

2019

## Screening Sexual Assault Evidence with Low Concentrations of Male DNA Utilizing the RapidHIT 200 and ParaDNA Intelligence Test

Taylor L. Koepfler  
West Virginia University, [tlkoepfler@mix.wvu.edu](mailto:tlkoepfler@mix.wvu.edu)

Follow this and additional works at: <https://researchrepository.wvu.edu/etd>



Part of the [Forensic Science and Technology Commons](#)

---

### Recommended Citation

Koepfler, Taylor L., "Screening Sexual Assault Evidence with Low Concentrations of Male DNA Utilizing the RapidHIT 200 and ParaDNA Intelligence Test" (2019). *Graduate Theses, Dissertations, and Problem Reports*. 7467.

<https://researchrepository.wvu.edu/etd/7467>

This Thesis is protected by copyright and/or related rights. It has been brought to you by the The Research Repository @ WVU with permission from the rights-holder(s). You are free to use this Thesis in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you must obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/ or on the work itself. This Thesis has been accepted for inclusion in WVU Graduate Theses, Dissertations, and Problem Reports collection by an authorized administrator of The Research Repository @ WVU. For more information, please contact [researchrepository@mail.wvu.edu](mailto:researchrepository@mail.wvu.edu).

**Screening Sexual Assault Evidence with Low Concentrations of  
Male DNA Utilizing the RapidHIT<sup>®</sup> 200 and ParaDNA<sup>®</sup> Intelligence  
Test**

**Taylor L. Koepfler**

**Thesis submitted  
to the Eberly College of Arts and Sciences  
at West Virginia University**

**in partial fulfillment of the requirements for the degree of**

**Master of Science in  
Forensic and Investigative Science**

**Jeremy Dawson, Ph.D., Chair  
Tina Moroosse, M.S.  
Casey Jelsema, Ph.D.**

**Department of Forensic and Investigative Science**

**Morgantown, West Virginia  
2019**

**Keywords: ParaDNA<sup>®</sup>, Intelligence Test, SAK, Screening, RapidHIT<sup>®</sup>  
Copyright 2019 Taylor L. Koepfler**

## ABSTRACT

### **Screening Sexual Assault Evidence with Low Concentrations of Male DNA Utilizing the RapidHIT<sup>®</sup> 200 and ParaDNA<sup>®</sup> Intelligence Test**

Taylor L. Koepfler

Over the last several years, crime laboratories have largely focused on the sexual assault kit (SAK) backlog, where they are often confronted with many low-quality samples. The lack of available screening techniques has prevented analysts from gaining insight into the disposition of the sample earlier on in the testing process; requiring analysts to rely on visual observations and little background information. Consequently, resulting in the hindrance of probative STR profiles while expending a large amount of time and resources to gain this result. In recent years, crime laboratories have explored a male screening technique, recommended by SWGDAM, to combat this problem. However, it is important to investigate alternative screening methods that can be performed prior to receiving the evidence at a crime laboratory, allocating time and resources toward enhanced testing of probative samples.

Mock sexual assault admixtures were prepared at different mixture ratios and split into two data sets. The first data set was comprised of 24 total samples that were prepared at the 1:10 and 1:20 mixture ratio, then run on the RapidHIT<sup>®</sup> 200. The second data set consisted of 91 total samples that were prepared at the following mixture ratios: 1:5, 1:10, 1:15, 1:20, 1:25. These samples were then run on the ParaDNA<sup>®</sup> System with the Intelligence Test assay.

The results of this study showed that the RapidHIT<sup>®</sup> 200 outperformed the ParaDNA<sup>®</sup> Intelligence Test when utilized to screen sexual assault admixtures in a non-laboratory setting. The RapidHIT<sup>®</sup> 200 was successful at detecting a mixture profile in 75% of the mock sexual assault cases. Whereas, the ParaDNA<sup>®</sup> Intelligence Test had less than one in four chance of detecting the presence of a mixture. In other words, the linear mixed-effects model demonstrated that the instrument used for screening sexual assault samples in a non-laboratory setting does have a significant effect on the proportion of loci that exhibit a mixture. In conclusion, the ParaDNA<sup>®</sup> Intelligence Test is not recommended for use in a non-laboratory setting. Although the RapidHIT<sup>®</sup> 200 outperformed this test, further research should be conducted before use in any setting.

## Acknowledgments

First, I would like to recognize and sincerely thank my research advisor, Tina Moroose, for her everlasting patience and guidance over the last four years. Without her encouragement and positivity, I would have never embarked on this journey and challenged myself to achieve another degree, despite the odds not being in my favor. Thank you for always believing in me and never giving up!

I would also like to thank my committee chair, Dr. Jeremy Dawson, for supporting another interdisciplinary partnership with the Forensic and Investigative Science Department. Thank you for believing in my ability to perform this research effort and for helping me develop skillsets in grant writing.

Next, I would like to express my overwhelming gratitude to Dr. Casey Jelsema, for always going above and beyond to help me succeed. I am truly thankful for your ability to educate others on the complex theories behind statistical analyses and for being able to explain the proper usage of statistical models to laypersons. Without you, I would never have been able to complete this masters' thesis!

Finally, I would like to thank Jacob Nead and Matt Green, two undergraduate students who have spent countless hours by my side in the DNA laboratory implementing the experimental design of this study. I truly could not have completed this thesis without your help and support!

# Table of Contents

|  |            |
|--|------------|
| <b>Abstract.....</b>                                   | <b>ii</b>  |
| <b>Acknowledgments.....</b>                            | <b>iii</b> |
| <b>List of Figures.....</b>                            | <b>v</b>   |
| <b>List of Tables.....</b>                             | <b>vi</b>  |
| <b>Introduction.....</b>                               | <b>1</b>   |
| <b>Chapter 1: Background</b>                           |            |
| Chapter 1.1 STR Analysis.....                          | 2          |
| Chapter 1.2 Two-Person Mixture Interpretation.....     | 5          |
| Chapter 1.3 ParaDNA <sup>®</sup> System.....           | 8          |
| Chapter 1.4 RapidHIT <sup>®</sup> 200.....             | 10         |
| <b>Chapter 2: Materials and Methodology</b>            |            |
| Chapter 2.1 Sample Collection.....                     | 11         |
| Chapter 2.2 Sample Preparation.....                    | 11         |
| Chapter 2.3 Sample Processing and Testing.....         | 12         |
| <b>Chapter 3: Results and Discussion</b>               |            |
| Chapter 3.1 Qualitative Approach.....                  | 14         |
| Chapter 3.2 Linear Mixed-Effects Model.....            | 16         |
| Chapter 3.3 Logistic Regression Model.....             | 19         |
| <b>Chapter 4: Conclusions and Recommendations.....</b> | <b>23</b>  |
| <b>Chapter 5: Literature Cited.....</b>                | <b>24</b>  |

## List of Figures

- Figure 1.** *Top:* ParaDNA<sup>®</sup> profile demonstrating allele sharing. At least three loci (D16, D8, THO) contain alleles (12, 13, 6) shared by both the male and female participants. *Bottom:* ParaDNA<sup>®</sup> profile demonstrating unequal sampling. Unequal sampling is apparent at the THO and Amelogenin loci because the florescent probes located in well D of the reaction plate that is used to target alleles at those locations did not identify alleles, whereas other loci did.....14
- Figure 2.** A schematic of the ParaDNA<sup>®</sup> Intelligence Test reaction plate containing the source of DNA, the allele call, and the chromosome location of each well for two admixtures (6343:2964, 1396:7721) .....15
- Figure 3.** A boxplot representing the proportion of missing observations across five mixture ratios, testing the effect of the mixture ratio.....17
- Figure 4.** A boxplot representing the proportion of missing observations at different mixture ratios, testing the effect of the instrument.....18
- Figure 5.** *Top:* A bar graph representing the probability of obtaining a single-source profile versus a mixture when using the ParaDNA<sup>®</sup> Intelligence Test in a non-laboratory setting and *Bottom:* in a laboratory setting.....20
- Figure 6.** A bar graph representing the probability of obtaining a single-source profile versus a mixture when using the RapidHIT<sup>®</sup> 200 in either a non-laboratory or a laboratory setting.....21

## List of Tables

|  |           |
|--|-----------|
| <b>Table 1.</b> Summary of the Intelligence Test assay.....  | <b>8</b>  |
| <b>Table 2.</b> Experimental data sets utilized for sample preparation.....  | <b>12</b> |
| <b>Table 3.</b> Allele charts for single-source DNA profiles demonstrating shared alleles (bold and purple)<br>..... | <b>15</b> |
| <b>Table 4.</b> Summary statistics for the fixed effects variables of the logistic regression model.....             | <b>19</b> |

## Introduction

When a sexual assault occurs, some victims choose to seek justice by reporting the incident, where he or she must further endure an invasive and time-consuming examination to collect any residual physical evidence [1]. Sexual assault kits (SAKs), which are compiled during this examination, commonly include clothing, hair, and swabs used to sample regions of the victim's body that may contain bodily fluid deposited from the suspect [1-4]. However, the quality of each sample collected drastically declines as the post-coital interval continues to increase [5]. As SAKs stored in evidence locker rooms are exposed to various environmental conditions, such as heat and humidity for great lengths of time, the difficulty with sample quality proceeds to grow. Low-quality samples then result in the hindrance of generating a probative STR (short tandem repeat) profile [6], jeopardizing the prosecution of the incident.

Furthermore, the prosecution of sexual assaults has been halted due to the immense SAK backlog that has been accumulating over the last three decades, containing approximately 500,000 untested kits [7-8]. Upon the national recognition of the backlog and the development of DNA (deoxyribonucleic acid) technology, this issue has become two-fold. First, law enforcement agencies have begun to hastily submit untested SAKs to forensic laboratories for DNA analysis throughout the nation causing a bottleneck to occur. The rapid influx in cases being received has resulted in a secondary backlog due to the lack of available resources needed to maintain caseloads of this magnitude [7]. Second, the accumulating backlog cannot be reduced with the utilization of the current serological techniques used by laboratories due to the lack of sensitivity required to screen samples of reduced quality [9]. Therefore, pre-screening samples to measure the suitability for STR analysis has become time and cost-ineffective owing to the high potential for a false negative. For this reason, there is a heightened demand for enhanced screening techniques to rapidly and efficiently process evidence.

To address both concerns, many government agencies have created "Sexual Assault Initiatives" that provides forensic laboratories with more funding in an effort to reduce the SAK backlog by investing in enhanced screening techniques. With funding from these initiatives, laboratories throughout the United States have begun to implement a novel male screening technique, recommended by SWGDAM (Scientific Working Group of DNA Analysis Methods) to replace current serological methods [9]. The male screening technique enables a small portion (~1/8 of a swab or stain) of the evidence to be tested and analyzed through DNA quantitation, which identifies the concentration of male DNA in a sample. The ability to detect low-quantity male DNA within a sample is significant as 90% of sexual assault perpetrators are men [10]. Furthermore, the quality of each sample is also evaluated using the degradation index (DI) that is obtained through the DNA quantitation method utilized in this process. This information combined enables analysts to limit the number of samples further examined by the STR analysis process. With the capability of limiting samples, analysts are reporting negative cases (samples with quantities below the validated detection limit) quicker, allowing more time and resources to be spent on positive cases.

On the other hand, rapid DNA instrumentation such as ParaDNA<sup>®</sup> and RapidHIT<sup>®</sup> 200 can improve upon the male screening technique by enabling SANEs (sexual assault nurse examiners) to pre-screen sexual assault samples prior to the submission to a forensic laboratory. These novel



screening techniques can provide laboratory analysts with information regarding the suitability of each sample for male screening without any sample preparation. The samples that do not exhibit sufficient quality for male screening can then be eliminated from the workflow, thus limiting the number of samples tested, saving both time and resources for laboratories. The goal of this research effort was to evaluate the use of the ParaDNA<sup>®</sup> Intelligence Test and the RapidHIT<sup>®</sup> 200 to screen sexual assault evidence with low-quantity male DNA. The results from this work could directly impact the processing workflow of SAKs by offering insight into the suitability of the ParaDNA<sup>®</sup> and RapidHIT<sup>®</sup> 200 for the screening of evidence at the time of collection. These pre-screening techniques can help to minimize indiscriminate sample submission and ultimately reduce the amount of time and money spent on processing unsuitable SAK swabs and allowing more timely processing of useful evidence.

## Chapter 1: Background

### 1.1 STR Analysis

Early pioneers of forensic DNA typing discovered a microsatellite residing in the non-coding region of the human genome that is currently being used as the foundation for identifying individuals within a population for the prosecution of a crime. This microsatellite is referred to as a short tandem repeat (STR). STRs are highly polymorphic segments of DNA comprised of one to thirteen nucleotides that have been found to consecutively repeat up to 48 times [11-12]. Additionally, a large degree of variability exists amongst STRs due to the enhanced mutation rate in the non-coding region of the genome where discriminate genotypes are introduced, thereby advancing DNA typing [13-14]. To generate an STR profile for the identification of an individual, evidence consisting of DNA must undergo the following process: extraction, quantitation, normalization, amplification, and capillary electrophoresis.

The DNA extraction process is utilized to physically remove DNA from the nucleus of a cell by chemical and thermal treatments while isolating and purifying the free DNA from all other cellular material [15]. Depending upon the substrate, size, and type of cell being tested there are different extraction protocols available to effectively execute this process under various circumstances. For instance, sexual assault evidence consists of swabs that are typically comprised of DNA from multiple contributors where more than one cell type may be present (i.e. sperm cells and non-sperm cells). Therefore, a differential extraction must be utilized to separate and individually lyse the different cell types by altering the chemical and thermal treatments, in hopes to separate the contributors. First, the non-sperm cells are lysed by applying a detergent that disrupts the continuity of the plasma membrane, enabling the exposure of the nucleus and permitting the removal of DNA [16, 17]. The sperm cells are then suspended at the bottom of the sample creating a physical separation from the newly released DNA. However, the physical separation of cellular material may only occur if the pelleted sperms cells are not disturbed during this process. Thereafter, a reducing agent such as Dithiothreitol (DTT), is applied to the spermatozoa causing the disulfide bonds that encompass the acrosome to break, enabling the removal of DNA [16, 18]. At the completion of this extraction procedure, the DNA retrieved from both types of cells will progress through the STR analysis process independently, as “E” (epithelial) and “S” (sperm) fractions.

Next, the concentration of purified DNA is quantified through a process known as real-time polymerase chain reaction (RT-PCR), to approximate the amount of DNA present in a sample [19-20]. To prevent the quantification of competing DNA from bacteria and viruses commonly encountered on sexual assault evidence, human-specific assays were developed [16, 21]. The development of human assays enabled the amplification of three specific targets by hybridizing a dye-labeled TaqMan<sup>®</sup> probe to the complementary sequences. Taq (*Thermus aquaticus*) DNA polymerase then replicates the targeted strand by cleaving the hybridized TaqMan<sup>®</sup> probe causing the reporter dye to detach and separate from the non-fluorescent quencher (NFQ) [22]. This separation along with the directed light from a tungsten-halogen lamp enables a fluorescent signal to radiate from the reporter dye. The fluorescent emission produced is then detected by the CCD (charged-couple device) camera and extrapolated by the HID Real-Time PCR Analysis Software [22].

The software then generates an amplification plot of the extrapolated fluorescent signal ( $R_n$ ) detected during each PCR cycle (E.g.  $c, n, m$ ), illustrating the progression of the process. The user can further evaluate this illustration by assessing the four primary stages: Lag, Exponential, Linear, and Plateau to ensure proper amplification has occurred [23]. The exponential phase, however, is the most significant out of the four stages because PCR amplification should achieve optimal efficiency ( $E$  and  $E_x$ ) and produce a copious amount of amplified product. Therefore, enabling the concentration of amplified product ( $N_c$ ) to be used to calculate the initial concentration of template DNA ( $N$ ), using equation one [22, 23]. The threshold cycle ( $C_T$ ) can then be computed (equation 2) to create a baseline that when reached demonstrates an increase in signal intensity of a target molecule at a given cycle ( $X_n$  and  $X_m$ ) above the software's internal baseline [22, 23].

$$N_c = N(1 + E)^c \quad (1)$$

$$X_n = X_m(1 + E_x)^{n-m} \quad (2)$$

Finally, a quantitative value illustrating the starting concentration (ng/ $\mu$ L; nanograms/microliter) of template DNA is generated when plotted against a linear regression line, referred to as a standard curve (equation 3) [22]. The regression line is created using commercially produced genomic DNA that is serially diluted to form five data points, representing standardized concentrations (50 ng/ $\mu$ L, 5 ng/ $\mu$ L, 0.5 ng/ $\mu$ L, 0.05 ng/ $\mu$ L and 0.005 ng/ $\mu$ L) [22]. Although the DNA quantitation process was developed to estimate concentrations of template DNA between 50 ng/ $\mu$ L and 0.005 ng/ $\mu$ L, the sensitivity has overcome PCR limitations for the production of an STR profile. Therefore, forensic laboratories have conducted internal validations to define a stop at quantification (SAQ) threshold to quantitatively predict the amplification limit where an STR profile will no longer be generated [16, 23]. Thus, a sample with a concentration below this threshold would not produce a genetic profile and would not continue to progress through the STR analysis process.

$$C_T = m[\log(Qty)] + b \quad (3)$$

Subsequently, samples that were chosen to progress through the STR analysis process, but generated a lower or higher concentration than the optimal amplification range (0.5 ng to 2 ng) must be normalized before PCR (polymerase chain reaction) [16, 24]. Sample normalization is a

two-fold process, where a sample can either be concentrated or diluted to reach the desired amplification target. A sample must be concentrated through the use of a Microcon<sup>®</sup> Centrifugation Filter when the initial concentration falls below the minimum amplification target (0.5 ng), due to the presence of excessive elution buffer [16]. This part of the normalization process assists with preventing allelic dropout and/or locus imbalance in the genetic profile. Conversely, a sample must be diluted when the concentration of DNA is above the maximum target to prevent the saturation of the CCD camera and the reduction of artifacts such as stutter, pull-up, and -A [16].

Next, the normalized samples will undergo a multiplex polymerase chain reaction, a method developed to exponentially replicate multiple segments of targeted DNA simultaneously [16]. This replication process is capable of generating 1.07 billion copies of each target loci by denaturing each DNA strand, annealing fluorescently labeled primers, and elongating the DNA strand through the addition of deoxynucleotide triphosphates (dNTPs), recreating fragments of double-stranded DNA [16, 26-28]. Thus, the PCR process is comprised of 29-32 cycles that alter heating and cooling settings to perform each task.

Lastly, a tetra-mutagenic component such as formamide is applied to the amplified product promoting chemical denaturation, enabling the single-stranded DNA to be separated by size through capillary electrophoresis [16, 29]. In order to execute capillary electrophoresis, a viscous polymer must first propel through the system enabling the amplified product to then enter the uncoated capillary through the application of an electrokinetic injection. The applied voltage allows the negatively charged DNA particles to migrate toward the anode, whereas the polymer retards the larger molecules permitting the smaller molecules to pass through the capillary quicker [30]. Thus, separating the DNA fragments by size. The DNA particles will then reach a detection window where an Argon laser beam will cause excitation in the fluorescent dye, formerly attached to the 5' oligonucleotide during PCR amplification [16]. The fluorescent emittance from the excitation of the dye encompasses different wavelengths, where a sequence of lenses is then utilized to direct the fluorescent light into the Applied Biosystems Invitrogen (ABI) Prism<sup>®</sup> spectrograph [30]. The diffraction grating within the spectrograph causes the light to then be dispersed by wavelength, which is subsequently detected by the CCD panel. The fluorescent dyes selected throughout this process were chosen to maximize the regions of interest while minimizing spectral overlap. Therefore, allowing the CCD camera to detect the wavelengths of interest with the assistance of programmed virtual filters and matrix files [30]. A matrix file is used to correct any "spectral overlap" that may occur when the emission of light is detected outside of the virtual filter [30].

Once the spectra are resolved, the data collection software will record the fragment size by tracking the fluorescent intensity over time producing a data file (.fsa or .HID) [31]. The data file can then be examined by an analyst through utilizing data analysis software, such as GeneMapper-*IDX*. Data analysis software has been developed as an expert system to size data based on internal size standards [32], generate allele calls by comparing the sizing data to an allelic ladder [33], and allowing for the application of a set of rules to assist with data interpretation. For example, an analytical threshold is an established rule determined by the laboratory to define real data from background noise that can be stored within the software [34]. Therefore, data that falls below the analytical threshold will not be sized nor will an allele call be noted unless overturned by the reviewing analyst. Another important rule that must be defined

before further analysis can occur is the establishment of a stochastic threshold. A stochastic threshold is a limit at which the laboratory can reasonably assume that allelic dropout may have occurred through empirical observation of single-source samples during the internal validation process [34]. Assuming the data file contained sizable data that could be converted into an allele call and all necessary rules are established within the software, data interpretation can now proceed. Once the genetic profile has been interpreted, the genotypes from one profile can be compared to genotypes of another for the generation of a court-admissible report.

## 1.2 Two-Person Mixture Interpretation

Sexual assault evidence is normally comprised of DNA from more than one individual, where the victim is one contributor to the admixture due to the collection process. Assuming that there is one suspect and ejaculation occurred, samples that presumably contain spermatozoa will undergo the differential extraction process. The primary goal of a differential extraction is to separate the male contributor (“S” Fraction) from the female contributor (“E” Fraction) for the production of two different single-source profiles [16]. However, this outcome is not always achieved because male epithelium in a sample may be lysed in conjunction with the female epithelium causing the “E” fraction to produce a mixture profile [35]. Additionally, carry-over is likely to occur as a consequence of the differential extraction process giving rise to a mixture profile in the “S” fraction when the female DNA is not entirely removed during the first separation [36]. Regardless of the differential outcome, DNA analysts will interpret all results by identifying 1) true alleles, 2) the existence of a mixture, 3) the number of potential contributors, 4) each donor’s contribution to the admixture, and 5) determining the appropriate statistical analysis [37-38].

First, the DNA analyst must distinguish true alleles from extraneous peaks, which may arise from artifacts created during capillary electrophoreses, such as pull-up. Pull-up is the result of peaks generated when the matrix file does not sufficiently correct for spectral overlap between dye colors [39]. Pull-up can have the same morphology as a true allele peak, is often identified directly underneath a peak of another dye channel, and if strong enough could be present across multiple channels [40]. Allele designations based on pull-up peaks can be eliminated from interpretation but must be properly documented. Stutter, on the other hand, is an amplification artifact that results when DNA polymerase missteps on the template strand during synthesis within the first eleven cycles of amplification [41-45]. This misstep will then cause either an insertion or deletion of a repeated motif to occur and will continue to amplify throughout the remaining cycles. Therefore, the stutter is primarily visualized one repeat away (+/- 4 bp) from a true allele peak and can be characterized by size [44, 45]. During the developmental and internal validation processes, the stutter percentages are established and documented to serve as a guide to aid analysts in determining if an allele peak is in the stutter position. Stutter percentages are calculated by dividing the RFU of the stutter peak by the RFU of the true allele peak and multiplying by 100 to express the result as a percentage (equation 4) [44, 46]. Any peak in the stutter position that falls below the validated percentages can be eliminated as a true allele and will not be utilized for genotype interpretations or statistical analysis.

$$\text{Stutter \%} = \frac{\text{RFU of the } n-4 \text{ peak}}{\text{RFU of the true allele peak}} * 100 \quad (4)$$

However, one challenge that may arise with distinguishing between true allele peaks and stutter is the possibility of allele stacking. Allele stacking occurs when an allelic peak is shared between the stutter peak of one contributor and the true allelic peak of another contributor causing overlap to exist [47-48]. The analyst will then attempt to differentiate the true allele by determining the allele's contribution to the shared peak by calculating the minimum allele contribution (MAC). The MAC is calculated by subtracting the stutter threshold value (STV) from the stutter peak height (SPH; equation 5) [48]. If the MAC is greater than the analytical and stochastic thresholds, the peak is considered a true allele and will be used for interpretation. If the MAC is greater than the analytical threshold but below the stochastic threshold, the locus may be considered incomplete and not suitable for comparison to reference samples or statistical analysis [47-48]. If the MAC is less than the analytical threshold, it is possible that the peak is *not* a true allele and the analyst may remove the label from the peak in the analysis software.

$$\text{MAC} = \text{SPH} - \text{STV} \quad (5)$$

Next, the DNA analyst will evaluate the STR profile for the presence of a mixture by observing the total number of allelic peaks at a locus across the entire profile. Generally, a sample is considered a mixture when more than one locus contains at least three alleles [38]. In order to identify which pair of sister alleles belong to a contributor, the peak height ratio (equation 6) is calculated at each locus individually [34, 47, 49]. If the peak height ratio of two allele peaks is less than the validated threshold, which is typically 70% [50], an imbalance exists suggesting the alleles may not belong to the same heterozygous pair. Alternatively, if two allele peaks are greater than the validated threshold percentage, the alleles are assumed to belong to the same contributor because heterozygous alleles are theorized to amplify equally [44].

$$\text{PHR} = \frac{\text{RFU of the smaller peak height}}{\text{RFU of the larger peak height}} \times 100 \quad (6)$$

Once a mixture has been identified and all extraneous peaks have been eliminated, the number of potential contributors must be estimated to provide accurate conclusions and make informed decisions. Not only can the number of contributors provide weight to the source level proposition, but this information can give the analyst insight into missing information. Genetically an individual can only inherit two alleles (except for tri-allelic individuals), one from each parent [44]. Analysts can determine the minimum number of contributors by counting all of the alleles observed at a locus and dividing that number by 2 (equation 7) [44, 47, 51]. The locus that should be chosen for this calculation must exhibit the largest number of true alleles to prevent underestimating the number of contributors [44]. For example, a mixture profile consisting of two-contributors should exhibit no more than four alleles per locus (assuming heterozygosity) and upon dividing this number by 2; the analyst should conclude the presence of two individuals [44]. The analyst will then examine the peak heights and peak height ratios to determine if this a reasonable assumption prior to making a conclusion. If the analyst suspects that information could be missing from multiple loci based on the observed peak heights, the mixed DNA profile may be considered *not* suitable for comparison to reference samples or statistical analysis.

$$\text{Minimum \# of contributors} = \frac{\text{Total Number of Alleles at a Locus}}{2 \text{ (Highest Number of Alleles per Contributor)}} \quad (7)$$



If the mixed DNA profile is determined to be from exactly two individuals, the DNA analyst will attempt to classify distinct groups of alleles to determine each donor's contribution to the admixture. Again, there are several ways to classify and calculate a donor's contribution, but only the scenario of one intimate and one foreign donor will be explained, as it applies to this research effort. First, the intimate donor's (known contributor) alleles will be identified, labeled and categorized into a distinct group [47]. Then the foreign contributor's alleles will be deduced by subtracting the alleles from the known contributor. However, instances, where the known and foreign contributor share an allele at a locus the percent peak height contribution (equation 8) for each donor must be calculated at all other loci to properly categorize the alleles at this location [38, 47, 52-53]. The percent peak height contribution will then be used to calculate each donor's total contribution to each locus of interest (LOI) through multiplying by the locus total. Once the total contribution of the LOI is calculated, the analyst must consider all allele combinations by analyzing the peak height ratios of the sister alleles prior to making any conclusions [47]. Finally, a deduced foreign profile is generated and ready for statistical analysis.

$$\% \text{ Peak Height Contribution} = \frac{(\text{PH of allele 1} + \text{PH of allele 2})}{(\text{Total PH of all alleles at a locus})} \times 100 \quad (8)$$

The DNA analyst must then determine the appropriate statistical analysis, if any, for the information obtained during the interpretation of a mixture. Continuing with the information above, the random match probability (RMP; equation 9) [45] is the most appropriate statistical calculation given a single-source profile was deduced from the two-person mixture of an evidentiary sample. The RMP is an extension of the Hardy-Weinberg Equilibrium (HWE), where allele frequencies (p and q) in a population are utilized to calculate the genotype frequency (2pq and p<sup>2</sup>), which can then be multiplied across each locus to obtain the RMP [45, 54-55]. In instances where a population has exhibited effects of subdivision (inbreeding), a theta (θ) correction must be employed because some of the HWE assumptions will not be met [45, 54]. In addition to the traditional RMP calculation, a DNA analyst may further employ this calculation under two approaches, restricted and unrestricted. The unrestricted approach assumes all genotype combinations are possible despite differences in peak heights. Whereas, the restricted approach assumes specific genotype combinations are possible based upon calculating peak height ratios for the determination of sister alleles [34].

$$\begin{aligned} \text{HWE: } & (p^2 + 2pq + q^2) = 1 \\ & P = 2pq \text{ (Heterozygous genotype)} \\ & P = p^2 + p(1-p)\theta \text{ (Homozygous genotype)} \\ \text{RMP} & = (P_1)(P_2) \dots (P_n) \end{aligned} \quad (9)$$

Regardless, the RMP is a statement provided to explain the probability of obtaining a DNA profile from a random individual within a population and that profile matching the DNA profile obtained from an evidentiary sample [45, 54]. The purpose of providing the jury with a statistical conclusion, such as the RMP, is to offer a scientifically valid estimate for the frequency of obtaining the same information by chance. This estimate will then assist the jury in providing significance or weight to the evidence in question.

### 1.3 ParaDNA<sup>®</sup> System

The LGC ParaDNA<sup>®</sup> System is a screening instrument that was developed to provide the forensic science community with a novel method to triage various types of biological evidence. This instrument is light-weight and does not require any sample preparation, enabling the flexibility to test evidence while at a crime scene, in a hospital, or at a laboratory. The ParaDNA<sup>®</sup> System has the capability of analyzing up to four samples simultaneously, where each unit can complete analysis, thus providing an STR profile, in under 90 minutes [56]. Unlike traditional techniques, biological evidence must be sampled directly or indirectly using an individually packaged sample applicator. Direct sampling occurs when the sample applicator physically contacts the stain and/or item of evidence [57]. Whereas, indirect sampling occurs when a cotton swab is used to collect the sample which is then sampled with the applicator—therefore the applicator never comes into contact with the original stain [57]. Both sampling methods are advantageous because only 2-5% of the sample is consumed, leaving enough evidence behind for traditional STR analysis. Following the collection process, the four-pronged head of the applicator is injected into one of the following reaction cartridges: Intelligence Test, Body Fluid ID, and/or Screening test. Each unit of the instrument may contain a different assay which can be analyzed concurrently to expand upon the types of evidence that can be screened.

The Intelligence Test assay was developed to sequence STRs at five chosen loci (TH01, D3S1358, D8S1179, D16S539, D18S51, and Amelogenin), where a condensed range of alleles are detected [58]. The reaction cartridge for this assay contains four wells, where LGC's proprietary HyBeacon<sup>™</sup> (Hybridization Beacon) probes are enclosed. The HyBeacon<sup>™</sup> probe is comprised of linear oligonucleotides containing a fluorescent dye (fluorophore), which covalently binds to internal nucleotides [59-61]. In order to capture overlapping STRs at various loci within a single well, three spectrally different fluorophores are utilized. The HyBeacon<sup>™</sup> probe will hybridize to a complementary DNA sequence located on one of the strands of DNA throughout the exponential phase of the PCR process. While the DNA is in a single-stranded conformation, low levels of fluorescence are emitted by the fluorophore; whereas measurable levels are emitted when the complementary DNA strand anneals reconstructing the double-strand causing the probe to dissociate [59-61]. Table 1 details each locus being targeted, the detectable range of alleles (Low or High), the fluorescent channels, and the wells of the cartridge where the reaction takes place.

Table 1. Summary of the Intelligence Test assay.

| <b>Locus</b>   | <b>Well A</b> | <b>Well B</b> | <b>Well C</b> | <b>Well D</b> |
|----------------|---------------|---------------|---------------|---------------|
| <i>D16S539</i> | 8-12+ (L)     | 12-15+ (H)    | -             | -             |
| <i>D18S51</i>  | 14-17+ (L)    | 17-21+ (H)    | 10-14+        | -             |
| <i>D8S1179</i> | 8-11+ (L)     | 11-14+        | 14-18+ (H)    | -             |
| <i>D3S1358</i> | -             | -             | 13-16+ (L)    | 16-19+ (H)    |
| <i>TH01</i>    | -             | -             | -             | 5-9.3+        |
| <i>AMEL</i>    | -             | -             | -             | X, Y          |

Next, melting curve analysis is used to determine the presence and identity of the target DNA sequences by assessing the melting temperature ( $T_m$ ) of the hybridization probes [62-63]. The

melting temperature is captured when 50% of the probe has dissociated from the template strand, resulting in a decrease of fluorescence [63]. Therefore, various repeat units and/or single nucleotides are differentiated by the variation in  $T_m$ . Conversely, there is a small difference in the melting temperature of STR alleles of similar lengths causing unreliable differentiation upon the use of elongated probes [59]. The number of repeat units available for probe hybridization is therefore limited through the use of a non-fluorescent blocker. The blocker operates by binding to the template strand preventing the Hybeacon™ probe from binding to the repeat units in that area, shortening the length of the allele(s) and allowing the melting temperature to be reduced for better separation [59]. In order to ensure the specificity of hybridization, an anchor is then attached to the non-repetitive region of the probe preventing probe slippage and lower resolution of melting peaks [59].

Subsequently, the ParaDNA® Software converts the data obtained from the melting curve into genotypes at all of the detectable loci. Alleles that are confidently typed and fall within all established thresholds are visualized with a green background [64]. However, when alleles are not typed and/or a mixture is present, a grey background will appear indicating further analysis may be required [64]. Once an STR profile is obtained the genotypes can then be compared to genotypes from another sample. If two samples contain similar genotypes at a given locus, a match probability will be calculated using an internal allele frequency database. The match probability is indicative of how likely it is for another individual within the selected population to have the same STR profile as the sample being analyzed [65]. Conversely, if any of the typed alleles are not concordant a match probability cannot be established.

Furthermore, the data displayed by the ParaDNA® Software is formatted to allow non-expert users to quickly and comprehensively analyze STR profiles at any of the five locations. For further analysis at an expert level, the ParaDNA® Data Analysis Software provides an alternative way to visualize the data by generating melting curve plots, allele tables, and bar charts. The bar charts display all of the observed data for each fluorescent channel in comparison to the internally developed thresholds [66]. This information will assist expert users in determining whether the data observed on the melting curve is real data, an artifact, or a false positive. Real data is indicated when the bar surpasses the green dot visualized on each bar chart and the fluorescent data can be fitted with the internal references [66]. False positives, on the other hand, are potentially indicated when the bar errs on the side of the red dot and the fluorescent data falls outside of the internal references.

Additionally, a melting curve plot is constructed by comparing the change in temperature (°C) to the relative fluorescent units (RFUs) of each fluorescent channel [66]. When the melting temperature for a specific repeat unit is reached, the Hybeacon™ Probe begins to dissociate causing a reduction in RFUs, where a “dip” is displayed on the melting curve plot [67]. For instance, the melting temperature for a 12 repeat allele is 40° C, therefore a “dip” would be exhibited at that location on the plot. This information is then translated into a first derivative plot enabling fluorescent peaks to be visualized similar to the electropherogram commonly analyzed by experts during traditional STR analysis.

Finally, mixture STR genotypes are deduced through assessing the data at each locus. Upon selecting a locus using the ParaDNA® Data Analysis Software the possible allele calls are



displayed in an allele table, along with the melting curve plots and bar charts. The forensic analyst will then review the data provided and make a conclusion on the possible genotypes for each contributor.

#### 1.4 RapidHIT<sup>®</sup> 200

The RapidHIT<sup>®</sup> 200 is a mobile bench-top instrument that was developed by IntegenX to rapidly produce an STR profile for single-source (reference) samples [68]. The primary goal of this development was to permit law enforcement officers to collect and rapidly test buccal swabs at a booking station for making an arrest. However, the RapidHIT<sup>®</sup> 200 has exhibited greater success in the field as a rapid technique to identify victims of mass disasters. Regardless of the instruments' intended use, flexibility exists for various situations and experience levels. In addition to the system's flexibility, the instrument does not require sample preparation has the capability of analyzing up to seven samples simultaneously, thus providing an STR profile in 90 minutes [68-69].

Unlike the ParaDNA<sup>®</sup> System, biological evidence must be sampled directly using a traditional cotton swab and placed directly into one of the RapidHIT<sup>®</sup> GlobalFiler Express Cartridges. The sampling cartridge utilized by this system is a microfluidic device composed of two primary layers, the fluidic and analytical. The fluidic layer of the device was designed to store aliquot reagents within the storage chambers and move the reagents through the transport channels when needed, using a pneumatic pressure-driven interface [70]. The pneumatic pressure-driven interface is a series of valves and pumps hardwired to move and mix fluidic streams throughout the device. The air supply system will then control the flow rate of the reagents being transported through the channels to the different ports, regulating the volume of each reaction mixture [71-73]. Furthermore, the analytical layer of the device was designed for the PCR reaction chamber and the capillary electrophoresis channels.

Similar to traditional methodologies, the RapidHIT<sup>®</sup> GlobalFiler Express Cartridge was designed to extract, amplify, and separate DNA for the production of an STR profile, using NDIS (National DNA Index System) approved chemistry [74-75]. The DNA entrapped within the substrate is first extracted at the bottom of the sample well. The lysate is then transported through the channels of the microfluidic device to the magnetic bead chamber, which is used to isolate and purify the DNA from all other cellular material. The magnetic bead chemistry was also designed to enable up to 50 ng of DNA to bind to the beads, partially quantifying the amount of DNA that can be amplified [76]. Next, amplification of all 24 loci is performed utilizing the GlobalFiler<sup>®</sup> Express v2.1 chemistry with slightly varied thermal cycling parameters for optimal replication [74]. Lastly, the amplified DNA is separated by size during capillary electrophoresis, where the MapMarker<sup>®</sup> DY632-500bp size standard is used to obtain sizing results. The IntegenX trace analysis software is then utilized to assess the output electropherogram, where the analytical and stochastic thresholds (AT; ST) are computed per run and per locus [74]. The data can be further analyzed by exporting the results to GeneMapper<sup>®</sup> ID-X or GeneMarker<sup>®</sup> HID.

Lastly, STR profiles generated by the RapidHIT<sup>®</sup> 200 have the capability of being searched against various internal and external databases using the RapidLINK software v1.0. The

RapidLINK software contains three databases: match, elimination, and familial that can be selectively searched. The respective databases then have the capability of synchronizing with the Combined DNA Index System (CODIS) database for a more thorough search, where a rapid match may be generated [77].

## Chapter 2: Materials and Methods

### 2.1 Sample Collection

Ten female donors affiliated with West Virginia University volunteered to donate liquid saliva and four buccal samples. The female donors were instructed to deposit approximately 1.5 mL (milliliters) of liquid saliva into a 2 ml micro-centrifuge tube. The buccal swab samples were then collected using a 6-inch sterile cotton-tipped applicator from Dukal Corporation and Dynarex Corporation (Ronkonkoma and Orangeburg, NY). Each donor was instructed to vigorously rub the inside of both cheeks and along the gum line while twisting the swab for approximately 10-15 seconds. The tip of the swab was then removed and placed into a 2 ml micro-centrifuge tube. The micro-centrifuge tube was labeled with a unique identification number, then stored at -16.6 °C. The unique identification number assigned to each participant's sample was randomly generated using R version 3.4.3.

Ten male donors also affiliated with West Virginia University volunteered to donate semen samples. The male donors were instructed to ejaculate into a sterile 50 mL conical tube (Thermo Fisher Scientific, Waltham, MA). The semen samples were then stored in a Styrofoam cooler underneath disposable ice packs and returned to the DNA laboratory within 24-48 hours. Upon receipt, the samples were labeled with a unique identification number then stored at -16.6 °C. All donors throughout this study were consenting individuals under the West Virginia University Institutional Review Board protocol number 1705604512.

### 2.2 Sample Preparation

The unique identification numbers assigned to each participant were paired together by using a random number generator. This pairing correlates to the samples that will be combined for the creation of each admixture. Prior to the physical preparation of the pairwise admixtures, a portion of each sample was extracted and quantified in five replicates. The quantities obtained were documented in R (version 3.4.3) where equations 10 and 11 were used to calculate the average concentration of each sample.

$$\bar{X} \pm t_{\frac{\alpha}{2}} \sqrt{s^2 + \frac{s^2}{n}} \quad (10)$$

$$\sum \frac{|x_i - \bar{x}|}{n} \quad (11)$$

Table 2. Experimental data sets utilized for sample preparation for each instrument.

| <i>Data Set:</i> | <i>Instrument:</i>        | <i>Sample Size (n):</i> | <i>Mixture Ratio:</i> |
|------------------|---------------------------|-------------------------|-----------------------|
| 1                | RapidHIT <sup>®</sup> 200 | 20                      | 1:10                  |
| 1                | RapidHIT <sup>®</sup> 200 | 4                       | 1:20                  |
| 2                | ParaDNA <sup>®</sup>      | 20                      | 1:5                   |
| 2                | ParaDNA <sup>®</sup>      | 20                      | 1:10                  |
| 2                | ParaDNA <sup>®</sup>      | 20                      | 1:15                  |
| 2                | ParaDNA <sup>®</sup>      | 20                      | 1:20                  |
| 2                | ParaDNA <sup>®</sup>      | 11                      | 1:25                  |

Mock sexual assault mixtures were then prepared in two different data sets (Table 2) using the established pairwise combinations. The first data set was comprised of 24 total samples, where 10 samples were prepared at a 1:10 mixture ratio in duplicate using a 6-inch sterile cotton-tipped applicator. Two of the ten pairwise combinations were then randomly selected and prepared at a 1:20 mixture ratio in duplicate (4 total samples). The mixture ratios were created by first diluting the semen samples with nuclease-free water (lot #: 1509080), where 5  $\mu$ L of the dilution was pipetted into a new micro-centrifuge tube. Then, the appropriate amount of saliva (50  $\mu$ L or 100  $\mu$ L) was added to each micro-centrifuge tube containing the diluted semen to ensure homogeneity. Next, the admixtures were vortexed, centrifuged, and pipetted onto a new 6-inch sterile cotton-tipped swab, covering as much of the surface as evenly as possible. Finally, the prepared swabs were placed into the RapidHIT<sup>®</sup> GlobalFiler Express Cartridge.

The second data set was comprised of 91 total samples that were prepared at the following mixture ratios: 1:5, 1:10, 1:15, 1:20, and 1:25 (Male: Female). Each mixture ratio was prepared in duplicate except for the 1:25 ratio, due to a shortage in saliva. First, the semen samples were diluted with Invitrogen TE<sup>-4</sup> buffer (ThermoFisher Scientific, Waltham, MA), where 5  $\mu$ L of the dilution was pipetted into a new micro-centrifuge tube for each sample. The saliva samples were then added to the appropriate micro-centrifuge tube containing the diluted semen to create the proper mixture ratio. Next, the mixtures were vortexed, centrifuged, and pipetted onto a 3.14 inch flocked nylon swab (Fisher Scientific, Waltham, MA), covering as much of the surface as evenly as possible. The flocked nylon swab was then sampled for approximately 1 minute with the ParaDNA<sup>®</sup> sample applicator while applying moderate pressure. The four-pronged head of the sample applicator was snapped into the Intelligence Test<sup>™</sup> reaction plate and the handle was broken off.

### 2.3 Sample Processing

The RapidHIT<sup>®</sup> GlobalFiler Express Cartridges containing up to seven samples were placed onto the RapidHIT<sup>®</sup> 200 for testing. The buccal protocol was selected in the IntegenX trace analysis software to designate a 70°C lysis temperature and a 28 cycle PCR parameter. Once finished, the data was imported into the GeneMarker<sup>®</sup> HID software for analysis. The cartridges were stored in the refrigerator at approximately 4°C. Later, the swabs were removed from the cartridges and rehydrated with 50  $\mu$ L of nuclease-free water. Each swab was then directly sampled with the ParaDNA<sup>®</sup> sample applicator and placed into the Intelligence Test<sup>™</sup> reaction plate.

The Intelligence Test™ reaction plate containing the applicator head was placed onto the ParaDNA® heating block. The ParaDNA® software was opened and the application parameters were set for the Intelligence Test™ chemistry. The case number and item number was also inputted into the software before the run was started. Once finished, the data was imported into the ParaDNA® Data Analysis software for expert analysis and the reaction plates were discarded.

Next, one of the four buccal swabs collected from each of the female participants was extracted with the Qiagen DNA Investigator kit (lot #: 157036909) to obtain a reference DNA profile necessary for deconvolution. The “Isolation of Total DNA from Surface and Buccal Swabs” manufacturer protocol was followed to perform a manual standard extraction. The purified DNA was then eluted from the QIAamp MinElute Column with 100 µL of Buffer ATE. Additionally, five of the pairwise samples prepared in the following ratios, 1:5, 1:15, 1:20 and 1:25, were differentially extracted using the “Isolation of Total DNA from Sexual Assault Specimens” manufacturer protocol. The female (“E”) fraction was stored in the refrigerator and no further analysis was completed on this portion of the sample.

The concentration of the purified DNA was then estimated on the 7500 Real-Time PCR Instrument (Applied Biosystems, Foster City, CA) using the Quantifiler™ Trio DNA Quantification Kit (ThermoFisher Scientific, Waltham, MA). The manufacturer’s protocol was followed. A virtual standard curve created within the HID Real-Time PCR Analysis Software v1.3 (Applied Biosystems, Foster City, CA) was applied to the run method and used to approximate the concentration of each sample. Samples with a concentration greater than 1.00 ng/µL were diluted using either a 1:10 or 1:100 dilution factor.

Thereafter, the normalized samples were amplified on the 9700 GeneAmp® Thermal Cycler using the GlobalFiler™ PCR Amplification Kit (ThermoFisher Scientific, Waltham, MA). The manufacturer’s protocol was followed. Lastly, the 3500 Genetic Analyzer was used to separate the amplified DNA by size through the process of capillary electrophoresis. The manufacturer’s protocol was followed. A 10-second injection was applied to the female reference samples to reduce the amount of DNA entering the capillary. This reduction was necessary to prevent saturating the CCD camera and reduce the production of artifacts associated with large amounts of input DNA. However, a 15-second injection was applied to the male reference samples obtained from the differential extraction. The increase in injection time was due to a lower concentration of DNA. The data was then analyzed and compared to the data obtained from the ParaDNA® Instrument and the RapidHIT® 200, using the GeneMapper® ID-X v1.4 software.

### Chapter 3: Results and Discussion

One of the primary challenges that confront forensic scientists of any discipline is the inability to gain insight into the disposition of the sample before arriving at the crime laboratory. The forensic biology discipline, however, is deeply affected by this issue forcing the analyst to make decisions about processing evidence based upon little information and visual observations. Currently, many laboratories have begun to utilize the male screening technique, recommended by SWGDAM, as a solution to combat this problem. However, this novel screening method still poses many challenges as only ~1/8 of the swab is utilized for testing potentially causing DNA to

be overlooked [9]. Additionally, laboratories are still required to expend more time and resources to perform additional testing just to obtain this information for enhanced decision-making.

The main objective of this research effort was to evaluate the use of the ParaDNA<sup>®</sup> Intelligence Test and the RapidHIT<sup>®</sup> 200 as a screening method for sexual assault evidence, in a hospital setting. It is believed that the ability to obtain information regarding the disposition of the sample at the time of collection may provide DNA analysts with greater insight earlier in the analytical process. Thus, enabling better decision-making regarding the manner in which the sample should be tested. The data presented here demonstrate that the ParaDNA<sup>®</sup> Intelligence Test is not suitable for non-laboratory testing, as the instrument produced a large failure rate in terms of identifying the presence of a mixture. The RapidHIT<sup>®</sup> 200, on the other hand, is a suitable option for triaging sexual assault evidence in a hospital setting and should be further researched prior to use.

### Chapter 3.1 A Qualitative Approach

Mock sexual assault admixtures were tested using the ParaDNA<sup>®</sup> Intelligence Test with a direct sampling approach at five mixture ratios (1:5, 1:10, 1:15, 1:20, and 1:25). As part of a qualitative evaluation, 40 samples out of 91 total samples run on this instrument produced a single-source profile. Consequently, the instrument was only able to detect a “possible mixture” in two other samples, where incomplete profiles were obtained for the remaining samples preventing further analysis.

Previous expectations were that the instrument would exhibit difficulty with identifying the presence of a mixture across different ratios because of the scientific foundation the instrument was based upon. The founding principle of this instrument is to designate allele calls from a confidence assessment derived from an internal algorithm originating from the melting curve analysis. Therefore, a hypothesis was formulated that the instrument could potentially select allele information from two separate individuals based upon the greater confidence of those allele calls. Thus, the instrument would return the result of a single-source profile, when a mixture is actually present.

|                   | D16                              |   | D18 |    | THO |   | D8 |    | Amelogenin |   | D3 |     |
|-------------------|----------------------------------|---|-----|----|-----|---|----|----|------------|---|----|-----|
| <b>2964:63432</b> | 12                               | - | 13  | 14 | 6   | 8 | 10 | 14 | X          | Y | 16 | -   |
|                   | <b>Possible mixture detected</b> |   |     |    |     |   |    |    |            |   |    |     |
| <b>2964:6343</b>  | 12                               | - | 14+ | 18 | -   | - | 13 | 14 | -          | - | 15 | 16+ |

|                   | D16 |    | D18 |   | THO  |   | D8 |     | Amelogenin |   | D3 |   |
|-------------------|-----|----|-----|---|------|---|----|-----|------------|---|----|---|
| <b>5175:80972</b> | 11  | 12 | 14+ | - | -    | - | 15 | -   | -          | - | 15 | - |
| <b>5175:8097</b>  | 12  | -  | 15  | - | 9.3+ | - | 10 | 14+ | -          | - | 15 | - |

Figure 1. ParaDNA<sup>®</sup> profile demonstrating allele sharing. At least three loci (D16, D8, THO) contain alleles (12, 13, 6) shared by both the male and female participants (top). ParaDNA<sup>®</sup> profile demonstrating unequal sampling. Unequal sampling is apparent at the THO and Amelogenin loci because the fluorescent probes located in well D of the reaction plate that are used to target alleles at those locations did not identify alleles, whereas other loci did (bottom).

However, the data demonstrated a high proportion of false negatives due to allele sharing between participants (*Figure 1, top*) and unequal sampling (*Figure 1, bottom*). The concept of allele sharing and unequal sampling was identified upon delineating the data back to the appropriate well of the ParaDNA® Intelligence Test reaction plate (*Figure 3*). This was accomplished with the use of reference profiles and the ParaDNA® Data Analysis Software. A “possible mixture” was identified for samples that exhibited more than two non-shared alleles at a locus, similar to sample 6343:2964 (*Figures 1 and 2A*). Additionally, superior results were observed when cells were evenly distributed to each reaction well in comparison to cartridges where unequal sampling occurred. A mixture could not be identified by the instrument for samples that had a maximum of two alleles at a given locus and when the participants had many overlapping alleles (*Figures 1 and 2B*).

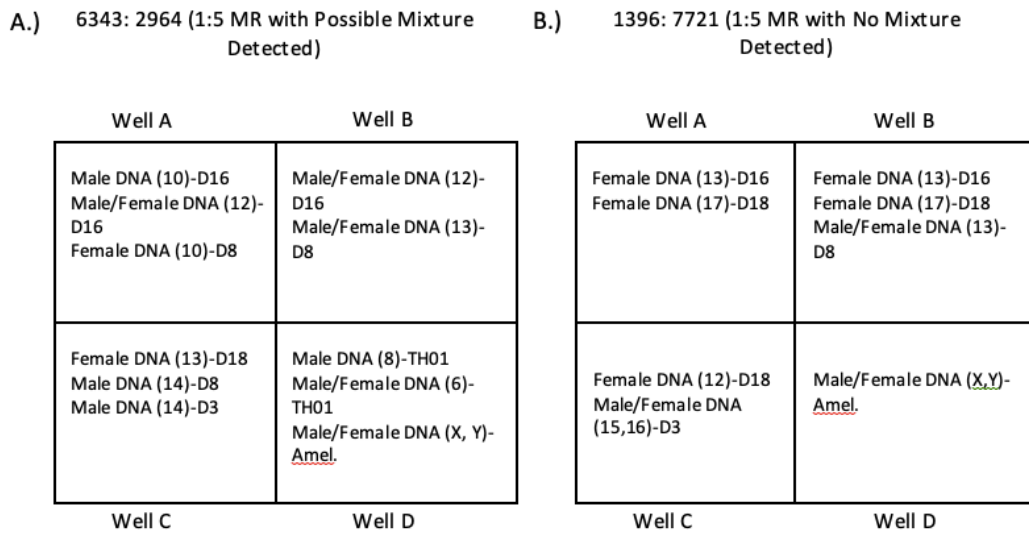


Figure 2. A schematic of the ParaDNA® Intelligence Test reaction plate containing the source of DNA, the allele call, and the chromosome location of each well for two admixtures (6343:2964, 1396:7721).

Upon further investigation, four of the admixtures utilized for this testing contained  $\geq 4$  shared alleles out of 13-18 possible alleles, explaining the instances of false negatives (*Table 3*). Overall these findings are in accordance with findings reported by Regan et. Al during the internal validation of the ParaDNA® Intelligence System [78]. The authors, however, contested allele sharing by selecting new donors, while analyzing the data with the ParaDNA® Data Analysis Software. As a non-expert system being utilized in a hospital setting the solution provided by the authors is problematic because nurses cannot exchange donors when allele sharing occurs and the training would be limited to use, where the additional software would not be utilized for analysis purposes at the time of collection. Thus, sexual assault nurse examiners may incorrectly compile documentation or make decisions based upon false perceptions.

Table 3. Allele charts for single-source DNA profiles demonstrating shared alleles (bold and purple).

|            | 1396 (Female) | 7721 (Male)  | 3685 (Female) | 5303 (Male)  | 6314 (Female) | 4663 (Male) | 5175 (Female) | 8097 (Male)  | 2964 (Female) | 6343 (Male)  |
|------------|---------------|--------------|---------------|--------------|---------------|-------------|---------------|--------------|---------------|--------------|
| D3S1358    | <b>15,16</b>  | <b>15,16</b> | 18            | 14,16        | 16,18         | 15,16       | 15            | 16,18        | 14, <b>15</b> | <b>15,16</b> |
| D16S539    | 13            | 12           | <b>12,13</b>  | <b>12,13</b> | <b>9,11</b>   | <b>9,11</b> | 12,13         | 11           | <b>8,12</b>   | <b>10,12</b> |
| Amelogenin | <b>X</b>      | <b>X,Y</b>   | <b>X</b>      | <b>X,Y</b>   | <b>X</b>      | <b>X,Y</b>  | <b>X</b>      | <b>X,Y</b>   | <b>X</b>      | <b>X,Y</b>   |
| D8S1179    | <b>13</b>     | <b>13</b>    | <b>11,12</b>  | <b>12,15</b> | 8,10          | 12,13       | <b>11,15</b>  | <b>10,15</b> | <b>10,13</b>  | <b>13,14</b> |
| D18S51     | 12,17         | 14,20        | <b>16,17</b>  | <b>14,17</b> | 14,17         | 12,13       | 14,18         | 12,15        | 12,13         | 16,18        |
| TH01       | <b>8,9,3</b>  | <b>9,3</b>   | <b>6,9,3</b>  | <b>9,3</b>   | 9             | 6           | 9             | 9,3          | <b>6,9,3</b>  | <b>6,8</b>   |



One concern about the findings correlated to unequal sampling was that the mixtures were not prepared with the use of a Hemocytometer rather a non-parametric prediction interval was utilized. In theory, a non-parametric prediction interval is a sufficient measure, as this mathematical model provides statistical boundaries for the distribution of concentrations that probabilistically predicts a future concentration. Practically, this model is not believed to have been sufficient for the preparation of mixtures because the data appeared relatively consistent throughout each ratio; whereas a significant decrease in male DNA would be expected as the proportion of female DNA increases, with proper preparation. This model may have been more effective if the diluted sperm fraction was quantified prior to the preparation of the mixtures, as this would have ensured the proper concentration was being achieved. Since the diluted sperm fraction was not quantified prior to the preparation of the admixtures a reasonable conclusion is that unequal sampling is more likely to occur, supporting the results obtained in this study. In the future, the use of a Hemocytometer would be recommended for the preparation of admixtures at varying ratios and/or the collection of post-coital samples from different intervals to ensure proper admixtures are prepared.

### Chapter 3.2 Linear Mixed-Effects Model

Likelihood ratio tests (LRT) and bootstrapping statistical inferences were performed on mock sexual assault admixtures from each of the datasets, under the linear mixed-effects model. The linear mixed-effects model was utilized to describe the relationship between the proportion of loci that suggests a mixture based upon the significance of fixed variables. Throughout this study, a mixture was determined by calculating the proportion of missing observations ( $y_i$ ) at a given locus across an entire DNA profile. A DNA profile obtained from the ParaDNA<sup>®</sup> Intelligence Test is comprised of six loci, differing from the GlobalFiler<sup>®</sup> Amplification Kit used in conjunction with the RapidHIT<sup>®</sup> 200 and the 3500 Genetic Analyzer to produce a 24 loci profile, respectively. Therefore, the proportion of missing observations was scaled to fit the instrument used for testing.

Furthermore, the model predictors intrinsically used as fixed variables include the mixture ratio, the instrument, and the swab type/diluent, where the variation originating from these variables is contributed independently [79-80]. The participants ( $1|ID_j$ ) used throughout this study, however, are repeated measures contributing to the model as random effects variables, where baseline variations may differ due to the variability in DNA concentration and cellular composition [79-80]. The respective model is explained by equation 12, which is first describing the intercept ( $\beta_0$ ), followed by the slope for each of the fixed variables ( $\beta_1$ - $\beta_3$ ), lastly characterizing the random error ( $\epsilon_i$ ) of the model.

$$y_i = \beta_0 + \beta_1 \times \text{Mixture Ratio} + \beta_2 \times \text{Instrument} + \beta_3 \times \text{Swab Type/Diluent} + (1|ID_j) + \epsilon_i \quad (12)$$

Initially, the mixture ratios prepared during this study were expected to follow a linear regression where the proportion of missing observations would increase as the ratio of male DNA decreased. Therefore, the mixture ratio was expected to have an effect on the determination of a mixture, a concept commonly observed during DNA analysis. The instrumentation utilized for testing was also expected to have an effect because the current use is designated for single-source samples only. Additionally, rapid DNA instrumentation does not require any sample

preparation, which could affect the yield of DNA, and only contains reagents known to effectively test samples with the presence of one type of cellular material. Lastly, the swab type and diluent were also expected to have an effect on the determination of a mixture in sexual assault samples because spermatozoa tend to entangle in the cotton fibers of a swab head [81]. Since both of the rapid instruments utilized do not have a way to disrupt this entanglement the yield was expected to decrease.

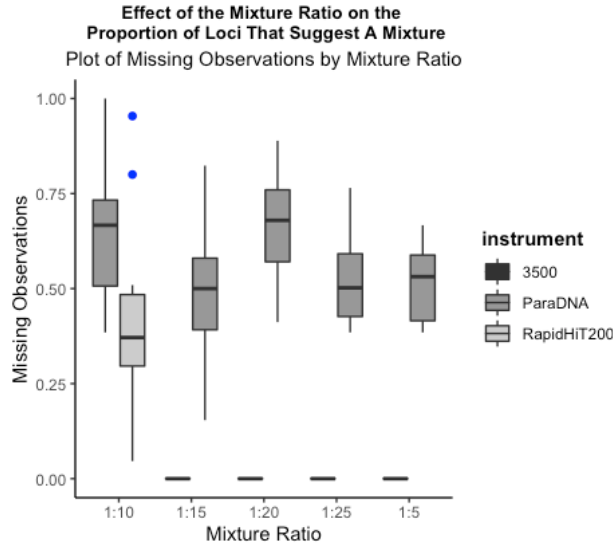


Figure 3. A boxplot representing the proportion of missing observations across five mixture ratios, testing the effect of the mixture ratio.

The findings demonstrated a relatively constant proportion of missing observations for the ParaDNA<sup>®</sup> Intelligence Test and the 3500 Genetic Analyzer across the five mixture ratios (Figure 3). To determine if the mixture ratio had a significant effect on the proportion of loci that suggest a mixture, a bootstrapped LRT was performed in R using the lme4 package [80, 82]. The LRT is performed by testing two hypothesis models. The first model is the null hypothesis (equation 14,  $H_0$ ), which omits the mixture ratio reducing the full model. The assumption under this hypothesis is there is no difference in the proportion of missing observations between each mixture ratio. The second model is the alternative hypothesis (equation 13,  $H_1$ ), which utilizes the full model and assumes there is a difference. In 10,000 bootstrapped samples, the null hypothesis was rejected ( $p= 0.012699$ ,  $LRT= 0.009034$ ), thus supporting the alternative hypothesis that there is a significant effect.

$$y_i = \beta_0 + \beta_1 \times \text{Mixture Ratio} + (1|ID_j) + \varepsilon_i \quad (13)$$

$$y_i = \beta_0 + (1|ID_j) + \varepsilon_i \quad (14)$$

As seen in figure 3, the 3500 Genetic Analyzer did not exhibit any missing observations across the four mixture ratios tested. This result appeared to significantly affect the bootstrapped LRT when testing this effect. Therefore, data generated at each mixture ratio by the 3500 Genetic Analyzer was removed from the model, along with other co-variants, such as the diluent. The results now fail to reject the null hypothesis ( $p= 0.2032$ ,  $LRT= 0.1703$ ). Thus, the mixture ratios could not be determined to have a significant effect on the determination of a mixture when the 3500 data were removed from the model, as previously believed. Although the RapidHIT<sup>®</sup> 200



produced a smaller proportion of missing observations at the 1:10 mixture ratio in comparison to the ParaDNA® Intelligence Test, there was no notable significant difference arising from the respective concentration.

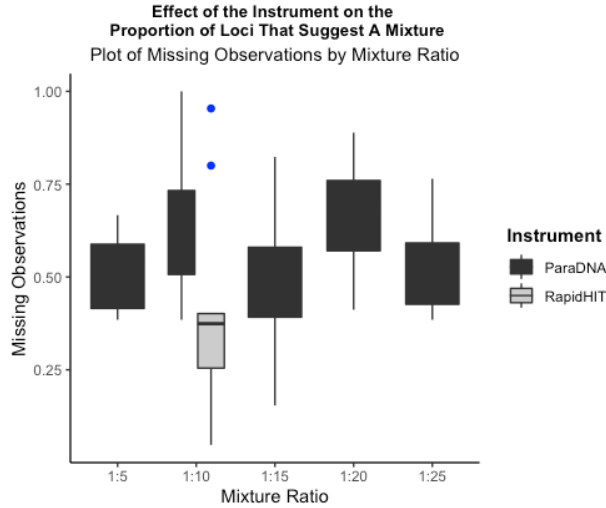


Figure 4. A boxplot representing the proportion of missing observations at different mixture ratios, testing the effect of the instrument.

However, the RapidHIT® 200 did appear to outperform the ParaDNA® Intelligence Test, in terms of an instrument detecting a greater number of alleles that contributed to the admixture (Figure 4). Equations 14 and 15 were then utilized to test the effect of the instrument on the proportion of loci that exhibit a mixture. The results of this test demonstrated the instrument does have a significant effect ( $p=0.0020998$ ,  $LRT=0.0008232$ ).

$$y_i = \beta_0 + \beta_1 \times Instrument + (1|ID_j) + \varepsilon_i \quad (14)$$

$$y_i = \beta_0 + (1|ID_j) + \varepsilon_i \quad (15)$$

Lastly, the swab type (cotton versus nylon) and the diluent (water versus  $TE^{-4}$  buffer) were tested using equations 16 and 17. Opposed to other studies [81, 83-84], the swab type and the diluent combination could not be determined to have a significant effect on the proportion of loci that suggested a mixture ( $p=0.3227$ ,  $LRT=0.2781$ ) in this study.

$$y_i = \beta_0 + \beta_1 \times Swab\ Type/Diluent + (1|ID_j) + \varepsilon_i \quad (16)$$

$$y_i = \beta_0 + (1|ID_j) + \varepsilon_i \quad (17)$$

Aside from the limitation outlined under the qualitative analysis chapter regarding the preparation of the admixtures, several limitations still exist in the present studies. The main limitation affecting the linear mixed-effects model is the lack of samples run on the RapidHIT® 200 across different mixture ratios. Although the instrument was found to have an effect on the determination of a mixture when testing sexual assault admixtures, comparison across multiple mixture ratios would have been a more comprehensive approach and should be considered for future studies. However, the expense of the RapidHIT® GlobalFiler Express Cartridges, the lack of grant funding, and the hard-stop expiration of the cartridges prevented further testing in the current study.

Additionally, the lack of reference profiles available eliminated a portion of the data from statistical analysis. Typically, in sexual assault cases, the victim is a known contributor to the sample and reference samples are often collected for comparison. The suspect, however, has not always been identified and a reference sample is not always available. Thus, male reference profiles were only obtained for a few of the participants to assimilate real case-type scenarios. This limitation specifically affected the ParaDNA<sup>®</sup> Intelligence Test because allele sharing could not be identified, preventing the determination of missing observations. In the future, reference samples should be tested in conjunction with the admixtures to further understand the relationship between obtaining a mixture profile and allele sharing.

Lastly, the swab type and diluent were changed during the study to achieve a better yield of male DNA based upon findings from other literature. Therefore, the samples run on the RapidHIT<sup>®</sup> 200 were not prepared with the same treatment as the samples from the ParaDNA<sup>®</sup> Intelligence Test. One solution to determining if this limitation significantly affected the data would have been to retest the RapidHIT<sup>®</sup> 200 samples with the new treatment parameters.

### Chapter 3.3 Logistic Regression Model

A parametric bootstrapped logistic regression model was applied to the data obtained from the previous mock sexual assault admixtures, where the suitability of the instrument was tested in two settings, laboratory, and non-laboratory. In this study, logistic regression is modeling the probability of detecting a mixture as a function of the admixture ratio. Unlike the last statistical model, a mixture was determined by two categorical predictors, mixture and single-source, which are dichotomous variables used to generate binary outcomes [85]. The two categorical predictors serve as the fixed effects variables, whereas the random-effects variables were removed. For stability, the replicated data was then aggregated to ensure a mixture was implied if either of the replicate admixtures suggested the presence of a mixture. The ‘glm’ function in R was used to perform this statistical analysis. Table 4 is an outline of the summary statistics obtained.

Table 4. Summary statistics for the fixed effects variables of the logistic regression model.

|                | <i>Instrument</i>     | <i>Setting</i> | <i>Estimate</i> | <i>Std. Error</i> | <i>Z-Value</i> | <i>P-Value</i> |
|----------------|-----------------------|----------------|-----------------|-------------------|----------------|----------------|
| <i>MR 1:5</i>  | ParaDNA <sup>®</sup>  | Non-Lab        | -2.1972         | 1.0541            | -2.084         | <b>0.0371</b>  |
| <i>MR 1:10</i> | ParaDNA <sup>®</sup>  | Non-Lab        | -1.3863         | 0.7906            | -1.754         | <b>0.0795</b>  |
| <i>MR 1:15</i> | ParaDNA <sup>®</sup>  | Non-Lab        | -19.5661        | 3400.7175         | -0.006         | <b>0.9954</b>  |
| <i>MR 1:20</i> | ParaDNA <sup>®</sup>  | Non-Lab        | -2.1972         | 1.0541            | -2.084         | <b>0.0371</b>  |
| <i>MR 1:25</i> | ParaDNA <sup>®</sup>  | Non-Lab        | -19.5661        | 3400.7175         | -0.006         | <b>0.9954</b>  |
| <i>MR 1:5</i>  | ParaDNA <sup>®</sup>  | Laboratory     | 0.8473          | 0.6901            | 1.228          | <b>0.2195</b>  |
| <i>MR 1:10</i> | ParaDNA <sup>®</sup>  | Laboratory     | 1.0986          | 0.6667            | 1.648          | <b>0.0994</b>  |
| <i>MR 1:15</i> | ParaDNA <sup>®</sup>  | Laboratory     | 0.4055          | 0.6455            | 0.628          | <b>0.5299</b>  |
| <i>MR 1:20</i> | ParaDNA <sup>®</sup>  | Laboratory     | 2.1972          | 1.0540            | 2.085          | <b>0.0371</b>  |
| <i>MR 1:25</i> | ParaDNA <sup>®</sup>  | Laboratory     | 0.0000          | 0.6325            | 0.000          | <b>1.000</b>   |
| <i>MR 1:10</i> | RapidHIT <sup>®</sup> | Both           | 1.030           | 0.521             | 1.976          | <b>0.0481</b>  |
| <i>MR 1:20</i> | RapidHIT <sup>®</sup> | Both           | 1.099           | 1.155             | 0.951          | <b>0.3414</b>  |

Prior to implementing the logistic regression model, the data was expected to demonstrate that the ParaDNA<sup>®</sup> Intelligence Test is not a suitable screening tool for use in a non-laboratory setting. This result was expected because an in-depth analysis of the data is not expected to occur in this type of setting. The main goal of utilizing the instrument at a hospital is to help provide SANEs with information that could aid in determining if re-collection needs to occur before the patient departs. However, it was expected that the ParaDNA<sup>®</sup> Intelligence and RapidHIT<sup>®</sup> 200 would be a suitable screening technique for use in a laboratory setting. In a laboratory setting, trained analysts have the ability to use education, experience, and specialized knowledge to analyze data regarding mixture DNA profiles.

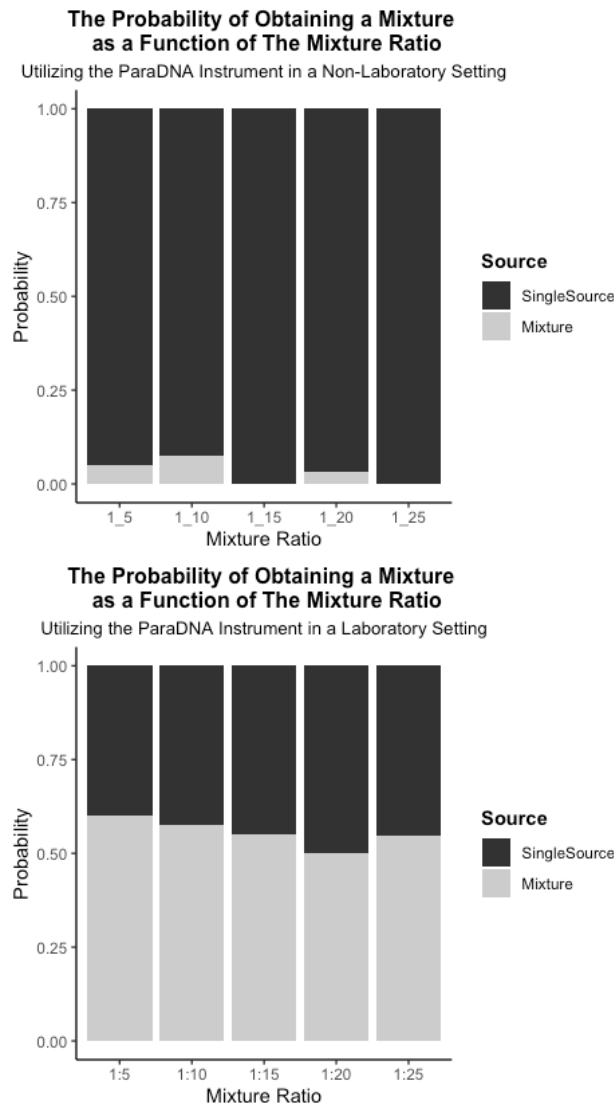


Figure 5. A bar graph representing the probability of obtaining a single-source profile versus a mixture when using the ParaDNA<sup>®</sup> Intelligence Test in a non-laboratory setting (top) and a laboratory setting (bottom).

The results of this study demonstrated that the replicate data obtained from a non-laboratory setting at mixture ratios 1:5, 1:10, and 1:20, has less than a one in four chance of suggesting the presence of a mixture (Figure 5, top). Thus, the probability of detecting a mixture as a function

of the mixture ratio in a non-laboratory setting is extremely low. The replicate data obtained from a laboratory setting, however, demonstrated that each mixture ratio had nearly an equal probability of suggesting a single-source and mixture profile, respectively (*Figure 5, bottom*).

Additionally, the parametric bootstrap statistical inference was applied to the ParaDNA<sup>®</sup> Intelligence Test data to observe whether the log-odds of a mixture being detected in each setting was equal for all of the mixture ratios. In 10,000 bootstrapped samples, the mixture ratios could not be statistically differentiated in either the laboratory or non-laboratory settings ( $p=0.3014$  and  $p=0.3246$ ).

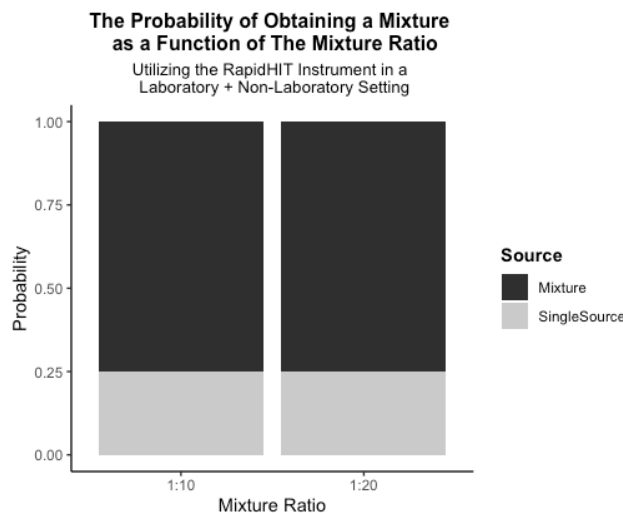


Figure 6. A bar graph representing the probability of obtaining a single-source profile versus a mixture when using the RapidHIT<sup>®</sup> 200 in either a non-laboratory or a laboratory setting.

On the other hand, the RapidHIT<sup>®</sup> 200 successfully detected a mixture profile in 75% of the cases when testing sexual assault admixtures at the 1:10 and 1:20 ratios, in either setting (*Figure 6*). Whereas, a single-source profile was detected in 25% of the cases. The same parametric bootstrap inference was applied and the log-odds between the two mixture ratios could not be statistically differentiated ( $p=0.9511$ ).

The two main limitations that effected the logistic regression model were the large failure rate of the ParaDNA<sup>®</sup> Intelligence Test when implemented in a non-laboratory setting before aggregating the data and the variability in the number of replications performed at each mixture ratio. The 1:10 mixture ratio had four replications per admixture because the swabs ran on the RapidHIT<sup>®</sup> 200 were also run on the ParaDNA<sup>®</sup> Intelligence Test. Additionally, the 1:25 mixture ratio only had one replicate for each admixture except for one, due to the lack of cellular material available to create this ratio in duplicate. Therefore, aggregating the data enabled the implementation of statistical analysis by changing the model to allow one of the replicates to exhibit a mixture to demonstrate success. Future studies should test mixtures with other cellular compositions to determine the suitability of the instrument for screening other types of evidence.

## Chapter 4: Conclusions and Recommendations

This study has found that the ParaDNA<sup>®</sup> Intelligence Test is not a suitable technique for screening sexual assault evidence in a hospital setting. Allele sharing and unequal sampling are two of the major weaknesses experienced by the ParaDNA<sup>®</sup> instrument that significantly affected the proportion of loci that suggest a mixture. Thus, non-experts should avoid using this instrument as a screening tool for better decision making until improvements can be made and further research is conducted. Although the ParaDNA<sup>®</sup> Intelligence Test greatly improved when the data was analyzed in a laboratory setting, the ParaDNA<sup>®</sup> Data Analysis Software should always be used when evaluating mixture profiles to prevent false negative conclusions. On the other hand, the RapidHIT<sup>®</sup> 200 outperformed the ParaDNA<sup>®</sup> Intelligence Test as a suitable technique for screening sexual assault evidence in any setting. This technique is comparable to traditional methods and provides superior information regarding the disposition of the sample. The results outlined in this study attempt to bridge the gap between blindly testing evidence after submission to a crime laboratory and gaining objective knowledge about the disposition of the sample to properly determine the best testing pathway. However, further research should be implemented prior to exploring either technique for screening sexual assault evidence regardless of the setting.

Future studies should attempt to discover better ways to screen sexual assault evidence in a hospital setting while combating the instruments' major weaknesses. One way this can be accomplished is by studying the ParaDNA<sup>®</sup> sample applicators' efficacy of evenly collecting cellular material. The manufacturer recommends direct sampling for one minute while applying moderate pressure, but it remains uncertain as to how many cells are being recovered on each prong. Testing different types of swabs on both the ParaDNA<sup>®</sup> and RapidHIT<sup>®</sup> 200 is also recommended for better release of cellular material since sample preparation is not required. Additionally, post-coital samples obtained at several intervals should be used to determine the suitability of the instrument in real case-type scenarios. Lastly, the ParaDNA<sup>®</sup> and RapidHIT<sup>®</sup> 200 instruments should be utilized by numerous participants to gauge any intra-subject variability that may arise.

## Chapter 5: Literature Cited

### Introduction

1. Campbell R, Feeney H, Fehler-Cabral G, Shaw J, Horsford S. The national problem of untested sexual assault kits (SAKs): Scope, causes, and future directions for research, policy, and practice. *Trauma Violence Abuse* 2015 Dec 23.
2. Campbell R, Patterson D, Lichty LF. The effectiveness of sexual assault nurse examiner (SANE) programs: A review of psychological, medical, legal, and community outcomes. *Trauma Violence Abuse* 2005 Oct;6(4):313-29.
3. Du Mont J, White D, World Health Organization, Sexual Violence Research Initiative. The uses and impacts of medico-legal evidence in sexual assault cases: A global review. 2007.
4. Ledray LE. Sexual assault nurse examiner, SANE: Development & operation guide. US Department of Justice, Office of Justice Programs, Office for Victims of Crime; 1999. .
5. Hall, A., & Ballantyne, J. (2003). Novel Y-STR typing strategies reveal the genetic profile of the semen donor in extended interval post-coital cervicovaginal samples. *Forensic Sciences International*, 136, 58-72.
6. Diegoli TM, et al. "An Optimized Protocol for Forensic Application of the PreCr™ Repair Mix to Multiplex Str Amplification of Uv-Damaged Dna." *Forensic Science International. Genetics*, vol. 6, no. 4, 2012, pp. 498–503., DOI: 10.1016/j.fsigen.2011.09.003.
7. Peterson J, Johnson D, Herz D, Graziano L, Oehler T. Sexual assault kit backlog study. Washington, DC: National Institute of Justice, Office of Justice Programs, US Department of Justice 2012.
8. Peterson JL, Hickman MJ. Census of publicly funded forensic crime laboratories, 2002. US Department of Justice, Office of Justice Programs, Bureau of Justice Statistics Washington, DC; 2005..
9. Scientific Working Group on DNA Analysis Methods. Recommendations for the Efficient DNA Processing of Sexual Assault Evidence Kits. National Institute of Justice 2016.
10. Smith, S.G., Chen, J., Basile, K.C., Gilbert, L.K., Merrick, M.T., Patel, N., et al. (2017). The National Intimate Partner and Sexual Violence Survey: 2010-2012 State Report. Atlanta, GA: National Center for Injury Prevention and Control, Centers for Disease Control and Prevention.

### Chapter 1.1 STR Analysis

11. "short tandem repeat." *Segen's Medical Dictionary*. 2011. Farlex, Inc. 23 Sep. 2018 <https://medical-dictionary.thefreedictionary.com/short+tandem+repeat>
12. "short tandem repeat." *Collins Dictionary of Biology*, 3rd ed., 2005. W. G. Hale, V. A. Saunders, J. P. Margham 23 Sep. 2018 <https://medical-dictionary.thefree dictionary.com /short+tandem+repeat>
13. Panneerchelvam, S and M N Norazmi. "Forensic DNA profiling and database" *Malaysian journal of medical sciences : MJMS* vol. 10,2 (2003): 20-6.
14. Fan, Hao, and Jia-You Chu. "A Brief Review of Short Tandem Repeat Mutation." *Genomics, Proteomics & Bioinformatics* 5.1 (2007): 7–14. *PMC*. Web. 23 Sept. 2018.
15. DNA Isolation Methods [Internet] *Encyclopedia: World of Forensic Science*; c2016 [cited 2017 June 18]. Available from: <http://www.encyclopedia.com/science-and-technology/biology-and-genetics/genetics-and-genetic-engineering/dna-isolation-methods>.
16. Butler, John M. *Advanced Topics in Forensic DNA Typing: Methodology*. 3rd ed., 3rd ed., Elsevier Science, 2011. *Elsevier Science and Technology*. <https://ebookcentral.proquest.com/lib/wvu/reader.action?docID=741334>. Accessed 30 Jan. 2019.

17. Rudin N, Inman K. An introduction to forensic DNA analysis. CRC press; 2001. .
18. "Product Information DTT, Molecular Biology Grade (Dithiothreitol, Cleland's Reagent)." *Document Connect*, Thermo Fisher Scientific Inc. , 2012, [www.thermofisher.com/document-connect/document-connect.html?url=https%3A%2F%2Fassets.thermofisher.com%2Fassets%2Fmanuals%2FMAN0013160\\_DTT\\_UG.pdf&title=VXNlciBHdWlkZTogRFRU](http://www.thermofisher.com/document-connect/document-connect.html?url=https%3A%2F%2Fassets.thermofisher.com%2Fassets%2Fmanuals%2FMAN0013160_DTT_UG.pdf&title=VXNlciBHdWlkZTogRFRU).
19. Barbisin, M., & Shewale, J.G. 2010. Assessment of DNA extracted from forensic samples prior to genotyping. *Forensic Science Review*, 22, 199-214.
20. Nicklas, J.A., & Buel, E. 2003. Quantification of DNA in forensic samples. *Analytical and Bioanalytical Chemistry*, 376, 1160-1167.
21. *Quality Assurance Standards (QAS) for Forensic DNA Testing Laboratories*. 2009. Available at <<http://www2.fbi.gov/hq/lab/html/testinglab.htm>>
22. Quantifiler™ HP and Trio DNA Quantification Kits User Guide. G. AppliedBiosystems, 2017. ThermoFisher Scientific Inc. <http://tools.thermofisher.com/content/sfs/manuals/4485354.pdf>. Accessed 5 Feb. 2019.
23. Thomas, Jacqueline T., et al. "Qiagen's Investigator™ Quantiplex Kit as a Predictor of STR Amplification Success from Low-Yield DNA Samples." *Journal of forensic sciences* 58.5 (2013): 1306-1309.
24. Bowyer VL. Real-time PCR. *Forensic Science, Medicine, and Pathology* 2007;3(1):61-3.
25. Forensic DNA: Human DNA Quantitation [Internet]: National Institute of Justice; c2014 [cited 2017 June 10]. Available from: <https://www.nij.gov/topics/forensics/evidence/dna/research/pages/human-quantitation.aspx>.
26. Friedman AL. Forensic DNA profiling in the 21st century. *Int J Offender Ther Comp Criminol* 1999;43(2):168-79.
27. Quick Reference: GlobalFiler™ PCR Amplification Kit--PCR Amplification and CE [Internet]; c2016 [cited 2017 06/01]. Available from: <https://tools.thermofisher.com/content/sfs/manuals/4477593.pdf>
28. Mullis KB. The unusual origin of the polymerase chain reaction. *Sci Am* 1990;262(4):56-61.
29. National Center for Biotechnology Information. PubChem Database. Formamide, CID=713, <https://pubchem.ncbi.nlm.nih.gov/compound/713> (accessed on Mar. 25, 2019)
30. "Applied Biosystems 3500/3500xL Genetic Analyzer User Guide." *Applied Biosystems | Hitachi, Life Technologies Corporation*, 2010, [tools.thermofisher.com/content/sfs/manuals/4401661.pdf](http://tools.thermofisher.com/content/sfs/manuals/4401661.pdf).
31. "Applied Biosystems 3500/3500xL Genetic Analyzer User Guide with 3500 Series Data Collection Software 3.1" *Applied Biosystems | ThermoFisher Scientific Inc.*, 2018, [100031809\\_3500\\_3500xL\\_Software\\_v3\\_1\\_UG.pdf](http://100031809_3500_3500xL_Software_v3_1_UG.pdf)
32. McLaren, R.S. et al. *Forensic Sci. Intl. Genet.* 2, 257–73. 2008.
33. Schumm JW. Why use a size marker and allelic ladders in STR analysis? *Profiles in DNA* 1997; 1:11-3.
34. Scientific Working Group on DNA Analysis Methods. Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories. SWGDAM 2017

## Chapter 1.2 Mixture Interpretation

35. McKiernan, H. E., and P. B. Danielson. *Chapter 21 - Molecular Diagnostic Applications in Forensic Science*. Ed. George P. Patrinos. Academic Press, 2017. Web.
36. Timken, Mark D., Sonja B. Klein, and Martin R. Buoncristiani. "Improving the efficacy of the standard DNA differential extraction method for sexual assault evidence." *Forensic science international: genetics* 34 (2018): 170-177.
37. Clayton, T. M., et al. "Analysis and interpretation of mixed forensic stains using DNA STR profiling." *Forensic Science International* 91.1 (1998): 55-70.



38. Gill, Peter, et al. "DNA commission of the International Society of Forensic Genetics: recommendations on the interpretation of mixtures." *Forensic science international* 160.2-3 (2006): 90-101.
39. Bowyer VL. Real-time PCR. *Forensic Science, Medicine, and Pathology* 2007;3(1):61-3.
40. Cotton, Robin W., et al. "Post-Amplification Artifacts." *Boston University Biomedical Forensic Sciences DNA Mixtures*, Boston University School of Medicine, <http://www.bu.edu/dnamixtures/7/>
41. Christian Schlötterer, Diethard Tautz, Slippage synthesis of simple sequence DNA, *Nucleic Acids Research*, Volume 20, Issue 2, 25 January 1992, Pages 211–215, <https://doi.org/10.1093/nar/20.2.211>
42. X.Y. Hauge, M. Litt, A study of the origin of 'shadow bands' seen when typing dinucleotide repeat polymorphisms by the PCR, *Human Molecular Genetics*, Volume 2, Issue 4, April 1993, Pages 411–415, <https://doi.org/10.1093/hmg/2.4.411>
43. P. Sean Walsh, Nicola J. Fildes, Rebecca Reynolds, Sequence Analysis and Characterization of Stutter Products at the Tetranucleotide Repeat Locus VWA, *Nucleic Acids Research*, Volume 24, Issue 14, 1 July 1996, Pages 2807–2812, <https://doi.org/10.1093/nar/24.14.2807>
44. Cotton, Robin W., et al. "Stutter." *Boston University Biomedical Forensic Sciences DNA Mixtures*, Boston University School of Medicine, [www.bu.edu/dnamixtures/8/](http://www.bu.edu/dnamixtures/8/).
45. Butler, John M. *Fundamentals of forensic DNA typing*. Academic press, 2009.
46. Globalfiler™ PCR Amplification Kit User Guide. AppliedBiosystems 2016. ThermoFisher Scientific Inc.
47. MUFSC Forensic Interpretation Guidelines-04.12.18
48. Bieber, Frederick R et al. "Evaluation of forensic DNA mixture evidence: protocol for evaluation, interpretation, and statistical calculations using the combined probability of inclusion." *BMC genetics* vol. 17,1 125. 31 Aug. 2016, DOI:10.1186/s12863-016-0429-7
49. Butler, John, and Simone Gittelsohn. "Basic STR Interpretation Workshop." ISFG. IFSG 2015, 27 June 2019, Krakow, Krakow.
50. Gilder, J.R., Inman, K., Shields, W. et al. *Int J Legal Med* (2011) 125: 87. <https://doi.org/10.1007/s00414-009-0411-2>
51. Paoletti, David R., et al. "Inferring the number of contributors to mixed DNA profiles." *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)* 9.1 (2012): 113-122.
52. Gill, Peter, et al. "Interpreting simple STR mixtures using allele peak areas." *Forensic science international* 91.1 (1998): 41-53.
53. Bill, Martin, et al. "PENDULUM—a guideline-based approach to the interpretation of STR mixtures." *Forensic science international* 148.2-3 (2005): 181-189.
54. Planz, John V. "Forensic Statistics." 15th International Symposium on Human Identification. 30 June 2019.
55. National Research Council (US) Committee on DNA Technology in Forensic Science. *DNA Technology in Forensic Science*. Washington (DC): National Academies Press (US); 1992. 3, DNA Typing: Statistical Basis for Interpretation. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK234541/>

### Chapter 1.3 ParaDNA® System

56. Blackman, Stephen, et al. "Developmental validation of the ParaDNA® Intelligence System—A novel approach to DNA profiling." *Forensic Science International: Genetics* 17 (2015): 137-148.
57. Hampshire, Toby. ParaDNA® Body Fluid ID Test. LGC 2017; 1-36.
58. Ball G, Dawnay N, Stafford-Allen B, Panasiuk M, Rendell P, Blackman S, Duxbury N, Wells S. Concordance study between the ParaDNA® intelligence test, a rapid DNA profiling assay, and a



- conventional STR typing kit (AmpFISTR<sup>®</sup> SGM plus<sup>®</sup>). *Forensic Science International: Genetics* 2015; 16:48-51.
59. French D, Howard R, Gale N, Brown T, McDowell D, Debenham P. Interrogation of short tandem repeats using fluorescent probes and melting curve analysis: A step towards rapid DNA identity screening. *Forensic Science International: Genetics* 2008;2(4):333-9.
  60. French D, Archard C, Brown T, McDowell D. HyBeacon<sup>™</sup> probes: A new tool for DNA sequence detection and allele discrimination. *Mol Cell Probes* 2001;15(6):363-74.
  61. HyBeacon<sup>®</sup> probes: Fast, accurate, reliable, detection of sequence variants. LGC Limited, 2014.
  62. French DJ, McDowell DG, Thomson JA, Brown T. and Debenham PG. 2006. HyBeacons<sup>®</sup>: A novel DNA probe chemistry for rapid genetic analysis. International congress series Elsevier. 707 p.
  63. Lyon E. Mutation detection using fluorescent hybridization probes and melting curve analysis. *Expert Review of Molecular Diagnostics* 2001;1(1):92-101.
  64. LGC. "ParaDNA Software Overview." 15 January 2018. Microsoft PowerPoint file.
  65. Foreman, Lindsey A., and Ian W. Evett. "Statistical analyses to support forensic interpretation for a new ten-locus STR profiling system." *International Journal of Legal Medicine* 114.3 (2001): 147-155.
  66. LGC. "ParaDNA Data Analysis Software Overview." 15 January 2018. Microsoft PowerPoint file.
  67. LGC. "Technical Overview of ParaDNA and Validation Studies." 15 January 2018. Microsoft PowerPoint file.

#### Chapter 1.4 RapidHIT<sup>®</sup> 200

68. ThermoFisher Scientific. *Introducing the New RapidHIT ID System. Introducing the New RapidHIT ID System*, Applied Biosystems, 2018.
69. IntegenX. *The Difference You Can Make In Minutes. The Difference You Can Make In Minutes*, ThermoFisher Scientific , 2018.
70. Le Roux D, Root BE, Hickey JA, Scott ON, Tsuei A, Li J, Saul DJ, Chassagne L, Landers JP, De Mazancourt P. An integrated sample-in-answer-out microfluidic chip for rapid human identification by STR analysis. *Lab on a Chip* 2014;14(22):4415-25.
71. Jovanovich S, Bogdan G, Belcinski R, Buscaino J, Burgi D, Butts EL, Chear K, Ciopyk B, Eberhart D, El-Sissi O. Developmental validation of a fully integrated sample-to-profile rapid human identification system for processing single-source reference buccal samples. *Forensic Science International: Genetics* 2015;16:181-94.
72. Grover WH, Skelley AM, Liu CN, Lagally ET, Mathies RA. Monolithic membrane valves and diaphragm pumps for practical large-scale integration into glass microfluidic devices. *Sensors Actuators B: Chem* 2003;89(3):315-23.
73. Jovanovich, S.B., I.I., Blaga. Microfluidic Devices. US Patent 7745207, June 29, 2010.
74. Hennessy LK, Mehendale N, Chear K, Jovanovich S, Williams S, Park C, Gangano S. Developmental validation of the GlobalFiler<sup>®</sup> express kit, a 24-marker STR assay, on the RapidHIT<sup>®</sup> system. *Forensic Science International: Genetics* 2014; 13:247-58.
75. Date-Chong M, Hudlow WR, Buoncristiani MR. Evaluation of the RapidHIT<sup>™</sup> 200 and RapidHIT GlobalFiler<sup>®</sup> express kit for fully automated STR genotyping. *Forensic Science International: Genetics* 2016; 23:1-8.
76. Workstations A. Walk-away magnetic bead-based DNA purification using the JANUS automated workstation...
77. Figarelli V, Press K [Presentation]. The Arizona DPS Rapid DNA Program. IntegenX; 2014.

### Chapter 3.1 Qualitative Approach

78. Regan, Stephanie, et al. [Poster Presentation] “Internal Validation and Implementation of the ParaDNA<sup>®</sup> Screening and Intelligence System.” LGC Limited (2017).

### Chapter 3.2 Linear Mixed-Effects Model

79. Winter, B. (2013). Linear models and linear mixed effects models in R with linguistic applications. arXiv:1308.5499.
80. Jelsema, Casey. Mixed Model Overview/demo. West Virginia University.
81. Field, J.C. “A Comparison of the Efficacy of Different Swab Types in the Absorption and Elution of Spermatozoa.” Boston University Theses & Dissertations. Boston University School of Medicine, 2013, <https://hdl.handle.net/2144/21150>.
82. Bates D, Maechler M, Bolker B, Walker S. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*. 67(1):1-48.
83. Adamowicz MS, Stasulli DM, Sobestanovich EM, Bille TW (2014) Evaluation of Methods to Improve the Extraction and Recovery of DNA from Cotton Swabs for Forensic Analysis. *PLoS ONE* 9(12): e116351. DOI:10.1371/journal.pone.0116351
84. Marshall, Pamela L., et al. "Evaluation of a novel material, Diomics X-Swab<sup>™</sup>, for collection of DNA." *Forensic Science International: Genetics* 12 (2014): 192-198.

### Chapter 3.3 Logistic Regression Model

85. Chandrayan, Pramod. “Logistic Regression For Dummies: A Detailed Explanation.” Medium, 2019, <https://towardsdatascience.com/logistic-regression-for-dummies-a-detailed-explanation-9597f76edf46>.