CONSTRUCTING A DNA PROFILE FREQUENCY DATABASE FOR SOUTH AFRICA USING THE QIAGEN INVESTIGATOR® 24 PLEX GO! KIT

by

LORRAINE NEL
NLXLOR001

SUBMITTED TO THE UNIVERSITY OF CAPE TOWN
In partial fulfilment of the requirements for the degree
MPhil (Biomedical Forensic Science)

Faculty of Health Sciences
UNIVERSITY OF CAPE TOWN

August 2017

Supervisor: Laura Heathfield

Division of Forensic Medicine and Toxicology

Department of Pathology

University of Cape Town

Declaration

I, Lorraine Nel, hereby declare that the work on which this dissertation/thesis is based is my original work (except where acknowledgements indicate otherwise) and that neither the work nor any part of it has been, is being, or is to be submitted for another degree in this or any other university.

I empower the university to reproduce for the purpose of research either the whole or any portion of the content in any matter whatsoever.

Signed by candidate

Signature

Date: 2017-08-28

Abstract

DNA profiling is routinely used in the forensic setting to identify individuals during criminal and medico-legal investigations. Its principle is based on the molecular analysis of DNA to produce a string of alpha-numeric characters which can be matched to a known reference sample. The use of allele frequencies from the background population aids the statistical interpretation of a match and can be used to calculate the random match probability. In South Africa, allele frequency data for the background population is currently limited, which can hinder the discriminatory value of DNA evidence, particularly when only a partial profile is obtained. Therefore, the aim of this study was to generate DNA allele frequency data for four South African population groups using the QIAGEN Investigator® 24PLEX GO! Kit, which has six markers for which data does not yet exist for the South African population. Full forensic DNA profiles were generated from 655 unrelated individuals from four population groups in South Africa: Black African (n = 172), Coloured (n = 195), Indian/Asian (n = 88) and White (n = 200). A 98% first time success rate was observed using the direct PCR approach. Allele frequencies were significantly different between all four population groups at three markers (D8S1179, D2S1338 and D2S441) after a Bonferroni correction ($\rho < 0.001$) and sixteen novel alleles were observed. Two genetic anomalies were observed, namely tri-allelic patterns at the TPOX marker (n = 9) and a null allele at amelogenin (n = 1). While the sample size for the Indian/Asian population group was limited in this study, the data generated here nevertheless prospects to contribute towards the data currently published for South Africa. This, in turn, will allow for more DNA markers to be analysed during forensic casework in South Africa, as the data for its statistical interpretation is now available.

Acknowledgements

I would like to thank my supervisor, Laura Heathfield, for all her advice, knowledge, support, trust, motivation, time and patience which assisted me during this project.

I would like to thank Maryvonne Abrahams for all the assistance at Unistel.

I would like to thank Kate Reid (UCT) with her assistance with sample collection at UCT and lysing of those and other samples.

I would like to thank Calvin Mole (UCT) with his assistance regarding statistical calculations.

I would like to thank my partner, parents, family, friends, personnel at the division of Forensic Medicine and Toxicology (UCT) and colleagues at the Forensic Science Laboratory (SAPS) for their moral support, motivation and for displaying continuous interest in this project.

I would also like to thank God who carried and guided me through this project.

Table of contents

List of abbreviations

%	Percentage
bp	Base pairs
CODIS	Combined DNA Index System
CPE	Combined probability of exclusion
CPI	Combined probability of inclusion
DNA	Deoxyribose nucleic acid
ENFSI	European Network of Forensic Science Institutes
ESS	European Set of Standard markers
FBI	Federal Bureau of Investigation
ILS	Internal Lane Standard
InDels	Insertion-deletion polymorphisms
ISO	International Organisation for Standardisation
kV	Kilovolts
LR	Likelihood ratio
MSDS	Material Safety Data Sheets
n	number
NDIS	National Data Index System
NFDD	National Forensic DNA Database of South Africa
NIST	National Institute of Standards and Technology
PCR	Polymerase Chain Reaction
PPE	Personal Protective Equipment
QS	Quality Sensor
RFU	Relative Fluorescence Units
RMP	Random Match Probability
SAPS	South African Police Services
SNPs	Single nucleotide polymorphisms
STR	Short tandem repeat
SWGDAM	Scientific Working Group on DNA Analysis methods
UCT	University of Cape Town

List of figures

List of tables

Chapter 1: Introduction

1.1 Background

The National Forensic DNA Database of South Africa (NFDD) was officially established on 31 January 2015 when the Criminal Law (Forensic Procedures) Amendment Act, Act No. 37 of 2013 became operational [1]. This breakthrough in South African legislation has allowed for the regulation of deoxyribose nucleic acid (DNA) evidence to be used in South Africa. This law permits the use of forensic DNA profiles in the investigation of crime and provides for different categories of forensic DNA profiles to be collected and stored; each with their own set of stipulations, and is compulsory for arrestees and convicted offenders of Schedule 8 offences. The NFDD is anticipated to grow exponentially and the need will arise for more informative forensic DNA profiles for which allele frequency data needs to be generated.

A forensic DNA profile results from the analysis of a DNA molecule using well-established genetic techniques. It consists of a selection of short tandem repeat (STR) loci which are polymorphic regions that are wide spread throughout the human genome [2]. An STR is a microsatellite marker with short repeat units, which can be highly variable between unrelated individuals and multiple alleles are possible for each locus [2]. These characteristics make STRs an appropriate marker for crime investigations. Upon forensic analysis of DNA, STR loci are described using alpha numerical characters with the result being a forensic DNA profile [2, 3]. Regulatory scientific bodies ensure that this science can be applied to forensic investigations and paternity testing by regularly reporting on forensic genetic standards and recommendations [4 - 8].

These alpha-numeric characters are compared to a database containing DNA profiles of known individuals, as well as other unidentified profiles from crime scenes. If a match occurs, the probability of that match is calculated using background frequencies of alleles from that specific population group. In 1995, the United Kingdom was the first country to use a DNA database. Their crime detection rate increased from 26% to 40% after DNA samples were loaded in their DNA database [9]. In January 2016, more than 15 million DNA profiles were included in the National DNA Index System (NDIS) in the United States of America, and by January 2017, it was suggested to include more STR loci to reduce the likelihood of incorrect

matches in large databases and to increase the international compatibility and discrimination power [10, 11]. The expansion of forensic DNA profiling kits to include more markers occurred alongside, which enable the analysis of more markers [12]. In turn, this improved discriminatory power of DNA evidence in forensic casework [13].

Currently, the Forensic Science Laboratory in South Africa utilises 15 STR loci and the amelogenin sex marker to calculate the probability of a DNA profile match [14]. The analysis of more markers in the local context is similarly motivated by the potential increase of statistical probability; however, this is currently hindered due to the absence of published allele frequencies for more markers for the South African population groups.

1.2. Forensic DNA profiling

1.2.1 Principles

Polymerase chain reaction (PCR) is used in the DNA analysis of forensic evidence to generate forensic DNA profiles. PCR multiplex chemistry allows for the simultaneous amplification of multiple regions of a genome (STRs). The resulting profile consists of alleles, one or two per STR marker, indicating the number of repeat motifs within the STR [2].

The samples become labelled with fluorescent dyes during amplification, and subsequently separated and detected via analytical software during capillary electrophoresis. The software deciphers the fluorescent intensity data into electropherograms and then labels the peaks using an allelic ladder as reference. The length, measured in base pairs (bp), is on the x-axis and the peak height, measured in relative fluorescent units (RFU), is on the y-axis of the electropherogram [2, 6]. Template 'panel' and 'bin' files specific to the chemistry used can be imported into the software to aid interpretation, in accordance with pre-defined quality settings. Accuracy is further ensured by an internal lane standard (ILS), allelic ladder, positive control and no template control [2, 6].

Microvariants, and potentially novel alleles, arise when there is a variation within a STR repeat motif, resulting in an allele with one incomplete motif; these are typically observed as peaks which do not correspond to the peaks of the allelic ladder. Many microvariants have been reported in South Africa and other populations [15 - 17].

1.2.2 Anomalies

As early as 1997, guidelines were published for creating forensic DNA profiles, including designation of alleles using a STR multiplex system [4, 7], validation for forensic applications [5, 18, 19], and interpretation of DNA profiles [4]. However, several factors influence the complexity of forensic DNA profile interpretation, including the diversity of sample types and the potential presence of artefacts in the electropherograms [19 - 21]. Specific examples include mixtures and partial profiles, which are commonly encountered in forensic casework, as well as instrumental or biological anomalies. Split peaks, stutter peaks, allelic drop-in and drop-out are anomalies created during PCR, whereas spikes and pull-ups are created during capillary electrophoresis. True biological anomalies include null alleles and tri-allelic patterns.

A null allele is when one allele of a heterozygote marker fails to amplify due to genetic variation in the primer binding region [2]. This can lead to erroneous interpretations, especially if a null allele is observed in the sex marker, amelogenin (Figure 1.1). Lane *et al.* (2008) reported the complication in paternity testing due to null alleles in South African individuals where the occurrence of null alleles in three STR markers (D13S317, vWA and TPOX) were observed [22].

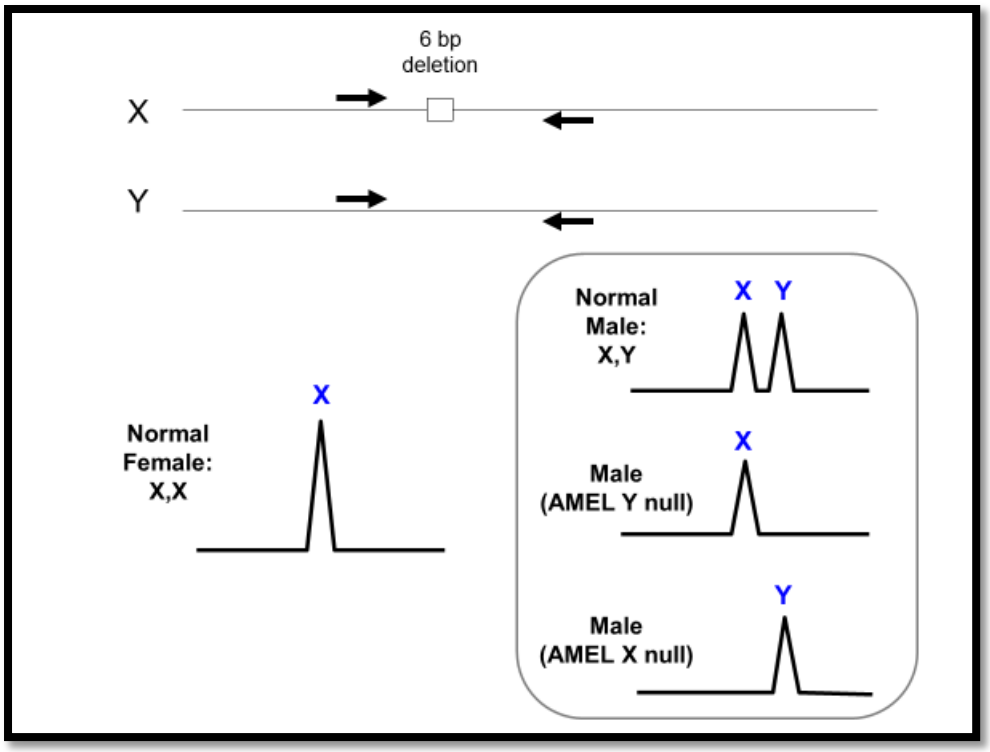


Figure 1.1 Amelogenin null allele

Amelogenin is used as a sex marker as there is a 6 bp deletion on the X-chromosome; the X-chromosome have 106 bp and the Y-chromosome 112 bp [23]. Two peaks in this locus indicates a male profile whereas one peak indicates a female profile. If a null allele is observed in amelogenin an unknown sample can be incorrectly be identified as a female profile as indicated above and both have been reported before [Figure obtained from reference 24]

Tri-allelic patterns are when there are three alleles from a single contributor at a specific STR marker. Tri-alleles can be classified as type one or type two, based on the sum of the peak heights as illustrated (Figure 1.2) [25]. Lane (2008) and Ristow *et al.* (2016) reported the occurrence of tri-alleles in the South African population. Both of these studies observed this genetic anomaly at the TPOX STR marker whereas Ristow *et al.* (2016) also observed this anomaly at the D1S1656 and vWA markers [16, 26]. In Brazil Piçanto *et al.* (2016) investigated the tri-allele at the TPOX marker in 105 tri-allelic individuals; their results were in agreement with Lane (2008) and Ristow *et al.* (2016) that allele 10 was always present [16, 26, 27]. Allele 10, which is actually present on the X-chromosome, is therefore considered to be the additional allele.

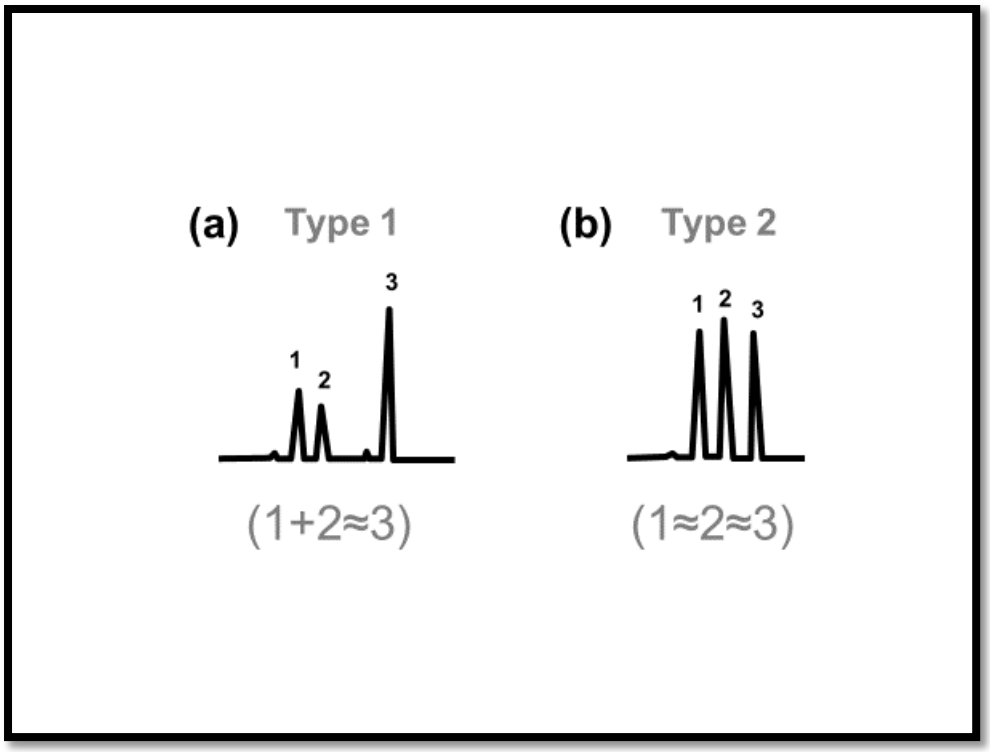


Figure 1.2 Tri-allelic pattern classification

Type one is when the sum of the peak height of the two smaller alleles are ±equal to the peak height of the third allele. Type two is when all three alleles have similar heights [Figure is obtained from reference 24].

1.2.3 Chemistries and commercial kits

Numerous commercial kits have been developed to streamline forensic DNA profiling across laboratories and even countries. Different laboratories have adopted different chemistries according to their needs, legislation and caseload.

One of the more recent kits developed was the QIAGEN Investigator® 24 Plex GO! Kit (Hilden, Germany), which follows the Scientific Working Group on DNA Analysis Methods (SWGDAM) and European Network of Forensic Science Institutes (ENFSI) guidelines. It is a multiplex STR system that amplifies 21 autosomal STR markers (TH01, D3S1358, vWA, D21S11, TPOX, D1S1656, D12S391, SE33, D10S1248, D22S1045, D19S433, D8S1179, D2S1338, D2S441, D18S51, FGA, D16S539, CSF1PO, D13S317, D5S818, D7S820) and two sex STR markers (amelogenin and DYS391) [28]. Of these markers, SE33 has been reported to be a particularly polymorphic STR marker [29].

In addition, this chemistry contains a Quality Sensor (QS) which is an internal PCR control which determines PCR efficiency and enables the forensic analyst to discern the reasons for sub-optimal results and differentiate between failed PCR amplification, no DNA present, degraded DNA and PCR inhibition [30]. The chemistry utilises a six-dye fluorescence channel with BT6 matrix with decreased spectral overlapping, and is available for "direct PCR' as well as conventional PCR on extracted samples.

While the analysis of autosomal STRs is perhaps the most common in forensic laboratories currently, other commercial kits are also available which include autosomal single nucleotide polymorphisms (SNPs), autosomal insertion-deletion polymorphisms (InDels), X-Chromosome STRs and Y-Chromosome STRs. Several next generation sequencing kits have combined all these markers into single reactions, enabling advanced analysis of severely degraded or compromised samples [29].

1.3. Data interpretation and Databases

1.3.1 Data interpretation

For DNA profile interpretation to occur, thresholds need to first be determined to ensure reliability and accuracy of data. The analytical threshold is the minimum height (RFU) at which detected peaks can be differentiated from background noise [6] and the stochastic threshold is the RFU value above which it is reasonable to assume that allelic drop-out of a sister allele has not occurred within a single source sample [6, 31]. By applying these thresholds, alleles at each STR marker can be designated and analysed. The alleles pertaining to a DNA profile are then added to a national DNA database to search for matches against DNA profiles from known individuals or other samples from crime scenes; rendering the database an intelligence tool in forensic investigations.

Should a match occur between the forensic DNA profile of interest and a reference sample, a calculation is performed to determine the random match probability (RMP) and/or likelihood ratio (LR) for that match. In cases of mixed DNA profiles, the combined probability of exclusion/inclusion (CPE/CPI) is calculated instead. In order to calculate these statistics, the frequency of each allele (and genotype) occurring in the forensic DNA profile of interest needs

to be determined and then multiplied together to obtain the DNA profile frequency. The inverse of the DNA profile frequency give rise to the RMP or LR.

These calculations are based on three major factors: (i) the presence of allele frequencies of the background population, in order to determine genotype frequency; (ii) Hardy Weinberg equilibrium, which evaluates allele independence within a genetic locus; and (iii) linkage equilibrium, which evaluates locus independence of alleles between loci. Since the loci in forensic DNA profiles are indeed independent, the product rule is used in calculations. Only the upper match probability is reported which leads to the most conservative and objective DNA profile frequency [6, 32]. These statistics then lead to the final conclusion of inclusion, exclusion or inconclusive match between a reference (known) and a crime scene (unknown) sample.

1.3.2 Population databases

Forensic-related population studies were performed to establish allele frequencies within a population group and to understand the structure of population groups. Many studies have shown that allele frequencies of markers within DNA profiling kits were different between population groups, and that relatively minor differences were observed between individuals from the same population, but who lived in different geographical locations [16, 33 - 35]. An allele frequency 'database' generated from a population study is not an intelligence tool as described previously; rather the population dataset is used to facilitate the calculation of a RMP.

The current allele frequency database used for South African forensic casework and some paternity testing consists of 15 STR markers (D3S1358, FGA, vWA, D18S51, D21S11, D8S1179, D13S317, D5S818, D7S820, CSF1PO, D19S433, TPOX, D16S539, D2S1338 and TH01) and amelogenin as a sex marker. This data was generated using the AmpFℓSTR® Identifiler Plus™ PCR amplification kit (Applied Biosystems, Foster City, CA) on four major population groups in South Africa [14]. Similar studies were performed on African populations including, but not limited to individuals from Equatorial Guinea, Mozambique, Namibië (Ovambo population) as well as Africans and Europeans from South Africa [36 - 39]. Buckleton *et al.* (2016) observed that populations with African ancestry were more diverse than other populations in the world [40].

South Africa went through major changes since their first census in 1865 [41]. The geographical position and complex history of South Africa has led to the unique admixed population, known locally as the Coloured population [42]. This population group exhibits mixed ancestry from the first European settlers, the indigenous KhoiKhoi, San and African individuals as well as immigrant Asian populations [42 - 45]. The South African Black population group is from African descent; the White population group is from European descent while the Indian/Asian population group is from the South Asian, Indian subcontinent [46].

As per the census of 2016 South Africa has a population of 55 653 654 individuals which consists of four major population groups; Black African (80.66%), Coloured (8.75%), White (8.12%) and Indian/Asian (2.47%) [47]. There are eleven official languages in South Africa where some individuals can speak more than one language. The Black African group comprises nine different cultural groups, each with their own language. Lane *et al.* (2002) reported on the substructure within the South African Black population groups and showed that there was relatively little STR allele frequency variation between the sub-population groups [48].

In 2014 Lucassen *et al.* reported some of the first population data for South African population groups, which is the population data utilised by the Forensic Science Laboratory at the South African Police Services currently [14]. A more recent study by Ristow *et al.* (2016) evaluated the GlobalFiler® Express kit (PROMEGA, Madison, Wisconsin, US) on the Afrikaner, amaXhosa, amaZulu, Asian/Indian and Coloured population/ethnic groups [14, 15]. The latter study showed significant differences between African and non-African population groups in terms of allele frequencies, but no significant differences between the amaXhosa and amaZulu population groups. Ristow *et al.* (2016) observed substructure in the Coloured population group, which aligned with similar findings from Lucassen (2014) [14].

LongLi *et al.* (2007) reported on the Chinese population and showed that there were differences in allele frequencies between different areas of China. Similarly, Šimková *et al.* (2016) reported significant differences in allele frequencies between Czech Republic population groups [49, 50]. Fujihara (2007) reported the Namibian Ovambo population has significantly different allele frequencies compared to African-Americans, Equatorial Guinea, Guinea Bissau,

Mozambique and South African Black Africans [38]. Amir et al. (2016) however, showed that there was no significance in allele frequencies between two North African populations when grouped into their respective linguistic and geographical locations. They also reported that 40% (56/140) of all the alleles in this population had a frequency of less than 0.05 [51].

A recent correspondence by Chishti *et al.* (2016) also reported about the relationship between allele frequencies of DNA profiles and ethnic groups within the Pakistan population. This study showed deviations from Hardy-Weinberg equilibrium which were attributed to inbreeding and consanguinity within the population. The author also concluded that while STRs were sufficient for individual discrimination, they should not be used in reverse to predict ethnicity for a sample within unknown origin [52]. The same recommendation was made by Ristow *et al.* (2016) for the South African population [16].

1.3.3 Legislation

In the United States of America, the Federal Bureau of Investigation (FBI) use the Combined DNA Index System (CODIS) to assist data interpretation, and in 1997, it was stipulated that 13 core loci must be present in every forensic DNA analysis in order to facilitate comparison over time as kits developed [2]. A similar notion was implemented in Europe in 2006, whereby the European Set of Standard markers (ESS), comprising seven markers, was proposed for similar reasons, and in 2009 five additional markers were added [32].

Before 31 January 2015, South African law did not have a database that regulated DNA evidence, which meant that DNA evidence could only be used for the case it "belonged" to which limited the value of the DNA evidence. Now, with the new legislation, South Africa can use DNA evidence and the NFDD as an intelligence tool and hopefully link crimes, similar to that of Brazil. The new law does not stipulate core loci that need to be used for forensic profiling. The law does however regulate that DNA profiles must be stored in different indices which are: convicted offender, arrestee, crime scene investigations, elimination, and missing persons and unidentified human remains. As mentioned before, the collection, storage and use of these samples/profiles are regulated for each index respectively, in order to ensure protection of human rights. The implementation of this new law is managed by the National Forensics Oversight Ethics Board [1].

1.3.4 Ethical considerations of databases

The fact that individuals can be identified with their DNA profile, and that DNA profiles are stored on a database, gives rise to ethical concerns. Wallace *et al.* (2014) described the importance of protecting human rights, where they attempted the establishment of a world-wide set of standards for human rights in respect of DNA databases [53]. Human rights need to be protected at all times; some of these standards include the destruction of biological samples, deletion of innocent individuals' profiles, using the best practice of scientific standards to obtain the results, use of DNA in court, collection of DNA, as well as access to, and protection of, forensic DNA profiles [54].

Individuals may be concerned that the police services might use genetic information to reveal aspects about them which is private or unknown; for example, a medical disposition or phenotypical associations. Furthermore, this genetic information is stored on a database which is searchable and individuals do not know if the confidentiality of this information is maintained; anyone might infiltrate the database.

While identification is based on a matching principle and not what an individual looks like [55], the public still have fears about DNA, and perhaps misunderstandings about how the identification through DNA is achieved. While STR markers may have links to specific diseases (e.g. tri-allelic pattern at marker D21S11 may suggest Down syndrome), policies and laws exist in many countries to protect the handling and storage of this information [1, 56].

Fortunately, for South Africa, many of these concerns have been addressed in the new legislation. Katsanis *et al.* (2013) clarified the genomic significance of the key identification markers, which are intended to identify individuals based on a matching principle and not through phenotypical information, except for sex [55].

Another ethical consideration is the use of an alleged innocent family member's DNA to obtain information regarding a forensic investigation; not all countries regulates familial searching [9, 54, 57]. If a body beyond recognition is found and the individuals' identity can only be confirmed by familial searching, this can assist the forensic services and the family can get closure from the results. On the contrary, if a homicide was committed and unknown DNA is found on the scene and the forensic services use familial searching to try and identify the suspect; the privacy of the family members can be compromised. Some family members might

also not be comfortable to be linked to a police investigation. Genetic linkage between individuals or the absence of it resulting from familial searches can also be devastating.

1.4 Rationale

Since the Criminal Law (Forensic Procedures) Amendment Act, Act No. 37 of 2013 become operational, the National Forensic DNA Database of South Africa is anticipated to grow exponentially. As such, more STR markers will be needed to obtain higher statistical probabilities in court. The integrity of DNA from a crime scene cannot be controlled and the quality and/or quantity might be compromised, leading to a partial profile. However, if more STR markers are analysed, then even a partial profile may give rise to a high random match probability.

Leading companies in the field such as QIAGEN, Thermo Fischer Scientific and Promega have all released additional kits for forensic DNA profiling, which comprise more than the 15 autosomal STR markers which are currently being analysed in South Africa. These later kits include 22 - 27 markers and therefore have potentially increased discriminatory value. In particular, the Investigator® 24plex GO! kit (QIAGEN, Hilden, Germany) also includes a quality sensor, which indicates if a sample is degraded, has low copy number or contains PCR inhibitors. However, such kits cannot be utilised to their full potential in South Africa until the allele frequencies of the markers in such kits have been established.

1.5 Aim and Objectives

The aim of the study was to create a forensic DNA profile frequency database for South Africa using the QIAGEN Investigator® 24plex GO! kit (Hilden, Germany). To reach this goal there were four major objectives namely to:

- recruit 100 - 200 individuals from each of the four major population groups in South Africa and to collect their buccal swabs and demographic information;
- optimise the direct PCR workflow and set up quality thresholds;
- subject the buccal swabs to the optimised workflow to generate forensic DNA profiles for each participant
- calculate the allele frequencies for each marker for each population group and evaluate the data statistical.

Chapter 2: Materials and methods

2.1 Design and setting

In order to broaden the forensic applicability of the QIAGEN Investigator® 24 PLEX GO! (Hilden, Germany) kit to South Africa, a new database of allele frequencies for South African populations needed to be created. A prospective, quantitative, cross-sectional study, situated in South Africa, was therefore designed to fulfil this aim. The diagram in figure 2.1 summarises the multi-step workflow used to construct the DNA profile allele frequency database for four major population groups in South Africa.

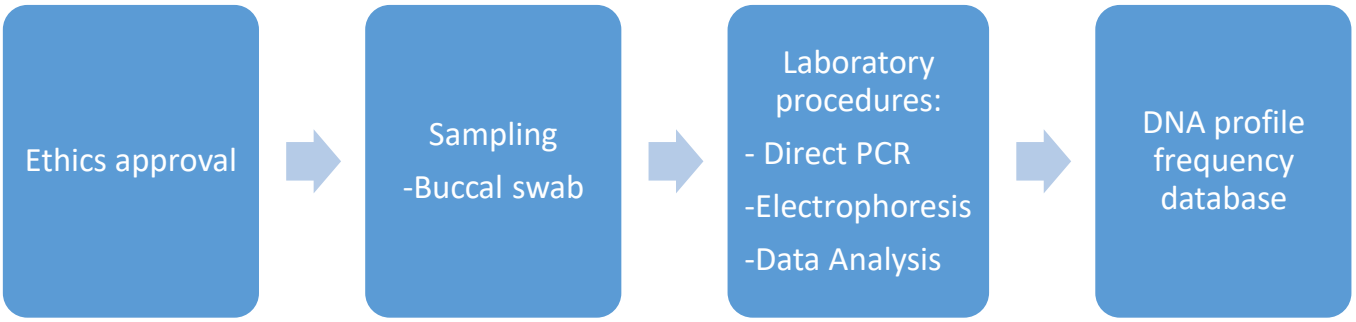


Figure 2.1 Diagram to demonstrate study flow

Ethics approval was obtained before samples could be collected. Buccal swabs were used to collect (DNA) from the four main population groups of South Africa. Several laboratory procedures were applied to the samples to obtain forensic DNA profiles using the QIAGEN Investigator® 24plex GO! Kit. A DNA profile allele frequency database for South Africa was created after the forensic DNA profiles were analysed.

2.2 Ethical considerations

This research project was approved by the Faculty of Health Science Human Research Ethics Committee of the University of Cape Town (HREC REF: 342/2016) (Appendix 2.1). Participation was voluntary and informed consent from all the participants was obtained before sample collection.

While there was little direct benefit of participating in the study, there was also minimal risk to the participants as samples were obtained non-invasively, and the data generated could not

reveal personal information about them, except for their sex. Further, all participants' information was kept private and confidential and only authorised personnel had access to the raw data. The participants' personal information was kept separate from their DNA sample to further ensure that the identity of the participants was protected.

2.3 Study populations

2.3.1 Population sampling

To create a DNA profile allele frequency database at least 100 individuals (200 alleles) from a specific population is required which is the lower threshold to capture most common variation [58]. Sampling was carried out to be representative of the population and demographics of South Africa. Due to some population groups in South Africa exhibiting mixed ancestry, an effort was made to recruit more than 100 (100 – 200) individuals from each of the four major South African population groups [47]. The four major population groups were: Black African, Coloured, Indian/Asian and White; and were selected based on the census groups in South Africa, as well as on previous studies in order to facilitate comparison [14, 47].

2.3.2 Participant selection and recruitment

Participants were recruited at various sites in South Africa, including academic and forensic related centres in the Western Cape, Gauteng and Kwa-Zulu Natal. While recruitment was limited to these provinces due to logistical reasons, the participants were representative of other parts of South Africa too, as some participants were only temporarily living in a province for study and/or employment purposes, and were originally from other provinces in the country.

Participants were healthy individuals over the age of eighteen and every effort was made to recruit unrelated individuals. Patients who received organ transplants were excluded from the study.

Participants were approached orally and/or electronically and information regarding the project was given (Appendix 2.2). Time was allocated for questions and/or concerns from participants prior to collection. The collection took place at a pre-arranged time and venue, which was private and convenient to the participants. After informed consent was obtained,

the participants completed a short questionnaire regarding their demographics, sex and self-reported population group (Appendix 2.2).

A biological sample in the form of a buccal swab sample was then obtained from each participant, subsequently to explaining the health and safety risks. This method of collection was quick and non-invasive and was performed by the participant under the supervision of the researchers. A sterile cotton-tipped swab with a wooden shaft (COPAN, Italia) was used to collect buccal cells from the inside of the participant's cheek. The sterile swab adhered to quality standards (Appendix 2.3). The swab was swiped five times on the one side of the inner cheek, as previously described by Jovanovich *et al.* (2015) [59].

The sample collection method for the buccal swabs was first optimised before collection from the participants. Six different approaches were considered: swabs were collected and lysed with and without being air-dried; swabs were collected and kept in the tube for two hours at room temperature and alternatively in a polystyrene cooler box with ice, again with and without being air-dried. The most suitable approach was to place the swabs in a polystyrene cooler box containing ice blocks immediately after collection and to add the lysis buffer to the swab within two hours. While air-drying of the swab completely before downstream processing is generally recommended, it was observed here that the absence of air-drying did not affect the quality of results if lysis buffer was added within two hours [60]. It was noted that it was better to keep the swabs in a cooler box after collection, rather than at room temperature. This optimised method was then adopted as the collection procedure. A total of 428 DNA profiles were obtained from this collection method. When samples were not collected at the University of Cape Town (UCT), the samples were transported to UCT for further processing.

In addition to buccal swabs being obtained from participants, extracted DNA from other participants (n = 227) was also available for this study. Such DNA were samples from similar previous and/or current studies within the Division of Forensic Medicine and Toxicology, from participants who gave explicit consent that their DNA and information may be used in this research study, in addition to their primary research study. These extracted DNA also originated from buccal cells, but in the form of a saliva mouth-rinse sample. DNA was extracted from saliva using the Quick gDNA™ MiniPrep kit (Zymo Research; Irvine, CA, USA). These samples were quantified using the NanoDrop® 2000 spectrophotometer (Thermo Fisher Scientific; Wilmigton, Delaware, USA) and diluted to a final concentration of 5 ng/µl

using AccuGene® molecular biology grade water (Lonza, Basel, Switzerland) which adhered to quality standards (Appendix 2.4).

2.4 Laboratory procedures

2.4.1 Pre-Polymerase Chain Reaction (PCR) sample preparation

The swab containing the buccal cells was broken off into a clear, Click-Fit 1.5 ml micro centrifuge tube (Argos technologies, Elgin, IL). A volume of 500 µl of Investigator® STR GO! Lysis buffer (QIAGEN, Hilden, Germany) was added to each micro centrifuge tube containing a swab/sample and tubes were then incubated at 95 degrees Celsius (ºC) while shaken at 1200 revolutions per minute (rpm) for 5 minutes in the Eppendorf® Thermomixer compact thermal shaker (Sigma-Aldrich, Cleveland, OH). The samples were left to reach room temperature and then stored at -20ºC until the next step. Samples were transported on ice and processed further by the researcher at Unistel Medical Laboratory; this laboratory had the necessary instrumentation for DNA profiling and adheres to the International Organisation for Standardisation (ISO) 17025 requirements for technical laboratories.

2.4.2 Direct Polymerase Chain Reaction (PCR)

The QIAGEN Investigator® 24plex GO! Kit was a chemistry which aligned with the SWGDAM and ENFSI guidelines (Appendices 2.5 - 2.8) [28]. Direct amplification was performed on the buccal swabs, which correctly implies that no DNA extraction or quantification was performed on these samples.

Direct PCR was prepared in final volume of 22 µl, as per the manufacturer's protocol, with no deviations [61]. Cycling conditions for buccal swab lysates was according to the manufacturer's protocol; however, the cycling protocol for the extracted DNA samples had 27 cycles instead of 24 cycles. Amplification was performed on the 9700 ProFlex™ 96-well PCR system (Applied Biosystems, Foster City, CA) which were calibrated regularly (Appendix 2.9).

A positive (Ref: 9948) and negative control were included in the amplification process. The positive control was commercially produced and included in the kit where an extraction blank

was used for the negative control for the extracted samples and a clean swab in lysis buffer were used as the negative control for the lysates.

2.4.3 Capillary electrophoresis

The amplified PCR products were separated and detected through capillary electrophoresis on the 8-capillary 3500 Genetic Analyser with POP-4 polymer (Applied Biosystems, Foster City, CA). This instrument has been validated by the National Institute of Standards and Technology (NIST) [8] and local services were performed regularly (Appendix 2.10). A six-dye fluorescence channel with a BT6 matrix (Matrix Standard BT6) was used to accommodate the visualisation of the number of loci amplified during PCR. The six fluorescence labels were 6-FAM (Blue), BTG (Green), BTY (Yellow), BTR2 (Red), BTP (Purple) and BTO (Orange). The instrument setup was performed per the manufacturer's instructions before the samples were processed. This included the spectral calibration, instrument protocol setup, size standard setup, quality control protocol setup and the optimised assay setup [61].

A volume of 12 µl of Hi-Di™ Formamide (Applied Biosystems, Foster City, CA) was added to 0.5 µl of DNA Size Standard (BTO) for each reaction and 12 µl of this was added to each appropriate well, together with 1 µl of amplified PCR product. Samples were then denatured for 3 minutes at 95ºC, followed by a rapid snap-freeze step on ice for 3 minutes. The 96-well microtiter plate with the samples was then loaded onto the Genetic Analyser for capillary electrophoresis, per the parameters optimised during the 'instrument protocol' setup. The injection time for electrophoresis was optimised to be 15 seconds for the extracted samples and 10 seconds for the buccal lysates. The samples were injected at 1.2 Kilovolts (kV) and analysed at a run voltage of 13 kV. The allelic ladder and ILS were included in all electrophoresis batches; one allelic ladder per batch and ILS for each sample.

2.4.4 Data analysis

2.4.4.1 Software

Data were initially analysed using the GeneMarker® HID version 2.8.2 (35-day free trial) software (Softgenetics, Walla Walla, WA), and subsequently analysed with GeneMapper® ID-X software version 1.5 (Applied Biosystems, Foster City, CA). The Investigator® template files for the software were downloaded from the QIAGEN website per the instructions of the

software setup of dye set BT6 in the QIAGEN Investigator® 24 PLEX GO! Handbook [61]. Both software systems have been reported to be accepted in the forensic community [62, 63].

The alleles were automatically called using the allelic ladder, ILS and calculated thresholds (see section 2.4.4.2). All results were checked manually and microvariants were manually calculated.

2.4.4.2 Analytical and stochastic threshold

The analytical and stochastic thresholds were determined before analysis of samples was carried out. Thresholds can be calculated per dye (dye-specific thresholds) or as a single threshold. For this project, a single threshold for each sample type were calculated [6].

This was imperative, since thresholds may differ for each instrument, chemistry, laboratory and sample type used. The analytical threshold was defined as the minimum height in relative fluorescent units (RFU) at which detected peaks can be differentiated from background noise [6]. This was established by subjecting the no template control lysis buffer to the direct PCR and capillary electrophoresis workflow, for a total of five times. The analytical threshold was then calculated as the highest detected peak from these five electropherograms, multiplied by two.

The stochastic threshold was defined as the RFU value above which it is reasonable to assume that allelic drop-out of a sister allele has not occurred within a single source sample [6, 31]. The stochastic threshold for the extracted samples was calculated from a dilution series from 1 ng/µl - 16 pg/µl. The stochastic threshold for the lysates was calculated as a dilution series from 1:1 - 1:32. The dilution series underwent capillary electrophoresis in triplicate. The average peak height and standard deviation of the observed false homozygotes were calculated and the stochastic threshold was calculated as the average peak height plus three standard deviations (See appendix 2.11) [31].

2.4.5 Statistics

Allele frequencies for all population groups for each marker were calculated using Microsoft Excel 2010. The chi-square test was used to establish if the population was in Hardy Weinberg equilibrium. If the ρ-value was less than 0.05, then that marker for that specific population group was considered to not be in Hardy-Weinberg equilibrium.

Chi-square tests were applied to investigate any significant differences in allele frequencies between population groups. One of the limitations using the chi-square test in this dataset was the possibility of false positives due to a small number of observations of each allele within each marker. To be conservative, alleles with an expected frequency of less than five were excluded before the calculations were performed. Chi-square tests were also performed to determine if there were any significant differences in allele frequencies between sex as well as between geographical locations.

Due to multiple testing between pairwise population groups, the Bonferroni correction was applied to the ρ-value, therefore $\rho < 0.0083$ was considered significantly different [64]. Lastly, the allele frequencies of Lucassen *et al.* (2014) [14] were changed to counts and then compared with the counts in this study. STATA software (Statistica 13.2) was used to perform chi-square tests [65].

2.5 Health, safety and quality control

The work benches were always cleaned before and after use. The appropriate personal protective equipment (PPE), including a laboratory coat, gloves, face mask and hair net was always used. The appropriate pipettes and aerosol-resistant pipette tips were used to prevent cross contamination. Work was performed in dedicated areas for pre- and post-amplification, respectively. All appropriate Material Safety Data Sheets (MSDS) of relevant reagents were read before use. All the instruments and equipment used were calibrated and serviced according to manufacturer's requirements.

Chapter 3: Results

3.1 Participants

Forensic DNA profiling was performed on DNA samples from unrelated South African participants and in total, full forensic DNA profiles were obtained from 655 individuals, yielding an overall success rate of 99.2%. Of the successful forensic DNA profiles, the number of participants per population group was: 172 Black African, 195 Coloured, 88 Indian/Asian and 200 White (Figure 3.1).

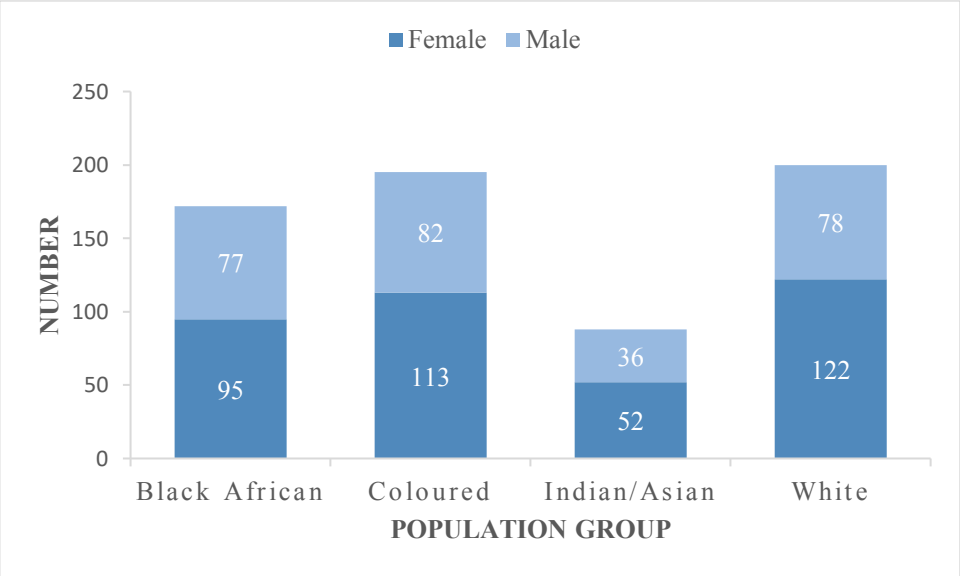


Figure 3.1 Numbers of participants per population group included in this study.

In total, forensic DNA profiles were obtained from 382 (58%) females and 273 (42%) males; 172 (26%) Black African, 195 (30%) Coloured, 88 (13%) Indian/Asian and 200 (31%) White volunteers.

Participants represented all nine provinces in South Africa, however, 47% of participants were from the Western Cape and 30% of participants' hometown information was unreported (Figure 3.2). Participants reported their home language to be English (n = 238), Afrikaans (n = 120), both Afrikaans and English (n = 3), isiXhosa (n = 35), isiZulu (n = 16), Sesotho (n = 9), Setswana (n = 6), Sepedi (n = 5), isiNdebele (n = 1), Tsonga (n = 1), Creole (n = 1), Mandarin (n = 1) and Spanish (n = 1), while 218 participants' home language were not available.

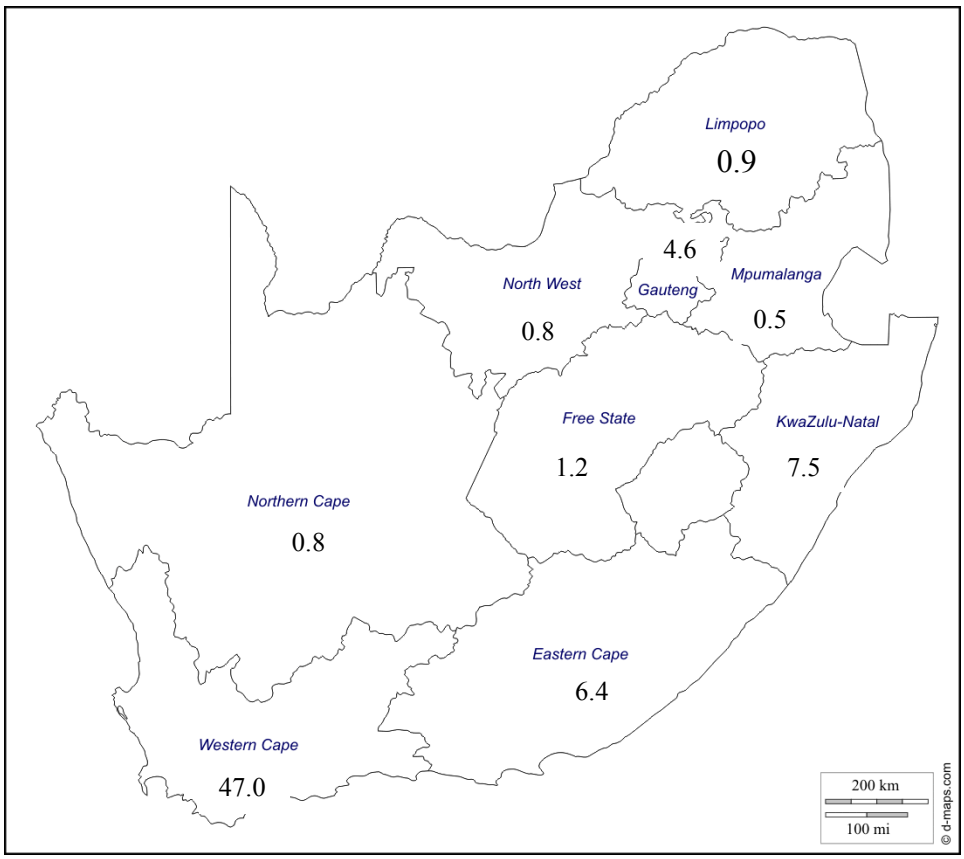


Figure 3.2: Percentages of individuals per province in this study

The chart demonstrates the percentages of the individuals from each of the nine provinces in South Africa. This information was not available for 30% of the participants.

3.2 Internal validation and optimisation of workflow

3.2.1 Analytical and stochastic thresholds

As part of the internal validation process the analytical and stochastic thresholds for each sample type (extracted DNA and buccal lysates) were determined before the samples were processed. The analytical threshold was established as 246 RFU and 102 RFU for the extracted DNA samples and the buccal lysates respectively. The stochastic threshold was established as 291 RFU and 200 RFU for the extracted DNA samples and buccal lysates respectively.

3.2.2 PCR and capillary electrophoresis optimisation

During the establishment of the analytical and stochastic thresholds, it was noted that the PCR cycle number of 27 were too high for the buccal lysates as split peaks were observed at multiple

markers (Figure 3.3.B). The PCR cycle number was therefore decreased to 24 PCR cycles for buccal lysates, which was in accordance with the recommendations of the manufacturer's instructions (Figure 3.3.C) [61]. The injection time for capillary electrophoresis was also adjusted from 15 seconds for the extracted DNA samples to 10 seconds for the buccal lysates. Figure 3.3 demonstrates the impact on the profile quality through optimization.

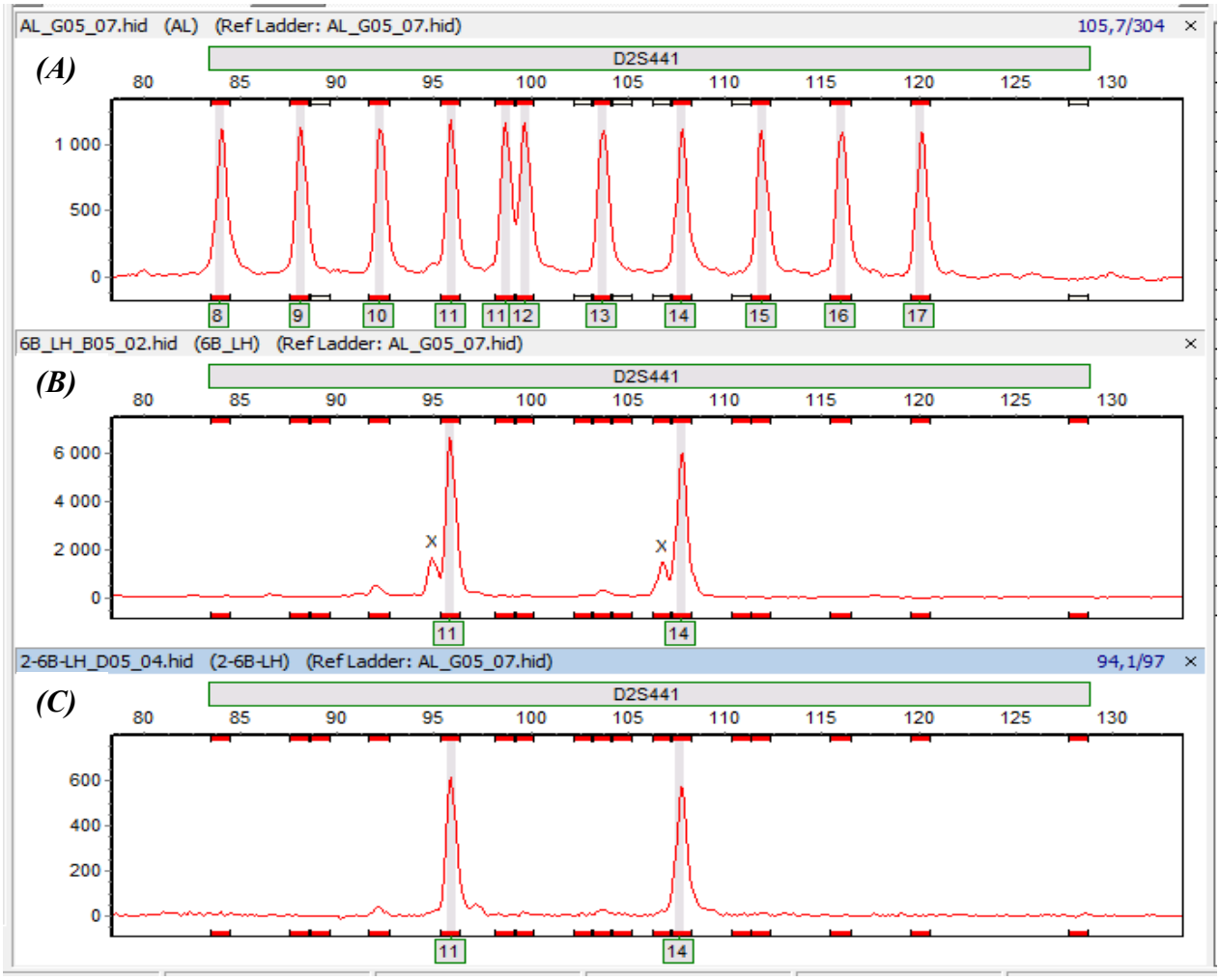


Figure 3.3: PCR cycle and capillary electrophoresis optimisation

(A) Allelic ladder and (B) split peaks observed at STR-marker D2S441 when 27 PCR cycles and 15 seconds' injection time were used to process a buccal lysate (C) no split peaks observed at STR marker D2S441 when 24 PCR cycles and 10 seconds' injection time were used to process the same buccal lysate. The split peaks are marked with an "x". (The electropherograms were obtained from GeneMarker® Software).

3.3 Allele frequencies

Forensic DNA profiling was subsequently performed using the optimised workflow and electropherograms were analysed using the established thresholds. To protect the confidentiality and identity of the participants, no single forensic DNA profile is included in the results. Allele frequencies for each marker were calculated and the combined results together with allele frequencies for each respective population group are presented in Appendix 3.1.

Overall, most markers were in Hardy-Weinberg equilibrium, except for D13S317, D19S433 D5S818, D10S1248 and D2S441. Some additional markers within individual population groups were also not in Hardy-Weinberg equilibrium, and these are indicated in the allele frequency tables in Appendix 3.1.

The allele frequencies were then statistically compared for each marker between each pairwise population group and corrected for multiple testing. The results showed that there were numerous significant differences of allele frequencies for markers between population groups (Table 3.1). The allele frequencies for three markers (D8S1179, D2S1338 and D2S441) were significantly different between all population groups, while one marker (FGA) were similar between all but one paired population group (Black African and White). The most prominent difference between allele frequencies for paired population groups was between Black African and White population groups (21 markers) where the least difference was observed between the Coloured and Indian/Asian population groups (7 markers).

Table 3.1: Statistical comparison of allele frequencies for each marker between population groups

The ρ-values from the Chi-square tests are portrayed in the table below. An asterisk indicates a significant difference (ρ < 0.0083) of alleles in a marker between paired population groups.

	Black African and Coloured	**Black African and Indian/Asian**	**Black African and White**	**Coloured and Indian /Asian**	**Coloured And White**	**Indian/Asian and White**
TH01	< 0.001*	< 0.001*	< 0.001*	0.011	< 0.001*	< 0.001*
D3S1358	0.264	0.101	< 0.001*	0.597	0,004*	0.055
vWA	0.008*	< 0.001*	< 0.001*	0.023*	0.034	0.104
D21S11	< 0.001*	< 0.001*	< 0.001*	< 0.001*	0.550	< 0.001*
TPOX	< 0.001*	< 0.001*	< 0.001*	0.014	< 0.001*	0.244
DYS391	0.001*	< 0.001*	< 0.001*	0.493	0.018	0.014
D1S1656	< 0.001*	0.002*	< 0.001*	0.325	< 0.001*	< 0.001*
D12S391	< 0.001*	< 0.001*	< 0.001*	0.305	0.016	0.017
SE33	< 0.001*	< 0.001*	< 0.001*	0.084	0.393	0.177
D10S1248	< 0.001*	< 0.001*	< 0.001*	< 0.001*	0.149	< 0.001*
D22S1045	0.017	< 0.001*	< 0.001*	0.005*	< 0.001*	0.001*
D19S433	0.013	< 0.001*	< 0.001*	0.377	< 0.001*	< 0.001*
D8S1179	< 0.001*	< 0.001*	< 0.001*	0.002*	< 0.001*	< 0.001*
D2S1338	< 0.001*	< 0.001*	< 0.001*	0.004*	< 0.001*	< 0.001*
D2S441	< 0.001*	< 0.001*	< 0.001*	< 0.001*	< 0.001*	< 0.001*
D18S51	< 0.001*	< 0.001*	< 0.001*	0.048	0.005*	0.525
FGA	0.113	0.515	< 0.001*	0.319	0.079	0.011
D16S539	0.111	< 0.001*	< 0.001*	0.021	0.001*	0.032
CSF1PO	0.001*	< 0.001*	< 0.001*	0.035	0.208	0.001*
D13S317	< 0.001*	< 0.001*	< 0.001*	0.073	0.559	0.630
D5S818	< 0.001*	< 0.001*	< 0.001*	0.063	< 0.001*	0.030
D7S820	< 0.001*	0.004*	0.040	0.159	0.099	0.111

Since the DNA profiles were mostly from females and from the Western Cape, statistical analysis was performed to assess if there were differences in allele frequencies between sex as well as between the Western Cape and all the other geographical provinces grouped together. A significant difference between sexes was observed at only two markers (D2S441; $\rho = 0.002$ and D13S317; $\rho = 0.042$) and a significant difference was observed for geographical differences only at one marker (D13S317; $\rho = 0.048$). The rest of the markers showed no significant difference in allele frequencies between sexes and geographical location.

During analysis, numerous alleles (including microvariants) were observed which were not included in the allelic ladder (Table 3.2). Some of these alleles had respective bins included in the template files for the kit, but 30 samples had off ladder alleles that were calculated manually. Sixteen of these alleles were not yet reported on the STRbase NIST website www.cstl.nist.gov/biotech/strbase/; date accessed 2017.06.21 [66].

Table 3.2 Novel alleles observed in the four population groups (BA=Black African, C=Coloured, IA=Indian/Asian, W=White)

The occurrence of the microvariants and alleles not included in the allelic ladder are portrayed in this table. The asterisk indicates alleles not reported to the National Institute for Standards and Technology.

Marker	Allele	BA	C	IA	W	Marker	Allele	BA	C	IA	W
TH01	8.3		1			**D10S1248**	7	1			
vWA	18.2		1			**D22S1045**	13.2				1
D21S11	24.3	1					20*		1		
	29.3		2			**D19S433**	7	10	11		
	33.1	3	1				10.3*		1		
	34.1	2	1	1			14.1		2		
	35.1	1	3				14.3				1
	36.1		2				20.2			1	
	37.1		1			**D2S441**	9.1		1	1	
	37.2		1				10.3	1			
D1S1656	13.3	1					12.3	8	8		3
	15.2				1		13.3*	5	2		
	19			2			14.3*	2	2		
D12S391	19.1	3	2			**D18S51**	15.2	3	5		
	19.2		1				16.3	1			
	19.3		1		2	**FGA**	16.1	4	1		
SE33	6.2*	1					26.2		1		
	7.2*	1					29.2	1			
	10.2*	1					38.3*				1
	11.2	2	2			**D16S539**	4		1		
	12.2		3			**D13S317**	8.3*		1		
	16.2	1	2		1		11.2*		1		3
	16.3			1			12.1*		1		
	17.2	1					13.1*				1
	17.3	2		1		**D5S818**	8.1*		2		
	18.3				2	**D7S820**	7.3	2			
	20.1*			1			10.3	1			
	21.1		2								
	21.3*	2									
	23	1			1						
	23.1*			1							
	30		1		1						

For the comparison of allele frequencies to the study performed by Lucassen *et al.* (2014), the allele frequencies established in Lucassen *et al.* (2014) study were changed to counts and then compared with the data from this study [14]. The results from this analysis are in Appendix 3.2.

3.4 Quality Sensor

One of the features of the QIAGEN Investigator® 24 PLEX GO! [Hilden, Germany] kit was the inclusion of the Quality Sensor (QS1 & QS2) markers [67]. In instances where the electropherograms were not optimal, these markers indicated the subsequent step to be performed. In three buccal lysate samples, no peaks were observed at any of the markers except for QS1 and QS2, indicating that no DNA was present during PCR (Figure 3.4). Despite repeating these samples to negate pipetting error as the reason, the same results transpired and thus it could be concluded that insufficient epithelial cells were collected from the participants.

In another two extracted DNA samples, the quality sensor peaks showed that the DNA was degraded and thus repeating the forensic DNA profiling on those samples would not improve the results (Figure 3.5). In total, the samples from five participants were not successfully profiled, and the inclusion of the Quality Sensor markers revealed the reasons for this. The first-time success rate for the extracted DNA samples was 89.6% (number of repeats = 24) after which a final success rate of 99.1%. For buccal lysates, there was a 98.4% first-time success rate (number of repeats = 7), with a final success rate of 99.3%.

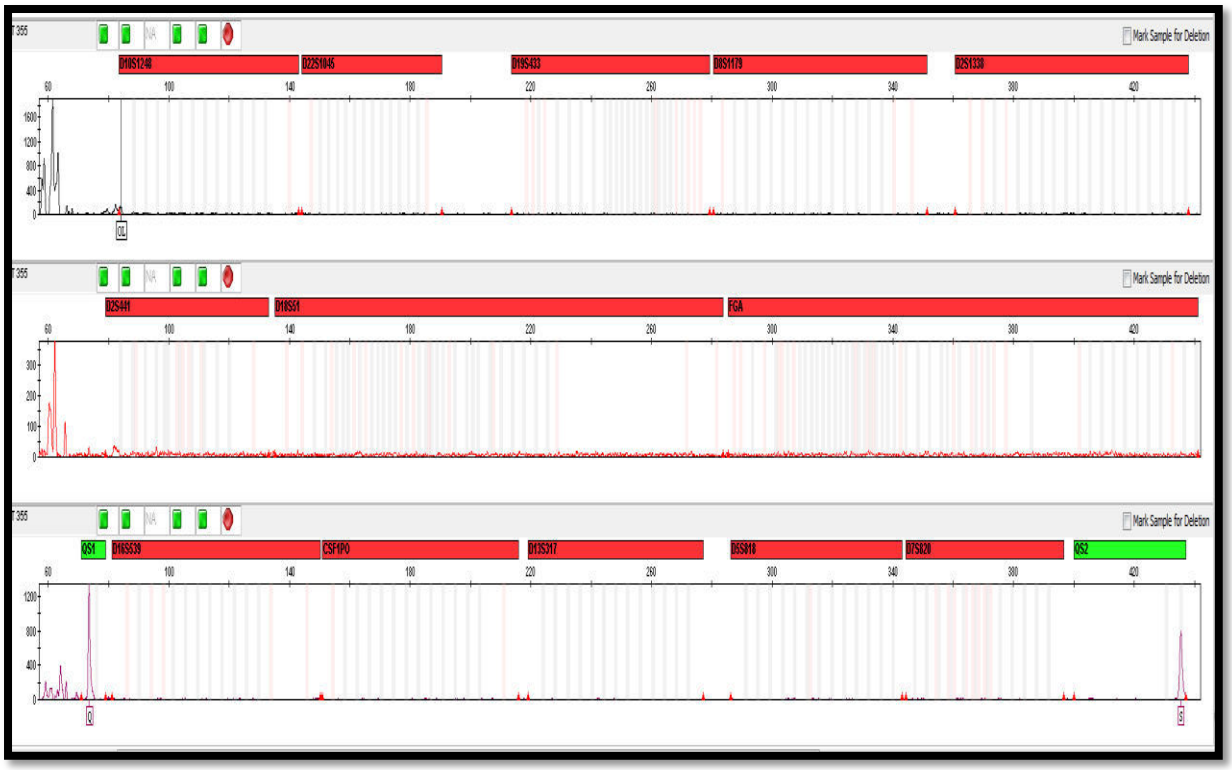


Figure 3.4 Quality Sensor marker indicates successful amplification

This internal performance control (QS1 and QS2) were useful in this study; in this instance, it indicated that no DNA was present on this buccal lysate sample (The electropherogram was obtained from GeneMapper® Software).

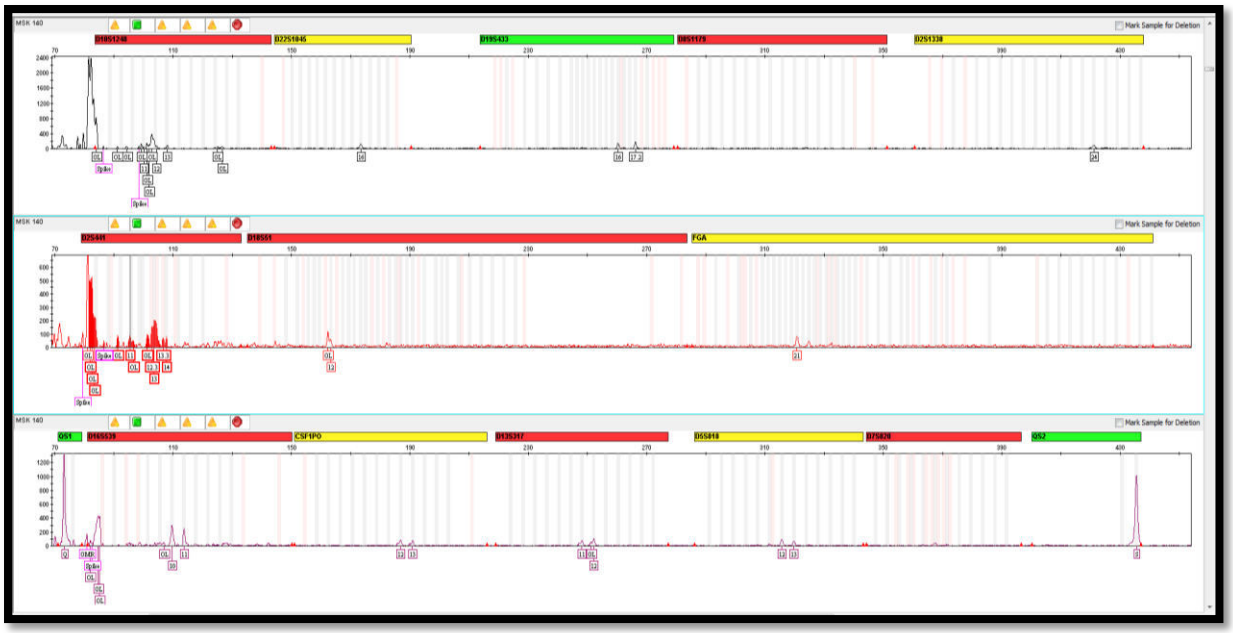


Figure 3.5 Quality sensor indicates degraded DNA

PCR amplification also successful, but ski-slope effect of alleles indicated degraded DNA of this extracted DNA sample (The electropherogram was obtained from GeneMapper® Software).

3.5 Anomalies

3.5.1 Tri-allelic patterns

During the analysis of results, a tri-allelic pattern at the marker TPOX was detected. Nine tri-allelic patterns were observed from the 655 participants; two in the Coloured population (males = 0; females = 2) and seven in the Black African population (males = 4; females = 3) (Figure 3.6). The patterns were 8, 9, 10 (n = 2); 8, 10, 11 (n = 2), 6, 10, 11 (n = 2); 6, 9, 10 (n = 1); 7, 10, 11 (n = 1) and 9, 10, 11 (n = 1).

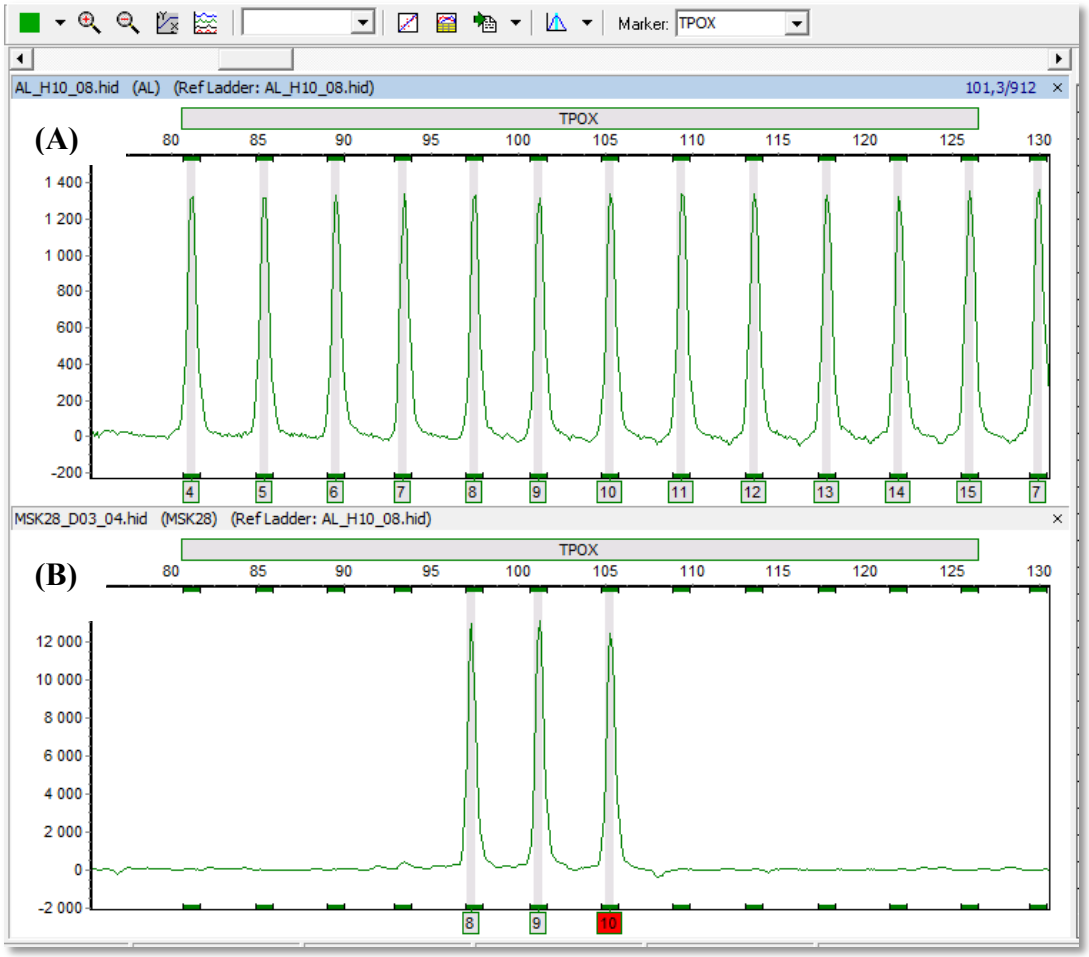


Figure 3.6: Tri-allelic pattern at TPOX

(A) Allelic ladder. (B) A genetic anomaly of three alleles at TPOX was observed for 9/655 participants (paragraph 3.4.1) (The figure above was obtained from GeneMarker® Software).

3.5.2 Null allele

In one instance, a single peak was observed at the amelogenin marker (indicating a female) and a single peak was also observed at the additional sex marker, DYS391 (indicating a male). This profile belonged to an Indian/Asian male, revealing a null allele at amelogenin (Figure 3.7.A) [2]. However, this occurred another 35 instances as well, but in these cases the profile belonged to female participants (Figure 3.7.B).

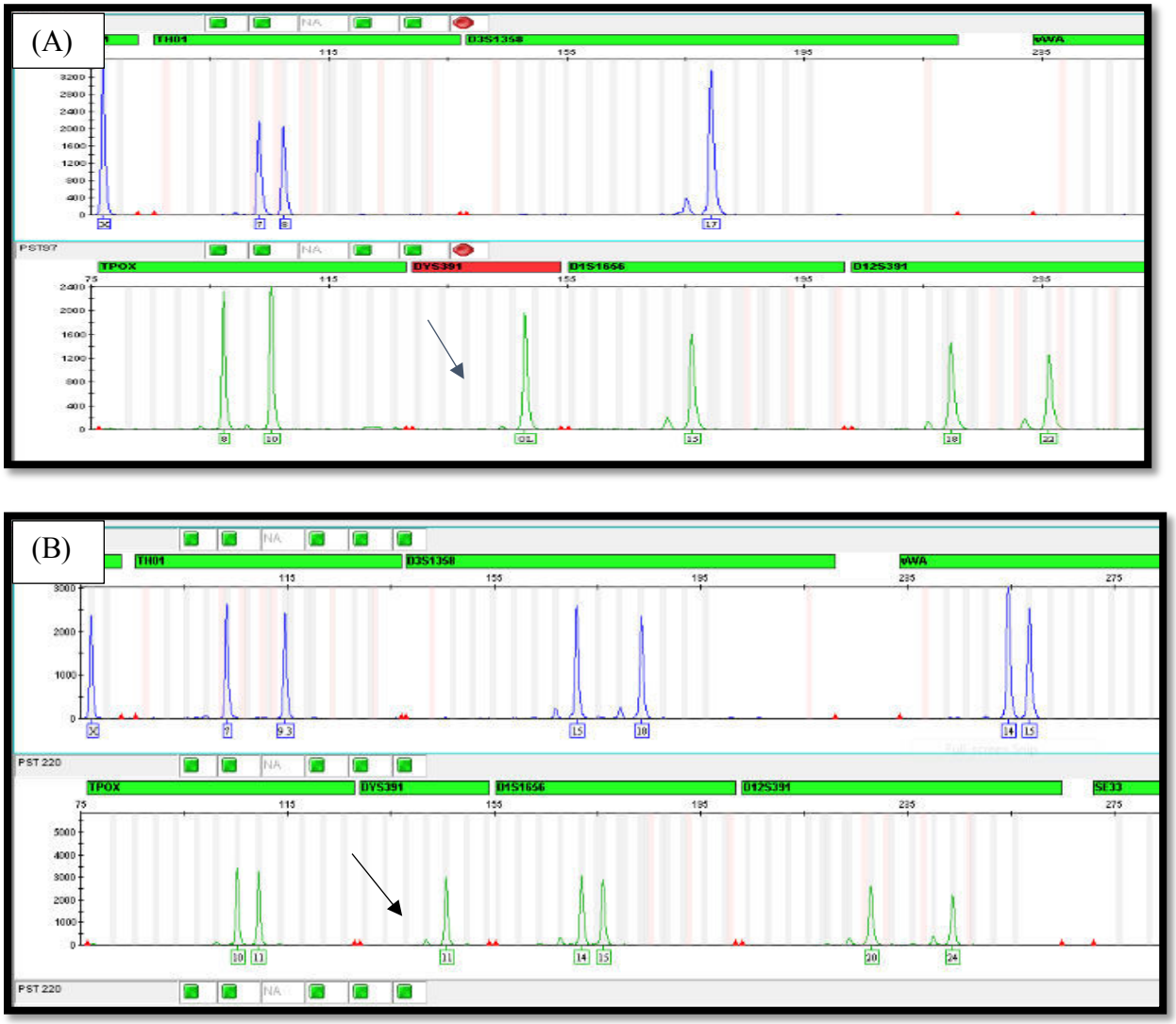


Figure 3.7 Additional sex marker DYS391

(A) A male profile with the null amelogenin and (B) is a female profile (The figures above were obtained from GeneMapper® Software).

Chapter 4: Discussion

4.1 Background and optimisation

The aim of this study was to generate DNA profile allele frequencies from South African individuals. To this end, forensic DNA profiles were successfully generated from 655 South African individuals. The objective was to collect 100 - 200 participants from each population group and to subject these samples to an optimised workflow using the Investigator® 24plex GO! kit (QIAGEN, Hilden, Germany) to generate the forensic DNA profiles and subsequently calculate the allele frequencies.

Forensic DNA profiling was performed using the Investigator® 24plex GO! kit (QIAGEN, Hilden, Germany) [28]. This chemistry featured six more autosomal STR-markers (D1S1656, D12S391, SE33, D10S1248, D22S1045 and D2S441) and an additional Y-STR sex marker (DYS391) than the current kit utilised by the Forensic Science Laboratory in the South African Police Service. Since this chemistry and workflow was new to the academic laboratory, quality thresholds and PCR optimisation had to be performed before submitting the samples to the process.

The thresholds of the extracted DNA samples were higher than those for the buccal lysates which may be attributed to the different PCR protocols for the two sample types. During the establishment of thresholds, an artefact was observed at D10S1248 in the lysate samples (28 times) at 101.04 – 101.19 bp at an average height of 368 RFU (Appendix 4.1). The height of the artefact was not affected by the serial dilution. A similar artefact in this marker was previously described when the PowerPlex® ESX 16 and PowerPlex® ESX 17 systems (PROMEGA, Madison, Wisconsin, US) were used; the peak was observed at 60 - 64 bp [68] and 86 bp [69] in the two kits respectively, and was subsequently described as a low-level DNA- dependent PCR artefact. Given that this artefact was reported before, it was excluded from the calculation of the analytical threshold during this study. Further, the presence of this artefact did not affect the interpretation of results.

During optimisation, split peaks were observed when 27 PCR cycles were used to amplify the buccal lysates and when the injection time for capillary electrophoresis was 15 seconds. Split

peaks are characterised by two peaks that differ by one base pair only, which arises due to incomplete adenylation during the amplification process, either due to an inactive or suboptimal *Taq* DNA polymerase enzyme or an excessive amount of template DNA [2, 60]. This effect could be exacerbated by a high number of PCR cycles and/or excessive injection time for capillary electrophoresis [2]. When the PCR cycle number for buccal lysates was reduced to 24, and the injection time for capillary electrophoresis decreased to 10 seconds, split peaks were no longer observed (Figure 3.3). An alternative solution may have been to increase the final extension step of amplification as this would allow adenylation to complete [60].

4.2 Alleles

4.2.1 Allele frequencies

South Africa's population is characterised by diversity which arose from the country's dynamic history and has resulted in large genetic variation and admixed population groups [42]. The South African population census divides the population into four major groups this being Black African, Coloured, Indian/Asian and White [47]. Each of these population groups can further be classified into subgroups, per language, religion or culture [46].

Allele frequencies for all markers in each population group were calculated and some markers' alleles were not in Hardy-Weinberg equilibrium (Appendices 3.1.1 - 3.1.5). Departures from Hardy Weinberg equilibrium are a result from aspects such as inbreeding, natural selection, mutations in STR markers, population stratification, genotyping errors or random chance [70, 71]. In the current study, the reason for this were most likely due to the admixed nature of the South African groups, possibly rendering the recommended 100 – 200 individuals too small in our context.

The reason for possible relatedness between participants, especially between those from the same geographical area, must also be considered. While every effort was made to ensure sampling did not occur within blood lines, the small possibility of cousins or distant relationships cannot be discounted fully, due to the confidential nature of participation. Sampling from unrelated participants is important to ensure precision of allele frequency estimates, by increasing the number of independent alleles sampled [2]. This variable was

challenging to control for in a population study of this size, and may also be an unreported factor in other population studies; however, if the population groups are sufficiently large, the impact of a few distant relationships should not influence the outcome substantially.

Since 47% of the participants were from the Western Cape province, allele frequencies between those from the Western Cape were compared to those from the other provinces. Except for D13S317 ($\rho = 0.048$), no significant difference was observed between the two groups, suggesting that the data here were not skewed towards the Western Cape province. This was also observed in other population studies in Africa, whereby allele frequencies from different geographical locations within a country were not significantly different in Northern Algeria (Berber and Arab language; 10 districts), Botswana (Bantu and KhoiSan; 9 districts), Guinea Bissau (13 Ethnicities) and South Africa (Zulu, Xhosa, Tsonga/Shangaan, Southern Sotho, Pedi, Tshwana, Venda) [33, 51, 48, 72].

The marker which had the highest number of different alleles was SE33 (54 different alleles); which was observed in the four populations as follows: Black African (34 different alleles), Coloured (38 different alleles), Indian/Asian (29 different alleles) and White (33 different alleles), rendering it one of the more informative markers in the South African population.

Novel alleles, which were not included in the allelic ladder, nor recorded on the STRbase NIST database, were observed in this study (Table 3.2). Of the 59 novel alleles observed here, it was found that 28 had previously been observed by Lane *et al.* (2002), Lucassen *et al.* (2014) and Ristow *et al.* (2016) but appear to not have been confirmed or submitted to the global repositories [14, 16, 48]. These include markers D21S11 (alleles 24.3, 29.3, 33.1, 34.1, 35.1, 36.1), D12S11 (allele 19.1), SE33 (alleles 6.2, 11.2, 12.2, 16.2, 17.2, 21.1, 21.3, 23), D10S1248 (allele 7), D22S1045 (alleles 13.2, 20), D2S441 (alleles 9.1, 12.3, 13.3, 14.3), D18S51 (Allele 15.2), FGA (alleles 16.1 and 29.2), and D7S820 (allele 10.3). Since these alleles have now been observed here, in an independent study, there is a strong motivation to report these novel alleles to manufacturers and global repositories for their inclusion in kits and databases alike. The novel alleles observed in this study, but not in previous studies, need to be confirmed by repetition and then Sanger sequencing before reporting to STRbase NIST [66]. Nevertheless, these findings accentuate the genetic diversity of South Africa and provide further support for the use of local allele frequency data.

4.2.2 Tri-allelic patterns

Tri-allelic patterns were observed at the STR-marker TPOX in this study (Figure 3.6). In this study, the occurrence was 4.1% in the Black African and 1.03% in the Coloured population groups, which was slightly higher than what Ristow *et al.* (2016) observed (Black African: 1.5%; Coloured: 0.9%) [15].

Lane (2008) reported that this anomaly was common in South Africa and it occurred approximately 2.4% in the Black African population (sample size: n = 6662) with the 'additional allele' 10 occurring more in females (3.4%) than in males (1.4%) [26]. In this study, however, the prevalence of the tri-allele in females and males was 3.2% and 5.2% respectively, for the Black African population. The difference in results may be attributed to the relatively smaller sample size in this study compared to Lane (2008) [26].

4.2.3 Null allele

Another genetic anomaly observed in this study was a null allele at the sex marker amelogenin which results in a male 'appearing' as a female on the forensic DNA profile (Figure 3.7.A). This anomaly occurs was observed once in the Indian/Asian population (1/88 = 2.7%). The deletion of the Y-chromosome amelogenin region has been reported to be more common in Indian populations than those of European and African origin [2]. Amelogenin X allele dropout has also been observed in males [73]. As a result, Thangaraj *et al*. (2002) queried the reliability of this marker for sex identification and suggested the inclusion of an additional Y chromosome marker [74].

In the Investigator® 24plex GO! kit (QIAGEN, Hilden, Germany), a second sex marker (DYS391) is included. However, this too needs to be interpreted with caution since it has been reported previously and in this study (Figure 3.7.B), that an excessive amount of female DNA can present as an additional peak. Therefore, the inclusion of more than one sex marker is recommended in DNA profiling [75].

Another reason for the DYS391 peak in female samples has been attributed to sequence homology in the primer binding regions between the Y and X chromosomes, but most of these artefacts have been removed over the years by modifications in primer design and PCR conditions [74]. However, these issues may not be fully resolved as evidenced by these results

in under-researched South African populations. This hypothesis requires further investigation by sequencing the primer binding regions of these individuals.

4.3 Comparison to other studies

Numerous population studies have been conducted previously, to which the results of this study have been compared and discussed below. Of particular interest is the study by Lucassen *et al.* (2014), which currently constitutes the largest forensic population data available [14]. The sample size of Lucassen *et al.* (2014) study was larger than this study with a total of 1543 individuals consisting of Black African (n = 634), Coloured (n = 218), Indian/Asian (n = 229) and White (n = 462) from four provinces in South Africa (Gauteng, Eastern Cape, KwaZulu Natal and Western Cape) [14]. The AmpFℓSTR® Identifiler Plus™ PCR kit (Applied Biosystems, Foster City, CA) was utilised to generate DNA profiles and create the allele frequency data.

The fifteen STR markers common to the current study were statistically compared. Only in two instances was there a significant difference between allele frequencies between the two studies (i) the vWA marker was significantly different in the White population groups between the two studies and (ii) the D19S433 marker was significantly different in the Indian/Asian population groups between the two studies (Appendix 3.2); however, once the Bonferroni correction was applied, these markers' allele frequencies were no longer statistically different between the two studies. This result suggested that even though the sample size, forensic DNA profiling chemistry, and sampling sites were different in this study, the forensic DNA profiles for these 15 markers were reproducible and the sample size in this study was representative of a larger sample size.

The most recent population data available for South Africa however is from Ristow *et al.* (2016) when a validation for the GlobalFiler® Express kit (PROMEGA, Madison, Wisconsin, US) was performed [16]. The sample size was 541 and the South African population were divided into five different ethnic groups, namely amaXhosa (n = 120), amaZulu (n = 100), Cape Coloured (n=113), Asian Indian (n = 102) and Afrikaner (n = 106); where the amaXhosa and amaZulu groups belong to the Black African population and the 'Afrikaner' group being Afrikaans-speaking individuals from Dutch settler progenies. Even though the sample size of Ristow *et al.* (2016) is comparable to the current study, all STR-markers in their study were

reported to be in Hardy Weinberg equilibrium [16]. While the population classification differs from those used in this study, significant differences in allele frequencies were also observed between all population groups, except for the two Black African populations.

Similar to this study, Ristow *et al.* (2016), also reported that SE33 was the most informative STR marker with a total of 50 different alleles; with the highest occurrence in the Coloured population group (35 different alleles) [16]. Likewise, Butler *et al.* (2009) observed a total of 58 alleles for this marker in the United States populations (n = 1443) with the highest occurrence in the African American population (43 different alleles) [28]. This suggests that SE33 would be an important marker to include in DNA profiling kits in African populations in the future.

Schlebusch *et al.* (2012) and Tau *et al.* (2017) used the same chemistry as Lucassen *et al.* (2014) to generate DNA profiles for several Southern African countries, including Botswana, Namibië and Angola, and 12 novel alleles similar to this study were observed. In another study by Kido *et al.* (2006), the South African population was grouped into South African Africans and South African Europeans, and profiled using the AmpFℓSTR® Identifiler Plus™ PCR kit (Applied Biosystems, Foster City, US) [14, 39, 72, 76]. This study reported that the African group from South Africa differed significantly in terms of their allelic frequencies compared to Mozambique (4 loci); Cabinda (4 loci), Equatorial Guinea (7 loci), African Americans (10 loci) and South African Europeans (all loci); whereas the European South Africans' allele frequencies were not significantly different to the Belgium population (13 loci) as well as to the United States Caucasians (13 loci) but were different to the Polish (7 loci) [39]. In this study allele frequencies were, however significantly different between pair-wise population groups (Table 3.1). This result, together with these studies, emphasises the genetic diversity of the population groups in South Africa and accentuates the importance of local population studies, particularly in previous under-researched African population groups.

4.4 Strengths and limitations

The major strength of this study was that it has provided data for six new markers; and an additional sex marker, for which data did not yet exist for these specific population groups. In addition, it has reported 59 novel alleles, which may be of statistical value during interpretation of forensic casework samples. With the addition of new markers, the discriminatory power has

potential to be higher; therefore, even partial and mixed profiles may be more useful than before. The inclusion of the additional internal performance control is anticipated to improve interpretation of forensically challenging profiles (PCR inhibition, degradation, low copy number DNA) and will exclude unnecessary repeats during forensic casework.

The main limitation of this study was the sample size, especially for the Indian/Asian population group (n = 88). This study can be expanded to include more individuals in the future, particularly due to the presence of admixed population groups in South Africa, evidenced by some markers departing from Hardy Weinberg Equilibrium. Another limitation was that some demographic data was not available for all participants which limited the extent ethnic, language and geographical comparisons.

The way forward in this study should focus on the confirmation of novel alleles and subsequent reporting to NIST STRbase. To this end, results must be duplicated and then Sanger sequencing must be performed. Future research may also include the stratification of the Black African populations into respective ethnic groups, to determine if there are any statistical significances and what the implications might be in forensic case work. Future studies should also investigate the difference in allele frequencies between provinces to confirm the finding in this study on a larger sample size.

A concordance study can be performed between the Investigator® 24plex GO! kit (QIAGEN, Hilden, Germany) and the AmpFℓSTR® Identifiler Plus™ PCR kit (Applied Biosystems, Foster City, CA) in future. A concordance study will establish if the same results will be obtained if a different DNA profiling kit is used and is needed to assess potential null alleles due to primer hybridisation problems specific to a kit [2]. Hill *et al.* (2012) published data for all the differences that were observed between different kits at NIST and reported a concordance rate of 99.88% when 948301 allele comparisons were made [77].

4.5 Conclusion

Participant recruitment and the analysis of 655 samples from the four major population groups in South Africa was successfully performed using an optimised Investigator® 24plex GO! kit (QIAGEN, Hilden, Germany) workflow. Forensic DNA profiles were analysed and were used to establish allele frequency data. Genetic studies for the South African population; and for African countries are currently limited and this study illustrates how this new information and knowledge can contribute to the forensic community. The population data for the seven

additional markers, including the highly variable STR-marker SE33, has been additionally established for the four major South African population groups. The departure from Hardy Weinberg equilibrium in the current study is likely due to admixture within population groups and might be resolved if more samples are added to the study.

The major limitation of this study was the small number of Indian/Asian participants but sample collection and processing will continue. However, when the results from this study were compared to Lucassen *et al.* (2014), it was found that the allele frequencies for common markers were similar. The consistency of the results here with Lucassen *et al.* (2014) enhances the reliability of these allele frequencies for use by the forensic community in South Africa [14].

Novel alleles from the current study were also observed in other population studies performed in South Africa [14, 16, 48] as well as in neighbouring countries [72, 76]; the number of novel alleles highlighted the unique diverse genetic makeup of South Africa and accentuates the importance of local allele data. The presence of the novel alleles in this study will be confirmed through repetition and Sanger sequencing, before reporting them to STRbase NIST [66].

The skewed geographical data was initially thought to be a limitation, however, after calculations were performed, it appeared to have no significance. This was in line with other studies that were performed in Africa where allele frequencies did not differ geographically within a country [33, 51]; however different countries in Africa showed significant differences to each other as well as to South Africa [35, 78]. Overall, this study is considered to be of value to the forensic community.

References (UCT Harvard style)

1. South African Government Gazette; Act No 37 of 2013: Criminal Law (Forensic Procedure) Amendment Act of 2013
2. Butler, J.M. 2009: 155, 148, 68, 236. Fundamentals of Forensic DNA Typing. Elsevier, London, United Kingdom. [ISBN: 978-0-12-374999-4]
3. Silva N.M., Pereira L., Poloni E.S. and Currat M. 2012 Human Neutral Genetic Variation and Forensic STR Data. PLoS ONE 7 (11): e49666.
   DOI: 10.1371/journal.pone.0049666
4. Bär, W., Brinkmann, B., Budowle, B., Carracedo, A., Gill, P., Lincoln, P., Mayr, W and Olaisen, B. 2007. DNA recommendations; Further report of the DNA Commission of the ISFH regarding the use of short tandem repeat systems. International Journal of Legal Medicine (110) 175-176
5. Lygo, J. E., Johnson, P. E., Holdaway, D. J., Woodroffe, S., Whitaker, J. P., Clayton, T. M., Kimpton, C. P. and Gill, P. 1994. The validation of short tandem repeat (STR) loci for the use in forensic casework. International Journal of Legal Medicine (107) 77-89
6. SWGDAM Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories. Available online.
   https://www.fbi.gov/about-us/lab/biometric-analysis/codis/swgdam-interpretation-guidelines
   [Date accessed 2015.08.03]
7. Gill, P., Sparkes, R. and Kimpton, C. 1997. Development of guidelines to designate alleles using an STR multiplex system. Forensic Science International (90) 185-197
8. Schneider, P. M. 2007. Scientific standards for studies in forensic genetics. Forensic Science International (165) 238-243
9. Kumar, S., Verma, A.K., Singh, P. & Singh, R. 2016. Current scenario of forensic DNA databases in or outside India and their relative risk. Egyptian Journal of Forensic Sciences (6), 1-5
   DOI: 10.1016\j.ejfs.2015.03.002
10. Hares, D. A. 2011. Expanding the CODIS core loci in the United States. Forensic Science International: Genetics (6) e52 – e54
    DOI: 10.1016/j.fsigen.2011.04.012

11. Moretti, T. R., Moreno, L. I., Smerick, J. B., Pignone, M. L., Hizon, R., Buckleton, J. S., Bright, J. and Onorato, A. J. 2016. Population data on the expanded CODIS core STR loci for eleven populations of significance for forensic DNA analyses in the United States. Forensic Science International (25) 175-181
DOI: 10.1016/j.fsigen.2016.07.022
12. Taylor, D., Bright, J., McGoven, C., Hefford, C., Kalafut, T. and Buckleton, J. 2016. Validating multiplexes in conjunction with modern interpretation strategies. Forensic Science International: Genetics (20) 6-19
DOI: 10.1016/j.fsigen.2015.09.011
13. Gill, P., Fereday, L., Morling, N. and Schneider, P. M. 2006. The evolution of DNA databases – Recommendations for new European STR loci. Forensic Science International. (156) 242-244
DOI: 10.1016/j.forsciint.2005.05.036
14. Lucassen, A., Ehlers, K., Grobler, P. J. & Shezi, A. L. 2014. Allele frequency data of 15 autosomal STR loci in four major population groups of South Africa. International Journal of Legal Medicine (128) 275-276
DOI: 10.1007/s00414-013-0898-4
15. Gill, P. 2002. Role of Short Tandem Repeat DNA in forensic casework in the UK- Past, present and future perspectives. BioTechniques (32) 366-385
16. Ristow, P. G., Cloete, K. W., D'Amato, M. E. 2016 GlobalFiler Express DNA amplification kit in South Africa: Extracting the past from the present. Forensic Science International: Genetics (24)194-201
17. Gettings, K. B., Aponte, R. A., Vallone, P. M. and Butler, J. M. 2015. STR allele sequence variation: Current knowledge and future issues. Forensic Science International: Genetics (18) 118-130
DOI: 10.1016/k.fsigen.2015.06.005
18. Moretti, T. R., Baumstarl, A. L., Defenbaugh, D. A., Keys, K. M., Serick, J. B. and Budowle, B. 2001. Validation of Short Tandem Repeats (STRs) for forensic usage: Performance testing of fluorescent multiplex STR systems and analysis of authentic and simulated forensic samples. Journal of Forensic Science (46) 3: 647-660
19. Gill, P., Sparkes, R., Fereday, L. and Werrett, D. J. 2000. Report of the European Network of Forensic Science Institutes (ENSFI): formulation and testing of principles to evaluate STR multiplexes. Forensic Science International (108) 1-29

20. Gill, P., Gusmão, L., Haned, H. Mayr, W. R., Morling, N., Parson, W. Prieto, L., Prinz, M., Schneider, H., Schneider, P. M. and Weir, B. S. 2012. DNA commission of the International Society of Forensic Genetics: Recommendations on the evaluation of STR typing results that may include drop-out and/or drop-in using probabilistic methods. Forensic Science International: Genetics (6)679-688
DOI: 10.1016/j.fsigen.2012.06.002
21. National recommendations of the Technical UK DNA working group on mixture interpretation for the NDNAD and for court going purposes. 2008. Forensic Science International: Genetics (2)76-82
DOI: 10.1016/j.fsigen.2007.08.008
22. Lane, A. B. 2013. STR null alleles complicate parentage testing in South Africa. Molecular Genetics. (103) 12
23. Mannucci, A., Sullivan, K. M., Ivanov, P.L. and Gill, P. 1994. Forensic application of a rapid and quantitative DNA sex test by amplification of the X-Y homologous gene amelogenin. International Journal of Legal Medicine (106) 190-193
24. http://www.cstl.nist.gov/strbase/training/Figures-Advanced-Topics-Forensic-DNA-Typing-Interpretation.pptx [accessed 2017.07.29]
25. Clayton, T. M., Guest, J. L., Urquhart, A. J. and Gill, P. D. 2004. A genetic basis for anomalous band patterns encountered during DNA STR profiling. Journal of forensic science (49)
DOI: 10.1520/JFS2003145
26. Lane, A.B. 2008 The nature of tri-allelic TPOX genotypes in African populations. Forensic Science International: Genetics (2) 134-137
27. Picanco, J. B., Raimann, P. E., Ares, C. H., Da Motta, S., Rodenbusch, R., Gusmão, L. and Alho, C. S. 2015. Identification of the third/extra allele for forensic application in cases with TPOX tri-allelic pattern. Forensic Science International (16) 88-93
DOI: 10.1016/j.fsigen.2014.11.016
28. Kreamer, M., Prochnow, A., Bussmann, M., Scherer, M., Peist, R. and Steffen, C. 2017. Developmental validation of QIAGEN Investigator®24plex QS Kit and Investigator® 24plex GO! Kit: two 6-dye multiplex assays for the extended CODIS core loci (29) 9-20
DOI: 10.1016/j.fsigen.2017.03.012
29. Butler, J. M., Hill, C. R., Kline, M. C., Duewer, D. L., Cynthia J. Sprecher, C. J., McLaren, R. S., Rabbach, Benjamin E. Krenke, B. E., and Storts, D. R. 2009. The single

most polymorphic STR locus: SE33 performance in U.S. populations. Forensic Science International: Genetics supplement series. (2) 23-24
DOI: 10.1016/j.fsigss.2009.08.173

30. Scherer, M., Konig, M., Breitbach, M., Cornelius, S., Bussmann, A., Prochnow, A., Peist, R. 2015. Improving sample analysis and interpretation using QIAGEN's latest Investigator® STR multiplex PCR assays with a novel quality sensor. Forensic Science International: Genetics Supplement Series (5) e308-e309
DOI: 1.1016/j.fsigss.2015.09.122
31. Promega Reference Manual. Internal validation guide of autosomal STR systems for forensic laboratories. Available online. https://www.promega.jp/-/media/files/resources/validation-guides/internal-validation-of-autosomal-str-systems.pdf?la=ja-jp [Date accessed: 2015.08.03]
32. Welch, L. A., Gill, P., Phillips, C., Ansell, R., Morling, N., Parson, W., Palo, J. U., Bastisch, I. 2012. European Network of Forensic Science Institutes (ENFSI): Evaluation of new commercial STR multiplexes that include the European Standard Set (ESS) of markers. Forensic Science International: Genetics (6) 819-826
DOI: 10.1016/j.fsigen.2012.03.005
33. Gonçalves, R., Jesus, J., Fernandes, A. T. and Brehm, A. 2002. Genetic profile of a multi-ethnic population from Guiné-Bissau (west African coast) using the new PowerPlex® 16 system kit. Forensic Science International (29) 78-80
34. Tomas, C., Skitsa, I., Steinmeier, E., Poulsen, L., Ampati, A., Børsting, C. and Morling, N. 2015. Results for five sets of forensic genetic markers studied in a Greek population sample. Forensic Science International: Genetics (16) 132-137
DOI: 10.1016/j.fsigen.2015.01.001
35. Beleza, S., Alves, C., Reis, F., Amorim, A., Carracedo, A. and Gusmão, L. 2004. 17 STR data (AmpF/STR Identifiler and Powerplex 16 system) from Cabinda (Angola). Forensic Science International (141) 193 – 196
DOI: 10.1016/j.forsciint.2004.01.008
36. Alves, C., Gusmao, L., Lopez-Parra, A. M., Soledad Mesa, M., Amorim, A., Arroyo-Pardo, E. 2004. STR allelic frequencies for an African population sample (Equatorial Guinea) using AmpFℓSTR Identifiler and Powerplex 16 kits. Forensic Science International (148) 239-242
DOI: 10.1016/j.forsciint.2004.05.007

37. Alves, C., Gusmão, L., Damasceno, A., Soares, B. and Anorim, A. 2004. Contribution for an African autosomic STR database (AmpF/STR Identifiler and Powerplex 16 system) and a report on genotypic variations. Forensic Science International (139) 201-205
DOI: 10.1016/j.forsciint.2003.11.002
38. Fujihara, J., Muro, T., Nakamura, H., Imamura, S. and Takeshita, H. 2007. Allele frequencies for nine STR loci in Ovambo population using AmpFlSTR® Profiler kit. Forensic Science International (169) e7-e9
39. Kido, A., Dobashi, Y., Fujitani, N., Hara, M., Susukida, R., Kimura, H., Oya, M. 2006. Populltion data on the AmpFℓSTR Identifiler loci in Africans and Europeans from South Africa. Forensic Science International (168) 232-235
DOI: 10/1016/j.forsciint.2006.02.040
40. Buckleton, J., Curran, J., Goudet, J., Taylor, D., Thiery, A. & Weir, B.S. 2016. Population-specific F_{ST} values for forensic markers: A worldwide survey. Forensic Science International: Genetics (23) 91-100
DOI: 10.1016/j.fsigen.2016.03.004
41. Christopher, A. J. 2009. Delineating the nation: South African censuses 1865-2007. Political Geography. (28) 101-109
DOI: 10.1016/j.polgeo.2008.12.003
42. De Wit, E., Delport, W., Rugamika, C. E., Meintjes, A., Möller, M., Van Helden, P. D., Seoighe, C. and Hoal, E. G. 2010. Genome-wide analysis of the structure of the South African Coloured population in the Western Cape. Human Genetics (128) 145-153
DOI: 10.1007/s00439-010-0836-1
43. Greeff, J.M. 2007. Deconstructing Jaco: Genetic Heritage of an Afrikaner. Annals of Human Genetics (71) 674-688
DOI: 10.1111/j.1469-1809.2007. 00363.x
44. Quintana-Murci, L., Harmant, C., Quach, H., Balanovsky, O., Zaporozhchenko, V., Bormans, C., Van Helden, P. D., Hoal, E. G. and Behar, D. M. 2010. Strong maternal Khoisan contribution to the South African Coloured population: A case of Gender-biased admixture. The American Society of Human Genetics (86) 611-620
DOI: 10.1016/j.ajhg.2010.02.014
45. Petersen, D. C., Libiger, O., Tindall, E. A., Hardie, R, Hannick, L. I., Glashoff, R. H., Mukerji, M., Indian Genome Variation Consortium, Fernandez, P., Haacke, W.,

Schork, N. J. and Hayes, V. M. 2013. Complex patterns of genomic admixture within Southern Africa. PLoS Genetics (9) 3: e1003309
DOI: 10.1371/journal.pgen.1003309

46. Adams, B. G., Van de Vijver, F. J. R. & De Bruin, G. P. 2012. Identity in South Africa: Examining self-descriptions across ethnic groups. International Journal of Intercultural Relations (36) 377-388
DOI: 10.1016\j.ijintrel.2011.11.008
47. 2016. Community Survey. Available online. www.statssa.gov.za [Date accessed: 2016-07-14]
48. Lane, A. B., Sooyall, H., Arndt, S., Ratshikhopha, M. E., Jonker, E., Freeman, C., Young, L., Morar, B. and Toffie, L. 2002. Genetic substructure in South African Bantu-speakers: Evidence from autosomal DNA and Y-chromosome markers. American Journal of physical anthropology (119) 175-185
49. LongLi, K., Dongya, Y., Fengying, Y., Kai, L. and Xi, Z. 2007. Genetic polymorphisms of 15 STR loci in two Tibetan populations from Tibet Changdu and Naqu, China. Forensic Science International (169) 239-243
DOI: 10.1016/j.forsciint.2006.03.012
50. Šimková, H., Faltus, V., Marvan, R., Pexa, T., Stenzl, V., Brouček, J., Hořínek, A., Mazura, I. and Zvárová. 2009. Allele frequency data for 17 short tandem repeats in a Czech population sample. Forensic Science International: Genetics (4) e15-e17
DOI: 10.1016/j.fsigen.2009.01.003
51. Amir, N., Sahnoune, M., Chikhi, L. and Atmani, D. 2015. STR-based genetic structure of the Berber population of Bejaia (Northern Algeria) and its relationships to various ethnic groups. Gene. (574) 140-148
DOI: 10.1016/j.gene.2015.08.002
52. Correspondence. 2016. Forensic Science International: Genetics.2016. (22) e7-e8
DOI: 10.1016/j.fsigen.2016.01.006
53. Wallace, H.M., Jackson, A.R., Gruber, J. & Thibedeau, A.D. 2014. Forensic DNA Database-Ethical and legal standards: A global review. Egyptian Journal of Forensic Sciences. (4) 57-63
DOI: 10.1016/j.ejfs.2014.04.002
54. Machado, H. & Silva, S. 2014. "Would you accept having your DNA profile inserted in the National Forensic DNA database? Why?" Results of a questionnaire applied in Portugal. Forensic Science International: Genetics (8) 132-136

DOI: 10.1016/j.fsigen.2013.08.014

55. Katsanis, S. H., and Wagner, J. K. 2013. Characterization of the standard and recommended CODIS markers. Journal of Forensic Science (58) No. S1
DOI: 10.1111/j.1556-4029.2012. 02253.x
56. Heathfield, L. J. 2014. Policy required for entry of DNA profiles onto the National Forensic DNA Database of South Africa. South African Journal of Science 110 (7/8), Article 2013-0374
DOI: 10.1590/sajs.2014/20130374
57. Maguire, C. N., McCallum, L. A., Storey, C. and Whitaker, J. P. 2014. Familial searching: A specialist forensic DNA profiling service utilising the National DNA Database® to identify unknown offenders via their relatives – The UK experience. Forensic Science International: Genetics (8) 1-9
DOI: 10.1016/j.fsigen.2013.07.004
58. Chakraborty, R. Requirements for addressing the population genetic issues of forensic use of DNA typing. Human Biology Vol 64, No.2 April 1992 pp 141-159
59. Jovanovich, S., Bogdan, G., Belcinski, R., Buscaino, J., Burgi, D., Butts, E. L. R., Chear, K., Ciopyk, B., Eberhart, D., El-Sissi, O., Franklin, H., Gangano, S., Gass, J., Harris, D., Hennessy, L., Kindwall, A., King, D., Klevenberg, J., Li, Y., Mehendale, N., McIntosh, R., Nielsen, B., Park, C., Pearson, F., Schueren, R., Stainton, N., Troup, C., Vallone, P. M., Vangbo, M., Woudenberg, T., Wyrick, D., Williams, S. 2015. Developmental validation of a fully integrated sample-to-profile rapid human identification system for processing single-source reference buccal samples. Forensic Science International: Genetics (16) 181-194.
DOI: 10.1016/j.fsigen.2014.12.004
60. Goodwin, W., Linacre, A., Hadi, S. (2011) An introduction to forensic genetics. John Wiley & sons Ltd, Oxford, United Kingdom. [ISBN 978-0-470-71018-0]
61. QIAGEN Investigator® 24 PLEX GO! Handbook. 2015. Available online. www.qiagen.com\24plex [Date accessed 2016-08-28]
62. Holland, M. M., Parson, W. (2011) GeneMarker® HID: A reliable software tool for the analysis of forensic STR data. Journal of forensic sciences (56).
DOI: 10.1111/j.1556-4029.2010. 01565.x
63. Hansson, O., Gill, P. (2011) Evaluation of GeneMapper® ID-X mixture analysis tool. Forensic Science International: Genetics supplement series (3) e11-e12. DIO: 10.1016/j.fsigss.2011.08.005

64. Bonferroni CE (1936) Teoria Statistica Delle Classi Ecalcolo Delle Probabilita. Pubblicazioni del R Instituto Superiore di ScienzeEconomiche e Commerciali di Firenze 8:3–62
65. STATA (Statistica version 13.2) software. Program by Mario A. Cleves. Licenced to University of Cape Town
66. National Institute of Standards and Technology (NIST) website (www.cstl.nist.gov/biotech/strbase/). Date accessed; 2017.05.07.
67. Scherer, M., Konig, M., Breitbach, M., Cornelius, S., Bussmann, A., Prochnow, A., Peist, R. 2015. Improving sample analysis and interpretation using QIAGEN's latest Investigator® STR multiplex PCR assays with a novel quality sensor. Forensic Science International: Genetics Supplement Series (5) e308-e309. DOI: 1.1016/j.fsigss.2015.09.122
68. Tucker, V. C., Hopwood, A. J., Sprecher, C. J., McLaren, R. S., Rabbach, D. R., Ensenberger, M. G., Thompson, J. M. and Storts, D. G. 2012. Developmental validation of the PowerPlex® ESX 16 and PowerPlex® ESX 17 Systems. Forensic Science International: Genetics (6) 124-131
DOI: 10.1016/j.fsigen.2011.03.009
69. Neuhuber, F., Klausriegler, E., Dunkelmann, B., Kreindl, G. and Cemper-Kiesslich, J. 2011. Experiences and surprises with PowerPlex ESI 17 and AmpflSTR NGM Select in routine casework. Forensic Science International: Genetics Supplement series (3) e23-e24
DOI: 10.1016/j.fsigss.2011.08.011
70. Wigginton, J. E., Cutler, D. J. and Abecasis, G. R. 2005. A note on exact test of Hardy-Weinberg equilibrium. American Journal of Human Genetics. (76)887–883
71. Wittke-Thompson, J., Pluzhnikov, A and Cox, N. J. 2005. Rational inferences about departures from Hardy-Weinberg equilibrium. Am. J. Hum. Genet. (76) 967-986
72. Tau, T., Wally, A., Fanie, T. P., Ngono, G. L., Mpoloka, S. W., Davison, S. and D'Amato, M. E. 2017. Genetic variation and population structure of Botswana populations as identified with AmpFLSTR Identifiler short tandem repeat (STR) loci. Scientific reports (7)
DOI:10.1038/s41598-017-06365-y
73. Borovlo, S., Shyla, A., Korban, V. and Borovko, A. 2015. Amelogenin test abnormalities revealed in Belarusian population during forensic DNA analysis. Forensic Science International: Genetics (15) 98-104

DOI: 10.1016/j.fsigen.2014.10.014

74. Thangaraj, K., Reddy, A. G. and Singh, L. 2002. Is the amelogenin gene reliable for gender identification in forensic casework and prenatal diagnosis? International Journal of legal medicine (116) 121-123
75. Moore, D., Clayton, T. and Thomson, J. 2016. Description of artefacts in the PowerPlex Y23® system associated with excessive quantities of background female DNA. Forensic Science International: Genetics (24) 44-50
    DOI: 10.1016/j.fsigen.2016.05.009
76. Schlebusch, C. M., Soodyal, H. and Jakobssen, M. 2012. Genetic variation of 15 autosomal STR loci in various populations from southern Africa. Forensic Science International: Genetics (6) e20-e21
77. Hill, C. R., Kline, M. C., Duewer, D. L. and Butler, J. M. 2011. Concordance testing comparing STR multiplex kits with a standard data set. Forensic Science International: Genetics supplement series (3) e188 – e189
78. Poetsch, M., Ergin, Z., Bayer, K., El-Mostaqim, D., Rakotomava, N., Browne, E. N. L., Timmann, C., Horstmann, R. D., Schwark, T. and Von Wurmb-Schwark, N. 2011. International Journal of Legal Medicine (125) 149-154
    DOI: 10.1007/s00414-010-0502-0

Appendix 2.1 Ethics approval

000

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/fhslresearch/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 wi// 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

Yours sincerely

PROFESSOR M BLOCKMAN
CHAIRPERSON FHS HUMAN RESEARCH ETHICS COMMITTEE Federal Wide Assurance Number: FWA00001637.
Institutional Review Board (IRB) number: IRB00001938

This serves to confirm that the University of Cape Town Human Research Ethics Committee complies to the Ethics Standards for Clinical Research with a new drug in patients, based on the Medical Research Council (MRC-SA), Food and Drug Administration (FDA-USA), International Convention on Harmonisation Good Clinical Practice (ICH GCP), South African Good Clinical Practice Guidelines (DOH 2006), based on the Association of the British Pharmaceutical Industry Guidelines (ABPI),and Declaration of Helsinki (2013) guidelines.
The Human Research Ethics Committee granting this approval is in compliance with the ICH Harmonised Tripartite Guidelines E6: Note for Guidance on Good Clinical Practice (CPMP/ICH/135/95) and FDA Code Federal Regulation Part 50, 56 and 312.

Appendix 2.2 Information form, consent form and questionnaire

Information form, consent form and questionnaire

Information form

Title: Constructing a DNA profile frequency database for South Africa using the QIAGEN Investigator 24plex GO! Kit

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 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

Participant number: _____

Questionnaire

1. What is your sex? ☐ Male ☐ Female

2. Are you older than 18 years? ☐ Yes ☐ No

3. Please confirm by filling in your date of birth: _____/_____/_____
   YYYY MM DD

4. Are you a South African citizen? ☐ Yes ☐ No

5. What is your home language?

   _____

6. What is your population group according to South African Census categories?

   ☐ Black African ☐Coloured ☐Indian/Asian ☐White

7. Where do you live <u>now</u>? _____

8. Where is your <u>home</u> town?

   Province: ☐ Eastern Cape ☐ Free State ☐Gauteng ☐ Kwa-Zulu Natal

   ☐Limpopo ☐ Mpumalanga ☐North west ☐ Northern Cape

   ☐Western Cape

   City:_____
   Suburb/village:_____

Please place a cross on the map indicating approximately where you live:

Appendix 2.3 Swab certificate

COPAN SGR11.1 DOC 8 Rev 04 09/2012	**CERTIFICATE OF STERILITY AND QUALITY ASSURANCE** *CERTIFICADO DE ESTERILIDAD Y GARANTIA DE CALIDAD*	**N°** **14/05663**

CODE */CÓDIGO*		LOT.N°*/NO. DE LOTE*	**EXP. DATE** */FECHA CADUCIDAD*
150C	150C	QT1400	31/10/2017
DEVICE DESCRIPTION: *DESCRIPCIÓN DEL PRODUCTO*:	PLAIN SWAB STERILE WOODEN APPLICATOR COTTON TIPPE HISOPO SENCILLOS SECOS MANGO DE MADERA		

Goods have been irradiated by gamma radiation.
The applied dose range is: 25-35 KGY
Sterilisation procedure validated in conformity with ISO 11137.
Producto ha sido irradiado por irradiación gamma.
La gama de dose aplicada es: 25-35 KGY.
El procedimiento de esterilización es validado en conformidad a la norma ISO 11137.

Goods has been manufactured according to the ISO 13485 system. *El material ha sido fabricado de acuerdo al sistema ISO 13485.*

Copan Italia SpA certifies that all product quality requirements have been met and that all information stated above is correct.
Copan Italia SpA certifica estar conforme con todos los requerimientos arriba mencionados y que toda la información es correcta.

21/10/2014	
Certificate Date of Issue *Fecha de Expedición del Certificad*	Quality Assurance Copan Italia SpA

Appendix 2.4 Certificate of analysis – Molecular biology grade water

Lonza

Lonza Rockland, Inc.
191 Thomaston Street
Rockland, ME 04841, USA
Tel (207) 594-3400
Fax (207) 594-3491

CERTIFICATE OF ANALYSIS

Product Code:	51200	**Lot Number:**	0000467168
Product:	ACCUGENE WATER-1L	**Manufacture Date:**	04-Mar-2015
		Expiration Date:	03-Mar-2016

TEST (Method)	SPECIFICATIONS Min.	Max.	Results
RNase/DNase, not detected	PASS	---	Pass
Protease, meets specification	PASS	---	Pass

This lot has been reviewed by Quality Assurance in compliance with requirements of Lonza's Quality System.

This document was generated from a validated Part 11-compliant electronic system and thus handwritten signatures are not required.

For Technical Assistance, call 1-800-521-0390

Appendix 2.5 Validation certificate – Investigator® 24plex GO! Kit

ValidationCertificate

The following QIAGEN products are validated for use in forensic applications:

Investigator 24plex GO! Kit (200), cat. no. 382426
Investigator 24plex GO! Kit (1000), cat. no. 382428

QIAGEN performed a developmental validation study based on the recommendations of:

The European Network of Forensic Science Institutes (ENFSI)

The Revised Validation Guidelines of the Scientific Working Group on DNA Analysis Methods (SWGDAM)

The results of this study show that the kit is fit for the purposes of forensic casework, paternity testing, and other human identity testing applications.

QIAGEN®

Sample & Assay Technologies

Appendix 2.6 Certificate of analysis - Investigator® 24plex GO! Kit

QIAGEN Certificate of Analysis

Product Name: Investigator 24plex GO! Kit

Catalog Number: **382428**

Lot Number: **151041004**

Content	Reagents for
Fast Reaction Mix	PCR
Primer Mix	PCR
Control DNA	PCR Control
Allelic ladder	Post PCR
DNA size standard 550 (BTO)	Post PCR

Quality Control

Passed

Test description

All reagents were tested concerning performance and reproducibility. The kit's PCR components were tested for absence of human genomic DNA by PCR and subsequent STR analysis. PCR controls were tested for amplifying the correct alleles. Allelic ladder and DNA size standard 550 were tested for correct allele and size calling.

Storage

All components of the Investigator 24plex GO! Kit should be stored at –15 to –30°C in a constanttemperature freezer.

Dr. Henning Plücken

Issue Date: 2015-08-20

Quality Assurance

This Certificate is a computer printout and therefore valid without signature.

1/1

COA-1114-001

Appendix 2.7 Certificate of analysis - Investigator® STR GO! Lysis buffer

QIAGEN Certificate of Analysis

Product Name: **Investigator STR GO! Lysis Buffer**

Catalog Number: **386516**

Lot Number: **154017439**

Content	Test description
Investigator STR GO! Lysis Buffer	All Investigator STR GO! Lysis Buffers were tested for absence of human genomic DNA by PCR and subsequent STR analysis. Passed:

Storage

The Investigator STR GO! Lysis Buffer should be stored at room temperature (15 to 25°C).

Dr. Henning Plücken Issue Date: 2015-08-20

_____ Quality Assurance

This Certificate is a computer printout and therefore valid without signature.

COA-1118-001

Appendix 2.8 Certificate of analysis – Matrix standard BT6

QIAGEN Certificate of Analysis

Product Name: **Matrix Standard BT6**

Catalog Number: **386224**

Lot Number: **154015045**

Content	Reagents for	Test description
Matrix Standard BT6	Post PCR	Matrix standards were tested concerning performance, reproducibility and correct fluorescent fragments. Passed:

Fluorescent labels
6-FAM/ BTG/ BTY/ BTR2/ BTO / BTP

Storage

The Matrix Standard BT6 should be stored at –15 to –30°C in a constant-temperature freezer.

Dr. Henning Plücken

Issue Date: 2014-10-02

_____ Quality Assurance

This Certificate is a computer printout and therefore valid without signature.

COA-0956-001

Appendix 2.9 Calibration certificate – Thermal cycler

Certificate of Temperature Verification

Certificate identification	805S1140121_241020160706
Date of Measurement	24-Oct-2016 ; 07:24 AM
Instrument type	9700 Blk - Aluminium
Base Serial Number	805S1140121
Sample/Block Serial	A96S0180332
Customer Tag #	
Raw Data Filename	805S1140121_241020160706.csv

Site identification

Company/Institute	Unistell
Contact person	
Department Address	
Tygerberg	
Postal Code	
City	
State/Province	
Country	
Phone	

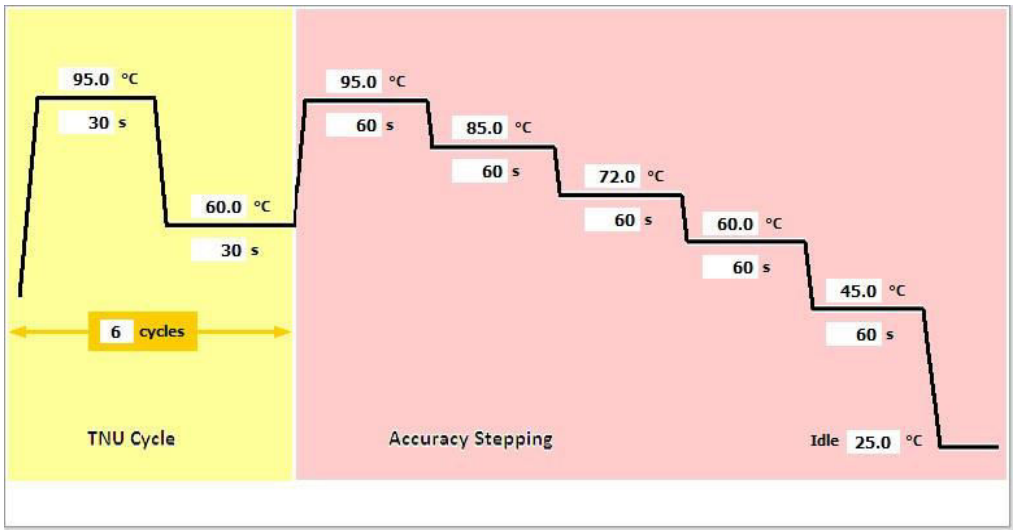
E-mail

User identification

User name	Darryl Barron (field service)
Service order	

Device identification				
Model	Serial number	Calibration date	Calibration due date	Calibration Certificate
t-POD	14D5X08-02	01/06/2016	1/6/2017	MDTV160106-028A_G
Probe Module	T06088	01/06/2016	1/6/2017	MDTV160106-028B

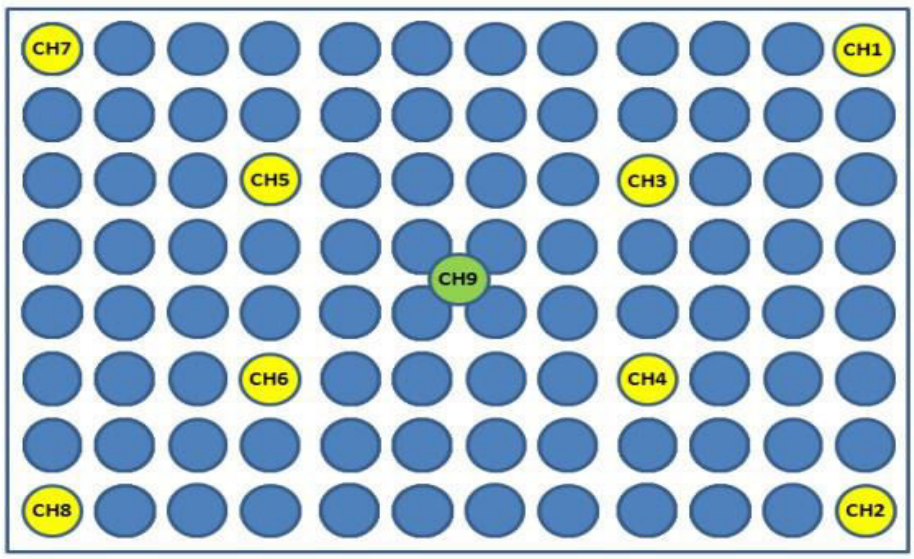
Overall Performance summary	
Measurement parameter	Pass/Fail
Sample block accuracy	Pass
Sample block uniformity	Pass
Sample block over and undershoot	Pass
Sample block ramp rates	Pass
Sample block stability	Pass
Heated cover accuracy	Pass



Probe Legend:

Channel = Well Position

CH1	=	A12
CH2	=	H12
CH3	=	C9
CH4	=	F9
CH5	=	C4
CH6	=	F4
CH7	=	A1
CH8	=	H1
CH9	=	HC



Sample block accuracy Set point accuracy (°C) for Ch1 - **Ch8 average**			
Set point	Limits	Measured value	Pass/Fail
95.0°C	±0.50	95.10	Pass
85.0°C	±0.50	85.06	Pass
72.0°C	±0.50	71.96	Pass
60.0°C	±0.50	59.93	Pass
45.0°C	±0.50	44.88	Pass

Sample block uniformity Set point uniformity (°C) of Ch1 - **Ch8**					
Cycle#	Limits	TNU at 60°C	Pass/Fail	TNU at 95°C	Pass/Fail
1	≤0.50	0.12	.	0.35	.
2	≤0.50	0.18	.	0.29	.
3	≤0.50	0.21	.	0.25	.
4	≤0.50	0.20	.	0.22	.
5	≤0.50	0.23	.	0.20	.
6	≤0.50	0.25	Pass	0.20	Pass

Sample block overshoot (OS) Set point temperature overshoot (°C) at 95.0°C (Ch1 - Ch8)			
Cycle#	Limits	Average OS	Pass/Fail
1	≤10.00	7.28	.
2	≤10.00	7.91	.
3	≤10.00	7.89	.
4	≤10.00	7.98	.
5	≤10.00	7.64	.
6	≤10.00	7.64	.
Average 4th-6th	≤10.00	7.75	Pass

Sample block undershoot (US) Set point temperature undershoot (°C) at 60.0°C (Ch1 - Ch8)			
Cycle#	Limits	Average US	Pass/Fail
1	≤10.00	8.25	.
2	≤10.00	7.77	.
3	≤10.00	7.73	.
4	≤10.00	7.88	.
5	≤10.00	7.63	.
6	≤10.00	7.95	.
Average 4th-6th	≤10.00	7.82	Pass

Sample block down-ramp rate Down-ramp rate (°C/s) from 95.0°C - 60.0°C (Ch1 - Ch8)			
Cycle#	Limits	Down-ramp rate	Pass/Fail
1	≥2.00	3.11	.
2	≥2.00	2.98	.
3	≥2.00	2.97	.
4	≥2.00	2.97	.
5	≥2.00	2.95	.
6	≥2.00	2.97	.
Average 4th-6th	≥2.00	2.96	Pass

Sample block up-ramp rate Up-ramp rate (°C/s) from 60.0°C - 95.0°C (Ch1 - Ch8)			
Cycle#	Limits	Up-ramp rate	Pass/Fail
1	≥2.00	2.60	.
2	≥2.00	2.93	.
3	≥2.00	3.00	.
4	≥2.00	3.01	.
5	≥2.00	3.02	.
6	≥2.00	3.01	.
Average 4th-6th	≥2.00	3.01	Pass

Sample block stability Set point stability (°C) for CH1 - Ch8			
Set point	Limits	Calculated stability (Max - Min)	Pass/Fail
95.0°C	≤0.10	0.03	Pass
85.0°C	≤0.10	0.03	Pass
72.0°C	≤0.10	0.04	Pass
60.0°C	≤0.10	0.03	Pass
45.0°C	≤0.10	0.04	Pass

Heated cover accuracy Set point accuracy measurement (°C) of Ch9			
Set point	Limits	Measured value	Pass/Fail
105.0°C	±5.00	102.66	Pass

Appendix 2.10 Service certificate – Genetic analyser

Field Service Report

Thermo Fisher

S C I E N T I F I C

Applied Biosystems South Africa (Pty) Ltd
200 Smit Street
Fairlands 2195
Johannesburg
South Africa

Tel: 27 11 478 0411

Fax: 27 11 478 0349

Customer		SeNiCe order Ngu
Unistel Medical Laboratories		00031105468f
Purchase Order No	Qr.der_lype	Maintenance Activity
	SMOI	PM

Address	Unjstel Medical Laboratories
City	TYGERBERG
, Post Code,	ZA, 7536, ZA Leonora Theart
Contact	2721 938 9215
Tel	Itheart sun.ac.za

Model No	3500	Contract No	0035385200
Serial No	22113-191	Contract Type	AB Assurance 1 PM
Equipment No	10632982	Contract Expiry	09/07/2017
			09/Jul/2017

Order Description

Date	22/Jun/2016	Engineer	Eamon D OKennedy - 4108225

Customer Signature	00/6/16 Date
50Z4'tPtedg Engineer Signature	22/6/16 Date

Customer		Service Order No.
Unistel MedicalLaboratories		000311054681
Purchase Order No	Order Type SMOI	Maintenance Activitv PM

Work Description

21/06/2016 3500 PM

Replaced needed parts

Ran instrument tests all passed

Cleaned optics and ran optic tests all passed

Completed PM as per PM protocol

En ineer Parts Consumption

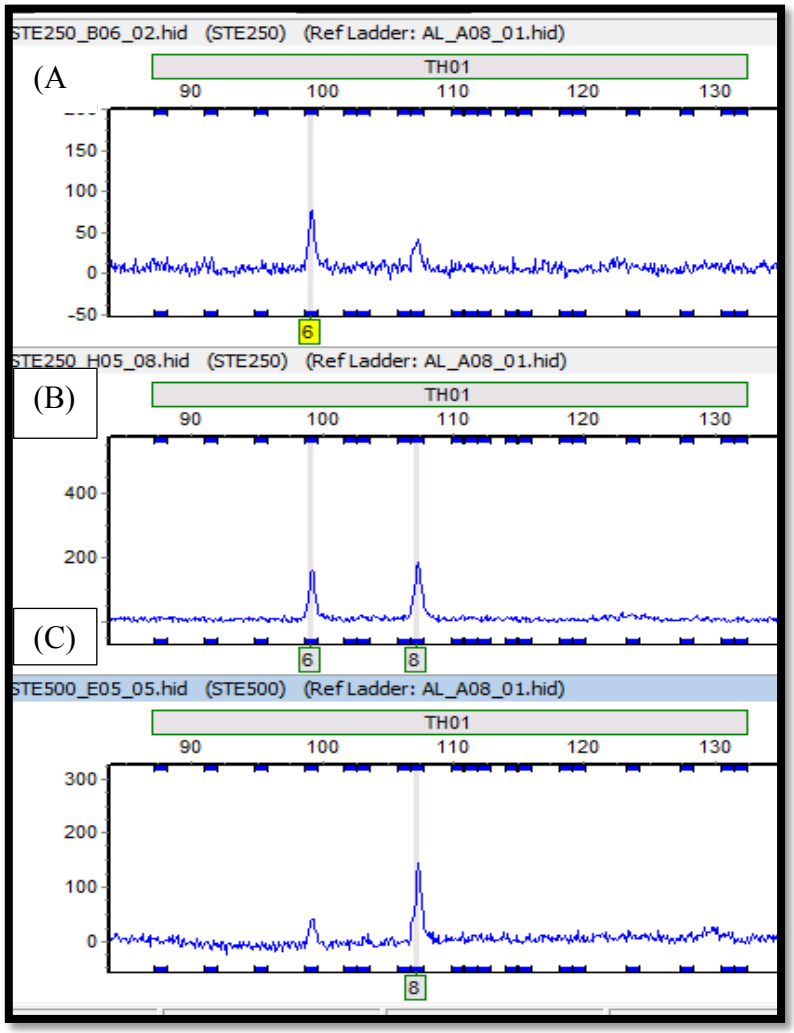
Material	Material Description	Qty
4454928	KITPM 3500 RUO	1
		1

Local Office Parts Consum tion

Material	Material Description	Qty

iv	Start Date	End Date	Hours
PM Labor	21/06/2016	21/06/2016	4
Travel - Non Billable	21/06/2016	21/06/2016	2
			6

Appendix 2.11 Stochastic threshold calculation



Calculation of Stochastic threshold

Here are three different results from a serial dilution of the same sample at STR-marker TH01.Dilutions were set-up in triplicate. (A) At 250 ng/µl allele 6 is observed as a false homozygote. (B) At another 250 ng/µl dilution both alleles are present and (C) at 500 ng/µl allele 8 is observed as a false homozygote. The RFU values of the false homozygotes of this marker were used to calculate the stochastic threshold (Figure obtained from GeneMarker® Software).

Appendix 3.1.1 Allele Frequencies: South African population

Allele	TH01	D3S1358	vWA	D21S11	TPOX	D1S1656	D12S391	SE33	D10S1248	D22S1045	D19S433	D8S1179	D2S1338	D2S441	D18S51	FGA	D16S539	CSF1PO	D13S317	D5S818	D7S820
3																					
4																	0.0008				
4.2																					
5	0.0015																				
6	0.1748				0.031084													0.013			
6.2								0.0008			0.0015										
6.3																					
7	0.2412				0.007581				0.0008		0.016							0.0137	0.0015	0.0015	0.013
7.2								0.0008													
7.3																					0.0015
8	0.2038				0.394238			0.0015		0.0008	0.0038	0.0092					0.0336	0.0092	0.1031	0.0267	0.1595
8.1																				0.0015	
8.3	0.0008																		0.0008		
9	0.1916				0.172858							0.0069			0.0015		0.1573	0.026	0.055	0.0344	0.1336
9.1														0.0015							
9.3	0.1733																				
10	0.013				0.08794	0.0069		0.0023	0.0038	0.0458	0.0053	0.0748		0.1458	0.0099		0.1122	0.2618	0.0481	0.0947	0.2794
10.2								0.0008							0.0053						
10.3											0.0008			0.0008							0.0008
11			0.0031		0.287338	0.0832			0.0252	0.1901	0.0267	0.0626		0.3534	0.0137		0.3176	0.2718	0.3038	0.3015	0.2176
11.2								0.0031											0.0031		
11.3														0.0489							
12		0.0015			0.018196	0.0809		0.0038	0.0672	0.0206	0.1107	0.1221		0.1046	0.1031		0.2122	0.3328	0.3069	0.3359	0.1557
12.1																			0.0008		
12.2								0.0023			0.0198										
12.3														0.0145							
13		0.0015	0.0031			0.0802		0.0092	0.2427	0.0038	0.2718	0.2527		0.0328	0.0947		0.1374	0.0595	0.116	0.1878	0.0351
13.1										0.0008									0.0008		
13.2								0.0023		0.0025	0.0351				0.0015						
13.3						0.0008								0.0053							
14		0.1008	0.1099		0.000758	0.1672		0.0282	0.2916	0.0611	0.2473	0.2435	0.0008	0.2481	0.142		0.0282	0.0115	0.0588	0.016	0.0031
14.1											0.0015										
14.2								0.0046			0.0595										
14.3						0.016					0.0008			0.0031							
15		0.3099	0.1397			0.1748	0.0496	0.0374	0.1847	0.3191	0.1099	0.1618		0.0389	0.1412		0.0008	0.0008	0.0008		0.0008
15.2						0.0008		0.0031			0.0573				0.0061						
15.3						0.0305															
16		0.2901	0.2084			0.1389	0.0496	0.0634	0.1496	0.2397	0.0191	0.0573	0.0511	0.0023	0.1412				0.0008		
16.1																0.0038					
16.2								0.0031			0.0076										
16.3						0.0618		0.0008							0.0008						
17		0.1985	0.2351			0.0511	0.1321	0.0748	0.0328	0.1076	0.0046	0.0076	0.1191		0.116	0.0008					
17.2								0.0008													
17.3						0.0664	0.0115	0.0023													
18		0.0931	0.2031			0.0053	0.2214	0.1031	0.0015	0.0099		0.0008	0.0847		0.0786	0.0107					
18.2			0.0008					0.0023								0.0053					
18.3						0.0275	0.0023	0.0015													
19		0.0038	0.0771			0.0015	0.1405	0.0947				0.0008	0.1473		0.0809	0.0687					
19.1							0.0038														
19.2							0.0008	0.0015								0.0046					
19.3						0.0053	0.0023														
20		0.0008	0.0176				0.1153	0.0763		0.0008			0.1099		0.0427	0.113					
20.1								0.0008													
20.2								0.0061			0.0008					0.0015					
20.3						0.0008															
21			0.0023				0.0908	0.0351					0.0885		0.0122	0.1252					
21.1								0.0015													
21.2								0.0115							0.0008	0.0061					
21.3								0.0015													
22							0.0634	0.0069					0.1092		0.0061	0.1611					
22.2								0.016								0.0076					
23							0.058	0.0015					0.113		0.0015	0.1618					
23.1								0.0008													
23.2								0.0198								0.0008					
24							0.0344						0.0733			0.158					
24.2								0.0267								0.0023					
24.3				0.0008																	
25							0.0145						0.0817			0.0908					
25.2								0.0389								0.0008					
26				0.0038			0.0061						0.0198			0.0489					
26.2								0.0573								0.0008					
27				0.0458			0.0038						0.0015			0.0176					
27.2								0.0634													
28				0.2137												0.0015					
28.2								0.0542													
29				0.1702												0.0023					
29.2				0.0023				0.0534								0.0008					
29.3				0.0015																	
30				0.1702				0.0015								0.0008					
30.2				0.0229				0.0397								0.0008					
31				0.0847																	
31.2				0.0878				0.0198								0.0023					
32				0.0137				0.0023													
32.2				0.0992				0.0069													
33				0.0053				0.0008													
33.1				0.0031																	
33.2				0.0298				0.0038													
34				0.0053				0.0015													
34.1				0.0031																	
34.2				0.0046				0.0015													
35				0.013				0.0008													
35.1				0.0031																	
35.2				0.0015																	
36				0.0076																	
36.1				0.0015																	
36.2																					
37				0.0031				0.0015													
37.1				0.0008																	
37.2				0.0008																	
38				0.0008																	
38.3																0.0008					
39																					
42																					
42.2																					
43.2																0.0008					
HW	0.8573	0.9793	0.9878	<0.0001	0.4034	0.9884	0.0658	1.0000	<0.0001	0.998	<0.0001	0.7862	0.1422	0.0154	0.977	0.9486	0.4169	0.4468	<0.0001	<0.0001	0.9295

HW Hardy-Weinberg (ρ-value)

Not in Hardy-Weinberg equilibrium

Appendix 3.1.2 Allele Frequencies: South African Black African population

Allele	TH01	D3S1358	vWA	D21S11	TPOX	D1S1656	D12S391	SE33	D10S1248	D22S1045	D19S433	D8S1179	D2S1338	D2S441	D18S51	FGA	D16S539	CSF1PO	D13S317	D5S818	D7S820
3																					
4																					
4.2																					
5	0.0029																				
6	0.0785				0.0969													0.0291			
6.2								0.0029													
6.3																					
7	0.3430				0.0142				0.0029		0.0291							0.0494			
7.2								0.0029													
7.3																					0.0058
8	0.3198				0.2678					0.0029	0.0145						0.0291	0.0145	0.0029	0.0872	0.1570
8.1																					
8.3																					
9	0.2035				0.1681												0.1890	0.0378	0.0029	0.0174	0.1395
9.1																					
9.3	0.0378																				
10	0.0145				0.1282	0.0145			0.0145	0.1017	0.0116	0.0029		0.0465	0.0174		0.1395	0.2587	0.0203	0.0872	0.3547
10.2								0.0029							0.0145						
10.3														0.0029							0.0029
11			0.0058		0.3162	0.0552			0.0610	0.2035	0.0640	0.0669		0.3547	0.0029		0.3866	0.2326	0.3372	0.2297	0.2209
11.2								0.0058													
11.3														0.0203							
12		0.0029			0.0085	0.0349			0.1337	0.0378	0.1424	0.0988		0.1890	0.0262		0.1453	0.3372	0.3692	0.3081	0.0988
12.1																					
12.2											0.0465										
12.3														0.0233							
13			0.0058			0.0959		0.0087	0.2384		0.2733	0.2238		0.0320	0.0407		0.0843	0.0407	0.1512	0.2587	0.0203
13.1																					
13.2								0.0058			0.0494										
13.3						0.0029								0.0145							
14		0.0988	0.0872			0.2703		0.0523	0.3140	0.0727	0.1977	0.3052	0.0029	0.2616	0.0610		0.0262		0.1163	0.0116	
14.1																					
14.2								0.0087			0.0727										
14.3						0.0233								0.0058							
15		0.3081	0.2326			0.1977	0.0988	0.0640	0.1192	0.2326	0.0581	0.2151		0.0465	0.1366						
15.2								0.0058			0.0320				0.0087						
15.3						0.0029															
16		0.3605	0.2209			0.1628	0.0756	0.1163	0.1047	0.1715	0.0087	0.0814	0.0814	0.0029	0.1715						
16.1																0.0116					
16.2								0.0029													
16.3						0.0756									0.0029						
17		0.1686	0.1890			0.0407	0.1308	0.0872	0.0116	0.1570		0.0058	0.0523		0.1453						
17.2								0.0029													
17.3						0.0029		0.0058													
18		0.0610	0.1657			0.0058	0.2267	0.1221		0.0203			0.0407		0.1337	0.0058					
18.2								0.0058								0.0116					
18.3						0.0116															
19			0.0814				0.1657	0.1337					0.1773		0.1424	0.0669					
19.1							0.0087														
19.2																0.0116					
19.3						0.0029															
20			0.0087				0.1512	0.1076					0.0843		0.0610	0.0872					
20.1																					
20.2								0.0087								0.0029					
20.3																					
21			0.0029				0.0407	0.0233					0.1628		0.0262	0.0785					
21.1																					
21.2								0.0087							0.0029						
21.3								0.0058													
22							0.0174	0.0058					0.1686		0.0058	0.1541					
22.2								0.0029								0.0029					
23							0.0407	0.0029					0.1134			0.1744					
23.1																					
23.2																					
24							0.0174						0.0349			0.1802					
24.2								0.0087													
24.3				0.0029																	
25							0.0174						0.0494			0.1192					
25.2								0.0291													
26				0.0116			0.0058						0.0291			0.0465					
26.2								0.0552													
27				0.0901			0.0029						0.0029			0.0349					
27.2								0.0436													
28				0.2558												0.0029					
28.2								0.0291													
29				0.1076																	
29.2								0.0145								0.0029					
29.3																					
30				0.0959																	
30.2				0.0116				0.0145								0.0029					
31				0.1017																	
31.2				0.1076																	
32				0.0058																	
32.2				0.0581																	
33				0.0145																	
33.1				0.0087																	
33.2				0.0349																	
34				0.0087																	
34.1				0.0058																	
34.2				0.0087																	
35				0.0349				0.0029													
35.1				0.0029																	
35.2				0.0058																	
36				0.0233																	
36.1																					
36.2																					
37				0.0029																	
37.1																					
37.2																					
38																					
38.3																					
39																					
42																					
42.2																					
43.2																0.0029					
HW	0.6541	0.7803	0.9736	0.9978	0.4823	0.9998	0.8839	0.3767	0.0119	0.8794	0.814	0.1735	0.243	0.2563	0.811	0.3697	0.474	0.0388	0.0001	0.1922	0.8901

HW Hardy-Weinberg (ρ-value)

Not in Hardy-Weinberg equilibrium

Appendix 3.1.3 Allele Frequencies: South African Coloured population

Allele	TH01	D3S1358	vWA	D21S11	TPOX	D1S1656	D12S391	SE33	D10S1248	D22S1045	D19S433	D8S1179	D2S1338	D2S441	D18S51	FGA	D16S539	CSF1PO	D13S317	D5S818	D7S820
3																					
4																	0.0026				
4.2																					
5																					
6	0.1615				0.0128													0.0179			
6.2																					
6.3																					
7	0.2718				0.0051						0.0282								0.0026		0.0205
7.2																					
7.3																					
8	0.2128				0.3724			0.0026				0.0077					0.0256	0.0103	0.1154	0.0128	0.1821
8.1																				0.0051	
8.3	0.0026																		0.0026		
9	0.1821				0.2500							0.0128			0.0026		0.1872	0.0308	0.0564	0.0256	0.1256
9.1														0.0026							
9.3	0.1590																				
10	0.0103				0.0740	0.0026		0.0077		0.0615	0.0026	0.0641		0.1462	0.0128		0.1205	0.2769	0.0667	0.1179	0.2231
10.2															0.0051						
10.3											0.0026										
11			0.0026		0.2704	0.0974			0.0231	0.2154	0.0231	0.0667		0.3795	0.0051		0.3128	0.2641	0.2872	0.2949	0.2154
11.2								0.0051											0.0026		
11.3														0.0692							
12					0.0153	0.0615		0.0103	0.0641	0.0205	0.1103	0.1333		0.1179	0.1179		0.1795	0.3333	0.3026	0.3128	0.1718
12.1																			0.0026		
12.2								0.0077			0.0205										
12.3														0.0205							
13		0.0026				0.0744		0.0051	0.2590	0.0051	0.2897	0.2205		0.0282	0.1026		0.1436	0.0590	0.1205	0.2103	0.0564
13.1																					
13.2								0.0026			0.0436										
13.3														0.0051							
14		0.0897	0.1205			0.1564		0.0256	0.2769	0.0769	0.2179	0.2308		0.2026	0.1436		0.0282	0.0077	0.0410	0.0205	0.0026
14.1											0.0051										
14.2								0.0026			0.0692										
14.3						0.0256								0.0051							
15		0.3436	0.1308			0.2077	0.0436	0.0231	0.2000	0.3128	0.0923	0.1872		0.0205	0.1282						0.0026
15.2								0.0051			0.0692				0.0128						
15.3						0.0333															
16		0.2872	0.2462			0.1718	0.0487	0.0513	0.1410	0.2000	0.0077	0.0615	0.0308	0.0026	0.1436						
16.1																0.0026					
16.2								0.0051			0.0154										
16.3						0.0513															
17		0.1897	0.2256			0.0513	0.1231	0.0795	0.0333	0.1000	0.0026	0.0128	0.1333		0.1000	0.0026					
17.2																					
17.3						0.0359	0.0103														
18		0.0795	0.1872			0.0051	0.2333	0.1026	0.0026	0.0051			0.1026		0.0744	0.0103					
18.2			0.0026													0.0077					
18.3						0.0179															
19		0.0077	0.0615				0.1385	0.0769				0.0026	0.1667		0.0769	0.0718					
19.1							0.0051														
19.2							0.0026									0.0051					
19.3						0.0051	0.0026														
20			0.0205				0.1205	0.0872		0.0026			0.1103		0.0590	0.1103					
20.1																					
20.2								0.0026													
20.3						0.0026															
21			0.0026				0.0795	0.0385					0.0923		0.0077	0.1179					
21.1								0.0051													
21.2								0.0077								0.0077					
21.3																					
22							0.0692	0.0051					0.1077		0.0051	0.1564					
22.2								0.0231								0.0103					
23							0.0590						0.1051		0.0026	0.1641					
23.1																					
23.2								0.0179								0.0026					
24							0.0410						0.0615			0.1769					
24.2								0.0256								0.0026					
24.3																					
25							0.0077						0.0821			0.0692					
25.2								0.0385													
26				0.0026			0.0077						0.0077			0.0538					
26.2								0.0667								0.0026					
27				0.0385			0.0077									0.0128					
27.2								0.0718													
28				0.2179																	
28.2								0.0462													
29				0.1667												0.0077					
29.2				0.0026				0.0590													
29.3				0.0051																	
30				0.1718				0.0026								0.0026					
30.2				0.0231				0.0410													
31				0.1026																	
31.2				0.0692				0.0179								0.0026					
32				0.0205				0.0026													
32.2				0.1000				0.0179													
33																					
33.1				0.0026																	
33.2				0.0205				0.0051													
34				0.0103				0.0026													
34.1				0.0026																	
34.2				0.0051				0.0026													
35				0.0051																	
35.1				0.0077																	
35.2																					
36				0.0051																	
36.1				0.0051																	
36.2																					
37				0.0077																	
37.1				0.0026																	
37.2				0.0026																	
38				0.0026																	
38.3																					
39																					
42																					
42.2																					
43.2																					
HW	0.291	0.5516	0.8595	< 0.0001	0.7473	0.1221	0.0219	< 0.0001	0.7255	0.9982	< 0.0001	0.9624	0.3389	0.5737	0.9861	0.0005	0.5849	0.9978	< 0.0001	< 0.0001	0.9371

HW Hardy-Weinberg (ρ-value)

Not in Hardy-Weinberg equilibrium

Appendix 3.1.4 Allele Frequencies: South African Indian/Asian population

Allele	TH01	D3S1358	vWA	D21S11	TPOX	D1S1656	D12S391	SE33	D10S1248	D22S1045	D19S433	D8S1179	D2S1338	D2S441	D18S51	FGA	D16S539	CSF1PO	D13S317	D5S818	D7S820
3																					
4																					
4.2																					
5																					
6	0.2216																				
6.2																					
6.3																					
7	0.1534				0.0057														0.0057	0.0057	0.0341
7.2																					
7.3																					
8	0.2102				0.4261			0.0057				0.0114					0.0795	0.0057	0.1705		0.1534
8.1																					
8.3																					
9	0.2557				0.1307										0.0057		0.1080	0.0114	0.1080	0.0227	0.0795
9.1														0.0057							
9.3	0.1477																				
10	0.0114				0.0682	0.0170					0.0057	0.1761		0.3182	0.0114		0.1193	0.2159	0.0568	0.1250	0.2670
10.2																					
10.3																					
11			0.0057		0.3295	0.1307			0.0057	0.2557	0.0057	0.0398		0.3409	0.0170		0.2898	0.2216	0.2784	0.3750	0.2273
11.2																					
11.3														0.0398							
12					0.0341	0.0795			0.0227		0.0739	0.0966		0.0625	0.1023		0.2330	0.4148	0.2727	0.3295	0.2102
12.1																					
12.2											0.0057										
12.3																					
13		0.0057	0.0057			0.1250		0.0057	0.1364	0.0057	0.3239	0.1591		0.0170	0.1250		0.1420	0.1080	0.0682	0.1250	0.0227
13.1																					
13.2											0.0341										
13.3																					
14		0.1648	0.1648		0.0057	0.1875			0.2614	0.0398	0.1875	0.2557		0.1932	0.2102		0.0284	0.0227	0.0284	0.0170	0.0057
14.1																					
14.2											0.0852										
14.3																					
15		0.3466	0.0568			0.1705	0.0170	0.0284	0.2386	0.3864	0.1250	0.1989		0.0227	0.1875				0.0057		
15.2											0.1080										
15.3						0.0170															
16		0.2670	0.1705			0.1477	0.0057	0.0398	0.2727	0.2159	0.0170	0.0511	0.0341		0.1250				0.0057		
16.1																					
16.2											0.0114										
16.3						0.0341		0.0057													
17		0.2159	0.2727			0.0682	0.1364	0.0511	0.0625	0.0795	0.0114	0.0057	0.0511		0.1023						
17.2																					
17.3						0.0057	0.0057	0.0057													
18		0.0909	0.2159				0.2784	0.0909		0.0170		0.0057	0.0852		0.0455						
18.2								0.0057													
18.3						0.0057															
19		0.0057	0.0739			0.0114	0.1534	0.0795					0.0966		0.0341	0.0511					
19.1																					
19.2																					
19.3																					
20			0.0341				0.1136	0.0511					0.1307		0.0227	0.1193					
20.1								0.0057													
20.2								0.0057			0.0057										
20.3																					
21							0.0852	0.0568					0.0511		0.0114	0.1250					
21.1																					
21.2																0.0170					
21.3																					
22							0.0909	0.0057					0.1420			0.1534					
22.2								0.0170								0.0057					
23							0.0568						0.1648			0.1250					
23.1								0.0057													
23.2								0.0341													
24							0.0511						0.1080			0.1932					
24.2								0.0739								0.0057					
24.3																					
25							0.0057						0.1080			0.1364					
25.2								0.0568								0.0057					
26													0.0284			0.0511					
26.2								0.0625													
27				0.0057												0.0114					
27.2								0.0455													
28				0.1420																	
28.2								0.0398													
29				0.1818																	
29.2				0.0114				0.0682													
29.3																					
30				0.2330																	
30.2				0.0227				0.0682													
31				0.0341																	
31.2				0.0909				0.0511													
32				0.0057																	
32.2				0.2102				0.0114													
33																					
33.1																					
33.2				0.0568				0.0170													
34																					
34.1																					
34.2				0.0057				0.0057													
35																					
35.1																					
35.2																					
36																					
36.1																					
36.2																					
37																					
37.1																					
37.2																					
38																					
38.3																					
39																					
42																					
42.2																					
43.2																					
HW	0.7607	0.96	0.8329	0.9988	0.011	0.8784	0.8248	0.7306	0.7908	0.6185	0.0015	0.7033	0.4755	0.2783	0.9999	0.9551	0.0473	0.9198	0.03	0.0002	0.6876

HW Hardy-Weinberg (ρ-value)

Not in Hardy-Weinberg equilibrium

Appendix 3.1.5 Allele Frequencies: South African White population

Allele	TH01	D3S1358	vWA	D21S11	TPOX	D1S1656	D12S391	SE33	D10S1248	D22S1045	D19S433	D8S1179	D2S1338	D2S441	D18S51	FGA	D16S539	CSF1PO	D13S317	D5S818	D7S820
3																					
4																					
4.2																					
5	0.0025																				
6	0.2500				0.0050																
6.2											0.0050										
6.3																					
7	0.1625				0.0050													0.0025		0.0025	0.0075
7.2																					
7.3																					
8	0.0925				0.5125							0.0175					0.0250	0.0050	0.1475		0.1425
8.1																					
8.3																					
9	0.1625				0.1200							0.0100					0.1225	0.0175	0.0750	0.0625	0.1600
9.1																					
9.3	0.3150																				
10	0.0150				0.0750					0.0025	0.0025	0.1025		0.1550			0.0775	0.2700	0.0500	0.0650	0.2750
10.2																					
10.3																					
11					0.2600	0.0725			0.0050	0.1250	0.0075	0.0650		0.3325	0.0300		0.2750	0.3350	0.3025	0.3375	0.2125
11.2																			0.0075		
11.3														0.0575							
12		0.0025			0.0225	0.1400		0.0025	0.0325	0.0150	0.1000	0.1425		0.0375	0.1550		0.2925	0.2925	0.2725	0.3850	0.1650
12.1																					
12.2											0.0025										
12.3														0.0075							
13			0.0025			0.0525		0.0150	0.2775	0.0050	0.2300	0.3500		0.0450	0.1200		0.1750	0.0550	0.1025	0.1325	0.0325
13.1																			0.0025		
13.2										0.0025	0.0150				0.0050						
13.3																					
14		0.1275	0.0950			0.0800		0.0225	0.3000	0.0450	0.3450	0.1975		0.3050	0.1800		0.0300	0.0200	0.0400	0.0150	0.0050
14.1																					
14.2								0.0050			0.0275										
14.3						0.0075					0.0025										
15		0.2625	0.1050			0.1250	0.0275	0.0325	0.2025	0.3700	0.1650	0.0750		0.0575	0.1375		0.0025	0.0025			
15.2						0.0025					0.0450										
15.3						0.0575															
16		0.2425	0.1775			0.0825	0.0475	0.0400	0.1425	0.3475	0.0400	0.0350	0.0525	0.0025	0.1200						
16.1																					
16.2								0.0025			0.0050										
16.3						0.0725															
17		0.2250	0.2675			0.0525	0.1400	0.0700	0.0375	0.0850	0.0075	0.0050	0.1925		0.1125						
17.2																					
17.3						0.1775	0.0250														
18		0.1350	0.2450			0.0075	0.1800	0.0925	0.0025	0.0025			0.1050		0.0500	0.0200					
18.2																					
18.3						0.0600	0.0075	0.0050													
19		0.0025	0.0900				0.1150	0.0850					0.1250		0.0525	0.0750					
19.1																					
19.2								0.0050													
19.3						0.0100	0.0050														
20		0.0025	0.0150				0.0800	0.0500					0.1225		0.0200	0.1350					
20.1																					
20.2								0.0075								0.0025					
20.3																					
21			0.0025				0.1475	0.0325					0.0375		0.0050	0.1725					
21.1																					
21.2								0.0225								0.0050					
21.3																					
22							0.0850	0.0100					0.0450		0.0100	0.1750					
22.2								0.0200								0.0100					
23							0.0725	0.0025					0.0975		0.0025	0.1650					
23.1																					
23.2								0.0325													
24							0.0350						0.1025			0.1050					
24.2								0.0225								0.0025					
24.3																					
25							0.0225						0.0975			0.0675					
25.2								0.0400													
26							0.0075						0.0200			0.0450					
26.2								0.0475													
27				0.0325			0.0025						0.0025			0.0100					
27.2								0.0800													
28				0.2050												0.0025					
28.2								0.0900													
29				0.2225																	
29.2								0.0750													
29.3																					
30				0.2050				0.0025													
30.2				0.0325				0.0475													
31				0.0750																	
31.2				0.0875				0.0250								0.0050					
32				0.0175				0.0050													
32.2				0.0850																	
33				0.0050				0.0025													
33.1																					
33.2				0.0225																	
34								0.0025													
34.1				0.0025																	
34.2																					
35				0.0075																	
35.1																					
35.2																					
36																					
36.1																					
36.2																					
37								0.0050													
37.1																					
37.2																					
38																					
38.3																0.0025					
39																					
42																					
42.2																					
43.2																					
HW	0.5688	0.9634	0.3895	0.7191	0.946	0.9982	0.7667	0.9992	0.9854	0.9966	0.0002	0.8029	0.6413	< 0.0001	0.0463	0.2632	0.3221	0.7664	< 0.0001	0.2771	0.972

HW Hardy-Weinberg (ρ-value)

Not in Hardy-Weinberg equilibrium

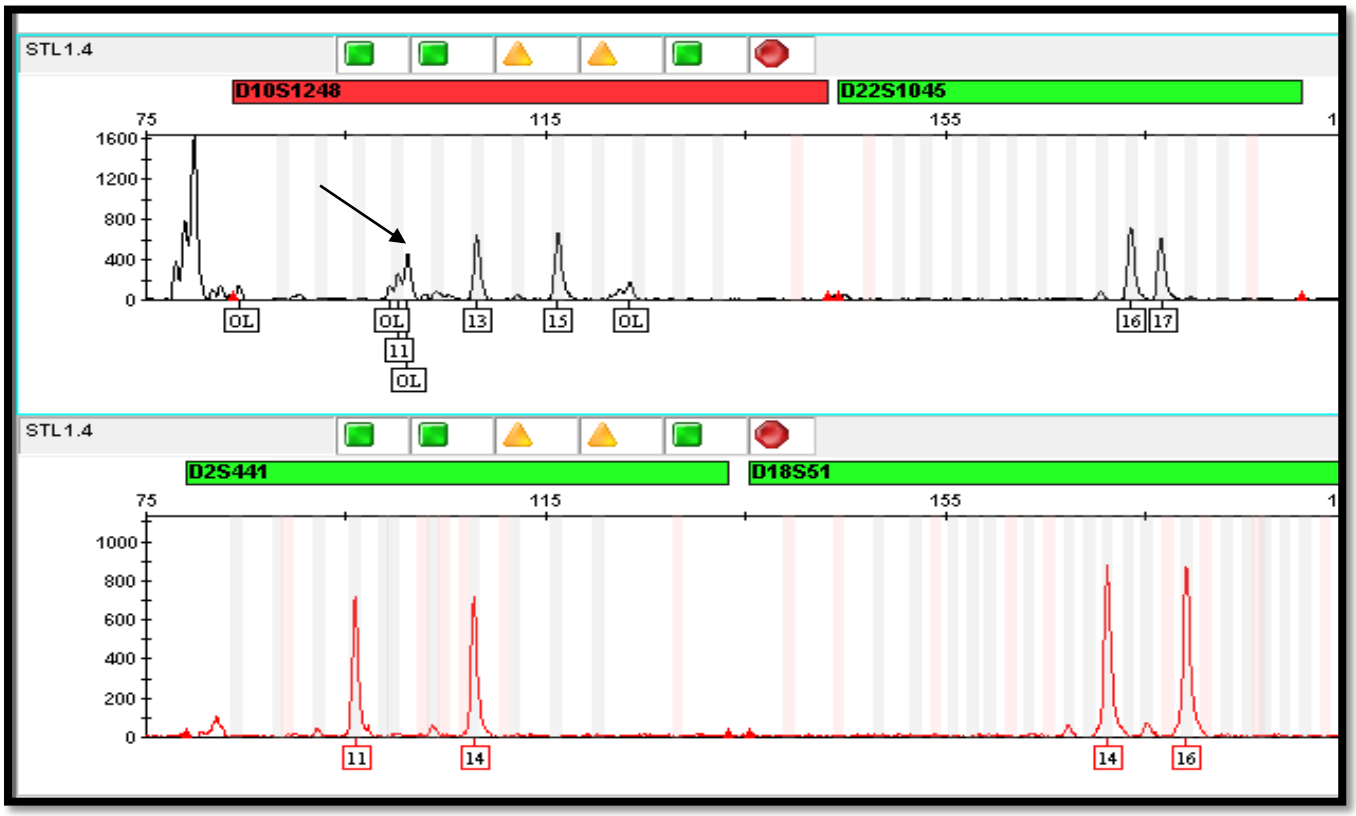
Appendix 3.2 Statistical comparison of allele frequencies between Nel and Lucassen

Table 3.3 Statistical comparison of allele frequencies between Nel and Lucassen

The ρ-values from the F-exact/Chi-square tests are portrayed in the table below. An asterisk indicates a significant difference (ρ < 0.005) of alleles in a marker between Nel and Lucassen. The Chi-square results are indicated with †.

	Black African	**Coloured**	**Indian/Asian**	**White**
TH01	0.112	0.250	0.066	0.489
D3S1358	0.227	0.977	0.518	0.286
vWA	0.839	0.243	0.518	0.048*
D21S11 †	0.248	0.058	0.961	0.799
TPOX	0.348	0.372	0.083	0.438
D19S433 †	0.287	0.263	0.019*	0.107
D8S1179	0.312	0.996	0.060	0.957
D2S1338 †	0.312	0.133	0.013	0.106
D18S51 †	0.912	0.142	0.574	0.011
FGA †	0.190	0.451	0.141	0.493
D16S539	0.106	0.114	0.722	0.275
CSF1PO	0.674	0.053	0.222	0.323
D13S317	0.161	0.600	0.289	0.140
D5S818	0.906	0.293	0.593	0.175
D7S820	0.304	0.681	0.341	0.447

Appendix 4.1 Artefact at STR marker D10S1248



Artefact at STR-marker D10S1248

At the region of allele 11 is a non-specific artefact at the D10S1248 marker; indicated with an arrow. This is a 1:4 dilution used during serial dilution for calculation of stochastic threshold for buccal lysates (The figure above was obtained from GeneMapper® Software).