Sage, Kelly A., Multiplex of INDELs for Human Identification Markers. Master of Science (Forensic Genetics), April 2016, pp. 30, 2 tables, 10 illustrations, 29 references.

Forensic scientists commonly use short tandem repeat (STR) loci when comparing an evidentiary profile to that of a reference profile. In commercially available STR kits, the amplified products tend to range from 100- 500 base pairs (bp) in length. For genomic DNA of degraded biological samples, the fragments are usually 180-200bps or less. Therefore, degraded biological samples may not produce a full STR profile. Another viable option has been proposed to enable successful typing of some degraded DNA samples. Insertion/ deletion (INDEL) polymorphisms are intergenic regions of the genome in which amplified products can be smaller in length than most STRs. Using highly discriminating markers is desirable to distinguish individuals. A multiplex panel of human identification (HID) INDEL markers that can individualize people would be beneficial. This project tested the hypothesis that INDELs, which can be used to identify individuals with high discriminatory power, can be developed as a multiplex PCR approach. To test this hypothesis, primers were designed and multiplexed together to amplify specific INDELs that have been previously identified to be suitable for human identity testing purposes.

Multiplex of INDELs

for Human Identification Markers

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

Kelly A. Sage, B.S.

Fort Worth, TX

April 2016

ACKNOWLEDGMENTS

I would like to express my appreciation to Dr. Bobby LaRue for his encouragement and support. His "pep talks" pushed me to give my best and you taught me that a lot can get accomplished when Starbucks is involved at a meeting. His help and guidance inspired me to produce a quality project I am proud to have been a part of. I would also like to thank Dr. Bruce Budowle for always going above and beyond what was required of him. All of his help, advice and guidance throughout this project and throughout my schooling have been greatly appreciated. A special thanks to my committee members, Drs. Patricia Gwirtz and Kunlin Jin for their valuable comments and advice which helped make my paper a success. A big thank you to Jonathan King, M.S. for humoring all of my questions and providing direction throughout my project. If it weren't for him, I would still be staring at 98 tubes of dried primers and not know where to start. I would like to thank Rachel Wiley for always being there for me, I would not have survived graduate school without her. Also a huge thanks to my partner in crime, Sarah Sturm. I could not think of anyone else I would have wanted to share lab space with. Finally, many thinks to my parents and family for their continuous support for my education and continual encouragement for my success.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

CHAPTER 1

INTRODUCTION

For several decades, short tandem repeats (STRs) have been the generally accepted markers of DNA analysis. STRs are highly polymorphic due to high variability with regards to the number of repeats among alleles in a population (1). These repeat length markers are able to be amplified with ease using the method known as the polymerase chain reaction (PCR) (1, 2). After amplification is completed, the fragments can be separated using capillary electrophoresis (CE) followed by genotype analysis accomplished *in silica* (3, 4). To provide statistical support for the analysis of STRs, population databases have been established to assist forensic analysts in determining the probative value of the genetic evidence (5).

The process of PCR allows for millions of copies of specific DNA sequences to be generated. By using primers, i.e., short oligonucleotide sequences, designed to be complementary to target regions flanking the area of interest, hybridization can occur to a specific location in the DNA. A heat-stable DNA polymerase, Taq polymerase, is then able to copy the DNA sequence by adding nucleotides to the primer. The nucleotides added are complementary to the target strand in which the primer has hybridized. This process repeated many times results in an exponential increase in the number of copies of the specific DNA target sequence (2). For STR amplification, the primers produce amplicons that are from 100- 500 base pairs (bp) in size. These primers are labeled with a fluorescent dye and thus incorporate the dye into the amplicons (1).

After amplification is complete, genotype analysis is accomplished using the CE. The DNA molecules are electrokinetically injected into a capillary in which voltage is applied causing the negatively charged DNA fragments to move through a liquid polymer. The fragments are separated by size as they migrate through the capillary with the smaller fragments migrating faster than the larger fragments (3, 6). At a particular point in the capillary a laser excites the fluorescent dye and a camera captures the emission wavelength, matching the time of the amplicon's passage. An internal lane standard is used to determine the size of the amplicons (6). The various emission wavelengths of the dyes allow for the analysis of several PCR amplicons at the same time. The wavelength peaks can be visualized by using a software system such as Gene Mapper ID-X (Figure 1).

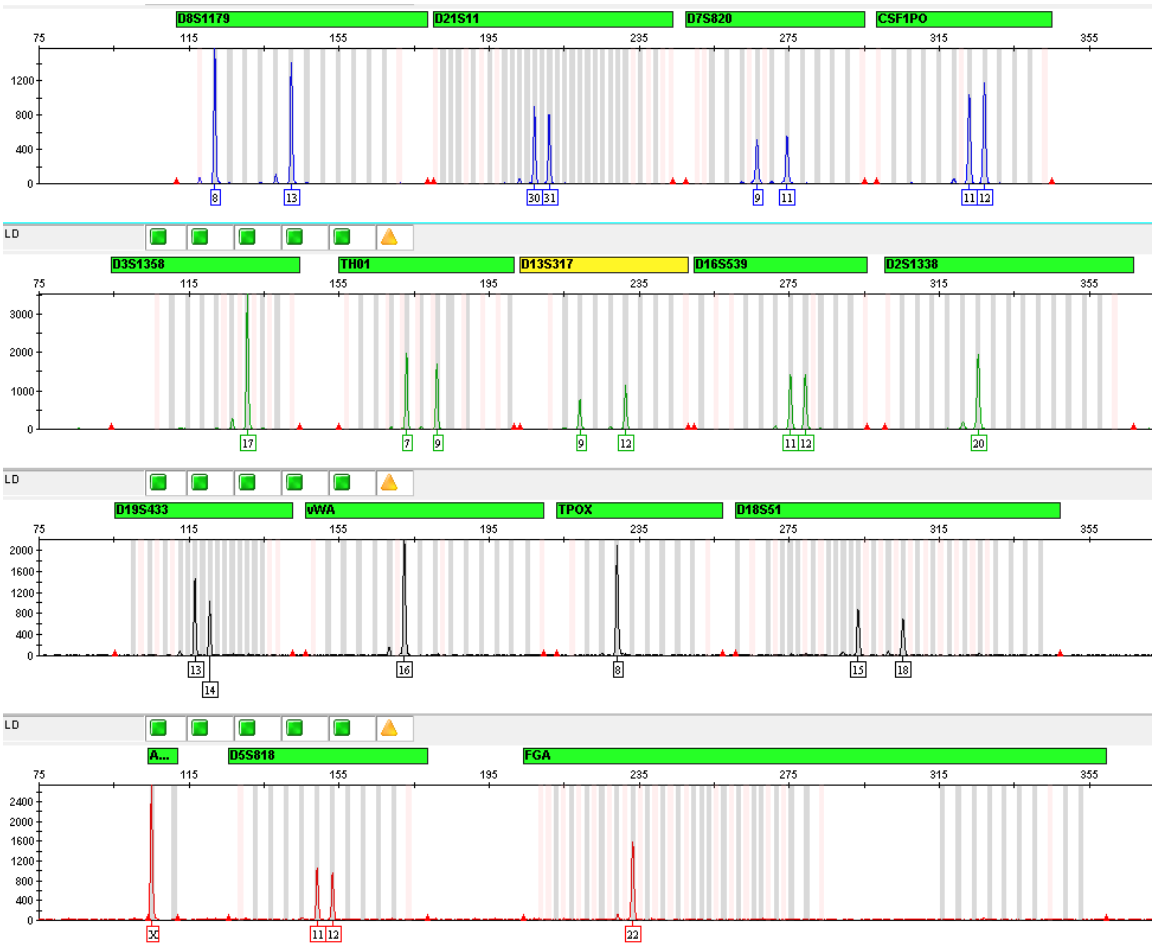


Figure 1. An image from GeneMapper ID-X v1.2 of an STR profile amplified with Identifiler Plus on a 3500xL Genetic Analyzer. The different labeled amplicons are detected by their attached fluorescent dyes and represented in an electropherogram.

Upon genetic analysis, a profile can be generated. The profiles of evidence and reference samples can be analyzed and compared to determine whether there is a match or there is an exclusion. If a match is discovered and the profile is from a single source, a random match probability (RMP) can be calculated. The RMP is the probability of a randomly selected person in a population having an identical genetic profile as that seen in the evidentiary profile. By using allele frequencies generated from a population database, the RMP can be calculated for each locus. A combined profile frequency is calculated by multiplying the locus genotype frequencies together (7, 8).

With time and exposure to environmental conditions, DNA from biological samples will begin to degrade. Degraded DNA is a challenge for forensic analysis because allele drop-out (disappearance of an allele) can occur and lead to an incomplete STR profile (9, 10). Figure 2 shows a STR profile of DNA from a bone sample from the 1890's. Due to degradation, only a few loci were able to be typed.

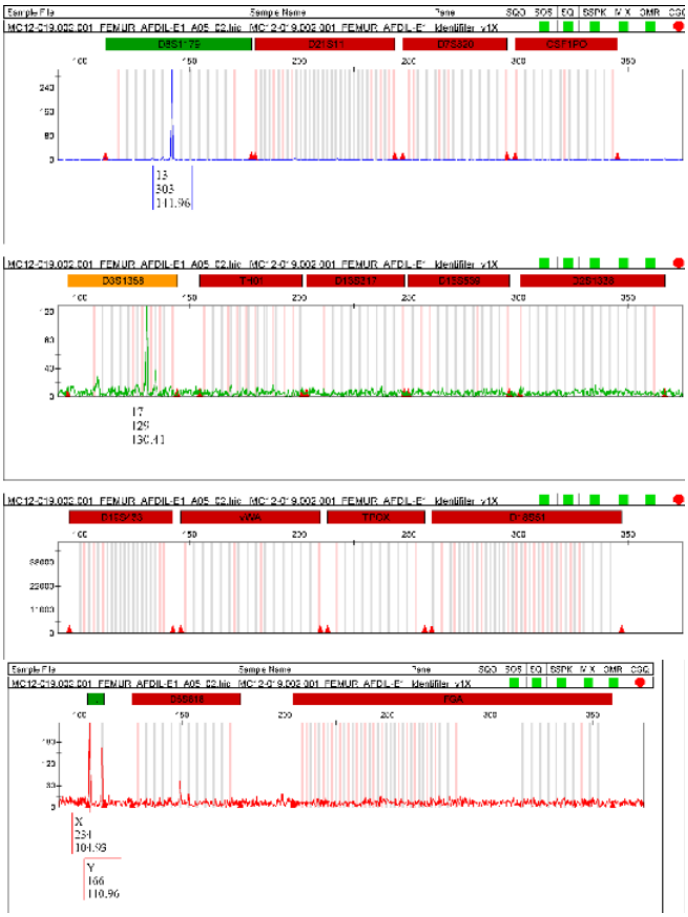


Figure 2. 120 year old bone typed with STR markers from the Applied Biosystems AmpFlSTR® Identifiler® PCR Amplification Kit (11).

Multiple marker systems that have been developed to try to overcome the problem of analyzing degraded samples. Single Nucleotide Polymorphisms (SNPs) are bi-allelic marker systems that are single nucleotide changes in the DNA sequence (12, 13). The amplicons for SNPs are able to be designed to be much smaller in length than STRs, approximately as short as

55 bp, which is the size limit for current PCR systems. Since SNPs are not repeat elements, there are no slippage artifacts (i.e., stutter) generated. However, SNP typing requires detection of a single base substitution which relies on burdensome chemistries and/or instrumentation not available in forensic casework laboratories.

Another marker system that was developed for the analysis of degraded samples is insertion/deletion polymorphisms (INDELs). INDELs are length polymorphisms of insertions or deletions in the DNA sequence as determined by comparison to a known sequence (14). There have been multiple INDEL panels developed for use in forensic identification (15-19). Like SNPs, the amplicons can be designed to be very small, and slippage artifacts are non-existent since INDELs are also non-repeating polymorphisms. However, unlike SNPs, INDELs have amplification products of varying sizes which allow them to be analyzed by size separation, similar to that of STRs, using CE (14). With CE being the present method for STR analysis in forensic laboratories, the current infrastructure would not have to change to employ the use of INDELs. Degraded samples have been analyzed using this marker system (16). The same 120 year old bone sample that was displayed in Figure 2 is shown in Figure 3 using an insertion/ null allele (INNUL) marker system, which is a type of INDEL.

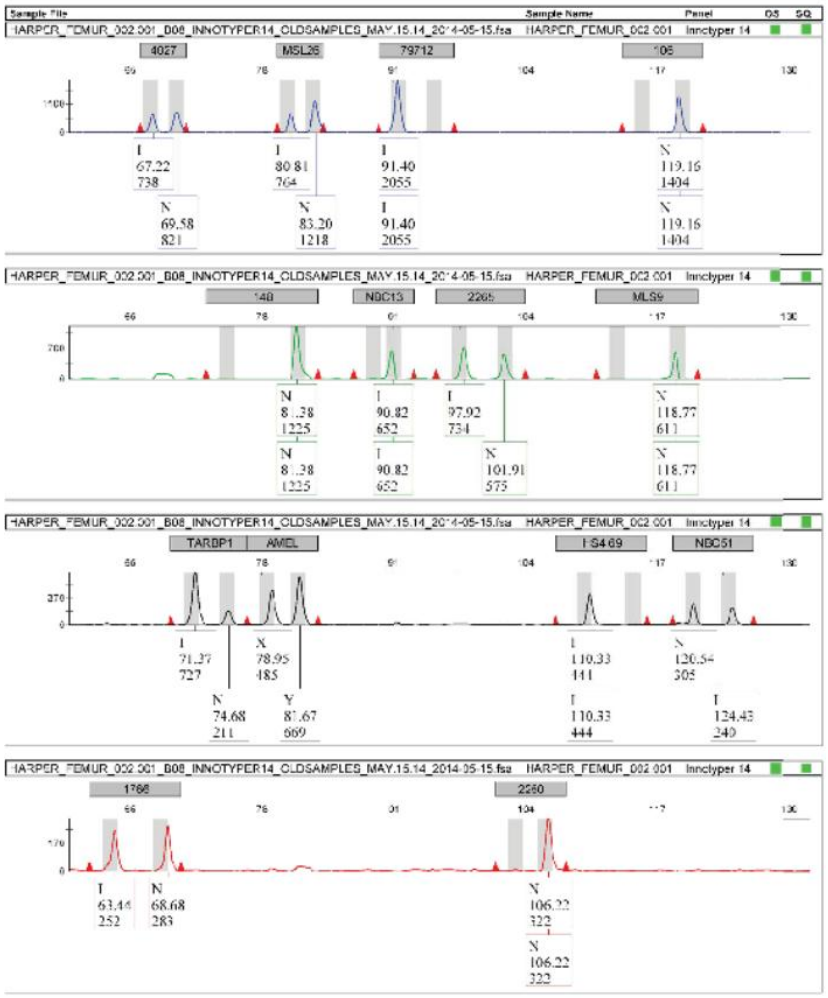


Figure 3. An electropherogram of a120 year old bone sample amplified with an INNUL multiplex composed of 14 different markers.

INDELs have been used successfully for human identification (HID) purposes. Pereira *et al*. (14) were able to generate complete profiles using a low DNA concentration and had full genotyping accomplished on degraded samples with their indel-plex. LaRue *et al*. (15) described a panel of 38 INDELs that have a RMP of at least 10^{-16} and an extended panel containing 49 INDEL markers that have a RMP of at least 10^{-19}. Seong *et al*. (19) performed population genetics analyses of a South Korean sample population using Investigator DIPplex (Qiagen) and found an RMP of 2.84×10^{-11}. An example of a degraded DNA sample analyzed using an INDEL system is shown in Figure 4 (15).

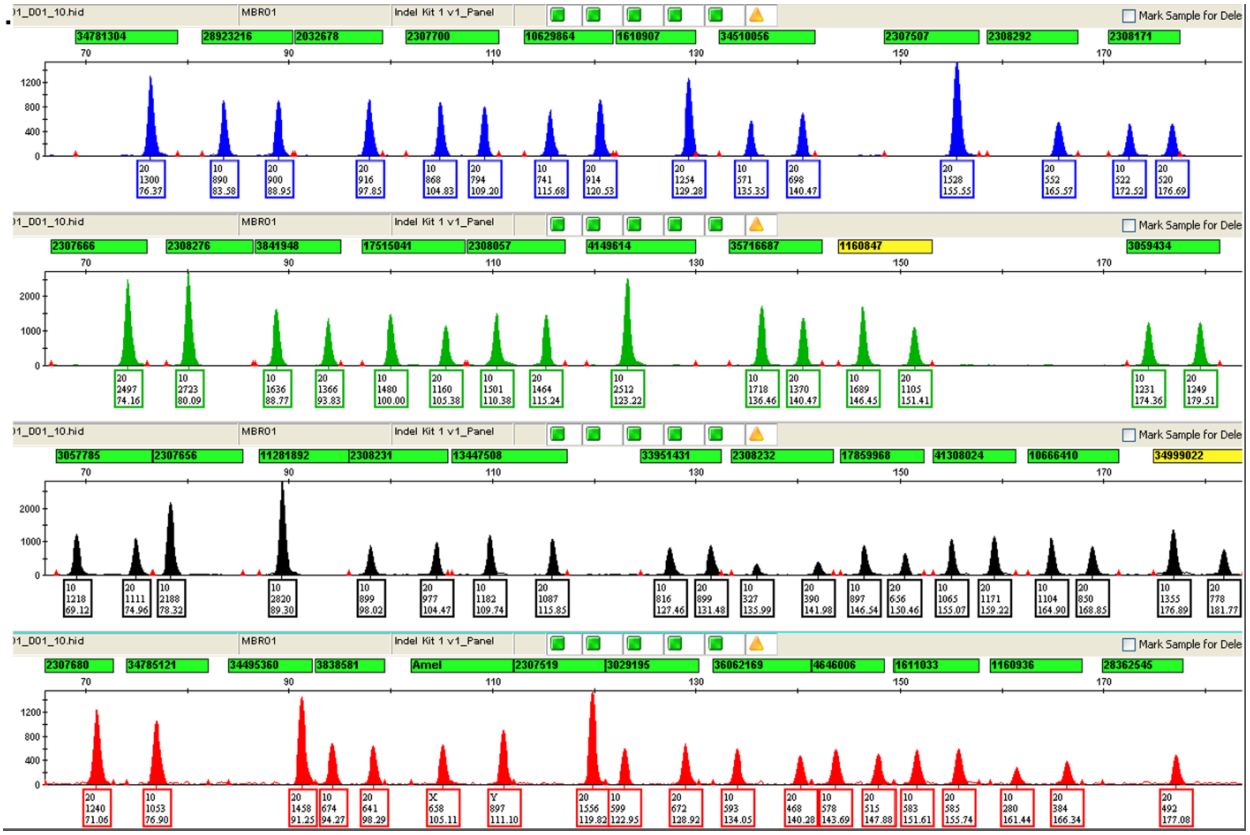


Figure 4. An electropherogram of DNA amplified using an INDEL system of 42 markers and analyzed by CE from Larue et al (15).

HID markers are useful in cases where a suspect has been identified and the suspect's reference profile can be compared to that of the profile from the evidentiary sample. They also are beneficial when unidentified remains have been found and an identification needs to be made. HID INDELs are a marker system that can individualize a person similar to that of the STRs that are currently in use. However, it would be an inefficient use of time and resources to use such markers individually typing all 49 INDELs. Of the resources, DNA is the most limited and precious. By running the DNA sample 49 separate times, there is a risk of consuming the evidence preventing any further testing or not having sufficient quantity of DNA to type all 49 markers. All 49 markers potentially can be combined into one assay, consuming less precious limited sample, simplifying the process and making the workload comparable to that of the STR kits currently being used in forensic laboratories.

A multiplex is defined as the simultaneous amplification of multiple DNA regions using primer pairs designed for each specific marker. Multiplexing is a cost efficient and labor saving way to perform genotype analysis of multiple regions of the genome without compromising the results. However, designing multiplex assays requires attention to multiple competing physical and molecular constraints that may occur among the various primers in the reaction (20-23). Primer design is a critical factor for a successful multiplex. There are six main criteria that are important when designing primers. These factors are: a) the primers should be specific to one region of the genome so as to not amplify other unintended areas; b) the forward and reverse primers should be between 18-30 nucleotides in length; c) the product size should not exceed 500 bp; d) the melting temperature of both primers should be between 58-65 ^{0}C and should not exceed 3 ^{0}C difference between the forward and reverse primers; e) the GC (guanine and cytosine) content should be between 40-60% of the primers; and f) the ΔG on the 3$'$ end should be greater than or equal to -9 kcal/ mol (21). These criteria together will help optimize and increase the specificity of the primers to their intended target. However, when making a multiplex PCR, other factors are necessary to ensure optimal design. One factor is that the primers should not form a dimer. Dimerization occurs when two primers bind to each other rather than to the DNA template. This dimer competes for PCR reagents and can potentially reduce the amplification yield of the intended targeted sequence. Another major factor is that the melting temperatures should be similar between all the primers involved (21-23).

The generally accepted method for generating a DNA profile for human identification purposes in forensics has been by typing STR markers. However, when the DNA in samples is degraded to fragments of less than 250 base pairs, in which some loci in STR amplification kits fail to yield amplification products (9, 10). By using an INDEL marker system, the amplification

issues with degraded samples can be reduced. Using the proposed 49 HID INDEL markers described by LaRue et al (15), a high power of discrimination can be achieved to successfully individualize people. Designing a multiplex assay that can amplify all intended targets in degraded biological samples would enhance successful characterization of forensic casework. This project tested the hypothesis that a multiplex of informative INDELs can be designed to identify individuals with high discriminatory power.

CHAPTER 2

RESEARCH DESIGN AND METHODOLOGY

Primer Design

For 49 INDEL markers described in the paper by LaRue *et al*. (15) forward and reverse primers were designed to amplify each marker. Publically available software tools through the websites dbSNP (24), the UCSC Genome Browser (25) and Primer-BLAST (26), were used to assist in designing the primers. Each of the chosen primer pairs were checked for potential dimerization with the other primer pairs using the publically available software tool, PriDimerCheck (27). After determining that no major issues should occur, unlabeled primers were ordered from Invitrogen™.

Unlabeled Primers

The unlabeled primers were run individually to ensure that each primer pair was performing as intended and would successfully amplify the DNA. For amplification, each sample contained the PCR mix made up of 5.5 µL of water, 2.5 µL of 10x buffer, 2.5 µL of BSA (10 mg/ mL), 2 µL of 50 mM $MgCl_2$, 1 µL of 10 mM dNTPs, 0.5 µL Taq polymerase (5 U/ µL), 0.5 µL of the forward primer at a concentration of 10 µM, 0.5 µL of the reverse primer at a concentration of 10 µM, and 10 µL of DNA (1 ng/ µL). The samples were amplified on the Applied Biosystems® GeneAmp® PCR System 9700 thermocycler under the parameters of 95˚C for 11 minutes, 36 cycles of 95˚C for 10 seconds, 61˚C for 30 seconds, 72˚C for 30 seconds, and a final extension of 70˚C for 10 minutes. Once amplified, each marker was assessed using the

Agilent© 2200 Tapestation using 2 µL of the TapeStation buffer with 2 µL of sample, following laboratory protocol (28). All individually run primer sets were evaluated to ensure each worked and produced a product around its estimated base pair range. After the primer pairs were run individually, they were arranged in groups of 5 primer pairs for an initial multiplex design. The amplification of the multiplex used the Qiagen® Multiplex PCR kit where each sample tube contained 5 µL of 10X primer mix, 10 µL of DNA (1 ng/ µL), 10 µL of water, and 25 µL of 2X multiplex PCR master mix, as instructed by the protocol (29). The 10X primer mix was made by 10 µL of the forward primer (10 µM) and 10 µL of the reverse primer (10 µM) added to 400 µL of water. Each sample was amplified on the thermocycler under the parameters of 95°C for 5 minutes, 35 cycles of 95°C for 30 seconds, 60°C for 90 seconds, and 72°C for 90 seconds, then 68°C for 10 minutes. The products of this initial multiplex were run on the TapeStation to assess if the primers could be amplified together and still produce intended results.

Fluorescently Labeled Primers

After the initial multiplex was evaluated to determine if it would perform as intended, fluorescently labeled primers were ordered from Applied Biosystems®. The fluorescent labeled primer pairs were run first individually to ensure they were functioning properly. For amplification of the individual primer pairs, each sample contained the PCR mix made up of 5.5 µL of water, 2.5 µL of 10x buffer, 2.5 µL of BSA (10 mg/ mL), 2 µL of 50 mM $MgCl_2$, 1 µL of 10 mM dNTPs, 0.5 µL Taq polymerase (5 U/ µL), 0.5 µL of the forward primer (10 µM), 0.5 µL of the reverse primer (10 µM), and 10 µL of DNA (0.5 ng/ µL). After amplification, the samples were analyzed using the Applied Biosystems® 3500 Genetic Analyzer. Each sample well for the CE contained 9.6 µL of HiDi formamide, 0.4 µL of Liz 600, and 1 µL of the amplified sample. The results from the CE were analyzed using GeneMapper ID-X (version 1.2). The next step was

to create an initial multiplex of amplicons separated into 5 groups based on their fluorophore color. Using the Qiagen® Multiplex PCR kit where each sample tube contained 5 µL of 10X primer mix, 10 µL of DNA (0.05 ng/ µL), 10 µL of water, and 25 µL of 2X multiplex PCR master mix. The 10X primer mix was prepared by adding 10 µL of the forward fluorescent primer (20 µM) and 2 µL of the reverse unlabeled primer (100 µM) and diluting to 100 µL. Each sample was amplified on the thermocycler under the parameters of 95°C for 5 minutes, 35 cycles of 95°C for 30 seconds, 60°C for 90 seconds, and 72°C for 90 seconds, then 68°C for 10 minutes. The amplified products were analyzed on the CE with each well containing 9.6 µL of HiDi formamide, 0.4 µL of Liz 600, and 1 µL of the amplified sample. After evaluating the profiles using GeneMapper ID-X, 1:10, 1:20, 1:50 and 1:100 dilutions of the amplified product were analyzed to reduce the amount of fluorescent dye overlap among amplicons. A new multiplex of the fluorophores was created with the amounts of each primer pair in the dye channel based on generating interlocus peak height balance. The new multiplex was amplified and run on the CE under the same parameters as described previously but with the 10X primer mix (20 µM) described depending on each primer pair and using a 1:100 dilution of the 0.5 ng/ µL DNA.

CHAPTER 3

Results and Discussion

Primer Design

The FASTA sequence for each marker was found by using DbSNP, a free online program (24). For some of the markers, the DbSNP website either did not provide a FASTA sequence or the sequence available was too small for forward and reverse primers to be designed. When this limitation was the case, the UCSC genome browser (25) was used to obtain the FASTA sequence. The FASTA sequence then could be entered into Primer Blast (26) to design the primer pairs for each marker *in silica*. Primers were designed to amplify the INDEL markers by using the Primer-BLAST software. The primers were designed to generate PCR product sizes between 60 and 200 bp. For one marker, RS2307579, the PCR product size was 219 bp; all others were in the intended range. The results of the primer design from Primer-BLAST are listed in Table 1, with the primer pairs numbered in numerical order, with primer pair 1 corresponding to RS 4646006. After primers were selected for each marker, they were tested for possible dimerization using the PriDimerCheck option from the MPprimer website (27). This software compares all the primers against each other and provides the alignment, the matches, the 3'-3' dimerization, and the ΔG.

Unlabeled Primers

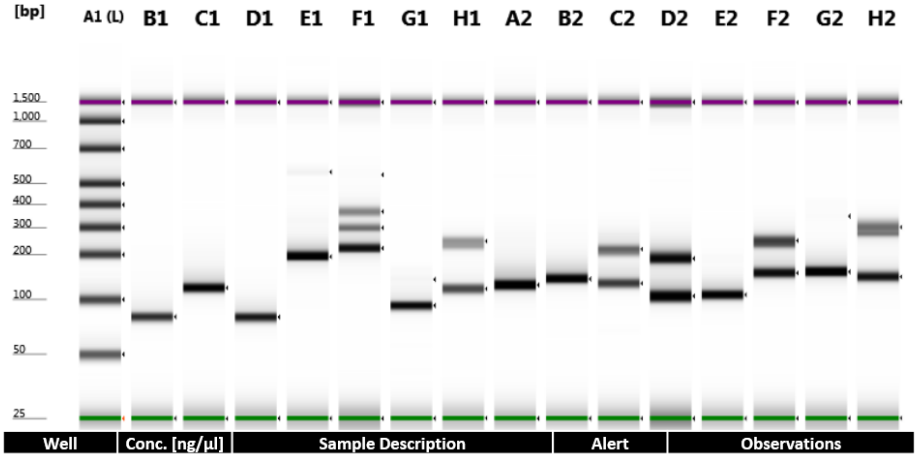
Since no major problems were observed when checking for dimerization, unlabeled primers were ordered. After amplification, the amplicons of the primers were run on the Agilent

TapeStation. The TapeStation is a simple to use, automated electrophoresis platform that enables analysis of DNA fragments between 35 and 1000 bp in length. The TapeStation utilizes a ScreenTape gel matrix, similar to that of an agarose gel, to separate samples by molecular weight.

Table 1. Using Primer-BLAST, primers were designed for all 49 INDELs which are represented by their RS number. The estimated product length for each primer pair is also given along with the melting temperatures for the forward and reverse primers. The primer pairs are numbered in numerical order, with primer pair 1 being RS4646006.

RS Number	Forwar Primer (5'-3')	Reverse Primer (5'-3')	Product Length	TM Forward	TM Reverse
4646006	GCTGGGAAATGGGGAGACAA	GCCCGCTGTTTGGAAAGAAA	83	59.96	59.61
13447508	ATGGTTCAGTGGAAATAGCATGA	CATGTGGTCCAATCCCCTCA	118	58.14	59.37
3047269	TCATTCCCATGCTGGGTGAG	TGCACTGTACTTGCATGCTG	83	59.74	59.12
2307507	TGAAGGTGGGGCTATTGAGAAA	TTTCTCTTAGTTTGCATAAAACCCT	181	59.35	57.17
2307579	TTGTGACTGTGTCTCAGCAGTTAT	CACTGACTTGACTGAACCTTTCAAC	219	60.2	60.45
3838581	AGCATATGGAGAATGATTACTGGTG	TGCTCAAGATTTGTATGAGGAAGT	96	58.76	58.19
2308276	CTGAGAGACAATGGGATTTGCC	TTGCATGGAATTTCTCCATTTGA	120	59.31	57.26
3042783	TTCCCTGAGCTTACCGGAGTT	GATATTGACCTGAAGGCACACTG	132	60.83	59.38
3841948	AAACTACATGGCCCACAAGT	ATCCCATGGCACATTCCAGT	138	57.32	59.37
35716687	GGATGCAGTAGAGGCAGGTT	CATGCCATCATTAGGGGACT	127	59.46	57.03
2307603	AGTGTGCCTACAGATACCACTT	ACAGTCTTCATAGAACTATCTCACA	110	58.83	57.05
60901515	TGTGGATACCAAGCACTCCTG	TGCTGGTTCCAACCGGAAG	113	59.72	60.23
2308292	ACTCTGTCTCCACTGGGAATGT	CTATCTGTTAGGCGCACTGTGTCA	152	60.76	62.69
2307526	TGTTGGAGCCACATCAATGAC	GAGAAAGATCAAATTAATGCCAGGA	159	58.84	57.66
2307656	TCTGTGGGCAGAAGGCAAG	ACCAGGTTTGAAAAATGACATGCTA	146	59.93	59.7
2308196	AGCCTGTAAAAATTCCCCTCTTGT	AACGAACATCTTTTTCCACCACA	151	60.45	59.3
2067140	ACCCACCAAATGTCCCTGAC	TAGCTCACCTTGCAGTGCTC	67	59.89	60.04
2067191	ATTTCAGGGTAATCGGATTCTGTA	TGGCCTGTTTATCTTCTAAAGGG	141	57.51	58.14
1610871	TCACCTTCCTCCCAGTAACCA	TCCATTTCCCCTGCTACTCC	62	60.13	58.79
2307710	GCCCATACCTACTGTGACCA	AGGCTTGTCTACAAAATGAATGAA	79	58.8	57.1
2307839	TGCATGTAGGACAAGAGGTAGTT	GGTCTTGCAAAATTAATCACACTC	149	59.16	57.07
34510056	TAGATCCCGGCCCAAAGTCA	CGGTGGAATGCAAAACGACT	114	60.62	59.41
16458	AGCTCCCCAAAGACATGGTT	TGTAAGACTCAGAAGTTATAGGGCA	144	59.22	59.04
34535242	CTACAGACAGGTTTAAAATGAGCAA	ATTTACATAAGCCTCCTTCTGTGG	133	57.8	58.56
10623496	TCAGAGCAGGCTTATCTTAAACA	CTTGCTAAGACAGAAAGAAGAAACA	97	57.58	57.75
33951431	ACAAAGCCTCGGCGATAGAC	ACTCACAGCATGTGGGAGAAC	79	60.18	60.27
16402	ATGCGCCTTTTTGGTTTTGGT	GCATCAGGACTGTATGGGGC	106	60.13	60.54
2308112	CAGAAGAGGCGGTGCTGATG	TCTGGAGGGACCCAAGGTAT	72	61.09	59.28
2307850	TCACCGTTTCCTCCGCACT	GCCCAACCTGCGTGGAAAG	60	61.5	61.92
140809	AGGCTTTCAGATGTTCTTAGCC	CTCCTGAGTGACCACAGCG	87	58.38	60.08
1160886	TTCCCATTGTGCTTAAAACTCCT	CCAGTCTACCCAAATGTATTCCA	74	58.52	57.89
34051577	GTCATCCAGATTATCGAGTGAGA	GCTGCACTTTAGTCTTCCTGA	137	57.3	57.95
10688868	TTCCATCCCTCCTCTTGCCT	CGCTCTGCACATGCGTAAAA	80	60.25	59.83
34811743	ACACTTCGTACCCAGGATGC	GCCTCTCCTTTTTGTTCAACCC	69	59.75	59.96
2307696	CACTGACAGCAATCAGAACAC	CTGAGCCCATCTGACTGCTC	116	57.47	60.18
34528025	GTCTTGGAGAGGAGTCAAATCAGA	CTGGAACTCAAAGCAAAACGAG	75	59.78	58.38
3045264	CTTACCTACGTGGTTGGTGAC	GTACACGAGTAGCCGATGGA	94	58.32	58.98
2308232	GATTGATGCAATCTCACACTACC	CTTCCTATTCTCCTTGCTTCGT	95	57.58	57.87
4187	TCAAATAAGAGTTGTCATATCCTGC	TGGCAGTGAAGAGAACAGGTC	119	57.21	59.93
3038530	GGCAATGAATTCCTCCATATCAAAA	TCTGCAGAAATCGCTTTGTAAAT	80	58.18	57.37
2308189	ACAAAGGAACGACAAGAACAAAA	TGGAACCTGATTCATGCTGCT	78	57.62	59.99
34795726	GGAGAAAACATGGATAGGTAGCAA	TCTCTTCACTAACAGGATGAAGTAT	114	58.56	57.04
17859968	GCTGTCTTAAAAGATTGTGGGG	GGGTCCTACTAAATGCCATGTG	68	57.56	58.79
28923216	GTGAATTGATCACTTTGTTTCTTGC	GCCATTAGCTCAGATTCTCAGGA	128	58.13	59.93
36062169	ACTATTCTACTGCCATTTACCACA	AGAGGATATCTCAGAAGGATGGACT	74	57.69	59.92
34511541	TGGAGGACTTTAGTAGAAGAGGA	ACCCTTCTTAGGTTCAAAGACACT	70	57.47	59.59
34495360	TGTGGTTTGGTCTCAGTACTTGTTC	TTTCTAAGCTGAGTGGCAAGATG	74	61.14	58.99
35605984	ATAGTTTTCCTGCATTATCCCCAT	GCACAAAGAAGCTTATGTCATAGTA	146	58.03	57.46
2307700	CTGGCAGGGCCAGAGC	TCCTTCCTCGGAATCCCCAT	76	59.71	60.03

This analysis was performed to assess whether all the primers worked as intended. Figure 5 shows the results of the first 15 primer pairs and an allelic ladder along with the resulting bp sizes. The allelic ladder is a defined set of fragments of known sizes that is used to help determine the size of the unknown samples. All 49 primer pair amplicons were able to be amplified and produced results similar to those seen in Figure 5.

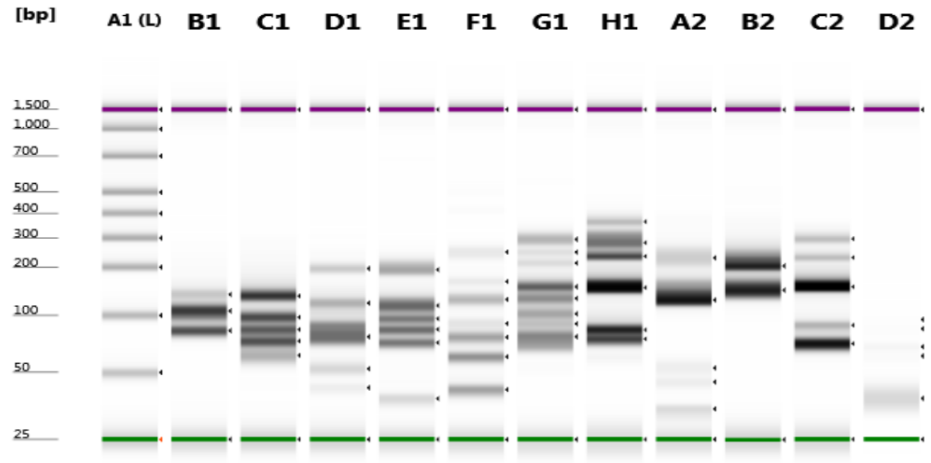


Well	Conc. [ng/µl]	Sample Description	Alert	Observations
A1	71.4	Ladder		Ladder
B1	11.4	1		Size bp: 84
C1	18.1	2		Size bp: 126
D1	13.6	3		Size bp: 84
E1	24.2	4		Size bp: 196
F1	19.9	5		Size bp: 224
G1	17.1	6		Size bp: 95
H1	16.3	7		Size bp: 123
A2	26.8	8		Size bp: 133
B2	22.1	9		Size bp: 146
C2	18.6	10		Size bp: 136
D2	31.1	11		Size bp: 108
E2	18.8	12		Size bp: 111
F2	28.9	13		Size bp: 159
G2	22.7	14		Size bp: 162
H2	20.5	15		Size bp: 150

Figure 5. The TapeStation gel image and summary table from the first 15 primer pairs and an allelic ladder. The table shows the base pair sizes for the respective primer pairs.

By running each of the primer pairs individually, the primers could be examined to ensure they were able to amplify the DNA and produce results. Since the TapeStation provided the approximate size for the amplicons, the primer pairs also were evaluated to determine if their bp size was approximately what was estimated during the design process. A review of the results

given from the TapeStation showed that the bp sizes of the amplicons were as designed. After the primer pairs were run individually, they were arranged in groups of 5 primer pairs for a general multiplex design. This multiplex then was run on the TapeStation to assess if the primers could be amplified together and still produce intended results. Figure 6 shows the results produced by the TapeStation of the multiplex. Although not every multiplex group produced 5 distinct bands, most banding that was observed was within the expected bp range for the primer pairs. Some of the primer pairs in the multiplex groups may have been too close in bp length to produce a distinct separation by the TapeStation.



Well	Conc. [ng/µl]	Sample Description	Alert	Observations
A1	34.4	Ladder		Ladder
B1	47.5	1		Size bp: 87, 109, 144
C1	76.6	2		Size bp: 65, 78, 88, 98, 140
D1	44.0	3		Size bp: 44, 53, 81, 126, 197
E1	62.8	4		Size bp: 76, 88, 97, 120, 195
F1	34.6	5		Size bp: 64, 81, 93, 133, 170
G1	65.9	6		Size bp: 81, 93, 103, 135, 159
H1	117	7		Size bp: 79, 87, 158, 237
A2	56.5	8		Size bp: 132
B2	68.5	9		Size bp: 153, 205
C2	98.0	10		Size bp: 75, 91, 159
D2	8.25	-		Water

Figure 6. The TapeStation results from the general multiplex groupings. Group 1 contained primer pairs 25, 27, 49, 40, and 1; group 2 = 3, 6, 8, 19, and 28; group 3 = 26, 33, 30, 35, and 47; group 4 = 11, 22, 20, 34, and 37; group 5 = 29, 7, 14, 12, and 31; group 6 = 38, 36, 39, 43, and 15; group 7 = 18, 41, 46, 16, and 5; group 8 = 42, 13, 2, 9, and 44; group 9 = 10, 24, 4, and 32; group 10 = 48, 17, 21, 23, and 45.

With multiple bands being observed within the expected bp ranges for the primer pairs in the multiplex groups, these results showed that the primer pairs were working, were able to be multiplexed together and still produce results. Since no major issues were found, fluorescently labeled forward primers were ordered for each of the primer pairs.

Fluorescently Labeled Primers

The 49 primer pairs were arranged into 5 different groups, each labeled with a different fluorophore: blue, green, yellow, red, and purple. The primer pairs were separated based on their resultant PCR product bp size. Amplicons labeled with the same fluorophore need to be different sizes to separate sufficiently and to have distinct products observed during analysis. Table 2 shows the arrangement of each of the amplicons by their fluorophore. The highlighted primer pair amplicons were alternates with those respective fluorophore labels because they were too close in base pair length to the other amplicons, leaving only 43 primer pairs being used. Once the primer pair amplicons were separated by their fluorophores, they were again run individually to make sure they still worked properly with the new forward primer. By running the amplicons individually, it also provided a reference of where the peaks should be seen when multiplexed. After being run on the CE and analyzed using GeneMapper ID-X, all 43 primer pairs were found to work. Figure 7 shows an example of one of the amplified primer pairs for each of the dye channels.

Table 2. Fluorophores by which each amplicon was labeled. The highlighted primer pair amplicons were ordered as alternates with those respective fluorophore labels because they were too close in base pair length to the other amplicons.

Dye Channel	Primer Pair	Base pair length	INDEL length
Blue= 6-FAM	29	60/64	4
	46	70/75	5
	40	76/80	4
	37	90/94	4
	27	102/106	4
	2	118/124	6
	10	123/127	4
	32	132/137	5
	48	146/151	5
	14	155/159	4
	4	176/181	5
	5	216/219	3
	20	116/120	4
	7	74/79	5
Green= VIC	19	58/62	4
	28	72/77	5
	33	78/80	2
	38	95/101	6
	11	110/115	5
	39	119/125	6
	9	133/138	5
	21	147/149	2
	31	71/74	3
Yellow= NED	17	63/67	4
	47	70/74	4
	1	79/83	4
	6	92/96	4
	12	109/113	4
	44	128/133	5
	18	137/141	4
	16	147/151	4
	45	74/80	6
Red= TAZ	43	64/68	4
	41	73/78	5
	3	79/83	4
	25	93/97	4
	22	114/119	5
	8	127/132	5
	23	140/144	4
	13	147/152	5
	36	71/75	4
Purple= SID	34	67/69	2
	26	75/79	4
	30	84/87	3
	42	110/114	4
	35	112/116	4
	24	129/133	4
	15	146/151	5
	49	72/76	4

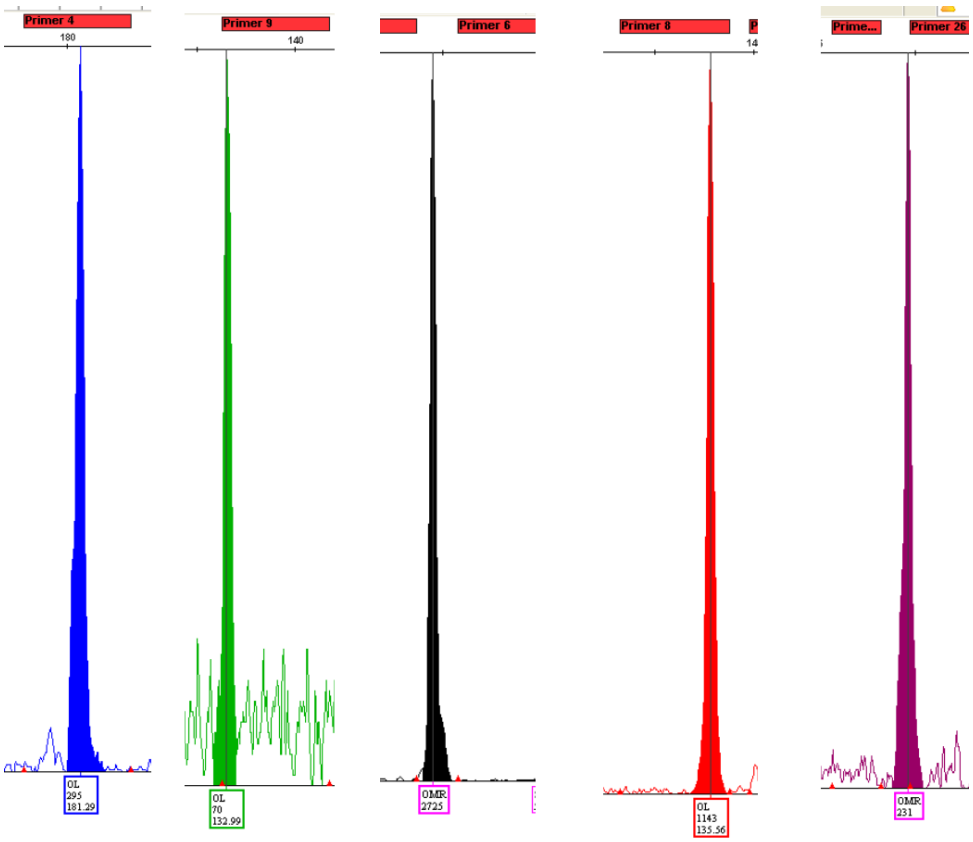


Figure 7. Electropherograms of amplicons 4 (RS2307507), 9 (RS3841948), 6 (RS3838581), 8 (RS3042783), and 34 (RS34811743) (left to right) are shown above. Each of the examples shows one peak indicating a homozygous insertion or deletion at those markers.

When creating a multiplex, it is expected that the size products seen with the individual amplicons would be consistent with the products observed in the multiplex. With the single amplicons known to be working, the amplicons were multiplexed into 5 groups based on the fluorophore label. The fluorophore multiplexes were designed to ensure no substantial spectral overlap from other amplicons. This assessment confirmed the fluorophore multiplexes could be amplified and multiplexed together and produce the desired results of product sizes remaining consistent. After running dilutions of the fluorophore multiplex to improve the primer overlap between amplicons, it was determined that the 1:100 dilution of the amplified product provided the best results. High signal produced by the amplicons can cause saturation and spectral overlap

(pull up) between the amplicons, diluting the sample helps to reduce this oversaturation. Figure 8 shows the electropherograms of the fluorophore multiplexes using the 1:100 dilution of PCR products (0.5 ng/ µL of input DNA).

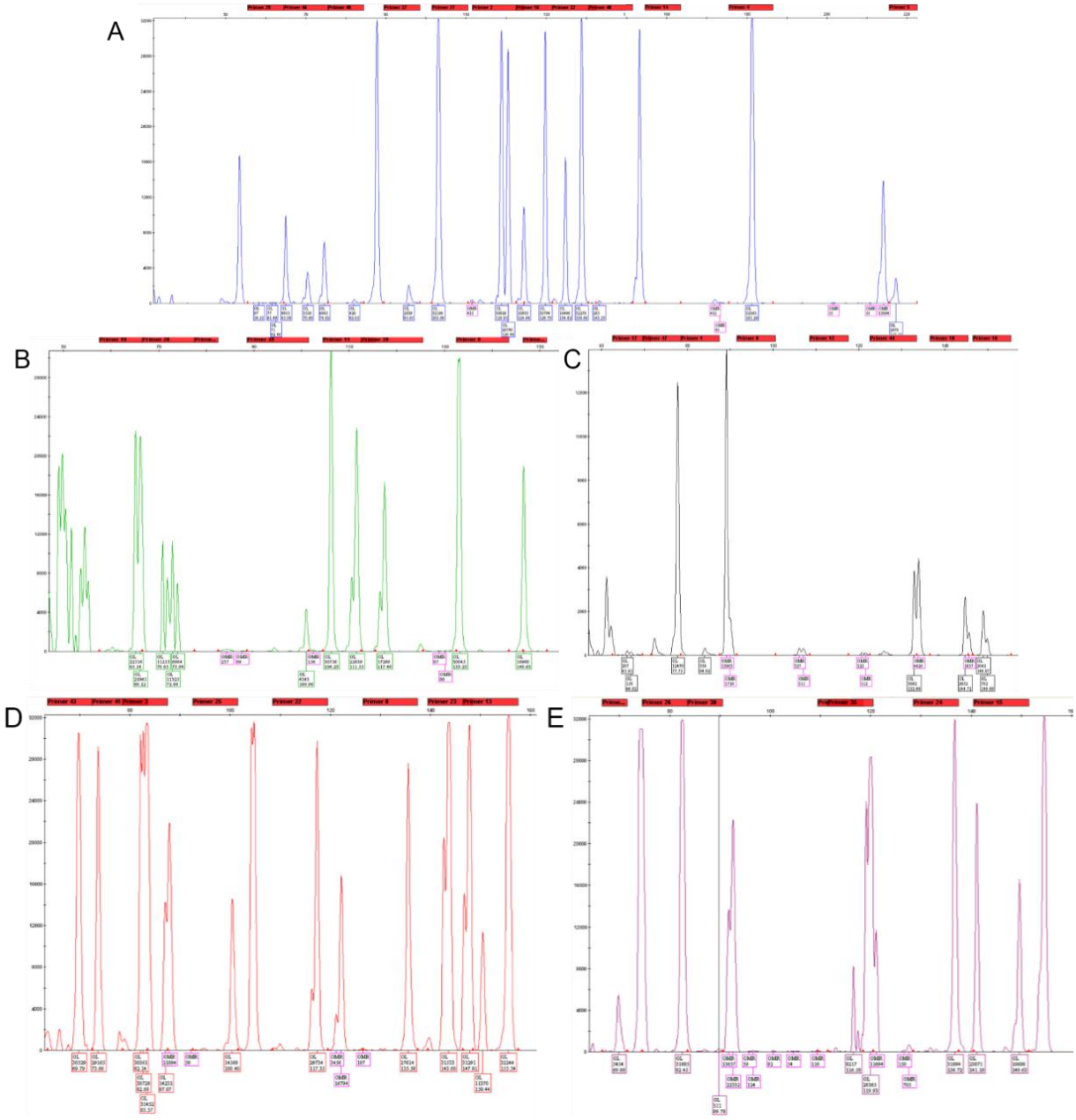


Figure 8. The electropherograms for each of the fluorophore multiplexes, with the x-axis representing the size in bp and the y-axis representing the reflective fluorescent units (RFUs). A) The blue fluorophore multiplex contains 12 amplicons. B) The green fluorophore multiplex contains 8 amplicons. C) The yellow fluorophore multiplex contains 8 amplicons. D) The red fluorophore multiplex contains 8 amplicons. E) The purple fluorophore multiplex contains 7 amplicons.

The fluorophore multiplexes show that most of the amplicons were able to amplify. The amplicon for primer pair 38 in the green fluorophore group did not generate a product in this multiplex (Figure 8). After adding 10 µL more of the forward (20 µM) and reverse (20 µM)

primers, a peak could be detected indicating that it is possible to amplify this locus in the multiplex. Figure 9 shows a 1:1000 dilution of the 0.5 ng/ µL of DNA, where more of the primer pair 38 was added.

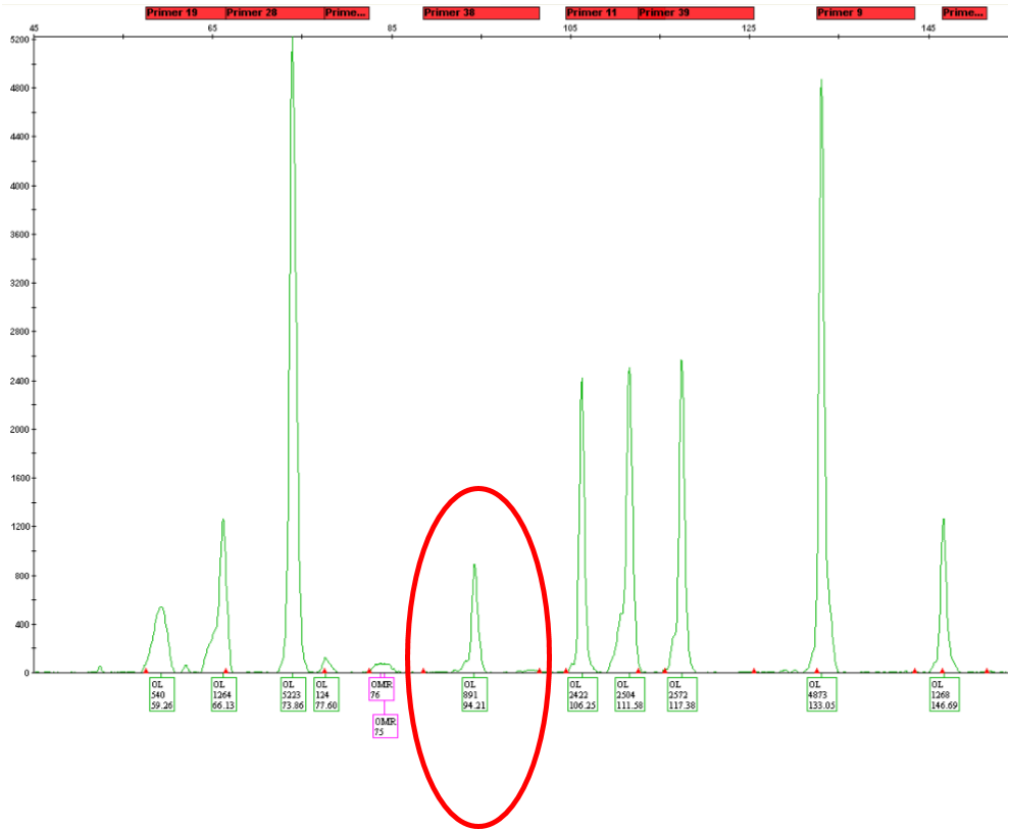


Figure 9. A 1:1000 dilution of 0.5 ng/ µL of DNA with 10 µL more of the forward (20 µM) and reverse (20 µM) primers of primer pair 38 added, with the x-axis representing the size in bp and the y-axis representing RFUs. A peak, circled in red, was seen, indicating this locus could be amplified in the multiplex. Amplicons included in this fluorophore multiplex are 19 (RS1610871), 28 (RS2308112), 33 (RS10688868), 38 (RS2308232), 11 (RS2307603), 39 (RS4187), 9 (RS3841948), 21 (RS2307839) (left to right).

In the yellow fluorophore group, two of the primer pair (12 and 44) amplicons did not produce peaks. An additional 10 µL of the forward (20 µM) and reverse (20 µM) primers for both pairs were added to the primer mix, but no improvement was seen. As a way to check if the amplicons were producing products of the same size in the fluorophore multiplex, the single amplicon

images were evaluated next to the multiplex results. Figure 10 shows a representative comparison of some of the amplicons for each fluorophore group.

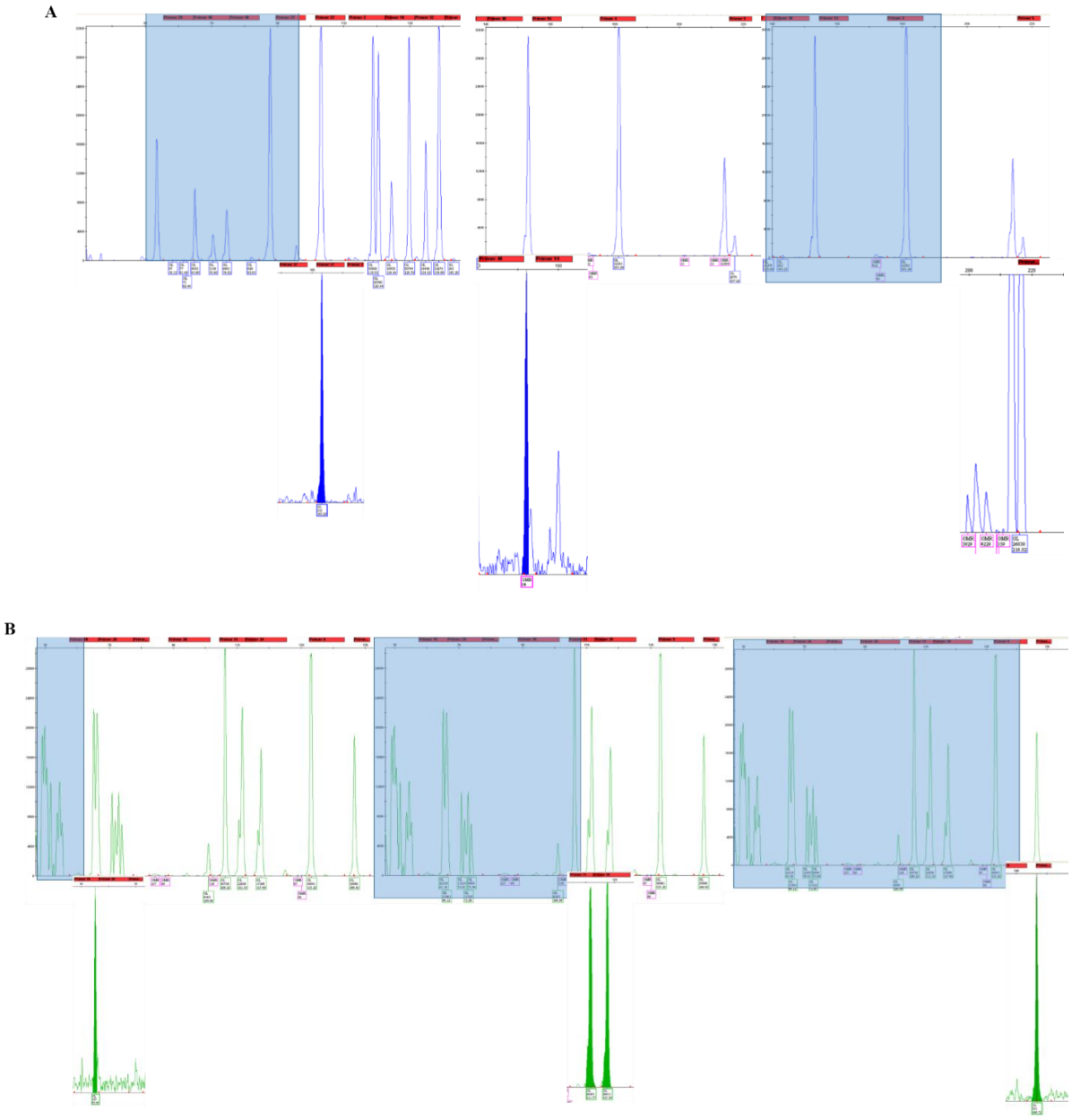


Figure 10. Fluor multiplex and single amplicon electropherograms comparing product size results, with the x-axis representing the size in bp and the y-axis representing RFUs. **A)** Blue fluorophore amplicon results. The INDELs represented are primer pairs 27 (RS16402), 14 (RS2307526), 4 (RS2307507). **B)** Green fluorophore amplicon results. The INDELs represented are primer pairs 28 (RS2308112), 39 (RS4187), 21 (RS2307839).

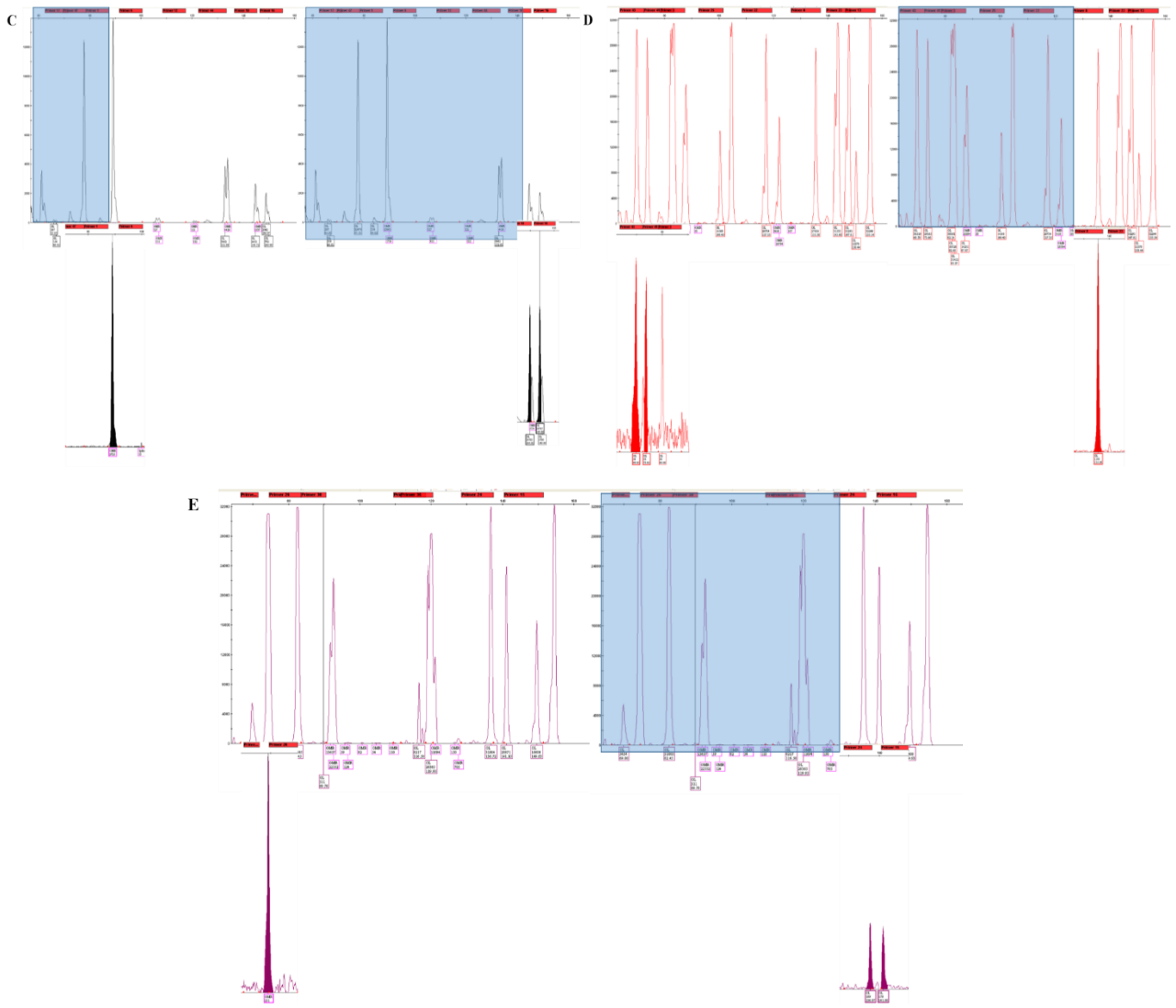


Figure 10. Fluor multiplex and single amplicon electropherograms comparing product size results, with the x-axis representing the size in bp and the y-axis representing RFUs. **C)** Yellow fluorophore amplicon results. The INDELs represented are primer pairs 6 (RS3838581) and 18 (RS2067191) **D)** Red fluorophore amplicon results. The INDELs represented are primer pairs 43 (RS17859968) and 8 (RS3042783). **E)** Purple fluorophore amplicon results. The INDELs represented are primer pairs 34 (RS34811743) and 24 (RS34535242).

The fluorophore multiplex products line up to what was observed with the single amplicons.

These results are the desired outcomes running the multiplex.

CHAPTER 4

Conclusion

Primers were designed for the 49 INDEL markers described by LaRue *et al.* (15). By using a combination of online tools, the optimal amplicon bp size could be selected for as well as ensuring the melting temperatures for the forward and reverse primers did not differ by more than 3 ^{0}C. Publically available software also made it possible to check the designed primers for any potential issues, such as dimerization, that may arise during the amplification process. With the 49 INDEL markers described by LaRue *et al*. (15) a power of discrimination of at least 10^{-19} can be achieved. The 49 designed primer pairs were able to successfully amplify 5 picograms of DNA. All 49 unlabeled amplicons were visualized using the TapeStation and shown to be functioning properly. However for the fluorophore labeled primer pairs, due to the amplicon bp sizes of 6 of them being too close to the other amplicons, only 43 of the INDELs were included in the multiplex to ensure clear product separation would be observed. As was seen in Figure 10, the individually run amplicon products were able to line up with the peaks present in the initial multiplex of amplicons separated into the 5 groups based on their fluorophore. By using a 1:100 dilution of PCR products generated with 0.5 ng/ µL of input DNA, the overlap between amplicons was reduced and the product peak heights were decreased.

Further optimization needs to be performed to achieve better interlocus peak height balance and to further reduce oversaturation. More testing should be done with primer pairs 12 and 44 to find out what may be causing them to not produce products when multiplexed. Once

the initial fluorescent amplicons have been optimized, a complete multiplex with all the fluorophores combined together should be made. This multiplex would further save time and save DNA compared to that of the 5 multiplex groups. This multiplex will need to be run using different samples of DNA to ensure the markers being amplified are able to distinguish individuals. Population studies need to be performed, in addition, to confirm the amplicons are able to distinguish people of the same and of different populations. Other testing designs that should be performed after the multiplex has been optimized are testing the sensitivity using lower amounts of input DNA, testing with degraded samples, testing with contaminated samples, and testing with nonprobative case samples.

In conclusion, primers were designed for all 49 INDEL markers described for human identification by LaRue *et al*. (15). Creating an INDEL multiplex system for human identification will benefit the forensic community by providing an alternate or adjunct system to STRs without the need to change the current laboratory infrastructure. Using the INDEL system could provide a better way to analyze degraded samples and obtain a more informative profile than that achieved by STR kits. The initial multiplex designed using the INDEL primers was able to successfully amplify and produce the predicted results.

CHAPTER 5

References

1. Edwards A, Civitello A, Hammond HA, Caskey CT. DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. Am J Hum Genet 1991 Oct;49(4):746-56.

2. Mullis KB, Faloona FA. Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. Meth Enzymol 1987;155(0):335-50.

3. Wang Y, Ju J, Carpenter BA, Atherton JM, Sensabaugh GF, Mathies RA. Rapid sizing of short tandem repeat alleles using capillary array electrophoresis and energy-transfer fluorescent primers. Anal Chem 1995;67(7):1197-203.

4. Buel E, Schwartz MB, LaFountain. Capillary electrophoresis STR analysis: comparison to gel-based systems. J Forensic Sci 1998;43(1):164-70.

5. Budowle B, Shea B, Niezgoda S, Chakraborty R. CODIS STR loci data from 41 sample populations. J Forensic Sci 2001;46(3):453-89.

6. Applied Biosystems. 2010. Applied Biosystems 3500/3500xL Genetic Analyzer User Guide.

7. National Research Council. The evaluation of forensic DNA evidence. Washington D.C.: National Academy Press; 1996. Report No.: 2.

8. Hammond HH, Jin L, Zhong Y, Caskey CT, Chakraborty R. Evaluation of 13 short tandem repeat loci for use in personal identification applications. Am J Hum Genet 1994;55(1):175-89.

9. Burger J, Hummel S, Herrmann B, Henke W. DNA preservation: a microsatellite-DNA study on ancient skeletal remains. Electrophoresis 1999;20(8):1722-8.

10. Golenberg EM, Bickel A, Weihs P. Effect of highly fragmented DNA on PCR. Nucleic Acids Res 1996;24(24):5026-33.

11. Applied Biosystems. August 2012. AmpFlSTR® Identifiler® PCR Amplification Kit.

12. Kidd KK, Pakstis AJ, Speed WC, Grigorenko EL, Kajuna SLB, Karoma NJ, et al. Developing a SNP panel for forensic identification of individuals. Forensic Sci Int 2006 Dec;164(1):20-32.

13. Pakstis AJ, Speed WC, Kidd JR, Kidd KK. Candidate SNPs for a universal individual identification panel. Hum Genet 2007;121:305-17.

14. Pereira R, Phillips C, Alves C, Amorim A, Carracedo Á, Gusmão L. A new multiplex for human identification using insertion/deletion polymorphisms. Electrophoresis. 2009;30(21):3682-90.

15. LaRue BL, Lagacé R, Chang C, Holt A, Hennessy L, Ge J, et al. Characterization of 114 insertion/deletion (INDEL) polymorphisms, and selection for a global INDEL panel for human identification. Leg Med 2014 Jan;16(1):26-32.

16. Fondevila M, Phillips C, Santos C, Pereira R, Gusmao L, Carracedo A, Butler JM, Lareu MV, Vallone PM. Forensic performance of two insertion-deletion marker assays. Int J Legal Med 2012;126:725-37.

17. Oka K, Asari M, Omura T, Yoshida M, Maseda C, Yajima D, et al. Genotyping of 38 insertion/deletion polymorphisms for human identification using universal fluorescent PCR. Mol Cell Probes 2014 Feb;28(1):13-8.

18. Wei Y, Qin C, Dong H, Jia J, Li C. A validation study of a multiplex INDEL assay for forensic use in four Chinese populations. Forensic Sci Int Genet 2014 Mar;9(0):e22-5.

19. Seong KM, Park JH, Hyun YS, Kang PW, Choi DH, Han MS, et al. Population genetics of insertion–deletion polymorphisms in South Koreans using investigator DIPplex kit. Forensic Sci Int Genet 2014 Jan;8(1):80-3.

20. Nicodeme P, Steyaert JM: Selecting optimal oligonucleotide primers for multiplex PCR. Proc Int Conf Intell Syst Mol Biol 1997, 5:210-213.

21. Chamberlain J S, Chamberlain J R. Optimization of multiplex PCRs. In: Mullis K B, Ferre F, Gibbs R A, editors. The polymerase chain reaction. Boston, Mass: Birkhauser; 1994. pp. 38–46.

22. Chamberlain, Jeffrey S., Richard A. Gibbs, Joel E. Rainer, Phi Nga Nguyen, and C. Thomas. "Deletion Screening of the Duchenne Muscular Dystrophy Locus via Multiplex DNA Amplification." *Nucl Acids Res Nucleic Acids Research* 16.23 (1988): 11141-1156. Web.

23. Rachlin J, Ding C, Cantor C, Kasif S: Computational tradeoffs in multiplex PCR assay design for SNP genotyping. BMC Genomics 2005, 6:102.

24. *"dbSNP." NCBI*. U.S. National Library of Medicine, n.d. Web. <http://www.ncbi.nlm.nih.gov/SNP/>.

25. UCSC Genome Browser: Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. The human genome browser at UCSC. *Genome Res*. 2002 Jun; 12(6):996-1006. https://genome.ucsc.edu/.

26. "Primer-BLAST." *NCBI*. U.S. National Library of Medicine, n.d. Web. 27 Nov. 2015. <http://www.ncbi.nlm.nih.gov/tools/primer-blast/>.

27. Shen Z, Qu W, Wang W, Lu Y, Wu Y, Li Z, Hang X, Wang X, Zhao D, Zhang C. MPprimer: a program for reliable multiplex PCR primer design. BMC Bioinformatics 2010;11:143. < http://biocompute.bmi.ac.cn/MPprimer/primer_dimer.html>.

28. Agilent Technologies. 2013. Agilent 2200 TapeStation User Guide.

29. Qiagen®. October 2010. Qiagen® Multiplex PCR Handbook.