGATGATAGATAGGTAGATAGAT
3	B20917	13	00414:01962	ATTCTTTAGAATAGTAATAGAGAGGATTATCTATATATACTTTGTTTTTCTGTTATTCTTACATGCAACTAGGATAAGGCAATCACTTGCCAAAAAAACTACAGAATGGAAAGTAAGTTGGCTGCTAGTTGGATAT***GATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATA*** TAGAT
3	D40120	12	00132:01455	CGCTTGATAACGGATTGACTGTAGAACAAGTAGATGTCAAGGATTGTTATAGGTTTCTGGATAGATAGTGGCTATCTGGATGGATGGATGGATGGATG***GATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATA*** GAT
3	9010hH	12	01275:02242	ACAAATATGACTGATGTTCAGGCACCAGGCACTGGGACAAACACAT***GATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATA*** GACA
3	B20917	12	00316:00564	TCTTGATCTCAGCTCAGGTCTTGATCTCAAGGTCATGAGCTCAAGCCCTATGTTGGGCTCCACACCCAACATTAAATAGACA***GATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATA*** GGATTGGGAATTCGAACTAAAGC
4	9010hH	12	00538:02095	GTCAACGAATGTG***AATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATG*** AATCGATGGCAGAGAGAGAGGAAAAGAGGGAGAAGGA
4	B20917	13	00492:00964	CTAGCCCTAAGCAGGAGGGAACCAAGTATTGGCTG***AATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATG*** ACGCCTGCAAAGAGAAGCATTTCCA

Bait	Sample	Allele	Location	FASTA Sequence
4	B20917	13	00756:00465	CTAGCCCTAAGCAGGAGGGAACCAAGTATTGGCCG***AATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATG*** ACGCCTGCAAAGAGAAGCATTTCCAGTTCAGGTTTAGGAAGTGGCTTGCTCAGTCACAGCCCTCTTGGAGTTTAGTG
4	B20917	13	00713:01683	CTAGCCCTAAGCAGGAGGGAACCAAGTATTGGCTG***AATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATG*** ACGCCTGCAAAGAGAAGCATTTCCAGTTCAGGTTTAGGAAGTGGCTTGCTCAGTCACAGCCCTCTTGGAGTTTAGTG
4	D40120	12	01382:00868	CTTAGAGAATGGAAGGATACTTTCAATATTAACAAAAGATG***AATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATG*** AGATTTAGAGTACACACTCAACTCATTCCTGTGGGTTTAACCTTGGATAAATCACCTAACTTCTCTGACTCTGTGCC
4	9010hH	13	01581:02150	CCCTCGTACGCTGTGCTAGGCTGCAAACGTCTTTATGTCTGTGCTGATCAGTGCTGTAGCATCCAGTCAACGAATGTG***AATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATG*** ATGGCAGAGAGAGAGGAAAAGAGGGAGAAGGAGTTCCAACTGAAACTGCTTAAAGTCAGCTCGGCCCTCCATTAGCCTGCGGGTCCATCACCATCGACTTCCTATATAGCTGGGCTTTC
4	B20917	12	00373:01725	ATGGCATATAGGATGCTACCCGATGTGAGCTATGGTAGAGCAGAGTCCTCAGCTTTCTGTTTCATGTCCCTGATCCCTAGAACAGTGCCTGGC***AATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATG*** AAATGATC
4	D40120	12	01611:02792	ATCTTTGAAAGCCTGATCAGAGAGGTTCTCCTGGAGGAGGTGACATATGATCAGGGAATAAGGGAACAAATGATGTGAAAATCTGAAGAATGAAGCTCTGGGCAGAAGATAGAAAGTGCCAAGGGTACAAGATG***AATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATG*** AAAAGGAAGAAA
4	H30201	12	00590:02025	CAGAGGACATGCAGTAGCTGTTTGCAG***AATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATG*** AAAAGCACTGCAAAATTGTGAATGGTCTTATTTTATTACCACTCATTTCTAAATAGTAAAAGAGATTTCCTTATGTTTTGTTCCATGTC
4	D40120	12	01405:00731	GCCTAGACTAGTACCTGGTATATAATTGGCAGTGCATAAACGCTTATGGAACGAATGGATG***AATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATG*** AATTTTCCATTAT

Bait	Sample	Allele	Location	FASTA Sequence
4	D40120	14	00330:00439	TCAAATGGGGGCTGGGAGAAAATTATTTCTCCTGTCAGACCCCAGGAGCCTTCTTTATG***AATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATG***
4	D40120	14	00453:01820	TCAAATGGGGGCTGGGAGAAAATTATTTCTCCTGTCAGACCCCAGGAGCCTTCTTTATG***AATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATG***
4	D40120	14	01020:01996	TCAAATGGGGGCTGGGAGAAAATTATTTCTCCTGTCAGACCCCAGGAGCCTTCTTTATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATG
4	D40120	13	01851:02429	TCAAATGGGGGCTGGGAGAAAATTATTTCTCCTGTCAGACCCCAGGAGCCTTCTTTATG***AATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATG***AAGTGAGGG
6	H30201	12	00424:01228	GTGGTCCTCCAGCTTGCACAGGGTAGGTACTCACTAGGGCTTTGTGA***TGGATGGATGGATGGATGGATGGATGGATGGATGGATGGATGGATGGA***AGAAAAATAAAGGGACCTCCTGTGACCTCTATGGCTTTGTTCTCATGGATACCACTATTTGTGGACTAGAAGCTTTGAGGAACTTGGATCAGTAAAGCCAAGGGAAGTCAGAGGACTGGGGAATCCAATTTTTAACA
6	B20917	12	00271:01307	CCTAGTACAATGGCTGGTACACAATAAGCACTGAATAGATGCATGGGTGATGGATAGAAAAA***TGGATGGATGGATGGATGGATGGATGGATGGATGGATGGATGGA***AGATGACTCTACCCTAAATGGACCAACATGAATTTAAGAAAAGGAATG
6	H30201	12	00209:00355	AGTCTCTGTTTGGGAGGAGGAAGCAGAATTTGAGAGATATTGGAGTTATGGAGCATTGGGAGTGGGGAGTTGTAAGAGCATTGGGTCCAGACTACAGAAATCCAGTGCTATGACACCATCCAGGGCCAAGGAT***TGGATGGATGGATGGATGGATGGATGGATGGATGGATGGATGGA***AGGATGGACGAGTGGATGGATGGATGGG
6	D40120	12	00072:00038	GATTGTTGATTACATATGTGTCTGTCTCCTTCTCTAGACAGTGAGCTATTTGAGGGCAGATGCCCTCCATTTGTCTTCGCATCCTCAGCACTGAGAATAAGATCTGGAGAATGTTTGAGGAGTGAAGAGA***TGGATGGATGGATGGATGGATGGATGGATGGATGGATGGATGGATGGATGGATGGA***CAGATGGACGGATGATGGATGGTGAGTGGATGGATGATGGATGGTGAATGCTTGGAT

Bait	Sample	Allele	Location	FASTA Sequence
6	9010hH	12	00915:01341	GAGACACTGGGTATACAGCAGTGAATGGAATCCCCACTCTCCTGAGGCCTA CATTCTATTGAAGGGACAGTAAACAAAATGAT***TGGATGGATGGATGGATGG ATGGATGGATGGATGGATGGATGGATGGA***CGGTTGAATATCTTTACGTGCCT CCCATTGAAAGACTAAACATGTGTCCAGGTAAGCTTGACTAATGCAGGATA GAAGAGGAGCTTCTTGGGTGCCTGGGTGGCTCAGTTGGTTAAG
6	H30201	12	00467:01607	GTGGGAAATTTTTTGGATGGATGGGTGGATGGG***TGGATGGATGGATGGATG GATGGATGGATGGATGGATGGATGGATGGA***CTGACAGGTAGGTGAATGGAT GGGTGGATGGATAGATGGACAGATGGGTGGATGGACAGATGTGTGGATGG GATGAATGGGCAGATGG
6	H30201	12	00721:00428	GTGGGAAATTTTTTGGATGGATGGGTGGATGGG***TGGATGGATGGATGGATG GATGGATGGATGGATGGATGGATGGATGGA***CTGACAGGTAGGTGAATGGGA TGGGTGGATGGATAGATGG
6	9010hH	12	01075:01620	TTGGACAGATGGATGGACAGTTGGATGGGTGGG***TGGATGGATGGATGGATG GATGGATGGATGGATGGATGGATGGATGGA***GGGCGAGCAGACATCCTAATG AACCGATAGCTTCCAGGTTCTCGGTGCAGGTTGGCATCAGCCTTCATCATG GAGGCTCACGGAGGCCATCCCCTGCAGCGATGGACTGCTTGCTGCCCACTT CCCGATGTGATACAATGCATCCATGAAACTAGGCCCCTTGTGCTTTTGCCTG TCAAACCACAAG
6	B20917	13	00609:01670	GGTGTTTATTGGGATTGAGATGTTCTCTCGAAGCTCAGCAGCAGCACCAAT GGGAGAAAGAGCTCAGGGTTACGGACAAGAGAGCCTGGGTCTTTGTAGAT ACTCACAGGCAATTCCAACCAATGTTGGCTGAATCCAGAAAGGGA***TGGATG GATGGATGGATGGATGGATGGATGGATGGATGGATGGATGGATGGA***TGAAT GTG
6	B20917	12	00441:00074	CTAAATATTTTTGTGTTTCACAGCCCCTAGATAGATGCTACCAAAAGGAAG GTACT***TGGATGGATGGATGGATGGATGGATGGATGGATGGATGGATGGATG GA***TGAATGA
6	9010hH	12	00535:00439	CTCACTGGAACAGAATCTGGCACTTGTCAGATAATGGATGGAGAGA***TGGAT GGATGGATGGATGGATGGATGGATGGATGGATGGATGGATGGA***TGGGTGGG T

Bait	Sample	Allele	Location	FASTA Sequence
6	9010hH	12	00278:02401	CAAAGAATGAATCTCTGTCTTTACCTTAGGGCCTTTCCCGAAGGAGTATCTC AGTGTCTTCTGAGTAAAAGGG***TGGATGGATGGATGGATGGATGGATGGATG GATGGATGGATGGATGGA***CAAATGAAAGAATAAAGGATTGAATACATATA GAGTCAGTTAA
6	D40120	12	00087:00590	ACTGGTGAGGATGCCGTGAAGACTGCTGAAATGATAACA***TGGATGGATGGATGGA TGGATGGATGGATGGATGGATGGATGGATGGA***TGGTTAGAATATTATGTAACTTTA GTTGACAAAGCAACAGCAGGGTTGGAG
6	B20917	12	00322:01269	CTTCTTCCTTAAAAATTAGTTACTAATATGCGGATGGATGGAAGGATGCAT GGAAAGACAATAGGA***TGGATGGATGGATGGATGGATGGATGGATGGATGG ATGGATGGATGGA***GTACTGGCTGTTTCATTATATGCAGTAAATACAATTAA ATCCCTGATTAACTCCT
6	9010hH	12	00380:02349	GATGGATGGATGGAGGGATGGATGGAAGGAAGGATAGA***TGGATGGATGGA TGGATGGATGGATGGATGGATGGATGGATGGATGGA***T
6	H30201	12	00305:02286	CACTCCTGTTCTCCTGCTTAGTAAAGTGCCCTGCACACTGTAGATGCTCCCA AATAC***TGGATGGATGGATGGATGGATGGATGGATGGATGGATGGATGGATG GA***CTGAGGGAGGGAGGGATGGAT
6	H30201	12	00477:01292	CTCAACAGTGCAAGGAAGGAAGTGCCAAGATGTTCTGGGATTGCATGTCTC CTGGACACTGAAAGAAAGAGGTCCATTGTTATAAATATGCATGTGTGGGTG GTTGAGTGGATGGATGGATGGGTGGGTGGTTGAGTGGATGGGTATATGGCT GTGAATGGGTGAGTAGATGGGTGAATGGATAAGTGGGTTGATGAATGTG***T GGATGGATGGATGGATGGATGGATGGATGGATGGATGGATGGATGGA***TGAG TAGGTGGGTGGA
6	B20917	12	00404:02338	GGTTAGATAGGTAGGAGGAAGGATGGAAGGA***TGGATGGATGGATGGATGG ATGGATGGATGGATGGATGGATGGATGGA***CGAGATGGA
6	9010hH	12	01141:01505	TTTAATCCCTCCTAGCAGCTGAGGTGTTATACATCAAAGTGCCTGCTACAGT GGCCCATAGTAAGCCTCAACAAATATTTGTTATGGATGGATGGGTGGGTGA GTGGATGGATGGTTGGATGGCTGGATGGTTGGATGGATGGAGAGATGGAT GGAAGAA***TGGATGGATGGATGGATGGATGGATGGATGGATGGATGGATGGA TGGA***TGGTGGATGGATGATGG

Bait	Sample	Allele	Location	FASTA Sequence
6	H30201	12	00297:02142	GAGCCTCATCAATGTTTGATGAGTGGATGCACCTATGAACAAC***TGGATGGATGGATGGATGGATGGATGGATGGATGGATGGATGGA***CAAAA
6	H30201	12	00062:00492	AGCTGGGGGATCTAAGTATGCAGCAGTGAACACAAAATTTTTCTCACTCCTTGTTTTATATTTTCTGTGTGCTCTTCTAACCACTTGACTT***TGGATGGATGGATGGATGGATGGATGGATGGATGGATGGATGGA***TTCATGTCTCTTTCCTAA
6	B20917	12	00455:01402	AGGGAGGGAAGGCAAAAATAATCTACATAGGCTACATTCACTGGTTTGTCCTGGACACACACAGTGCTTAGTAC***TGGATGGATGGATGGATGGATGGATGGATGGATGGATGGATGGA***CGGACAGGATGGGTGGGTGGATGGATGGATGGATGG
6	9010hH	12	00898:02606	ACTCAATGCACTTTCAGTTTAACCACAAAAGTATAGGA***TGGATGGATGGATGGATGGATGGATGGATGGATGGATGGATGGA***TAAATGGATGGATGGATCAATGGATGGTCGACAATAATAGATACATGAT
6	9010hH	12	00512:01591	ATGGATATT***TGGATGGATGGATGGATGGATGGATGGATGGATGGATGGATGGATG GATGGA***TAAGTAGATGGGATGATGGATGGAAGGATGGACGGACAGATGGATGATGGATGGAGGGATGGAGAATAAATGGTAAACTAGTAATGTGTACCTGGTTCATCTGGATGCCATCCATG
6	9010hH	12	01556:02582	AGATAGATGGATGGGTGGATGAATGGATAGATGGAAGGAA***TGGATGGATGGATGGATGGATGGATGGATGGATGGATGGATGGATGGA***TAGAAGGGTGGATGGATAAAACTCAAACTCAAGAGTTC
6	H30201	12	00021:01613	AGAAATATACCACACTACCCATAGATGGATGGATGCA***TGGATGGATGGATGGATGGATGGATGGATGGATGGATGGATGGATGGA***TGAGTGGATGGGTGGATGGATAAGCAAATGAATGGGTGAATGGATGGGTGTATGAATGGATAACAACGTATGTGGTCCATAGGACTGCAGCGTCAGAAAAACAAGGAAACTGGCAGAAGTTGAAAAAG
6	9010hH	12	00618:01236	ATGGATATT***TGGATGGATGGATGGATGGATGGATGGATGGATGGATGGATGGATG GATGGA***TGGGGACAGCTGCACAGCTAGGGGA

Bait	Sample	Allele	Location	FASTA Sequence
6	9010hH	12	00707:02717	CTCCTCTGAATCTTCACCATGCTGAAGCCACATGGTGCCAGGACCAGTTGT AGTCACCACTGAATCACCCACCCTCAGCAGGGTCATCTCATAGCTAGTGAT ACTCAACAGGTATTCACTGGT***TGGATGGATGGATGGATGGATGGATGGATG GATGGATGGATGGATGGA***TGGTTAGATGGATGGATAGATGGATGGATGGAT GGATGGATGTGTGGATGGGTGGATGGACGGAT
6	D40120	13	00116:01009	ATACCTACTGTCACCCCCACCCCATTCCTGACCTACGTAATTCCTACTTGAGCTTCC CCAGCAATGAAGATTGCCACGGATTCCTTCATTGCACCCCCAGTGTACCCAAGGGA ***TGGATGGATGGATGGATGGATGGATGGATGGATGGATGGATGGATGGA***TGGATTGA TGGATGGATGGATGATGTGCAAAGATGCTGAGAGCACCCCAGGCATAGCTAAC
7	H30201	12	00259:00735	CTTCAGGCTCTGCAGTATCCAAAGGTGAACGAGAAGGAAGTTGGAATGGA TGTGAGGTAATGCACCGTATCTATCATAATGTTCTCAACATCTGAATAACA AAGTCTCAGAA***AGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATA GATAGAT***AACAGACAGAGACAGAGAGACATGGTGTTGGGTGGAAGGGGGT TATGCAAGAAATGAATTCAAGAAAATTGGAAGAAATGAATTTCAAC
7	D40120	13	00111:01688	ATTATAGTTTGTAGGTAGCCCTATAGTTAGCCATGTACCCTGAAAGTTAGGCTTGA ACATAT***AGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGA T***ACATGCACACACACACACACACACACACACACA
7	D40120	12	00056:00530	GCTATAGAAAGAAACAAAACCTTAGATATAT***AGATAGATAGATAGATAGATAGATA GATAGATAGATAGATAGATAGAT***ATACACGCATACACACACACACAACTTTGGGAG AT
7	H30201	12	00174:00856	GTTGATTAGTTAATTAAATCTAGAGTGGCGTTAAGTTTAATCTATGAGATAT AGATAGATGATAGATAGATAGATGATAGATAGATGAT***AGATAGATAGATAG ATAGATAGATAGATAGATAGATAGATAGATAGAT***GATAGAAAGAAAGATGG ATGGAGGGATGGATGGATGGATGGATGGATGGATGGATGGATGGGTGGAT GGATGGATGGATGGTTGGTTGGATGGATTGGAGAGATAGATGATAGATAG GTAGATAGAT
7	H30201	12	00200:00431	CTTCCAAGTGGTGCTGCCTGCTGGCCCAAGGACCACACTTCGTGTTGTATTG GTATGGAAATAAATAAGTAGGTAGGTAGGTAGATT***AGATAGATAGATAGATAGAT AGATAGATAGATAGATAGATAGATAGATAGAT***GATAGATAGATGATAGATA GAGCGAGCCAGCCGTGAGCAGGGTAGAAGCAGGCATTCCTGTGCATCTCG TTTCCCAGCTCAGTGTGTGTGCTTC

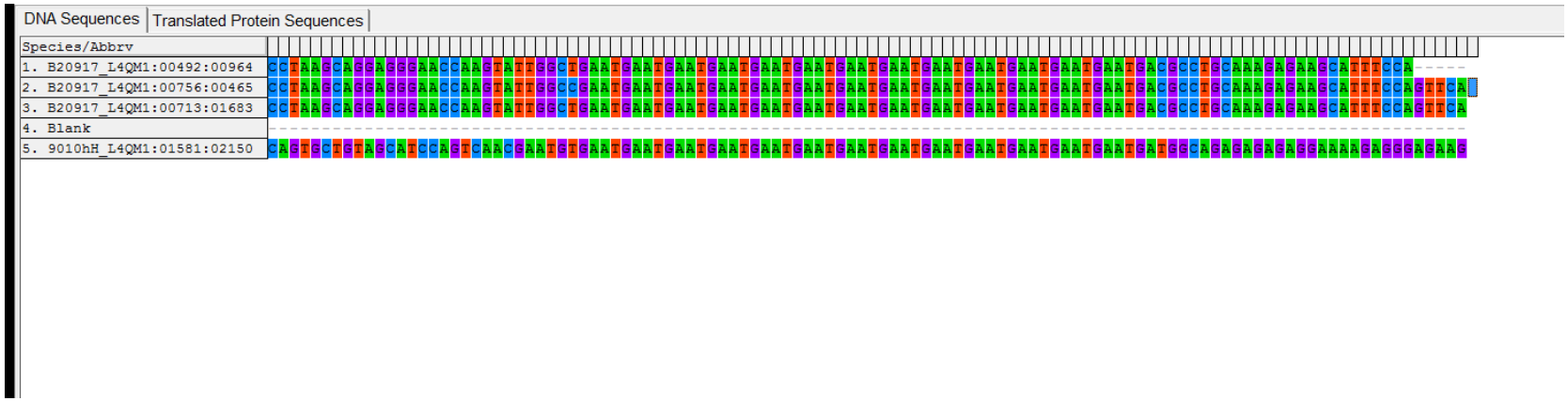
Bait	Sample	Allele	Location	FASTA Sequence
7	9010hH	12	00807:02415	ATAGATAGATAAT***AGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGAT***GATATAGTGATAGAGATGGATGAATGGATGGATGGATGGATGGATGGATGAGTAGATA
7	9010hH	12	00705:01286	ATTATAGTTAGATAAAT***AGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGAT***GATGGCTGATAGATGGATGGATGGATGGATGGATGGATGGTTGGATGGATGGTTAGGTGGGTGGGTGGGTGGATGGATGGATGAACAGACGGATAGATGGATGGATGGATAGGT
7	9010hH	12	01118:01382	GGTGGTTGGTAGGGTGAAGAGACAGATGTAGATAGATGAT***AGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGAT***ATAGATAGATAGATAGATAGATAGATAGATAGATAGAT
7	9010hH	12	02058:02175	GGTGGTTGGTAGGGTGAAGAGACAGATGTAGATAGATGAT***AGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGAT***ATAGATAGATAGATAGATAGATAGATAGATAGATAGA
7	D40120	12	00136:00688	ATAGCCAAAATATGAAAACAATCTAAGTGTTCACAGATGGGTGAATGGGTAAAGA***AGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGAT***AGACAGATGATAGATTGACTATAGGTATAACTAGACATATAGATATAA
7	D40120	12	00132:01455	CGCTTGATAACGGATTGACTGTAGAACAAGTAGATGTCAAGGATTGTTATAGGTTTCTGGATAGATAGTGGCTATCTGGATGGATGGATGGATGGATGGAT***AGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGAT***
7	D40120	12	00067:00331	GCTTAACAACTGCACCACCAGGTGCCCCTGGAGATTTTTTAATATTGTAAATTGTCATCTGGACTCTGGGATGCCAGGAGATAGATAGATGAT***AGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGAT***GATAGATCACTTATTTAGAG
7	B20917	12	00363:00761	AGTCCTTTCTGCTTTAGAAACAAAGCCAGTAAAGAATTTAATGTCTTGGGCAATATGCAGGAACTTGGAGTTAACTTATCTGAATGACTCTTCCCATGACCTATCAGACTCACAT***AGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGAT***ATGAT
7	9010hH	13	01560:03045	ATATAGATAGATGATAGATAATGATAGCTAGCTAGCTAGCTAGCTAGATGAT***AGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGAT***GATAGATAGATAGATACAGATAGATGATAAGAAAATGATAGATGATAGATAATGATTGAT

Bait	Sample	Allele	Location	FASTA Sequence
7	9010hH	12	01753:01282	ATATAGATAGATGATAGATAATGATAGCTAGCTAGCTAGCTAGCTAGATGA T***AGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGAT***GAT AGATAGATAGATACAGATAGATGATAGAAAATGATAGATGATAGATAATG ATTGAT
7	B20917	12	00316:00564	TCTTGATCTCAGCTCAGGTCTTGATCTCAAGGTCATGAGCTCAAGCCCTATG TTGGGCTCCACACCCAACATTAAATAGAC***AGATAGATAGATAGATAGATAG ATAGATAGATAGATAGATAGATAGAT***AGGATTGGGAATTCGAACTAAAGC
7	D40120	12	00106:01578	GCAGGCTTTGGGGTAGCCCTTCCCTGATTAGAACCCTATGGTGAAGAGAGGGACTT TTTTAAAATTAAGTTTAGTTATGAGAGTGTATTGGTTAGGGCTTTCCTAAGAAGAG AACCAATATG***AGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGA T***G
7	9010hH	12	00220:02501	CTCTTGATTGATTTTTGGCTCAGGTCATGATCTCTCTCTCCCTGTGCCCCTCC CCTACTCTCTAAAAT***AGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATA GATAGATAGAT***AAACA
7	B20917	13	00473:02343	CCAGGATATTCAGCTTGAACTTCCGGTGTGAGATAGACTGCATTTCCCTGA AAGAGCTACAGAGGAGTTTCTAGTCTCATGTTCTTCCAGAGCCTTGAGGGG CCTTGAGGTGCCCCATCAAGGTAGAGCACTGGG***AGATAGATAGATAGATAG ATAGATAGATAGATAGATAGATAGATAGATAGAT***AGACAGACA
7	B20917	13	00345:00976	TGCCAATAAGCTTTTGTTTTGAATGGTAGGTTGTGGGTGCTTGGTAGGTACA T***AGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGAT***AGA TCGACCCACCAACCAGGCCAGGCATGCCACTAATGATGAGTGTCAGGTA
11	B20917	15	01624:01466	GGCTCAGTTGGTAGAGTATACGGCTCTTGATCTTAGGGTCGTGAGTTCAGG CCCCACGTTGGCCATAGAGATTACTG***AAAGAAAGAAAGAAAGAAAGAAAGA AAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG***AAA
11	B20917	14	01784:01790	GGCTCAGTTGGTAGAGTATACGGCTCTTGATCTTAGGGTCGTGAGTTCAGG CCCCACGTTGGCCATAGAGATTACTG***AAAGAAAGAAAGAAAGAAAGAAAGA AAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG***AAA
11	D40120	13	00396:01267	ACCAGCAATTCACAGAAGAAACCGGAGTATCCAGTTAACAGCAGATGCTCAATCA CACTTGCAAGCAAGAGAATTAG***AAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGA AAGAAAGAAAGAAAGAAAG***AAA

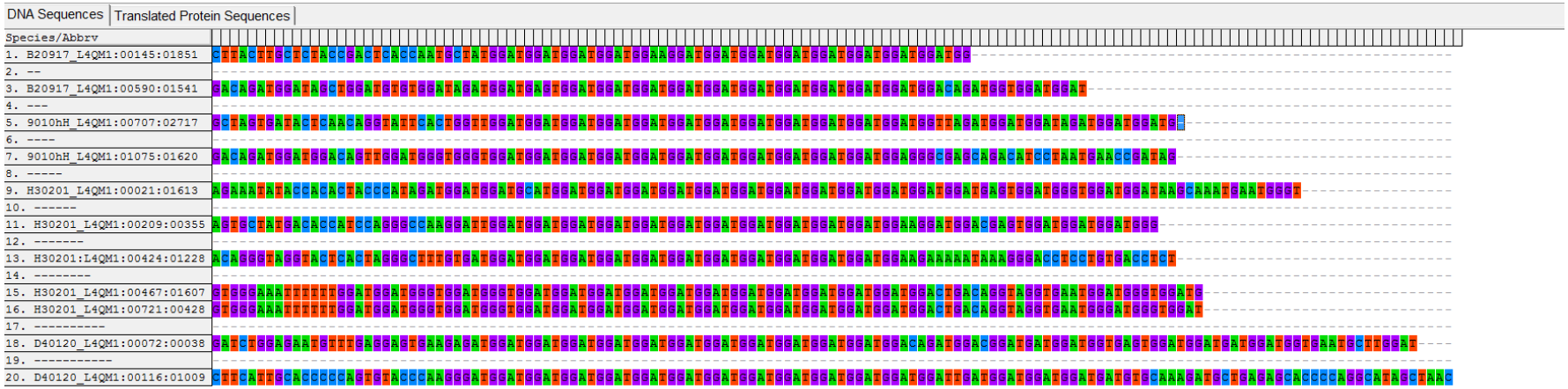
Bait	Sample	Allele	Location	FASTA Sequnce
12	D40120	12	00835:00580	CAGACTCCCTATTGAACAGAGAACCCAATGCAGGGCTTGATCCCAGGACCCTGAG CAGGGTGCACACACACACTCTCTCTCTCAAA***TACATACATACATACATACATACATA CATACATACATACATACATACA***TTAAAATCTTTAAAAATTAATCAAAGG
12	D40120	12	01315:02002	CAGACTCCCTATTGAACAGAGAACCCAATGCAGGGCTTGATCCCAGGACCCTGAG CAGGGTGCACACACACACTCTCTCTCTCAAA***TACATACATACATACATACATACATA CATACATACATACATACATACA***TTAAAATCTTTAAAAATTAA
12	H30201	12	01426:01321	ATGTCCCTTCACATGTGTCCCTTCTACAGTAAGTTAAAAGGAAGCTGGAAC TGGATTTCACCCTGGGTTCATCGTTCCTGTCCCTGGGAAAA***TACATACATAC ATACATACATACATACATACATACATACATACATACA***TAAAGTCACGCAGGC CTTCTGCTTGCCAGAATGGAC
12	H30201	12	01664:03211	ATGTCCCTTCACATGTGTCCCTTCTACAGTAAGTTAAAAGGAAGCTGGAAC TGGATTTCACCCTGGGTTCATCGTTCCTGTCCCTGGGAAAA***TACATACATAC ATACATACATACATACATACATACATACATACATACA***TAAAGTCAC
12	*D40120*	*12*	*01855:01624*	*ACTAGAGCCTGCTTCTTCCTCTCCCACTCCCCCTGCTTGTGTTCCCTCTCTTGCTGGCTC TCTGTCTCTGT**TACATACATACATACATACATACATACATACATACATACATACATACATACA** TAAAAATCTTTAAA*
12	*H30201*	*12*	*00140:01192*	*ACTCCCCCTGCTTGTGTTCCCTCTCTTGCTGGCTCTCTGTCTCTGT**TACATACAT ACATACATACATACATACATACATACATACATACATACATACA**TAAAAAT*

APPENDIX K

Alignment of Flanking Regions and Repeat Unit Using the MEGA software v 5.2



Appendix Figure 1: FASTA sequences that contain both start and end flanking regions in individuals B20917 9010hH, D40120, and H30201 for repeat motif AATG (bait TNRSTR-4). Sequences from different individuals have been separated by a space between sequences



Appendix Figure 2: FASTA sequences that contain both start and end flanking regions in individuals B20917 9010hH, D40120, and H30201 for repeat motif TGGA (bait TNRSTR-6). Sequences from different individuals have been separated by a space between sequences

APPENDIX L

Computer script used for algorithm mining of twelve RNA bait repeat motifs

Algorithm mining was performed using a three-step process:

1. FASTA files converted to custom reference libraries
2. Repeat motif sequences converted to FASTQ files
3. Custom reference libraries and FASTQ files containing repeat motif sequences were aligned using the Bowtie alignment software for each individual genome

In step 1, custom reference libraries are built manually. Each motif contains four lines.

Line 1 begins with a '@' character and is followed by a sequence identifier and an *optional* description.

Line 2 is its raw sequence letters.

Line 3 begins with a '+' character and is *optionally* followed by the same sequence identifier (and any description) again.

Line 4 encodes the quality values for the sequence in Line 2, and must contain the same number of symbols as letters in the sequence.

For example, a motif in motif.fastq contains four lines.

@TNRSTR-1

TGTCTGTCTGTC

+

!"#$%& !"#$%&

In step 2, bowtie2 was used for building index for customized reference library.

bowtie2-build B20917.fasta B20917

In step 3, bowtie2 was used to align motif with the customized reference library which is original FASTA file.

bowtie2 -x B20917 -U motif.fastq -S B20917.sam

The sam files are the output files.

REFERENCES

[1] Q. Santana, M. Coetzee, E. Steenkamp, O. Mioyeni, G. Hammond, M. Wingfield, and B. Wingfield. Microsatellite discovery by deep sequencing of enriched genomic libraries. BioTechniques. 2009. **46**(3); 217-223.

[2] D.M. Bornman, M.E. Hester, J.M. Schuetter, M.D. Kasoji, A. Minard-Smith, C.A Barden, S.C. Nelson, G.D Godbold, C.H. Baker, and B. Yang. Short-read, high-throughput sequencing technology for STR genotyping. BioTechniques. 2012. 1-6.

[3] L. Waits, D. Paetkau, and C. Strobeck. Genetics of the Bears of the World. In: International Union for the Conservation of Nature and Natural Resources, editors. Status survey and conservation action plan: bears. Cambridge, UK: Publications Services Unit, 1999, 25-32

[4] M. Gymrek, Y. Erlich. Profiling short tandem repeats from short reads. Methods Mol Biol. 2013. **1038**; 113-135.

[5] D. Paetkau, L. Waits, P.L. Clarkson, L. Craighead, E. Vyse, R. Ward, and C. Strobeck. Variation in genetic diversity across the range of North American brown bears. Conserv Biol. 1998. **12**(2); 418-429.

[6] D. Paetkau, G.F. Shields, and C. Strobeck. Gene flow between insular, coastal, and interior populations of brown bears in Alaska. Mol Ecol. 1998. **7**(10); 1283-1292.

[7] R. Andreassen, J. Schredel, A. Kopatz, C. Tobiassen, P.M. Knappskog, S.B. Hagen, O. Kleven, M. Schneider, I. Kojola, J. Aspi, A. Rykov, K.F. Tirronen, P.I Danilov, and H.G. Eiken. A forensic DNA profiling system for Northern European brown bears (*Ursus arctos*). Forensic Sci Int Genet. 2012. **6**(6); 798-809.

[8] E.E. Puckett, T.V. Kristensen, C.M. Wilton, S.B. Lyda, K.V. Noyce, P.M. Holahan, D.M. Leslie Jr., J. Beringer, J.L. Belant, D. White Jr., and L.S. Eggert. Influence of drift and admixture on population structure of American black bears (*Ursus americanus*) in the Central Interior Highlands, USA, 50 years after translocation. Mol Ecol. 2014. **23**(10); 2414-2427.

[9] R. Ogden. Unlocking the potential of genomic technologies for wildlife forensics. Mol Ecol Resour. 2011. **11 Suppl 1**; 109-116.

[10] D. Paetkau and C. Strobeck. Microsatellite analysis of genetic variation in black bear populations. Mol Ecol. 1994. **3**(5); 489-495.

[11] D. Paetkau, W. Calvert, I. Stirling, and C. Strobeck. Microsatellite analysis of population structure in Canadian polar bears. Mol Ecol. 1995. **4**(3); 347-354.

[12] L. Craighead, D. Paetkau, H.V. Reynolds, E.R. Vyse, and C. Strobeck, Microsatellite analysis of paternity and reproduction in Arctic grizzly bears. J Hered. 1995. **86**(4); 255-261.

[13] D. Paetkau, L. Waits, P.L. Clarkson, L. Craighead, and C. Strobeck. An empirical evaluation of genetic distance statistics using microsatellite data from bear (Ursidae) populations. Genetics. 1997. **147**(4); 1943-1957.

[14] H.G. Eiken, R.J. Andreassen, A. Kopatz, S.G. Bjervamoen, I. Wartiainen, C. Tobiassen, P.M. Knappskog, P.E. Aspholm, M.E. Smith, and J. Aspi. Population data for 12 STR loci in Northern European brown bear (*Ursus arctos*) and application of DNA profiles for forensic casework. Forensic Sci Int Genet Suppl Series. 2009. **2**(1); 273-274.

[15] V.E. Kutschera, T. Bidon, F. Hailer. J.L. Rodi, S.R. Fain, and A. Janke. Bears in a forest of gene trees: phylogenetic inference is complicated by incomplete lineage sorting and gene flow. Mol Biol Evol. 2014. **31**(8); 2004-2017.

[16] F.W. Allendorf, F.B. Christiansen, T. Dobson, W.F. Eanes, and O. Frydenberg. Electrophoretic variation in large mammals. I. The polar bear, *Thalarctos maritimus*. Hereitas. 1979. **91**(1); 19-22.

[17] Y.-P. Zhang and O.A. Ryder. Mitochondrial DNA sequence evolution in the Arctoidea. Proc Natl Acad Sci. 1993. **90**(20); 9557-9561.

[18] S.L. Talbot and G.F. Shields. A phylogeny of the bears (Ursidae) inferred from complete of three mitochondrial genes. Mol Phylogenet Evol. 1996. **5**(3); 567-575.

[19] L. Yu, Q.-W. Li, O.A. Ryder, and Y.-P. Zhang. Phylogeny of the bears (Ursidae) based on nuclear and mitochondrial genes. Mol Phylogenet Evol. 2004. **32**(2); 480-494.

[20] C.R. Miller and L.P. Waits. The history of effective population size and genetic diversity in the Yellowstone grizzly (*Ursus arctos*): Implications for conservation. Proc Natl Acad Sci USA. 2003. **100**(1); 4334-4339.

[21] Where do bears live in North America? http://geology.com/stories/13/bear-areas/. Downloaded on 28 February 2015.

[22] L.P. Waits, P. Taberlet, J.E. Sweson, F. Sandegren, and R. Franzen. Nuclear DNA microsatellite analysis of genetic diversity and gene flow in Scandinavian brown bear (*Ursus arctos*). Mol Ecol. 2000. **9**(4); 421-431.

[23] D. F. Williamson. In the black: status, management, and trade of the American black bear (*Ursus americanus*) in North America. TRAFFIC North America. Washington D.C.: World Wildlife Fund. 2002.

[24] H. D. Marshall and K. Ritland. Genetic diversity and differentiation of Kermode bear population. Mol Ecol. 2002. **11**(4); 685-697.

[25] D.L. Garchelis, D. Crider, and F. van Manen. IUCN SSC Bear Specialist Group. 2008. *Ursus americanus*. The IUCN Red List of Threatened Species. v. 2014.3. www.iucnredlist.org. Downloaded on 2 March 2015.

[26] D. Paetkau, S.C. Amstruo, E.W. Born, W. Calvert, A.E. Derocher, G.W. Garner, F. Messier, I. Stirling, M.K. Taylor, Ø. Wiig, and C. Strobeck. Genetic structure of the world's polar bear population. Mol. Ecol. 1999. **8**(10); 1571-1584.

[27] S. Schliebe, Ø. Wiig, A. Derocher, and N. Lunn. IUCN SSC Polar Bear Specialist Group. 2008. *Ursus maritimus*. The IUCN Red List of Threatened Species. v. 2014.3. www.iucnredlist.org. Downloaded on 2 March 2015.

[28] M.A. Cronin, S.C. Amstrup, S.L. Talbot, G.K. Sage, and K.S. Amstrup. Genetic variation, relatedness, and effective populations size of polar bears (*Ursus maritimus*) in the southern Beaufort Sea, Alaska. J Hered. 2009. **100**(6); 681-690.

[29] World Wildlife Fund. Polar Bear. 2015. http://www.worldwildlife.org/species/polar-bear. Downloaded on 3 March 2015.

[30] U.S. Fish and Wildlife Service. Environmental Conservation Online System. 2015. Polar bear (*Ursus maritimus).* The U.S. Fish and Wildlife Service List of Endangered Species. http://ecos.fws.gov/speciesProfile/profile/speciesProfile.action?spcode=A0IJ. Downloaded on 3 March 2015.

[31] M.A. Cronin, S.C. Amstrup, G.W. Garner, and E.R Vyse. Interspecific and intraspecific mitochondrial DNA variation in North American bears (Ursids). Can J Zool. 1991. **69**(12); 2985-2992

[32] G.F. Shields and T.D. Kocher. Phylogenetic relationships of North American Ursids based on analysis of mitochondrial DNA. Evolution. 1991. **45**; 218-221.

[33] E.C. Berglund, A. Kiialainen, and A.C. Syvanen. Next-generation sequencing technologies and applications for human genetic history and forensics. Investig Genet. 2012. **6**(6); 798-809.

[34] E. Zietkiewicz, A. Rafalski, and D. Labuda. Genome fingerprinting by simple sequence repeat (SSR)-anchored polymerase chain reaction amplification. Genomics. 1994. **20**(2); 176-183.

[35] L. Zane, L. Bargelloni, and T. Patarnello. Strategies for microsatellite isolation: a review. Mol Ecol. 2002. **11**(1); 1-16.

[36] A.D. Ellington and J.W. Szostak. In vitro selection of RNA molecules that bind specific ligands. Nature. 1990. **346**(6287); 818-822.

[37] P. Vos, R. Hogers, M. Bleeker, M. Reijans, T. van de Lee, M. Hornes, A. Frijters, J. Pot, J. Peleman, and M. Kuiper. AFLP: a new technique for DNA fingerprinting. Nucleic Acid Res. 1995. **23**(21); 4407-4414.

[38] H. Ellegren. Microsatellites: simple sequences with complex evolution. Nat Rev Genet. 2004. **5**(6); 435-445

[39] A. Urquhart, C.P. Kimpton, T.J. Downes, and P. Gill. Variation in short tandem repeat sequences-a survey of twelve microsatellite loci for use as forensic identification markers. Int J Legal Med. 1994. **107**(1); 13-20.

[40] B. Langmead, C. Trapnell, M. Pop, and S.L. Salzberg. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009. **10**(3); R25.1-.10

[41] M Gymrek, D. Golan, S. Rosset, and Y. Erlich. lobSTR: a short tandem repeat profiler for personal genomes. Genome Res. 2012. **22**(6); 1154-1162.

[42] D.H. Warshauer, D. Lin, K. Hair, R. Jain, C. Davis, B. Larue, J.L. King, and B. Budowle. STRait Razor: a length-based forensic STR allele-calling tool for use with second generation sequencing data. Forensic Sci Int Genet. 2013. **7**(4); 409-417.

[43] J.A. McElhoe, M.M. Holland, K.D. Makova, M.S. Su, I.M. Paul, C.H. Baker, S.A. Faith, and B. Young. Development and assessment of an optimized next-generation DNA sequencing approach for the mtgenome using the Illumina MiSeq. Forensic Sci Int Genet. 2014. **13**; 20-29.

[44] L. Liu, Y. Li, S. Li, N. Hu, Y. He, R. Pong, D. Lin, L. Lu, and M. Law. Comparison of next-generation sequencing systems. J Biomed Biotech. 2012. **2012**(3); 251364.

[45] M. Kreutzer. Allele characterization of ten short tandem repeat loci of North American bears (ursids) using next-generation sequencing.

[46] MyBaits: Sequence Enrichment for Targeted Sequencing, v1.3.8 [Manual]. 2013.

[47] Organic DNA Extraction: Blood, Tissue and Stains (non-sperm), Rev 4 [Protocol]. 2008.

[48] Qubit® dsDNA BR Assay Kits [Protocol]. 2011.

[49] Agilent High Sensitivity DNA Kit Guide [Manual]. 2013.

[50] NEBNext® fast DNA fragmentation and library prep set for Ion Torrent™. [Manual]. 2013.

[51] Select-a-Size DNA Clean and Concentrator™ [Manual]. 2014.

[52] Qiagen. QIAquick PCR purification. [Manual]. 2010.

[53] Ion PGM™ Template OT2 400 Kit [Protocol]. 2013.

[54] Ion PGM™ Sequencing 400 Kit [Manual]. 2013.