

2022

# Analysis of mock-degraded and mixed DNA samples using the ForenSeq™ MainstAY kit

---

<https://hdl.handle.net/2144/45671>

*"Downloaded from OpenBU. Boston University's institutional repository."*

BOSTON UNIVERSITY  
SCHOOL OF MEDICINE

Thesis

**ANALYSIS OF MOCK-DEGRADED AND MIXED DNA SAMPLES USING THE  
FORENSEQ™ MAINSTAY KIT**

by

**TANYA RAJA TANNOUS**

B.A., University of California, Berkeley, 2020

Submitted in partial fulfillment of the  
requirements for the degree of  
Master of Science

2022

© 2022 by  
TANYA RAJA TANNOUS  
All rights reserved

Approved by

First Reader

---

Robin W. Cotton, Ph.D.  
Associate Professor of Anatomy & Neurobiology  
Program in Biomedical Forensic Sciences

Second Reader

---

Bobby LaRue, Ph.D.  
Senior CSM, Human Identification and Forensics  
Qiagen LLC

Third Reader

---

Fabio Oldoni, Ph.D.  
Assistant Professor of Forensic Science  
Arcadia University

## **ACKNOWLEDGMENTS**

First and foremost, I would like to sincerely thank Dr. Robin Cotton for her invaluable guidance and expertise in completing this work. It has been an honor to learn from both an incredible professor and mentor. This project would not have been possible without the members of my thesis committee, Dr. Bobby LaRue and Dr. Fabio Oldoni, who provided their time and knowledge that have allowed for the completion of this work. I am also grateful to the students and faculty of the Boston University Biomedical Forensic Sciences Program, as the experiences and opportunities provided to me have further cultivated my passion for forensic science. A special thank you to Verogen for providing training and support for proper use of all sequencing materials. Lastly, I would like to thank my family and friends for their unwavering support and motivation throughout my educational career.

**ANALYSIS OF MOCK-DEGRADED AND MIXED DNA SAMPLES USING THE  
FORENSEQ™ MAINSTAY KIT**

**TANYA RAJA TANNOUS**

**ABSTRACT**

For more than two decades, the polymerase chain reaction (PCR) amplification of short tandem repeat (STR) polymorphisms followed by detection with capillary electrophoresis (CE) has been the standard method in forensic deoxyribonucleic acid (DNA) analysis. While direct PCR to CE is a relatively inexpensive and highly efficient approach, it is limited in resolution, throughput, and multiplexing ability. The advent of massively parallel sequencing (MPS) offers the potential of overcoming limitations of CE-based approaches due to the increased overall discrimination power, multiplexing capacity of various forensically-relevant markers and number of samples, and the ability to reveal true sequence variation. This offers forensic practitioners the potential to solve the challenges faced by PCR-CE when analyzing forensically challenging samples, such as degraded and mixed DNA. The Verogen ForenSeq™ MainstAY Kit is a MPS library preparation method with a streamlined workflow that allows for the targeted sequencing of 53 standard STR loci on the Illumina MiSeq™ FGx Sequencing System.

This study evaluated the repeatability and performance of the Verogen ForenSeq™ MainstAY Kit in analyzing mock-degraded and mixed DNA samples. The MainstAY Kit produced reliable and repeatable results across replicates. When compared to the GlobalFiler™ Kit for mock-degraded samples, the MainstAY Kit outperformed the

standard method with more consistent intralocus balance and signal intensity, along with less dropout observed with increasing degradation. Complete and partial Combined DNA Index System (CODIS) STR profiles were generated using MPS across a wide range of degradation severity, with an 86% complete single-source profile obtained for the most degraded sample (degradation index: 2041.6). The MainstAY Kit successfully detected minor alleles for low level contributor mixtures as low as 6% minor contribution. This study proves the viability of the ForenSeq MainstAY Kit as an alternative solution to conventional fragment analysis, and may be useful to laboratories assessing MPS technologies.

## TABLE OF CONTENTS

ACKNOWLEDGMENTS .....	iv
ABSTRACT .....	v
TABLE OF CONTENTS .....	vii
LIST OF TABLES .....	ix
LIST OF FIGURES .....	x
LIST OF ABBREVIATIONS.....	xi
1. INTRODUCTION .....	1
1.1 Conventional Forensic DNA Analysis .....	1
1.1.1 Short Tandem Repeat Analysis.....	2
1.1.2 Mitochondrial DNA Analysis .....	4
1.2 Next-Generation Sequencing Technologies .....	6
1.2.1 Capabilities of Next-Generation Sequencing .....	8
1.2.2 Challenges of Next-Generation Sequencing .....	10
1.3 Analysis of Forensically Challenging Samples .....	11
1.4 MiSeq FGx™ System and ForenSeq™ MainstAY Kit .....	12
1.5 Objective .....	14
2. MATERIALS AND METHODS.....	16
2.1 Repeatability Study.....	16
2.2 Mock-Degradation Study.....	16
2.2.1 DNA Degradation and Quantification .....	16
2.2.2 PCR Amplification and Capillary Electrophoresis .....	17

2.3 Mixture Study .....	18
2.4 ForenSeq™ MainstAY Library Preparation and MiSeq FGx™ Sequencing.....	19
2.5 Data Analysis .....	24
3. RESULTS AND DISCUSSION.....	25
3.1 Controls and Quality Assessment.....	25
3.2 Repeatability Study.....	26
3.3 Mock-Degradation Study .....	27
3.3.1 Effects of Varying Degrees of Degradation .....	27
3.3.2 Performance of the ForenSeq™ MainstAY Kit versus the GlobalFiler™ Kit	31
3.4 Mixture Study .....	42
4. CONCLUSIONS .....	47
5. FUTURE CONSIDERATIONS .....	49
APPENDIX.....	50
LIST OF JOURNAL ABBREVIATIONS .....	51
BIBLIOGRAPHY.....	52
CURRICULUM VITAE.....	59

## LIST OF TABLES

Table 1. Proportions of contributors and DNA input for sequencing of mixtures .....	19
Table 2. Quality metrics for the MiSeq FGx™ sequencing run .....	25
Table 3. Discordant alleles observed in the repeatability experiment.....	26
Table 4. Average degradation index and concentration across all time points .....	29
Table 5. Average percent of imbalanced loci for mock-degraded DNA .....	33
Table 6. Dropouts of loci and alleles for mock-degraded DNA .....	37
Table 7. Effect of increasing DNA incubation time at 95°C on allele dropout .....	38
Table 8. Unique alleles detected for minor contributors of mixture study .....	43

## LIST OF FIGURES

Figure 1. A simplified overview of the differences between the CE and MPS analyses of STRs .....	9
Figure 2. Outline of mock-degradation study workflow .....	16
Figure 3. Overview of the Verogen ForenSeq™ MainstAY protocol .....	21
Figure 4. Variation in total read count for repeatability study .....	27
Figure 5. Effect of increasing DNA incubation time at 95°C on degradation index .....	28
Figure 6. Effect of increasing DNA incubation time at 95°C on total read count	30
Figure 7. Effect of increasing DNA incubation time at 95°C on average total read count .....	31
Figure 8. Effect of increasing DNA incubation time at 95°C on intralocus balance .....	32
Figure 9. Effect of increasing DNA incubation time at 95°C on average allele signal intensity .....	34
Figure 10. Average allele signal intensity per locus for mock-degraded DNA ...	35
Figure 11. Effect of amplicon size on allele dropouts for mock-degraded samples .....	40
Figure 12. Effect of increasing degradation on percent of CODIS loci typed .....	41
Figure 13. Percent of unique minor alleles detected across a range of minor contributions .....	44
Figure 14. ForenSeq™ UAS Output for locus D3S1358 of a three-person mixture .....	46

## LIST OF ABBREVIATIONS

ADI	Allele Drop-In
ADO	Allele Dropout
ARC	Allele Read Count
AT	Analytical Threshold
Au-STR	Autosomal Short Tandem Repeat
bp	Base Pair
CE	Capillary Electrophoresis
CODIS	Combined DNA Index System
ddNTP	Dideoxynucleotide Triphosphate
DNA	Deoxyribonucleic Acid
DNL	Denatured Normalized Libraries
dNTP	Deoxynucleoside Triphosphate
EPG	Electropherogram
FBI	Federal Bureau of Investigation
FSP	ForenSeq Sample Plate
HSC	Human Sequencing Control
HV1	Hypervariable Region 1
HV2	Hypervariable Region 2
IT	Interpretation Threshold
kV	Kilovolt

MPS	Massively Parallel Sequencing
mtDNA	Mitochondrial DNA
ng	Nanograms
NGS	Next-Generation Sequencing
NLP	Normalization Library Plate
NWP	Normalization Working Plate
PBP	Purification Bead Plate
PCR	Polymerase Chain Reaction
Pg	Picogram
PH	Peak Height
PLP	Purification Library Plate
PNL	Pooled Normalized Libraries
RE	Restriction Endonuclease
RFLP	Restriction Fragment Length Polymorphism
RFU	Relative Fluorescent Unit
SBS	Sequencing By Synthesis
UAS	Universal Analysis Software
VNTR	Variable Number of Tandem Repeats
Y-STR	Y Chromosome Short Tandem Repeat
μL	Microliter

## **1. INTRODUCTION**

### **1.1 Conventional Forensic DNA Analysis**

Since its introduction into the criminal justice system in 1987, the analysis of deoxyribonucleic acid (DNA) from biological evidence has revolutionized forensic investigations. The analysis of biological evidence including blood, saliva, and semen has been an invaluable tool in the identification of victims of crime and mass disasters, conviction of perpetrators of crime, and exoneration of the wrongly accused (1). The advancement of DNA technologies over the past 30 years has allowed for the study of human variability with increased speed, sensitivity, reproducibility, and discrimination power (2,3).

The emergence of DNA typing, or “fingerprinting”, in 1985 by Sir Alec Jeffreys began with his discovery of repetitive blocks of DNA called minisatellites, or variable number of tandem repeats (VNTR). Restriction endonucleases (RE) were utilized to fragment DNA and produce restriction fragment length polymorphism (RFLP) patterns (4). Following DNA extraction, this technique allowed for the separation of fragment lengths by electrophoresis, transfer using Southern blotting, and hybridization with radioactive or chemiluminescent probes for visualization using autoradiography (5). This method became the standard in forensic labs until the late 1990s, providing a high power of discrimination and used in paternity, immigration, and criminal cases (6–8). However, the

time-consuming nature and high DNA input requirement of this method posed limitations for forensic applications (9).

Forensic DNA analysis was transformed by the use of the polymerase chain reaction (PCR), discovered by Kary Mullis in 1985 (10). PCR enables the exponential amplification of minute traces of DNA in a manner similar to its duplication in the cell (11). This three-step process begins with the separation of double-stranded DNA by heating. These single-stranded segments are then hybridized to oligonucleotide primers, or short DNA fragments that are complementary to the sequences targeted for amplification. The final step utilizes deoxynucleotide triphosphates (dNTPs) and a thermostable DNA polymerase to replicate the target sequence for double-stranded products. Upon repetition of this cycle, millions of copies of specific DNA fragments are generated, allowing for the analysis of small quantities of DNA often found in forensic cases. The advent of PCR decreased the DNA input amount required by previous methods from 500 nanograms (ng) to less than 1ng for successful analysis (12).

Ultimately, the discovery of the individualization potential of hypervariable minisatellite polymorphisms by Jeffries, concurrent with the discovery of PCR, produced the forensic DNA analysis methods presently used (4,10).

#### 1.1.1 Short Tandem Repeat Analysis

PCR-based short tandem repeat (STR) analysis has been considered the standard method in forensic DNA analysis (13,14). STRs are highly polymorphic, stably inherited segments of DNA that consist of 2-7 nucleotides tandemly repeated at a locus, often in

non-coding regions of the genome (15). The number and sequence of repeats at a specific locus is variable among individuals (15). Multiple loci can be multiplexed with PCR, offering the substantial power to distinguish between individuals (16). The STRs used in forensic DNA typing are mostly tetranucleotide repeats, selected due to their high variation and compatibility with PCR amplification due to their lengths (17). As of 2017, 20 core STR loci are required in DNA profile analysis for the Federal Bureau of Investigation (FBI) Combined DNA Index System (CODIS) (18). These loci were selected due to their high discriminating power, allele lengths between 90-500 base pairs (bp), unlinked chromosomal locations, robustness, reproducibility, and low stutter characteristics (19).

The PCR-based STR typing generally follows a four-step process. First, DNA is extracted from the source material and purified (20). The amount of DNA present is then quantified to ensure sufficient amounts are added to the subsequent PCR reaction (20). Following PCR amplification of the targeted STR loci, amplicons are separated by size using capillary electrophoresis (CE) (20). Fluorochromes on each primer set permit laser-induced fluorescence with multiwavelength detection to output relative fluorescent units (RFUs) that are then visualized as peaks on electropherograms (EPGs) (21). An allelic ladder containing characterized alleles is run alongside samples, and an internal size standard containing DNA fragments of known size is added to each allelic ladder and sample (21). Genotyping software then uses this information to identify alleles present in

each sample (21). Guidelines for STR profile interpretation are continuously revised and published by the Scientific Working Group on DNA Analysis Methods (SWGDM) (22).

Analysis of STRs are the mainstay of forensic DNA profiling due to the standardized loci analyzed, development of commercial kits that aid in standardization across laboratories, compatible databases containing allelic frequencies across human sub-populations, guidelines for statistical reporting, and general acceptance in courts of law (12,14,15). Furthermore, the ability to multiplex allows for the analysis of both autosomal STRs (Au-STRs) for individualization, and Y chromosome STRs (Y-STRs) that can infer biological sex and paternal lineage of male samples (23). The analysis of Y-STRs is commonly employed to aid in the separation of male DNA from female-dominant mixtures often related to sexual assault cases.

#### 1.1.2 Mitochondrial DNA Analysis

Although STRs are the most widely used genetic marker for forensic identification, mitochondrial DNA (mtDNA) has been an additional tool valuable to forensic DNA typing. Unlike nuclear DNA, mtDNA offers the forensic value of high copy number, lack of recombination, and matrilineal inheritance (24). Thus, the use of mtDNA can aid in generating biological profiles from samples that fail with standard STR analysis due to insufficient template quantity and quality (25,26).

Compared to the nuclear genome, the mtDNA genome has a higher mutation rate due to the low fidelity of mtDNA polymerase and a lack of repair mechanisms (26). Regions of the genome that evolve at higher rates are of interest for human identification

due to their hypervariability. The two regions used most in forensic analysis that possess the highest interpersonal sequence variation are the hypervariable region 1 (HV1) and hypervariable region 2 (HV2) found in the control region (27). While human cells contain two copies of DNA in the nucleus, hundreds of copies of mtDNA genomes are often present, offering advantages when analyzing compromised or degraded samples (28). MtDNA is maternally inherited without recombination, meaning the mtDNA lineage is shared between siblings and maternal relatives. Therefore, maternal relatives generations apart can serve as reference samples for comparison in forensic cases such as missing person and mass disaster identifications (25,29). MtDNA analysis can be performed on a wide variety of biological samples, including bones, teeth, and hair shafts (30). The robust circular nature of mtDNA allows for a higher likelihood of preservation and greater resistance to degradation compared to nuclear DNA (31). These characteristics make mtDNA testing a valuable, sensitive, and reliable method of forensic analysis.

The traditional method of mtDNA analysis often used in war casualty identification and challenging forensic samples such as bones and teeth, employs Sanger sequencing following PCR amplification of the HV1 and HV2 regions (24). While some forensic laboratories are implementing the amplification of the entire control region, mtDNA analysis is limited by a discrimination power lower than STR analysis due to the lack of recombination (24,32). However, the implementation of new technologies such as

massively parallel sequencing (MPS) offer the potential for higher discrimination power, speed, and resolution for the analysis of mtDNA (33).

## **1.2 Next-Generation Sequencing Technologies**

Determining the order of nucleic acids in biological samples has allowed for significant advances in molecular biology and specifically, forensic genetics. Since the introduction of Sanger sequencing in 1977, DNA sequencing technology has advanced to overcome the high cost, low-throughput, and low speed challenges that originally limited its use (34,35). The recent introduction of next-generation sequencing (NGS), or MPS, has aided in overcoming these limitations with its capacity to sequence billions of DNA molecules in parallel (36). Since the completion of the Human Genome Project in 2003, NGS has been integrated into routine clinical diagnostics and other biological fields, its application to forensic science offers both advantages and challenges when compared to the traditional forensic DNA analysis methods (37).

The sequencing of DNA has seen three generations of methodologies that have allowed for evolution from sequencing short oligonucleotides to the sequencing of whole genomes (38). First generation sequencing, known as Sanger sequencing, was based on the di-deoxynucleotide triphosphate (ddNTP) chain terminating method (34). In a manner similar to PCR, a DNA template, DNA primer, DNA polymerase, and fluorescently labeled dNTPs are used to synthesize strands of DNA (34). These reactions include the four labeled ddNTPs which, when added to the reaction and incorporated into the strand, cause termination of the elongation process (34). The advancement of the visualization of these

synthesized DNA fragments from slab gel electrophoresis to CE platforms decreased the cost of sequencing, and increased sensitivity and throughput (37). However, the time-consuming and cost inefficient nature of Sanger sequencing has rendered it an impractical method for routine forensic DNA analysis (39).

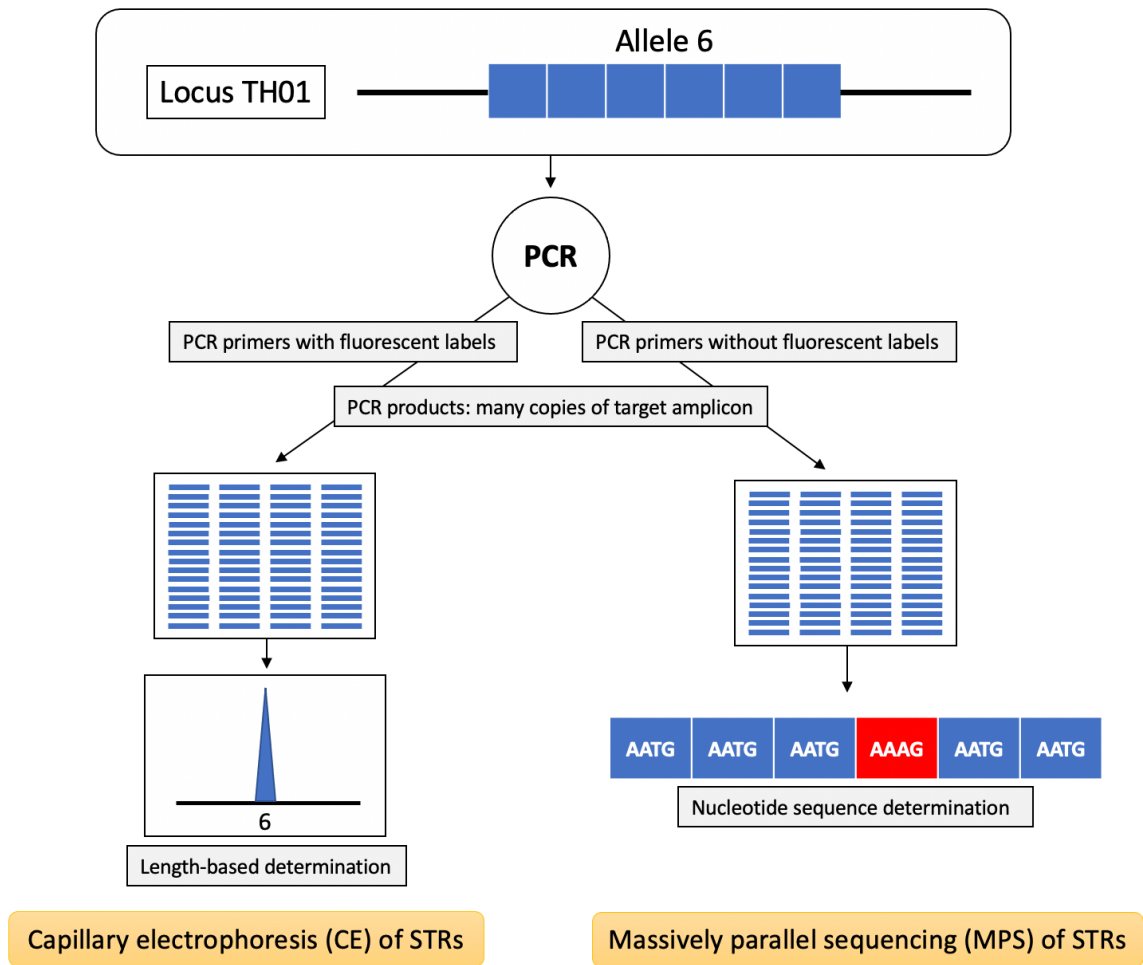
The development of PCR and instrument innovation led to the development of low-cost, high-throughput second-generation sequencing systems, also known as NGS. The first successful commercial NGS machine was released in 2005 by 454 Life Sciences (Roche), utilizing MPS to perform large numbers of parallel sequencing reactions to significantly increase the amount of DNA sequenced on one run (38). Currently, three bench-top sequencers are most often employed in forensic analyses: Illumina's MiSeq FGx™, and Thermo Fisher's Ion S5™ (40). While Illumina's platform utilizes a sequence-by-synthesis (SBS) technique that includes the reversible termination of fluorescently labeled dNTPs, the Thermo Fisher Ion Torrent platforms use semi-conductor technology to measure the pH change associated with the hydrogen ions released during DNA synthesis (41,42).

While second-generation sequencing techniques are amplification-based, recent efforts in third-generation sequencing methods allow for faster sequencing with longer reads generated (43). These platforms are capable of reading single molecules in real time, based on SBS technology or the use of nanopores (43). As the advancement of these

sequencing technologies continues to develop, advantages and disadvantages to the implementation of NGS in forensic science will evolve.

### 1.2.1 Capabilities of Next-Generation Sequencing

The use of NGS for forensic applications offers a powerful tool with the potential to considerably expand the field's capabilities by overcoming limitations of CE-based approaches. As opposed to fragment length-based analysis by the gold standard PCR-CE methods, sequencing of STR markers reveals the full level of variation of these loci (37). Analysis by NGS determines the full sequence of PCR products, including the STR repeat and flanking regions of alleles. This allows for the discrimination of otherwise indistinguishable alleles determined by length-based typing, as alleles known as isoalleles or isometric alleles may be identical in length but differ in sequence, as demonstrated in Figure 1. Previous studies have shown that sequence-based STR analysis demonstrates an increased allele variability of forensically significant STRs, and revealed the potential to prove true heterozygosity of PCR-CE homozygous genotype calls (44–46). Not only does capturing sequence variation increase the power of discrimination by resolving isoalleles, but can also be used to more accurately predict stutter behaviors (40).



**Figure 1. A simplified overview of the differences between the CE and MPS analyses of STRs.** In this example, an individual has a 6 allele at the TH01 locus known to have an AATG motif. While CE analysis would simply make a length-based determination of six repeats, the increased power of discrimination of MPS to provide nucleotide sequences can reveal sequence-based variants at loci, or isoalleles. This isoallele at allele 6 is denoted by the AAAG present within the repeat indicated in red.

Although PCR-CE can be performed in one day and NGS in two to three days, an advantage of NGS is the ability to combine all PCR-CE assays into one if all relevant loci can be targeted (37). The ability to simultaneously analyze large numbers of genetic markers for a number of samples in a single reaction increases the power of discrimination,

and may increase sample processing speed, decrease the cost per sample, and limit the amount of sample consumed for analysis (37,47). In addition, and in contrast to PCR-CE, combinations of genetic markers can be analyzed in one experiment, including Au-STRs, Y-STRs, along with identity-, phenotype-, and ancestry-informative single nucleotide polymorphisms (SNPs) (48,49). Therefore, with a single amplification, data can be obtained from CODIS loci, Y loci, and many other relevant markers using a single aliquot of sample.

### 1.2.2 Challenges of Next-Generation Sequencing

Various obstacles face NGS becoming a routine forensic analytical tool as PCR-CE has been. Despite the increased capabilities of MPS, adoption of new technologies by the forensic DNA community is slow due to the need for method validation and ability to yield results comparable to current methods (15). A 2017 survey of forensic laboratories identified the high cost of sequencing machines and kits as a factor preventing the implementation of MPS in forensics (50). Thus, the ability to decrease costs and the need for training and validation will dictate how quickly NGS may eventually replace PCR-CE (37). The lack of uniformity in nomenclature, workflow, and data analysis must be addressed to ensure the ability to compare STR typing results between laboratories (51). For example, recommendations must be made regarding the minimum number of reads that are required to reliably determine the presence of an STR allele using sequencing technologies (52). Furthermore, as forensic labs begin to implement the use of NGS, the admissibility of results generated will be challenged in courts of law, as with other new

methods (48). Overcoming these challenges will allow for the positive impact of MPS on forensic DNA analysis.

### **1.3 Analysis of Forensically Challenging Samples**

Degraded and mixed DNA samples are often encountered in forensic cases, causing difficulties in analyses and interpretation. Biological samples typically found at crime scenes may be inadequate in both quantity and quality, as they may contain limited or degraded DNA due to exposure to heat, light, humidity, and microorganisms (53). Although conventional PCR-CE analysis offers a greater chance of finding intact primer binding sites on small STR amplification product sizes, STR loci with larger amplicon sizes tend to dropout of degraded DNA profiles (15,53). Various stochastic effects which occur during PCR amplification can complicate the PCR-CE analysis of degraded DNA. Allele drop-in (ADI) causes additional alleles to be observed in the DNA profile, and allele dropout (ADO) causes the absence of true alleles in the profile if they fail to amplify (15). Increased stutter may occur in low template samples, causing the presence of alleles with repeat units less or greater in size than the true allele due to strand slippage (15). Additionally, heterozygote peak imbalance can be observed in heterozygous alleles that should be very similar in peak height (PH) (15). An advantage of NGS analysis over PCR-CE of degraded DNA is the lack of separation by size, providing the ability to amplify all STRs using the smallest viable amplicon length (40). Like current CE methods, the use of MPS for forensic analysis has shown that full profiles can be obtained from limited samples containing as low as 25 to 100 picograms (pg) of DNA (54).

The interpretation of samples with multiple known and unknown contributors can be challenging in forensic DNA analysis. Two or more individuals contributing to a sample may be expected in finger clippings, sexual assault cases, or touch DNA samples, for example, or during accidental sample contamination. In conventional PCR-CR analysis STR PHs in an EPG are directly related to the amount of DNA present, allowing for the estimation of relative contributor proportions in mixtures (55). However, preferential amplification and stutter can complicate this estimation, with the detection of minor contributors being difficult to separate from stutter and noise (56). Similarly, mixture proportions can be accurately estimated using read numbers in NGS analysis, with several minor contributor sequences having been successfully detected in 1:100 mixtures (37,51). The ability of NGS to discriminate sequence-specific alleles aids allows for the identification of allele sharing and stutter to permit the deconvolution of mixed DNA samples (40). Further, the NGS analysis of polymorphisms found in STR flanking regions has been shown to be informative in conducting mixture analysis (57).

#### **1.4 MiSeq FGx™ System and ForenSeq™ MainstAY Kit**

The MiSeq FGx™ Forensic Genomics System (Illumina Inc., San Diego, CA) is a benchtop NGS platform validated for forensic genomics. Using SBS technology, fluorescently labeled reversible-terminator nucleotides are utilized on clonally amplified DNA templates immobilized on the surface of a glass flow cell (58). Depending on sample complexity, a single sequencing run can process up to 96 samples at once, consisting of 398 total sequencing cycles comprised of the following: Read 1, Index 1, Index 2, and Read

2. The first read (Read 1) is 351 SBS cycles where one nucleotide base per cycle is sequenced in each target DNA amplicon. The first index read (Index 1) and second index read (Index 2) are 8 cycles each, determining the i7 and i5 indices, respectively. The second read (Read 2) is 31 SBS cycles that determine the last 31 nucleotides of the amplicon in the direction reverse of Read 1. The ForenSeq™ Universal Analysis Software (UAS, Verogen, San Diego, CA) is a platform interfaced with the MiSeq FGx™ System to perform run setup, sample management, allele and genotype calling, and report generation for sequencing data analysis (59).

The UAS assesses the quality of sequencing runs using positive and negative amplification controls, and four quality metrics with target ranges. Cluster density (K/mm<sup>2</sup>), clusters passing filter (%), phasing rate (%), and pre-phasing rate (%) values within target range ensure proper performance, although metrics that occasionally fall out of range can still provide sufficient data quantity and quality (60). Similar to CE-based software such as GeneMapper™ ID-X, analysis setting thresholds such as analytical thresholds (ATs), interpretation thresholds (ITs), and stutter filters are applied by the UAS during genotyping and can be altered for system optimization. While CE measures RFU, the UAS measures intensity via number of reads, or read count. The sequencing coverage level (read count) helps determine whether an allele or variant call is made with confidence.

In conjunction with the MiSeq™ FGx System, the ForenSeq™ MainstAY Kit (Verogen, San Diego, CA) targets 27 Au-STRs, 25 Y-STRs, and Amelogenin in a single reaction, enabling end-to-end sequencing in under 30 hours. The combination of 53 forensically-informative loci (Table A) focus on sequencing the most essential STRs and

omitting the burden of analyzing supplementary markers, such as SNPs. The ForenSeq™ MainstAY Kit creates dual-indexed libraries, as samples are combined with a primer mix containing a pair of tagged oligos for each target region. PCR is then used to attach these unique tags to each target copy, forming DNA templates with universal primer sequences flanking each region of interest. These tags then attach to index adapters, generating libraries to be amplified, purified, and pooled for sequencing. Using Unique Dual Indices (UDIs), a high density of short amplicons (<200 bp) is included to increase the likelihood of detecting low quality and quantity samples. The utilization of a streamlined data analysis software and assay workflow designed to mirror current DNA analysis workflows with equipment common to forensic labs, this kit is designed to address barriers to NGS being adopted as a routine casework tool.

### **1.5 Objective**

The analysis of forensically challenging samples such as samples containing degraded and mixed DNA is often limited with conventional PCR-based DNA typing. The advent of MPS technologies offers capabilities that can improve the analysis of DNA limited in quantity or containing multiple contributors. This study will evaluate the repeatability of the ForenSeq™ MainstAY Kit on the MiSeq™ FGx Sequencing System. The ForenSeq™ MainstAY Kit will be utilized to analyze mock-degraded samples alongside the GlobalFiler™ PCR Amplification Kit to compare the performance of MPS analysis with the conventional PCR-CE method. The capabilities of the MainstAY Kit to

analyze complex mixed samples will be assessed by the sequencing of two- and three-person DNA mixtures.

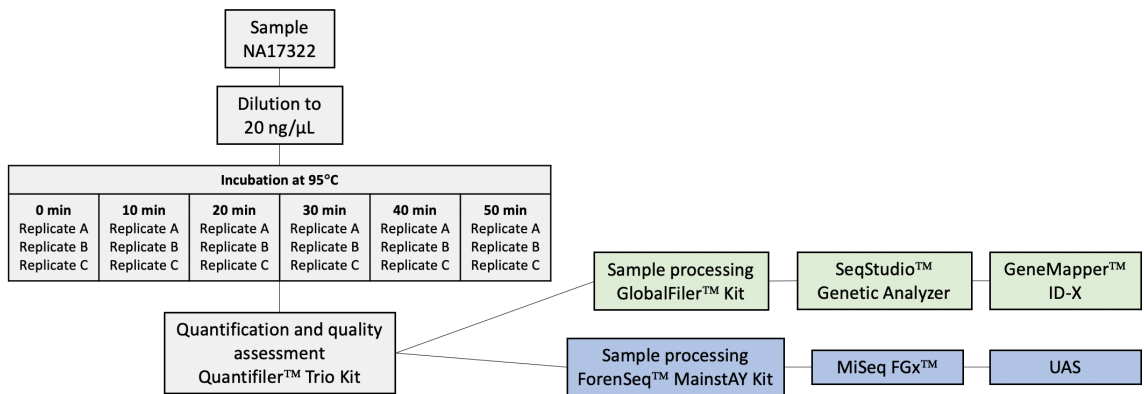
## 2. MATERIALS AND METHODS

### 2.1 Repeatability Study

The repeatability of the ForenSeq™ MainstAY Kit on the MiSeq™ FGx System was assessed using seven replicates of a commercially available single-source male sample, NA17325 (Coriell Cell Repositories, Coriell Institute for Medical Research, Camden, NJ), run at 1 ng of total input DNA following the manufacturer’s protocol.

### 2.2 Mock Degradation Study

A commercial single-source female sample, NA17322 (Coriell Cell Repositories, Coriell Institute for Medical Research, Camden, NJ) was used for this study. The workflow followed is outlined in Figure 2.



**Figure 2. Outline of mock-degradation study workflow.** Steps in green boxes are specific to the conventional PCR-CE workflow, and blue boxes to the NGS workflow.

#### 2.2.1 DNA Degradation and Quantification

Based on a degradation study by *Sharma et al.*, 10 microliter ( $\mu\text{L}$ ) samples of 20 ng/ $\mu\text{L}$  DNA were degraded by incubation at  $95^{\circ}\text{C}$  in triplicate for six time periods (0 min,

10 min, 20 min, 30 min, 40 min, 50 min) using a SimpliAmp™ Thermal Cycler (Applied Biosystems, Foster City, CA) (61). To assess the level of degradation and quantify the samples for further processing, quantitation was performed in duplicate using the Quantifiler™ Trio Kit (Applied Biosystems, Foster City, CA) on a 7500 Real-Time PCR System (Applied Biosystems, Foster City, CA) using 2 µL of sample, according to the manufacturer's established protocols. Sample concentrations were determined using a calibrated standard curve. Sample concentrations from duplicates were averaged using small autosomal quantities. The degradation index (DI) provided a qualitative measure of the degradation of each sample, and was calculated using the ratio of the small autosomal target to the large autosomal target. These degraded and quantified samples were then used for PCR-CE and sequencing.

### 2.2.2 PCR Amplification and Capillary Electrophoresis

The 18 samples processed for DNA degradation were amplified using the GlobalFiler™ PCR Amplification Kit (Applied Biosystems, Foster City, CA) on a Veriti™ Thermal Cycler (Applied Biosystems, Foster City, CA) for comparison testing, following manufacturer's protocol. This 6-dye kit tests for 24 STR loci listed in Table A. Sample amplifications targeted 0.75 ng, and included a positive control of DNA Control 007 (5 µL) and a negative control of master mix (5 µL) for quality control. The PCR amplification parameters on the Veriti™ with a ramp rate set to 9600 were as follows: denaturation for 1 minute at 95°C, amplification for 29 cycles of (10 sec at 94°C, 90 sec at 59°C), extension for 10 min at 60°C, and a final soak at 4°C.

Following the Boston University laboratory validated “Capillary Electrophoresis using SeqStudio™ and GlobalFiler™” protocol, 1  $\mu$ L PCR product was added to the appropriate wells of a 96-well reaction plate. Each well also received 1  $\mu$ L of allelic ladder, along with a Hi-Di formamide and LIZ 600 size standard master mix (10  $\mu$ L). The samples were then denatured at 95°C for 3 minutes on a GeneAmp™ PCR System Thermal Cycler (Applied Biosystems, Foster City, CA), and immediately chilled at -20°C for 3 minutes. Capillary electrophoresis was performed using a SeqStudio™ Genetic Analyzer (Applied Biosystems, Foster City, CA) with parameters for injection of 1.2 kilovolts (kV) for 5 seconds. Resulting EPGs were analyzed using the GeneMapper™ ID-X version 1.6 (Applied Biosystems, Foster City, CA) with stutter filter on and an AT of 50 RFU.

### **2.3 Mixture Study**

Human genomic DNA mixtures were prepared from the following three commercial single-source samples (Coriell Cell Repositories, Coriell Institute for Medical Research, Camden, NJ): NA17323 (female, Italian), NA17326 (male, Italian), and NA17330 (female, Italian). Following dilutions in nuclease-free water, two-person and three-person mixtures were created at varied contributor ratios, detailed in Table 1. Mixtures were sequenced using less than 1 ng of total input DNA due to limited sample availability (Table 1). Each single-source sample used as a “contributor” was sequenced for use in later mixture deconvolution using 1 ng of total input DNA following the manufacturer’s protocol.

**Table 1. Proportions of contributors and DNA input for sequencing of mixtures.** Two-person and three-person mixture proportions along with the amount of DNA used for sequencing are listed in decreasing order of minor contribution percentage. Minor contributions for each mixture ratio are bolded.

	Sample	Mixture Ratio	Proportion of DNA (ng)			Minor Contribution	DNA Input for Sequencing (ng)
			Contributor 1	Contributor 2	Contributor 3		
Two-Person Mixtures	C1	4:1	0.80	0.20	-	20%	0.54
	D1	8:1	0.89	0.11	-	11%	0.53
Three-Person Mixtures	C3	4:1:1	0.67	0.17	0.17	17%	0.54
	B3	4:2:1	0.57	0.29	0.14	14%	0.56
	C4	4:1:2	0.57	0.14	0.29	14%	0.56
	C5	4:1:4	0.44	0.11	0.44	11%	0.56
	D2	8:1:1	0.80	0.10	0.10	10%	0.54
	C2	8:2:1	0.73	0.18	0.09	9%	0.54
	D3	8:1:2	0.73	0.09	0.18	9%	0.54
	B2	8:4:1	0.62	0.31	0.08	8%	0.54
	D4	8:1:4	0.62	0.08	0.31	8%	0.54
	A2	8:8:1	0.47	0.47	0.06	6%	0.56
D5	8:1:8	0.47	0.06	0.47	6%	0.56	

#### 2.4 ForenSeq™ MainstAY Library Preparation and MiSeq™ FGx Sequencing

The Verogen ForenSeq™ MainstAY Kit was used to prepare all samples for sequencing according to the manufacturer’s protocol detailed in the ForenSeq™ MainstAY Kit Reference Guide, outlined in Figure 3 (62). A sample sheet was created to record the positions of each sample, control, and index adapter. In a pre-PCR environment to avoid cross-contamination, each well of the ForenSeq Sample Plate (FSP) received 7 µL of master mix (4.7 µL PCR1, 0.3 µL FEM, 2 µL DPMC). Purified DNA input of 1 ng (8 µL)

was added to the appropriate wells for all samples, with the exception of mixture samples that received lower DNA inputs (8  $\mu$ L) detailed in Table 1 due to limited sample availability. A positive amplification control of diluted NA24385 control DNA (8  $\mu$ L) and negative amplification control of nuclease-free water (8  $\mu$ L) were also added to the FSP for sequencing quality control, and the plate was sealed with Microseal 'A' and centrifuged at 1000 x g for 30 seconds. Target amplification and tagging was achieved by placing the samples on a Veriti™ Thermal Cycler using the following PCR1 program parameters with 8% ramp mode: 98°C for 3 minutes, 8 cycles of (96°C for 45 seconds, 80°C for 30 seconds, 54°C for 2 minutes with ramp, 68°C for 2 minutes with ramp), 10 cycles of (96°C for 30 seconds, 68°C for 3 minutes with ramp), 68°C for 10 minutes, and an indefinite hold at 10°C.



**Figure 3. Overview of the Verogen ForenSeq™ MainstAY protocol.** The six steps of the workflow to prepare libraries, with estimated hands-on and total times taken to complete each step.

Target enrichment amplified the DNA and added UDI adapters required for cluster generation and sample identification for analysis. The FSP and UDI plate were centrifuged at 1000 x g for 30 seconds, and each sample received 8 µL of a distinctive UDI adapter that later permitted the discrimination between samples during analysis. Each well of the FSP received 27 µL of PCR2, followed by sealing with Microseal ‘A’ and centrifugation at 1000 x g for 30 seconds. The plate was placed on the thermal cycler for a second amplification (PCR2m) according to the following parameters: 98°C for 30 seconds, 15

cycles of (98°C for 20 seconds, 66°C for 120 seconds), 68°C for 10 minutes, and an indefinite hold at 10°C.

The amplified libraries were then purified using purification beads and an enzyme. Each well of a Purification Bead Midi Plate (PBP) received 45 µL of SPB2 beads, and 45 µL of each sample in the FSP were transferred to the corresponding well of the PBP. The PBP was then sealed with Microseal 'B' and shaken at 1800 rpm for 2 minutes, incubated at room temperature for 5 minutes, and placed on a magnetic stand until the liquid was transparent (approximately 5 minutes). All supernatant was removed from each well, and the plate was washed twice as follows: 200 µL of fresh 80% ethanol was added to each well, the PBP was incubated for 30 seconds, and all supernatant was removed and discarded. Residual ethanol was discarded from each well, and the PBP was removed from the magnetic stand. Each well received 52.5 µL of RSB, and the PBP was sealed and shaken at 1800 rpm for 2 minutes. Following incubation at room temperature for 2 minutes, the PBP was placed on the magnetic stand until the liquid became clear (approximately 2 minutes). The supernatant from each well (50 µL) was then transferred to the corresponding well of the Purified Library Plate (PLP), and the PLP was sealed and centrifuged at 1000 x g for 30 seconds.

The libraries were then normalized to ensure even representation, optimize the resolution of each library in the pool, and achieve consistent cluster density without an additional quantification step. Each well of a Normalization Working Midi Plate (NWP) received 45 µL of master mix (46.8 µL LNA1, 8.5 µL LNB1). The PLP was placed on a magnetic stand until the liquid was clear to remove any beads that may have aspirated, and

20  $\mu\text{L}$  of supernatant from each well of the PLP was transferred to the corresponding well of the NWP. The NWP was sealed and shaken at 1800 rpm for 30 minutes, and immediately placed on a magnetic stand until the liquid became clear. All supernatant was removed and discarded, and the NWP was removed from the magnetic stand. The NWP was then washed twice as follows: 45  $\mu\text{L}$  of LNW1 was added to each well, the plate was sealed and shaken at 1800 rpm for 5 minutes, the plate was placed on the magnetic stand until the liquid became clear, the supernatant was removed, and the plate was removed from the magnetic stand. Following sealing and centrifugation at 1000 x g for 30 seconds, the NWP was placed on the magnetic stand again until the liquid became clear, and residual LNW1 was removed from each well. Fresh 0.1 N HP3 was prepared by combining 33.3  $\mu\text{L}$  of nuclease-free water and 1.8  $\mu\text{L}$  of HP3, which was then added (32  $\mu\text{L}$ ) to each well. The NWP was sealed and shaken at 1800 rpm for 5 minutes, and placed on the magnetic stand until the liquid became clear. Each well of the Normalization Library PCR Plate (NLP) received 30  $\mu\text{L}$  of LNS2, followed by 30  $\mu\text{L}$  of supernatant from the NWP. Finally, the NLP was sealed and centrifuged at 1000 x g for 30 seconds.

Equal volumes of each normalized library were then combined to create a library pool to be sequenced. Pooling was accomplished by transferring 5  $\mu\text{L}$  of each library to a 1.5 mL microcentrifuge tube labeled PNL for Pooled Normalized Libraries, which was vortexed and centrifuged.

Before being placed on the MiSeq FGx<sup>TM</sup> instrument, the libraries were denatured and diluted to the loading concentration. The PNL tube was heated at 96°C for 2 minutes, and 12  $\mu\text{L}$  was immediately transferred from the PNL tube to the DNL tube for Denatured

Normalized Libraries containing 600  $\mu\text{L}$  of HT1. A denatured HSC tube containing 2  $\mu\text{L}$  of Human Sequencing Control (HSC), 2  $\mu\text{L}$  of HP3, and 36  $\mu\text{L}$  of nuclease-free water was vortexed, centrifuged, and incubated at room temperature for 5 minutes. The HSC is a DNA library pool of 23 STRs that serves as a positive sequencing control for the MiSeq FGx<sup>TM</sup> instrument. Finally, the DNL tube received 4  $\mu\text{L}$  of denatured HSC, and was vortexed and centrifuged. The entire DNL tube volume was immediately pipetted into the MiSeq FGx<sup>TM</sup> Reagent Cartridge (Illumina Inc., San Diego, CA) for sequencing and placed on the instrument with a micro flow cell, according to manufacturer's protocol.

## **2.5 Data Analysis**

MiSeq FGx<sup>TM</sup> sequencing data was analyzed using the MainstAY analysis module in the ForenSeq<sup>TM</sup> Universal Analysis Software (version 2.0) with default settings. The UAS provided allele and genotype calling, genotype sequences, flags, and read coverage information. Default analysis parameters used were a 1.5% AT and 1.5% IT determined as a percentage of the total reads per locus, for all loci except for DYS481 (2% AT, 2% IT), DYS612 (2% AT, 2% IT), and DYS38911 (4.5% AT, 4.5% IT). A default intralocus balance of 60% was used for all loci for single-source samples. If the balance of read counts typed between alleles at a heterozygous locus falls below this value, an "imbalanced" flag is assigned. Stutter filtering ranged from 0-47% between loci, based on developmental validation by Verogen for the ForenSeq<sup>TM</sup> for the MainstAY Kit (63). Once all results were analyzed in the UAS, secondary analyses were performed using Microsoft Excel.

### 3. RESULTS AND DISCUSSION

#### 3.1 Controls and Quality Assessment

The sequencing run passed all run quality metrics (Table 2), along with all read and index quality metrics. Cluster density (K/mm<sup>2</sup>) is the number of clusters per square millimeter for the run, and clusters passing filter (%) is the percentage of clusters that can pass the chastity filter, an internal quality filtering procedure. The phasing rate (%) and pre-phasing rate (%) are the percentage estimates of molecules within a cluster that fall behind or jump ahead of the current cycle in Read 1 (59). The Human Sequencing Control showed no discordant loci, and passed in overall intensity. Both amplification controls performed as expected, as the positive amplification control showed 100% typing of STRs (n=53 loci), and the negative amplification control showed no reads. All samples met the sample representation guideline provided by Verogen of 15,000 reads for this kit, except for three samples due to user error during library preparation. Thus, the following samples were omitted from data analysis: one replicate sample for the repeatability study, one replicate from the 50 min incubation in the degradation study, and a single-source sample for the mixture study. This guideline is recommended to achieve optimal coverage of all loci.

**Table 2. Quality metrics for the MiSeq FGx™ sequencing run.**

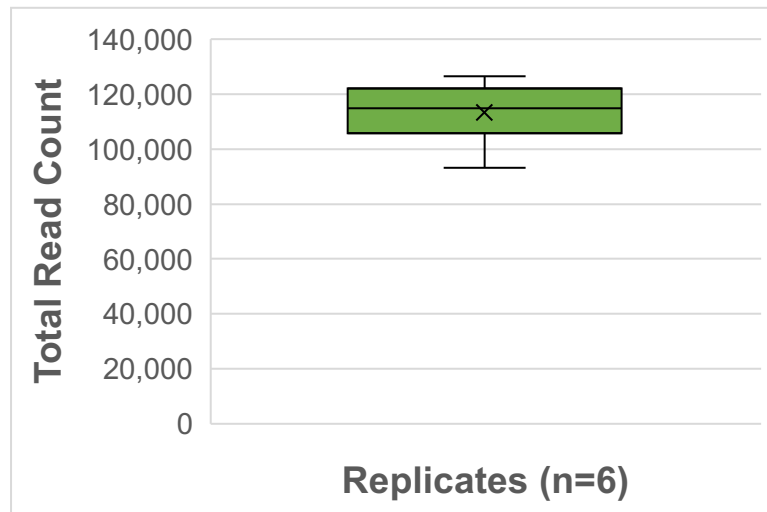
	<b>Cluster Density (K/mm<sup>2</sup>)</b>	<b>Clusters Passing Filter (%)</b>	<b>Phasing (%)</b>	<b>Pre-Phasing (%)</b>
Target Range	400 - 1650	≥ 80	≤ 0.25	≤ 0.15
Run Metric	1407	88.48	0.187	0.091

### 3.2 Repeatability Study

A repeatability study was performed using six replicates of a commercial male single-source DNA sample to assess the performance of the ForenSeq™ MainstAY Kit. Concordance of the genotypes called by the UAS was determined by comparing it to the CE data. For the 22 loci shared between the ForenSeq™ MainstAY and GlobalFiler™ Kits, high accuracy was observed for all replicates for 38 Au-STR alleles (98.7%) and 1 Y-STR allele (100%). Three discordant alleles were detected out of 231 Au-STR alleles. These discrepant alleles can be attributed to stutter alleles exceeding the default stutter filters at the D1S1656 and D16S539 loci (Table 3). Uniformly high total read count (intensity) was observed for the six replicates, with a range of 93,122 to 126,525 reads, and an average of 113,331 reads (Figure 4). Proving the repeatability of results using this kit further establishes the kit’s ability to produce accurate and repeatable genotype calling for all experimentation, also further demonstrated by Verogen (64).

**Table 3. Discordant alleles observed in the repeatability experiment.** Three discrepant alleles and respective read counts observed out of 231 autosomal STR alleles for six replicates of a commercial sample analyzed using the ForenSeq™ MainstAY Kit. These discordant alleles are results of stutter alleles exceeding default stutter filters.

Locus	True Allele	Read Count	Discordant Allele	Read Count	Stutter Filter	Observed Stutter
D1S1656	15	434	14	85	19%	19.6%
D16S539	12	858	11	141	16%	16.4%
D16S539	12	1079	11	185	16%	17.1%



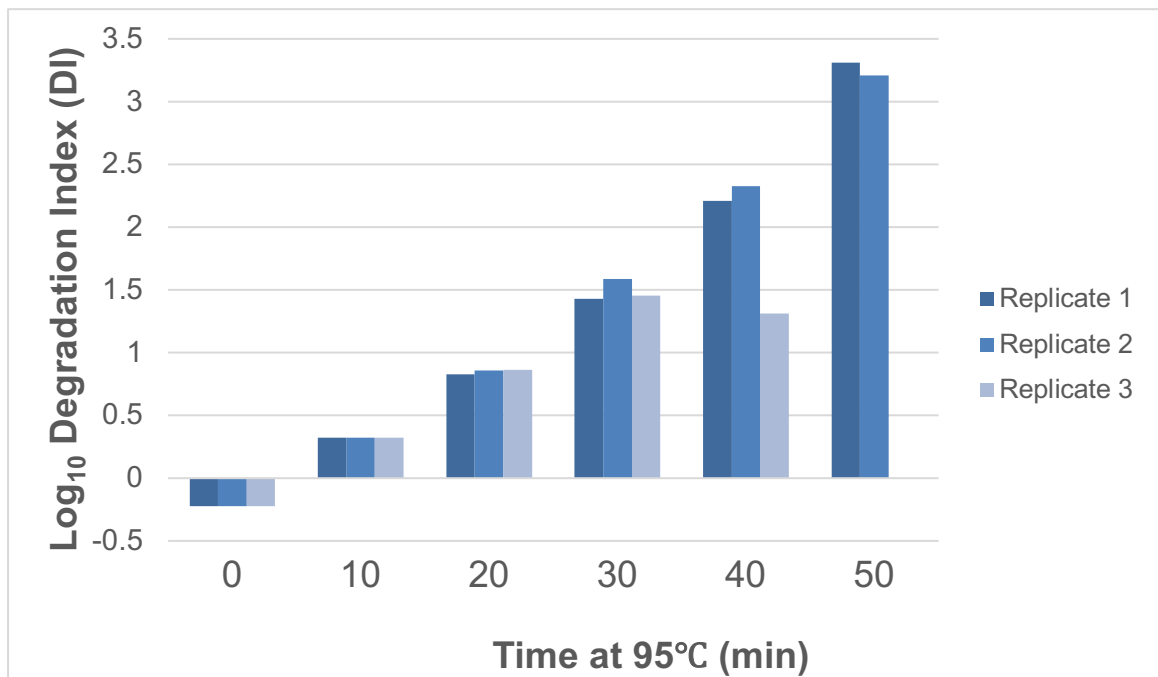
**Figure 4. Variation in total read count for repeatability study.** A boxplot of total number of reads (intensity) for six replicates analyzed using the ForenSeq™ MainstAY Kit.

### 3.3 Mock-Degradation Study

#### 3.3.1 Effects of Varying Degrees of Degradation

Pristine DNA of high concentration (20 ng/μL) was degraded by heating for several periods of time with three replicates per time point, followed by quantity and quality assessment using the Quantifiler™ Trio Kit. As the time the DNA was heated at 95°C increased, the DIs consistently increased (Figure 5). Average DI and concentration were calculated for each of six time points, along with standard deviations (Table 4). The average DI between replicates increased from 0.6 for untreated samples to 1832.6 for samples heated for the longest time period (Table 4). Untreated samples showed no degradation ( $DI < 1$ ) as expected, samples heated for 10 and 20 minutes showed moderate degradation ( $DI < 10$ ), and samples heated for 30-50 minutes showed severe degradation ( $DI > 10$ ). There was minimal variation in DI between replicates at each timepoint, with

the exception of samples heated for 40 minutes ( $131.6 \pm 99.4$ ). Despite all samples initially having the same qPCR input concentration, increased incubation at 95°C caused a decrease in the measured concentration also observed in a similar experiment conducted by *Sharma et al.* and additional degradation studies (Table 4) (61,65,66). This is explained by the qPCR assay amplifying regions of longer lengths that are more susceptible to degradation. That is, the effective DNA concentration of longer fragments is reduced in comparison to the effective concentration of shorter DNA fragments. The DI values observed at 40- and 50-minute heat exposure were much higher in this study than previously referenced, likely due to the heat-induced decay of large DNA fragments driving up the DI values calculated.



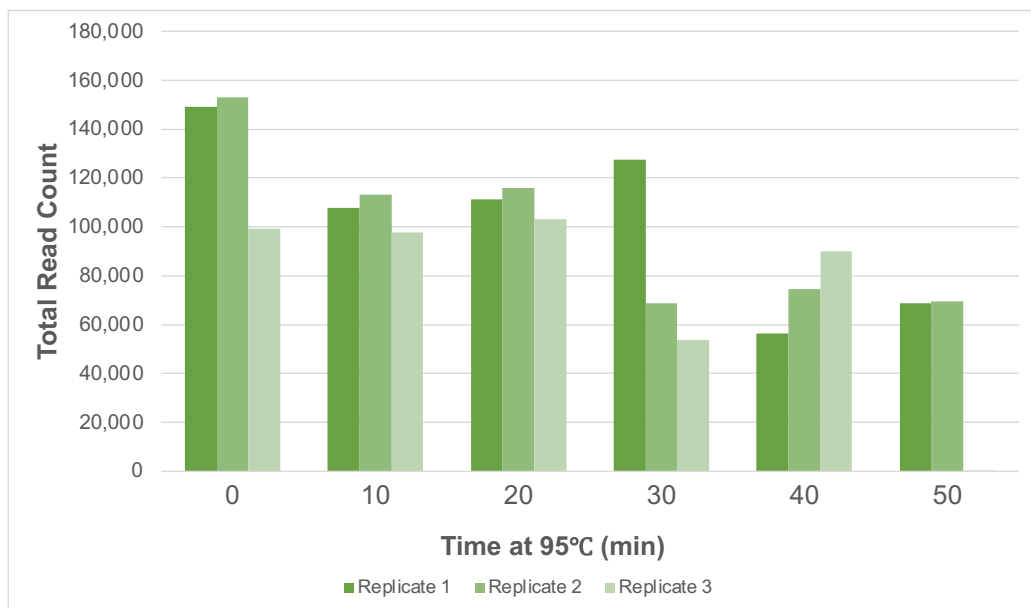
**Figure 5. Effect of increasing DNA incubation time at 95°C on degradation index.** Degradation indices for replicates at each time point are plotted on a log<sub>10</sub> scale.

**Table 4. Average degradation index and concentration across all time points.**

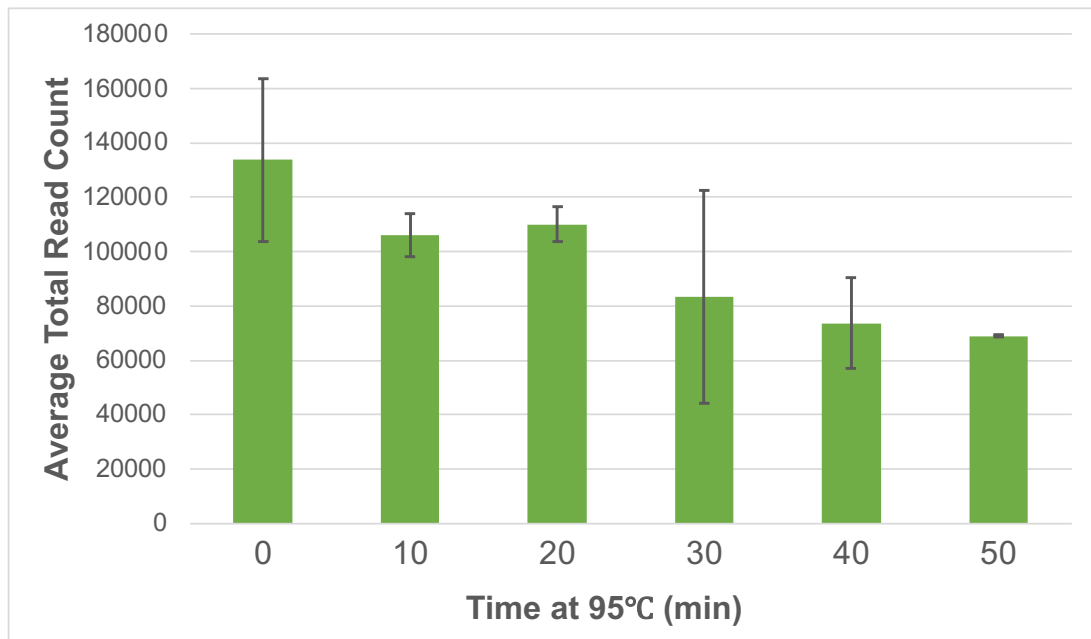
<b>Time at 95°C (min)</b>	<b>Degradation Index (DI) (Mean ± SD)</b>	<b>Concentration (ng/μL) (Mean ± SD)</b>
<b>0</b>	0.6 ± 0.0	19.4 ± 1.4
<b>10</b>	2.1 ± 0.0	9.6 ± 0.9
<b>20</b>	7.1 ± 0.3	4.7 ± 0.5
<b>30</b>	31.3 ± 6.3	2.5 ± 0.2
<b>40</b>	131.6 ± 99.4	1.0 ± 0.1
<b>50</b>	1832.6 ± 295.5	0.5 ± 0.0

The total read counts observed using the ForenSeq™ MainstAY Kit on the MiSeq FGx™ System for replicates at each time point were compared in this study. As the time the DNA was heated at 95°C increased, the total read count decreased overall (Figure 6). Average total read counts calculated across replicates were compared at each period of incubation at 95°C (Figure 7). The average total read count for the untreated DNA was 133,756 reads, which decreased to 68,989 reads for samples incubated for 50 minutes. Untreated samples and samples incubated at 95°C for 30 minutes had the greatest variation in average total read count between replicates, shown by the error bars in Figure 7. As replicates at these time points do not vary significantly in DI, the difference in total read counts could be a result of library preparation or instrument variation. The nearly two-fold decrease in read counts observed from untreated to longest incubation time would be expected, as read count is a measure of signal in DNA sequencing. In conventional PCR-CE analysis, peak heights typically follow a downward trend as PCR amplicon length increases, also known as a “degradation slope” (67). Similarly, the decrease in overall read count in sequencing would be expected as samples become severely degraded, thus having

decreasing signal for longer target amplicons. However, the ForenSeq™ MainstAY Kit showed success in meeting the sample representation guideline of 15,000 reads for all samples, including those that were severely degraded. This guideline is recommended to achieve optimal coverage for all loci, and was even met 4.5-fold for the sample with the highest DI of 2041.6. This is likely a benefit of the kit's small number of markers multiplexed in a single sequencing reaction with overall shorter amplicons targeted, resulting in a more consistent average total read count despite the high level of degradation observed by samples incubated for longer periods of time.



**Figure 6. Effect of increasing DNA incubation time at 95°C on total read count.** The total of numbers of sequencing reads (intensity) are plotted for replicates analyzed using the ForenSeq™ MainstAY Kit.



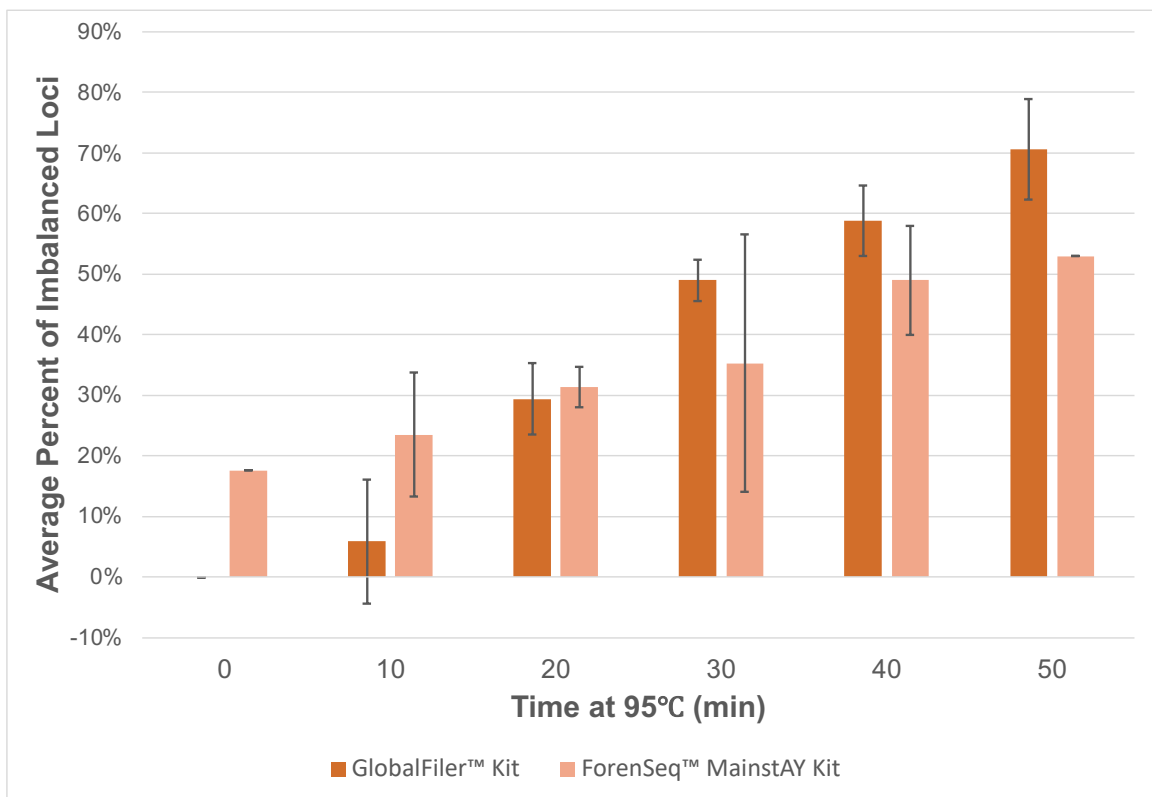
**Figure 7. Effect of increasing DNA incubation time at 95°C on average total read count.** The average total of numbers of sequencing reads (intensity) are plotted for samples analyzed using the ForenSeq™ MainstAY Kit. Standard deviation per time point is included as error bars.

### 3.3.2 Performance of the ForenSeq™ MainstAY Kit versus the GlobalFiler™ Kit

To assess the performance of NGS technology in comparison to the standard PCR-CE method, the ForenSeq™ MainstAY Kit was evaluated alongside the GlobalFiler™ PCR Amplification Kit for the mock-degraded samples.

To compare the heterozygous balance of amplicons between the two kits for degraded samples, the percent of imbalanced loci was determined across samples heated at each time point (Figure 8). The intralocus balance, or heterozygous balance, was calculated as a ratio of the lower to higher number of allele intensity at a heterozygous locus. The ForenSeq™ UAS flags loci with read count ratios lower than 60%, and these flags were counted for samples as degradation increased. Similarly, peak heights ratios

were calculated for the samples analyzed by the GlobalFiler™ Kit and considered imbalanced if the ratio did not exceed 60%. As expected, the percent of imbalanced loci generally increased for both kits as heating of samples increased (Figure 8). The average percent of imbalanced loci increased from 0% for untreated samples to 70.6% for samples heated for 50 minutes (DI: 1832.6) using the GlobalFiler™ Kit (Table 5). While imbalanced loci increased less drastically from 17.6% to 52.9% for the most heated samples using the MainstAY Kit, the increase in imbalance observed using both kits can be explained by the amplification biases for small amplicons (64).

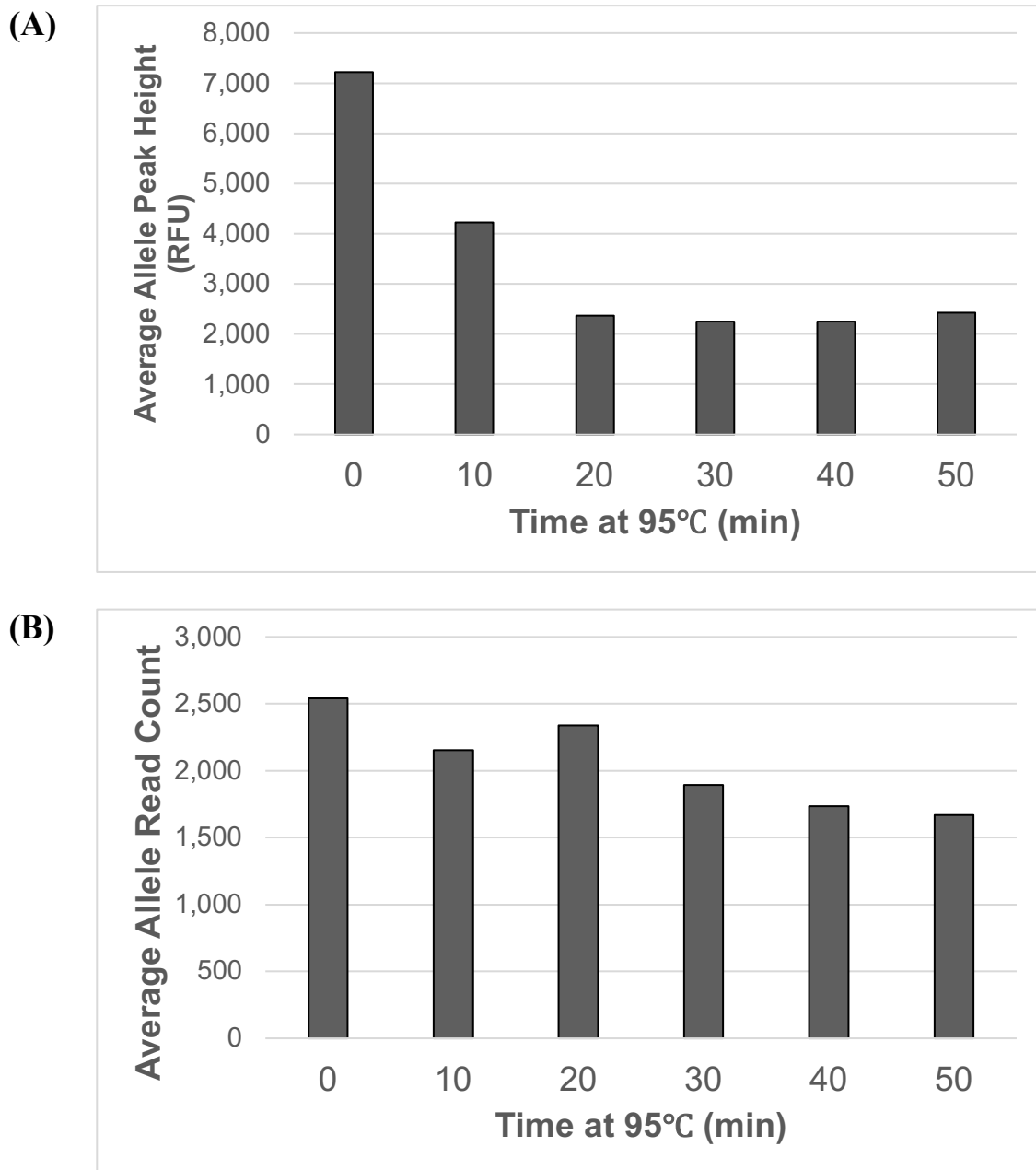


**Figure 8. Effect of increasing DNA incubation time at 95°C on average intralocus balance.** The average percent of imbalanced loci for replicates analyzed using the GlobalFiler™ and ForenSeq™ MainstAY Kits across 17 shared, heterozygous loci. Locus imbalance was identified if the balance of peak heights or read counts between alleles was lower than 60%. Standard deviation per time point is included as error bars.

**Table 5. Average percent of imbalanced loci for mock-degraded DNA.** Comparison of average intralocus imbalance (<60%) observed between the GlobalFiler™ and ForenSeq™ MainstAY Kits across 17 shared, heterozygous loci.

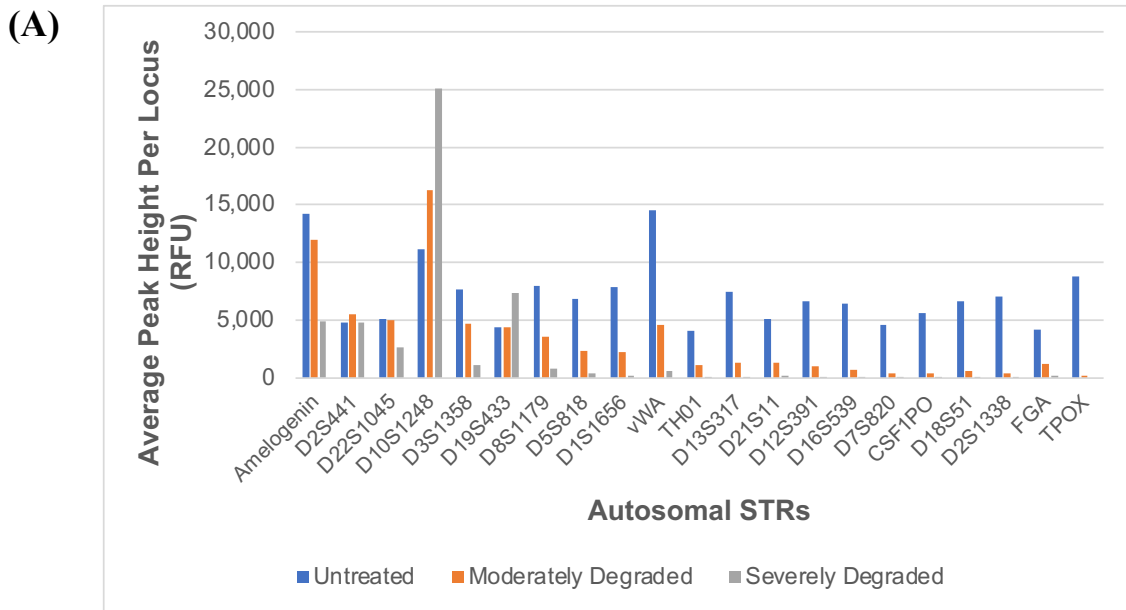
<b>Time at 95°C (min)</b>		0	10	20	30	40	50
<b>Average DI</b>		<b>0.6</b>	<b>2.1</b>	<b>7.1</b>	<b>31.3</b>	<b>131.6</b>	<b>1832.6</b>
<b>Percent Imbalanced Loci (Mean ± SD)</b>	GlobalFiler™ Kit	0.0 ±	5.9 ±	29.4 ±	49.0 ±	58.8 ±	70.6 ±
		0.0	10.2	5.9	3.4	5.9	8.3
	ForenSeq™ MainstAY Kit	17.6 ±	23.5 ±	31.4 ±	35.3 ±	49.0 ±	52.9 ±
		0.0	10.2	3.4	21.2	9.0	0.0

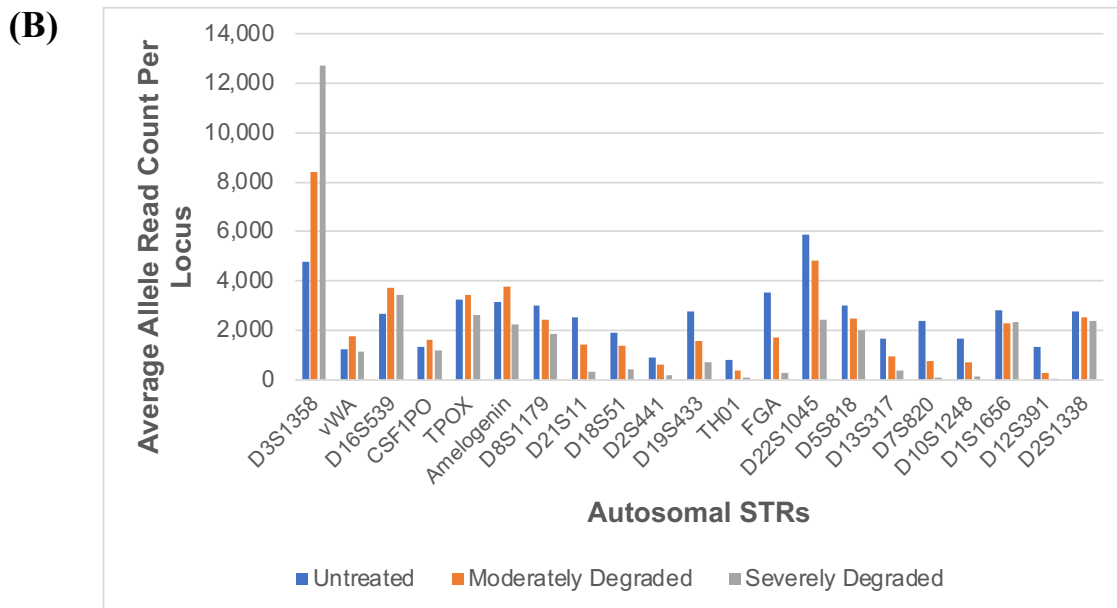
The average true allele signal intensities across 21 autosomal loci shared between the two kits were calculated across replicates at each incubation time point (Figure 9). While signal intensity is measured by peak height (RFU) for CE-based analysis, the sequencing UAS measures allele read count (ARC). The overall average peak height significantly decreased after 10 minutes of heating for the GlobalFiler™ Kit, with an approximate three-fold decrease between untreated and most degraded samples. However, average ARC for the ForenSeq™ MainstAY Kit remained more consistent as heating time increased, with an approximate 1.5-fold decrease between untreated and most degraded samples. Therefore, heating the DNA at 95°C caused a two-fold greater impact on average allele signal intensity for CE-based versus NGS analysis, suggesting the MainstAY Kit’s increased capability of detecting signal for samples with varying degrees of degradation.



**Figure 9. Effect of increasing DNA incubation time at 95°C on average allele signal intensity.** Comparison of the effect of increased incubation periods for samples analyzed using the (A) GlobalFiler™ Kit and (B) ForenSeq™ MainstAY Kit, calculated for 21 shared autosomal loci.

To assess the impact of amplicon size and increasing degrees of degradation on allele signal intensity, the average peak heights and ARCs were calculated for the 21 autosomal loci shared between both kits (Figure 10). Plotting the average peak height per locus for the GlobalFiler™ Kit revealed that as the amplicon length increased, the average peak heights decreased for all degrees of degraded samples. This inverse relationship became more apparent as the degree of degradation became more severe. However, this “degradation slope” phenomenon was less apparent for both moderately and severely degraded samples analyzed using the ForenSeq™ MainstAY Kit, as average ARC per locus remained more consistent as amplicon size increased. These results highlight the benefit of the MainstAY Kit’s targeting of a high density of short amplicons (<200 bp).





**Figure 10. Average allele signal intensity per locus for mock-degraded DNA.** Comparison of the average allele signal intensity for samples analyzed using the (A) GlobalFiler™ Kit and (B) ForenSeq™ MainstAY Kit, calculated for 21 shared autosomal loci. Loci are listed in increasing amplicon size, using the longest amplicon size per locus for sorting. Untreated samples (n=2) were incubated for 0 minutes, moderately degraded (n=6) for 10-20 minutes, and severely degraded (n=9) for 30-50 minutes at 95°C.

Locus and allele dropouts observed for mock-degraded samples were determined by comparison to untreated samples (Table 6). For samples analyzed with the GlobalFiler™ Kit, locus dropout was first observed for samples heated for 20 minutes (DI: 7.1) at D2S1338 (Table 6A). In comparison, locus dropouts were first observed for samples heated for 40 minutes (DI: 131.6) when analyzed with the ForenSeq™ MainstAY Kit (Table 6B). Across all incubation time points, GlobalFiler™ exhibited more locus and allele dropouts than the MainstAY Kit, especially for loci with amplicon lengths >300 bp (Table 6).

**Table 6. Dropouts of STR loci and alleles for mock-degraded DNA.** Dropouts observed for samples analyzed using the (A) GlobalFiler™ Kit and (B) ForenSeq™ MainstAY Kit, across 21 shared autosomal loci. Loci with longest amplicons >300 bp are underlined.

(A)

GlobalFiler™ PCR Amplification Kit (CE)			
Time at 95°C (min)	Average DI	Locus Dropout	Allele Dropout
0	0.6		
10	2.1		
20	7.1	<u>D2S1338</u>	<u>CSF1PO, D7S820, D2S1338</u>
30	31.3	<u>CSF1PO, TPOX, D7S820, D2S1338</u>	D16S539, <u>CSF1PO, D18S51, TH01, D13S317, D7S820</u>
40	131.6	D16S539, <u>CSF1PO, TPOX, D18S51, D7S820, D2S1338</u>	<u>CSF1PO, D18S51, TH01, D13S317, D1S1656, D12S391, D2S1338</u>
50	1832.6	D16S539, <u>CSF1PO, TPOX, D21S11, D18S51, TH01, D13S317, D7S820, D12S391, D2S1338</u>	D21S11, TH01, D5S818, D13S317, D12S391

(B)

ForenSeq™ MainstAY Kit (NGS)			
Time at 95°C (min)	Average DI	Locus Dropout	Allele Dropout
0	0.6		
10	2.1		
20	7.1		
30	31.3		D12S391
40	131.6	D12S391, D22S1045	D1S1656, D12S391
50	1832.6	D1S1656, D12S391, D22S1045	D1S1656, D16S539, D21S11

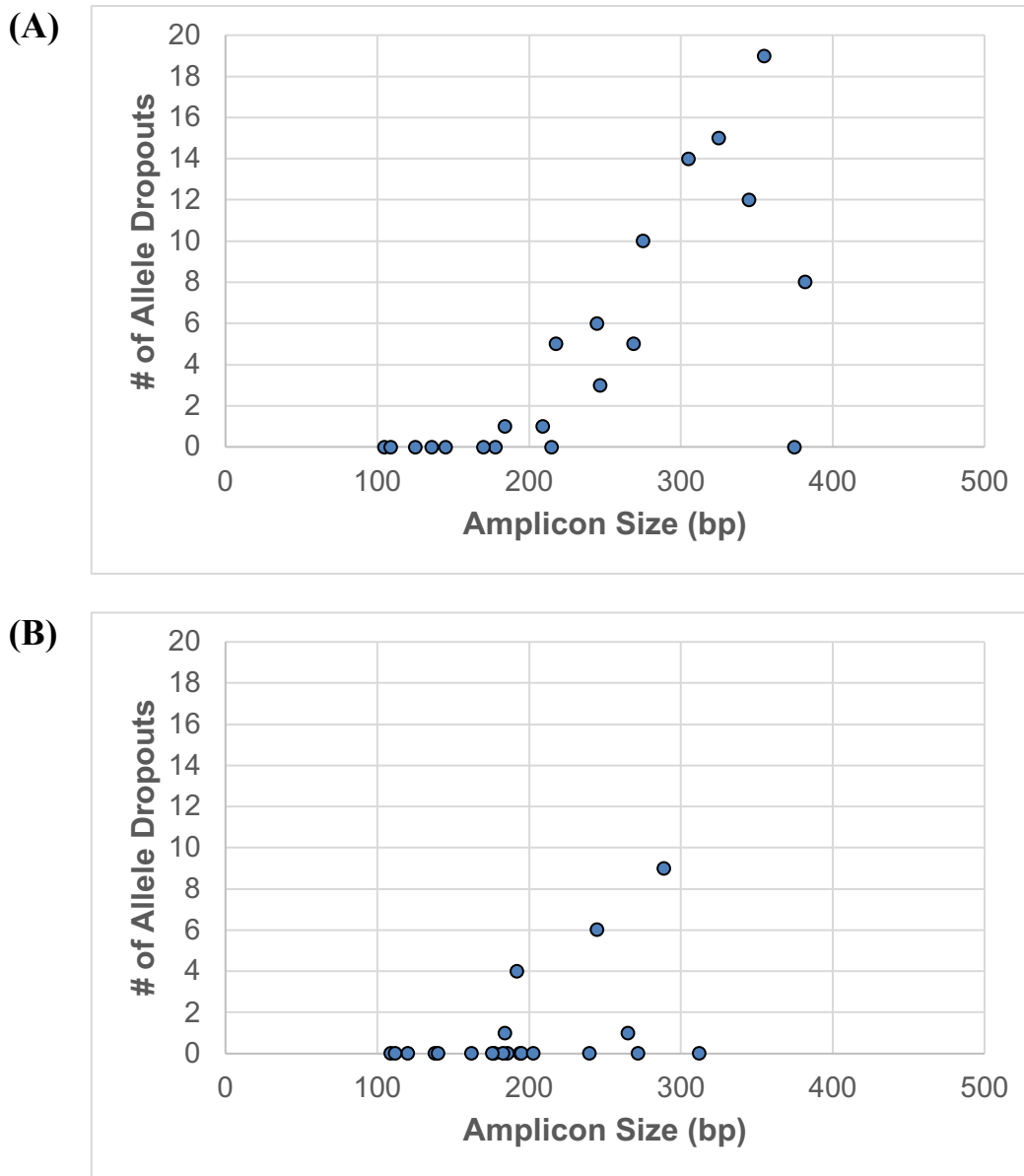
Allele dropouts were calculated as percentages of the total true alleles across all replicates at each time point (Table 7). As previously observed by *Zhang et al.*, increasing DIs resulted in greater allele dropouts (65). For the GlobalFiler™ Kit, no ADO was observed until incubation for 20 minutes (DI: 7.1) where 5.3% of alleles dropped out, which increased to 46.1% for the highest incubation time (DI: 1832.6). However, no ADO was observed until incubation for 30 minutes (DI: 31.3) for the ForenSeq™ MainstAY Kit, which only increased to 17.1% for the highest incubation time. More loci being affected by increasing DIs (longer incubation times) and greater percentages of overall ADO suggest that GlobalFiler™ is considerably more sensitive to degraded DNA than the ForenSeq™ MainstAY Kit, likely due to a greater amount of long amplicons (>200 bp) targeted by the kit.

**Table 7. Effect of increasing DNA incubation time at 95°C on allele dropout.** Comparison of allele dropouts observed as percentages of total true alleles across replicates for 21 autosomal loci shared between the GlobalFiler™ and ForenSeq™ MainstAY Kits.

<b>Time at 95°C (min)</b>		0	10	20	30	40	50
<b>Average DI</b>		<b>0.6</b>	<b>2.1</b>	<b>7.1</b>	<b>31.3</b>	<b>131.6</b>	<b>1832.6</b>
<b>Allele Dropout</b>	GlobalFiler™ Kit	0.0%	0.0%	5.3%	23.7%	27.2%	46.1%
	ForenSeq™ MainstAY Kit	0.0%	0.0%	0.0%	1.8%	5.3%	17.1%

Literature has demonstrated that in the analysis of highly degraded DNA, poor amplification of longer loci (300-500 bp) in multiplex typing kits is common (68–71). Thus, as peak height has been proven to be inversely proportional to amplicon length,

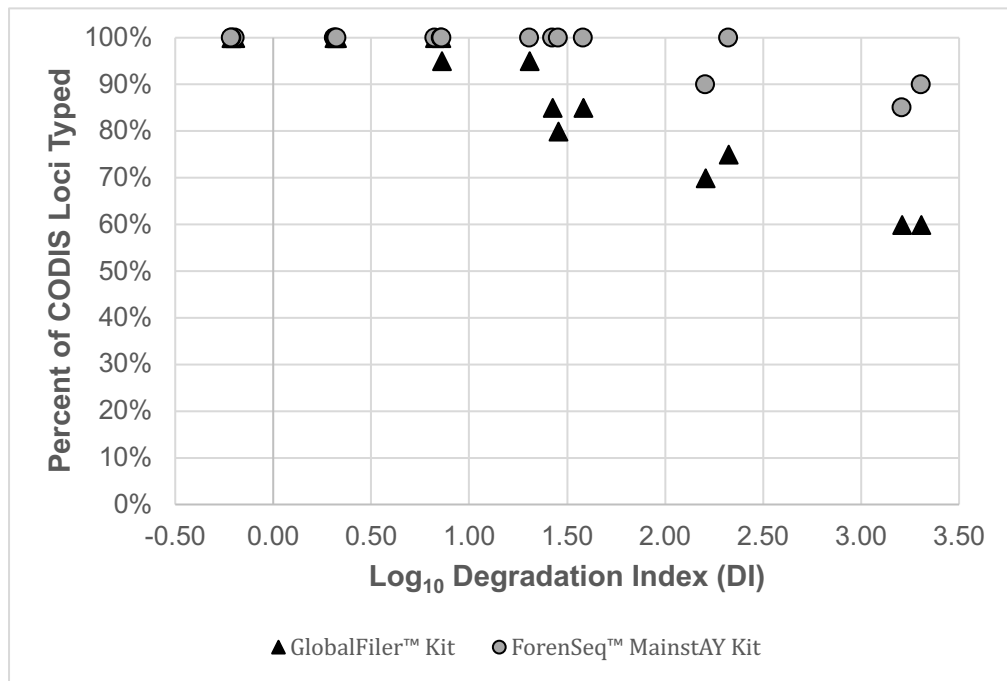
longer amplicons can fall below detection thresholds causing increased ADO (68,72). The total number of ADOs observed across all mock-degraded samples was plotted against the longest amplicon size for loci shared between the two kits (Figure 11). The longest amplicon lengths range from 105 to 382 bp for the GlobalFiler™ Kit, and 109 to 312 bp for the ForenSeq™ MainstAY Kit. The number of observed ADOs increased as amplicon size increased for GlobalFiler™, especially as lengths exceeded 200 bp (Figure 11A). However, this relationship was considerably less apparent for the MainstAY Kit, due to the significantly lower number of ADOs overall and more long amplicons with no dropout observed (Figure 11B). The 63 total ADOs observed for amplicons >200 bp in GlobalFiler™ compared to the 16 observed in the MainstAY Kit highlight the kit's increased capability in detecting alleles for degraded samples. Unlike PCR-CE, NGS has no limit on amplicon size, allowing for the smallest possible amplicon sizes to be targeted (73,74). This advantage is demonstrated by the ForenSeq™ MainstAY Kit's increased signal detection from loci of higher lengths.



**Figure 11. Effect of amplicon size on allele dropouts for mock-degraded samples.** The total number of ADOs observed across all mock-degraded samples (n=17) plotted against the longest amplicon size per locus for the 21 autosomal loci shared between both the (A) GlobalFiler™ Kit and (B) ForenSeq™ MainstAY Kit.

In order to compare the ForenSeq™ MainstAY Kit with conventional PCR-CE analysis used in casework and DNA databases, the two kits were evaluated at the 20 CODIS

loci. Figure 12 shows the percentages of the 20 CODIS loci that could be reported for the GlobalFiler™ and ForenSeq™ MainstAY Kits. While both kits demonstrated a decrease in percentages of CODIS loci typed as degradation increased, the MainstAY Kit outperformed GlobalFiler™ for all samples heated for 20 minutes or longer (DI > 6.7). For the sample that exhibited the highest level of degradation (DI: 2041.6), the MainstAY Kit provided an 86% complete CODIS profile, compared to the 51% complete profile resulting from GlobalFiler™. Previous studies evaluating serially degraded samples similarly found that MPS returned considerably higher percentages of CODIS loci than traditional CE kits for increasingly degraded DNA (61,65).



**Figure 12. Effect of increasing degradation on percent of CODIS loci typed.** Percent of reportable CODIS loci after increased incubation at 95°C using the GlobalFiler™ and ForenSeq™ MainstAY Kits (n=17 samples). Degradation index is plotted on a log10 scale.

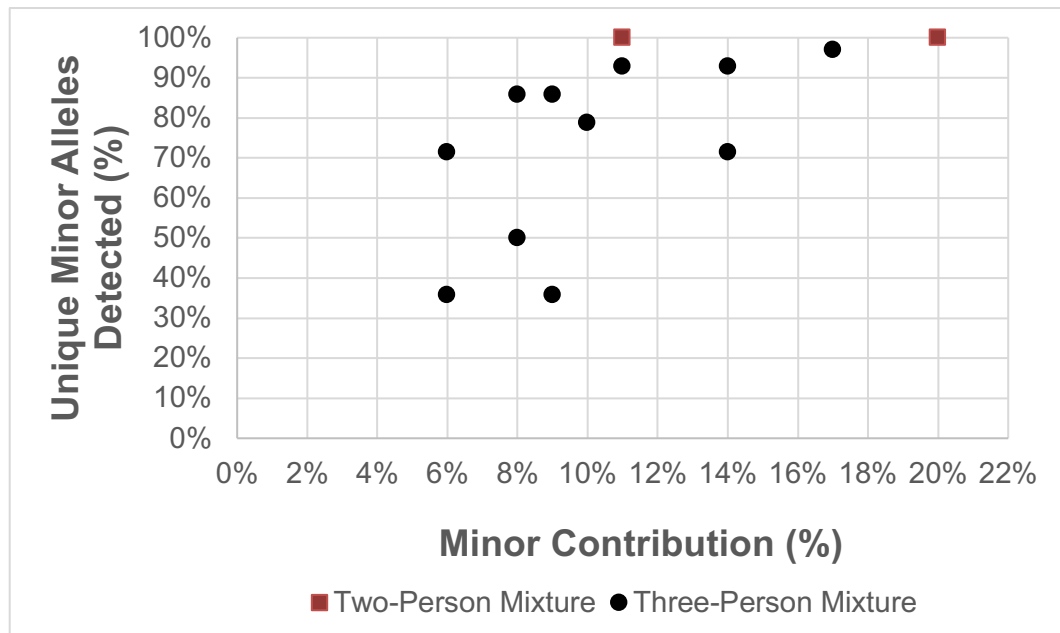
### 3.4 Mixture Study

To evaluate the ability of the ForenSeq™ MainstAY to detect minor alleles of low-level contributors, two- and three-person mixtures with a range of minor contributions were evaluated. The percentage of minor contributors ranged from 11% to 20% for two-person mixtures, and 6% to 17% for three-person mixtures. The number and percent of unique, unshared minor alleles across 22 loci detected for each mixture sequenced are listed in Table 8. The MainstAY Kit detected 100% of unique alleles for both two-person mixtures with low level contributions of 20% and 11%. At 17% contribution from each of two minor contributors in a three-person mixture, 32 of 33 (97%) unique minor alleles were detected (Table 8). Additionally, the kit detected 86% of unique minor alleles for two mixtures with 8% and 9% contribution. The minor contribution percent was plotted against the percent of unique minor allele for the two- (n=2 samples) and three-person (n=11 samples) mixtures in Figure 13. While there was a general increase in percent of unique minor alleles detected as the minor contributor percent increased, the MainstAY Kit demonstrated the ability to detect > 50% of unique minor alleles in all but two mixtures with 6% and 9% minor contributions (Figure 13). Although some minor alleles were not detected at the lowest minor contributions, the kit detected minor allelic information that could be valuable in mixture analysis. It is important to note that these results were observed with mixture DNA inputs of approximately 0.5 ng of DNA due to limited sample availability, which is half of the recommended DNA input for the ForenSeq™ MainstAY protocol. Thus, these findings demonstrate the kit's potential in successfully analyzing low-level sample mixtures often encountered in forensic casework, and may provide more genetic

information at the recommended DNA input. This capability addresses recent challenges of crime laboratories being tasked with evaluating more poor-quality, low-template DNA mixtures (67).

**Table 8. Unique alleles detected for minor contributors of mixture study.** Unique, unshared minor contributor alleles across 22 loci (common to the GlobalFiler™ and ForenSeq™ MainstAY Kits) detected using the ForenSeq™ MainstAY Kit listed in decreasing percentages of minor contributions. Minor contributions for each mixture ratio are bolded.

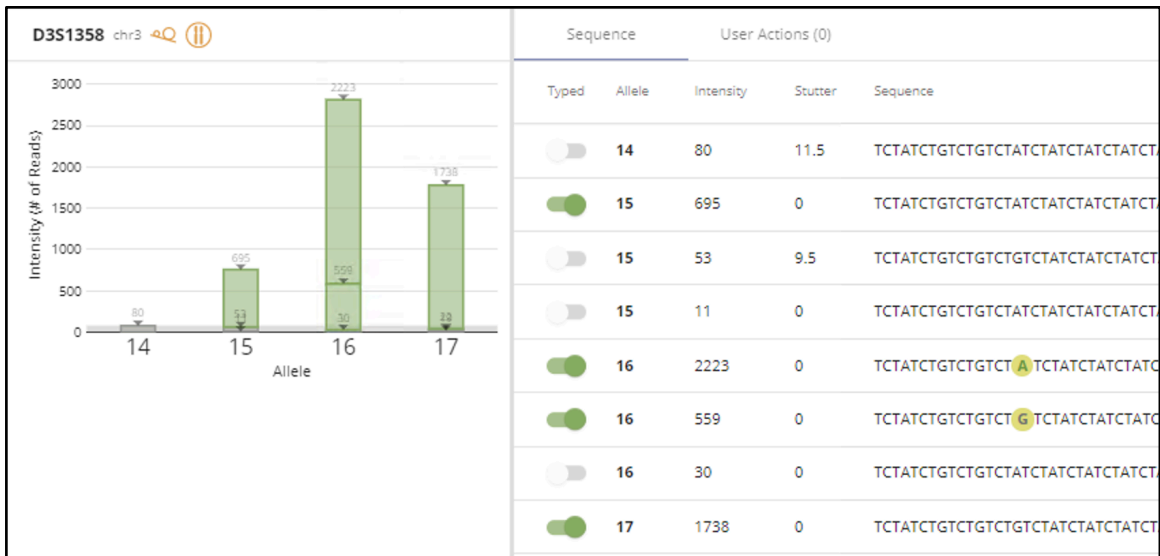
	Sample	Mixture Ratio	Contribution	Minor Contributor		
				# Total Unique Alleles	# Unique Alleles Detected	% Unique Alleles Detected
Two-Person Mixtures	C1	<b>4:1</b>	20%	19	19	100%
	D1	<b>8:1</b>	11%	19	19	100%
Three-Person Mixtures	C3	<b>4:1:1</b>	17%	33	32	97%
	B3	<b>4:2:1</b>	14%	14	10	71%
	C4	<b>4:1:2</b>	14%	14	13	93%
	C5	<b>4:1:4</b>	11%	14	13	93%
	D2	<b>8:1:1</b>	10%	33	26	79%
	C2	<b>8:2:1</b>	9%	14	5	36%
	D3	<b>8:1:2</b>	9%	14	12	86%
	B2	<b>8:4:1</b>	8%	14	7	50%
	D4	<b>8:1:4</b>	8%	14	12	86%
	A2	<b>8:8:1</b>	6%	14	5	36%
D5	<b>8:1:8</b>	6%	14	10	71%	



**Figure 13. Percent of unique minor alleles detected across a range of minor contributions.** Percentages of unique, unshared minor contributor alleles calculated across 22 loci using the ForenSeq™ MainstAY Kit for two- and three-person mixtures.

The integration of the Universal Analysis Software with the MiSeq FGx™ Sequencing System allows for a streamlined method for evaluation of sequencing data. The software’s preconfigured analysis settings provide for a quick and user-friendly way to analyze challenging samples such as mixtures. For example, the software output of the D3S1358 locus of a three-person mixture sample (4:1:2) is shown in Figure 14. The UAS allows for the automated typing of alleles that meet the preconfigured IT, and provides users the ability to manually type alleles that are below the IT but above the AT. This capability can prove to be useful in the detection of minor alleles in mixtures that may have fallen below the IT, with the ability for laboratories to customize ITs and ATs upon validation. The ability to manually customize thresholds for individual loci is not possible

using software commonly used in PCR-CE analysis such as GeneMapper™, where a custom AT must be applied to each dye color in the profile. Instead of a fixed value for the AT applied to dye colors with GeneMapper™, the UAS determines ATs based on a customizable percent of the total number of reads at a locus, allowing for the detection of minor alleles that may have been masked in low level mixtures. A major advantage of NGS technologies over the standard PCR-CE method is the ability to capture sequence variation, which provides a higher level of discrimination over length-based determination (40). The UAS provides the sequences of each allele, and possesses the ability to prove the true heterozygosity of genotypes that would have otherwise been called as homozygous by PCR-CE, or differentiate stutter from true alleles in some cases. For example, the sequence polymorphism present in the 16 allele for the B3 mixture sample with contributor ratios of 4:1:2 was highlighted by the software, revealing a single nucleotide difference in sequence in two allele 16 isoalleles (Figure 14). As allele sharing of contributors leading to allele stacking is a challenge faced in complex mixture analysis, the UAS addresses this challenge by displaying an allele peak denoted with a line to show the signal contribution from each sequence-based allelic variant (Figure 14). Thus, sample sequencing using the ForenSeq™ MainstAY Kit in conjunction with the features available with analysis using the UAS provide enhanced capabilities when analyzing DNA mixture samples.



**Figure 14. ForenSeq™ UAS Output for locus D3S1358 of a three-person mixture.** The left is the software output plot for a three-person mixture (4:1:2) that mimics a conventional STR electropherogram with the allele call (x-axis) plotted against the signal intensity (y-axis). The right shows the sequence output where uncalled alleles that did not meet the interpretation threshold can be manually added (gray). Typed alleles (green) and their respective intensities and sequences are displayed. An isoallele at the D3S1358 locus can be seen for allele 16, as the nucleotide difference is highlighted in yellow by the software.

#### 4. CONCLUSIONS

The purpose of this study was to evaluate the performance of the ForenSeq™ MainstAY Kit on the MiSeq FGx™ Forensic Genomics System for the analysis of mock-degraded and mixed DNA samples. As no literature has been published to date for this novel kit, this study provides data proving the success of the MainstAY Kit in analyzing forensically challenging samples. Performance evaluations such as this are useful for assessing the strengths and limitations of MPS in the analysis of forensic samples as technologies continue to evolve.

The data presented in this study proved accuracy and repeatability of results, with uniformly high read counts. The successful degradation of DNA by incubation at a high temperature provided for the analysis of samples with a wide range of degradation indices (DI: 0.6-2041.6). The sequencing of these samples using this kit demonstrated the overall decrease in concentration and total read count as samples became increasingly degraded. The ForenSeq™ MainstAY Kit outperformed the GlobalFiler™ Kit in the analysis of mock-degraded samples, as intralocus balance, allele signal intensities, and number of dropouts remained more consistent as degradation increased in severity. Moderately and severely degraded samples provided more complete CODIS profiles using the MainstAY Kit than GlobalFiler™, highlighting the advantage of the kit's backward compatibility of allele calling with both national and international forensic databases. The MainstAY Kit demonstrated an increased ability over standard methods to effectively type samples spanning a wide range of degradation, targeting a larger multiplex with smaller amplicon sizes. As crime laboratories have been increasingly receiving poor-quality DNA mixtures,

the MainstAY Kit's ability to detect unique minor alleles in low-level mixtures as low as 6% minor contribution observed prove the kit's value as a forensic analytical tool (67).

The MPS of STR markers offers a sequence-based approach that eliminates artifacts encountered in traditional PCR-CE, multiplexing a greater number of smaller amplicons that allows for improved allele detection, sample conservation, and efficiency. The streamlined workflow and software integration to automate sequencing data evaluation allow the ForenSeq™ MainstAY Kit to be a viable candidate as a routine forensic analytical tool. While the replacement of PCR-CE methods in the near future might not be likely due to the challenges associated with the cost and adoption of MPS into operational workflows, this kit has proven to be a valuable alternative tool. This study offers promising results in the analysis of forensically challenging samples that may be helpful in the assessment and validation studies of the ForenSeq™ MainstAY Kit for criminal laboratories.

## 5. FUTURE CONSIDERATIONS

Rather than the commercial DNA samples utilized in this study, future experimentation may allow for the sequencing of a greater number of forensically challenging samples, including mock case-type samples derived from bones and teeth, for example. A sensitivity study conducted by diluting commercial DNA in replicates would be beneficial in assessing the system's allele detection capabilities at decreased DNA inputs. To compare the power of discrimination of the MainstAY Kit with the standard method, the analysis of a number of samples from unrelated individuals would allow for the comparison of the number of alleles detected by both systems.

The repeatability and reproducibility of the system could be further assessed by sequencing sample replicates using multiple MainstAY Kit lots, along with multiple operators on various MiSeq FGx™ instruments. The inclusion of more incubation time points and temperatures in the mock-degradation study would provide a wider range of DIs that offer more information regarding the range of degradation level successfully analyzed by the ForenSeq™ MainstAY Kit. While each sample analyzed in the mixture study had a DNA input lower than the recommended amount, future studies could include the analysis of mixtures with replicates at the standard 1 ng input to better assess the variation in the system's performance. To test the limits of the analysis of low-level contributor mixtures, samples with minor contributions lower than the 6% lowest of this study could be tested, along with more complex mixtures with four and five contributors.

**APPENDIX:**

**Table A. Loci targeted in each analysis method.** The 24 loci targeted from the CE-based method using the GlobalFiler™ PCR Amplification Kit. The 53 loci targeted from the NGS-based method using the ForenSeq™ MainstAY Kit. The 22 loci common to both kits are highlighted in gray. The 20 CODIS loci are bolded.

GlobalFiler™	ForenSeq™ MainstAY	
	Autosomal STRs	Y-STRs
Y indel	<b>D2S441</b>	DYS391
Amelogenin	<b>D22S1045</b>	DYS392
<b>D2S441</b>	<b>D10S1248</b>	DYS393
<b>D22S1045</b>	<b>D3S1358</b>	DYS437
<b>D10S1248</b>	<b>D19S433</b>	DYS438
<b>D3S1358</b>	<b>D8S1179</b>	DYS439
<b>D19S433</b>	<b>D5S818</b>	DYS448
<b>D8S1179</b>	<b>D1S1656</b>	DYS460
<b>D5S818</b>	<b>VWA</b>	DYS481
<b>D1S1656</b>	<b>TH01</b>	DYS505
<b>VWA</b>	<b>D13S317</b>	DYS522
<b>TH01</b>	<b>D21S11</b>	DYS533
<b>D13S317</b>	<b>D12S391</b>	DYS549
<b>D21S11</b>	<b>D16S539</b>	DYS570
<b>D12S391</b>	<b>D7S820</b>	DYS576
<b>D16S539</b>	<b>CSF1PO</b>	DYS612
<b>D7S820</b>	<b>D18S51</b>	DYS635
<b>CSF1PO</b>	<b>D2S1338</b>	DYS643
<b>D18S51</b>	<b>FGA</b>	DYS390
<b>D2S1338</b>	<b>TPOX</b>	DYS389II
<b>FGA</b>	D4S2408	DYS385a-b
<b>TPOX</b>	D6S1043	DYF387S1
DYS391	D9S1122	DYS19
SE33	PentaE	DYS389I
	PentaD	Y-GATA-H4
	D17S1301	
	D20S482	
	Amelogenin	

## LIST OF JOURNAL ABBREVIATIONS

Annu Rev Genom Hum Genet	Annual Review of Genomics and Human Genetics
BMC Genet	BMC Genetics
Cold Spring Harb Symp Quant Biol	Cold Spring Harbor Symposia on Quantitative Biology
Croat Med J	Croatian Medical Journal
Emerg Top Life Sci	Emerging Topics in Life Sciences
Forensic Sci Int Genet	Forensic Science International. Genetics
Forensic Sci Med Pathol	Forensic Science, Medicine, and Pathology
Genome Res	Genome Research
Int J Leg Med	International Journal of Legal Medicine
Investig Genet	Investigative Genetics
J Biol Chem	Journal of Biological Chemistry
J Forensic Sci	Journal of Forensic Sciences
J Invest Dermatol	The Journal of Investigative Dermatology
Malays J Med Sci	The Malaysian Journal of Medical Sciences
Mol Hum Reprod	Molecular Human Reproduction
Nat Genet	Nature Genetics
Proc Natl Acad Sci USA	Proceedings of the National Academy of Sciences of the United States of America
Sci Rep	Scientific Reports

## BIBLIOGRAPHY

1. Amankwaa AO, McCartney C. The effectiveness of the current use of forensic DNA in criminal investigations in England and Wales. *WIREs Forensic Science* 2021;3(6):e1414. <https://doi.org/10.1002/wfs2.1414>.
2. National Research Council. *The Evaluation of Forensic DNA Evidence*. Washington, DC: The National Academies Press, 1996.
3. Roewer L. DNA fingerprinting in forensics: past, present, future. *Investig Genet* 2013;4(1):22. <https://doi.org/10.1186/2041-2223-4-22>.
4. Jeffreys AJ, Wilson V, Thein SL. Hypervariable ‘minisatellite’ regions in human DNA. *Nature* 1985;314(6006):67–73. <https://doi.org/10.1038/314067a0>.
5. Jeffreys AJ, Wilson V, Thein SL. Individual-specific “fingerprints” of human DNA. *Nature* 1985;316(6023):76–9. <https://doi.org/10.1038/316076a0>.
6. Jeffreys AJ. The man behind the DNA fingerprints: an interview with Professor Sir Alec Jeffreys. *Investigative Genetics* 2013;4(1):21. <https://doi.org/10.1186/2041-2223-4-21>.
7. Gill P, Jeffreys AJ, Werrett DJ. Forensic application of DNA “fingerprints.” *Nature* 1985;318(6046):577–9. <https://doi.org/10.1038/318577a0>.
8. Jeffreys AJ, Brookfield JF, Semeonoff R. Positive identification of an immigration test-case using human DNA fingerprints. *Nature* 1985;317(6040):818–9. <https://doi.org/10.1038/317818a0>.
9. Wyman AR, White R. A highly polymorphic locus in human DNA. *Proc Natl Acad Sci U S A* 1980;77(11):6754–8. <https://doi.org/10.1073/pnas.77.11.6754>.
10. Mullis K, Faloona F, Scharf S, Saiki R, Horn G, Erlich H. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb Symp Quant Biol* 1986;51 Pt 1:263–73. <https://doi.org/10.1101/sqb.1986.051.01.032>.
11. Garibyan L, Avashia N. Research Techniques Made Simple: Polymerase Chain Reaction (PCR). *J Invest Dermatol* 2013;133(3):e6. <https://doi.org/10.1038/jid.2013.1>.
12. Buckleton JS, Bright J-A, Taylor D. *Forensic DNA Evidence Interpretation*. CRC Press, 2018.

13. Kayser M. Forensic DNA Phenotyping: Predicting human appearance from crime scene material for investigative purposes. *Forensic Science International: Genetics* 2015;18:33–48. <https://doi.org/10.1016/j.fsigen.2015.02.003>.
14. Jordan D, Mills D. Past, Present, and Future of DNA Typing for Analyzing Human and Non-Human Forensic Samples. *Frontiers in Ecology and Evolution* 2021;9:172. <https://doi.org/10.3389/fevo.2021.646130>.
15. Butler JM. *Advanced topics in forensic DNA typing: methodology*. Waltham, MA: Elsevier/Academic Press, 2012.
16. Edwards A, Hammond HA, Jin L, Caskey CT, Chakraborty R. Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. *Genomics* 1992;12(2):241–53. [https://doi.org/10.1016/0888-7543\(92\)90371-X](https://doi.org/10.1016/0888-7543(92)90371-X).
17. Edwards A, Civitello A, Hammond HA, Caskey CT. DNA Typing and Genetic Mapping with Trimeric and Tetrameric Tandem Repeats. :11.
18. Hares DR. Selection and implementation of expanded CODIS core loci in the United States. *Forensic Science International: Genetics* 2015;17:33–4. <https://doi.org/10.1016/j.fsigen.2015.03.006>.
19. Gill P, Urquhart A, Millican E, Oldroyd N, Watson S, Sparkes R, et al. A new method of STR interpretation using inferential logic -development of a criminal intelligence database. *Int J Leg Med* 1996;109(1):14–22. <https://doi.org/10.1007/BF01369596>.
20. Nwawuba Stanley U, Mohammed Khadija A, Bukola AT, Omusi Precious I, Ayevbomwan Davidson E. Forensic DNA Profiling: Autosomal Short Tandem Repeat as a Prominent Marker in Crime Investigation. *Malays J Med Sci* 2020;27(4):22–35. <https://doi.org/10.21315/mjms2020.27.4.3>.
21. Butler JM, Buel E, Crivellente F, McCord BR. Forensic DNA typing by capillary electrophoresis using the ABI Prism 310 and 3100 genetic analyzers for STR analysis. *Electrophoresis* 2004;25(1011):1397–412. <https://doi.org/10.1002/elps.200305822>.
22. Approved S. *SWGDM Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories*. undefined 2010.
23. Gopinath S, Zhong C, Nguyen V, Ge J, Lagacé RE, Short ML, et al. Developmental validation of the Yfiler® Plus PCR Amplification Kit: An enhanced Y-STR multiplex for casework and database applications. *Forensic Science International: Genetics* 2016;24:164–75. <https://doi.org/10.1016/j.fsigen.2016.07.006>.

24. Amorim A, Fernandes T, Taveira N. Mitochondrial DNA in human identification: a review. *PeerJ* 2019;7:e7314. <https://doi.org/10.7717/peerj.7314>.
25. Holland MM, Mitchell LG, Rodriguez WC, Canik JJ, Merrill CR, Weedn VW. Mitochondrial DNA sequence analysis of human skeletal remains: identification of remains from the Vietnam War. *J Forensic Sci* 1993;38(3):542–53.
26. Budowle B, Allard MW, Wilson MR, Chakraborty R. Forensics and Mitochondrial DNA: Applications, Debates, and Foundations. *Annu Rev Genom Hum Genet* 2003;4(1):119–41. <https://doi.org/10.1146/annurev.genom.4.070802.110352>.
27. Syndercombe Court D. Mitochondrial DNA in forensic use. *Emerging Topics in Life Sciences* 2021;5(3):415–26. <https://doi.org/10.1042/ETLS20210204>.
28. Bogenhagen D, Clayton DA. The number of mitochondrial deoxyribonucleic acid genomes in mouse L and human HeLa cells. Quantitative isolation of mitochondrial deoxyribonucleic acid. *J Biol Chem* 1974;249(24):7991–5.
29. Ginther C, Issel-Tarver L, King MC. Identifying individuals by sequencing mitochondrial DNA from teeth. *Nat Genet* 1992;2(2):135–8. <https://doi.org/10.1038/ng1092-135>.
30. Wilson MR, DiZinno JA, Polansky D, Replogle J, Budowle B. Validation of mitochondrial DNA sequencing for forensic casework analysis. *Int J Leg Med* 1995;108(2):68–74. <https://doi.org/10.1007/BF01369907>.
31. Foran DR. Relative degradation of nuclear and mitochondrial DNA: an experimental approach. *J Forensic Sci* 2006;51(4):766–70. <https://doi.org/10.1111/j.1556-4029.2006.00176.x>.
32. Chaitanya L, Oven M van, Brauer S, Zimmermann B, Huber G, Xavier C, et al. High-quality mtDNA control region sequences from 680 individuals sampled across the Netherlands to establish a national forensic mtDNA reference database. *Forensic Science International: Genetics* 2016;21:158–67. <https://doi.org/10.1016/j.fsigen.2015.12.002>.
33. Kim H, Erlich HA, Calloway CD. Analysis of mixtures using next generation sequencing of mitochondrial DNA hypervariable regions. *Croat Med J* 2015;56(3):208–17. <https://doi.org/10.3325/cmj.2015.56.208>.
34. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* 1977;74(12):5463–7.

35. Fullwood MJ, Wei C-L, Liu ET, Ruan Y. Next-generation DNA sequencing of paired-end tags (PET) for transcriptome and genome analyses. *Genome Res* 2009;19(4):521–32. <https://doi.org/10.1101/gr.074906.107>.
36. Yang Y, Xie B, Yan J. Application of Next-generation Sequencing Technology in Forensic Science. *Genomics, Proteomics & Bioinformatics* 2014;12(5):190–7. <https://doi.org/10.1016/j.gpb.2014.09.001>.
37. Børsting C, Morling N. Next generation sequencing and its applications in forensic genetics. *Forensic Science International: Genetics* 2015;18:78–89. <https://doi.org/10.1016/j.fsigen.2015.02.002>.
38. Heather JM, Chain B. The sequence of sequencers: The history of sequencing DNA. *Genomics* 2016;107(1):1–8. <https://doi.org/10.1016/j.ygeno.2015.11.003>.
39. Gettings KB, Kiesler KM, Faith SA, Montano E, Baker CH, Young BA, et al. Sequence variation of 22 autosomal STR loci detected by next generation sequencing. *Forensic Sci Int Genet* 2016;21:15–21. <https://doi.org/10.1016/j.fsigen.2015.11.005>.
40. Ballard D, Winkler-Galicki J, Wesoly J. Massive parallel sequencing in forensics: advantages, issues, technicalities, and prospects. *Int J Legal Med* 2020;134(4):1291–303. <https://doi.org/10.1007/s00414-020-02294-0>.
41. Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, et al. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 2008;456(7218):53–9. <https://doi.org/10.1038/nature07517>.
42. Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, et al. An integrated semiconductor device enabling non-optical genome sequencing. *Nature* 2011;475(7356):348–52. <https://doi.org/10.1038/nature10242>.
43. Bruijns B, Tiggelaar R, Gardeniers H. Massively parallel sequencing techniques for forensics: A review. *Electrophoresis* 2018;39(21):2642–54. <https://doi.org/10.1002/elps.201800082>.
44. Dalsgaard S, Rockenbauer E, Buchard A, Mogensen HS, Frank-Hansen R, Børsting C, et al. Non-uniform phenotyping of D12S391 resolved by second generation sequencing. *Forensic Sci Int Genet* 2014;8(1):195–9. <https://doi.org/10.1016/j.fsigen.2013.09.008>.
45. Gelardi C, Rockenbauer E, Dalsgaard S, Børsting C, Morling N. Second generation sequencing of three STRs D3S1358, D12S391 and D21S11 in Danes and a new nomenclature for sequenced STR alleles. *Forensic Sci Int Genet* 2014;12:38–41. <https://doi.org/10.1016/j.fsigen.2014.04.016>.

46. Scheible M, Loreille O, Just R, Irwin J. Short tandem repeat typing on the 454 platform: strategies and considerations for targeted sequencing of common forensic markers. *Forensic Sci Int Genet* 2014;12:107–19. <https://doi.org/10.1016/j.fsigen.2014.04.010>.
47. Haddrill PR. Developments in forensic DNA analysis. *Emerg Top Life Sci* 2021;5(3):381–93. <https://doi.org/10.1042/ETLS20200304>.
48. Budowle B, Schmedes SE, Wendt FR. Increasing the reach of forensic genetics with massively parallel sequencing. *Forensic Sci Med Pathol* 2017;13(3):342–9. <https://doi.org/10.1007/s12024-017-9882-5>.
49. Alvarez-Cubero MJ, Saiz M, Martínez-García B, Sayalero SM, Entrala C, Lorente JA, et al. Next generation sequencing: an application in forensic sciences? *Annals of Human Biology* 2017;44(7):581–92. <https://doi.org/10.1080/03014460.2017.1375155>.
50. European survey on forensic applications of massively parallel sequencing - ClinicalKey. <https://www-clinicalkey-com.ezproxy.bu.edu/#!/content/playContent/1-s2.0-S1872497317301011?returnurl=null&referrer=null> (accessed December 27, 2021).
51. Fordyce SL, Mogensen HS, Børsting C, Lagacé RE, Chang C-W, Rajagopalan N, et al. Second-generation sequencing of forensic STRs using the Ion Torrent™ HID STR 10-plex and the Ion PGM™. *Forensic Science International: Genetics* 2015;14:132–40. <https://doi.org/10.1016/j.fsigen.2014.09.020>.
52. Weber-Lehmann J, Schilling E, Gradl G, Richter DC, Wiehler J, Rolf B. Finding the needle in the haystack: Differentiating “identical” twins in paternity testing and forensics by ultra-deep next generation sequencing. *Forensic Science International: Genetics* 2014;9:42–6. <https://doi.org/10.1016/j.fsigen.2013.10.015>.
53. Bender K, Farfán MJ, Schneider PM. Preparation of degraded human DNA under controlled conditions. *Forensic Science International* 2004;139(2):135–40. <https://doi.org/10.1016/j.forsciint.2003.10.003>.
54. Sharma V, Chow HY, Siegel D, Wurmbach E. Qualitative and quantitative assessment of Illumina’s forensic STR and SNP kits on MiSeq FGx™. *PLoS One* 2017;12(11):e0187932. <https://doi.org/10.1371/journal.pone.0187932>.
55. Gill P, Sparkes R, Pinchin R, Clayton T, Whitaker J, J. Buckleton. Interpreting simple STR mixtures using allele peak areas. *Forensic Science International* 1998;91(1):41–53. [https://doi.org/10.1016/S0379-0738\(97\)00174-6](https://doi.org/10.1016/S0379-0738(97)00174-6).

56. Børsting C, Morling N. Chapter 22 - Genomic Applications in Forensic Medicine. In: Kumar D, Antonarakis S, editors. *Medical and Health Genomics*. Oxford: Academic Press, 2016;295–309.
57. Phillips C, Fang R, Ballard D, Fondevila M, Harrison C, Hyland F, et al. Evaluation of the Genplex SNP typing system and a 49plex forensic marker panel. *Forensic Science International: Genetics* 2007;1(2):180–5. <https://doi.org/10.1016/j.fsigen.2007.02.007>.
58. Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, et al. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics* 2012;13(1):341. <https://doi.org/10.1186/1471-2164-13-341>.
59. Verogen. 2021e. *Universal Analysis Software v2.0 Reference Guide*.
60. Jäger AC, Alvarez ML, Davis CP, Guzmán E, Han Y, Way L, et al. Developmental validation of the MiSeq FGx Forensic Genomics System for Targeted Next Generation Sequencing in Forensic DNA Casework and Database Laboratories. *Forensic Science International: Genetics* 2017;28:52–70. <https://doi.org/10.1016/j.fsigen.2017.01.011>.
61. Sharma V, van der Plaats DA, Liu Y, Wurmbach E. Analyzing degraded DNA and challenging samples using the ForenSeq™ DNA Signature Prep kit. *Science & Justice* 2020;60(3):243–52. <https://doi.org/10.1016/j.scijus.2019.11.004>.
62. Verogen. 2021a. *ForenSeq™ MainstAY Reference Guide*.
63. Verogen. 2021. *Establishing robust thresholds and filters for the ForenSeq MainstAY Kit*.
64. Verogen. 2021. *ForenSeq™ MainstAY*.
65. Zhang Q, Zhou Z, Liu Q, Liu L, Shao L, Zhang M, et al. Evaluation of the performance of Illumina's ForenSeq™ system on serially degraded samples. *ELECTROPHORESIS* 2018;39(21):2674–84. <https://doi.org/10.1002/elps.201800101>.
66. Fattorini P, Previderé C, Carboni I, Marrubini G, Sorçaburu-Cigliero S, Grignani P, et al. Performance of the ForenSeq™ DNA Signature Prep kit on highly degraded samples. *ELECTROPHORESIS* 2017;38(8):1163–74. <https://doi.org/10.1002/elps.201600290>.
67. Bieber FR, Buckleton JS, Budowle B, Butler JM, Coble MD. Evaluation of forensic DNA mixture evidence: protocol for evaluation, interpretation, and statistical

- calculations using the combined probability of inclusion. *BMC Genet* 2016;17(1):125. <https://doi.org/10.1186/s12863-016-0429-7>.
68. Chung DT, Drábek J, Opel KL, Butler JM, McCord BR. A Study on the Effects of Degradation and Template Concentration on the Amplification Efficiency of the STR Miniplex Primer Sets. *J Forensic Sci* 2004;49(4):1–8. <https://doi.org/10.1520/JFS2003269>.
  69. Cotton EA, Allsop RF, Guest JL, Frazier RRE, Koumi P, Callow IP, et al. Validation of the AMPFISTR® SGM Plus™ system for use in forensic casework. *Forensic Science International* 2000;112(2):151–61. [https://doi.org/10.1016/S0379-0738\(00\)00182-1](https://doi.org/10.1016/S0379-0738(00)00182-1).
  70. Krenke BE, Tereba A, Anderson SJ, Buel E, Culhane S, Finis CJ, et al. Validation of a 16-locus fluorescent multiplex system. *J Forensic Sci* 2002;47(4):773–85.
  71. Sipes DE, Harris JA, Bethune MT, Cotton RW. Validation of the AMPFISTR® Identifiler™ PCR Amplification Kit using the ABI Prism® 310 Genetic Analyzer. :1.
  72. Piyamongkol W, Bermúdez MG, Harper JC, Wells D. Detailed investigation of factors influencing amplification efficiency and allele drop-out in single cell PCR: implications for preimplantation genetic diagnosis. *Mol Hum Reprod* 2003;9(7):411–20. <https://doi.org/10.1093/molehr/gag051>.
  73. Kwon Y-L, Kim BM, Lee EY, Shin K-J. Massively parallel sequencing of 25 autosomal STRs including SE33 in four population groups for forensic applications. *Sci Rep* 2021;11(1):4701. <https://doi.org/10.1038/s41598-021-82814-z>.
  74. Butler JM, Shen Y, McCord BR. The development of reduced size STR amplicons as tools for analysis of degraded DNA. *J Forensic Sci* 2003;48(5):1054–64.

## CURRICULUM VITAE

