ANALYSIS OF SHORT TANDEM REPEAT LOCI FOR FORENSIC USE IN NAMIBIA

A MINI THESIS SUBMITTED IN PARTIAL FULFILMENT

OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE (INDUSTRIAL BIOCHEMISTRY)

OF

THE UNIVERSITY OF NAMIBIA

BY

ANN-ANGELIQUE LUKAS

201126737

SEPTEMBER 2020

SUPERVISOR: PROFESSOR MARTHA KANDAWA-SCHULZ (DEPARTMENT OF CHEMISTRY AND BIOCHEMISTRY, UNIVERSITY OF NAMIBIA)

Abstract

In the genetics section of the forensics laboratory, in Windhoek, Namibia, forensic DNA (deoxyribonucleic acid) typing has become an integral part of establishing familial relations as well as providing weight to evidence during criminal investigations. Currently, there is limited existing knowledge regarding the ethnic and racial genetic diversity of the Namibian population. In the present study, biological samples from 150 individuals from three racial groupings selected for the paper were used to generate a reference DNA database. The data was analysed at 21 short tandem repeat (STR) loci to observe the genetic variation between individuals of different racial groupings. In order to obtain the STR profiles, the Globalfiler Express kit from Life Technologies was used. The extracted samples were amplified and detected by the Genetic Analyser 3500 as STR profiles, with its recommended allelic ladder. Allele frequencies from the STR profiles were used to evaluate the forensic efficiency parameters such as the power of discrimination (PD), power of exclusion (PE) and the expected heterozygosity (He). Locus point SE33 exhibited the highest polymorphism and was the most informative during the study, whereas TPOX and D13S317 were the least informative. The findings in this study would be useful in assisting the Namibian legal system as well as other relevant bodies in human identification. The findings provided a method for characterizing the genetic structure of the Namibian population and suggested valuable information required to generate a standard online reference DNA database, thus motivate population studies on the Y-chromosome and mitochondrial DNA which could be useful in human identification of complex cases.

Keywords: forensics, human identification, loci, Globalfiler, DNA

Table of Contents

List of Tables

List of Figures

List of Abbreviations and/ or Acronyms

AMEL:	Amelogenin
BP:	Base Pair
CODIS:	Combined DNA Index System
CSF1PO:	Human c-fms Proto-Oncogene for CSF-1 Receptor Gene
DNA:	DeoxyriboNucleic Acid
FBI:	Federal Bureau of Investigations
HID:	Human Identification
LIZ-600:	Local Southern Sizing 600
MEC:	Mean Exclusion Chance
NFSI:	National Forensics Science Institute
PI:	Paternity Index
PCR:	Polymerase Chain Reaction
PD:	Power of Discrimination
PM:	Probability Match
POP4-polymer:	Performance Optimized Polymer 4
RFLP:	Restriction Fragment Length Polymorphism
STR:	Short Tandem Repeat
THO1:	Human Tyrosine Hydroxylase Gene
TPOX:	Human Thyroid Peroxidase Gene
TPI:	Typical Paternity Index
USA:	United States of America
vWA:	Human von Willebrand Factor Gene

Acknowledgments

I would like to express my sincerest gratitude to the Chief of the Genetics section at the National Forensic Science Institute in Windhoek, Namibia, Mrs. Maryn Swart for her expert, in depth technical and moral support throughout my studies. The permission granted by Mrs. Swart to conduct my studies at the National Forensic Science Institute laboratory and the extensive hours that we spent working on the data analysis, has taught me that patience, consistency and perseverance does in fact allow one to accomplish their greatest desires.

Gratitude is also extended to my supervisor from the Department of Chemistry and Biochemistry at the university, Professor Martha Kandawa-Schulz for her continuous guidance. Thank you for words of encouragement, proposing solutions and for reviewing all my drafts over the years.

In addition, I am grateful for the ongoing assistance of independent statistics consultants, Mr. Jason Nakaluudhe and Mr. Naphtal Sakaria of Etameko Research Consultancy cc. Thank you for the collaborative efforts with the Information Technology specialist from Prestige Printing cc and the application specialists from qualitype GmbH.

Furthermore, I would like to acknowledge the Scene of Crime Officers of the Namibian Police force, as well as my colleagues in the Genetics section for their unlimited support during the entire duration of my studies. Lastly, to my closest friends and family, thank you, thank you so much for inspiring discussions during this time, I have truly come to appreciate it.

Dedication

This thesis is dedicated to my late grandfather, Sekulu Basilius Basia Lukas. Thank you for always providing me with opportunities to better myself and improve my skills set. Your strong presence in my life has taught me never to limit myself to my immediate surroundings.

Declaration

I, Ann-Angelique Lukas, hereby declare that this study is my own work and is a true reflection of my research, and that this work, or any part thereof has not been submitted for a degree at any other institution.

Date…….

Signature…….

CHAPTER 1

INTRODUCTION

1.1 Background of the study

In the forensic deoxyribonucleic acid (DNA) laboratories, DNA profiling has become an increasingly important aspect of criminal investigations, sibling relations, mass disasters, maternity and paternity testing [1]. Population genetic principles are applied to infer valuable information used to identify individuals that may have contributed to a specific biological sample [1]. The genetic make-up of human beings is unique to individuals; it is known as the deoxyribonucleic acid (DNA) of the person [2]. Within the field of forensic science, good sources of human DNA for identification purposes are biological samples such as blood, saliva, semen and sweat. It is possible to use these biological samples in efforts to identify genetic variations among individuals.

A practical approach to human identification (HID), would be one in which an individual can be either excluded or included in a biological sample whereby there is no chance that another person can yield the exact same DNA profile. However, the exception is made for cases involving identical twins.

Studies have shown that a technique called Short Tandem Repeat (STR) Analysis is a highly reliable method used in forensic DNA typing [3]. Researchers have identified core repeat units in the human genome for use as genetic markers [4]. These genetic markers are known as STR alleles. Forensic DNA profiles can be obtained from biological samples through these core repeat units using a specified set of STR loci. STR alleles are defined as units of repeated DNA sequences inside the human genome [2]. The repeated sequences

are of varying lengths and sizes, dependent on the number of tandem repeats of base pairs in a DNA fragment [5]. These STR regions of repeats are highly specific to individuals at a locus, enabling scientist to differentiate between human beings. The frequency at which the STR regions of interests occur in human beings must be as small as possible in order to eliminate results that do not reflect the true DNA profile [6]. The occurrence of STR alleles can be dependent on the general population, therefore, it is important to detail the genetic data of a particular population in efforts to increase the discriminating power of the DNA profiles before generating a reference database with all possible allele patterns represented by a given population. A rarity of certain alleles within a population can prove useful in identifying contributors of a biological sample [7]. This highlights how valuable it is to study genetic variations specific to the country's population, in efforts to directly influence future forensics investigations.

1.2 Statement of the Problem

To date, the forensic DNA laboratory in Namibia has observed an alarmingly increased rate of unsolved cases, especially that of rape and murder. A review of the laboratory's registry for DNA case submissions, observed a total of three thousand seven hundred and fourteen (3714) outstanding cases thus far. The backlog of the cases spans from as far back as 1993. These figures can be seen in Appendix 1. Many court proceedings in the criminal justice system, utilize biological evidence from crime scenes in efforts to draw conclusions in a criminal case. The absence of DNA results may delay the judgement of the case proceedings. Since STR alleles are highly specific to individuals, the frequency at which they differ can be measured. The allele frequency estimates generated from the

reference database, can be useful in identifying the perpetrator of a serious crime and or contributors to a particular biological sample, such as those demonstrating familial relations.

1.3 Objectives of the Study

The objectives of the study were:

a) To generate a DNA reference database representing three racial groupings in Namibia; black, white and coloured
b) To analyse biological samples from the reference database at 21 STR loci
c) To determine/analyse/characterize the variations between individuals at 21 STR loci (allele frequencies)

1.4 Significance of the Study

Reference databases can be used in investigations where the perpetrator of a serious crime is unknown. The uniqueness of STR alleles present in a biological sample can help link a suspect to a serious crime. The more similarities there are between STR profiles derived from a biological sample found at a scene in comparison to that from the reference database, the less likely the contributor (s) of said sample is from a random person, unrelated to the crime.

Furthermore, the population database can be used to generate an online DNA database for determining general inclusive probabilities. There is very little information regarding genetic studies specific to the Namibian population. It is important to note that allele

frequency estimates will help reduce the backlog in the forensic DNA laboratory as more cold cases and disaster victim investigations will have a better chance of possibly being concluded.

1.5. Limitation of the Study

Due to financial constraints, the study focused on biological samples that were brought forward to the National Forensic Science Institute (NFSI) and already in the databank, instead of casework that was yet to be submitted throughout the course of the year by officers of the Namibian Police force, Scene of Crime Unit. The biological samples gathered from the NFSI databank, reliably reduced the transport costs that could have been incurred during sampling, allowing resources to be redirected towards scientific analysis in the laboratory.

1.6 Delimitation of the Study

Although Namibia has a diverse population, many people identify themselves as mixed race. For the purposes of this study, the data to be collected was focused on three racial groupings instead of the various ethnic groups in the country. The racial groupings can help to narrow down the search for suspects of a specific crime. Inclusion of the various ethnic groupings in the reference database for this study would have been time consuming and require extensive operational support. Efforts to procure the consumables that were used to conduct the research proved challenging throughout the study. More time was allocated to interpret the allelic peaks in order to produce the genotype table. This increased the length of the study.

In addition, there was limited access to the software required to transfer the results from the genotype table onto worksheets representing the allele frequencies. Some of the data entry had to be done manually. It took several weeks of liaising with the qualified application specialists to acquire and gain access to the statistical software required to calculate the forensic efficiency parameters. Time had to be allocated to familiarize with the user interface design software once this access was granted.

Furthermore, the NFSI databank had a limited number of biological samples from the Caucasian racial grouping. The Caucasian racial grouping, only eighteen (18) out of the one hundred and fifty (150) biological samples that were required for the study, hence this group was excluded for all other purposes throughout the research conducted as it is insufficient to truly represent the most common alleles in this subset of the population.

CHAPTER 2

LITERATURE REVIEW

2.1 Human identification

The identification of individual human beings forms an integral part of the National Forensic Science Institute in Namibia. Over the years, scientists have discovered various methods to tell human beings apart at a genetic level [8]. Forensic DNA typing is based on the principle that the human genome is made up of repeating DNA sequences [9]. These repeated DNA sequences, are highly variable regions of sequences within a person's DNA termed as short tandem repeat (STR) units and can be used as genetic markers. The genetic markers are loci that are used to investigate how humans differ from one another [10], without affecting their genetic health [9]. The ability of STR alleles to vary in length and sizes of base pairs have made it possible for STRs to contribute towards studies relating to genetic diversity in human populations.

2.2 The history of DNA typing methods

Earlier experiments by scientists in the field of forensic DNA analysis demonstrated the use of a variety of techniques to distinguish between human beings at a genetic level [11]. Figure 1 shows the technological advances of forensic DNA typing over the past three decades.

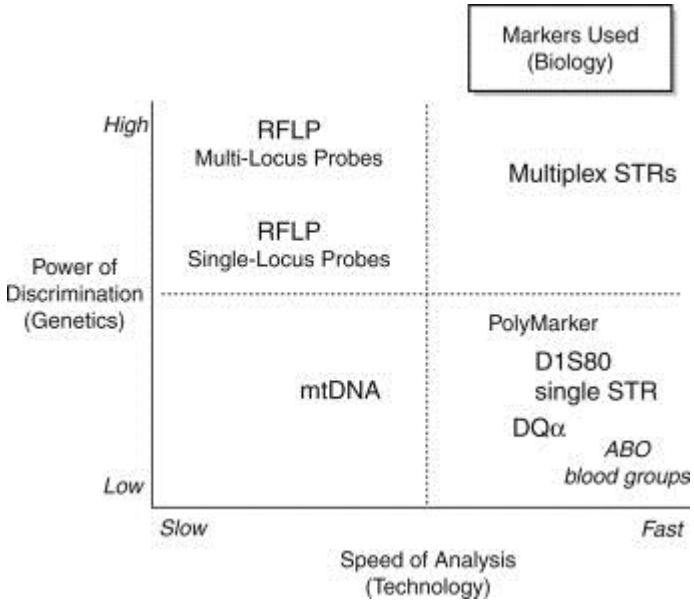

Figure 1: Comparison of DNA typing technologies over the past three decades.

Source: Butler JM, [11].

Based on the information in figure 1, the forensic DNA markers are plotted in relationship to four quadrants defined by the power of discrimination for the genetic system used and the speed at which the analysis for that marker is performed. Note that this diagram does not reflect the usefulness of these markers in terms of forensic cases. In the early 1900s, the four different blood groups, A, B, AB and O were used in laboratories for human identification. Even though the blood group system proved useful in identifying individuals, it was not very informative when individuals possessing the common blood types A and O were part of forensics investigations [11]. A few years later, the need to generate more effective techniques, led to discoveries by Alec Jeffrey's [11] based on DNA finger printing. The first decade of DNA fingerprinting (1985-1995) used multi-focus probes to visualize DNA patterns unique to individuals [11]. After the discovery of restriction fragment length polymorphism (RFLP), the sensitivity of DNA typing

increased considerably leading to reports of short tandem repeats using various PCR (polymerase chain reaction) methods in the mid-1990s [11].

The studies conducted around STR and PCR sensitive analysis in the 1990s by United Kingdom scientists, enabled investigators to produce the first-generation-quadruplex STR markers; TH01, vWA, FES/FPS and F13A1 [12]. The quadruplex system, was followed by a second-generation-multiplex STR system consisting of six genetic markers; TH01, vWA, FGA, D8S1179, D18S51 and D21S11, including the sex-determining marker Amelogenin (AMEL) [13].

2.3 The types of STR markers and the locus nomenclature

It is important that the genetic markers used in forensic DNA typing can discriminate between individuals. In order to perform STR analysis, the locus on the human chromosome exhibiting the highest polymorphic regions must be determined. The repeating STR units are comprised of 1-6base pairs (bp) that total approximately 3% of the human genome [14]. These bp's have a high mutation rate thus contributes towards genetic diversity in human populations [14].

The inherent variations in DNA sequences among human beings are named by the length of the STR repeat units. A study conducted by Urquhart and colleagues [15] emphasized that the type of STR is dependent on the varying length of the repeat units. Dinucleotides repeats possess two nucleotides next to each other. Trinucleotides have three repeat units, tetranucleotides have four nucleotides, pentanucleotides have five, and hexanucleotides have six STR repeat units in the core repeat [15]. There are various categories that exists

for the STR units which are not limited to their varying lengths instead the rigor with which the STR repeat units form in a pattern [15].

A simple repeat such as that from the genetic marker TPOX, contains DNA fragments of identical lengths and sequence [15]. In addition, compound repeat units comprise two or more adjacent simple repeats and the complex STR locus such as D21S11 may contain several repeating units with intervening sequences. Furthermore, the existence of complex hyperalleles that differ in both size and sequence have numerous non-consensus alleles: these alleles may be challenging to genotype reproducibility.

The last category of STR marker is a complex hypervariable locus; SE33. It is not commonly used by scientists for forensic DNA typing due to challenges with the STR nomenclature and measure of variability between DNA laboratories [15]. Table 1 depicts the different types of categories for the core locus points required for STR analysis:

Table 1: The categories for STR markers.

Category	Example of repeat structure	Example of 13 STR core loci
Simple repeats	(GATA)(GATA)(GATA)	TPOX, CSF1PO, D5S818, D13S317, D16S539
Simple repeats with non-consensus alleles	(GATA)(GAT-)(GATA)	TH01, D18S51, D7S820
Compound repeats	(GATA)(GATA)(GACA)	VWA, FGA, D3S1358, D8S1179
Complex repeats	(GATA)(GACA)(CA)(CATA)	D2S11

According to Butler [16], some of the STR nomenclature is based on the parts of a gene in which the genetic marker (STR allele) falls. For example, the name of the STR marker TH01 is derived from the human tyrosine hydroxylase gene located on chromosome 11. The '01' portion of TH01 comes from the fact that the repeat region in question is located within intron 1 of the tyrosine hydroxylase gene. Sometimes the prefix HUM- is included at the beginning of a locus name to indicate that it is from the human genome. Thus, the STR locus TH01 would be correctly listed as HUMTH01.

Butler [16] further emphasized that some genetic markers may fall outside the DNA regions and may be designated by their chromosomal position. An example would be for the locus point D3S1358, whereby D represents DNA, 3 signifies chromosome position,

the letter S mentions that it is single copy sequence and finally the order in which the marker was discovered and categorized for a particular chromosome; 1358.

2.4 The Combined DNA Index System (CODIS)

With the success of STR markers in the years following the discovery of second-generation-multiplex-STR system, the Federal Bureau of Investigations (FBI) laboratory in United States of America (USA) established a 13 core STR loci database in 1998 [17]. The 13 core loci were used to launch the combined DNA index system (CODIS), a database representing laboratories from across the world with the genetic markers; CSF1PO, FGA, TH01, TPOX, vWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S359, D18S51, and D21S11 [17]. A report by the European Network of Forensic Science Institute in April of 2009, voted to adopt an additional five STR loci [1]. It is important to note that an expansion of the STR loci used in DNA typing improves probability test results and can be useful for DNA profiling in more specific populations.

Since 2009, several genetic markers have been identified in efforts to improve the CODIS system. Yang and colleagues [18] have demonstrated that a variety of polymorphic markers inside the human genome can display a greater power of discrimination in genetic studies. Furthermore, the launch of CODIS software system, has motivated scientists to develop and validate commercially available STR kits, with expanded STR markers for robust and highly sensitive PCR analysis.

2.5 Commercial kits and studies in different countries

The development of commercially available STR kits allows for quality control and enables DNA data sharing regionally as well as nationally [19]. Currently, commercially available STR kits perform PCR on 15-24 of the most common STR loci. However, researchers have been verifying kits that expand on CODIS in efforts to include all world populations. The STR kits come readily prepared with the premixed reagents, including sample controls that target the specific regions of interest in the human genome [19].

Core Loci	Currently Available Commercial STR Kits		
	MiniFiler	Identifiler	Powerplex 16
D3S1358		D3S1358	D3S1358
D5S818		D5S818	D5S818
D7S820	D7S820	D7S820	D7S820
D8S1170		D8S1170	D8S1170
D13S317	D13S317	D13S317	D13S317
D16S539	D16S539	D16S539	D16S539
D18S51	D18S51	D18S51	D18S51
D21S11	D21S11	D21S11	D21S11
CSF1PO	CSF1PO	CSF1PO	CSF1PO
FGA	FGA	FGA	FGA
THO1		THO1	THO1
TPOX		TPOX	TPOX
vWA		vWA	vWA
	D2S1338	D2S1338	Penta D
		D19S433	Penta E
	Amelogenin	Amelogenin	Amelogenin

Figure 2: **Information on various commercially available STR kits [20].**

There have been several STR amplification kits in the past as depicted by the figure above. Some developed countries such as Korea use 13 STR loci for human identification purposes, these loci include most of the major allelic markers used to distinguish between individuals [21]. In Cape Town, South Africa, most DNA typing has been performed on

17 STR loci thus far [22]. Researchers in the USA investigated a Spanish population using 15 STR loci with the identifiler commercial STR kit [23].

In 2014, Life Technologies developed the Globalfiler kit for amplification in human identification [24]. Namibia recently phased out the Identifiler kit and introduced the Globalfiler amplification kit for forensic use. The Globalfiler STR kit amplifies 21 loci along with one Y-STR (DYS391) on the Y-indel chromosome and the sex-determining AMEL marker [24]. The Globalfiler amplification kit was developed as a 6-dye kit in order to amplify more loci in efforts to produce sufficiently amplified DNA for full human DNA profiles during polymerase chain reaction (PCR) process [24].

2.6 Variant alleles

Polymerase chain reaction can produce high levels of discrimination between human beings when using the Globalfiler amplification kit [25]. However, the amplified products from the PCR can yield artefacts; some highly variable STR alleles at the selected loci [26] in comparison to the recommended allelic ladder. The loci may yield allele sequences which differ by one or more base pairs at certain STR markers. These incomplete repeat units which often interfere with the interpretation of STR profiles and generation of the genotype tables are referred to as microvariant alleles [27].

CHAPTER 3

MATERIALS AND METHODS

3.1 Sample Collection and Preparation

The population database for this study was generated by collecting 150 biological samples from the National Forensic Science Institute databank. The samples gathered were selected in the hopes that there were no familial relationships between them, in efforts to reliably represent the most common alleles. The application for scientific examinations forms associated with every biological sample, was carefully reviewed to draw conclusions from the case summary that the samples were in fact unrelated. Each of the individual samples were given a unique identifier number marked in duplicates on clean, dry Eppendorf tubes. These unique codes were used throughout the study in order to prevent personal information from being tracked back to the biological sample donor.

Some of the biological samples that were used during the study were already extracted and stored on the computer software, ArmedXpert. This data was sampled and then characterized accordingly; that is into the appropriate racial grouping.

3.2 DNA Extraction

The biological samples provided for this study were extracted from either blood or buccal swabs using the Prep-n-Go extraction method in accordance with the Manufacturers guidelines. All possible measures were taken to maintain a clean, safe, sterile working environment.

Each biological swab was detached into an individually labelled Eppendorf tube. This was followed by the addition of 400µl of the Prep-n-Go buffer to each tube. Prep-n-Go buffer direct PCR method has high data reproducibility and minimizes the chances of artefacts in the biological sample thus yielding more interpretable peaks [28]. In addition, Prep-n-Go buffer assists with cell lysis, thus exposing the DNA inside the nucleus [28]. After adding the Prep-n-Go buffer solution, the tubes were then incubated at 90ºC for 20minutes on a preheated HLC thermal shaker. Afterwards, they were cooled for 15minutes at room temperature. The sample lysates were then transferred to different sterile, dry tubes labelled accordingly. The sample heads were discarded accordingly.

3.3 PCR Amplification

Amplification of the DNA extracts was carried out in duplicates by using the Life Technologies, Globalfiler Express Amplification kit. Firstly, a volume for each of the reaction components was calculated. The total PCR reaction volume consisted of: 6µl master mix, 6µl primer and then 3µl of the extracted DNA. It is important to note that the Globalfiler Express reaction mix was prepared in an Eppendorf tube and labelled accordingly. This was done in order to minimize the pipetting steps, thus reducing the chances of contamination between the individual samples.

Secondly, 12µl of the reaction master mix was dispensed into a MicroAmp Optical 96-well reaction plate. The plate was kept on a plate holder in order to prevent the bottom of the wells from being scratched: this would otherwise have affected the PCR product. Thirdly, 3µl of every DNA sample extract was dispensed into the MicroAmp Optical 96-well reaction plate, making up a final PCR volume 15µl. Afterwards, the positive and

negative controls were added in the appropriate wells based on the number of PCR cycles. The reaction plate was then covered with a MicroAmp optical adhesive film, vortexed for 3seconds on a mix mate and spun in a table top centrifuge at 3000rpm for 20seconds.

The PCR process was then performed in a GeneAmp PCR system 9700 from Life Technologies. The thermal cycling conditions for the Globalfiler Express kit were programmed as depicted in table 2 according to the manufacturer's guidelines:

Table 2: Thermal cycle conditions to perform PCR

Initial Incubation Step	**Cycle (28 or 29 cycles)**		**Final Extension**	**Final Hold**
	Denature	**Anneal/Extend**		
HOLD	Cycle		HOLD	HOLD
95°C 1minute	94°C 3seconds	60°C 30seconds	60°C 8minutes	4°C up to 24hours

All the extracted DNA were amplified at 24STR loci with the Globalfiler Express Amplification kit. The 24STR loci amplified in this study were D3S1358, vWa, D16S539, CSF1PO, TPOX, D8S1179, D21S11, D18S51, D2S441, D19S433, TH01, FGA, D22S1045, D5S818, D13S317, D72S820, SE33, D10S1248, D1S1656, D12S391 and D2S1338. The remaining genetic markers were the Y-STR, Y-INDEL and amelogenin (AMEL). Amelogenin is the sex determining allele.

3.4 STR Typing

Upon completion of the PCR cycles, the amplified products were subjected to capillary electrophoresis. Capillary electrophoresis was carried out by the Genetic Analyser 3500.

The internal standard LIZ-600 (Local Southern Sizing) (figure 3) was used to prepare a master mix in order to run the PCR products on the Genetic Analyser 3500 along with a POP-4 polymer. An internal standard is labelled with fluorescent dye that allows for the detection and interpretation of unknown peaks at a specified threshold [29].

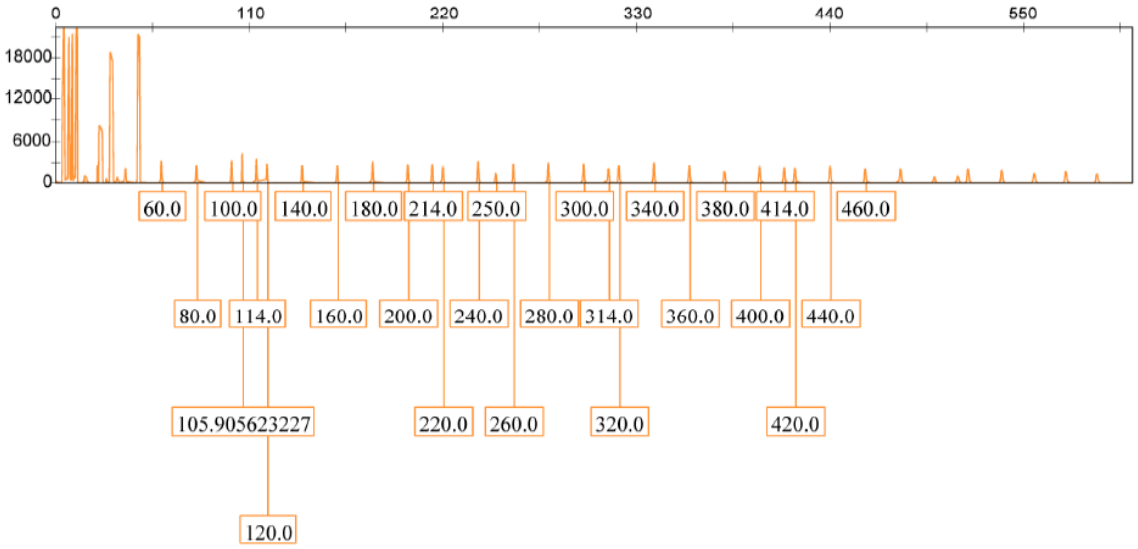

Figure 3: **An illustration of an electropherogram showing the fragments of the internal standard LIZ-600.**

The Genetic Analyser 3500 displayed the DNA fragments in the form of peaks that correspond to the various STR alleles amplified at the 24 locus points. The peaks representing the extracted DNA fragments were identified and associated with the appropriate fluorescent dye colour. The dye labelled DNA fragments of the DS-36 Matrix Standard (Dye Set J6) were 6-FAM, VIC, NED, SID, TAZ and LIZ. These DNA fragments were then internally sized in comparison with the Globalfiler Express allelic ladder that was added to the 96-well reaction plate, prior to the run. The Globalfiler Express allelic ladder was added to the 96 well plate according to Life Technologies

instructions for the biological samples. The allelic ladder provided a standard reference for the DNA sizes of each of the common alleles on the markers, in order to ensure consistent and precise allele designations [30].

GeneMapper ID-X (Life Technologies) was used to translate the fragments into full STR profiles in an electropherogram that was later converted to a genotype table representing all the allele combinations on the STR markers per sample analysed.

3.5 Statistical Analysis

The STR profiles obtained from the genotype table during data analysis were organized into the appropriate racial groupings using Microsoft excel software. Before the genetic parameters were applied, the data collected from the genotype table were used to determine the allele frequencies by counting the number of times each of the alleles were observed in the racial grouping: 150 for black, 150 for coloured's, the Caucasian group was not included in the study due to insufficient number of samples in the NFSI databank. A sample size of more than one hundred is a reliable representation for a population subset in population genetic studies [1].

For the forensic efficiency parameters, GenoProof version 3.0.7 (qualitype GmbH, Dresden, Germany) was used. The parameters included the observed heterozygosity (Ho), power of discrimination (PD), probability of exclusion (PE), polymorphism information content (PIC) and the mean exclusion chance (MEC). The same software was used to collect data on the expected heterozygosity and the Hardy-Weinberg equilibrium (p-value).

Following are some of the formulae that were used to calculate forensic efficiency parameters for the data analysis. Adopted from Stephenson [31].

3.5.1 The power of discrimination

The power of discrimination (PD) is equivalent to 1 minus the product of the match probability (PM) of all the markers. The formula used to determine the random match probability was however adopted from methods devised by Fisher [32]. The power of discrimination highlights the probability that two individuals will not have the exact same allele designations [33]:

$$PD=1-PM$$

3.5.2 Paternity Index

This parameter describes the probability of paternity (PI). It is calculated as

$$PI=X/Y$$

Where X is the probability the alleged father, is in fact, the father and Y is the probability that any randomly selected man of the same race is the father.

Furthermore, Edward and colleagues [34] derived a formula to calculate the expected heterozygosity as follows:

3.5.3 Expected heterozygosity

$$H=\frac{n}{(n-1)}\left[1-\sum_{j=1}^{k}\left(n_j/n\right)^2\right]=\frac{n}{(n-1)}\left[1-\sum_{j=1}^{k}\left(p_j\right)^2\right]$$

Whereby the n1, n2…nk are the number of alleles at a genetic marker in a sample under investigation and $p_{j\ is}$ the value of the allele frequency initially calculated.

3.5.4 Power of exclusion

In addition, the forensic efficiency parameter called the power of exclusion is the probability that a person selected at random is not inclusive of the alleles present at a locus. The following formula describes this probability:

$$\text{Power of Exclusion (PE)} = H^2(1-(1-H)H^2)$$

Whereby H represents the heterozygosity.

Adopted from Botstein D and colleagues [35].

CHAPTER 4

RESULTS

The DNA population database was compiled of 300 biological samples analysed at the NFSI laboratory. Forensic DNA typing was carried out with the Globalfiler Express PCR Amplification kit at 21 STR loci.

4.1 An example of the allele counts for STR locus

The Globalfiler STR profiles produced with the Genetic Analyser 3500 were genotyped for the 24 STR loci. Figure 4 is an indication of the allele designations that were observed after STR analysis. The allele designations from the STR profiles were used to generate the genotype tables using the GeneMapper ID-X software.

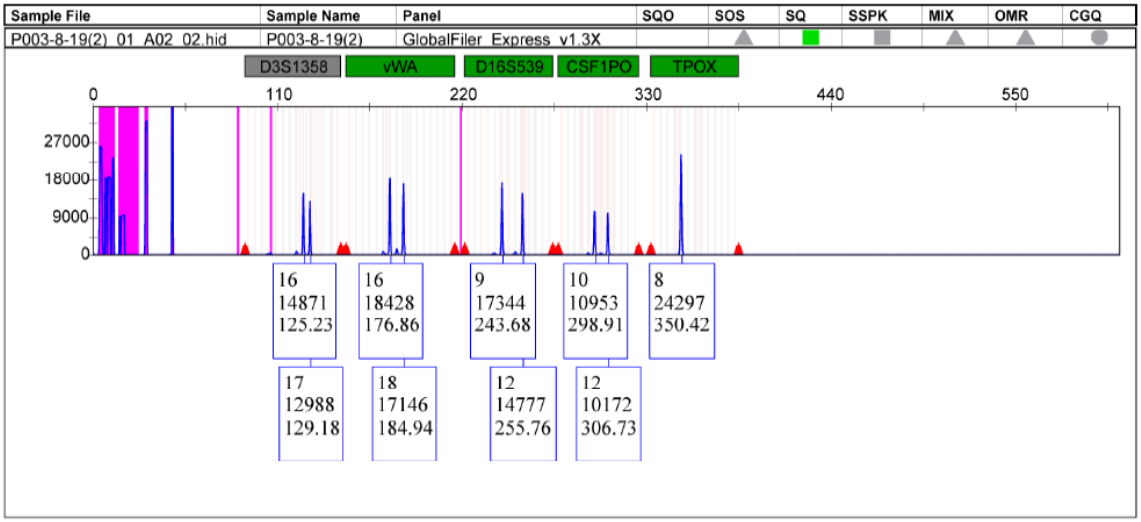

Figure 4a: **STR profile of sample P003-8-19 (2) amplified using the Globalfiler Express amplification kit.**

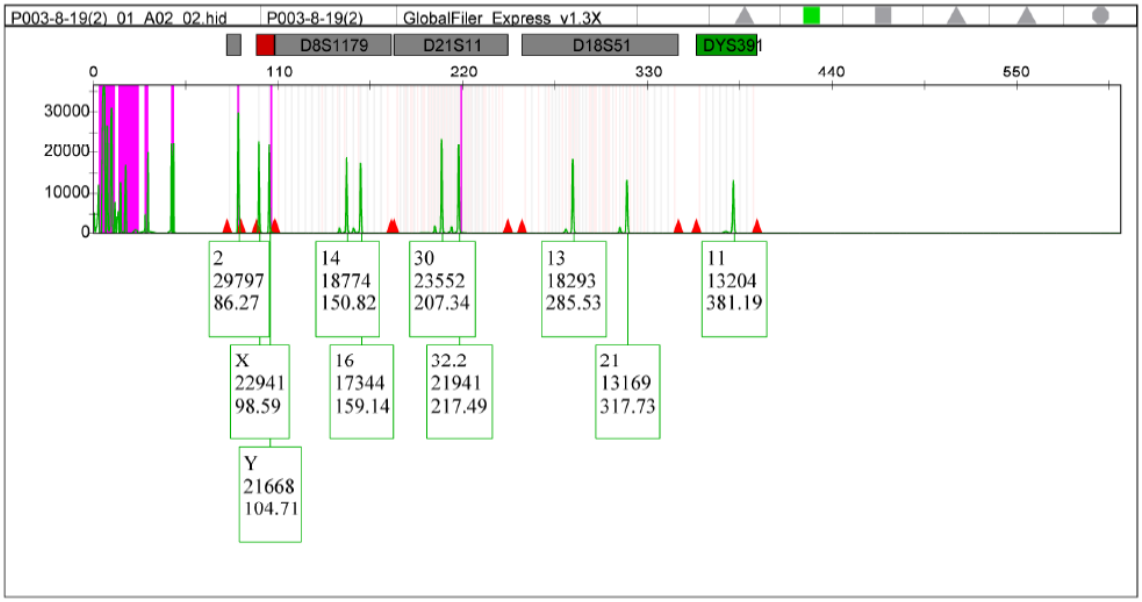

Figure 5b: **STR profile of sample P003-8-19 (2) amplified using the Globalfiler Express amplification kit.**

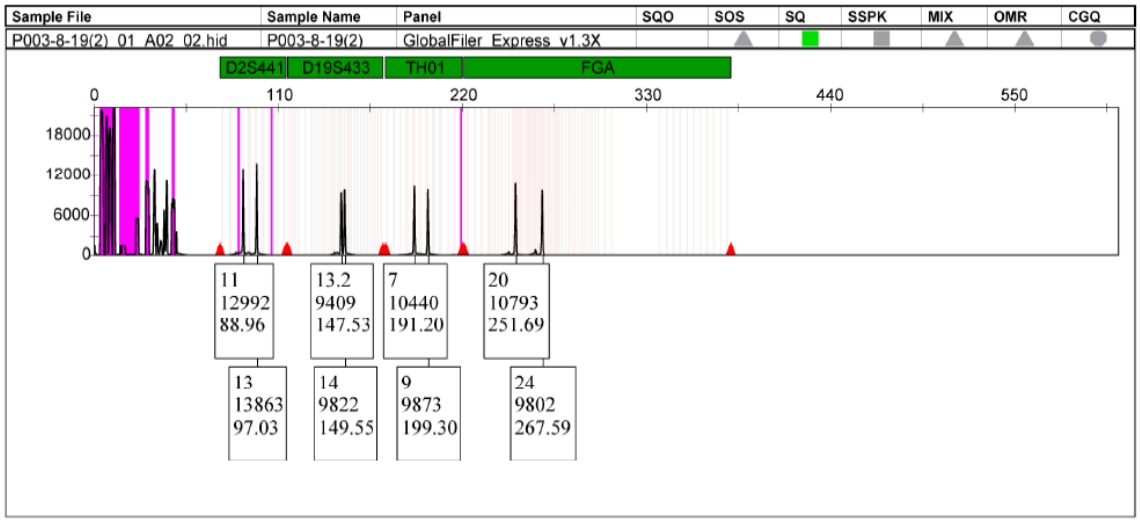

Figure 6c: **STR profile of sample P003-8-19 (2) amplified using the Globalfiler Express amplification kit.**

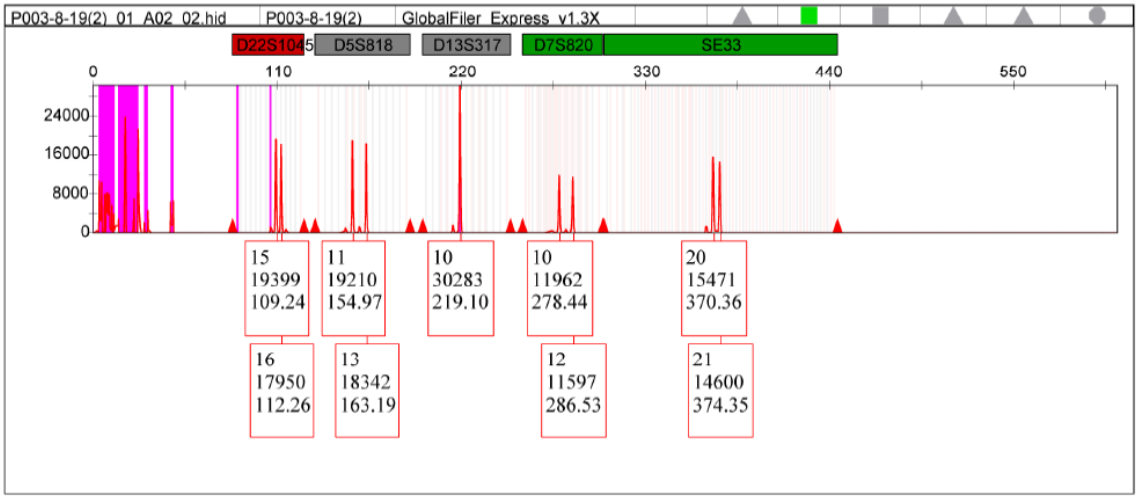

Figure 7d: **STR profile of sample P003-8-19 (2) amplified using the Globalfiler Express amplification kit.**

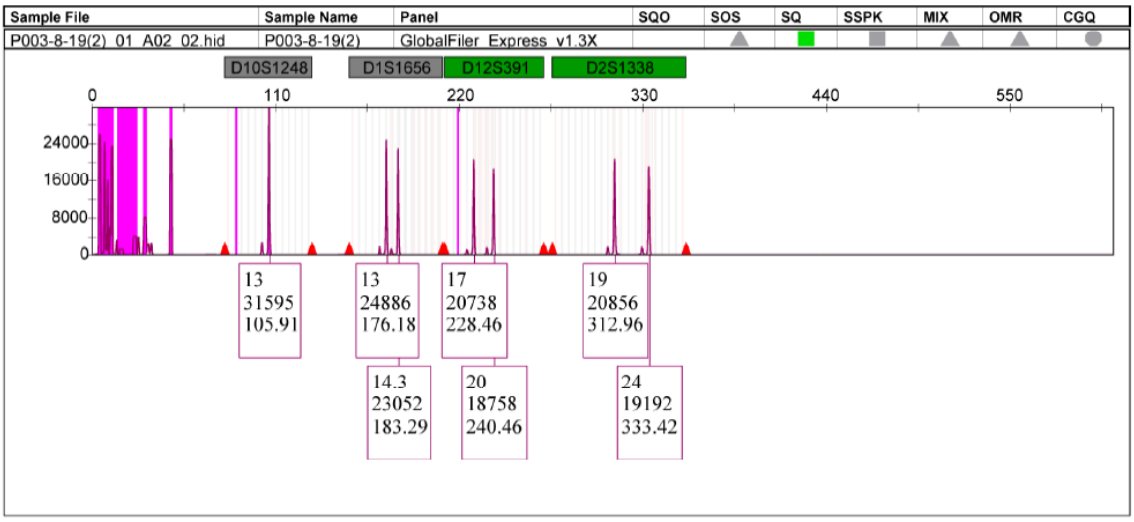

Figure 8e: **STR profile of sample P003-8-19 (2) amplified using the Globalfiler Express amplification kit.**

In figure 4 above, the electropherogram shows all the reportable loci for the sample examined. For every peak indicated, the three observed values are the; allele designations, the peak height and the base pair value for the marked peak. The GeneMapper ID-X

software recovers all the alleles in an interpretable source table. Information from the source table referred to as the genotype tables was converted to allele frequencies by counting the number of alleles observed at each of the set markers. Table 3 shows an example of the possible allele combinations and the number of alleles counted therein at STR marker D10S1248.

Table 3: Allele patterns, allele counts and allele frequency for STR marker D10S1248 from unrelated coloured individuals (sample size 150 or 300 chromosomes measured).

D10S1248	7	11	12	13	14	15	16	17	ALLELE COUNT	OBSERVED FREQUENCY
7	0	0	0	1	0	0	0	0	1	0.00333
11		0	0	4	2	1	3	0	10	0.03333
12			4	7	4	5	7	0	31	0.10333
13				11	27	17	12	1	91	0.30333
14					11	10	5	3	73	0.24333
15						5	4	2	49	0.16333
16							2	2	37	0.12333
									300	

Table 3 depicts the observed patterns of allele combinations for the marker D10S1248 listed above and across the table. At the intersection of the rows and columns in the table are the number of genotypes, for instance the allele combinations that were observed for the marker. The allele count was obtained by summing the columns and rows containing the alleles of interest. For example, the genotype array for 7/13 was observed once in the said marker, that of 13/13 was seen 11 times in the 150 individuals for the coloured racial grouping, but no allele combinations were observed for 12/17. The allele count for 13/13 genotype would thus be equivalent to 91. The value was obtained by summing the number

of chromosomes containing allele designation 13 for D10S1248; followed by the observed allele frequency by dividing the allele count (91) by the total number of chromosomes 300. The number of chromosomes for the study per racial grouping was said to be 300 since there are two chromosomes for each of 150 individuals.

4.2 Allele frequencies and genotypes for the racial groupings

The various allele patterns that were observed in the population for the black and the coloured racial groupings, along with the observed allele frequencies are shown in Appendix 2 and Appendix 3, respectively.

4.3 Heterozygosity

The results obtained for the observed heterozygosity and the expected heterozygosity can be seen in table 4 as average calculations for the coloured and black racial groupings. **Furthermore, reportable results that were detected for the behaviour of alleles can be seen in figure 5 and figure 6**. The number of homozygotes (h) plus the number of heterozygotes (H) equals 100% of the samples tested. Thus, since h +H= 1, then H = 1 − h and h = 1 −H. A higher heterozygosity means that more allele diversity exists and therefore there is less chance of a random sample matching.

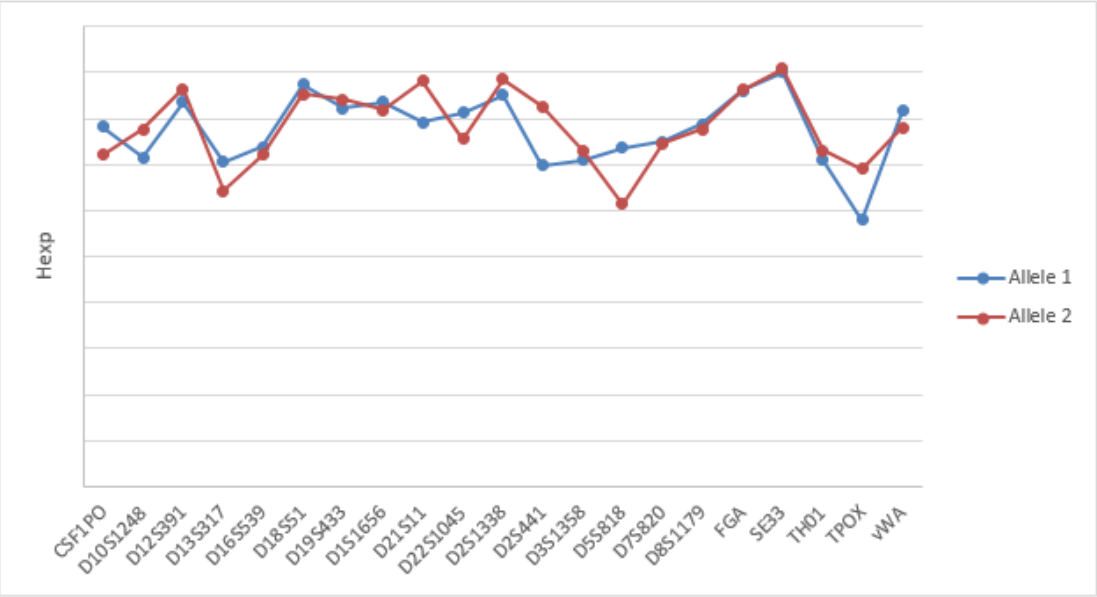

Figure 9: **Coloured autosomal markers expected heterozygosity by alleles**

Table 4: Average values for the observed and the expected heterozygosity

	Colored Group		Black group	
STR loci	**Ho**	**He**	**Ho**	**He**
CSF1PO	0.249	0.752	0.230	0.77
D10S1248	0.255	0.745	0.242	0.759
D12S391	0.150	0.850	0.171	0.830
D13S317	0.323	0.674	0.369	0.631
D16S539	0.271	0.730	0.253	0.747
D18S51	0.138	0.863	0.169	0.831
D19S433	0.169	0.832	0.178	0.823
D1S1656	0.173	0.827	0.156	0.844
D21S11	0.164	0.837	0.158	0.842
D22S1045	0.217	0.783	0.223	0.777
D2S1338	0.132	0.868	0.123	0.879
D2S441	0.238	0.762	0.297	0.704
D3S1358	0.281	0.720	0.230	0.704
D5S818	0.325	0.675	0.317	0.684
D7S820	0.253	0.748	0.266	0.734
D8S1179	0.219	0.782	0.308	0.693
FGA	0.139	0.861	0.143	0.858
SE33	0.0965	0.904	0.096	0.906
TH01	0.489	0.720	0.342	0.658
TPOX	0.366	0.635	0.293	0.707

Figure 5 depicts that the examined loci; SE33 for the Namibian coloured group was the most discriminating at an average of 0.9035, whereas the TPOX genetic marker was the least discriminating (0.6350). The black racial grouping for this study observed maximum and minimum expected heterozygosity values for markers SE33 (0.9055) and D13S317 (0.6310) respectively.

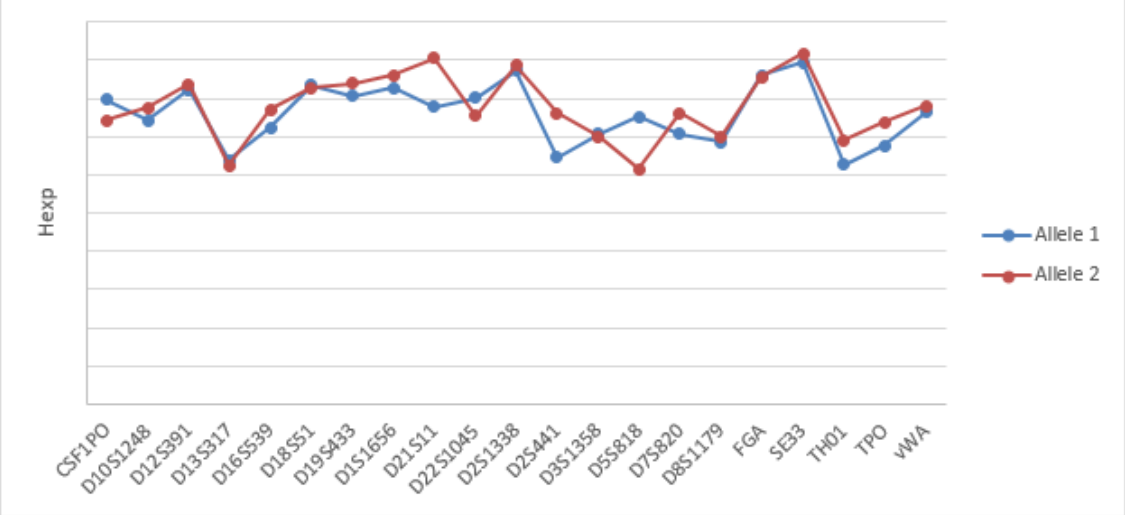

Figure 10: **Black autosomal markers expected heterozygosity by alleles**

4.4 Hardy-Weinberg equilibrium

There was no significant departure from the Hardy-Weinberg equilibrium observed throughout the study. However, values that deviated slightly, were corrected by the GenoProof software package ensuring the stability of all other calculations.

4.5 The most common genotypic frequencies

The frequencies of the most common DNA profiles would prove the least favourable when interpreting genotype tables for human identification [34]. The current findings used the most common alleles at every locus point to estimate theoretical genotype frequencies. Table 5 and table 6 reflects the theoretically most common genotype frequencies that were calculated based on the two most common allele designations recorded among the coloured and blacks for the Namibian population.

Table 5: The two most common observed alleles and their frequencies as well as the most common genotypic frequencies in Namibian Coloured people (n = 150).

Autosomal STR loci	Allele 1	Allele 2	Allele 1 frequency (p)	Allele 2 frequency (q)	Genotype frequency (2pq)
CSF1PO	10	12	0.380	0.420	0.319
D10S1248	13	14	0.453	0.287	0.260
D12S391	17	19	0.253	0.193	0.098
D13S317	11	12	0.413	0.533	0.441
D16S539	11	12	0.353	0.347	0.245
D18S51	17	19	0.193	0.227	0.088
D19S433	13	14	0.307	0.260	0.159
D1S1656	11	15	0.244	0.293	0.143
D21S11	28	32.2	0.300	0.220	0.132
D22S1045	15	18	0.253	0.407	0.206
D2S1338	19	22.2	0.233	0.173	0.081
D2S441	11	14	0.500	0.260	0.260
D3S1358	15	17	0.407	0.347	0.282
D5S818	11	13	0.373	0.407	0.304
D7S820	8	11	0.340	0.313	0.213
D8S1179	14	14	0.307	0.307	0.188
FGA	20	25	0.193	0.220	0.085
SE33	18	19.5	0.193	0.127	0.049
TH01	8	8	0.320	0.373	0.239
TPOX	8	11	0.600	0.380	0.456
vWA	17	18	0.240	0.327	0.157

The 22.2 repeat allele of locus D2S1338 among the coloured Namibian was not observed among the black racial grouping for the country. Instead on marker D2S1338 for the black racial grouping repeats 21-22 were observed on allele 1 and allele 2 for the locus. This suggests that the genetic marker D2S1338 is specific to the coloured people and can thus be very informative in identifying persons related to a crime, persons of disaster victim investigations, paternity cases as well as cases involving missing persons. Furthermore, from the 150 samples that were analysed for this study, it was discovered that SE33 held

a 19.5 allele repeat for the coloured people, an observation that held 16-20 repeat on allele 1 and allele 2 for the black people. Both 10-12 alleles were reported for the locus point CSF1PO in the racial grouping, more extensively, 8-11 repeat alleles were also observed on D21S11, which makes these two markers the least informative in comparison with all the other occurrences of alleles at various markers. The alleles were interpreted by the software as allele 1 and allele 2, as a representation of how chromosomes appear in pairs. The most common observed allele for the coloured racial grouping in this paper representing the Namibian population was TPOX, as it was discovered in 60% of the 150 biological samples analysed. Second most common observed allele for the coloured's as indicated in table 5 was D2S441 (50%), followed by D10S1248 at 45.3%. It is worth mentioning that the autosomal STR markers SE33, FGA and D18S51 showed the lowest frequency of 19.3%, a result which can be collaborated by 19.5 allele repeat which is unique to this group. In comparison to the occurrence of alleles in the black racial grouping as depicted by table 6, the alleles were detected in 53.3% of the 150 biological samples amplified at locus D2S441. An observation that is more than the 50% detected in the coloured grouping. Following is TH01 and TPOX with 52.7% and 51.3% occurrence respectively. The marker TPOX can therefore be reported as the least informative among the black and the coloured people for this paper because of its allele distribution in the population. The lowest percentage with respect to the number of alleles for the black racial grouping was however observed at the genetic marker SE33 (16%). The 16-20 allele repeats were seen at locus point SE33, an observation that deviated from the 18-19.5 at 19.3% for the biological samples extracted from the coloured people for the purposes of this paper. The most informative marker in terms of the alleles observed at every locus by locus for the Namibian population is SE33 as suggested by the finding in the paper.

Table 6: The two most common observed alleles and their frequencies as well as the most common genotypic frequencies in Namibian Black people (n = 150).

Autosomal STR loci	Allele 1	Allele 2	Allele frequency (p)	Allele frequency (q)	Genotype frequency (2pq)
CSF1PO	10	12	0.333	0.393	0.262
D10S1248	13	14	0.380	0.293	0.223
D12S391	17	20	0.240	0.273	0.131
D13S317	11	12	0.493	0.547	0.539
D16S539	9	11	0.407	0.273	0.222
D18S51	15	16	0.253	0.267	0.135
D19S433	13	14	0.333	0.260	0.173
D1S1656	14	15	0.267	0.247	0.132
D21S11	28	32.2	0.393	0.153	0.121
D22S1045	11	17	0.333	0.407	0.271
D2S1338	21	22	0.197	0.173	0.068
D2S441	11	14	0.533	0.367	0.391
D3S1358	15	17	0.373	0.407	0.304
D5S818	12	13	0.380	0.160	0.122
D7S820	8	11	0.407	0.320	0.260
D8S1179	14	14	0.467	0.433	0.404
FGA	23	24	0.220	0.233	0.103
SE33	16	20	0.160	0.153	0.049
TH01	7	8	0.527	0.440	0.463
TPOX	8	11	0.513	0.373	0.383
vWA	16	17	0.367	0.327	0.240

4.6 Allele variations

Even though variations in allele patterns can interfere with the interpretation of peaks during STR analysis, these unique profiles are ideal for forensic investigations. Of the 300 samples that were studied, variant alleles were observed among the coloured racial population at the markers D21S11 and D2S441 as triallelic peaks for two different samples. With regards to the black population, not only was a microvariant allele observed on the genetic marker D2S1338, the loci TPOX yielded PCR products with triallelic

patterns for three different samples. For purposes of this study, these samples with allelic variations were replaced, however results were recorded in order to emphasis the significance of such findings for further research.

4.7 Forensic efficiency parameters

In genetic studies, it is important to determine the significance of DNA typing results with statistical analysis. A statistical approach in the forensic sciences expands the knowledge on the value of the results obtained. The parameters that were calculated to investigate the weight of the data for this paper were as follows; power of exclusion (PE), typical paternity index (TPI), power of discrimination (PD), mean exclusive chance (MEC) and the polymorphism information content (PIC). Table 7 and table 8 shows the values for all loci that were investigated. Furthermore, table 9 is an indication of the PIC on all the examined loci.

Table 7: The forensic efficiency parameter values in Namibian coloured people (n =150).

Autosomal STR loci	Allele	PE	TPI	Ho	He	PD	MEC	P
CSF1PO	1	0.567	2.299	0.218	0.782	0.883	0.597	0.000
	2	0.462	1.795	0.279	0.721	0.810	0.488	0.000
D10S1248	1	0.451	1.750	0.286	0.714	0.792	0.494	0.000
	2	0.556	2.236	0.224	0.776	0.886	0.562	0.000
D12S391	1	0.667	3.041	0.164	0.836	0.939	0.675	0.000
	2	0.722	3.672	0.136	0.864	0.959	0.727	0.000
D13S317	1	0.436	1.696	0.295	0.705	0.786	0.468	0.000
	2	0.344	1.396	0.358	0.642	0.660	0.405	0.000
D16S539	1	0.489	1.906	0.262	0.738	0.836	0.507	0.000
	2	0.462	1.793	0.279	0.721	0.818	0.471	0.000

Autosomal STR loci	Allele	PE	TPI	Ho	He	PD	MEC	P
D18S51	1	0.740	3.931	0.127	0.873	0.965	0.743	0.000
	2	0.699	3.385	0.148	0.852	0.951	0.708	0.000
D19S433	1	0.640	2.808	0.178	0.822	0.925	0.658	0.000
	2	0.678	3.153	0.159	0.841	0.943	0.686	0.000
D1S1656	1	0.668	3.050	0.164	0.836	0.940	0.675	0.000
	2	0.634	2.752	0.182	0.818	0.923	0.650	0.000
D21S11	1	0.584	2.401	0.208	0.792	0.899	0.599	0.000
	2	0.756	4.195	0.119	0.881	0.967	0.767	0.000
D22S1045	1	0.621	2.656	0.188	0.812	0.920	0.627	0.761
	2	0.518	2.037	0.246	0.754	0.850	0.552	0.000
D2S1338	1	0.697	3.358	0.149	0.851	0.951	0.703	0.000
	2	0.765	4.354	0.115	0.885	0.972	0.767	0.000
D2S441	1	0.425	1.656	0.302	0.698	0.754	0.498	0.000
	2	0.648	2.874	0.174	0.826	0.931	0.658	0.000
D3S1358	1	0.442	1.715	0.292	0.708	0.795	0.462	0.000
	2	0.478	1.861	0.269	0.731	0.832	0.484	0.000
D5S818	1	0.487	1.897	0.264	0.736	0.832	0.514	0.000
	2	0.308	1.296	0.386	0.614	0.630	0.337	0.000
D7S820	1	0.508	1.991	0.251	0.749	0.852	0.522	0.000
	2	0.504	1.972	0.254	0.746	0.850	0.515	0.000
D8S1179	1	0.577	2.361	0.212	0.788	0.896	0.588	0.000
	2	0.554	2.225	0.225	0.775	0.882	0.571	0.000
FGA	1	0.714	3.567	0.140	0.860	0.957	0.718	0.000
	2	0.718	3.615	0.138	0.862	0.957	0.725	0.000
SE33	1	0.795	5.000	0.100	0.900	0.978	0.802	0.000
	2	0.810	5.388	0.093	0.907	0.981	0.814	0.000
TH01	1	0.443	1.721	0.291	0.709	0.800	0.456	0.000
	2	0.478	1.860	0.687	0.731	0.828	0.496	0.000
TPOX	1	0.268	1.191	0.420	0.580	0.515	0.353	0.000
	2	0.412	1.610	0.311	0.690	0.766	0.435	0.000
vWA	1	0.631	2.732	0.183	0.817	0.925	0.636	0.000
	2	0.562	2.268	0.220	0.780	0.886	0.578	0.000

KEYS-PE: probability of exclusion; TPI: typical paternity index; Ho: observed heterozygosity; He: expected heterozygosity; PD: power of discrimination; MEC: mean exclusive chance; P: p-value of the HWE exact test.

Short tandem repeat analysis based on the 21 STR loci investigated, exhibited the highest power of exclusion values at SE33 for the coloured people. The highest reported at allele 1 and allele 2 was 0.795 and 0.810 respectively. At TPOX the lowest power of exclusion value was calculated as 0.268 for allele 1 and 0.412 for allele 2, yielding an average value of 0.340. On the other hand, analysis of the power of exclusion parameter for the black racial grouping in Namibia, showed the minimum value (0.330) at the genetic marker, D13S317. The maximum value for the power of exclusion was seen at SE33 as well, however at an average of 0.807, higher than that calculated for the coloured people. The power of exclusion forensic efficiency parameter showed some of the highest indices at the markers D2S1338 (0.752), FGA (0.709) and D21S11 (0.684).

Analysis of the typical paternity index (TPI) yielded results situated between 1.546 (D13S317) and 5.194 (SE33) for the coloured people as depicted in table 6. As mentioned earlier, the typical paternity index compares the probability that an allele passed on from an alleged father to a child is similar to that of an unrelated man of similar background [34]. Of the 150 biological samples analysed, the highest typical paternity index for the black racial grouping was observed at the STR locus SE33 (4.719; 6.055) on allele 1 and allele 2, respectively. A good appreciation of genetic diversity was followed by markers D21S11 and FGA at average TPI values of 3.793 and 3.503. The lowest typical paternity index values for the black group was observed at locus TH01 at an average of 1.475, followed by a calculated average of 1.356 at the genetic marker D13S137.

Table 8: The forensic efficiency parameter values in Namibian black people (n =150).

Autosomal STR loci	Allele	PE	TPI	Ho	He	PD	MEC	P
CSF1PO	1	0.594	2.467	0.203	0.797	0.903	0.611	0.000
	2	0.498	1.944	0.257	0.743	0.840	0.522	0.000
D10S1248	1	0.495	1.934	0.259	0.742	0.840	0.517	0.000
	2	0.554	2.224	0.225	0.775	0.885	0.558	0.000
D12S391	1	0.643	2.832	0.177	0.823	0.930	0.648	0.000
	2	0.667	3.042	0.164	0.836	0.938	0.677	0.000
D13S317	1	0.339	1.381	0.362	0.638	0.666	0.385	0.000
	2	0.321	1.331	0.376	0.624	0.625	0.385	0.000
D16S539	1	0.465	1.805	0.277	0.723	0.811	0.497	0.000
	2	0.546	2.182	0.229	0.771	0.880	0.554	0.000
D18S51	1	0.664	3.013	0.166	0.834	0.937	0.675	0.000
	2	0.651	2.901	0.172	0.828	0.932	0.663	0.000
D19S433	1	0.611	2.583	0.194	0.806	0.911	0.632	0.000
	2	0.674	3.110	0.161	0.839	0.941	0.683	0.000
D1S1656	1	0.649	2.883	0.173	0.827	0.932	0.658	0.000
	2	0.716	3.594	0.139	0.861	0.956	0.726	0.000
D21S11	1	0.559	2.254	0.222	0.778	0.876	0.596	0.000
	2	0.808	5.332	0.094	0.906	0.981	0.811	0.000
D22S1045	1	0.600	2.505	0.200	0.800	0.906	0.616	0.000
	2	0.518	2.037	0.246	0.754	0.850	0.552	0.000
D2S1338	1	0.739	3.916	0.128	0.872	0.964	0.746	0.000
	2	0.765	4.354	0.115	0.885	0.972	0.767	0.000
D2S441	1	0.350	1.412	0.354	0.646	0.665	0.419	0.000
	2	0.529	2.091	0.239	0.761	0.863	0.553	0.000
D3S1358	1	0.440	1.709	0.293	0.707	0.795	0.456	0.000
	2	0.429	1.670	0.300	0.701	0.784	0.447	0.000
D5S818	1	0.512	2.013	0.248	0.752	0.851	0.534	0.000
	2	0.310	1.300	0.385	0.615	0.632	0.340	0.000
D7S820	1	0.439	1.706	0.293	0.707	0.791	0.463	0.000
	2	0.529	2.095	0.239	0.761	0.866	0.544	0.000
D8S1179	1	0.407	1.594	0.314	0.686	0.750	0.452	0.000
	2	0.427	1.662	0.301	0.699	0.776	0.460	0.000
FGA	1	0.714	3.560	0.140	0.860	0.957	0.718	0.000
	2	0.704	3.445	0.145	0.855	0.952	0.715	0.000
SE33	1	0.783	4.719	0.106	0.894	0.976	0.787	0.000
	2	0.831	6.055	0.086	0.917	0.985	0.834	0.000

Autosomal STR loci	Allele	PE	TPI	Ho	He	PD	MEC	P
TH01	1	0.323	1.336	0.374	0.626	0.638	0.366	0.000
	2	0.413	1.613	0.310	0.690	0.764	0.438	0.000
TPO	1	0.393	1.545	0.324	0.676	0.720	0.460	0.000
	2	0.489	1.908	0.262	0.738	0.837	0.505	0.000
vWA	1	0.532	2.108	0.237	0.763	0.866	0.551	0.000
	2	0.562	2.268	0.220	0.780	0.886	0.578	0.000

KEYS-PE: probability of exclusion; TPI: typical paternity index; Ho: observed heterozygosity; He: expected heterozygosity; PD: power of discrimination; MEC: mean exclusive chance; P: p-value of the HWE exact test.

With respect to the power of discrimination as presented in table 7 and table 8, the software computed high values for this parameter. The highest power of discrimination was observed on allele 1 and allele 2 of locus SE33; an average of 96.7% for the coloured racial grouping. Similar results were evident for the second group. As mentioned above, following is a table displaying the polymorphism information content for both the coloured and the black racial groupings that were studied for Namibia.

Table 9: Polymorphism information content (PIC) values in Namibian coloured and black people.

Autosomal STR loci	Allele	Namibian Coloured People	Namibian Black People
CSF1PO	1	0.757	0.771
	2	0.679	0.705
D10S1248	1	0.677	0.701
	2	0.740	0.739

Autosomal STR loci	Allele	Namibian Coloured People	Namibian Black People
D12S391	1	0.816	0.800
	2	0.849	0.816
D13S317	1	0.658	0.578
	2	0.597	0.576
D16S539	1	0.695	0.681
	2	0.670	0.734
D18S51	1	0.860	0.814
	2	0.836	0.806
D19S433	1	0.801	0.783
	2	0.823	0.820
D1S1656	1	0.816	0.805
	2	0.796	0.847
D21S11	1	0.763	0.754
	2	0.871	0.899
D22S1045	1	0.785	0.775
	2	0.724	0.724
D2S1338	1	0.834	0.860
	2	0.874	0.874
D2S441	1	0.670	0.604
	2	0.804	0.727
D3S1358	1	0.659	0.655
	2	0.683	0.648
D5S818	1	0.696	0.715
	2	0.538	0.540
D7S820	1	0.708	0.658
	2	0.703	0.724
D8S1179	1	0.757	0.641
	2	0.742	0.652
FGA	1	0.844	0.844
	2	0.847	0.839
SE33	1	0.892	0.885
	2	0.900	0.912
TH01	1	0.657	0.568
	2	0.687	0.638
TPOX	1	0.536	0.642
	2	0.634	0.695
vWA	1	0.792	0.728
	2	0.748	0.748

Table 9 above presents the polymorphism information content (PIC) estimates for both coloured and black people in Namibia. Among the investigated STR markers, TPOX possessed the least PIC value (0.536) on allele 1 and allele 2 (0.634), yielding an average of 58.8%, while the genetic marker SE33 exhibited the highest value of 89.6% for the coloured group. In addition, the highest PIC average for the black group recorded was on the SE33 locus as well, a value of 89.9%, whereas the least PIC value 57.7% was observed on locus D13S317. These PIC values serve as a good indicator of genetic variation for all the STR loci examined in this paper.

CHAPTER 5

DISCUSSION

The human genome has a wide range of sequence information that was developed by researchers. This information is constantly being updated, captured and preserved in archives which allow for information sharing between laboratories [36]. In depth sequencing of the human genome has enabled scientists to identify regions of interests showing variation between individuals. These regions with sequence variations are referred to as autosomal STR sequences [37]. They consist of repeating nucleotide units which facilitate the identification of human beings in forensic investigations [38-41]. As the number of the STR regions used during human identification increases, so does the uniqueness of an individual's genotypic profile [42].

Based on the characterized DNA sequences; that is the STR markers, the allele frequencies for the samples at these regions were determined. Guided by several forensic efficiency parameters, results from the allelic frequencies were used to deduce the significance of the findings. The informative genetic parameters for forensic efficiency included the observed and expected heterozygosity, power of discrimination, power of exclusion, mean exclusive chance, the typical paternity index and the polymorphism information content.

5.1 Allele frequency distribution

A locus-by-locus analysis of the allelic frequencies among the Namibian people, for the two racial groupings, are illustrated in Appendix 1 and Appendix 2 of the paper. At the genetic marker D7S820 a total of 7 alleles were observed with the repeat numbers 7, 8, 9,

10, 10.3, 11.2 and 12 for the black group. Allele 10 occurred with the highest frequency of 0.303, however for the coloured group, this value was lower with an allele frequency of 0.277. The repeat numbers 8, 10 and 10.3 were the most predominant alleles among the blacks, whereas that of the coloured group included allele 12, instead of 10.3. In comparison with other studies, allele 10 of the same locus was also found to be the most predominant in Angola, Mozambique and the American-Basques population [43, 44,45]. It is worth mentioning that instead of Globalfiler, the Identifiler direct amplification kit was used to conduct studies for the Angolan and Mozambique populations [46]. The use of different amplification kit may give rise to alleles that would otherwise not be identified when using the same kit.

There were 12 allele types recorded on D5S818 and vWA among the black population in Namibia and a total of 7 at TPOX marker. The highest frequency ranges among the coloureds was 0.290-0.383. Regarding D8S1179, allele 14 of the coloured group was the most predominant. In addition, at the locus CSF1P0, alleles 9 and 10 were observed for the coloured and black group respectively.

For the loci FGA, alleles observed on 31.2 were not found in the Angolan population. Furthermore, the allele 43.2 was observed in this paper, but not in the paper conducted for Mozambique. On the other hand, the American-Basque population showed no presence of allele 31.2 and 43.2 on the locus FGA. Since there was an absence of certain alleles at various markers, it can be said that such information proves useful in forensic investigations in Namibia.

In addition, the presence of more STR markers such as SE33, prove invaluable to human identification analysis during DNA typing. This is because the Globalfiler amplification

kit that was applied to this study, allowed for the detection of certain alleles unique to the Namibian population.

5.2 Heterozygosity and homozygosity

Results for the expected and observed heterozygosity can be observed in table 7 and table 8 of the forensic efficiency parameters for both the coloured and the black group respectively. The expected heterozygosity versus the observed heterozygosity is illustrated in figure 7 and figure 8 below. They are an indication of the degree of variation [47] among the black and the coloured people in Namibia for the STR loci studied.

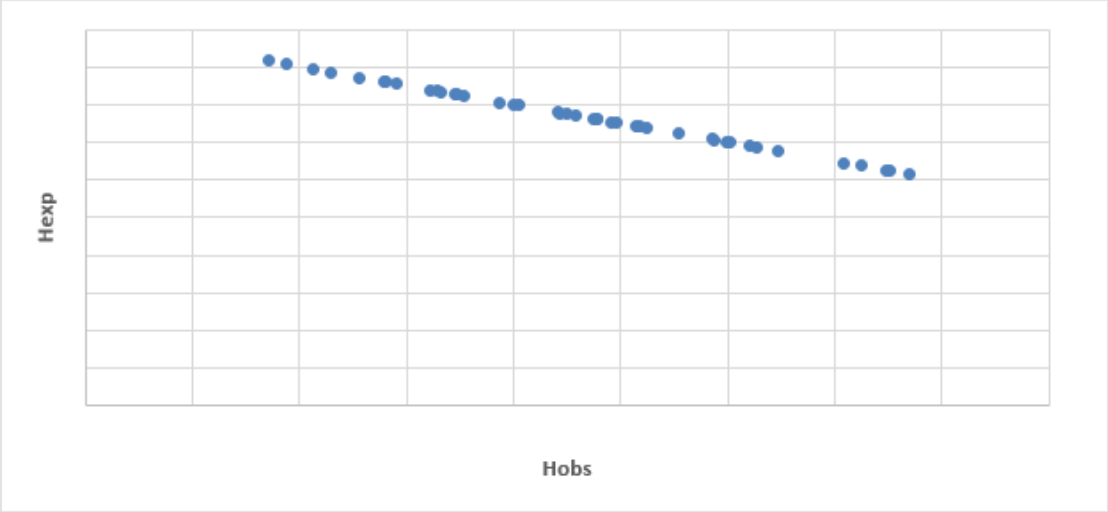

Figure 11: **Expected heterozygosity as function of observed heterozygosity per locus in Namibian black people.**

The figure above shows a linearity between the expected heterozygosity and the observed heterozygosity among the black people in Namibia. On the other hand, figure 8 below shows a non-linearity for the expected heterozygosity against the observed heterozygosity. It is worth noting that the presence of variant alleles among the 150 samples analysed for the coloured racial grouping could be a result of this. The value for the observed

heterozygosity depends greatly on the number of the genetic markers under investigation, as well as the allelic frequency results recorded for every locus-by-locus calculation [48]. The allelic variations in population studies can be identified as unique to a specific group of people; a uniqueness which can be used to extract more accurate DNA profiles. A study conducted by Dente and colleagues regarding African immigrants in Lisbon [49], suggested that the presence of these variations increases the probability that a biological sample under forensic investigation belongs to a specific person or suspect. Such variation can be observed in figure 8 for the coloured group.

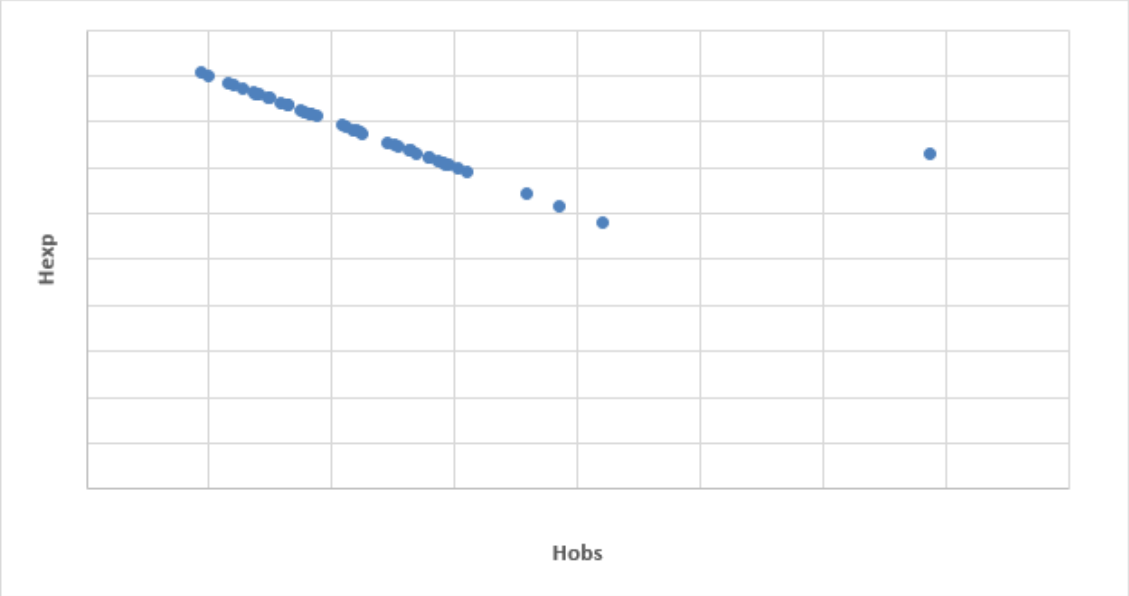

Figure 12: **Expected heterozygosity as function of observed heterozygosity per locus in Namibian coloured people.**

Previous data collected by researchers [50] for the Ovambo (Bantu) ethnic group in Namibia, yielded observed heterozygosity values for 15 STR loci. The loci under investigation were D8S1179, D21S11, D7S820, CSF1PO, D3S1358, TH01, D13S317, D16S539, D2S1338, D19S433, vWA, TPOX, D18S51, D5S818 and FGA. The STR

marker D18S51 yielded the highest heterozygosity of 0.887, whereas at locus D13S317 the lowest observed heterozygosity was seen.

Presented in table 10 that follows, are the observed heterozygosity values for some of the populations studied in different countries. The populations that are illustrated in comparison to Namibia for this study are Angola, Mozambique, Basque-Americans and Koreans.

Table 10: **Values for the observed heterozygosity in various population studies**

	HETEROZYGOSITY			
STR MARKER	**Angola**	**Mozambique**	**Basque-American**	**Korea**
D8S1179	0.760	0.766	0.763	0.834
D21S11	0.885	0.812	0.750	0.790
CSF1PO	0.808	0.786	0.699	0.704
D13S317	0.656	0.721	0.769	0.783
vWA	-	0.831	0.750	0.792
TPOX	0.765	0.812	0.673	0.640
D18S51	0.854	0.877	0.833	0.772
D5S818	0.738	0.721	0.660	0.772
FGA	0.854	0.864	0.846	0.842
SE33	-	-	-	-

Most of the values for heterozygosity were above 50% which is a good indicator for the selection of independent alleles. Furthermore, the researchers Crow and Kimura explained that the expected heterozygosity was the probability that two alleles of the same locus, taken at random from a population under study would have different designations, an explanation that would later emphasize genetic diversity [51]. However, in some countries such as those published by Pérez-Miranda AM and colleagues, the observed heterozygosity was as little as 0.339 [48]. These low values would be influenced by the

country’s geographical location, population size in that area, including but not limited to the inbreeding cultural preferences [52,53].

5.3 Polymorphism information content

Similar to the observed heterozygosity, the genetic diversity of a population under investigation can also be determined by utilizing the polymorphism information content for every STR locus point. The diversity of PIC values can be classified as monomorphic to highly informative at a set marker; with the highest polymorphism defined by values closer to 1 and the least either values closer to zero or including it [54]. These PIC values are all dependent on the occurrence of the most frequent alleles and allele frequencies in the population of interest [55]. Relative to this paper, the PIC values for the black and the coloured group as seen in table 9 for the 21 STR loci indicated a high polymorphism in the Namibian population. The genetic marker SE33 is located on the long arm of chromosome 6 (6q14) and codes a b-actin related pseudogene which is believed to be functionless [56]. It is however noticeable that the degree of polymorphism as presented in table 9 of this paper, showed that TPOX was the least polymorphic marker for the coloured group and that the STR locus SE33 exhibited the highest polymorphism. The selected markers for a study conducted in 2013 by Thanakiatkrai and colleagues were in accordance with this paper that TPOX revealed the lowest variation among the 150 biological samples for the Namibian individuals [57]. It is worth mentioning that the behaviour regarding the most polymorphic marker was similar for the black population, despite the least genetic diversity observed at the autosomal STR marker D13S317. Furthermore, the PIC values for a study conducted by researchers at the University of Zimbabwe on the native ethnic group Shona revealed results with highly informative markers on similar positions [58]. A similar degree of polymorphism presented values on previously mentioned markers for results obtained for a Japanese population study [59]. A population study on the Basque-Americans [45], used the software Powerstats version

1.2 (United States of America) [60], to report polymorphism information content values ranging from 0.600 on TPOX with the least polymorphic to 0.860 at the autosomal STR marker FGA showing the highest polymorphism. On the contrary, a study conducted on Koreans reported that the highest polymorphic regions were located on the STR loci D18S51 and D2S1338 with recoded PIC values of 0.8402 and 0.8585, respectively [61]. During the polymerase chain reaction process, challenges can occur when analysing scientific data whereby mutations at primer-binding sites lead to allele-dropouts; additionally, some of the commercially designed STR multiplex amplification kits could be unable to detect allele designations that are either too large or too small for a given genetic marker such as FGA [62, 63, 64]. However, an examination of the STR locus by Yang and other researchers concluded that the number of variations exhibited by short tandem repeat units during STR typing, plays an important role in determining human genetic diversity in specific populations and subpopulations [65], thus facilitating many investigations involving human identification.

5.4 Power of discrimination

In forensics investigations, the examination of short tandem repeat loci has become the "golden standard" for human identification over twenty years [66, 67, 68]. The genetic markers for all the tested groups in this study used the Globalfiler express amplification kit which revealed the highest and the lowest polymorphic regions of autosomal loci for the racial groupings. The calculated values for the power of discrimination for all the investigated STR loci for the coloured and black people are displayed in table 7 and table 8, respectively. Fisher [32] explained that the power of discrimination is the probability

that two persons chosen at random would have genotype tables that were not identical. In this study, the STR alleles produced with full genetic information revealed that the highest power of discrimination was at the following loci: SE33, D2S1338, D18S51 and D12S391, with average values of 0.980, 0.962, 0.958 and 0.949 respectively for the coloured group. The lowest power of discrimination was observed at the marker TPOX; 0.641. The STR marker TPOX stands for the human thyroid peroxidase gene located on chromosome 2. Simple AATG repeat units and sequence alleles of various ranges has contributed towards the variation in this STR region [69, 70]. Additionally, for the black group the minimum and maximum power of discrimination values 0.645 and 0.981 were observed at SE33 and D13S317 respectively. Locus D13S317 is found on the long arm chromosome 13. The sequenced alleles from the GeneBank reference contains between 5 and 15 repeat units [71, 72], with length-based repeats extending to 17 repeat units [73]. These values mentioned for the Namibian population imply that the loci investigated can contribute towards the generation of an allele frequency database for Namibia. With regards to the Angolan population study, the highest power of discrimination was observed at D2S1338. This locus was also among the highest recorded power of discrimination values for the Namibian coloured people, a similarity worth mentioning. Furthermore, the Mozambique population compared to this study revealed D21S11 (0.864) and D13S317 (0.714) as the maximum and minimum values exhibiting discrimination. The Globalfiler express amplification kit was also used in a Japanese population and the findings of the locus exhibiting the highest discriminating power were similar to that of the coloured and black Namibian racial grouping [74]. On the other hand, TPOX was observed as the least discriminating in a study conducted by Zgonjanin and colleagues [75]; a result that varies from that of the black racial grouping as compared to

the Namibian people in this study. This marker was also observed as the least discriminating in previously researched papers [76]. An increasing power of discrimination value, is not only useful for criminal investigations, but can also aid in missing person cases [77], further providing useful information on patterns of genetic diversity and population structures that can be used to aid the development of allele frequency databases [78]. The polymorphism which exists at the loci describes the probability that an individual who is innocent of an offense, will not be included as a contributor of a particular biological sample, thus considered as an excluded DNA profile. A higher discriminating power could therefore be very useful in forensic applications.

5.5 Power of exclusion

A STR marker is considered informative for paternity testing under the assumption that a "random man" can be excluded as a father of a child, assuming that there is no sudden change in the genetic structure of a population. This probability is known as the power of exclusion and can be measured either in the absence or presence of one of the biological parents of the child [79, 80]. Moreover, the power of exclusion provides statistical information useful in cases where individuals are falsely accused. In the present study, the power of exclusion was calculated and presented in table 7 and table 8 of the forensic efficiency parameters. Regarding table 7 (coloured racial grouping), it was found that the power of exclusion values ranged from 0.340 (TPOX) to 0.803 (SE33). The marker TPOX also showed the least genetic diversity in some studies previously captured by the CODIS system [76]. However, the combined power of exclusion was above 90% which is a good indicator that the loci investigated is useful in parentage testing. Relative to the black

racial grouping, the minimum power of exclusion was however revealed by locus D13S317 (0.330) and maximum at SE33 (0.807). Making a comparison between populations for the power of exclusion, a study conducted in Angola reported D2S1338 as the maximum and D13S317 as the minimum. These results were in accordance with reports on the Mozambique population which revealed D2S1338 with the highest exclusion value, but D3S1358 as the lowest power of exclusion.

The two most common alleles in the two racial groupings were used to determine genotype frequencies. These frequencies can be used to estimate theoretical values for general inclusion probabilities which mean that good forensic techniques should give a small inclusive probability when the suspect and offender are the same person and a large inclusive probability when they are different [81]. Regarding the variability of the autosomal STR markers, it was observed that the loci D21S11, FGA, D12S391 and D2S1388 reported the greatest with respect to the number of alleles for the black population. Furthermore, the least variability was observed at marker TH01 for allele 1 and allele 2. Among the results conveyed for the coloured Namibian people the markers showing greatest genetic diversity were similar to that of the coloured group; D21S11, D2S1338, FGA and SE33, whereas the least variable marker was also indicated to be TH01 with allele 8 being the most prevalent.

Among the 21 STR loci investigated, some of the samples yielded DNA profiles with what is called variant alleles. Variant alleles consist of partial repeat sequences in the individual [82]. Micro variants can cause challenges with forensic investigations when reporting STR profiles as inconclusive. However, the rarity of these alleles at certain loci may prove useful in identifying potential suspects related to a crime by cataloguing results.

Within the human genome, the markers presenting the highest polymorphism can be found in the flanking regions of the microvariant alleles; on the basis that they are noncoding regions of no significant importance to phenotypic traits [82]. As mentioned in the results for this paper, microvariant alleles were observed among the coloured racial population at the markers D21S11 and D2S441 as triallelic peaks 35-36-36.1 and 10-12-13 respectively. Identification of the variations in the short tandem repeat units such as that of trialleles and variant alleles [83, 84, 85, 86] allow for supportive data entry into the STRidER (STR for Identity ENFSI Reference database) [87]. Among the black people for the Namibian population, the variant allele 19.343 was observed at D2S1338, furthermore, three triallelic patterns were observed at TPOX; 8-9-10, 8-10-11 and 9-10-11. The resulting variant allele at D2S1338 may have been due to a sequence variation in the base pair repeats ranging from 19-27 repeats [88, 89, 90]. The triallelic variation at TPOX among the black people for the Namibian population study presented similar results conducted on a Ghanaian population in 2019 [91]. The patterns 8-9-10 and 9-10-11 were observed in both studies. However, the 8-10-11 pattern has only been reported in the Namibian population. The presence of microvariant alleles in a population, especially unique to said group can provide an advantage when exploring the genetic diversity between individuals.

5.6 Typical paternity index

Any DNA analysis in the forensic sciences is accompanied by a statistical approach in efforts to add weight of evidence to the findings. These calculations are reported and presented to the court or person (s) who requested for the investigation to be conducted. Regarding the calculations for paternity tests, Gjertson and colleagues published bio

statistical recommendations for probability calculations specific to genetic investigations [92]. It is however up to the internal validation of all forensics laboratories to set up their threshold values for STR profile inclusions based on the paternity index values [92, 93]. Allele frequency data can be used on various STR markers to identify alleged fathers. The probability that an alleged father is in fact related to the child can be measured by the likelihood ratio (paternity index) [94, 95, 96, 97, 98]. In this paper and from the genotype data collected, the paternity index values were calculated and presented in table 7 and 8, along with the power of exclusion and the rest of the forensic efficiency parameters. The information regarding a population structure can actively facilitate the quantification of genetic differentiation among subpopulations relative to a country's total population [99, 100]. Every locus-by-locus typical paternity index value that is above 1, is a good indicator of a parameter used during paternity test calculations [101, 102] in forensics investigations at the National Forensic Science Institute in Namibia and globally. For the Namibian population, the typical paternity index values ranged from 1.475 to 6.055. The data in this paper revealed that the maximum TPI is at SE33 (5.194) and the minimum at D13S317 (1.546) for the coloured people, whereas among the black people the minimum typical paternity index was recorded as 1.356 at the STR marker 1.356. Of all the 21 STR markers that were investigated, SE33 yielded results that show the most usefulness and relativeness in terms of paternity calculations. Furthermore, FGA and D21S11 showed the least degree of reassurance for the Namibian population. During a study published on a population in the south of Portugal [103], the typical paternity index values showed a wide range of TPI; the least significant at the genetic markers D13S317 (2.9186) and at TPOX (1.410). However, SE33 was still yielding the highest TPI, even when compared with population data from a Circassian subpopulation in Jordan [104]. When the Circassian subpopulation

was compared to the data collected on the Namibian population in this paper, it was found that the lowest TPI values were at the loci D2S441 (1.47) and TPOX (1.32). The marker D2S441 is a compound tetranucleotide repeat found on the short arm of chromosome 2 [105]. The sequenced alleles at this marker are ranging from 10-13 repeats [106]; however, smaller repeats of 8 and 9 have been noticeable in some populations [105,107]. More extensively, various microvariant and other rare alleles have also been reported for this locus point [107, 108]. In the case of D5S818, a population study on the Basque-Americans showed the minimum TPI for this marker and the maximum at D2S1338; an observation that was significantly different from the Namibian population as well as studies conducted on Saudi-Arabians of African descent [109,110, 111, 112, 113]. The TPI values in the current paper, show high levels of genetic diversity among the coloured and the black people in Namibia, with the most informative marker for closely related individuals among the 21 STR loci being SE33.

5.7 Mean exclusion chance

To characterize the genetic diversity of the 21 STR loci that were investigated, the mean paternity exclusion chance (MEC) also had to be evaluated as one of the forensic statistical parameters. Allele calls have been achieved with the use of various commercially available STR kits that have been developed over the years. These kits have proven the locus-by-locus variations that can exist among populations [114-120]. Many authors have developed formulae suitable for use in probability testing related to forensic casework [121]. These formulae were derived for exclusion purposes and can be applied in the following instances; if one of the alleged biological parents of the child has been wrongly

identified, when both parents are wrongly identified as that of the offspring and (or) when the STR profile for one of the parents is unavailable [122, 123-127]. The mean exclusion chance is therefore used in paternity cases to describe the probability of exclusion for a certain STR locus under investigation [128-129]. Thus, determining the probability that a certain DNA profile obtained from STR typing is in fact that of the biological mother or alleged father of the child. Szibor and colleagues (2003, 2007) as well as Desmarais and other authors (1998) also designed formula to describe the MEC which is equivalent to the probability of exclusion for a given mother/child genotype and (or) alleged father/child [130, 131].

The efficiency parameter investigated in this work ranged from 0.802 to 0.353 for the coloured racial grouping, whereas that for the black group ranged from 0.340 to 0.834. For the coloured group, in all the STR loci examined except for TPOX, D5S818, D13S317, D3S1358 and D16S539, the MEC values were below 50%: 0.394, 0.426, 0.437, 0.473 and 0.489, respectively. The highest mean exclusion chance value was observed at the genetic marker SE33. The high values for the forensic efficiency parameter, demonstrates the usefulness of the set of markers to provide supplementary information for forensic and kinship analysis [132]. Additionally, the calculations based on the black racial group, showed that the marker SE33 (0.812) also had the highest MEC for allele 1 and allele 2. This was followed by the loci FGA and D2S1338. The genetic locus FGA is a tetranucleotide located in the human alpha fibrinogen locus on the long arm of chromosome 4 and sometimes referred to as FIBRA or HUMFIBRA. The sequence motif of the locus ranges from 13-29 total repeats [71, 134, 135], however micro variants have been observed in some forensic literature [136-137]. Furthermore, the markers with the

lowest mean exclusion chance values were D13S317 and TH01. Varying patterns of genetic relationships were observed in population study on Northern German unrelated individuals, whereby the highest MEC was on the marker D2S441 and the lowest MEC value on D14S1434 [138]. There are very few studies reported on mean exclusion chance values. The mean exclusion chance results from this paper can therefore provide supplementary information required for parentage identification, such as in cases involving father/daughter relationships, paternity in close relatives, deficiency cases without access to the putative father, maternity testing, and in rape or incest cases [139].

CHAPTER 6

CONCLUSION

There have been a great number of cases which have remained inconclusive in the Namibian forensic DNA system. Some of these cases are related to mass victim disaster investigations; plane crashes and roadside accidents, cases involving identified and/ or unidentified suspects related to crimes such as burglary, rape and murder. Maternity and paternity cases, as well as missing person's investigations. The identification of human beings often relies on the analysis of short tandem repeat loci. However, the increasing number of unknown contributors of a biological sample poses as one of the greatest challenges during DNA analysis.

The study presented a method for characterizing the genetic structure of the Namibian population using 21 forensic STR loci. Relying on the number of STR loci, various forensic efficiency parameters were tested to determine the relativeness of this genetic information. The Globalfiler loci for this study, yielded the following conclusions:

- The highest allele frequency for the coloured racial grouping occurred in the allele 8 of the genetic marker TH01 (36.3%) and for the black racial grouping the observation was made in allele 13 and in allele 14 for the STR markers vWA and D8S1179 (44.7%), respectively.
- The heterozygosity of the 21 STR loci for the Namibian population ranged from 63.1% to 90.5% average. The genetic marker SE33 had the highest heterozygosity for both coloured and the black racial grouping, while the lowest values were

presented by TPOX and D13S317 for the coloured and the black group, respectively.

- The maximum power of exclusion was observed at SE33 for both the groups under investigation. With a minimum power of exclusion value exhibited by TPOX (34%) and D13S317 (33%) by the coloured and the black racial grouping, respectively.
- The most power of discrimination was generated by the STR locus SE33 for both the typing groups at an average value of 96.7%, whereas the least values were at the loci TPOX (64.1%) and D13S317 (average value of 0.645) for the coloured and the black racial groupings, respectively.
- Where the Hardy-Weinberg equilibrium was calculated, there were no significant departures seen for most of the studied STR loci.
- Consequently, it can be concluded that the forensic efficiency parameters evaluated for all the 21 STR loci in this paper, are highly informative in forensic casework applications involving human identification. This in turn assisting the Namibian legal system and all other relevant bodies including but not limited to insurance groups and motor vehicle accident funds.

RECOMMENDATIONS

There are many aspects that can play beneficiary roles towards drawing conclusions in criminal and all other forensics investigations. Highly polymorphic STR loci can be used to generate genetic information with a greater discriminatory power. This information will be highly diverse and can prove effective in human identification. There is potential for short tandem repeat loci to be used in fully equipped forensics laboratories. The evaluation of the genetic markers in this paper showed a great statistical power among the racial groupings. A power which can be associated with the generation of DNA reference sample databases. Furthermore, there was a presence of various variants; micro variants and an occurrence of triallelic patterns in some of the analysed biological samples. Some of these patterns are believed to be occurrences of rare alleles unique to the genetic diversity of the Namibian population. A study using commercial amplification kits that investigate more STR markers can be conducted to demonstrate the extend of this polymorphism.

In efforts to have a faster DNA analysis turnover, thus reducing the backlog that has accumulated in the forensic DNA laboratory over the years, it will be more informative and effective to establish search parameters most suitable for a DNA reference sample database. Since the forensic parameters defined in the present study yielded great discriminatory power, it is recommended to develop a strategy to investigate the statistical power of the various independent ethnic groupings in Namibia. These frequencies would be used in cases whereby the ethnic group of the suspect is known to limit the number of hours spend in identifying an individual linked to a particular case. A race-based DNA reference database alone, may not be accountable enough if the population size is too large and when interracial relationships are observed. Additionally, a comparison between

newer isolation methods can be researched to investigate the genetic diversity of the geo-socio structures of the Namibian population. Such a paper would focus on evaluating the male specific region of the Y-chromosome which would prove most useful in gang rape cases. Moreover, the mitochondrial DNA frequency of the Namibian population can also be investigated in order to compare the mean difference in polymorphism. Therefore, choosing the most informative genetic loci and search parameters which can be used to generate not just a DNA reference frequency database but also motivate the generation of an online DNA reference database.

REFERENCES

1. Butler JM, Hill CR. Biology and genetics of new autosomal STR loci useful for forensic DNA analysis. CRC press; 2012.
2. Widyanto MR, Hartono RN, Soedarsono N. A novel human STR similarity method using cascade statistical fuzzy rules with tribal information inference. International Journal of Electrical and Computer Engineering. 2016 Dec 1;6(6):3103.
3. Zahra A, Hussain B, Jamil A, Ahmed Z, Mahboob S. Forensic STR profiling based smart barcode, a highly efficient and cost-effective human identification system. Saudi Journal of Biological Sciences. 2018 Dec 1;25(8):1720-3.
4. Gettings KB, Aponte RA, Vallone PM, Butler JM. STR allele sequence variation: current knowledge and future issues. Forensic Science International: Genetics. 2015 Sep 1; 18:118-30.
5. Butler JM. Chapter 5 - Short tandem repeat (STR) loci and kits. In: Butler JM, editor. Advanced Topics in Forensic DNA Typing. San Diego: Academic Press. 2012;148-157.
6. Butler J.M. (2001). Forensic DNA Typing: Biology and Technology behind STR Markers. Academic Press.
7. Butler JM. Statistical interpretation: evaluating the strength of forensic DNA evidence. Fundamentals of forensic DNA typing. Academic Press; 2009.242-244.
8. Kholief M, El Shanawany S, Gomaa R. Sex determination from dental pulp DNA among Egyptians. Egyptian Journal of Forensic Sciences. 2017 Dec 1;7(1):29.
9. Ellegren H. Microsatellites: simple sequences with complex evolution. Nature reviews genetics. 2004 Jun;5(6):435.

10. Butler JM. Chapter 5 - Short tandem repeat (STR) loci and kits. In: Butler JM, editor. Advanced Topics in Forensic DNA Typing. San Diego: Academic Press. 2012;99-139.
11. Butler JM. Fundamentals of forensic DNA typing. San Diego, CA Elsevier Academic Press; 2010. Chapter 3-Historical methods; p44-68.
12. Kimpton C, Fisher D, Watson S, Adams M, Urquhart A, Lygo J, Gill P. Evaluation of an automated DNA profiling system employing multiplex amplification of four tetrameric STR loci. International Journal of Legal Medicine. 1994 Nov 1;106(6):302-11.
13. Sparkes R, Kimpton C, Watson S, Oldroyd N, Barnett L, Arnold J, Thompson C, Hale R, Chapman J, Urquhart A, Gill P. The validation of a 7-locus multiplex STIR test for use in forensic casework. International Journal of Legal Medicine. 1996 Dec 1;109(4):186-94.
14. Moreno LI, Brown AL, Callaghan TF. Internal validation of the DNAscan/ANDE™ Rapid DNA Analysis™ platform and its associated PowerPlex® 16 high content DNA biochip cassette for use as an expert system with reference buccal swabs. Forensic Science International: Genetics. 2017 Jul 1; 29:100-8.
15. Urquhart A, Kimpton CP, Downes TJ, Gill P. Variation in short tandem repeat sequences—a survey of twelve microsatellite loci for use as forensic identification markers. International Journal of Legal Medicine. 1994 Jan 1;107(1):13-20.
16. Butler JM. Forensic DNA typing: biology, technology, and genetics of STR markers. Elsevier; 2005 Feb 8.

17. Butler JM. Genetics and genomics of core short tandem repeat loci used in human identity testing. Journal of Forensic Science. 2006;51(2):253-65.
18. Yang X, Zhang X, Zhu J, Chen L, Liu C, Feng X, Chen L, Wang H, Liu C. Genetic analysis of 19 X chromosome STR loci for forensic purposes in four Chinese ethnic groups. Scientific Reports. 2017 Feb 17; 7:42782.
19. Moretti TR, Moreno LI, Smerick JB, Pignone ML, Hizon R, Buckleton JS, Bright JA, Onorato AJ. Population data on the expanded CODIS core STR loci for eleven populations of significance for forensic DNA analyses in the United States. Forensic Science International: Genetics. 2016 Nov 1; 25:175-81.
20. Basic Medical Key. Cited 19 October 2019. From: https://basicmedicalkey.com/wpcontent/uploads/2016/11/B978141606164900041X_t0015.png
21. Hares DR. Selection and implementation of expanded CODIS core loci in the United States. Forensic Science International: Genetics. 2015 Jul 1; 17:33-4.
22. Cloete K, Ehrenreich L, D'Amato ME, Leat N, Davison S, Benjeddou M. Analysis of seventeen Y-chromosome STR loci in the Cape Muslim population of South Africa. Legal Medicine. 2010 Jan 1;12(1):42-5.
23. Besecker J, Peri G, Davis M, Zubizarreta J, Hampikian G. Allele frequencies of 15 STR loci (Identifiler™ kit) in Basque-Americans. Legal Medicine. 2018 Mar 1; 31:17-9.
24. Flores S, Sun J, King J, Budowle B. Internal validation of the GlobalFiler™ Express PCR Amplification Kit for the direct amplification of reference DNA samples on a high-throughput automated workflow. Forensic Science International: Genetics. 2014 May 1; 10:33-9.

25. Li R, Wu R, Li H, Zhang Y, Peng D, Wang N, Shen X, Wang Z, Sun H. Characterizing stutter variants in forensic STRs with massively parallel sequencing. Forensic Science International: Genetics. 2020 Mar 1; 45:102225.
26. Dente Â, Amorim A, Costa HA, Porto MJ, Cunha E, Real FC, da Silva CV. Study of Y chromosome markers with forensic relevance in Lisbon immigrants from African countries–Allelic variants study. Forensic Science International: Genetics Supplement Series. 2019 Nov 27.
27. Butler JM. Chapter 3-STR alleles and amplification artefacts. Advanced Topics in Forensic DNA Typing: Interpretation. 2015:47-86.
28. Cavanaugh SE, Bathrick AS. Direct PCR amplification of forensic touch and other challenging DNA samples: a review. Forensic Science International: Genetics. 2018 Jan 1; 32:40-9.
29. Butler J. STR genotyping and data interpretation. Fundamentals DNA Typing. 2010:205-28.
30. Ciavaglia S, Linacre A. OzPythonPlex: An optimised forensic STR multiplex assay set for the Australasian carpet python (Morelia spilota). Forensic Science International: Genetics. 2018 May 1; 34:231-48.
31. Stephenson FH. Calculations for molecular biology and biotechnology. Academic press; 2016 Jun 16; 13: 423-446.
32. Fisher RA. 243: Standard Calculations for Evaluating a Blood-Group System. Heredity. 1951; 5: 95-102.
33. Hameed IH, Ommer AJ, Murad AF, Mohammed GJ. Allele frequency data of 21 autosomal short tandem repeat loci in Mesan and Basra provinces in South Iraq. Egyptian Journal of Forensic Sciences. 2015 Dec 1;5(4):150-6.

34. Edwards AL, Hammond HA, Jin L, Caskey CT, Chakraborty R. Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. Genomics. 1992 Feb 1;12(2):241-53.
35. Botstein D, White RL, Skolnick M, Davis RW. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. American Journal of Human Genetics. 1980 May;32(3):314.
36. Shaw J. Documenting genomics: Applying archival theory to preserving the records of the Human Genome Project. Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences. 2016 Feb 1; 55:61-9.
37. Xie B, Chen L, Yang Y, Lv Y, Chen J, Shi Y, Chen C, Zhao H, Yu Z, Liu Y, Fang X. Genetic distribution of 39 STR loci in 1027 unrelated Han individuals from Northern China. Forensic Science International. Genetics. 2015 Nov; 19:205.
38. Ball G, Dawnay N, Stafford-Allen B, Panasiuk M, Rendell P, Blackman S, Duxbury N, Wells S. Concordance study between the ParaDNA® Intelligence Test, a Rapid DNA profiling assay, and a conventional STR typing kit (AmpFlSTR® SGM Plus®). Forensic Science International: Genetics. 2015 May 1; 16:48-51.
39. Blackman S, Dawnay N, Ball G, Stafford-Allen B, Tribble N, Rendell P, Neary K, Hanson EK, Ballantyne J, Kallifatidis B, Mendel J. Developmental validation of the ParaDNA® Intelligence System—A novel approach to DNA profiling. Forensic Science International: Genetics. 2015 Jul 1; 17:137-48.
40. Butler JM, Shen Y, McCord BR. The development of reduced size STR amplicons as tools for analysis of degraded DNA. Journal of Forensic Sciences. 2003 Sep 1;48(5):1054-64.

41. Cornelis S, Gansemans Y, Deleye L, Deforce D, Van Nieuwerburgh F. Forensic SNP genotyping using nanopore MinION sequencing. Scientific Reports. 2017 Feb 3;7(1):1-5.
42. Silva NM, Pereira L, Poloni ES, Currat M. Human neutral genetic variation and forensic STR data. PloS one. 2012;7(11).
43. Melo MM, Carvalho M, Lopes V, Anjos MJ, Serra A, Vieira DN, Sequeiros J, Corte-Real F. Genetic study of 15 STRs loci of Identifiler system in Angola population. Forensic Science International: Genetics. 2010 Oct 1;4(5): e153-7.
44. Semo AC, Carvalho MR, Bogas V, Serra A, Lopes V, Brito P, Sá FB, Porto MJ, Gonçalves IM, Corte-Real F. Allelic frequencies of 15 autosomal STRs from two main population groups (Makua and Changana) in Mozambique. Forensic Science International: Genetics Supplement Series. 2017 Dec 1;6: e286-8.
45. Besecker J, Peri G, Davis M, Zubizarreta J, Hampikian G. Allele frequencies of 15 STR loci (Identifiler™ kit) in Basque-Americans. Legal Medicine. 2018 Mar 1; 31:17-9.
46. Wang DY, Chang CW, Lagacé RE, Oldroyd NJ, Hennessy LK. Development and Validation of the AmpFℓSTR® Identifiler® Direct PCR Amplification Kit: A Multiplex Assay for the Direct Amplification of Single- Source Samples. Journal of Forensic Sciences. 2011 Jul;56(4):835-45.
47. Al Janaahi N, Al Ghafri R, Qamar SA. Forensic evaluation of VeriFiler™ Plus 6-dye chemistry kit composed of 23 loci with casework samples. Forensic Science International: Genetics Supplement Series. 2019 Dec 1;7(1):892-6.
48. Pérez-Miranda AM, Alfonso-Sánchez MA, Peña JA, Herrera RJ. Qatari DNA variation at a crossroad of human migrations. Human heredity. 2006;61(2):67-79.

49. Dente Â, Amorim A, Costa HA, Porto MJ, Cunha E, Real FC, da Silva CV. Study of Y chromosome markers with forensic relevance in Lisbon immigrants from African countries–Allelic variants study. Forensic Science International: Genetics Supplement Series. 2019 Nov 27.
50. Muro T, Fujihara J, Imamura S, Nakamura H, Yasuda T, Takeshita H. Allele frequencies for 15 STR loci in Ovambo population using AmpFlSTR® Identifiler Kit. Legal Medicine. 2008 May 1;10(3):157-9.
51. Crow JF, Kimura M. An introduction to population genetics theory. An introduction to population genetics theory. 1970: 83-98.
52. Khubrani YM, Wetton JH, Jobling MA. Extensive geographical and social structure in the paternal lineages of Saudi Arabia revealed by analysis of 27 Y-STRs. Forensic Science International: Genetics. 2018 Mar 1; 33:98-105.
53. Khubrani YM, Wetton JH, Jobling MA. Analysis of 21 autosomal STRs in Saudi Arabia reveals population structure and the influence of consanguinity. Forensic Science International: Genetics. 2019 Mar 1; 39:97-102.
54. Lemos SC, Silveira RL, Buuron SK, dos Santos RS, Moro SC. Determining the Polymorphism Information Content of a Molecular Marker. Gene. 2019 Nov 11:144175.
55. Phillips C, Rodriguez A, Mosquera-Miguel A, Fondevila M, Porras-Hurtado L, Rondon F, Salas A, Carracedo A, Lareu MV. D9S1120, a simple STR with a common Native American-specific allele: forensic optimization, locus characterization and allele frequency studies. Forensic Science International: Genetics. 2008 Dec 1;3(1):7-13.

56. Moos M, Gallwitz D. Structure of a human β-actin-related pseudogene which lacks intervening sequences. Nucleic Acids Research. 1982 Dec 11;10(23):7843-9.

57. Thanakiatkrai P, Kitpipit T, Gupta R, Sharma S, Babahan I, Coban EP, Biyik H, Alizadeh K, da Silva JA, Babnik A, Petkovsek R. Current STR-based techniques in forensic science. Maejo International Journal of Science and Technology. 2013 Jan 1;7(01):1-5.

58. Shonhai M, Nhiwatiwa T, Nangammbi T, Mazando S. Genetic analysis of 27 Y-chromosomal STR loci in a Zimbabwean Shona ethnic group. Legal Medicine. 2020 Mar 1; 43:101660.

59. Hashiyada M, Itakura Y, Nata M. PowerPlex™ 16 analysis in the Japanese population. In International Congress Series 2003 Jan 1 (Vol. 1239, pp. 125-130). Elsevier.

60. Tokdemir M, Tunçez FT, Vicdanli NH. Population genetic data for 15 autosomal STR markers in eastern Turkey. Gene. 2016 Jul 15;586(1):36-40.

61. Kim S, Park HC, Kim JS, Nam Y, Kim HY, Park J, Chung U, Lee J, Lim SK, Park SJ. Allele frequency data of 20 STR loci in 2000 Korean individuals. Forensic Science International: Genetics Supplement Series. 2017 Dec 1;6: e65-8.

62. Tsuji A, Ishiko A, Umehara T, Usumoto Y, Hikiji W, Kudo K, Ikeda N. A silent allele in the locus D19S433 contained within the AmpFℓSTR® Identifiler™ PCR Amplification Kit. Legal Medicine. 2010 Mar 1;12(2):94-6.

63. Boutrand L, Egyed B, Füredi S, Mommers N, Mertens G, Vandenberghe A. Variations in primer sequences are the origin of allele drop-out at loci D13S317 and CD4. International Journal of Legal Medicine. 2001 Apr 1;114(4-5):295-7.

64. Negi DS, Alam M, Bhavani SA, Nagaraju J. Multistep microsatellite mutation in the maternally transmitted locus D13S317: a case of maternal allele mismatch in the child. International Journal of Legal Medicine. 2006 Sep 1;120(5):286-92.
65. Yang Y, Ren H, Chen W, Xie B, Wang Y, Shi Y, Chen C, Li C, Yi L, Fang X, Yan J. Concurrent copy number variations on chromosome 8 and 22 combined with mutation at FGA locus revealed in a parentage testing case. Forensic Science International: Genetics. 2015 Nov 1; 19:81-5.
66. Fregeau CJ, Fourney RM. DNA typing with fluorescently tagged short tandem repeats: a sensitive and accurate approach to human identification. Bio techniques. 1993 Jul;15(1):100-19.
67. Jobling MA, Gill P. Encoded evidence: DNA in forensic analysis. Nature Reviews Genetics. 2004 Oct;5(10):739-51.
68. Moretti TR, Baumstark AL, Defenbaugh DA, Keys KM, Smerick JB, Budowle B. Validation of short tandem repeats (STRs) for forensic usage: performance testing of fluorescent multiplex STR systems and analysis of authentic and simulated forensic samples. Journal of Forensic Science. 2001 May 1;46(3):647-60.
69. Kline MC, Hill CR, Decker AE, Butler JM. STR sequence analysis for characterizing normal, variant, and null alleles. Forensic Science International: Genetics. 2011 Aug 1;5(4):329-32.
70. Allor C, Einum DD, Scarpetta M. Identification and characterization of variant alleles at CODIS STR loci. Journal of Forensic Science. 2005 Aug 17;50(5): JFS2005024-6.
71. Wang L, Zhao XC, Ye J, Liu JJ, Chen T, Bai X, Zhang J, Ou Y, Hu L, Jiang BW, Wang F. Construction of a library of cloned short tandem repeat (STR) alleles as

universal templates for allelic ladder preparation. Forensic Science International: Genetics. 2014 Sep 1; 12:136-43.

72. Lins AM, Micka KA, Sprecher CJ, Taylor JA, Bacher JW, Rabbach DR, Bever RA, Creacy SD, Schumm JW. Development and population study of an eight-locus short tandem repeat (STR) multiplex system. Journal of Forensic Science. 1998 Nov 1;43(6):1168-80.
73. Huel RL, Bašić L, Madacki-Todorović K, Smajlović L, Eminović I, Berbić I, Miloš A, Parsons TJ. Variant alleles, triallelic patterns, and point mutations observed in nuclear short tandem repeat typing of populations in Bosnia and Serbia. Croatian Medical Journal. 2007 Aug;48(4):494.
74. Fujii K, Watahiki H, Mita Y, Iwashima Y, Kitayama T, Nakahara H, Mizuno N, Sekiguchi K. Allele frequencies for 21 autosomal short tandem repeat loci obtained using GlobalFiler in a sample of 1501 individuals from the Japanese population. Legal Medicine. 2015 Sep 1;17(5):306-8.
75. Zgonjanin D, Almheiri R, Alghafri R, Nedić D, Stojiljković G. Forensic and population genetic analysis of Serbian population using 21 STR loci of GlobalFiler™ PCR amplification kit. Forensic Science International: Genetics Supplement Series. 2019 Sep 23.
76. Thanakiatkrai P, Kitpipit T, Gupta R, Sharma S, Babahan I, Coban EP, Biyik H, Alizadeh K, da Silva JA, Babnik A, Petkovsek R. Current STR-based techniques in forensic science. Maejo International Journal of Science and Technology. 2013 Jan 1;7(01):1-5.
77. Hares DR. Expanding the CODIS core loci in the United States. Forensic Science International: Genetics. 2012 Jan 1;6(1): e52-4.

78. Buckleton J, Curran J, Goudet J, Taylor D, Thiery A, Weir BS. Population-specific FST values for forensic STR markers: A worldwide survey. Forensic Science International: Genetics. 2016 Jul 1; 23:91-100.
79. Pinto N, Gusmão L, Egeland T, Amorim A. Paternity exclusion power: comparative behaviour of autosomal and X-chromosomal markers in standard and deficient cases with inbreeding. Forensic Science International: Genetics. 2013 Feb 1;7(2):290-5.
80. Chakraborty R, Stivers DN. Paternity exclusion by DNA markers: effects of paternal mutations. Journal of Forensic Science. 1996 Jul 1;41(4):671-7.
81. Balding DJ, Greenhalgh M, Nichols RA. Population genetics of STR loci in Caucasians. International Journal of Legal Medicine. 1996 Jun 1;108(6):300-5.
82. McBeth M. A Study of the Molecular Basis of Micro variants at the FGA and D21S11 Loci Used in Forensic DNA Testing.
83. Gill P, Brinkmann B, d'Aloja E, Andersen J, Bar W, Carracedo A, Dupuy B, Eriksen B, Jangblad M, Johnsson V, Kloosterman AD. Considerations from the European DNA profiling group (EDNAP) concerning STR nomenclature. Forensic Science International. 1997 Jun 23;87(3):185-92.
84. Bar W, Brinkmann B, Budowle B, Carracedo A, Gill P, Lincoln P, Mayr W, Olaisen B. DNA recommendations further report of the DNA commission of the ISFH regarding the use of short tandem repeat systems. International Journal of Legal Medicine. 1997;110(4):175-6.
85. Gettings KB, Aponte RA, Vallone PM, Butler JM. STR allele sequence variation: current knowledge and future issues. Forensic Science International: Genetics. 2015 Sep 1; 18:118-30.

86. Børsting C, Morling N. Next generation sequencing and its applications in forensic genetics. Forensic Science International: Genetics. 2015 Sep 1; 18:78-89.

87. Röck A, Irwin J, Dür A, Parsons T, Parson W. SAM: String-based sequence search algorithm for mitochondrial DNA database queries. Forensic Science International: Genetics. 2011 Mar 1;5(2):126-32.

88. Kline MC, Hill CR, Decker AE, Butler JM. STR sequence analysis for characterizing normal, variant, and null alleles. Forensic Science International: Genetics. 2011 Aug 1;5(4):329-32.

89. Heinrich M, Felske-Zech H, Brinkmann B, Hohoff C. Characterization of variant alleles in the STR systems D2S1338, D3S1358 and D19S433. International Journal of Legal Medicine. 2005 Sep 1;119(5):310-3.

90. Dauber EM, Kratzer A, Neuhuber F, Parson W, Klintschar M, Bär W, Mayr WR. Germline mutations of STR-alleles include multi-step mutations as defined by sequencing of repeat and flanking regions. Forensic Science International: Genetics. 2012 May 1;6(3):381-6.

91. Wepeba PP, Iyengar A, Goodwin W. Heterozygous 21 STR loci and triplet alleles observed in population genetic analysis of the GlobalFiler STR loci in the Ghanaian population. Forensic Science International: Genetics Supplement Series. 2019 Oct 22.

92. Gjertson DW, Brenner CH, Baur MP, Carracedo A, Guidet F, Luque JA, Lessig R, Mayr WR, Pascali VL, Prinz M, Schneider PM. ISFG: recommendations on biostatistics in paternity testing. Forensic Science International: Genetics. 2007 Dec 1;1(3-4):223-31.

93. Hallenberg C, Morling N. A report of the 2000 and 2001 paternity testing workshops of the English-speaking working group of the international society for forensic genetics. Forensic Science International. 2002 Sep 10;129(1):43-50.
94. Brenner CH. Symbolic kinship program. Genetics. 1997 Feb 1;145(2):535-42.
95. Liem M, Barber C, Markwalder N, Killias M, Nieuwbeerta P. Homicide–suicide and other violent deaths: An international comparison. Forensic Science International. 2011 Apr 15;207(1-3):70-6.
96. Budowle B, Bieber FR, Eisenberg AJ. Forensic aspects of mass disasters: strategic considerations for DNA-based human identification. Legal Medicine. 2005 Jul 1;7(4):230-43.
97. Ge J, Budowle B, Chakraborty R. Choosing relatives for DNA identification of missing persons. Journal of Forensic Sciences. 2011 Jan;56: S23-8.
98. Scally A, Durbin R. Revising the human mutation rate: implications for understanding human evolution. Nature Reviews Genetics. 2012 Oct;13(10):745-53.
99. Wright S. The interpretation of population structure by F- statistics with special regard to systems of mating. Evolution. 1965 Sep;19(3):395-420.
100. National Research Council. The evaluation of forensic DNA evidence. National Academies Press; 1996 Dec 12.
101. Bentayebi K, Abada F, Ihzmad H, Amzazi S. Genetic ancestry of a Moroccan population as inferred from autosomal STRs. Meta gene. 2014 Dec 1; 2:427-38.
102. Jones RJ, Al Tayaare W, Tay GK, Alsafar H, Goodwin WH. Population data for 21 autosomal short tandem repeat markers in the Arabic population of the United Arab Emirates. Forensic Science International: Genetics. 2017 May 1;28: e41-2.

103. Almeida C, Ribeiro T, Oliveira AR, Porto MJ, Santos JC, Dias D, Dario P. Population data of the GlobalFiler® Express loci in South Portuguese population. Forensic Science International: Genetics. 2015 Nov 1; 19:39-41.

104. Al-Eitan LN, Darwish NN, Hakooz NM, Dajani RB. Investigation of the forensic GlobalFiler loci in the genetically isolated Circassian subpopulation in Jordan. Gene. 2020 Apr 5; 733:144269.

105. Gettings KB, Kiesler KM, Faith SA, Montano E, Baker CH, Young BA, Guerrieri RA, Vallone PM. Sequence variation of 22 autosomal STR loci detected by next generation sequencing. Forensic Science International: Genetics. 2016 Mar 1; 21:15-21.

106. Coble MD, Butler JM. Characterization of new miniSTR loci to aid analysis of degraded DNA. Journal of Forensic Science. 2005 Jan 1;50(1): JFS2004216-11.

107. Phillips C, Fernandez-Formoso L, Garcia-Magarinos M, Porras L, Tvedebrink T, Amigo J, Fondevila M, Gomez-Tato A, Alvarez-Dios J, Freire-Aradas A, Gomez-Carballa A. Analysis of global variability in 15 established and 5 new European Standard Set (ESS) STRs using the CEPH human genome diversity panel. Forensic Science International: Genetics. 2011 Jun 1;5(3):155-69.

108. Hill CR, Duewer DL, Kline MC, Sprecher CJ, McLaren RS, Rabbach DR, Krenke BE, Ensenberger MG, Fulmer PM, Storts DR, Butler JM. 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. 2011 Aug 1;5(4):269-75.

109. Alsafiah HM, Goodwin WH, Hadi S, Alshaikhi MA, Wepeba PP. Population genetic data for 21 autosomal STR loci for the Saudi Arabian population using the

GlobalFiler® PCR amplification kit. Forensic Science International: Genetics. 2017 Nov 1;31: e59-61.

110. Alshamali F, Alkhayat AQ, Budowle B, Watson ND. STR population diversity in nine ethnic populations living in Dubai. Forensic Science International. 2005 Sep 10;152(2-3):267-79.

111. Garcia-Bertrand R, Simms TM, Cadenas AM, Herrera RJ. United Arab Emirates: Phylogenetic relationships and ancestral populations. Gene. 2014 Jan 1;533(1):411-9.

112. Ali Alhmoudi O, Jones RJ, Tay GK, Alsafar H, Hadi S. Population genetics data for 21 autosomal STR loci for United Arab Emirates (UAE) population using next generation multiplex STR kit. Forensic Science International: Genetics. 2015 Jul 11; 19:190-1.

113. Alshamali F, Alkhayat AQ, Budowle B, Watson ND. STR population diversity in nine ethnic populations living in Dubai. Forensic Science International. 2005 Sep 10;152(2-3):267-79.

114. Gill P, Haned H, Bleka O, Hansson O, Dørum G, Egeland T. Genotyping and interpretation of STR-DNA: low-template, mixtures and database matches—twenty years of research and development. Forensic Science International: Genetics. 2015 Sep 1; 18:100-17.

115. Westen AA, Kraaijenbrink T, Clarisse L, Grol LJ, Willemse P, Zuniga SB, de Medina EA, Schouten R, van der Gaag KJ, Weiler NE, Kal AJ. Analysis of 36 Y-STR marker units including a concordance study among 2085 Dutch males. Forensic Science International: Genetics. 2015 Jan 1; 14:174-81.

116. Rolf B, Bulander N, Wiegand P. Insertion-/deletion polymorphisms close to the repeat region of STR loci can cause discordant genotypes with different STR kits. Forensic Science International: Genetics. 2011 Aug 1;5(4):339-41.

117. Ochiai E, Osawa M, Tamura T, Minaguchi K, Miyashita K, Matsushima Y, Kakimoto Y, Satoh F. Effects of using the Globalfiler™ multiplex system on parent–child analyses of cases with single locus inconsistency. Legal Medicine. 2016 Jan 1; 18:72-4.

118. Zhang L, Wu H, Wang K, Fan A, Niu J, Zhu Z. Identification of a discordant genotype at the D2S1338 locus between Identifiler Plus® 16 kit and Power Plex® 21 system kit. Journal of Forensic and Legal Medicine. 2016 Nov 1; 44:174-7.

119. Li F, Xuan J, Xing J, Ding M, Wang B, Pang H. Identification of new primer binding site mutations at TH01 and D13S317 loci and determination of their corresponding STR alleles by allele-specific PCR. Forensic Science International: Genetics. 2014 Jan 1;8(1):143-6.

120. Vanderheyden N, Mai A, Gilissen A, Cassiman JJ, Decorte R. Identification and sequence analysis of discordant phenotypes between AmpFlSTR SGM Plus™ and PowerPlex® 16. International Journal of Legal Medicine. 2007 Jul 1;121(4):297-301.

121. Ayres KL, Powley WM. Calculating the exclusion probability and paternity index for X-chromosomal loci in the presence of substructure. Forensic Science International. 2005 May 10;149(2-3):201-3.

122. Graber RA, Morris JW. General equations for the average power of exclusion for genetic system of n co-dominant alleles in one-parent cases of disputed parentage. Inclusion probabilities in parentage testing. 1983:277-80.

123. Gundel H, Reetz I. Exclusion probabilities obtainable by biochemical polymorphisms in dogs. Animal Blood Groups and Biochemical Genetics. 1981 Apr;12(2):123-32.

124. Jamieson A. Electromorphs and erroneous pedigrees. In Proceedings of the XVIth International Conference on Animal Blood Groups and Biochemical Polymorphism. The National Committee of the USSR 1978 (p. 27).

125. Jamieson A. The effectiveness of using co- dominant polymorphic allelic series for (1) checking pedigrees and (2) distinguishing full- sib pair members. Animal Genetics. 1994 Jun;25(S1):37-44.

126. Jamieson A, CS Taylor S. Comparisons of three probability formulae for parentage exclusion. Animal Genetics. 1997 Dec;28(6):397-400.

127. Jamieson A. The genetics of transferrins in cattle. Heredity. 1965 Aug;20(3):419-41.

128. Lee HS, Lee JW, Han GR, Hwang JJ. Motherless case in paternity testing. Forensic Science International. 2000 Nov 13;114(2):57-65.

129. Lee HS, Lee JW, Han GR, Hwang JJ. Motherless Case in Paternity Testing. Korean Journal of Legal Medicine. 1999 May 1;23(1):72.

130. Desmarais D, Zhong Y, Chakraborty R, Perreault C, Busque L. Development of a highly polymorphic STR marker for identity testing purposes at the human androgen receptor gene (HUMARA). Journal of Forensic Science. 1998 Sep 1;43(5):1046-9.

131. Szibor R. X-chromosomal markers: past, present and future. Forensic Science International: Genetics. 2007 Jun 1;1(2):93-9.

132. Szibor R, Krawczak M, Hering S, Edelmann J, Kuhlisch E, Krause D. Use of X-linked markers for forensic purposes. International Journal of Legal Medicine. 2003 Apr 1;117(2):67-74.

133. Zambrano AK, Vaca-Pólit M, Boada L, Vela M, Rodríguez-Pólit C, Fiallos G, Gruezo C, Paz-y-Miño C, Gaviria A. X-STR decaplex study in the population of Imbabura-Ecuador. Forensic Science International: Genetics Supplement Series. 2019 Oct 1.

134. Barber MD, McKeown BJ, Parkin BH. Structural variation in the alleles of a short tandem repeat system at the human alpha fibrinogen locus. International Journal of Legal Medicine. 1996 Dec 1;108(4):180-5.

135. Jiang W, Kline M, Hu P, Wang Y. Identification of dual false indirect exclusions on the D5S818 and FGA loci. Legal Medicine. 2011 Jan 1;13(1):30-4.

136. Griffiths RA, Barber MD, Johnson PE, Gillbard SM, Haywood MD, Smith CD, Arnold J, Burke T, Urquhart AJ, Gill P. New reference allelic ladders to improve allelic designation in a multiplex STR system. International Journal of Legal Medicine. 1998 Jul 1;111(5):267-72.

137. Rolf B, Schürenkamp M, Junge A, Brinkmann B. Sequence polymorphism at the tetranucleotide repeat of the human beta-actin related pseudogene H-Beta-Ac-psi-2 (ACTBP2) locus. International Journal of Legal Medicine. 1997 Apr 1;110(2):69-72.

138. Herzog U, Augustin C, Püschel K. Allele frequencies of six miniSTR markers in a population sample from Northern German and its application on forensic stain analysis. Forensic Science International: Genetics Supplement Series. 2008 Aug 1;1(1):331-3.

139. Kalaoglu E, Ay M, Ulubay A, Canan H, Serin A, Alper B. Population data of 12 X-STRs in Turkey. Forensic Science International: Genetics Supplement Series. 2019 Oct 18.

APPENDIX 1

Appendix 1: Backlog of DNA cases at the NFSI.

Year	Total
1993	2
1998	5
1999	86
2000	33
2001	8
2002	137
2003	146
2004	136
2005	143
2006	130
2007	6
2008	49
2009	67
2010	146
2011	81
2012	120
2013	134
2014	173
2015	237
2016	241
2017	275
2018	569
2019	511
2020	279
Total	**3714**

APPENDIX 2

Appendix 2: Allele frequency distribution of 21 STR loci in Black population (n = 150).

Allele	D8S1179	D21S11	D7S820	CSF1PO	D3S1358	TH01	D13S317	D16S539	D2S1338	D19S433	VWA	TPOX	D18S51	D5S818	FGA	D2S441	D22S1045	SE33	D10S1248	D1S1656	D12S391
5					3			0.01													
6				0.04		0.1						0.073									
7			0.007	0.043		0.373						0.03							0.013		
8	0.03		0.247	0.08		0.347	0.013	0.013		0.037				0.383						0.033	
9			0.123	0.02		0.14															
10			0.303	0.243		0.003	0.033	0.033		0.097	0.01		0.073	0.003		0.077	0.047			0.08	
10.2	0.03													0.003							
10.3			0.207																		
11				0.213		0.003	0.297	0.03		0.296	0.283	0.04	0.01	0.24		0.037	0.163		0.037	0.037	0.203
11.2	0.067		0.093	0.267														0.003			
12	0.177		0.02	0.08			0.093	0.267		0.243		0.44	0.163	0.1		0.047	0.01	0.063	0.127	0.047	0.243
12.3							0.03														
13	0.447			0.01	0.07			0.177	0.02	0.08			0.153	0.153		0.01	0.273			0.053	0.3
13.2									0.08	0.007											
14	0.177			0.003	0.243		0.033	0.023	0.253	0.09	0.447		0.01	0.073			0.033	0.093	0.25	0.017	0.043
14.2										0.107	0.003	0.01									
15	0.083				0.313		0.167	0.147	0.19	0.003	0.147			0.177			0.003	0.003	0.147	0.16	
15.2												0.057	0.017			0.007					
15.3																	0.013				
16	0.02				0.283		0.247		0.007	0.17	0.107	0.097	0.313	0.107				0.247	0.107	0.107	0.267
16.2															0.013						
17					0.087			0.07	0.003	0.02	0.27		0.177		0.003		0.23	0.003	0.027	0.27	0.003
17.3							0.017	0.003													
18									0.08		0.087			0.007	0.043			0.157		0.08	
18.2											0.017										
18.3													0.013								
19											0.05			0.11	0.183			0.05			
19.2									0.003												
19.3											0.003		0.003					0.003			
20											0.083		0.003	0.087	0.023			0.087			
20.2		0.017																			
21		0.01							0.197						0.003			0.197			
21.2		0.23							0.007						0.003			0.003			
21.3									0.003									0.003			

Allele	D8S1179	D21S11	D7S820	CSF1PO	D3S1358	TH01	D13S317	D16S539	D2S1338	D19S433	VWA	TPOX	D18S51	D5S818	FGA	D2S441	D22S1045	SE33	D10S1248	D1S1656	D12S391
22		0.153							0.007						0.167						
22.2		0.01													0.003			0.003			
22.3															0.003			0.003			
23		0.063													0.003						
23.2		0.093																			
24		0.003													0.043						
24.2		0.023													0.177						
24.3		0.023													0.017			0.017			
25		0.047																			
25.2															0.033			0.033			
26		0.003																			
26.2															0.02			0.02			
27		0.003													0.023						
28															0.01			0.23			
28.2															0.043						
29															0.153						
29.2															0.017						
30															0.123						
30.2																		0.003			
31																		0.063			
32																					
33																		0.03			

APPENDIX 3

Appendix 3: Allele frequency distribution of 21 STR loci in Coloured population (n = 150).

Allele	D8S1179	D21S11	D7S820	CSF1PO	D3S1358	TH01	D13S317	D16S539	D2S1338	D19S433	VWA	TPOX	D18S51	D5S818	FGA	D2S441	D22S1045	SE33	D10S1248	D1S1656	D12S391
6				0.027		0.11						0.02									
7			0.01	0.04		0.23				0.017		0.01							0.003		
8	0.007		0.187	0.057		0.363	0.033	0.02				0.383		0.023						0.007	
9			0.11	0.04		0.187	0.037	0.173				0.29		0.047							
9.3						0.09															
10	0.013		0.277	0.263		0.02	0.063	0.103		0.007		0.057	0.003	0.05		0.1	0.043			0.007	
10.3																0.003					
11	0.08		0.21	0.187			0.247	0.33		0.053	0.007	0.22	0.013	0.227		0.36	0.187		0.03	0.123	
11.2																0.057		0.007			
12	0.103		0.177	0.29	0.003		0.427	0.23		0.11		0.02	0.067	0.4		0.16	0.027	0.003	0.103	0.06	
12.2										0.017								0.023			
12.3																0.02					
13	0.167		0.023	0.09			0.12	0.133	0.003	0.24	0.017		0.07	0.23		0.063	0.007		0.307	0.133	
13.2										0.087											
14	0.303		0.007	0.007	0.1		0.073	0.01		0.206	0.11		0.053	0.02		0.16	0.073	0.017	0.247	0.143	
14.2										0.09								0.013		0.003	
14.3																				0.023	
15	0.203				0.283					0.11	0.153		0.107			0.053	0.257	0.03	0.163	0.23	0.047
15.2										0.033			0.03					0.017			
15.3																				0.023	
16	0.097				0.293				0.05	0.007	0.237		0.113			0.017	0.217	0.043	0.123	0.14	0.057
16.2										0.017								0.013			
16.3																				0.017	
17	0.017				0.223				0.11	0.003	0.247		0.167	0.003			0.177	0.063	0.023	0.057	0.157
17.3																				0.013	
18	0.01				0.097				0.07		0.173		0.133		0.003		0.003	0.127		0.003	0.17
18.2										0.003					0.003						
18.3																				0.01	0.007
19									0.163		0.037		0.127		0.08			0.143			0.2
19.1																		0.003			0.003
19.2															0.003						
19.3																		0.003		0.003	
20									0.12		0.02		0.043		0.097		0.01	0.123			0.127
20.2															0.003			0.013			
21									0.133				0.037		0.077			0.043			0.027
21.2															0.003			0.007			

Allele	D8S1179	D21S11	D7S820	CSF1PO	D3S1358	TH01	D13S317	D16S539	D2S1338	D19S433	VWA	TPOX	D18S51	D5S818	FGA	D2S441	D22S1045	SE33	D10S1248	D1S1656	D12S391
21.3															0.003						
22									0.127				0.023		0.123			0.003			0.083
22.2															0.003			0.003			
23									0.067						0.183						0.053
23.2															0.003			0.013			
24									0.06				0.013		0.147						0.03
24.2																		0.027			
25									0.063						0.143						0.03
25.2																		0.02			
26									0.027						0.043						0.003
26.2																		0.023			
27		0.055							0.007						0.043						0.007
27.2																		0.077			
28		0.176																			
28.2																		0.087			
29		0.138													0.01						
29.2																		0.013			
30		0.245																			
30.2		0.017													0.003			0.027			
31		0.052																			
31.2		0.031													0.003			0.007			
32		0.031																			
32.2		0.124																0.007			
33		0.021																			
33.1		0.003																			
33.2		0.017																			
34		0.034																			
34.1		0.007																			
34.2		0.007																			
35		0.007																			
35.1		0.017																			
36		0.007																			
36.1		0.007																			
37																					
37.2																					
38		0.003																			
39																					
40																					
41																					
42																					
42.2																					

Allele	D8S1179	D21S11	D7S820	CSF1PO	D3S1358	TH01	D13S317	D16S539	D2S1338	D19S433	VWA	TPOX	D18S51	D5S818	FGA	D2S441	D22S1045	SE33	D10S1248	D1S1656	D12S391
43.2															0.01						