

FORENSIC HUMAN IDENTIFICATION: GENERATING Y-STR DATA FOR THE SOUTH AFRICAN POPULATION

Kate Megan Reid

RDXKAT001

MSc (Med) Forensic Genetics

Supervisor: Laura Heathfield

Co-supervisor: Professor Lorna Martin

August 2018

Word count (body): 30 273



*Division of Forensic Medicine and Toxicology, Department of Pathology, Faculty of Health
Sciences, University of Cape Town*

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

PLAGIARISM DECLARATION

1. I know that plagiarism is wrong. Plagiarism is using another's work and to pretend that it is one's own work.
2. I have followed the referencing style according to the Forensic Science International journal and I have used the Forensic Science International referencing style as the convention for citation and referencing. Each significant contribution to, and quotation in, this proposal from the work, or works of other people has been attributed, cited and referenced.
3. This proposal is my own work.
4. I have not allowed, and will not allow, anyone to copy my work with the intention of passing it off as his or her own work.
5. I acknowledge that copying some else's assignment or essay, or part of it, is wrong, and declare that this is my own work.

Name: Kate Megan Reid

Date: 2018/08/28

Signature:

Signed by candidate

ABSTRACT

Salt River Mortuary (SRM), Cape Town, investigates ~3500 cases of unnatural death annually, with an apparent burden of unclaimed bodies. A retrospective review was first undertaken to assess the number of these individuals who remained unidentified. Medico-legal records were examined (2010-2017), and ~9% of cases remained unidentified each year. DNA analysis was performed in 23.5% of cases. At the time of this study, unidentified bodies were in storage for up to two years, pending pauper burial.

DNA profiling assists forensic human identification, and the analysis of markers on the Y-chromosome has particular importance in kinship analysis. To evaluate the statistical probability of DNA profiles matching between samples, reference data from the background population is required. Such data for the Y-chromosome is lacking for some population groups in South Africa (SA). As such this study aimed to generate Y-chromosome data relevant to SA. Second to this, the obtainability of DNA profiles from unidentified decedents at SRM, prior to pauper burial, was investigated.

Biological samples were obtained from 653 SA individuals (living: n=480; deceased: n=173) belonging to four major population groups. Following internal validation, samples were processed using the Promega PowerPlex® Y23 System. A cohort-representative subset of DNA profiles were also generated using the forensically validated Next Generation Sequencing (NGS) assay on the MiSeq FGx™ system, to assess concordance. Statistical analysis was performed using Arlequin and STATA packages.

Full DNA profiles (i.e. haplotypes) were obtained from 626 samples (African: n=183; Coloured: n=170; Indian/Asian: n=111; White: n=162), with 599 haplotypes being unique to a single individual. Following optimisation, haplotypes were obtained from >99% and 85% of living and deceased individuals, respectively. Haplotypes were generated from numerous individuals stored for over one year, and DNA profile quality was not associated with time between death declaration and sample collection. NGS results confirmed the presence of one micro-variant and resolved allele-calling in five instances where the capillary electrophoresis assay was incorrect. Thus, concordance was observed in 98% of loci reviewed.

Overall, haplotypes were successfully obtained for four different SA population groups, including refrigerated decedents, even 887 days after death declaration. This demonstrates that DNA profiling can be successful for decedents and efforts should be made to store DNA profiles for the possibility of familial searching and identification, even after burial. Identification of the multitude of unclaimed bodies at forensic facilities nationwide holds immense value for living family members, and provides closure for the acceptance of death and life thereafter.

ACKNOWLEDGEMENTS

I would not have been able to complete this Master's dissertation without the support I received over the past year and a bit. This degree was made possible through the funding I received from the Oppenheimer Memorial Trust, and the National Research Foundation (disclaimer below).

Firstly, I would like to say a huge thank you to my supervisor, Laura Heathfield, for the constant support, guidance and advice over my time in the lab. I cannot say how much I appreciate all the time and effort you have invested in both the project and me.

Secondly, to my co-supervisor, Professor Lorna Martin, your guidance and knowledge helped me so much and I appreciate the insight you provided.

To my collaborator, Stefan Kutranov at Promega Corporation, your technical assistance and support was invaluable to the success of the project.

I would also like to thank Dr Vicky Gibbon (FACT), Devin Finaughty (CapeForte) and Wayne Mitten (FPS) for their assistance in obtaining relevant information, and providing context of the data received. You all provided key points to consider that have aided in strengthening the value of the research.

Lastly, I would like to thank my family for the love, enthusiasm, support and guidance provided throughout this journey. You have helped me to stay focused, motivated and excited over this year. To my brother, I appreciate all the time you selflessly sat and assisted with computer and statistics problems.

The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged, expressed and conclusions arrived at are those of the author and are not necessarily to be attributed to the NRF.

TABLE OF CONTENTS

Abbreviations.....	iv
Glossary.....	v
Table of figures.....	viii
Table of tables.....	ix
Table of equations.....	ix
CHAPTER 1: Introduction.....	1
1.1 Synopsis.....	1
1.2 Literature review.....	2
1.2.1 Development of DNA profiling.....	2
1.2.2 Applications of Y-STR profiling.....	5
1.2.2.1 Sexual offence cases.....	5
1.2.2.2 Kinship analysis.....	9
1.2.3 DNA profiling in the deceased.....	13
1.2.4 Y-STR profiling performed in South Africa.....	19
1.3 Rationale.....	21
1.4 Aims and objectives.....	23
1.4.1 Aim.....	23
1.4.2 Objectives.....	23
1.4.3 Ethics approval.....	24
CHAPTER 2: Retrospective review of unidentified decedents at Salt River Mortuary (2010 – 2017) ..	25
2.1 Introduction.....	25
2.1.1 Identification of decedents in South Africa.....	25
2.1.2 The burden of unidentified persons worldwide.....	26
2.2 Methods.....	29
2.3 Results.....	31
2.3.1 Overview.....	31
2.3.2 Demographics.....	32
2.3.3 Location of recovery.....	33
2.3.4 Admission circumstances, post-mortem investigation and cause of death.....	35

2.4 Discussion.....	37
2.4.1 Number of unidentified per annum, and circumstances surrounding death.....	37
2.4.2 Attempts to identify the decedent	39
2.4.3 Recommendations and future research studies.....	42
2.5 Conclusion.....	45
CHAPTER 3: Methods and Materials.....	46
3.1 Study design.....	46
3.2 Cohort and sample collection	46
3.2.1 Sample collection from living participants.....	47
3.2.2 Sample collection from unidentified decedents.....	47
3.3 Sample processing	48
3.4 Sample quantification	48
3.5 Internal Validation	49
3.5.1 Spectral calibration	50
3.5.2 Amplification reaction and capillary electrophoresis	50
3.5.3 Threshold analyses.....	51
3.6 Standardised DNA profiling assay of samples.....	54
3.7 Data analysis	54
3.8 Next generation sequencing (MiSeq FGx™ system)	57
CHAPTER 4: Results	59
4.1 Cohort	59
4.2 Collection and quantification of extracted DNA samples.....	61
4.3 DNA profiling.....	62
4.3.1 Success rate of Y-STR profiling.....	62
4.3.2 Additional findings	64
4.3.3 Haplotype Frequencies	66
4.4 Gene Diversity and Linkage.....	68
4.5 Population variation.....	70
4.6 Next generation sequencing.....	72
CHAPTER 5: Discussion.....	74
5.1 Cohort composition	74
5.2 Performance of DNA profiling.....	76

5.3 DNA profiling in the deceased	77
5.3.1 DNA degradation and PCR inhibition	77
5.3.2 Flocked nylon and cotton swabs.....	80
5.4 Profile frequencies and gene diversity	82
5.4.1 Profile frequency comparison to South African literature	82
5.4.2 Profile frequency comparison to international database and literature.....	83
5.4.3 Gene diversity comparisons to local and international literature	85
5.5 The value of next generation sequencing technology.....	86
5.5.1 Coverage	86
5.5.2 Concordance analysis.....	87
5.6 Limitations and future studies	89
5.7 Conclusion.....	91
REFERENCES	93
APPENDICES	a
Appendix A.....	a
Appendix B	b
Appendix C	f
Appendix D.....	i
Appendix E	j
Appendix F	m
Appendix G.....	n
Appendix H.....	n
Appendix I	o
Appendix J.....	q
Appendix K.....	r

ABBREVIATIONS

Abbreviation	Meaning
%	Percentage
~	Approximately
<	Less than
=	Equals to
>	More than
A	Adenine
AMOVA	Analysis of molecular variance
bp	Base pairs
C	Cytosine
DNA	Deoxyribonucleic acid
e.g.	Example
Etc	Etcetera
FBI	Federal Bureau of Investigation
FFPE	Formalin Fixed Paraffin Embedded
FPS	Forensic Pathology Services
G	Guanine
mtDNA	Mitochondrial DNA
N	Sample size
ng/pg	Nanogram/pictogram
NGS	Next Generation Sequencing
PCR	Polymerase chain reaction
RFLP	Restriction fragment length polymorphism
RM Y-STR	Rapid mutation Y-chromosome STR
SA	South Africa
SAPS	South African Police Service
SNP	Single nucleotide polymorphism
STRs	Short tandem repeats
SWGDM	Scientific working group on DNA analysis methods
T	Thymine
UK	United Kingdom
USA	United States of America
VIC	Victim Identification Centre
VNTR	Variable number tandem repeat
YHRD	Y-chromosome haplotype reference database
Y-STR	Y chromosomal STR

GLOSSARY

Within the definition, terms written in italics indicate those included in the glossary list as well

Term	Meaning
Admixed (ure)	Population group showing genetic composition originating from multiple ancestral origins.
Allele	Genetic identity of a specific region of DNA specific to an individual. In this study, alleles refer to the number of repeats observed for each <i>STR</i> marker.
Allele frequency	Percentage of proportion that an <i>allele</i> has within the reference population data.
Allele drop out	An anomaly observed in capillary electrophoresis, wherein amplification of the <i>target region</i> fails due to genetic variability in the <i>primer</i> binding region.
Allelic Ladder	Synthetically manufactured sample containing fragments of DNA representing the most commonly observed <i>alleles</i> for specific <i>markers</i>
Amelogenin	Genetic region on the sex chromosomes (X and Y) that is used to type the sex of an individual during <i>DNA profiling</i> .
Analytical threshold	This represents the height (relative fluorescent units) that a peak on an <i>electropherogram</i> should meet in order for the peak to be assigned an allelic identity. The analytical threshold differentiates background noise from true peaks with a level of confidence
Anthropometric	Use of the principles of anthropology (measurement of human remains) to predict <i>biogeographical ancestry</i> , sex and age.
Autosomal	Chromosomes that contain genetic information that is not specific to a certain sex.
Bin-set	Analysis file that contains information for known alleles possible for a specific <i>marker</i> , which is used to assign allelic identity to peaks of an <i>electropherogram</i> . Alignment for the assignment of allelic identity is based off the <i>allelic ladder</i> .
Biogeographical ancestry	Prediction of an individual's appearance or demographic group based on genetic regression analysis. Often misreported as an individual's race.
Cluster density	Number of clusters per square millimetre of the flow cell used in the <i>NGS</i> run. Optimal value for the workflow used in this study is between 400 and 1650 K/mm ²
Cluster passing filter	Percentage of clusters passing the chastity filter (>80%) during <i>NGS</i>
Concordance	In this study, concordance refers to the similarity of allelic identities for markers assigned between <i>DNA profiling</i> and <i>NGS</i>
Coverage	Number of reads per fragment generated by <i>NGS</i> for analysis
Crude lysate	Refers to the solution of lysis buffer and cotton swab, without the performance of any purification or extraction
Cycle threshold	During real-time PCR, the cycle threshold (Ct) is the point at which the amount of fluorescence in a sample reaching the pre-set analysis minimum for the determination of DNA concentration.
Direct PCR (assay)	Processing of <i>crude lysate</i> immediately into an amplification reaction
DNA degradation	Fragmentation of DNA due to the presence of adverse environmental, enzymatic and bacterial conditions

DNA profile	String of allele identities obtained for an individual following the performance of DNA profiling. Results are viewed as an <i>electropherogram</i> . In this study the following types of profiles were find: Full: 23/23 <i>markers</i> met the <i>analytical threshold</i> Partial: 3-22 <i>markers</i> met the <i>analytical threshold</i> Failed: <3 <i>markers</i> met the <i>analytical threshold</i>
Dye channel	Separation of <i>markers</i> fluorescently tagged according to the type of fluorescence used, resulting in an <i>electropherogram</i> with five different lines.
Dye-blob	Capillary electrophoretic anomaly where the fluorescent dye used to tag <i>primers</i> specific to a <i>marker</i> results in a peak near the beginning of the <i>dye channel</i> in an <i>electropherogram</i> .
Electropherogram	Visual output of DNA profiling, where the X axis represents fragment length and the Y axis indicates relative fluorescent units.
Extracted DNA (assay)	Performance of DNA profiling (PCR and CE) on biological samples that have undergone DNA extraction and quantification
Gene diversity	Refers to the number of different <i>alleles</i> observed at a particular <i>marker</i> analysed in this study. Calculation thereof is according to equation 3.3.
Haplotype	Set of allelic identities for various <i>markers</i> specific to an individual
Internal positive control	A synthetic piece of DNA included in qPCR assay, used to assess the presence of PCR inhibitors from the sample being quantified.
Internal validation	Process in which a laboratory determines optimal assay conditions for the performance of specific assay on a specific instrument
Kinship	The study of familial relationships (e.g. paternity)
Locus/Markers/Target region	Section of DNA that is of interest and therefore amplified and analysed using molecular techniques
Micro-variant	An incomplete repeat unit within an <i>STR</i>
Mutation rate	Rate at which the genetic composition of a specific <i>marker</i> changes at the sequence level. Measured between populations when referring to Y-chromosomal data
Negative control	A sample that undergoes the full molecular assay that is devoid of any biological material. In many cases molecular biology grade water is used.
Next generation sequencing (NGS)	Molecular process whereby multiple <i>target regions</i> (SNPs and <i>STRs</i>) are amplified and typed through the use of massively parallel sequencing.
Non-coding regions	Areas of the DNA that do not encode for an amino acid change, and therefore do not produce a protein.
Non-viable foetus	A foetus under 28 weeks of gestation
Null allele	Absence of a <i>marker</i> due to genetic variation and mutation (also referred to as a deletion)
Off-ladder	Peaks falling between bins of the <i>bin set</i> (on an <i>electropherogram</i>) and thereby not being assigned an allelic identity. Usually these off ladder peaks are <i>microvariants</i>
PCR inhibition (ors)	Presence of severely degraded DNA (debris) or other biological artefacts that prevent successful amplification of the <i>target region</i>
Phasing	Percentage of molecules within a cluster falling behind the current read cycle in NGS. A low percentage (<0.25%) is desirable.

Phenotype (ic)	Observable or physical characteristics of an individual determined by the genetic composition of the individual
Positive control	A DNA sample of known genetic composition, used to assess the accuracy of the assay performed
Post-mortem interval (PMI)	The amount of time elapsing between death and performance of the post-mortem investigation
Prephasing	Percentage of molecules within a cluster occurring ahead of the current read cycle during <i>NGS</i> . A low percentage (<0.15%) is desirable.
Primer (binding region)	Primers are synthetically generated DNA sequences that are complementary to the <i>target region</i> and assist in the amplification. The primer binding region is the area where the primer attaches to the template DNA through sequence complementarity.
Short tandem repeat (STR)	A repetitive region of DNA consisting of a repetitive unit of between 3 and 6 bases pairs.
Singleton	A <i>haplotype</i> that has only been observed in one individual within a population
Standard operating procedure	A uniform validated protocol established within a laboratory or facility, and based off of international recommendation, specific to the performance of particular assay or assessment.
Stutter	Erroneous peaks visible on an <i>electropherogram</i> either one repeat unit above or below that of the true allele
Success rate	The number of full profiles obtained as a ratio of the total number of samples analysed
Y-STR	<i>STRs</i> located on the Y-chromosome (male-specific)

TABLE OF FIGURES

Figure 2-1: Diagram indicating all sixteen forensic pathology service facilities in the Western Cape, South Africa.....	27
Figure 2-2: Total caseload observed at Salt River Mortuary and the number of unidentified decedents across the eight year review period (2010-2017).....	31
Figure 2-3: Distribution of unidentified cases at Salt River Mortuary from 2010 to 2017 according to age category of decedent	32
Figure 2-4: Heat maps indicating the area in which the decedent was retrieved as per the SAPS station to which the case was reported	33
Figure 2-5: Distribution of cases as per the scene from which the decedent was recovered.....	34
Figure 2-6: Schematic diagram of the category of death under which the cases are admitted to and resolved at Salt River Mortuary	35
Figure 2-7: Type of biological sample obtained at autopsy by the pathologist for the performance of DNA analysis.....	36
Figure 3-1: Schematic workflow summary of internal validation performed for the Promega PowerPlex®Y23 system	50
Figure 4-1: Heatmap indicating province from which participants, recruited in this study, originated from.....	59
Figure 4-2: Number of participants recorded according to home language reported.....	60
Figure 4-3: Electropherograms showing variable quality of DNA profiles obtained from a) lysate sample from living individual, b) post-mortem extracted DNA and c) post-mortem lysate sample....	63
Figure 4-4: Images of the DYS576 marker from three different samples indicating the presence of the dye blob (indicated by the arrow)	64
Figure 4-5: Sample showing deletion at DYS448 and consequent duplication at DYS576.....	64
Figure 4-6: Linage disequilibrium scores obtained for the full cohort of full DNA profiling (n=626) ...	69
Figure 4-7: Matrix of pairwise F_{ST} values determined using Arlequin v3.1 when comparing the four population groups.....	70
Figure 4-8: Scatter plot of number of markers meeting the analytical threshold compared to the time between death and sample collection.....	71
Figure 4-9: Electropherogram results for one dye channel depicting the variation in peak height and ski slope effect in a (A) flocked and (B) cotton swab sample	72
Figure 5-1: Line diagram indicating the various time intervals of importance when processing post-mortem samples	78

TABLE OF TABLES

Table 3-1: Amplification reaction conditions for both direct PCR and extracted DNA approaches.....	51
Table 3-2: Polymerase Chain Reaction setup for lysate and extracted DNA samples.....	54
Table 3-3: Summary of statistical tests performed between different variables of interest	56
Table 4-1: Number of samples obtained as categorised by life status, self-reported population group and sample type.....	61
Table 4-2: Number and type of Y-STR profile obtained from living and deceased individuals across the two sample type groups (extracted DNA and lysate).....	62
Table 4-3: Microvariants and novel allelic variants observed in eight different Y-STR markers, and their number of observations	65
Table 4-4: Allelic combination of haplotypes generated using the Promega PowerPlex® Y23 system, which are shared between multiple individuals	66
Table 4-5: Summary of shared haplotypes per population group and commonness on YHRD	67
Table 4-6: Gene diversity (GD) determined for each marker and the DYS385a/b haplotype	68
Table 4-7: Summarised NGS phenotype prediction and ancestry prediction results obtained using the ForenSeq DNA Signature Prep assay (NA = not available).....	73

TABLE OF EQUATIONS

Equation 3.1: Stutter ratio calculation.....	52
Equation 3.2: Marker specific stutter ratio calculation	52
Equation 3.3: Gene diversity.....	56

CHAPTER 1: INTRODUCTION

1.1 SYNOPSIS

In forensic science, the use of DNA evidence focuses on the identification of both victims and suspects, with the aim of placing/linking an individual on/with a scene; reuniting a missing or deceased individual with their family; or identifying an unknown individual [1]. For this to occur, a DNA profile for an individual is generated and matched to reference DNA profile from a known individual. However, for the DNA profile to have evidentiary value in court, and assist in criminal and/or medico-legal proceedings, statistical analyses need to be performed to determine the likelihood of two DNA profiles matching by chance [2]. These calculations rely on frequency data from the background population, specifically pertaining to the regions analysed during DNA profiling.

The majority of DNA profiles generated and referred to in legal proceedings, are those obtained using non-coding regions (with no associated phenotypic information) on autosomal (non-sex) chromosomes, except for the sex of an individual from the amelogenin marker [1]. In some situations DNA profiles are challenging to generate and/or interpret, due to numerous biological and environmental factors which will be discussed later in this chapter. These situations would benefit from the generation of an alternative type of DNA profile, such as Y-chromosome profiling, X-chromosome profiling or mitochondrial DNA sequencing. Y-chromosome profiling utilises a similar workflow and principles as autosomal DNA profiling and hence also requires relevant reference data for the performance of statistical analyses. However, reference data pertaining to the Y-chromosome is severely lacking in South Africa (SA) which limits its use in forensic applications.

This literature review has two main aims; the first is to describe the applications of Y-STR profiling and highlight the importance of Y-chromosome profiling. To this end, the principles and development of conventional DNA and Y-chromosome profiling will be described. Secondly, value of DNA profiling in the forensic post-mortem setting will be discussed, as well as factors which affect the quality of results. The purpose of discussing these various topics, is to provide an overview of how Y-chromosome DNA profiles can be utilised in criminal investigations and forensic applications. Furthermore, it is important that complications faced with DNA profiling in the deceased be discussed, so as to provide an understanding of what problems may be encountered. Thereafter, the rationale, aim and objectives of this research study will be provided.

1.2 LITERATURE REVIEW

1.2.1 Development of DNA profiling

Historically the performance of genotyping studies relied on the identification of suitable restriction enzymes and performance of restriction fragment length polymorphism (RFLP) analyses. Most genetic studies focused their research questions on a particular disease and resolution of the causes thereof. In 1980, Wyman *et al.* identified a polymorphic region of DNA, devoid of encoding any phenotypic information, that would be suitable for the tracing of human pedigrees and potentially the identification of individuals [3]. This was the first time that any genetically based study recognised the potential to identify individuals. The principles investigated in this study were developed and led to the start of DNA “fingerprinting” [4]. DNA “fingerprinting” involved the amplification and separation of mini-satellite regions of DNA, in order to obtain a gel electrophoresis image specific to an individual [4,5]. This process was noted to have significant value in the identification of cell lines, pedigree analyses, paternity testing and various uses in genetic medicine [4].

RFLP and DNA “fingerprinting” have since been used in the investigation of sexual assault cases, wherein biological samples obtained often contained DNA originating from multiple individuals [6]. Traditionally, the investigation of sexual assault cases involved the performance tests to confirm the presence of spermatozoa and thus confirmation of the occurrence of sexual activity, however these tests provided no assistance in the identification of the perpetrator [7]. Giusti *et al.* (1986) investigated the use of RFLP to match DNA from semen samples to DNA from blood samples, when obtained from the same individual [7]. It was found that the DNA banding patterns obtained for the two sample types were concordant, therein providing value to the potential use in future criminal investigations [7].

Continual development of RFLP aimed to reduce the amount of input DNA required, as well to simultaneously co-amplify multiple fragments. The advent and use of thermostable *Taq* polymerase improved the capabilities of polymerase chain reaction (PCR) and DNA “fingerprinting” [8]. In 1988, the successful co-amplification reaction of six mini-satellite markers determined the likelihood of two unrelated individuals having fully concordant DNA “fingerprints” to be approximately one in every 49 000 individuals (2.0×10^{-5}) [8]. While significantly lower than the discriminatory value of locus-specific mini-satellite probes (1 in every 100 000), no obvious challenges were identified that would prevent the inclusion of more mini-satellite markers in PCR-based DNA “fingerprinting”, in order to increase the evidentiary power [8].

Following the advances in RFLP and extensive use in forensic casework, the Technical Working Group on DNA Analysis Methods (TWGDAM, now known as SWGDAM), was established in 1988 in North America. Attendees of this meeting represented 31 laboratories in the United States of America (USA) and Canada, and proposed guidelines for the standardisation of the performance of RFLP [9]. The release of these guidelines aimed to ensure the validation of laboratory procedure and to assist in the sharing of general practices.

Concurrently to the development of DNA “fingerprinting” techniques, studies were being conducted to evaluate the use of microsatellite markers, or short tandem repeats (STRs) [10]. Microsatellite markers consists of smaller repeat motifs (3-6 bp) compared to minisatellite regions (10-100 bp) [1]. The term “DNA profiling” was subsequently adopted to refer to applications of STR profiling, so as to avoid confusion with the nature of DNA “fingerprinting”.

Additionally, much research was being conducted in to the development and use of mini and micro satellite markers specific to the Y-chromosome (Y-STRs). Motivation for these studies focussed on improving exclusion criteria in paternity and forensic case work, as well as to improve understanding of genealogies and linkage studies [11,12]. These studies focused on the (GATA)_n repeat unit on the Y-chromosome and identified up to eight different alleles amongst males [11]. Use of the Y-GATA marker alone determined the exclusion of paternity to be 0.66, which means that in any particular paternity case, the alleged father could be excluded with a confidence value of 66% [12].

Throughout the advances in DNA profiling, the need for a reference database was always supported. The first national DNA database was established in the United Kingdom (Britain and Wales) in 1995, (UK NDNAD) [13]. Eight autosomal markers were included in the database, in accordance with the markers included in the commercially available Second Generation Multiplex (SGM) kit. Thereafter, the Federal Bureau of Investigation (FBI) released the CODIS (Combined DNA Index System) thirteen autosomal markers, for the incorporation into the United States of America National DNA index System (NDIS) [14].

As national databases were being established for autosomal markers, continuous development was underway for the use of Y-STR markers in various applications. The first forensic Y-User workshop was held in Berlin in 1996, and assisted with the development of the Y-STR Haplotype Reference Database (YHRD). Therein it was also discussed what the minimal haplotype (YHRD1) should entail, which was subsequently announced in 1999. Since then, three updates of the YHRD have occurred in 2004, 2008 and 2014 respectively, owing to the increasing amounts of research conducted on the topic. Various other Y-STR

databases now exist, including the USA Y-STR database which was last updated in February 2017, and contains 35 660 Y-STR haplotypes, categorised according to four commercially available kits [15].

With specific attention to Y-STRs, Hanson *et al.* (2006) conducted a study to comprehensively review all the markers that have been reported and recorded on various public databases [16]. In the study it was concluded that a minimum of 417 individual Y-STR markers were available for potential use in forensic application and casework. However, it was noted that a significant number of these markers may prove unsuitable for forensic work due to the presence of overlapping regions, low genetic variance or high homology with the X-chromosome. The location of these markers was annotated onto a Y-chromosome thereby indicating the chromosomal segment on which the marker lies. Evaluation of marker location revealed that in some cases, markers overlapped and as such primer design should be optimised to flank such regions if required. Furthermore, the authors identified a number of markers that fell within coding regions of the Y-STR and should thus be excluded from Y-STR profiling kits due to their ability to reveal phenotypic information of the individual, which may contravene various guidelines and legislation [16]. It was distinguished that the location of the markers depended on the accuracy of the databases utilised, and thus updates to the annotated Y-chromosome should be conducted annually [16].

Throughout the development of human identification techniques, there was an underlying goal of developing automated machines, so as to reduce risk of human error and to improve turnaround time. The first automated DNA profiling system made use of the DNA sequencer 370 (Applied Biosystems) which was released in 1986. This system incorporated fluorescently tagged primers for amplification of target regions and separation of amplicons using polyacrylamide gel electrophoresis [17]. While this system allowed for better analysis, determination of precision and standard deviation, it was still a long procedure, wherein electrophoresis could take up to eight hours [17].

Currently, commercial kits are available for the performance of autosomal, Y- and X-chromosome profiling, from various leading manufacturers such as Promega Corporation, Applied Biosystems (Thermo Fisher Scientific) and Qiagen. Autosomal DNA profiling kits now include up to 35 markers [18], some with additional quality sensors and controls built in [19]. Similarly, Y-STR profiling kits now contain up to 27 markers, designed to resolve more complex mixtures and work on trace amounts of evidence [20].

Various working groups have emerged that govern the performance, validation and interpretation of DNA profiling. These include those mentioned before (SWGDM and FBI) as well as the European Network of Forensic Science Institutes (ENSFI, founded 1995) and

the International Society for Forensic Genetics (ISFG, founded 1968) [21,22]. These working groups promote the sharing of laboratory procedures with the aim of creating standardised operating procedures that are used internationally, allowing for the comparison and inclusion of data from different institutes or countries.

1.2.2 Applications of Y-STR profiling

The use of Y-STR markers in DNA profiling has developed significantly since its first conception in 1986 [23]. A review (1992 – 2017) performed by Kayser *et al.* (2017), summarised and identified the various applications in which Y-STR data is used in forensic and legal setting [24]. Herein it was identified that Y-STR data can be used in predicting biogeographical ancestry of an individual, which allows for the narrowing down of which population groups to search in for a suspect [24]. Kinship analysis and tracing of family trees may also benefit from using Y-STR data in conjunction with autosomal data. However, the areas in which Y-STR data is most valuable is the investigation of sexual offence cases and paternity testing [24], which will be discussed in more detail.

1.2.2.1 Sexual offence cases

In South Africa, 49 660 sexual offence cases were reported between April 2016 and March 2017 [25], of which approximately 80% (39 828) were suspected rape cases [25]. When such cases are reported to the police, an initial investigation is performed to determine whether ejaculation of semen occurred, and following this identification of the perpetrator is investigated through the use of DNA analysis [1,26].

DNA analysis of such samples typically follows presumptive identification of semen and/or cytological confirmation of the presence of spermatozoa [27]. Following successful differential extraction of DNA from semen, the DNA can then be subjected to a DNA profiling assay. However, the performance of presumptive and cytological analyses faces numerous limitations and thus potentially, has a negative impact on resulting legal outcomes. In recent years, researchers have begun to explore a combined approach for investigation of evidence in sexual offence cases, wherein DNA profiling, other molecular analyses and cytology are used in conjunction to better improve the confidence in results obtained [1,26]. While the performance of such newer molecular analyses to identify body fluids is not validated for forensic applications yet, the value thereof is noted.

Presumptive testing for semen

In most countries, the standard evidence processing procedure for sexual offence cases begins with the performance of a presumptive test for semen; e.g. the Brentamine Fast Blue test which detects the presence of acid phosphatase, which is present in high concentrations in semen. This test relies on the observation of a colour change as an indication of the presence of possible semen [28]. The outcome of this test guides the selection of samples for DNA profiling assays.

Performance of presumptive testing initially relies on subjective analyses, including whether or not a presumptive test should be performed, and secondly whether or not the presumptive test yields a positive result. Not only has the time cut-off for a positive reaction been challenged [29], it is also widely known that the performance of presumptive tests can easily result in false positive (other substances contain acid phosphatase) and false negative (inhibitors of the colour change reaction performed) results, once again diminishing the amount of evidential value of these results [28].

Confirmatory testing of semen

In some countries, confirmatory tests are performed prior to DNA profiling to provide further confidence of the presence of sperm [28]. Cytological analysis to confirm the presence of spermatozoa is performed in cases where suspected semen stains are thought to be present, most commonly following a positive result on the presumptive test [27]. However, confirmatory tests also pose challenges as male individuals who have had a vasectomy or who do not ejaculate during sexual assault will not leave a trace of spermatozoa resulting in cytological negative results [7,30]. In such instances, criminal investigations based solely on the cytological results would be incorrectly halted.

Fortunately, alternative confirmatory tests for semen are available, including the RSID™ Semen Strip Test (Independent Forensics DNA, Lombard, IL. USA). This test is specific to human semen and relies on the detection of semenogelin, a seminal vesicle specific antigen [31]. This antigen is reportedly specific to human semen and thus reduces the risk of false positive results, and has a high sensitivity which allows detection of semen in samples that tested negative in the presumptive test [31]. While this method is aligned with exploring 'activity-level' propositions, identification of possible semen does not investigate who the sample may have belonged to (i.e 'source-level' propositions). Conventional DNA profiling may be used in this regard, so as to assist in identifying the person from whom the semen sample originated, and thus is still an essential aspect in resolving sexual offence cases.

A significant proportion of sexual assault cases, in particular rape cases, in SA are performed by multiple individuals or gangs [32–35]. Autosomal DNA profiles obtained from gang rape cases often contain complex mixtures that are not easily or cannot be resolved, and thus provides little useful information to the investigation [6]. This means that the identities of individual contributors cannot easily be distinguished or identified, which has significant legal ramifications.

Dealing with mixture samples

While various recommendations and advances have been made for the analysis of multiple origin DNA samples, it is suggested that the use of Y-STR data could be of value in such instances [36]. Firstly, due to the specificity of Y-STR profiling to male individuals, the contribution of a female victim would immediately be excluded. This would significantly improve and simplify the interpretation of mixtures, which is otherwise a complex and time consuming process. Secondly, Y-STR profiles specific to more than one individual may be distinguished should the proportion of contribution differ between individuals [37]. Major and minor profiles obtained can aid in investigations and potentially match to a suspect should a reference sample from the suspect be obtained. The ability to discern between contributors within a mixed DNA profile is simplified when using Y-STR data, as generally one allele per marker (represented by a single peak on an electropherogram) is observed due to the single copy number of the Y-chromosome itself.

Prinz *et al.* (1997) investigated the ability to differentiate between male/male and male/female mixtures in ratios of up to 1:99 and 1:2000 respectively [37]. With use of a Y-STR triplex system, the male/male contributors could be accurately identified up to 1:4 (total DNA 5 ng), which was significantly lower than that observed when using an autosomal marker in a male/male mixture (1:24, total DNA 5 ng) [37]. However, when using the same Y-triplex system to accurately distinguish a male/female ratio, male DNA could be identified in a ratio of 1:2000, which indicated great improvement of the use of an autosomal marker (1:99, 50 pg male DNA) [37]. It was suggested that the incorporation of more Y-STRs, particularly those that are highly discriminatory, would improve this type of mixture analysis, particularly in male/male mixtures. Despite the low ratio at which minor male contributors could be identified, it was stated that this type of analysis would be advantageous when differential extraction is too risky, DNA is suspected to be degraded (aged samples), and when cytology revealed a low sperm count in the sample. Furthermore, it was stated that Y-STRs allow for the determination of how many male perpetrators there were and can aid in the exclusion of wrongly accused suspects [37].

Additional advantages of Y-STR data

Y-STR profiling also carries advantages in comparison to cytological testing. Sibille *et al.* (2001), investigated the performance of Y-STR profiling of 104 cytological negative evidence swabs obtained from 79 alleged sexual assault cases [38]. In ~29% of swabs (30/104) Y-chromosome DNA was present in cases of vaginal, anal and oral penetration. However, the authors still recommended the use of differential DNA extraction and autosomal DNA profiling when cytology tests are positive [38], due to the increased random match probability observed when using autosomal markers versus Y-STR markers. It should be noted, however, that this might be due to the low number of Y-STR markers (n=3) utilised and thus the observed variation in detection success rates thereof [38].

In addition to motivating for the use of Y-STR data in the investigation of certain sexual assault cases, the authors also investigated how long after the event could DNA evidence be obtained [38]. The results indicated that Y-STR data could be obtained in one third of samples more than 48 hours post assault (<4 hours – 192 hours). This is of significant value to the scientific community and the investigation of sexual assault cases as >48 hours post assault is typically used as the cut off for examination of biological evidence [38,39]. This is usually motivated due to the enzymatic digestion of spermatozoa over time, especially in anal and oral swabs [38].

McDonald *et al.* (2015) observed that when penile penetration of the vagina has occurred, DNA profiles of 10 or more markers were obtained 21% of cases, which was lower than previous studies [38,40]. Despite these results, it was found that male DNA capable of providing profiles of 10 or more markers can be obtained up to 48 hours post event. This cut off time period was used in accordance with generally accepted standardised procedure, however it was suggested that future studies investigate this in depth and determine a true cut-off period [39]. This is of noteworthy importance as it would assist in preventing the loss of valuable evidence, and has potential to improve the outcome of legal proceedings.

A similar study conducted by Neuhuber *et al.* (2013), managed to identify perpetrators from Y-STR data (n=17 markers), in 38 of 239 sexual assault cases [40]. The findings of this study were motivation for the expansion of the Austrian National DNA database to include Y-STR data. The addition of a mere 40 Y-STR profiles to the database led to linking of multiple previously unrelated sexual assault cases and the identification of the perpetrator [38]. This study was the first of its kind to actively support and show the insurmountable value of a Y-STR database in generating new evidence for criminal investigation of these cases.

Key findings and motivation for the use of Y-STR data

Overall, the limitations of cytology-negative results and challenges of interpreting mixed DNA profiles can be partially or completely overcome through the performance of DNA profiling specific to the Y-chromosome. While Y-STR analysis has significant value in determining presence of male DNA and potentially identifying the individual, it does *not* provide evidence of ejaculation (presence of semen), as Y-STR data is not limited to semen [41]. As such these DNA based methods should be used in conjunction with other body fluid identification methods in cases of sexual assault.

Currently, the Promega PowerPlex® Y23 System is one of the most advanced Y-STR profiling assays with the incorporation of 23 Y-STR markers. This assay is only superseded by the Y-Filer Plus (Thermo Fisher Scientific, MA, USA) assay which incorporates 27 markers [42]. As part of the developmental validation of the PowerPlex® Y23 system, in accordance with the FBI and SWGDAM guidelines, Thompson *et al.* (2013) conducted a mixture study [43]. It was determined that unique minor male contributor profiles could be obtained up to ratios of 3:1 and vice versa. However, it was important to note that at ratios of 9:1 and 19:1 there was ~90% and ~60% of the unique minor contributors profile obtained respectively [43]. Furthermore, in the presence of 400ng of female DNA, ~92% and ~85% of alleles were identified or male 1 and 2 respectively, when their DNA concentration was <32 pg. This indicated a marked improvement from previous studies, and could be attributed to the incorporation of more markers, improved primer specificity and PCR efficiency, as well as the development of interpretational tools [43]. While this kit has been shown to be the most advanced in its field, its statistical power and evidentiary value lies in there being a sufficient reference database, which is not true for the South African population wherein only 114 African (Xhosa descent) profiles have published to the YHRD [44]. This will be further discussed in Section 1.2.3.

1.2.2.2 Kinship analysis

Kinship analysis refers to the process by which a relationship is either confirmed or excluded following the performance of genetic testing, usually DNA profiling [45]. During this process a reference sample is obtained from the individuals in question. Due to inheritance of autosomal genetic composition, the alleles represented in the offspring's DNA profile can be traced through the maternal or paternal line [45]. In situations where paternity needs to be confirmed, the use of Y-STR profiling may prove beneficial.

Confirmation of paternity is most commonly required in cases where disputes have arisen between the two parents of a child, and the cost of maintenance is needed (e.g. in divorce).

In such cases where a dispute of paternity has arisen, the alleged father will likely be requested to undergo paternity testing. Similarly, the need for paternity testing is sometimes used in monetary inheritance claims by offspring to the alleged father who has passed away. However, if the father has passed away, retrospective DNA samples are often unattainable and thus autosomal DNA profiling cannot be used to show paternity.

Utility of Y-STR analysis for kinship analysis

Due to the patrilineal inheritance (shared paternal genetics) and conservation of the Y-chromosome, blood-related male individuals have almost completely concordant Y-STR profiles [46]. This allows for a reference and 'proxy' male sample to be obtained from a paternal uncle, brother, grandfather (etc.), without losing significance in the findings. This does not hold true for autosomal DNA profiling, as the amount of shared DNA and thus inference of paternity is reduced the more distant the relationship [47]. For example, an uncle or grandfather would have a relatedness coefficient of $\frac{1}{4}$ to the individual in question, when using autosomal DNA profiling. However, with using low mutation rate Y-STR markers a relatedness coefficient of approximately one would be expected [46].

Chakraborty *et al.* (1985) conducted a study to review the efficiency of using Y-STR markers, in place of autosomal markers to exclude paternity [45]. It was discussed that the use of Y markers, even those that are not highly polymorphic, were more valuable than autosomal markers for the exclusion of paternity. However, when trying to show paternity, autosomal markers were more reliable [45].

The reasons for this conclusion are most likely based on the patrilineal inheritance pattern of the Y-STR. This results in related male individuals almost always having fully concordant Y-haplotypes, which in turn limits the ability to explicitly identify a single individual as the father [20]. Furthermore, analysis of Y-STR data allows for the performance of familial searching, whereby related individuals (through the paternal line) can be used to link family members with unidentified male remains. However, the incorporation of more Y-STRs may improve the discriminatory power, particularly when markers with a significant mutation rate are incorporated [20,24].

Mutation frequency of Y-STRs and value thereof in paternity testing

Heyer *et al.* (1997) estimated the mutation frequency of Y-STR tetranucleotide repeats (4 bp) to be in the order of 2×10^{-3} , following investigation of nine commonly used Y-STRs [48]. These findings were further supported by Kayser *et al.* (2001) who investigated 15 markers commonly used in forensic applications. Mutation rates ranged from 0 to 8.6×10^{-3} , with an overall average of 2.8×10^{-3} [49].

In 2009, Goedbloed *et al.* (2009), investigated the mutation rates of the seventeen Y-STR markers included in the AmpFLSTR Yfiler kit (Thermo Fisher Scientific, Waltham, MA, USA) commonly used in paternity and forensic casework [50]. Mutation rates for the seventeen markers were determined to range between 3×10^{-4} and 7.4×10^{-3} , with an overall average of 2.5×10^{-3} [50]. These findings were in accordance with previous studies [48,49]. While it is a requirement that the mutational stability of Y-STRs be known prior to clinical and forensic application, many Y-STR markers have not been investigated in terms of mutation rate. Furthermore, the mutation rates determined in the abovementioned studies fail to discriminate between close paternal relatives [51]. As a result, Ballantyne *et al.* (2010) investigated mutation rates of 186 known Y-STR markers, in 1966 father son pairs [51]. The authors investigated nucleotide repeats (tri to hexa repeat units), and identified mutation rates that spanned a range of 3.4×10^{-4} to 7.7×10^{-2} , with the overall mutability rate to be 3.4×10^{-3} [51]. In contrast to the mutation rates recorded by Heyer *et al.* (1997) and Goedbloed *et al.* (2009), which averaged at approximately 2×10^{-3} , particularly for tetranucleotide repeats, Ballantyne *et al.* (2010) found the average mutation rate of tetranucleotide repeats to be 4.3×10^{-3} . This underestimation in previous studies could be due to the small sample size, sampling error, as well as the limited number of markers investigated.

In the same study, Ballantyne *et al.* (2010), identified and investigated various factors that may contribute to variation of mutation rates observed. Firstly, the nature of the repeat motif is of importance in terms of length and complexity of the repeat sequence [51]. A significant difference in mutation rates (an increase in mutation rate) was observed when a simple repeat motif was interrupted by a single base pair. Mutation rates were also greater for repeat loci where there are multiple different repeating motifs [51]. A decrease in mutation rate was observed when repeat motif increased in length, however it was recommended that the small number of hexa and penta repeat motifs have skewed the statistical analysis [51]. While the authors concluded that the mutation rate decreases with increased length of repeat motif, the distribution of mutation rate versus motif length appears to be more bell curved. In both the Ballantyne (2010) and Kayser (2001) studies, tetranucleotides repeat motifs has greater mutation rates than di and tri nucleotide repeats [49,51]. Thus, it appears that the mutation rate peaks at 4 bp repeat motifs and decreased before and after this point.

Secondly, the number of repeats present may also affect the mutation rates, however no statistical difference was observed in the study [51]. An association was observed between mutational direction and length of repeat, wherein longer repeats were more likely to contract, whereas shorter repeats were shown to undergo expansion [51].

Ballantyne *et al.* (2010) identified thirteen Y-STRs that mutate at rates above 1×10^{-2} , explaining the name Rapid Mutating Y-STRs (RM Y-STRs) [51]. The discriminatory power of this set of thirteen RM Y-STRs was investigated in 305 related male individuals. The results found that in 66% of cases the two individual could be distinguished, with 50% of father son pair and 60% of brothers being differentiated, which was significantly greater than those observed using the generic Yfiler kit [46]. These findings were further confirmed when the same study was conducted on a larger sample size following collaboration with 51 other research centres [52]. Furthermore, it was found that the use of the 13 RM-YSTRs was able to distinctly identify 99% of unrelated males [52].

These findings have significant impact in paternity testing, but also in the other applications in which Y-STRs are utilised, including investigation of sexual assault cases (discussed above) and the identification of unclaimed deceased males or male remains [51]. However, some concern has arisen regarding the inclusion of the RM Y-STRs in paternity testing, as this may lead to false exclusion of the father should a mutation be observed [24]. As such it is advised that markers with low to medium mutations rates be incorporated in kinship analyses, and the RM Y-STRs be used in situations where a specific individual needs to be identified, such as in sexual assault cases [24].

Application of sequencing analysis of STR markers in kinship analysis

Further to incorporating more markers, particularly those with a high mutation rate, the use of a sequencing assay should be considered. Sequencing of STRs allows for the investigation of more variants as well as concordance assessment of the alleles identified through STR analysis. A targeted next generation sequencing (NGS) system has been released and validated by Illumina (San Diego, USA), that is specific to forensic casework [53]. Validation was completed under four main topics; 1) accuracy and reproducibility, 2) sensitivity, 3) stability and 4) mock case samples [53,54]. Findings of the validation studies indicated that MiSeq FGx™ Forensic Genomics system (Illumina Corp., San Diego, USA) is able to detect and identify minor component contributors with high sensitivity, with the ability to distinguish between three different genetic profiles in some instances [43]. This has huge potential in altering the way sexual assault cases are investigated, especially when mixed DNA profiles are obtained, and standard capillary electrophoresis is unable to discern the different profiles.

The ForenSeq DNA Signature Prep assay, in conjunction with the MiSeq FGx™™ Forensic Genomics System (Illumina Corp., USA), incorporates 95 identity informative SNPs (iiSNPs), 56 biogeographical ancestry-informative SNPs (aiSNPs), 22 phenotypic-informative SNPs (piSNPs), 27 Global autosomal STRs, 7 X chromosome STRs and 24 Y chromosome STRs

[55]. The incorporation of these markers allows for testing of concordance with, and validation of, the results obtained using the Promega PowerPlex® Y23 system, and other forensic DNA profiling systems (e.g. Qiagen 24 Go! Plex and the Promega PowerPlex® ESI 16) [19,56]. Furthermore, the aiSNPs data may be used to confirm the self-reported ancestry of the participants, and could lead to possible reclassification of samples into genetically accurate population groups.

In addition to paternity testing, Y-STR data may be used in other kinship analyses such as prediction of ancestry, biogeographical estimates and familial searching [24]. Ancestral and biogeographical prediction may assist in identifying population groups to focus on when searching for a suspect; this is due to the conservation of Y-STR data throughout generations. Furthermore, familial searching of the Y-STR data from the evidence with a pre-existing database may potentially identify family members who could assist in the criminal investigation. Finally, Y-STR data and familial searching has potential to identify human remains that would have previously remain unidentified and unclaimed [24,57].

1.2.3 DNA profiling in the deceased

DNA profiling in forensic science is a vital investigative tool used for human identification [1]. DNA profiling is performed on samples obtained from deceased individuals with the goal of identifying the individual. The phenomenon of individuals being admitted to a medico-legal forensic facility without a known identity, and subsequently remained unidentified, is a problem faced worldwide [58–62].

The use of DNA profiling in attempts to identify these individuals, particularly when visual identification is not possible (e.g. severe decomposition, burning or skeletonisation), often provides little assistance. This can largely be attributed to the lack of a 'match' to a pre-existing reference profile on the country's national reference database, as well as poor quality of DNA profiles due to environmental and bacterial degradation of the DNA samples [63]. In such instances, the performance of Y-STR profiling may be of assistance as it provides the opportunity for familial matching, as well as due to the majority of unidentified individuals are male [58–61].

These complication and issues (poor quality DNA samples, and no reference samples for comparison) are also faced in the event of mass disasters, be it environmental or physical in nature. A significant proportion of research into identification of human remains, has focused on application in mass disaster events, such as the severe bushfires experienced in Victoria, Australia [64]. Mass disasters may also encompass mass transport accidents, such as an

aeroplane crash, however in these situations identification is simplified due to prior knowledge of passengers and numbers of individuals [65].

Throughout these research studies, various biological samples have been identified as alternatives to standard sample types for DNA profiling. However, two unique and pressing scientific questions regarding DNA profiling of post-mortem samples remain: (i) what the effect is of time since death on DNA quality and quantity, and (ii) how long after death and/or storage at a forensic services facility can suitable DNA profiles be generated from buccal swabs? These key questions will be discussed in more detail below, with reference to existing literature.

1.2.3.1 Post-mortem changes and the effect on DNA

Following death, various events occur that result in the degradation of DNA. Firstly, bacterial colonisation of the body reduces accessibility to good quality DNA of a single source and activity thereof may enzymatically degrade DNA [66]. Secondly, cells undergo autolysis (cell death) and rupture, causing DNA to be exposed to harsh unfavourable conditions and accelerating DNA degradation processes [67]. Thirdly, exposure to the elements, such as fire or damp conditions, results in decomposition of the body which has in turn been shown to reduce DNA output and thus quality of downstream analyses [64,68,69].

Schwark *et al.* (2010) noted that full DNA profiles can still be obtained from challenging samples, albeit with the performance of PCR optimisation [70]. It was also recommended that microsatellite repeats and conventional STRs be preferentially used over minisatellite repeats (VNTRs), as well as the use of alternative DNA profiling schemes such as those focusing on mitochondrial DNA (mtDNA) or Y-STR markers [70]. Hughes-Stamm *et al.* (2011) longitudinally investigated the level of degradation that a DNA sample has and the ability of the sample to completely resolve the identity of various STRs, mini-STRs and SNPs [63]. They found that in extreme cases of degradation (DNA fragment <200 bp) and low yield of DNA, the use of mini-STRs and SNPs was more informative than conventional STRs [63]. This supports the development of the QIAmp DNA Mini Kit and the Type-it Microsatellite PCR Kit by Qiagen Corporation.

Even under harsh conditions, where the body was decomposed and DNA degradation was noted, Hartman *et al.* (2011) showed that post-mortem DNA profiles were still in concordance with ante mortem DNA profiles [64]. This suggests that while physical integrity of the DNA is lost, chemical or sequence level information is not altered, for non-ancient DNA samples. However, the question of whether allele frequencies differ between ante and

post-mortem populations has not been investigated on a large scale. Furthermore, it is unclear whether post-mortem samples will require different validation, processing and interpretation, particularly with respect to reduced quality of profiles obtained.

1.2.3.2 Alternative prospective biological samples for use in DNA profiling

Traditionally for the performance of DNA based molecular analyses, blood is preferred for both deceased and living individuals [1]. This is motivated by research and forensic laboratories consistently obtaining high yield and good quality DNA from blood. Studies have also supported the collection and use of buccal swabs, particularly with living individuals due to its non-invasive nature [66,71]. This is of particular importance in a South African context under the Criminal Law (Forensic procedures) Act 58 of 2013, which allows for the collection of buccal cells for the generation of the National Forensic DNA Database [72]. Tredoux *et al.* (2015) found that full forensic DNA profiles could be obtained from buccal cells collected via the Whatman EasiCollect™ device, from deceased individuals, with little PCR inhibition and degradation observed [71].

Hartman *et al.* (2010), noted that blood samples obtained from decomposed bodies which were brown or green in colour failed to produce detectable levels of DNA [64]. As such, it was recommended that when collecting samples and preparing them for DNA extraction, the colour and consistency of the sample be noted, in efforts to predict possible complications [64]. Shintani-Ishida *et al.* (2009) investigated the quality of STR profiles obtained from dismembered and decomposed bodies, where tissue fragments were found between one and 49 days following death [73]. In two different and unrelated cases multiple samples, both soft and hard tissue, were obtained and subjected to a full STR analysis workflow. Additionally, the performance of muscle tissue and blood vessels was reviewed through the use of nine sets of matched samples from six cases. The results showed that hard tissue samples (nail, bone) are more suitable samples for severe decomposition cases due the ability of these sample types to resist DNA degradation better than soft tissue samples (tissue). Furthermore, it was determined that DNA obtained from blood vessels was of better quality, and resulted in more successful STR profiles than that obtained from muscle tissue [73]. These findings were in support of previous research conducted by Sato *et al.* (2009), in which the use of aortic tissue yielded sufficient DNA for the performance of STR typing [74].

Long bones, such as the femur, have typically been the preferred bone for DNA analyses, however recent studies have shown that smaller bones such as the metatarsals or patella are better suited for DNA profiling [75]. This finding may be attributed to the protected and

sheltered nature of these bones from external environmental degrading factors, and thus the quality of DNA is better than that obtained from other bone types.

Hartman *et al.* (2011) showed that full DNA profiles (9/9 markers and Amelogenin) were obtained for all post-mortem bone samples investigated [64]. More recent studies have shown that petrous bone, parietal bone and temporal bone (skull) provide DNA that is sufficient for obtaining full forensic DNA profiles following STR analysis [75,76]. This is of particular interest when dealing with old, decomposed or skeletonised remains wherein previously preferred samples are unavailable.

Soft tissue, in particular muscle tissue, is only a valuable DNA source if it is still entwined with a significant proportion of blood vessels [68]. This is often not the case in fire related deaths as the exposure to high temperatures over an extended time period desiccates and destroys the sample integrity [68]. Schwark *et al.* (2011) investigated DNA profiling of other soft tissue sample, kidney and aorta, and found that 17/18 samples produced profiles of eight markers or more [77]. However, in accordance with international guidelines, a minimum of twelve markers is required for kinship analysis, which was obtained in 15/18 samples.

Owen *et al.* (2013) investigated the use of bladder swabs in fire related deaths to identify individuals. Motivation for such sample type lay with the notion that internal organs were more protected from heat damage than bones and skin [68]. Results obtained showed that bladder swabs yield DNA of similar or better quality than typical sample types (blood, bone or tissue), with DNA profiling results either confirming or showing improvement of the typical sample DNA profiles [68].

Fingernails and toenails have been examined as biological samples due to the resilience to decomposition, when compared to tissue or blood [78]. Fingernails obtained were cleaned soon after collection and stored at -20 °C until processing, even though the nails could be kept at room temperature. Full forensic DNA profiles, of single source origin, were obtained for all fingernail samples obtained from post-mortem cases (n=10), with post-mortem interval (PMI) ranging from five days to approximately seven months. DNA profile quality was comparable between the alternative sample types investigated (muscle tissue, blood or bone), as well as to results obtained for samples from living participants (n=7) [78]. Interestingly, the authors found that in cases older than three weeks, sufficient quality and quantity of blood for molecular analyses could no longer be obtained, and similar findings were determined for muscle tissue with cases older than two months [78]. While the study has clearly elucidated the potential value of nails as a DNA source for decomposed cases, the sample size was small and observations made may be subject to sampling bias.

A similar study, conducted in Victoria, Australia, completed a comparative study between the suitability of bone and nail samples for DNA profiling [79]. Full DNA profiles could be obtained from 26/30 samples. Two cases showed signs of mixed source DNA, but this was overcome through thorough wash steps. The remaining two cases yielded partial profiles, and the observation was attributed to DNA degradation [79]. Additionally, DNA obtained from toenail samples was compared to that obtained from conventional samples (bone or blood) in terms of peak height imbalance, peak height and level of degradation. While Allouche *et al.* (2008) observed no significant difference in peak height imbalance between toenail samples and conventional samples [78], this was not observed by Schlenker *et al.* (2016) [79]. Furthermore DNA extracted from toenail samples was overall shown to be more degraded than conventional reference samples (blood stains or bone), and to yield lower peak heights in DNA profiling assays, though it should be mentioned that in some instances the toenail sample performed better than conventional samples [79]. These findings supported the implemented of toenails as an alternative sample type for the identification of decomposed cases.

While unconventional, Motani *et al.* (2006) determined that the dura mater is a sample type of value when performing DNA analysis of a decedent [80]. Dura mater was collected from 50 autopsy cases, of which 30 cases had successful typing of 15 STR markers and Amelogenin. It was noted that the success of the DNA typing procedure relied on the nature of death, particularly in cases of drowning and mummification [80]. It was concluded that the dura mater would prove valuable in cases of decomposition, due to slow decomposition which may be attributed to the compositional contribution of nerves, blood vessels and fibroblasts. Furthermore, this samples proves useful in circumstances where dismemberment has occurred and the head is retrieved, and in some cases of burning [80]. However, it should be considered that in events of head trauma it is possible that the dura mater is damaged and may be exposed to environmental factors should a puncture wound be present.

Various sample types have been, and are, utilised in forensic analysis. These include samples obtained via non-invasive and invasive procedure; and hard and soft tissue. While guidelines can be established regarding type of sample to be collected in various types of cases, the ultimate collection of samples depends on the availability of samples and the nature of the case. Furthermore, efforts should be made to optimise the sample collection procedure and processing thereof, in order to obtain DNA that is useful and can aid in the identification procedure.

1.2.3.3 Time interval after death and the effect on DNA profiling

Shintani-Ishida *et al.* (2010) found that tissue fragments (nail, muscle, hair, blood vessel) obtained full forensic DNA profiles within four days post-mortem [73]. While those found at nine days produced full profiles in 73% of samples, and the majority (85%) of tissue fragments found at 23 days failed to provide a useable DNA profile [73].

Tissue samples that are fixed in formalin and embedded in paraffin (FFPE) have been shown to provide DNA that is suitable for molecular analyses, even after 40 years of storage [81]. However, the ability to obtain sufficient yield of DNA, that is not severely degraded, is difficult and generally a significant decline in DNA integrity is noted after four years of storage [82]. A review of DNA extraction from FFPE tissue identified various parameters that could be optimised to improve the quality of results obtained [83]. These included tissue type, tissue thickness, deparaffinisation method and storage conditions [83]. Bone specimens and teeth samples are able to provide DNA decades after death, but the process of DNA extraction can be challenging and time-consuming [84,85]. As with FFPE tissue, the DNA obtained from bone specimens depends on various factors including storage conditions and age of sample.

In a study which assessed the suitability of toenails as an alternative sample type for human identification in decomposed cases, the time between death and admission to the forensic facility was estimated [79]. This time period was estimated based on when last the individual was seen alive and when the body was found and subsequently admitted to the facility. This time period ranged between two days and nine months, with the majority of cases (24/30) having estimated times of less than two weeks [79]. However, no association between amount of time and success of DNA profiling was observed, as the two samples which failed to give full DNA profiles had estimated time periods of four days and nine months respectively [79].

The amount of time elapsing prior to refrigeration is hypothesised to negatively impact the quality of DNA obtained and reproducibility of downstream applications, particularly in ancient specimens [86]. This has also been mentioned by Fondevila *et al.* (2008), wherein DNA degradation was most severe in cases of long term exposure of samples to adverse environmental conditions [87]. With decomposition cases, studies have shown that success of DNA profiling decreases with advanced decomposition, particularly after the bloating stage [88,89]. Furthermore, these studies have highlighted the need for storage of the samples at either 4 °C or -20 °C [88,89].

The paucity of research actively looking at this variable, strongly highlights the need for future studies to investigate this and develop guidelines as to how long after death DNA can

still be obtained from different sample types. Furthermore, if DNA can be obtained, focus on the quality and suitability of the DNA for molecular analyses needs to be incorporated, as this carries weight and importance in the identification of human remains.

Time periods investigated differed between studies depending on sample type and thus provided little conclusion. However, all studies did note that forensic DNA profiles could be obtained from post-mortem samples even if the sample has been exposed to adverse environmental conditions for extended periods of time. Therefore, the recommendation still stands that the collection of samples and appropriate storage needs to be implemented as soon as possible, to reduce the effects of decomposition and degradation.

1.2.3.4 Summary of literature

Forensic DNA profiling has rapidly developed over the past few decades and its applications are significantly valuable in forensic science. While much research has been conducted into improving the quality of DNA profiles from living individuals, less has been conducted specifically on post-mortem samples. When handling human remains, various samples have been identified as potential sources of DNA depending of the nature of the body, extent of decomposition and availability of samples. Additionally, while a general trend of increased DNA degradation is suspected to occur with increased storage age of decedent and time of exposure to the elements, future studies should aim to incorporate time data into their design so as to improve understanding of this occurrence. Nevertheless, while DNA profiling and other genetic analyses performed on post-mortem samples can be complicated, it is still possible and thus more efforts should be made to use DNA profiling as a means to identify decedents who have not been claimed by next-of-kin.

1.2.4 Y-STR profiling performed in South Africa

Currently, Y-STR data is not routinely used in criminal investigations conducted in SA. This may be attributed to a number of reasons, but primarily lies with the lack of reference data pertaining to Y-STR haplotypes from the background population, thereby limiting or preventing the performance of necessary statistical analyses. As such, while Y-STR profiling can be performed it provides no evidentiary value as a random match probability cannot be determined, and is thus often perceived as a waste of resources.

Allele and haplotype frequencies of various Y-STR markers, and combinations thereof, have been published for some South African population groups including Caucasian, Xhosa,

Indian and Cape Muslim [90–93]. This data has been collected over decades of research and due the various developments in forensic science the markers investigated vary extensively between the difference research papers. For this reason, the data available is not fully comparable and thus the data cannot be collated into a single reference database.

Nevertheless, many of the obtained haplotypes have contributed to the development of the worldwide Y Haplotype Reference Database (YHRD), with the addition of 946 Minimal Haplotypes, 132 Haplotypes according to the PowerPlex® Y kit and YFiler kits and 114 haplotypes for the PowerPlex® Y23 Kit [94].

1.3 RATIONALE

Currently, there is limited data in SA regarding frequencies of Y-haplotypes. This inhibits the ability to use Y-STR genotyping as a forensic identification tool in sexual assault cases. Without the use of background haplotype and allele frequencies, the Y-STR forensic profiles generated have little evidentiary value as there is no statistic associated with the results.

While the value of a Y-STR database lies significantly within the context of sexual assault cases and kinship analysis, there is potential for the database to aid in identification of male individuals. In circumstances when a body has undergone decomposition or other disfiguring processes, the use of DNA profiling is necessary for identification. Under circumstances of disfiguration or severe decomposition, the likelihood of obtaining a full autosomal forensic DNA profile is reduced. In such instances, especially when a male relative is available and willing to provide a DNA sample, Y-STR profiling can be used to link the family members and assist in the identification of the deceased individual. Two such examples are discussed below.

Following devastating bushfires in Australia, disaster victim identification was inundated with a large number of post-mortem samples of victims and ante-mortem samples either of the victims or family members [64]. Through the use of DNA profiling 67 individuals were identified, twelve of which were due to direct profile matches between ante and post-mortem samples, the remainder was as a result of kinship analysis [64]. While 52% of identifications had a multidisciplinary approach (odontology and other forensic measurements), 48% cases had DNA as the primary evidentiary source for identification [64]. Y-chromosome and mitochondrial DNA profiling were not used in this event, due to the availability of assisting information, such as last known location, but results still indicate the value of DNA profiling and kinship analysis for identification purposes.

A Indian woman went missing from her village, and her father suspected a local home owner in her disappearance [95]. Upon investigation of the household, skeleton remains were found, including 19 skulls or portions thereof. Identification of the numerous victims followed a two-step procedure, wherein anthropological reconstruction (with the aid of radiology) of the skeleton was initially performed followed by DNA analysis. Due to degradation and poor sample condition not all individuals could be identified, however genetic analysis did confirm the presence of 19 sets of remains and led to the identification of eight individuals [95].

More case reports exist, but the overarching message of all such articles is that identification of human remains relies on multidisciplinary approaches, and that all available evidence should be used in support of each other. Advances in the analysis of DNA evidence has significantly improved the process of identification, and provides insight into the persons

familial relations [64,95–97]. Furthermore, when samples are severely compromised and DNA quality and yield is low, the use of kinship analyses and thus mitochondrial or Y-STR analysis is of importance [64,97].

While some Y-STR data is available that is specific to the South African population, the data reported in these research papers has not made use of the Promega PowerPlex® Y23 System [90–93]. The commercially available kit named previously, includes 23 Y-STR markers, which is more than the 9-17 markers previously investigated in SA. However, one study has published PowerPlex® Y23 data to YHRD for 114 Xhosa individuals [94]. The incorporation and use of more Y-STR markers will allow for a greater discriminatory power, thereby increasing the value of such evidence in court [98]. However, the true value and potential of a Y-STR genetic profile cannot be realised until a background frequency data is generated.

Further to the establishment of baseline frequency data, it is important to evaluate the suitability of Y-STR DNA profiling within the deceased population for a number of reasons. Firstly, it is known that following death, quality and quantity of the DNA is compromised due to the post-mortem factors. These factors include that of decomposition which is suspected to worsen over extended periods of times. Secondly, reference frequency data has not been obtained from deceased individuals in South Africa, and as such it is unknown whether separate reference databases would be required for living and deceased individuals.

As such the collection of samples from deceased individuals is valuable to not only investigate the points made above, but also to establish guidelines for the performance of DNA profiling on post-mortem samples. Such guidelines may include the ability to determine a timeline regarding when samples are no longer likely to yield sufficient DNA for downstream molecular analyses such as forensic DNA profiling.

While there is extensive literature regarding internal validation processes, there is little guidance published regarding how to carrying this out in a local and resource-constrained setting. Internal validation guidelines differ between kits and methodologies, and consequently many facilities hire external specialists to perform and document the validation. As such, a standard operating protocol for internal validation needs to be established for the UCT Biomedical Forensic Science Research Laboratory that draws from all pre-existing recommendations. The establishment of such a protocol would increase reliability and reproducibility of results obtained from the laboratories, and also allow for confidence in the collation of data obtained from multiple forensic service laboratories. Furthermore this SOP may also be valuable to other local and international laboratories which are resource-constrained and cannot afford to employ external individuals to do the validation.

The University of Cape Town Biomedical Forensic Science laboratory, in collaboration with Western Cape Government Forensic Pathology Services procured and hosts the first MiSeq FGx™ instrument on the African continent. In the field of forensic science, the MiSeq FGx™ Forensic Genomics System is the first to be fully validated in compliance with the guidelines established by SWGDAM [99]. To date no NGS or DNA profiling data generated using this instrument is available for the South African population. This data may show to be more diverse than traditional means of genetic analysis, through the ability to view sequence level DNA profiling results. Utilisation of this instrument in the current study will primarily be to assess concordance between capillary electrophoresis based assays and NGS data. Secondly, the use of NGS sequence data allows for deeper analysis of alleles and microvariants through characterisation and verification thereof. However, for the purposes of this research study the use of NGS was focussed on concordance assessment with CE based methods.

1.4 AIMS AND OBJECTIVES

1.4.1 Aim

This research study aims to establish Y-STR forensic DNA profile frequency data for the South African population using the PowerPlex® Y23 System (Promega Corp, USA), subsequent to establishing an internal validation procedure, as well as assess of a subset of results using next generation sequencing technology.

1.4.2 Objectives

This research project incorporated two research phases, namely a retrospective review and a prospective assessment. The retrospective component was undertaken to establish and contextualise the burden of unidentified persons at Salt River Mortuary, while the prospective component generated Y-STR data for the South African population. The aim of this research project was achieved through the following objectives:

Retrospective component:

1. Assess the burden of unidentified persons at Salt River Mortuary
2. Evaluate types of cases that remain unidentified at Salt River Mortuary based on demographic details (sex, age), and cause of death details
3. Determine the number of cases where scientific means of identification were undertaken

Prospective component:

1. Design and perform a feasible internal validation procedure to establish an optimised forensic DNA profiling workflow
2. Obtain DNA samples from South African male individuals representative of the country's population group demographics and generate Y-STR profiles
3. Analyse and statistically evaluate haplotype frequencies, gene diversity and linkage disequilibrium within and between population groups
4. Investigate the effect of various post-mortem variables (e.g. storage temperature, time since death declaration) on the quality of Y-STR profiles obtained
5. Select a subset of samples showing anomalies ('off-ladder' peaks, microvariants) for verification using next generation sequencing technology
6. Assess concordance between the capillary electrophoresis based fragment analysis (DNA profiling) and next generation sequencing technology

1.4.3 Ethics approval

For the performance of this research study, both the retrospective and prospective components required ethical approval. For the performance of the retrospective review additional ethics approval was obtained to the access the local office autopsy database (HREC REF: R036/2014). For the prospective research component approval for the collection of buccal swabs from participants was obtained from the University of Cape Town Human Research Ethics Committee (UCT HREC REF 342/2016, Appendix A). Additionally, approval from the Director of Forensic Pathology Services in the Western Cape, Salt River Mortuary Facility Manager and head of the UCT Division of Forensic Medicine and Toxicology was obtained for the collection of samples from unidentified decedents at Salt River Mortuary.

CHAPTER 2: RETROSPECTIVE REVIEW OF UNIDENTIFIED DECEDENTS AT SALT RIVER MORTUARY (2010 – 2017)

2.1 INTRODUCTION

When a deceased individual is found, and an unnatural cause of death is suspected, a medico-legal investigation is undertaken [67]. The main purpose of a medico-legal investigation is to determine the cause of death, and aid the court in criminal proceedings [67]. Forensic pathologists conduct these medico-legal investigations, and depending on the nature of the case, the investigation may require the utilisation of ancillary examinations such as microbiology, histology, toxicology, amongst others [67].

2.1.1 Identification of decedents in South Africa

In SA, the performance of forensic post-mortem investigations is governed by the National Health Act 61 of 2003 (Regulation regarding the rendering of forensic pathology service) and The Inquest Act 58 of 1959, wherein an autopsy is required for cases where the circumstances at death are considered to be unnatural [100–102]. In addition to determining the cause of death, deceased individuals need to be identified. The process of identification falls within the mandate of SAPS, but is assisted by Forensic Pathology Services [100]. Identification is primarily performed through visual confirmation by a next-of-kin member or legal guardian. However, in circumstances of significant physical trauma (e.g. burns, decomposition) or where next-of-kin is unknown, alternative scientific means of identification such as fingerprints, DNA testing and/or anthropometric analysis must be employed [1]. Furthermore, it is mandated that fingerprints of unidentified individuals be collected within seven days of death [100], at this time biological samples should be obtained if DNA analysis is required.

In SA, the collection of biological samples for DNA analysis and identification is governed by the Criminal Procedure (Forensic Procedures) Amendment Act 37 of 2013 [72]. Herein, buccal swabs are the preferred sample type due to the non-invasive manner and ease in which samples can be obtained from living individuals [72]. In the past, blood was typically obtained from deceased individuals, however, the use of the EasiCollect™ device has been evaluated in a forensic mortuary setting and is now being incorporated into practise [71]. In SA, should a decedent not be identified within 30 days after death, the body becomes the responsibility of the state and is thereafter considered ‘unidentified’ [102].

2.1.2 The burden of unidentified persons worldwide

The occurrence of unidentified decedents is a problem faced in SA, but also globally [58–62,103–105]. These unidentified bodies can become a burden on the forensic facility and the state for various reasons; including the expense of storage and subsequent burial or cremation [106], as well as the strain placed on forensic facilities to store large numbers of bodies over extended periods of time, particularly in developing countries (Mitten, W., 2018, oral communication, 7 June 2018).

The burden of unidentified bodies however, differs greatly between countries. This was demonstrated in a retrospective review (1994-1998) of eight European countries (Denmark, Finland, Greece, Ireland, Portugal, Luxembourg, Spain and Germany), whereby numbers of unidentified bodies varied from between one (Luxembourg) and 263 (Germany) unidentified bodies, depending on the total population size and caseload [107]. While the definition of unidentified was noted to differ between countries reviewed, a total of 3035 decedents were admitted without identification, of which 807 individuals remained unidentified across all eight countries [107]. Another European study at the Institute of Legal Medicine (Milan, Italy) observed that 454 individuals (3% of total caseload) arrived at the facility without identification and 17% (<1% of total caseload) of those remained unidentified [60].

While studies pertaining to unidentified bodies in developing countries are scarce, two studies conducted in different regions of India, both highlighted that unidentified bodies were a burden on their forensic facilities. They noted variations in their respective caseloads and percentages of unidentified bodies: 20% of cases over 2 years [58] and 4% of cases of five years [105]. These percentages are higher than those reported in the European studies mentioned above, which suggests that the burden of unidentified remains may be associated with socio-economic status, with developed countries having lower rates of unidentified bodies compared to developing countries [58,60,105,107].

Only one study in SA has been performed pertaining to the investigation of unidentified individuals at a forensic mortuary, which was conducted in the county's capital city, Pretoria [61]. This study, conducted by Evert *et al.* (2011), undertook a four year retrospective review (2005-2008) at the Pretoria Medico-Legal Laboratory. Over the four year period, a total caseload of 9417 was observed. Of these cases, 848 (9%) remained unidentified despite attempts at identification [61].

2.1.3 Burden of Unidentified persons in the Western Cape of South Africa

Information pertaining to unidentified cases in the Western Cape, and Cape Town is limited. Within the Western Cape, there are eighteen forensic pathology service centres (Figure 2.1). The forensic facility which serves the western metropole of the City of Cape Town is Salt River Mortuary. This mortuary is one of the two forensic facilities in the Western Cape which provides academic and skills training through close association with the University of Cape Town. This is similar in nature to the relationship between the Tygerberg forensic facility (Western Cape) and Stellenbosch University [108]. Salt River Mortuary has a catchment area with a population size of approximate 3.7 million [108,109].

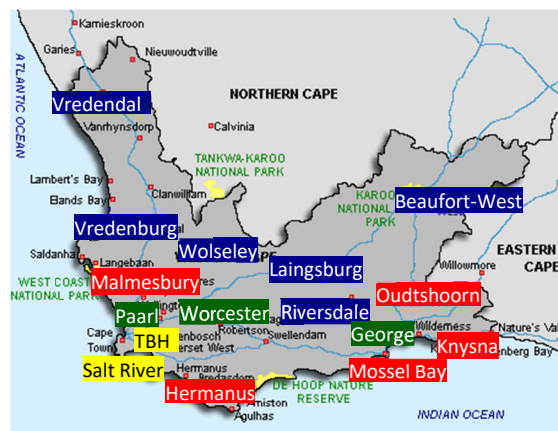


Figure 2-1: Diagram indicating all sixteen forensic pathology service facilities in the Western Cape, South Africa. Facility names written in different colours represent the caseload per annum and ranking (Blue = <250 cases, Red = 250-500 cases, Green = 500-1000 cases, Yellow = >2000 cases)

Figure adapted from [108]

Several retrospective studies have been conducted into various aspects of cases admitted to Salt River Mortuary [110–112], however, none have specifically reported on the burden of unidentified persons. One study reported that of abandoned live births, non-viable foetuses and stillbirths the majority remain unidentified, at Salt River Mortuary [111]. Reasons for this were attributed to the lack of economic security and education in these lower socio-economic areas, which may result in a significant number of unwanted pregnancies [109].

Barring this data and the study in Pretoria [61,111], there is very little published information pertaining to unidentified bodies in SA. Understanding the factors surrounding the identification process as well as the types of cases which remain unidentified may aid in the development of better processes to identify individuals in the future. This has important implications, not only for the state, but also for families who need closure and social justice. As such a retrospective review was undertaken at Salt River Mortuary with the aim of

determining of the number and demographics of cases which remained unidentified at Salt River Mortuary, as well as to describe the circumstances surrounding death and various elements of identification. This was undertaken in order to contextualise the scope of unidentified remains at Salt River Mortuary and the potential benefits of using Y-STR DNA profiling, the latter which forms the focus of this dissertation.

2.2 METHODS

For the years 2010 to 2017 selection and inclusion of cases at Salt River Mortuary was performed using records regarding 'state-pauper' burials (with no confirmed identification), kept by mortuary staff. Thereafter the local Office Autopsy Database (HREC: R036/2014) was consulted to obtain information pertaining to the demographics of the decedent, circumstances surrounding death and the death investigation, as well as the cause of death. This information on the Office Autopsy Database was originally gathered from the hard copy medico-legal case files which include a post-mortem report, death scene documentation and any other relevant documentation to the case.

The age of decedent was categorised, due to some bodies having an estimated age, as opposed to others which were determined from an identity document or medical records. Additional information regarding post-mortem interval (time since death and discovery of remains), as estimated by entomology and anthropology, was obtained through consultation with members from CapeForte and the Forensic Anthropology Cape Town (FACT) initiatives, respectively.

In this study, 'unidentified' was defined according to the "Regulations regarding the rendering of Forensic Pathology Service" [100], whereby individuals who have not been identified within seven days are classified as unidentified. At this time, scientific means of identification (e.g. fingerprinting) must be implemented [100]. Further, if the individual was not identified through visual or scientific means after 30 days, then they are legally 'unidentified' in accordance with the National Health Act 61 of 2013 [102]. Burial records available distinguished between unclaimed and unidentified decedents, through use of the terminology 'pauper' and 'state pauper' burials, respectively. The former (pauper burials) are requested from destitute family members who have identified the decedent, but cannot afford the burial ((Mitten, W., 2018, oral communication, 10 May), whereas 'state pauper' burials are those where the individual was unidentified after 30 days, despite scientific means of identification. As such, cases were included in this study if they were reported to have had a state pauper burial. Due to the cross-sectional nature of this research, there were limitations to the ability to define at what procedural stage a decedent is classified as 'unidentified', and was thus based on the most up to date documentation available.

All variable for analysis were obtained from the office autopsy database. These included demographic information (age and sex), scene of death (location of recovery, area of recovery), and circumstances of death (suspected manner and mechanism of death). Additionally, any data pertaining to efforts made to scientifically identify the individual were recorded as well.

All variables were collected into a separate excel (Microsoft office) spreadsheet for each year. Thereafter results were collated to another excel document, wherein descriptive statistics were summarised, tabulated and graphically represented. Heatmaps were generated using a free online tool (simpleheatmap.com).

2.3 RESULTS

2.3.1 Overview

The total caseload experienced at Salt River Mortuary was between 2904 and 3886 cases per annum, and was found to increase across the eight years (Figure 2.2). A total of 2476 out of 27060 cases (9.22%) met the criteria for inclusion in this study, and were determined to be unidentified from the data reviewed (Figure 2.2). While total caseload increased, the percentage of unidentified bodies fluctuated, with an overall increase from 2010 to 2017.

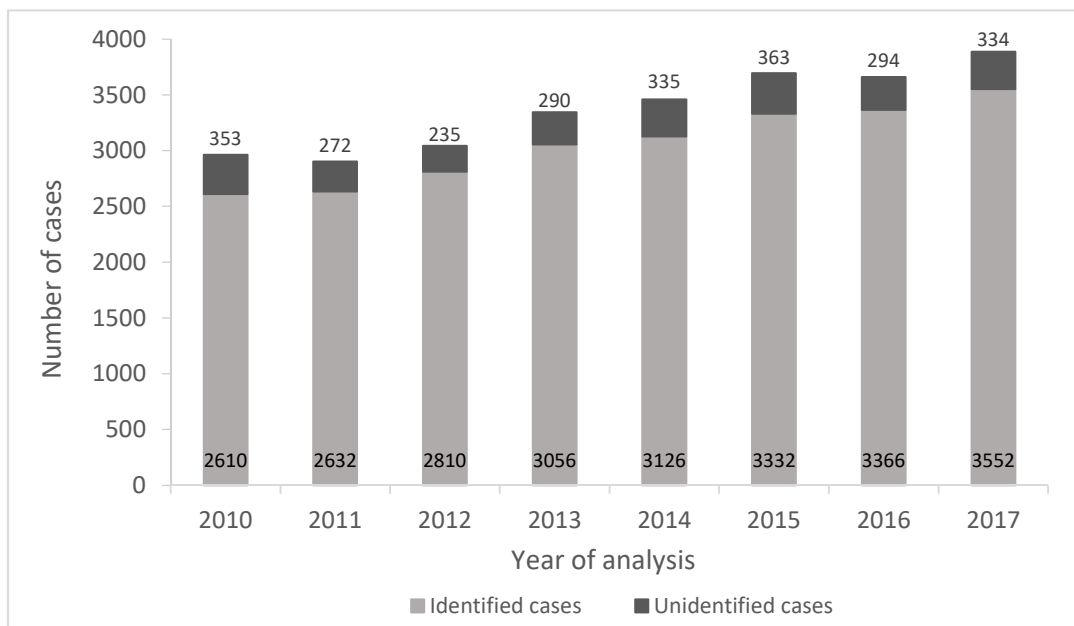


Figure 2-2: Total caseload observed at Salt River Mortuary and the number of unidentified decedents across the eight year review period (2010-2017)

Light grey boxes indicate the total number of cases admitted to Salt River Mortuary in that year, and dark grey boxes indicate the number of cases that are unidentified persons. The total caseload ranged from 2904 to 3886, wherein a general trend of increasing number of cases was observed. On average 9.2% of cases were unidentified individuals.

The number of days that elapsed between date of death declaration and date of post-mortem performance ranged between zero and 33 days, and the average was determined to be three days (mean= 3.80 days). A noteworthy finding was the median for 2017 which was determined to be 5 days (mean= 5.94 days), while every other year had a median of 3 days (mean = 3.54 days).

2.3.2 Demographics

The majority of individuals who remained unidentified were male (n=1954; 78.72%), with some cases (n=88; 3.66%) having an undetermined sex due to the gestational age of foetus or physical damage to the body.

Almost half (n=1148; 46%) of the cases were aged 20 – 40 years, with the third highest category (n= 266; 10.74%) being non-viable foetuses or stillbirths (Figure 2.3). The neonatal age category included infants who were born alive, but subsequently demised within 24 hours following birth.

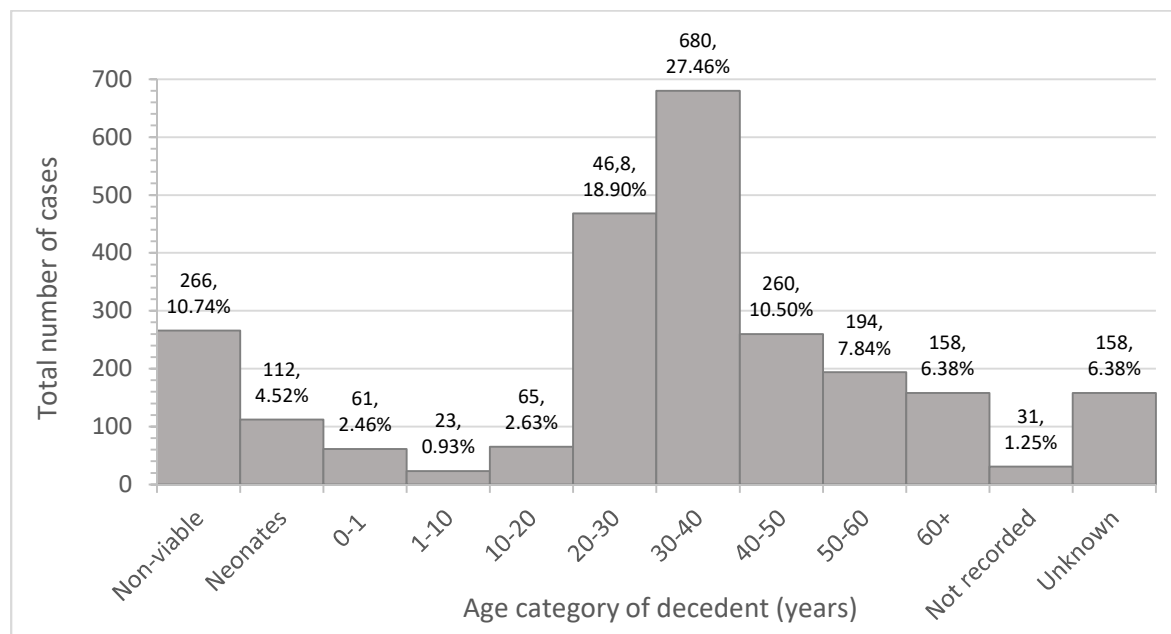


Figure 2-3: Distribution of unidentified cases at Salt River Mortuary from 2010 to 2017 according to age category of decedent

Almost half (46%) of the unidentified person cases fell between the ages of 20 and 40 years, thereafter the category of non-viable (non-viable foetus and stillborn births) was common. Neonates were defined as livebirths that demised within 24 hours since birth.

2.3.3 Location of recovery

In SA, deaths and crimes are mandated to be reported to the closest police station [113]. As such, this information was used to establish estimates as to areas where the majority of unidentified individuals were found (Figure 2.4a). In each year the highest number of cases per SAPS station was found to be Nyanga (n=426, 17.21%), followed by other lower income areas (informal settlements) of the Cape Town Metropole, including but not limited to Phillipi-East (n=165, 6.66%), Mitchell's Plain (n=154, 6.21%) and Gugulethu (n=164, 6.62%) (Figure 2.4b).

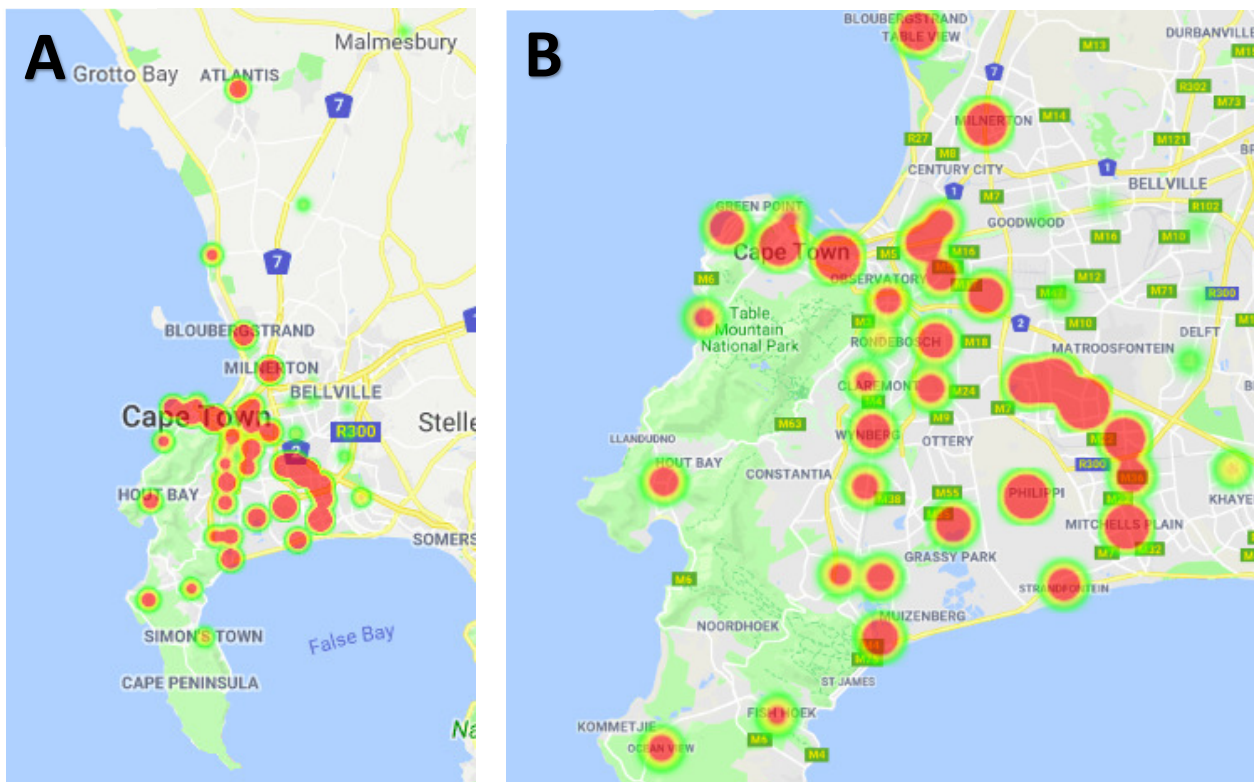


Figure 2-4: Heat maps indicating the area in which the decedent was retrieved as per the SAPS station to which the case was reported

a) Map indicating all areas (SAPS stations) from which cases were received at Salt River Mortuary, b) map of the West Metropole of the City of Cape Town indicating origin of cases as per SAPS station.

The size of the circle indicates the relative number of cases originating from that area, with a large number of cases coming from the Nyanga/Gugulethu region. (Heatmaps generated using simpleheatmap.com; accessed 10 May 2018)

Roads (n=708; 28.29%) were the most common area in which bodies were found (Figure 2.5). Subsequent to this, medical centres (n=353; 14.37%) and open land (n=352; 14.21%) areas were the other most common areas where bodies were recovered. 'Other' areas include scenes such as sewers, garbage bins, drainpipes and various additional scenes that are not encompassed by the other options provided. 'Public property' refers to any area that is open to any member of the public to visit, such as shopping malls, and restaurants/ bars.

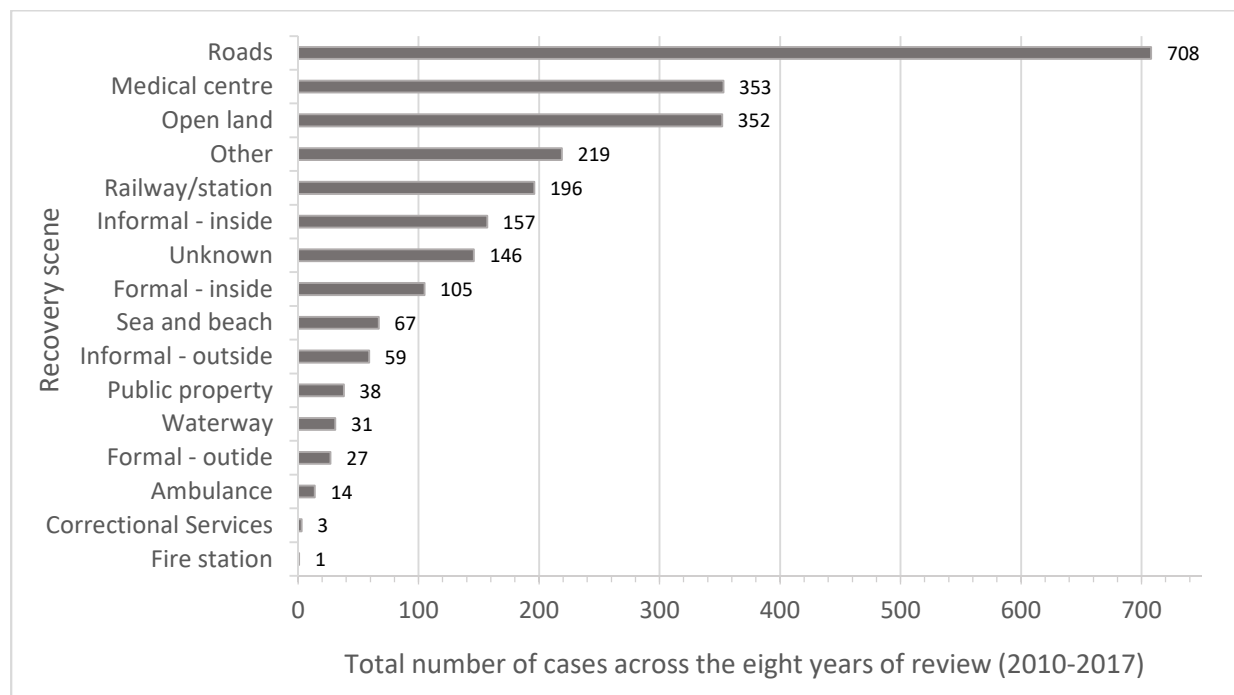


Figure 2-5: Distribution of cases as per the scene from which the decedent was recovered
 Scene of recovery was obtained according to the information reported on the FPS002 document.
 Road areas were the scene in which the majority of cases were recovered, followed by medical centres and open land areas.

2.3.4 Admission circumstances, post-mortem investigation and cause of death

Deaths investigated at Salt River Mortuary with an alleged unnatural manner of death were the most dominant category amongst unidentified individuals (Figure 2.6). Approximately 79% of road traffic crash victims, at the time of the incident, were pedestrians or cyclists. Within cases of alleged homicide, it was determined that the average age of victims was 29 years, and over 90% of decedents were male. Of these cases, the most prevalent methods of death were determined to be gunshots, assault and stabbings, with percentages of 30.58%, 29.29% and 28.74% respectively.

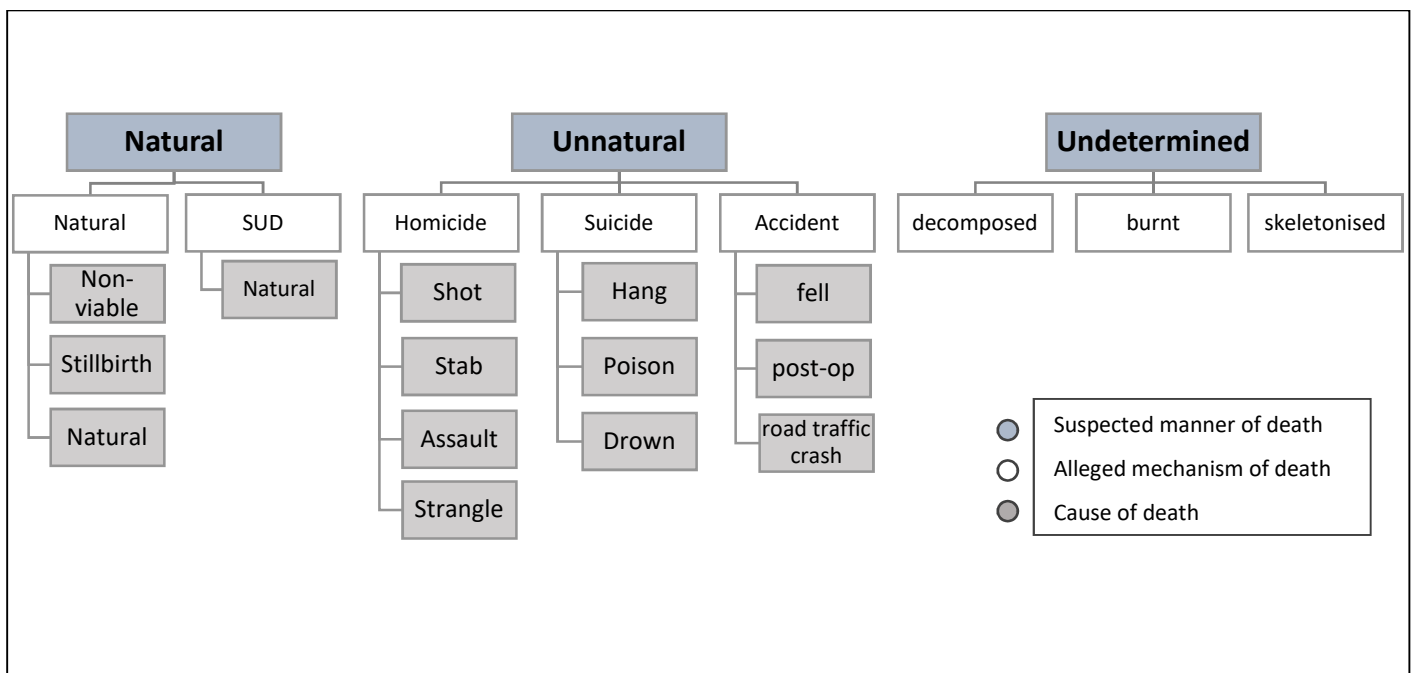


Figure 2-6: Schematic diagram of the category of death under which the cases are admitted to and resolved at Salt River Mortuary

The greatest proportion of cases was admitted to Salt River Mortuary with homicide as the alleged mechanism of death.

It was observed that 14.08% (n=345) of the cases included in this study, were physically unidentifiable due to decomposition (n=134; 5.53%), burns (n=109; 4.40%), skeletonisation (n=41; 1.64%) and other means (n=61; 2.46%) such as scavenging and dismemberment or a combination of the above. Of these cases, 44 (12.75%) were recovered from aquatic environments such as the ocean, lakes and rivers.

The determination of post-mortem interval (PMI) (i.e. time between death and discovery of the body) was requested via entomological analysis in four cases. However, in all four cases PMI estimation could only be made with sufficient accuracy and reliability in terms of season of year. Similarly, of the 36 cases in total that were referred for anthropological analysis, PMI

was not reported in any of these, as this type of analysis does not fall within their mandate. The performance of fingerprinting for purposes of human identification was not noted in this review, due to the unavailability of electronic records.

Across all eight years, 23.5% of cases (n=582) had biological samples obtained for DNA analysis by the forensic pathologist. It is possible that SAPS Victim Identification Centre (VIC) may have come and collected oral swabs for identification; however records of this were unavailable. In these cases, the most commonly obtained samples was blood (n=485; 91.8%), followed by tissue specimens (n=84; 15.9%) (Figure 2.7). In a small proportion of cases, both blood and tissue specimens (n=7; 1.2%) and/or bone (n=6, 1.03%) was obtained for identification purposes via DNA analysis.

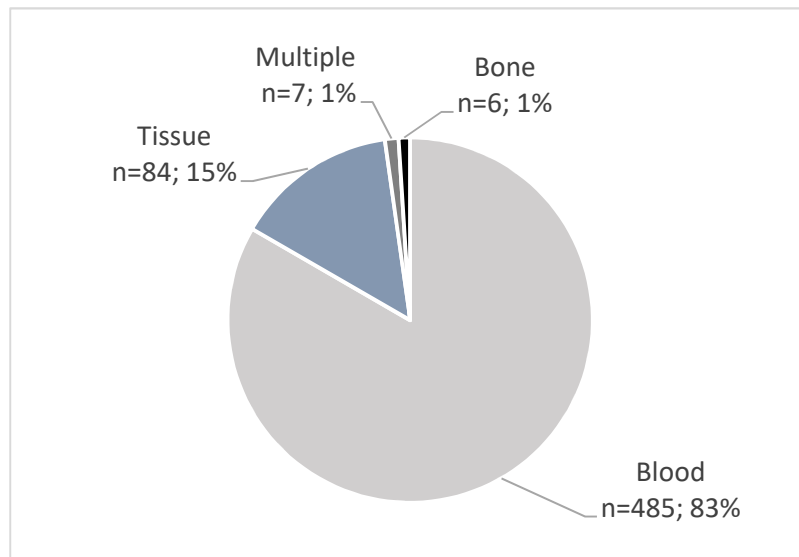


Figure 2-7: Type of biological sample obtained at autopsy by the pathologist for the performance of DNA analysis

Blood was the most commonly obtained sample type, followed up tissue specimens. In 1% of the cases, bone sections or multiple samples (blood and tissue) were obtained

2.4 DISCUSSION

The problem of unidentified persons is a significant burden not only on Salt River Mortuary itself, but also on Forensic Pathology Services (FPS). The number of bodies remaining unidentified at Salt River Mortuary and possibly other forensic facilities creates complications with regards to storage of the bodies, as many facilities (like Salt River Mortuary) are not equipped to house so many unclaimed bodies. Furthermore, the removal of these bodies, either through burial or cremation can be difficult to arrange, as is the case at Salt River Mortuary where burial is the joint responsibility of the City of Cape Town, SAPS and FPS [100]. This can burden the staff as the selection of bodies to be removed is done on a case by case basis, but also requires the co-operation and communication between multiple parties such as the City of Cape Town, and SAPS.

2.4.1 Number of unidentified per annum, and circumstances surrounding death

The percentage of cases remaining unidentified at Salt River Mortuary was determined to be 9.22% (range: 7.72% - 11.91%) per annum (Figure 2.2). This was not dissimilar to Pretoria (Gauteng, South Africa) Medico-legal mortuary which reported an average 9% (range of 7% - 10%) per annum, [61].

Of particular interest is that the total caseload experienced at the Chandigarh Mortuary (India) for the 5 year period (n=3165) was equivalent to one year's caseload at Salt River Mortuary [105]. Chandigarh (India), which has a population size of approximately 1.05 million individuals, which is less than one third of the Cape Town metropole (approximately 3.78 million) [109,114]. The catchment area and population size for the mortuary where research was conducted is unknown, as there are reportedly 52 medical institutions in the Chandigarh area [115]. Nevertheless, the percentage of unidentified bodies experienced at Salt River mortuary is more than 2-fold of that observed at the Chandigarh Department of Forensic Medicine & Toxicology, Government Medical College & Hospital Mortuary. Compared to a different district of India, the percentage of unidentified bodies observed at Salt River Mortuary was less than that observed at the Calcutta Police Morgue. The Calcutta Mortuary reported that over a two period, approximately 25% of cases were unidentified at the time of autopsy [58], with a final of 20% of cases remaining unidentified.

In comparison to developed countries, the caseload for Salt River Mortuary is more alarming. Following the completion of a 26 year (1979 – 2004) retrospective review of unidentified persons across the United States of America, it was determined that on average 413 individuals remained unidentified nationwide each year [59], whereas the number

experienced at Salt River Mortuary alone, was on average 310 individuals per annum. This is a significant finding, considering that the USA population size (approximately 326 million) is about 86-fold greater than that of the Cape Town Metropole (approximately 3.78 Million) [109,116].

It could also be hypothesised that the higher number of cases remaining unidentified in SA might be attributed to the low socio-economic status of the vast proportion of individuals living in SA (Figure 2.4). This could be due to individuals (either deceased or next-of-kin) not having a formal address that can be revisited, lack of active contact numbers, and the inability of family member to afford the cost of transport to the relevant forensic facility. According to the South African Census 2011, almost half (47%) of the population in the Cape Town City Metropole, had an annual income less than R 38 200 (minimum wage as of 2011), and 78.6% of individuals had not completed secondary school [109]. In addition to this, regulation required that identification of a decedent be performed by an authorised person in possession of an original identity document [100], however it is suspected that approximately 40% of the South African population is not registered at the Department of Home Affairs, and thus does not have an official identity document.

Male decedents accounted for 78.72% of the unidentified cases, which is slightly higher than the overall number of males admitted to Salt River Mortuary (75.1%), with the most individuals falling within the age categories of 20 – 30 years (18.73%) and 30 – 40 years (27.44%). The age distribution observed amongst unidentified decedents is similar to that of the caseload experienced at Salt River Mortuary annually with approximately 45.9% of all decedents falling between 20 and 40 years of age (Figure 2.3). These findings are in accordance with the Pretoria Mortuary findings wherein 75% of unidentified individuals were male, and approximately 40% of individuals falling between 20 and 40 years of age [61]. Similar age distributions were found for the studies conducted by Kumar *et al.* (2014) in India and Paulozzi *et al.* (2008) in the United States of America, where 39.8% and 46.6% of cases were determined to fall into the age category of 20 – 40 years, respectively [59,105].

The rate of non-vehicular (stab, shot, assault and strangulation) injury cases (31.84%), was almost twice of that observed at the Pretoria Mortuary (16%) [61]. Of particular interest is the large disparity in the firearm related homicides, where Salt River Mortuary was ~2.5 fold greater than that at Pretoria. This may be attributed to the higher rates of gang violence in Cape Town [117,118]. The general makeup of gangs (males between the ages of 20 and 40 years [117]), was also consistent with age and sex demographics of unidentified persons at Salt River Mortuary (Figure 2.3). In contrast, Paulozzi *et al.* (2008) determined that strangulation cases were most common amongst female victims [59], but only a slight female (53.8%) predominance was observed in this study for these types of cases. This

however, could likely be attributed to the small number (n=13) of strangulation cases recorded over the eight years.

Secondary to non-vehicular trauma, natural causes (18.23%) were the most common. In this study, the natural causes were made up of non-viable foetus, stillbirths, natural diseases and pre-existing medical conditions. When subdividing natural deaths into type, it was determined that stillbirth/non-viable foetuses accounted for 11.01% of the deaths. While the percentage of these cases was less than that determined by Evert *et al.* (2011) (Pretoria, South Africa), it should be noted that the number of non-viable foetus cases (Figure 2.3) recorded at Salt River Mortuary and elsewhere is thought to be under-reported, due to the manner in which the foetuses are 'disposed' of [111]; thus it is largely suspected that this statistic is unreliable both locally and internationally [119–122]. In cases where the newborn is disposed of in efforts to hide the birth, a charge of concealment of birth may be laid [123]. However, in many circumstances these charges are not formally laid or pursued due to the lack of evidence as to whom the infant was, and consequently against whom the charge should be laid.

The large number of decedents being recovered from roads (n=708) or transportation areas (n=196) (Figure 2.4), is not surprising due to the high rate of rate traffic crashes in SA. According the Road Traffic Report 2011, over 13 000 fatalities occurred as a result of road traffic crashes in 2011, which equates to approximately 38 fatalities nationwide every day [124]. The impact of these road traffic crashes has been investigated significantly over recent years, due to the detrimental effect these events places on a developing economy, like SA [125–127].

A significant proportion of decedents who remained unidentified were recovered from medical centres (Figure 2.5). This could be attributed to the individual being referred to a forensic facility following futile attempts of resuscitation. In this particular study, these deaths were most often reported to the Milnerton and Mowbray SAPS stations, as these police stations service the geographical areas in which two major hospitals are located in Cape Town (Figure 2.4).

2.4.2 Attempts to identify the decedent

When a decedent is received at a forensic facility their identity needs to be confirmed via standard or scientific means. At Salt River Mortuary, the process followed is detailed within the Western Cape Government Forensic Pathology Services Standard Operating Procedure (WCG FPS SOP) entitled "identification process". In this procedure, initial attempts at

identification require the communication with next-of-kin as to the arrangement of an appointment for visual identification of the decedent (SOP no: WC/P004 v2). However, if no next-of kin is contacted for identification within seven days of death, fingerprints needs to be obtained from the decedent by a member of SAPS VIC. Hereafter, other scientific means of identification (e.g. odontology, DNA, anthropology) should be utilised if it may aid in the identification process. This standardised approach to identification is similar to that implemented in Milan, Italy [60]. However, in SA a single individual is required to confirm identity through visual means, but in Italy two individuals are needed to confirm identity through visual identification [60].

A record of fingerprinting of unidentified remains is maintained by the SAPS VIC, and according to Chapter 3 of Criminal Procedure Act 51 of 1977, fingerprints must be obtained from decedents to assist in the identification of remains [113]. Recent updates to the 'regulations for rendering forensic pathology service', now mandate that these fingerprints be obtained within seven days of admittance to the mortuary, which is in agreement with the SAPS mandate [100]. However, this register was not available for review at the time of this study. It should also be noted, that while fingerprints are presumed to be taken from all unidentified individuals, it does not guarantee identification. Firstly, if the individual is under the age of 16, they will not have fingerprint on system at the South African Department of Home Affairs. This is further complicated when the individual is not a South African citizen, in which case the Automated Fingerprint Identification System (AFIS) is the only available comparison algorithm available for analysis [128]. Another means of identification would be to note identifying features, such as birthmarks or tattoos. This was evaluated by Kumar *et al.* (2014), wherein it was found that 65% of unidentified persons cases had marks of identification present on the body [105]. The photographing of such features is described in the WCG FPS SOP entitled "Management of Unidentified Bodies" (SOP no: WC/P013).

Scientific identification, through the use of DNA analysis, appeared to be used in 23.61% of cases, wherein biological specimens were retained (Figure 2.7). This was half of the number of cases compared to the study by Evert *et al.* [61], but substantially more than recorded (8%) by Kumar *et al.* [105]. An increase in the number of cases having specimens retained for DNA analysis was observed in 2015 and 2016. This peak could be attributed to the enactment of the South African Criminal Law (Forensic Procedures Act) 31 of 2013, wherein a national DNA database was approved for creation [72]. This database was established to not only act as an investigative tool in solving criminal cases, but also in efforts to improve the identification of unclaimed decedents (Chapter 5B, subsection 15M, of the Criminal Law Forensic Procedures Amendment Act 37 of 2013) via the 'missing persons' index, which, in

theory, would allow for the DNA profiles to be cross checked against other indices making use of both traditional and familial matching procedures.

While it is promising that specimens for DNA human identification are being retained, there is no available knowledge as to the outcome of these samples, despite numerous attempts at follow up with SAPS since February 2017. Had feedback been obtained and it was indicated the majority of individuals were not being identified, even after DNA analyses, this could suggest (1) the inability of generating a DNA profile that is useable for comparative searches (this could be related to the sample collection process or the technical laboratory process), (2) the infancy of the South African National Forensic DNA Database and inability to perform 'familial' searches and/or (3) the possible lack of procedure for family members of missing persons to provide reference DNA samples. Given the short time period (median 3 days) elapsing from when the individual was declared dead to performance of post-mortem, it is hypothesised that DNA degradation would not have been a major cause of DNA analysis failure (if this was indeed the case) [63]. This turn-around time was less than that determined for the Mortuary in Chandigarh (India) [105], and comparable to that determined in the Pretoria, South Africa study [61]. It should be noted that the increase in time in 2017 (median=5 days) could be attributed to increased caseload and consequent service pressures. However, the absence of feedback regarding DNA profiling motivates for the improved collaboration and communication between all parties involved (SAPS, FPS and the Department of Home Affairs) is essential to the improvement and optimisation of identification attempts.

In this study it was determined that physical inhibitors of visual identification were present in 14.08% (n=345) of cases. This was attributed to decomposition (5.53%), burns (4.40%) and skeletonisation (1.64%). In these cases, the value of DNA evidence becomes increasingly important for identification purposes, as visual identification and fingerprints are not always feasible. However, it is known that PMI affects the quality and quantity of DNA that can be obtained from human remains [63,79].

While it is known that the time since death declaration to autopsy was short, it is the time that the body is in the environment which is of more important in the obtaining of suitable biological samples for DNA analysis. The short amount of time between death declaration and post-mortem is commendable in the resource-constrained environment and the large caseload experienced at Salt River Mortuary; however, it is concerning that no data regarding post-mortem interval (PMI) was available for cases with physical inhibitors of visual identification (Section 2.3.4). Upon consultation with members of the FACT and CapeForte initiatives, it was revealed that PMI estimations do not fall within the FACT mandate, or is not requested in a timeous fashion. However, when entomology was

requested albeit late, especially in circumstances of decomposition, the most scientifically sound estimate as to true time of death could only be provided with regards to season of year. This was attributed to changes in insect infestation over time as well as the effects of refrigeration and removal of the body from scene prior to entomological review. The absence of this PMI data not only complicates interpretation and analysis of DNA results, but also fails to supply useful investigate tools for identification. This was in significant contrast to results obtained by Cattaneo *et al.* (2010), wherein a “biological profile” of all poor preservation cases was generated through the use of scientific methods in Milan, Italy [60]. These methods included the performance of anthropological methods (sexing, aging and ancestry determination) as well as examination of all recorded personal identifiers. Furthermore, criticism was placed on cases where scientific identification of poorly preserved bodies was replaced with means of suspected identity as a result of personal belongings [60]. The disparity in number of cases with PMI estimations between the current retrospective review and that conducted by Cattaneo *et al.* (2010), may be attributed to the developing status of SA versus the developed nature of Milan, as well as the differences in total caseload.

2.4.3 Recommendations and future research studies

The largest difficulty faced in this study, as with other retrospective studies, was the lack of detailed or compiled documentation relating to all the factors pertinent to the problem of unidentified persons at Salt River Mortuary. Second to this was the lack of compliance with the pre-existing standardised procedure of human identification. This can be attributed firstly to the lack of communication between FPS and SAPS. The under-utilisation of DNA, anthropology and entomology suggest that either these options are under-valued or not utilised appropriately to obtain useful and informative information. Furthermore, the SOPs in place need to be re-examined and updated in such a manner that all parties are aware of their and the others’ responsibilities, wherein all individual work together towards a common goal.

Databases and records systems in place should be better utilised and maintained so that the most up to date information is available for reference. This recommendation supports that suggested by Evert *et al.* (2011), and Paulozzi *et al.* (2008) [59,61]. The Fulton County Medical Examiner’s office (USA) [129], developed a national records system which constitutes the unidentified persons tab on the National Missing and Unidentified Persons System (NamUs) [130]. Similar databases or template databases exist in other countries such as Spain and Brazil [104,131]. Through these systems, a standardised record of

demographic data, specimens retained and physical identifiers exists and also allows for the tracking of unidentified cases, and subsequent removal of the decedent's profile should they be identified [129].

The South African National Forensics DNA database (NFDD) contains an index for missing persons, which should consist of DNA profiles of next-of-kin of missing individuals and unidentified persons [72]. This index is allowed to be searched against other DNA profiles in the database which hopes to find a full match or familial match. However this database, only contains DNA profile information, and does not record identifying features or results of ancillary examinations performed (e.g. odontology). Therefore, a possible area of development would be to expand the records of unidentified persons, to include information pertaining to all efforts made towards identification (e.g. presence of tattoos, DNA results, fingerprinting results). Efforts to this end have been attempted through the establishment of the Victim Identification Board, however co-operation and participation from all involved parties is limited, and might be attributed the board still being in its infancy.

The Victim Identification Board was established in 2016, to hopefully improve the sharing of information between FPS, SAPS and the Department of Home Affairs with regards to identification of human remains. This board aims to help track the number of cases of unidentified persons, but also allow for optimisation of the sampling procedures and ultimate identification of the decedents. In addition, the standardised procedure for dealing with unidentified decedents needs to be updated and compliance needs to be urged. Through this SOP the application of all necessary and available analyses can be monitored and streamlined, especially the use of scientific means of identification (DNA, anthropology etc.).

During data capturing and analysis, various research questions were identified that should be investigated further in future studies. Through this research, the need for a detailed review of the cases which employed anthropology and entomology for consultation was identified, and in cases where not employed reasons to this end. In this regard, issues surrounding the limitations of the analyses as a result of incomplete or inadequate crime scene analysis can be evaluated, and the value of the reports generated added to the investigation can be assessed. The dissemination of the information contained in these reports to parties other than the requesting individual (usually the pathologist) can be reviewed, and hopefully protocols can be improved and standardised.

Secondly, the SAPS VIC assists in the identification procedure through the collection of buccal swabs from unidentified bodies wherein decomposition and skeletonisation is noted. However, it appears that the collection and use of these samples has been unsuccessful and was suspended Mitten, W., 2018, oral communication, 7 June 2018). However, this was not

actively recorded or reviewed in this study. This further motivated the need for the larger research project that has been undertaken, in which the success rate of DNA profiling of deceased individuals is evaluated, and suggestions for improved quality and success are identified. These observations highlight the need for collaboration between FPS, SAPS and research facilities to improve the level of scientific identification being implemented. Furthermore this collaboration can help to ensure that the best and most applicable sample type be obtained timeously.

Lastly, it is known that the majority of cases at Salt River Mortuary are identified visually; it could be of interest to quantitatively investigate how soon after death, and admittance to the facility, this occurs. This could possibly highlight possible case profiles or time periods that are associated with no visual identification, and thus scientific means of identification can be implemented immediately. Additionally, in cases where visual identification did not occur, or was not possible, how successful scientific methods of identification (fingerprinting and DNA analysis) were. Again, this requires collaboration between SAPS, FPS and other parties such as researchers at UCT.

2.5 CONCLUSION

The alarmingly high percentage of cases (9.2%) that remained unidentified in this study was similar to that reported in Pretoria, South Africa in 2011, but is at least 2-fold higher than other countries of both developing and developed status. The majority of bodies did not have physical inhibitors of identification, and on average a short time interval between death declaration and post-mortem elapsed. These findings suggest that biological methods of identification (specifically DNA profiling) should be used more routinely in efforts to assist the human identification process.

Nevertheless, it was determined that the most common 'profile' of unidentified persons at Salt River Mortuary is a male between the ages of 20 and 40 years, most likely of low socio-economic status and residing in an informal settlement. The individual probably died from sharp or blunt force trauma, with the body remaining intact, arriving at the forensic facility within one day after death, and autopsy being performed on average three days after death. In such cases, it is suspected and motivated that both fingerprinting and DNA analysis could be beneficial in the identification process, provided that the individual is thought to be South African.

This retrospective review highlights that unidentified persons are a burden on the Salt River Mortuary Forensic facility. Furthermore, the large number of individuals remaining unidentified means that there are just as many families missing a loved one and lacking closure thereof. While efforts are undertaken to identify these individuals, they are often perceived as futile and time consuming. As such there is a need for research to investigate an easier, less expensive and quicker sampling procedure. Ultimately, it is important that communication between FPS, SAPS, VIC and scientific researchers be improved to ensure social justice is improved and that the burden of unidentified persons is lessened.

CHAPTER 3: METHODS AND MATERIALS

3.1 STUDY DESIGN

Following the performance of the retrospective study (Chapter 2), a prospective study was undertaken to not only generate Y-STR data for the South African population, but also to evaluate the success of DNA profiling in deceased individuals. Due to the high number of cases remaining unidentified at Salt River Mortuary each year (~9.2%, Chapter 2), samples were obtained from some of these individuals in order to determine whether DNA profiling was still a feasible tool in forensic human identification. Additionally, samples were obtained from living individuals for the purposes of frequency data generation.

3.2 COHORT AND SAMPLE COLLECTION

The Forensic Science International: Genetics Journal specifies that a minimum of 400 samples be obtained from unrelated male individuals for Y-STR population studies [132]. As such, this was the aim for recruitment in this study. Furthermore, the sample size was decided upon based on international literature which states that a minimum of 100 individuals should be recruited per population group evaluated [1,8]. This recommendation of population group size is expanded to 150 – 200 persons per population group, when admixture is displayed, as in the case in some South African population groups.

In this study, individuals were recruited from the four major population groups in SA (White, African, Coloured, and Indian/Asian) as defined by the national census [133]. The population groups used were in concordance with previous studies conducted at the University of Cape Town as well as comparable to those implemented in previous SA studies, to accommodate comparisons thereof [134]. Samples were prospectively collected from living individuals, wherein consent was provided. Additional samples were obtained from deceased individuals, to assess post-mortem factors affecting the performance of DNA profiling for forensic human identification.

The following selection criteria of participants (both living and deceased) were utilised:

1. Living individuals over the age of 18 years were included
2. Deceased individuals of all ages were recruited provided that the individual was unidentified or next-of-kin provided consent to do so
3. Only male participants were included
4. Non-humans samples, related individuals and persons who had received organ transplants were excluded

3.2.1 Sample collection from living participants

Participants were required to complete an informed consent form (Appendix B) as well as a questionnaire pertaining to demographic information (Appendix B). Each participant then swabbed their inner cheeks, five times per side, using a cotton swab (Copan Diagnostic Inc., CA, USA). The swab was packaged/sealed and placed into an ice box for transport to the Biomedical Forensic Science laboratory for processing.

Samples obtained during previous studies within the Division of Forensic Medicine and Toxicology (UCT HREC REF: 342/2016 and 158/2016) were utilised in this study as well. The sample or DNA extracted therefrom, was used in this research project, *only if* the participant had consented to the use of the sample in other research conducted at UCT. The number of people who gave consent from these studies, and who would meet the inclusion criteria for this study totalled to 300, with 259 and 41 lysate and extracted DNA samples respectively. For these samples, the availability of all demographic data required (e.g. home language) was not always available as it was not required in the original research study. Furthermore, the DNA quality assessment results differed between studies depending on what was required in the original research study (see Section 3.3).

Recruitment of participants predominantly occurred at academic and forensic-related facilities in Cape Town (Western Cape), Durban (KwaZulu-Natal) and Pretoria (Gauteng). This resulted in recruiting individuals from all nine provinces in SA.

3.2.2 Sample collection from unidentified decedents

Collection of samples was carried out using cotton swabs (Copan Diagnostic Inc., CA, USA), wherein the researcher swabbed the inner cheek of the decedent five times per side. Towards the end of the project, and in attempts to improve DNA profiling success from deceased individuals, flocked nylon swabs (Copan Diagnostics Inc., CA, USA) were also obtained from the last 18 individuals. DNA which had been extracted from blood samples was also available for 72 individuals, wherein next-of-kin had provided consent for use of samples in other research studies. These blood samples were obtained during previous research studies conducted within the Division of Forensic Medicine and Toxicology at UCT (HREC REF: 445/2017 and 110/2017).

Cases were selected through collaboration/communication with Forensic Pathology Services, Salt River Mortuary. Demographic details (e.g. age, population group) as well information pertaining to the time and circumstances surrounding death which may affect the quality of DNA, (e.g. burns, decomposition etc.) was obtained following review of the post-

mortem report. For deceased participants home language was presumed based on population group assigned by the Forensic Pathologist.

3.3 SAMPLE PROCESSING

Swabs obtained were inserted into micro-centrifuge tubes containing either 1000 µL of SwabSolution™ (Promega Corp., WI, USA) or 500 µL of STR GO! Lysis Buffer (Qiagen, Hilden, Germany) and then processed according the SwabSolution™ technical manual (Promega Corp., WI, USA) [135] or Investigator 24plex GO! QuickStart Protocol (Qiagen, Hilden, Germany) [136], respectively. Incubation of the cotton swab heads was performed using the ThermoMixer F2.0 (Eppendorf, Hamburg, USA), swabs which were placed in the STR GO! Lysis Buffer (Qiagen, Hilden, Germany) were incubated at 95 °C for 5 minutes with a rotational speed of 1000 rpm, and swabs placed in SwabSolution™ (Promega Corp., WI, USA) were incubated at 70 °C for 30 minutes with no rotation. Samples were subsequently stored at -20 °C. The two abovementioned lysis buffers, and swab processing procedures, were utilised to allow for some samples to also be included in a parallel population study being conducted within the Division of Forensic Medicine and Toxicology.

3.4 SAMPLE QUANTIFICATION

Extracted DNA samples (mouth rinse or blood) were quantified using the Quantifiler Human DNA Quantification kit (Thermo Fisher Scientific, MA, USA) using the Applied Biosystems (CA, USA) 7500 Real-Time PCR system. Some samples obtained from previous research studies had quantification information available, which had been performed using the Quantifiler Trio Kit (Thermo Fisher Scientific, MA, USA).

A standard curve of cycle threshold (Ct) versus DNA concentration was established through the performance of a serial dilution of the 2800M DNA standard, in duplicate, for each assay. Concentrations of samples were subsequently determined by using the observed Ct value for each sample. Following DNA quantification using qPCR assay, samples were diluted to 0.5ng/µL using molecular biology grade water (Lonza, Basel, Switzerland).

3.5 INTERNAL VALIDATION

Internal validation refers to the process by which the optimal assay conditions are determined for a specific instrument and chemistry/kit/assay within a specific laboratory. The purpose of internal validation in this study was to ensure that the Promega PowerPlex® Y23 system (WI, USA) performed in accordance with the developmental validation completed by the manufacturer [43]. In this study, internal validation was performed separately for the two sample types: extracted DNA and lysate DNA samples, as they each perform slightly differently and require different optimised conditions. Extracted DNA would undergo quantification and conventional PCR, whereas lysate samples require direct PCR, and thus cycling conditions and CE injection time needed to be optimised accordingly. The design and procedure for internal validation followed in this study was based on literature [137], and guided by developmental validation of the kit by the manufacturer [43].

The internal validation therefore consisted of four main aspects: 1) amplification and capillary electrophoresis parameters, 2) threshold analyses, 3) reproducibility of analysis and 4) sensitivity analysis. Each parameter is briefly discussed below, and the data obtained from these analyses is depicted in the form of Appendices. The results from internal validation guided subsequent processing and analysis of experimental samples, and thus are referred to in this methods chapter (or appendices) and not in a separate chapter.

Following the performance of internal validation, a detailed standard operating procedure (SOP) for internal validation was compiled for the laboratory, which can be made available upon request. All data analyses were performed using GeneMapper® ID-X software (Applied Biosystems, CA, USA) and Microsoft excel 2013 (Microsoft Corporation, NM, USA).

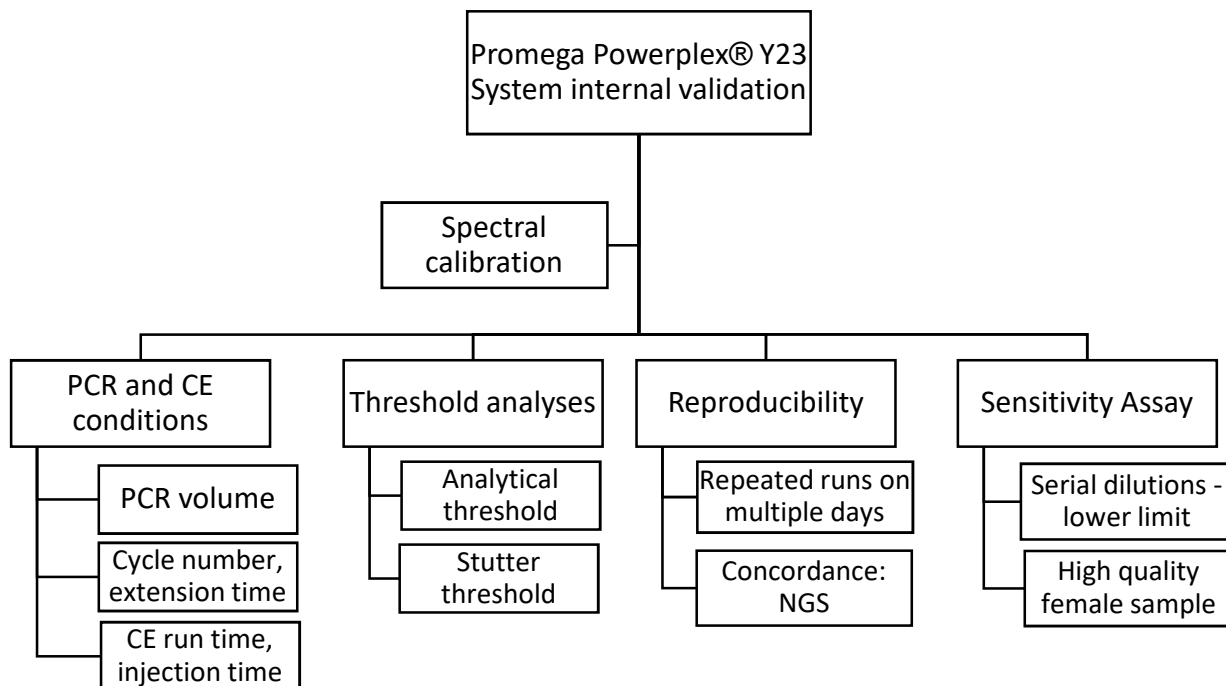


Figure 3-1: Schematic workflow summary of internal validation performed for the Promega PowerPlex®Y23 system

This flowchart indicates the various subcategories used to guide the performance of internal validation and the tests performed within each category

3.5.1 Spectral calibration

Prior to the validation and optimisation of amplification and capillary electrophoresis of samples, calibration of the electrophoretic spectral detector was performed using the PowerPlex®5C Matrix standard, according to the manufacturer’s protocol (corresponding Section 5 for the ABI 3130xl instrument, with POP7 polymer and a 36 cm array) [138]. During this process the initial conditions for injection time and voltage were optimised to 18 seconds and 1.5 kV respectively. Quality parameters of the calibration were set as stipulated in the manual, so as to ensure that ‘pass’ results were within acceptable limits [138].

3.5.2 Amplification reaction and capillary electrophoresis

A reduced amplification reaction volume of 6.25 µL was selected for the optimisation and processing of samples. This reduced amplification volume was determined according to an amended protocol obtained from Promega Corporation (WI, USA).

Amplification and capillary electrophoresis conditions for the two sample assays (extracted DNA and lysate samples) were initially analysed as per the guidelines stipulated in the PowerPlex® Y23 System Technical Manual [139]. No alterations to PCR conditions were made to either assay type (Table 3.1); however the capillary electrophoresis conditions were optimised and set to injection voltage 1.5 kV, injection time of 25 s and run time of 1800 s for both sample types. These alterations were made to improve the quality of resultant electropherograms, with peaks not being too high or too low. Increasing the run time from 1200 s to 1800 s ensures that migration of fragments through the detection system is fully completed prior to initialising the subsequent sample set.

Table 3-1: Amplification reaction conditions for both direct PCR and extracted DNA approaches

	Direct PCR		Extracted DNA approach	
	Condition	No. cycles	Condition	No. cycles
Initial denaturation	2 min @ 96 °C	1	2 min @ 96 °C	1
Denaturation	10 sec @ 94 °C	<u>27</u>	10 sec @ 94 °C	<u>30</u>
Annealing	1 min @ 61 °C		1 min @ 61 °C	
Extension	30 sec @ 72 °C		30 sec @ 72 °C	
Final extension	20 min @ 60 °C	1	20 min @ 60 °C	1
Hold	10 min @ 4 °C	1	10 min @ 4 °C	1

It was noted that Promega PowerPlex® Y23 kit was optimised for use with POP4 polymer by the manufacturer. However it was found that the POP7 polymer proved sufficient for the completion of STR analysis, provided that the appropriate software and analysis tools were used to view results.

3.5.3 Threshold analyses

3.5.3.1 Analytical threshold

The analytical threshold is the minimum height a peak must reach for the peak to be assigned an allelic identity. This accounts for any background noise appearing on the electropherogram due to the fluorescent dye itself and the medium in which samples are kept.

During this internal validation three blank samples, devoid of any biological material, (cotton swabs placed directly into lysis buffer or water) were subjected to the DNA profiling workflow which consists of DNA extraction (if applicable), PCR amplification and capillary electrophoresis. In order to determine the most conservative analytical threshold, five

methods were used: three utilised formulae and two incorporated the results of a DNA dilution series (Appendix C). From these methods it was determined that the analytical threshold for direct PCR and extracted DNA assays was 54 relative fluorescent units (RFU) and 68 RFU, respectively.

3.5.3.2 Stutter peak threshold

Stutter peaks refer to anomalous peaks that occur on an electropherogram either one repeat unit above or below that of the true allele. These stutter peaks occur as a result of PCR inefficiency. The presence of stutter peaks can be confirmed and negated if they fall below a certain percent height of the true peak [139]. Conservatively this percent height is set at 15%, but varies for different markers due to the nature of the repeat unit.

Internal validation of stutter peaks included the performance of DNA dilution series of multiple samples (n=3) performed in triplicate. During data analysis the analytical threshold was set at 25 RFU, which is half of the internationally accepted conservative analytical threshold (50 RFU) for STR analysis [1,133]. All true allele peaks were identified, and stutter peak ratios were determined by dividing the stutter peak height by the height of the true peak (equation 3.1). Sutter thresholds were determined for each marker, each channel and overall, by summing the average stutter ratio for each peak and three times the overall marker-specific standard deviation (equation 3.2). The determined stutter ratios and thresholds were in line with the developmentally validated values, and as such the manufacturer's stutter thresholds were selected for analysis of all proceeding samples.

$$\text{Stutter ratio} = \frac{\text{Height of stutter peak (RFU)}}{\text{Height of true peak (RFU)}}$$

Equation 3.1: Stutter ratio calculation

$$\text{Stutter threshold} = \text{Average marker stutter ratio} + 3(\text{Standard Deviation})$$

Equation 3.2: Marker specific stutter ratio calculation

3.5.3.3 Reproducibility

To assess reproducibility of the PowerPlex® Y23 kit, alleles identified for an individual's profile were checked for concordance between replicates within and between a capillary electrophoresis application on the same and different days. This was performed for the 2800M positive control (Promega Corp, WI, USA), as well as all samples used in the internal validation process using PCR and CE conditions established earlier in the internal validation process. This consisted of the analysis of three lysate samples (two in Promega SwabSolution™ and one in Qiagen lysis buffer), three samples where DNA extraction was performed from mouth rinse samples, and three samples where DNA was extracted from blood samples.

Following the evaluation of profile identity between repeats of the same sample processed in different assays, on different days, it was found the reproducibility of the kit was successful and in accordance with the developmental validation (Appendix D). It was noted an allelic ladder should be included per every 16 sample injections to account for small changes in the migration of fragments through the capillary during an electrophoresis sample.

Reproducibility was further evaluated in a subset of samples (n=24) with NGS, using the ForenSeq DNA Signature Prep library preparation assay on the MiSeq FGx™ Forensic Genomics System (Illumina, CA, USA) (Section 3.8).

3.5.3.4 Sensitivity analysis

Sensitivity analysis was performed in order to determine the ability of the Promega PowerPlex® Y23 System to accurately identify and assign allele identities to peaks observed in the electropherogram. To this end, a 10X and 12X dilution series was performed for direct PCR and extracted DNA samples, respectively. The extracted DNA dilution series had an initial starting concentration of 0.5 ng/μL (as determined by qPCR). The number of true allele peaks meeting the determined analytical thresholds (54 RFU and 68 RFU for direct PCR and extracted DNA assays) were noted for each concentration.

Sensitivity was further evaluated by subjecting a single origin female sample to the complete work flow. For the direct PCR assay a swab obtained from a female participant (stored in Promega SwabSolution™) was used, whereas for validation of the extracted DNA workflow a sample of concentration 100 ng/μL was used. All 23 markers were accurately assigned up to a 1 in 16 dilution (C5) for direct PCR assay. For the extracted DNA assay, all 23 markers met the analytical threshold up to an input concentration of 0.125 ng/μL (C3) (appendix E).

This was in accordance with the findings/performance of the kit when tested by the manufacturer during developmental validation [43].

3.6 STANDARDISED DNA PROFILING ASSAY OF SAMPLES

The performance of the DNA profiling workflow (PCR and CE) was performed for experimental samples according to the internally validated procedure, as summarised in Section 3.5.2. In summary, the PCR for lysate and extracted DNA samples was setup as per Table 3.2, and subjected to cycling conditions recommended by the manufacturer [139]. The CE conditions for both assays were set at an injection voltage 1.5 kV, injection time 25 s and run time of 1800 s. Extracted DNA was quantified using qPCR and diluted to 0.5 ng/μL using molecular biograde water (Lonza, Basel, Switzerland).

Table 3-2: Polymerase Chain Reaction setup for lysate and extracted DNA samples

Reagent	Lysate sample 1X (μL)	Extracted DNA 1X (μL)
5X Master Mix	1.25	1.25
10X Primer Pair Mix	0.625	0.625
5X AmpSolution™ Reagent	-	1.25
Water, amplification grade	3.375	1.125
DNA template (extracted)	1 (at a concentration of 0.5 ng/μL)	-
DNA template (lysate)	-	2
Total volume	6.25	6.25

Samples which happened to fail according to these optimised parameters were analysed and the DNA profiling workflow was repeated with slightly altered conditions, wherein the cycle number, amount of input DNA, and injection voltage were altered when appropriate.

3.7 DATA ANALYSIS

Analysis of data captured from the genetic analyser (ABI 3130xl, Thermo Fisher Scientific, MA, USA) was performed using GeneMapper v4.1 and electropherograms were subsequently visualised using GeneMapper ID-X software (Thermo Fisher Scientific, MA, USA). GeneMapper ID-X software includes an analysis panel that incorporates the analytical and stutter thresholds determined during internal validation, and contains a 'bin-set' which allows for the accurate assigning of allele numbers to observed peaks, provided that the peak falls within a 10% match window [140]. The PowerPlex®Y23_Bins_IDX_v2.0.txt file

was used for the software analysis. This bin set is automatically adjusted to align with peaks in the allele ladder based on the migration of the allelic ladder during each capillary electrophoresis session. The allelic ladder contains DNA fragments corresponding to the majority of the possible alleles for each of the 23 markers analysed in the Promega PowerPlex®Y23 System. Data were exported into Microsoft Excel 2013 spreadsheet for further analysis.

For analysis purposes a full DNA profile was defined as those with any 22 or more markers meeting the analytical threshold. A partial DNA profile was defined as any sample that had between 3 and 21 markers (any of the 23 markers) meeting the analytical threshold, and a failed profile was characterised by the presence of fewer than three markers that met the internally validated analytical threshold.

Statistical analysis of haplotype frequencies, variability between and within population groups, and other analyses were performed using Microsoft Excel 2013 (Microsoft Corporation, NM, USA), STATA v14 (StataCorp, TX, USA) and population genetics statistical package Arlequin 3.5 (release date August 2015) [141].

Correlations between numerical datasets were performed using Pearson's correlation which assesses if a linear relationship between two variables exists. Chi-Square test was employed to evaluate statistical differences between categorical data sets. Fixation index (F_{ST}) was used to evaluate genetic variability between different cohort sub-populations, and AMOVA was utilised to assess molecular diversity between and within different population groups (Table 3.3).

Table 3-3: Summary of statistical tests performed between different variables of interest

Statistical comparison between:		Statistical Test	Output result
Variable One	Variable Two		
Success rate of DNA profiling of extracted DNA samples	Success rate of DNA profiling of lysate samples	Chi-square test	p-value
Success rate of DNA profiling of samples in Promega lysis buffer	Success rate of DNA profiling of samples in Qiagen lysis buffer	Chi-square test	p-value
Colour of swab collected from decedent	Type of DNA profile obtained	Chi-square test	p-value
Time between death declaration and sample collection	Type of DNA profile obtained	Pearson's correlation	R ² value p-value
DNA profile obtained from samples stored at 4 °C	DNA profile obtained from samples stored at -20 °C	Chi-square test	p-value
Haplotype identity and composition for different population groups		F _{ST} Fisher's Exact	F _{ST} value p-value
Haplotype identity and composition for living population versus deceased population		F _{ST}	F _{ST} value p-value
Molecular variation between population groups		AMOVA	Percentage variability

Due to the patrilineal inheritance of the Y-chromosome, markers were expected to be linked. This level of shared inheritance (i.e linkage disequilibrium) was assessed using the online tool SHEsis [142]. Haplotype frequency was determined using the counting method, and gene diversity was calculated using equation 3.3. Bonferroni corrections were applied to p-values obtained for haplotype comparisons.

$$GD = 1 - \sum P_i^2$$

Equation 3.3: Gene diversity

GD is gene diversity and P_i is the frequency of each allele

3.8 NEXT GENERATION SEQUENCING (MISEQ FGX™ SYSTEM)

NGS was performed using a validated Illumina workflow, which includes the MiSeq FGx™. The performance of NGS encompasses three main components namely: 1) library preparation, 2) sequencing reaction and 3) data analysis.

The library preparation component made use of the ForenSeq DNA Signature Prep assay [143]. This assay analyses 27 autosomal STRs, seven X-chromosome STRs, 24 Y-STRs and 94 identity-informative SNPs, and for ease of reference the results output will be referred to as the “NGS profile” in later chapters. The related manufacturer’s protocol was utilised for this setup [143]. Briefly the library preparation included the amplification and tagging of the abovementioned targets. Thereafter the fragments were enriched and barcoded specific to each individual sample. Samples were purified and normalised using bead-based methods. Finally the samples were pooled, denatured and diluted prior to sequencing setup. Deviations made to the manufacturer’s protocol included increasing the volume of library loaded onto the sequencing cartridge to 10 µL. Each assay also included a positive control of known DNA composition (2800M), a negative control and a human sequencing control [143]. The human sequencing control (HSC), allows for the evaluation of the sequencing reaction alone, as this does not undergo the library prep steps. Conversely, the positive control (2800M) and the negative control, are used to determine if errors were encountered during the library prep stage

Sequencing reaction was performed according to the manufacturer’s protocol. This setup makes use of a MiSeq FGx™ v3 600 cycles cartridge, flow cell and incorporation buffer. The sequencing reaction is monitored through the MiSeq FGx™ Control Software [144]. This software captured images of the flow cell periodically for sequence analysis, instructed the release of reagents and controlled temperature of the flow cell. Real-Time Analysis software analysed the images captured to allow for the performance of base calling and quality assessment (cluster density, cluster passing filter, prephasing and phasing). A quality score was assigned to each base of each fragment in every cycle of the analysis.

Hereafter data analysis was performed using the Universal Analysis Software, which is intrinsically integrated into the system itself [143,145]. Universal Analysis Software aligns the sequences of each clonal fragment with a reference sequence for the assignment of allele and genotype identity. Additionally, quality indicators were assigned to the interpretation of these genotypes and alleles. These include, but are not limited to, assessing inter and intra locus balance, threshold parameters, presence of mixtures and the presence of stutter peaks [145]. The positive, negative and HSC controls were flagged according to the number of alleles and genotypes meeting the analytical and interpretation thresholds, as well

depending on concordance with the known genetic composition of the control. Concordance of samples was evaluated by looking at the allele identity obtained for each marker from DNA profiling, and comparing it to that obtained from NGS. This was therefore performed using the 20 Y-STR markers common between the two assays.

CHAPTER 4: RESULTS

4.1 COHORT

The research cohort evaluated in this study comprised 653 individuals in total. Biological samples were obtained predominantly from living participants (n=480, 73.62%) from all nine provinces in SA, with the majority being from the Western Cape (46.69%) (Figure 4.1). The samples collected from deceased (n=173) individuals were all from the Western Cape.

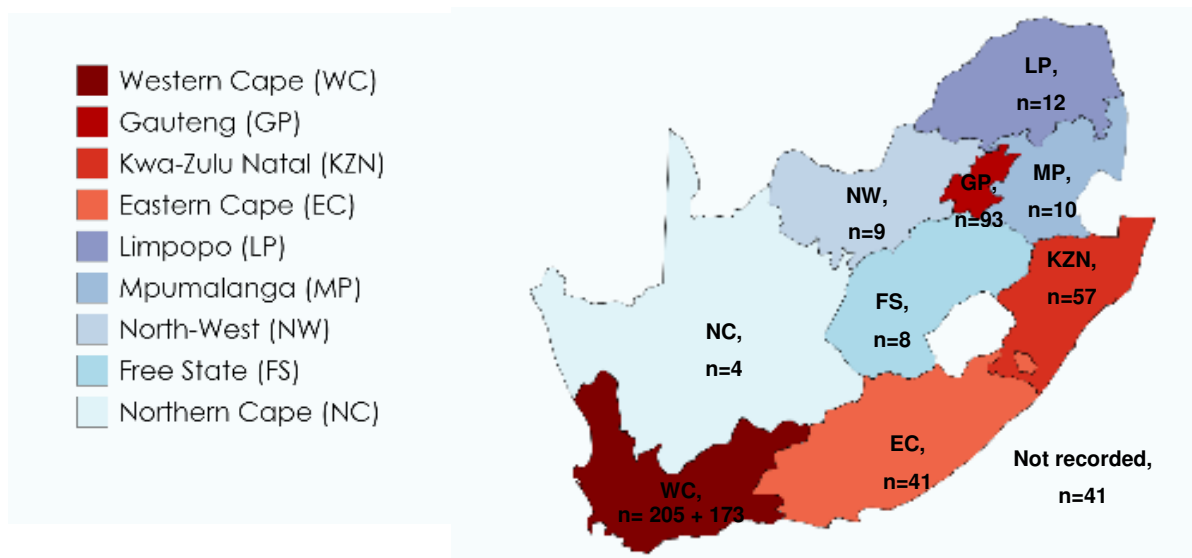


Figure 4-1: Heatmap indicating province from which participants, recruited in this study, originated from

Western Cape (WC) was the most common South African province in terms of reported home province, followed by Gauteng (GP) and Kwa-Zulu Natal (KZN). Heatmap generated using <https://mapchart.net/africa-detailed.html>

Living participants were requested to complete a questionnaire regarding their demographic information (Appendix B), including population group as per the categories utilised by the South African census. For deceased individuals, this information was obtained from medico-legal documentation or next-of-kin. Overall, the minimum of 100 samples was obtained from each of four major population groups, with three of the four population groups exceeding 150 samples each: African (n=200), Coloured (n=175), Indian/Asian (n=112) and White (n=165). One sample obtained from a deceased individual could not be assigned to a population group.

Across the cohort of living individuals, English was the most common home language, followed by African languages (IsiXhosa, IsiZulu and other) and Afrikaans (Figure 4.2). Home language was not recorded for 41 individuals, as this was not required in previous research studies conducted. Within the deceased population, home language was presumed

according to racial group (Section 3.2.2). No home language was presumed for the individual of unknown population group.

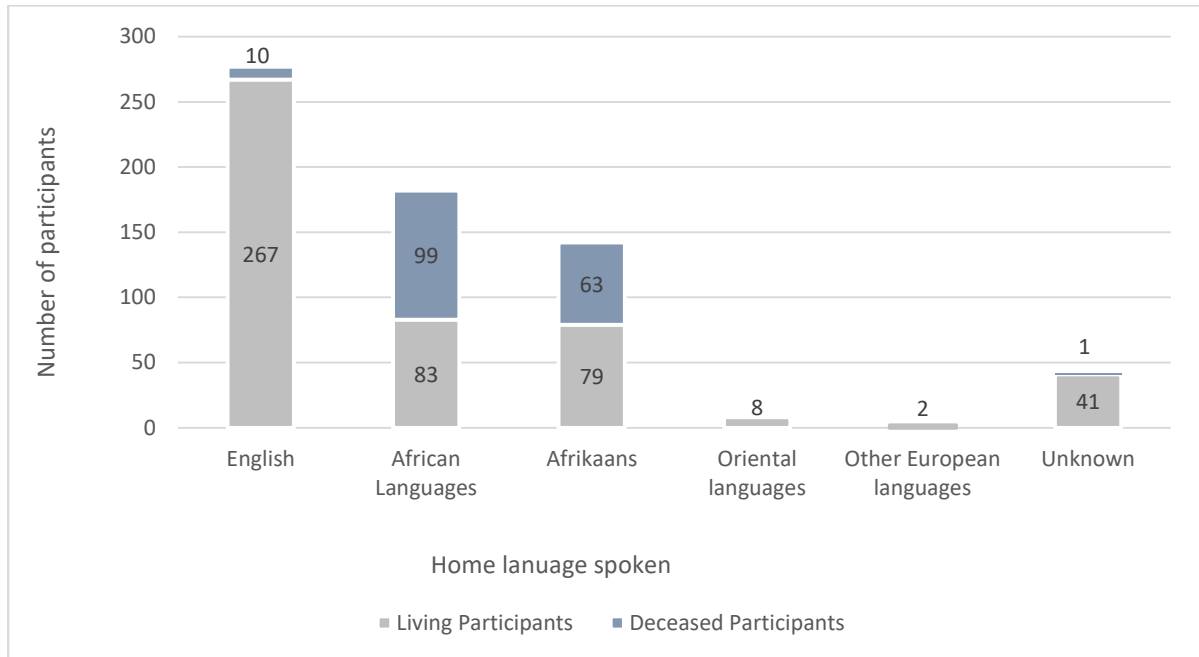


Figure 4-2: Number of participants recorded according to home language reported
Oriental languages include Mandarin, Korean and Gujarati. African languages included isiXhosa, IsiZulu, Sotho and Sepedi. Other European languages included Spanish and German

When stratifying home language according to population group, English remained the most common language in the White, Coloured and Indian/Asian populations (Appendix G). Afrikaans was the second most common language in both the White and Coloured groups.

4.2 COLLECTION AND QUANTIFICATION OF EXTRACTED DNA SAMPLES

A total of 653 samples were obtained from the four main population groups in SA. Two sample types were evaluated: extracted DNA (n=113, 17.30%) and lysate buccal swabs (n=540, 82.70%) (Table 4.1, Appendix F).

Table 4-1: Number of samples obtained as categorised by life status, self-reported population group and sample type

	Sample type	Number of samples					
		African	Coloured	Indian/ Asian	White	Undet.	TOTAL
Living participants	Lysate	84	107	108	140	0	439
	Extracted DNA	17	5	3	16	0	41
	Total	101	112	111	156	0	480
Deceased participants	Lysate	57	35	0	8	1	101
	Extracted DNA	42	28	1	1	0	72
	Total	99	63	1	9	1	173
	TOTAL	200	175	112	165	1	653

Extracted DNA was quantified using qPCR, and results obtained are summarised in Appendix I. Four mouth rinse samples were 'flagged' due to the internal positive control showing no or late amplification (Ct>30), indicating the presence of inhibitors. The blood samples showed variable DNA concentrations, but all showed little or no degradation. One sample extracted from a buccal swab showed severe degradation with an index of 31.08.

4.3 DNA PROFILING

4.3.1 Success rate of Y-STR profiling

Y-STR profiling was performed following the internally validated Promega PowerPlex® Y23 protocol discussed in Section 3.5. Overall, the first time success rate for obtaining a full DNA profile was 91.88% (600 full profiles out of 653 samples). When reviewing samples obtained from living individuals, a first time success rate of 97.08% (n=466/480) was calculated, whereas 77.46% (n=134/173) was observed for the post-mortem samples (Table 4.2). However extracted blood samples (n=61/72, 84.7%) appeared to perform significantly better than lysate samples (n=73/101, 72.3%), (p=0.010). No difference (p=0.26) in first time success rate was observed for swabs collected into Promega SwabSolution™ (n=234/242, 96.7%) or Qiagen Lysis Buffer (n=194/197, 98.5%). Quality of DNA profiles obtained mirrored that of the success rates, with samples from living individuals having the best quality, followed by post-mortem extracted DNA samples and then post-mortem lysate samples (Figure 4.3).

Table 4-2: Number and type of Y-STR profile obtained from living and deceased individuals across the two sample type groups (extracted DNA and lysate)

	Number of Y-STR profiles (n)				
	Living participants		Deceased participants		Total
	Extracted DNA	Lysate	Extracted DNA	Lysate	
Full Profile (1 st time)	38	428	61	73	600
Full Profile (2 nd time)	3	9	8	6	26
Total full profile	41	437	69	79	626
Partial	0	1	2	16	19
Failed	0	1	1	6	8

Performance of an optimised repeat assay (Section 3.6) resulted in 26 additional full Y-STR profiles. These additional full Y-STR profiles were from both living (n=12) and deceased (n=14) individuals (Table 4.2). In total, following optimisation and repeated assays, 626 full Y-STR profiles were obtained. Therefore, the overall success rate of DNA profiling was 95.87% (n=626/653).

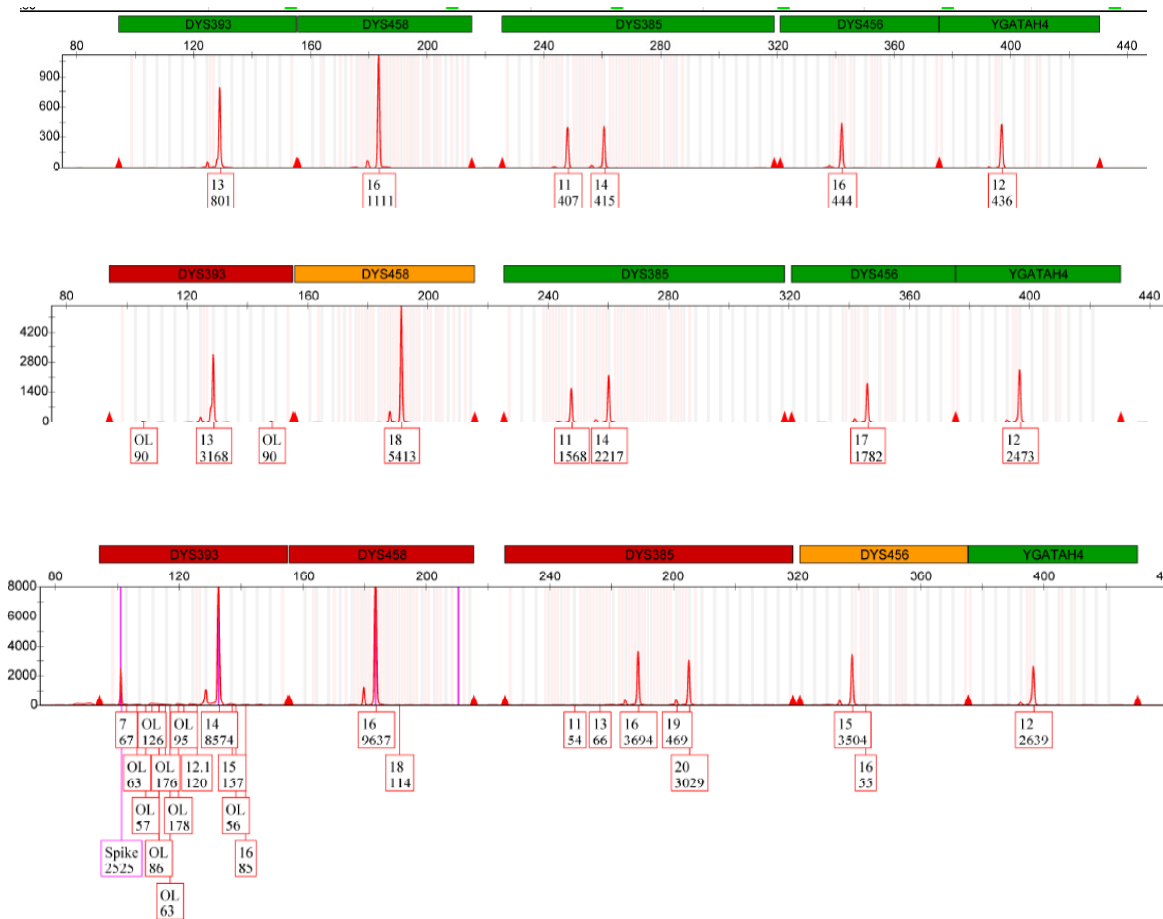


Figure 4-3: Electropherograms showing variable quality of DNA profiles obtained from a) lysate sample from living individual, b) post-mortem extracted DNA and c) post-mortem lysate sample. Samples obtained from living individuals showed little degradation, and inhibition, whereas the opposite was seen for post-mortem samples. Extracted DNA from post-mortem samples yielded electropherograms with less stutter and degradation compared to lysate samples from deceased individuals. A single peak would be expected for each marker, except for DYS385 which may have two peaks due to the duplicated nature of the marker.

4.3.2 Additional findings

Across the majority of samples, a dye-blob was observed within the blue dye channel at the beginning of the DYS576 marker (Figure 4.4). In some samples the dye-blob had a fluorescent intensity great enough to be assigned an ‘off-ladder’ (OL) allelic identity (Figure 4.4a/b).

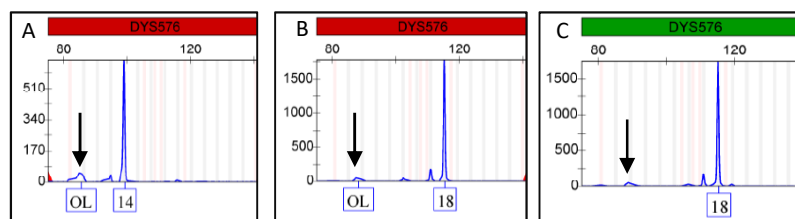


Figure 4-4: Images of the DYS576 marker from three different samples indicating the presence of the dye blob (indicated by the arrow)

In the majority of samples, at the beginning of the blue dye channel a dye blob was observed, which in some cases exceeded the determined analytical thresholds determined and was subsequently assigned an “off-ladder” allele

Five samples appeared to have a deletion at the DYS448 marker, which results in a duplication at DYS576 (Figure 4.5). This is a known deletion that has been characterised and identified by the manufacturer previously [139]. Four of these individuals belonged to the Coloured population group, and one to the White population group. These samples were included in the NGS assay for confirmation; however, NGS was unable to confirm the presence of the deletion. This was as a result of the poor sample quality and subsequent obtainment of partial or failed “NGS profiles”. The absence of sequence data at this marker could not confidently be attributed to the presence of the deletion, or low sequencing coverage by chance.

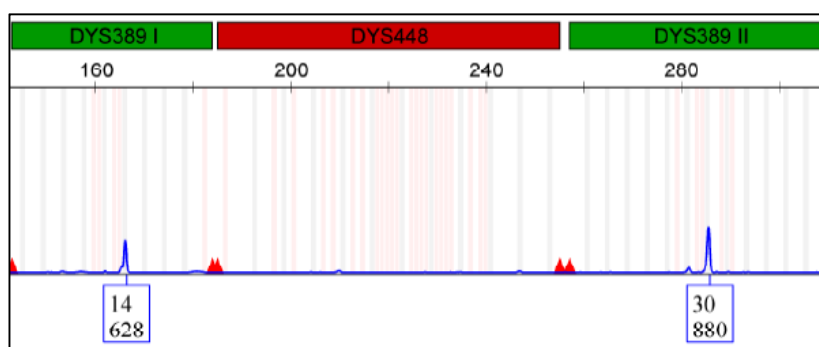


Figure 4-5: Sample showing deletion at DYS448 and consequent duplication at DYS576
Deletion at the DYS448 marker, results in a duplication of the DYS 576 marker. This is observed through the more than two-fold height difference between the DYS76 allele 16 and other alleles (DYS389I/II) in the blue dye channel

Micro-variants (n=8) were observed and automatically typed at five loci in seven different individuals. Of these, three appeared to be novel as they have not been reported in YHRD and NIST STRbase (<https://strbase.nist.gov/>). In addition, one OL allele was observed in each of the following three markers YGATA-H4, DYS391 and DYS635; and a possible null allele was observed in DYS456, DYS576 and DYS481. These micro-variants or anomalies were confirmed or resolved through the use of NGS in six samples (Table 4.3). Metrics and conditions of the NGS assay will be discussed further in Section 4.7.

One micro-variant (DYS389I, allele 13.3) and one OL allele (DYS635, allele OL=21.2) observed using the PowerPlex® Y23 System (Promega Corp, WI, USA) could not be confirmed using NGS due to the samples being depleted. Additionally, micro-variants at the DYS458 (alleles 17.1 and 20.1) and DYS456 (null allele) could not be verified or confirmed as the markers were not included within the ForenSeq DNA Signature Prep library preparation.

Table 4-3: Microvariants and novel allelic variants observed in eight different Y-STR markers, and their number of observations

¹ reported on YHRD, ² reported on NIST, ³ reported in prior literature, “Undet”= undetermined allele due to low read number, brackets indicate allelic identities determined from datapoint analysis of the electropherograms

Marker	Allele	Number of observations	Allele	NGS outcome
	Y23		NGS	
DYS570	13.2	1	18	Resolved
DYS385	17.2 ^{2,3}	1	17	Resolved
	18.2 ³	1	18	Resolved
DYS391	OL (7.1)	1	10	Resolved
DYS576	Null	1	18	Resolved
DYS448	18.4 ¹	2	18.4 Undet.	1 confirmed 1 not confirmed
DYS643	9.1	1	Undet.	Failed NGS profile
Y-GATAH4	OL (10.2) ¹	1	Undet.	Failed NGS profile
DYS481	Null	1	Undet.	Failed NGS profile

4.3.3 Haplotype Frequencies

The 626 full Y-STR profiles were utilised for the haplotype frequency, allele frequency and gene diversity analyses. Of the 200 African, 175 Coloured, 112 Indian/Asian and 165 White individuals sampled in this study, full DNA profiles were obtained from 183, 170, 111, and 162 samples respectively. The sample obtained from the deceased individual of undetermined population group, did not yield a full DNA profile.

Of these 626 full Y-STR profiles, 599 (95.69%) were unique in the sense that they were only observed once within this cohort (haplotypes available on request, and will be submitted to an online repository). Within the different population groups, the numbers of shared haplotypes differed, however only one haplotype was shared between individuals in two different population groups (White and African) (Table 4.4).

*Table 4-4: Allelic combination of haplotypes generated using the Promega PowerPlex® Y23 system, which are shared between multiple individuals
A=African, C= Coloured, IA= Indian/Asian, W=White*

Haplotype 576-389I-448-389II-19-391-481-549-533-438-437-570-635-390-439-392-643- 393-458-385a/b-456-YGATAH4	Reference	Population Group			
		A	C	IA	W
15-12-19-28-14-10-25-13-12-11-14-18-24-26-11-11-11-13-18-15-20-15-11	A	3			
15-12-19-28-14-10-26-13-12-11-14-18-24-25-11-11-11-13-18-15-20-15-11	B	2			
17-14-21-31-16-10-27-11-11-11-14-18-21-21-12-11-13-15-16-16-20-15-11	C	4			
15-14-21-30-15-10-26-12-11-11-14-19-21-21-12-11-15-13-17-15-17-15-12	D	5			1
14-12-21-30-15-10-28-11-11-11-14-20-21-21-11-11-14-13-16-17-18-15-12	E	2			
17-13-19-29-14-10-22-11-12-12-15-16-24-24-13-13-10-13-17-11-16-15-12	F				2
17-12-20-28-14-10-26-12-12-10-16-18-22-23-11-11-12-13-15-14-15-14-11	G				2
15-12-19-28-14-10-25-12-12-11-14-18-24-26-11-11-11-13-18-15-20-15-11	H		2		
19-12-18-28-15-10-24-11-11-10-14-15-20-24-11-11-13-13-16-15-15-14-11	I			2	
18-12-20-28-15-10-23-13-13-11-15-18-20-24-12-14-11-12-18-14-21-15-12	J			2	

These shared haplotypes were compared to the YHRD database in order to further assess commonness with pre-existing haplotype data. Results obtained stated the number of matches to the database as well as the random match probability (Table 4.5). Five haplotypes (B, F, G, I and J) had no match to YHRD, thereby indicating that these haplotypes have not previously been reported to the database. Additionally, through the use of minimal haplotype data, comprised of 9 Y-STR markers, YHRD can predict bio-

geographic ancestry of the haplotype. This was not possible for one haplotype, which indicates that the minimal haplotype of this full 23 marker haplotype has not been reported to the database either.

Table 4-5: Summary of shared haplotypes per population group and commonness on YHRD

Haplotype Reference	Number of matches	Ancestry prediction
A	7 matches ~1 match in 6,289 Haplotypes	Sub-Saharan African
B	No match	Sub-Saharan African Afro-Caribbean
C	4 matches ~1 match in 11006 Haplotypes	Sub-Saharan African
D	3 matches ~1 match in 14674 Haplotypes	Sub-Saharan African
E	1 match ~1 match in 44022 Haplotypes	Sub-Saharan African
F	No Match	Western European & Afro-American Admixed
G	No Match	Uralic-Yukaghir (Eastern European) & Western European Admixed
H	2 matches ~1 match in 22011 Haplotypes	Sub-Saharan African Admixed
I	No Match	No prediction available
J	No Match	Chinese (Han)

4.4 GENE DIVERSITY AND LINKAGE

Gene diversity was calculated according to equation 3.3 (Section 3.7), which makes use of allele frequencies (raw data available on request). DYS385a/b was treated as both a single marker and as two individual markers. Gene diversity scores were categorised into low ($GD < 0.5$), intermediate ($0.5 < GD < 0.9$) and high ($GD > 0.9$) diversity.

When reviewing all population groups together, the combined marker DYS385a/b had 16 alleles in 66 different combinations (Appendix H), resulting in the greatest overall gene diversity of 0.9435. Thereafter, 12 markers had gene diversities greater than 0.7, and only one fell below 0.5 (Table 4.6). Across all population groups, the combined DYS385a/b marker had the greatest gene diversity, however it fell below 0.9 (0.8547) in the White population group.

Table 4-6: Gene diversity (GD) determined for each marker and the DYS385a/b haplotype

¹ Gene Diversity calculated according to equation 3.3, in the overall population DYS385 had 16 alleles observed in 66 different haplotype combinations. Green highlights indicate $GD > 0.9$, and grey highlights represent $GD < 0.5$

Marker	Overall		African		Coloured		Indian/Asian		White	
	# alleles	GD ¹	# alleles	GD ¹	# alleles	GD ¹	# alleles	GD ¹	# alleles	GD ¹
DYS385(a/b)	16 (66)	0.944	13 (33)	0.93469	16 (50)	0.93356	13 (38)	0.93613	12 (33)	0.85467
DYS481	14	0.857	12	0.818	11	0.85765	11	0.7514	10	0.73457
DYS385b	14	0.857	12	0.820	13	0.85211	12	0.83435	10	0.70256
DYS385a	13	0.831	10	0.757	11	0.8263	7	0.79831	10	0.65813
DYS576	11	0.811	8	0.782	9	0.8264	11	0.76341	8	0.77045
DYS570	11	0.800	9	0.781	9	0.7919	10	0.84344	10	0.75324
DYS390	8	0.797	8	0.511	7	0.80997	7	0.7962	6	0.68275
DYS643	10	0.796	8	0.733	8	0.78893	8	0.65774	9	0.59511
DYS389II	9	0.775	6	0.754	8	0.77633	9	0.76731	6	0.67276
DYS635	12	0.772	10	0.650	9	0.77024	10	0.78955	8	0.62742
DYS458	12	0.768	6	0.716	9	0.77792	10	0.79815	8	0.75011
DYS448	11	0.731	7	0.638	10	0.74311	7	0.6519	6	0.59846
DYS19	6	0.705	5	0.658	5	0.70734	5	0.704	6	0.51303
DYS438	7	0.698	5	0.321	5	0.71958	6	0.70952	6	0.66331
DYS549	6	0.693	6	0.550	5	0.69772	6	0.67998	4	0.60326
DYS439	7	0.670	5	0.511	5	0.67869	7	0.72819	5	0.66796
DYS456	9	0.638	8	0.507	7	0.62388	7	0.63988	6	0.69852
DYS533	8	0.634	5	0.450	8	0.65806	4	0.61002	6	0.56569
DYS389I	6	0.630	6	0.675	4	0.6117	3	0.64865	5	0.49939
YGATAH4	8	0.624	5	0.565	5	0.61287	5	0.5933	7	0.64274
DYS393	7	0.605	6	0.643	6	0.59709	6	0.66878	5	0.39087
DYS392	10	0.561	6	0.165	7	0.62844	7	0.55012	5	0.58078
DYS437	5	0.532	4	0.145	5	0.53522	4	0.56018	4	0.598
DYS391	4	0.405	4	0.153	4	0.43958	4	0.33552	4	0.53765

The diversity of markers differed depending on the population group evaluated. A number of markers were determined to be of low gene diversity ($GD < 0.5$). DYS391 was the lowest marker in the overall population group, and only exceeded the 0.5 point in the White population group ($GD = 0.5376$). Within the African population group, four markers (DYS533, DYS438, DYS437, DYS392) had low (< 0.5) GD. Interestingly, these four markers all had intermediate (between 0.5 and 0.7) GD score within the three other population groups (White, Coloured and Indian/Asian), as well as the overall population group. Similarly, DYS393 and DYS389I had low GD in the White population, but intermediate scores within all other and the overall population groups. All markers (except for DYS391) were determined to have a GD greater than 0.5 Coloured and Indian/Asian population.

Linkage disequilibrium assessments showed that in the full cohort markers were not in complete linkage disequilibrium (scores < 100 , Figure 4.7). The level of linkage was variable between population groups with the African population group showing most linkage, and thus a lesser genetic diversity than the other population groups (Appendix K).

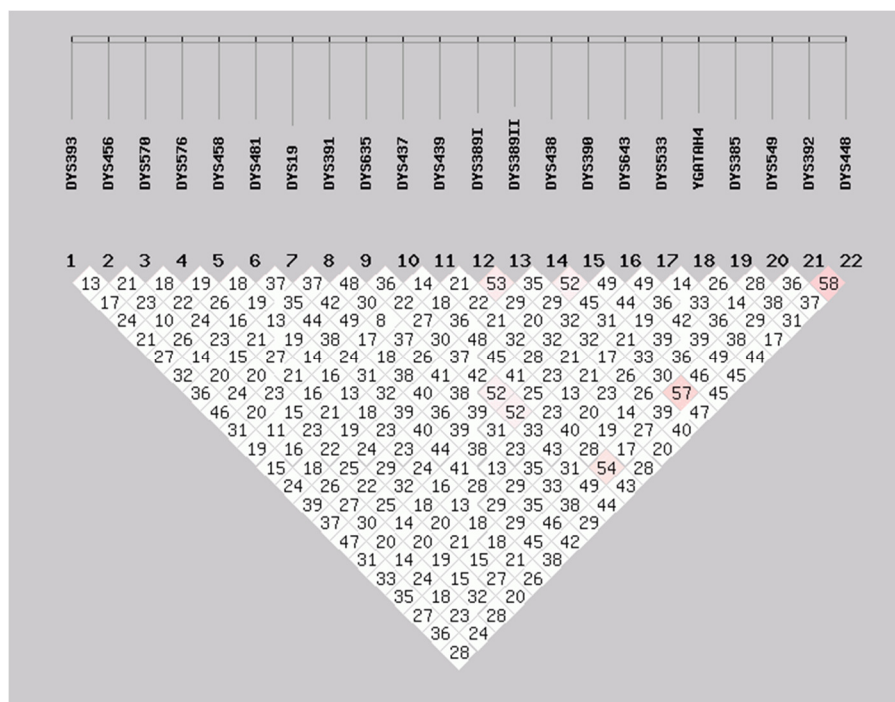


Figure 4-6: Linkage disequilibrium scores obtained for the full cohort of full DNA profiling ($n = 626$). DYS385a/b was treated as a single marker and labelled as DYS385 herein. A high score (closer to 100) indicates complete linkage.

4.5 POPULATION VARIATION

Differences in genetic composition between the different population groups were determined using F_{ST} values (Figure 4.7). Significant differences in haplotype frequencies ($p < 0.05$) was observed between all population groups, except between the Coloured and White population groups. Following application of Bonferroni corrections, significant differences ($p < 0.0083$) was only observed between the African and Coloured and Indian/Asian population groups; and the Indian/Asian and Coloured population groups (Appendix J). However, when population group variation was evaluated using the Fisher's Exact test, significant differences ($p < 0.05$) were only observed between the African and all other group (White, Coloured, Indian/Asian). Furthermore, results from AMOVA indicated that 99.96% of molecular variability existed among the population groups and only 0.04% between the population groups.

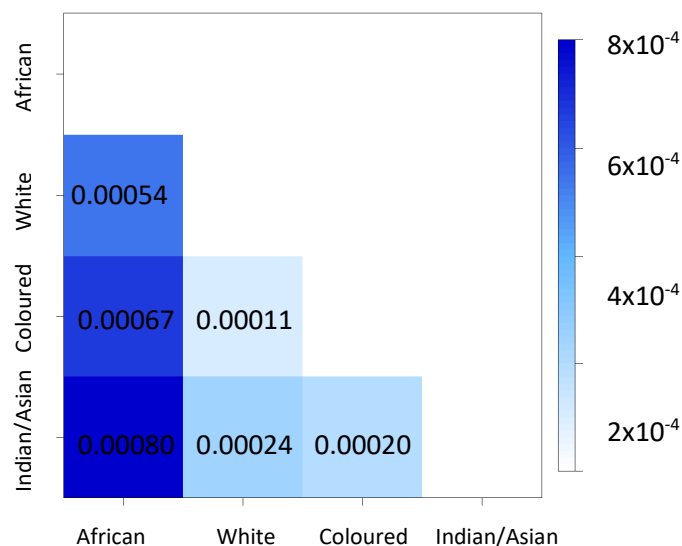
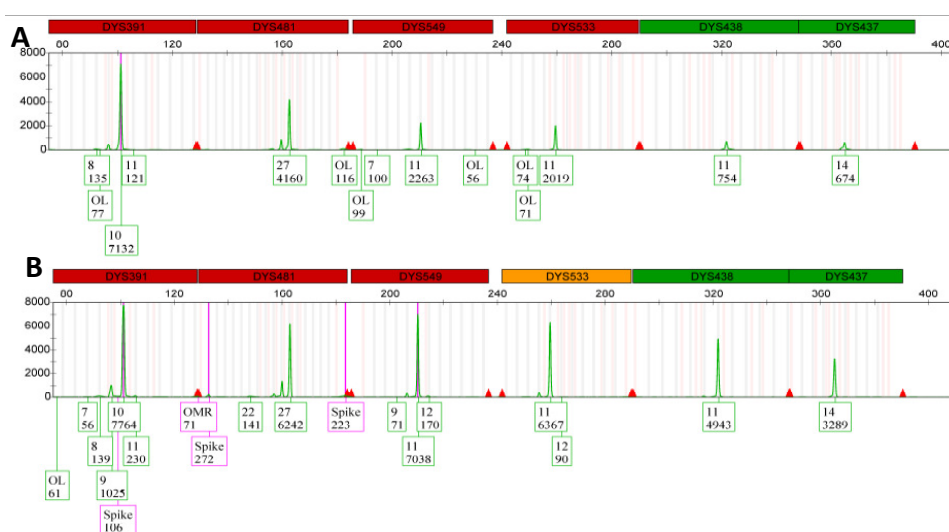


Figure 4-7: Matrix of pairwise F_{ST} values determined using Arlequin v3.1 when comparing the four population groups

Scores depicted on the figure represent the calculated F_{ST} values, wherein a higher score (darker blue) indicates greater genetic variability between groups

DNA profiles obtained from deceased individuals were sub-categorised into two population groups (African and Coloured), and the haplotype compositions thereof were compared to that of the corresponding living population group. No significant differences ($p = 0.405 \pm 0.03$ African vs African deceased, and $p = 0.991 \pm 0.003$ Coloured vs Coloured deceased) were observed between the haplotype composition of deceased and living individuals belonging to the same population group.

However, overall it was observed that the use of a flocked nylon swab resulted in electropherograms with a greater ski slope effect, more stutter peaks and lower peak height (Figure 4.9).



*Figure 4-9: Electropherogram results for one dye channel depicting the variation in peak height and ski slope effect in a (A) flocked and (B) cotton swab sample
Flocked swabs showed a greater ski slope effect with overall lower peak heights when compared to cotton swabs*

4.6 NEXT GENERATION SEQUENCING

The NGS assay performed met all quality and metric requirement parameters, in terms of cluster density (777 k/m²), phasing (0.312%), pre-phasing (0.093%) and cluster passing filter (93.25%). Additionally the HSC, library preparation positive control and the negative control all passed requirements.

Of the 24 samples subjected to the NGS workflow, ten samples yielded full assay results and eight partial assay results which could be used for the evaluation of concordance. The remaining six samples (five lysate samples and one extracted DNA sample) failed to yield any results on the NGS assay. Coverage (number of reads) of typed SNPs and STRs varied significantly between samples, and within samples. Variation depended on the quality of sample, quantification of sample, and length of target for typing.

Concordance was observed in all samples except one, in which the first marker of each dye channel in the GeneMapper ID-X electropherogram was incorrectly identified using the PowerPlex® Y23 kit and associated analysis software. NGS was able to confirm the presence of one micro-variant, and resolve allele identity of off-ladder (OL), null or micro-variant observations (Table 4.3).

Additionally, NGS results indicated that one sample may be mixed, and comprise of multiple DNA sources, as was also seen by the DNA profiling results. This DNA profile was obtained from a deceased individual who had been stored in the refrigerator for 48 days at the time of sample collection. Two other samples, obtained from deceased individual, also appeared to have mixed DNA profiles according to the DNA profiling results. These decedents had been stored for 303 and 382 days at the time of sample collection. These two potentially mixed DNA profiles were not confirmed as the one sample was not subjected to the NGS workflow, and the other failed to produce an NGS profile.

Performance of the NGS assay selected in this study allowed for the assessment of phenotypic prediction (Table 4.7) in a subset of the samples analysed (n=6/24). These samples were either obtained from deceased individuals or participants who consented to use of samples in a molecular phenotyping research project at the University of Cape Town. Biogeographical ancestry predicted through the NGS assay, matched that of the reported population group in all six samples. Samples A-E were experimental samples selected based on the purpose of showing the value of NGS phenotype prediction, and its correspondence to self-reported phenotypes. Sample F was a case example, whereby the individual was unidentified, and thus the use of NGS phenotype prediction may be a valuable tool for the identification process.

Table 4-7: Summarised NGS phenotype prediction and ancestry prediction results obtained using the ForenSeq DNA Signature Prep assay (NA = not available)

Sample ID	NGS predicted phenotype			Reported phenotype			Result
	Eye colour	Hair colour	Population group	Eye colour	Hair colour	Population group	
A	100% Brown	83% Black 17% Brown	African	NA	NA	African	Match
B	100% Brown	84% Black 16% Brown	African	NA	NA	African	Match
C	67% brown 17% blue 16% Intermediate	43% Black 39% Brown 18% Blond	Ad Mixed American	Hazel	Brown	Coloured	Match
D	100% Brown	80% Black 20% Brown	African	Brown	Black	African	Match
E	98% Brown 2% Intermediate	77% Black 21% Brown 2% Blond	Ad Mixed American	Brown	Black	Coloured	Match
F	92% blue 4% brown 4% intermediate	57% blond 38% brown 5% black	European	NA	NA	White	Case

CHAPTER 5: DISCUSSION

The use of Y-STR DNA profiling in SA, by official forensic services, is limited. This may be attributed to the early stages of the National Forensic DNA Database, but could also be due to the lack of reference data available that is specific to the local population. To date, only seven articles have been published that look specifically at Y-chromosomal data in a forensic setting in SA. However, the majority of these publications did not report on all population groups in the country, and have evaluated different forensic markers, making it challenging to collate the data in these articles [44,90–93,146,147].

5.1 COHORT COMPOSITION

One of the biggest considerations when conducting a population study is the size of the cohort [1]. As discussed in Section 3.2, the minimum size of each population group must be 100 unrelated individuals [1], and this requirement was met for all four population groups evaluated in this study (Table 4.1). However, for admixed populations such as those in SA, a minimum of 150 individuals is recommended [1]. This was achieved in three of the four population groups in this study, with the Indian/Asian population group only having 112 individuals. As encountered in previous studies, difficulties were encountered when sampling individuals of the Indian/Asian population group within the Cape Town Metropolitan Area. This is mainly attributed to the small national percentage of Indian/Asian individuals in SA, and even lower percentage within the City of Cape Town (1.4%) [109]. Efforts were made to overcome this through the collection of samples from three provinces in the country, as well as to obtain samples from individuals originating from all nine provinces in SA. Overall, the sample sizes obtained in this study are comparative to, and exceed, other studies conducted elsewhere in SA, where population size implemented ranged between 77 and 144 individuals per population group [44,92,93].

The overall population size ($n=652$) in this study however is comparable, though generally smaller, than that of international studies [46,146]. Davis. (2013) evaluated six different population groups in three native Alaskan and three American population groups, wherein sample size of each population group ranged from 141 to 175 individuals [148]. Similarly, a study conducted by Coble *et al.* (2013) analysed four US population groups with sub-population group sizes ranging from 96 to 359 [146]. Furthermore, larger variation in sample size may be attributed to the collaborative nature of Y-STR population studies in the United States of America, and Europe, wherein various research groups and collection sites were included. For example, Kayser *et al.* (1997) conducted a multicentre study to characterise

thirteen Y-STRs in terms of mutation rate and gene diversity [149]. In this multicentre study, 3825 unrelated male samples, representing 48 population groups, were utilised. These samples were obtained from a number of institutions across the world including Japan, the United Kingdom and numerous European countries [149]. The value of such collaborative studies lies in the access to greater sample numbers, and potentially greater genetic variation in samples obtained. While in the current research study, efforts were made to establish various sampling sites nationwide, future studies should aim to collaborate with other academic institutions for the sampling phase. By doing so, additional statistical power and value could be added to the results obtained in such population studies, and the genetic composition of samples analysed may be more representative of the national population. Furthermore, a limitation exists when deciding upon cohort sizes in a population study. Most genetic studies rely on the performance of a statistical power calculation, wherein the minimum population size is determined in order for confidence to exist within the results obtained. However, no such statistical calculation exists for forensic genetic population studies of this nature, and thus it is suggested that the basis of a minimum of 100 individuals per population group [1], may need to be revisited when suitable statistical calculations are developed.

SA has eleven official languages, of which nine are African languages. The two most commonly observed African languages in this study, IsiXhosa and IsiZulu, are related to the geographical location from where the majority of samples came from (Figure 4.1). IsiXhosa is the most common home language spoken by African individuals in the Western Cape and the Eastern Cape. IsiZulu is the common African home language in Gauteng and Kwa-Zulu Natal.

The admixed nature of the Coloured population is attributed to inter-racial relationships, and is often considered an intermediary 'racial' group between African and White [150]. Investigation of the genomic structure of 959 self-reported Coloured individuals revealed that the genetic origins stem from between four and seven different populations including the Khoesan, European and Asian groups [151]. Within the Coloured community, the high number of Afrikaans speaking individuals is attributed to the Cape Dutch origin of settlers of the Western Cape [152].

5.2 PERFORMANCE OF DNA PROFILING

The Promega PowerPlex® Y23 System (Promega Corp., WI, USA), is a developmentally validated chemistry and utilises the performance of a direct PCR approach through the use of the SwabSolution™ Kit [135]. Within the Division of Forensic Medicine and Toxicology, UCT, several sample types are obtained, and stored in various mediums (different lysis buffers), depending on the nature of the research being conducted. This is a similar occurrence to that observed in other research studies conducted at different institutions. As such it was important to evaluate the suitability of the Promega PowerPlex® Y23 System for all samples types. The use of two different lysate preparation kits showed no significant difference in overall success rate within the living population ($p=0.267$; Promega: 96.7%; Qiagen: 98.5%). Success rates obtained for the direct PCR approach assays of samples obtained from living individuals (reference samples) was comparable to other internal validation studies conducted in other countries which observed first time success rate of over 95% [153,154].

When using SwabSolution™ (Promega Corp., WI, USA), there was a significant difference in first time success rates between samples from the living ($n=234/242$, 96.7%) and deceased individuals ($n= 72/101$, 71.3%) ($p<0.05$). However, this is likely due to the poor quality of post-mortem samples as opposed to the buffer itself [63,71,155]. It has been reported that the success rate of forensic post-mortem samples differs based on the type of material and substrate being used, with some studies showing between 33% and 100% first time success rates [155]. As such the observed overall success rate of 85.5% ($n=148/173$) (Table 4.2) for post-mortem samples is comparable to international findings. While these results showed that lysate sample types are an alternative to conventional sampling and processing procedures (DNA extraction from blood) in forensic applications, the results still indicate that extracted DNA is the better sample type ($p=0.01$).

The use of a direct PCR approach carries potential advantages internationally, particularly in developing countries, such as SA, because of the reduced cost and hands-on time required [155]. This has been acknowledged through recent South African legislature changes which now allow for reference DNA samples to be obtained through buccal cell sampling [72]. A study at the University of Cape Town found that use of the EasiCollect™ device produced full DNA profiles from deceased individuals that were comparable to those obtained with blood samples, with a success rate of 100% [71]. This is greater than that observed for casework samples experienced at an Australian Laboratory [156]. However, this may have been due to the types of cases included in Tredoux *et al.* (2015) study (e.g. decomposition cases were excluded) as well as the small sample size used ($n=30$) [71].

5.3 DNA PROFILING IN THE DECEASED

A unique feature of this study is that it included both living (n=480) and deceased (n=173) individual. As discussed above (Section 5.2), the first time success rate of Y-STR profiling within the deceased population group was less than that observed for the living population group. While expected, there are a number of possible reasons for the difference in performance. One of the most apparent issues in post-mortem samples is DNA degradation. This has been attributed to bacterial activity, enzymatic activity and exposure to adverse environmental conditions [63,64,66–69]. While often unavoidable, DNA degradation can be managed through optimisation of input DNA concentration (qPCR assessment of degradation, see Section 5.3.1) and cycle number [70], and/or selection of suitable genetic markers [63].

In this study, PCR optimisation had to be conducted and it improved the success rate of DNA profiling in the overall cohort by approximately 4% (n=26/653; 8% in deceased population n=14/173; 2.5% in living population n=12/480, Table 4.2). The remaining DNA profiles were either partial in nature or failed completely. This suggests that the remaining partial DNA profiles may have been improved with further optimisation. However, it is possible that if partial profiles were due to severe fragmentation of DNA, and not low peak height, it is unlikely that extensive optimisation would have yielded full DNA profiles. The observation of failed DNA profiles may be due to poor DNA quality or sampling complications, such as inability to access the oral cavity due to the presence of intubation tubes, or damage of the oral cavity through exposure to water or fire.

While a direct PCR approach was similar in success to extracted DNA analysis in the living population, the same did not hold for the deceased population (Section 4.3.1 and 5.2). The appeal of a direct PCR approach lies in the ability to eliminate two costly steps, namely DNA extraction and quantification. However, the DNA quantification and quality assessment steps are of importance when processing post-mortem samples because the occurrence of DNA degradation and PCR inhibition is more likely (discussed in Section 5.3.1).

5.3.1 DNA degradation and PCR inhibition

Degradation of DNA is characterised by the presence of allele drop out at larger marker lengths (i.e. ski slope effect). DNA degradation is hypothesised to increase with a longer time period between death and sample collection [63,79], however no significant associations between these two variables have been identified due to the complex nature

and interplay with environmental influences. Studies have suggested that perhaps more important, is the duration of time between death and refrigeration of the body [63,87].

Furthermore, the interplay between extended time periods before refrigeration and adverse environmental conditions (extreme heat, dry surroundings etc.), is thought to further compound the effects of decomposition and DNA degradation [87]. As discussed in Chapter 2, it was found that the amount of time between death and refrigeration was not actively recorded at Salt River Mortuary and thus was not possible to include in this analysis. Nevertheless it should be noted that all bodies are admitted to Salt River Mortuary within 24 hours of first call of the forensic pathology officers to the crime scene (Figure 5.1).

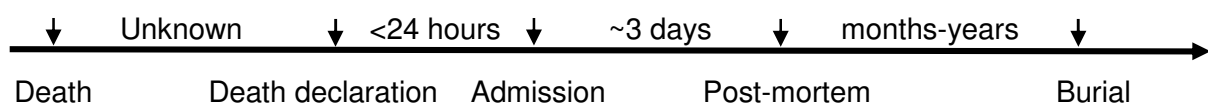


Figure 5-1: Line diagram indicating the various time intervals of importance when processing post-mortem samples

Of particular importance is the time from death to death declaration, often unknown, as this is suspected to affect DNA degradation most actively (results adapted from Chapter 2)

The time between death declaration and sample collection ranged between 1 and 887 days (Figure 4.8), and no significant association was found between time and number of markers meeting the analytical threshold (54RFU) ($p=0.6485$, Section 4.5). Of note, a full DNA profile was obtained from the individual who had been stored at Salt River Mortuary for 887 days; however, the time between death and death declaration, as well as environmental conditions of this case, are unknown.

DNA degradation was quantitatively measured for samples undergoing DNA extraction through the use of the Quantifiler Trio Kit (Thermo Fisher Scientific, MA, USA). Only two post-mortem samples showed high levels of degradation (>4.00) and these were those extracted from buccal swabs, wherein a resulting partial DNA profile was obtained. The poor quality of DNA extracted was buccal swabs was in agreement with that found by Livy *et al.* (2012) [157]. The remaining samples were collected through buccal swabs for processing as lysates, as such no quantitative measure of degradation was utilised.

One advancement that could be developed into new DNA profiling kits, particularly those designed for a direct PCR approach, is the inclusion of a quality sensor, similar to that included in the Qiagen 24plex GO! Kit [136], and Investigator Argus Y-12 QS Kit [158]. This quality sensor provides an added assessment of PCR efficiency and possible presence of

inhibitors thereof. This is of particular value in circumstances where a direct PCR approach is utilised, wherein a quality assessment step is not included.

The higher success rate of extracted DNA samples obtained from blood (Section 4.3.1, $p=0.01$), compared to that of lysate samples, suggests that blood remains the gold standard sample type for forensic human identification for deceased individuals [1]. Human identification through the performance of DNA profiling has become more valued and utilised in SA following the enactment of the Criminal Law (Forensic Procedures) Amendment Act in 2015 [72]. This Act allows for the collection of buccal swabs for development of the National Forensic DNA Database; however this has not fully been adopted into the post-mortem setting, due to the infancy of the new law. Within the post-mortem setting the collection of biological samples from decedents falls under the mandate of the VIC and SAPS. VIC is responsible for the identification of unidentified remains wherein decomposition or skeletonisation is at play, whereas the SAPS investigating officer (I/O) is required to obtain fingerprints and buccal swabs from fresh bodies. As such the collection of biological samples by the forensic pathologists are considered additional samples for identification, which are provided to SAPS should the body remain unidentified after seven days of death [100].

This process is often complicated and delayed due to the large caseload experienced at Salt River Mortuary as well as the high crime rate in SA, and thus caseload on the I/O as well. Decedents then are required to be stored for long periods of times (up to years), as a state burial cannot take place unless samples have been taken for identification purposes (Chapter 2), and a result has been obtained from the SAPS forensic science laboratory. For these reasons the current research question arose regarding the success of DNA profiling of deceased individuals who have been stored at the mortuary for extended periods of time.

The ability to obtain single source DNA profiles from lysate buccal swabs collected from deceased individuals shows the value of collection of biological samples for identification processes. Furthermore, full DNA profiles were obtained in a number of cases that had been stored for over the required 30 days of storage prior to burial, with some having been stored for over two years (887 days since death declaration) (Figure 4.8).

Storage of DNA samples, either extracted or crude is very important for the maintenance of DNA integrity [67]. Similarly, the storage of bodies is as important in order to slow down decomposition and to preserve the state of the body for autopsy and visual identification [86,87,89,159]. Prior to autopsy at Salt River Mortuary, bodies are generally stored at 4 °C, and following post-mortem at -20 °C [67]. It has been hypothesised that the ability to obtain DNA through buccal swabs is complicated when the body is frozen. In this research study,

the temperature at which bodies were stored did not appear to affect the quality of DNA profiles obtained ($p=0.203$) (Section 4.3, Figure 4.8).

Another complication faced with post-mortem DNA samples, is the presence of inhibitors of PCR amplification [160]. Inhibitors prevent the amplification of the target region either through binding to the *Taq* polymerase or to the DNA itself, and has been shown to affect larger amplicons first [160]. This is characterised by the presence of a high peak in the first marker of each dye channel and the gradual failure of larger markers to reach the analytical threshold or be amplified at all [160]. Inhibition was found to be an issue with samples extracted from mouth rinse samples and thus was presumed to affect the quality of buccal swab DNA. However, no severe cases of inhibition were observed, as was observed by Tredoux *et al.* (2015) [71]. It is possible that the inhibition was masked by the presence of DNA degradation, as it is often difficult to discern the two when full DNA profiles are being obtained albeit with the presence of a ski-slope [160]. Nevertheless, the suspected presence of inhibition was overcome in a small number of samples ($n=4$) through performance of further dilution of the stock sample so as to reduce the concentration of inhibitors [160]. However, with this additional dilution step, the amount of input DNA is also reduced and therefore an additional cycle was included in the PCR.

Due to the observation and quantitative assessment of PCR inhibition in extracted DNA samples from mouth rinse samples, it can be presumed that lysate buccal samples may also be affected by PCR inhibition. However, the first time success rate of 72.3% ($n=73/101$) observed for the lysate samples (Table 4.2), suggests that perhaps a direct PCR approach should be implemented initially. Thereafter, for the samples that failed, DNA extraction from the lysate could be applied, followed by a quantification assay. Another possible avenue would be to assess if a qPCR assay can be performed directly on the lysate samples, wherein the Quantifier Trio DNA Quantification kit (Thermo Fisher Scientific, MA, USA) is used to assess both DNA degradation and the presence of PCR inhibition, without extracted the DNA.

5.3.2 Flocked nylon and cotton swabs

In a subset of cases ($n=18$), both cotton swabs and flocked nylon swabs were collected from the same individual (Figure 4.9). The use of flocked swabs yielded three additional full DNA profiles when compared to the cotton swabs. This finding is most likely due to the material and composition of the swab, but may have also been influenced by the order of sampling. Cotton swabs were obtained before the flocked swab; it is thus possible that the mechanical

action of buccal cell collection with the cotton swab may have released cells which were subsequently available for collection with the second (flocked) swab. The first collection may have also warmed the collection area and thawed the cells further (especially in the stored bodies at -20 °C) prior to swabbing with the flocked swab, thereby possibly further releasing cells for the second (flocked) sample collection.

One interesting observation was that flocked swabs produced greater ski slope effects and peak imbalance than that obtained with the cotton swab (Figure 4.9). A possible explanation for this is that the flocked nylon swab releases more cellular content than cotton swabs, but with this there is increased chance of releasing microbial organisms which degrade DNA and possible inhibitors of DNA such as amylase [161]. Additionally, flocked nylon swabs have been noted to leave fibres on rough substrates swabbed, indicating the possibility of nylon fibres dislodging during extended mixing steps of processing, which may inhibit PCR [162]. Verdon *et al.* (2014) noted that flocked swabs have a poor ability to preserve DNA quality [162], thereby supporting the findings within this research project.

Brownlow *et al.* (2012) identified that both cotton and flocked nylon swabs yield sufficient yields of DNA. However the yields varied substantially based on the extraction kit used, and in all combinations cotton swabs always obtained greater yields of DNA [163]. Another study evaluated a number of different swabs and concluded that flocked swabs produce lower yields of DNA than other swab types [162], which was not evaluated in this study.

One criticism of the flocked nylon swabs for forensic use, was the inability to effectively swab hard surfaces due to the flexible nature of the swab shaft [163]. This was observed in this study as well, where difficulty was encountered swabbing the inner cheek of frozen bodies. While far more variable and of reduced quality, the obtaining of a poor quality DNA profile from a flocked swab is better than obtaining no DNA profile at all from a cotton swab. However, the increased cost of flocked nylon swabs compared to cotton swabs was a factor in the decision to first use cotton swabs in both this research project and with forensic service providers [162].

5.4 PROFILE FREQUENCIES AND GENE DIVERSITY

5.4.1 Profile frequency comparison to South African literature

The number of shared haplotypes (Table 4.4) found in this research study was less than that observed for a previous South African Y-STR population study, with an overall 599/626 (95.68%) singletons as opposed to the 113/199 (56.78%) singletons observed by Leat *et al.* (2004) [93]. The large variation in Y-STR profile resolution between the two studies is attributed to the inclusion of 23 Y-STR markers in this study and only nine (minimal haplotype) in the previous study [93]. Leat *et al.* (2004) reviewed two population groups of specific linguistic background (English Caucasian and Xhosa African) [93], explaining the variation in sample size. Within the African population group the number of singletons was slightly lower (92%; n=184/200) using the Promega PowerPlex® Y23 System in the current research study. However, when looking at only the White population (n=165) the use of 23 Y-STR markers resulted in 161 singletons (97.57%). Of particular interest was that one individual reported themselves as a member of the White population, had the most commonly shared Y-STR profile observed in the African population (Table 4.4). A possible reason for this is that the individual has mixed ancestry, where a distant paternal relation (e.g. paternal great grandfather) was of African ancestry.

A study investigating the genetic composition of the Cape Muslim population in the Western Cape, using 17 Y-STR markers resulted in 104/105 (99.05%) different haplotypes [92], which is comparable to the number of different haplotypes observed in the current research study which made use of 23 Y-STR markers (n=174/175, 99.42%) (Table 4.4, Section 4.3.3). In both studies only one haplotype was shared between two different Coloured individuals, indicating the genetic diversity within the coloured population group, and furthermore within sub-populations therein (Cape Muslim) [92]. Within the current study, comparisons to other population groups were made in terms of gene diversity and number of alleles per population group, these comparisons will be discussed in Section 5.4.3. However, this analysis was not done in the previous study as the aim of that research was to identify discriminatory markers specific to a population group [92].

F_{ST} values and associated p-values determined for full DNA profiles, showed significant difference between some population groups in this study (Figure 4.7). However, Fisher's exact test for inter-population variation only showed significant differences between the African population and each of the other three population groups, even after the application of Bonferroni corrections. The differences in findings may be attributed to fundamental differences in algorithmic calculations, wherein the Fisher's exact test incorporated 100000 step Markov Chain analysis, and F_{ST} looked at the population variation using 100

permutations of population pairwise analyses [141,164]. The population variation scores obtained through F_{ST} and Fisher's Exact test were supported by AMOVA results, wherein it was determined that 99.96% of molecular diversity existed within the population groups, and only 0.04% between population groups.

The findings mentioned above indicate that while separate reference platforms may be required for each population group, the discriminatory value is significant even when all population groups are reviewed as a whole. It is still recommended that for greater confidence to exist with these results, a larger population size should be investigated to see if any differences are obtained.

No other study conducted in a local context has yet evaluated the difference between population groups in terms of full haplotypes, but rather focused on selection of informative markers for each population group [44,90,92,93]. Only one other study has reported South African haplotype frequencies for the PowerPlex® Y23 System to YHRD [165]. This study only published data for 144 IsiXhosa African individuals, though the aim of the study was to evaluate the suitability of genetic markers in different population groups, and not to assess haplotype variation [44].

5.4.2 Profile frequency comparison to international database and literature

At the time of analysis, YHRD consisted of 44 022 haplotypes for the Promega PowerPlex® Y23 System [94]. To further evaluate the commonness of shared haplotypes seen in this study, they were compared to the YHRD database. Of the 10 full profiles shared between 27 individuals in this study (Table 4.4), five of them have not been reported on YHRD (Table 4.5). The percentage of singletons observed in this study ($n=599/626$; 95.69%) was less than that observed by Coble *et al.* (2013) in a Y23 population study in the United States of America (USA) ($n=1026/1032$; 99.42%) [146]. In the USA data, only three haplotypes were shared between two individuals each. The 'high' number of shared haplotypes in this study may simply be because of the genetic composition of the country, or sampling error in which related individuals were accidentally recruited. Nevertheless, the number of singletons and apparent subdivision of genetic composition of the Y-chromosome indicates that the South African population is genetically diverse.

In addition to checking commonness of the DNA profiles to YHRD, the ancestry prediction tool was implemented, making use of minimal haplotype (nine STR markers) data (Table 4.5). The shared haplotype belonging to the Coloured population was predicted to be predominantly of Sub-Saharan descent, followed by admixed genetic composition. The

inability to match certain haplotypes to those existing on YHRD emphasises the need for this research and the addition of South African and African genetic data for international reference. While a number of researchers have contributed minimal haplotype data to the database, there is limited South African data for the Promega PowerPlex® Y23 system on YHRD. It is the intention of this study to contribute this data to the YHRD following implementation of the necessary additional quality control assessments.

Within the White population group, two haplotypes (Table 4.4 and 4.5 F and G) were shared between two individuals each, as was the case with the Indian/Asian population group (Table 4.4 and 4.5 I and J). In each of the above situations, the individuals who shared the same haplotype were found to share a surname as well. The two shared haplotypes in the White population appear to belong to individuals of White Afrikaans ancestry [44], possibly explaining the lack of a match to any PowerPlex® Y23 haplotypes on YHRD. It is unknown whether these people are direct relatives or not, as at the time of collection no known brothers or father son pairs were recruited. Furthermore, due to the confidential manner in which samples were obtained the knowledge of somewhat related individuals would only be known had the participant known and said so themselves.

ForenSeq DNA Signature Prep assay, has the ability to genotype bio-geographical ancestry and phenotype informative SNPs concurrently with STRs [53], thereby generating a report containing estimated biogeographical ancestry, and hair and eye colour (Table 4.7) [145]. These estimations have value in forensics, particularly when physical appearance is impaired, such as with skeletal or severely decomposed remains. However, in SA the performance of such analyses is prohibited legally [72]. The Criminal Law Forensic Procedures Amendment Act of 2013, mandates that no medical, historical, behavioural or phenotypic (except for biological sex) data may be obtained from a forensic DNA profile [72]. In this study, consent for the performance of such phenotypic and bio-geographical ancestry predictions was obtained from 41 living individuals.

Within the deceased population, the obtaining of consent often poses various ethical concerns. The individual may have consented to use of their deceased body and/or samples for research during their lifetime; however, if they did not, the decedent themselves is no longer able to provide consent; as such next-of-kin are consulted for consent. However, this is not always possible, particularly in the context of unidentified and unclaimed individuals, as no next-of-kin is available or known. In such instances, the ability to estimate a phenotypic profile for the individual may be useful. These phenotypic features can be circulated to the general public through the use of various platform such as social media or online tools similar to NamUS [130], with the intention of generating possible investigative

leads. While legally not allowed in SA, it cannot be ignored the generation of such prediction phenotype profiles could assist in the identification procedure. It is therefore recommended that studies continue to evaluate the value of such analyses in a research context in hopes of changing legislature to allow for this under specific circumstances.

5.4.3 Gene diversity comparisons to local and international literature

Of the 23 markers included in the PowerPlex® Y23 System (Promega, WI, USA), nine were the markers for the YHRD minimal haplotype (DYS389I/II, DYS19, DYS391, DYS390, DYS392, DYS391, DYS393, DYS385a/b) [94]. DYS391 had low gene diversity ($GD < 0.5$) in three of the four population groups, and had the lowest gene diversity overall (Table 4.6). Additionally, DYS392 ($GD = 0.165$) and DYS393 ($GD = 0.391$) had low gene diversity scores within the African and White population groups respectively, but had good gene diversity scores ($GD > 0.5$) in all other population groups. Gene diversity of DYS393 observed here was the same as the White (Caucasian) population group evaluated in a USA study, however, the gene diversity of DYS392 in the African population in the current study was 2.6 fold less than that in the Afro-American population ($GD = 0.442$) in USA [146]. This variation in gene diversity of DYS392 suggests that the genetic composition of African individuals in SA is different to that of African Americans.

Whilst having high gene diversity in all population groups, DYS385a/b in the White population had the lowest gene diversity (0.854) of all population groups (Table 4.6). This was in agreement with the findings of Cloete *et al.* (2010) and Coble *et al.* (2013) [92,146]. The DYS385a/b marker had 17 alleles in 69 different combinations, which is one more allele in 3 more combinations than in this study (Appendix H). However, the overall gene diversity for this marker was slightly higher in the current research study ($GD = 0.944$) than that observed by Coble *et al.* (2013) ($GD = 0.929$) [146].

The additional fourteen markers (DYS576, DYS448, DYS481, DYS549, DYS533, DYS438, DYS437, DYS570, DYS635, DYS439, DYS643, DYS458, DYS456, YGATAH4) had overall diversities ranging from 0.532 to 0.811. However, within the African population three of these markers [(DYS533; $GD = 0.45$), (DYS437; $GD = 0.145$), (DYS438, $GD = 0.321$)] had gene diversities below 0.5. This observation suggests that these three markers are not informative within the African population group in SA. This was also observed by D'Amato *et al.* (2009) who determined a gene diversity of 0.09 for DYS437 within the isiXhosa African population [90]. The slight difference in gene diversity observed for DYS437 between studies, is likely as a result of this study considering haplotypes from African individuals in South African irrespective of home language. Nevertheless the interpretation of these gene diversities is

that these specified markers are not discriminatory or informative within the African population group.

As found by Coble *et al.* (2013), four additional markers (DYS481, DYS576, DYS570, DYS643) fell within the top third of overall gene diversity, with comparable values [146]. While DYS456 was found to show variable gene diversity depending on US population group [146], the same was not observed in this study. These differences indicate that variation of gene diversity is dependent on population group. As such selection of markers for kit development should potentially be based on which markers are the most informative across all groups, if being designed for international use.

Few markers analysed were common between the current research and previous South African research, with at most one study sharing 10 markers with the PowerPlex® Y23 system used [92]. One (DYS481) and five markers were shared between this research and publications by D'Amato *et al.* in 2010 and 2009 respectively [90,91]. Gene diversities calculated for these common markers was comparable between studies.

5.5 THE VALUE OF NEXT GENERATION SEQUENCING TECHNOLOGY

5.5.1 Coverage

NGS technology allows for the identification of alleles based on sequence variation, whereas capillary electrophoresis based application (e.g. DNA profiling) identifies alleles according to total fragment size [166]. In this study, complete NGS-profiles were obtained for 10 of the 24 samples analysed. It should also be considered that some of the samples that failed on NGS, were selected due to failure in the DNA profiling assay, and thus failure may be attributed to poor sample quality.

Coverage (number of reads) largely depends on the quality of the sample, wherein samples with high levels of inhibitors or degradation are expected to yield less coverage for larger targets [53]. During developmental validation of the ForenSeq DNA Signature Prep Assay, with use of the MiSeq FGx™ instrument, it was found that tannic acid and haematin were the greatest inhibitors of loci genotyping [53]. Tannic acid is commonly found in many plant based foods, alcohols, and some pharmaceutical products, whereas haematin is an active component of blood [160]. As such the presence of tannic acid is not unexpected when obtaining buccal swabs from individuals (living or deceased), who may have eaten, drank or taken medication prior to swabbing. Haematin is expected in post-mortem samples where blood has been released from vessels and dries. These findings are possible reasons for the

observation of partial or failed NGS assays in four deceased, and ten living individuals (Table 4.3).

Secondly, the performance of accurate dilutions for input DNA concentration is vital to the success of the assay. In a previous NGS assay lysate samples that failed were not diluted, and in subsequent assays were successful following the performance of dilution. Of the samples that failed to produce full genotype reports in this study, thirteen were lysate samples wherein no quantification is performed. It is therefore possible that the variable coverage of samples is due to the variable nature of input DNA concentration (possibly too concentrated) and quality. Furthermore, with the Primer Set B assay, a maximum of 32 samples can be processed at a single time to ensure successful coverage [53]. While only 24 samples from this cohort were subjected to the NGS assay, they were sequenced together with eight other samples from the laboratory. It has been suggested that in compromised samples, such as those of mixed source, where a lower limit of detection is required, fewer samples should be processed at a single time [53]. It is therefore recommended that internal validation of the ForenSeq DNA Signature Prep assay be performed so as to determine the optimal number of samples to process simultaneously, as well as to improve the quality of NGS profile reports obtained for post-mortem samples.

5.5.2 Concordance analysis

Recent advances in NGS technologies have allowed for the typing and analysis of multiple types of genetic markers simultaneously, and in some cases is motivated to replace traditional DNA profiling assays. This is not always a feasible option, and as such it is important that results obtained from CE based assays is the same as that from NGS [148,166]. Concordance was evaluated based on 20 shared loci between the PowerPlex® Y23 and ForenSeq DNA Signature Prep assays. A concordance level of 98.47% (258/262 loci) was observed in this study, which is slightly lower than that observed by Davis *et al.* (2013) and Devesse *et al.* (2018) [148,166]. Discordance observed was within a single sample and corresponded to the first marker of each dye channel from the electropherogram, suggesting the occurrence of a migration error during CE and not within the NGS assay. In this instance, the NGS technology was found to be superior to that of the CE-based assay that is DNA profiling.

Due to only 20 of 23 loci being shared between the two assay types implemented, it is suggested that various avenues be used to assess concordance of the Y23 kit. These should include NGS, DNA profiling using different chemistries and possibly Sanger

Sequencing. Furthermore, it has been noted that the results be interpreted with all knowledge of the difference chemistries and migration patterns [148]. The findings of this research study support that of Devesse *et al.* (2018) that NGS applications may be of particular importance when resolving DNA profiles of known or suspected mixture samples [166].

This was noted in the current research study wherein one sample failed the single source assessment on the ForenSeq DNA Signature Prep Assay (Illumina, CA, USA). Four Y-chromosome markers showed two typed alleles, where only one should be obtained for such markers. However, for all four of these markers low coverage was obtained. It is possible that this 'mixed' source identity, could simply be due to the typing and inability to discern between stutter and true allele identities. It should be noted, that this possible mixture sample also showed to be of multiple DNA origin on the DNA profiling electropherogram. This was a post-mortem samples that was obtained from decedent who had been stored at 4°C and 48 days at the time of sample collection. There were an additional two samples obtained from deceased individuals that may be mixtures following electropherogram analysis, however these were not confirmed using NGS. Nevertheless, the observation of possible mixed source DNA samples from post-mortem samples suggest that storage conditions, or handling of the bodies may be introducing contamination. This could be through the improper cleaning of tools, as well as failure to change gloves regularly or at appropriate times. Further, this may have occurred prior to admission to Salt River Mortuary, during the post-mortem, storage or removal of the body for burial.

A known deletion at the DYS448 marker results in the duplication of DYS576. This was suspected in five samples following the performance of Y-STR profiling (Figure 4.5). These samples all underwent NGS and did not show any sequence at the DYS448 marker; however this did not constitute as confirming concordance in this study, as at least one other marker also failed to amplify and sequence using NGS in these samples; and thus it cannot be said if the absence of the sequence at the DYS448 marker was due to the deletion or poor sample quality. Therefore it is recommended that either optimisation of the NGS assay be undertaken, or to implement an alternative sequencing technology such as Sanger Sequencing to confirm these results.

5.6 LIMITATIONS AND FUTURE STUDIES

Due to the limited Y-STR data available for the South African population, the value of this research was clear in the sense that this data is needed for the statistical interpretation of Y-STR data in a forensic context. The results obtained in this study, potentially indicate that the population groups in SA are statistically different in terms of the Y-chromosome. In order to confirm this, the research should be continued and sample sizes should be enlarged. The need for a larger population size is motivated due to the admixture of South African population groups, as well as complex sub-populations and cultural or religious groupings.

The observation of eight individuals, who shared four haplotypes and surnames, highlights a possible limitation in this study at the sampling phase. Firstly, it is difficult to track whether individuals who are related (paternal cousins or brother) may have both provided sample if they are not in contact with each. Furthermore, it is possible that individuals may not know that they are biologically related, or that the shared haplotype is due to chance and genetic mutation over many generations.

This study confirmed that the Promega PowerPlex® Y23 System is suitable for DNA profiling of both extracted DNA and lysate samples, with little difference between success rates depending on sample type. However, in SA legislation allows for the collection of buccal cells, recommended to be with the EasiCollect™ Device [72]. This device makes use of an FTA card for the preservation of DNA integrity, which was not a sample type validated or evaluated in this study. As such future research should validate the use of FTA cards (blood spotted and buccal cells), and assess the quality of this sample type in comparison to others.

Within the deceased population, it was found that the amount of time since death declaration did not affect the quality of the DNA profile obtained, provided that appropriate storage at either 4 °C or -20 °C is implemented (Figure 4.8). This indicates the value of collecting DNA specimens months or years after death for the purpose of identification. This is particularly important in the local context, both Salt River Mortuary and South Africa, due to the large number of bodies that remain unidentified each year, and who remained stored for lengthy periods of time before pauper burial (approximately 9% per annum at Salt River Mortuary, Chapter 2).

Lastly, a limitation of this study was the small sample size for NGS analysis. Given the variability in the quality of results obtained, a larger sample size would have aided in assessment of the effect different sample types have on the quality of NGS. Furthermore, had more samples been included in the NGS assessment, more weight and confidence could be placed on the concordance results obtained. Nevertheless these results are some

of the first NGS results generated from South African individuals, and provide insight into the next steps that should be undertaken. These results identified the need to dilute lysate samples during library preparation, and to research methods of quantification thereof. While not analysed in this study, the generation of sequence data for STRs analysed provides the opportunity for future research. Herein fragments typed identically during capillary electrophoresis based DNA profiling (e.g. allele 10) may actually differ on the sequence level, which could invariably aid in the resolution of mixture samples.

5.7 CONCLUSION

DNA profiling is a valuable tool for criminal and forensic investigations. Following the enactment of the Criminal Law (Forensic Procedures) Amendment Act of 2013, a national forensic DNA database was established in South Africa. This database aims to aid in combating crime as well as assist in forensic human identification of missing persons and unidentified remains. South Africa currently only uses autosomal DNA profiling for identification purposes. However, the high number (approximately 310 individual annually) of unidentified persons at Salt River Mortuary, and the large male predominance thereof, suggests that Y-chromosomal data may be of value. The retrospective study showed that DNA analysis for human identification was only performed in 23.61% of these unidentified cases.

This research study observed a high discriminatory potential of the Promega PowerPlex® Y23 System for South African population groups, with 599/626 individuals having unique haplotypes. This was further seen through variation in gene diversity indices between population groups, and the low linkage scores observed. Full DNA profiles were obtainable from 85% of deceased individuals irrespective of storage temperature (4 °C or -20 °C), and time from death declaration to sample collection. This was shown through the ability to obtain a DNA profile from a body which had been stored for 887 days. However, the assessment of time effects on DNA profile quality was limited by the lack of data regarding true post-mortem interval, and as such this should be actively researched in future studies.

In order for the current results to be viewed with confidence, an internal validation procedure was established and performed prior to sample processing. This method was developed into a standard operating procedure which will be valuable to resource-stricken laboratories worldwide which are unable to employ external persons to perform the validation assay.

Additionally a subset of samples was processed using next generation sequencing, to verify presence of anomalies and assess concordance between the two chemistries. Concordance was observed in 98.47% of loci analysed, and where discordance was observed NGS proved to be the more reliable molecular assay. The application of the NGS biogeographical ancestry prediction in experimental and casework samples indicates this may be a valuable resource in the identification procedure. Future studies should expand on this aspect of NGS, as well as the suitability of NGS generated STR sequence data in resolving mixture samples.

Overall, this study has generated much needed Y-STR data for South Africa and has demonstrated the value of forensic genetic testing in a post-mortem setting. It has also

revealed some stark gaps and absence of collaborative efforts within the current human identification workflows in South Africa. While previous Y-STR data in South Africa has shown to be of lower discriminatory value, the haplotype data generated here, using the Promega PowerPlex® Y23 system, shows a high level of discrimination with 599 unique haplotypes from 626 individuals. It is anticipated that these results will form the foundation for kinship analyses in South Africa and that other forensic genetics laboratories in South Africa and Africa will be able to draw from these results for the benefit of their own casework. Through a collaborative effort, it is hoped that at least some of the hundreds of unclaimed and unidentified bodies stored at mortuaries nationwide will be able to be reunited with their families, bringing social justice and closure to the living family members.

REFERENCES

- [1] J.M. Butler, *Fundamentals of forensic DNA typing*, Academic Press, Gaithersburg, Maryland, USA, 2009.
- [2] Scientific Working Group on DNA Analysis Methods, *Scientific Working Group on DNA Analysis Methods Interpretation Guidelines for Autosomal STR typing by Forensic DNA Testing Laboratories*, (2013) 1–23.
- [3] A.R. Wyman, R. White, A highly polymorphic locus in human DNA., *Proc. Natl. Acad. Sci. U. S. A.* 77 (1980) 6754–8. doi:10.1073/pnas.77.11.6754.
- [4] A.J. Jeffreys, V. Wilson, S.L. Thein, Hypervariable minisatellite regions in human DNA, *Nature*. 314 (1985) 67–73. doi:10.1038/314067a0.
- [5] A.J. Jeffreys, V. Wilson, S.L. Thein, Individual-specific ‘fingerprints’ of human DNA, *Nature*. 316 (1985) 76–79. doi:10.1038/316076a0.
- [6] L. Roewer, Y chromosome STR typing in crime casework, *Forensic Sci. Med. Pathol.* 5 (2009) 77–84. doi:10.1007/s12024-009-9089-5.
- [7] A. Giusti, M. Baird, S. Pasquale, I. Balazs, J. Glassberg, Application of Deoxyribonucleic Acid (DNA) Polymorphisms to the Analysis of DNA Recovered from Sperm, *J. Forensic Sci.* 31 (1986) 12270J. doi:10.1520/JFS12270J.
- [8] A.J. Jeffreys, V. Wilson, R. Neumann, J. Keyte, Hypervariable minisatellites can be amplified from human DNA by the polymerase chain reaction, using primers from DNA flanking the minisatellite, *Nucleic Acids Res.* 16 (1988) 10953–10971.
- [9] SWGDAM, *History of SWGDAM*, (2017). <https://www.swgdam.org/about-us> (accessed 18 July 2017).
- [10] C.J. Frégeau, R.M. Fourney, DNA typing with fluorescently tagged short tandem repeats: a sensitive and accurate approach to human identification., *Biotechniques*. 15 (1993) 100–19.
- [11] L. Roewer, J. Amemann, N.K. Spurr, K.H. Grzeschik, J.T. Epplen, Simple repeat sequences on the human Y chromosome are equally polymorphic as their autosomal counterparts, *Hum. Genet.* 89 (1992) 389–394. doi:10.1007/BF00194309.
- [12] F.R. Santos, S.D.J. Pena, J.T. Epplen, Genetic and population study of a Y-linked tetranucleotide repeat DNA polymorphism with a simple non-isotopic technique, *Hum Genet.* 90 (1993) 655–656. doi:10.1007/BF00202486.
- [13] The Parliamentary Office of Science and Technology, *The National DNA Database*,

London, United Kingdom, 2010.

- [14] FBI, Frequently Asked Questions on CODIS and NDIS, (2017). <https://www.fbi.gov/services/laboratory/biometric-analysis/codis/codis-and-ndis-fact-sheet> (accessed 19 July 2017).
- [15] US YSTR database organisation, Database Descriptive Statistics, (n.d.).
- [16] E.K. Hanson, J. Ballantyne, Comprehensive annotated STR physical map of the human Y chromosome: Forensic implications, *Leg. Med.* 8 (2006) 110–120. doi:10.1016/j.legalmed.2005.10.001.
- [17] C.P. Kimpton, P. Gill, A. Walton, A. Urquhart, E.S. Millican, M. Adams, Automated DNA profiling employing multiplex amplification of short tandem repeat loci., *Genome Res.* 3 (1993) 13–22.
- [18] Promega Corporation, Spectrum CE system, (2016).
- [19] Qiagen, Investigator 24plex Kits, (2015) 1–4.
- [20] S. Gopinath, C. Zhong, V. Nguyen, J. Ge, R.E. Lagacé, M.L. Short, J.J. Mulero, Developmental validation of the Yfiler® Plus PCR Amplification Kit: An enhanced Y-STR multiplex for casework and database applications, *Forensic Sci. Int. Genet.* 24 (2016) 164–175. doi:10.1016/j.fsigen.2016.07.006.
- [21] ISFG, ISFG History, (2017). <https://www.isfg.org/About/History> (accessed 19 July 2017).
- [22] ENSFI, ENSFI History, (2017). <http://ensfi.eu/history/> (accessed 19 July 2017).
- [23] K.Y. Ngo, G. Vergnaud, C. Johnsson, G. Lucotte, J. Weissenbach, A DNA probe detecting multiple haplotypes of the human Y chromosome., *Am. J. Hum. Genet.* 38 (1986) 407–18.
- [24] M. Kayser, Forensic use of Y-chromosome DNA: a general overview, *Hum. Genet.* 136 (2017) 621–635. doi:10.1007/s00439-017-1776-9.
- [25] Statistics South Africa, Victims of Crime Survey 2016/2017, Pretoria, South Africa, 2017. doi:10.1016/S0022-5223(12)00629-0.
- [26] W. Goodwin, A. Linacre, H. Sibte, *An introduction to Forensic Genetics*, Second, John Wiley and Sons, Oxford. UK, 2011.
- [27] L. Gusmão, M. Brion, A. González-Neira, M. Lareu, A. Carracedo, Y chromosome specific polymorphisms in forensic analysis, *Leg. Med.* 1 (1999) 55–60.

doi:10.1016/S1344-6223(99)80013-3.

- [28] L. Gefridis, K. Welch, *The Forensic Laboratory Handbook Procedures and Practice*, Humana Press, Totowa, NJ, 2011. doi:10.1007/978-1-60761-872-0.
- [29] P. Redhead, M.K. Brown, The acid phosphatase test two minute cut-off: An insufficient time to detect some semen stains, *Sci. Justice*. 53 (2013) 187–191. doi:10.1016/j.scijus.2012.09.004.
- [30] A.F. Dekairelle, B. Hoste, Application of a Y-STR-pentaplex PCR (DYS19, DYS389I and II, DYS390 and DYS393) to sexual assault cases, *Forensic Sci. Int.* 118 (2001) 122–125. doi:10.1016/S0379-0738(00)00481-3.
- [31] B.C.M. Pang, B.K.K. Cheung, Identification of human semenogelin in membrane strip test as an alternative method for the detection of semen, *Forensic Sci. Int.* 169 (2007) 27–31. doi:10.1016/j.forsciint.2006.07.021.
- [32] R. Jewkes, Y. Sikweyiya, R. Morrell, K. Dunkle, Gender Inequitable Masculinity and Sexual Entitlement in Rape Perpetration South Africa: Findings of a Cross-Sectional Study, *PLoS One*. 6 (2011) e29590. doi:10.1371/journal.pone.0029590.
- [33] R. Jewkes, N. Abrahams, The epidemiology of rape and sexual coercion in South Africa: an overview, *Soc Sci Med.* 55 (2002) 1231–1244. doi:10.1016/S0277-9536(01)00242-8.
- [34] R. Jewkes, K. Dunkle, M.P. Koss, J.B. Levin, M. Nduna, N. Jama, Y. Sikweyiya, Rape perpetration by young, rural South African men: Prevalence, patterns and risk factors, *Soc. Sci. Med.* 63 (2006) 2949–2961. doi:10.1016/j.socscimed.2006.07.027.
- [35] K. Wood, Contextualizing group rape in post-apartheid South Africa, *Cult. Heal. Sex.* 7 (2005) 303–317. doi:10.1080/13691050500100724.
- [36] J. Kreike, A. Lehner, Sex determination and DNA competition in the analysis of forensic mixed stains by PCR, *Int. J. Legal Med.* 107 (1995) 235–238. doi:10.1007/BF01245480.
- [37] M. Prinz, K. Boll, H. Baum, B. Shaler, Multiplexing of Y chromosome specific STRs and performance for mixed samples, *Forensic Sci. Int.* 85 (1997) 209–218. doi:10.1016/S0379-0738(96)02096-8.
- [38] I. Sibille, C. Duverneuil, G. Lorin de la Grandmaison, K. Guerrouache, F. Teissière, M. Durigon, P. De Mazancourt, Y-STR DNA amplification as biological evidence in sexually assaulted female victims with no cytological detection of spermatozoa, *Forensic Sci. Int.* 125 (2002) 212–216. doi:10.1016/S0379-0738(01)00650-8.

- [39] A. McDonald, E. Jones, J. Lewis, P. O'Rourke, Y-STR analysis of digital and/or penile penetration cases with no detected spermatozoa, *Forensic Sci. Int. Genet.* 15 (2015) 84–89. doi:10.1016/j.fsigen.2014.10.015.
- [40] F. Neuhuber, E. Klausriegler, G. Kreindl, W. Zahrer, B. Dunkelmann, I. Pickrahn, J. Cemper-Kiesslich, The efficiency of Y-chromosome markers in forensic trace analysis and their inclusion in the Austrian National DNA Database, *Forensic Sci. Int. Genet. Suppl. Ser. 4* (2013) e172–e173. doi:10.1016/j.fsigss.2013.10.089.
- [41] L. Romero-Montoya, H. Martínez-Rodríguez, M.A. Pérez, R. Argüello-García, Relationship of spermatoscopy, prostatic acid phosphatase activity and prostate-specific antigen (p30) assays with further DNA typing in forensic samples from rape cases, *Forensic Sci. Int.* 206 (2011) 111–118. doi:10.1016/j.forsciint.2010.07.012.
- [42] Applied Biosystems, Yfiler® Plus PCR Amplification Kit User Guide, (2016).
- [43] J.M. Thompson, M.M. Ewing, W.E. Frank, J.J. Pogemiller, C.A. Nolde, D.J. Koehler, A.M. Shaffer, D.R. Rabbach, P.M. Fulmer, C.J. Sprecher, D.R. Storts, Developmental validation of the PowerPlex® Y23 System: A single multiplex Y-STR analysis system for casework and database samples, *Forensic Sci. Int. Genet.* 7 (2013) 240–250. doi:10.1016/j.fsigen.2012.10.013.
- [44] N. Leat, L. Ehrenreich, M. Benjeddou, K. Cloete, S. Davison, Properties of novel and widely studied Y-STR loci in three South African populations, *Forensic Sci. Int.* 168 (2007) 154–161. doi:10.1016/j.forsciint.2006.07.009.
- [45] R. Chakraborty, J.M. Opitz, J.F. Reynolds, Paternity testing with genetic markers: Are Y-linked genes more efficient than autosomal ones?, *Am. J. Med. Genet.* 21 (1985) 297–305. doi:10.1002/ajmg.1320210212.
- [46] K.N. Ballantyne, V. Keerl, A. Wollstein, Y. Choi, S.B. Zuniga, A. Ralf, M. Vermeulen, P. De Knijff, M. Kayser, A new future of forensic Y-chromosome analysis: Rapidly mutating Y-STRs for differentiating male relatives and paternal lineages, *Forensic Sci. Int. Genet.* 6 (2012) 208–218. doi:10.1016/j.fsigen.2011.04.017.
- [47] B.S. Weir, A.D. Anderson, A.B. Hepler, Genetic relatedness analysis: modern data and new challenges, *Nat. Rev. Genet.* 7 (2006) 771–780. doi:10.1038/nrg1960.
- [48] E. Heyer, J. Puymirat, P. Dieltjes, E. Bakker, P. de Knijff, Estimating Y Chromosome Specific Microsatellite Mutation Frequencies using Deep Rooting Pedigrees, *Hum. Mol. Genet.* 6 (1997) 799–803. doi:10.1093/hmg/6.5.799.
- [49] M. Kayser, A. Sajantila, Mutations at Y-STR loci: Implications for paternity testing and

- forensic analysis, *Forensic Sci. Int.* 118 (2001) 116–121. doi:10.1016/S0379-0738(00)00480-1.
- [50] M. Goedbloed, M. Vermeulen, R.N. Fang, M. Lembring, A. Wollstein, K. Ballantyne, O. Lao, S. Brauer, C. Krüger, L. Roewer, R. Lessig, R. Ploski, T. Dobosz, L. Henke, J. Henke, M.R. Furtado, M. Kayser, Comprehensive mutation analysis of 17 Y-chromosomal short tandem repeat polymorphisms included in the AmpFSTR® Yfiler® PCR amplification kit, *Int. J. Legal Med.* 123 (2009) 471–482. doi:10.1007/s00414-009-0342-y.
- [51] K.N. Ballantyne, M. Goedbloed, R. Fang, O. Schaap, O. Lao, A. Wollstein, Y. Choi, K. van Duijn, M. Vermeulen, S. Brauer, R. Decorte, M. Poetsch, N. von Wurmb-Schwark, P. de Knijff, D. Labuda, H. Vézina, H. Knoblauch, R. Lessig, L. Roewer, R. Ploski, T. Dobosz, L. Henke, J. Henke, M.R. Furtado, M. Kayser, Mutability of Y-Chromosomal Microsatellites: Rates, Characteristics, Molecular Bases, and Forensic Implications, *Am. J. Hum. Genet.* 87 (2010) 341–353. doi:10.1016/j.ajhg.2010.08.006.
- [52] K.N. Ballantyne, A. Ralf, R. Aboukhalid, N.M. Achakzai, M.J. Anjos, Q. Ayub, J. Balažic, J. Ballantyne, D.J. Ballard, B. Berger, C. Bobillo, M. Bouabdellah, H. Burri, T. Capal, S. Caratti, J. Cárdenas, F. Cartault, E.F. Carvalho, M. Carvalho, B. Cheng, M.D. Coble, D. Comas, D. Corach, M.E. D’Amato, S. Davison, P. de Knijff, M.C.A. De Ungria, R. Decorte, T. Dobosz, B.M. Dupuy, S. Elmrghni, M. Gliwiński, S.C. Gomes, L. Grol, C. Haas, E. Hanson, J. Henke, L. Henke, F. Herrera-Rodríguez, C.R. Hill, G. Holmlund, K. Honda, U.-D. Immel, S. Inokuchi, M.A. Jobling, M. Kaddura, J.S. Kim, S.H. Kim, W. Kim, T.E. King, E. Klausriegler, D. Kling, L. Kovačević, L. Kovatsi, P. Krajewski, S. Kravchenko, M.H.D. Larmuseau, E.Y. Lee, R. Lessig, L.A. Livshits, D. Marjanović, M. Minarik, N. Mizuno, H. Moreira, N. Morling, M. Mukherjee, P. Munier, J. Nagaraju, F. Neuhuber, S. Nie, P. Nilasitsatporn, T. Nishi, H.H. Oh, J. Olofsson, V. Onofri, J.U. Palo, H. Pamjav, W. Parson, M. Petlach, C. Phillips, R. Ploski, S.P.R. Prasad, D. Primorac, G.A. Purnomo, J. Purps, H. Rangel-Villalobos, K. Rębała, B. Rerkamnuaychoke, D.R. Gonzalez, C. Robino, L. Roewer, A. Rosa, A. Sajantila, A. Sala, J.M. Salvador, P. Sanz, C. Schmitt, A.K. Sharma, D.A. Silva, K.-J. Shin, T. Sijen, M. Sirker, D. Siváková, V. Škaro, C. Solano-Matamoros, L. Souto, V. Stenzl, H. Sudoyo, D. Syndercombe-Court, A. Tagliabracci, D. Taylor, A. Tillmar, I.S. Tsybovsky, C. Tyler-Smith, K.J. van der Gaag, D. Vanek, A. Völgyi, D. Ward, P. Willemse, E.P.H. Yap, R.Y.Y. Yong, I.Z. Pajnič, M. Kayser, Toward Male Individualization with Rapidly Mutating Y-Chromosomal Short Tandem Repeats, *Hum. Mutat.* 35 (2014) 1021–1032. doi:10.1002/humu.22599.
- [53] A. Jager, M. Alvarez, C.P. Davis, E. Guzmán, Y. Han, L. Way, P. Walichiewicz, D.

- Silva, N. Pham, G. Caves, J. Bruand, F. Schlesinger, S. Pond, J. Varlaro, C.L. Holt, K.M. Stephens, Developmental Validation of the MiSeq FGx Forensic Genomics System for Targeted Next Generation Sequencing in Forensic DNA Casework and Databasing, *Forensic Sci. Int. Genet.* 28 (2017) 52–70. doi:10.1016/j.fsigen.2017.01.011.
- [54] S.B. Lee, J. Varlaro, C.L. Holt, The Future of Forensic Genomics : Developmental Validation of NGS, *Forensic Mag.* (2017) 1–11.
- [55] Illumina Corp., Targeted Next-Generation Sequencing for Forensic Genomics, (2015) 1–12.
- [56] Promega Corporation, PowerPlex ESI 16 System Technical Manual, Revised 5., Maddison, WI, USA, 2014.
- [57] L. Roewer, DNA fingerprinting in forensics: past, present, future, *Investig. Genet.* 4 (2013) 22. doi:10.1186/2041-2223-4-22.
- [58] S. Chattopadhyay, B. Shee, B. Sukul, Unidentified bodies in autopsy – A disaster in disguise, *Egypt. J. Forensic Sci.* 3 (2013) 112–115. doi:10.1016/j.ejfs.2013.05.003.
- [59] L.J. Paulozzi, C.S. Cox, D.D. Williams, K.B. Nolte, John and Jane Doe: The epidemiology of unidentified decedents, *J. Forensic Sci.* 53 (2008) 922–927. doi:10.1111/j.1556-4029.2008.00769.x.
- [60] C. Cattaneo, D. Porta, D. De Angelis, D. Gibelli, P. Poppa, M. Grandi, Unidentified bodies and human remains: An Italian glimpse through a European problem, *Forensic Sci. Int.* 195 (2010) 167.e1-167.e6. doi:10.1016/j.forsciint.2009.11.008.
- [61] L. Evert, S.H. Rossouw, Unidentified Bodies in Forensic Pathology Practice in South Africa, University of Pretoria, 2011.
- [62] S. Cavard, J.C. Alvarez, P. De Mazancourt, F. Tilotta, P. Brousseau, G.L. de la Grandmaison, P. Charlier, Forensic and police identification of ‘X’ bodies. A 6-years French experience, *Forensic Sci. Int.* 204 (2011) 139–143. doi:10.1016/j.forsciint.2010.05.022.
- [63] S.R. Hughes-stamm, K.J. Ashton, A. Van Daal, Assessment of DNA degradation and the genotyping success of highly degraded samples, *Int. J. Legal Med.* (2011) 341–348. doi:10.1007/s00414-010-0455-3.
- [64] D. Hartman, O. Drummer, C. Eckhoff, J.W. Scheffer, P. Stringer, The contribution of DNA to the disaster victim identification (DVI) effort, *Forensic Sci. Int.* 205 (2011) 52–58. doi:10.1016/j.forsciint.2010.09.024.

- [65] J. Manhart, A. Bittorf, A. Büttner, Disaster victim identification-experiences of the 'Autobahn A19' disaster, *Forensic Sci. Med. Pathol.* 8 (2012) 118–124. doi:10.1007/s12024-011-9307-9.
- [66] S.M. Beckett, S.J. Laughton, L.D. Pozza, G.B. McCowage, G. Marshall, R.J. Cohn, E. Milne, L.J. Ashton, Buccal Swabs and Treated Cards: Methodological Considerations for Molecular Epidemiologic Studies Examining Pediatric Populations, *Am. J. Epidemiol.* 167 (2008) 1260–1267. doi:10.1093/aje/kwn012.
- [67] P. Saukko, B. Knight, *Knight's Forensic Pathology*, 3rd ed., Arnold: A member of the Hodder Headline Group, London, UK, 2005.
- [68] R. Owen, P. Bedford, J. Leditschke, A. Schlenker, D. Hartman, Post mortem sampling of the bladder for the identification of victims of fire related deaths, *Forensic Sci. Int.* 233 (2013) 14–20. doi:10.1016/j.forsciint.2013.07.018.
- [69] M. Graw, H.J. Weisser, S. Lutz, DNA typing of human remains found in damp environments, *Forensic Sci. Int.* 113 (2000) 91–95. doi:10.1016/S0379-0738(00)00221-8.
- [70] T. Schwark, A. Heinrich, A. Preuße-Prange, N. Von Wurmb-Schwark, Reliable genetic identification of burnt human remains, *Forensic Sci. Int. Genet.* 5 (2011) 393–399. doi:10.1016/j.fsigen.2010.08.008.
- [71] S. Tredoux, S. Mfolozi, K. Shires, Efficiency of Buccal DNA Sampling Device in the Mortuary, *J. Forensic Investig.* 3 (2015). doi:10.13188/2330-0396.1000023.
- [72] Department of Justice. South Africa, Act No. 37 of 2013: Criminal Law (Forensic Procedures) Amendment Act, 2013, <http://www.justice.gov.za/legislation/acts/2013-037.pdf>, 2014.
- [73] K. Shintani-Ishida, K. Harada, M. Nakajima, K. Yoshida, Usefulness of blood vessels as a DNA source for PCR-based genotyping based on two cases of corpse dismemberment, *Leg. Med.* 12 (2010) 8–12. doi:10.1016/j.legalmed.2009.09.004.
- [74] Y. Sato, H. Motani, H. Inoue, M. Hayakawa, D. Yajima, S. Nagasawa, K. Kobayashi, K. Sato, K. Otsuka, H. Iwase, Multiplex STR typing of aortic tissues from unidentified cadavers, *Leg. Med.* 11 (2009) S455–S457. doi:10.1016/j.legalmed.2009.02.064.
- [75] G. Kulstein, T. Hadrys, P. Wiegand, As solid as a rock—comparison of CE- and MPS-based analyses of the petrosal bone as a source of DNA for forensic identification of challenging cranial bones, *Int. J. Legal Med.* (2017) 1–12. doi:10.1007/s00414-017-1653-z.

- [76] S.M. Edson, A.F. Christensen, S.M. Barritt, A. Meehan, M.D. Leney, L.N. Finelli, Sampling of the cranium for mitochondrial DNA analysis of human skeletal remains, *Forensic Sci. Int. Genet. Suppl. Ser.* 2 (2009) 269–270. doi:10.1016/j.fsigss.2009.09.029.
- [77] T. Schwark, A. Heinrich, N. Von Wurmb-Schwark, Genetic identification of highly putrefied bodies using DNA from soft tissues, *Int. J. Legal Med.* 125 (2011) 891–894. doi:10.1007/s00414-010-0537-2.
- [78] M. Allouche, M. Hamdoum, P. Mangin, V. Castella, Genetic identification of decomposed cadavers using nails as DNA source, *Forensic Sci. Int. Genet.* 3 (2008) 46–49. doi:10.1016/j.fsigen.2008.07.008.
- [79] A. Schlenker, K. Grimble, A. Azim, R. Owen, D. Hartman, Toenails as an alternative source material for the extraction of DNA from decomposed human remains, *Forensic Sci. Int.* 258 (2016) 1–10. doi:10.1016/j.forsciint.2015.10.025.
- [80] H. Motani, K. Sakurada, T. Akutsu, H. Ikegaya, M. Hayakawa, Y. Sato, D. Yajima, K. Sato, K. Kobayashi, H. Iwase, Usefulness of dura mater in providing DNA samples for identifying cadavers, *J. Forensic Sci.* 51 (2006) 888–892. doi:10.1111/j.1556-4029.2006.00157.
- [81] N. Ludyga, B. Grünwald, O. Azimzadeh, S. Englert, H. Höfler, S. Tapio, M. Aubele, Nucleic acids from long-term preserved FFPE tissues are suitable for downstream analyses, *Virchows Arch.* 460 (2012) 131–140. doi:10.1007/s00428-011-1184-9.
- [82] S. Paireder, B. Werner, J. Bailer, W. Werther, E. Schmid, B. Patzak, M. Cichna-markl, Comparison of protocols for DNA extraction from long-term preserved formalin fixed tissues, *Anal. Biochem.* 439 (2013) 152–160. doi:10.1016/j.ab.2013.04.006.
- [83] K.M. Reid, S. Maistry, R. Ramesar, L.J. Heathfield, A review of the optimisation of the use of formalin fixed paraffin embedded tissue for molecular analysis in a forensic post-mortem setting, *Forensic Sci. Int.* 280 (2017) 181–187. doi:10.1016/j.forsciint.2017.09.020.
- [84] G.C. Calacal, D.L.T. Apaga, J.M. Salvador, J.A.D. Jimenez, L.J. Lagat, R.P.F. Villacorta, M.C.F. Lim, R.D.R. Fortun, F.A. Datar, M.C.A. De Ungria, Comparing different post-mortem human samples as DNA sources for downstream genotyping and identification, *Forensic Sci. Int. Genet.* 19 (2015) 212–220. doi:10.1016/j.fsigen.2015.07.017.
- [85] P.E. Raimann, J.B. Picanço, D.S.B.S. Silva, T.C.K. Albuquerque, F.J.O. Paludo, C.S.

- Alho, Procedures to recover DNA from pre-molar and molar teeth of decomposed cadavers with different post-mortem intervals, *Arch. Oral Biol.* 57 (2012) 1459–1466. doi:10.1016/j.archoralbio.2012.08.014.
- [86] J. Burger, S. Hummel, B. Hermann, DNA preservation: A microsatellite DNA based study on ancient skeletal remains, *Electrophoresis.* 20 (1999) 1728–1729.
- [87] M. Fondevila, C. Phillips, N. Naverán, M. Cerezo, A. Rodríguez, R. Calvo, L.M. Fernández, Á. Carracedo, M. V. Lareu, Challenging DNA: Assessment of a range of genotyping approaches for highly degraded forensic samples, *Forensic Sci. Int. Genet. Suppl. Ser. 1* (2008) 26–28. doi:10.1016/j.fsigss.2007.10.057.
- [88] A. Sorensen, E. Rahman, C. Canela, D. Gangitano, S. Hughes-Stamm, Preservation and rapid purification of DNA from decomposing human tissue samples, *Forensic Sci. Int. Genet.* 25 (2016) 182–190. doi:10.1016/j.fsigen.2016.05.013.
- [89] A.S. Holmes, M.G. Roman, S. Hughes-Stamm, In-field collection and preservation of decomposing human tissues to facilitate rapid purification and STR typing, *Forensic Sci. Int. Genet.* 36 (2018) 124–129. doi:10.1016/j.fsigen.2018.06.015.
- [90] M.E. D’Amato, M. Benjeddou, S. Davison, Evaluation of 21 Y-STRs for population and forensic studies, *Forensic Sci. Int. Genet. Suppl. Ser. 2* (2009) 446–447. doi:10.1016/j.fsigss.2009.08.091.
- [91] M.E. D’Amato, L. Ehrenreich, K. Cloete, M. Benjeddou, S. Davison, Characterization of the highly discriminatory loci DYS449, DYS481, DYS518, DYS612, DYS626, DYS644 and DYS710, *Forensic Sci. Int. Genet.* 4 (2010) 104–110. doi:10.1016/j.fsigen.2009.06.011.
- [92] K. Cloete, L. Ehrenreich, M.E. D’Amato, N. Leat, S. Davison, M. Benjeddou, Analysis of seventeen Y-chromosome STR loci in the Cape Muslim population of South Africa, *Leg. Med.* 12 (2010) 42–45. doi:10.1016/j.legalmed.2009.10.001.
- [93] N. Leat, M. Benjeddou, S. Davison, Nine-locus Y-chromosome STR profiling of Caucasian and Xhosa populations from Cape Town, South Africa, *Forensic Sci. Int.* 144 (2004) 73–75. doi:10.1016/j.forsciint.2004.02.022.
- [94] S. Willuweit, L. Roewer, Database composition, YHRD. (2017). <https://yhrd.org/pages/resources/composition>.
- [95] A. Raina, T.D. Dogra, A. a Leenaars, B. Yadav, C. Bhera, S. Lalwani, L. Leenaars, Identity of victims from fragmented and decomposed remnants by DNA profiling in a case of serial killings., *Med. Sci. Law.* 50 (2010) 220–223.

doi:10.1258/msl.2010.010106.

- [96] Y. Fujita, S. Kubo, Y. Yoshida, I. Tokunaga, T. Gotohda, A. Ishigami, Application of AmpFISTR Profiler PCR Amplification kit for personal identification of a putrefied cadaver., *J. Med. Invest.* 51 (2004) 243–6. doi:10.1097/00000542-196009000-00017.
- [97] O.M. Loreille, R.L. Parr, K.A. McGregor, C.M. Fitzpatrick, C. Lyon, D.Y. Yang, C.F. Speller, M.R. Grimm, M.J. Grimm, J.A. Irwin, E.M. Robinson, Integrated DNA and fingerprint analyses in the identification of 60-year-old mummified human remains discovered in an alaskan glacier, *J. Forensic Sci.* 55 (2010) 813–818. doi:10.1111/j.1556-4029.2010.01356.
- [98] B. Budowle, J. Ge, R. Chakraborty, Basic Principles for Estimating the Rarity of Y-STR Haplotypes Derived from Forensic Evidence, *Eighteenth Int. Symp. Hum. Identif.* (2007) 1–15.
- [99] Scientific Working Group on DNA Analysis Methods, *Validation Guidelines for DNA Analysis Methods*, (2016) 1–13. https://media.wix.com/ugd/4344b0_813b241e8944497e99b9c45b163b76bd.pdf (accessed 26 March 2017).
- [100] South Africa Department of Health, *Regulations regarding the rendering of forensic pathology service*, 2018.
- [101] Department of Justice. South Africa, *The Inquests Act 58 of 1959*, <http://www.justice.gov.za/legislation/acts/1959-58.pdf>, 1996.
- [102] South African Government, *The National Health Act No 61 of 2003*, *Gov. Gaz.* 469 (2004) 1–94. doi:102GOU/B.
- [103] E. Andreev, W.A. Pridemore, V.M. Shkolnikov, O.I. Antonova, An investigation of the growing number of deaths of unidentified people in Russia, *Eur. J. Public Health.* 18 (2008) 252–257. doi:10.1093/eurpub/ckm124.
- [104] J.A. Lorente, C. Entrala, J.C. Alvarez, B. Arce, B. Heinrichs, M. Lorente, F. Carrasco, B. Budowle, E. Villanueva, Identification of missing persons: the Spanish ‘Phoenix’ program., *Croat. Med. J.* 42 (2001) 267–70.
- [105] A. Kumar, D. Harish, A. Singh, Cause of death in ‘John Doe & Jane Doe’: A 5 year review, *J. Clin. Diagnostic Res.* 8 (2014) 20–23. doi:10.7860/JCDR/2014/8876.4661.
- [106] Z. Mukhuthu, Pauper burials for unclaimed bodies, *Dispatch Live.* (2015). <https://www.dispatchlive.co.za/news/2015-11-18-pauper-burials-for-unclaimed-bodies/> (accessed 17 June 2018).

- [107] C. Cattaneo, S. Ritz-Timme, H.W. Schutz, M. Collins, E. Waite, H. Boormann, M. Grandi, H.J. Kaatsch, Unidentified cadavers and human remains in the EU: an unknown issue., *Int. J. Legal Med.* 113 (2000) N2-3.
- [108] Western Cape Government, Forensic Pathology Services - Our Facilities, (2018). <https://www.westerncape.gov.za/general-publication/our-facilities> (accessed 2 May 2018).
- [109] Statistics South Africa, City of Cape Town, (2011). http://www.statssa.gov.za/?page_id=993&id=city-of-cape-town-municipality (accessed 16 April 2017).
- [110] T.A.T. Majero, C. Mole, Retrospective analysis of blunt force trauma associated with road traffic accidents in Cape Town (South Africa) over a two-year period., University of Cape Town, 2017.
- [111] C. du Toit, L. Martin, L. Heathfield, Retrospective analysis of abandoned live births, stillbirths and non-viable fetuses admitted to Salt River Mortuary, Cape Town, University of Cape Town, 2017.
- [112] M.M. Kruger, L.J. Martin, S. Maistry, L.J. Heathfield, A systematic review exploring the relationship between infection and sudden unexpected death between 2000 and 2016: A forensic perspective, *Forensic Sci. Int.* 289 (2018) 108–119. doi:10.1016/j.forsciint.2018.05.023.
- [113] Republic of South Africa, Criminal Procedure Act 51 of 1977, 1977.
- [114] Office of the Registrar General & Census Commissioner, India: Chandigarh, (2011). Office of the Registrar General & Census Commissioner, India%0AMinistry of Home Affairs,%0AGovernment of India (accessed 16 June 2018).
- [115] Office of the Registrar General Indian Government Ministry of Home Affairs, Report on Medical Certification of Cause of Death, 2012, (2012) 1–204.
- [116] United States Census Bureau., United States| Population estimates, (2017). <https://www.census.gov/search-results.html?q=population+size&page=1&stateGeo=none&searchtype=web&cssp=SE RP>.
- [117] I. Kinnes, Monograph No 48 From urban street gangs to criminal empires: The changing face of gangs in the Western Cape , June, Africa (Lond). (2000).
- [118] S. Jensen, The security and development nexus in cape town: War on gangs, counterinsurgency and citizenship, *Secur. Dialogue.* 41 (2010) 77–97.

doi:10.1177/0967010609357038.

- [119] N. Abrahams, S. Mathews, L.J. Martin, C. Lombard, N. Nannan, R. Jewkes, Gender Differences in Homicide of Neonates, Infants, and Children under 5 y in South Africa: Results from the Cross-Sectional 2009 National Child Homicide Study, *PLOS Med.* 13 (2016) e1002003. doi:10.1371/journal.pmed.1002003.
- [120] S. Mathews, L.J. Martin, D. Coetzee, C. Scott, Y. Brijmohun, Child deaths in South Africa: Lessons from the child death review pilot, *South African Med. J.* 106 (2016) 851–852. doi:10.7196/SAMJ.2016.v106i9.11382.
- [121] A. Gheorghe, J. Banner, S.H. Hansen, U. Stolborg, N. Lynnerup, Abandonment of newborn infants: A Danish forensic medical survey 1997-2008, *Forensic Sci. Med. Pathol.* 7 (2011) 317–321. doi:10.1007/s12024-011-9253-6.
- [122] L. Du Toit-Prinsloo, C. Pickles, Z. Smith, J. Jordaan, G. Saayman, The medico-legal investigation of abandoned fetuses and newborns — a review of cases admitted to the Pretoria Medico-Legal Laboratory , South Africa, (2016) 569–574. doi:10.1007/s00414-015-1198-y.
- [123] Republic of South Africa, Act No. 66 of 2008: Judicial Matters Amendment Act, 2008, 2009.
- [124] Road Traffic Management Corporation, Road Traffic Report 31 March 2011, (2011) 107.
- [125] F. Parkinson, S. Kent, C. Aldous, G. Oosthuizen, D. Clarke, Road traffic crashes in South Africa: The burden of injury to a regional trauma centre, *South African Med. J.* 103 (2013) 850–852. doi:10.7196/SAMJ.6914.
- [126] P.N. Moeketsi, Prevention and Control of Road Traffic Injuries: A South African Rural Experience, *South African Transp. Conf.* (2000) 8.
- [127] F. Labuschagne, E. De Beer, D. Roux, K. Venter, The Cost of Crashes in South Africa 2016, 36th South African Transp. Conf. 2004 (2017) 474–485.
- [128] D.L. Meadows II, A.J. Pouratian, Automated Fingerprint Identification System, 5869822, 1999. doi:US005485919A.
- [129] R. Hanzlick, S. Clark, The unidentified decedent reporting system: A model national website registry for the unidentified deceased, *Am. J. Forensic Med. Pathol.* 29 (2008) 106–113. doi:10.1097/PAF.0b013e318174f098.
- [130] NamUs, Unidentified Persons Database, (2010). <https://www.identifyus.org/en>

(accessed 17 April 2018).

- [131] L.A.F. da Silva, W. Vilaça, D. Azevedo, G. Majella, I.F. Silva, B.F. Silva, Missing and unidentified persons database, *Forensic Sci. Int. Genet. Suppl. Ser. 2* (2009) 255–257. doi:10.1016/j.fsigss.2009.08.090.
- [132] L. Gusmão, J.M. Butler, A. Linacre, W. Parson, W. Parson, P.M. Schneider, A. Carracedo, Revised guidelines for the publication of genetic population data, *Forensic Sci. Int. Genet.* 30 (2017) 160–163. doi:10.1016/j.fsigen.2017.06.007.
- [133] Scientific Working Group on DNA Analysis Methods, Scientific Working Group on DNA Analysis Methods Interpretation Guidelines for Y-Chromosome STR Typing, (2014) 1–20. https://www.thermofisher.com/content/dam/LifeTech/Documents/PDFs/SWGDAM_YSTR_Guidelines_APPROVED_01092014_v_02112014_FINAL.pdf (accessed 12 February 2017).
- [134] A. Lucassen, K. Ehlers, P.J. Grobler, A.L. Shezi, Allele frequency data of 15 autosomal STR loci in four major population groups of South Africa, *Int. J. Legal Med.* 128 (2014) 275–276. doi:10.1007/s00414-013-0898-4.
- [135] Promega Corporation, SwabSolution™ Kit SwabSolution™ Kit, Madison, WI, United States of America, 2016.
- [136] Qiagen, Investigator 24plex Go! Kit, (2015).
- [137] J. Tokarz, S. Foster, S. Milne, P. Staton, K. Press, Validation of PowerPlex® Y23 PCR Amplification System for Forensic Casework, Colorado Springs, Colorado, n.d.
- [138] Promega Corporation, PowerPlex® 5C Matrix Standard, Madison, WI, USA, 2015.
- [139] Promega Corporation, PowerPlex® Y23 System, (2016).
- [140] W.K. Fung, 10% or 5% match window in DNA profiling, *Forensic Sci. Int.* 78 (1996) 111–118. doi:10.1016/0379-0738(95)01876-X.
- [141] L. Excoffier, H.E.L. Lischer, Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows, *Mol. Ecol. Resour.* 10 (2010) 564–567.
- [142] Y.Y. Shi, L. He, SHEsis, a powerful software platform for analyses of linkage disequilibrium, haplotype construction, and genetic association at polymorphism loci, *Cell Res.* 15 (2005) 97–98. doi:10.1038/sj.cr.7290272.
- [143] Illumina, ForenSeq™ DNA Signature Prep Reference Guide, 2015.

- [144] Illumina, MiSeq FGx Instrument Reference Guide, (2015).
- [145] Illumina, ForenSeq™ Universal Analysis Software User Guide, (2016).
- [146] M.D. Coble, C.R. Hill, J.M. Butler, Haplotype data for 23 Y-chromosome markers in four U.S. population groups, *Forensic Sci. Int. Genet.* 7 (2013) e66–e68. doi:10.1016/j.fsigen.2013.03.006.
- [147] M.E. D’Amato, V.B. Bajic, S. Davison, Design and validation of a highly discriminatory 10-locus Y-chromosome STR multiplex system, *Forensic Sci. Int. Genet.* 5 (2011) 122–125. doi:10.1016/j.fsigen.2010.08.015.
- [148] C. Davis, J. Ge, C. Sprecher, A. Chidambaram, J. Thompson, M. Ewing, P. Fulmer, D. Rabbach, D. Storts, B. Budowle, Prototype PowerPlex® Y23 System: A concordance study, *Forensic Sci. Int. Genet.* 7 (2013) 204–208. doi:10.1016/j.fsigen.2012.06.005.
- [149] M. Kayser, A. Caglià, D. Corach, N. Fretwell, C. Gehrig, G. Graziosi, F. Heidorn, S. Herrmann, B. Herzog, M. Hidding, K. Honda, M. Jobling, M. Krawczak, K. Leim, S. Meuser, E. Meyer, W. Oesterreich, A. Pandya, W. Parson, G. Penacino, A. Perez-Lezaun, A. Piccinini, M. Prinz, C. Schmitt, P.M. Schneider, R. Szibor, J. Teifel-Greding, G. Weichhold, P. de Knijff, L. Roewer, Evaluation of Y-chromosomal STRs: a multicenter study, *Int. J. Legal Med.* 110 (1997) 125–133. doi:10.1007/s004140050051.
- [150] K. Brown, Coloured and black relations in South Africa: The burden of racial hierarchy, *Macalester Int.* 9 (2000) 198.
- [151] E. De Wit, W. Delpont, C.E. Rugamika, A. Meintjes, M. Möller, P.D. Van Helden, C. Seoighe, E.G. Hoal, Genome-wide analysis of the structure of the South African Coloured Population in the Western Cape, *Hum. Genet.* 128 (2010) 145–153. doi:10.1007/s00439-010-0836-1.
- [152] T. Nurse, G. T., Weiner, J. S. & Jenkins, *The Peoples of Southern Africa and their Affinities.*, Oxford University Press., Oxford, 1985.
- [153] M. Caputo, M.C.C. Bobillo, A. Sala, D. Corach, Optimizing direct amplification of forensic commercial kits for STR determination, *J. Forensic Leg. Med.* 47 (2017) 17–23. doi:10.1016/j.jflm.2017.01.003.
- [154] B.A. Myers, J.L. King, B. Budowle, Evaluation and comparative analysis of direct amplification of STRs using PowerPlex® 18D and Identifiler® Direct systems, *Forensic Sci. Int. Genet.* 6 (2012) 640–645. doi:10.1016/j.fsigen.2012.02.005.

- [155] S.E. Cavanaugh, A.S. Bathrick, Direct PCR amplification of forensic touch and other challenging DNA samples: A review, *Forensic Sci. Int. Genet.* 32 (2018) 40–49. doi:10.1016/j.fsigen.2017.10.005.
- [156] K. Gray, D. Crowle, P. Scott, Direct amplification of casework bloodstains using the Promega PowerPlex®21 PCR Amplification System, *Forensic Sci. Int. Genet.* 12 (2014) 86–92. doi:10.1016/j.fsigen.2014.05.003.
- [157] A. Livy, S. Lye, C.K. Jagdish, N. Hanis, V. Sharmila, L.W. Ler, B. Pramod, Evaluation of Quality of DNA Extracted from Buccal Swabs for Microarray Based Genotyping, *Indian J. Clin. Biochem.* 27 (2012) 28–33. doi:10.1007/s12291-011-0154-y.
- [158] Qiagen, Investigator Argus Y-12 QS Handbook, 2015.
- [159] A.S. Holmes, R. Houston, K. Elwick, D. Gangitano, S. Hughes-Stamm, Sample to Insight Comparison of four commercial qPCR kits for analyzing inhibited and degraded forensic samples, (n.d.) 1–10.
- [160] B.R. Mccord, K. Opel, M. Funes, L.M. Jantz, An Investigation of the Effect of DNA Degradation and Inhibition on PCR Amplification of Single Source and Mixed Forensic Samples, Proj. 2006-DN-BX-K006. Final Rep. to US Dep. Justice. (2011) 1–66.
- [161] F.M. Dalmaso G, Bini M, Paroni R, Qualification of high-recovery, flocked swabs as compared to traditional rayon swabs for microbiological environmental monitoring of surfaces, *PDA J Pharm Sci Technol.* 62 (2008) 191–199.
- [162] T.J. Verdon, R.J. Mitchell, R.A.H. van Oorschot, Swabs as DNA collection devices for sampling different biological materials from different substrates, *J. Forensic Sci.* 59 (2014) 1080–1089. doi:10.1111/1556-4029.12427.
- [163] R.J. Brownlow, K.E. Dagnall, C.E. Ames, A Comparison of DNA Collection and Retrieval from Two Swab Types (Cotton and Nylon Flocked Swab) when Processed Using Three QIAGEN Extraction Methods, *J. Forensic Sci.* 57 (2012) 713–717. doi:10.1111/j.1556-4029.2011.02022.x.
- [164] N. Ryman, S. Palm, C. André, G.R. Carvalho, T.G. Dahlgren, P.E. Jorde, L. Laikre, L.C. Larsson, A. Palmé, D.E. Ruzzante, Power for detecting genetic divergence: Differences between statistical methods and marker loci, *Mol. Ecol.* 15 (2006) 2031–2045. doi:10.1111/j.1365-294X.2006.02839.x.
- [165] S. Willuweit, L. Roewer, Information on the National Database ‘South Africa’, YHRD. (2017).
- [166] L. Devesse, D. Ballard, L. Davenport, I. Riethorst, G. Mason-Buck, D. Syndercombe

Court, Concordance of the ForenSeq™ system and characterisation of sequence-specific autosomal STR alleles across two major population groups, *Forensic Sci. Int. Genet.* 34 (2018) 57–61. doi:10.1016/j.fsigen.2017.10.012.

APPENDICES

APPENDIX A

Ethics approval letter (HREC 342/2016): Amendment to add Kate Megan Reid to research study was submitted on 31 March 2017, and approved on 7 April 2017.



UNIVERSITY OF CAPE TOWN
Faculty of Health Sciences
Human Research Ethics Committee



Room E52-24 Old Main Building
Groote Schuur Hospital
Observatory 7925
Telephone [021] 404 7682 • Facsimile [021] 406 6411
Email: nosi.tsama@uct.ac.za
Website: www.health.uct.ac.za/fhs/research/humanethics/forms

17 October 2016

HREC REF: 342/2016

Ms L Heathfield
Pathology
Forensic Medicine
Falmouth Building
Entrance 3, Level 1

Dear Ms Heathfield

PROJECT TITLE: CONSTRUCTING A DNA PROFILE FREQUENCY DATABASE FOR SOUTH AFRICA USING THE QIAGEN INVESTIGATOR 24PLEX GO! KIT (MPhil candidate- L Nel)

Thank you for submitting your study to the Faculty of Health Sciences Human Research Ethics Committee for review.

It is a pleasure to inform you that the HREC has **formally approved** the proof of concept for phase 1 of the above-mentioned study.

Approval is granted for one year until the 30th October 2017.

Please submit a progress form, using the standardised Annual Report Form if the study continues beyond the approval period. Please submit a Standard Closure form if the study is completed within the approval period.

(Forms can be found on our website: www.health.uct.ac.za/fhs/research/humanethics/forms)

Please submit an FHS007 form for the addition of Yusasha Pillay and the removal of Laura Pirjol.

The HREC note that amendments will be submitted for 317/2015 and 158/2016 for the use of samples collected as part of these studies.

We acknowledge that the student L Nel will be involved in this study.

Please note that for all studies approved by the HREC, the principal investigator **must** obtain appropriate institutional approval before the research may occur.

Please quote the HREC REF in all your correspondence.

Please note that the ongoing ethical conduct of the study remains the responsibility of the principal investigator.

HREC 342/2016

APPENDIX B

Information form, consent form and questionnaire



Information form

Title: Constructing a DNA profile frequency database for the South African population

We are inviting you to participate in a research study.

What is the research about?

In forensic science, DNA samples from a scene can be analysed to produce a DNA profile. A DNA profile is a set of numbers which represent variations in the DNA. The DNA profile itself does not reveal any information about you, except for your sex. This is because the variations that are analysed do not code for physical information. How it actually works, is that the DNA profile is searched against a database in hope for a match to a known individual. If it matches, then the probability of this DNA profile occurring in the background population must be calculated. In order to calculate this, the background frequencies of DNA profiles in the population need to be determined. The purpose of this study is to construct a DNA profile frequency database of the background population in South Africa. The value of this information will be that probabilities of DNA profiles will be able to be calculated.

What we are asking from you?

To participate, we ask that you provide the following:

- Informed consent
- Fill in a short questionnaire
- Two buccal swabs (these are swabs from the inside of your cheek) - one on each side.

Before taking these, please rinse your mouth out with water. You may take these samples yourself or under guidance from the researcher.

The buccal swabs that you provide will be analysed in the laboratory using molecular techniques, to generate a DNA profile. The default is that your samples will be destroyed once the DNA profile has been generated, unless you specify that you would like your DNA to be stored for other similar research. In this case, the DNA sample will be stored in an access controlled freezer at UCT for 20 years.

Are there any risks?

No. Buccal swabs are non-invasive and it does not hurt. There will be no risk of any incidental findings regarding your genetics, as the parts of the DNA which are analysed do not code for any physical information about you, except for your sex. The samples you provide will not be able to link you to any crime scenes either. Participation does not have the potential for you to be implicated in any crime, whatsoever. This project is completely separate from the forensics unit in South Africa. This study will provide *anonymous* data upon which other DNA profiles can be interpreted accurately.

Are there any benefits?

There are no direct benefits for participating in the study. The community at large however will benefit as these results will allow for more confidence in DNA evidence presented in South African forensic cases. You will not be compensated for participating in the study.

Will information be confidential?

Yes, your DNA profile will be kept confidential. Your samples will be allocated a unique participation number, which will be used to identify your samples during processing the laboratory. No individual name will be linked to specific samples. The results of the study at large will be anonymised and published, but there is no way you could be linked to the data that will be published.

Voluntary participation and withdrawal

Participation is completely voluntary and your allocated participation number will ensure traceability of your sample, and therefore if you wish to withdraw from the project, you may do so without providing a reason and your sample will be discarded. This will be possible up until the anonymous publishing of the results, at which point, it will not be possible to withdraw your data from the overall results.

Any questions?

If you may have any questions with regards to the rights and welfare of a research subject in the study, please contact the Chairperson of the University Of Cape Town Faculty Of Health Science Human Research Ethics Committee, **Professor Marc Blockman** on (021) 406 6496. If you require any further information about this study please contact Laura Heathfield at (021) 406 6569 or email at laura.heathfield@uct.ac.za.

Participant number: _____



Consent form

1. I confirm that I am South African and 18 years or older.
2. I confirm have read and understand the research information form for the above study and have had the opportunity to ask questions.
3. I agree to participate in this study and understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason.
4. I consent to:
(Tick the appropriate box)
 - The use of my DNA and information to be used ONLY for this study and to be destroyed after conclusion of this project.
 - The use of my DNA and information to be used for this study AND stored for the ONLY purpose of possible future continuation of this specific research only if approved by HREC.
 - The use of my DNA and information to be used for this study AS WELL AS my samples to be stored and used for future research of similar nature that is approved by the HREC.

Printed Name of Subject

Signature of Subject

Date

Printed Name of Witness

Signature of Witness

Date

APPENDIX C

Calculations used for the determination of analytical threshold

Method	Equation	Symbol meaning
1: IUPAC (1976)	$AT_{method\ 1} = \bar{Y}_{blank} + kS_{blank}$	\bar{Y}_{blank} Average RFU for blanks k Kaiser value of 3, >89% confidence S_{blank} standard deviation of blank peaks
2: IUPAC (1995)	$AT_{method\ 2} = \bar{Y}_{blank} + t_{1-\alpha} \frac{S_{blank}}{\sqrt{n}}$	\bar{Y}_{blank} Average RFU for blanks $t_{1-\alpha}$ student's t-table, 99% confidence $\frac{S_{blank}}{\sqrt{n}}$ standard deviation of full data set
3: SWGDAM guideline	$AT_{method\ 3} = 2(Y_{max} - Y_{min})$	Y_{max} Maximum peak height of blank Y_{min} Minimum peak height of blank
4: Miller-DNA dilution	$AT_{method\ 4} = c + 3S_y$	S_y Standard regression error c Y-intercept of RFU vs input DNA
5: IUPAC CEAC (1997)	$AT_{method\ 5} = c + t_{n-1,\alpha} S_y$	S_y Standard regression error $t_{n-1,\alpha}$ Standard regression error c Y-intercept of RFU vs input DNA

Determined analytical thresholds for direct PCR approach

	Overall average Peak height	Standard deviation	Value of constant	Analytical threshold
AT_{method 1}	2.504	1.37891	$k = 3$	6.641
AT_{method 2}	2.504	1.37891	$t_{1-a,v} = 2.326$	5.711
AT_{method 3}	2.504	1.37891	$\bar{Y}_{min} = 0$ $\bar{Y}_{max} = 27$	54

Assay number	Equation of trendline	R ² value	Standard error of regression	AT _{method 4}	AT _{method 5} $t=2.896$
Dilution series 1	$y = 36.385x + 31.646$	0.758	16.81	82.076	80.328
Dilution series 2	$y = 31.894x + 26.22$	0.965	6.40	45.42	44.754
Dilution series 3	$y = 27.704x + 8.353$	1.00	0.20	8.953	8.932
Overall average	$y = 31.11x + 7.9608$	0.892	11.21	41.591	40.424

Determined analytical threshold for extracted DNA approach (blood and mouth rinse sample types)

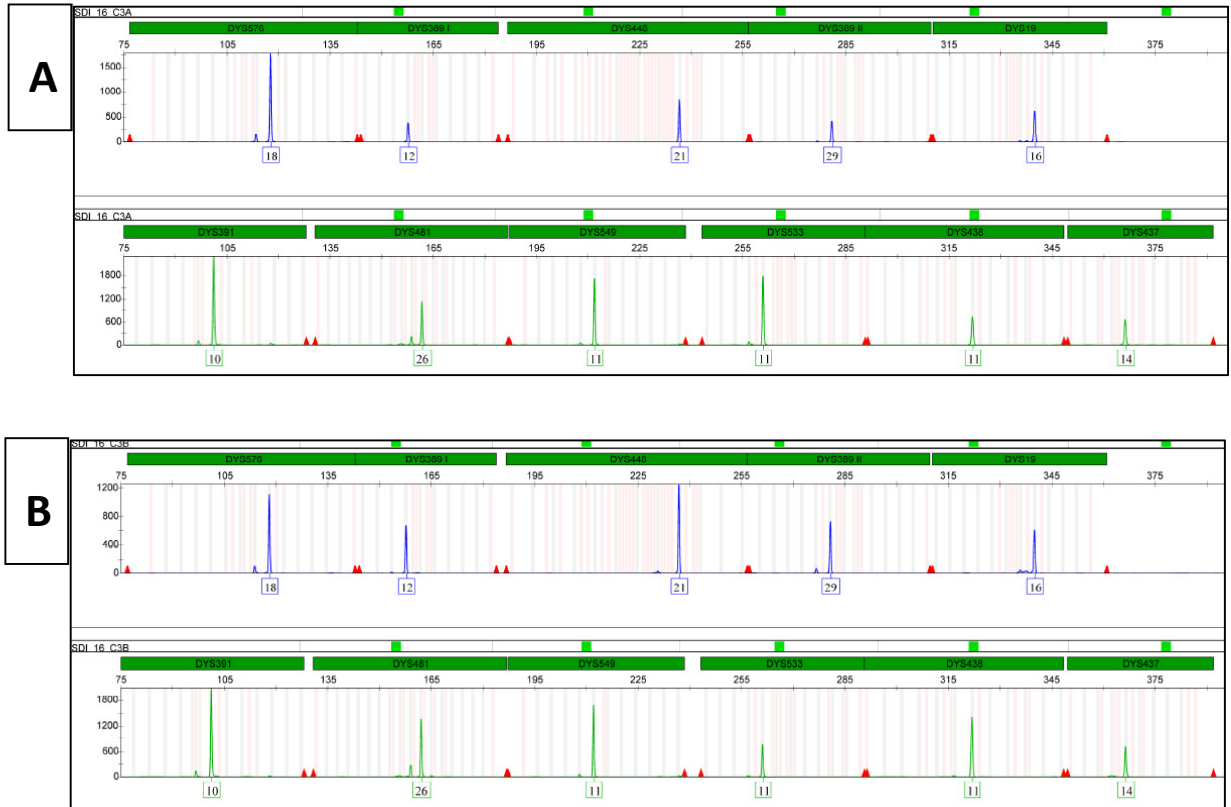
Extracted DNA: combined	Overall average Peak height	Standard deviation	Value of constant	Analytical threshold
AT_{method 1}	3.234	1.691812	$k = 3$	8.321436
AT_{method 2}	3.234	1.691812	$t_{1-a,v} = 2.326$	7.181177712
AT_{method 3}	3.234	1.691812	$\bar{Y}_{min} = 0$ $\bar{Y}_{max} = 34$	68

Extracted DNA: Mouth rinse sample	Equation of trendline	R ² value	Standard error of regression	AT _{method 4}	AT _{method 5} <i>t=2.896</i>
Dilution series 1	$y = 3346.7x - 16.794$	0.9856	69.00443	190.22	183.04
Dilution series 2	$y = 8797.2x - 2.5247$	0.9999	20.36871517	58.58	56.45
Dilution series 3	$y = 7764.6x - 67.768$	0.9939	151.6821	387.28	371.50
Overall average	$y = 6250.2x - 101.04$	0.9565	227.698	582.05	558.37

Extracted DNA: blood samples	Equation of trendline	R ² value	Standard error of regression	AT _{method 4}	AT _{method 5} <i>t=2.896</i>
Dilution series 1	$y = 9541.4x - 191.91$	0.9655	321.8889	773.76	740.28
Dilution series 2	$y = 9165.7x - 25.661$	0.9996	46.34363	113.36	108.54
Dilution series 3	$y = 9366.4x + 167.23$	0.9694	411.217	1400.88	1358.11
Overall average	$y = 9535.9x - 18.443$	1.00	15.47266	28.05	26.44

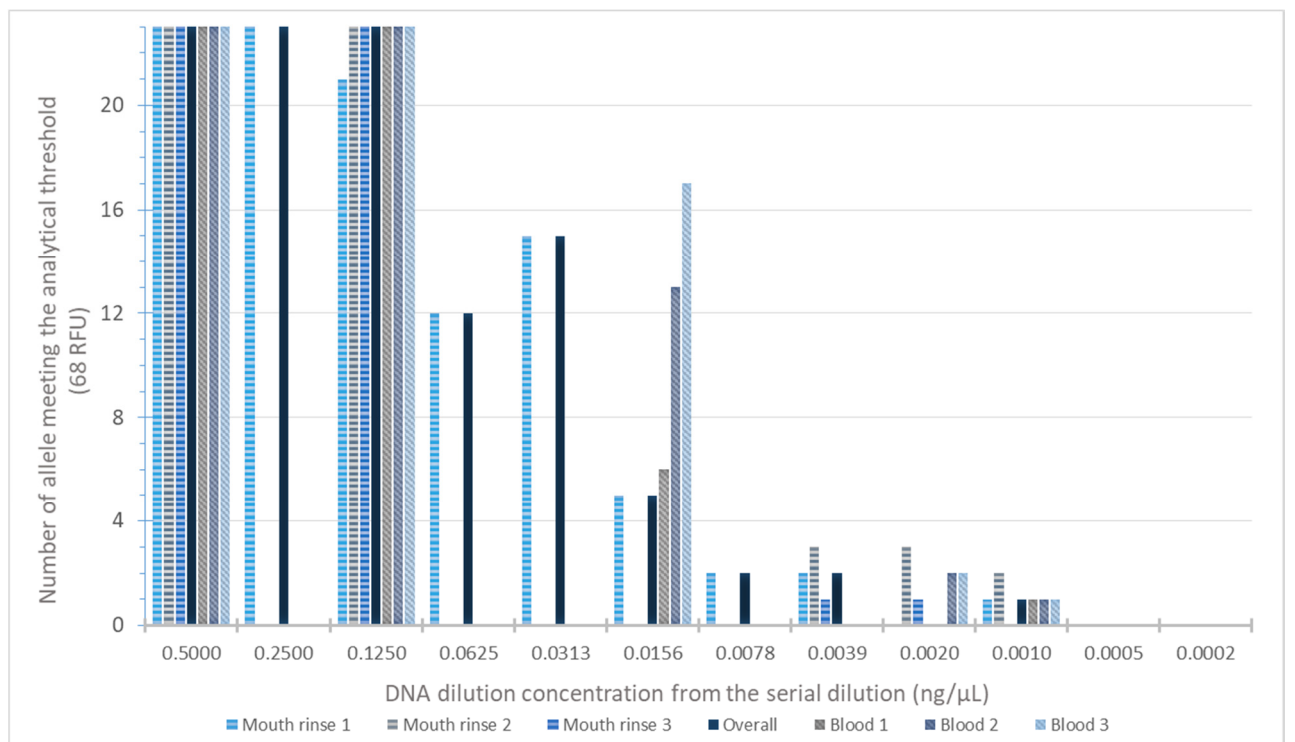
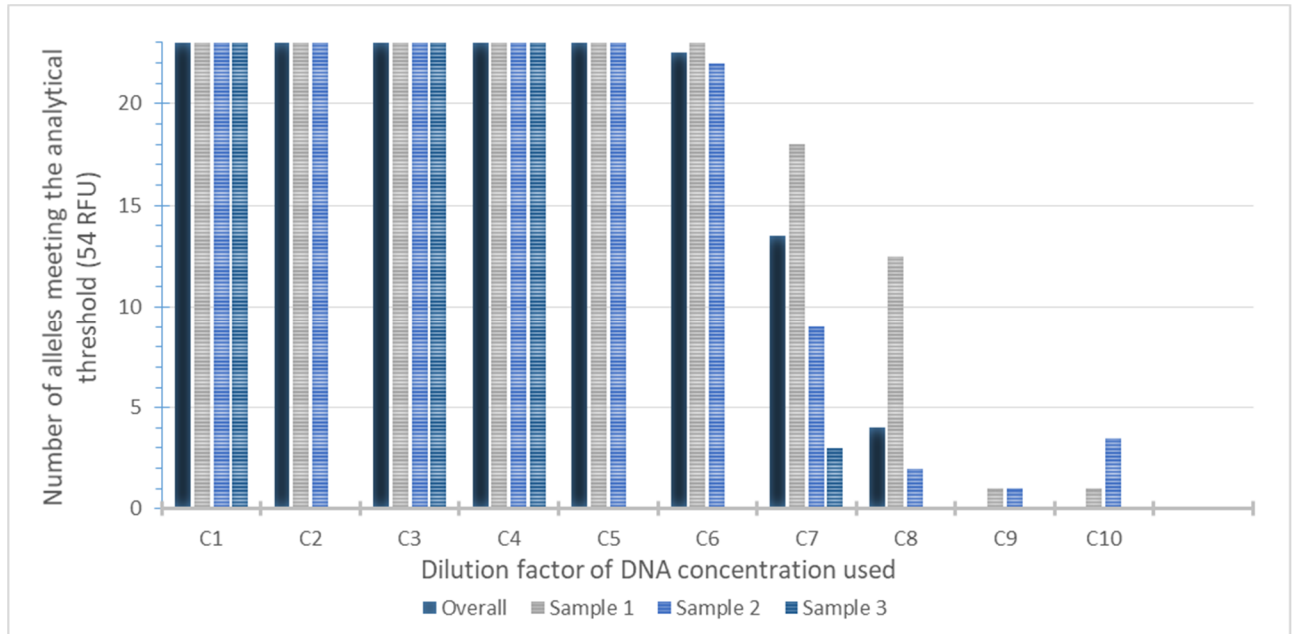
APPENDIX D

Reproducibility results: Two channels from an electropherogram for a single sample processed on different days, where in the allele identities are the same



APPENDIX E

Sensitivity results: showing the number of markers meeting the determined analytical threshold for direct PCR approach (54 RFU) and extracted DNA approach (68 RFU)



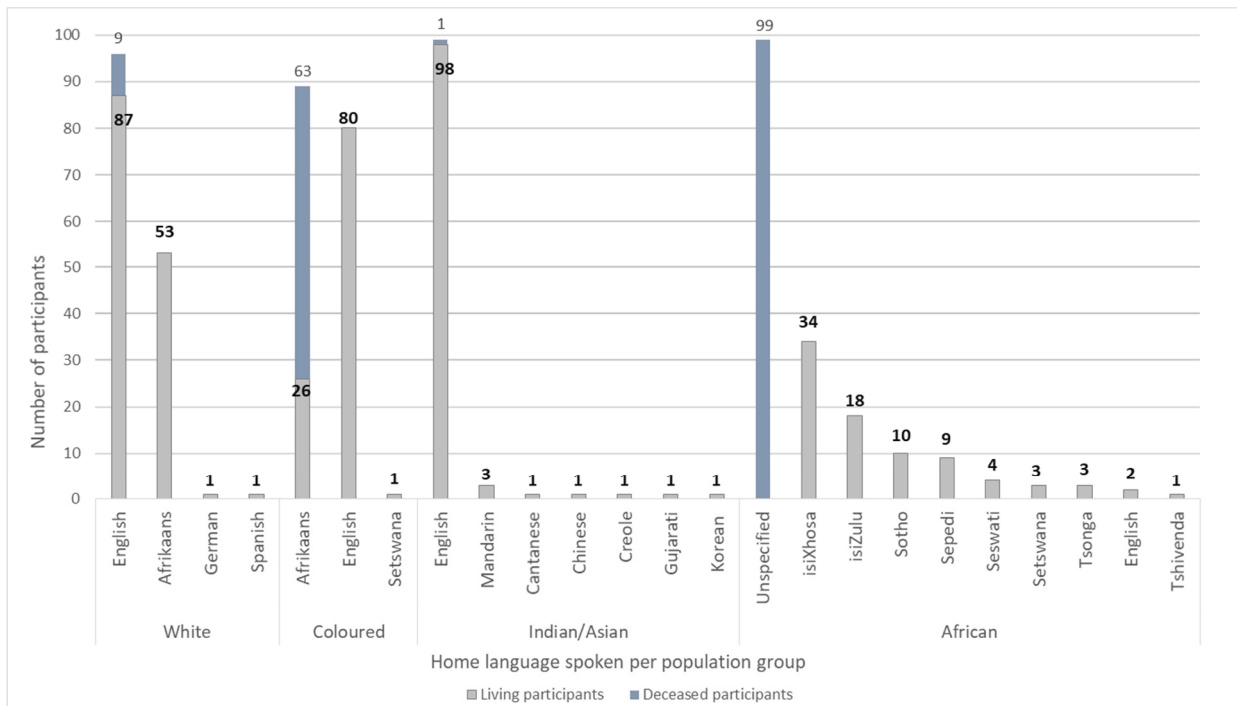
APPENDIX F

Sample numbers per life status, population group and sample type

		Black	Coloured	Indian/ Asian	White	Undet.	Total
Deceased	Promega Swab solution	57	35	0	8	1	101
	Qiagen lysis buffer	0	0	0	0	0	0
	Extracted DNA (blood)	42	28	1	1	0	72
	<i>Total</i>	99	63	1	9	1	173
Living	Promega Swab solution	40	46	71	85	0	242
	Qiagen lysis buffer	44	61	37	55	0	197
	Extracted DNA (mouth wash)	17	5	3	16	0	41
	Extracted DNA (blood)	0	0	0	0	0	0
	<i>Total</i>	101	112	111	156	0	480
TOTAL		200	175	112	165	1	653

APPENDIX G

Home languages reported differed between population groups.



APPENDIX H

Number of observations of the DYS385/b allele combinations across all population groups. DYS385a is indicated on the Y axis and DYS385b on the X axis

Allele	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
8				1												
9									3							
10			1	2		1	7			1						
11				11	4	19	119	29	4			1				
12					4	5	10	5	5	6	3	2	1	1		
13						6	18	6	8	14	8	3	1	1		1
14							8	17	6	7	5	7		4	1	
15								11	24	27	10	4	28		1	1
16									17	24	23	4	15	2		
17										16	26	6	8	3	1	
18											6	1				
19												1				
20													1			

APPENDIX I

q PCR results obtained using the Quantifiler Human and Quantifiler Trio assays

Sample ID	Target	Quantity	Degradation index
MSK_010	Quant Human	18,68	NA
MSK_012	Quant Human	3,84	NA
MSK_019	Quant Human	2,41	NA
MSK_021	Quant Human	25,86	NA
MSK_031	Quant Human	0,64	NA
MSK_037	Quant Human	2,26	NA
MSK_042	Quant Human	17,77	NA
MSK_043	Quant Human	15,40	NA
MSK_044	Quant Human	14,52	NA
MSK_045	Quant Human	7,99	NA
MSK_047	Quant Human	14,81	NA
MSK_048	Quant Human	14,09	NA
MSK_050	Quant Human	9,69	NA
MSK_058	Quant Human	24,15	NA
MSK_059	Quant Human	22,03	NA
MSK_060	Quant Human	5,23	NA
MSK_062	Quant Human	26,93	NA
MSK_071	Quant Human	8,90	NA
MSK_072	Quant Human	8,52	NA
MSK_074	Quant Human	0,40	NA
MSK_078	Quant Human	3,55	NA
MSK_084	Quant Human	9,93	NA
MSK_089	Quant Human	2,72	NA
MSK_090	Quant Human	24,11	NA
MSK_091	Quant Human	6,12	NA
MSK_093	Quant Human	8,88	NA
MSK_094	Quant Human	1,90	NA
MSK_106	Quant Human	7,04	NA
MSK_118	Quant Human	10,36	NA
MSK_123	Quant Human	8,17	NA

Sample ID	Target	Quantity	Degradation Index
MSK_124	Quant Human	3,97	NA
MSK_130	Quant Human	7,90	NA
MSK_132	Quant Human	0,20	NA
MSK_137	Quant Human	1,88	NA
MSK_138	Quant Human	6,85	NA
MSK_145	Quant Human	4,34	NA
MSK_149	Quant Human	8,58	NA
MSK_161	Quant Human	21,77	NA
MSK_171	Quant Human	7,63	NA
MSK_176	Quant Human	13,88	NA
MSK_184	Quant Human	22,20	NA
SDI_001	T.Large Autosomal	311,34	1,00
SDI_002	T.Large Autosomal	2107,23	2,42
SDI_003	T.Large Autosomal	15,63	31,08
SDI_004	T.Large Autosomal	27,46	2,51
SDI_007	T.Large Autosomal	258,77	0,80
SDI_014	T.Large Autosomal	82,49	0,74
SDI_016	T.Large Autosomal	134,82	0,96
SDI_017	T.Large Autosomal	7,73	0,86
SDI_020	T.Large Autosomal	2622,44	0,74
SDI_022	T.Large Autosomal	82,09	1,20
SDI_024	T.Large Autosomal	128,12	0,89
SDI_025	T.Large Autosomal	476,42	0,70
SDI_030	T.Large Autosomal	128,18	1,16
SDI_031	T.Large Autosomal	6,74	1,22
SDI_033	T.Large Autosomal	1082,82	0,88
SDI_036	T.Large Autosomal	391,94	0,68
SDI_037	T.Large Autosomal	154,82	0,81
SDI_038	T.Large Autosomal	1423,13	1,09

Sample ID	Target	Quantity	Degradation Index
SDI_040	T.Large Autosomal	301.79	0.74
SDI_041	T.Large Autosomal	3205.80	1.09
SDI_042	T.Large Autosomal	190.85	0.90
SDI_045	T.Large Autosomal	777.72	0.64
SDI_049	T.Large Autosomal	696.43	0.70
SDI_057	T.Large Autosomal	178.93	1.31
SDI_058	T.Large Autosomal	695.05	0.66
SDI_059	T.Large Autosomal	453.92	0.83
SDI_061	T.Large Autosomal	1806.84	0.94
SDI_063	T.Large Autosomal	1785.07	0.72
SDI_067	T.Large Autosomal	307.51	1.36
SDI_068	T.Large Autosomal	3855.92	0.62
SDI_074	T.Large Autosomal	4.46	0.88
SDI_075	T.Large Autosomal	4889.67	0.54
SDI_078	T.Large Autosomal	2102.07	0.35
SDI_079	T.Large Autosomal	842.89	0.85
SDI_085	T.Large Autosomal	144.22	0.70
SDI_087	T.Large Autosomal	2244.52	0.68
SDI_089	T.Large Autosomal	2996.74	0.65
SDI_093	T.Large Autosomal	1799.84	0.62
SDI_095	T.Large Autosomal	659.61	1.56
SDI_096	T.Large Autosomal	850.08	0.77
SDI_100	T.Large Autosomal	226.88	0.89
SDI_101	T.Large Autosomal	40.95	0.91
SDI_103	T.Large Autosomal	127.64	0.76
SDI_107	T.Large Autosomal	450.25	0.71
SDI_108	T.Large Autosomal	2223.90	0.56
SDI_109	T.Large Autosomal	1060.27	0.66
SDI_113	Quant Human	541.75	NA

Sample ID	Target	Quantity	Degradation Index
SDI_119	Quant Human	1049.45	NA
SDI_120	T.Large Autosomal	370.44	0.87
SDI_122	T.Large Autosomal	92.08	1.12
TXG_05	T.Large Autosomal	13.14	0.82
TXG_07	T.Large Autosomal	27.17	0.71
TXG_08	T.Large Autosomal	27.48	0.82
TXG_10	T.Large Autosomal	9.28	0.65
TXG_11	T.Large Autosomal	32.05	0.90
TXG_13	T.Large Autosomal	12.17	0.89
TXG_14	T.Large Autosomal	19.53	0.79
TXG_15	T.Large Autosomal	22.32	0.77
TXG_16	T.Large Autosomal	53.33	1.07
TXG_17	T.Large Autosomal	8.23	0.57
TXG_18	T.Large Autosomal	38.49	0.81
TXG_19	T.Large Autosomal	28.79	0.96
TXG_20	T.Large Autosomal	40.22	1.15
TXG_21	T.Large Autosomal	25.49	1.06
TXG_22	T.Large Autosomal	16.91	0.87
TXG_25	T.Large Autosomal	5.91	0.88
TXG_26	T.Large Autosomal	31.89	1.10
TXG_27	T.Large Autosomal	31.29	0.97
TXG_28	T.Large Autosomal	67.18	0.89
TXG_29	T.Large Autosomal	14.51	0.80
TXG_31	T.Large Autosomal	10.34	0.65
TXG_33	T.Large Autosomal	12.38	0.74

APPENDIX J

F_{ST} associated P-values obtained for the four population groups analysed.

	African	White	Coloured	Indian/Asian
African				
White	0.02703±0.0139			
Coloured	<0.0005±<0.0005	0.09009±0.0192		
Indian/Asian	<0.0005±<0.0005	0.00901±0.0091	<0.0005±<0.0005	

APPENDIX K

Linkage Disequilibrium scores obtained from SHEsis for the a) African, b) Coloured, c) Indian/Asian, d) White

