Characterisation of apparent mismatches detected during routine Short Tandem Repeat analysis in parentage investigations

Dissertation

by

Researcher: Ms Afika Soldati

Supervisor: Prof André de Kock

Co-Supervisor: Mr Jean F. Kloppers

Submitted in fulfilment of the requirements in respect of the Master of Medical Sciences (MMedSc) degree in the department of Haematology and Cell Biology in the Faculty of Health Sciences at the University of the Free State

28 July 2023

UNIVERSITY OF THE FREE STATE
UNIVERSITEIT VAN DIE VRYSTAAT
YUNIVESITHI YA FREISTATA

UFS·UV
HEALTH SCIENCES

DECLARATION

I, Afika Soldati, declare that the Master's Degree research dissertation or interrelated publishable manuscripts/published articles, or casework Master's Degree mini dissertation that I herewith submit for the Master's Degree qualification in Medical Science in Haematology and Cell Biology at the University of the Free State is my independent work, and that I have not previously submitted it for a qualification at another institution of higher education.

A Soldati

Date: 28/07/2023

ACKNOWLEDGEMENTS

I would like to extend my heartfelt appreciation and gratitude to my wonderful supervisors:

- Prof André de Kock, for his expertise, leadership, motivation and contribution to the study.
- Mr Jean F. Kloppers, for his invaluable advice and constructive contribution to the study. Thank you for your patience, consistent guidance, and support throughout the years.

I would like to further express my gratitude to the following institutions and persons:

- To the National Health Laboratory Service (NHLS) Research Grant for providing funding for this study
- To the JB Marks Education Trust Fund, for awarding me the bursary that made my whole academic career possible. I am grateful for the financial assistance as well as the support for my personal growth and well-being.
- To the University of the Free State (UFS) and National Research Foundation (NRF) for providing me with partial financial support that made it possible to complete this study.
- To my colleagues and close friends in the department, for the constant motivation and encouragement. Thank you for always believing in me and for making this journey so memorable.
- To my loving family and friends, thank you for the continuous support and understanding even at the most trying times.
- To my dearest mother Neliswa Soldati, thank you for your prayers and wisdom, this achievement would be meaningless without your support.
- To my precious little sisters, Busisiwe Soldati and Ntandokazi Soldati, thank you for always looking up to me, you have been my greatest source of motivation. I dedicate this dissertation to you. May I continue to be your role model.

To God Almighty, thank you for blessing me more than I deserve.

CONTACT DETAILS

RESEARCHER

Ms Afika Soldati

University of the Free State

Department of Haematology and Cell biology

2nd floor, Block B, Room 409A

Email: soldatiafika@gmail.com

Contact no: 084 453 8554

SUPERVISOR

Prof André de Kock

University of the Free State and NHLS

Department of Haematology and Cell biology

Email: deKockA@ufs.ac.za

Contact no: 051 405 3283

CO-SUPERVISOR

Mr Jean F. Kloppers

University of the Free State and NHLS

Department of Haematology and Cell biology

Email: KloppersJF@ufs.ac.za

Contact no: 051 405 3283

TABLE OF CONTENTS

LIST OF ABBREVIATIONS AND ACRONYMS

°C	**-**	Degrees Celsius
%	**-**	Percentage
(-)	**-**	Loss of STR repeat unit(s)
(+)	**-**	Gain of STR repeat unit(s)
(-/+)	**-**	Loss or gain of STR repeat unit(s)
3'	**-**	3 prime
5'	**-**	5 prime
AAFS	**-**	American Academy of Forensic Sciences
bp	**-**	base pair(s)
BLAST	**-**	Basic Local Alignment Search Tool
BLAT	**-**	BLAST-like alignment tool
dNTP	**-**	Deoxynucleotide triphosphate
CE	**-**	Capillary Electrophoresis
CPA	**-**	Criminal Procedures Act
CPI	**-**	Combined Paternity Index
CODIS	**-**	Combined DNA Index System
DNA	**-**	Deoxyribonucleic Acid
dbSNP	**-**	database for Single Nucleotide Polymorphisms
EDNAP	**-**	European DNA Profiling
ENFSI	**-**	European Network of Forensic Science Institutes
EQA	**-**	External Quality Assessment
FBI	**-**	Federal Bureau of Investigation

FSL - Forensic Science Laboratory

FSS - Forensic Science Services

g - Grams

GDB - Genome Database

HLA - Human Leukocyte Antigen

HSREC - Health Science Research Ethics Committee

ISFG - International Society of Forensic Genetics

kb - Kilo base pair(s)

LR - Likelihood Ratio

Mb - Mega base pair(s)

min - Minute(s)

mM - Milimolar

MPS - Massively Parallel Sequencing

n - number

NCBI - National Center for Biotechnology Information

NDDSA - National DNA Database of South Africa

NDNAD - National DNA Database

NHLS - National Health Laboratory Services

NIST - National Institute of Science and Technology

ng - Nanograms

NGS - Next Generation Sequencing

nM - Nanomolar

NTC - non-Template Control

PCR - Polymerase Chain Reaction

PI	**-**	Paternity Index
PI	**-**	Principal Investigator
pH	**-**	Potential of Hydrogen
RFLP	**-**	Restriction Fragment Length Polymorphism
Rh	**-**	Rhesus factor
rpm	**-**	revolutions per minute
rs	**-**	Reference SNP
rSAP	**-**	Shrimp Alkaline Phosphate
SAPS	**-**	South African Police Services
SB	**-**	Sodium Borate
sec	**-**	Second(s)
SGM	**-**	Second-Generation Multiplex
SNP	**-**	Single Nucleotide Polymorphism
SNV	**-**	Single Nucleotide Variant
SRD	**-**	Standard Reference Database
STR	**-**	Short Tandem Repeat
STRBase	**-**	Short Tandem Repeat Database
SSR	**-**	Simple Sequence Repeat
T_m	**-**	Melting temperature
™	**-**	Trademark
U	**-**	Units
UK	**-**	United Kingdom
µL	**-**	Microliter
µM	**-**	Micromolar

USA	**-**	United States of America
US	**-**	United States
UV	**-**	Ultraviolet
V	**-**	Volt
VNTR	**-**	Variable Number Tandem Repeats
W	**-**	Probability of Paternity
ZR	**-**	Zymo Research

LIST OF FIGURES AND TABLES

Figures:

Tables:

SUMMARY

Background: Short Tandem Repeat (STR) analysis has proven effective for establishing parentage and biological relatedness. There are commercially available STR kits that allow for reliable PCR amplification and genotyping of STR loci. However, one or two STR loci mismatches may be identified in non-exclusion cases. In routine analysis, these discrepancies are classified as apparent STR loci mismatches. The mismatches result from various mutational mechanisms. However, the mechanisms that drive these mutations are poorly understood. Several STR loci mismatches have previously been reported to impact parentage analysis. The alleles involved in the mismatch affect the interpretation of genetic profiles and can sometimes lead to false parentage exclusions. As such, it is essential to identify and characterise the underlying cause of STR loci mismatches for further validation of the genotypic data produced within a specific DNA profiling laboratory.

Methods: A laboratory-based descriptive-comparative study was conducted. This study consisted of 100 parentage cases with one or two STR loci mismatches from the DNA testing facility, Universitas Academic Unit, National Health Laboratory Services (NHLS) Bloemfontein from 1 January 2021 to 31 March 2022. The following 15 STR autosomal loci were included in the analysis: CSF1PO, FGA, vWA, D2S1338, D3S1358, D5S818, D6S1043, D7S820, D8S1179, D10S1248, D13S317, D16S539, D18S51, D19S433, and D21S11. Both published and designed study primers were used to optimise the PCR assay conditions for the amplification of the selected STR loci using commercially available control DNA. The optimised PCR assay conditions were used to screen the samples across the 15 STR loci. Sanger sequencing and sequence analysis was conducted for each parentage case to identify and characterise the underlying cause of the observed apparent STR mismatches. Furthermore, the sequence-based alleles were evaluated for concordance with genotypes determined by Capillary Electrophoresis-based (CE-based) STR typing previously reported by the facility.

Results: An average concordance of 82% was observed between STR profiling and Sanger sequencing across the 15 STR loci studied. In 11 of the loci, a 100%

concordance was obtained. In contrast, no concordance was observed for the D19S433 locus. The stepwise mutations observed at the various loci were 70% more frequent than other mutation models; these were attributed to DNA polymerase slippage. In comparison, 30% of the mutations were as a result of allelic dropouts, accounted for by primer-binding site sequence variants. It was observed that there were more mutations originating from paternal (n=76) rather than maternal (n=26) lineages.

Conclusion: The observation of one or two STR loci mismatches in parentage analysis should not be overlooked; all the studied allelic mismatches between the parent and child were characterised successfully. The findings revealed that most of the apparent mismatches occurred due to DNA polymerase slippage. The results of this study provide evidence that sequencing of the core STR repeat and the flanking regions can provide valuable information to characterise STR loci mutational events when inconclusive parentage or kinship results are obtained. Because of the limited sample size, the findings of this study provide evidence that STR mutations are more prevalent in males than females. Furthermore, this study demonstrates the need for DNA testing facilities to have a method in place to characterise and confirm inconclusive genotypic data obtained using the available commercial STR kits.

Keywords: Short tandem repeats (STRs), non-exclusion cases, apparent STR loci mismatches, DNA profiling, Capillary Electrophoresis-based (CE-based), DNA polymerase slippage, primer-binding site.

CHAPTER 1: INTRODUCTION

Human genetic identification is the process of identifying individuals by their deoxyribonucleic acid (DNA) sequence (Negi *et al*., 2014). For forensic reasons, human genetic identification is accomplished by interpreting genetic profiles. A genetic profile, often known as a genetic fingerprint, is the genotypic description of a group of genomic loci unique to that individual. At present, genetic profiles are obtained by analysing genetic markers called Short Tandem Repeats (STRs). STRs are utilised for human genetic identification (Budowle *et al*., 2003). This is in accordance with international recommendations made by several coordinating bodies including, the International Society of Forensic Genetics (ISFG), the European DNA Profiling Group (EDNAP), the National Institute of Science and Technology Short Tandem Repeat Internet Database (NIST-STRBase), and the Scientific Working Group on DNA Analysis Methods (SWGDAM) (Budowle *et al*., 2003; Jordan and Mills, 2021).

Human genetic variation has been the primary source of evolution in forensic genetics for more than a century (Jobling and Gill, 2004). The study of human genetic variation involves the analysis of observed polymorphic markers. This has permitted a better understanding and knowledge of the diversity and history of human populations (Pereira *et al.*, 2009). Additionally, it has provided a method for human genetic identification, paternity testing, kinship analysis, identification of perpetrators of major crimes like rape and murder, as well as the identification of missing persons' remains (Tomas, 2012).

Prior to the DNA era, the analysis of surface polymorphic antigen systems of ABO blood groups, the Human Leucocyte Antigen (HLA) complex, serum proteins, and red cell polymorphic isozymes was the basis of human identification (Jobling and Gill, 2004). More recently, STR analysis is considered a standard approach for human identification (Pereira *et al.*, 2009). STRs, also known as microsatellites or simple sequence repeats (SSRs), are tandem repeat units of 2-6 base pairs (bp) in length (Thomson *et al.*, 1999; Butler and Hill, 2012). These genetic markers are highly polymorphic and follow Mendelian patterns of inheritance. As a result, STR typing is

the method of choice in genetic analysis, providing a high level of discrimination to adequately address most human identification challenges (Pereira *et al.*, 2009).

DNA profiling of STR loci is a powerful tool in parentage analysis (biological maternity and paternity) and kinship analysis (Dakin and Avise, 2004). In paternity and maternity testing, STR analysis is performed to determine whether or not a particular male is the biological father or a particular female is the biological mother of a child, respectively (Flanagan and Jones, 2019). Parentage assignment is based on Mendelian principles; therefore, each individual inherits genetic material from both parents, at each locus, one allele is inherited from both biological parents (Flanagan and Jones, 2019). Deviations from this concept could result in the exclusion of a putative parent as the biological parent of a child (Negi *et al.*, 2006).

The human genome contains thousands of STR markers. However, a core set of 13 STR loci had been selected for application in parentage and kinship analysis (Butler *et al.*, 2007). This set is commonly known as the Federal Bureau of Investigation Combined DNA Index System (FBI CODIS STR loci) (Butler and Reeder, 1997). Over the years, more loci have been added to this system. Polymerase Chain Reaction (PCR) multiplex commercial STR kits are now widely available to DNA testing laboratories to generate DNA profiles containing these core STR loci. Some of the commonly used commercial STR kits from leading brands include PowerPlex 1.1, PowerPlex 1.2, PowerPlex 2.1, PowerPlex 16, PowerPlex 16 BIO, PowerPlex 18D and PowerPlex Fusion) from Promega Corporation and (AmpFLSTR VeriFiler, AmpFLSTR VeriFiler Plus AmpFLSTR Identifiler, AmpFLSTR SGM Plus, AmpFlSTR Profiler Plus, and AmpFlSTR COfiler from Applied Biosystems (Butler and Reeder, 1997, Welch *et al.*, 2012; Green *et al.*, 2021). PCR-based STR typing has become a standard technique in DNA analysis because of associated features such as specificity, sensitivity, the ability to amplify multiple loci (multiplexing) simultaneously, with automation capability. However, STR loci are prone to naturally occurring biological mutations which may result in STR mismatches. These are mostly accounted for by DNA polymerase slippage during DNA replication. Population specific-variants can result in primer binding-mutations leading to allelic dropout and null alleles being

observed in some commercially available STR kits (Westen *et al.*, 2014). These are mostly accounted for by DNA polymerase slippage during DNA replication (Lareu, 2013).

Numerous STR loci mismatches have been previously reported. These may be as a result of microsatellite variants (loss or gain of repeat unit in alleles) (Ali *et al.*, 2009) or primer binding site variants (null alleles as a result of allelic dropout) (Negi *et al.*, 2006). In addition, they can result in parent-child allelic mismatches during parentage analysis. Such allelic mismatches are incompatible with Mendelian patterns of inheritance (Ali *et al.*, 2009). As a result, this can contribute to the erroneous exclusion of a putative father in a paternal allelic mismatch and exclusion of a putative mother in a maternal allelic mismatch (Jia *et al.*, 2015). However, a single allele mismatch is not usually considered as an exclusion criterion. As a general rule, STR mismatches at more than two loci are sufficient to exclude a putative parent as a child's biological parent (Akhteruzzaman *et al.*, 2012). When STR mismatches at one or two loci are observed, the possibility of variants is considered. Therefore, this study will focus on the characterisation of apparent STR loci mismatches observed during parentage analysis.

CHAPTER 2: LITERATURE REVIEW

2.1 Evolution of forensic genetics

Previously, different blood groups were not considered; however, in 1900, Landsteiner discovered the human ABO blood groups polymorphisms (Watkins, 2001). Furthermore, Landsteiner and his colleague Alexander Weiner discovered the Rhesus factor (Rh factor) in 1940, another red blood cell protein (Schwarz, 2003; Mokoma, 2017). Together with the ABO system, this system allowed individuals to be identified based on their blood group. As a result, the first genetic tool applicable to solving criminal cases was developed (Mokoma, 2017). The main disadvantage of these markers was that they degraded quickly or were compromised by bacterial enzymes (Jobling and Gill, 2004).

Another highly polymorphic genetic system used in forensic laboratories was the Human Leukocyte Antigen (HLA) system. In the late 1970s, HLA typing for paternity testing was introduced (Decorte and Cassiman, 1993). Despite being a serological system, it was considered the human genetic system with the most polymorphisms (Choo, 2007). The HLA class II genes were widely used for forensic purposes. HLA *DRB1* was mainly used in paternity tests, while HLA *DQA1* was used in forensic cases. The *DRB1* gene is the most polymorphic of the HLA class II genes, with a 98% discrimination power and an 80% probability of paternity exclusion (Decorte and Cassiman, 1993). Due to challenges in evaluating alleles, it was difficult to distinguish each genotype present in a mixture, so its use in forensic casework was limited. (Decorte and Cassiman, 1993). **Figure 2.1** summarises some of the developments made in forensic genetics.

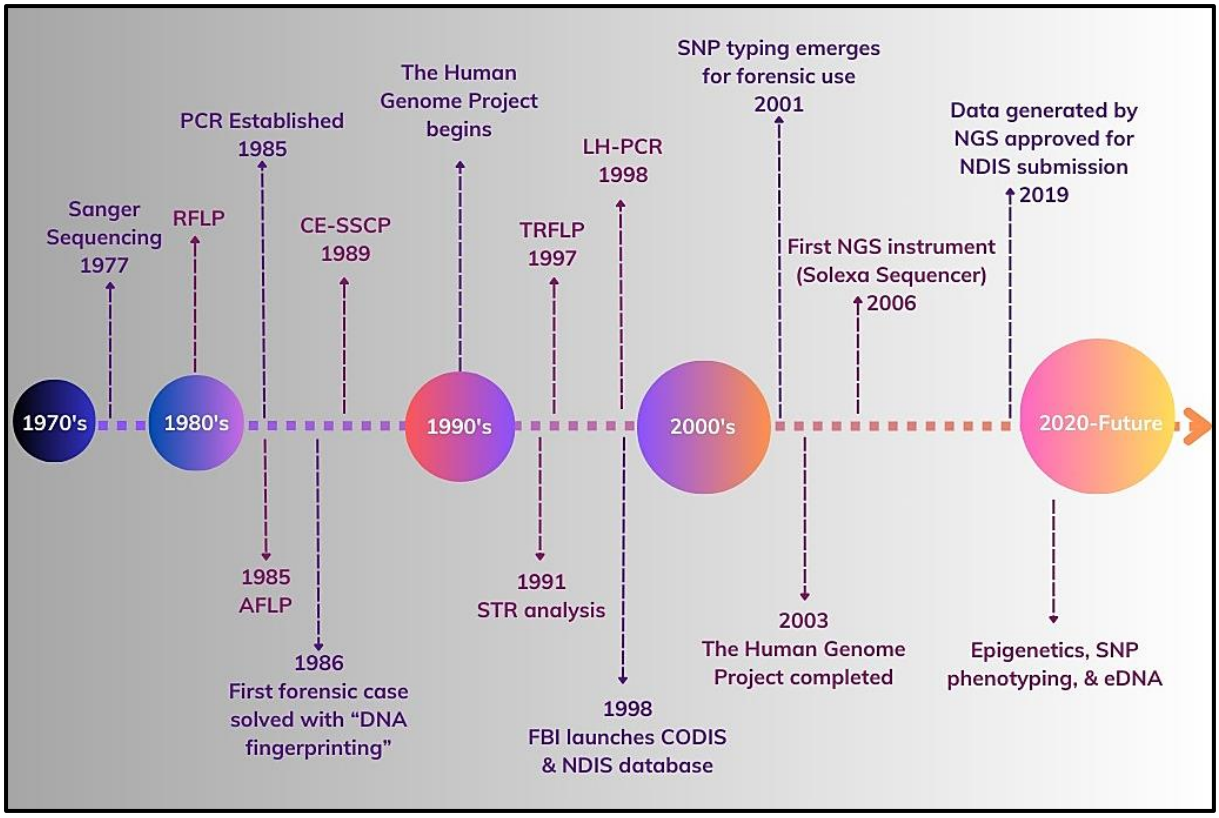


Figure 2.1: A timeline to summarise some of the developments of DNA typing technologies from the 1970's to the present. Adapted from (Jordan and Mills, 2021).

2.2 History of DNA analysis

In 1984, Sir Alec Jeffreys discovered variable number tandem repeats (VNTRs), also referred to as minisatellites. In the following year, he used the restriction fragment length polymorphism (RFLP) approach to create a "DNA fingerprint" (Jeffreys *et al*., 1985). Because the DNA fingerprint was unique to each individual, this technology represented a significant development in the field of forensic science because individual identification could be made with some confidence (Van Camp and Dierickx, 2007), and the profiles were reproducible even with dried blood stains (Decorte and Cassiman, 1993). The main disadvantage of this approach was that it required a significant amount of intact input DNA for a "fingerprint" to be helpful, which posed a significant challenge for forensic samples. Furthermore, this method required

extensive labour, and it was time-consuming to produce a result, a resource that is most often unavailable in forensic casework. It was also challenging to resolve mixtures, where more than one genetic profile was present in a sample (Decorte and Cassiman, 1993).

Kary Mullis' invention of the PCR in 1986 (Mullis *et al*., 1986) increased the efficiency of DNA testing regardless of which system was used (Jeffreys *et al*., 1988). PCR improved both HLA and VNTR analysis because it allowed smaller amounts of input DNA and analysis took less time. As a result, PCR-based methods were preferred over RFLP-based methods. In the late 1980s, challenges with the use of PCR for the amplification of VNTRs led to the discovery of STRs. Shorter alleles were preferentially amplified during VNTR PCR analysis. Consequently, the longest alleles tended not to be detected (Decorte and Cassiman, 1993). Microsatellites became a solution to this problem due to the smaller size of the region to be amplified and thus the ability to be detected efficiently (Decorte and Cassiman, 1993).

2.3 Parentage analysis

Genetic systems to establish parentage began shortly after the ABO blood groups were described (Watkins, 2001). Since then, STR markers have become the cornerstone of parentage analysis. Mendel's principles of inheritance provide the fundamental concept underlying parentage analysis; each individual inherits genetic material from both parents, at each locus one allele is from the mother and the other one is from the father (Flanagan and Jones, 2019). Since STRs are inherited in a Mendelian manner (unlike multi-locus DNA fingerprinting) and have a large number of alleles per locus (unlike allozymes), this makes them the genetic makers of choice for parentage analysis (Flanagan and Jones, 2019).

At present, most parentage testing laboratories perform parentage analysis using panels of 9-23 unlinked autosomal STRs (Lane, 2013). One additional locus called Amelogenin (AMEL) is used for sex determination (Jia *et al.*, 2015). The general rule

in paternity testing is that when more than two STR loci mismatches are observed between the child and the putative father, the putative father is excluded as the child's biological father (Lareu, 2013; De Kock and Kloppers, 2021). Furthermore, the putative father is not excluded on one or two observed STR loci mismatches due to possible STR mutations (Lareu, 2013). When uncertain, the paternity index (PI) should be calculated (Gjertson *et al.*, 2007). The PI indicates the probability of paternity between the tested individuals. It is defined by contrasting two likelihoods (X/Y), where X is the chance of passing the obligate allele from the tested man, and Y is the chance of passing the obligate allele from a random man (Gjertson *et al.*, 2007; Akhteruzzaman *et al.*, 2012). In essence, the first likelihood (X) is the genetic probability of observing the genotypes of the tested individuals based on the condition that they are a biological family, and the second likelihood (Y) is the probability of observing the same genotypes on condition that someone other than the putative parent is the child's biological parent (Lane, 2013). Since the autosomal STR loci used are unlinked, the combined paternity index (CPI) can be determined (Lane, 2013). Additionally, a probability of paternity (*W*) can be determined using CPI and a prior probability of 0.5, using the formula shown in **Figure 2.2** (Jia *et al.*, 2015). In most countries, for a paternity inclusion, a CPI of 1000 and probability of paternity of 99.9% is the recommended threshold. A CPI of 1 indicates that the chance of parentage is equal to that of non-parentage. A CPI that is less than 1 indicates evidence of non-parentage, while a CPI that is greater than 1 is evidence of parentage (Akhteruzzaman *et al.*, 2012).

A typical paternity test consists of the biological mother, child, and putative father, commonly referred to as a standard trio (Jia *et al.*, 2015). In some cases, a paternity test can involve the child and the putative father only (motherless paternity testing) due to numerous socio-economic reasons, such as death or abandonment (Nothnagel *et al*., 2010). However, motherless paternity testing can lead to the false inclusion of a putative father due to coincidental STR loci matches between two unrelated individuals (De Kock and Kloppers, 2021). Furthermore, knowledge of the mother's genotype improves the chances of identifying the biological father (Nothnagel *et al*., 2010).

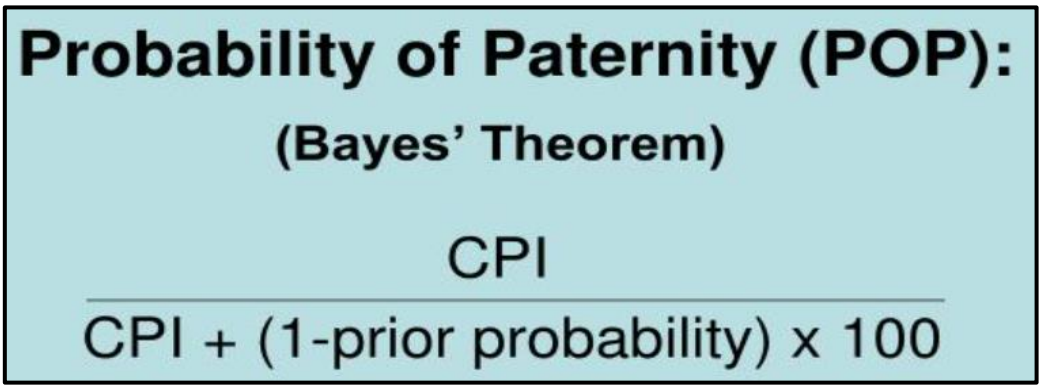


Figure 2.2: Probability of Paternity equation, adapted from (Press and Shigemasu, 1989; Manas, 1994).

2.4 Kinship analysis

A kinship test is performed to establish and provide genetic evidence as to whether two individuals are biologically related or not. Kinship cases are common and are generally required to confirm pedigree information for genetic studies. Furthermore, kinship analysis assists in reconnecting families torn apart by death, war, adoption, or immigration, as well as in the investigation of potential heirs in estate claims (Krawczak and Nothnagel, 2009).

Kinship analysis uses the level of genotypic similarity between individuals at the commonly used STR loci to determine their degree of familial relationship. Furthermore, it works on the principle that the genotypes of close relatives are more similar than those of unrelated individuals. The assessment is done by analysing existing genotypic data under various kinship assumptions, represented by different pedigree structures, or by directly determining kinship coefficients using computer algorithms and statistical methods. (Krawczak and Nothnagel, 2009). There are formulas established to determine biological relatedness between two individuals (Wenk *et al*., 1996). The principle and formulas used to determine kinship are described by Wenk *et al*. (1996). The k coefficient (k) is used to describe all possibilities of any two relatives sharing alleles by decent: probability of (k2+2k1+k0) =1.0. The k coefficient terms can be explained using two samples (S1 and S2) as

follows (Wenk, Traver and Chiafari, 1996): k2 = probability that both alleles of S2 are identical by descent of to those of S1
k1 = probability that one allele of S2, chosen at random, is identical by descent to one allele of S1, but the second allele is not 2k1 = probability that one or the other of the alleles in S2 is identical by descent to one of the alleles of S1, but the second is not
k0 = probability that neither alleles in S2 is identical by decent to alleles of S1.

2.5 Short tandem repeats (STRs)

There are various forms of DNA repeat sequences, composing nearly half of the human genome, which contribute to gene function, genome structure and evolution (Murat *et al*., 2020). STRs are among the most variable types of DNA repeat sequences (Ali *et al.*, 2009). Approximately 4,500,000 STR loci are estimated to be present in the human genome. They represent about 2.5% of the human genome (Negi *et al.*, 2006; Murat *et al*., 2020). STRs are scattered evenly throughout the genome. On average, they appear in every 10,000 nucleotides (Goodwin *et al.,* 2007; Wyner *et al.*, 2020). STRs are non-coding in nature and are, therefore, not implicated in gene expression (Wyner *et al.*, 2020).

The repeat structure of each STR marker is defined following the recommendations for nomenclature made by the International Society of Forensic Genetics (ISFG). The basic structure of a simple STR repeat is illustrated in **Figure 2.3**. STR sequences can be classified based on the number of repeats and the repeat unit length (Gill *et al.*, 1997). The sequence of an STR repeat motif can contain two to six nucleotides, such as di-, tri-, tetra-, and penta-, hexa-nucleotide repeats. However, in DNA testing laboratories, tetranucleotide repeats are the most commonly used STRs (Butler and Hill, 2012).

The number of repeats in STR markers can vary significantly among individuals, thus conferring different allele size ranges. Furthermore, STR repeats can be classified as simple, compound, and complex. Simple STR repeats contain identical units in

sequence and length. Compound STR repeats are made up of two or more contiguous simple repeats. Complex STR repeats may have repeat blocks of varying unit lengths and different intervening sequences (Gill *et al.*, 1997; Lareu, 2013).

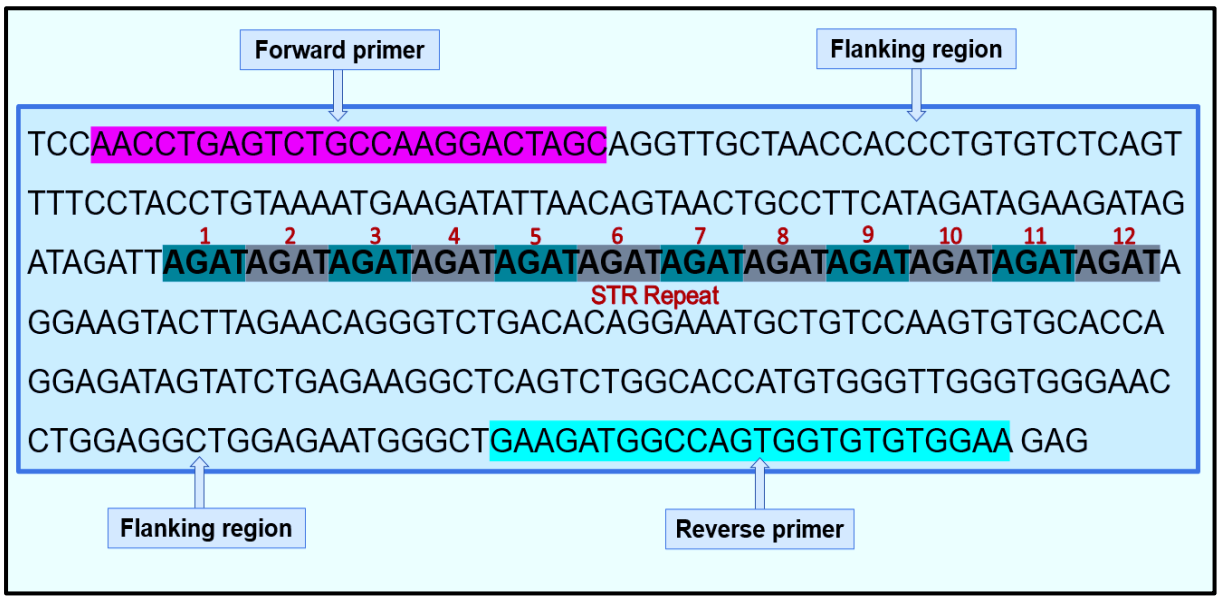


Figure 2.3: An illustration of the basic structure of an STR repeat sequence consisting of 12 repeats. Indicating the forward and reverse primers and the flanking regions. Adapted from (Lareu, 2013).

2.6 The utilisation of Short Tandem Repeats

STR analysis is a useful tool in various applications, including the construction of genetic linkage maps, population genetics, disease diagnosis, individual identification, and paternity testing (Fan and Chu, 2007). STR-based DNA typing has gained worldwide acceptance as a credible method for human identification. Over the years, the use of STRs in forensic DNA analysis has played a crucial role in the conviction or exoneration of suspects and in identifying victims of crimes, accidents, and mass disasters. Thus, aiding in emotional closure for bereaved families (Jobling and Gill, 2004).

STR markers have become the golden standard in resolving paternity disputes and determining biological relatedness. The broad genomic distribution and high variation has led to use of STRs in parentage and kinship testing (Jia *et al*., 2015). Thus, parentage tests using STR loci can exclude most falsely accused individuals by revealing STR loci mismatches between the putative parent and the child (Lane, 2013). The results of parentage or a kinship test can have various implications, including medical, judicial, and home affairs decisions (De Kock and Kloppers, 2021).

2.7 Short Tandem Repeat mutation mechanisms

Following the Mendelian principle, a child inherits one allele from each parent. However, deviations from this principle may be observed when using commercially available STR kits for parentage analysis. STR loci are prone to mutations due to their repetitive nature. They exhibit higher mutation rates than other types, such as single nucleotide polymorphisms (SNP) or copy number variants. However, the mechanisms that drive these mutations are poorly understood (Murat *et al*., 2020). The STR loci mismatches are usually caused by *de novo* mutations in one of the parental germlines (Dauber *et al.*, 2012). DNA polymerase slippage is the mutation mechanism mainly responsible for STR mutations (Lareu, 2013). Other mutational mechanisms contributing to the loss or gain of repeat units include recombinational events such as unequal crossing over or gene conversion (Dauber *et al*., 2012).

2.7.1 DNA polymerase slippage

During DNA replication, the DNA polymerase enzyme can dissociate from the nascent strand, and the displaced loops reanneal incorrectly to the repeat sequence. As a result, an insertion or deletion of tandem repeat units in the nascent strand occurs (Dauber *et al.*, 2012). This process is depicted in **Figure 2.4**. These repeat units are countered by repair pathways, notably mismatch repair, but a small fraction of events become fixed and transmitted across cell division (Murat *et al*., 2020).

Traditional models, such as the stepwise mutation model, have been theorised to explain further how STR alleles mutate due to DNA polymerase slippage (Lareu, 2013). This model has proven beneficial in analysing the differences in STR length between individuals (Murat *et al*., 2020). Sequencing data of the repeat motif and flanking regions can provide valuable information to characterise mutational events regarding the parental origin, change of repeat number (mutational step size), and the direction of the mutation (loss or gain of repeats). When using the stepwise mutation model, a single-step or one-step mutation results from a loss or gain of only one repeat unit, and a double-step or two-step mutation is because of a loss or gain of two repeats, and so forth. The study by Dauber *et al*. (2012) concluded that one-step mutations occur more frequently than multi-step mutations.

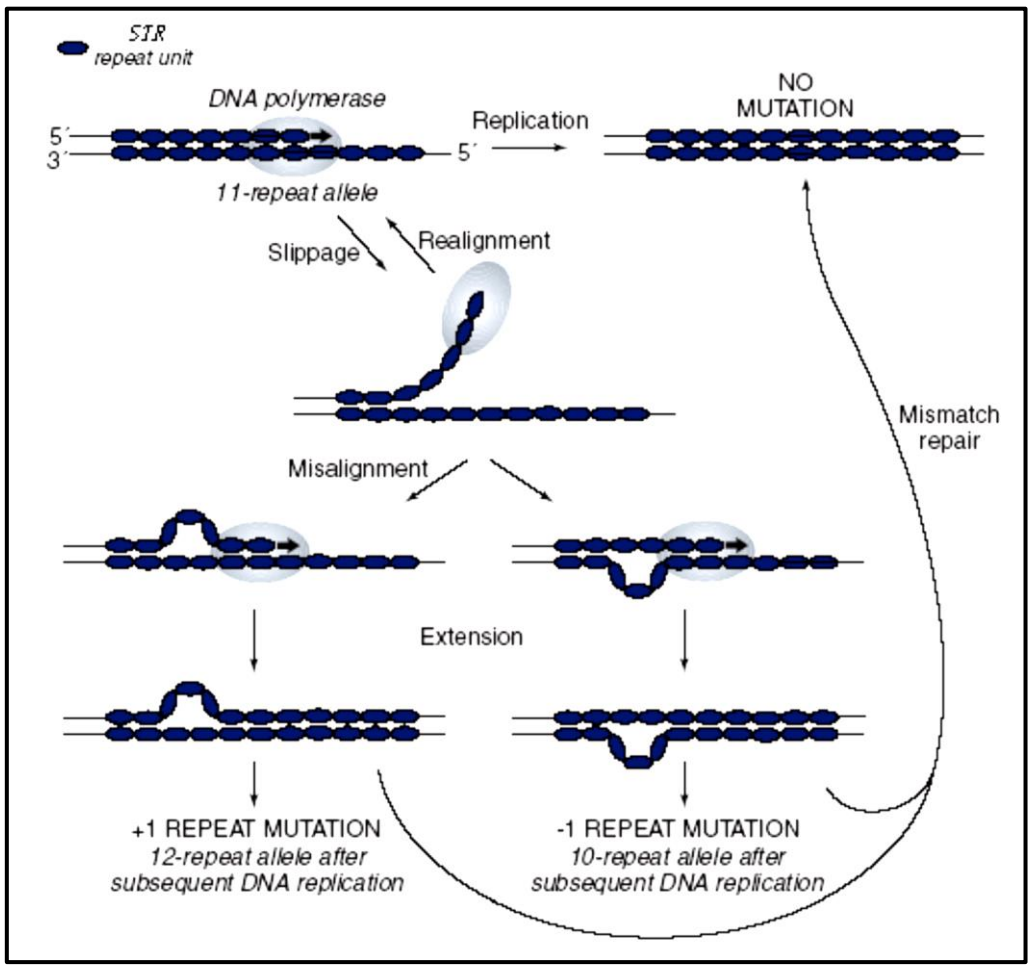


Figure 2.4: An illustration of DNA polymerase slippage, adapted from (Fan and Chu, 2007).

2.7.2 Short Tandem Repeat Primer-binding site mutations

Primer-binding site mutations have previously been reported to impact some of the PCR primers used in commercially available kits (Lareu, 2013). Sequence variations at STR loci primer-binding sites are known to arise at random. Such variations may be population-specific or unique to an individual. Sequence variation in the STR loci flanking regions results in primer design differences by commercial kit manufacturers. Thus, STR loci mismatches can be observed. PCR amplification of STR alleles with primer-binding site variants can result in little or no amplification. An allele that fails to amplify due to an allele dropout is a null or silent allele (Butler, 2006; Wang, 2010). Therefore, a heterozygous genotype will show a single peak as though it were a homozygote for the detectable allele. In contrast, a homozygous genotype will show no peak (Wang, 2010; Kline *et al*., 2011). This is illustrated in **Figure 2.5**.

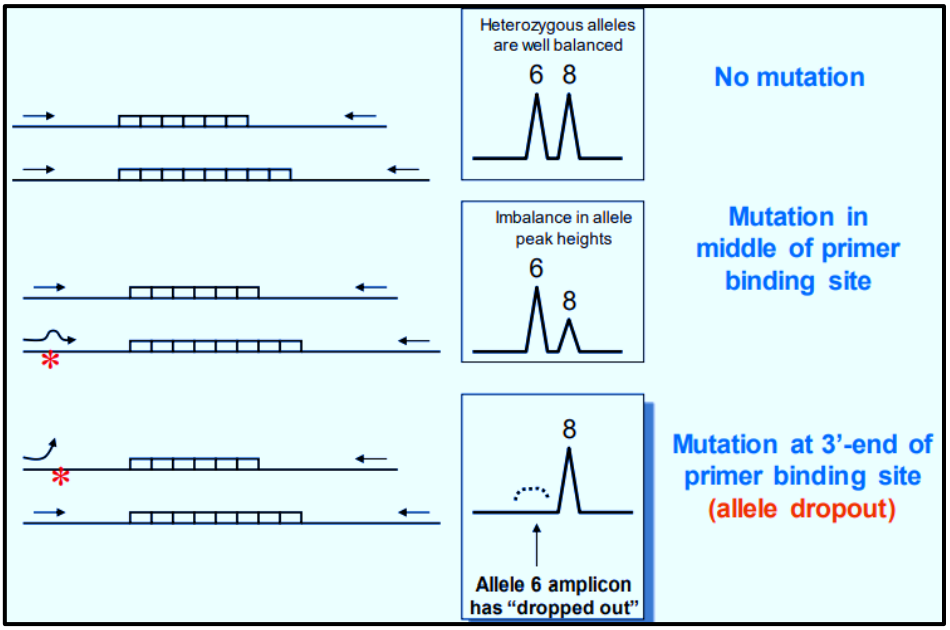


Figure 2.5: Diagram to illustrate the primer-binding site variant mechanisms. It indicates when there is no mutation affecting the alleles, a mutation present in the middle of the primer-binding site and a mutation located at the 3'- end of the primer- binding site. Adapted from (Coble and Hill, 2012).

2.7.3 Microvariant alleles

STR typing is performed with standardised allelic ladders containing the most common alleles to make size comparisons. These alleles have been sequenced to obtain the actual number of repeats (Butler, 2006). Allelic ladders with slightly varied size ranges may be supplied by manufacturers of commercially available STR kits. New alleles that do not size precisely with the ladder alleles are constantly discovered. It is important to note that not all STR alleles contain complete repeat units (Rockenbauer *et al*., 2011). In addition, nonconsensus alleles that fall between alleles with complete repeat units can be found in an STR locus. These are termed microvariant alleles or off-ladder alleles (Butler, 2006). The most common microvariant allele is the 9.3 allele found on the TH01 locus, which occurs in 33% of the US Caucasian population (Lareu, 2013).

2.7.4 Tri-allelic patterns/ genotypes

Although tri-allelic patterns are generally rare, they have been previously observed for many commonly used STR loci (Butler and Reeder, 1997). Tri-allelic patterns can result from triplicated chromosomes (trisomy), somatic variants, or localised chromosomal rearrangements. The tri-allelic patterns can be classified as type 1 and type 2 (Lareu, 2013). A type 1 tri-allelic pattern occurs as an imbalance between three alleles. The imbalance is usually observed when the sum of the peak heights for two alleles is equivalent to the third one (Butler, 2006). A type 2 tri-allelic pattern results from localised chromosomal rearrangement, and all three peak heights are equal. Usually, tri-allelic patterns do not complicate parentage analysis (Lareu, 2013). **Figure 2.6** illustrates the difference between the two types of tri-allelic patterns and the STR loci in which they are most commonly found. Data from the NIST STRBase website suggests that TPOX locus which occurs closest to the tip of chromosome 2 has the highest number of observed Tri-allelic patterns. Most of the observed tri-allelic patterns at the TPOX locus are type 2. Lane (2007) reported that approximately 2.4% of indigenous South Africans have tri-alleles at the TPOX loci. Furthermore, this study

indicated that the extra allele is almost always allele 10 and it segregates independently of those at the main TPOX locus.

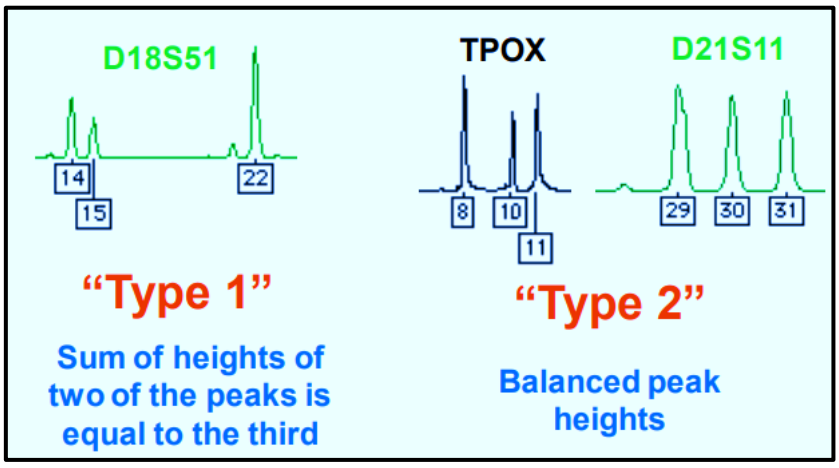


Figure 2.6: An illustration of the two types of tri-allelic pattern and the STR loci in which they most commonly occur, copied from (Coble and Hill, 2012)

2.8 History of Short Tandem Repeat marker selection

In the early 1990s, STR markers were described as effective and practical tools for human identity testing. In the search for new STR loci, the Forensic Science Services (FSS) began studies on population variation. The Royal Canadian Mounted Police and numerous European laboratories contributed to the early STR typing efforts. The first multiplex to be applied to forensic investigations consisted of four simple STRs: TH01, vWA, FES/FPS, and F13A1 (Jobling and Gill, 2004). Following this was a second-generation multiplex (SGM) with the TH01, vWA, D8S51, and D21S11 loci. The United Kingdom (UK) National DNA Database (NDNAD) was launched in April 1995, using the SGM loci and the amelogenin sex-typing test. Due to the success obtained by the UK in STR typing, the FBI laboratory led the United States' (US) efforts to establish the core STR loci, which formed the backbone of the US national database system, the Combined DNA Index System (CODIS) (Butler, 2006; Van Camp and Dierickx, 2007).

2.8.1 The Combined DNA Index System (CODIS)

In November 1997, 13 core STR loci were selected as the primary CODIS markers in the US national database. These were the CSFP1O, TPOX, FGA, vWA, TH01 D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, and D21S11 loci. The UK used ten core STR loci, including the D2S1338 and D19S433 loci and eight overlapping CODIS loci, FGA, vWA, TH01 D3S1358, D8S1179, D16S539, D18S51, and D21S11 (Butler, 2006)**.** The core STR loci have been located within the reference human genome sequence using a BLAST-like alignment tool (BLAT) (available at http://genome.ucsc.edu). The reference sequences of the core STR loci are available at http://www.cstl.nist.gov/biotech/strbase/seq_ref.htm (Butler, 2006). **Figure 2.7** depicts the human chromosome positions of the 52 autosomal STR makers that are currently used in forensic analysis systems (Fan *et al*., 2022).

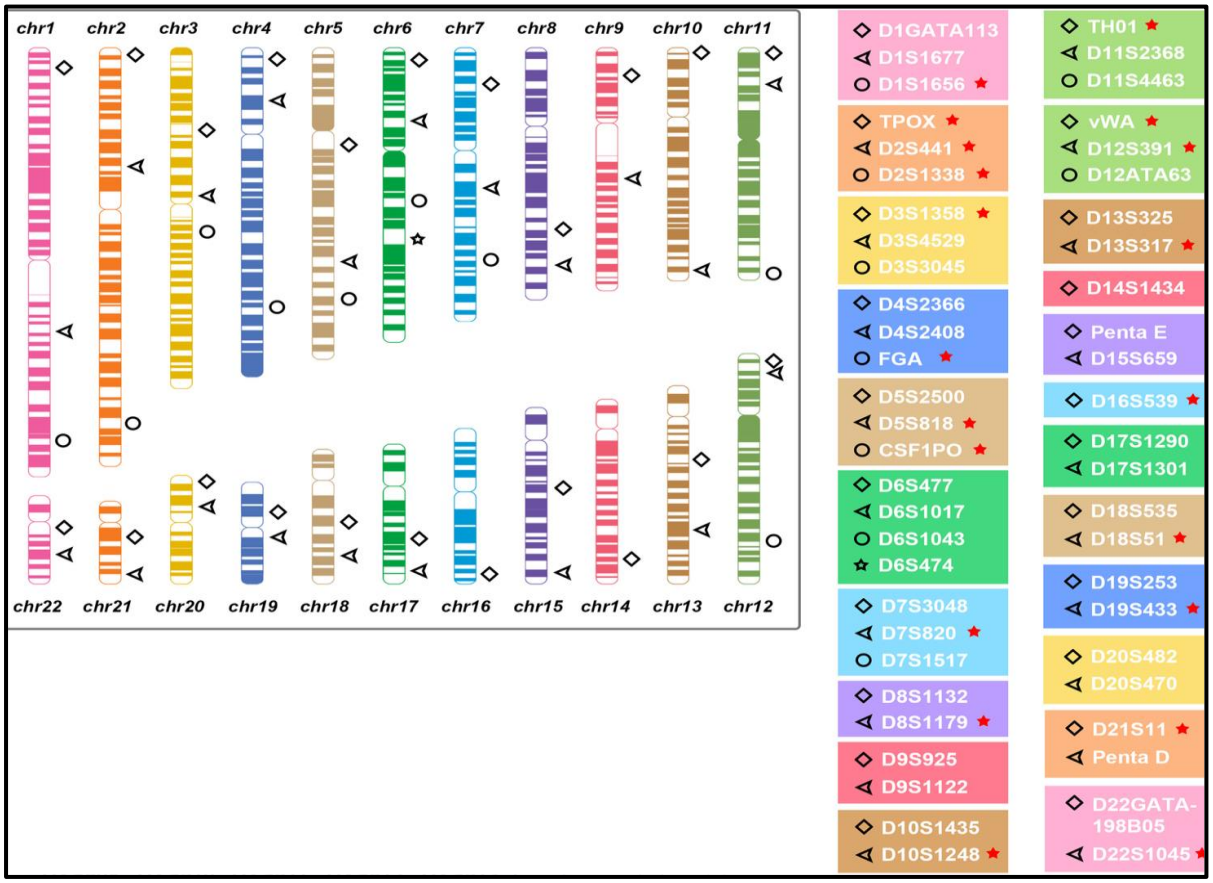


Figure 2.7: The human chromosome positions of the 52 autosomal STR makers that are currently used in forensic analysis systems (Fan *et al*., 2022).

2.8.2 National Institute of Standards and Technology Short tandem repeat database (NIST-STRBase)

The National Institute of Standards and Technology Short Tandem Repeat Internet Database, or STRBase (http://www.cstl.nist.gov/biotech/strbase), is one of the most extensive and frequently used internet resources on STR loci involved in human identification (Butler, 2006). This database aims to combine and organise the abundant scientific literature on STRs useful for the ongoing efforts in DNA typing. This database provides detailed information on the commonly used STR DNA markers. This includes information about the repeat structure, commercially available allelic ladders, PCR primer sets, standard multiplexes, and variant rates (Ruitberg *et al*., 2001). In 2005, the STRBase website was adopted by the NIST as an official Standard Reference Database (SRD), promoting the credibility of the information on the website (Butler, 2007).

2.8.3 International and national regulatory and governing bodies

Utilising a standard set of core STR loci allows national and international data exchange for human identity testing in paternity testing and forensic casework. This is mainly because of the availability of commercial STR kits. However, there are differences in practice jurisdictions because of historical, social, and legal circumstances. Despite these differences, the rapid developments and universal acceptance of new DNA-based technologies in forensic genetics are primarily due to the collaborations between international groups coordinated under various academic and government-sponsored institutions (Jobling and Gill, 2004). The international coordinating bodies in forensic genetics are described in **Table 2.1.** In South Africa, the collection and storage of DNA profiles for criminal intelligence purposes is governed by the Criminal Procedures Act (CPA) of 1977 and the Criminal Law (Forensic Procedures) Amendment Act 37 of 2013 (referred to as the 'DNA Act') (https://dnaproject.co.za/legislation-homepage/legislation/a-new-dna-act/). In 1989, South Africa set up its first DNA profile database called the National DNA Database of

South Africa (NDDSA). Currently, the South African Police Services (SAPS) Forensic Science Laboratory (FSL) undertakes all forensic investigations, such as, crime scenes and missing persons' remains. A number State and private laboratories do paternity tests, and individuals at educational institutions conduct research on a wide range variety of organisms using DNA profiling as a tool (www.pub.ac.za).

Table 2.1: International coordinating bodies in forensic genetics

Organisation/ subgroup	Purpose
International Society of Forensic Genetics (ISFG) • DNA commission • European DNA Profiling (EDNAP)	Promotes scientific knowledge in forensic genetics • Makes recommendations for the use of DNA • Harmonisation of European DNA technologies
European Network of Forensic Science Institutes (ENFSI)	Mainly represents government institutions; coordinates efforts to develop European DNA database
American Academy of Forensic Sciences (AAFS)	Academic body for North American forensic scientists
Federal Bureau of Investigation (FBI)	Responsible for setting standards, training, and developing the national DNA database
National Institute of science and technology (NIST) • STRBase	Supports the forensic community by organising collaborative proficiency exercises • Database giving characteristics of forensically useful STRs and SNPs

Adapted from (Jobling and Gill, 2004)

2.8.4 The VeriFiler™ Express PCR amplification kit

Currently, the VeriFiler™ Express PCR amplification kit (ThermoFisher Scientific, USA) is used for STR analysis in our setting (DNA testing facility, Universitas Academic Unit, NHLS, Bloemfontein). This kit uses a 6-dye chemistry formulation for paternity and kinship analysis. The VeriFiler Plus kit amplifies 23 autosomal STR loci: D3S1358, vWA, D16S539, CSF1PO, D6S1043, D8S1179, D21S11, D18S51, D5S818, D2S441, D19S433, FGA, D10S1248, D22S1045, D1S1656, D13S317, D7S820, Penta E, Penta D, TH01, D12S391, D2S1338, and TPOX, two internal quality control markers (IQCS and IQCL), one insertion/deletion polymorphic marker on the Y chromosome (Yindel), and Amelogenin (sex determining marker). A summary of the characteristics of the 15 STR loci included in this study is provided in **Table 2.2.**

Table 2.2: A summary of the characteristics of the 15 STR loci included in this study adapted from (Butler and Reeder, 1997; Gettings *et al*., 2015)

Locus designation	Chromosomal location	GenBank Accession	Repeat type classification	Repeat motif	Allelic range	Mutation rate (%)
1. CSF1PO	5q33.3-34	X14720	simple tetranucleotide	[AGAT]	5-16	0.16
2. FGA	4q28	M64982	Compound tetranucleotide	$[TTTC]_3$ [TTTT] [TTCT] $[CTTT]_n$ [CTCC] $[TTCC]_2$	12.2-51.2	0.28
3. vWA	12p13.31	M25858	Compound tetranucleotide	[TCTA] $[TCTG]_n$ $[TCTA]_n$ [TCCA] [TCTA]	10-25	0.17
4. D2S1338	2q35	AC010136/ AC010136	Compound tetranucleotide	$[TGCC]_n$ $[TTCC]_n$	11-28	0.12
5. D3S1358	3p21.31	AC099539	Compound tetranucleotide	[TCTA]/[TCTG]	8-20	0.12
6. D5S818	5q21-31	G08446/ AC008512	simple tetranucleotide	[AGAT]	6-18	0.11
7. D6S1043	6q16.1	G08539/ NT_007299	Compound tetranucleotide	[AGAT]	9-25	0.06

Table 2.2 continued

8. D7S820	7q11.21-22	G08616/ AC004848	Simple tetranucleotide	[GATA]	5-16	0.10
9. D8S1179	8q24.13	G08710/ AF216671	Compound tetranucleotide	[TCTA]	7-20	0.14
10. D10S1248	10q26.3	AL391869	Simple tetranucleotide	[GGAA]	8-19	-
11. D13S317	13q22-31	G09017/ AL353628.2	Simple tetranucleotide	[TATC]	5-17	0.14
12. D16S539	16q24.1	G07925/ AC024591.3	simple tetranucleotide	[GATA]	4-16	0.11
13. D18S51	18q21.33	X91254/ AP001534	simple tetranucleotide	AGAA	7-28.3	0.22
14. D19S433	19q12	G08036/ AC008507.6	Compound tetranucleotide	[AAGG][AAAG][AAGG][TAGG] $[AAGG]_n$	5.2-20	0.11
15. D21S11	21q11.2-q21	M84567	Complex tetranucleotide	$\{[TCTA]_3$ TA $[TCTA]_3$ TCA $[TCTA]_2$ TCCA TA}	12-41.2	0.19

2.8.5 Advantages and limitations of STR analysis

STR markers are highly polymorphic in nature, as the number of repeats at a given locus can vary among individuals. This variation owes to independent chromosomal assortment, recombination, and variants (Jobling and Gill, 2004). However, the significant advantage of using PCR-based methods in STR analysis is that the STR loci can be typed with a high level of specificity and sensitivity at an improved turn-around-time because PCR can amplify DNA from a template of a few cells (Goodwin *et al*., 2007). This high level of specificity and sensitivity makes STR analysis feasible and reproducible even with forensic samples of low DNA quantity and poor quality (Dauber *et al.*, 2012). Moreover, PCR amplification of degraded DNA does not present the problem of preferential amplification, but results in the dropout of larger alleles. PCR-based STR multiplex analysis allows STR detection to be easily automated (Butler, 2015a).

The availability of commercial STR kits that use a uniform set of core STR loci offers the advantage of national and international data sharing for forensic casework and parentage analysis (Butler, 2006). In addition, the commercially available STR multiplex kits allow for more STR loci to be examined and compared. This results in STR profiles with high discriminating power (the chance that two unrelated individuals will have different genotypes) among both related and unrelated individuals (El-Alfy and Abd El-Hafez, 2012)

The main limitation of STR analysis is that STRs are prone to mutations (Akhteruzzaman *et al.*, 2012). The relatively high mutation rate has proven to be a disadvantage in parentage and kinship analysis (Lane, 2013). Furthermore, most STRs are less polymorphic than VNTRs, thus having less discrimination power per locus due to the small number of alleles and less heterozygosity per locus than VNTRs. Therefore, more STR loci must be typed for comparable results with VNTRs. However, VNTRs require a considerable amount of input DNA and are currently not used frequently (Kebede, 2002).

Although using a uniform set of core STR loci as commercially available kits has provided national and international data sharing (Butler, 2006), genetic discrepancies

can be observed. This is due to differences in primer design for the same loci in the different commercial kits (Westen *et al.*, 2014). Additionally, the contents of the commercial kits may not be completely known to the user, for example, the primer sequences may be proprietary (Coble and Hill, 2012).

2.8.6 Methods of detection of Short tandem repeats

STR typing generally involves isolating DNA from biological material such as blood, semen, saliva, urine, teeth, bone, tissue, and hair strands, or the use of dried blood stains on Fast Technology for Analysis (FTA) of nucleic acids. This step is followed by DNA quantification or direct amplification when FTA cards are used, multiplex PCR amplification of STR loci, separation of the PCR amplicons on a genetic analyser, and bioinformatics to analyse the resulting data (Udogadi *et al.*, 2020).

DNA analysis utilises PCR to amplify STR loci. In one PCR reaction, more than one set of primers may be added to target multiple regions. This is commonly referred to as multiplex PCR (Lareu, 2013). In PCR-based STR multiplexing, primers are labelled with different fluorescent dyes. This allows for the resolution of STR loci with alleles that fall in the same size range (Butler *et al.*, 2004). Several different STR multiplex kits are commercially available. Two major suppliers of these kits are Applied Biosystems (Foster City, California) and Promega Corporation (Madison, Wisconsin), and several STR assays have been developed by both companies (Lareu, 2013). Capillary electrophoresis (CE) is performed to separate STR alleles using systems such as, the ABI PRISM system, the Spectrum CE System, and the Spectrum Compact CE System. Applied Biosystems and Promega Corporation provides the most commonly used CE systems. Following CE, data analysis and interpretation is performed using software, such as, GeneMapper® ID, GeneMapper® ID-X, GeneMarker® and GeneMarker®HID (Butler and Reeder, 1997).

Massively parallel sequencing (MPS) technology, or next-generation sequencing (NGS), offers a new high-throughput research tool for biological sciences.

Furthermore, because NGS technology can produce hundreds of sequences in a single reaction, a growing number of researchers are using it in paternity testing and for forensic purposes (Zhang *et al*., 2018). The most important feature of NGS is that it can sequence STR loci and offer accurate composition information, such as product length and repeat structure (Eduardoff *et al*., 2015). Unlike conventional capillary electrophoresis-based (CE-based) STR typing, NGS technology is not limited by the number of fluorescent dyes or loci with overlapping size ranges. As a result, in principle, NGS might result in more STR locus variations than CE-based STR typing (Poethe *et al*., 2023). However, the available commercial NGS systems have the drawback of being developed for high throughput sequencing, often requiring specially customised equipment with specialised software for allele identification, making analysis less flexible and costly, and unsuitable for low throughput analysis, this poses major challenges to resource-limited laboratories (Poethe *et al*., 2023).

2.8.6.1 Possible methodology-related events that may result in apparent STR loci mismatches

During routine STR analysis in parentage and kinship investigations, methodology-related events may complicate the detection and interpretation of STR profiles (Goodwin *et al.,* 2007). These events may result in apparent erroneous exclusions. Some commonly observed events include using DNA of low quantity or poor quality obtained during DNA extraction (Butler, 2015a). This may lead to incomplete STR profiles. Moreover, preferential amplification may be observed, where one allele at a locus is amplified with greater efficiency than the other allele (Goodwin *et al.,* 2007).

The methodology-related events that are most likely to occur are observed during the visualisation and analysis of STR profiles. These include the formation of stutter peaks, split peaks, pull-ups, and incorrect allele sizing (Goodwin *et al.,* 2007). A stutter peak can be produced during multiplex PCR amplification of an STR allele (Martinez *et al.,* 2017). Stutter peaks are formed by strand slippage during the extension of the template DNA strand; this results in an STR allele that is one repeat unit smaller or

larger than the true allele (Butler, 2015a). Therefore, the interpretation of STR profiles requires experienced individuals to ensure the results are robust and consistent (Goodwin *et al*., 2007).

2.9 Conclusion

STRs are widely used for forensic investigations and parentage analysis because the selected STR loci show a high degree of polymorphism. STR analysis is feasible and reproducible even with low amounts of degraded DNA. Moreover, the application of STR markers in parentage and kinship analysis requires a solid understanding of the patterns of inheritance and reasonable estimation of the mutation rates. While STR markers are the most effective genetic markers for determining parentage and kinships, numerous STR loci mismatches have previously been reported, including microsatellite and primer-binding site variants. These variants could potentially result in apparent erroneous exclusions of a putative parent as the biological parent in parentage analysis. Additionally, they can impact kinship calculations and thus the outcome of the relationship in question. Therefore, it is essential to characterise STR loci variants and validate the commercial STR genotyping kits used within a specific DNA profiling laboratory.

2.10 Rationale, aim, and objectives

2.10.1 Rationale

Short tandem repeat analysis is widely used to perform parentage analysis due to the polymorphic nature of STRs. The commercially available STR kits allow for reliable PCR amplification and genotyping of STR loci. However, one or two STR loci mismatches may be observed in non-exclusion cases. In routine analysis, these discrepancies are classified as apparent STR loci mismatches. They are attributed to possible variants or possible methodology-related events. The STR alleles involved in these mismatches are rarely sequenced. Thus, the cause of the parent-child allelic

mismatches remains unclear. Moreover, DNA sequencing of the core STR repeat and the flanking regions can provide valuable information to characterise the mutational events. Therefore, mismatched STR genotypic data obtained during parentage investigations using commercially available kits needs to be confirmed by alternative methodologies including DNA sequencing, particularly in inconclusive outcomes.

2.10.2 Aim

The aim of this study was to characterise apparent STR loci mismatches observed during parentage investigations at the DNA testing facility, Universitas Academic Unit, NHLS Bloemfontein.

2.10.3 Objectives

Objective 1:

The first objective was to optimise PCR assays to amplify each of the selected STR loci where mismatches were observed by evaluating primers published on the STRBase website for target specificity, as they are used in some commercially available kits, as well as to design primers encompassing a larger flanking region to capture the full spectrum of repeats at a particular locus.

Objective 2:

The second objective was to perform Sanger sequencing of the selected STR loci observed in parentage cases with one or two STR loci mismatches.

Objective 3:

The last objective was to characterise the underlying cause of the apparent STR loci mismatches by determining whether the mismatches were attributed to possible variants or methodology-related events.

CHAPTER 3: METHODOLOGY

3.1 Study design

A laboratory-based descriptive-comparative study was conducted in the Department of Haematology and Cell Biology at the Faculty of Health Sciences, University of the Free State, Bloemfontein, South Africa.

3.2 Ethical considerations

This study was approved by an Evaluation Committee appointed by the Faculty of Health Sciences at the University of the Free State. Furthermore, study approval was obtained from the head of the Department of Haematology and Cell Biology and the Business Manager at the Universitas Academic Unit of the NHLS. Ethical clearance was obtained from the Health Science Research Ethics Committee (HSREC) in the Faculty of Health Science at the University of the Free State (UFS-HSD2021/1815/2501) **(Appendix E)**. No approval was requested from the Department of Health in the Free State as no patients were included in this study, only data from previous clients of the NHLS.

No informed consent was requested from the clients, as there was no interaction with the clients. According to the paternity exclusion rule, the parentage cases included in this study are non-exclusion cases; therefore, the outcome of this study will have no impact on the clients' paternity results and well-being. Furthermore, confidentiality was maintained as only pseudo-anonymisation samples with unique study numbers that cannot be traced back to the original parentage investigation were used.

In health sciences research, the use of personal information is essential. According to the Protection of Personal Information Act of 2013 (POPI Act), personal information is any information that can be used to identify an individual. The purpose of the POPI Act is to give effect to the constitutional right to privacy by ensuring that personal information is protected when it is collected, processed, kept, shared, and disposed of

by a responsible party. In compliance with the POPI Act, voluntary informed consent is required from an individual prior to using their personal information when required. Unless stated otherwise, the obtained information may only be used for the purpose for which received. Security preventive measures must be put in place to safeguard obtained personal information. In health sciences research, personal information is de-identified by assigning unique study numbers. Thus this information cannot be traced back to a specific individual. This study complied with all the regulations and guidelines of the POPI Act 4 of 2013.

3.3 Study population and selection

The NHLS DNA Testing Facility reviewed and identified 100 parentage cases with one or two STR loci mismatches from previous paternity and maternity reports. According to the facility's records, 99% of the individuals tested are of African ancestry. The ethnicity of this individuals is self-reported. The facility then performed the pseudo-anonymisation process of samples. Unique study numbers were assigned to each of the included parentage cases. A list of unique study numbers with the corresponding genotypic data was provided to the PI. The PI analysed all samples and genotype data pertaining to the included cases according to the unique study numbers. The PI had no knowledge of the clients' personal information and could not trace the unique study number to the original case number. However, the facility kept a separate list on a password-protected computer with the original parentage case numbers and genotypes to enable them to trace the samples back to the original parentage case. No informed consent was requested as there was no interaction with the paternity clients.

3.3.1 Inclusion criteria

This study included the following:

- A minimum of 100 parentage cases with one or two STR loci mismatches from 1 January 2021 to 31 March 2022.
  - 66 Standard trio cases
  - 25 Paternity duo cases
  - 9 Maternity duo cases
- Of these 100 parentage cases, a minimum of 200 samples with parent-child allelic mismatches were included.
- 15 STR loci that the paternity testing facility identified as the most frequently involved in apparent STR loci mismatches.

3.3.2 Exclusion criteria

This study excluded the following:

- Kinship investigations
- Parentage cases with more than two STR loci mismatches

3.4 Overview of the study procedure

A brief overview of the study procedure is illustrated in **Figure 3.1** below.

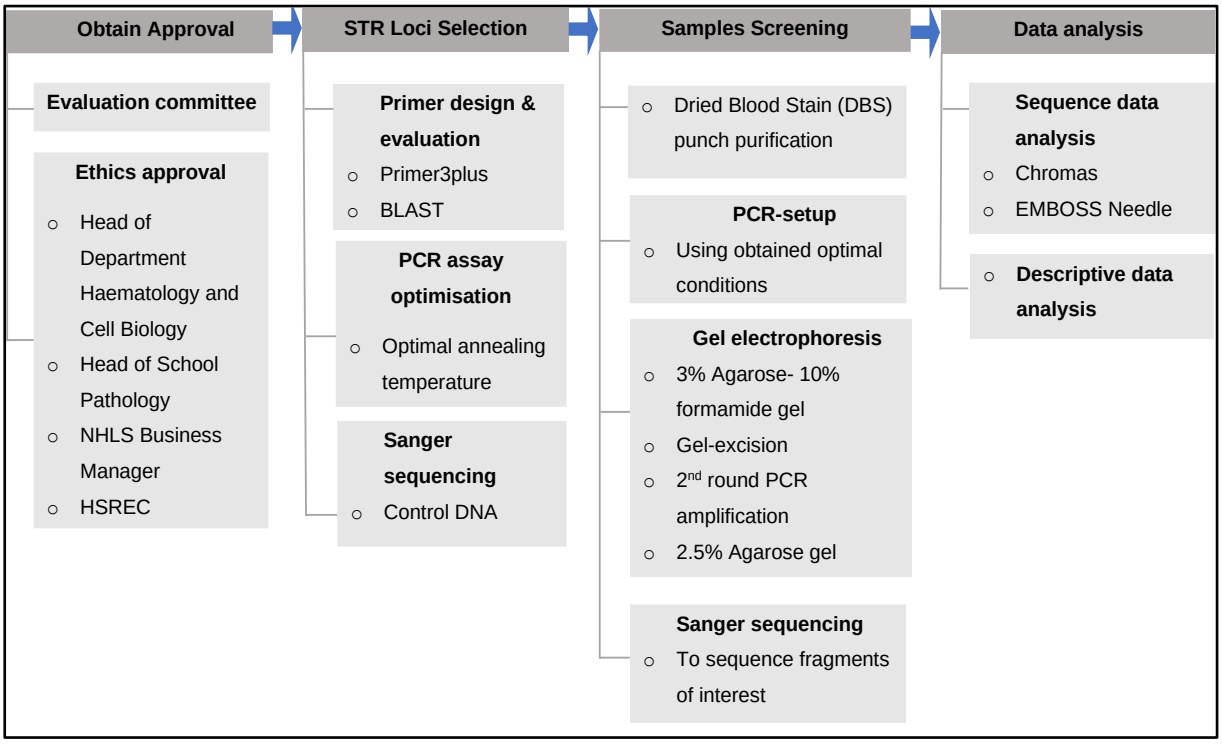


Figure 3.1: Schematic illustration of the study procedure that was followed

3.5 Primer design and evaluation

3.5.1 Evaluation of published primers

For each of the STR loci included in this study, published primers from the STRBase website were evaluated for target specificity using the Basic Alignment Search Tool (BLAST) (Altschul *et al*., 1990). **Table 3.1** indicates the primer sequences for each locus.

Table 3.1: Published primers sequences for each locus from the STRBase website

STR Loci	Primer sequences	Published by
CSF1PO	**F:**5'-AACCTGAGTCTGCCAAGGACTAGC-3' **R:**5'-TTCCACACACCACTGGCCATCTTC-3'	Hammond *et al*., 1994
FGA	**F:**5'-GCCCCATAGGTTTTGAACTCA-3' **R:**5'-TGATTTGTCTGTAATTGCCAGC-3'	Urquhart *et al*., 1995
vWA	**F:**5'CCCTAGTGGATGATAAGAATAATCAGTATG-3' **R:**5'-GGACAGATGATAAATACATAGGATGGATGG-3'	Urquhart *et al*., 1995
D2S1338	**F:**5'- CCAGTGGATTTGGAAACAGA-3' **R:**5'- ACCTAGCATGGTACCTGCAG-3'	NCBI
D3S1358	**F:**5'- ACTGCAGTCCAATCTGGGT-3' **R:**5'- ATGAAATCAACAGAGGCTTG -3'	Li *et al*., 1993
D5S818	**F:**5'-GGGTGATTTTCCTCTTTGGT-3' **R:**5'-TGATTCCAATCATAGCCACA-3'	Jin *et al*., 1997
D6S1043	**F:**5'-CAAGGATGGGTGGATCAATA-3' **R:**5'-TTGTATGAGCCACTTCCCAT-3'	Shin *et al*., 2004
D7S820	**F:**5'-TGTCATAGTTTAGAACGAACTAACG-3' **R:**5'-CTGAGGTATCAAAAACTCAGAGG-3'	Jin *et al*., 1997
D8S1179	**F:**5' - TTTTTGTATTTCATGTGTACATTCG-3' **R:**5' - CGTAGCTATAATTAGTTCATTTTCA-3'	Barber *et al*., 1996
D10S1248	**F:**5'-GGAATAAGTGCAGTGCTTGG-3' **R:**5'-ACCAATCTGGTCACAACCAT-3'	Genome Database (GDB)
D13S317	**F:**5'-ACAGAAGTCTGGGATGTGGA-3' **R:**5'-GCCCAAAAAGACAGACAGAA-3	Jin *et al*., 1997
D16S539	**F:**5'-GATCCCAAGCTCTTCCTCTT-3' **R:**5'-ACGTTTGTGTGTGCATCTGT-3'	CHLC website
D18S51	**F:**5'-GAGCCATGTTCATGCCACTG-3' **R:**5'-CAAACCCGACTACCAGCAAC-3'	Barber *et al*., 1996
D19S433	**F:**5'-CCTGGGCAACAGAATAAGAT-3' **R:**5'-TAGGTTTTTAAGGAACAGGTGG-3'	NCBI
D21S11	**F:**5'-ATATGTGAGTCAATTCCCCAAG-3' **R:**5'-TGTATTAGTCAATGTTCTCCAG-3'	Urquhart *et al*., 1994

*F: Forward primer; R: Reverse primer.

3.5.2 Primer design

For each selected STR locus, primers for PCR amplification were designed using the primer-design software (Primer3Plus) (Untergasser *et al*., 2007). These primers were evaluated using the Basic Alignment Search Tool (BLAST) (Altschul *et al*., 1990) to ensure target specificity to each STR locus. **Table 3.2** indicates the primer sequences for each STR locus. The designed primers were subsequently used for Sanger sequencing.

Table 3.2: Primer sequences of the designed primers

STR Loci	Primer Sequences 5' -3'	T_m (°C)	GC content (%)	PCR fragment size
FGA	F: CATCTTAACTGGCATTCATGG R: TCACGGTCTGAAATCGAAAA	58.1 59.2	42.9 40	294 bp
vWA	F: CTCCTCAGACTGATCCTATAAGGTA R: AGAGATAGGATAGATGATAGATACAAAGGA	58.2 58.6	44 33	328 bp
D2S1338	F: AGCTCTCACCAGGGGGTGT R: TGGCTTCTTCCCTGTCTCAC	62.5 60.4	63 55	398 bp
D3S1358	F: CAGCGAGACCCCATCTCTTA R: ATTCACTTGCCCACTTCTGC	60.4 60.3	55 50	351 bp
D5S818	F: CTGGCTTACCCCCTCATTTT R: CCATCTGGATAGTGGACCTCA	60.3 59.9	50 52	470 bp
D6S1043	F: TTCGGTATTCTCCACATGGTT R: CAGAACTTTGGCACAAGCAG	59.3 59.6	42.9 50	472 bp
D7S820	F: GCTTCTTGGGTCAAGTGGTT R: CAGAAGGAATAAAAACAGGCAAA	59.2 59.7	50 33.8	498 bp
D8S1179	F: TGGCCAGAAACCTCTGTAGC R: GTGCATTGTTGTTGGGAATG	60.4 59.8	55 45	379 bp
D13S317	F: CAAATGGTAATTCTGCCTACAGC R: TTTGGGTAGGAAAAAGAGTGGA	60 60	43.5 40.9	412 bp
D16S539	F: TGTGCACAAATCTAAATGCAG R: CAAGCGAAAGTGATGCCATA	59.8 57.9	40.9 45.5	274 bp
D18S51	F: AAAATTAGTTGGGCATGGTG R: CCGACTACCAGCAACAACAC	57.5 59.2	40 55	398 bp
D19S433	F: TCACTTGAGGGAAGGAGTTCA R: GCTCCTGGGGTTCTAGGAAT	59.8 59.5	47.6 55	487 bp
D21S11	F: TCAGACTTGGACAGCCACAC R: CACTGAGAAGGGAGAAACACTG	59.9 59	55 50	399 bp

*F: forward primer; R: Reverse primer; T_m: Melting temperature; GC content%: Guanine and Cytosine content and bp: base pairs.

3.6 Optimisation of PCR assays to amplify the selected STR loci

PCR assays were optimised using commercially available control DNA (HLA-EQA Control DNA # 169, UCLA) for the published and designed primers. For each assay, the optimal primer annealing temperature of the primer pairs was determined using a temperature gradient. Each primer pair was tested at six different temperatures. The temperatures ranged from 55°C to 65°C.

For each STR locus, PCR assays were prepared separately. A reaction Master Mix was prepared by mixing 14.25 µL of nuclease-free water, 5 µL of 5X Green GoTaq® Flexi buffer, 2 µL $MgCl_2$ (25 mM), 0.5 µL dNTPs (10 mM) (Promega, Wisconsin, USA), 1 µL of the forward primer, 1 µL of the reverse primers (100 nM), 0.25 µL of GoTaq® G2 Flexi DNA Polymerase (5 u/µL) and 1 µL of the control DNA (650 ng/ µL). These reaction mixtures were vortexed and centrifuged briefly. The PCR cycling conditions for each reaction were: one cycle at 95°C for 5 minutes, 25 cycles at 95°C for 45 seconds for denaturation, annealing for 30 seconds at 55°C to 65°C as per the obtained optimal annealing temperature, and elongation at 72°C for 30 seconds. The final elongation step was performed at 72°C for 10 minutes (2720 Thermal Cycler, Applied Biosystems, USA).

3.7 Sample screening

3.7.1 Dried bloodstain punch purification

For each of the cases included in this study, archived samples in the form of dried bloodstains (Copan NUCLEIC-CARD™) were used. A nucleic card is a chemically treated card that lyse cells, denature proteins, and protect nucleic acids from nucleases, oxidative, and UV damage. It is an easy-to-use system for collection, preservation, and long-term storage of nucleic acids at room temperature. The dried bloodstains were purified and used for PCR amplification. Firstly, 200 µL of nuclease-free water was added to PCR eppendorf tubes containing the nucleic-card punches

(1.2 mm discs). Only one punch was used per reaction. The nucleic-card punches were incubated at room temperature for 15 min, and thereafter the water was discarded. This washing step was repeated to ensure properly purified nucleic-card punches. Subsequently, the nucleic-card punches were dried at 65°C for 1 hour in the 2720 Thermal Cycler (Applied Biosystems, USA). Thereafter, PCR amplification was performed as described in **section 3.9**, where 24 µL the reaction master mix was added directly to the purified nucleic-card punches.

3.7.2 Agarose-formamide gel electrophoresis

Agarose gel electrophoresis was used to evaluate the PCR products obtained. The PCR products were resolved on a 3% SeaKem LE agarose gel (Lonza, USA) containing 10% formamide (Sigma-Aldrich, USA) stained with 10 mg/mol ethidium bromide (12 µL) in 1X Sodium Borate buffer (SB buffer) (pH 8.5). Gel electrophoresis was performed at 180 V for approximately 6 hours. The 3% agarose- 10% formamide gel was required for the separation of heterozygous samples. This would allow for single alleles to be prepared for the Sanger sequencing part of the study. PCR fragment sizes were confirmed using a 100 bp DNA molecular weight marker (New England Biolabs, USA) resolved on the 3% agarose gel. The gels were visualised under UV light on the Kodak Gel Logic documentation system (Carestream Molecular Imaging, Rochester, NY, USA). Excision and purification of fragments of interest was performed. This was done by manually excising the fragment from the gel, which was subsequently submerged and incubated in 40 µL of nuclease-free water overnight. This incubation step allows for the PCR amplicon to diffused out of the gel and into the solution. Thereafter, 1 µL of the PCR amplicon-containing solution was used as input DNA for the second round of PCR amplification as described in **section 3.9**. Subsequently, the PCR products were resolved on a 2.5% agarose gel to ensure that single fragments are obtained prior to Sanger sequencing.

3.7.3 Sanger Sequencing

Pre-sequencing purification of PCR products was performed using the Exonuclease I and Shrimp Alkaline (rSAP) clean-up method (New England Biolabs, USA). A mixture was prepared by mixing 5 µL of PCR product, 0.5 µL of the Exonuclease I enzyme, and 1 µL rSAP enzyme. Excess primers and dNTPs were degraded by incubating the mixture at 37°C for 15 minutes. Following incubation, the enzymes were inactivated at 80°C for 15 min (GeneAmp 2720, Applied Biosystems, USA).

The sequencing reactions were performed using the purified PCR products. The BigDye™ Terminator v3.1 Kit (Applied Biosystems, USA) was used to prepare the sequencing reactions. The forward and reverse sequencing reactions each contained 2 µL of the BigDye™ Terminator v3.1 Ready Reaction Mix, 1 µL of BigDye™ Terminator v3.1 Sequencing Buffer, 1 µL of the 100 nM forward primer or reverse primer, respectively, 2.5 µL of the purified PCR product and 3.5 µL of nuclease-free water for a total reaction volume of 10 µL. The total reaction volume used was half the volume recommended by the manufacturer (20 µL). However, the use of halved sequencing reactions was previously validated by the DNA testing laboratory at our institution. The sequencing cycle conditions were one cycle at 96°C for 1 min, 25 cycles at 96°C for 10 sec, 50°C for 5 sec and 60°C for 4 min (GeneAmp 2720 Applied Biosystems, USA).

The ZR DNA sequencing Clean-Up Kit™ (Zymo Research, USA) was used to purify the sequencing products. A mixture containing 240 µL of Sequencing Binding Buffer and the sequencing product was prepared, and vortexed briefly. Thereafter, the mixture was transferred to a column in a collection tube and centrifuged in a Sigma 1-14 mini centrifuge (Germany) at 13,000 rpm for 30 seconds. The flow-through was discarded, and 300 µL of Sequencing Wash Buffer was added to the column in a collection tube and centrifuged at 13,000 rpm for 30 seconds. An additional centrifugation step at 13,000 rpm for 1 min was included to ensure the removal of all excess Sequencing Wash Buffer from the filter cartridge. The sequencing product was eluted by adding 10 µL of the Elution-buffer to the column in a clean collection tube.

Lastly, the column was centrifuged at 13,000 rpm for 15 sec to collect the purified sequence-containing elute. Sequencing was performed on the 3500 Genetic Analyser (Applied Biosystems, USA). The Sequencing Analysis Program v.7 was then used to analyse the raw sequence reads. Thereafter, sequence data analysis was performed using Chromas v2.6.6. The EMBOSS Needle online sequence alignment tool (Huang & Miller, 1991) (Available at: https://www.ebi.ac.uk/Tools/psa/emboss_needle/) was used to align and compared the obtained sequences the reference sequences. The reference sequences were obtained from the National Center for Biotechnology Information (NCBI) website and the STRBase website.

3.8 Data analysis

In the present study, descriptive data analysis was conducted by the researcher and the study supervisors. DNA sequence analysis was conducted for each parentage case studied to identify and characterise the underlying cause of the observed apparent STR mismatches. Furthermore, the sequence-based alleles were evaluated for concordance with alleles determined by Capillary Electrophoresis (CE) -based STR typing previously reported by the facility. The characterisation process involved the classification of the apparent mismatches by proposing a most likely presumed mutation model and mutation origin, as well as describing other identified variants.

CHAPTER 4: RESULTS AND DISCUSSION

4. Introduction

In this chapter, the findings of this study are reported and discussed. The aim was to characterise apparent STR loci mismatches detected in 100 parentage cases in the DNA testing facility, Universitas Academic Unit, National Health Laboratory Services (NHLS) Bloemfontein. A total of 200 samples (including only the parent and child with the allelic mismatch) were analysed across the following 15 autosomal STR loci: CSF1PO, FGA, vWA, D2S1338, D3S1358, D5S818, D6S1043, D7S820, D8S1179, D10S1248, D13S317, D16S539, D18S51, D19S433, and D21S11. The study's limitations, recommendations and impact are discussed.

4.1 Capillary-based STR analysis of the control DNA

In this study, PCR assays were optimised using commercially available control DNA (HLA-EQA Control DNA # 169, UCLA). Prior to PCR optimisation, the genetic profile of the control DNA was determined using the VeriFiler™ Express PCR amplification kit. The resulting genetic profile of the 25 STR loci included in the kit are shown in **Figure 4.1.**

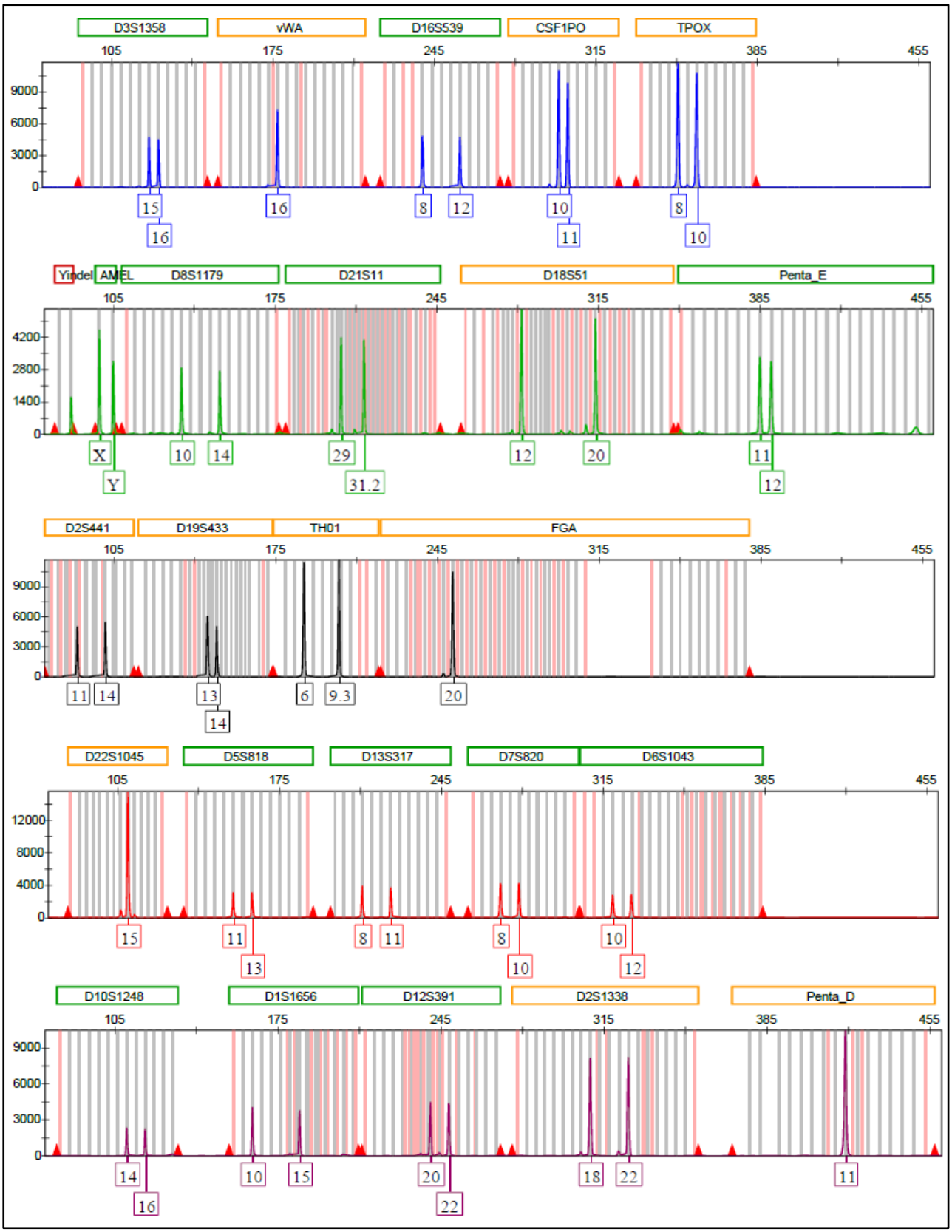


Figure 4.1: The genetic profile of the control DNA obtained using the VeriFiler™ Express PCR amplification kit

4.2 Optimisation of PCR assays to amplify the selected STR loci using the published primers

Primers published on the STRBase website **(Table 3.1)** were evaluated to determine the optimal primer annealing temperature. PCR fragments with the expected sizes were successfully obtained at the optimal primer annealing temperatures, defined as the highest tested temperature where no nonspecific amplification was observed. A gel electrophoresis image depicting an example of results obtained at the D13S317 STR locus for the temperature gradient conducted is depicted in **Figure 4.2.** The optimal primer annealing temperatures obtained for each tested locus are summarised in **Table 4.1**. For the D8S1179 and D19S433 loci, 57°C was found to be the optimal temperature, while 60°C was optimal for the FGA, vWA, D2S1338, D3S1358, D5S818, D6S1043, D10S1248, D13S317 and D16S539 loci. For the CSF1PO, D7S820, D18S51 and D21S11 loci, 65°C was the optimal temperature. As a practical consideration, 60°C was chosen as the suited annealing temperature for most of the STR loci to allow for PCR amplification to be performed simultaneously. Subsequently, the PCR fragments were resolved using agarose gel electrophoresis to enable the separation of heterozygous alleles for sequencing. Unfortunately, this process was unsuccessful; see **section 4.5** for the attempted troubleshooting. As a result, all heterozygous alleles involved in the study were sequenced unresolved. Moreover, the resulting double-sequence data (overlapping electropherogram peaks) was analysed and presented as is. Although this was a more challenging undertaking, the double sequences could be analysed successfully.

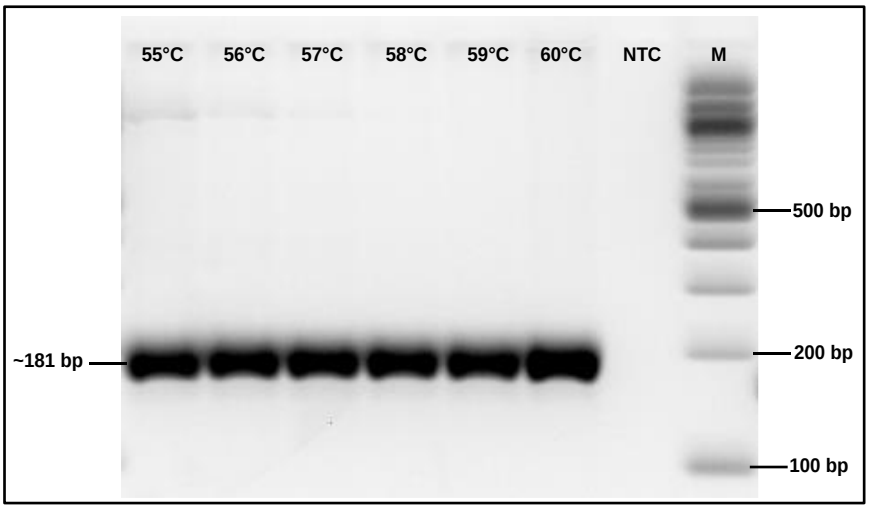


Figure 4.2: Agarose gel electrophoresis image depicting an example of results obtained at the D13S317 STR locus for the temperature gradient conducted at 55°C to 60°C. NTC: non-template control, M: 100 bp molecular weight marker, bp: base pairs.

Table 4.1: The optimal primer annealing temperature obtained for each tested STR locus using published primers.

STR locus	Control DNA alleles	Expected PCR fragment size (bp)	Obtained optimal annealing temperature (°C)
CSF1PO	10/11	311/315	65
FGA	20/20	196	60
vWA	16/16	138	60
D2S1338	18/22	192/208	60
D3S1358	15/16	127/131	60
D5S818	11/13	153/161	60
D6S1043	10/12	106/114	60
D7S820	8/10	202/210	65
D8S1179	10/14	173/189	57
D10S1248	14/16	251/259	60
D13S317	8/11	173/185	60
D16S539	8/12	157/173	60
D18S51	12/20	263/295	65
D19S433	13/14	128/132	57
D21S11	29/31.2	222/232	65

4.2.1 Sanger sequencing of the 15 selected STR loci using published primers

Following the optimisation of PCR assays using published primers, Sanger sequencing of the 15 STR loci mentioned above was conducted. However, only two of the 15 loci (CSF1PO and D10S1248) were sequenced successfully. The electropherograms of the sequences for the CSF1PO alleles 10/11 and the D10S1248 alleles 14/16 of the control DNA are shown in **Appendix A**. Moreover, for the other 13 loci, the sequencing resulted in incomplete STR repeat motifs. An example of an incomplete repeat motif of 10 repeats produced at the D5S818 locus is shown in **Figure 4.3**, where the CE-based genotype was indicated as 11/13. Therefore, the sequence composition of these loci could not be determined. It was identified that the PCR fragments of these loci were too short in size to allow analysis of the repeat motifs. The PCR fragments ranged from 106 bp to 295 bp, and although some of the STR loci were reasonably long, incomplete repeat motif sizing persisted. To overcome this issue, primer sets for the 13 loci mentioned above were redesigned and assessed for target specificity using BLAST as described in **section 3.8.2**.

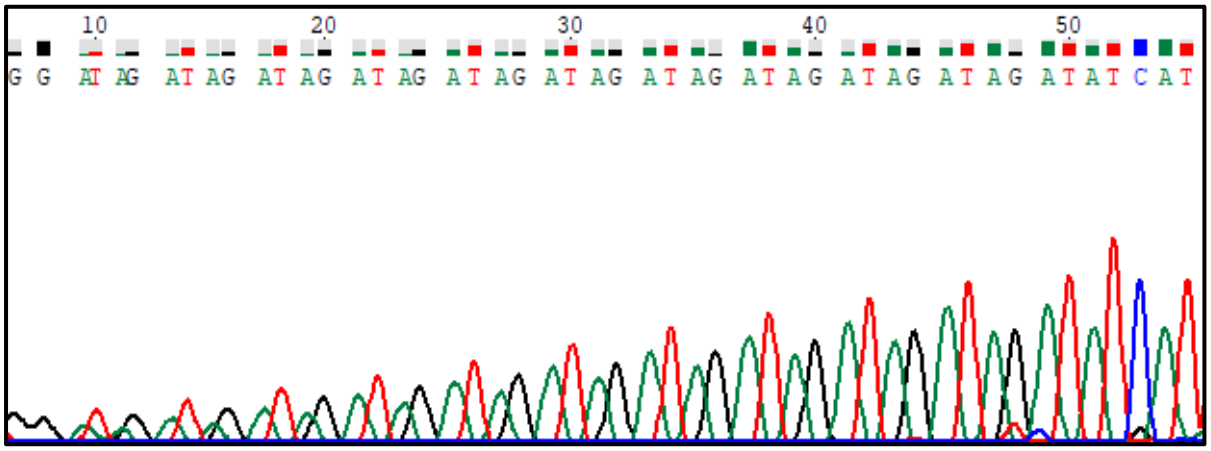


Figure 4.3: A sequence electropherogram depicting an example of an incomplete [AGAT] sequence motif at the D5S818 containing 10 repeats instead of 11 or 13 as indicated by the CE-based genotype.

4.3 Optimisation of PCR assays to amplify the selected STR loci using the designed primers

To address the sequencing problem of incomplete STR repeat motifs encountered when using published primers **(see section 4.2.1)**, new primer sets were designed for the FGA, vWA, D2S1338, D3S1358, D5S818, D6S1043, D7S820, D8S1179, D13S317, D16S539, D18S51, D19S433, and D21S11 loci. These primers encompassed larger flanking regions to allow for complete STR repeat motifs to be obtained from Sanger sequencing. The expected fragments ranged from 274 to 498 bp. The primers were tested to determine the optimal primer annealing temperature using a temperature gradient ranging from 55°C to 65°C. **Table 4.2** shows the optimal primer annealing temperature for each locus investigated. The optimal temperature for the D13S317 was 55°C, whereas 60°C was optimal for the FGA, vWA, D2S1338, D3S1358, D5S818, D6S1043, D7S820, D8S1179, D16S539, and D21S11 loci; and 64°C was optimal for the D18S51 and D19S433 loci. Subsequently, the obtained fragments for each locus were sequenced and analysed. **Figures 4.4- 4.10** depicts the sequencing results obtained. A 100% concordance between the genotypic data of the control DNA obtained by CE and the sequencing results was obtained.

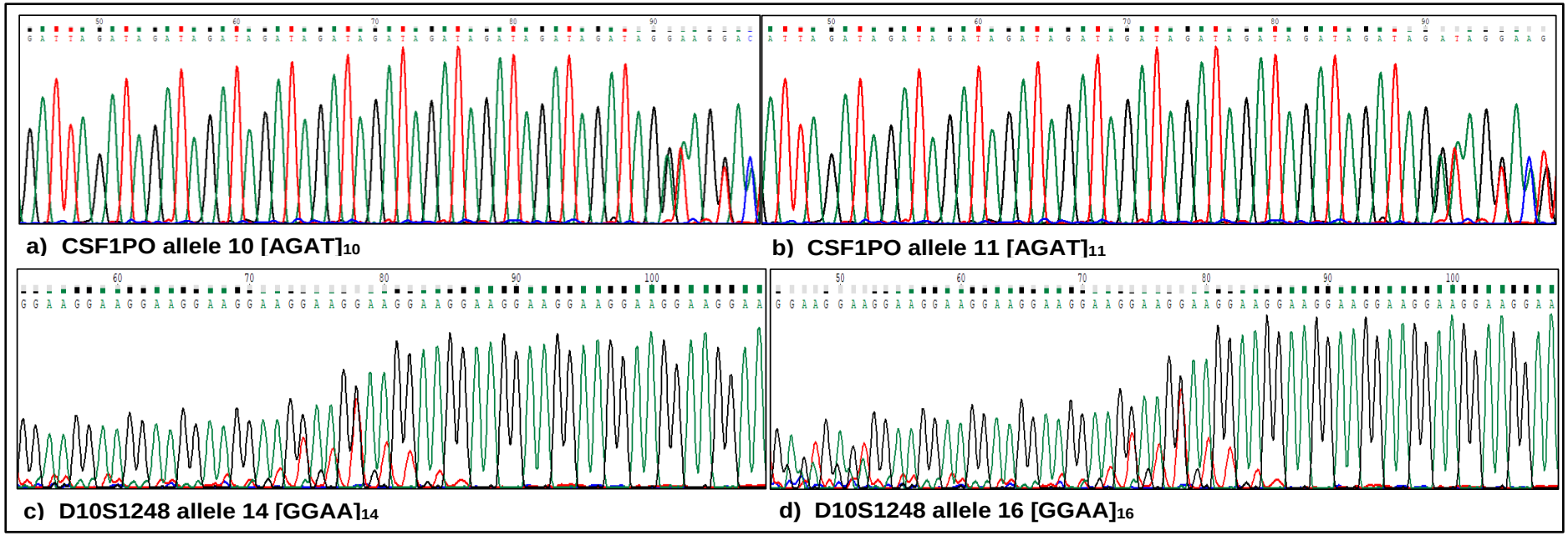


Figure 4.4: A sequence electropherogram depicting the sequence for the CSF1PO alleles a) CSF1PO allele 10, b) CSF1PO allele 11 and the D10S1248 alleles, c) D10S1248 allele 14, d) D10S1248 allele16 of the control DNA.

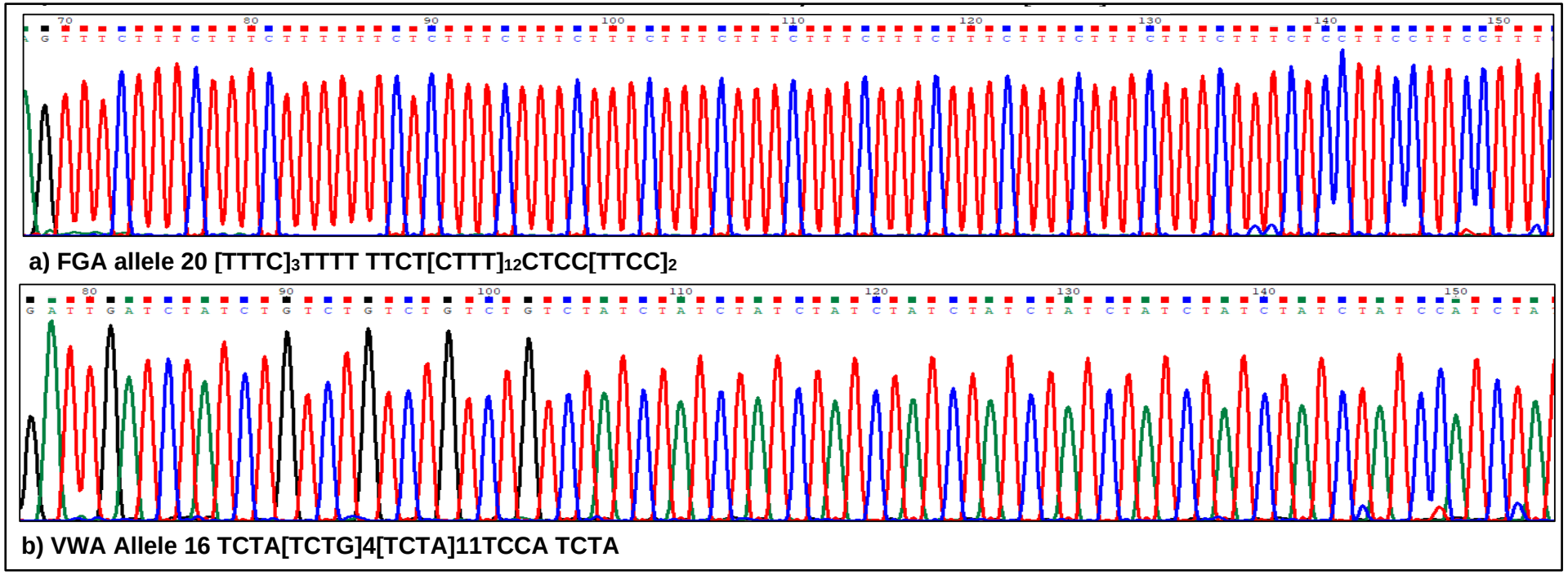


Figure 4.5: A sequence electropherogram depicting the sequence for: a) FGA allele 20, b) vWA allele 16

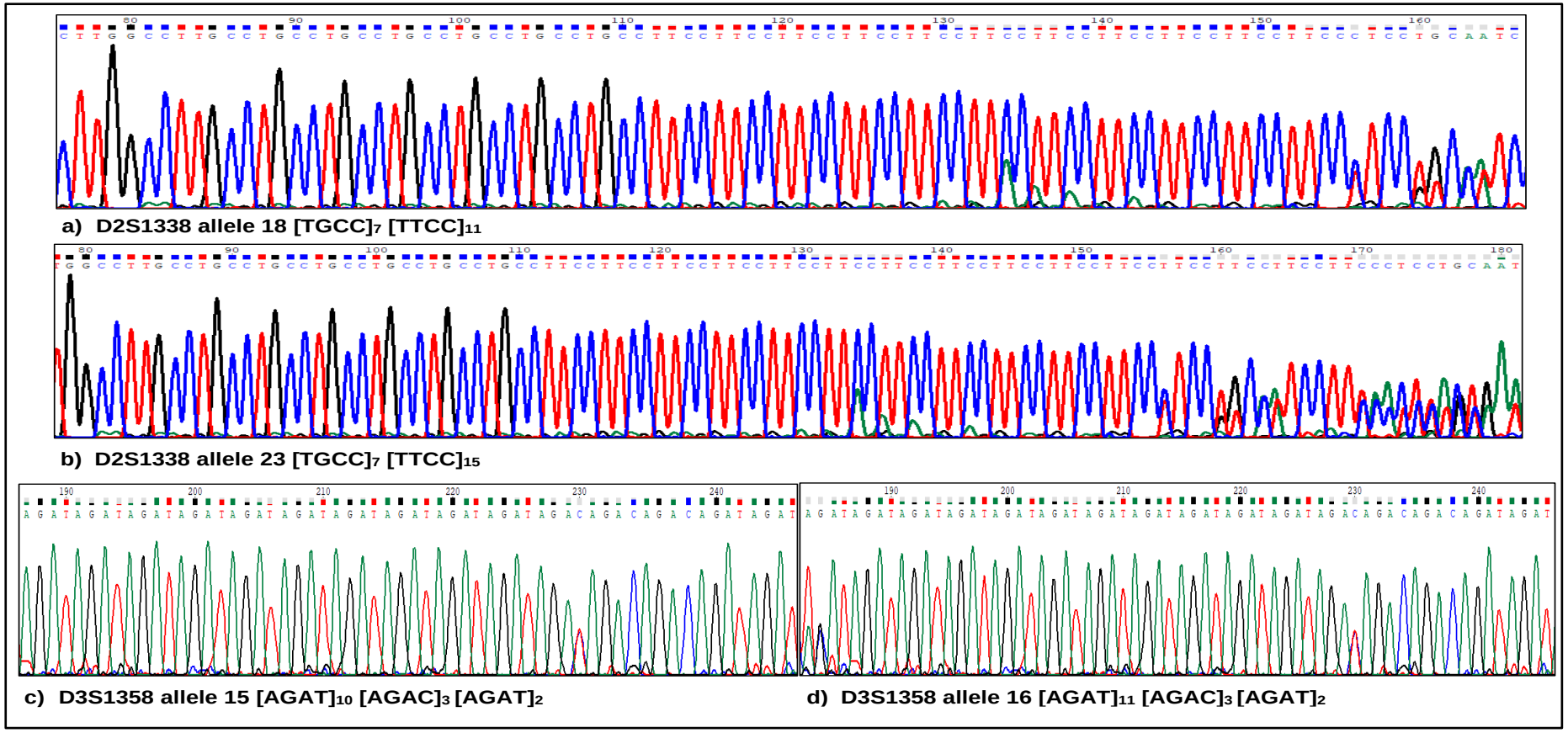


Figure 4.6: A sequence electropherogram depicting the sequence for: a) D2S1338 allele 18, b) D2S1338 allele 23, c) D3S1358 allele 15, d) D3S1358 allele 16.

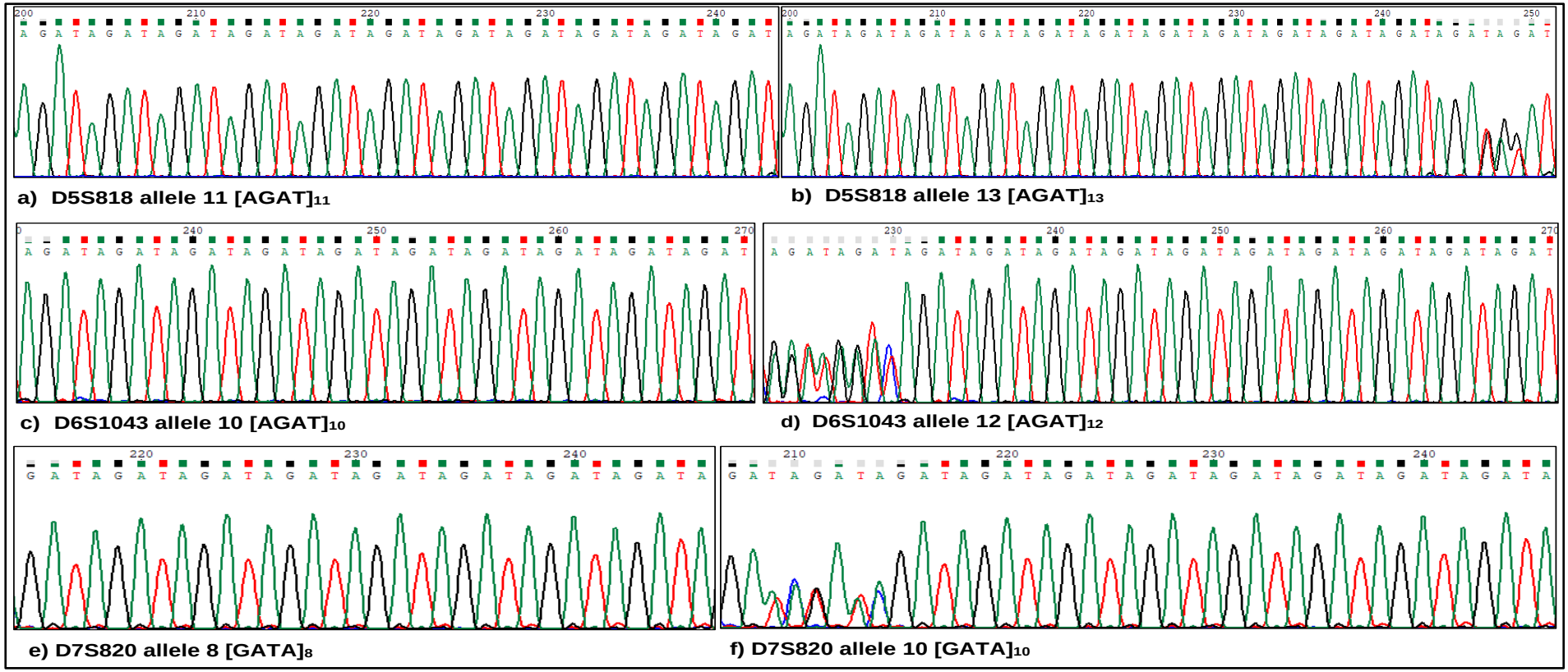


Figure 4.7: A sequence electropherogram depicting the sequence for: a) D5S818 allele 11, b) D5S818 allele 13, c) D6S1043 allele 10, d) D6S1043 allele 12, e) D7S820 allele 8, and f) D7S820 allele 10.

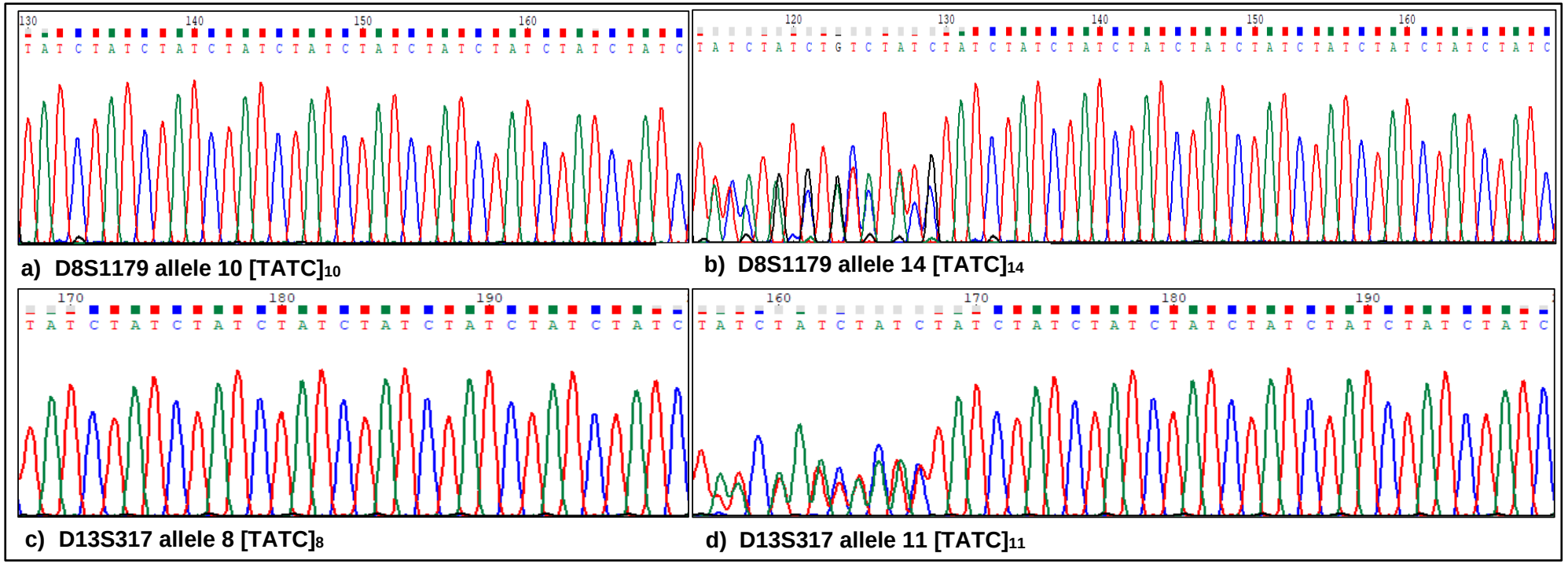


Figure 4.8: A sequence electropherogram depicting the sequence for: a) D8S1179 allele 10, b) D8S1179 allele 14, c) D13S317 allele 8, d) D13S317 allele 11.

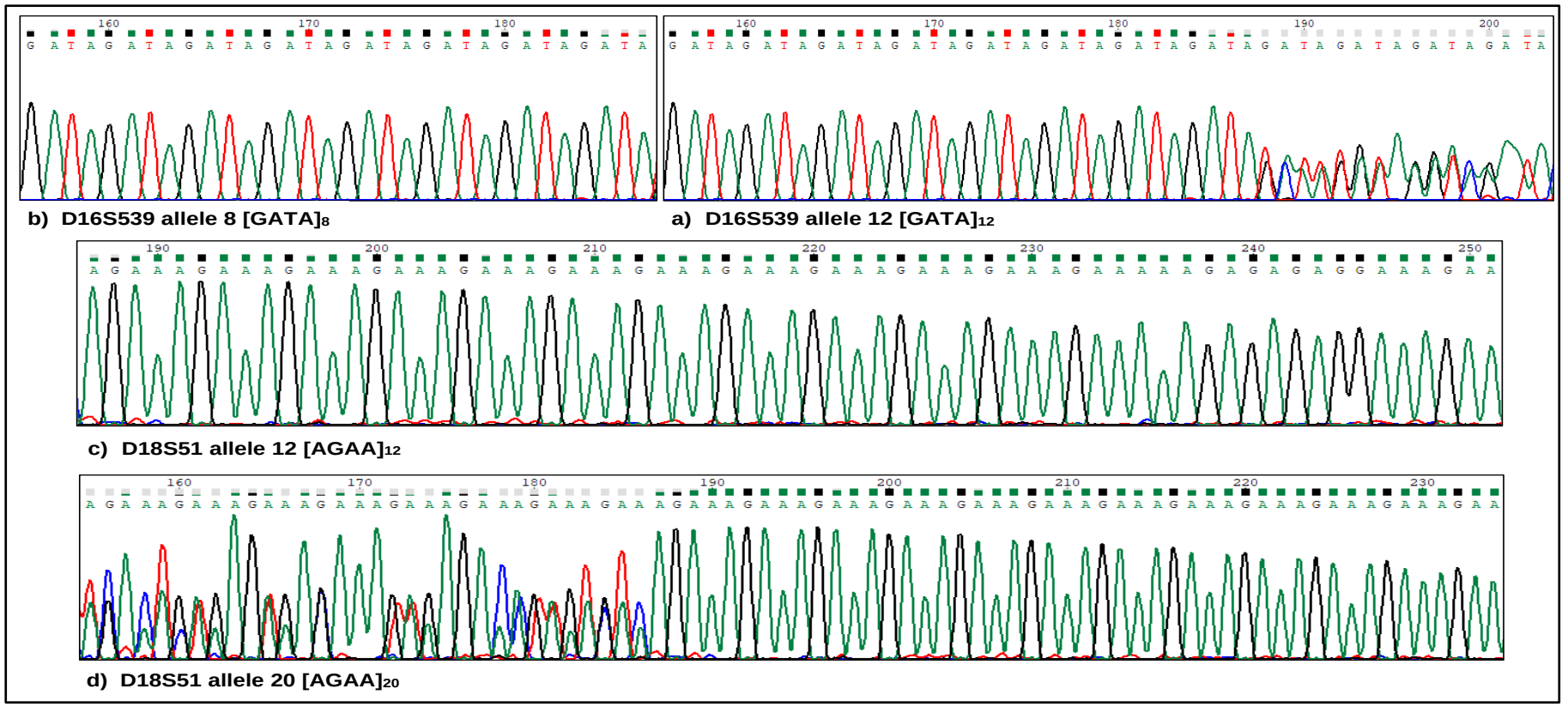


Figure 4.9: Sequence electropherogram depicting the sequence for: a) D16S539 allele 6, b) D16S539 allele 12, c) D18S51 allele 12, and d) D18S51 allele 20.

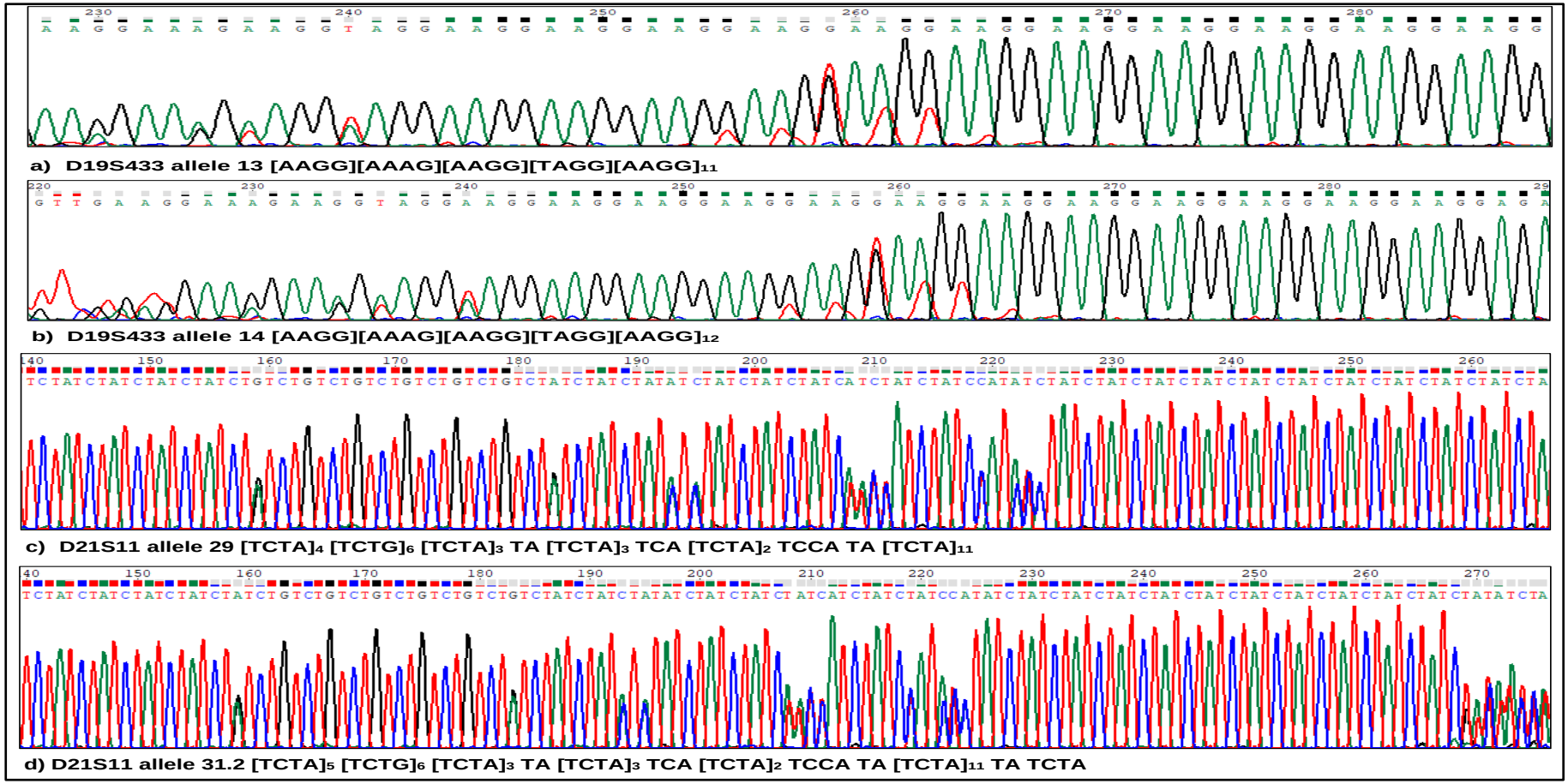


Figure 4.10: A sequence electropherogram depicting the sequence for: a) D19S433 allele 13, b) D19S433 allele 14, c) D21S11 allele 29, d) D21S11 allele 31.2

Table 4.2: Results of the obtained optimal primer annealing temperature and PCR fragment lengths for each tested locus using the redesigned primers.

STR locus	Control DNA alleles	Expected PCR fragment size (bp)	Obtained optimal annealing temperature (°C)
FGA	20/20	290	60
vWA	16/16	320	60
D2S1338	18/22	278/394	60
D3S1358	15/16	347/351	60
D5S818	11/13	470/478	60
D6S1043	10/12	464/472	60
D7S820	8/10	470/478	60
D8S1179	10/14	367/383	60
D13S317	8/11	400/412	55
D16S539	8/12	262/278	60
D18S51	12/20	358/390	64
D19S433	13/14	475/479	64
D21S11	29/31.2	399/409	60

4.5 Troubleshooting measures implemented for resolution of heterozygous alleles using gel electrophoresis

The 3% agarose- 10% formamide gel was used to attempt separation of heterozygous samples as described in **section 3.10.1**. This would allow for the preparation of single alleles for the Sanger sequencing part of the study. However, after several attempts this method did not yield the expected results. The concentrations were then adjusted to 7% agarose- 50% formamide following recommendations by Sun *et al*. (1982). The PCR fragments were resolved at 180 V for 6 hours. Sub-optimal resolution of the fragments was obtained. However, with these results gel excision posed a challenge as double sequence profiles were obtained indicating the presence of the two alleles. Following this, the voltage and time was adjusted to 200 V and 9 hours. This allowed for adequate resolution of the heterozygous alleles to be obtained **(Appendix B)**. This separation method was not efficient and feasible in a routine setting as it was laborious and time consuming.

Considering the number of samples included in this study (200 samples) this method was not feasible to meet the aim and objectives within the prescribed timeframe for an MMedSc degree. Furthermore, alternative methods were not explored due to time constraints and limited resources. Irrespective of this challenge, the double sequences could be analysed successfully.

4.4 Characterisation of apparent mismatches across the 15 selected STR loci

4.4.1 The CSF1PO locus

The CSF1PO locus contains a simple [AGAT] tetranucleotide repeat with reported alleles ranging from 6 to 15 repeats in size. The average mutation rate is estimated at 0.16% (Gettings *et al*., 2015; Butler, 2017). There were four (n=4) parentage cases with parent-child allelic mismatch studied in this locus. The obtained results are summarised in **Table 4.3** below**.** A 100% concordance was obtained between the previously reported CE-based genotypic data and the obtained sequence-based alleles in all the parentage cases investigated.

CSF1PO case 1 involved a male child and a father, where the genotypes previously established by CE were 10/12 and 11/13, respectively. However, the allele responsible for this could not be determined as this case was a paternal duo, and the mother's genotype was unknown. Therefore, three possible scenarios were proposed. The first possibility was a loss of one [AGAT] repeat in allele 11 of the father leading to allele 10 in the child. Secondly, the lost one [AGAT] repeat, resulted in allele 12 in the child. The most likely presumed mutation model for these possibilities was a loss (-) one-step mutation. Lastly, there could have been a gain in one [AGAT] repeat unit, resulting in allele 12 in the child. In this instance, the proposed model would be a gain (+) one-step mutation. Therefore, the allelic mismatch, in this case, was characterised as a (-/+) one-step mutation attributed to DNA polymerase slippage. Furthermore, a one-step mutation model was proposed rather than a two-step mutation, as one-step mutations occur more frequently than two-step mutations (Xu *et al*., 2022).

In CSF1PO case 2, a mother, a female child, and a father were studied. The genotypes previously obtained by CE were 10/12, 10/12, and 9/13, respectively. Although the mother's genotype was available, the alleles involved in the mismatch could not be identified. Therefore, a loss or gain (-/+) one-step mutation was presumed to be the most likely mutation model from the paternal lineage. There were two explanations for this: the father's allele 13 could have lost one repeat unit, resulting in allele 12 in the

child, and the father's allele 9 could have gained one [AGAT] unit, resulting in allele 10 in the child depending on which allele was maternally inherited. The reported mismatch was caused by DNA polymerase slippage.

CSF1PO case 3 involved a female child and a father. The CE-based genotypes were 11/13 and 10/12, respectively. The most like mutation model was presumed to be a (+) one-step mutation attributed to DNA polymerase slippage. Similar to CSF1PO case 1, the mother was not previously tested. Therefore, allele 10 or 12 of the father could have gained one [AGAT] repeat unit, leading to allele 11 or 13 in the child. In CSF1PO case 4, a mother, a male child, and a father were tested. The previously obtained CE-based genotypes were 10/12, 10/13, and 10/11, respectively. The allelic mismatch was of maternal origin, where allele 12 of the mother gained one repeat and was passed to the child as allele 13. The most likely presumed mutation model was a (+) one-step mutation.

At this locus, we characterised 75% (n=3/4) of the mismatches were of paternal origin, while only 25% (n=1/4) were of maternal origin. In literature, it is widely accepted that STR loci mutations in males occur at a rate 5 to 6 times higher than in females (Zhao *et al*., 2015). Although this observation is supported by literature, the assumption that STR loci mutations are more prevalent in males could not be deduced in this study due to the small sample size in this locus. Furthermore, all the identified mismatches were attributed to DNA polymerase slippage, as the sequencing data indicated no other sequence variation that could suggest alternative mutation models.

Table 4.3: Results obtained for the CSF1PO locus parentage cases. The table indicates the genotype previously obtained by CE, genotypes obtained by sequencing, sequence structure, most likely presumed mutation model and presumed mutation origin

Study case no.	Genotype previously obtained by CE	Genotype obtained by sequencing	Sequence structure	Most likely presumed mutation model	Presumed mutation origin
CSF1PO Case 1	Male child: 10/12	Male child: 10/12	$[AGAT]_{10}$	(-/+) One-step mutation	Paternal
			$[AGAT]_{12}$		
	Father: 11/13	Father: 11/13	$[AGAT]_{11}$		
			$[AGAT]_{13}$		
CSF1PO Case 2	Mother: 10/12	N/A	N/A	(+/-) One-step mutation	Paternal
	Female child: 10/12	Female child: 10/12	$[AGAT]_{10}$		
			$[AGAT]_{12}$		
	Father: 9/13	Father: 9/13	$[AGAT]_{9}$		
			$[AGAT]_{13}$		
CSF1PO Case 3	Female child: 11/13	Female child: 11/13	$[AGAT]_{11}$	(+) One-step mutation	Paternal
			$[AGAT]_{13}$		
	Father: 10/12	Father: 10/12	$[AGAT]_{10}$		
			$[AGAT]_{12}$		
CSF1PO Case 4	Mother: 10/12	Mother: 10/12	$[AGAT]_{10}$	(+) One-step mutation	Maternal
			$[AGAT]_{12}$		
	Male child: 10/13	Male child: 10/13	$[AGAT]_{10}$		
			$[AGAT]_{13}$		
	Father: 10/11	N/A	N/A		

4.4.2 The FGA locus

The FGA locus is characterised by a compound tetranucleotide repeat of $[TTTC]_3$ [TTTT] [TTCT] $[CTTT]_n$ [CTCC] $[TTCC]_2$, with allele sizes ranging from 13 to 51.2 repeats. The mutation rate at this locus is reported at 0.28 (Butler, 2017; Hamester *et al*., 2019). A total of 10 parentage cases with an apparent mismatch at the FGA locus were examined. The obtained results are presented in **Table 4.4**. In all cases for the FGA locus, the genotypic data were concordant (100%) with the sequencing data obtained.

All parent-child allelic mismatches at this locus were found to be due to DNA polymerase slippage. In approximately 50% of the parentage cases (cases 1, 2, 3, 4, 5, and 9) at this locus, the allelic mismatches were characterised as (-/+) one-step mutations, of these, cases 2, 3, 4, 5, and 9 were transmitted from the paternal lineage. In contrast, the mutation origin in case 1 was presumed to be from the paternal or maternal lineage. Furthermore, the (+) one-step mutation was identified as the second most common mutation model in this locus, accounting for 30% (n=3/10) of the parent-child mismatches observed (cases 5, 7, and 10). The (+) one-step mutations characterised in these cases were passed from the paternal lineage. In FGA cases 6 and 8, the allelic mismatches were presumed to be most likely (-) one-step mutation.

In the literature, one-step mutation models have been reported to occur more frequently than other models (Hamester *et al*., 2019). Notably, the [CTTT] repeat motif was the only unit affected by the slippage in all the cases studied in this locus. This finding can be attributed to the length (the number of times it is repeated) within the FGA repeat motif itself: $[TTTC]_3$ [TTTT] [TTCT] $[CTTT]_n$ [CTCC] $[TTCC]_2$ as compared to the other repeats. Other studies have reported similar trends (Lu *et al*., 2012; Zhang *et al*., 2018). Furthermore, in 90% (n=9/10) of the cases, the observed mutations were determined to be from the paternal lineage, while there was only one case where the origin of the mutation was not determined. Similarly, Paredes (2011) found 97 mutations in this locus, of which 74% were shown to come from the paternal lineage and 11% were attributed to maternal origin. However, in 15% of the cases, it was not

possible to establish the origin. The origin of mutations characterised in this locus were found to be predominantly from the paternal lineage, which was in keeping with published literature.

Table 4.4: Results obtained for the FGA locus parentage cases. The table indicates the genotype previously obtained by CE, genotypes obtained by sequencing, sequence structure, most likely presumed mutation model and presumed mutation origin.

Study case no.	Genotype previously obtained by CE	Genotype obtained by sequencing	Sequence structure	Most likely presumed mutation model	Presumed mutation origin
FGA Case 1	Mother: 21/23	N/A	N/A	(-/+) One-step mutation	Paternal/ Maternal
	Male child: 21/22	Male child: 21/22	$[TTTC]_3\,[TTTT]\,[TTCT]\,[CTTT]_{13}\,[CTCC]\,[TTCC]_2$		
			$[TTTC]_3\,[TTTT]\,[TTCT]\,[CTTT]_{14}\,[CTCC]\,[TTCC]_2$		
	Father: 21/27	Father: 21/27	$[TTTC]_3\,[TTTT]\,[TTCT]\,[CTTT]_{13}\,[CTCC]\,[TTCC]_2$		
			$[TTTC]_3\,[TTTT]\,[TTCT]\,[CTTT]_{19}\,[CTCC]\,[TTCC]_2$		
FGA Case 2	Mother: 19/22	N/A	N/A	(-/+) One-step mutation	Paternal
	Male child: 19/22	Male child: 19/22	$[TTTC]_3\,[TTTT]\,[TTCT]\,[CTTT]_{11}\,[CTCC]\,[TTCC]_2$		
			$[TTTC]_3\,[TTTT]\,[TTCT]\,[CTTT]_{14}\,[CTCC]\,[TTCC]_2$		
	Father: 21/23	Father: 21/23	$[TTTC]_3\,[TTTT]\,[TTCT]\,[CTTT]_{13}\,[CTCC]\,[TTCC]_2$		
			$[TTTC]_3\,[TTTT]\,[TTCT]\,[CTTT]_{15}\,[CTCC]\,[TTCC]_2$		
FGA Case 3	Male child: 25/27	Male child: 25/27	$[TTTC]_3\,[TTTT]\,[TTCT]\,[CTTT]_{17}\,[CTCC]\,[TTCC]_2$	(-/+) One-step mutation	Paternal
			$[TTTC]_3\,[TTTT]\,[TTCT]\,[CTTT]_{19}\,[CTCC]\,[TTCC]_2$		
	Father: 24/26	Father: 24/26	$[TTTC]_3\,[TTTT]\,[TTCT]\,[CTTT]_{16}\,[CTCC]\,[TTCC]_2$		
			$[TTTC]_3\,[TTTT]\,[TTCT]\,[CTTT]_{18}\,[CTCC]\,[TTCC]_2$		
FGA Case 4	Female child: 22/26	Female child: 22/26	$[TTTC]_3\,[TTTT]\,[TTCT]\,[CTTT]_{14}\,[CTCC]\,[TTCC]_2$	(-/+) One-step mutation	Paternal
			$[TTTC]_3\,[TTTT]\,[TTCT]\,[CTTT]_{18}\,[CTCC]\,[TTCC]_2$		
	Father: 23/25	Father: 23/25	$[TTTC]_3\,[TTTT]\,[TTCT]\,[CTTT]_{15}\,[CTCC]\,[TTCC]_2$		
			$[TTTC]_3\,[TTTT]\,[TTCT]\,[CTTT]_{17}\,[CTCC]\,[TTCC]_2$		

Table 4.4 continued

Study case no.	Genotype previously obtained by CE	Genotype obtained by sequencing	Sequence structure	Most likely presumed mutation model	Presumed mutation origin
FGA Case 5	Mother: 19.2/ 21	N/A	N/A	(+) One-step mutation	Paternal
	Female child: 19.2/22	Female child: 19.2/22	$[TTTC]_3$ [TTTT] [TT] $[CTTT]_{12}$ [CTCC] $[TTCC]_2$		
			$[TTTC]_3$ [TTTT] [TTCT] $[CTTT]_{14}$ [CTCC] $[TTCC]_2$		
	Father: 21/25	Father: 21/25	$[TTTC]_3$ [TTTT] [TTCT] $[CTTT]_{13}$ [CTCC] $[TTCC]_2$		
			$[TTTC]_3$ [TTTT] [TTCT] $[CTTT]_{17}$ [CTCC] $[TTCC]_2$		
FGA Case 6	Mother: 19/22	N/A	N/A	(-) One-step mutation	Paternal
	Female child: 19/23	Female child: 19/23	$[TTTC]_3$ [TTTT] [TTCT] $[CTTT]_{11}$ [CTCC] $[TTCC]_2$		
			$[TTTC]_3$ [TTTT] [TTCT] $[CTTT]_{15}$ [CTCC] $[TTCC]_2$		
	Father: 21/24	Father: 21/24	$[TTTC]_3$ [TTTT] [TTCT] $[CTTT]_{13}$ [CTCC] $[TTCC]_2$		
			$[TTTC]_3$ [TTTT] [TTCT] $[CTTT]_{16}$ [CTCC] $[TTCC]_2$		
FGA Case 7	Mother: 20/25	N/A	N/A	(+) One-step mutation	Paternal
	Female child: 20/27	Female child: 20/27	$[TTTC]_3$ [TTTT] [TTCT] $[CTTT]_{12}$ [CTCC] $[TTCC]_2$		
			$[TTTC]_3$ [TTTT] [TTCT] $[CTTT]_{19}$ [CTCC] $[TTCC]_2$		
	Father: 26/45.2	Father: 26/45.2	$[TTTC]_3$ [TTTT] [TTCT] $[CTTT]_{18}$ [CTCC] $[TTCC]_2$		
			$[TTTC]_3$ [TTTT] [TT] $[CTTT]_{37}$ [CTCC] $[TTCC]_2$		

Table 4.4 continued

Study case no.	Genotype previously obtained by CE	Genotype obtained by sequencing	Sequence structure	Most likely presumed mutation model	Presumed mutation origin
FGA Case 8	Male child: 24/24	Male child: 24/24	$[TTTC]_3$ [TTTT] [TTCT] $[CTTT]_{16}$ [CTCC] $[TTCC]_2$	(-) One-step mutation	Paternal
	Father: 22/25	Father: 22/25	$[TTTC]_3$ [TTTT] [TTCT] $[CTTT]_{14}$ [CTCC] $[TTCC]_2$		
			$[TTTC]_3$ [TTTT] [TTCT] $[CTTT]_{17}$ [CTCC] $[TTCC]_2$		
FGA Case 9	Male child: 22/26	Male child: 22/26	$[TTTC]_3$ [TTTT] [TTCT] $[CTTT]_{14}$ [CTCC] $[TTCC]_2$	(-/+) One-step mutation	Paternal
			$[TTTC]_3$ [TTTT] [TTCT] $[CTTT]_{18}$ [CTCC] $[TTCC]_2$		
	Father: 23/25	Father: 23/25	$[TTTC]_3$ [TTTT] [TTCT] $[CTTT]_{15}$ [CTCC] $[TTCC]_2$		
			$[TTTC]_3$ [TTTT] [TTCT] $[CTTT]_{17}$ [CTCC] $[TTCC]_2$		
FGA Case 10	Mother: 19/25	N/A	N/A	(+) One-step mutation	Paternal
	Female child: 25/27	Female child: 25/27	$[TTTC]_3$ [TTTT] [TTCT] $[CTTT]_{17}$ [CTCC] $[TTCC]_2$		
			$[TTTC]_3$ [TTTT] [TTCT] $[CTTT]_{19}$ [CTCC] $[TTCC]_2$		
	Father: 24/26	Father: 24/26	$[TTTC]_3$ [TTTT] [TTCT] $[CTTT]_{16}$ [CTCC] $[TTCC]_2$		
			$[TTTC]_3$ [TTTT] [TTCT] $[CTTT]_{18}$ [CTCC] $[TTCC]_2$		

4.4.3 The vWA locus

The vWA locus contains a compound tetranucleotide repeat motif [TCTA] $[TCTG]_n$ $[TCTA]_n$. The allele ranges from 10-25 repeats (Gettings *et al*., 2015). The mutation rate at this locus is reported to be 0.17% (Butler, 2015b). A 100% concordance between the CE-based and the sequence-based genotypic data was maintained for all parentage cases studied.

This study evaluated thirteen (n=13) parentage cases with apparent mismatches at this locus. The most prevalent mutation model in this locus was determined to be a (+) one-step mutation, identified in 38% (n=5/13) of the cases (vWA cases 2, 4, 6, 8, and 9). These mismatches were most likely caused by the insertion of one [TCTA] repeat unit. Furthermore, mismatches in cases 2, 4, and 6 were noted to originate from the paternal lineage, whereas the mismatches in cases 8 and 9 were passed maternally.

In 30% (n=4/13) of these cases (vWA cases 1, 3, 5, and 12), the parent-child mismatches were most likely to be (-) one-step mutations resulting from the loss of one [TCTA] repeat unit. Furthermore, mismatches in cases 1, 5, and 12 were identified to be from the paternal lineage, and the mismatch in case 3 was maternally transmitted. In vWA case 11, a (+) two-step mutation originating from the paternal lineage was identified. The allelic mismatch in case 7 was characterised as a (+) three-step mutation resulting from the gain of three [TCTA] repeats.

Surprisingly, in cases 10 and 13, a (-) five-step mutation model was observed. vWA case 10 was a standard trio, with CE-based genotypic data reported as a homozygous 21/21 for the mother, 16/16 for the female child and a heterozygous 16/18 for the father. These genotypes indicated that the child inherited one allele 16, from the father. Following this, a (-) five-step mutation was presumed as the most likely mutation model. The multi-step mutation resulted from a loss of five [TCTG] repeat units in the mother's allele 21, thus, accounting for the second allele 16 in the child. A similar finding was observed in case 13. The mismatch was between the father's allele 20 and the child's allele 15. However, in this case, the loss of one [TCTG] and four [TCTA] repeats were observed. In this locus, multi-step mutations were identified to occur less

frequently than one-step or two-step mutations, with a frequency of 15% (n=2/13). This is consistent with reports in the literature (Zhao *et al*., 2015). However, in a study by Sun *et al*. (2014), a single multi-step mutation out of 41 mutations (2.4%) was identified at the vWA locus. Although the findings by Sun *et al*. (2014) were differed significantly compared to the current study, it can be appreciated that multi-step mutation are less frequent than other mutation models. Moreover, in parentage cases where no sequence variants were detected, which might provide alternative mutation models, the CE-based genotypic data should be assigned with certainty.

Table 4.5: Results obtained for the vWA locus parentage cases. The table indicates the genotype previously obtained by CE, genotypes obtained by sequencing, sequence structure, most likely presumed mutation model and presumed mutation origin.

Study case no.	Genotype previously obtained by CE	Genotype obtained by sequencing	Sequence structure	Most likely presumed mutation model	Proposed mutation origin
vWA Case 1	Mother: 15/15	Mother: 15/15	N/A	(-) One-step mutation	Paternal
	Male child: 15/15	Male child: 15/15	[TCTA] $[TCTG]_4$ $[TCTA]_{10}$		
	Father: 16/16	Father: 16/16	[TCTA] $[TCTG]_4$ $[TCTA]_{11}$		
vWA Case 2	Male child:15/15	Male child:15/15	[TCTA] $[TCTG]_3$ $[TCTA]_{11}$	(+) One-step mutation	Paternal
	Father: 14/14	Father: 14/14	[TCTA] $[TCTG]_3$ $[TCTA]_{10}$		
vWA Case 3	Male child: 15/15	Male child: 15/15	[TCTA] $[TCTG]_4$ $[TCTA]_{12}$	(-) One-step mutation	Maternal
	Mother: 16/16	Mother: 16/16	[TCTA] $[TCTG]_3$ $[TCTA]_{12}$		
vWA Case 4	Female child:15/15	Female child:15/15	[TCTA] $[TCTG]_3$ $[TCTA]_{11}$	(+) One-step mutation	Paternal
	Father: 14/14	Father: 14/14	[TCTA] $[TCTG]_3$ $[TCTA]_{10}$		
vWA Case 5	Female child:14/14	Female child:14/14	[TCTA] $[TCTG]_3$ $[TCTA]_{10}$	(-) One-step mutation	Paternal
	Father: 15/15	Father: 15/15	[TCTA] $[TCTG]_4$ $[TCTA]_{10}$		
vWA Case 6	Male child: 16/18	Male child: 16/18	[TCTA] $[TCTG]_3$ $[TCTA]_{12}$	(+) One-step mutation	Paternal
			[TCTA] $[TCTG]_4$ $[TCTA]_{13}$		
	Father: 15/15	Father: 15/15	[TCTA] $[TCTG]_3$ $[TCTA]_{11}$		
vWA Case 7	Mother: 15/15	Mother: 15/15	[TCTA] $[TCTG]_4$ $[TCTA]_{10}$	(+) Three-step mutation	Maternal
	Female child:18/18	Female child:18/18	[TCTA] $[TCTG]_4$ $[TCTA]_{13}$		
	Father: 16/18	N/A	N/A		

Table 4.5 continued

Study case no.	Genotype previously obtained by CE	Genotype obtained by sequencing	Sequence structure	Most likely presumed mutation model	Presumed mutation origin
vWA Case 8	Mother: 16/16	Mother: 16/16	[TCTA] $[TCTG]_4$ $[TCTA]_{11}$	(+) One-step mutation	Maternal
	Female child:17/17	Female child:17/17	[TCTA] $[TCTG]_4$ $[TCTA]_{12}$		
	Father: 17/20	N/A	N/A		
vWA Case 9	Male child:15/15	Male child:15/15	[TCTA] $[TCTG]_4$ $[TCTA]_{10}$	(+) One-step mutation	Maternal
	Mother: 14/14	Mother: 14/14	[TCTA] $[TCTG]_4$ $[TCTA]_9$		
vWA Case 10	Mother: 21/21	Mother: 21/21	[TCTA] $[TCTG]_6$ $[TCTA]_{14}$	(-) Five-step mutation	Maternal
	Female child: 16/16	Female child: 16/16	[TCTA] [TCTG] $[TCTA]_{14}$		
	Father: 16/18	N/A	N/A		
vWA Case 11	Male child: 16/16	Male child: 16/16	[TCTA] $[TCTG]_4$ $[TCTA]_{11}$	(+) Two-step mutation	Paternal
	Father: 14/14	Father: 14/14	[TCTA] $[TCTG]_4$ $[TCTA]_9$		
vWA Case 12	Mother:15/17	N/A	N/A	(-) One-step mutation	Paternal
	Male child: 14/17	Male child: 14/17	[TCTA] $[TCTG]_3$ $[TCTA]_{10}$		
			[TCTA] $[TCTG]_4$ $[TCTA]_{12}$		
	Father: 15/18	Father: 15/18	[TCTA] $[TCTG]_4$ $[TCTA]_{10}$		
			[TCTA] $[TCTG]_4[TCTA]_{13}$		
vWA Case 13	Mother: 15/15	N/A	N/A	(-) Five-step mutation	Paternal
	Male child: 15/15	Male child: 15/15	[TCTA] $[TCTG]_4$ $[TCTA]_{10}$		
	Father: 20/20	Father:20/20	[TCTA] $[TCTG]_5$ $[TCTA]_{14}$		

4.4.4 The D2S1338 locus

The D2S1338 locus contains a compound tetranucleotide repeat. Two primary sequence motifs are used in allele designation. The first motif is composed of a $[TGCC]_{4-9}$ and a $[TTCC]_{6-19}$ motif, collectively accounting for allele sizes ranging from 10 to 28 repeats: $[TGCC]_n$ $[TTCC]_n$. The second motif is an analogue of the first, but with a single interruption by a [GTCC] repeat: $[TGCC]_n$ $[TTCC]_n$ [GTCC] $[TTCC]_2$, with alleles containing 19-29 repeats (Gettings *et al*., 2015). The mutation rate at this locus is 0.12% (Butler, 2015). Five (n=5) parentage cases were studied at this locus. There was a 100% concordance between the obtained sequencing data and the CE-based genotypic data. The results obtained are summarised in **Table 4.6.**

The first case studied involved a mother, a male child and a father, with CE-based genotypic data reported as homozygous 24/24, heterozygous 23/24, and heterozygous 21/22, respectively. The allele allelic mismatch was between the father's allele 22 and the child's allele 23. The allelic mismatch was characterised to be most likely a (+) one-step mutation model, resulting from the gain of one [TTCC] repeat unit because of DNA polymerase slippage. Furthermore, this mutation was presumed to be from the paternal lineage as the mother had passed allele 24 to the child. The second case at this locus was a maternal duo. The genotypic data was reported as heterozygous 18/19 for the female child and heterozygous 16/21 for the mother. The mutation model responsible for the mother-child allelic mismatch was presumed to be a (+/-) two-step mutation model attributed to DNA polymerase slippage. The presumed mutation model could be explained by either a gain of two [TTCC] repeats in allele 16 of the mother or a loss of the two repeats in the mother's allele 21, resulting in these alleles being passed to the child as alleles 18 and 19, respectively.

The last three cases (cases 3, 4 and 5) studied in this locus were standard paternity trios. The parent-child allelic mismatches identified in these cases were presumed to be (-) one-step mutation model, owing to DNA polymerase slippage. Moreover, the characterised mutation model was presumed to be from the paternal origin. Interestingly, a base substitution of guanine to thymine (G>T) was detected in individuals tested in cases 2 (mother), 3 (child), and 5 (child) (**Figure 4.11)**. The

identified SNV is located 35 bp from the 3'-end of the D2S1338 repeat motif in the flanking region. However, there was no impact associated with this SNV on the sequence-based alleles obtained. However, it has been characterised as rs6736691 (NC_000002.11:g.218879547C>A) in the 3'-5' complementary sequence on the database for single nucleotide polymorphisms (dbSNP) (available at https://www.ncbi.nlm.nih.gov/snp/rs6736691). The SNV was also reported in a review by Gettings *et al*. (2015). The authors reviewed STR allelic sequence variation in 24 loci most commonly used in forensic DNA investigations. Furthermore, the authors reported the rs6736691 SNV to occur in 20% of cases, making it the second most frequent SNV after the rs72951992 SNV occurring in 29% of cases. The SNV occurred at a frequency of 30% (n=3/10) in this study, which was higher than the study by Gettings *et al.* (2015). Consequently, it is important for DNA testing laboratories and researchers to take note of SNVs detected at the STR flanking region as these could potentially impact some of the PCR primers used in commercially available kits.

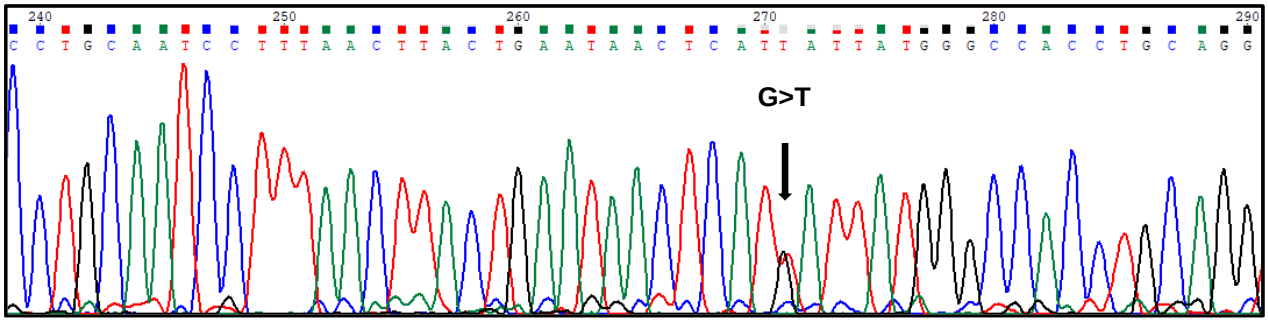


Figure 4.11: A sequence electropherogram depicting the G>T SNV detected at the D2S1338 STR locus.

Table 4.6: Results obtained for the D2S1338 locus parentage cases. The table indicates the genotype previously obtained by CE, genotypes obtained by sequencing, sequence structure, most likely presumed mutation model and presumed mutation origin.

Study case no.	Genotype previously obtained by CE	Genotype obtained by sequencing	Sequence structure	Most likely presumed mutation model	Presumed mutation origin	Other variants
D2S1338 Case 1	Mother: 24/24	N/A	N/A	(+) One-step mutation	Paternal	None
	Male child: 23/24	Male child: 23/24	$[TGCC]_7$ $[TTCC]_{13}$ [GTCC] $[TTCC]_2$			
			$[TGCC]_7$ $[TTCC]_{14}$ [GTCC] $[TTCC]_2$			
	Father: 21/22	Father: 21/22	$[TGCC]_7$ $[TTCC]_{11}$ [GTCC] $[TTCC]_2$			
			$[TGCC]_7$ $[TTCC]_{12}$ [GTCC] $[TTCC]_2$			
D2S1338 Case 2	Female child:18/19	Female child: 18/19	$[TGCC]_7$ $[TTCC]_{11}$	(+/-) Two-step mutation	Maternal	None
			$[TGCC]_7$ $[TTCC]_{12}$			
	Mother: 16/21	Mother: 16/21	$[TGCC]_5$ $[TTCC]_{11}$			G>T
			$[TGCC]_7$ $[TTCC]_{14}$			
D2S1338 Case 3	Mother: 16/19	N/A	N/A	(-) One-step mutation	Paternal	N/A
	Male child:16/20	Male child:16/20	$[TGCC]_4$ $[TTCC]_{12}$			G>T
			$[TGCC]_4$ $[TTCC]_{16}$			
	Father: 19/21	Father: 19/21	$[TGCC]_4$ $[TTCC]_{15}$			G>T
			$[TGCC]_4$ $[TTCC]_{17}$			

Table 4.6 continued

Study case no.	Genotype previously obtained by CE	Genotype obtained by sequencing	Sequence structure	Most likely presumed mutation model	Presumed Mutation origin	Other variants
D2S1338 Case 4	Mother: 19/25	N/A	N/A	(-) One-step mutation	Paternal	N/A
	Male child: 20/25	Male child: 20/25	$[TGCC]_8$ $[TTCC]_9$ [GTCC] $[TTCC]_2$			None
			$[TGCC]_8$ $[TTCC]_{14}$ [GTCC] $[TTCC]_2$			
	Father: 18/21	Father: 18/21	$[TGCC]_7$ $[TTCC]_9$ [GTCC] $[TTCC]_2$			None
			$[TGCC]_8$ $[TTCC]_{10}$ [GTCC] $[TTCC]_2$			
D2S1338 Case 5	Mother: 16/18	N/A	N/A	(-) One-step mutation	Paternal	N/A
	Female child: 16/23	Female child:16/23	$[TGCC]_6$ $[TTCC]_7$ [GTCC] $[TTCC]_2$			G>T
			$[TGCC]_5$ $[TTCC]_{14}$ [GTCC] $[TTCC]_2$			
	Father: 20/24	Father: 20/24	$[TGCC]_7$ $[TTCC]_{10}$ [GTCC] $[TTCC]_2$			None
			$[TGCC]_6$ $[TTCC]_{15}$ [GTCC] $[TTCC]_2$			

4.4.5 The D3S1358 locus

The D3S1358 locus has a $[AGAT]_n[AGAC]_n[AGAT]_2$ compound tetranucleotide repeat. The allele size range is from 9 to 20 repeats (Gettings *et al*., 2015). The mutation rate is estimated to be 0.12% (Butler, 2015). The D3S1358 locus included two standard trio cases. The obtained sequence-based genotypes were 100% concordant with the CE-based genotypic data. The summary of the results is shown in **(Table 4.7)**.

In D3S1358 case 1, the previously reported genotypes were homozygous 16/16, 16/16, and 17/17 for the mother, female child, and father, accordingly. The genotypic data was verified by the sequencing results. There was a loss of one [AGAC] repeat unit from allele 17 of the father, resulting in allele 16 in the child. The loss of one repeat unit suggested a (-) one-step mutation as the most likely presumed mutation model. The mutation was from the paternal lineage. In D3S1358 case 2, the genotypic data was previously determined to be heterozygous 15/17, 17/19, and 16/18 for the mother, male child, and father, respectively. The child and the mother shared allele 17; thus, the mismatch was between allele 18 of the father and allele 19 of the child. There was a gain of one [AGAT] repeat unit, suggesting a (+) one-step mutation model from the paternal origin. In this locus, the loss or gain of either the [AGAT] or [AGAC] repeat units has previously been described in the literature (Heinrich *et al*., 2005; Shao *et al*., 2021), and is thus not novel in this study.

Table 4.7: Results obtained for the D3S1358 locus parentage cases. The table indicates the genotype previously obtained by CE, genotypes obtained by sequencing, sequence structure, most likely presumed mutation model and presumed mutation origin.

Study case no.	Genotype previously obtained by CE	Genotype obtained by sequencing	Sequence structure	Most likely presumed mutation model	Presumed mutation origin
D3S1358 Case 1	Mother: 16/16	N/A	N/A	(-) One-step mutation	Paternal
	Female child: 16/16	Female child: 16/16	$[AGAT]_{12}[AGAC]_2[AGAT]_2$		
	Father: 17/17	Father: 17/17	$[AGAT]_{12}[AGAC]_3[AGAT]_2$		
D3S1358 Case 2	Mother:15/ 17	N/A	N/A	(+) One-step mutation	Paternal
	Male child:17/19	Male child:17/19	$[AGAT]_{14}$ [AGAC] $[AGAT]_2$		
			$[AGAT]_{15}$ $[AGAC]_2$ $[AGAT]_2$		
	Father: 16/18	Father: 16/18	$[AGAT]_{12}$ $[AGAC]_2$ $[AGAT]_2$		
			$[AGAT]_{14}$ $[AGAC]_2$ $[AGAT]_2$		

4.4.6 The D5S818 locus

The D5S818 locus contains a simple [AGAT] tetranucleotide repeat with allele sizes ranging from 7 to 18 repeats (Gettings *et al*., 2015). The estimated mutation rate at this locus is 0.11% (Butler, 2015). The D5S818 locus included two parentage cases. A 50% concordance between the CE-based genotypic data and sequence-based alleles was obtained in this locus. **Table 4.8** shows an overview of the findings.

In D5S818 case 1, a mother, a female child and a father were analysed. The previously obtained genotypes using CE were reported as heterozygous 11/13 for the mother, 13/14 for the female child and a homozygous 13/13 for the father were confirmed by sequencing. Two possibilities were considered to identify the alleles responsible for the mismatch. Firstly, if the child inherited allele 13 from the father, the mismatch would be between the mother's allele 13 and the child's allele 14. Secondly, if the child inherited allele 13 from the mother, the mismatch would be between allele 13 of the father and the child's allele 14. The latter was considered to be more likely, as STR mutations are seen more frequently in males than in females. Only the child's and the father's alleles were sequenced. Interestingly, a thymine to cytosine (T>C) SNV located 13 bp upstream of the [AGAT] repeat motif was detected in both individuals **(Figure 4.12a)**. This base substitution has been characterised and reported as rs25768 on the dbSNP (available at https://www.ncbi.nlm.nih.gov/snp/rs25768). A study by Gettings *et al*. (2015) reported the T>C SNV to have a frequency of 16%. The T>C SNV did not impact the previous CE-based genotypic data, as there was concordance with the sequence-based genotypes. In contrast, Edwards and Allen (2005) and Fujii *et al*. (2016) have reported this SNV to affect the CE-based genotypic data of the D5S818 locus. As a result, it is critical for DNA testing facilities and researchers to be aware of SNVs at STR primer-binding sites, as this provides useful information for developing more efficient primer sets for STR typing. Furthermore, the most likely mutation model responsible for the allelic mismatch observed was presumed to be a (+) one-step mutation model from the paternal or maternal lineage.

D5S818 case 2 involved a female child and a father. The genotypes obtained by CE were a homozygous 12/12 for the child and a 13/13 for the father. However, upon sequencing, the child's genotype was determined to be heterozygous 12/13 and the father's genotypes remained as homozygous 13/13. The child was further determined to have a guanine to thymine (G>T) base substitution 4 bp downstream of the repeat region **(Figure 4.12b)**. This resulted in an additional [AGAT] repeat unit, leading to heterozygous allele 12/13. The child also had a G>T SNV. In the father, a G>C SNV interrupting the [AGAT] repeat motif was detected **(Figure 4.12c)**. The resulting sequence motif was $[AGAT]_9$ [ACAT] $[AGAT]_3$. A similar sequence motif was described by Allor *et al*. (2005). However, the sequence was observed at D5S818 allele 18. When following the recently updated nomenclature based on the STRBase website (STRBase Version 2.0, last update to data content: June 2023) the [ACAT] repeat motif is included in D5S818 allele designation. Therefore, the father's genotype remained as homozygous 13/13. Furthermore, the T>C SNV identified in case 1 was also detected in individuals of case 2. A study by Edwards and Allen (2005) strongly suggested that SNVs affect alleles 11,12,13 and 14 of the D5S818 locus. This finding was confirmed in this study as all the genotypes involved in the mismatches in this locus were composed of similar alleles. DNA testing facilities should take this into consideration when dealing with cases containing the alleles involved in this phenomenon, as this may impact the resulting genotypes.

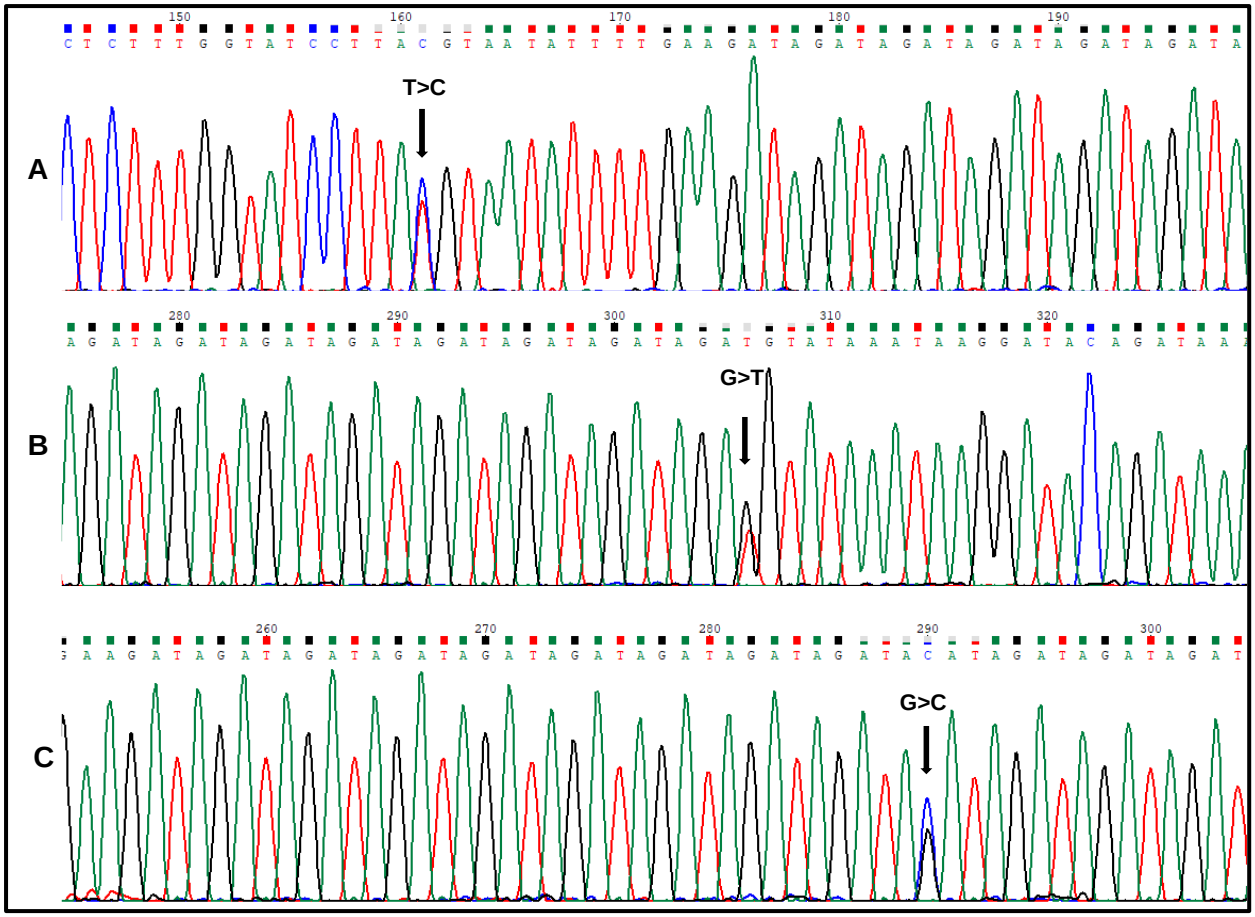


Figure 4.12: A sequence electropherogram depicting the SNVs detected at the D5S818 STR locus. a) shows the T>C SNV, b) shows the G>T SNV, and c) shows the G>C SNV.

Table 4.8: Results obtained for the D5S818 locus parentage cases. The table indicates the genotype previously obtained by CE, genotypes obtained by sequencing, sequence structure, most likely presumed mutation model and presumed mutation origin.

Study case no.	Genotype previously obtained by CE	Genotype obtained by sequencing	Sequence structure	Presumed mutation model	Presumed mutation origin	Other variants
D5S818 Case 1	Mother: 11/13	N/A	N/A	(+) One-step mutation	Paternal	N/A
	Female child: 13/14	Female child: 13/14	$[AGAT]_{13}$			T>C
			$[AGAT]_{14}$			
	Father: 13/13	Father: 13/13	$[AGAT]_{13}$			T>C
D5S818 Case 2	Female child:12/12	Female child:12/13	$[AGAT]_{12}$	Primer-binding site mutation	Paternal	T>C; G>T
			$[AGAT]_{13}$			
	Father: 13/13	Father: 13/13	$[AGAT]_{9}$ [ACAT] $[AGAT]_{3}$			T>C G>C

4.4.7 The D6S1043 locus

The D6S1043 locus has a simple [AGAT] tetranucleotide repeat with allele sizes ranging from 9 to 25 repeats (Gettings *et al.,* 2015). The mutation rate in this locus is reported to be approximately 0.03% (Lu *et al*., 2012). A total of 18 parentage cases with parent-child allelic mismatches at this locus were evaluated. A 28% concordance was obtained between the CE-based genotypic data and the sequence-based genotypes. **Table 4.9** provides a summary of the findings.

Following sequence analysis of samples studied in this locus, an important observation was made: all samples contained one [ACAT] repeat, interrupting the previously described D6S1043 core repeat [AGAT]n structure on the reference sequences G08539 and AL132766.13 (STRBase Version 1.0), causing the core repeat to become compound $[AGAT]_n$ [ACAT] $[AGAT]_n$. However, this repeat pattern was reported as “other patterns” on the recently updated version of the STRBase website. Furthermore, the D6S1043 core repeat sequence pattern is reported as $[AGAT]_n$ $[ACAT]_n$ $[AGAT]_n$ (STRBase Version 2.0, last update to data content: June 2023). It is of importance that DNA testing facilities as well as forensic units take note of updates of STR loci nomenclature, as this can influence the CE-based genotypic data and sequence-based STR allele designation.

In the parentage cases studied in this locus, it was observed that 72% (n=13/18) of cases harboured a cytosine to thymine (C>T) base substitution located 87 bp upstream of the D6S1043 core repeat **(Figure 4.13)**. The observed SNV resulted in the dropout of alleles 19, 20 and 21 in cases (6, 10, 15, and 16), (1, 7, 8, 9, 13, 14, 17, and 20), and (11), respectively. Notably, all the cases with the null alleles had opposite homozygous genotypes, further emphasising the importance of investigating parent-child allelic mismatches with opposite homozygous alleles observed in parentage analysis. To our knowledge, there is currently no record of the C>T SNV in the literature, making this SNV a novel finding. The relatively high frequency of the detected SNV suggests that it is likely to be a population-specific variant. It would be interesting to note this SNV in other African regions as it may be ancestry-specific. At

present, there are no SNVs reported at greater than 5% frequency for the D6S1043 locus (Gettings *et al*., 2015; Jian *et al*., 2021).

Approximately 77% (n=10/13) of the allelic dropouts were of paternal origin, while only 23% (n=3/13) were maternally transmitted. However, no associations can be made in the literature as there is minimal information on null alleles reported at the D6S1043 locus. The reported data mainly focus on stutter peak formation and detected micro-deletions as attributes to mutations identified at the D6S1043 locus (Bright *et al*., 2014; Martinez *et al.,* 2017; Wu *et al*., 2023).

In four of the 18 cases studied, we identified parent-child allelic mismatches attributed to DNA polymerase slippage. The first instance was cases 4, 5, and 12, where the most likely mutation model was characterised as a (+) one-step mutation. The mismatch in case 4 originated from the maternal lineage, while in cases 5 and 12, it was from the paternal line. Secondly, in case 2 of this locus, a (-) one-step mutation from the paternal origin was presumed. Lastly, in case 3, we observed a (-) five-step mutation model transmitted from the maternal lineage. The CE-based genotypic data and sequence-based alleles indicated that the child inherited allele 14 from the father. Therefore, there was a loss of five [AGAT] repeats in the mother's allele 19, leading to allele 14 in the child. In this study, the D6S1043 locus was observed to be the locus with the highest number of apparent parent-child allelic mismatches. It would be of interest to further study this locus in an attempt to determine the mutation rate in our region. This could only be feasible if the study population is larger for this specific STR locus. Nevertheless, important differences were noted between the CE-based genotypic data and sequencing data that needs to be taken into consideration when DNA analysis is performed.

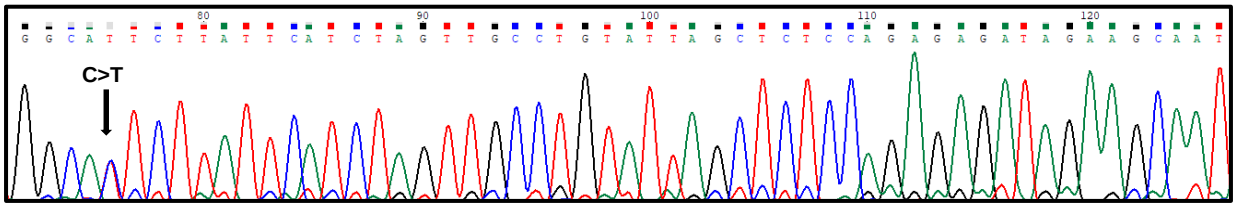


Figure 4.13: A sequence electropherogram depicting the C>T SNV detected at the D6S1043 STR locus

Table 4.9: Results obtained for the D6S1043 locus parentage cases. The table indicates the genotype previously obtained by CE, genotypes obtained by sequencing, sequence structure, most likely presumed mutation model and presumed mutation origin.

Study case no.	Genotype previously obtained by CE	Genotype obtained by sequencing	Sequence structure	Most likely presumed mutation model	Presumed mutation origin	Other variants
D6S1043 Case 1	Mother: 12/16	N/A	N/A	Primer-binding site mutation	Paternal	N/A
	Male child: 16/16	Male child: 16/20	$[AGAT]_{10}$ $[ACAT]$ $[AGAT]_{5}$			C>T
			$[AGAT]_{10}$ $[ACAT]$ $[AGAT]_{9}$			
	Father: 11/11	Father: 11/20	$[AGAT]_{5}$ $[ACAT]$ $[AGAT]_{5}$			C>T
			$[AGAT]_{14}$ $[ACAT]$ $[AGAT]_{5}$			
D6S1043 Case 2	Mother: 12/12	N/A	N/A	(-) One-step mutation	Paternal	N/A
	Female child: 21/21	Female child: 20/21	$[AGAT]_{14}$ $[ACAT]$ $[AGAT]_{5}$			None
			$[AGAT]_{15}$ $[ACAT]$ $[AGAT]_{5}$			
	Father: 19/22	Father: 19/22	$[AGAT]_{13}$ $[ACAT]$ $[AGAT]_{5}$			None
			$[AGAT]_{16}$ $[ACAT]$ $[AGAT]_{5}$			
D6S1043 Case 3	Mother: 19/19	Mother: 19/19	$[AGAT]_{13}$ $[ACAT]$ $[AGAT]_{5}$	(-) Five-step mutation	Maternal	None
			$[AGAT]_{15}$ $[ACAT]$ $[AGAT]_{5}$			
	Male child: 14/14	Male child: 14/14	$[AGAT]_{9}$ $[ACAT]$ $[AGAT]_{5}$			None
	Father: 11/14	N/A	N/A			N/A
D6S1043 Case 4	Mother: 11/15	Mother: 11/15	$[AGAT]_{6}$ $[ACAT]$ $[AGAT]_{5}$	(+) One-step mutation	Maternal	None
			$[AGAT]_{10}$ $[ACAT]$ $[AGAT]_{5}$			
	Male child: 12/18	Male child: 12/18	$[AGAT]_{6}$ $[ACAT]$ $[AGAT]_{5}$			None
			$[AGAT]_{12}$ $[ACAT]$ $[AGAT]_{5}$			
	Father: 18/20	N/A	N/A			N/A

Table 4.9 continued

Study case no.	Genotype previously obtained by CE	Genotype obtained by sequencing	Sequence structure	Most likely presumed mutation model	Presumed mutation origin	Other variants
D6S1043 Case 5	Mother: 11/15	N/A	N/A	(+) One-step mutation	Paternal	C>T
	Male child:15/19	Male child:15/19	$[AGAT]_9$ [ACAT] $[AGAT]_5$			C>T
			$[AGAT]_{13}$ [ACAT] $[AGAT]_5$			
	Father: 15/18	Father: 15/18	$[AGAT]_5$ [ACAT] $[AGAT]_5$			
			$[AGAT]_9$ [ACAT] $[AGAT]_5$			N/A
D6S1043 Case 6	Mother: 11/11	Mother: 11/19	$[AGAT]_5$ [ACAT] $[AGAT]_5$	Primer-binding site mutation	Maternal	C>T
			$[AGAT]_{13}$ [ACAT] $[AGAT]_5$			
	Male child:12/12	Male child:12/19	$[AGAT]_6$ [ACAT] $[AGAT]_5$			C>T
			$[AGAT]_{13}$ [ACAT] $[AGAT]_5$			
	Father: 12/12	N/A	N/A			N/A
D6S1043 Case 7	Mother: 13/17	N/A	N/A	Primer-binding site mutation	Paternal	
	Male child: 13/13	Male child: 13/20	$[AGAT]_7$ [ACAT] $[AGAT]_5$			C>T
			$[AGAT]_{14}$ [ACAT] $[AGAT]_5$			
	Father: 12/12	Father: 12/20	$[AGAT]_6$ [ACAT] $[AGAT]_5$			C>T
			$[AGAT]_{14}$ [ACAT] $[AGAT]_5$			
D6S1043 Case 8	Mother: 12/15	N/A	N/A	Primer-binding site mutation	Paternal	N/A
	Female child: 12/12	Female child: 12/20	$[AGAT]_6$ [ACAT] $[AGAT]_5$			C>T
			$[AGAT]_{14}$ [ACAT] $[AGAT]_5$			
	Father: 11/11	Father: 11/20	$[AGAT]_5$ [ACAT] $[AGAT]_5$			C>T
			$[AGAT]_{14}$ [ACAT] $[AGAT]_5$			

Table 4.9 continued

Study case no.	Genotype previously obtained by CE	Genotype obtained by sequencing	Sequence structure	Most likely presumed mutation model	Presumed mutation origin	Other variants
D6S1043 Case 9	Mother: 13/21	Mother: 13/21	N/A	Primer-binding site mutation	Paternal	
	Female child: 13/13	Female child: 13/20	$[AGAT]_7$ [ACAT] $[AGAT]_5$			C>T
			$[AGAT]_{14}$ [ACAT] $[AGAT]_5$			
	Father: 16/16	Father: 16/20	$[AGAT]_{10}$ [ACAT] $[AGAT]_5$			C>T
			$[AGAT]_{14}$ [ACAT] $[AGAT]_5$			
D6S1043 Case 10	Male child:18/18	Male child:18/19	$[AGAT]_{12}$ [ACAT] $[AGAT]_5$	Primer-binding site mutation	Paternal	C>T
			$[AGAT]_{13}$ [ACAT] $[AGAT]_5$			
	Father: 12/18	Father: 12/19	$[AGAT]_6$ [ACAT] $[AGAT]_5$			C>T
			$[AGAT]_{13}$ [ACAT] $[AGAT]_5$			
D6S1043 Case 11	Mother: 11/11	Mother: 11/21	$[AGAT]_5$ [ACAT] $[AGAT]_5$	Primer-binding site mutation	Maternal	C>T
			$[AGAT]_{15}$ [ACAT] $[AGAT]_5$			
	Female child: 14/14	Female child 1: 14/21	$[AGAT]_8$ [ACAT] $[AGAT]_5$			C>T
			$[AGAT]_{15}$ [ACAT] $[AGAT]_5$			
	Father: 14/17	N/A	N/A			N/A
D6S1043 Case 12	Mother: 15/17	Mother: 15/17	N/A	(+) One-step mutation	Paternal	N/A
	Female child: 17/19	Female child: 17/19	$[AGAT]_{11}$ [ACAT] $[AGAT]_5$			None
			$[AGAT]_{13}$ [ACAT] $[AGAT]_5$			
	Father: 11/18	Father: 11/18	$[AGAT]_5$ [ACAT] $[AGAT]_5$			None
			$[AGAT]_{12}$ [ACAT] $[AGAT]_5$			

Table 4.9 continued

Study case no.	Genotype previously obtained by CE	Genotype obtained by sequencing	Sequence structure	Most likely presumed mutation model	Presumed mutation origin	Other variants
D6S1043 Case 13	Mother: 12/12	Mother: 12/20	$[AGAT]_6$ [ACAT] $[AGAT]_5$	Primer-binding site mutation	Maternal	C>T
			$[AGAT]_{14}$ [ACAT] $[AGAT]_5$			
	Male child: 14/14	Male child: 14/20	$[AGAT]_8$ [ACAT] $[AGAT]_5$			C>T
			$[AGAT]_{14}$ [ACAT] $[AGAT]_5$			
	Father: 13/13	N/A	N/A			N/A
D6S1043 Case 14	Male child: 17/17	Male child: 17/20	$[AGAT]_{11}$ [ACAT] $[AGAT]_5$	Primer-binding site mutation	Paternal	C>T
			$[AGAT]_{14}$ [ACAT] $[AGAT]_5$			
	Father: 11/11	Father: 11/20	$[AGAT]_5$ [ACAT] $[AGAT]_5$			C>T
			$[AGAT]_{14}$ [ACAT] $[AGAT]_5$			
D6S1043 Case 15	Mother: 11/16	N/A	N/A	Primer-binding site mutation	Paternal	N/A
	Male child: 16/16	Male child: 16/19	$[AGAT]_{10}$ [ACAT] $[AGAT]_5$			C>T
			$[AGAT]_{13}$ [ACAT] $[AGAT]_5$			
	Father: 15/15	Father: 15/19	$[AGAT]_9$ [ACAT] $[AGAT]_5$			C>T
			$[AGAT]_{13}$ [ACAT] $[AGAT]_5$			
D6S1043 Case 16	Mother: 19/19	N/A	N/A	Primer-binding site mutation	Paternal	N/A
	Male child: 19/19	Male child: 19/19	$[AGAT]_{13}$ [ACAT] $[AGAT]_5$			C>T
	Father:12/12	Father:12/19	$[AGAT]_6$ [ACAT] $[AGAT]_5$			C>T
			$[AGAT]_{13}$ [ACAT] $[AGAT]_5$			

Table 4.9 continued

Study case no.	Genotype previously obtained by CE	Genotype obtained by sequencing	Sequence structure	Most likely presumed mutation model	Presumed mutation origin	Other variants
D6S1043 Case 17	Mother: 14/14	Mother: 14/20	$[AGAT]_8$ [ACAT] $[AGAT]_5$	Primer-binding site mutation	Maternal	C>T
			$[AGAT]_{14}$ [ACAT] $[AGAT]_5$			
	Female child: 19/19	Female child: 19/20	$[AGAT]_{13}$ [ACAT] $[AGAT]_5$			C>T
			$[AGAT]_{14}$ [ACAT] $[AGAT]_5$			
	Father: 12/19	N/A	N/A			N/A
D6S1043 Case 18	Mother: 17/18	N/A	N/A	Primer-binding site mutation	Paternal	N/A
	Male child: 18/18	Male child: 18/20	$[AGAT]_{12}$ [ACAT] $[AGAT]_5$			C>T
			$[AGAT]_{14}$ [ACAT] $[AGAT]_5$			
	Father: 12/12	Father: 12/20	$[AGAT]_6$ [ACAT] $[AGAT]_5$			C>T
			$[AGAT]_{14}$ [ACAT] $[AGAT]_5$			

4.4.8 The D7S820 locus

The D7S820 locus contains a simple [GATA] tetranucleotide. The locus has alleles ranging in size from 6 to 15 repeats (Gettings *et al*., 2015). The mutation rate at this locus is approximately 0.10% (Butler, 2015). The D7S820 locus included two parentage cases. The obtained sequence-based alleles were 100% concordant with the previously reported CE-based genotypic data. **Table 4.10** shows an overview of the findings.

D7S820 case 1 had a mother, a female child, and a father. The genotypic data obtained by CE was reported as heterozygous 8/10, 10/13 and 11/12, respectively. Identical genotypes were observed by sequencing. The child inherited allele 10 from the mother; hence the mismatch was between allele 12 of the father and allele 13 of the child. A gain of one [GATA] repeat motif in the father's allele 12 leading to allele 13 in the child was observed. This mutation was from the paternal lineage presumed to be most likely a (+) one-step mutation. However, in the child's sequence data, three SNVs were identified; Arginine to Thymine (A>T) **(Figure 4.14a)**, Thymine to Guanine (T>G) **(Figure 4.14b)**, and Guanine to Arginine (G>A) **(Figure 4.14c)**. In contrast, the father's sequences data showed only the A>T SNV. These SNVs were located at 21 bp, 66 bp, and 116 bp downstream of the D7S820 repeat region. Interestingly, the T>G SNV identified in the child is located in the primer-binding site of the D7S820 reverse primer published on STRBase. However, it did not affect the heterozygous allelic profile of the child as it remained unchanged. Notably, the A>T and G>A SNVs are located downstream of the D7S820 repeat motif flanking region, which may possibly be primer-binding sites of the D7S820 reverse primers in other STR commercial kits.

The second case studied for this locus involved a female child and a father. The genotypes were a heterozygous 8/12 for the child and a homozygous 11 for the father. The genotypic data was confirmed by the sequencing results. The most likely presumed mutation model was a (+) one-step mutation, indicating that allele 11 of the father was inherited as allele 12 by the child. Similar to the female child of case 1, the child in case 2 also harboured the A>T, T>G, and G>A SNVs. The father only had the T>G and G>A SNVs.

A total of three SNVs were identified in individuals tested at this locus. These SNVs were characterised by Gettings *et al*. (2015) and are now included in updated STRBase website. Firstly, the A>T SNV is characterised as rs7789995 with a frequency of 7%. The T>G SNV was characterised as rs7786079 with a frequency of 8%, and the G>A SNV characterised as rs59186128 a frequency of 8%. Numerous SNVs have been reported to impact this locus (De Bruyn *et al*., 2019). However, in this study, the detected SNVs had no impact on the previously reported CE-based genotypic data as the repeat length did not change due to these variants. However, it is important to characterise and report such variants detected at STR loci as this may aid in developing more efficient primers for STR typing.

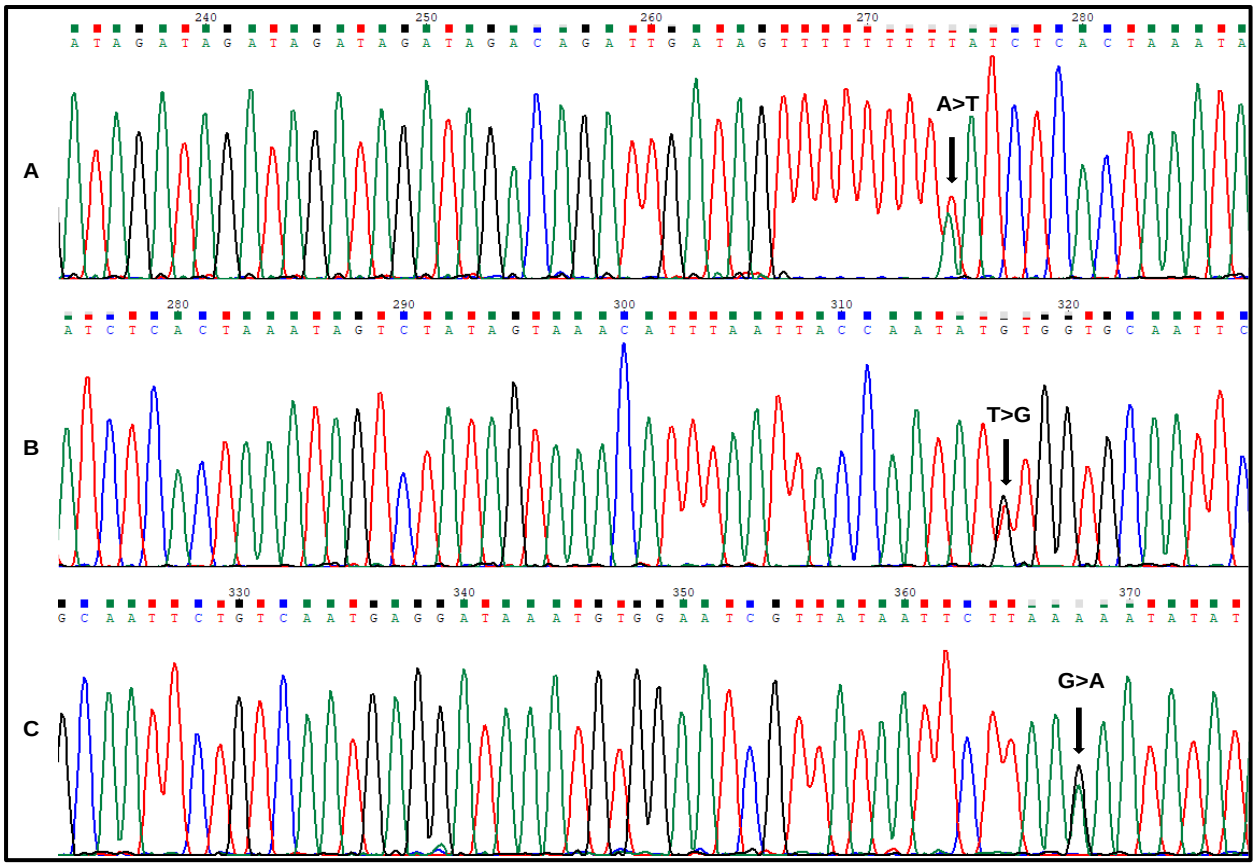


Figure 4.14: A sequence electropherogram depicting the SNVs detected at the D7S820 STR locus. a) shows the A>T SNV, b) shows the T>G SNV, and c) shows the G>A SNV.

Table 4.10: Results obtained for the D7S820 locus parentage cases. The table indicates the genotype previously obtained by CE, genotypes obtained by sequencing, sequence structure, most likely presumed mutation model and presumed mutation origin.

Study case no.	Genotype previously obtained by CE	Genotype obtained by sequencing	Sequence structure	Most likely presumed mutation model	Presumed mutation origin	Other variants
D7S820 Case 1	Mother: 8/10	N/A	N/A	(+) One-step mutation	Paternal	N/A
	Female child: 10/13	Female child: 10/13	$[GATA]_{10}$			A>T; T>G; G>A
			$[GATA]_{13}$			
	Father: 11/12	Father: 11/12	$[GATA]_{11}$			A>T
			$[GATA]_{12}$			
D7S820 Case 2	Female child: 8/12	Female child: 8/12	$[GATA]_{8}$	(+) One-step mutation	Paternal	A>T; T>G; G>A
			$[GATA]_{12}$			
	Father: 11/11	Father: 11/11	$[GATA]_{11}$			T>G; G>A

4.4.9 The D8S1179 locus

The D8S1179 locus contains a compound tetranucleotide. It primarily consists of alleles containing a [TCTA] repeat. However, a [TCTG] repeat joins the motif for alleles larger than 13 repeats. This is usually observed at the second or third position from the 5'-end of the repeat sequence (Gettings *et al*., 2015). The allele size range is between 5 and 19 repeats. The mutation rate at this locus is estimated at 0.14% (Butler, 2015). A total of six parentage cases were studied in this locus, and the findings are presented in **Table 4.11**. A concordance of 50% was obtained between the previously reported CE-based genotypic data and the obtained sequence-based alleles.

In three parentage cases (1, 2, and 4), the parent-allelic mismatches were characterised to be most likely one-stepwise mutation models attributed to DNA polymerase slippage. The first case studied was a paternity duo. The CE-based genotypic data was reported as a homozygous 15/15 for the male child and a heterozygous 14/16 for the father. The obtained sequencing data supported the CE-based genotypic data. Moreover, it was possible that the allelic mismatch was a (+) one-step mutation observed between the father's allele 14 and the child's allele 15, resulting from a gain of a [TCTA] repeat. It was also possible that the mismatch was because of a loss of one [TCTA] repeat, leading to allele 16 of the father being inherited as allele 15 in the child. The allelic mismatch was, therefore, characterised as a (+/-) one-step mutation model from the paternal lineage.

The second case studied in this locus was a standard trio, with CE-based genotypic data noted as homozygous 13/13 for the mother, heterozygous 13/15 for the female child and heterozygous 12/14 for the father. The sequencing results confirmed these alleles. It appeared that the child inherited allele 13 from the mother; thus, a paternal mismatch was identified between the child's allele 15 and the father's allele 14. The paternal allelic mismatch was characterised as a (+) one-step mutation due to a gain of a [TCTA] repeat. D8S1179 case 4 consisted of a mother, a male child and a father. Their genotypes were 14/14, 14/15, and 11/16, respectively. The allelic mismatch was identified between the child's allele 15 and the father's allele 16. This apparent

mismatch was characterised as a (-) one-step mutation model, resulting from a loss of one [TCTA] repeat unit.

D8S1179 case 3 involved a male child and a mother, with CE-based genotypic data determined as homozygous 14/14 and 15/15, respectively. The sequencing data only confirmed the mother's genotype. However, there was no concordance between the child's CE-based alleles and the obtained sequence-based alleles (14/15). Moreover, this observation could not be explained by the stepwise mutation model. Further sequence analysis identified a five base pair deletion (AATTA) located 38 bp downstream of the [TCTA] repeat structure in both the child and the mother **(Figure 4.15)**. This finding suggested the presence of an allelic dropout caused by the deletion of a sequence fragment at the [TCTA] repeat flanking region.

Interestingly, an identical observation was made for individuals evaluated in cases 5 and 6, where the deletion led to the dropout of allele 15. Therefore, an allelic dropout due to delAATTA was presumed to be the most likely mutation model for these cases. This deletion has not previously been described in the literature, making this finding novel. A study by Hering *et al*. (2005) identified a four base pair deletion (TAAA) occurring 36 bp downstream of the D8S1179 core repeat. The authors described an outcome similar to our findings: the four base pair deletion resulted in an allele dropout. Furthermore, a paternity case study by Xiao *et al*. (2019) characterised a microdeletion at position 8q24.13, about 2.99 to 49.76 kb, detected in both the father and his son. These findings suggest that parent-child allelic mismatches should not be rashly attributed to DNA polymerase slippage, especially in cases with opposite homozygotes. Moreover, this phenomenon was identified in 50% (n=3/6) of the cases included in this locus, which had opposite homozygous genotypes.

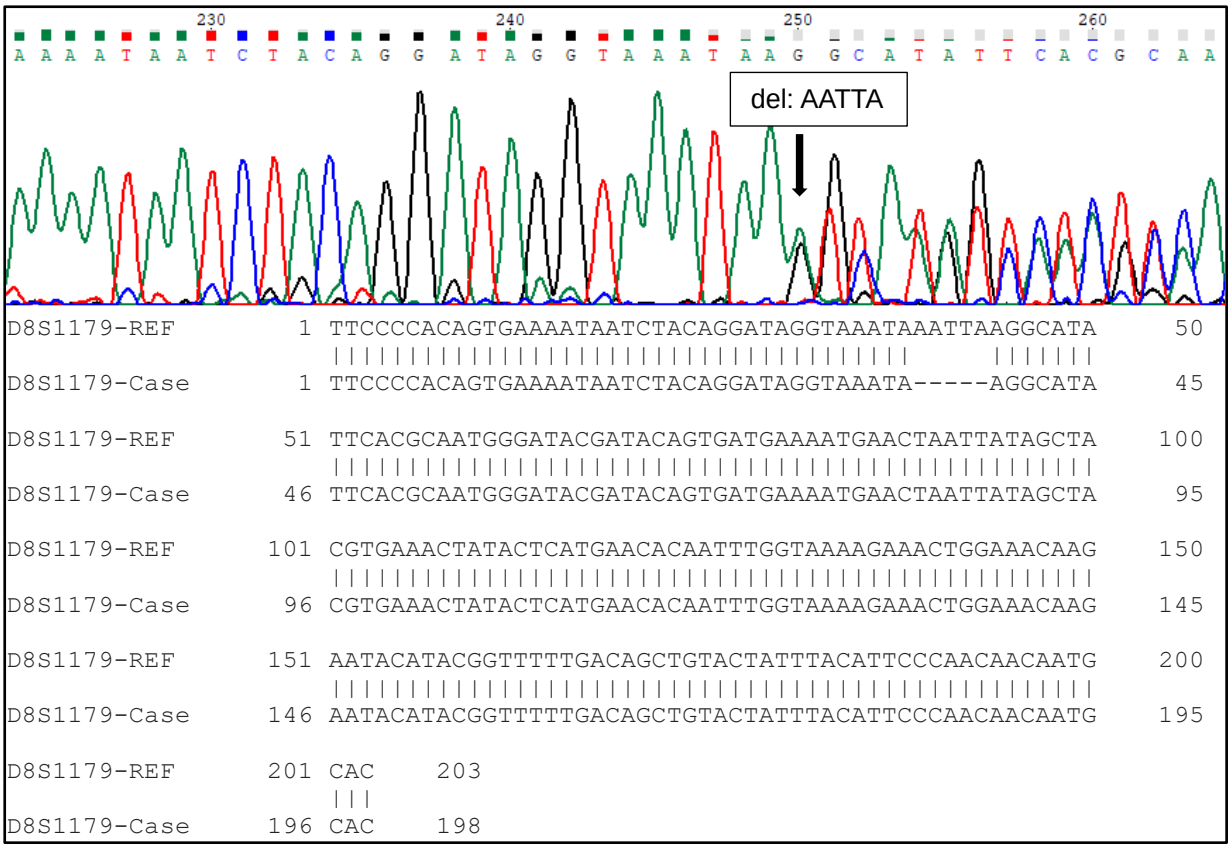


Figure 4.15: A sequence electropherogram and sequence alignment depicting the delAATTA detected at the D8S1179. D8S1179-ref: reference sequence (AF216671), D8S1179-Case: a sequence obtained at one of the D8S1179 cases with the deletion.

Table 4.11: Results obtained for the D8S1179 locus parentage cases. The table indicates the genotype previously obtained by CE, genotypes obtained by sequencing, sequence structure, most likely presumed mutation model and presumed mutation origin.

Study case no.	Genotype previously obtained by CE	Genotype obtained by sequencing	Sequence structure	Most likely presumed mutation model	Presumed mutation origin	Other variants
D8S1179 Case 1	Male child:15/15	Male child:15/15	$[TCTA]_2$ [TCTG] $[TCTA]_{12}$	(+/-) One-step mutation	Paternal	None
	Father: 14/16	Father: 14/16	$[TCTA]_2$[TCTG] $[TCTA]_{11}$			
			$[TCTA]_2$[TCTG] $[TCTA]_{13}$			
D8S1179 Case 2	Mother:13/13	N/A	N/A	(+) One-step mutation	Paternal	None
	Female child:13/15	Female child:13/15	[TCTA] [TCTG] $[TCTA]_{11}$			
			[TCTA] [TCTG] $[TCTA]_{13}$			
	Father: 12/14	Father: 12/14	$[TCTA]_2$ [TCTG] $[TCTA]_9$			
			$[TCTA]_2$ [TCTG] $[TCTA]_{11}$			
D8S1179 Case 3	Male child: 14/14	Male child: 14/15	[TCTA] [TCTG] $[TCTA]_{12}$	Allelic dropout	Maternal	delAATTA
			[TCTA] [TCTG] $[TCTA]_{13}$			
	Mother: 15/15	Mother:15/15	[TCTA] [TCTG] $[TCTA]_{13}$			
D8S1179 Case 4	Mother: 14/14	N/A	N/A	(-) One-step mutation	Paternal	None
	Male child:14/15	Male child:14/15	[TCTA] [TCTG] $[TCTA]_{12}$			
			[TCTA] [TCTG] $[TCTA]_{13}$			
	Father:11/16	Father:11/16	$[TCTA]_2$ [TCTG] $[TCTA]_8$			
			[TCTA] [TCTG] $[TCTA]_{14}$			

Table 4.11 continued

Study case no.	Genotype previously obtained by CE	Genotype obtained by sequencing	Sequence structure	Most likely presumed mutation model	Presumed mutation origin	Other variants
D8S1179 Case 5	Mother:12/14	Mother:12/14	N/A	Allelic dropout	Paternal	delAATTA
	Female child 1:12/12	Female child 1:12/15	[TCTA] [TCTG] $[TCTA]_{10}$			
			[TCTA] [TCTG] $[TCTA]_{13}$			
	Female child 2:14/14	Female child 2:14/15	[TCTA] [TCTG] $[TCTA]_{12}$			
			[TCTA] [TCTG] $[TCTA]_{13}$			
	Father: 13/13	Father: 13/15	[TCTA] [TCTG] $[TCTA]_{11}$			
			[TCTA] [TCTG] $[TCTA]_{13}$			
D8S1179 Case 6	Mother: 13/15	Mother: 13/15	N/A	Allelic dropout	Paternal	delAATTA
	Female child:15/15	Female child:15/15	[TCTA] [TCTG] $[TCTA]_{13}$			
	Father:14	Father:14/15	[TCTA] [TCTG] $[TCTA]_{12}$			
			[TCTA] [TCTG] $[TCTA]_{13}$			

4.4.10 The D10S1248 locus

The D10S1248 locus contains a simple [GGAA] tetranucleotide repeat. The allele size ranges from 8 to 19 repeats, with a mutation rate reported at 0.25% (Hill *et al*., 2011). In this locus, three paternity cases were studied, and the results a presented in **Table 4.12**. A 100% concordance was obtained between the CE-based genotypic data and sequence-based alleles.

D10S1248 case 1 consisted of a male child and a father. Previously determined genotypes were a heterozygous 11/13 for the child and a homozygous 14/14 for the father. These genotypes were confirmed by sequencing performed in this study. It was assumed that the father's allele 14 was passed to the child as allele 13 because of a loss of one [GGAA] repeat. Therefore, the most likely presumed mutation model was a (-) one-step mutation originating from the paternal lineage rather than a (-) two-step mutation (14>11). Moreover, the literature supports that one-step mutations are observed more frequently than two-step and multi-step mutation models as previously mentioned, because of the increased rate at which one-step mutations are reported to occur in STR loci (Hamester *et al*., 2019).

In the second case, a female child and father were involved. Their genotypic data was homozygous 13/13 and 14/14, respectively. There was a loss of one repeat unit in allele 14 of the father, resulting in allele 13 in the child, suggesting a (-) one-step mutation model. In D10S1248 case 3, a mother, female child and father were tested accordingly with genotypes 12/14, 12/14 and 13/15. The mother's genotype provided no information in determining the affected allele in the child; therefore, a (-) one-step mutation model was suggested as most likely. The (-) one-step mutation was possible for both alleles 13 and 15 of the father, leading to alleles 12 or 14 in the child, respectively. The observed results at this locus for the three parentage cases studied was likely due to (-) one-step mutations resulting from DNA polymerase slippage. All the observed mutations were of paternal origin.

Table 4.12: Results obtained for the D10S1248 locus parentage cases. The table indicates the genotype previously obtained by CE, genotypes obtained by sequencing, sequence structure, most likely presumed mutation model and presumed mutation origin.

Study case no.	Genotype previously obtained by CE	Genotype obtained by sequencing	Sequence structure	Most likely presumed mutation model	Presumed mutation origin
D10S1248 Case 1	Male child: 11/13	Male child: 11/13	$[GGAA]_{11}$	(-) One-step mutation	Paternal
			$[GGAA]_{13}$		
	Father: 14/14	Father: 14/14	$[GGAA]_{14}$		
D10S1248 Case 2	Female child: 13/13	Female child: 13/13	$[GGAA]_{13}$	(-) One-step mutation	Paternal
	Father: 14/14	Father: 14/14	$[GGAA]_{14}$		
D10S1248 Case 3	Mother: 12/14	N/A	N/A	(-) One-step mutation	Paternal
	Female child: 12/14	Female child: 12/14	$[GGAA]_{12}$		
			$[GGAA]_{14}$		
	Father: 13/15	Father: 13/15	$[GGAA]_{13}$		
			$[GGAA]_{15}$		

4.4.11 The D13S317 locus

The D13S317 locus contains a simple [TATC] tetranucleotide repeat. Alleles at this locus range between 5 to 16 repeat units (Gettings *et al*., 2015). The mutation rate at this locus is reported to be 0.14% (Butler, 2015). A total of four cases with apparent mismatches at the D13S317 locus were examined. A 100% concordance was observed between the CE-based genotypic data and sequence-based alleles. **Table 4.13** summarises the obtained results.

In the first case, a mother, a male child and a father were analysed. The genotypes obtained by CE were a homozygous allele 11/11 for the mother, heterozygous alleles 11/12 for the male child and heterozygous 8/13 for the father. Upon sequencing, it was determined that the child inherited allele 13 from the father with a loss of a [TATC] repeat, resulting in allele 12. Therefore, a (-) one-step mutation was presumed to be the most likely mutation model.

In the second case, a female child and a father were studied. The alleles obtained by CE were reported as heterozygous 12/13 and 11/12, respectively. The allelic mismatch was identified between the child's allele 13 and the father's allele 12, assuming the child inherited allele 12 from the mother. The mismatch was then characterised as a (+) one-step mutation due to DNA polymerase slippage. According to the D13S317 reference sequences (G09017/ AL353628.2) (available at https://strbase.nist.gov/), there are two [AATC] repeats located directly after the TATC repeat motif. An adenine to thymine (A>T) SNV located directly after the [TATC] repeat motif was identified in both the heterozygous alleles of the child and father in case 2 **(Figure 4.16)**. The identified SNV resulted in an additional [TATC] repeat unit and only one [AATC] repeat. Therefore, alleles determined by sequencing appeared to have an additional repeat leading to alleles 13/14 for the child and 12/13 for the father. The SNV has previously been characterised as rs9546005 on the database for single nucleotide polymorphisms (dbSNP). Moreover, the observed sequence variation did not impact the length-based genotypic data previously obtained, as length-based methods would not count these additional repeats. This is because there is no difference in the generated PCR amplicon. Youn *et al*. (2022) identified the same SNV in a paternity case where they investigated a successive single-step mutation at the

D13S317 locus across three generations of a Korean family. The detected rs9546005 SNV causes discrepancy between CE-based genotypic data and sequence-based genotypes, as the sequence-based alleles will appear to have an additional repeat unit. Therefore, it is important to review the literature for nomenclature in such instances.

The third case tested at the D13S317 locus consisted of a female child and a father. Their genotypic data was previously reported as a heterozygous 13/14 for the child and a homozygous 12/12 for the father. The child presumably inherited allele 14 from the mother with heterozygous alleles 12/14. The sequence data revealed that the child and father harboured the same rs9546005 SNV identified in D13S317 case 2. According to the sequencing, it was noted that the child's genotype was 14/15 and 13/13 for the father. Similar to case 2, this observation does not change the previously reported length-based genotypic data, thus the mutation model was presumed to be most likely a (+) one-step mutation from the paternal lineage.

The last case investigated involved a male child and a father with alleles 12/12 and 11/13, respectively. The sequencing data confirmed this genotypic data. Without knowing the mother's alleles, a (+/-) one-step mutation from the paternal lineage was presumed the most likely mutation model.

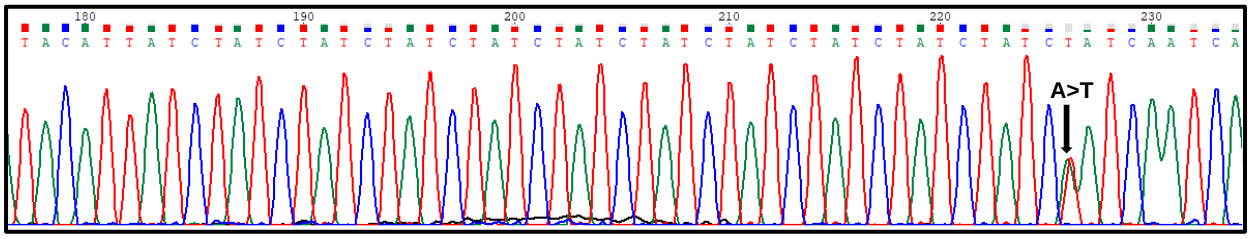


Figure 4.16: A sequence electropherogram depicting the A>T SNV detected at the D13S317 STR locus.

Table 4.13: Results obtained for the D13S317 locus parentage cases. The table indicates the genotype previously obtained by CE, genotypes obtained by sequencing, sequence structure, most likely presumed mutation model and presumed mutation origin.

Study case no.	Genotype previously obtained by CE	Genotype obtained by sequencing	Sequence structure	Most likely presumed mutation model	Presumed mutation origin	Other variants
D13S317 Case 1	Mother: 11	N/A	N/A	(-) One-step mutation	Paternal	N/A
	Male child: 11/12	Male child: 11/12	$[TATC]_{11}$ $[AATC]_2$ $[ATCT]_3$			None
			$[TATC]_{12}$ $[AATC]_2$ $[ATCT]_3$			
	Father: 8/13	Father: 8/13	$[TATC]_8$ $[AATC]_2$ $[ATCT]_3$			None
			$[TATC]_{13}$ $[AATC]_2$ $[ATCT]_3$			
D13S317 Case 2	Mother:10/12	N/A	N/A	(+) One-step mutation	Paternal	N/A
	Female child:12/13	Female child:13/14	$[TATC]_{13}$ $[AATC]$ $[ATCT]_3$			A>T
			$[TATC]_{14}$ $[AATC]$ $[ATCT]_3$			
	Father: 11/12	Father: 12/13	$[TATC]_{12}$ $[AATC]$ $[ATCT]_3$			A>T
			$[TATC]_{13}$ $[AATC]$ $[ATCT]_3$			
D13S317 Case 3	Mother: 12/14	N/A	N/A	(+) One-step mutation	Paternal	N/A
	Female child: 13/14	Female child:14/15	$[TATC]_{14}$ $[AATC]$ $[ATCT]_3$			A>T
			$[TATC]_{15}$ $[AATC]$ $[ATCT]_3$			
	Father: 12/12	Father:13/13	$[TATC]_{13}$ $[AATC]$ $[ATCT]_3$			A>T
D13S317 Case 4	Male child: 12/12	Male child:12/12	$[TATC]_{12}$ $[AATC]_2$ $[ATCT]_3$	(+/-) One-step mutation	Paternal	None
	Father: 11/13	Father: 11/13	$[TATC]_{11}$ $[AATC]_2$ $[ATCT]_3$			None
			$[TATC]_{13}$ $[AATC]_2$ $[ATCT]_3$			

4.4.12 The D16S539 locus

The D16S539 locus has a simple [GATA] tetranucleotide repeat. The allele sizes at this locus range from 5 to 15 repeats (Gettings *et al*., 2015). The mutation rate at this locus is estimated at 0,11% (Butler, 2015). The previously reported CE-based genotypic data was 100% concordant to the sequence-based genotypes. We investigated 6 cases with instances of mismatches at this locus. **Table 4.14** presents an overview of the findings.

The first parentage case studied was a paternal duo. The previously reported alleles were heterozygous 9/14 for the male child and 10/13 for the father. The obtained sequencing data supported the previously reported alleles. It was proposed that the most likely mutation model was a (-/+) one-step mutation from the paternal lineage. The two possible mechanisms were either that there was a loss of [GATA] repeat leading to the father's allele 10 being passed as allele 9 to the child. Furthermore, a gain of one repeat could have occurred in allele 13 of the father resulting in allele 14 in the child.

In the second case, a female child and father were studied. Sequencing data confirmed the allele previously obtained by CE as heterozygous 12/13 for the child and 10/12 for the father. The mother's alleles were reported as 10/12. To determine the alleles involved in the mismatch, two assumptions were made. Firstly, if the child inherited allele 12 from the mother, a (+) one-step mutation model was presumed; the paternal allele 12 being passed as allele 13 in the child. The same mutation model would be supported if the child inherited the father's allele 12. However, it would result from the maternal allele 12 being passed as allele 13 in the child.

The third case investigated in this locus involved a male child and a mother. The previously assigned alleles were homozygous 9/9 and heterozygous 10/11, respectively. The sequencing data confirmed these alleles. The mutation model responsible for the mismatch was presumed to be most likely a (+) one-step mutation. The sequence analysis revealed a thymine to adenine (T>A) substitution 38 bp downstream of the last [GATA] repeat motif in both the child and the mother. The T>A SNV did not affect the genotypes previously obtained by CE. This SNV was characterised as rs148505465, with a 0.7% frequency in the 1000 Genomes Phase 3

dataset (Gettings *et al*., 2015). It has previously been described to impacted reverse primers of CE-based kits, but redesigned primer sets have been used to overcome this issue (Krenke *et al*., 2002; Wallin *et al*., 2002). The child in case 3 also had and adenine to cytosine (A>C) SNV located 16 bp downstream of the D16S539 repeat sequence previously described as rs11642858 (Gettings *et al*., 2015) with a 26% frequency. The observed SNV also had no impact on the genotypic data previously described.

D16S539 case 4 consisted of a paternal duo. The STR genotypes at this locus were as follows: the male child (11/12) and the father (13/13). The sequencing data verified these genotypes and, therefore, the mutation model that caused the mismatch was assumed to be a (+) one-step mutation. D16S539 case 5 was a standard trio, the alleles determined by CE were heterozygous 9/10 for the mother, heterozygous 10/12 for the child and homozygous 11 for the father. The sequencing findings supported the alleles obtained by CE. Given that the child inherited allele 10 from the mother, the allelic mismatch was identified to be between the child's allele 12 and the father's allele 11. The mismatch was due to a gain of a [GATA] repeat. Furthermore, a (+) one-step mutation model presumed and was attributed to DNA polymerase slippage.

The last case at this locus was a standard trio, with the alleles determined by CE being heterozygous 11/14 for the mother, 12/13 for the child, and 9/12 for the father. Sequencing results matched the alleles detected by CE. The allelic mismatch was between the child's allele 13 and the mother's allele 14, provided that the child inherited allele 12 from the father. It was determined that the mismatch was caused by a loss of one [GATA] repeat unit owing to DNA polymerase slippage. Additionally, the mismatch was characterised as a (-) one-step mutation model from the maternal lineage.

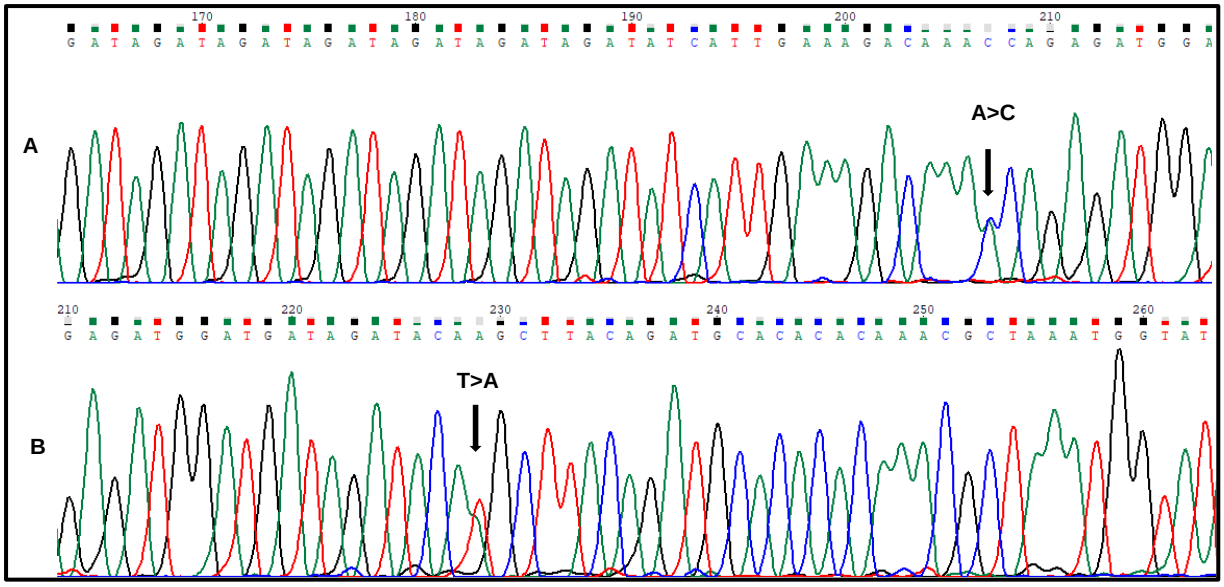


Figure 4.17: A sequence electropherogram depicting the SNVs detected at the D16S539 STR locus. a) shows the A>C SNV and b) shows the T>A SNV.

Table 4.14: Results obtained for the D16S539 locus parentage cases. The table indicates the genotype previously obtained by CE, genotypes obtained by sequencing, sequence structure, most likely presumed mutation model and presumed mutation origin.

Study case no.	Genotype previously obtained by CE	Genotype obtained by sequencing	Sequence structure	Most likely presumed mutation model	Presumed mutation origin	Other variants
D16S539 Case 1	Male child: 9/14	Male child: 9/14	$[GATA]_9$	(+/-) One-step mutation	Paternal	None
			$[GATA]_{14}$			
	Father:10/13	Father:10/13	$[GATA]_{10}$			
			$[GATA]1_3$			
D16S539 Case 2	Mother: 10/12	N/A	N/A	(+) One-step mutation	Paternal/ maternal	None
	Female child: 12/13	Female child: 12/13	$[GATA]_{12}$			
			$[GATA]_{13}$			
	Father: 10/12	Father: 10/12	$[GATA]_{10}$			
			$[GATA]_{12}$			
D16S539 Case 3	Male child: 9/9	Male child: 9/9	$[GATA]_9$	(-) One-step mutation	Maternal	A>C; T>A
	Mother: 10/11	Mother: 10/11	$[GATA]_{10}$			A>C
			$[GATA]_{11}$			

Table 4.14 continued

Study case no.	Genotype previously obtained by CE	Genotype obtained by sequencing	Sequence structure	Most likely presumed mutation model	Presumed mutation origin	Other variants
D16S539 Case 4	Male child: 11/12	Male child: 11/12	$[GATA]_{11}$	(-) One-step mutation	Paternal	None
			$[GATA]_{12}$			
	Father: 13/13	Father: 13/13	$[GATA]_{13}$			None
D16S539 Case 5	Mother: 9/10	Mother: 9/10	N/A	(+) One-step mutation	Paternal	None
	Male child: 10/12	Male child: 10/12	$[GATA]_{10}$			T>A
			$[GATA]_{12}$			
	Father: 11/11	Father: 11/11	$[GATA]_{11}$			None
D16S539 Case 6	Mother: 11/14	Mother: 11/14	N/A	(+) One-step mutation	Maternal	N/A
	Male child:12/13	Male child:12/13	$[GATA]_{12}$			None
			$[GATA]_{13}$			
	Father: 9/12	Father: 9/12	$[GATA]_{9}$			None
			$[GATA]_{12}$			

4.4.13 The D18S51 locus

The D18S51 locus contains a simple tetranucleotide [AGAA] repeat. The allele size range is 7 to 27 repeats. More than 70 alleles have been reported for this locus, making it one of the most polymorphic loci commonly used (Gettings *et al*., 2015). Its mutation rate is reported to be 0.22% (Butler, 2015). The present study evaluated 12 parentage cases with parent-child allelic mismatches at this locus. The was 100% concordance between the previously reported CE-based genotypic data and the obtained sequence-based data. The findings a reported in **Table 4.15**.

Three mutation models attributed to DNA polymerase slippage were identified in this locus. In five (cases 1, 4, 8, 9, and 10) of the 12 cases, we characterised the mutation models as (-) one-step mutations. The mutations in cases 1, 4, 8, and 10 were determined to be from the paternal lineage, and only case 9 was maternally transmitted. Furthermore, in six cases (2, 3, 5, 6, 11, and 12), the mutation model was presumed to be most likely a (+) one-step mutation. There was no significant difference in the (+) one-step mutation to (-) one-step mutations ratio (1.2:1.0). This finding was similar to that of Zhao *et al*. (2015), where the authors reported that gains in repeat units are more frequent than losses at the D18S51 locus. Moreover, mismatches in cases 3, 6, 11, and 12 were from the paternal origin. Only cases 2 and 5 had maternally transmitted mismatches. Case 7 of this locus was a paternal duo, where the CE-based genotypic data was reported as homozygous 18/18 for the male child and heterozygous 17/19 for the father. The alleles involved in the mismatch could not be determined; therefore, a (+/-) one-step mutation model was assumed. This suggests that there was a gain or loss of one repeat unit [AGAA] from the father to the child. In this locus there were 9 paternal and 3 maternal parent of origin mutations identified. this observation was consistent with the literature (Zhao *et al*., 2015; Hamester *et al*., 2019)

Table 4.15: Results obtained for the D18S51 locus parentage cases. The table indicates the genotype previously obtained by CE, genotypes obtained by sequencing, sequence structure, most likely presumed mutation model and presumed mutation origin.

Study case no.	Genotype previously obtained by CE	Genotype obtained by sequencing	Sequence structure	Most likely presumed mutation model	Presumed mutation origin
D18S51 Case 1	Mother: 18/20	N/A	N/A	(-) One-step mutation	Paternal
	Female child: 18/19	Female child: 18/19	$[AGAA]_{18}$		
			$[AGAA]_{19}$		
	Father: 18/20	Father: 18/20	$[AGAA]_{18}$		
			$[AGAA]_{20}$		
D18S51 Case 2	Female child: 14/21.2	Female child: 14/21.2	$[AGAA]_{14}$	(+) one-step mutation	Maternal
			$[AGAA]_{21}AG$		
	Mother: 18/20.2	Mother: 18/20.2	$[AGAA]_{18}$		
			$[AGAA]_{20}AG$		
D18S51 Case 3	Mother: 18/21.2	N/A	N/A	(+) one-step mutation	Paternal
	Male child: 21.2/23	Male child: 21.2/23	$[AGAA]_{21}AG$		
			$[AGAA]_{23}$		
	Father: 18/22	Father: 18/22	$[AGAA]_{18}$		
			$[AGAA]_{22}$		
D18S51 Case 4	Male child: 12/20	Male child: 12/20	$[AGAA]_{12}$	(-) One-step mutation	Paternal
			$[AGAA]_{20}$		
	Father: 15/21	Father: 15/21	$[AGAA]_{15}$		
			$[AGAA]_{21}$		

Table 4.15 continued

Study case no.	Genotype previously obtained by CE	Genotype obtained by sequencing	Sequence structure	Most likely presumed mutation model	Presumed mutation origin
D18S51 Case 5	Mother: 16/18	Mother: 16/18	$[AGAA]_{16}$	(+) One-step mutation	Maternal
			$[AGAA]_{18}$		
	Male child:16/19	Male child:16/19	$[AGAA]_{16}$		
			$[AGAA]_{19}$		
	Father: 16/16	N/A	N/A		
D18S51 Case 6	Mother: 15/19	N/A	N/A	(+) One-step mutation	Paternal
	Male child: 19/22	Male child: 19/22	$[AGAA]_{19}$		
			$[AGAA]_{22}$		
	Father: 14/21	Father: 14/21	$[AGAA]_{14}$		
			$[AGAA]_{21}$		
D18S51 Case 7	Male child: 18/18	Male child: 18/18	$[AGAA]_{18}$	(+/-) One-step mutation	paternal
			[AGAA]		
	Father: 17/19	Father: 17/19	$[AGAA]_{17}$		
			$[AGAA]_{19}$		
D18S51 Case 8	Mother: 15/16	N/A	N/A	(-) One-step mutation	Paternal
	Female child: 15/18	Female child: 15/18	$[AGAA]_{15}$		
			$[AGAA]_{18}$		
	Father: 16/19	Father: 16/19	$[AGAA]_{16}$		
			$[AGAA]_{19}$		

Table 4.15 continued

Study case no.	Genotype previously obtained by CE	Genotype obtained by sequencing	Sequence structure	Most likely presumed mutation model	Presumed mutation origin
D18S51 Case 9	Mother: 18/18	Mother: 18/18	$[AGAA]_{18}$	(-) One-step mutation	Maternal
	Female child: 14/19	Female child: 14/19	$[AGAA]_{14}$		
			$[AGAA]_{19}$		
	Father: 14/15	N/A	N/A		
D18S51 Case 10	Mother: 18/20	N/A	N/A	(-) One-step mutation	Paternal
	Male child: 14/20	Male child: 14/20	$[AGAA]_{14}$		
			$[AGAA]_{20}$		
	Father: 15/15.2	Father: 15/15.2	$[AGAA]_{15}$		
			$[AGAA]_{15}AG$		
D18S51 Case 11	Mother: 12/14	N/A	N/A	(+) One-step mutation	Paternal
	Male child: 14/22	Male child: 14/22	$[AGAA]_{14}$		
			$[AGAA]_{22}$		
	Father: 20/21	Father: 20/21	$[AGAA]_{20}$		
			$[AGAA]_{21}$		
D18S51 Case 12	Mother: 13/16	N/A	N/A	(+) One-step mutation	Paternal
	Female child: 13/21	Female child: 13/21	$[AGAA]_{13}$		
			$[AGAA]_{21}$		
	Father: 19/20	Father: 19/20	$[AGAA]_{19}$		
			$[AGAA]_{20}$		

4.4.14 The D19S433 locus

The D19S433 locus contains a compound tetranucleotide repeat located on the long arm of chromosome 19 It has a [AAGG] [AAAG] [AAGG] [TAGG] $[AAGG]_n$ repeat. However, only the [AAGG] repeats are counted towards the allele designation. The allele size ranges from 6 to 19.2 (Gettings *et al*., 2015). A mutation rate of 0.11% has been reported (Butler, 2015). In this study, 14 parentage cases containing a parent-child allelic mismatch at the D19S433 locus were analysed. There was no concordance between the obtained sequence-based genotypic data and the previously reported CE-based genotypic data. The results obtained are summarised in **Table 4.16**.

Notably, all the samples harboured a guanine to thymine (G>T) SNV 21 bp downstream from the D19S433 repeat region **(Figures 4.18 and 4.19)**. The identified variant led to an allelic dropout of allele 8 with the following repeat motif: [AAGG] [AAAG] [AAGG] [TAGG] $[AAGG]_6$. Furthermore, this allele was not detected when genotyping was performed using Capillary Electrophoresis (CE). **Figure 4.20** depicts an electropherogram of case 9 as an example where allele 8 dropped out.

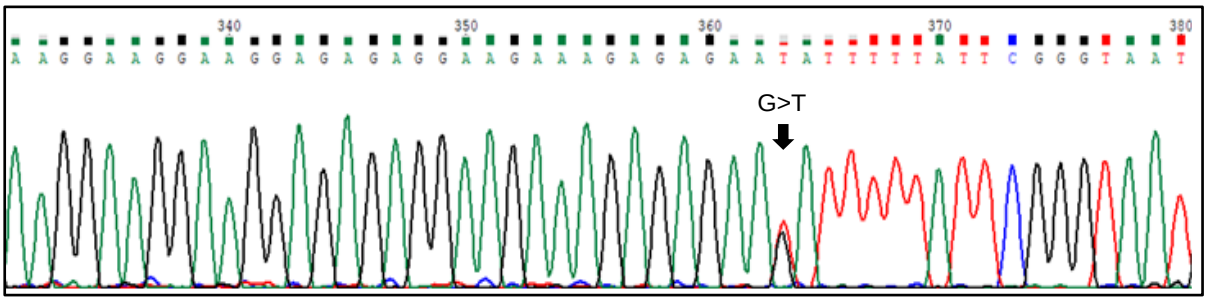


Figure 4.18: An electropherogram depicting the guanine to thymine (G>T) point variant

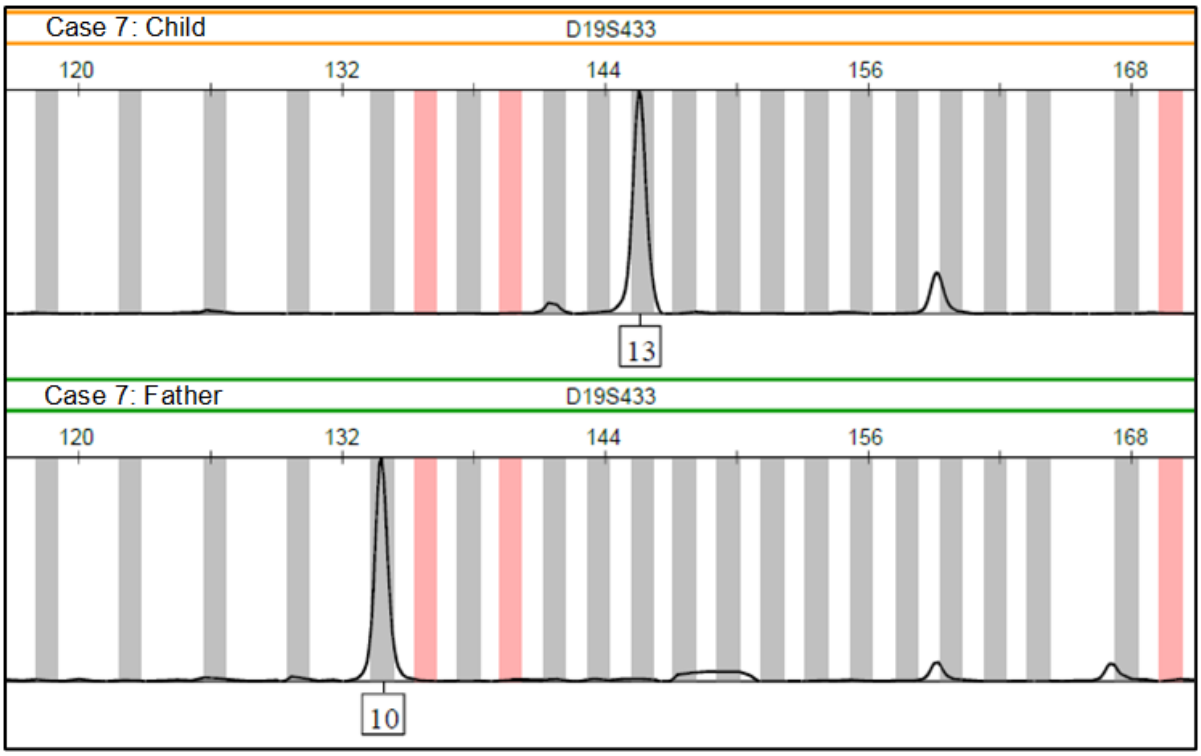


Figure 4.19: An electropherogram depicting Case 7 as an example where allele 8 was dropped out.

The possible underlying cause of the null allele at the D19S433 was determined to be due to a primer-binding site mutation (G>T) at the binding site of the reverse primer used in the VeriFiler™ Express PCR amplification kit. However, this cannot be said with absolute certainty as the kit's manufacturer does not share the D19S433 primer sequences. Nevertheless, it can be acknowledged that genotyping this locus using the VeriFiler™ Express PCR amplification kit produced apparent opposite homozygous genotypes for all the investigated parentage cases. When both the parent and child harbour a single allele at one STR locus with a discrepant allele transfer, null alleles should be considered (Sun *et al*., 2014).

An identical SNV was reported in a study by Pitterl *et al*. (2010). The authors evaluated the discrimination power of forensic STRs in indigenous South African and Central Asian populations. However, the authors reported the SNV to cause the dropout of allele 6. Nevertheless, allele 6 corresponds to allele 8 when using updated nomenclature on the STRBase website. Furthermore, they identified the variant in 13

individuals within the African sample set (N=108). As per records, 99% of the individuals tested at the NHLS DNA testing facility are of African ancestry. Considering these findings, we can hypothesise that the identified G>T SNV is specific to the population of African descent. However, an in-depth study with a larger and diverse population group should be conducted to confirm this assumption.

Notably, the allelic dropout reported by Pitterl *et al*. (2010) was observed when using the Identifiler™ PCR amplification kit. Similar to the VeriFiler™ Express PCR amplification kit, the primer sequences are not published. Both kits are manufactured by Applied Biosystems (Foster City, CA), It is therefore, important to notify the manufacturer of these findings, as allele dropouts have a major impact on CE-based genotyping during parentage analysis.

In case 9, a male child and mother were tested. The D19S433 genotype previously obtained by CE was homozygous alleles 8 and 13, respectively. Interestingly, when using the designed sequencing primers, the genotypes were obtained as a homozygous allele 8 for the child and heterozygous 8/13 for the mother. In both individuals, the G>T variant was identified. At first, it appears that this variant did not cause the dropout of allele 8 in the child, as seen in the other cases in this study. The CE-based genotypic data was re-evaluated for any discrepancies **(Figure 4.20)**. The peak height of the child's allele 8 is shorter in height compared to the that of the mother's allele 13 and does not represent a homozygous allele. Moreover, for homozygous alleles, the peak height is expected to be double in size (Bright *et al*., 2014). Therefore, it can be concluded that the identified G>T point variant led to the dropout of the second allele 8. Although the father was not tested in this case, we can assume that the allele 8 detected by CE was of paternal origin.

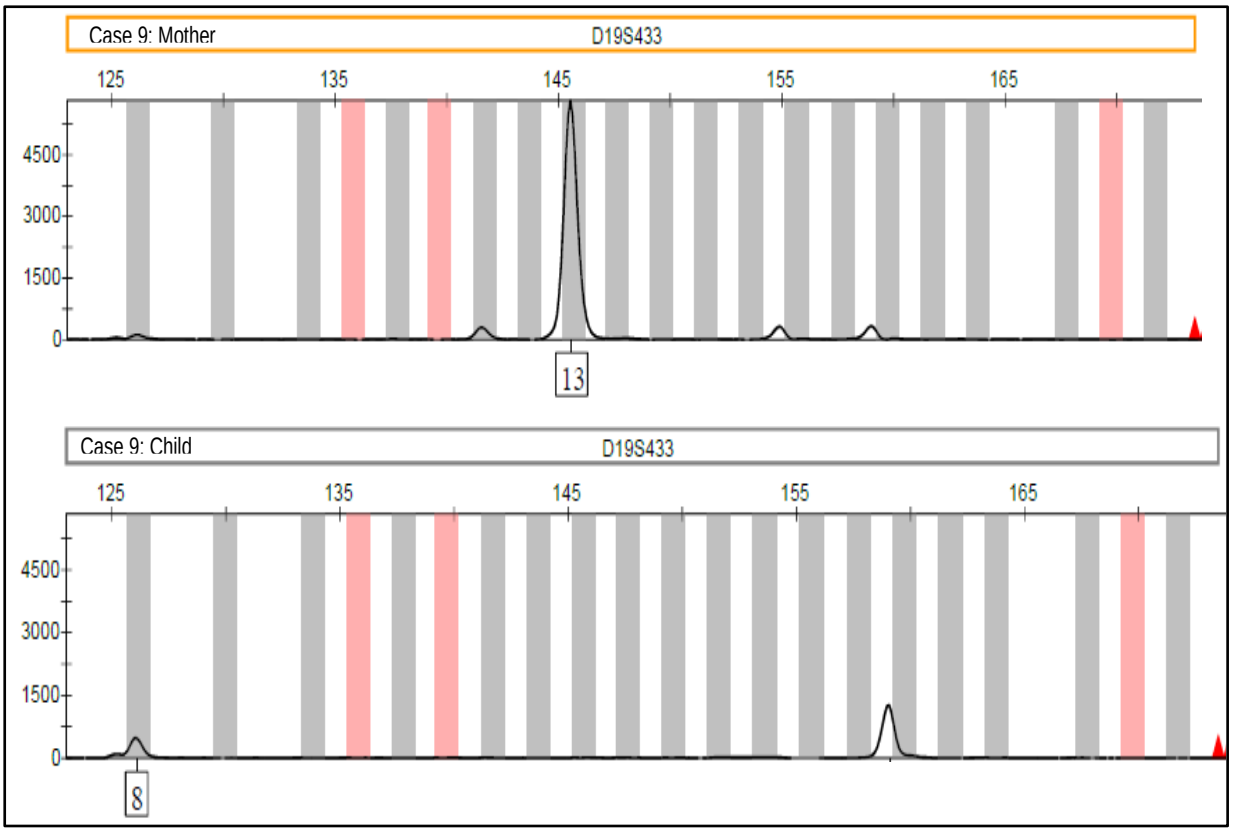


Figure 4.20: An electropherogram showing the results of Case 9 with the D19S433 genotype previously obtained by CE.

In the D19S433 locus, mutations of paternal origin were more common than maternal mutations, at a ratio of ~3:1 between males and females. This observation is supported by literature, as it is widely accepted that STR loci mutation rates are higher in the male germline than in the female germline, because the number of germ-cell divisions per generation is much larger in males than in females (Shimmin *et al*., 1993; Zhao *et al*., 2015). However, owing to the small number of cases studied, an association could not be made in this study. Thus, more parentage cases must be analysed to draw this conclusion confidently.

Table 4.16: Results obtained for the D19S433 locus parentage cases. The table indicates the genotype previously obtained by CE, genotypes obtained by sequencing, sequence structure, most likely presumed mutation model and presumed mutation origin.

Study case no.	Genotype previously obtained by CE	Genotype obtained by sequencing	Sequence structure	Most likely presumed mutation model	Presumed mutation origin	Other variants
D19S433 Case 1	Mother: 12/14.2	N/A	N/A	Primer-binding site mutation	Paternal	N/A
	Male child: 14.2/14.2	Male child: 8/14.2	[AAGG] [AAAG] [AAGG] [TAGG] [AAGG]$_6$			G>T
			[AAGG] [AAAG] [AAGG] [TAGG] [AAGG]			
	Father: 12/12	Father: 8/12	[AAGG] [AAAG] [AAGG] [TAGG] [AAGG]$_6$			G>T
			[AAGG] [AAAG] [AAGG] [TAGG] [AAGG]$_{10}$			
D19S433 Case 2	Male child: 12.2/12.2	Male child: 8/12.2	[AAGG] [AAAG] [AAGG] [TAGG] [AAGG]$_6$	Primer-binding site mutation	Paternal	G>T
			[AAGG] [AAAG] [AAGG] [TAGG] [AAGG]			
	Father: 13/14.2	Father: 8/13	[AAGG] [AAAG] [AAGG] [TAGG] [AAGG]$_6$			G>T
			[AAGG] [AAAG] [AAGG] [TAGG] [AAGG] $_{11}$			
D19S433 Case 3	Male child: 12/12	Male child: 8/12	[AAGG] [AAAG] [AAGG] [TAGG] [AAGG]$_6$	Primer-binding site mutation	Maternal	G>T
			[AAGG] [AAAG] [AAGG] [TAGG] [AAGG]$_{10}$			
	Mother: 15/15	Mother: 8/15	[AAGG] [AAAG] [AAGG] [TAGG] [AAGG]$_6$			G>T
			[AAGG] [AAAG] [AAGG] [TAGG] [AAGG]$_{13}$			
D19S433 Case 4	Mother: 13/14	N/A	N/A	Primer-binding site mutation	Paternal	N/A
	Male child: 13/13	Male child: 8/13	[AAGG] [AAAG] [AAGG] [TAGG] [AAGG]$_6$			G>T
			[AAGG] [AAAG] [AAGG] [TAGG] [AAGG]$_{11}$			
	Father: 14.2/14.2	Father: 8/14.2	[AAGG] [AAAG] [AAGG] [TAGG] [AAGG]$_6$			G>T
			[AAGG] [AAAG] [AAGG] [TAGG] [AAGG]			

Table 4.16 continued

Study case no.	Genotype previously obtained by CE	Genotype obtained by sequencing	Sequence structure	Most likely presumed mutation model	Presumed mutation origin	Other variants
D19S433 Case 5	Mother: 13/13	N/A	N/A	Primer-binding site mutation	Paternal	N/A
	Female child: 13/13	Female child: 8/13	[AAGG] [AAAG] [AAGG] [TAGG] $[AAGG]_6$			G>T
			[AAGG] [AAAG] [AAGG] [TAGG] $[AAGG]_{11}$			
	Father: 14/14	Father: 8/14	[AAGG] [AAAG] [AAGG] [TAGG] $[AAGG]_6$			G>T
			[AAGG] [AAAG] [AAGG] [TAGG] $[AAGG]_{12}$			
D19S433 Case 6	Male child: 13/13	Male child: 8/13	[AAGG] [AAAG] [AAGG] [TAGG] $[AAGG]_6$	Primer-binding site mutation	Paternal	G>T
			[AAGG] [AAAG] [AAGG] [TAGG] $[AAGG]_{11}$			
	Father: 12/12	Father: 8/12	[AAGG] [AAAG] [AAGG] [TAGG] $[AAGG]_6$			G>T
			[AAGG] [AAAG] [AAGG] [TAGG] $[AAGG]_{10}$			
D19S433 Case 7	Male child: 13/13	Male child: 8/13	[AAGG] [AAAG] [AAGG] [TAGG] $[AAGG]_6$	Primer-binding site mutation	Paternal	G>T
			[AAGG] [AAAG] [AAGG] [TAGG] $[AAGG]_{11}$			
	Father: 10/10	Father: 8/10	[AAGG] [AAAG] [AAGG] [TAGG] $[AAGG]_6$			G>T
			[AAGG] [AAAG] [AAGG] [TAGG] $[AAGG]_8$			
D19S433 Case 8	Female child: 13/13	Female child: 8/13	[AAGG] [AAAG] [AAGG] [TAGG] $[AAGG]_6$	Primer-binding site mutation	Paternal	G>T
			[AAGG] [AAAG] [AAGG] [TAGG] $[AAGG]_{11}$			
	Father: 15/15	Father: 8/15	[AAGG] [AAAG] [AAGG] [TAGG] $[AAGG]_6$			G>T
			[AAGG] [AAAG] [AAGG] [TAGG] $[AAGG]_{13}$			

Table 4.16 continued

Study case no.	Genotype previously obtained by CE	Genotype obtained by sequencing	Sequence structure	Most likely presumed mutation model	Presumed mutation origin	Other variants
D19S433 Case 9	Male child: 8/8	Male child: 8/8	[AAGG] [AAAG] [AAGG] [TAGG] $[AAGG]_6$	Primer-binding site mutation	Maternal	G>T
			[AAGG] [AAAG] [AAGG] [TAGG] $[AAGG]_6$			
	Mother: 13/13	Mother: 8/13	[AAGG] [AAAG] [AAGG] [TAGG] $[AAGG]_6$			G>T
			[AAGG] [AAAG] [AAGG] [TAGG] $[AAGG]_{11}$			
D19S433 Case 10	Mother: 7/12	N/A	N/A	Primer-binding site mutation	Paternal	N/A
	Male child: 7/7	Male child: 7/8	[AAGG] [AAAG] [AAGG] [TAGG] $[AAGG]_5$			G>T
			[AAGG] [AAAG] [AAGG] [TAGG] $[AAGG]_6$			
	Father: 13/13	Father: 8/13	[AAGG] [AAAG] [AAGG] [TAGG] $[AAGG]_6$			G>T
			[AAGG] [AAAG] [AAGG] [TAGG] $[AAGG]_{11}$			
D19S433 Case 11	Mother: 13/14	N/A	N/A	Primer-binding site mutation	Paternal	N/A
	Male child: 13/13	Male child: 8/13	[AAGG] [AAAG] [AAGG] [TAGG] $[AAGG]_6$			G>T
			[AAGG] [AAAG] [AAGG] [TAGG] $[AAGG]_{11}$			
	Father: 7/7	Father: 7/8	[AAGG] [AAAG] [AAGG] [TAGG] $[AAGG]_5$			G>T
			[AAGG] [AAAG] [AAGG] [TAGG] $[AAGG]_6$			
D19S433 Case 12	Male child: 14/14	Male child: 8/14	[AAGG] [AAAG] [AAGG] [TAGG] $[AAGG]_6$	Primer-binding site mutation	Maternal	G>T
			[AAGG] [AAAG] [AAGG] [TAGG] $[AAGG]_{12}$			
	Mother: 13/13	Mother: 8/13	[AAGG] [AAAG] [AAGG] [TAGG] $[AAGG]_6$			G>T
			[AAGG] [AAAG] [AAGG] [TAGG] $[AAGG]_{11}$			

Table 4.16 continued

Study case no.	Genotype previously obtained by CE	Genotype obtained by sequencing	Sequence structure	Most likely presumed mutation model	Presumed mutation origin	Other variants
D19S433 Case 13	Mother: 13.2/13.2	Mother: 8/13.2	[AAGG] [AAAG] [AAGG] [TAGG] $[AAGG]_6$	Primer-binding site mutation	Maternal	G>T
			[AAGG] [AAAG] [AAGG] [TAGG] [AAGG]			
	Male child: 13/13	Male child: 8/13	[AAGG] [AAAG] [AAGG] [TAGG] $[AAGG]_6$			G>T
			[AAGG] [AAAG] [AAGG] [TAGG] $[AAGG]_{11}$			
	Father: 11/13	N/A	N/A			N/A
D19S433 Case 14	Female child: 14/14	Female child: 8/14	[AAGG] [AAAG] [AAGG] [TAGG] $[AAGG]_8$	Primer-binding site mutation	Paternal	G>T
			[AAGG] [AAAG] [AAGG] [TAGG] [AAGG]			
	Father: 13/13	Father: 8/13	[AAGG] [AAAG] [AAGG] [TAGG] $[AAGG]_6$			G>T
			[AAGG] [AAAG] [AAGG] [TAGG] $[AAGG]_{11}$			

4.4.15 The D21S11 locus

The D21S11 locus contains a {[TCTA]$_3$ TA [TCTA]$_3$ TCA [TCTA]$_2$ TCCA TA} complex tetranucleotide repeat. Alleles ranges in size from 24 to 38 (Butler and Reeder, 1997). The 11 non-repeat unit bases in this region (unbracketed) are not counted for length-based allele designation. Many of the alleles in this locus have the same length but different internal sequence structure. This is because some of the repeat units are switched around. Sequence variation in this region has shown to result in initial three [TCTA] repeats becoming two repeats. Furthermore, complete loss of the initial {[TCTA]$_3$ TA} segment can occur. Therefore, specific differences in the D21S11 allele structures can only be determined by DNA sequencing (Gettings *et al*., 2015).

Four parentage cases were analysed in this locus. A summary of the findings is depicted in **Table 4.17**. In all cases studied, the CE-based genotypic data was 100% concordant with the obtained sequence-based alleles. In the first case, a female child and a mother were tested. Their alleles were previously determined as heterozygous 32/35 and 30/31, respectively. These alleles were confirmed by the obtained sequencing data. The alleles involved in the mismatch were identified to be allele 31 of the mother and allele 32 of the child. There was a gain of one TCTA repeat, leading the mother's allele 31 being passed to the child as allele 32. The mutation model responsible for the mismatched alleles was presumed to be most likely a (+) one-step mutation model attributed to DNA polymerase slippage.

D21S11 case 2 consisted of a female child and a mother. The genotypic data obtained by CE was 29/32.2 and 28/30, accordingly. The genotypic data was supported by the sequencing results. In this case, it was assumed that the child inherited allele 32.2 from the father (29/32.2), because two-step and multi-step mutations occur less frequently compared to one-step mutations. Therefore, a (-/+) one-step mutation was presumed as the most likely mutation model responsible for the mismatch between allele 29 of the child and alleles 28 and 30 of the mother.

In D21S11 case 3, a male child and father were analysed, with previous genotypes reported as 28/33 and 31/34, respectively. these genotypes were confirmed by the sequencing data. Furthermore, the most likely mutation model was proposed to be a

(-) one-step mutation, resulting from a loss of a TCTA repeat; leading to the father's allele 34 being passed as allele 33 in the child.

The last case investigated in this locus involved an allelic mismatch between a male child and a mother. Their alleles determined by CE were heterozygous 29/36 and 27/30, respectively. With the knowledge that the child inherited allele 36 from the father (30/36), it was possible to identify the mismatch to be between allele 30 of the mother and allele 29 of the child; suggesting a (-) one-step mutation as the most likely mutation model.

Although the mutation rate at D21S11 locus is estimated at 0.19% (Butler and Reeder, 1997), no correlations could be made in this regard due to the limited number of cases studied in this locus. However, it was noteworthy that 75% (n=3/4) of the identified mutations were of maternal lineage while only 25% (n=1/4) were of the paternal lineage. Furthermore, 100% (4/4) the identified mutations followed the one-step mutation model, and they were all accounted for by DNA polymerase slippage.

Table 4.17: Results obtained for the D21S11 locus parentage cases. The table indicates the genotype previously obtained by CE, genotypes obtained by sequencing, sequence structure, most likely presumed mutation model and presumed mutation origin.

Study case no.	Genotype previously obtained by CE	Genotype obtained by sequencing	Sequence structure	Most likely presumed mutation model	Presumed mutation origin
D21S11 Case 1	Female child: 32/35	Female child: 32/35	[TCTA]5 [TCTG]6 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCA TA [TCTA]13	(+) One-step mutation	Maternal
			[TCTA]11 [TCTG]5 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCA TA [TCTA]11		
	Mother: 30/31	Mother: 30/31	[TCTA]5 [TCTG]6 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCA TA [TCTA]11		
			[TCTA]5 [TCTG]6 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCA TA [TCTA]12		
D21S11 Case 2	Mother: 28/30	Mother: 28/30	[TCTA]4 [TCTG]6 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCA TA [TCTA]10	(+/-) One-step mutation	Maternal
			[TCTA]4 [TCTG]6 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCA TA [TCTA]12		
	Female child: 29/32.2	Female child: 29/32.2	[TCTA]4 [TCTG]6 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCA TA [TCTA]11		
			[TCTA]4 [TCTG]6 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCA TA [TCTA]13 TA TCTA		
	Father: 29/32.2	N/A	N/A		

Table 4.17 continued

Study case no.	Genotype previously obtained by CE	Genotype obtained by sequencing	Sequence structure	Most likely presumed mutation model	Presumed mutation origin
D21S11 Case 3	Mother: 28/31	N/A	N/A	(-) One-step mutation	Paternal
	Male child: 28/33	Male child: 28/33	[TCTA]5 [TCTG]6 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCA TA [TCTA]9		
			[TCTA]5 [TCTG]6 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCA TA [TCTA]14		
	Father: 31/34	Father: 31/34	[TCTA]5 [TCTG]6 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCA TA [TCTA]12		
			[TCTA]5 [TCTG]6 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCA TA [TCTA]15		
D21S11 Case 4	Mother: 27/30	Mother: 27/30	[TCTA]6 [TCTG]5 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCA TA [TCTA]8	(-) One-step mutation	Maternal
			[TCTA]6 [TCTG]5 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCA TA [TCTA]11		
	Female child: 29/36	Female child: 29/36	[TCTA]6 [TCTG]5 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCA TA [TCTA]10		
			[TCTA]11 [TCTG]5 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCA TA [TCTA]12		
	Father: 30/36	Father: 30/36	N/A		

4.5 Characterisation of parent-child allelic mismatches across the 15 STR loci studied

In the present study, a total of 100 parentage cases with parent-child allelic mismatches across 15 STR loci were studied. Of the 100 cases, 95 cases had an allelic mismatch at one locus, whereas five cases had mismatches at two loci, resulting in a total of 105 parent-child allelic mismatches. **Table 4.18** below summarises the results obtained across the 15 STR loci studied. The number of mismatches examined at each locus ranged from 2 to 18, with the D6S1043 locus containing the most mismatches. For all the parentage cases included, the sequence-based allelic data was compared to the previously reported CE-based genotypic data to established concordance between the two genotypes. Out of the 15 STR loci studied, a 100% concordance was maintained in 11 loci (CSF1PO, FGA, vWA, D2S1338, D3S1358, D7S820, D8S1179, D10S1248, D13S317, D16S539, D18S51, and D21S11). There was only 50% concordance in the D5S818 and D8S1179. At the D6S1043 locus, a concordance of 28% was observed, with no concordance in the D19S433 locus. Therefore, an average of 82% concordance was determined over all the investigated loci. The observed non-concordance was determined to be as a result of either primer-binding site mutation or other detected sequence variants, which led to allelic-dropouts. Furthermore, the allelic dropouts contributed to 30% (n=31/105) of the total parent-child mismatches observed.

The parent-child allelic mismatches were further characterised by presumption of the most likely mutation model responsible for each observed mismatch. Firstly, the stepwise mutation model was used to describe mutation attributed to DNA polymerase slippage. The stepwise mutation model was 70% (n=74/105) more frequent compared to other mutation models. As expected, one-step mutation models were more common than the other stepwise models, accounting for 93% (n=69/74) of all cases. A similar observation was made in a study by Jin *et al*. (2016), where the authors performed mutational analysis of 33 autosomal STR loci and reported the one-step mutation model at a frequency of 98.3%. Furthermore, the one-step mutation gain (+) was the most frequent one-step model accounting for 45% (n=31/69) of the one-step mutations, followed by the one-step loss (-) at 36% (n=25/69) and the one-step loss/gain (+/-) at 19% (n=13/69). Moreover, two-step and multi-step mutations were

less frequent than one-step mutations, accounting for only ~7% (n=5/74) of all the stepwise mutations.

For all the characterised parent-child allelic mismatches, the mutation origin was further presumed to be either of the paternal or maternal lineage. There were 76 paternal-origin and 26 maternal-origin mutations observed in this study. The overall difference in the paternal-to-maternal mutation ratio (~3:1) was not supported by the literature, Zhao *et al*. (2015) reported paternal mutation to occur at 5-6 times higher compared to maternal mutations, while Jin *et al*. (2016) reported a 5.02:1 male to female mutation ratio. Although this study's findings are lower in comparison to the literature in this regard, they provide evidence that most of these mutations occur more frequently in males than in females. This could potentially be linked to various types and amounts of cell division in germ-cell genesis (Brinkmann *et al*., 1998; Mardini *et al*., 2013; Zhang *et al*., 2018).

Table 4.18: An overview of allelic mismatches characterised across the 15 STR loci studied. The table indicates the number of mismatches, CE-based and Sequence-based genotype concordance (%), presumed mutation model, and mutation origin.

STR Loci	Number of mismatches	CE-based and Sequence-based genotype concordance (%)	Presumed mutation model								Presumed mutation origin		
			One-step mutation			Two-step mutation		Multi-step mutation		Primer-binding site/ allelic dropout			
			-1	+1	-/+1	-2	+2	-n	+n		Paternal	Maternal	Paternal/ maternal
CSF1PO	4	100	0	2	2	0	0	0	0	0	3	1	0
FGA	10	100	2	3	5	0	0	0	0	0	9	0	1
vWA	13	100	4	5	0	0	1	2	1	0	8	5	0
D2S1338	5	100	3	1	1	0	0	0	0	0	4	1	0
D3S1358	2	100	1	1	0	0	0	0	0	0	1	0	1
D5S818	2	50	0	1	0	0	0	0	0	1	2	0	0
D6S1043	18	28	1	3	0	0	0	1	0	13	12	6	0
D7S820	2	100	0	2	0	0	0	0	0	0	2	0	0
D8S1179	6	50	1	1	1	0	0	0	0	3	5	1	0
D10S1248	3	100	3	0	0	0	0	0	0	0	3	0	0
D13S317	4	100	1	2	1	0	0	0	0	0	4	0	0
D16S539	6	100	2	3	1	0	0	0	0	0	3	2	1
D18S51	12	100	5	6	1	0	0	0	0	0	9	3	0
D19S433	14	0	0	0	0	0	0	0	0	14	10	4	0
D21S11	4	100	2	1	1	0	0	0	0	0	1	3	0
Total	**105**	**Avg= 82%**	**25**	**31**	**13**	**0**	**1**	**3**	**1**	**31**	**76**	**26**	**3**
			105								**105**		

CHAPTER 5: CONCLUSION

5.1 Conclusion

DNA analysis of Short tandem repeats (STRs) is a reliable tool for determining parentage (biological maternity and paternity) and kinships (Dakin and Avise, 2004). Commercial multiplex polymerase chain reaction (PCR) STR kits are widely available to DNA testing facilities. The PCR-based STR typing method has become a common approach in parentage analysis because of characteristics such as specificity, sensitivity, the capacity to amplify several loci concurrently, and automation capabilities. However, discrepancies in STR genotypes have previously been reported due to differences in primer design across commercial kit manufacturers (Westen *et al*., 2014). Furthermore, one or two STR loci mismatches between a parent and a child may be observed in non-exclusion cases. In routine analysis, these are known as “apparent STR loci mismatches” resulting from STR mutation occurring via several mutational mechanisms. Numerous apparent STR mismatches have been reported (Negi *et al*., 2006; Ali *et al*., 2009). Therefore, one or two STR loci mismatches detected in parentage analysis should not be dismissed.

The present study included 100 paternity cases with one or two STR loci mismatches from the DNA testing facility, Universitas Academic Unit, National Health Laboratory Services (NHLS) Bloemfontein. The samples were examined across 15 autosomal STR loci. Published primers and designed study primers were used to optimise the PCR conditions to amplify the targeted STR loci using commercially available control DNA. The optimised PCR conditions were utilised to screen the samples across the 15 STR loci. Sanger sequence analysis was conducted to characterise the underlying cause of the observed parent-child allelic mismatches. Additionally, the sequence data was compared to the previously determined CE-based genotypic data to evaluate concordance.

All allelic mismatches between parent(s) and child(ren) were successfully described. The findings indicated that most parent-child allelic mismatches were caused by DNA polymerase slippage, expressed using the stepwise mutation model. The stepwise

mutation model had a 70% higher frequency than other mutation models, with one-step mutation models being more prevalent (93%) than the other stepwise models. Allelic dropouts occurred at a relatively high prevalence of 30%. This mutation model was predominantly identified to affect the D19S433 (n=14/14), D6S1043(n=13/18), D8S1179 (n=3/6) and D5S818 (n=1/2).

A total of 13 Single Nucleotide Variants (SNVs) and one sequence deletion (delAATA) was detected across the 15 STR loci. These were located in the STR flanking region, which could possibly be primer-binding sites for primers included in the commercial STR kit used in this study. Sequence variants at the STR loci primer-binding sites are known to occur randomly and thus cannot be avoided completely. However, the high number sequence variation obtained in this study demonstrates that sequencing the core STR repeat and surrounding areas can provide useful information for determining STR locus mutational events. Moreover, this study emphasises the need to characterise variants at primer-binding sites of STRs within multiplex typing systems used in parentage analysis. It is important for DNA testing laboratories and researchers to take note of SNVs detected at the STR flanking region as these could potentially impact some of the PCR primers used in commercially available kits. Furthermore, this may aid in developing more efficient primer sets for STR typing. Moreover, it would encourage manufacturers to provide degenerate primers to the commercially available STR kits in order to recover the alleles that are dropped out due to primer-binding site variants.

STR mutations originating from the paternal lineage were more frequent than in females, with a paternal-to-maternal mutation ratio of (~3:1). These findings were approximately 2 times lower compared to the literature (Zhao *et al*.,2015; Jin *et al*., 2016), possibly because this study's sample size was small compared to the studies in the literature. However, this finding demonstrates that STR mutations are more common in males than in females.

In routine parentage analysis, the possibility of STR mutations should be considered when STR allelic mismatches at one or two loci are observed between a child and a parent. All the studied mismatches were attributed to STR mutation events, while no

methodology-related event was characterised. The results of this study reiterates the need to confirm STR genotypic data obtained using commercially available kits when ambiguous or inconclusive parentage or kinship results are obtained. Furthermore, the study emphasises the need for DNA testing facilities to have a method in place to characterise and confirm genotypic data produced using commercial STR kits.

In our setting, we recommend that when ambiguous or inconclusive parentage/kinship results are obtained in future, the optimised sequence analysis conditions, as determined in this study, be used to confirm the genotypic data obtained during routine STR analysis. Furthermore, this will allow for the implementation of a Sanger sequence method to investigate STR loci for which genotypic data is unclear or ambiguous. Additionally, it will allow for the resolution of inconclusive paternity and kinship investigations.

5.2 Study limitations and recommendations

The number of parentage cases studied per locus was a limiting factor of this study. The cases per locus ranged from 2 to 18. This expected due to the current STR loci respective mutation rates. The small number of cases per locus may be limiting for making conclusive observations regarding the characterisation of mutations. Moreover, the findings at those loci may not be an accurate reflection of the mutations generally affecting these loci.

The unavailability of CE-based genotypic data of the untested parent in parentage duos (mother-child and father-child) posed a limiting factor, as in some cases, the alleles involved in the parent-child allelic mismatch could not be determined, and characterisation of mutations in such cases was based solely on assumptions.

The use of parentage duos and trios with similar genotypes due to shared common alleles in a population can provide non-informative genotypes for mutation characterisation. (Example FGA case 1: mother: 21/23, child: 21/22, and father: 21/27). In such cases, the alleles implicated in the mismatch cannot be established

confidently; thus, no definite conclusions can be drawn on the mutation model responsible for the mismatch and the mutation origin.

We recommend that future studies consist of larger study cohort, with even distribution of parentage cases across the studied loci as this would allow for conclusive deductions. It would also allow for the comparison of characterised mutation models between the studied STR loci. We further suggest that future studies use standard parentage trios with informative genotypic data. This would allow for the alleles involved in the mismatch to be defined with certainty, and firm inferences can be reached about the mutation model responsible for the mismatch and the mutation origin.

Failure of the proposed method for separating heterozygous alleles posed a challenge in the study. The major challenge was experienced during the analysis of the sequencing data for allele designation. We suggest that a different method of heterozygous allele separation be explored in future studies. For similar studies, we recommend that technologies such as Next Generation Sequencing (NGS) be used should the resources be available, as this can overcome this challenge.

REFERENCES

Akhteruzzaman, S., Majumder, A.K., Ferdous, A. and Ali, M.E., 2012. False paternity with one or two mismatches using commercial STR kits. *Australian Journal of Forensic Sciences*, 44(3), pp.253-259.

Ali, M.E., Alam, S., Ferdous, A., Hasan, M., Ali, Eunus, Shafiul Alam, Ahmad Ferdous, Mahamud Hasan, T. H. and Akhteruzzaman, S. 2009. Mutation in paternally transmitted alleles at FGA microsatellite locus: A case of allele mismatch in the child, *Indian Journal of Forensic Medicine and Toxicology*, 3(2), pp. 41-43.

Allor, C., Einum, D. D., & Scarpetta, M. 2005. Identification and characterization of variant alleles at CODIS STR loci. *Journal of Forensic Sciences*, 50(5), pp.1128-1133.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J., 1990. Basic local alignment search tool. *Journal of molecular biology*, 215(3), pp.403-410.

Barber, M.D. and Parkin, B.H., 1996. Sequence analysis and allelic designation of the two short tandem repeat loci D18S51 and D8S1179. *International journal of legal medicine*, 109, pp.62-65.

Bright, J.A., Stevenson, K.E., Coble, M.D., Hill, C.R., Curran, J.M. and Buckleton, J.S., 2014. Characterising the STR locus D6S1043 and examination of its effect on stutter rates. *Forensic Science International: Genetics*, 8(1), pp.20-23.

Brinkmann, B., 1998. Overview of PCR-based systems in identity testing. *Forensic DNA Profiling Protocols*, pp.105-119.

Budowle, B., Allard, M.W., Wilson, M.R. and Chakraborty, R., 2003. Forensics and mitochondrial DNA: applications, debates, and foundations. *Annual review of genomics and human genetics*, 4(1), pp.119-141.

Butler, J.M., 2015a. The future of forensic DNA analysis. *Philosophical transactions of the royal society B: biological sciences*, 370(1674), p.20140252.

Butler, J.M., 2015b. US initiatives to strengthen forensic science & international standards in forensic DNA. *Forensic Science International: Genetics*, 18, pp.4-20.

Butler, J.M., 2015c. Introduction and issue summary: New trends in forensic genetics. *Forensic Science International: Genetics*, 18, pp.1-3.

Butler, J.M., 2017. Recent activities in the United States involving the National commission on forensic science and the organization of scientific area committees for forensic science. *Australian Journal of Forensic Sciences*, 49(5), pp.526-540.

Butler, J.M., 2006. Genetics and genomics of core short tandem repeat loci used in human identity testing. *Journal of forensic sciences*, 51(2), pp.253-265.

Butler, J.M., 2009. Fundamentals of forensic DNA typing. *Academic press*.

Butler, J.M., Buel, E., Crivellente, F. and McCord, B.R., 2004. Forensic DNA typing by capillary electrophoresis using the ABI Prism 310 and 3100 genetic analyzers for STR analysis. *Electrophoresis*, 25(10-11), pp.1397-1412.

Butler, J.M., Coble, M.D. and Vallone, P.M., 2007. STRs vs. SNPs: thoughts on the future of forensic DNA testing. *Forensic science, medicine, and pathology*, 3, pp.200-205.

Butler JM. and Hill CR. 2012. Biology and Genetics of New Autosomal STR Loci Useful for Forensic DNA Analysis, *Forensic Science Review*, 24(15), pp. 16-26.

Choo, S.Y., 2007. The HLA system: genetics, immunology, clinical testing, and clinical implications. *Yonsei medical journal*, 48(1), pp.11-23.

Coble, M. and Hill, B., 2012. Application of Thresholds for Interpretation. The Copenhagen Forensic Genetic Summer School Advanced Topics in STR DNA Analysis. *NIST Applied Genetics Group*, 27.

Dakin, E.E. and Avise, J.C., 2004. Microsatellite null alleles in parentage analysis. Heredity, 93(5), pp.504-509.

Dauber, E.M., Kratzer, A., Neuhuber, F., Parson, W., Klintschar, M., Bär, W. and Mayr, W.R., 2012. Germline mutations of STR-alleles include multi-step mutations as defined by sequencing of repeat and flanking regions. Forensic Science International: Genetics, 6(3), pp.381-386.

De Bruyn, M., Kotze, A., Harris, Y., Dalton, D.L. and Welgemoed, D.J., 2019. Nucleotide sequence analysis to identify a one-step mutation in a STR DNA profile during paternity testing at locus D7S820. The *Journal of Medical Laboratory Science and Technology of South Africa*, 1(3), pp.20-24.

De Kock, A. and Kloppers, J.F. 2021. The impact of motherless paternity testing in a South African population. African Health Sciences, 21(1), pp. 379-384.

Decorte, R. and Cassiman, J.J., 1993. Forensic medicine and the polymerase chain reaction technique. Journal of medical genetics, 30(8), p.625.

Eduardoff, M., Santos, C., De La Puente, M., Gross, T.E., Fondevila, M., Strobl, C., Sobrino, B., Ballard, D., Schneider, P.M., Carracedo, A. and Lareu, M.V., 2015. Inter-laboratory evaluation of SNP-based forensic identification by massively parallel sequencing using the Ion PGM™. Forensic Science International: Genetics, 17, pp.110-121.

Edwards, M. and Allen, R.W., 2004. Characteristics of mutations at the D5S818 locus studied with a tightly linked marker. Transfusion, 44(1), pp.83-90.

El-Alfy, S.H. and Abd El-Hafez, A.F., 2012. Paternity testing and forensic DNA typing by multiplex STR analysis using ABI PRISM 310 Genetic Analyzer. Journal of Genetic Engineering and Biotechnology, 10(1), pp.101-112.

Fan, H. and Chu, J.Y., 2007. A brief review of short tandem repeat mutation. *Genomics, proteomics & bioinformatics*, 5(1), pp.7-14.

Fan, H., Wang, L., Liu, C., Lu, X., Xu, X., Ru, K., Qiu, P., Liu, C. and Wen, S.Q., 2022. 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), pp.447-464.

Flanagan, S.P. and Jones, A.G., 2019. The future of parentage analysis: From microsatellites to SNPs and beyond. *Molecular ecology*, 28(3), pp.544-567.

Fujii, K., Watahiki, H., Mita, Y., Iwashima, Y., Kitayama, T., Nakahara, H., Mizuno, N. and Sekiguchi, K., 2016. Typing concordance between PowerPlex® Fusion and

GlobalFiler® based on 1501 Japanese individuals and the causes of typing discrepancies. *Forensic Science International: Genetics*, 25, pp.e12-e13.

Gettings, K.B., Aponte, R.A., Vallone, P.M. and Butler, J.M., 2015. STR allele sequence variation: current knowledge and future issues. *Forensic Science International: Genetics*, 18, pp.118-130.

Gill, P., Brinkmann, B., D'Aloja, E., Andersen, J., Bar, W., Carracedo, A., Dupuy, B., Eriksen, B., Jangblad, M., Johnsson, V., Kloosterman, A. D., Lincoln, P., Morling, N., Rand, S., Sabatier, M., Scheithauer, R., Schneider, P. and Vide, M. C. 1997. Considerations from the European DNA profiling group (EDNAP) concerning STR nomenclature, *Forensic Science International*, 87(3), pp. 185-192.

Gjertson, D. W., Brenner, C. H., Baur, M. P., Carracedo, A., Mayr, W. R., Guidet, F., Luque, J. A., Pascali, V. L., Prinz, M., Schneider, P. M. and Morling, N. 2007. ISFG: Recommendations on biostatistics in paternity testing. *Forensic Science International: Genetics*, 1(3-4), pp. 223-231.

Goodwin, W., Linacre, A. and Hadi, S., 2007. *An introduction to forensic genetics*. John Wiley & Sons.

Green, R., Elliott, J.L., Norona, W., Go, F., Nguyen, V.T., Ge, J., Short, M.L., Mulero, J.J. and Zhong, C., 2021. Developmental validation of VeriFiler™ Plus PCR Amplification Kit: A 6-dye multiplex assay designed for casework samples. *Forensic Science International: Genetics*, 53, p.102494.

Hamester, F.I.R., da Silva, D.S., Leboute, A.P.M., Motta, C.H.A. and Alho, C.S., 2019. Slippage mutation rates in 15 autosomal short tandem repeat loci for forensic purposes in a Southeastern Brazilian population. *Electrophoresis*, 40(21), pp.2873-2876.

Hammond, H.A., Jin, L., Zhong, Y., Caskey, C.T. and Chakraborty, R., 1994. Evaluation of 13 short tandem repeat loci for use in personal identification applications. *American journal of human genetics*, 55(1), p.175.

Heinrich, M., Felske-Zech, H., Brinkmann, B. and Hohoff, C., 2005. Characterisation of variant alleles in the STR systems D2S1338, D3S1358 and D19S433. *International journal of legal medicine*, 119, pp.310-313.

Hering, S., Nixdorf, R. and Dreßler, J., 2005. Identification of more sequence variations in the D8S1179 locus. *Forensic science international*, 149(2-3), pp.275-278.

Hill, C.R., Duewer, D.L., Kline, M.C., Sprecher, C.J., McLaren, R.S., Rabbach, D.R., Krenke, B.E., Ensenberger, M.G., Fulmer, P.M., Storts, D.R. and Butler, J.M., 2011. Concordance and population studies along with stutter and peak height ratio analysis for the PowerPlex® ESX 17 and ESI 17 Systems. *Forensic Science International: Genetics*, 5(4), pp.269-275.

Jeffreys, A.J., Wilson, V. and Thein, S.L., 1985. Hypervariable 'minisatellite'regions in human DNA. *Nature*, 314(6006), pp.67-73.

Jeffreys, A.J., Wilson, V., Neumann, R. and Keyte, J., 1988. Amplification of human minisatellites by the polymerase chain reaction: towards DNA fingerprinting of single cells. *Nucleic Acids Research*, 16(23), pp.10953-10971.

Jobling, M.A. and Gill, P., 2004. Encoded evidence: DNA in forensic analysis. *Nature Reviews Genetics*, 6(3), pp.246-247.

Jordan, D. and Mills, D., 2021. Past, present, and future of DNA typing for analyzing human and non-human forensic samples. *Frontiers in Ecology and Evolution*, 9, p.646130.

Jia, Y.S., Zhang, L., Qi, L.Y., Mei, K., Zhou, F.L., Huang, D.X. and Yi, S.H., 2015. Multistep microsatellite mutation leading to father-child mismatch of FGA locus in a case of non-exclusion parentage. *Legal Medicine*, 17(5), pp.364-365.

Jian, H., Wang, L., Lv, M., Tan, Y., Zhang, R., Qu, S., Wang, J., Zha, L., Zhang, L. and Liang, W., 2021. A novel SNP-STR system based on a capillary electrophoresis platform. *Frontiers in Genetics*, 12, p.636821.

Jin, B., Su, Q., Luo, H., Li, Y., Wu, J., Yan, J., Hou, Y., Liang, W. and Zhang, L., 2016. Mutational analysis of 33 autosomal short tandem repeat (STR) loci in

southwest Chinese Han population based on trio parentage testing. *Forensic Science International: Genetics*, 23, pp.86-90.

Jin, L., Underhill, P.A., Buoncristiani, M. and Robertson, J.M., 1997. Defining microsatellite alleles by genotyping global indigenous human populations and non-human primates. *Journal of Forensic sciences*, 42, pp.496-499.

Kebede E.A. 2002. *DNA characterisation of the FGA locus in the human genome*. (Master's dissertation, University of the Free State).

Kline, M.C., Hill, C.R., Decker, A.E., and Butler, J.M., 2011. STR sequence analysis for characterizing normal, variant, and null alleles. *Forensic Science International: Genetics*, 5(4), pp.329-332.

Krawczak, M. and Nothnagel, M., 2009. Kinship testing goes linkage. *Annals of Human Genetics*, 73, pp.665-665.

Krenke, B.E., Tereba, A., Anderson, S.J., Buel, E., Culhane, S., Finis, C.J., Tomsey, C.S., Zachetti, J.M., Masibay, A., Rabbach, D.R. and Amiott, E.A., 2002. Validation of a 16-locus fluorescent multiplex system. *Journal of Forensic Sciences*, 47(4), pp.773-785.

Lane, A.B., 2013. STR null alleles complicate parentage testing in South Africa: Festschrift: Professor Trefor Jenkins-molecular genetics. *South African Medical Journal*, 103(12), pp.1004-1008.

Lareu, M. (2013). Short Tandem Repeats. Forensic Sciences: Second Edition, *Elsevier Inc*., pp. 219-226.

Li, H., Schmidt, L., Wei, M.H., Hustad, T., Lerman, M.I., Zbar, B. and Tory, K., 1993. Three tetranucleotide polymorphisms for loci: D3S1352; D3S1358; D3S1359. *Human molecular genetics*, 2(8), pp.1327-1327.

Lu, D., Liu, Q., Wu, W. and Zhao, H., 2012. Mutation analysis of 24 short tandem repeats in Chinese Han population. *International Journal of Legal Medicine*, 126, pp.331-335.

Manas, G.J., 1994. Calculating the probability of paternity in genetic testing. South African Journal of Science, 90(1), pp.33-36.

Mardini, A.C., Rodenbusch, R., Schumacher, S., Chula, F.G.L., Michelon, C.T., Gastaldo, A.Z., Maciel, L.P., de Matos Almeida, S.E. and da Silva, C.M.D., 2013. Mutation rate estimates for 13 STR loci in a large population from Rio Grande do Sul, Southern Brazil. *International journal of legal medicine*, 127, pp.45-47.

Martinez, J., Braganholi, D.F., Ambrósio, I.B., Polverari, F.S., and Cicarelli, R.M.B., 2017. Mutation rates for 20 STR loci in a population from São Paulo state, Southeast, Brazil. *Annals of human biology*, 44(7), pp.659-662.

Mokoma, P.M. 2017. *Implications of secondary DNA transfer & public perceptions on the national DNA database of South Africa* (Master's dissertation, University of the Free State).

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G. and Erlich, H., 1986. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harbor symposia on quantitative biology*, 51, pp. 263-273.

Murat, P., Guilbaud, G., and Sale, J.E., 2020. DNA polymerase stalling at structured DNA constrains the expansion of short tandem repeats. *Genome biology*, 21(1), pp.1-26.

Negi, D.S., Shrivastava, P. and Das, S.P., 2014. Biochemical characterization of Molecular Markers for Human Genetic Identification in Paternity testing by DNA profiling. *Asian Journal of Biomedical and Pharmaceutical Sciences*, 4(37), pp. 52-56.

Negi, D.S, Alam, M., Bhavani, S.A. and Nagaraju, J., 2006. Multistep microsatellite mutation in the maternally transmitted locus D13S317: a case of maternal allele mismatch in the child. International Journal of Legal Medicine, 120, pp.286-292.

Nothnagel, M., Schmidtke, J. and Krawczak, M., 2010. Potentials and limits of pairwise kinship analysis using autosomal short tandem repeat loci. *International journal of legal medicine*, 124, pp.205-215.

Paredes, M., 2011. Mutational dynamics of STRs: Analysis of FGA-FIBRA locus. *Forensic Science International: Genetics Supplement Series*, 3(1), pp.e261-e262.

Pereira, R., Phillips, C., Alves, C., Amorim, A., Carracedo, Á. and Gusmão, L., 2009. A new multiplex for human identification using insertion/deletion polymorphisms. *Electrophoresis*, 30(21), pp.3682-3690.

Pitterl, F., Schmidt, K., Huber, G., Zimmermann, B., Delport, R., Amory, S., Ludes, B., Oberacher, H. and Parson, W., 2010. Increasing the discrimination power of forensic STR testing by employing high-performance mass spectrometry, as illustrated in indigenous South African and Central Asian populations. *International journal of legal medicine*, 124, pp.551-558.

Poethe, S.S., Holtel, J., Biermann, J.P., Riemer, T., Grabmüller, M., Madea, B., Thiele, R. and Jäger, R., 2023. Cost-Effective Next Generation Sequencing-Based STR Typing with Improved Analysis of Minor, Degraded and Inhibitor-Containing DNA Samples. *International Journal of Molecular Sciences*, 24(4), p.3382.

Press, S.J. and Shigemasu, K., 1989. Bayesian inference in factor analysis. In Contributions to probability and statistics: Essays in honor of Ingram Olkin (pp. 271-287). New York, NY: Springer New York.

Rockenbauer, E., Holgersson, M.B., Fordyce, S.L., Arcos, M.C.Á., Børsting, C., Hansen, A.J., Frank-Hansen, R., Willerslev, E., Gilbert, M.T.P. and Morling, N., 2011. Sequences of microvariant/ “off-ladder” STR alleles. *Forensic Science International: Genetics Supplement Series*, 3(1), pp.e204-e205.

Press, S.J. and Shigemasu, K., 1989. Bayesian inference in factor analysis. In Contributions to probability and statistics: Essays in honor of Ingram Olkin (pp. 271-287). New York, NY: Springer New York.

Ruitberg, C.M., Reeder, D.J. and Butler, J.M., 2001. STRBase: a short tandem repeat DNA database for the human identity testing community. *Nucleic acids research*, 29(1), pp.320-322.

Schwarz, H.P. and Dorner, F., 2003. Karl Landsteiner and his major contributions to haematology. *British Journal of Haematology*, 121(4), pp.556-565.

Shao, C., Yao, Y., Pan, X., Wu, M., Zhang, B., Xu, H., Xie, J. and Sun, K., 2021. Variants in linkage status at D5S818 detected by multiple STR kits comparison and Sanger sequencing. *Molecular Genetics & Genomic Medicine*, 9(9), p.e1765.

Shimmin, L.C., Chang, B.H.J. and Li, W.H., 1993. Male-driven evolution of DNA sequences. *Nature*, 362(6422), pp.745-747.

Shin, C.H., Jang, P., Hong, K.M. and Paik, M.K., 2004. Allele frequencies of 10 STR loci in Koreans. *Forensic Science International*, 140(1), pp.133-135.

Sun, H., Liu, S., Zhang, Y. and Whittle, M.R., 2014. Comparison of southern Chinese Han and Brazilian Caucasian mutation rates at autosomal short tandem repeat loci used in human forensic genetics. *International journal of legal medicine*, 128, pp.1-9.

Sun, Y.L., Xu, Y.Z. and Chambon, P., 1982. A simple and efficient method for the separation and detection of small DNA fragments by electrophoresis in formamide containing agarose gels and Southern blotting to DBM-paper. *Nucleic Acids Research*, 10(19), pp.5753-5763.

Thomson, J.A., Pilotti, V., Stevens, P., Ayres, K.L. and Debenham, P.G., 1999. Validation of short tandem repeat analysis for the investigation of cases of disputed paternity. *Forensic science international*, 100(1-2), pp.1-16.

Tomas, C., Børsting, C. and Morling, N., 2012. A 48-plex autosomal SNP GenPlex™ assay for human individualization and relationship testing. *DNA Electrophoresis Protocols for Forensic Genetics*, pp.73-85.

Udogadi, N.S., Abdullahi, M.K., Bukola, A.T., Imose, O.P. and Esewi, A.D. 2020 Forensic DNA profiling: Autosomal short tandem repeat as a prominent marker in crime investigation. *Malaysian Journal of Medical Sciences*, 27(4), pp. 22-35.

Urquhart, A., Kimpton, C.P., Downes, T.J. and 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, pp.13-20.

Urquhart, A., Oldroyd, N.J., Kimpton, C.P. and Gill, P., 1995. Highly discriminating heptaplex short tandem repeat PCR system for forensic identification. *Biotechniques*, 18(1), pp.116-8.

Van Camp, N., and Dierickx, K., 2007. The expansion of forensic DNA databases and police sampling powers in the post-9/11 era: ethical considerations on genetic privacy. *Ethical Perspectives: Journal of the European Ethics Network*, 14(3), pp. 237-268.

Wallin, J.M., Holt, C.L., Lazaruk, K.D., Nguyen, T.H. and Walsh, P.S., 2002. Constructing universal multiplex PCR systems for comparative genotyping. *Journal of Forensic Sciences*, 47(1), pp.52-65.

Wang, J., 2010. Effects of genotyping errors on parentage exclusion analysis. *Molecular Ecology*, 19(22), pp.5061-5078.

Watkins, W.M., 2001. The ABO blood group system: historical background. *Transfusion medicine*, 11(4), pp.243-265.

Wei, T., Liao, F., Wang, Y., Pan, C., Xiao, C. and Huang, D., 2018. A novel multiplex assay of SNP-STR markers for forensic purpose. *PLoS One*, 13(7), p.e0200700.

Welch, L.A., Gill, P., Phillips, C., Ansell, R., Morling, N., Parson, W., Palo, J.U. and Bastisch, I., 2012. European Network of Forensic Science Institutes (ENFSI): Evaluation of new commercial STR multiplexes that include the European Standard Set (ESS) of markers. *Forensic Science International: Genetics*, 6(6), pp.819-826.

Wenk, R.E., Traver, M. and Chiafari, F.A., 1996. Determination of sibship in any two persons. *Transfusion*, 36(3), pp.259-262.

Westen, A.A., Kraaijenbrink, T., de Medina, E.A.R., Harteveld, J., Willemse, P., Zuniga, S.B., van der Gaag, K.J., Weiler, N.E., Warnaar, J., Kayser, M. and Sijen, T.,

2014. Comparing six commercial autosomal STR kits in a large Dutch population sample. *Forensic Science International: Genetics*, 10, pp.55-63.

Wu, H., Zhang, L., Fan, A., Wu, H. and Wang, K., 2023. Genetic inconsistency at the D6S1043 locus caused by microdeletion at 6q15. I*nternational Journal of Legal Medicine*, pp.1-7.

Wyner, N., Barash, M. and McNevin, D., 2020. Forensic autosomal short tandem repeats and their potential association with phenotype. *Frontiers in genetics*, 11, p.884.

Xiao, C., Wang, Y., Liao, F., Yi, S. and Huang, D., 2019. Microdeletion at 8q24. 13 rather than multistep microsatellite mutation resulting in the genetic inconsistency at the D8S1179 locus in a true trio. *International Journal of Legal Medicine*, 133, pp.999-1006.

Xu, W., Yao, N., Li, X., Ma, Z., Zhou, H., Fu, S. and Chen, X., 2021. Mutation Analysis of STR Locus on 23 Autosomes in Hainan Population*. Forensic Medicine and Anatomy Research*, 10(1), pp.1-6.

Youn, B.J., Lee, K. and Kim, C.H., 2022. A Case of Single-step Mutations at Two Short Tandem Repeat loci (D13S317 and DXS10148) among Three Generations of a Korean Family. *Biomedical Science Letters*, 28(4), pp.327-333.

Zhang, Q. X., Yang, M., Pan, Y. J., Zhao, J., Qu, B. W., Cheng, F., & Yan, J. W. (2018). Development of a massively parallel sequencing assay for investigating sequence polymorphisms of 15 short tandem repeats in a Chinese Northern Han population. *Electrophoresis*, 39(21), 2725-2731.

Zhao, Z., Zhang, J., Wang, H., Liu, Z. P., Liu, M., Zhang, Y., & Zhang, H. (2015). Mutation rate estimation for 15 autosomal STR loci in a large population from Mainland China. *Meta Gene*, 5, 150-156.

Online reference sources

Basic Local Alignment Search Tool (BLAST). [Online]. Available at: https://blast.ncbi.nlm.nih.gov/Blast.cgi. (Accessed on: 10/11/2021)

Butler, J.M. and Reeder, D.J., 1997. National Institute of Science and Technology Short Tandem Repeat Internet Database Version 1.0 (NIST-STRBase). [Online]. Available at: https://strbase.nist.gov/index.htm (Accessed on: 01/02/21).

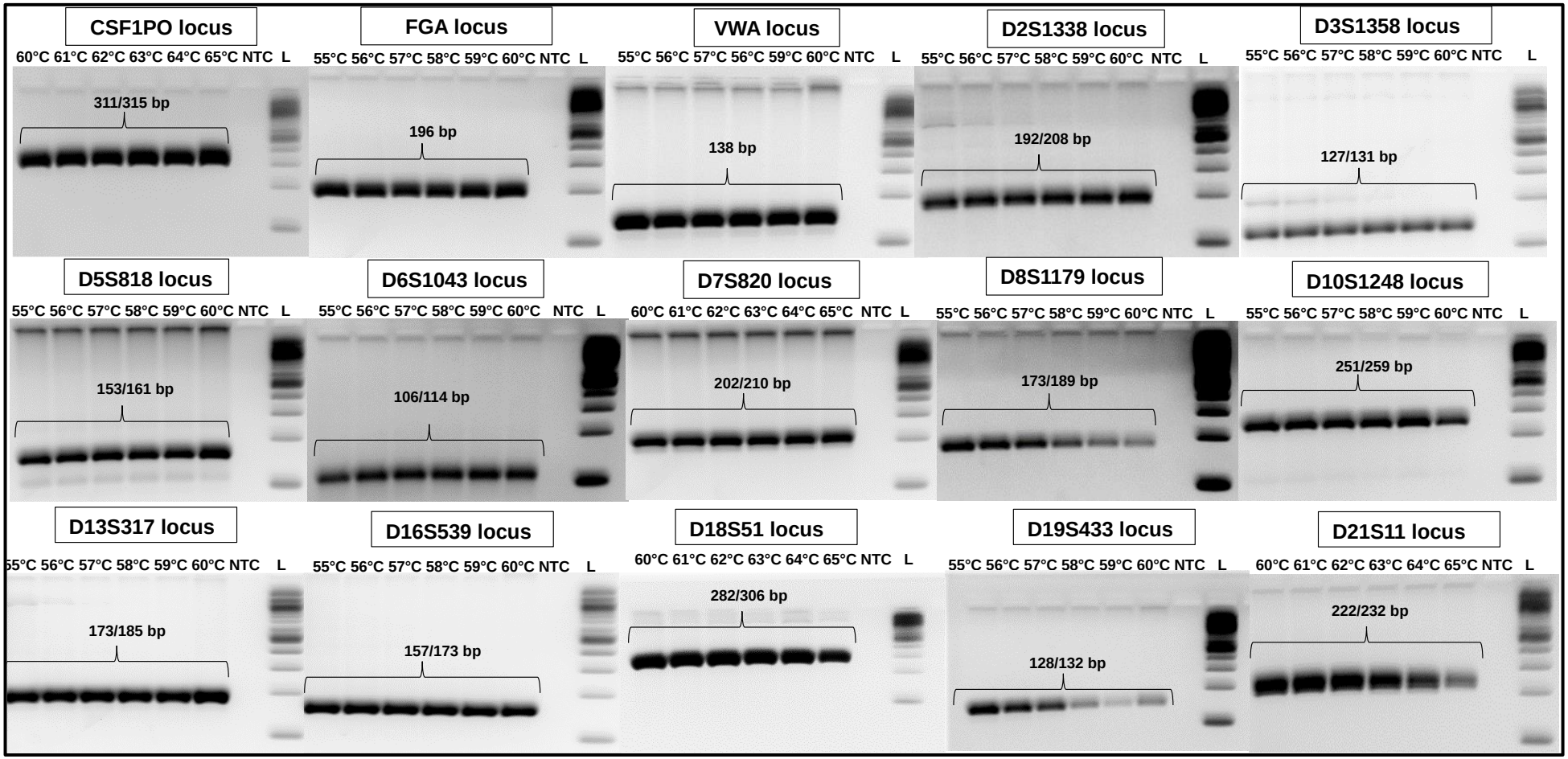
Butler J.M., Vallone P.M., Gettings K.B., Borsuk L.A., Ruitberg C.M, and Reeder D.J., 2023. National Institute of Science and Technology Short Tandem Repeat Internet Database Version 2.0 (NIST-STRBase). [Online]. Available at: https://strbase.nist.gov/. (Accessed on: 20/04/23).

Database for Single Nucleotide Polymorphisms (dbSNP). [Online]. Available at: https://www.ncbi.nlm.nih.gov/snp/. (Accessed on: 02/02/23).

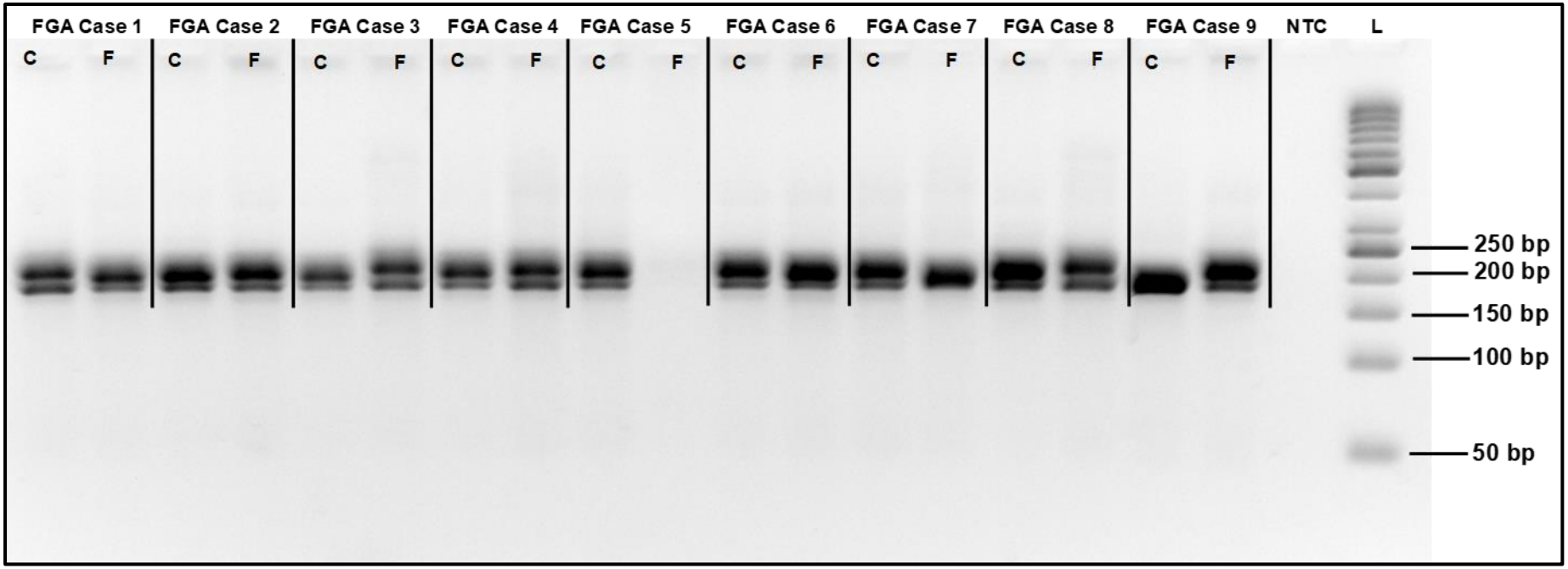
Untergasser, A., Nijveen, H., Rao X., Bisseling T., Geurts, R. and Leunissen J.A.M. 2007. Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Research*. [Online]

APPENDICES

Appendix A: Agarose gel electrophoresis images depicting the optimal primer annealing temperatures obtained for each tested STR locus using published primers

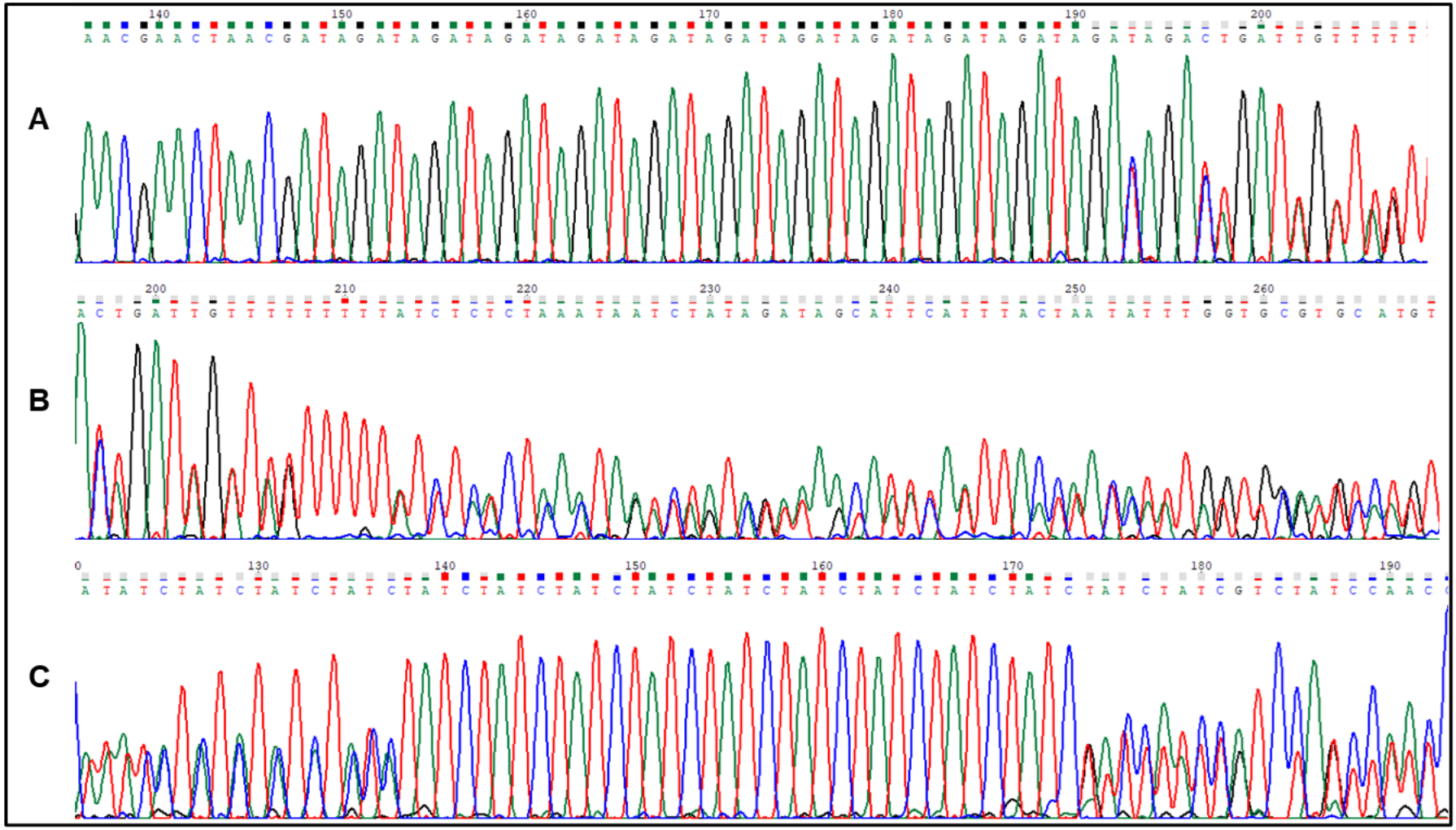


Appendix B: A 7% agarose-50% formamide gel electrophoresis image depicting an example separated heterozygous alleles for the FGA locus samples



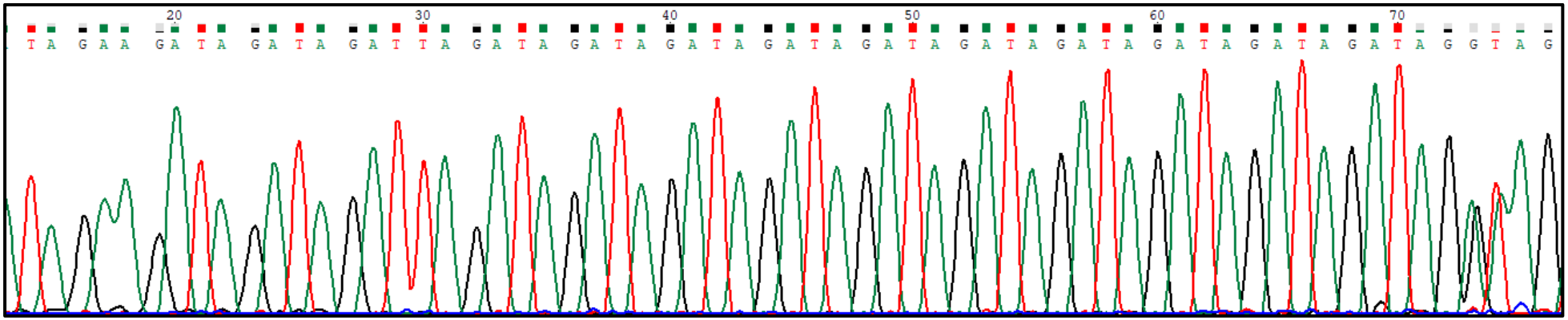
* C: Child, F: Father, NTC: non-template control, L: 50 bp molecular weight marker, bp: base pairs.

Appendix C: Sequence electropherograms depicting examples of the double-sequence data (overlapping electropherogram peaks) obtained

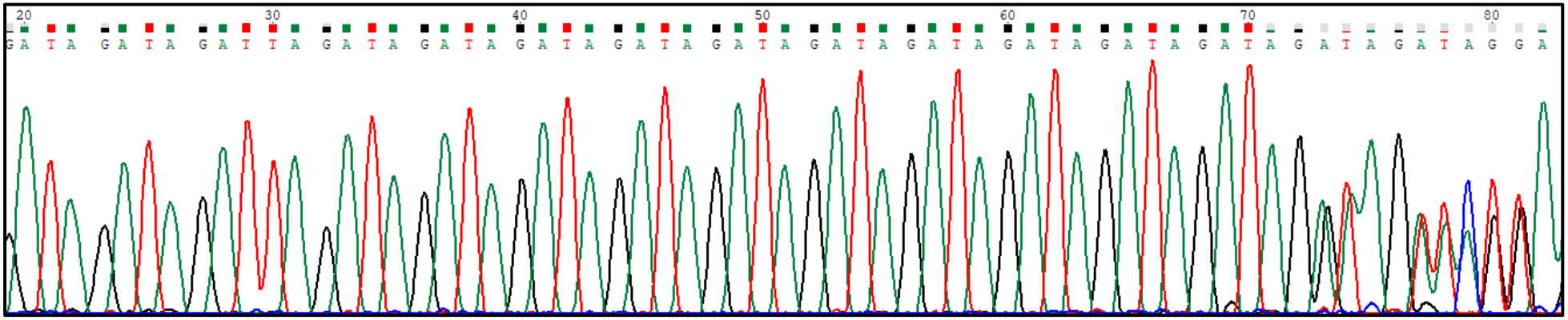


Appendix D: Sequence electropherograms of the 15 studied STR loci, presentation of one case per locus.

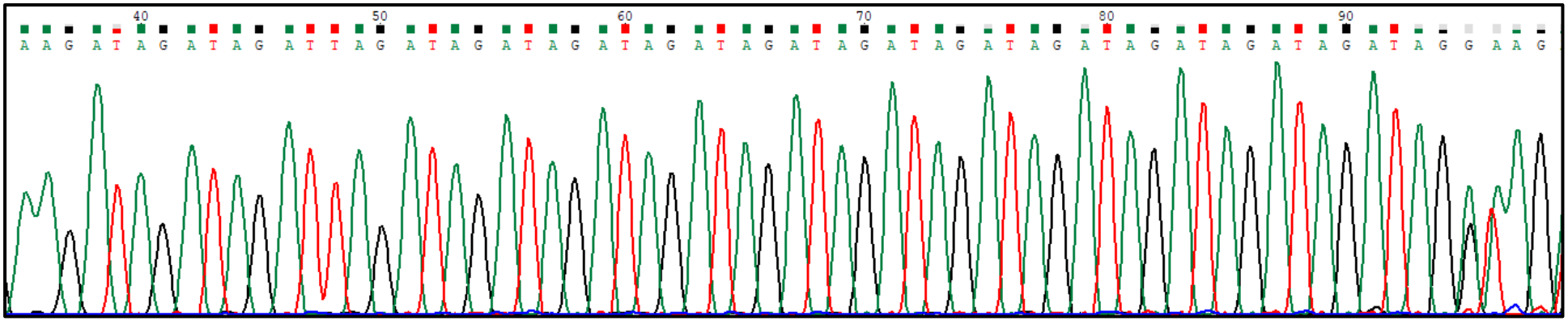
CSF1PO locus



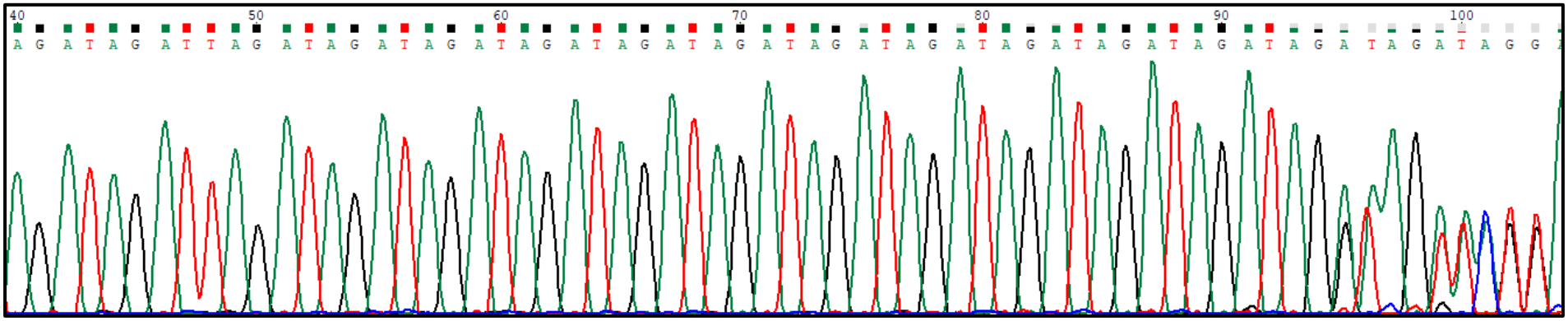
Case 1: Male child CSP1PO allele 10



Case 1: Male child CSP1PO allele 12

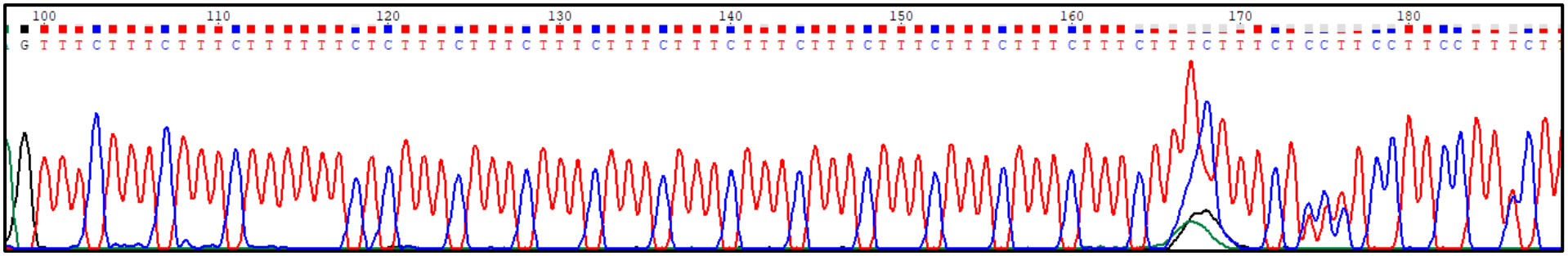


Case 1: Father CSP1PO allele 11

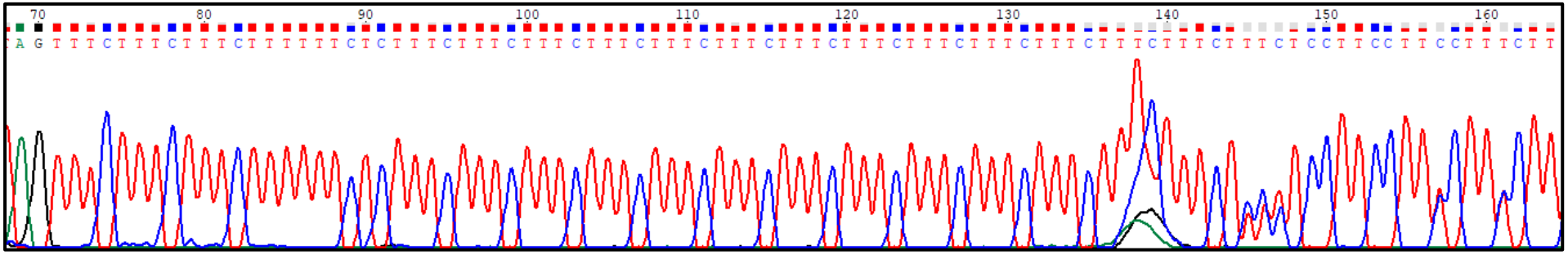


Case 1: Father CSP1PO allele 13

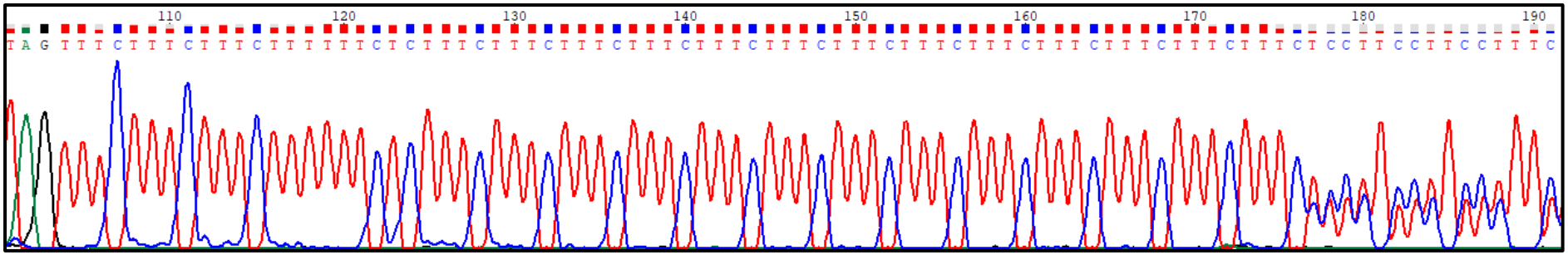
FGA locus



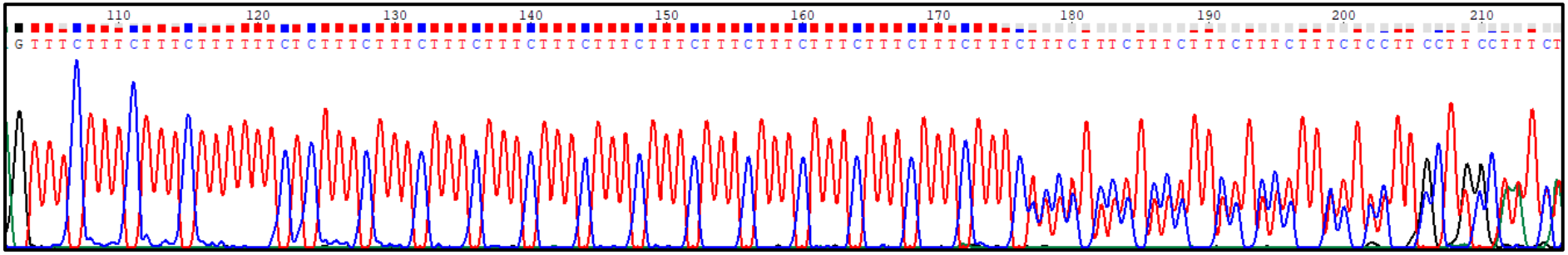
Case1: Male child FGA Allele 21



Case1: Male child FGA Allele 22

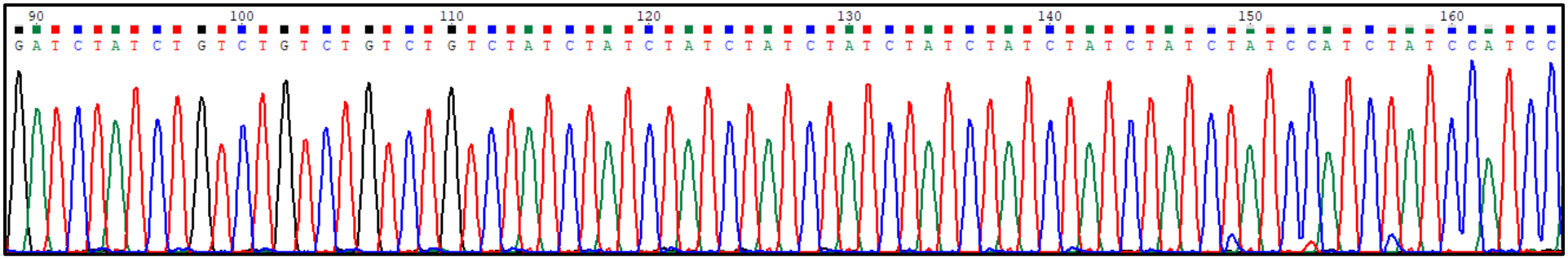


Case1: Father child FGA Allele 21

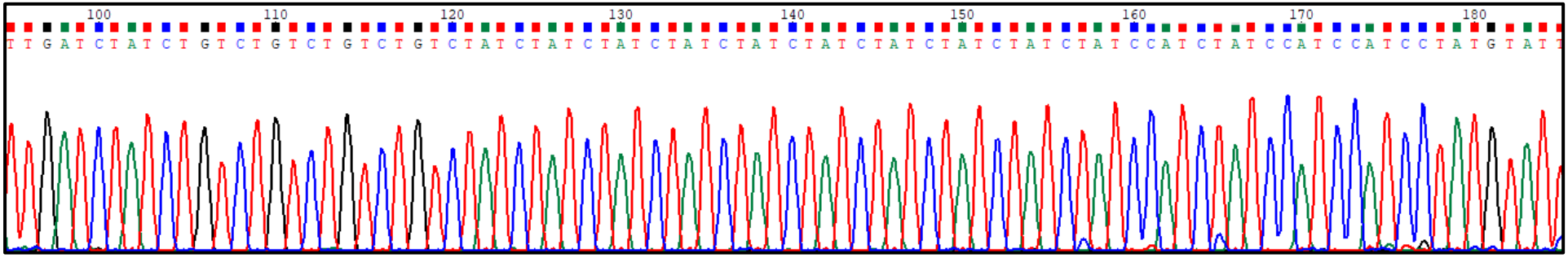


Case1: Father child FGA Allele 27

vWA locus

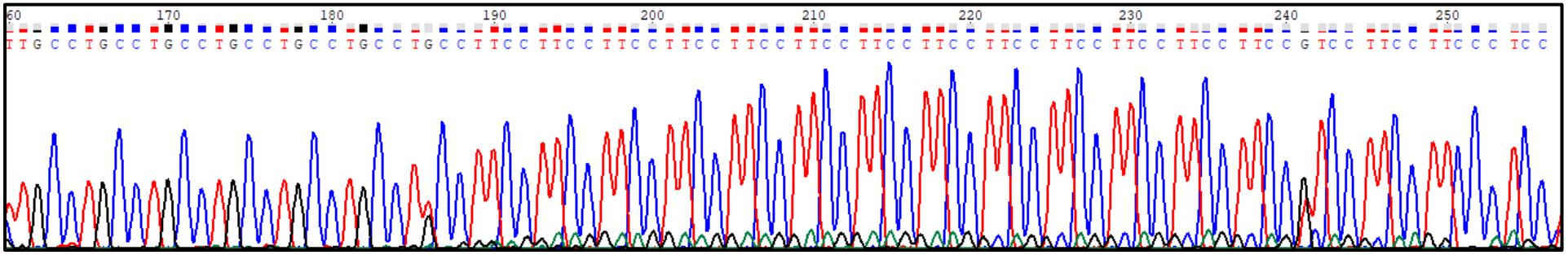


Case1: Male child vWA allele 15

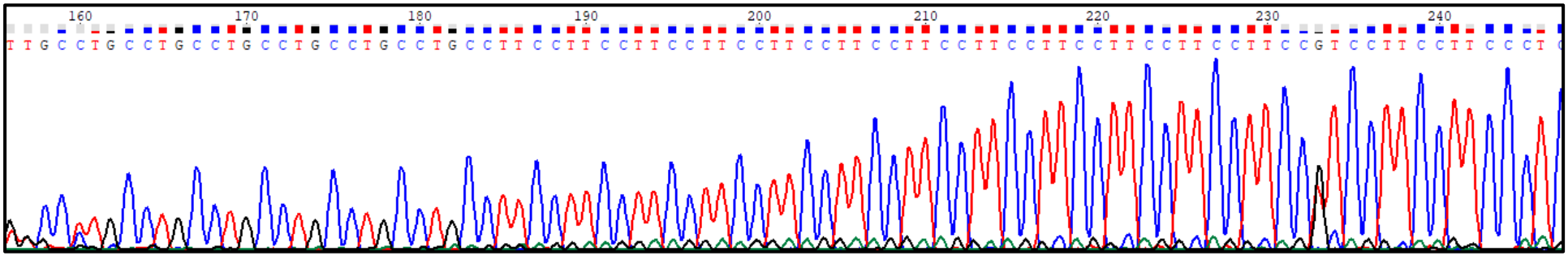


Case1: Father vWA allele 16

D2S1338 locus

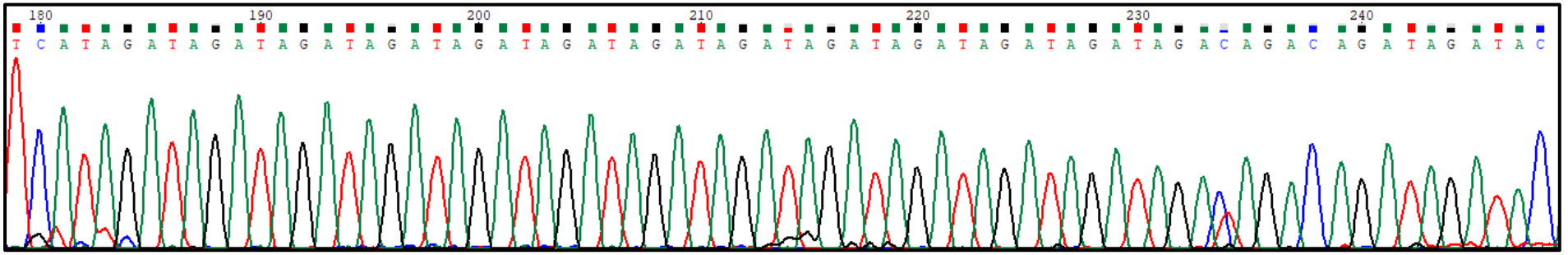


Case1: Male child D2S1338 allele 23

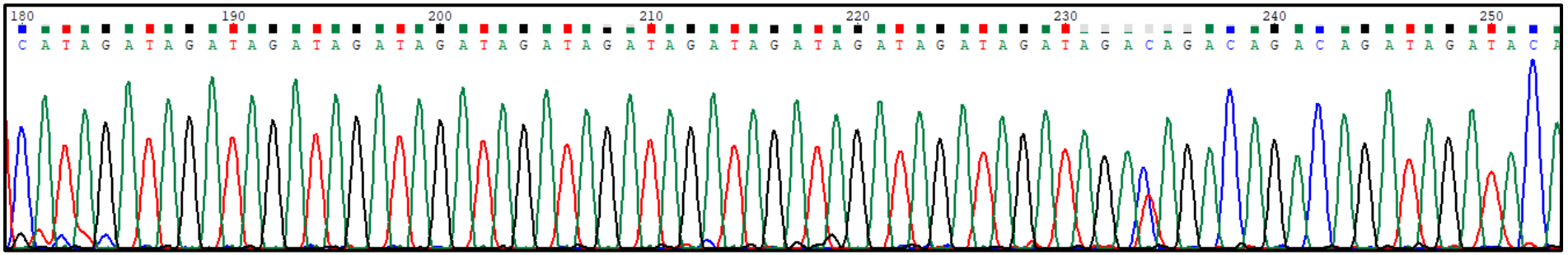


Case1: Father D2S1338 allele 22

D3S1358 locus

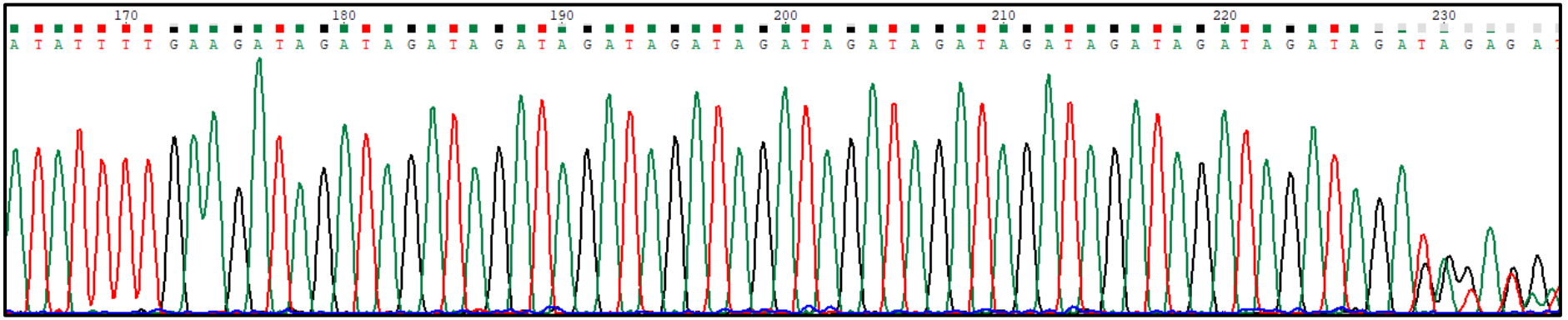


Case1: Female child D3S1358 allele 16

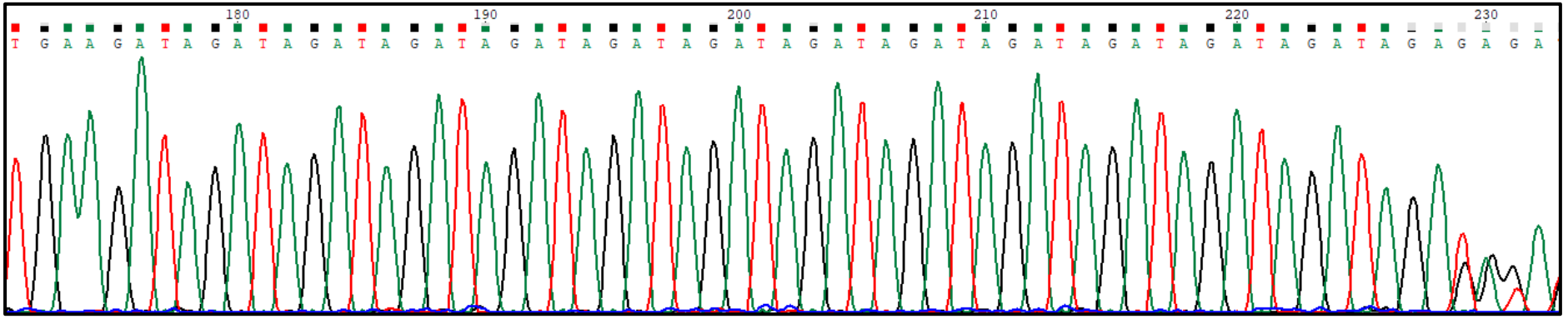


Case1: Father D3S1358 allele 17

D5S818 locus

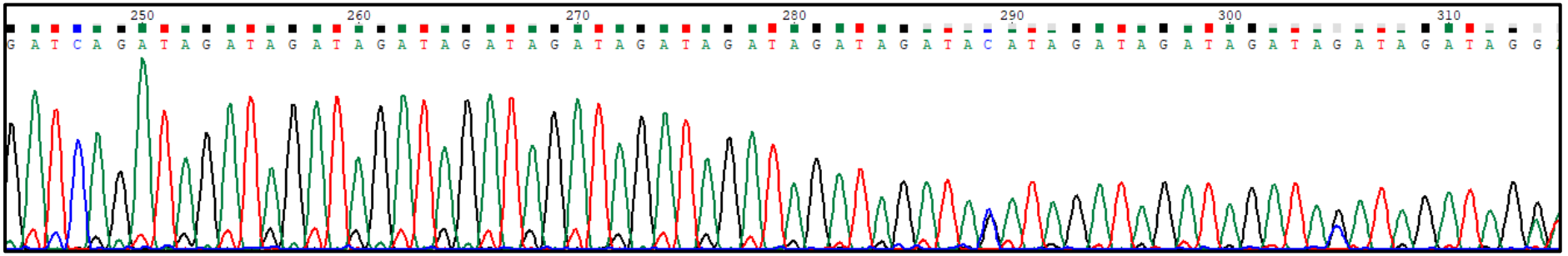


Case1: Female child D5S818 allele 14

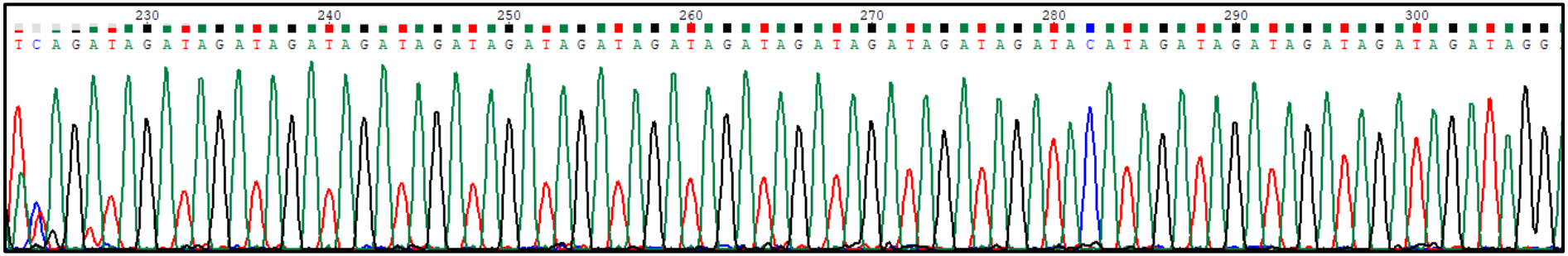


Case1: Father D5S818 allele 13

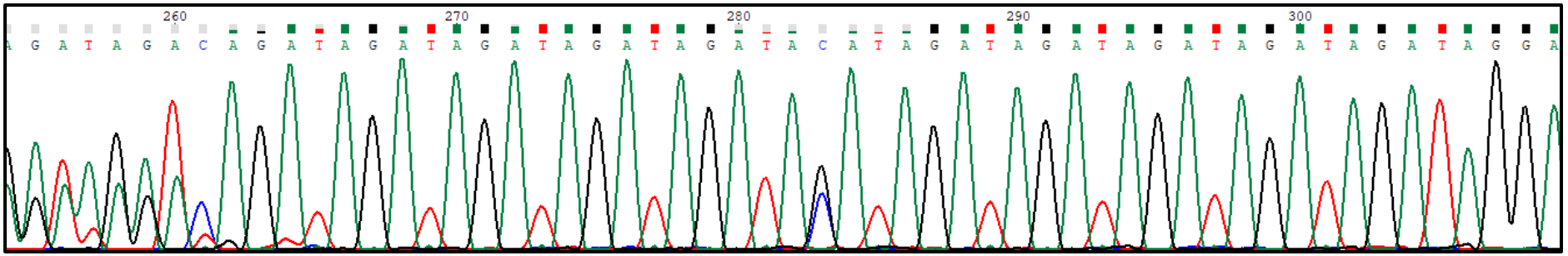
D6S1043 locus



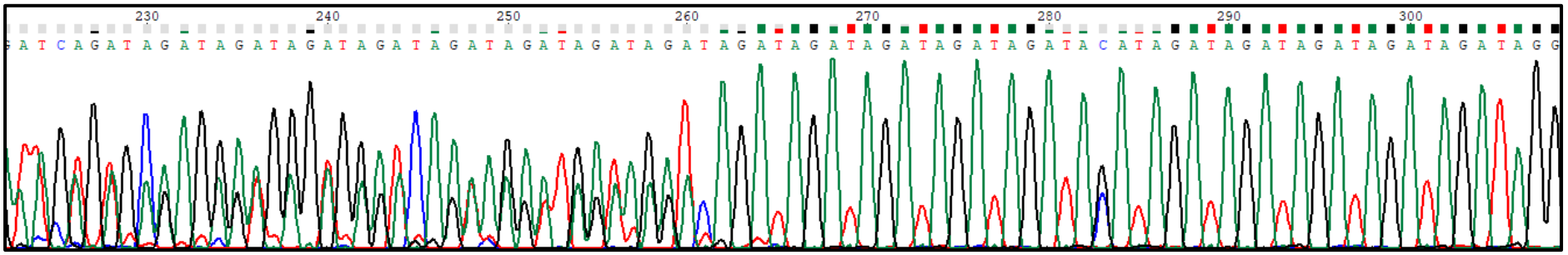
Case1: Male child D6S1043 allele 16



Case1: Male child D6S1043 allele 20

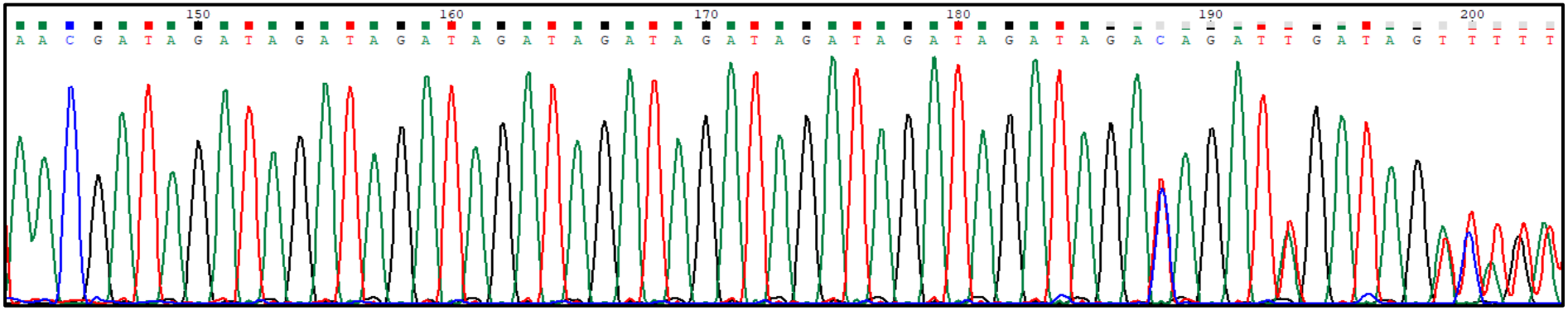


Case1: Father D6S1043 allele 11

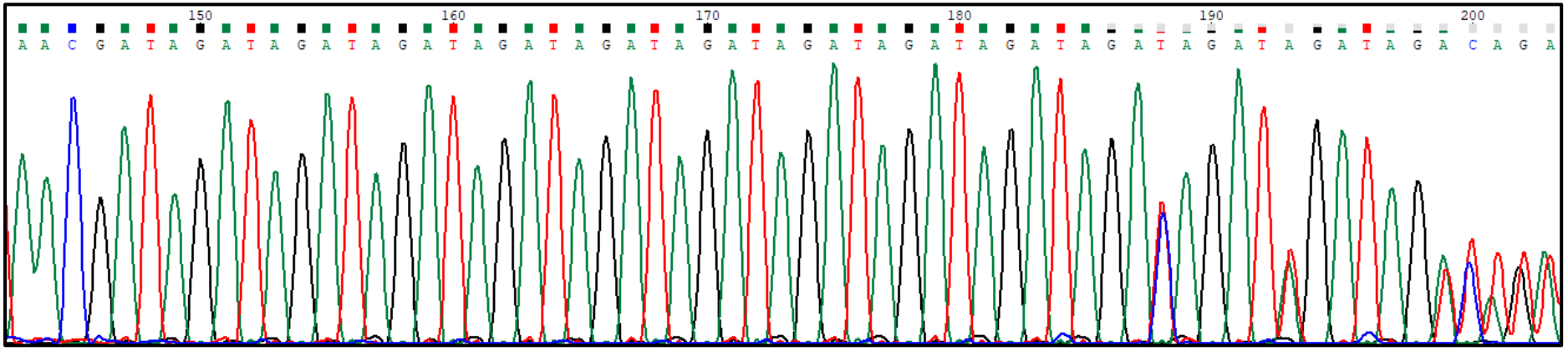


Case1: Father D6S1043 allele 20

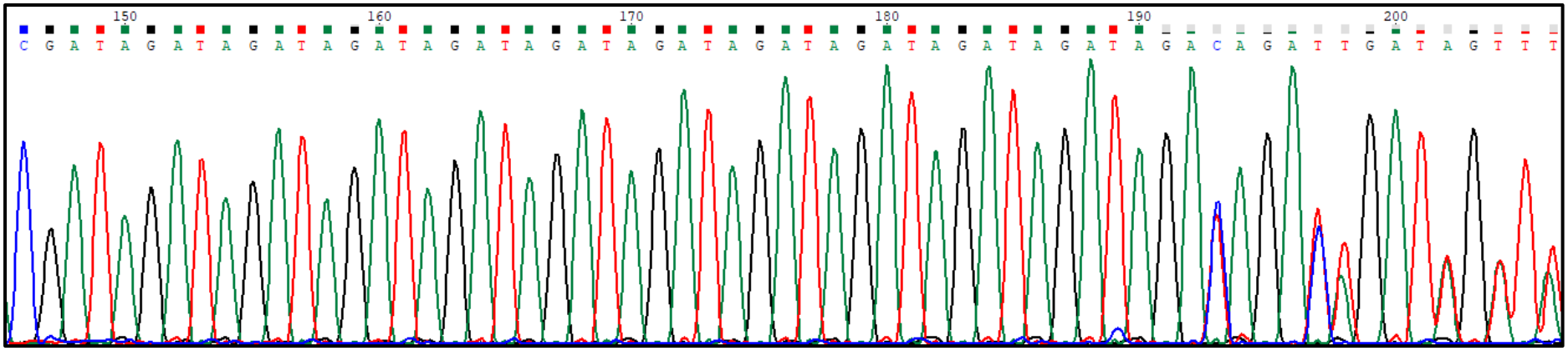
D7S820 locus



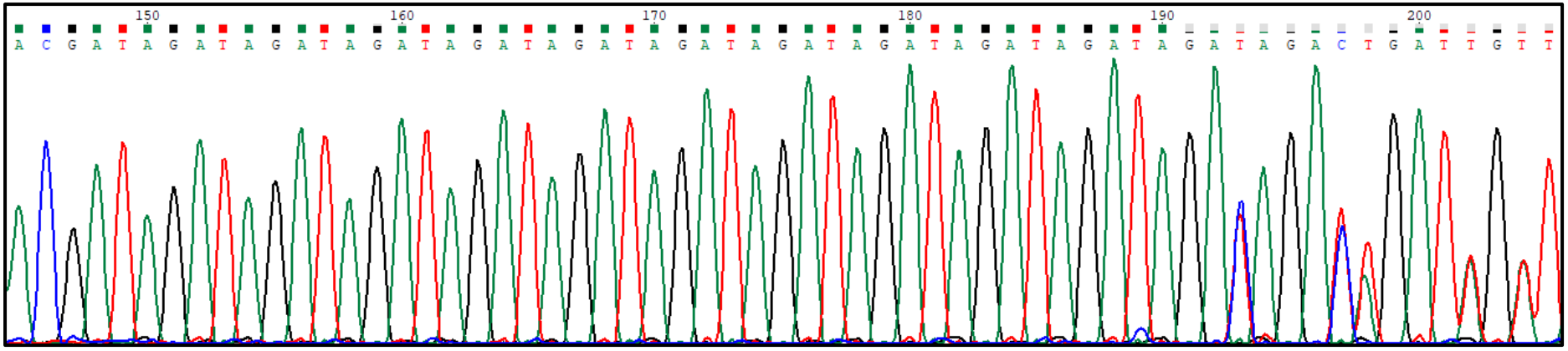
Case1: Female child D7S820 allele 10



Case1: Female child D7S820 allele 13

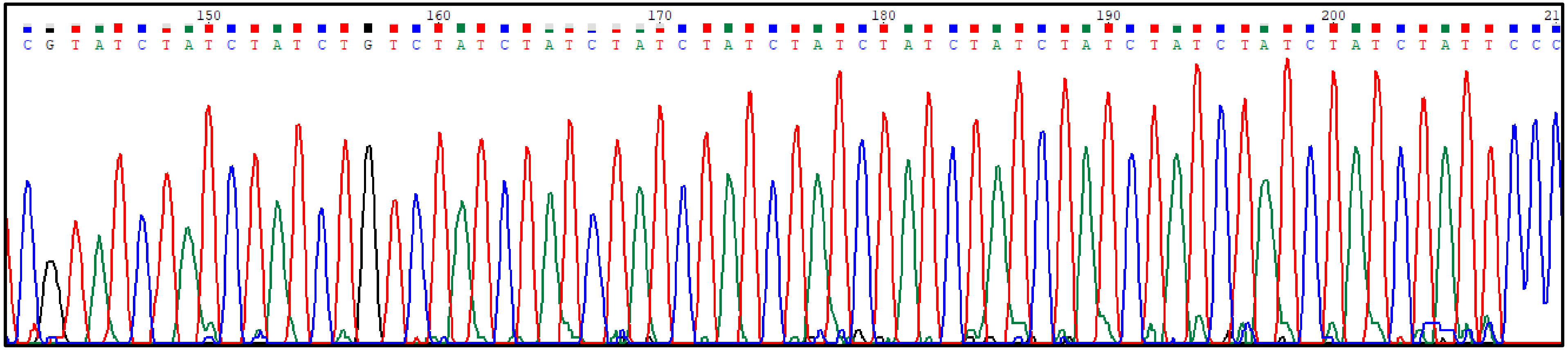


Case1: Father D7S820 allele 11

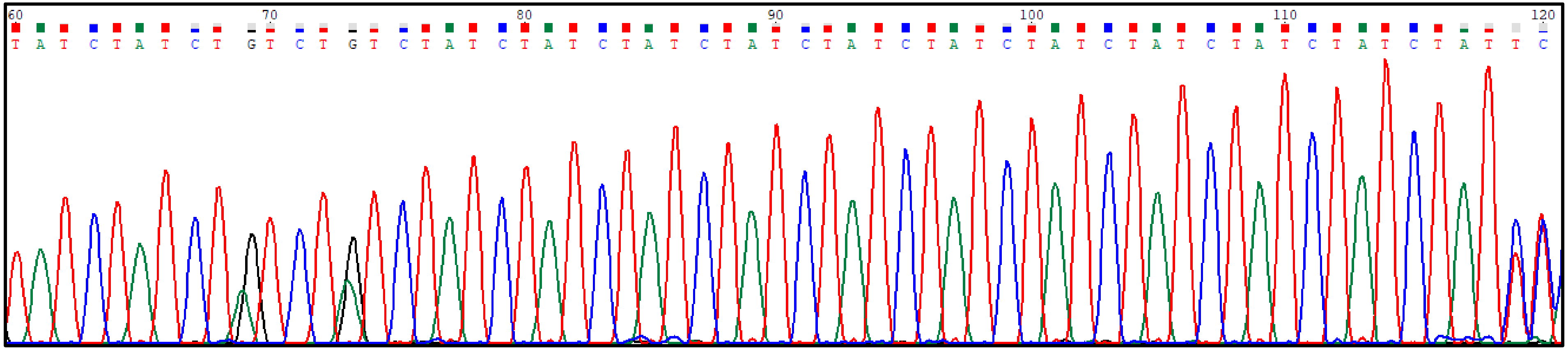


Case1: Father D7S820 allele 12

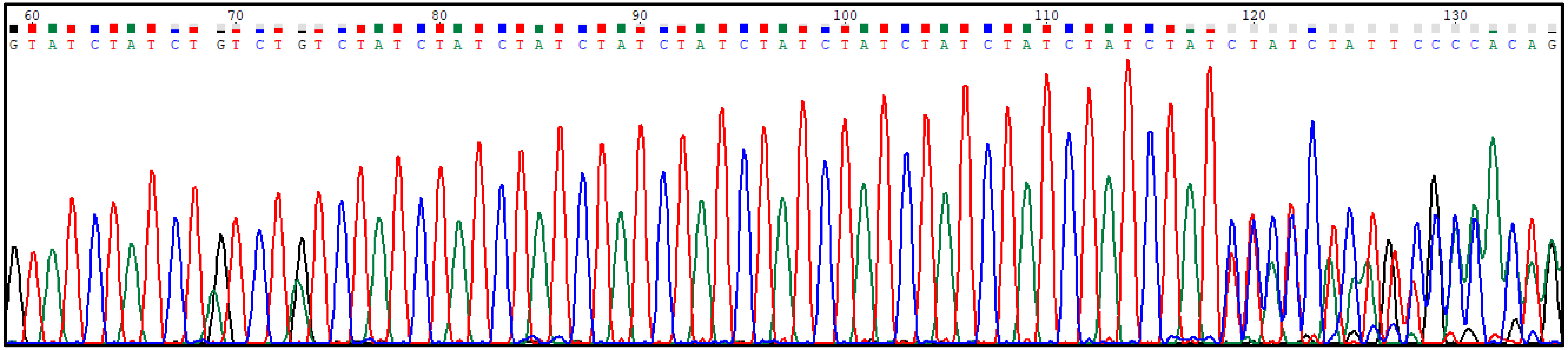
D8S1179 locus



Case1: Male child D8S1179 allele 15

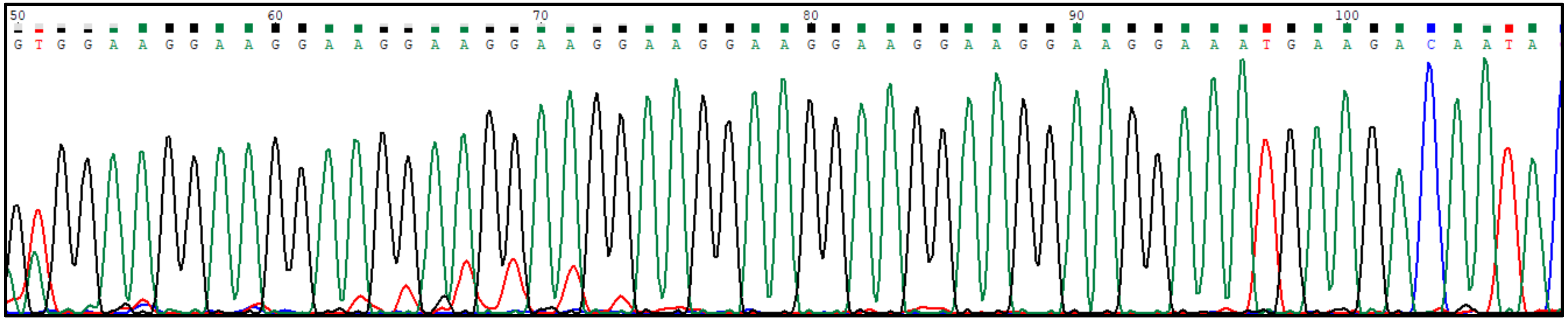


Case1: Father D8S1179 allele 14

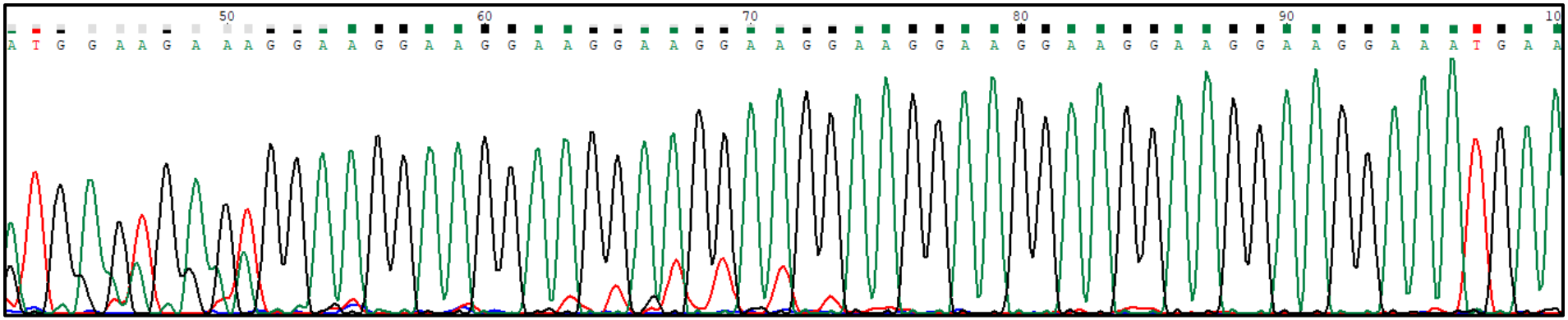


Case1: Father D8S1179 allele 16

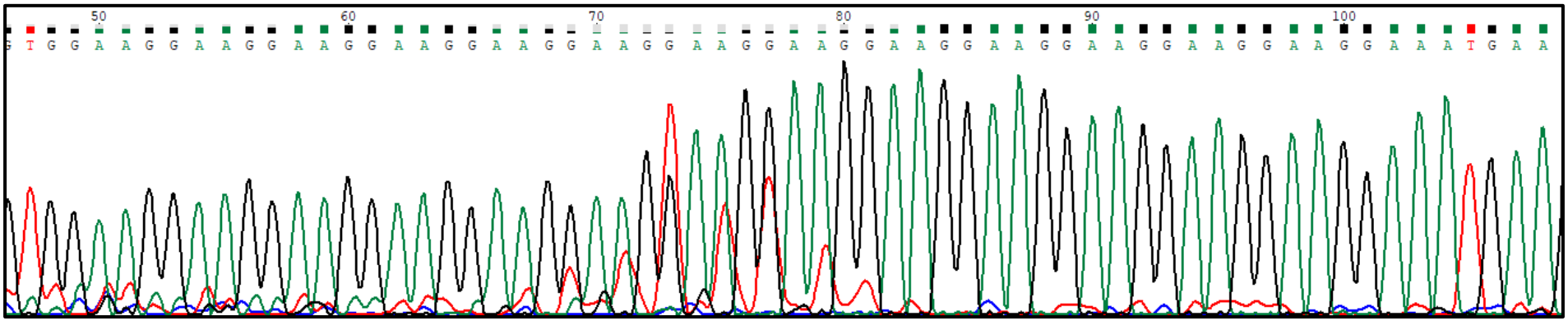
<u>D10S1248 locus</u>



Case 1: Male child D10S1248 allele 11

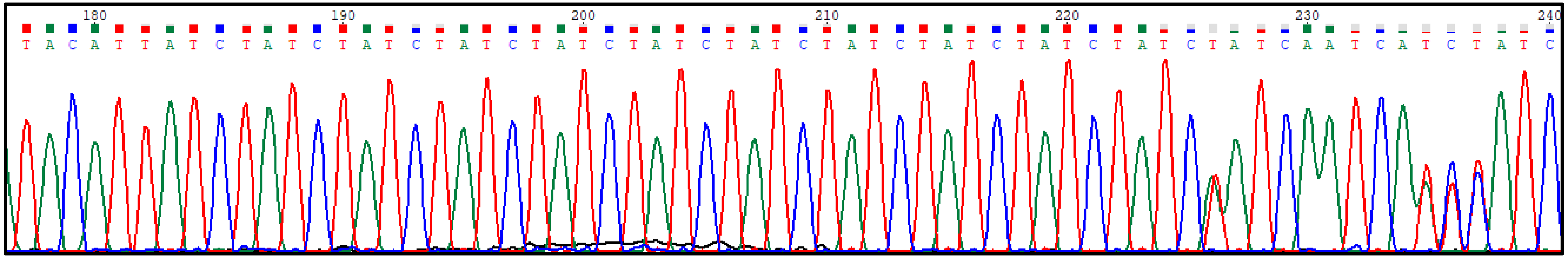


Case 1: Male child D10S1248 allele 13

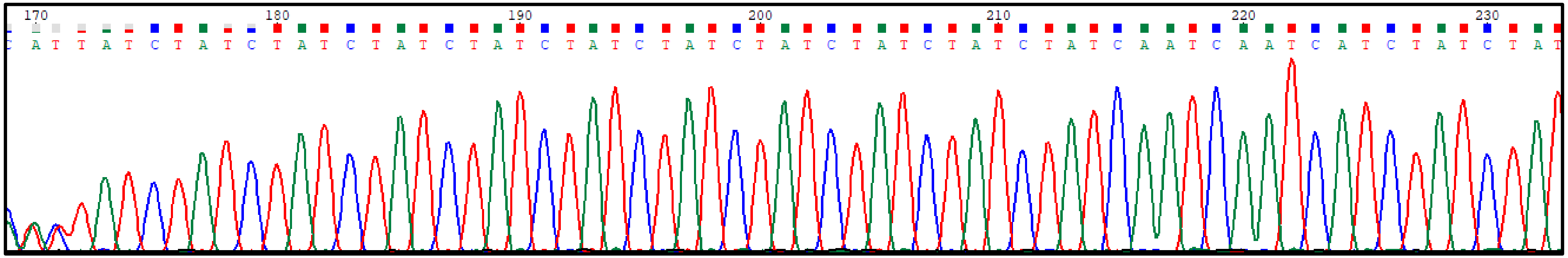


Case 1: Father D10S1248 allele 14

D13S317 locus

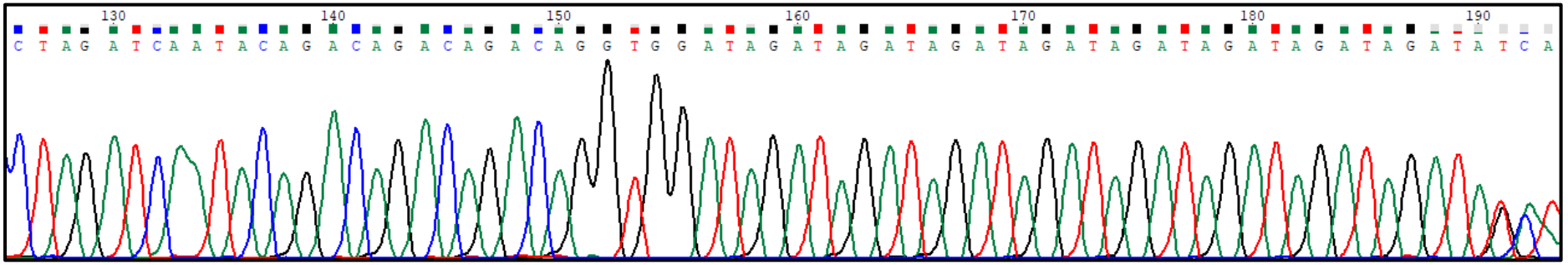


Case 1: Male child D13S317 allele 12

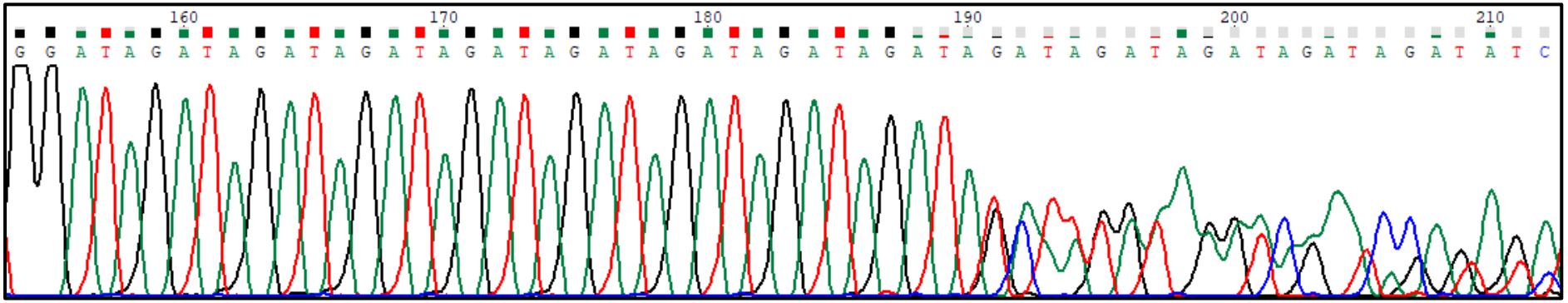


Case 1: Father D13S317 allele 13

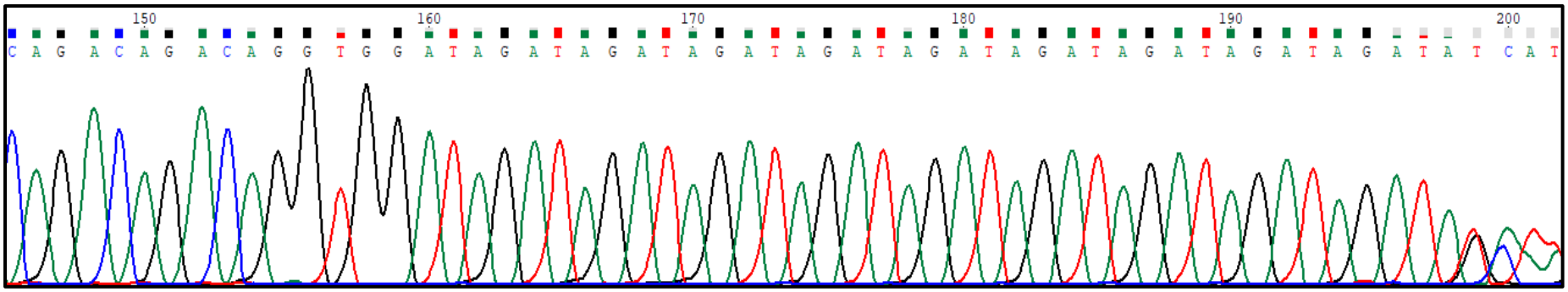
D16S539 locus



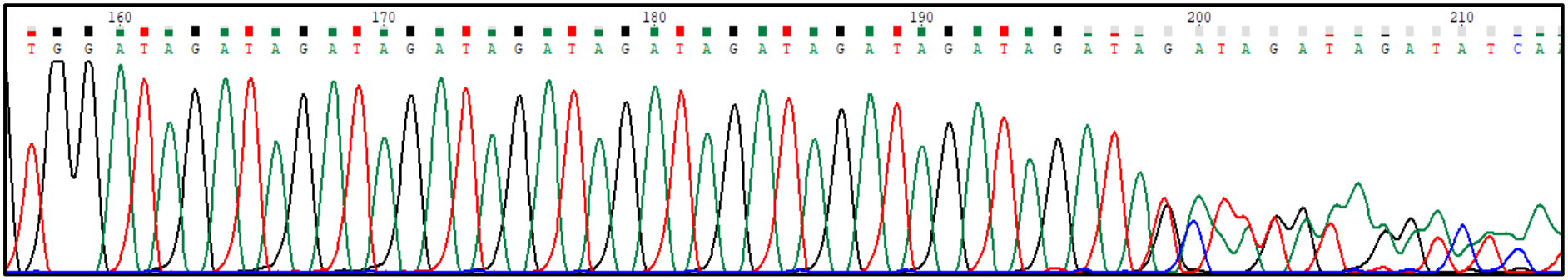
Case 1: Male child D16539 allele 9



Case 1: Male child D16539 allele 14

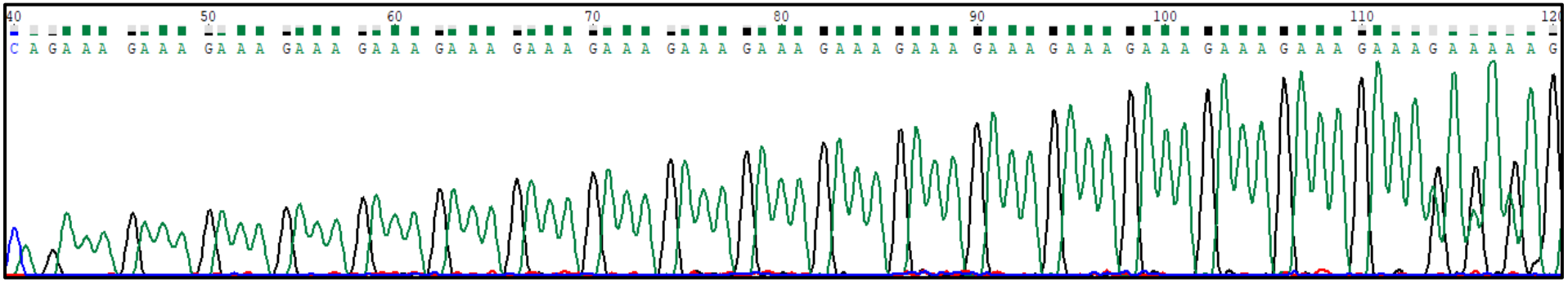


Case 1: Father D16539 allele 10

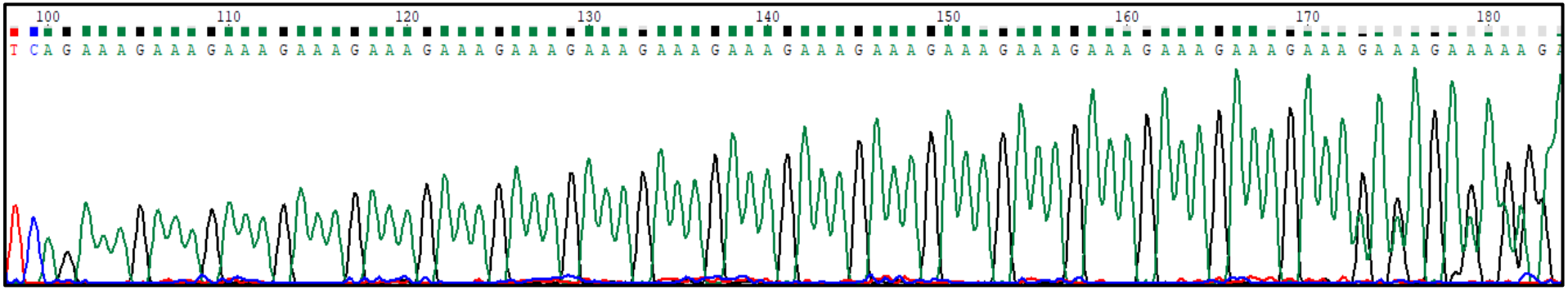


Case 1: Father D16539 allele 13

D18S51 locus

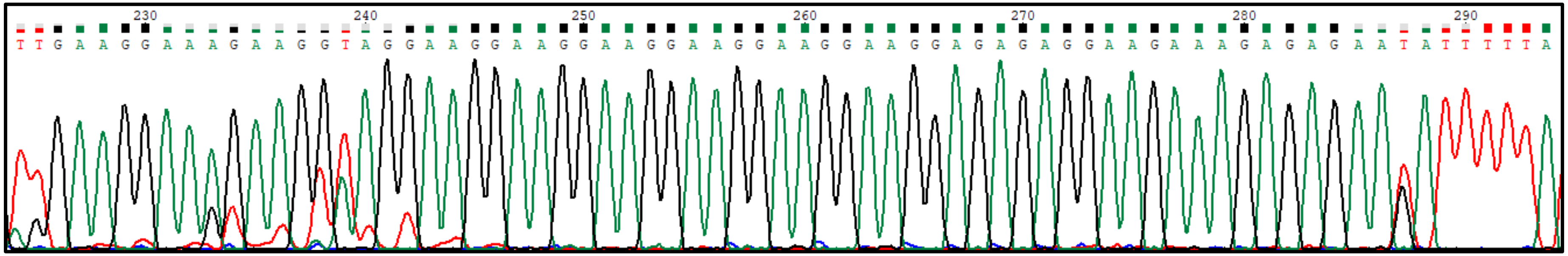


Case 1: Female Child allele 19

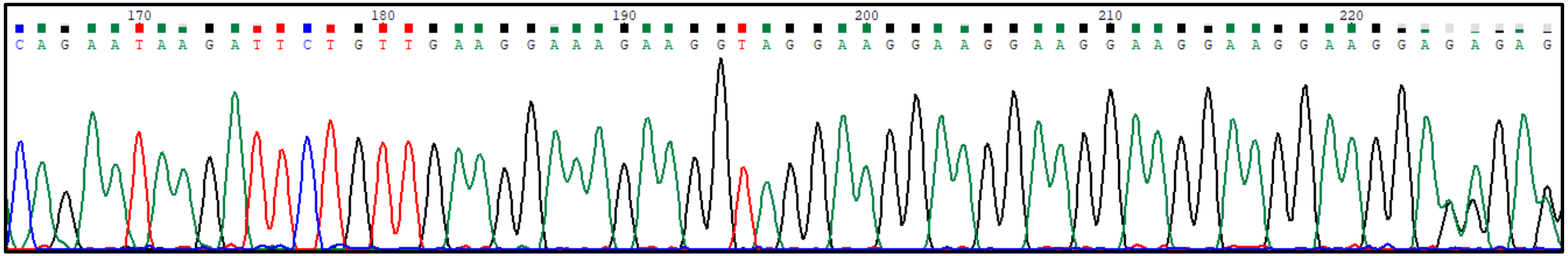


Case 1: Father allele 20

D19S433 locus

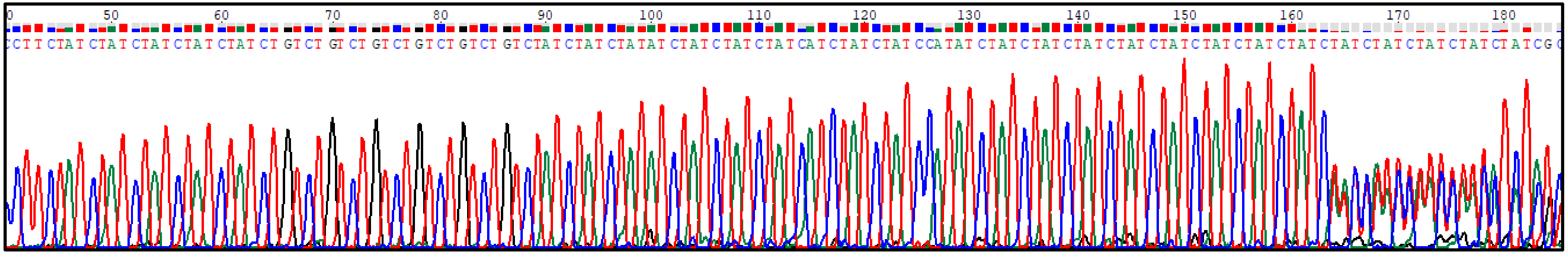


Case 1: Male Child D19S433 allele 8

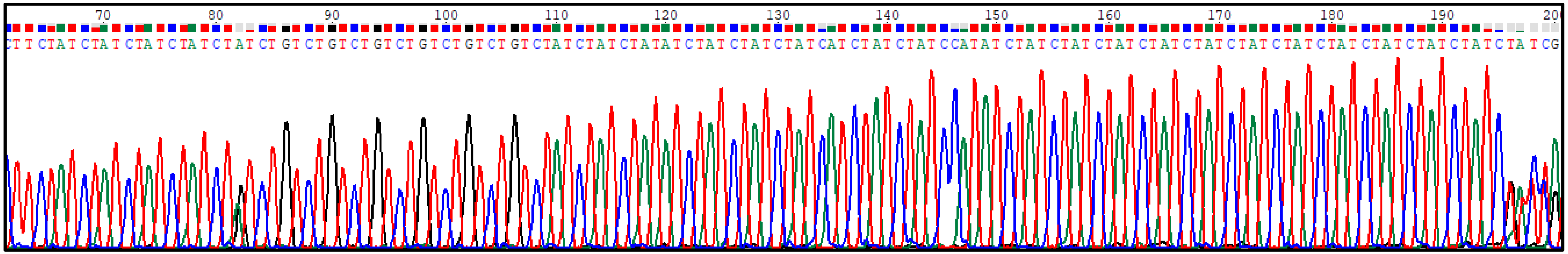


Case 1: Father D19S433 allele 8

D21S11 locus



Case 1: Female child D21S11 allele 32



Case 1: Mother D21S11 allele 31

Appendix E: HSREC Ethical Clearance letter

UNIVERSITY OF THE FREE STATE
UNIVERSITEIT VAN DIE VRYSTAAT
YUNIVESITHI YA FREISTATA

UFS·UV
HEALTH SCIENCES
GESONDHEIDSWETENSKAPPE

Health Sciences Research Ethics Committee

15-Dec-2021

Dear **Ms Afika Soldati**

Ethics Clearance: **Characterisation of apparent mismatches detected during routine Short Tandem Repeat (STR) analysis in parentage investigations**

Principal Investigator: **Ms Afika Soldati**

Department: **Haematology and Cell Biology Department (Bloemfontein Campus)**

Submission Page

APPLICATION APPROVED

Please ensure that you read the whole document

With reference to your application for ethical clearance with the Faculty of Health Sciences, I am pleased to inform you on behalf of the Health Sciences Research Ethics Committee that you have been granted ethical clearance for your project.

Your ethical clearance number, to be used in all correspondence is:**UFS-HSD2021/1815/2501**

The ethical clearance number is valid for research conducted for one year from issuance. Should you require more time to complete this research, please apply for an extension.

We request that any changes that may take place during the course of your research project be submitted to the HSREC for approval to ensure we are kept up to date with your progress and any ethical implications that may arise. This includes any serious adverse events and/or termination of the study.

A progress report should be submitted within one year of approval, and annually for long term studies. A final report should be submitted at the completion of the study.

Research conducted in any Department of Health facility: Researchers are required to sign and return the HSREC approval letters to the provincial Department of Health where they applied. It is also a requirement for researchers to submit electronic copies of their final research findings, and/or make a presentation of their findings and recommendations at departmental research days when and where indicated.

The HSREC functions in compliance with, but not limited to, the following documents and guidelines: The SA National Health Act. No. 61 of 2003; Ethics in Health Research: Principles, Structures and Processes (2015); SA GCP(2006); Declaration of Helsinki; The Belmont Report; The US Office of Human Research Protections 45 CFR 461 (for non-exempt research with human participants conducted or supported by the US Department of Health and Human Services- (HHS), 21 CFR 50, 21 CFR 56; CIOMS; ICH-GCP-E6 Sections 1-4; International Council for Harmonisation (ICH) Harmonised Guideline, Integrated Addendum to ICH E6(R1), Guideline for Good Clinical Practice (GCP) E6(R2), 2016, SAHPRA Guidelines as well as Laws and Regulations with regard to the Control of Medicines, Constitution of the HSREC of the Faculty of Health Sciences.

For any questions or concerns, please feel free to contact HSREC Administration: 051-4017794/5 or email EthicsFHS@ufs.ac.za.

Thank you for submitting this proposal for ethical clearance and we wish you every success with your research.

Yours Sincerely

Prof. A. Sherriff
Chairperson: Health Sciences Research Ethics Committee

Health Sciences Research Ethics Committee
Office of the Dean: Health Sciences
T: +27 (0)51 401 7795/7794 | E: ethicsfhs@ufs.ac.za
IRB 00011992; REC 230408-011; IORG 0010096; FWA 00027947
Block D, Dean's Division, Room D104 | P.O. Box/Posbus 339 (Internal Post Box G40) | Bloemfontein 9300 | South Africa
www.ufs.ac.za

Appendix F: Approval letter from the Head of Department of Haematology and Cell Biology

UNIVERSITY OF THE FREE STATE
UNIVERSITEIT VAN DIE VRYSTAAT
YUNIVESITHI YA FREISTATA

UFS·UV
HEALTH SCIENCES
GESONDHEIDSWETENSKAPPE

Head of Department Permission letter

The Chair: Health Sciences Research Ethics Committee
For Attention: Mrs M Marais
Block D, Room 104,
Francois Retief Building

22 October 2021

Dear Prof Sherriff,

Afika Soldati 2015173103
Project Title: Characterisation of apparent mismatches detected during routine Short Tandem Repeat (STR) analysis in parentage investigations

I hereby grant Afika Soldati permission to conduct the above-mentioned research project. The research will be completed in accordance with myself as Head of Department of Haematology, Dr A de Kock as supervisor, and Mr JF Kloppers as co-supervisor of this study.

Yours faithfully,

Dr Jaco Joubert
MBChB, MMed (Haem)
PGDip (Transfusion Medicine)
MP 0589055
Head: Haematology & Cell Biology
Speed Dial: 7104 / (051) 405 3043

Dr Jaco Joubert
Head: Haematology and Cell Biology
UFS and NHLS Universitas

Department of Haematology and Cell Biology/Faculty of Health Sciences
Departement Hematologie en Selbiologie/Fakulteit Gesondheidswetenskappe
T: +27(0)51 405 3043, F: +27(0)51 405 2923
205 Nelson Mandela Drive/Rylaan, Park West/Parkwes, Bloemfontein 9301, South Africa/Suid-Afrika
P.O. Box/Posbus 339 (G2), Bloemfontein 9300, South Africa/Suid-Afrika, www.ufs.ac.za

Appendix G: Approval letter from the Business Manager of the National Health Laboratory Service (NHLS), Universitas Academic Unit.

NATIONAL HEALTH LABORATORY SERVICE
Practice No. 5200296

Office of the Business Manager
Universitas Academic Laboratories
PO BOX 339(G3)
Faculty of Health Sciences
Chemical Pathology Block C
1st Floor, Office 301
University of Free State
Bloemfontein
9301

REQUEST FOR APPROVAL OF LABORATORY RESOURCES FOR ACADEMIC PURPOSES

Date: 18 October 2021

Requestor: Ms A Soldati

Project Name: **"CHARACTERISATION OF APPARENT MISMATCHES DETECTED DURING ROUTINE SHORT TANDEM REPEAT ANALYSIS IN PARENTAGE INVESTIGATION."**

Dear Ms A Soldati,

Your request for use of laboratory facilities / data is hereby granted under following conditions:

1) That University Ethical Committee approval and approval from the Universitas Hospital management is obtained
2) All laboratory data remain confidential to the patient and doctor (anonymity is maintained)
3) This Office must be notified before any publication of any results / findings are made.
4) NHLS is recognised in all publications

May your project be successful.

NATIONAL HEALTH LABORATORY SERVICE
2021 -10- 15
BUSINESS MANAGER
UNIVERSITAS ACADEMIC
LABORATORIES

Regards,

MR. PAKISO LETANTA
MANAGER (BUSINESS) - UNIVERSITAS ACADEMIC LABORATORIES