UNIVERSITY OF CENTRAL OKLAHOMA

JACKSON COLLEGE OF GRADUATE STUDIES

A PATH TOWARDS GENERAL ACCEPTANCE: THE FEASIBILITY OF MASSIVELY PARALLEL SEQUENCE TECHNOLOGY, NOMENCLATURE, AND STATISTICAL ANALYSIS IN FORENSIC CASEWORK

A THESIS

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

MASTER OF SCIENCE IN FORENSIC MOLECULAR BIOLOGY

By

TYLER PREWETT
Edmond, Oklahoma
2022

A Path Towards General Acceptance: The Feasibility of Massively Parallel Sequence Technology, Nomenclature, And Statistical Analysis in Forensic Casework

A THESIS APPROVED FOR THE
FORENSIC SCIENCE INSTITUTE

BY

Dr. James Creecy, Chair

Dr. Wayne Lord, Member

Dr. Matthew Parks, External Member

I dedicate this to my mother, Stacie McMillian. For whom I would not be the man I am today and who has been my rock and driving force.

Thank you, mom. I love you.

Acknowledgments

I want to thank Dr. James Creecy for helping me and guiding me throughout this project. To Dr. Matthew Parks and Dr. Wayne Lord, thank you for agreeing to be on my committee and for all the information they gave me regarding my thesis and the classroom. Dr. Dwight Adams, thank you for allowing me to follow my dream of learning forensic science. Thank you to the Forensic Science Institute faculty for everything I have learned and always being there for a friendly chat. Thank you to my family for pushing me forward and being there to be sounding boards. Thank you to all the Forensic Science Master's students who have gone through this journey with me; Whitney Mountain, Stephanie Harrison, Skye Wardell-Villarreal, Maddison Roberts, Kelly Riebesell, Rachael Kliewer, and Rebecca Brooks.

Table of Contents

List of Tables

List of Figures

Abstract

Throughout the short history of forensic DNA testing, advances in genetic analysis technologies have changed which gene markers are utilized and how these gene markers are analyzed. From Restriction Fragment Length Polymorphisms (RFLP) and major histocompatibility complex to Short Tandem Repeat (STR), each technological iteration changes how the loci are analyzed. In contrast, massively parallel sequencing (MPS) technology has reshaped the biological and forensic science fields over the last ten years. MPS technologies build off current STR analysis methods and permit multiple sequences to be analyzed simultaneously. This leads to fewer iteration changes and allows for the simultaneous analysis of STRs and single nucleotide polymorphisms (SNPs). Recently the STR Sequence working group (STRAND) has developed standardized nomenclature for this new type of DNA analysis and sequence data. STRAND's nomenclature uses bracketing repeats, relative positions to name the STR-SNPs, and standard STR repeat numbers. Along with the development of a nomenclature, MPS has been recently adjusted to include a statistical analysis procedure. This analysis combines the random match probability (RMP) of the STR repeats and the RMP of the STR-SNP to determine a likelihood ratio. This thesis work seeks to summarize the recent advances of MPS within forensic science and lay out the next steps for MPS following the formation of a standard reporting nomenclature and statistical analysis method. With MPS having the potential to be used in high-profile casework, complex mixtures, and highly degraded sample casework, MPS could become the next leading forensic DNA testing technology, especially through new developments including critical infrastructures, such as databases, and reduced run and technological costs.

Introduction

The history of forensic DNA analysis is only 38 years old, and in those 38 years, forensic DNA analysis has developed rapidly and possesses a rich history comprised of numerous changes. Forensic DNA analysis started in the early 1980s with Dr. Alec Jeffries's development of the DNA fingerprint using restriction fragment length polymorphisms (RFLP) analysis (1985a). While the start of forensic DNA analysis, RFLP was not the end. RFLP was quickly replaced by human leukocyte antigen DQ alpha (DQα), a new technology built on the understandings gained from RFLP and technological advancements at the time (Helmuth et al., 1990). DQα was also quickly replaced following further research by a forensic DNA technology developed due to studies in new innovative forms of analysis called short tandem repeat analysis. Short tandem repeats (STR) analysis was developed in the mid-1990s through breakthroughs in understanding the human genome (Lygo et al., 1994). STR analysis has continued to expand since the 1990s, and today has been the primary method used for forensic human identification. STR analysis has fixed the gaps in the previous generations of forensic DNA. However, it has started showing signs of issues with fulfilling the needs of forensic science as more complicated cases, such as cases with mixtures of multiple DNA sources, are appearing due to more sensitive equipment and a growing global population.

The intersection of DNA analysis and the courts has been the primary cause of the changes found in DNA analysis technology. Forensic science has always needed to fill the crossroads of scientific understanding and the court. Reporting information gained through scientific endeavors needs to be explained and defined so that judges, lawyers, and juries understand the methods and conclusions presented by forensic experts. The importance of accurate and effective reporting within the field of forensic DNA analysis was crystallized with

the release of the second National Research Council (NRC-II) report on forensic DNA evidence (National Research Council, 1996). The NRC-II report established that all forensic DNA analysis reports must use an established nomenclature and appropriate statistical methods. Nomenclature has been defined as the naming system used to report DNA analysis findings. Statistical analysis provides meaning to a conclusion based on the reported DNA analysis results. One of the greatest strengths of forensic STR analysis was the simple and concise nomenclature and the robust statistical analysis associated with the technology. This nomenclature and statistical analysis have made reporting DNA in court reliable and easy to understand. The nomenclature was developed so that it was easy to explain findings, and the statistical analysis helps further this understanding in the courts.

A new form of DNA technology has been developed, massively parallel sequencing (MPS). MPS analysis improves upon STR analysis, combining the genetic information gained from STR analysis with sequencing data from these exact genetic locations. This combination dramatically improves the amount of data acquired and strives to resolve some of the issues associated with STR analysis (Williamson, Laris, Romano, & Marciano, 2018). However, before MPS methods can be used in forensic casework, an effective nomenclature and statistical analysis method are required. Recently a nomenclature was developed by the STR Sequence working group (STRAND) (Gettings, 2022). As well as nomenclature, a statistical analysis method has been defined using the combinations of random match probabilities (RMP) of the STR and single nucleotide polymorphisms (SNP). With the recent formation of a standard reporting nomenclature and statistical analysis method, an outline of the following steps for MPS needs to be defined.

Background of Forensic DNA Analysis

The primary goal of forensic science has been to bridge the gap between scientific discovery and the legal community. Forensic science aids in establishing the circumstances of a crime, and DNA analysis was one of the most significant scientific breakthroughs in forensics to date. Since the inception of DNA analysis in forensic science, technological advancements have continued to push the field of forensic DNA analysis forward, and with each advancement, the techniques had to be organized in a manner understandable to the courts. Every advancement in DNA technology has added to what came before, but the critical elements of forensic human identification are conserved with each innovation. These conserved areas of forensic DNA analysis include a reproducible validated procedure, a method for the data to be visualized or verbalized in court so that it can be understandable to laypersons, a way for the data to be databased, and how the new technique improves the study of forensic DNA. For MPS, the nomenclature, the way the results are documented and presented in court, was one of the elements needed before MPS could be commonly used in forensic DNA. The nomenclature needs for MPS have been an area of discussion since 2016 by the International Society for Forensic Genetics (ISFG) (Parson et al., 2016) and again in 2019 by the STR STRAND working group (Getting, 2019). While the need for an accurate and effective nomenclature has been a priority of the Forensic DNA community for several years, the reality was that until recently, consensus on the matter was elusive.

The newest technology used in forensic DNA analysis is MPS, which has improved the amount of genetic information generated and has been determined to be a validated forensic technique. However, there are issues surrounding MPS technology that prevent it from being incorporated into casework. The primary issue was that the forensic DNA analysis field lacked a

way to discuss the data, a nomenclature that made sense to the jury, judge, lawyers, and civilians that would need to formulate conclusions about the DNA analysis. The gaps in MPS data reporting interfere with forensic science's primary function, bridging the court system and the scientific community. In addition, MPS raw data currently cannot be used in forensic databasing, so the developed nomenclature must also allow for databasing. The STR STRAND working group has recently advised a nomenclature at the 2022 American Academy of Forensic Science conference (Gettings, 2022). This nomenclature, if implemented, could lead to greater use of MPS in forensic science. To understand the importance of this nomenclature, an understanding of forensic DNA history, techniques, and nomenclatures needs to properly evaluate this advance in the context of the broader field of forensic DNA analysis. Including the different implementations of forensic DNA, how they met the requirements listed above, and a comparison to the technology that overtook them in the realm of forensic DNA.

DNA Fingerprinting and Restrictive Fragment Length Polymorphisms

The use of DNA in forensic science started in 1985 with Alec Jeffries et al. identifying the presence of "hypervariable microsatellite regions in human DNA." In the study, three microsatellites within the human genome were determined to have various versions of tandem repeats (1985a). These satellites were found using restriction enzyme probes to locate "cores" that were shared between individuals but had differences in the number of base pairs. In addition, it was determined that these satellites were inherited following basic patterns of Mendelian genetics. The inheritance of the satellites following Mendelian genetics means that the pattern inherited from the parents must be random, and the offspring has variations of the parents' alleles. The microsatellites identified were also polymorphic enough to be used for human identification. The definition of these microsatellites was then advanced in *Individual-specific*

fingerprints of human DNA. Finally, a larger population was used to test these hypervariable microsatellite regions, which showed that they were specific to an individual (Jeffreys, Wilson, & Thein, 1985b). With the microsatellites being both randomly inherited and polymorphic enough to be specific to individuals, identifying a specific person to the exclusion of all others was now possible.

DNA fingerprinting was first used in an immigration case in 1985, followed by a criminal case in 1986. The 1986 criminal case was the rape of two girls by Colin Pitchfork, where DNA analysis was the only avenue for finding the perpetrator. The Pitchfork case set DNA analysis up to become the standard method in forensic science for its ability to identify an individual (Gill, Jeffreys, & Werrett, 1985).

Technique for RFLP Uses Restriction Enzymes

In 1986 Jeffries et al. started looking into the segregation analysis of multiple hypervariable microsatellites (Jeffreys, Wilson, Thein, Weatherall, & Ponder, 1986). In this study, 34 genetically unlinked polymorphic minisatellite loci were found. DNA fingerprinting technology was then used to describe multiple locations where these polymorphisms could be found, changing the produced data. The type of analysis used in this study was called RFLP. For RFLP analysis to be successful, according to Jeffries et al., at least a "single drop of human blood" was needed (1985b).

To achieve results using this method, as shown in DNA "Fingerprints" and Segregation Analysis of Multiple Markers in Human Pedigrees, the DNA source, usually blood, must be diluted in a saline sodium citrate, sodium chloride (NaCl), trisodium citrate, and pH7 solution. Next, the solution must be either collected by centrifugation or nucleated cell plus nuclei pelleted by centrifugation (Jeffreys et al., 1986). RFLP used two different restriction enzymes, HinfI and

or Sau3A. HinfI was a restriction enzyme that cuts DNA between a G and A nucleotide in a GANTC DNA polydromic sequence, while Sau3A cuts before a G nucleotide in a GATC sequence. The HinfI restriction enzyme cleaves the DNA into two hybridized probes, the 33.15 and the 33.6, shown in Figure 1. The 33.15 probe was a cloned human minisatellite with 29 16-base pair (bp) core sequence repeats. In comparison, the 33.6 probe was a diverged trimer with 11 bp of the 3' end of the base pair that repeated three times. The difference in the sequence of the probes and the repeat lengths was how this system could determine identification (Jeffreys et al., 1986). The repeat lengths are found in specific frequencies in the human population. The frequencies can then be statistically analyzed, and a probability can then be determined of how likely it was for a set of repeat lengths to be found randomly in a population (Budowle et al., 1991). To determine the repeat lengths, first, they must be visualized.

The restriction digests were added to a gel loading mix and ethidium bromide for the Southern blot. The mixture was loaded into a horizontal agarose gel and run through electrophoresis until all fragments less than 1.5 kilobases long run off the gel. Then the mixture was then loaded onto a nitrocellulose filter with a labeled probe from the 33.6 and 33.5 recombinants. The resulting DNA fragment pattern can be transferred by blotting. The resulting fragments are then visualized depending on the loading mix, using either an X-ray or an

Figure 1- Sequence of the HinfI restriction enzyme Core and Probe: The figure above shows the core sequence, 33.15 repeat sequence, and the 33.6 repeat sequences described in Jeffries et al.'s study DNA "Fingerprints" and Segregation Analysis of Multiple Markers in Human Pedigrees

alternative light source. The resulting fragments are then compared to a ladder standard. This ladder standard was used to determine the length of the fragments.

RFLP Technique's usage in Court

The separation of DNA fragments by gel electrophoresis was essential to RFLP analysis. The bands of various lengths were separated, and the size of individual DNA fragments was determined using known DNA size markers that were run simultaneously. The fragment lengths observed would then be compared to the frequency at which the lengths appeared in the population. The population statistical analysis would then be conducted and communicated in court (Budowle et al., 1991).

RFLP used a fixed bin approach to determine the statistical relevance. The fixed bin approach used a standard where alleles of similar size would be binned together. This would then be compared to the total number of alleles found in a sample population that also fell within the bin. Thus, a frequency was determined. For RFLP, any allele from an unknown sample's frequency was determined via these bins. If an allele bled through multiple bins, whichever bin it fell in the most was used. In this way, the alleles themselves did not matter, just their size, and an overestimate for frequency could be determined (Budowle et al., 1991).

When RFLP was the standard for forensic DNA used in courts, there were two admissibility tests that forensic DNA analysis had to meet: Frye v United States and Rule 702. Under *Frye,* scientific evidence was only admitted into the court of law if experts generally accepted the technique in that field ("Frye v. United States," 1923). While Rule 702 states that:

> *A witness who is qualified as an expert by knowledge, skill, experience, training, or education may testify in the form of an opinion or otherwise, if the expert's scientific,*

> *technical, or other specialized knowledge will help the trier of fact to understand the evidence or to determine a fact in issue (Testimony by Expert Witnesses, 1975).*

With these standards, forensic DNA started to be introduced into the courts in the United States in the late 1980s. With these statutes, all forensic evidence in court had to meet the standards of Frye and rule 702. For evidence to be accepted under Frye, the techniques used and how the evidence was processed must be generally accepted within the scientific community. In this context, the technique was used in most labs and had adequate publications showing that the technique works as presented, while rule 702 was based on the role of the expert's testimony. This rule finds that an expert's opinion can be sought if it will help understand what the evidence and the analysis mean to help determine the facts of the case.

However, problems arose almost immediately regarding RFLP meeting these standards. In *People v Castro* (1989), the trial court of New York found that while the DNA evidence was generally accepted and reliable, the technique applied in that case was flawed, and the evidence was inadmissible. The court also found that the laboratory's population frequency databank could not accurately estimate the likelihood of identifying the suspect. In 1989 the Supreme Court of Minnesota was the first appellate court to reject the use of DNA evidence during the *State v Schwartz* Case (1989). The Schwarts Case was another case where the laboratory failed to show that it met standards, specifically falsely identifying samples, and had not met relevant, validated protocols set by the FBI. However, in *Cobey v State* (1987), the Maryland court found that the same lab in the *Schwartz* case did meet standards and that the DNA evidence for this case was admissible. For rule 702, RFLP had the same issues where it was admissible in some courts but failed to meet the standards in others. One of the issues found often was the presence of erroneous bands of DNA that could not be accounted for, see Figure 2. With RFLP having

encountered issues with its ability to meet the standards at the time, shown through its significant court losses, research into new forms of DNA testing was being investigated. RFLP's issues with the court show that a decisive nomenclature was needed for forensic DNA. Without strong nomenclature, issues associated with quantifying the data through population statistics and describing complications were frequently encountered. Along with nomenclature issues and the limited variability in the RFLP locus, a new DNA technology took over, with highly variable regions Major Histocompatibility Complex, Class II, DQ Alpha (DQα) became the new forensic DNA technology.

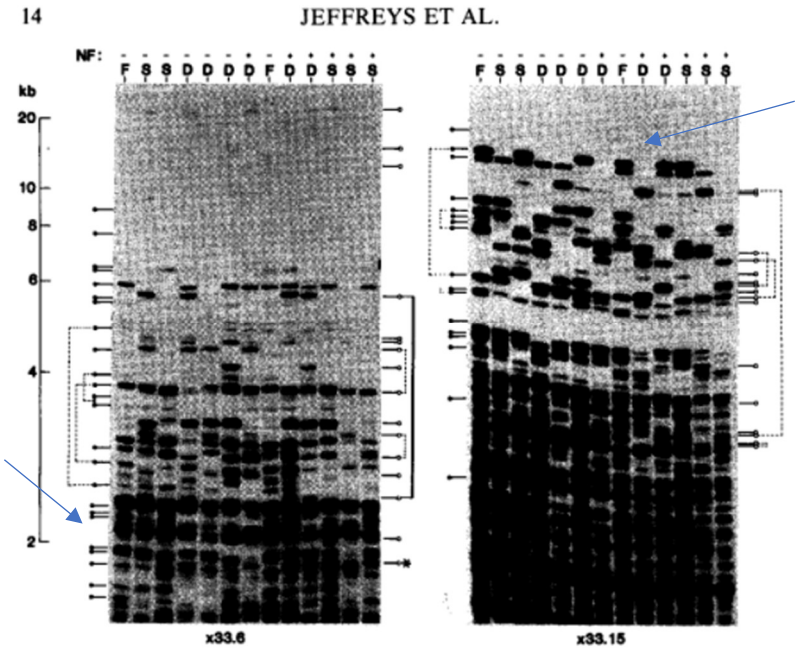


Figure 4 - Gel used in RFLP: The figure to the left shows two gels that demonstrate what an RFLP run looks like. These gels were used in kinship testing to test for neurofibromatosis (Jeffreys et al., 1986). The gels have unique issues. The first gel had unusual banding in the bottom part of the gel, while the second had a curve to the positioning of bands. Depending on the court, both issues could lead to the DNA results being inadmissible.

Major Histocompatibility Complex, Class II, DQ Alpha (DQα)

Major histocompatibility complex, class II, DQ alpha (DQα), was a known area of DNA polymorphism found in the human population, and this genetic location was used for a short time for forensic DNA analysis. The DQα generation of forensic DNA analysis directly followed the RFLP generation. The advent of the polymerase chain reaction (PCR) was a major scientific

breakthrough, and this development allowed for DQα to become a valid form of forensic DNA identification (Helmuth et al., 1990). With PCR, a smaller sample was needed, and DQα was found to have many variations than in the previous RFLP regions. These changes led to the induction of DQα as the new method of forensic DNA identification.

Polymerase Chain Reaction

PCR was a breakthrough in DNA forensics that has continued to be used in forensics science (Mullis and Faloona, 1987). PCR allowed a single area of the human genome to be copied millions of times, thereby allowing for better detection with a smaller amount of start DNA. The DNA amplification process was accomplished using forward and reverse primers. Primers are short fragments of DNA that are synthesized in the lab and are complementary to the sequences of interest. These primers designate what parts of the genome will be copied. After the primers bind to the genome via complementary base pairing, Taq DNA polymerase was used along with free nucleotides to generate a copy of the desired region of the genome. The PCR method has been incorporated into all forms of DNA testing since the 1990s when it was recognized that less starting DNA would be needed. Whereas RFLP without PCR needed large puddles of blood or large amounts of semen, would be needed.

HLA-DQα Technique and How the Results are Reported

DQα uses PCR to amplify six genetic loci. The loci are tagged by primers that inform the DNA copying mechanics, DNA polymerase, and what areas to copy. The loci are then copied multiple times. The solution containing the copied DNA was then pipetted onto a test strip. These test strips have multiple areas with different enzymes. Each enzyme will react with a different allele sub-type of the DQα locus. When the enzyme reacts, a color change happens on the test strip. The first test strips for the presence of the DQα locus while the other tests for the

other five loci. The strip for the five loci will indicate whether each locus was present. The DQα strip will test which allele of the DQα loci was present. These loci were named from one to four; there were also subtypes for DQα one and DQα four, as shown in Figure 3. Thus, the nomenclature for DQα was determined by which area on the strip had a color change; Figure 3 would be a 1, 2, 1.1, 1.2, 4. The frequency would then be determined based on a population frequency for each area that observed a color change (Menevse, Ulkuer, 1995). With the continued use of DNA in forensics, the frequency of alleles found in populations became an issue for DQα. More DNA was analyzed, and more alleles with more stable variations were needed to continue identification. DQα was found not to have enough variations in specific populations (Sullivan, Gill, Lingard, Lygo, 1992). Statistics were also difficult to calculate due to allele dropout and the inability to determine if two or more individuals contributed to the DNA sample. DQα was a short-lived DNA technology with very few studies being done using the technology. However, the use of PCR has continued in subsequent DNA technologies.

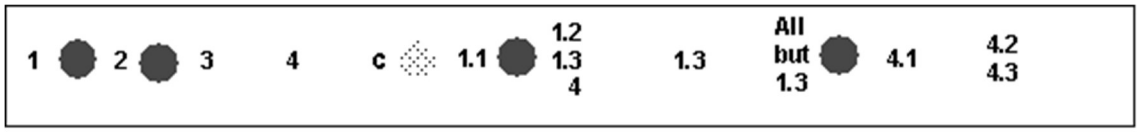


Figure 5 - DQα test strip: The figure above shows a DQα test strip that shows how the DQα analysis was complete with this individual having DQα 1,2,1.2,1.3,4, and 4.1 variations of the DQα loci

Short Tandem Repeat Analysis

Continuing the trend of new advances in genomic understanding leading to new forensic DNA technologies was the development of STR via capillary electrophoresis (CE) in the mid-1990s, which has since become the standard used worldwide (Tautz, 1989; Lygo et al., 1994). One of STR's greatest strengths was its robust statistics, primarily due to the many allelic

variations tested when running an STR analysis. With a more significant variation, the limitations of the previous DNA technology were surpassed. In previous generations of DNA technology, the statistical analysis was limited to determining discrimination power with low yields due to the low number of alleles that could be tested at once. With 21 or more loci utilized in forensic STR analysis, limited allelic variation was overcome, and statistical estimations of DNA profile rarity can be found many times greater than in the human population.

STRs are genetic loci found in the non-coding areas of all genomes and are made up of repeating strings of nucleotide sequences. These repeating sequences are conserved in the number of repeats passed from parent to child. STRs are arranged into three groups: simple STRs, which have identical lengths and sequences of repeats; compound STRs with two or more simple repeats; and complex STRs have several repeats of varying lengths and intervening sequences, as shown in Table 4 (Weber, 1993). The highly polymorphic nature of STR loci results in an analyst's ability to determine human identity based on the outcome of the analysis process (Frégeau, 1993). Capillary electrophoresis with STRs was the process used to determine each STR's length and can be used to identify a DNA's STR strains lengths to standards to determine the repeat number. Now, most forensic labs in the United States use STR analysis with

Categories for STR Markers

Category	Example Repeat Structure	13 CODIS Loci
Simple repeats – contain units of identical length and sequence	(GATA)(GATA)(GATA)	TPOX, CSF1PO, D5S818, D13S317, D16S539
Simple repeats with non-consensus alleles (e.g., TH01 9.3)	(GATA)(GAT-)(GATA)	TH01, D18S51, D7S820
Compound repeats – comprise two or more adjacent simple repeats	(GATA)(GATA)(GACA)	VWA, FGA, D3S1358, D8S1179
Complex repeats – contain several repeat blocks of variable unit length	(GATA)(GACA)(CA)(CATA)	D21S11

These categories were first described by **Urquhart *et al.* (1994) *Int. J. Legal Med.* 107:13-20**

Table 1 - Categories of STR Markers: The figure to the right is a table of the categories of STR Markers. The figure shows the categories of different STRs with the category of the STRs on the left-hand side, an example structure in the middle, and examples of 13 CODIS loci on the right (Butler, 2007b)

CE (Gettings, Aponte, Vallone, & Butler, 2015; Gettings et al., 2017; Zhang et al., 2018) for the individualization of forensic evidence and comparison with known reference samples. This technology was also the base for which the next generation of DNA technology builds via sequencing of the STRs.

Technological Advances behind Short Tandem Repeats

The understanding of the nature of DNA and how CE could be used in forensic DNA led to the development of STR technology. CE was developed in the 1930s primarily for chemistry use (Wätzig, Degenhardt & Kunkel, 1998). Due to CE's use of electrical currents, it was found that due to the charge inherent in DNA, the technology could also be used in forensic DNA to separate and identify different DNA segments with consistent results through runs. Capillary electrophoresis allowed forensic DNA analysis to move away from gel electrophoresis, which results were impossible to recreate on multiple runs. With capillary electrophoresis, no matter how many times a DNA profile goes through the instrument, the results will look the same, with known error rates. With STR via CE, multiple loci from one DNA source can be investigated simultaneously. CE uses a tube with a gel matrix to separate DNA by length into known patterns by forcing the DNA through with an electrical current. These separated strands can then be detected and reported. The longer the DNA strands, the longer it takes to travel through the capillary.

Along with differentiation by length, DNA can also be tagged through the amplification portion of PCR with different fluorescent tags. The areas or loci containing the STR alleles are copied during PCR using primers, with one primer adding a fluorescent tag. These primers signal where DNA was copied. In the United States, there are five different fluorescent tags used through PCR, with each color used to tag up to 5 different STRs. These fluorescent tags can be

identified with the camera used in a CE instrument and the different lengths. Thus, the CE instrument can run all the STRs simultaneously if there are sufficient fluorescent tags and differences in the lengths of the STR loci. With the instrument's ability to run large numbers of STRs, there needed to be a standard developed to categorize the DNA strands. Thus, a DNA ladder was developed (Pures, 1993). The DNA ladder can identify what allele a DNA strand was based on the length of the strand.

Along with the scientific advances, changes in the law also allowed for the advancement of STR over DQα. In the case of Daubert vs. Merrell Dow Pharmaceuticals, Inc (Daubert) in 1993, the rules of evidence changed in the United States. Daubert ruled that all evidence must meet five rules. These rules are 1) that the theory or technique employed by the expert was generally accepted in the scientific community, 2) the theory or technique has been subjected to peer review and publication, 3) the technique can be and has been tested, 4) the technique has a known error rate, and 5) the research was conducted independent of the litigation or dependent on an intention to provide the proposed testimony (Daubert, 1993). DQα failed to meet the new Daubert rules, while STR was shown to meet all Daubert's factors leading to STR becoming the standard DNA testing technique in the United States.

Short Tandem Repeats have an easy-to-understand nomenclature

The nomenclature associated with forensic STR analysis uses the information from CE to designate length values to the STRS. The CE instrument's software translates the camera information and compares the DNA strand length and the specific fluorescent tag to an allelic ladder. An allelic ladder was a set of information from the developer of the specific STR PCR primer set that tells the software what STRs are tagged, what color, and what lengths can correspond to STRs (see Figure 4). Each STR will have a specific range of DNA lengths that the

STR will fit in, called a bin (Frégeau, 1993). The information gathered from the camera and the information from the ladder will be interpreted by the software, producing a graph called an electropherogram.

The electropherogram has the DNA strands for each STR allele designated as peaks, with each peak labeled as a certain repeat number; this repeat number was based on a size standard of known sequence. The size standards were developed with DNA strand lengths of known repeat numbers (Griffiths, 1998). Thus, a relationship between the length of DNA can be used to determine the repeat number of the STRs. The determination that a relationship between length and repeat number allowed for a nomenclature to be developed where forensic analysts could use repeat numbers in court instead of describing lengths. This allowed for a more understandable testimony than one that was based on the number of nucleotides found in alleles. With this easier-to-understand nomenclature, the way the statistics were developed was also more understandable to the layperson.

Forensic STR analysis has a solid statistical analysis system using frequency-based statistics and the repeat numbers of the STRs. A statistical probability using the Hardy-Weinberg Principle can be done using the frequency data of each allele's repeat number. For each locus, if one peak were detected, the frequency would be squared (p^2), and if two peaks were detected, each peak's frequency would be multiplied and then by two (2pq). Then the frequency of all the alleles was multiplied together. The frequency calculation could then be converted into a likelihood ratio showing the likelihood of a specific profile being found in the population at random. With solid statistics and an easy-to-understand nomenclature already used for casework and databasing, STR analysis remains the standard for DNA analysis (de Knijff, 2019).

.

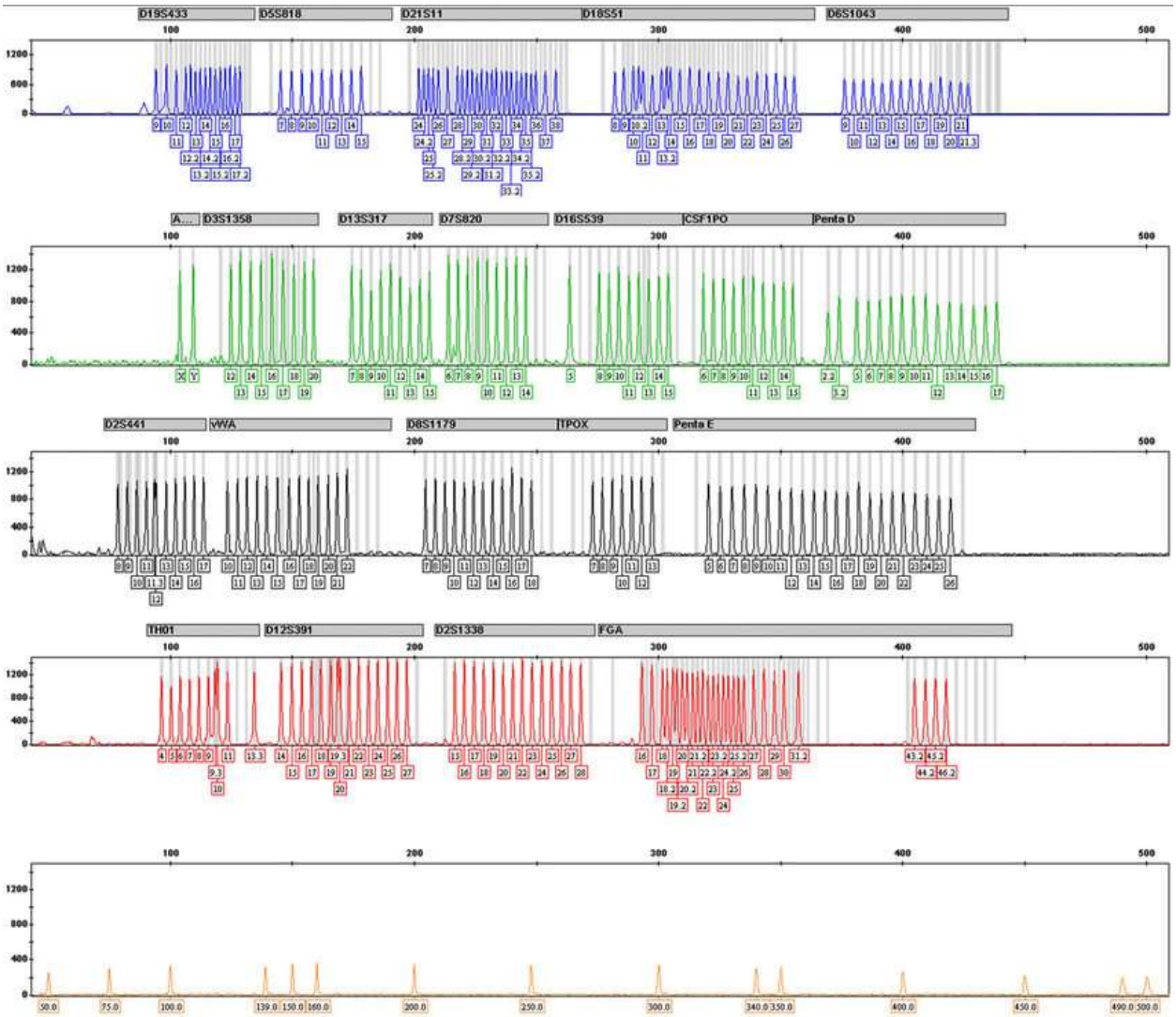


Figure 4 - Allelic Ladder: The above figure is that of an allelic later. The figure has five rows; the first four rows correspond to a color of a florescent tag used when copying the DNA during PCR. The final row is a size standard. (Yang et al., 2016). When a DNA profile is run, the length and fluorescent tag for the DNA strands are compared to this ladder. The CE instruments software will convert that information into the repeat number for the individual alleles. The DNA strands' available lengths for each STR are shown below the STR title.

Limitations of STR Analysis

STR has been the standard for forensic DNA analysis for over 20 years. This was due to its detection of variation in allelic repeats, high power of discrimination, and the standard format for running the samples and reporting results. However, STR analysis was not without its

drawbacks. These include artifacts often found during runs, difficulty distinguishing profiles in mixtures, inability to distinguish repeats of the same length but with differing sequences, and difficulty producing data from degraded DNA samples. Unfortunately, after 20 years of research and development, these weaknesses of STR analysis do not appear to be resolvable using CE techniques.

Another STR drawback includes the limited number of STRs that can be multiplexed simultaneously and the number of sequences amplified (Chamberlain & Chamberlain, 1994). Many commercial kits have been developed for STR amplification (Zhang et al., 2015). However, one of the most significant limitations of STR via CE was that less than 30 loci could be multiplexed at once (Jäger et al., 2017), with one study being able to multiplex 25 core STR loci at a time (Zhang et al., 2015). In the Zhang et al. (2015) study, an STR kit was developed for a specific group of individuals belonging to the Han population in China. The STRs used were grouped on the electropherogram so closely that any off-ladder would be designated the wrong loci. An off-ladder was caused by alleles mutations that lead to the allele being out of the range, or bin, of the STR allelic ladder. Off-ladder STRs were caused by adding a nucleotide to the STR loci. In the case of the Zhang et al. study (2015), an off-ladder could lead to an STR peak being mislabeled as an entirely different STR, as shown in Figure 5. With a growing global population, the need for either more STRs or a new way of analyzing the STRs was needed.

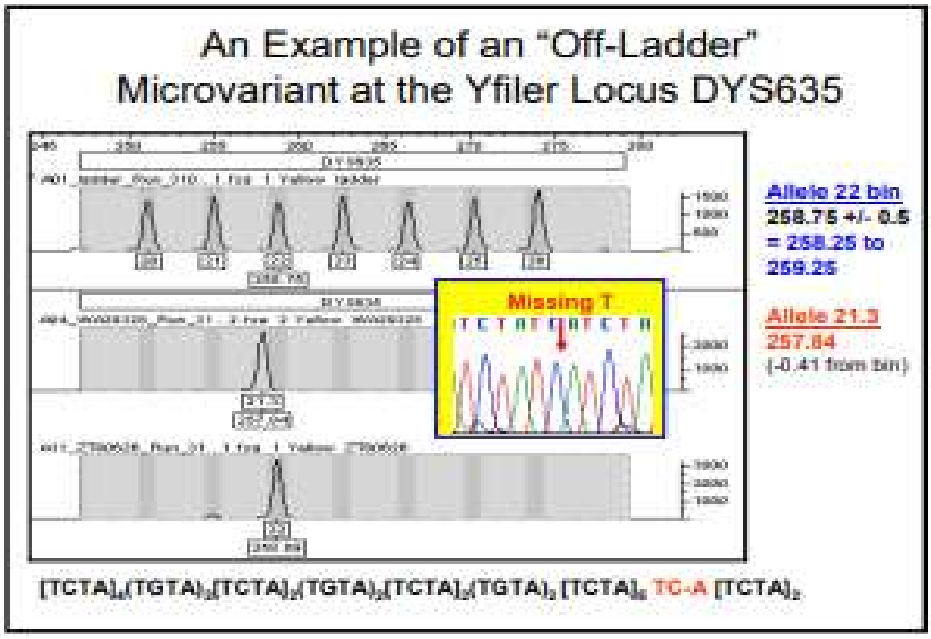


Figure 5 – Example of an Off-Ladder: The figure to the left depicts an off-ladder. The off ladder is shown on an electropherogram and the reason for the off-ladder is shown in the yellow box (Butler, 2007b).

Massively Parallel Sequencing

The newest advance in forensic DNA analysis is MPS which uses the base of STR loci and sequences them all at once. However, only one MPS case has gone to court ("Verogen Inc.; First Criminal Conviction Secured with Next-Gen Forensic DNA Technology," 2019). MPS technology has been researched since 2010 in clinical genetics, and tens of thousands of samples are routinely screened annually using either the entire genome or targeted genome sequencing through MPS (de Knijff, 2019; Koboldt et al., 2013). However, there are still some hurdles for MPS. MPS must be generally accepted, peer-reviewed, have known error rates, and be validated to meet the Daubert factors for use in court. The MiSeq® FGx™ Forensic Genomics System has met all the requirements.

In forensics, MPS is helpful in the sequencing of STRs. These years of study have allowed for the multiplex sequencing of STRs and Single Nucleotide Polymorphisms (SNPs)

detection. MPS technology can sequence multiple STR repeat regions, including their flanking region, and put several samples together for higher throughput using the barcodes in the MPS system. It also has application to determine additional variation between individuals' sequences. As with standard sequencing, only one STR can be done simultaneously, and in STR, multiple repeats and flanking regions are copied; however, they are not sequenced. Flanking region SNPs, also known as STR-SNPs, have previously been studied and have begun to be studied again to see if they are used to improve the power of discrimination and in mixture deconvolution, as seen in Table 2 (Williamson, Laris, Romano, & Marciano, 2018).

One of the most exciting prospects of MPS is its ability to deconvolute mixtures (Oldoni, 2019). Mixtures are an issue in forensic science when analyzed through STR. This is due to the difficulty of analyzing STR data to determine which person is contributing which peaks in mixtures, especially when two people share a single STR peak. With its ability to identify sequence variation, MPS can allow for distinguishing mixed DNA with the same STR repeat lengths. This can be done in cases where one contributor has a difference in sequence compared to the other with the same amount of repeats. Using the STR process for deconvolution and adding the sequencing variations of STRs with the exact repeat, an analyst can deconvolute complex mixtures. If each contributor has a difference in sequence for each allele, the sequence variation could be used to eliminate suspects or to eliminate the victim's DNA to have a single source of DNA in cases of DNA mixtures of two or more individuals (Oldoni, 2019).

Allele Name	STR Frequency	STR Sequence Frequency	Flanking Region Sequence Frequency
D3S1358	10	22	0
vWA	11	35	3
D16S539	8	19	8
CSF1PO	9	11	0
TPOX	9	9	0
D8S1179	12	30	1
D21S11	22	85	0
D18S51	21	37	9
DYS391	4	4	0
D2S441	14	24	2
D19S433	17	23	1
TH01	9	10	1
FGA	26	34	0
D22S104	11	17	7
D5S818	9	18	0
D13S317	9	26	2
D7S820	11	26	13
D10S124	9	10	0
D1S1656	18	34	1
D12S391	19	79	1
D2S1338	14	64	0

Table 3 - Allelic Differences found in STR, STR sequencing, and Flanking Regions: The table above uses data from a study by Novoroski et al. (2016) that shows the STRs and the different amounts of variation in three DNA analyses; STR via CE, STR sequencing, and flanking regions of each STR.

Massively Parallel Sequencing Technology

MPS analysis improves upon the number of areas of DNA, loci, and SNPs that can be multiplexed than by STR technology. Two forms of MPS technology have been developed for use in forensic DNA, as seen in Figure 6. One MPS technology uses fluorescent tagging nucleotides (Reversible Dye Terminator). Thus, as a nucleotide was added to a sequence, the exact nucleotide could be determined. The other method of MPS was releasing energy as a

nucleotide was added to a DNA strand (pyrosequencing). Thus, one single type of nucleotide was added at a time; if the nucleotide were the next in the sequence when added to the sequence, it would release energy that could be converted into light through enzymes. The light released by the enzyme can then be sensed by a camera that will then develop a peak. The peaks can then determine the DNA sequence (Michael, 2009). The reversible termination revealing sequencing has a lower error rate for the two MPS methods. Pyrosequencing has more errors due to the homopolymers, multiple of the same nucleotide in a row. Pyrosequencing can add or skip a nucleotide if there are too many of the same in a row. MPS data has also been more helpful in deconvolution and analyzing degraded samples than STR fragment analysis methods

MPS ability to simultaneously sequence STRs and SNPs is advantageous in analyzing challenging samples where DNA might be in low quantity and degraded. Votrubova et al. (2017) compared samples with previously undergone STR-CE in different cases at different labs. The samples varied in age and amount of degradation. This study found that MPS was more suitable for the degraded samples' genetic characterization due to its ability to obtain this information without consuming the limited sample further. At the same time, CE was slightly better at genotyping more STR loci.

MPS can discriminate between DNA sequencing of the same length but differ in sequence structure caused by mutations over multiple generations. The underlying DNA sequence variations of STRs can be determined through MPS, as shown in Figure 7. This data can then be translated into repeat numbers. The underlying sequence of STRs was becoming more necessary due to the number of profile mixtures collected using STR-CE. These underlying sequences can be used to deconvolute these mixtures. Complex STR locations were considered good candidates for MPS STR variate analysis because these STRs interrupted repeat sections

known to accumulate point mutations. Average mutation rates of the STRs loci are estimated at between 0.01% and 0.64 %, according to STRbase (Ruitberg, 2001).

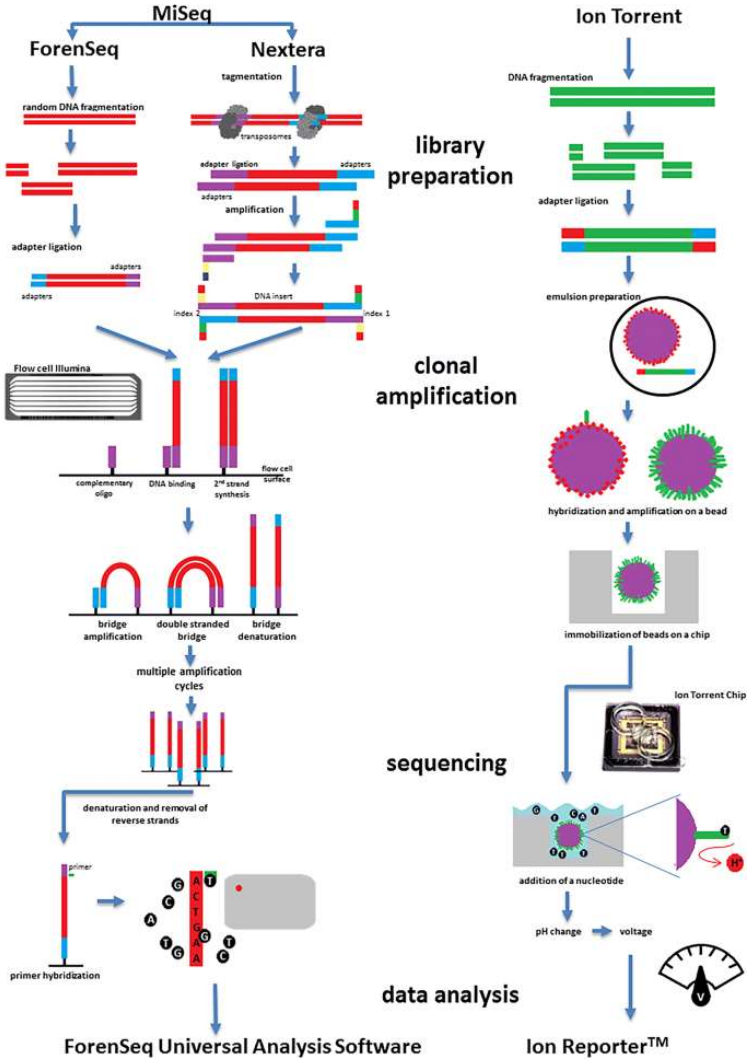


Figure 6 - The Difference Between Reversible Dye Terminator and Pyrosequencing: The figure to the left depicts the difference between reversable dye terminator sequencing shown through the Miseq flow chart and that of Pyrosequencing shown in the Ion Torrent flow chart (Ballard et al. (2020).

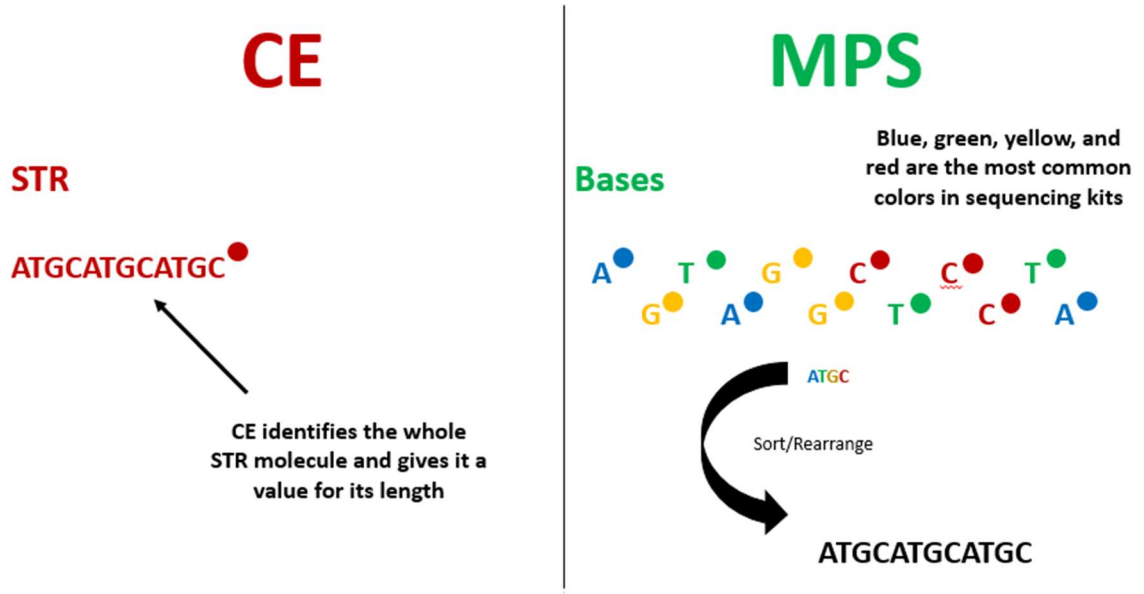


Figure 7 - Comparison of STR to MPS: The figure above shows the comparison information gained from STR, shown on the left as CE, to that of MPS, shown on the right (Dai et al., 2019).

Massively Parallel Sequencing Nomenclature

The standardization of nomenclature was a defining point in forensic DNA. Going from images and DNA strips to verbalization of allele numbers and now the need to be able to describe sequencing data. The standardization of nomenclature has been necessary for DNA analysis to fill the gaps of understanding between the scientist and the courts. The STRAND working group discussed a standard MPS nomenclature at the American Academy of Forensic Sciences conference in 2022 (Gettings, 2022). Before the STR STRAND nomenclature within forensic studies, the nomenclature for MPS was not standardized, leading to confusion about reported alleles. For example, the strand reporting difference may have occurred due to the forward strand matching the reference sequence, but the reverse strand was sequenced for some

STRs. Also, some STRs may, over time, have differences in reporting where the repeat begins (Gettings et al., 2017).

Since the introduction of MPS, a nomenclature for sequence differences in STR alleles has been needed. A study by Gelardi et al. (2014) came up with a naming system; this study recognized the need to specify the sequence variation in repeat sequences with the same repeat number and include SNPs of the STRs flanking sequence. To achieve this, they used a nomenclature with four elements. 1) a locus name commonly used in forensic science, 2) the length of the repeat region divided by the number of the repeat units, 3) the sequence of the sub-repeats followed by the number of sub-repeats, 4) variations in the flanking region using either the rs number and base call or chromosome position on the variation, as seen in Figure 8. However, the study found no SNP variations in the alleles in question, so there was no example of the fourth part of the nomenclature.

A study by Parson et al. (2016) works toward determining a naming system for the sequence differences in the flanking regions of STRs so that the forensic science community can adopt it for human identification. The nomenclature that Parson et al. proposed was the locus name followed by the CE sequence description, then the chromosome and reference sequence used, followed by the STR repeat coordinates, then the STR motifs, and finally, the flanking region variants. However, their study states that a more succinct naming system will need to be developed if MPS is in everyday use.

CE allele name	SGS allele name	Count	Frequency
14	D3S1358[14]TCTA[1]TCTG[3]TCTA[10]	1	0.0025
14	D3S1358[14]TCTA[1]TCTG[1]TCTA[12]	1	0.0025
20	D3S1358[20]TCTA[1]TCTG[4]TCTA[15]	1	0.0025
18	D12S391[18]AGAT[11]AGAC[7]	3	0.0076
18	D12S391[18]AGAT[12]AGAC[5]AGAT[1]	4	0.0102
19	D12S391[19]AGAT[11]AGAC[7]AGAT[1]	1	0.0025
19	D12S391[19]AGAT[11]AGAC[8]	6	0.0152
20	D12S391[20]AGGT[1]AGAT[10]AGAC[9]	2	0.0051
21	D12S391[21]AGAT[13]GGAC[1]AGAC[7]	1	0.0025
21	D12S391[21]AGGT[1]AGAT[11]AGAC[9]	1	0.0025
22	D12S391[22]AGAT[13]AGAC[8]AGAT[1]	7	0.0178

Figure 8 - Gelardi et al. (2014) MPS Nomenclature: The Figure above shows the nomenclature devised by Gelardi et al. (2014) for MPS; the figure has the STR name on the left followed by their proposed nomenclature followed by how often it was found and the frequency. The study however found no SNP differences so the SNP nomenclature is not shown.

The nomenclature that these studies formulated has a few issues. These include the length for naming differences found in the flanking region sequence, which can become very complicated due to the mixing of rs numbers and location values for SNPs. The system described by Parson et al. (2016) includes the numbering of all the bases and identifying the base ranges of the STRs based on this numbering system. Parson et al. argue that a complete human genome reference sequence should be used and that any found differences be named based on the reference. This was like mitochondrial DNA nomenclature and was already a familiar naming system in the forensic community. In a study by Zhang et al. (2018), data was compared to the

GRCh38 complete human genome reference sequence, while in a study by Gelardi et al., the sequences were compared to the GenBank reference sequences. Finally, in a study by Avila et al. (2019), the samples were compared to Hg19. The problem with this was that human genomic references are being updated continuously. Thus, the base ranges of STRs will need to be updated with each reference sample's reevaluation. Also, with each update, all previous DNA will need to be sequenced for further matches, which will need to be done for any nomenclature.

STRAND working group used the information from the previous studies to formulate the best nomenclature moving forward. A bracketed repeat nomenclature was developed on the ideas from Parson et al. (2016). The repeat region was "condensed into a descriptive, human-readable format" in this nomenclature" (Gettings, 2019). The repeat regions are described by the repeated nucleotide code, followed by a number that designates the number of times that sequence was repeated. This bracket format was then repeated with the different nucleotide codes in the repeat region in the order they appear. This condensed bracket will then be followed by the SNPs found in the flanking regions via a positive or negative with the distance from the repeat. Meaning that if a nucleotide change were 15 nucleotides before the sequence, it would be designated -15; if it were after the repeat, it would be +15. This number would then be followed by what the change was. If it were a c to an a, it would be a c>a designation. If there was more than one SNP for an allele, these are added as further positive or negative numbers. These differences would be found by comparing the sequence to a known one, as shown in Figure 9. Now that MPS has a nomenclature, there are fewer barriers to use in forensic DNA.

STR STRAND working group has developed a nomenclature that will become the nomenclature standard for MPS. This nomenclature builds off the previous STR nomenclature while allowing for the additional information found through MPS. Some of the issues MPS faces

is that the nomenclature has not been set across the board. Other issues are that there was not enough information on the number of sequence variations found in either the repeat region of STRs or the existing flanking regions. Studies need to be done to determine what STRs are best for deconvolution and how flanking region SNPs, or STR-SNPs, might help deconvolution. MPS' other issue is the price per run, which is becoming less expensive.

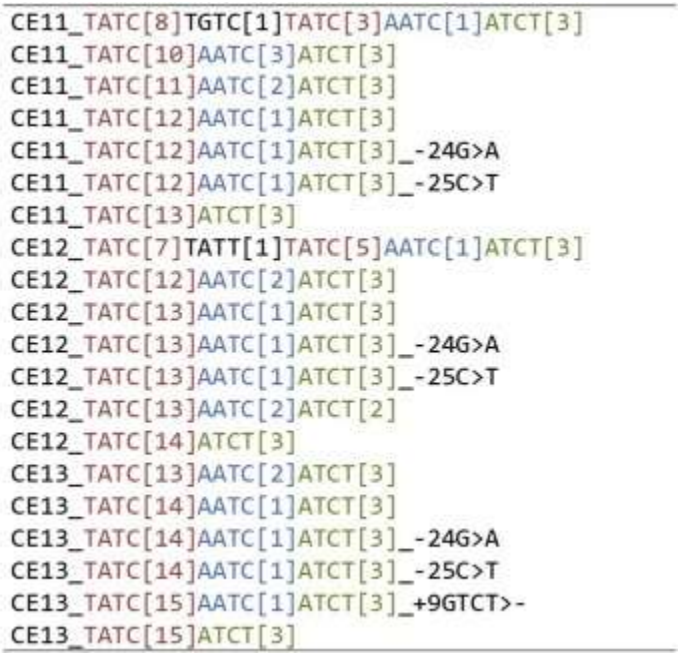


Figure 9- STR STRAND working group MPS nomenclature: The figure on the left shows from right to left an abbreviated name of the str, followed by the repeat sequences with a bracket [ ] showing the number of repeats for each sequence, finally an underline _ and either a positive or negative number designating the location of a SNP from the beginning of the repeat sequence and the nucleotide change shown by the nucleotide and a carrot > followed by the changed nucleotide or a dash – to designate a deletion.

Summary of Forensic DNA Analysis

Forensic DNA analysis has been an applied scientific skill that has evolved and changed to meet the needs of the time. The start of RFLP in the 1980s showed the breadth of knowledge that can be gained using DNA in forensic science. DQα and PCR development expanded on the information gained from the DNA analyzed. PCR also allowed less source DNA to be needed to complete an analysis. STR followed the trend of DNA techniques based on another. STR was built from RFLP and integrated PCR to develop a robust DNA analysis technique. STR is still the main form of DNA testing used in forensics. STR increased the discrimination ability of the

DNA and moved it into a form of DNA analysis that can be reproduced via CE technology. As with previous generations, a new generation of DNA analysis has been developed, building from STR and expanding on the information gained from the STR generation. MPS has started to make its way into the forensic DNA community. With its ability to sequence STRs and the more significant amount of data generated, MPS can quickly become the next form of forensic DNA technology. With a nomenclature being standardized by Gettings et al. (2022), the first major hurdle has been overcome. The next step for MPS technology was developing a statistical approach to defining what that data means when an analyst goes to court.

Massively Parallel Sequencing and the Daubert Standard

For forensic evidence to be used in a court of law in the United States, the evidence must meet the case law factors set forth by the case Daubert vs. Merrell Dow Pharmaceuticals, Inc (Daubert) in 1993. Daubert ruled that all evidence must meet five factors. These factors are 1) that the theory or technique employed by the expert is generally accepted in the scientific community, 2) the theory or technique has been subjected to peer review and publication, 3) the technique can be and has been tested, 4) the technique has a known error rate, and 5) the research was conducted independent of the litigation or dependent on an intention to provide the proposed testimony (Daubert, 1993).

According to a RAND Corporation study in 2002, after the implementation of Daubert, forensic evidence had a significant increase in rejection from trials across the United States. For evidence to be accepted in court, the theory for the analysis and the technique must meet the Daubert factors. Due to Daubert being the governing rule for forensic science, all newly developed forensic techniques and analyses must meet all the factors that Daubert set forth. So

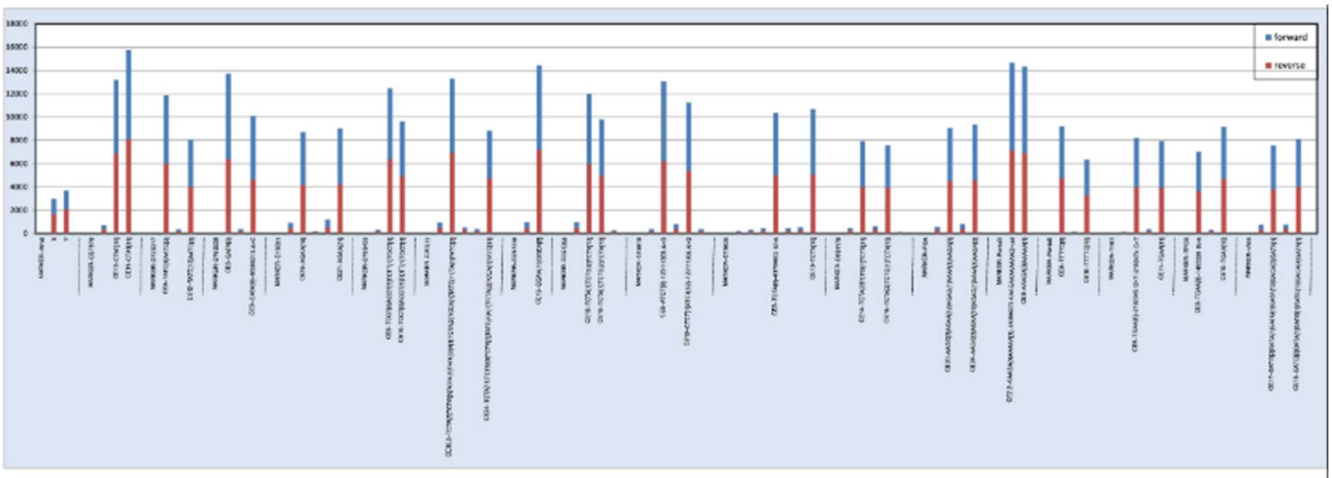
far, the MPS technology has met all the criteria set forth by the Daubert factors. However, the analysis of MPS data has yet to meet all the Daubert factors. With the determination of a set nomenclature, MPS data analysis is now on the path to overcoming the final hurdle of general acceptance, the one factor it has yet to meet.

Massively Parallel Sequencing Theory Tested and with Set Standards

As shown in the sections of MPS above, MPS theory has been thoroughly tested. MPS has been used in clinical research for over 20 years and tested in forensic usage for over 10 (de Knijff, 2019). Standards were at the forefront through the development of the MPS technology in forensic usage (Kastanis et al., 2019). Standards were also at the forefront in the development of the nomenclature for MPS data. With standards at the forefront of MPS technology, nomenclature data analysis standards have been quickly developed.

The development of MPS was based on a combination of STRs theory and sequencing. The STR and sequencing techniques are well-tested theories with well-developed standards. Specific aspects of MPS had to be developed and rigorously tested to build off these theories. The central aspect of the MPS technique that had to be developed was the multiplexing of the sequencing (Parson et al., 2016, Ganchow et al., 2019). In individual parts, the theory was sound; multiplexing came from STR theory, where multiple STRs are run simultaneously (Scheible et al., 2014). A study by Fan et al. (2022) explored developing a multiplex system for forensic STRs that sequenced 133 STRs, a mix of autosomal and Y STRs. This study built on previous studies such as a study by Wang et al. (2017) that sequenced 32 autosomal STRs, a study by Ganschow, Silvery, and Tiemann (2019) that focused on the 21 STRs used in forensic science in the USA, and a study by Fan et al. (2019) that sequenced 40 autosomal STR markers.

The theory and standards for analyzing MPS data had to be developed for the first steps of analysis. Due to the large volume of data gained through MPS, the old analysis techniques of STR and sequencing were not viable (Warshauer, 2015). The raw data yielded by MPS consists of every DNA sequence read by the instrument for each sample. This large amount of data is too large to be reviewed manually, like that of STR and sequencing data. An image of raw MPS data is shown below in figure 10. Due to the large amount of data obtained, software needed to be developed to help analyze the MPS data.



B. Sample read statistics

Read-category	Read-counts	Proportion of total reads
Total passed filter reads	537665	100,0%
Matched pairs	510409	94,9%
Known alleles (including stutters)	406437	75,6%
Genuine alleles (excluding stutter)	350294	65,2%
Reads with errors in the variant region (new alleles in TSSV analysis) *(Singletons)*	103972 *(27973)*	19,3% *5,2%*
Reads representing stutters	56143	10,4%
Primer dimers	27256	5,1%

Figure 106- Example MPS Read Data**:** The image above is an example of the amount of data obtained from running STRs through MPS technology. The graph on the top shows the number of reads for each individual STR. While the table below breaks down the total reads into matched pairs (with and without stutter), reads with errors in variant regions and reads representing stutter and primer dimers. The figure was obtained from a study by van der Garr et al. (2016)

SNP sequencing needs software that aligns and calls variants found in DNA sequences (Zhang et al., 2005). Bioinformatics software was developed to analyze the genomic data from MPS and determine the loci, allele, and SNP calls from the sequencing data. For use with MPS data, the developed software must analyze the raw data for the presence of each of the STR motifs. The software must assign the motifs to read the correct STR locus. The software for MPS analysis then needs to call the repeat numbers for the STR, as the standard STR CE software does. The software must finally align the sequences of the STR reads to a standard and call variants found therein (Warshauer, 2015). According to Van Neste et al. (2014), "Forensic bioinformaticians have been working on several algorithms to process MPS forensic STR data: lobSTR, RepeatSeq, STRait Razor, TSSV and the MyFLq-framework LobSTR and RepeatSe." However, some of the bioinformatic software that Neste et al. (2014) listed fall outside the scope of forensic analysis due to how they process the information. The listed bioinformatic software process genome-wide sequences.

Van Neste et al. (2014) continue that "STRait Razor, TSSV, and MyFLq" are locus-centric bioinformatic options focusing on forensic loci with the correct specifications. In contrast, forensic science requires analyzing a limited number of validated loci. The forensic adaptable software workflow is shown in figure 12 below.

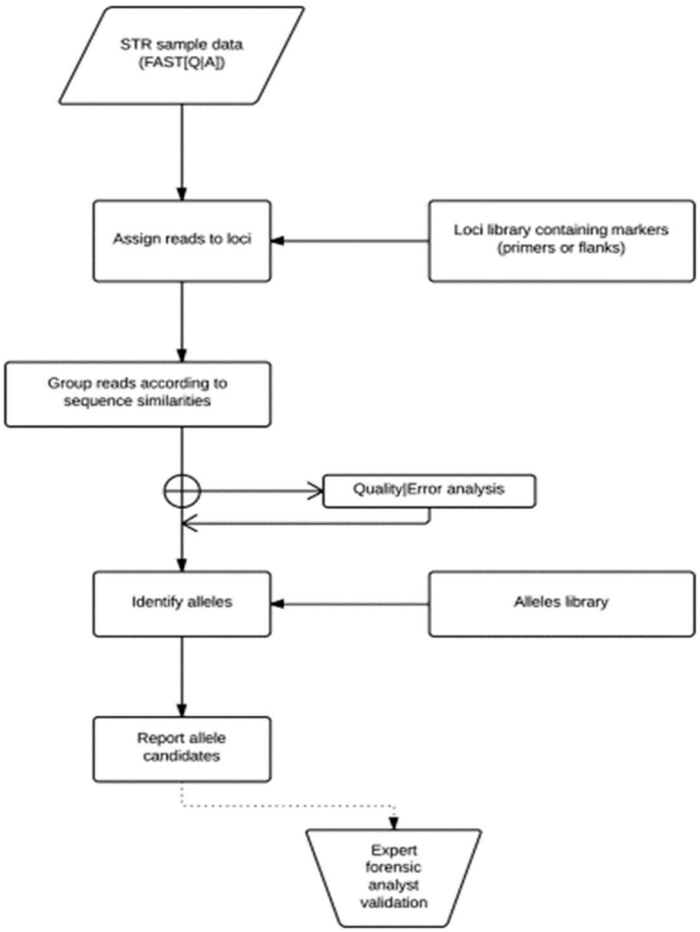


Figure 11 - Work Flow of MPS Analysis Software: The figure on the right shows the workflow of the bioinformatics software that analyses a large amount of MPS data and makes it legible for forensic analysis to make determinations on the data found from a study by Van Nest et al. (2014).

Massively Parallel Sequencing has been Peer Reviewed and Published

MPS technology can sequence multiple STR repeat regions, including their flanking region, and put several samples together for higher throughput using the barcodes in the MPS system. For MPS to be used in a court of law, the technique needs to have been peer-reviewed and published. MPS has been used in peer-reviewed research for many years. These years of study have allowed for the multiplex sequencing of STRs and Single Nucleotide Polymorphisms (SNPs) detection. It also has application to determine additional variation between individuals' sequences. In standard sequencing, currently, only one sequence of a single STR locus can be performed at a time (Budowle, 2003). In STR, multiple repeats and flanking regions are copied;

however, they are not sequenced. With the advancement of MPS, flanking region SNPs, also known as STR-SNPs, are being studied again. The flanking region of an STR is the region between the primer sites and the repeat region's start. There are 5' and 3' flanking region areas for all the STRs (Getting et al., 2015). Flanking region insertions and deletion (InDels) can make comparing STR and MPS difficult. This difficulty was because, in STR, the InDels are not noticed. Therefore, the InDels effect on the length of the DNA strand was disregarded (Getting et al., 2015). This can lead to cases where the repeat calls for STRs between STR-CE and MPS differ. These InDels are sequenced in MPS, and the repeat numbers can be accurately determined.

Flanking region SNPs, also known as SNP-STRs, are compound markers (Wang et al., 2013). The flanking region SNP was found to be tightly linked to the STR locus. SNP-STRs define STR alleles' genotypes based on the linked SNP found in the STR flanking region (Wang et al., 2015; Wang et al., 2013). In a study done by Wang et al. (2013), in 95 unrelated European individuals, 10 SNP alleles were found for D5S818, with only six alleles for D5S818 identified by STR-CE analysis. This study states that it might be possible to create allele-specific primers for minor contributing DNA using the SNP-STRs found in the minor contributor's DNA. Another study by Wang et al. (2015) looked at two SNPs, one located in the 5' flanking region of D10S12486 and another in the 5' flanking region of D13S317. Three SNP-STR genotypes were profiled successfully from the 9947A profile and 73 samples. Nevertheless, Tan et al. (2018) found 11 SNP-STR markers, and they also designed primers for these loci and got a 1-part to 100-part specificity in detecting mixtures.

Dalsgaard et al. (2013) reported three SNP variations in the flanking region in a study of four alleles. Two were found in D2S1338's flanking region and one in D12S391's flanking

region. This study concluded that the four alleles all had high sequence variation. Of the 24 loci examined in Gettings et al. (2015), three STRS out of 24 appeared to have significant flanking region variations. In the study by Novroski et al. (2016), flanking region variation was also gathered in addition to the 58 loci sequence variation. In a study by Wendt et al. (2017) of the Yavapai Native Americans, it was found that there were fewer variations found in the flanking region. Due to the lower genetic diversity in a population such as the Yavapai Native Americans, there was a lower variation of STR sequences. Wendt et al. (2017) postulate that population isolation and founding effects might lead to this lower genetic diversity. Different markers may need to be analyzed in similar groups for enough variations to be identified.

Known Potential Error Rate for Massively Parallel Sequencing

The following rule of evidence that MPS has met was that of having a known error rate. MPS reveals the entire spectrum of errors during PCR, including stutter, base-pair error due to DNA editing, strand slippage, and base-pair errors caused by substitution miscalls. Stutter was defined as an error in which a repeat region was missed during PCR that caused a call of one less in STR analysis. Stutter represents 10% of the actual peak call (Guo et al., 2017). For example, Guo et al. (2017) found that the percent of stutter for the MiSeq® FGx™ system was, on average, 6.71%, with the highest percent of stutter at DYS481 at 22.07%, not unlike STR analysis. The DYS481 STR was not one of the core STRs used in forensic analysis in the United States of America. Thus, analysts and software for MPS must distinguish between an error and an SNP mutation.

Through the development of the MiSeq® FGx™ system by Churchill et al. in 2015 using single source DNA profiles, DNA from one individual, the study found that in a source with

between 1 ng and 125 pg of DNA, there was a 99.6% to 99.9% type concordance. Using less DNA saw a loss of SNP with more than 50% loss at 7.82 pg. In independent studies by Xavier & Parson (2017) and Silva, Shugarts, & Smith (2017), all STRs studied through MPS were found to be in concordance with the STRs found through STR by CE, except for DXS10148, which has been found and noted as a problem in a previous study (Novoroski et al., 2016).

MPS has been studied for many years, and due to being developed after Daubert standards were in place, finding the known error rates was integral in its development. A study by Sharma et al. (2022) investigated the software tools associated with MPS. This study found that MPS had an error rate of 0.1% or higher due to artifacts such as stutter or indels. With MixtureAce, an MPS software, these artifacts could be reduced significantly, such as with possible allelic drop-in being reduced from 435 to fewer than two per sample, 57 STRs total. The allelic drop-in was defined as adding an allele that should not be there, usually caused by running PCR for too long. Thus, MPS meets the requirement of having a known error rate to be valid in court.

To determine the error rate in the analysis of MPS data, a minimum analytical coverage threshold must be set. A cover threshold was defined as the minimum amount of reads a single locus must meet to be analyzed. For example, in a study by Wang et al. (2017), the cover threshold was set to "500 reads for loci, and the minimum coverage threshold used for determination of a minor allele was 200 reads." Along with a coverage threshold, the depth of coverage (DoC), also known as read depth, must be calculated for each sample and each STR locus. The depth of coverage was determined to be the ratio of all allele reads to the length of the allele.

Massively Parallel Sequencing Theory is Validated

The next Daubert factor was that the theory in question must be validated. The theory behind the technology and analysis must be validated for MPS. MPS technology has two primary instruments available for forensic MPS. The two technologies for forensic science are the MiSeq® FGx™ Forensic Genomics System by Illumina and the HID-Ion Personal Genome Instrument (PGM)™ by Thermo Fisher Scientific (Dame Apaga et al., 2017). Only the MiSeq® FGx™ Forensic Genomics System has been approved for human forensic ID by The National DNA Index System (FBI, 2021), and the Scientific Working Group on DNA Analysis Methods (2019) focus was on the MiSeq® FGx™ system. Additionally, multiple laboratories have validated this instrument, and it is the only MPS system certified to upload to the national database for CODIS (Tabak, 2019).

The MiSeq® FGx™ forensic genomics system, made by Verogen, a subsidiary of Illumina, comprises a DNA library kit, MPS instrumentation, and data analysis software (Jäger et al., 2017). For use in forensic science, this system has been validated (Tabak, 2019). The validation of the MiSeq® FGx™ system required a range of studies to be conducted on many samples. The DNA analysis guidelines for validation, according to the Scientific Working Group on DNA Analysis Methods (SWGDAM) (2016), are shown in Table 3 below. Continuing the studies that showed that MPS meets the requirements for being peer-reviewed and published and that MPS has known error rates found in studies by Jager et al. (2017), Hollard et al. (2019), and Sukawutthiya, Sathirapatya, and Vongpaisarnsin (2017) and Tan et al. (2018) all have shown the validation of MPS technology. In addition, all studies listed above used known non-probative samples, meeting the first validation requirement.

	Extraction System	Quantitation System	Amplification System/Reaction Conditions	Detection System
Known / Non-Probative Samples	X	X	X	X
Precision and Accuracy: Repeatability	X	X	X	X
Precision and Accuracy: Reproducibility	X	X	X	X
Sensitivity Studies	X	X	X	X
Stochastic Studies		X	X	X
Mixture Studies	X*	X*	X	X
Contamination Assessment	X	X	X	X

Table 4 - Validation Requirements According to SWGDAM: The above table shows the requirements SWGDAM had defined as necessary for validating a DNA analysis technology (2016). The requirements are known samples, precision, accuracy (both in repeatability and reproducibility), and sensitivity studies. Stochastic studies, mixture studies, and contamination assessment. The astrics mean that the system only has to meet the guideline if it is going to be used in mixture analysis

The Jager et al. (2017) study shows that MPS technology meets the sensitivity standard set by the SWGDAM guidelines above. The study's sensitivity was over 99%, with more than 62.5 pg of DNA for autosomal, X, and Y STRs. The Jager et al. study (2017) tested 182 samples of four substrates: buccal swabs collected using sterile cotton swabs, FTA cards, and Bode Buccal DNA Collectors were amplified as well as 41 prepared human DNA lysates. Two higher primates, one avian, and nine non-primate animals were also tested, as two fungal and six bacterial samples. The study by Jager et al. (2017) observed partial data in seven samples, most likely due to the sampling collection method. The study found that some non-human DNA produced sequencing data, except for the ferret and the bacterial samples. None of the non-

human samples produced less than 3000 reads per loci other than the mouse, rat, and non-human primates, with the non-human primates producing the most reads.

Hollard et al. (2019) implemented an automated process developed with Hamilton Robotics to bypass five steps prone to human error in the preparation steps to run the MiSeq® FGx™ system. They found that their automation system significantly reduced the variation of the DOC across X-STRs. In contrast, the variations were similar for autosomal and Y STRs to previous studies. The benefit of the automated process was that it significantly reduced the time required to set up the DNA for analysis. This included the addition of two extra steps compared to the manual method. The automated process dropped hands-on time from 3 hours to 15 minutes for 96 samples. Hollard et al. (2019) study showed that massively parallel sequencing meets the precision and accuracy standards of SWGDAM and further exceeds it with the automated process.

In another validation study by Sukawutthiya, Sathirapatya, and Vongpaisarnsin (2017), the sensitivity of MPS technology was tested. The general guideline for reads per sample was 85,000, and each locus needs to have greater than 250 reads, according to the ForenSeq™ Universal Analysis Software's Quality Metric page (Illumina, 2016). The allele coverage ratio (ACR) for the autosomal STRs was between 0.81 and 0.95. Other than in D22S1045, which should be interpreted cautiously according to the ForenSeq™ DNA Signature Prep Kit guidebook. The study showed that the average DOC for all the STR locus was between 523-11,534 reads. These results were like the study by Hollard et al. (2019), which found an average of 0.82 and a DOC of 230-4317 reads for autosomal STRs. This study and studies by Guo et al. (2017) and Jager et al. (2017) also had read issues with D22S1045, showing that this STR should not be analyzed using this system.

The Tan et al. (2018) study explored the validation of mixed DNA samples. The development of MPS circumvents many of the limitations of cell separation. It could detect a minor allele even at a ratio of 100:1. However, assessing minor contributors with mixtures of less than 5% was challenging if no prior information was available (Tan et al., 2018). This was due to the major contributor of DNA overwhelming the minor contributor; thus, the information obtained from the minor contributor will be minimal or lost altogether. With the increase in possible STR loci, the forensic analysis might include more STRs that show significant variation in the flanking region, repeat region, or both for better mixture deconvolution (Novroski et al., 2016).

For MPS analysis, the bioinformatic software used must be validated. For validation, there needs to be a known error rate, as shown above, and the data be concordant with data gained by STR and traditional sequencing. For STRait Razor, a study by Warshauer et al. (2013) found that in 427 STR alleles, STRait Razor was in "complete concordance" with the STR genotype. In a study by Valle-Silva et al. (2022), HipSTR, STRait Razor, and toaSTR were compared. Valle-Silva et al. found that "HipSTR, STRait Razor, and toaSTR from [MPS] data in the Brazilian population are highly concordant and reliable."

General Acceptance of Massively Parallel Sequencing Theory

The final hurdle for MPS to meet the Daubert factor is general acceptance. General acceptance in this context was defined as acceptance of MPS by the forensic DNA community. The forensic science community has generally accepted the technology behind MPS; however, some areas of concern still remain for MPS analysis to be entirely accepted. The largest of these issues is that there is no set statistical approach to determine the combined statistical weight of

the STR and SNP data obtained through MPS. The statistical analysis of STR and SNP data has been studied multiple times independently of each other, but a transparent methodology for the combined statistic has not been articulated.

Statistical analysis of Massively Parallel Sequencing Data

In a study by Avila et al. (2019), their statistical approach met the Brazilian Federal Police requirements where the weight of genetic evidence was reported for criminal samples as likelihood ratios. To accomplish this, the likelihood ratio was calculated with validated, internal use computational tools and confirmed with Familias software v.3.2.2 for STR markers only. The likelihood ratio was developed using random match probabilities (RMP). The chance that the same DNA sample will occur in a random sample in the population. The random match probability was calculated using "IBM1 Statistical Package for the Social Sciences (SPSS1), version 22. Single-factor analysis of variance (ANOVA) tests were performed for genetic weight-of-evidence in the form of samples' RMP values to verify if resulting values differ for regional Brazilian populations and across worldwide frequency databases" (Avila et al., 2019). This RMP value was used to determine the RMP and the population data for STR and SNP markers following current BFP technical protocols. Avila et al. (2019) found RMP values for STR-SNP using the national allele frequencies found in Brazil. The study then compared sample genotypes and calculated fully equivalent profiles. The Avila et al. (2019) study found that the "average LR for all criminal samples was estimated at 1.0715×10^{39} ([standard deviation] = ±2.1345 × 1039) when the 90-SNP marker set was considered." The Avila study proved a higher power of discrimination using the combination of random match probabilities of the SNPs and the STRs.

In a study by Li et al. (2018), RMP, power of discrimination (PD), polymorphism information content (PIC), and power of exclusion (PE) were determined. The study accomplished this using modified PowerStats software. To determine the Linkage disequilibrium (LD), Li et al. (2018) used Arlequin v3.5.1.3 to test between all pairs of STR loci. An ANOVA test was used to estimate the interpopulation differentiation between the Chengdu Han population and the other 25 published populations.

The statistical analysis of these studies was completed by running the relevant data, RMP, through statistical software. The Avila et al. (2019) study used SPSS, while the Li et al. (2018) used a modified PowerStats software. These software packages function by the user imputing the relevant information as variables. Once the variables were defined, the method of analysis was then selected. For both studies, an ANOVA test was run. An ANOVA test was an F-test of main effects and interactions (Rouder, 2016). At this point, the statistical analysis of MPS data has gone past the point that an analyst can determine without the assistance of these statistical software packages.

Proposed Statistical Analysis of Massively Parallel Sequencing

While the studies above have touched on statistical options in evaluating MPS data, the statistics they use are advanced for use in court. When used in court, the statistical analysis must be understandable by a large audience with limited knowledge of statistics and forensic DNA (Chaudhuri, 2017). With the development of the nomenclature for MPS, a statistical approach can be developed using the nomenclature as a base which will be more readily understood. Building off the previous studies, the author proposes using a statistical approach that expands upon the information obtained from the research into the background of forensic DNA

techniques. A statistical model for MPS can be developed using the base of STR statistics to build off. Thus, the Hardy Weinberg Equation, $p^2+2pq+q^2$, and the frequency statistics for the alleles for each STR loci were used as the base for the new statistical equation for MPS (Guo & Thompson, (1992). The proposed statistical analysis continues the Hardy Weinberg equation usage found in studies by Churchill et al. (2017), King et al. (2018), Avila et al. (2019), and Li et al. (2018). Using mutation rates found through STRbase, an equation can be developed that builds off the STR equation. The new equation is as follows: 2pq*[m*x]. Whereas 2pq is the population frequency for both allele times, 2, m is the mutation rate for the STR, and x is the number of mutations found in the STR. The mutation rate for the STR loci was found through forensic paternity, using the rate at which each STR mutates from parent to child according to STRbase and the Association for the Advancement of Blood & Biotherapies. The idea of combining the population frequencies of STRs and SNPs was built on the suggestion of a study by Cortellini et al. (2019). This equation can be used to determine the RMP for each SNP-STR locus.

To test the statistical equation, data from a study by Wang et al. (2017) was used to make a composite MPS DNA source. The composite was developed from the 007 control, and the sequence was found at identical alleles. These alleles then had their frequency determined as just STRs using population statistics for the Asian population (Steffen et al., 2017) due to this being the population studied by Wang et al. (2017). This data was then put through the equation described above using mutation rates found in STRbase (Ruitberg, 2001). This information was compiled into a spreadsheet that is attached as addendum 1. The RMP for the STR of the composite was found to be $8.998*10^{-33}$ with a likelihood ratio of 1 in $1.111*10^{32}$. The RMP for the MPS equation was found to be $1.302*10^{-43}$ and a likelihood ratio of 1 in $7.677*10^{42}$. This

statistic only accounts for the RMP and Likelihood ratio but does not account for the PD, PIC, or PE found in the Avila et al. (2019) study.

The author calls on the Scientific Working Group on DNA Analysis Methods (SWGDAM) and the International Society for Forensic Genetics (ISFG) to set recommendations on the statistical approach needed for analyzing MPS data. Without the statistical approach being set, MPS will never meet the general acceptance threshold set forth by Daubert. Once the approach is defined, MPS will become the leading forensic DNA technique used in degraded DNA and complex mixture analysis cases within five years. With the continued decrease in the price of running MPS in 10 years, MPS will become the standard technique used in all forensic DNA testing in the United States.

Discussion

No one truly knows where forensic DNA will be in 25 years. Through the research above, MPS and STR-SNP are agreed upon as the next step in forensic DNA casework. The research conducted in this study strived to determine if MPS met Daubert standards and what areas of concern were still found. Using the lens of a forensic science student, areas of concern were found and expanded upon. These areas included a nomenclature and statistical approach and areas where MPS data and interpretation could be confusing in court. The study tested the now-defined nomenclature and a proposed statistical model to find other areas of concern.

With the development of nomenclature by the STR STRAND working group (Gettings, 2019), there is now a path forward for MPS to become the new forensic DNA testing standard. The next step needed for MPS is the development of statistical analysis. Once a statistical analysis is developed, MPS can be used for forensic cases involving highly degraded DNA and

complex mixtures. MPS can even be used for high-priority cases in the next few years. However, more research will be needed for MPS to be used in more varied forensic cases.

For MPS to be most helpful, understanding the underlying variation of STRs needs to be studied in relevant populations. This needs to be accomplished, so allelic frequency calculations can be accomplished to strengthen MPS data and to be able to use it in court. A study by Novroski et al. (2016) continued the work of Gettings et al. (2015) and studied the STR repeat region of 58 STRs for 777 individuals belonging to one of four significant populations: African American, Caucasian, Hispanic, and Chinese. These studies include large-scale studies to determine the variations in human populations in the STR repeat region and their flanking region. Studies must be completed to determine which STRs have alleles with enough sequencing variation to be useful for mixture deconvolution. Studies also need to be done to determine if flanking region variations can be helpful in mixture deconvolution. Finally, research needs to be done on the difference in repeat information gathered from STR via CE and MPS to see how best to compare the two results. Once the frequencies are compiled, MPS can be used for more extensive cases. In five years, with these studies done, MPS will start to propagate in more forensic labs in the United States of America.

The final advancement for MPS will be the development of a database for sequence data used in forensic cases. This database will need to be like CODIS used in STR cases now. The database will need to be national and connected to the regional labs. The database will also need to be compatible with previous STR data. With the development of the database, MPS could become the standard in forensic DNA testing in the United States in 10 years. For MPS to become the next standard of forensic DNA testing is all contingent on the price of each run of MPS continuing to decline to make it cost-effective compared to STR.

Conclusion

MPS will become the next forensic DNA testing technology in the United States. However, a few issues could limit MPS technology propagation to all labs. Through this thesis, the research in forensic science was combined and compared to the Daubert standard and the previous DNA testing technique. Through the comparison to the Daubert standard, issues of general acceptance for data analysis were determined. It was found that MPS had issues with its nomenclature and statistics. Through the comparison of how STR is used in court, further issues that MPS faced were discovered. These issues were the reeducation needs for MPS, the cost per run for MPS compared to STR, and the need to stress the limitations of MPS in forensic use due to how many uses MPS has in the greater genetic field.

The biggest issue surrounding MPS technology for full use is the cost of the technology and the cost of running price, with the cost of a platform being around $480,000 (von Budnoff, 2008) and the cost of a single run being in the hundreds of dollar range (Nimwegen et al., 2016). A study by Nimewgen et al. (2016) notes that the sequencing cost has decreased drastically as speed, sequence length, and throughput have increased.

The next challenge facing MPS is the over-saturation of DNA testing available in the United States can lead to confusion on what MPS in forensic science limitations are. With the increase in at-home genetic testing and focus on DNA phenotyping, the use of MPS data in court can become unclear to the jury. An article by Jorde and Bamshad (2020) estimated that “26 million people worldwide” have undergone at-home genetic ancestry testing. With more people worldwide knowing how ancestry DNA works, confusion between ancestry and forensic DNA can become commonplace. As with having to keep statistics simple, what and how DNA is used in forensic science also must be kept simple for a jury (Chaudhuri, 2017). With the basic

knowledge of ancestry and MPS also being used in phenotyping, confusion about what MPS is used for can be common (Chiara et al., 2019; Canales, 2020; Kukla-Bartoszek et al., 2020; Melchionda et al., 2022). Forensic phenotyping currently has many limitations (Serrano, 2020). With phenotyping's limitations and the use of MPS in phenotyping, a correlation between these limitations and MPS for forensic use can be developed by a jury. In a study by Delahunty & Hewson (2010), a pre-trial "cognitively-sequenced generic tutorial" helped a jury understand DNA evidence. If tutorials were allowed for MPS for forensic use, it would help limit the confusion caused by phenotyping and ancestry.

The final hurdle of MPS is that until SWGDAM and ISGF have set a statistical model, MPS will not meet all the Daubert factors. Without meeting all the Daubert factors, MPS will not be allowed to be used as evidence. Once the statistical model is defined, a database for the data gained through MPS technology will need to be developed and validated. The database must continue building off previous DNA technologies and the processes surrounding it. The Database for MPS will need to be able to upload all the information already found on the Combined DNA Index System currently used for STR data and handle all the information gained from MPS. Once these issues are resolved and studied, MPS will become the next leading DNA testing technology in the United States, meeting all Daubert factors for use in court.

References

Avila, E., Cavalheiro, C. P., Felkl, A. B., Graebin, P., Kahmann, A., & Alho, C. S. (2019). Brazilian forensic casework analysis through MPS applications: Statistical weight-of-evidence and biological nature of criminal samples as an influence factor in quality metrics. *Forensic science international, 303*, 109938-109938. doi:10.1016/j.forsciint.2019.109938

Ballard, D., Winkler-Galicki, J., & Wesoły, J. (2020). Massive parallel sequencing in forensics: advantages, issues, technicalities, and prospects. International Journal of Legal Medicine, 134(4), 1291–1303. https://doi.org/10.1007/s00414-020-02294-0

Børsting, C., Fordyce, S. L., Olofsson, J., Mogensen, H. S., & Morling, N. (2014). Evaluation of the Ion Torrent™ HID SNP 169-plex: A SNP typing assay developed for human identification by second generation sequencing. *Forensic Science International : Genetics*, *12*, 144–154. https://doi.org/10.1016/j.fsigen.2014.06.004

Budowle, B., Allard, M., Wilson, M., & Chakraborty, R. (2003). FORENSICS AND MITOCHONDRIAL DNA: Applications, Debates, and Foundations. Annual Review of Genomics and Human Genetics, 4(1), 119–141. https://doi.org/10.1146/annurev.genom.4.070802.110352

Budowle, B., Giusti, A. M., Waye, J. S., Baechtel, F. S., Fourney, R. M., Adams, D. E., Presley, L. A., Deadman, H. A., & Monson, K. L. (1991). Fixed-bin analysis for statistical evaluation of continuous distributions of allelic data from VNTR loci, for use in forensic comparisons. American journal of human genetics, 48(5), 841–855.

Budowle, B., Planz, J.V., Campbell, R., and Eisenberg, A.J. (2005). Molecular diagnostic applications in forensic science. In: Molecular Diagnostics, Patrinos, G. and Ansorge, W., (eds.), Elsevier, Amsterdam, pp. 267-280.

Butler, J. M. (2007a). Short tandem repeat typing technologies used in human identity testing. *BioTechniques, 43*(4), ii-v. doi:10.2144/000112582;

Butler, J. M. (2007b). *STR and Molecular Biology Artifacts* [PowerPoint Presentation]. Houston DNA Training Workshop, Huston, Texas, USA. https://strbase.nist.gov/pub_pres/2_STR_Artifacts.pdf

Canales S. (2020). Forensic DNA phenotyping: A promising tool to aid forensic investigation. Current situation. Spanish Journal of Legal Medicine, 46(4), 183–190. https://doi.org/10.1016/j.remle.2020.01.002

Chamberlain, J. S., & Chamberlain, J. R. (1994). Optimization of Multiplex PCRs. In K. B. Mullis, F. Ferré, & R. A. Gibbs (Eds.), *The Polymerase Chain Reaction* (pp. 38-46). Boston, MA: Birkhäuser Boston.

Chaudhuri, P. (2017). A Right To Rational Juries? How Jury Instructions Create The "Bionic Juror" In Criminal Proceedings Involving DNA Match Evidence. California Law Review, 105(6), 1807–1852. https://doi.org/10.15779/Z38183426H

Churchill, J., Novroski, N. M. ., King, J. L., Seah, L. H., & Budowle, B. (2017). Population and Performance Analyses of Four Major Populations with Illumina's FGx Forensic Genomics System. Forensic Science International : Genetics, 30, 81–92. https://doi.org/10.1016/j.fsigen.2017.06.004

Clarke, L., Fairley, S., Zheng-Bradley, X., Streeter, I., Perry, E., Lowy, E., Tassé, A.-M., & Flicek, P. (2017). The international Genome sample resource (IGSR): A worldwide

collection of genome variation incorporating the 1000 Genomes Project data. Nucleic Acids Research, 45(D1), D854–D859. https://doi.org/10.1093/nar/gkw829

Cobey v State 73 Md. 233 (1987)

Dai, W., Pan, Y., Sun, X. et al. (2019). High polymorphism detected by massively parallel sequencing of autosomal STRs using old blood samples from a Chinese Han population. Sci Rep 9, 18959 https://doi.org/10.1038/s41598-019-55282-9

Dalsgaard, S., Rockenbauer, E., Gelardi, C., Børsting, C., Fordyce, S. L., & Morling, N. (2013). Characterization of mutations and sequence variations in complex STR loci by second generation sequencing. *Forensic Science International: Genetics Supplement Series, 4*(1), e218-e219. doi:10.1016/j.fsigss.2013.10.112

Dame Apaga, L., Dennis, S., Salvador, J., Calacal, G., & De Ungria, M. (2017). Comparison of Two Massively Parallel Sequencing Platforms using 83 Single Nucleotide Polymorphisms for Human Identification. *Scientific Reports, 7*(1), 1. doi:10.1038/s41598-017-00510-3

Daubert v. Merrell Dow Pharmaceuticals, Inc. (1993) 509 US 579, 589.

de Knijff, P. (2019). From next generation sequencing to now generation sequencing in forensics. *Forensic Science International: Genetics, 38*, 175-180. doi:10.1016/j.fsigen.2018.10.017

Delahunty, J., & Hewson, L. (2010). Improving jury understanding and use of expert DNA evidence. Australian Institute of Criminology.

den Dunnen, J. T., Dalgleish, R., Maglott, D. R., Hart, R. K., Greenblatt, M. S., McGowan-Jordan, J., . . . Taschner, P. E. (2016). HGVS Recommendations for the Description of Sequence Variants: 2016 Update. *Hum Mutat, 37*(6), 564-569. doi:10.1002/humu.22981

Devesse, Davenport, L., Borsuk, L., Gettings, K., Mason-Buck, G., Vallone, P. M., Syndercombe Court, D., & Ballard, D. (2020). Classification of STR allelic variation using massively parallel sequencing and assessment of flanking region power. *Forensic Science International : Genetics*, *48*, 102356–102356.

DNA IQ™ System— Database Protocol. (2016). Promega. Madison, MI, USA

Fan, H., Wang, L., Liu, C., Lu, X., Xu, X., Ru, K., Qiu, P., Liu, C., & Wen, S.-Q. (2021). Development and validation of a novel 133-Plex Forensic STR panel (52 strs and 81 Y-strs) using single-end 400 BP massive parallel sequencing. International Journal of Legal Medicine, 136(2), 447–464. https://doi.org/10.1007/s00414-021-02738-1

Fan, H., Wang, Y., Zhao, J., Cheng, F., Wang, M., Liu, Z., Chen, M., Li, W., Shi, L., Zhang, X., Zhang, J., Zhang, G., & Yan, J. (2019). Development of a 40 linked autosomal STRs panel using massively parallel sequencing. Forensic Science International. Genetics Supplement Series, 7(1), 355–357. https://doi.org/10.1016/j.fsigss.2019.10.010

FBI, (2021) *National DNA Index System (NDIS) Operational Procedure Manual,*

Fordyce, S. L., Ávila-Arcos, M. C., Rockenbauer, E., Børsting, C., Frank-Hansen, R., Petersen, F. T., . . . Gilbert, M. T. P. (2011). High-throughput sequencing of core STR loci for forensic genetic investigations using the Roche Genome Sequencer FLX platform. *BioTechniques, 51*(2), 127-133. doi:10.2144/000113721

ForenSeq™ DNA Signature Prep Reference Guide. (2018). In: Verogen. Forensic Science International: Genetics, 43, 102165, ISSN 1872-4973, https://doi.org/10.1016/j.fsigen.2019.102165

Frégeau C., Fourney R. (1993) DNA typing with fluorescently tagged short tandem repeats: a sensitive and accurate approach to human identification. Biotechniques (1):100-119. PMID: 8103347.

Frye v. United States, No. 3968 (Court of Appeals District Court of Columbia 1923).

Ganschow, S., & Tiemann, C. (2019). Development of a multiplex forensic identity panel for massively parallel sequencing and its systematic optimization using design of experiments. Forensic Science International : Genetics, 39, 32–43. https://doi.org/10.1016/j.fsigen.2018.11.023

Gelardi, C., Rockenbauer, E., Dalsgaard, S., Børsting, C., & Morling, N. (2014). Second generation sequencing of three STRs D3S1358, D12S391 and D21S11 in Danes and a new nomenclature for sequenced STR alleles. *Forensic Science International: Genetics, 12*, 38-41. doi:10.1016/j.fsigen.2014.04.016

Gettings, K. (2022, March 3). A Short Tandem Repeat (STR) Sequence Nomenclature Update From the Short Tandem Repeat: Align, Name, Define (STRAND) Working Group [Conference Abstract]. 74th annual convention of the American Academy of Forensic Science, Seattle, WA, United States. https://www.aafs.org/sites/default/files/media/documents/2022Proceedings_Final.pdf

Gettings, K. B., Aponte, R. A., Vallone, P. M., & Butler, J. M. (2015). STR allele sequence variation: Current knowledge and future issues. *Forensic Science International: Genetics, 18*, 118-130. doi:10.1016/j.fsigen.2015.06.005

Gettings, K. B., Ballard, D., Bodner, M., Borsuk, L., King, J., Parson, W., and Phillips, C., (2019), Report from the STRAND Working Group on the 2019 STR sequence nomenclature meeting,

Gettings, K. B., Borsuk, L. A., Ballard, D., Bodner, M., Budowle, B., Devesse, L., . . . Vallone, P. M. (2017). STRSeq: A catalog of sequence diversity at human identification Short Tandem Repeat loci. *Forensic Science International: Genetics, 31*(C), 111-117. doi:10.1016/j.fsigen.2017.08.017

Gill, P., Jeffreys, A. J., & Werrett, D. J. (1985). Forensic application of DNA' fingerprints'. *Nature, 318*(6046), 577-579. doi:10.1038/318577a0

Gomes, C., Magalhães, M., Alves, C., Amorim, A., Pinto, N., & Gusmão, L. (2012). Comparative evaluation of alternative batteries of genetic markers to complement autosomal STRs in kinship investigations: autosomal indels vs. X-chromosome STRs. *International Journal of Legal Medicine, 126*(6), 917-921. doi:10.1007/s00414-012-0768-5

Griffiths, R., Barber, M. D., Johnson, P. E., Gillbard, S. M., Haywood, M. D., Smith, C. D., Arnold, J., Burke, T., Urquhart, A. J., & Gill, P. (1998). New reference allelic ladders to improve allelic designation in a multiplex STR system. International Journal of Legal Medicine, 111(5), 267–272. https://doi.org/10.1007/s004140050167

Gross, T., & Käufer, N. F. (1998). Cytoplasmic ribosomal protein genes of the fission yeast Schizosaccharomyces pombe display a unique promoter type: a suggestion for nomenclature of cytoplasmic ribosomal proteins in databases. *Nucleic acids research, 26*(14), 3319-3322. doi:10.1093/nar/26.14.3319

Grubwieser, P, Mühlmann, R, & Parson, W. (2003). New sensitive amplification primers for the STR locus D2S1338 for degraded casework DNA. International Journal of Legal Medicine, 117(3), 185–188. https://doi.org/10.1007/s00414-002-0359-y

Guo, F., Yu, J., Zhang, L., & Li, J. (2017). Massively parallel sequencing of forensic STRs and SNPs using the Illumina® ForenSeq™ DNA Signature Prep Kit on the MiSeq FGx™ Forensic Genomics System. *Forensic Science International: Genetics, 31*, 135-148. doi:10.1016/j.fsigen.2017.09.003

Guo, S., & Thompson, E. (1992). Performing the Exact Test of Hardy-Weinberg Proportion for Multiple Alleles. Biometrics, 48(2), 361–372. https://doi.org/10.2307/2532296

Helmuth, R., Fildes, N., Erlich, H. A., Blake, E., Luce, M. C., Madej, R., Gorodezky, C., Stoneking, M., Schmill, N., Klitz, W., & Higuchi, R. (1990). HLA-DQα allele and genotype frequencies in various human populations, determined by using enzymatic amplification and oligonucleotide probes. American Journal of Human Genetics, 47(3), 515–523.

Hollard, C., Ausset, L., Chantrel, Y., Jullien, S., Clot, M., Faivre, M., . . . Laurent, F.-X. (2019). Automation and developmental validation of the ForenSeq™ DNA Signature Preparation kit for high-throughput analysis in forensic laboratories. *Forensic Science International: Genetics, 40*, 37-45. doi:10.1016/j.fsigen.2019.01.010;

isoallele. (2005). In Cambridge Dictionary of Human Biology and Evolution.

Jäger, A. C., Alvarez, M. L., Davis, C. P., Guzmán, E., Han, Y., Way, L., . . . Holt, C. L. (2017). Developmental validation of the MiSeq FGx Forensic Genomics System for Targeted Next Generation Sequencing in Forensic DNA Casework and Database Laboratories. *Forensic Science International: Genetics, 28*, 52-70. doi:10.1016/j.fsigen.2017.01.011

Jeffreys, A. J., Wilson, V., & Thein, S. L. (1985a). Hypervariable' minisatellite' regions in human DNA. *Nature, 314*(6006), 67-73. doi:10.1038/314067a0

Jeffreys, A. J., Wilson, V., & Thein, S. L. (1985b). Individual-specific' fingerprints' of human DNA. *Nature, 316*(6023), 76-79. doi:10.1038/316076a0

Jeffreys, A. J., Wilson, V., Thein, S. L., Weatherall, D. J., & Ponder, B. A. (1986). DNA "fingerprints" and segregation analysis of multiple markers in human pedigrees. *Am J Hum Genet, 39*(1), 11-24.

Kastanis, G., Santana-Quintero, L. V., Sanchez-Leon, M., Lomonaco, S., Brown, E. W., & Allard, M. W. (2019). In-depth comparative analysis of Illumina® MiSeq run metrics: Development of a wet-lab quality assessment tool. Molecular Ecology Resources, 19(2), 377–387. https://doi.org/10.1111/1755-0998.12973

King, J., Churchill, J. D., Novroski, N. M. M., Zeng, X., Warshauer, D. H., Seah, L.-H., & Budowle, B. (2018). Increasing the discrimination power of ancestry- and identity-informative SNP loci within the ForenSeq™ DNA Signature Prep Kit. Forensic Science International : Genetics, 36, 60–76. https://doi.org/10.1016/j.fsigen.2018.06.005

Koboldt, Daniel c., Steinberg, Karyn m., Larson, David e., Wilson, Richard k., & Mardis, E. R. (2013). The Next-Generation Sequencing Revolution and Its Impact on Genomics. *Cell, 155*(1), 27-38. doi:10.1016/j.cell.2013.09.006

Kukla-Bartoszek, M., Szargut, M., Pośpiech, E., Diepenbroek, M., Zielińska, G., Jarosz, A., Piniewska-Róg, D., Arciszewska, J., Cytacka, S., Spólnicka, M., Branicki, W., & Ossowski, A. (2020). The challenge of predicting human pigmentation traits in degraded bone samples with the MPS-based HIrisPlex-S system. Forensic Science International : Genetics, 47, 102301–102301. https://doi.org/10.1016/j.fsigen.2020.102301

Li, Y., Li, H., He, G., Liang, W., Luo, H., Liao, M., Zhang, J., Yan, J., Li, Y., Hou, Y., & Wu, J. (2018). Genetic polymorphism of 21 non-CODIS STR loci in Chengdu Han population

and its interpopulation analysis between 25 populations in China. Legal Medicine (Tokyo, Japan), 31, 14–16. https://doi.org/10.1016/j.legalmed.2017.12.007

Life Technologies. (2018). *Quantifiler™ HP and Trio DNA Quantification Kits, edition H.* Warrington, UK.

Lygo, J., Johnson, P. E., Holdaway, D. J., Woodroffe, S., Whitaker, J. P., Clayton, T. M., Kimpton, C. P., & Gill, P. (1994). The validation of short tandem repeat (STR) loci for use in forensic casework. International Journal of Legal Medicine, 107(2), 77–89. https://doi.org/10.1007/BF01225493

Melchionda, F., Silvestrini, B., Robino, C., Bini, C., Fattorini, P., Martinez-Labarga, C., De Angelis, F., Tagliabracci, A., & Turchi, C. (2022). Development and Validation of MPS-Based System for Human Appearance Prediction in Challenging Forensic Samples. Genes, 13(10), 1688. https://doi.org/10.3390/genes13101688

Metzker, Michael L. (2010). Sequencing technologies - the next generation. Nature Reviews. Genetics, 11(1), 31–46. https://doi.org/10.1038/nrg2626

Michael, L. M. (2009). Sequencing technologies — the next generation. *Nature Reviews Genetics, 11*(1), 31. doi:10.1038/nrg2626

Mizuno, Kitayama, T., Fujii, K., Nakahara, H., Yoshida, K., Sekiguchi, K., Yonezawa, N., Nakano, M., & Kasai, K. (2008). A D19S433 Primer Binding Site Mutation and the Frequency in Japanese of the Silent Allele It Causes. *Journal of Forensic Sciences*, *53*(5), 1068–1073.

Mullis, K., Faloona, F. (1987). Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods Enzymol*. 155:335-50. doi: 10.1016/0076-6879(87)55023-6.

Nimwegen, K., Soest, R., Veltman, J., Nelen, M., van der Wilt, G., Peart-Vissers, L., & Grutters, J. (2016). Is the $1000 Genome as Near as We Think? A Cost Analysis of Next-Generation Sequencing. Clinical Chemistry (Baltimore, Md.), 62(11), 1458–1464. https://doi.org/10.1373/clinchem.2016.258632

Novroski, N., King, J., Churchill, J., Seah, L., & Budowle, B. (2016). Characterization of genetic sequence variation of 58 STR loci in four major population groups. *Forensic Science International: Genetics, 25*, 214-226. doi:10.1016/j.fsigen.2016.09.007

Oldoni, F., Bader, D., Fantinato, C., Wootton, S. C., Lagacé, R., Hasegawa, R., Chang, J., Kidd, K., & Podini, D. (2019). A massively parallel sequencing assay of microhaplotypes for mixture deconvolution. Forensic Science International. Genetics Supplement Series, 7(1), 522–524. https://doi.org/10.1016/j.fsigss.2019.10.075

Parson, W., Ballard, D., Budowle, B., Butler, J., Gettings, K., Gill, P., . . . Phillips, C. (2016). Massively parallel sequencing of forensic STRs: Considerations of the DNA commission of the International Society for Forensic Genetics (ISFG) on minimal nomenclature requirements. *Forensic Science International: Genetics, 22*, 54-63. doi:10.1016/j.fsigen.2016.01.009

People v. Castro, 144 NY 956, 970 (1989)

Puers, C., Hammond, H., Jin, L., Caskey, C., & Schumm, J. (1993). Identification of repeat sequence heterogeneity at the polymorphic short tandem repeat locus HUMTH01[AATG]n and reassignment of alleles in population analysis by using a locus-specific allelic ladder. *American journal of human genetics*, *53*(4), 953–958.

Reeder, D. (1999). Impact of DNA typing on standards and practice in the forensic community. *Archives of pathology & laboratory medicine, 123*(11), 1063-1065.

Rockenbauer, E., Hansen, S., Mikkelsen, M., Børsting, C., & Morling, N. (2014). Characterization of mutations and sequence variants in the D21S11 locus by next generation sequencing. *Forensic Science International: Genetics, 8*(1), 68-72. doi:10.1016/j.fsigen.2013.06.011

Rouder, J., Engelhardt, C. R., McCabe, S., & Morey, R. D. (2016). Model comparison in ANOVA. Psychonomic Bulletin & Review, 23(6), 1779–1786. https://doi.org/10.3758/s13423-016-1026-5

Ruitberg, C., Reeder, D., Butler, J. (2001) STRBase: a short tandem repeat DNA database for the human identity testing community. Nucleic Acids Res. 29(1): 320-322. DOI: doi.org/10.1093/nar/29.1.320

Scheible, M., Loreille, O., Just, R., & Irwin, J. (2014). Short tandem repeat typing on the 454 platform: Strategies and considerations for targeted sequencing of common forensic markers. Forensic Science International : Genetics, 12, 107–119. https://doi.org/10.1016/j.fsigen.2014.04.010

Schneider, P., Prager-Eberle, M., Rittner, C. (1991) Use of polymerase chain reaction of the HLA-DQ alpha system in forensic trace analysis. Archiv fur Kriminologie.

Scientific Working Group on DNA Analysis Methods. (2016, December). " Validation Guidelines for DNA Analysis Methods. https://www.swgdam.org/_files/ugd/4344b0_813b241e8944497e99b9c45b163b76bd.pdf

Scientific Working Group on DNA Analysis Methods. (2019, April). Addendum to "SWGDAM Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories" to Address Next Generation Sequencing. https://www.swgdam.org/_files/ugd/4344b0_91f2b89538844575a9f51867def7be85.pdf

Sharma, V., Young, B., Armogida, L., Khan, A., & Wurmbach, E. (2022). Evaluation of ArmedXpert software tools, MixtureAce and Mixture Interpretation, to analyze MPS-STR data. Forensic Science International : Genetics, 56, 102603–102603. https://doi.org/10.1016/j.fsigen.2021.102603

Silvia, A., Shugarts, N., & Smith, J. (2017). A preliminary assessment of the ForenSeq™ FGx System: next generation sequencing of an STR and SNP multiplex. International Journal of Legal Medicine, 131(1), 73–86. https://doi.org/10.1007/s00414-016-1457-6

State v Schwartz, 447 M.N. C5-89-460 (1989)

Steffen, C., Coble, M., Gettings, K., Vallone, P. (2017) Corrigendum to 'U.S. Population Data for 29 Autosomal STR Loci' [Forensic Sci. Int. Genet. 7 (2013) e82-e83]. Forensic Sci. Int. Genet. 31, e36–e40

Sukawutthiya, P., Sathirapatya, T., & Vongpaisarnsin, K. (2017). Considering a performance test of forensic genomics system on massively parallel sequencing technology. *Forensic Science International: Genetics Supplement Series, 6*(C), e599-e600. doi:10.1016/j.fsigss.2017.10.007

Sullivan K., Gill, P., Lingard, D., Lygo, J. (1992). Characterization of HLA DQ alpha for forensic purposes. Allele and genotype frequencies in British Caucasian, Afro-Caribbean and Asian populations. *Int J Legal Med*. 105(1):17-20. doi: 10.1007/BF01371231.

Tabak, J. (2019). FBI Approves Verogen's Next-Gen Forensic DNA Technology for National DNA Index System [Press release]

Tan, Y., Bai, P., Wang, L., Wang, H., Tian, H., Jian, H., . . . Zhang, L. (2018). Two-person DNA mixture interpretation based on a novel set of SNP-STR markers. *Forensic Science International: Genetics, 37*, 37-45. doi:10.1016/j.fsigen.2018.07.021

Testimony by Expert Witnesses, 28 USC 22 § 702 (1975)

Torres, Y., Flores, I., Prieto, V., López-Soto, M., Farfán, M. a. J., Carracedo, A., & Sanz, P. (2003). DNA mixtures in forensic casework: a 4-year retrospective study. *Forensic science international, 134*(2), 180-186. doi:10.1016/S0379-0738(03)00161-0

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

Urquhart, A., Kimpton, C. P., Downes, T. J., & Gill, P. (1994). Variation in short tandem repeat sequences--a survey of twelve microsatellite loci for use as forensic identification markers. International Journal of Legal Medicine, 107(1), 13–20. https://doi.org/10.1007/BF01247268

van der Gaag, K., de Leeuw, R.H., Hoogenboom, J., Patel, J., Storts, D.R., Laros, J.F.J., & de Knijff, P. (2016). Massively Parallel Sequencing of Short Tandem Repeats—Population data and mixture analysis results for the PowerSeq™ system. *Forensic Science International : Genetics*, *24*, 86–96. https://doi.org/10.1016/j.fsigen.2016.05.016

Van Neste, C., Gansemans, Y., De Coninck, D., Van Hoofstat, D., Van Criekinge, W., Deforce, D., & Van Nieuwerburgh, F. (2014). Forensic massively parallel sequencing data analysis tool: Implementation of MyFLq as a standalone web- and Illumina BaseSpace® - application. Forensic Science International : Genetics, 15, 2–7. https://doi.org/10.1016/j.fsigen.2014.10.006

Verogen Inc.; First Criminal Conviction Secured with Next-Gen Forensic DNA Technology. (2019). In (pp. 1024). Atlanta.

Von Bubnoff, A., (2008). Next-Generation Sequencing: The Race Is On. Cell, 132(5), 721–723. https://doi.org/10.1016/j.cell.2008.02.028

Votrubova, J., Ambers, A., Budowle, B., & Vanek, D. (2017). Comparison of standard capillary electrophoresis based genotyping method and ForenSeq DNA Signature Prep kit (Illumina) on a set of challenging samples. *Forensic Science International: Genetics Supplement Series, 6*, e140-e142. doi:10.1016/j.fsigss.2017.09.027

Wang, L., He, W., Mao, J., Wang, H., Jin, B., Luo, H. B., . . . Zhang, L. (2015). Development of a SNP-STRs multiplex for forensic identification. *Forensic Science International: Genetics Supplement Series, 5*(C), e598-e600. doi:10.1016/j.fsigss.2015.09.236

Wang, L., Schneider, P. M., Rothschild, M. A., Bai, P., Liang, W., & Zhang, L. (2013). SNP–STR polymorphism: A sensitive compound marker for forensic genetic applications. *Forensic Science International: Genetics Supplement Series, 4*(1), e206-e207. doi:10.1016/j.fsigss.2013.10.106

Wang, Z., Zhou, D., Wang, H., Jia, Z., Liu, J., Qian, X., Li, C., & Hou, Y. (2017). Massively parallel sequencing of 32 forensic markers using the precision ID globalfiler™ NGS STR panel and the ION PGM™ system. Forensic Science International: Genetics, 31, 126–134. https://doi.org/10.1016/j.fsigen.2017.09.004

Warshauer, D. (2015) Development of a Comprehensive Massively Parallel Sequencing Panel of Single Nucleotide Polymorphism and Short Tandem Repeat Markers for Human Identification. Doctor of Philosophy (Biomedical Sciences)

Warshauer, D., Lin, D., Hari, K., Jain, R., Davis, C., LaRue, B., King, J. L., & Budowle, B. (2013). STRait Razor: A length-based forensic STR allele-calling tool for use with second generation sequencing data. Forensic Science International : Genetics, 7(4), 409–417. https://doi.org/10.1016/j.fsigen.2013.04.005

Wätzig, H., Degenhardt, M. and Kunkel, A. (1998), Strategies for capillary electrophoresis: Method development and validation for pharmaceutical and biological applications. Electrophoresis, 19: 2695-2752. https://doi.org/10.1002/elps.1150191603

Weber, J., Wong, C. (1993), Mutation of human short tandem repeats, Human Molecular Genetics, 2, 8, Pages 1123–1128, https://doi.org/10.1093/hmg/2.8.1123

Wendt, F. R., King, J. L., Novroski, N. M. M., Churchill, J. D., Ng, J., Oldt, R. F., . . . Budowle, B. (2017). Flanking region variation of ForenSeq™ DNA Signature Prep Kit STR and SNP loci in Yavapai Native Americans. *Forensic Science International: Genetics, 28*, 146-154. doi:10.1016/j.fsigen.2017.02.014

Williamson, V. R., Laris, T. M., Romano, R., & Marciano, M. A. (2018). Enhanced DNA mixture deconvolution of sexual offense samples using the DEPArray™ system. *Forensic Science International: Genetics, 34*, 265-276. doi:10.1016/j.fsigen.2018.03.001

Xavier, & Parson, W. (2017). Evaluation of the Illumina ForenSeq™ DNA Signature Prep Kit – MPS forensic application for the MiSeq FGx™ benchtop sequencer. Forensic Science International : Genetics, 28, 188–194. https://doi.org/10.1016/j.fsigen.2017.02.018

Yang, M., Yin, C., Lv, Y., Yang, Y., Chen, J., Yu, Z., . . . Yan, J. (2016). Development of a rapid 21-plex autosomal STR typing system for forensic applications. Electrophoresis. 37. 10.1002/elps.201500498.

Zhang, H., Wheeler, D., Yakub, Wei, S., Sood, R., Rowe, W., Liu, P., Gibbs, R., & Buetow, K. (2005). SNPdetector: A software tool for sensitive and accurate SNP detection. PLoS Computational Biology, 1(5), 395–404. https://doi.org/10.1371/journal.pcbi.0010053

Zhang, Q., Wang, X., Cheng, P., Yang, S., Li, W., Zhou, Z., & Wang, S. (2022). Complex kinship analysis with a combination of STRs, SNPs, and indels. Forensic Science

International : Genetics, 61, 102749–102749. https://doi.org/10.1016/j.fsigen.2022.102749

Zhang, S., Bian, Y., Tian, H., Wang, Z., Hu, Z., & Li, C. (2015). Development and validation of a new STR 25-plex typing system. *Forensic Science International: Genetics, 17*, 61-69. doi:10.1016/j.fsigen.2015.03.008

Zhang, S., Niu, Y., Bian, Y., Dong, R., Liu, X., Bao, Y., . . . Li, C. (2018). Sequence investigation of 34 forensic autosomal STRs with massively parallel sequencing. *Scientific Reports, 8*(1), 6810-6810. doi:10.1038/s41598-018-24495-9

Appendix: Comparison of statistics using STR model and proposed MPS model

Comparison of statistics using STR model and proposed MPS model					
STR Locus	Control DNA 007*	Allele Bracketed Sequence	Allele 1 Freqeucy	STR Statistic	MPS Statistic
CSF1PO	11	[ATCT]11	0.2165	0.1673	0.0002677
	12	[ATCT]5 GTCT [ATCT]6	0.3866		
D10S1248	12	[GGAA]12	0.0876	0.0355	0.0355
	15	[GGAA]15	0.2062		
D12S391	18	[AGAT]10 [AGAC]8	0.2629	0.09217	?
	19	[AGAT]10 [AGAC]8 AGAT	0.1753		
D13S317	11	[TATC]11	0.2680	0.07182	0.07182
	11	[TATC]11	0.2680		
D16S539	9	[GATA]9	0.3557	0.1173	0.1173
	10	[GATA]10	0.1649		
D18S51	12	[AGAA]12	0.0361	0.01299	0.01299
	15	[AGAA]15	0.1804		
D19S433	14	[CCTT]12 CCTACCTT	0.2990	0.03701	0.03701
	15	[CCTT]13 CCTACCTT	0.0619		
D1S1656	13	CCTA [TCTA]12	0.1340	0.05386	0.05386
	16	CCTA [TCTA]15	0.2010		
D21S11	28	[TCTA]5 [TCTG]5 [TCTA]3 TA [TCTA]2 TCA [TCTA]2 TCCATA [TCTA]11	0.0567	0.01402	0.00002665
	31	[TCTA]5 [TCTG]6 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCATA [TCTA]12	0.1237		
D22S1045	11	[ATT]8 ACT [ATT]2	0.2010	0.09117	0.09117
	16	[ATT]13 ACT [ATT]2	0.2268		
D2S1338	20	[GGAA]13 [GGCA]7	0.1598	0.0527	0.0527
	23	[GGAA]2 GGAC [GGAA]14 [GGCA]6	0.1649		
D2S441	14	[TCTA]11 TTTA [TCTA]2	0.0825	0.001699	0.001699
	15	[TCTA]12 TTTA [TCTA]2	0.0103		
D3S1358	15	[TCTA]1 [TCTG]1 [TCTA]13	0.3660	0.02414	0.02414
	16	[TCTA]1 [TCTG]2 [TCTA]13	0.3299		
D5S818	11	[ATCT]11	0.2732	0.07463	0.07463
	11	[ATCT]11	0.2732		
D6S1043	12	[ATCT]12	0.1237	0.001275	0.001275
	14	[ATCT]14	0.0052		
D7S820	7	[TATC]7	0.0052	0.001807	0.001807
	12	[TATC]12	0.1753		
D8S1179	12	[TCTA]12	0.1186	0.04767	0.00006673
	13	TCTA TCTG [TCTA]11	0.2010		
FGA	24	[GGAA]2 GGAG [AAAG]16 AGAA AAAA [GAAA]3	0.1495	0.009246	0.009246
	26	[GGAA]2 GGAG [AAAG]18 AGAA AAAA [GAAA]3	0.0309		
TH01	7	[ATGA]7	0.2680	0.02211	0.02211
	9.3	[ATGA]5 ATG [ATGA]4	0.0412		
TPOX	8	[AATG]8	0.5464	0.2985	0.2985
	8	[AATG]8	0.5464		
VWA	14	[TAGA]3 TGGA [TAGA]3 [CAGA]4 TAGA CTGA TAGA	0.1959	0.05452	0.0001854
	16	[TAGA]11 [CAGA]4 TAGA	0.1392		
* Using Control 007 with random bracketting from matching lengths a composit MPS DNA nomenclature was developed.					