Gadean Brecht **December 2019**
3689509

Genetic analysis of mitochondrial DNA within Southern African populations.

UNIVERSITY *of the*
WESTERN CAPE

A thesis submitted in fulfilment of the requirements of Magister Scientiae in the Department of Biotechnology, University of the Western Cape.

Supervisor: Prof. Maria Eugenia D'Amato
Forensic DNA Laboratory, Department of Biotechnology,

University of the Western Cape.

Co Supervisor: Dr Peter Ristow

Keywords (10)

Southern African populations
Population genetics
Mitochondrial DNA (mtDNA)
Control region
Single Nucleotide Polymorphisms (SNPs)
Haplogroups
High Resolution Melting (HRM)
Ancestral testing
Diagnostic multiplex
Melting temperature (Tm)

Abstract

As human beings we are curious about our origin and ancestry. A curiosity has led to an investigation of human evolution and expansion across the world by means of population genetics and phylo-genetics by evaluating a region in Southern Africa that is largely unknown.

The objective of this study was to develop a quick, inexpensive and accurate hierarchical diagnostic screening system of the MtDNA phylogenetic tree, AI-SNPs in the mtDNA genome by using High Resolution Melting analysis to evaluate the population composition and ancestral haplogroups of Southern African populations in Limpopo.

The admixture between the 'Khoesan' hunter-gatherers, herders and the Bantu speaking populations led to population growth and expansion in Limpopo. This has contributed to populations settling in Limpopo and has thus shaped the ancestral contemporary populations. No research on these individuals residing in Limpopo has been done before, thus an investigation of their ancestral origin was necessary.

A total of 760 saliva samples were collected from individuals residing in Limpopo. Only 500 saliva samples were extracted by means of an optimized salting out technique. Five hundred extracted genomic samples were genotyped by means of a quick, inexpensive High-resolution melting analysis.

Of the 500 samples, the genotyping results showed 95 individuals derived for the L3 haplogroup which gives a 19% ratio of individuals screened with Multiplex 1. Only 56 individuals were derived for the L1 haplogroup, which gives a percentage of 11%. A total of 249 individuals were derived for the L0 haplogroup, making up a 50% of the total individuals genotyped. Only 100 samples were derived for L0a, making up 20% of individuals screened with Multiplex 1. Of the 95 samples derived for the L3 haplogroup, the results showed 87 individuals to be ancestral for both M and N, making up 91.57% of individuals screened with Multiplex 2.

In population genetics using SNPs to infer population history and ancestral origin has become significant, this study allowed researchers to evaluate population groups by investigating their genetic markers and the application of the results allowed for downstream analyses. Finally, this study provides a quick and simple screening method for the selection of lineages that are of interest for further studies.

UNIVERSITY of the
WESTERN CAPE

Declaration
University of the Western Cape

Private Bag X17, Bellville 7535, South Africa

Email: **3689509@myuwc.ac.za**

FACULTY OF NATURAL SCIENCE

I declare that the MSc thesis titled "*Genetic analysis of mitochondrial DNA within Southern African populations*" is my own work.

Name: Gadean Brecht

Student number: 3689509

I hereby declare that I know what plagiarism entails, namely, to use another's work and to present it as my own without attributing the sources in the correct way.

1. I know that plagiarism is a punishable offence because it constitutes theft.
2. I understand the plagiarism policy of the Faculty of Natural Science of the University of the Western Cape.
3. I know what the consequences will be if I plagiarise in any of the assignments for my course.
4. I declare that all work presented by me for every aspect of my course, will be my own, and where I have made use of another's work, I will attribute the source in the correct way.

Signature Date

G Brecht December 2019

Acknowledgements

I would like to acknowledge my parents, family and friends. This has been one of the most challenging tasks I had to complete, but your support has meant so much through this chapter in my life.

My parents, thank you for always encouraging me to do what I love and teaching me to never give up on my dreams and goals. I am where I am because of your sacrifices and your lessons. I thank you and I love you.

My family and friends, thank you for always being proud of my achievements and the work I have put in, for motivating me every step of the way. This degree is dedicated to each family member. I want to encourage all my younger family members wanting to study further and wanting to achieve and make something of themselves. Take every opportunity given and never give up!

Thank you to the FDL (Forensic DNA Laboratory) group of forensic scientists. The dynamic teamwork has contributed to a project that forms part of a greater goal and achievement within the Science community. Our brain storming sessions, interactive meetings and presentations were a learning experience. Wishing every FDL member the very best with their future endeavours.

I would like to thank my Co Supervisor, Dr Peter Ristow. Thank you for your help and support in every aspect. Thank you to my Supervisor, Prof M.E. D'Amato for giving me the opportunity and allowing me to do this project.

Lastly, I would like to thank my funders, the CPRR grant for funding this project and the NRF (National Research Foundation) for the financial assistance as a student and a special thank you to the voluntary donors who contributed their DNA to this study.

List of abbreviations

mtDNA- mitochondrial DNA

DNA- Deoxyribonucleic acid

HRM- High resolution melting

AI-SNP- Ancestry informative- single nucleotide polymorphism

kya- thousand years ago

STR- Short tandem repeat

INDEL- Insertion and deletion

MRCA- Most recent maternal common ancestor

ybp- years before present

rCRS- Cambridge reference sequence

PCR- Polymerase chain reaction

kb- kilo base

ml- millilitre

SDS- Sodium dodecyl sulphate

EDTA- Ethylenediamine tetra-acetic acid

pH- potential Hydrogen

tris-CL- Hydroxymethyl aminomethane hydrochloride

rpm- revolutions per minute

EtOH- Ethyl alcohol

$MgCL_2$- Magnesium chloride

dH_20- distilled water

°C- degrees Celsius

µl- microlitre

bp- base pair

µM- micromolar

ng- nanogram

AMH- anatomically modern humans

SSR- simple sequence repeat

VNTR- variable number tandem repeat

SNV- single nucleotide variant

RFLP- restriction fragment length polymorphism

NGS- next generation sequencing

AMOVA- analysis of molecular variance

Table of contents

List of figures:

List of tables:

Chapter 1

Literature review

1.1 Introduction

As human beings we are curious about our origin and ancestry. A curiosity has led to an investigation of human evolution and expansion across the world by means of population genetics and phylo-genetics by evaluating a region in Southern Africa that is largely unknown. The term "population genetics" refers to bringing an evolutionary change between populations (Relethford and Harding, 2001). Any scientific investigation would be based on evidence and a working hypothesis. A brief history on modern human origins and migrations will be introduced, genetics and SNP testing methods of which a quick, inexpensive screening method for African haplogroups is proposed.

Among various researchers, Relethford and Harding's (2001) hypotheses reported on human evolution and population genetics in a study focused on mutations, gene flow, genetic drift and natural selection. All these factors would ultimately influence the genetic diversity within and between human populations. This would question how the modern human species has evolved over the years and how evolutionary, demographic and social factors have significantly played a role in shaping our diversity.

By thus investigating the evolutionary changes within populations, we can apply the evidence (our results) and knowledge (literature) to test our research question, whether our diagnostic multiplex system can be used as a quick, inexpensive and accurate screening method to infer individuals to their haplogroups.

1.1.1 The origin and evolution of humans in Africa

The earliest fossil remains of modern humans originated from Ethiopia around 190 000 kya (White *et al*., 2003). These fossils have displayed anatomical features from Morocco which dates to 315 000 kya. The remains from Morocco now pushes the timeline back by 100,000 kya, suggesting that the modern humans did not only evolve in East Africa (Callaway E., 2017) but Ethiopia as well. This indicates that modern humans existed for a long period in Africa before the migration across the globe (Brown *et al*., 2009, 2012, Marean 2010).

The origin of modern humans has been a heavily debated topic and it was Cann *et al*., 1987 who first postulated an Africa origin, around 200 000 kya. Subsequent research has suggested that humans originated in Africa and have lived on the African continent longer than any other ever since (Campbell and Tishkoff, 2010). The exact origin within Africa is still debated by researchers who suggest that the human species originated from South Africa (Tishkoff *et al.*, 2009, Compton, 2011, Henn *et al.*, 2011), East Africa (Prugnolle *et al.*, 2005, Ray *et al.*, 2005) or more recently Botswana (Chan *et al*. 2019).

The most current fossil and chronological evidence seems to agree with DNA evidence. Using DNA as genetic evidence in modern human evolution has played a significant role in Southern African contemporary populations (Chen *et al*., 1995, Ingman *et al*., 2000, Lombard *et al*., 2013).

Multi-regional hypothesis was that a human ancestor dispersed throughout the world and that the modern humans evolved from this predecessor at various times and locations in the world (Gugliotta, 2008). Even though the mode of evolution might still be unclear, it could be said that modern humans did not suddenly emerge but rather continuously from archaic *H.erectus* to modern *H. sapiens* in a gradual morphological way (Callaway E., 2017 and Hublin *et al.*, 2017).

1.1.2 History on human migration in Africa

The migration in Africa involved various time points across different geographic areas (Campbell and Tishkoff 2010). Migration at around >20 kya involved the ancestors of the extant African hunter-gatherer populations and the Khoesan speakers, migrating into central, eastern and Southern Africa (Nurse, 1997, Ehret, 2002) as seen on Figure 1.1. However, Chan *et al*. (2019) proposed that the Khoesan and Khoe-san ancestral distribution diverged at 69 kya southwest of the Zambezi River.

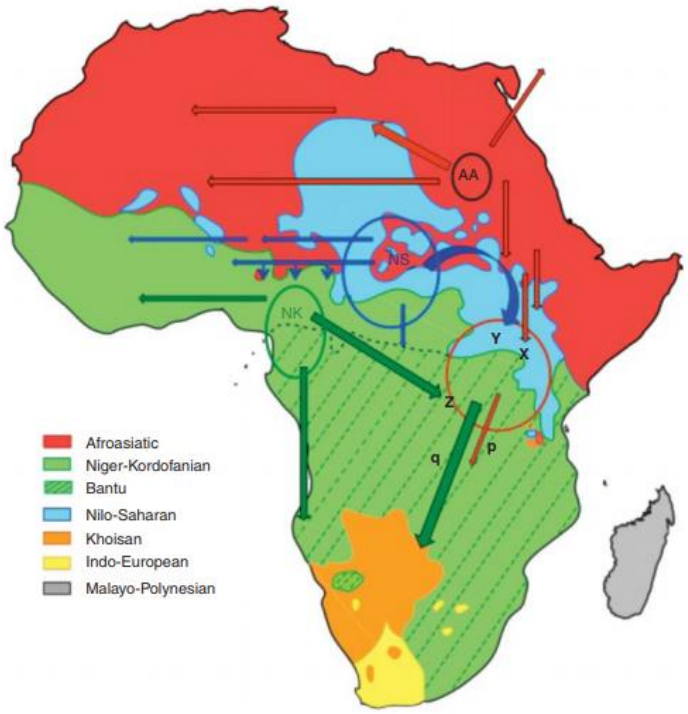


Figure 1. 1 Image taken from Gomez *et al*., (2014) Showing the geographic distribution of the major linguistic groups in Africa. AA: Afro-asiatic (Ehret 1995), NS: Nilo-Saharan (Blench 1993, Ehret 1993, Blench 2006), NK: Niger-Kordofanian (Nurse 1997, Ehret 2001) (Gomez *et al*., 2014).

The archaeological data suggests that this migration contributed to the genetic landscape in Africa, due to the expansion from Central-West Africa by agriculturalists and the migration of pastoralists' populations from North-eastern Africa (Gomez *et al.*, 2014). The following populations migrated in Africa according to Gomez *et al*., (2014).

The Afro-Asiatic-Agro pastoralists were the first to migrate to Ethiopia from the Nile Valley (8000-5000 years) (Ehret 1995, 1998). Followed by the Nilo-Saharan pastoralists who migrated westwards across Sahel (7000 years) and eastwards into Kenya and Tanzania (3000 years) from Chad/Sudan (Blench 1993, Ehret 1993).

The Bantu speakers (Niger-Kordofanian) migrated from Cameroon/Nigeria across sub-Saharan Africa at 5000 years (Figure 1.1) (Nurse 1997, Ehret 2001). Lastly, the migration of Southwestern Asians north-and eastwards (3000 years) into Africa (Phillipson, 2009) occurred. These migrations have allowed for a network of genetic variation to form within the African human genome.

1.1.3 History on human migration out of Africa

Researchers conclude a definite exit from Africa, arriving in India, China, Central Asia, Indochina, Sunda, Sahul and then America (Oppenheimer, 2012). Most of the archaeological and genetic models estimates an exit over 40 kya (Oppenheimer, 2012).

In the geographical setting, researchers propose a single southern dispersal out of Africa by African ancestry for anatomically modern humans (AMH) (Oppenheimer, 2012). They moved from the Red Sea towards the coastline of the Indian Ocean to Bali, eventually moving in a southwest route to Melanesia and Australia as seen in Figure 1.2. However, the morpho-genetic models argued two main exit routes:

Firstly a southern route from the Red Sea (50-45kya) along the coastline of South Asia towards Australia (Mellars 2006, Stanyon *et al*., 2009, Quintana-Murci 1999, where small populations such as Andaman islanders, Southeast Asian Negritos, Melanesians and Australians were (Underhill *et al*., 2001, Armitage *et al*., 2011). Genetic traces of Japan and North America was present. This was known as the Australian wave.

The Asian wave occurred secondly, a north-eastern route with Eurasian populations that moved from Africa to Levant (45kya), dispersing west towards Europe and east to Eurasia (40-20 kya) (Underhill *et al*., 2001)

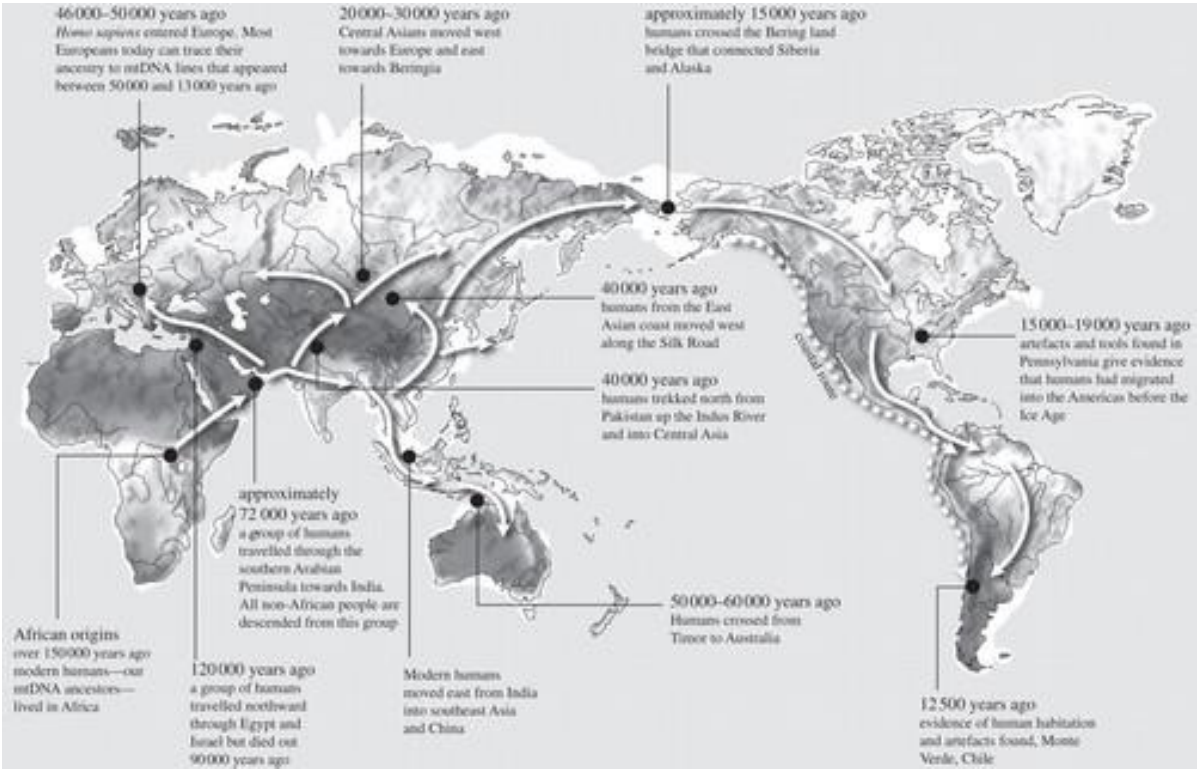


Figure 1. 2 Image taken from Oppenheimer (2012) showing the dispersal of humans out of Africa.

All the non-African groups that exist today, is because of the out-of-Africa migration, except for 7% of these groups derived from admixture with archaic non-African groups (Oppenheimer, 2012). Both Europe and Asia could have been colonized from the same African exit group (Oppenheimer, 2012). The L3-derived (M and N) diverged at 60-65 kya, giving rise to all non-Africans (Forster *et al*., 2001, Kong *et al*., 2003, Mishmar *et al*., 2003, Macaulay *et al*., 2005).

However, long-distance gene flow through the Atlantic slave trade in the 16th century, influenced Brazilians, who harbours the most important African maternal lineages outside of Africa (Rosa and Brehm, 2011). L1c and L3e lineages reported as nearly half of the African lineages (Alves-Silva *et al*., 2000). Studies have shown Afro-Americans residing in South America having 65% Central African origin.

1.1.4 Population structure in Southern Africa

The mtDNA population structure in Southern Africa, the earliest diverging maternal lineage is largely found in Khoesan inhabitants (Gonder *et al*., 2007, Tishkoff *et al*., 2007, Behar *et al*., 2008, Barbieri *et al*., 2013, Schlebusch *et al*., 2013, Petersen *et al*., 2013, Chan *et al*., 2019). The chronological order of historical migrations into Southern Africa can be seen on Figure 1.3. The first individuals to arrive in Southern Africa were the Proto-Nguni arrivals at 1000 ybp.

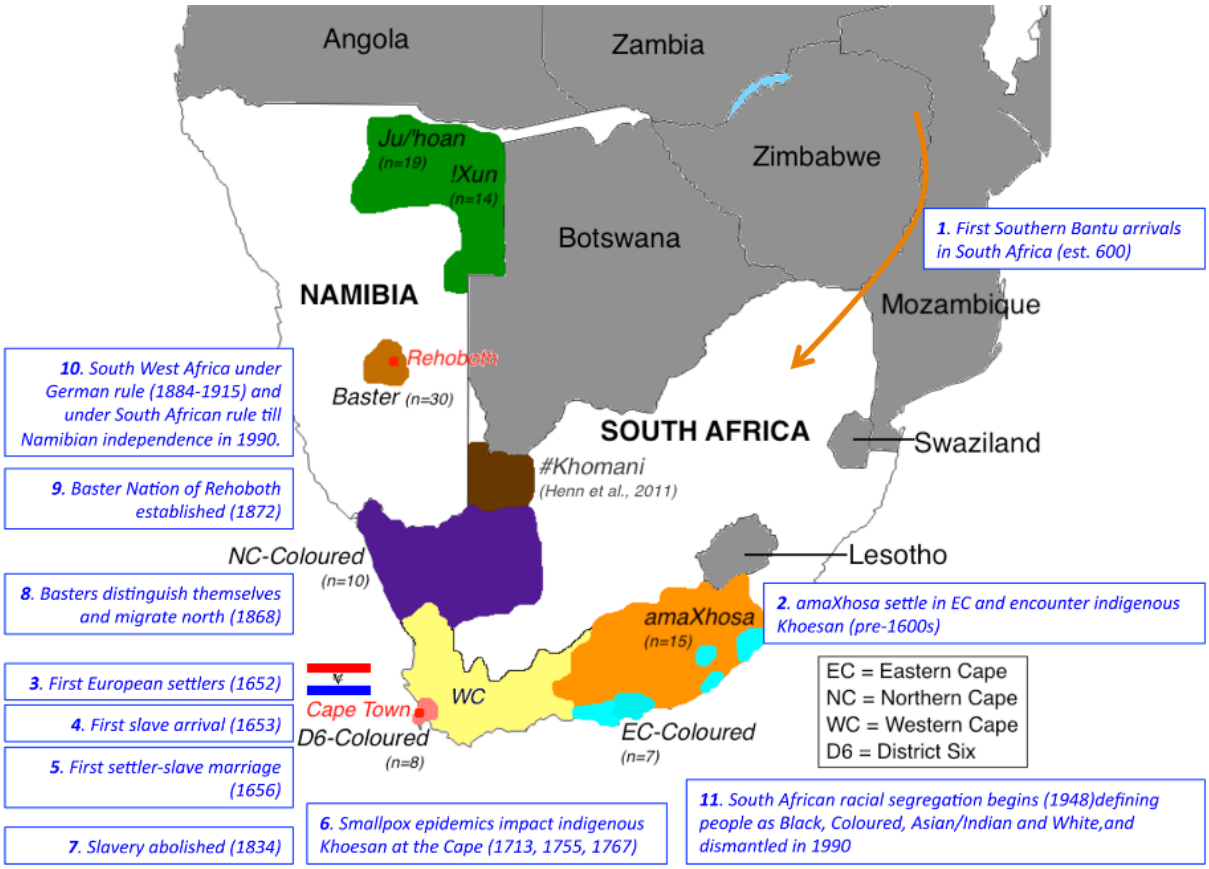


Figure 1. 3 Image taken from Petersen *et al*., (2013) showing the distribution of populations in Southern Africa with significant events that have shaped ancestral contributions. The number of participants per region is presented by "n".

The Southern Bantu populations migrated from a western/central African origin 1,500 years ago (Berniell-Lee *et al*., 2009), ending with amaXhosa settlements in Eastern Cape (South Africa). The Cape of Good Hope was populated by the Khoi-Khoi by the time the Europeans arrived in the 1600s. During the colonization the

eastern Bantu-speakers (specifically the Xhosa-speaking individuals) moved as far as the Fish River which is in the Eastern Cape (South Africa). The Europeans then settled (Dutch, German, French and later the British). The European colonists gave rise to the South African Coloured and Namibian Baster populations (Eurasian and indigenous descent) (Petersen *et al*., 2013).

KhoeSan people are situated in Namibia and Botswana (Petersen *et al*., 2013). The archaeological evidence shows that proto-Khoekhoe pastoralists migrated along a west coastal route southward through Namibia of which they crossed the Orange river into South Africa. Pre-pastoral Khoesan remains are thus found across these regions, south of the Orange River and south of the Zambezi River, the remains provide evidence of indigenous inhabitancy (Morris *et al*., 2014).

Researchers suggests that these herding migrants entered Nambia at 2,200 years (Robbins *et al*., 2005, Pleurdeau *et al*., 2012) along the coast to the Cape (Smith, 2006). The descendants from the modern Khoesan and the nonclick-derived Bantu pastoralists established themselves roughly 500 years later (Huffman 1992, Güldemann and Vossen 2000). The Agro-Pastoral Bantu populations migrated along the eastern coast. Researchers suggests a link between the pastoralists and East Africa (Pickrell *et al*., 2012) and a Eurasian contribution to Khoe-Kwadi people (Pickrell *et al*., 2014).

According to Li *et al.*, (2014) the Khoesan groups (hunter-gatherers) occupies the western parts of South Africa and the south and central parts of Namibia. Whereas the South-western Bantu groups are found in the north of Namibia co-existing with the hunter-gatherers (Ehret and Posnansky, 1982), the South-eastern Bantu groups occupies the entire eastern part of South Africa. The co-existence between Khoesan groups and Southern Bantu groups has led to admixture.

The Admixture between the ‘Khoesan’ hunter-gatherers, herders and the Bantu speaking populations led to population growth and expansion in Limpopo. The populations settling in Limpopo have contributed to the population composition and

shaped the ancestral contemporary populations in Southern Africa (Petersen *et al*., 2013).

1.1.4.1 History and statistics on the Limpopo population

According to the linguistics, history and physical anthropology, most South Africans are Bantu-speakers (Mitchell, 2010). The Agro-pastoralists settled into the eastern half of Southern Africa, where admixture occurred between hunter gatherers.

The archaeological evidence shows initial contemporary populations in southern Nambia, southwestern Zimbabwe and North-western Botswana. These farming communities settled during the first millennium in the southern parts of Africa. Bushmen living in the highland of Lesotho in 1870 and 1880s were surviving in their communities before the Bantu-speaking populations arrived (Jolly 1994, Vinnicombe, 2009), with Sotho-farmers arriving later in Lesotho (Mitchell, 2010).

Researchers have been investigating the North-eastern and North-western (Periera *et al*., 2001, Salas *et al.,* 2002) settlement of agropastoral populations. The genetic analysis of samples from Zimbabwe have shown that hunter-gatherers disappeared (due to assimilation perhaps) from most of Zimbabwe (Mitchell, 2002).

During the second millennium of farmers (Mitchell, 2010), the origin of the Sotho and Tswana and Nguni-speaking individuals, inhabitants of Botswana, Lesotho, South Africa and Swaziland (Mitchell, 2010) occurred. Thus, we expect population expansions and population growth, possible population's splits and admixture of populations that could have contributed to the population composition in Limpopo.

According to the Consensus (2011), the Limpopo province is in the northern region of South Africa (Statistics South Africa, 2019) neighbouring Botswana, Zimbabwe and Mozambique. The people in Limpopo form part of the fifth largest population in South Africa. With 5.4 million people, predominantly black and native black languages are spoken (Statistics South Africa, 2019). The common languages are **Sepedi** (52.9%), followed by **Xitsonga** (17%) and **Tshivenda** (16.7%). The populations consist of Black Africans (96.7%), White (2.6%), Indian or Asian

(0.3%) and Coloured (Admixture) populations (0.3%), (Statistics South Africa, 2019).

With this information we were curious about the population history and ancestral haplogroups of the people living in Limpopo. It is important to trace the phylogenetic lineages over a long period of the people today (Mitchell, 2010). This would answer questions as to how modern humans evolved. We are thus interested in tracing back the ancestral lineages of people living in Limpopo.

Researchers has been studying human migration and evolution in many ways, archaeology, languages and now more recently DNA. We can thus use mtDNA as a genetic marker to test the migration of the Limpopo populations and trace ancestral lineages.

The uniparental inheritance of mtDNA markers allows us to trace ancestral lineages of Limpopo populations through multiple generations. This brings us to the concept of maternal ancestry, which gets passed on over time.

1.2 Maternal ancestry

Ancestral information can be derived from DNA (deoxyribonucleic acid) that is known to be the hereditary material found within humans (NIH). Most of the cells in the human body contain DNA, nuclear DNA is in the cell's nucleus and mitochondrial DNA is in the mitochondria, outside the cell's nucleus (NIH).

During the 1980s, new tools allowed scientists to ask questions about the past. Studying DNA and the genetic information thereof in living human populations today, geneticists can trace lineages back in time (Gugliotta, 2008) and how humans have migrated over thousands of years. Thus, we can now look at the genetic variation found in specific nucleotide sequences of SNPs, INDELS, CNVs to give us the useful information we need to understand evolution and the migration of contemporary modern humans.

1.2.1 DNA polymorphisms

Genetic polymorphisms exist when at least two variants are present, such as gene sequences, chromosome structure or phenotypes (Daly, 2010). The probability of

polymorphic DNA in humans are present because of the large human genome causing genetic variability, this includes single base pair changes, many base pairs and repeated sequences. (Buckingham, 2012). These highly polymorphic markers have been used in human linkage mapping, forensics and paternity testing (Stone and Bornhorst, 2012). The types of polymorphisms are SNPs, INDELS, polymorphic repeats and CNVs.

SNPs: Single nucleotide polymorphisms (SNPs) are the most frequent DNA sequence variation to study (Zhang and Jia Min, 2005) and has made it easy to choose as markers for haplotype identification in this study due to being attractive markers (Brumfield *et al*., 2003, Morin *et al*., 2004). The availability of high numbers of annotated markers, the low-scoring error rates and the relative ease of calibration among laboratories makes SNPs useful in genetic and genomic studies (Helyar *et al*., 2011).

INDELS: An insertion/deletion polymorphism, can be used for ancestral testing but mutations are almost non-existent and have longer sequences, one to several 100 base pairs is inserted or deleted, thus time consuming (Gellman and Turner, 2013).

Polymorphic repetitive sequences: Tandem repeats are known to disperse across the genome within gene sequences. These repeats are ≥ 2 bp in length, divided into macro- , mini- and microsatellites (O'Dushlaine *et al*., 2005).

CNVs: Copy number variants (CNVs) involves segments of DNA > 1kb, involving a duplication or deletion of DNA (Feuk *et al*., 2006).

Studies have shown SNPS to be more favourable than microsatellites (Rosenberg *et al*., 2003) due it being highly informative having a very high information content for population structure analysis (Paschou *et al*., 2007, Helyar *et al*., 2011). Although microsatellites display greater allelic diversity per locus, the individual

SNPs can segregate strongly among populations (Freamo *et al*., 2011, Karlsson *et al*., 2011). SNPs are the best markers to use for ancestral testing because of the slow mutation rate.

1.3 Modes of inheritance influencing mutations

A defined feature of eukaryotes is the cytoplasmic organelles, these organelles have separate genomes undergoing different modes of inheritance (Neiman and Taylor, 2009). The modes of inheritance effects the evolutionary forces which allows for different genomes (Birkey *et al*., 1983). In contrast to the nuclear genome, we are interested in the mitochondrial genome, which undergo little or no recombination. This allows for a high mutation rate in mtDNA (Gabriel *et al*., 1993, Howell, 1996).

Mutations are known to fuel the variability in populations which will enable evolutionary change (Loewe and Hill, 2010). In population genetics, mutations are what will change populations in the long term (Loewe and Hill, 2010). With the concept of mutations and evolution that focuses on changes in populations due to natural selection. We can use the concepts of natural selection, genetic drift and gene flow to really understand the change in allele frequencies over time (Andrews, 2010).

By thus studying SNPs we can look at the mutations found within SNPs of today's populations and use the genetic information obtained from mtDNA to look at the lineages until we are able to reach an evolutionary trunk (Gugliotta, 2008). We can use the SNPs to test and assign individuals to their ancestral haplogroups and populations. To assign these haplogroups we need to look at genotyping methods used for SNP identification and testing and decide on the best method needed for ancestral testing.

1.4 Genotyping methods used for SNP identification.

The methods developed over the years to characterize and identify the genomic information at the specific sites of the human genome is what is important in this study, for this reason I discuss the historical overview of methods used to genotype.

1.4.1 Polymerase chain reaction

Kary Mullis developed the method, polymerase chain reaction (PCR), where the DNA sequences are exponentially amplified (Mullis *et al*., 1987). Specifically chosen primers and DNA polymerase enzyme is used to amplify fragments, these fragments are separated by gel electrophoresis and detected by means of different staining dyes (Teama, 2018). This allowed for technology to evolve and improve, which was the foundation in molecular techniques (Stone and Bornhorst, 2012).

1.4.2 Restriction Fragment length polymorphism analysis.

Restriction Fragment length polymorphism analysis (RFLP) is a method used for the identification of genetic variation within markers. This method was based on polymerase chain reaction (PCR) which amplifies DNA using primer pairs and DNA polymerase (Stone and Bornhorst, 2012).

Using restriction enzymes, the PCR product is digested, and the fragments are visualized by means of gel electrophoresis. It is time-consuming, low-throughput and inexpensive and you obtain a little information.

1.4.3 SNaPshot analysis

This method uses a primer-extension based method for genotyping, where the known SNP positions are analysed (Snapshot Multiplex System for Genotyping, [cd.genomics.]). Up to 10 SNPs can be analysed in a single reaction, detecting minor variants.

The advantages include high accuracy, cost effective and a quick turnaround time, high throughput level, easy to use and consistency within results. Multiplexing can be done, automated analysis and sensitive allele-frequency detection (Snapshot Multiplex System for Genotyping, [cd.genomics.]).

1.4.4 High resolution melting analysis

High Resolution melting (HRM) analysis has been used to identify DNA genotypes following PCR (Kanderian *et al*., 2015). Wittwer and Gundry *et al.,* (2003) introduced high-resolution DNA melting for the first time. This method of genotyping uses dyes particularly, which allows for PCR products with no additional processing to melt directly.

The melting curves are used for the investigation of mutations, due to the modifications of the GC/AT content of a stretch of the DNA resulting in a variation in melting temperature. When the amplicon melts the derived or ancestral state would show for the SNP site under investigation, these temperatures were then compared to a theoretical table of expected melting temperatures.

This method allows for the genotyping of small insertions or deletions, allowing for costs to be reduced and the sequencing load for analyses (Zhou *et al.,* 2013). HRM analysis is important for detecting DNA variants within research (Wittwer *et al.,* 2003) and was the preferred method used for testing SNPs within our study.

High resolution melting analysis is advantageous because of the low contamination risk, high analytical sensitivity and high speed (Wittwer *et al.,* 2009, Vossen *et al*., 2009). It does not require post-PCR separation, subsequent analysis such as sequencing can thus still be performed afterwards (Vossen *et al*., 2009). This method is easy to use, and you obtain a lot of data.

1.4.5 Sanger sequencing

Sanger sequencing is the method used for direct sequencing of DNA (Saumya and Walsh, 2016). This method allows for the amplification of specific DNA sequences with labelled nucleotides, detecting point mutations and small deletions or inserts.

The sequencing technology has evolved over the past years which resulted in reduced costs, enhanced accuracy and genetic coverage (Saumya and Walsh, 2016). However, it is time consuming and not a lot of data is obtained (1000bp).

1.4.6 Next generation sequencing

Studies have shown that the cost, speed and accuracy of next generation sequencing (NGS) to be revolutionizing the discovery when studying SNPs (Helyar *et al*., 2011). The potential of this high-throughput technology will allow for improved genotyping results of poor-quality samples (Smith *et al*., 2011).

The NGS data will most likely remain the basis of SNP development for the future (Helyar *et al*., 2011). Although NGS is a powerful tool, we consider the limitations of it being labour intensive with a large dataset to be analysed.

Yet we still focus on the advances it offers. We can resolve the difficulties of the previous systems with NGS as a powerful tool for the future, longer reads and more reliable and accurate data (Rusk, 2009). This method would be ideal for future work. We now look at how the mitochondrial genome is of importance within this study.

1.5 Mitochondrial genome

Mitochondrial DNA consists of a small fraction of the human genome, residing outside the cell's nucleus within the mitochondria producing energy (Van Oven and Kayser, 2009). Studying the uniparental inherited portions of the human genome is of interest due to the important applications of evolutionary studies as well as population history (Jobling and Tyler-Smith, 2003, Torroni *et al*., 2006, Underhill and Kivisild, 2007) and forensic sciences (Kayser, 2007). This molecule is a circular, double stranded DNA molecule, 16.6 kb pairs in length as seen in Figure 1.4.

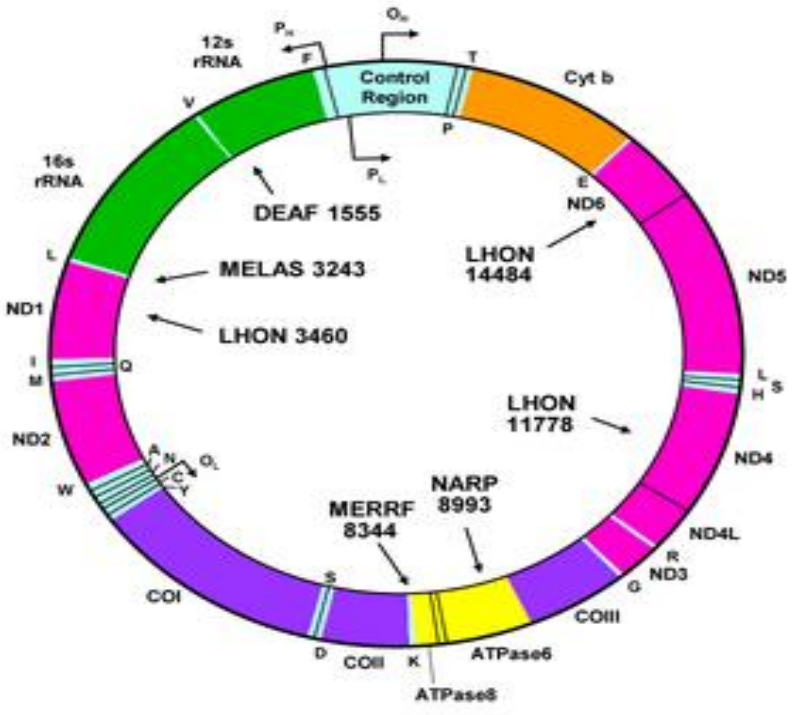


Figure 1. 4 Map of the Human Mitochondrial DNA (www.mitomap.org/mitomap) (28 September 2018, Marie Lott)

The distribution of the nucleotides in the mitochondrial genome can be separated into "light" and "heavy" strands (Scheffler, 1999). The "heavy" strand being the strand with the highest relative molecular mass with its guanine nucleotides compared to the light strand.

The replication of mtDNA begins with the heavy strand in the control referred as the "non-coding region" because it does not code for any gene products (Scheffler, 1999). The coding and the control (non-coding) region of the mitochondrial genome consists of nucleotide positions which range from 1 to 16569 according to the revised Cambridge Reference Sequence (rCRS) (Anderson *et al*., 1981, Andrews *et al*., 1999).

The mutation rate in the non-coding region (bases 16024-576) of 1.1kb is much higher, compared to the coding region (bases 577-16023) of 15.5kb (Howell *et al*.,

2007, Pakendorf and Stoneking, 2005), accumulating more mutations especially with sites prone to mutations (e.g. positions 709, 1719, 3010, 5460, 10398, 11914, 13105, 15884) which enriches sequence variation making the non-coding region particularly interesting to sequence (van Oven and Kayser, 2009).

In 1981 the human mtDNA was first sequenced in Cambridge, England (Anderson *et al*., 1981) and used as reference. The Anderson sequence is also known as the Cambridge Reference sequence (CRS) that was later re-sequenced and used to compare new sequences to Single nucleotide polymorphisms (SNPs) and insertions/deletion polymorphisms (INDELS) that are detected in the mitochondrial genome will be assigned as mutations to the rCRS (Van Oven and Kayser, 2009).

The human mtDNA anchors our most recent maternal common ancestor (MRCA) to sub-Saharan Africa, which suggested a single migration out of that continent resulting in the overall population of the world (Mellars 2006, Macaulay *et al*., 2005, Torroni *et al.,* 2006). Studying mitochondrial DNA is important because of the features which includes, matrilineal inheritance, high copy number and a higher evolutionary turnover rate compared to nuclear DNA, it is also non-recombinant (Van Oven and Kayser, 2009).

By looking at mtDNA types in the human gene pool, humans can be traced back to their matrilineal ancestral who lived approximately 200,000 years ago in Africa (Behar *et al*., 2008, Macaulay *et al*., 2005, Mishmar *et al*., 2003). MtDNA assists in understanding the history of man in terms of maternal migration and back migration of the populations that has shaped the human global population (Maji *et al*., 2009). The reverse migration of non-Africans back into Africa has contributed to the African mtDNA gene pool and gene flow of modern African populations (Tishkoff *et al*., 2009).

Examples of the back migration was shown in particularly Middle Eastern/European and East African Cushitic ancestry that was detected in Saharan

African Beja populations. We are thus able to also determine the genetic diversity that occurred within these populations (Tishkoff *et al*., 2009).

1.6 Genetic diversity

When we study genetic variation and diversity, it is crucial to identify the genetic markers that play a role in African populations as well as populations with non-African descent (Gomez *et al*., 2014). Thus, patterns of genetic variation in the present-day within genomes with African descent are a product of demographic and selective events (Campbell and Tishkoff, 2010).

When studying the characterization of the genetic diversity within Africa, the re-identification of modern human origins will be critical and will be informative for population specific variant identification and demographic history purposes (Campbell and Tishkoff, 2010). Previous studies based on genome wide data indicated high levels of genetic diversity in Africa compared to non-Africans resulting from mtDNA, X-chromosome and Y-chromosome analysis (Campbell and Tishkoff 2008), making the Africans particularly interesting to study.

A genome-wide study done by Tishkoff *et al.,* (2009) detected a larger more extensive population structure in Africa, 848 short tandem repeat polymorphisms (STRS), 476 small insertions and deletions (INDELs) and 3 single nucleotide polymorphisms (SNPS) genotyped 2,400 individuals which were from 121 geographically diverse populations, allowing to reveal 14 genetically distinct ancestral population clusters within Africa (Tishkoff *et al.,* 2009). This study proved that geographically diverse populations in Africa such as the Central African Pygmies share common ancestry with certain east- and Southern African KhoeSan-speaking populations (Tishkoff *et al.,* 2009).

This shows that Africa has the highest level of human genetic variation in the world. It is said that the African demographic history consists of extensive population structure, fluctuations in population size, short-and long-range migrations and admixture which has resulted in complex patterns of variation within the modern

populations (Campbell and Tishkoff, 2008, Manica *et al.*, 2007). Thus, the genetic diversity can be grouped into clusters by studying human mtDNA phylogeny.

1.7 MtDNA phylogeny

The reconstruction of the human mtDNA phylogeny demonstrates human migrations and demographic history of the human evolution (Behar *et al.*, 2008). The mtDNA phylogeny is presented as a branching structure shown in Figure 1.5, where researches reported these branches as being shaped in this manner due to the climate change that occurred within Southern Africa at different time frames (Barbieri *et al.*, 2013).

Each branch on the mitochondrial phylogenetic tree defines a specific mutation known as an ancestry informative single nucleotide polymorphism (AI-SNPs). The AI-SNPs represents the DNA sequence variation, in which the single nucleotide can undergo a transition or trans-version within every population.

Once the AI-SNPs information has been identified and obtained from the coding region on the mitochondrial genome, the genotypes of the screened samples can be used for haplogroup assignment. To accurately infer the mtDNA phylogenetic tree as part of mtDNA quality control, estimating site-specific mutation rates and haplogroup assignment, an up to date mtDNA phylogenetic tree is of importance (van Oven M., 2015).

Phylotree uses all the available sequences with variations along the entire 16.6 kb mtDNA genome, where it presents a world-wide phylogeny constructed and defined by a set of mutations in each branch (haplogroup) present (van Oven M., 2015) seen on Figure 1.5.

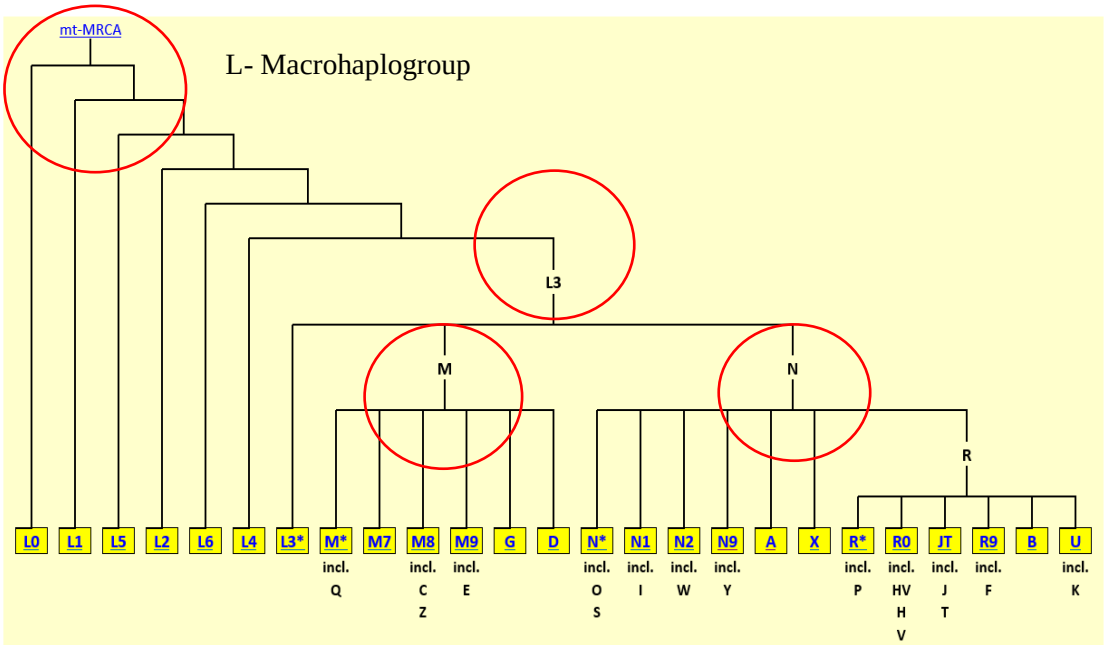


Figure 1. 5 MtDNA phylogenetic tree build 16 which is divided into 11 subtrees accessible through links on (http://www.phylotree.org/tree/main.htm.) The mitochondrial most recent common ancestor (mt-MRCA) is used to root the tree. Accumulation of polymorphisms over time separates individuals into various haplogroups. (Updated, 18 February 2016).

Haplogroups are based on the variants that share a common ancestor clustered hierarchically in clades with common mutations (Rosa and Brehm, 2011). Haplotypes are defined as a set of alleles at a linked locus, being present in one of the two homologous chromosomes (Bergstrom D.E., 2001). The haplotypes are defined by the mutations along the mtDNA molecule (Rosa and Brehm, 2011).

The tree is divided into four major branches (L, L3, M and N) known as Macrohaplogroups indicated by the red circles shown on Figure 1.5. These branches are then divided in sub-branches known as haplogroups. The sub-branches are divided into 8 major haplogroups (Van Oven and Kayser, 2009).

The diagnostic tool for screening was designed based on multiplexes of SNPs to determine the position of the individuals' ancestry on the mitochondrial DNA tree to assign haplogroups.

1.8 Mitochondrial haplotypes and haplogroups

The first discovery of mtDNA haplogroups in Native Americans, were made by Torroni *et al.*, (1993) who used A, B, C and D. The detected haplogroups were designated using other letters of the alphabet. As research developed, studies based on the mitochondrial genome was of importance and the haplogroup nomenclature changed over the years.

African Hg L (xM, N) is now used to describe African mtDNA haplogroups except L3 (M and N). The term 'branch' refers to the evolving sides if the root (L0 and L1′6) (Torroni *et al*., 2006). The L0 branch presents the L0 hg alone, whereas the L1′6 branch presents haplogroups (L1-L6) (Torroni *et al*., 1993). The two deep branches on the mtDNA phylogenetic tree, presents haplogroup L0 (Southern Africa) and the L1′6 (central and eastern Africa) which includes the out of Africa ancestral L3-sub-branch (Chan *et al*., 2019). Rito *et al*., (2019) reported on mtDNA analysis of a split within the northern side of the tree (L1′6) central/western African and eastern Africa (L2′3′4′5′6) around 150-135 kya (Soares *et al*., 2016).

MtDNA haplogroups are classified as Macrohaplogroup, namely L, M and N and are distinct by means of geographic origin (Maji *et al.*, 2009). The mtDNA macrohaplogroups are diverse and within different regions across the globe. Africa contributes to the highest percentage of diversity within the gene pool of humans, comprising of L haplogroups primarily (Maji *et al*., 2009).

1.8.1 Mitochondrial haplogroups relevant to South Africa

Macrohaplogroup L is the oldest and restricted to Africans which includes the Sub-Saharan African populations (L0, L1, L5, L2, L6, and L4). Fendt *et al.* (2012) classified 90.7% L0 haplotypes belonging to the Khoe-San (L0d/L0k) clusters (Behar *et al.*, 2008), in addition, they identified a 9.3% contribution of haplogroups (L2 and L3) within Khoe-San. Where L3 indicates out of Africa descent (Sub-haplogroup M and N) (Quintana-Murci *et al*., 1999, Mishmar *et al*., 2003). The MHg L can be found across many regions of Africa, central-west and south-east (Pereira *et al*., 2001, Salas *et al*., 2002.

Models predating the haplogroups L1 and L0 suggest that the mtDNA gene pool within the sub-Saharan continent is a result of the early expansion of the modern humans from their place of origin, often suggested as East Africa (Mishmar *et al*., 2003, Maca-Meyer *et al.*, 2001) to the rest of the African continents by exclusively L1 haplogroups. Fendt *et al*., (2012) confirms the dispersals of groups within Northern/Eastern Africa (L5, L6 and L4) and Central to Southeast Africa (L1) as well (Quintana-Murci *et al*., 2010).

A later expansion occurred, estimated at 60,000-80,000 years before present (ybp) by haplogroups L2 and L3 (Foster 2004, Watson *et al*., 1997). Another interesting mark was the geographical enrichment of the African gene pool during the early Palaeolithic stage when populations from M and U clades arrived in north and northeast Africa from Eurasia (Behar *et al.*, 2008). Sub-haplogroup M and N initially entered the Indian sub-continent or South Asia moving to Australia and Southeast Asia at a later stage (Maji *et al*., 2009).

Majority of the world falls within the L3 (M and N) haplogroups as non-Africans, marking the out of Africa dispersal at 50,000-65,000 years before present (Behar *et al.*, 2008, Macaulay *et al.*, 2005). As previously mentioned, we know that Agro-pastoralists settled within the neighbouring countries of Limpopo, admixture between Khoesan and Southern Bantu populations occurred, thus population expansion and population growth is expected. We are thus expecting to screen haplogroups with African-ancestry particularly focused on haplogroups L1 and L3 and L0 as we were curious about the population composition and ancestral haplogroups in Limpopo.

Haplogroup L1

MtDNA L1 lineages coalesced about 140-150kya (Torroni *et al*., 2006, Behar *et al*., 2008). When studying the L1 haplogroups, research has divided L1 into L1b and L1c, where L1b is in western, central and northern Africa (Watson *et al*., 1997, Rosa *et al*., 2004). L1c is present in central African Bantu populations (Vigilant *et al*., 1991, Destro-Bisol *et al*., 2004). L1b and L1c are both not present in eastern and Southern Africa (Gonder *et al*., 2006).

Reports stated an 18-25% in Anglo Bantu ethnic groups (Beleza *et al*., 2005, Coelho *et al*., 2009). The cultural transition from hunting-gatherers to agriculture has contributed to a shared ancestral Central African proto-population in L1e haplotypes (Quintana-Murci *et al*., 2008). These haplogroups have shown to present African Bantu populations that we now know are present within the Limpopo populations.

Haplogroup L3

The origin of the L3 haplogroup started in East Africa about 60-75 kya (Salas *et al*., 2002, Macaulay *et al*., 2005, Kivisild *et al*., 2006, Torroni *et al*., 2006, Behar *et al*., 2008, Soares *et al*., 2009). The L3 haplogroup composes all non-Africans sub-branching in M and N (60-65 kya) (Macaulay *et al*., 2005), these two groups are the only other groups found throughout the rest of the world (Metspalu *et al*., 2004).

This would suggest that the rest of the world has been colonized by descendants with an ex-African mtDNA haplotype (Oppenheimer S., 2012). However, the L3* branch is the exception, due to having African descent. When studying the L3 haplogroups, research has shown the highest proportions to be in north and east Africa (Harich *et al*., 2010). Consisting of sub-haplogroups L3b and L3d dominating west and north Africa. L3b is common among Bantu speakers, being a marker of the Bantu expansion (Watson *et al*., 1997). L3d forms part of the South African maternal pool and dominant in Angola and Tanzania (Tishkoff *et al*., 2007, Coelho *et al*., 2009).

Sub-haplogroup L3e is commonly found in central Sahara and South Africa (Harich *et al*., 2010). L3e has become quite frequent among southeast Bantu-speakers in Mozambique (Salas *et al*., 2002). The L3f sub-haplogroups is found across the Sahara with an eastern localization with Chadic-speaking groups (Cerný *et al*., 2009).

Haplogroup L0

One of the earliest mtDNA haplogroups L0 (L0a, L0d, L0f and L0k) estimated at 140-160 kya (Salas *et al*., 2004, Gonder *et al*., 2006, Kivisild *et al*., 2006, Behar *et al*., 2008).). L0d diverged around 100 kya (Behar *et al*., 2008) and is distributed within Khoesan people in South Africa, Tanzania and Angolan populations (Coelho *et al*., 2009).

L0k diverged around 40 kya (Behar *et al*., 2008) and found exclusively among Southern African Khoesan groups (Chen *et al*., 2000, Salas *et al*., 2002) being present in low-frequencies among click-speaking Tanzanian people (Tishkoff *et al*., 2007). L0f is rare and only present in East Africa, diverging around 85-90 kya, commonly found in Tanzanians (Castri *et al*., 2009). L0a diverged about 40-55 kya in East Africa (Soares *et al*., 2009) and found throughout South Africa. Specific L0a2 lineages have shown to trace the dispersal of Bantu-speakers towards South Africa around 3 kya (Soodyall *et al*., 1996).

1.9 Aims of the study

The aims of this study was to design and develop a quick, inexpensive and accurate hierarchical diagnostic screening system of mtDNA haplogroups according to the mtDNA phylogenetic tree, selecting AI-SNPs in the mtDNA genome by using High Resolution Melting analysis to evaluate the population composition and ancestral haplogroups of Southern African populations in Limpopo.

The objectives were to

-to use the hierarchical diagnostic system to infer mtDNA lineages

-to evaluate the maternal ancestry of a region in South Africa that is largely unknown

To achieve this aim, we have designed primers (Chevonne de Wet, Peter Ristow and Eugenia D'Amato, unpublished, 2015) to test each SNP under investigation at a specific site by establishing whether the site under investigation is derived or ancestral within a given state.

Once the SNPs information is identified and obtained from the coding region of the mitochondrial genome, the genotype results of the screened samples can be used for haplogroup assignment.

1.10 Significance of the study

The results of the quick, accurate and cost-effective diagnostic multiplex system were used to evaluate the population composition of 500 Southern African individuals from Limpopo (e.g. Khoesan, Bantu, out-of-Africa). HRM is a powerful tool used in population genetics and forensics to investigate AI-SNPs.

SNPs are ideal, due to its low error rates, the potential for high-throughput genotyping and downstream analysis, being high-informative markers for population analysis (Helyar *et al*., 2011). In population genetics it important to determine the ancestral origin of these individuals so that we can contribute to the knowledge of Southern African ancestry.

In forensic genetics using SNPs to infer population history and ancestral origin has become significant, this study allows us to study population groups by investigating their genetic markers and the application of our results allows us to do downstream analyses.

Chapter 2

Methodology

2.1 Ethical approval

To ensure that the correct ethical procedures were observed, the study was presented to and approved by the Ethics Board of the University of the Western Cape, Cape Town (Ethics clearance letter: 15-4-97) and consent to perform testing was obtained from each participant. Permits from the University of Limpopo (Project: TREC/09/2016) and the University of Venda (Directory of Research and Innovation, 25 November 2015) was obtained (See supplementary data).

2.2 DNA sample collection and preparation

The DNA control samples were collected (Chevonne de Wet, Peter Ristow, Eugenia D'Amato, Kevin Cloete, 2015) in Johannesburg. Donors were from the Gauteng Khoesan Descendant Communities, through the Gauteng Khoe & San Council under the Leadership of Commissioner John Stalin van Rooyen (! Kung-Koya-Quena).

The tested DNA samples were collected (Adria Burrows, Peter Ristow, Eugenia D'Amato, Kevin Cloete, 2016) in Limpopo. Donors were from the University of Limpopo and the University of Venda.

All consenting individuals provided DNA in the form of saliva and their ethnographic information regarding their ancestry, language and the parental place of birth were reported (See supplementary data for consent form). Briefly, each donor was given a 15 ml Falcon tube which contained 1 ml of an in-house mouth wash solution, which facilitates saliva production providing more DNA. The falcon tube was filled with saliva to the 5 ml mark and an equal amount of storage buffer (Burrows *et al.*, 2017) was added to every saliva sample and kept at room temperature. The samples were kept at room temperature until they could be stored at -20˚C for further analysis. Additional information of each participant was recorded including paternal and maternal population groups, home language, date and place of birth.

A total of 760 genetic samples were collected from donors within Limpopo using this method. The genetic samples were collected in the North eastern region of Limpopo.

2.3 DNA extractions

Only 500 of the 760 saliva samples were extracted by means of an optimized Medrano's (1990) salting out technique. The other 260 saliva samples were extracted by other members of the laboratory with the same technique. The extraction starts by combining 300 µl saliva with 250 µl Lysis buffer (50 mM Tris-CL, 10 mM EDTA, pH 8.0, 2% SDS) and 6 µl of 20mg/ml Proteinase K in a 2 ml Eppendorf tube, incubated overnight at 53˚C. The following day 300 µl of 4M NaCl was added to each tube and vortexed at high speed for 10 seconds. The samples were incubated on ice for 8 minutes and centrifuged at 14 000 rpm for 5 minutes and 600 µl of the supernatant was transferred to a new tube containing 600 µl isopropanol. The tube was then vortexed, incubated on ice for 30 minutes and centrifuged at 14 000 rpm for 7 minutes. The supernatant was carefully discarded without dislodgement of the pellet. The salts were removed by washing the DNA pellet with 300 µl EtOH (70%), vortexed and centrifuged at 14 000 rpm for 7 minutes. Once again, the supernatant was carefully discarded without dislodging the pellet. Lastly, the sample tubes were air dried for 15 minutes and the DNA was re-suspended in 50 µl 1x TE buffer pH 8 at 55 ˚C for 15 minutes in the incubator. The quantity and purity of the extracted genomic DNA was determined using a Nanodrop ND-2000 UV spectrophotometer (Thermo Fisher Scientific, Wilmington, USA). All samples were diluted down to 10ng/µl as working solutions. All the samples were stored at -20˚C until amplification occurred.

2.4 HRM diagnostic screening system

The pre-optimization of the HRM diagnostic system was done by Chevonne de Wet with control samples (positive controls) (Chevonne de Wet, unpublished, 2015). The control samples were validated by sequencing done by Peter Ristow which largely contributed to the experimental design of this study.

2.4.1 Experimental design

Using Phylotree v16 (http://www.phylotree.org/tree/main.htm) (Accessed, 2016) we identified 5 (AI-SNPs) within the mitochondrial genome that define African and non-African ancestries. For each AI-SNP we designed oligonucleotides sets that could be divided into two hierarchical multiplexes (Chevonne De Wet, Peter Ristow and Eugenia D'Amato, unpublished, 2015).The two multiplex systems were, Multiplex 1 which consists of 3 primer sets identifying African ancestry and Multiplex 2 which consists of 2 primer sets, evaluating non-African ancestry.

2.4.2 Multiplex design and primer efficiency

To test the efficiency and the compatibility of the primer's pairs, each of the 5 primer sets were designed (Chevonne de Wet, Peter Ristow, Eugenia D'Amato, unpublished, 2015) (Table 2.1) with Oligo v7 (Wojciech, 2010) primer design software and the rCRS (Andrews *et al*., 1999). The annealing temperatures were designed to be as uniform between sets as possible (Table 2.1) as determined by Oligo v7 (Wojciech, 2010). In order to predict the possible genotypes and to identify mutations in primer binding sites the primer sequences (White Scientific (Pty) Ltd, IDT) were aligned against full mtDNA sequences consisting of published and unpublished whole genome sequences using Bioedit v.7.0.5.3 (Hall, 1999). The selected sequences were evaluated using the BLAT (https://genome.ucsc.edu/cgi-bin/hgBlat) search against the human genome to confirm the specificity of each primer set. Finally, Oligo 7 was used to estimate the melting temperatures for the derived and ancestral state for each SNP (Refer to Table 2.1)

Table 2. 1 Five SNPs for every position under investigation. The table below contains information regarding the sequence combination, melting temperature for the primer and amplicon, the primer combination and amplicon size.

Position	Multiplex system	Primer Tm	Amplicon size	Allelic state	Amplicon Tm	Theoretical Tm	Sequence
A1018G (Haplogroup L3) **Motif: A/G**	**Multiplex 1**	58.5˚C 59.7˚C	65bp	Ancestral: A Derived: G	72.9˚C 73.5˚C	72 ˚C 73.5˚C	**F:** 5'TGAGTTGTAAAAAACTCCAGTTGACAC 3' **R**: 3'TGTTCAGATATGTTAAAGCCACTTTCG 5'
T14178C (Haplogroup L1) **Motif: T/C**		57.4˚C 58.9˚C	78bp	Ancestral: T Derived: C Derived: C+C	76.2 ˚C 76.7 ˚C 77.2˚C	74.5 ˚C 75.5 ˚C 75.5 ˚C	**F:** 5'CTACTCCTAATCACATAACCTATTCCC 3' **R:** 3' GGTTGTTTGTTACAAGTTGGTCA 5'
T5442C (Haplogroup L0) **Motif: T/C**		56.7˚C 58.2˚C	51bp	Ancestral: T Derived: C	80.1°C 80.9°C	80.5 ˚C 81.0 ˚C	**F:** 5' GAA CAT ACA AAA CCC ACC CC 3' **R:** 3' CGT GGT AAG GGC GAT GAG 5'
T14783C (Haplogroup M) **Motif: T/C**	**Multiplex 2**	62.1 ˚C 64.3 ˚C	68bp	Ancestral: T Derived: C	76.9°C 77.5°C	74 ˚C 75 ˚C	**F:** 5' ATG ACC CCA ATA CGC AAA ACT AAC C **R:** 3' GGG AGG TCG ATG AAT GAG TGG TTA A 5'
G8701A (Haplogroup N) **Motif: G/A**		53.6 ˚C 54.5 ˚C	51 bp	Ancestral: G Derived: A	72°C 71.2°C	86.7 ˚C 85 ˚C	**F:** 5' TCA AAC TAA CCT CAA AAC AAA TGA TA 3' **R:** 3' TTC GTC CTT TAG TGT TGT GT 5'

2.4.3 Controls for HRM testing

To ensure that the melting profiles are corrected, samples with known sequence data (Peter Ristow, 2017) was used as positive controls for HRM testing. The positive controls were validated to ensure the diagnostic system is reliable with reproducible results.

Table 2. 2 The following control samples were used to confirm the haplogroups for Multiplex 1 and Multiplex 2. Representing the five SNP sites.

Haplogroup	A1018G	T14178C	T5442C	G5231A	T14783C
L3	G	T	T	A	T
L1	A	T	T	A	T
L0d	A	C	T	A	T
L0a	A	T	T	G	T
L0a	A	T	C	A	T
M	G	T	T	A	C
N	G	T	T	A	T

The HRM technique was used to test the AI-SNPs of the 500 genomic samples from Limpopo using the KAPA™ HRM FAST PCR kit (Kapabiosystems), following the manufacter's protocol. The DNA polymerase and the EvaGreen® dye in combination allows for a quick, effective PCR product to amplify and for the discrimination of difficult sequences (Kapabiosystems, 2016). All reagents were thawed on ice and the reactions were made up according to Table 2.3.

The reactions were genotyped on the Rotor-Gene® Q (QIAGEN) real time thermocycler with the following cycle conditions. The enzyme activation at 95 ˚C for 3min, 35 cycles of 95 ˚C for 5 seconds and 62 ˚C for 30 seconds. A final denaturation at 95 ˚C for 1 min ensures that all the amplicons are fully melted. After a final denaturation, the amplicons were gradually heated from 65 ˚C up to 95 ˚C for 2 seconds per temperature with a ramp of 0.1 ˚C/s on HRM channel.

Table 2. 3 Reagents and concentrations for the PCR setup.

Reagents	Stock concentration	Volume (µl)	Final concentration
Kapa HRM kit	2X	5 µl	2X
$MgCl_2$	25mM	1 µl	2.5 mM
Primer set (Multiplex 1: L0, L1 and L3)	10X	1 µl	0.3 µM per primer set
Primer set (Multiplex 2: M and N)			0.4 µM per primer set
DNA	10 ng/µl	1 µl	10 ng/µl
$_dH_20$		2 µl	
Total volume		10 µl	

2.5 Data analysis

To determine that the unknown samples were assigned correctly, a visual confirmation and manual assignment was done. The amplification and melting results of each unknown sample screened against each control sample was done. The results of the amplification and melting was analysed on the Rotor-Gene® HRM software. This software was as analysis of the 500 Limpopo samples that were genotyped. The screening process was done with Multiplex 1, followed by Multiplex 2 as shown in Table 2.1. The data was normalized from 0% to 100% fluorescence for the sole purpose of ensuring that the pre and post-melting regions are horizontal. This would relate to the DNA helicity, which can be compared to predicted curves (Palais and Wittwer 2009, Dwight *et al.,* 2011). The instrument precision is important and allows for sensitivity with constant melting rates of 0.1˚C/s (Hermann *et al.,* 2007). This software is a powerful tool used for high resolution melting analysis.

The software that exist for the genotyping and clustering allows the user to select different temperature windows (before and after the change in fluorescence), thus allowing background fluorescence to be removed (Li *et al.,* 2014, Palais and Wittwer, 2009). The analytical results should also be consistent if different

genotyping software is used (Kanderian *et al.,* 2015), thus allowing results to be reproducible and valid.

2.5.1 Rotor-Gene® ScreenClust analysis

After the visual analysis of the screened samples, the melting curve data was imported from the Rotor-Gene ® HRM software to Rotor-Gene ® ScreenClust software. This software groups the screened samples and the control samples in clusters based on their melting temperatures.

This clustering method allowed users to use an automated assignment, known as a principle component analysis (PCA) to confirm the screened and control samples and identify the haplogroups shown in clusters. The comparison of an unsupervised and supervised mode for clustering was done. The unsupervised mode determines the ideal number of clusters and the supervised mode clusters unknown samples into known clusters (QIAGEN, 2009: Reja *et al*., 2010).

2.5.2 Algorithm to confirm haplogroup confirmation

A mathematical algorithm that confirms the screened samples and assigned haplogroups was designed. The algorithm is based on the theoretical melting temperatures for each haplogroup predicting expected haplogroups. This method is proposed as an additional assignment test to confirm the expected haplogroups by determining the difference in haplogroups = IF (logical test, (true value), (false value) X= predicted haplogroup as shown in Table 2.4.

Table 2. 4 Mathematical algorithm for the prediction of assigned haplogroups.

L0-L3	Pred L0	L3_D-L3	Pred L3	L1_D-L1	Pred L1
Difference in melting temperatures of controls = L0- L3* L3* could be L1, L0 or screened sample	X=IF (Difference in (L0-L3) < "mean average of "L0 control", **"L0" or "0"**	Difference in melting temperatures of controls= Derived L3- D-L3* could be L1, L0 or screened sample	X= IF (Difference in (L3_D-L3*) > "mean average of L3 control, **"0" or "L3"**	Average in melting temperatures for L1 control haplogroup, mean average	X= IF (Mean average of L1 control group (L1_D-L1) ≤ "L1_D1-L1", **"L1" or "0"**

Table 2.4 shows the formulas needed to predict assigned haplogroups as an additional confirmation test. It needs to be acknowledged that this method is not a required method for ancestral testing and that the melting temperatures would greatly influence the prediction of haplogroups compared to the other methods used for testing. However, this method can be used as an additional method of haplogroup assignment.

Thus, a discrepancy is that the formula is consistent within a set melting temperature range, if samples fall outside of this range, you are not able to confirm the predicted haplogroups with this equation.

Chapter 3

Results and discussion

3.1 Spectrophotometric analysis

As mentioned in Chapter 2, within the scope of this study, the 500 extracted DNA samples were quantified. The results of the extracted DNA samples were confirmed by using the Nano drop ND-2000 (Thermo Fisher Scientific).

This non-invasive method of sampling DNA has proven to reveal high quality and quantity DNA (Medandro, 1990). The quality and quantity of the DNA samples yielded pure, contaminant free DNA. The sample concentration input needed to be between 10- 20 ng/µl, for successful PCR to occur. Concentrations ranged from 5.5 ng/µl – 485 ng/µl, with a mean value of 28.9 ng/µl. The purity of the samples was within the range of 1.6 - 1.9, with a mean value of 1.8. All the samples that exceeded the expected input concentrations (10 ng/µl) were used as working stock samples and dilutions of 10 ng/µl, were made up of these samples (See the Supplementary data). The samples with very high DNA concentrations, were confirmed by testing the absorbance values to ensure that the purity was 1.8.

3.2 Population composition

Below the population metadata is shown, it was recommended by (Parson and Bandelt, 2007, Haga and Venter, 2003) to provide information which includes regional geographic information about the birthplace of the maternal grandmother before the process of mtDNA screening. This information was used to confirm the population composition of the screened samples.

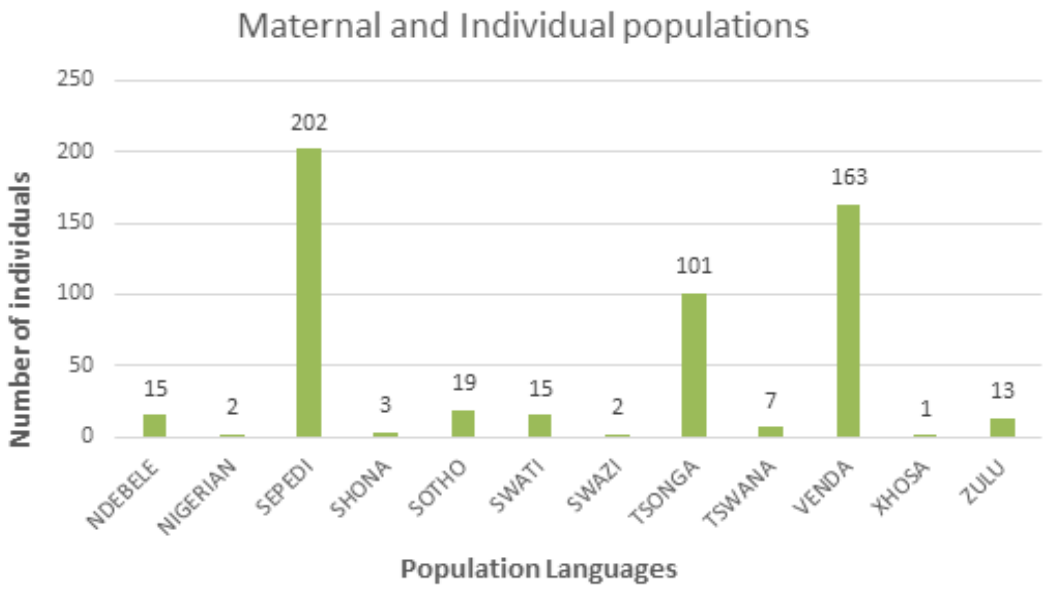


Figure 3. 1 The population meta-data of maternal and individual populations screened and their spoken languages within Limpopo.

Figure 3.1 shows the variation of languages within the group of screened individuals. The results show the number of individuals and three populations ranking the highest, Sepedi speaking individuals are just over 200 individuals. The second most spoken language in this sampled population is Venda which is a total of 163 individuals. Tsonga-speaking individuals' ranks third with just over 100 individuals of the total screened samples. Tswana, Xhosa, Zulu, Sotho, Swati, Swazi, Shona, Ndebele and Nigerian individuals are present within the screened samples as seen in Figure 3.1.

These results coincide with the most prevalent languages and populations in Limpopo (Consensus, 2011) according to (Statistics South Africa, 2019). Sepedi is the most spoken language in Limpopo and the most prevalent individuals. Predominantly black and native black languages are spoken (Statistics South Africa, 2019) in Limpopo. The common languages are Sepedi (52.9%), followed by Xitsonga (17%) and Tshivenda (16.7%).

Schuster *et al*., (2010) reported on the complete Khoesan and Bantu genomes in Southern Africa, where Bantu individuals are representatives of Sotho-Tswana and Nguni speakers (Niger-Congo languages). Based on the results in Figure 3.1 it can be deduced that Bantu individuals are present in Limpopo, presenting Sotho-Tswana individuals. The results confirm that African ancestry has contributed to the composition and population structure of the screened individuals from Limpopo.

3.3 HRM ancestral testing for Multiplex 1

The 500 extracted samples were amplified with the designed primers (Chevonne De Wet, Peter Ristow and Eugenia D'Amato, unpublished, 2015). As mentioned in Chapter 2, the samples were first genotyped with Multiplex 1. The first multiplex consists of the three informative SNPS under investigation, L3: A1018G, L1: T14178C and L0: T5442C (Refer to table 2.1), which gives an indication of in African (L1, L0 or L3*) and out of Africa (M and N) ancestry.

To test each SNP under investigation at a specific site, an in-housing screening method known as HRM was used to successfully assign South African individuals to a specific ancestral population group. This would establish whether the site under investigation is derived or ancestral within a given state. The screened samples were inferred from the control samples. These screened samples are manually assigned to their specific ancestral group. As stated in Chapter 2 (Table 2.1) the theoretical temperatures for each SNP is presented and used as reference to confirm the melting temperatures for each multiplex.

Figure 3. 2 Sequencing result for control L0d shown in Ugene of the SNP T5442C assigned to haplogroup L0 (Ristow, 2017).

Figure 3.2 shows the mutation within SNP T5442C, where the nucleotide position has changed. **T** (black circle) showing the derived state of the SNP. The controls have been sequenced to confirm the haplogroup assignment.

When each SNP is under investigation, the motif is studied and based on the modification of the nucleotide base T/C or A/G distinguishes the derived or ancestral state of the SNP. The sequencing results of the L0 control shown in Ugene confirms the SNP position. Individuals who are derived for the L0 haplogroup would have the exact same change in nucleotide position, T: derived state and C: ancestral state. Thus, the screened samples were compared to the sequenced controls of each multiplex system for inference.

Melting curve for the ancestral and derived state for a SNP

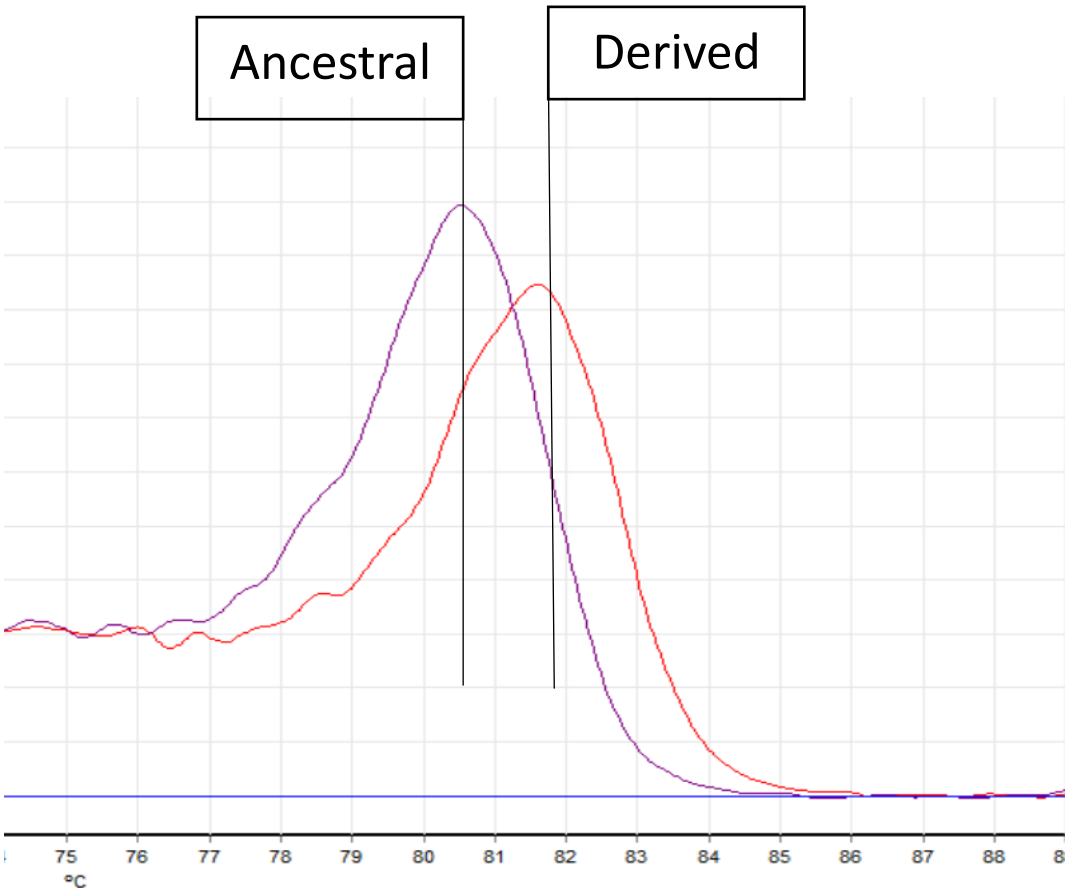
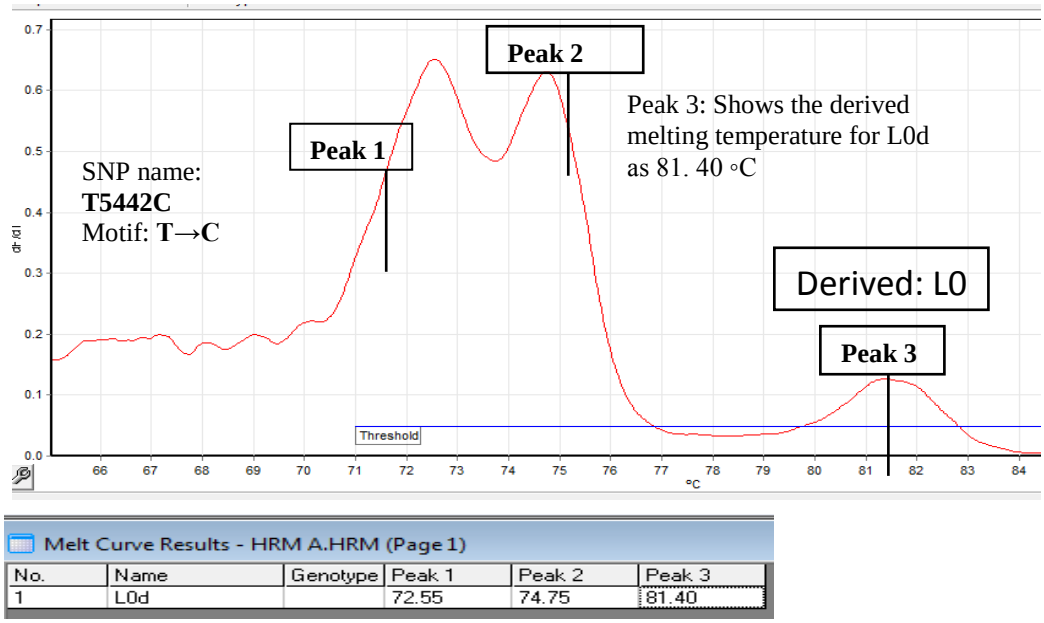


Figure 3. 3 Presents two distinctive groups (ancestral and derived) for the L0-haplogroup looking at SNP: T5442C.

As seen in Figure 3.3 the genotyping is facilitating in showing that there are two distinctive groups. The ancestral form (T5442) and the derived form (5442C). In other words the purple melting curve presents an individual who is ancestral for the L0 haplogroup. Whereas, the red melting curve presents an individual who is derived for the L0 haplogroup. In this experiment only primer L0 was used, thus two peaks for one SNP in this instance.

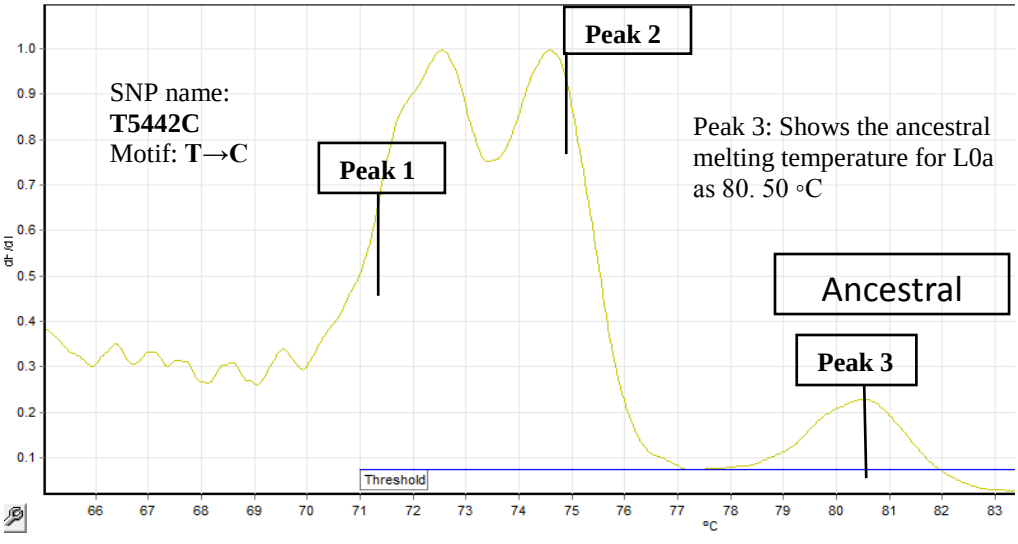


Melt Curve Results - HRM A.HRM (Page 1)

No.	Name	Genotype	Peak 1	Peak 2	Peak 3
1	L0d		72.55	74.75	81.40

Figure 3. 4 Multiplex 1 is represented by control sample L0d (red) derived for L0 (Khoesan) haplogroup.

Figure 3.4 merely indicates the melting curve for control L0d, which is important when there are multiple primers in one run you are able to differentiate between the different melting curves, in this instance Khoesan haplogroups.

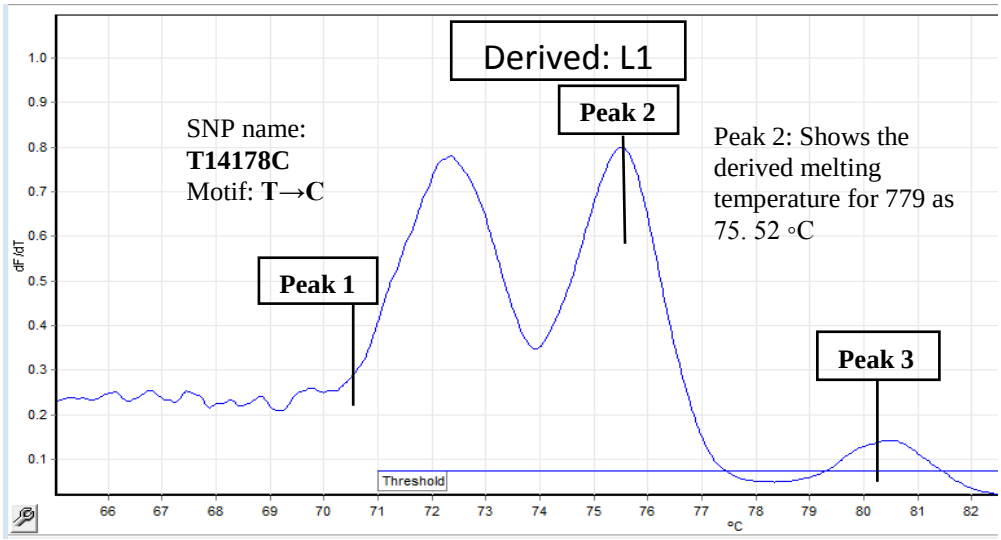


Melt Curve Results - HRM A.HRM (Page 1)

No.	Name	Genotype	Peak 1	Peak 2	Peak 3
2	L0a		72.55	74.60	80.50

Figure 3. 5 Multiplex 1 is represented by control sample L0a (yellow) ancestral for L0 haplogroup. The second derivate will differentiate between L0d (derived-Khoesan) and L0a (Khoesan) ancestry.

Figure 3.5 merely indicates the melting curve for control L0a, which is important when there are multiple primers in one run you are able to differentiate between the different melting curves. Loa control is also an indication of being ancestral for all SNPs used in Multiplex 1.

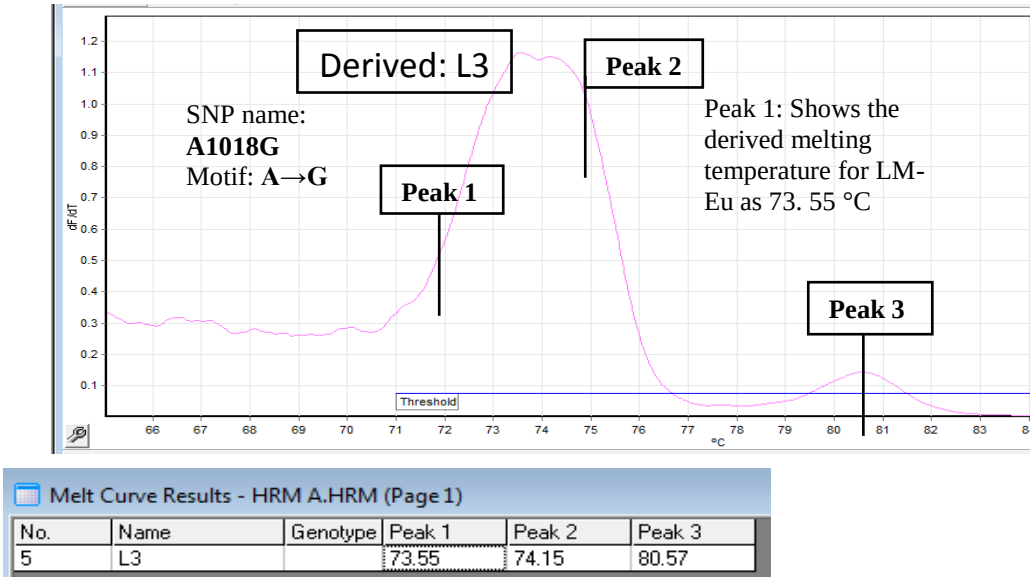


Melt Curve Results - HRM A.HRM (Page 1)

No.	Name	Genotype	Peak 1	Peak 2	Peak 3
3	779		72.35	75.52	80.50

Figure 3. 6 Multiplex 1 is represented by control sample 779 (blue) derived for haplogroup L1 (Bantu) ancestry.

Figure 3.6 merely indicates the melting curve for control L1, which is important when there are multiple primers in one run, you are able to differentiate between the different melting curves, in this instance Bantu haplogroups.



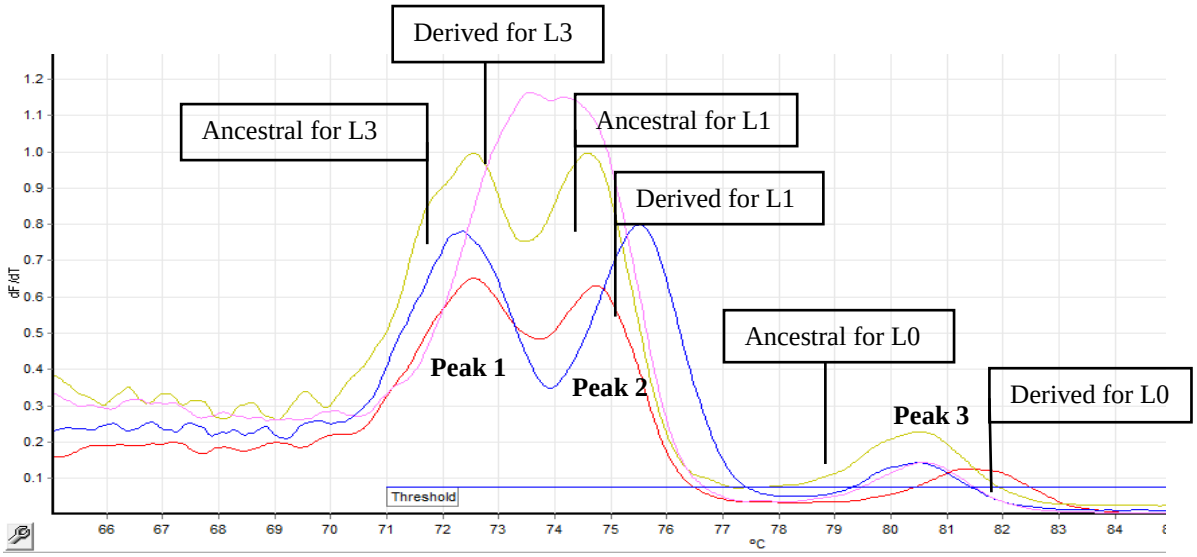
Melt Curve Results - HRM A.HRM (Page 1)

No.	Name	Genotype	Peak 1	Peak 2	Peak 3
5	L3		73.55	74.15	80.57

Figure 3. 7 Multiplex 1 is represented by control sample L3 (pink) derived for haplogroup L3 (Non-African ancestry or L3* possible African).

Figure 3.7 merely indicates the melting curve for control L3, which is important when there are multiple primers in one run, you are able to differentiate between the different melting curves, in this instance non-African or L3* haplogroups.

Showing the control melting curves individually seen in Figure 3.4 – Figure 3.7 as a single-plex allows researchers interpret a multiplex with ease. Multiplex 1 consists of three primer sets as shown below in Figure 3.8.



Melt Curve Results - HRM A.HRM (Page 1)

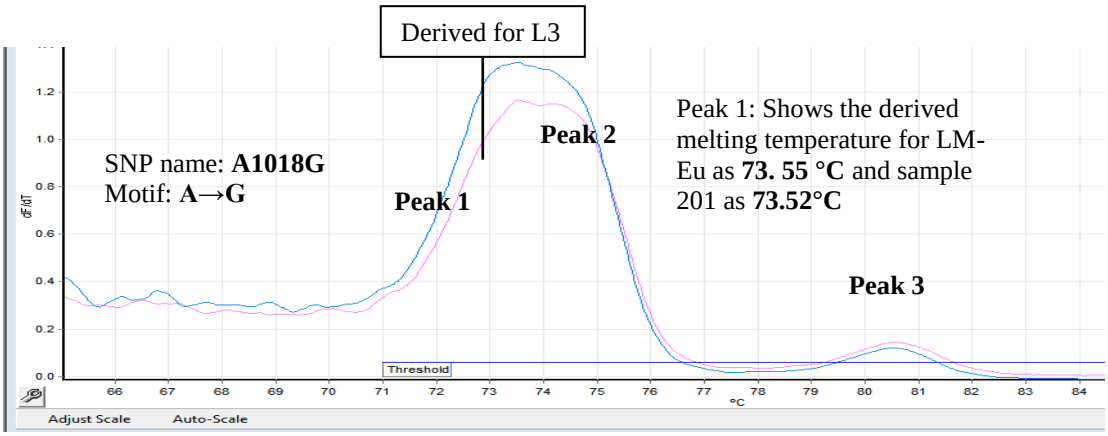
No.	Name	Genotype	Peak 1	Peak 2	Peak 3
1	L0d		72.55	74.75	81.40
2	L0a		72.55	74.60	80.50
3	779		72.35	75.52	80.50
5	L3		73.55	74.15	80.57

Figure 3. 8 Multiplex 1 is presented by the four control samples L0d (red), L0a (yellow), 779 (blue) and L3 (pink) of three SNPS which are derived for haplogroups L0, L1 and L3.

The melting curve for every SNP under investigation is shown above. The motif of each SNP is shown on the figures and each control curve and can be differentiated based on the melting differences between amplicons. Figure 3.8 presents the combined melting curves of the controls for Multiplex 1, assigning haplogroups L3, L1 and L0. The melting curve analysis was done on the 500 DNA samples that were amplified with the designed primers and the melting temperatures for every screened sample can be found in the supplementary data.

The melting curves are used for the investigation of mutations, due to the modifications of the GC/AT content of a stretch of the DNA resulting in a variation in melting temperatures. The melting temperature for every amplicon was determined by means of the melt curve which consists when relative fluorescence versus temperature constitutes (Vandersteen *et al.*, 2007).

The melting curves can be affected by the ionic strength, magnesium and the intercalating dye concentration but more importantly the DNA template concentration will affect the melting temperatures and the amplification curves (Zuccarelli *et al*., 2011). Below the results of the DNA samples screened against the control samples are shown. Every screened sample was compared to the control samples for that specific multiplex, by using the theoretical melting temperatures (Table 2.1) as reference when assigning individuals to their ancestral haplogroup.



Melt Curve Results - HRM A.HRM (Page 1)

No.	Name	Genotype	Peak 1	Peak 2	Peak 3
5	L3		73.55	74.15	80.57
6	201		73.52	80.55	

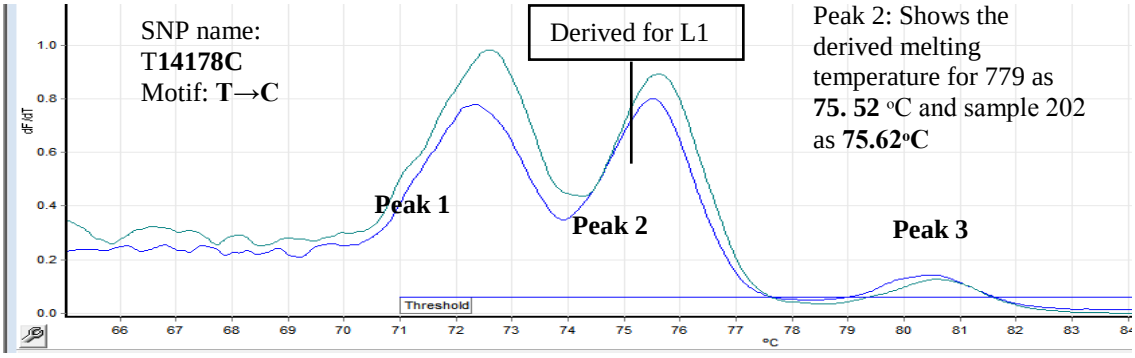
Figure 3. 9 Multiplex 1 is represented by control sample L3 (pink) derived for haplogroup L3, screened against sample 201.

Table 3. 1 The theoretical and melting temperatures of the control L3 sample and the screened sample 201.

	Derived	**Ancestral**
Theoretical temperature	73.5 °C	72.9 °C
Melting temperature (L3)	73.55 °C	74.15 °C
Melting temperature (201)	73.52 °C	80.55 °C

Figure 3.9 shows two DNA melting curves, where the pink melting curve indicates the sequenced control with the derived state for the L3 SNP at position A1018G. The pink melting curve has a Tm of 73.55°C. The blue melting curve

shows the screened sample 201 which melts within the same temperature range as the L3 control, with a Tm of 73.52°C. The difference in temperature is only 0.03°C as seen in Table 3.1 the theoretical temperature for this SNP melts within the same temperature range, which allows for the assignment of sample 201 to the L3 haplogroup having out of Africa or L3* African ancestry.



Melt Curve Results - HRM A.HRM (Page 1)

No.	Name	Genotype	Peak 1	Peak 2	Peak 3
3	779		72.35	75.52	80.50
7	202		72.60	75.62	80.57

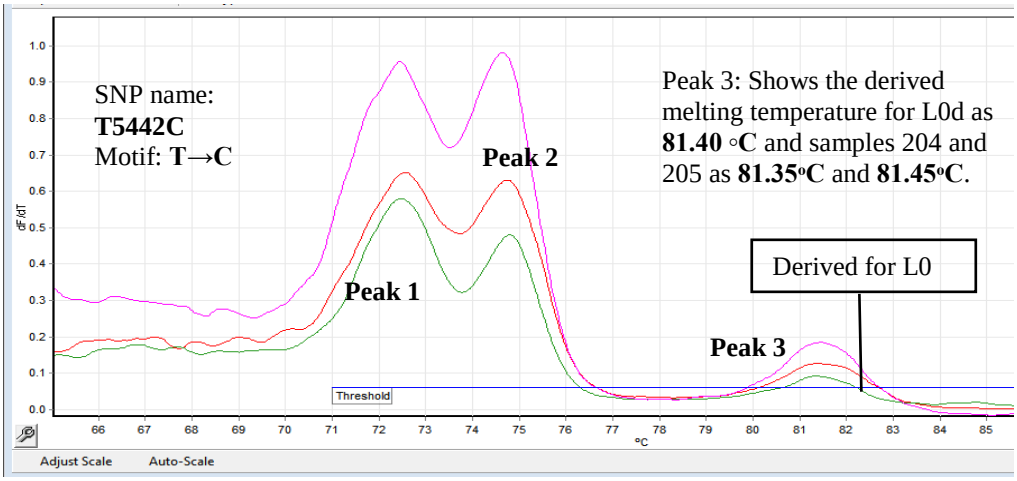
Figure 3. 10 Multiplex 1 is represented by sample 779 (blue) derived for haplogroup L1, screened against sample 202.

Table 3. 2 The theoretical and melting temperatures of the control L1 sample and the screened sample 202.

	Derived	**Ancestral**
Theoretical temperature	75.5 °C	74.5 °C
Melting temperature (L1)	75.52 °C	72.35 °C
Melting temperature (202)	75.62 °C	72.60 °C

Figure 3.10 shows two DNA melting curves, where the blue melting curve indicates the sequenced control with the derived state for the L1 SNP at position T14178C. The blue curve has a Tm of 75.52°C. The green melting curve shows sample 202 which melts within the same temperature range as the L1 control, with a Tm of 75.62°C. The difference in temperature is only 0.1°C as seen in Table 3.2 the theoretical temperature for this SNP melts within the same

temperature range, which allows for the assignment of sample 202 to the L1 haplogroup as having Bantu ancestry.



Melt Curve Results - HRM A.HRM (Page 1)

No.	Name	Genotype	Peak 1	Peak 2	Peak 3
1	L0d		72.55	74.75	81.40
9	204		72.47	74.80	81.35
10	205		72.45	74.65	81.45

Figure 3. 11 Multiplex 1 is represented by L0d (red) derived for L0 haplogroup, screened against sample 204 and 205.

Table 3. 3 The theoretical and melting temperatures of the control L0d sample and the screened samples 204 and 205.

	Derived	**Ancestral**
Theoretical temperature	81.0 °C	80.5 °C
Melting temperature (L0d)	81.40 °C	74.75 °C
Melting temperature (204)	81.35 °C	74.80 °C
Melting temperature (205)	81.45 °C	74.65 °C

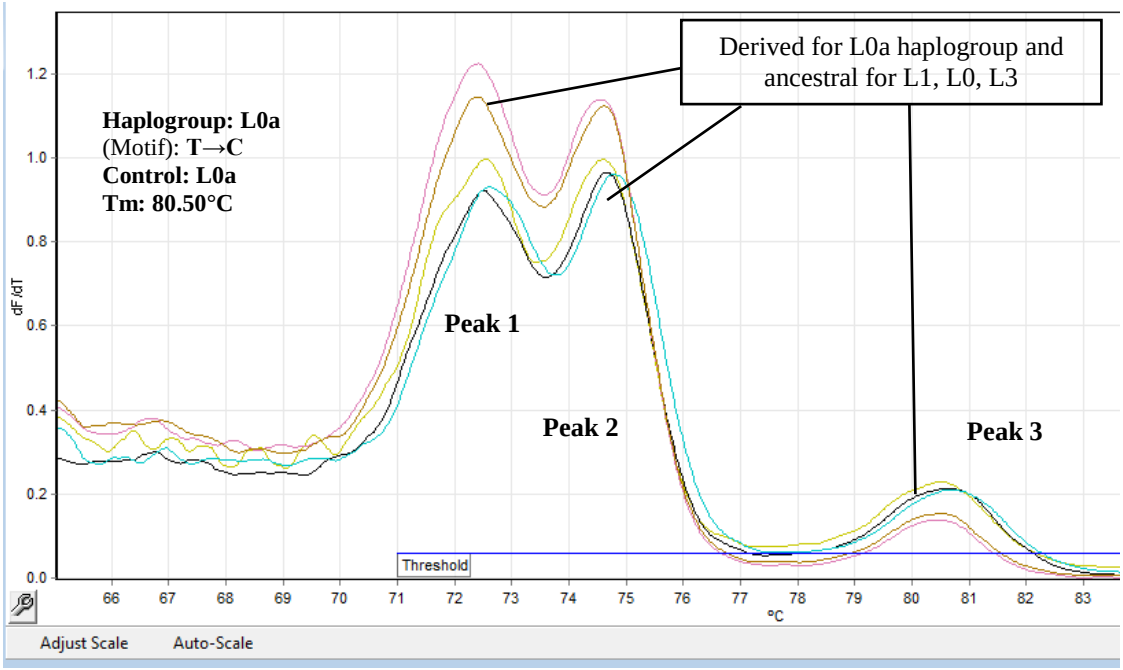
Figure 3.11 shows three DNA melting curves, where the red melting curve indicates the sequenced control with the derived state for the L0 SNP at position T5442C. The red curve has a Tm of 81.40°C. The pink melting curve shows the screened sample 205 which melts within the same temperature range as the L0d control, with a Tm of 81.45°C. The green melting curve shows the screened sample 204 which melts within the same temperature range as the L0d control,

with a Tm of 81.35°C. The difference in temperature for every screened sample compared to the control is -0.05°C and 0.05°C as seen in Table 3.3 the theoretical temperature for this SNP melts within the same temperate range as the control and screened samples, which allows for the assignment of sample 204 and 205 to the L0 haplogroup as having Khoesan ancestry.

The primers were designed in a way to obtain amplicons from the same set of primers with a 0.5 ˚C difference in melting temperatures between alternative amplicons. This meant that within a 0.5°C temperature range the SNP would be classified as ancestral or derived state. As seen above in (Table 3.3) the difference in melting temperatures between the control sample and the screened samples were less than the 0.5°C, which would be the melting difference for classification of the derived state.

If the difference in melting temperature is higher than the 0.5°C criteria, you cannot assign the screened sample to that specific haplogroup. In Figure 3.11 the individual would more likely be derived for the L0 haplogroup and not ancestral, as can see the melting temperature of the screened sample is closer to the L0 control.

Thus, the Multiplex 1 system is easy to understand once the sequenced controls are compared to each screened sample. Considering the temperature shifts between samples due to sample quality, however if the screened sample melts within the temperature range of theoretical and sequenced controls, the samples can be assigned to the specific haplogroup.



Melt Curve Results - HRM A.HRM (Page 1)

No.	Name	Genotype	Peak 1	Peak 2	Peak 3	Peak 4
2	L0a		72.55	74.60	80.50	
11	206		72.50	74.67	80.60	
12	207		72.60	74.80	77.65	80.68
19	208		72.40	74.57	80.45	
21	210		72.40	74.62	80.52	

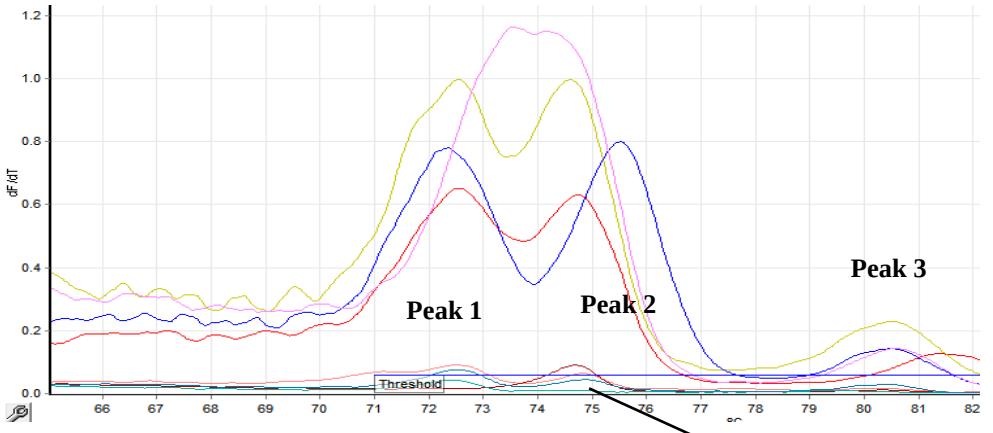
Figure 3. 12 Samples were screened using Multiplex 1 with control L0a and cannot be assigned as being derived for any haplogroup (L1, L0 or L3) thus Khoesan ancestry and ancestral for L1, L0 and L3.

Table 3. 4 The theoretical and melting temperatures of the control L0a and the screened samples 206, 207, 208 and 210.

	Derived	**Ancestral**
Theoretical temperature	81.0 °C	80.5 °C
Melting temperature (L0a)	74.60 °C	80.50 °C
Melting temperature (206)	74.67 °C	80.60 °C
Melting temperature (207)	74.80 °C	77.75 °C
Melting temperature (208)	74.57 °C	80.45 °C
Melting temperature (210)	74.62 °C	80.52 °C

Figure 3.12 shows the results of the screened samples that melts within the same temperature range of the sequenced control L0a at 80.50°C, as seen in Table 3.4 each screened sample's temperature range melts within the temperature range as the theoretical temperature for this specific SNP and cannot be assigned to the L0 haplogroup as derived. The reason why these samples cannot be assigned to the L0 haplogroup is because of the second derivate that differentiates between L0d derived (Khoesan) and L0a ancestral (Khoesan) ancestry.

Thus, each screened sample (206, 207, 208 and 210) melts within the ancestral melting Tm range and can be assigned to the L0a haplogroup, these samples are thus ancestral for all the haplogroups (L1, L0 and L3) within Multiplex 1. If screened samples are defined as ancestral for all the SNPs within Multiplex 1, a further screening would be necessary to assign L0a derived individuals to any other haplogroups within African ancestry. This however was not done as it did not fall within the scope of the study.



Melt Curve Results - HRM A.HRM (Page 1)

No.	Name	Genotype	Peak 1	Peak 2	Peak 3
1	L0d		72.55	74.75	81.40
2	L0a		72.55	74.60	80.50
3	779		72.35	75.52	80.50
5	L3		73.55	74.15	80.57
8	203		72.50		
20	209		74.70		
23	213				
24	215		72.53		

Figure 3. 13 Samples were screened using Multiplex 1 and cannot be assigned to any haplogroup due to the melting temperatures being too low to detect.

Figure 3.13 Shows the screened samples for Multiplex 1 that cannot be assigned to any haplogroup as the peaks are too low to detect. This can be an indication of sample degradation. It is important to consider the discrepancy that expected temperature values will not always be the exact same value as the theoretical temperature values due to a temperature shifts of samples which could be because of sample degradation. Samples 203, 209, 213 and 215 presented this incidence.

Zucarelli *et al.,* (2011) reported that the DNA template concentration affects the melting curve and the amplification curve. Thus, it was important to use quantified DNA template in every reaction tube (Zucarelli *et al*., 2011). Degraded samples would obscure the data and the melting temperatures would be too low to detect. If this occurred, the degraded samples were re-extracted, re-quantified and re-amplified. Inferring an ancestral haplogroup by analysing the melting curves for each sample allows for haplogroup identification for every individual under investigation. Below an automated method to assign individuals to their ancestral haplogroups is presented.

3.4 Screen cluster analysis for Multiplex 1

RotorGene® ScreenClust software was used to analyse and confirm the visual results from the melting curve analysis for Multiplex 1. This software groups the screened samples and the control samples in clusters based on their melting temperatures. The analysis will cluster each Limpopo sample that was genotyped, according to the specific haplogroup it can be assigned to.

Principal component analysis is an automated method used in SNP studies to evaluate human population structure (Patterson *et al*., 2006). This method is advantageous as it can be performed quickly, it is beneficial in identifying haplogroup clusters and markers that effectively describes differences in populations (Paschou *et al*., 2007).

PCA clustering: PCA clustering was used as an automated method to confirm the HRM results and as a second method to infer individuals to their ancestral haplogroups, by confirming the screened and control samples, the haplogroups

shown in clusters were identified. The same screened samples we show in the HRM results above is shown in the PCA results below.

3.4.1 Multiplex 1: Haplogroup clusters: Supervised mode

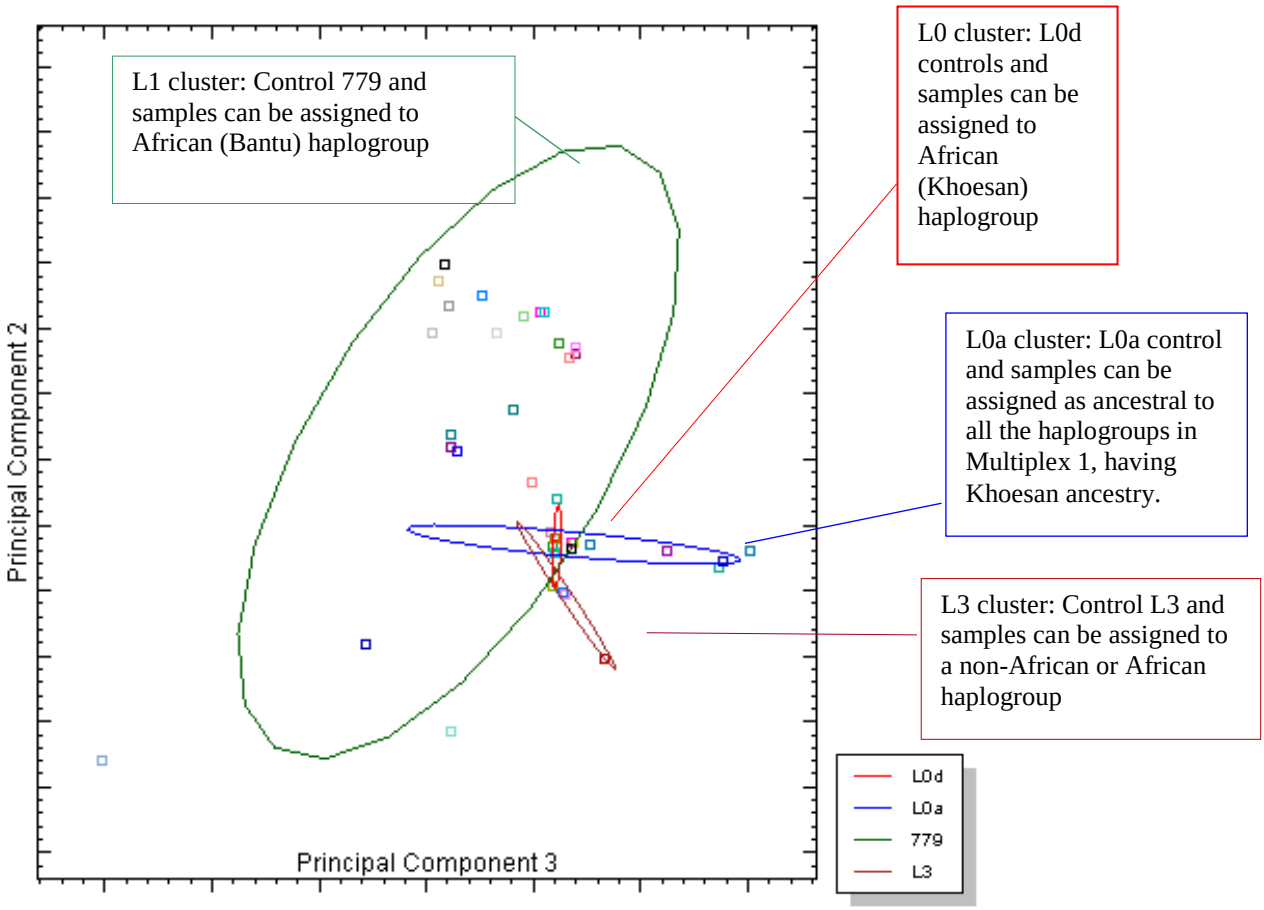


Figure 3. 14 Supervised cluster analysis for Multiplex 1. The clusters are separated in red (L0d), blue (L0a), green (L1) and maroon (L3) according to the key.

The supervised mode clusters the unknown samples into known clusters. The cluster centres are set up as the mean of the controls and the relatedness of the unknown samples that are allocated to clusters are based on how close these samples are to the mean points of the controls (QIAGEN, 2009, Reja *et al*., 2010).

Majority of the samples fall within the L1 cluster having Bantu African ancestry for this specific run. This is expected as it was shown that majority of the samples collected are individuals from African populations (Figure 3.1) that speak African languages. A few samples are clustered in L3 and would need to be screened further

with Multiplex 2 to establish African (L3*) or non-African descent (M and N). There are samples clustered within the L0 cluster having Khoesan ancestry. The fourth cluster is shown as being ancestral to the three SNPs within Multiplex 1, being Khoesan and these samples are clustered as ancestral.

Two of the outliers are non-template controls which would be expected to not fall in any cluster due to these samples not containing any template to amplify as discussed previously according to Zucarelli *et al.,* (2010) the DNA template concentration affects the melting curve and the amplification curve. Thus, it was important to use quantified DNA template in every reaction tube (Zucarelli *et al*., 2011). This would ensure that the melting curves are detected.

The two other outliers are samples that visually shows melting temperatures below the threshold thus could be out of the cluster due to PCA methods being sensitive to missing data, which effects the samples (Helyar *et al.,* 2011). This automated method of inferring individuals to population groups can limit the inference about underlining historical and demographic processes (Novembre *et al.,* 2008).

The detection of outliers in our study is truly subject to selection, which means the goal was to identify markers with a high discrimination power between different populations, thus it would not influence the results if outliers are present. The outliers can be an explorative exercise supporting downstream analyses (Helyar *et al*., 2011).

3.4.2 Multiplex 1: Haplogroup clusters: Unsupervised mode

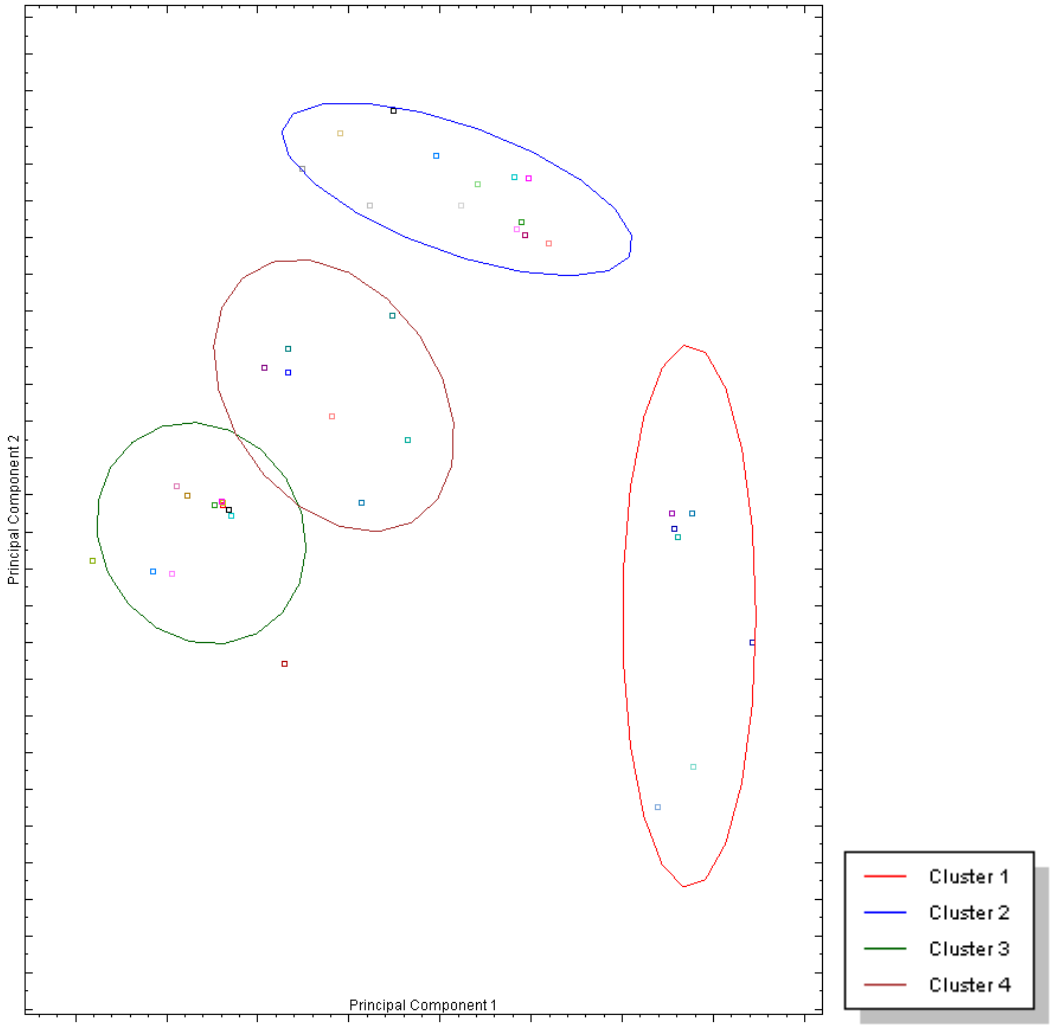


Figure 3. 15 Shows the unsupervised cluster analysis for Multiplex 1. The clusters are separated as red (Cluster 1), blue (Cluster 2), green (Cluster 3) and maroon (Cluster 4).

The unsupervised mode determines the ideal number of clusters (QIAGEN, 2009: Reja *et al*., 2010). Figure 3.15 shows samples clustered within 4 clusters. Majority of the samples clustered with unsupervised mode falls within the blue cluster. The other samples are distributed within the red, green and maroon cluster.

Comparing the results of the supervised and unsupervised mode of clustering, the supervised mode clusters the unknown within known which is a biased approach but still effective to confirm the visual analysis for Multiplex 1. The supervised

mode of clustering is sensitive and specific and would not cluster samples that detect below the threshold, more outliers are observed.

The unsupervised mode is unbiased and automatically clusters samples that are unknown and determines ideal clusters. Based on the output, this automated method of assigning individuals to their ancestral groups varies and shows only two samples that cannot be clustered.

3.4.3 Multiplex 1: Haplogroup clusters: Supervised mode

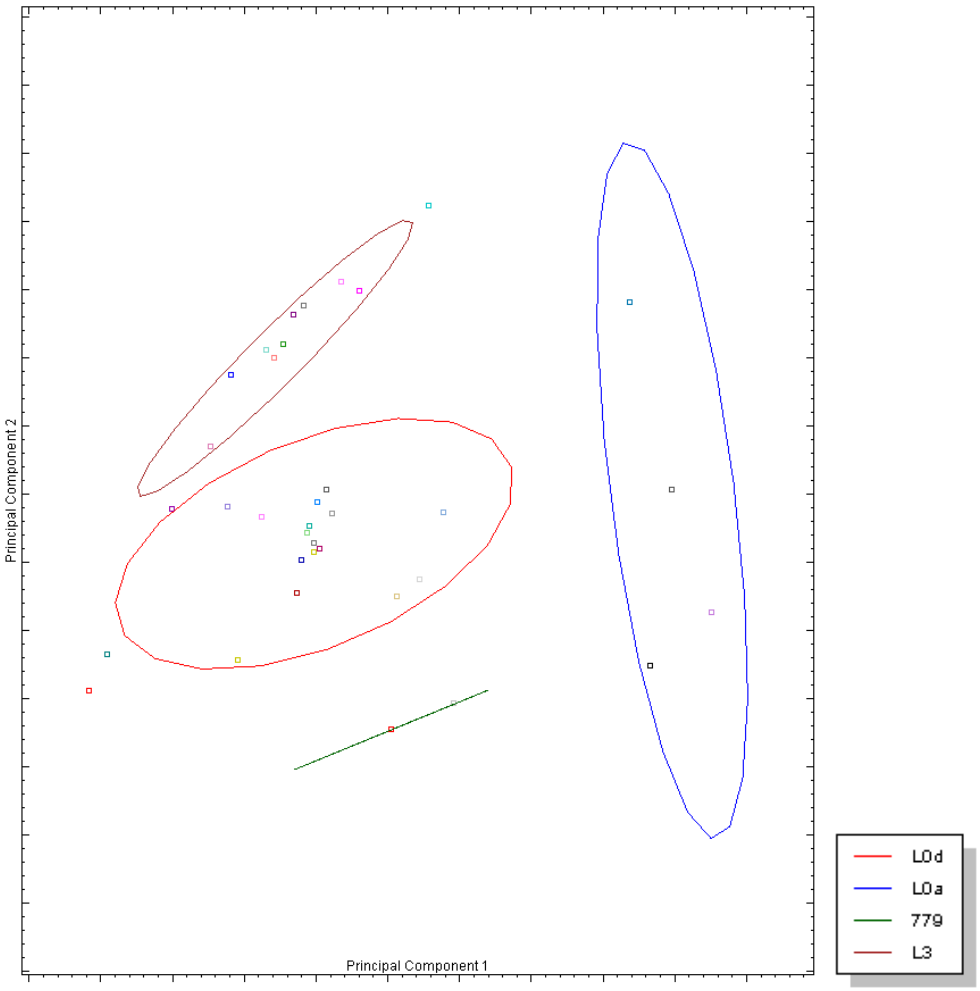


Figure 3. 16 Supervised cluster analysis for Multiplex 1. The clusters are separated in red (L0d), blue (L0a), green (L1) and maroon (L3) according to the key.

Figure 3.16 shows the supervised clustering of samples genotyped with Multiplex 1. With this run majority of the samples are clustered within the L0d cluster as

having Khoesan ancestry and the L3 cluster as having African (L3*) or non-African descent. Only 4 samples are clustered and assigned as being ancestral for all the SNPs of Multiplex 1. Two samples are shown to cluster with control sample 779 (L1) haplogroup with Bantu ancestry. Four samples are shown as outliers.

3.4.4 Multiplex 1: Haplogroup clusters: Unsupervised mode

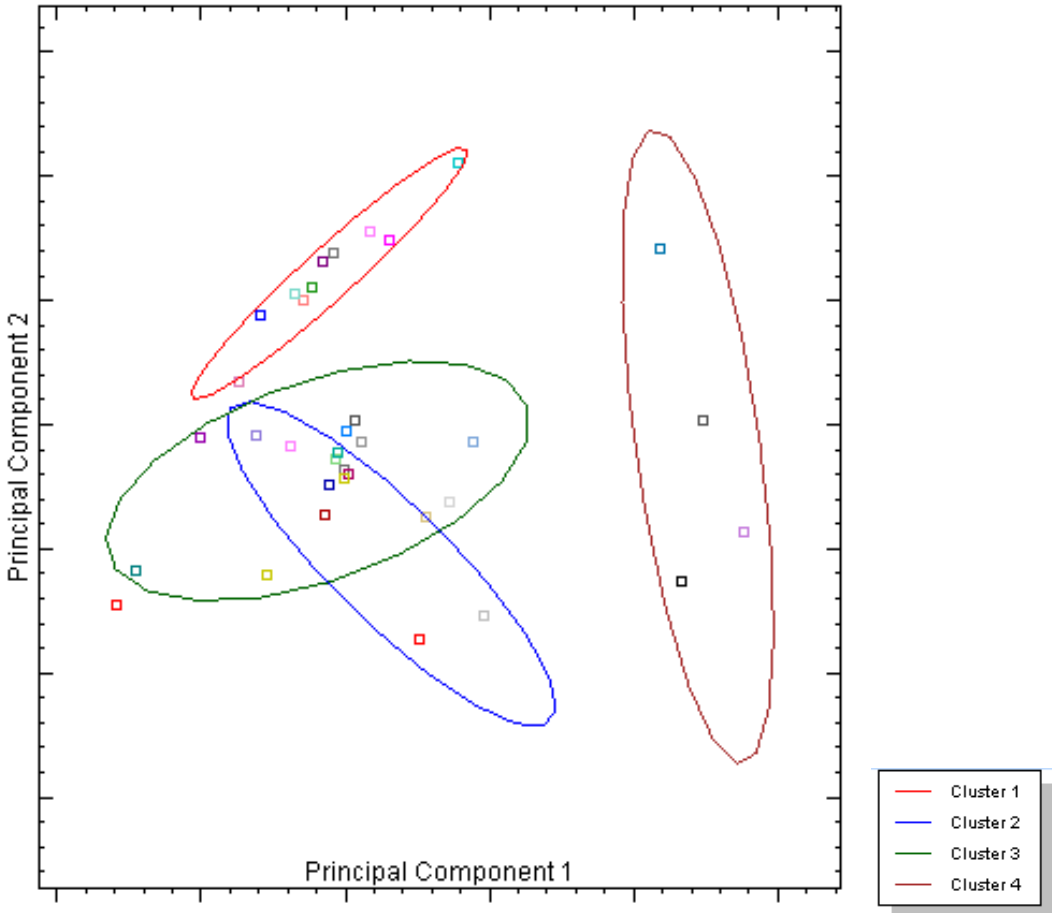


Figure 3. 17 Shows the unsupervised cluster analysis for Multiplex 1. The clusters are separated as red (Cluster 1), blue (Cluster 2), green (Cluster 3) and maroon (Cluster 4).

The unsupervised mode of clustering the samples for the same run as the above-mentioned presents four clusters. The blue and green cluster overlaps and clusters majority of the samples, whereas Figure 3.15 clusters majority of the samples within the red cluster. Once again only two samples are shown as outliers. Four samples assigned to the maroon cluster.

In a study done by Schuster *et al*., (2010) a population wide PCA was done which separates the Khoesan from the western and southern Bantu populations but clearly clusters the Sotho-Tswana representative as part of the Southern Bantu cluster. They predict gene flow between Khoesan ancestors and Southern Bantu confirming L0 type mitochondria. This would support our PCA results as seen in Figure 3.16, showing that majority of our samples are clustered within L0d haplogroup.

In summary PCA can be used as an automated method to confirm the visual analysis of the individuals assigned to haplogroups with HRM. Based on these results it can be shown that majority of the screened samples from Limpopo have African descent, comprising of Southern Bantu-, Khoesan contribution to the population structure and ancestry of these individuals.

3.5 Algorithm to confirm haplogroup confirmation

	Peak 1	Peak 2	Peak 3		31.05.16	L0-L3	Pred L0	L3_D-L3	Pred L3	L1_D- L1	Pred L1
L0d control		72.55	74.75	81.4		8.85	L0	1	0	0.77	0
L0a control		72.55	74.6	80.5		7.95	0	1	0	0.92	0
779 control (L1)		72.35	75.52	80.5		8.15	0	1.2	0	0	L1
L3		73.55	74.15	80.57		7.02	0	0	L3	1.37	0
202	72.6	75.62	80.57	L1		7.97	0	0.95	0	-0.1	L1
204	72.47	74	81.35	L0		8.88	L0	1.08	0	1.52	0
205	72.45	74.65	81.45	L0		9	L0	1.1	0	0.87	0
206	72.5	74.67	77.65	A		5.15	0	1.05	0	0.85	0
207	72.6	74.8	77.65	A		5.05	0	0.95	0	0.72	0
208	72.4	74.57	80.45	A		8.05	0	1.15	0	0.95	0
210	72.4	74.62	80.52	A		8.12	0	1.15	0	0.9	0

Figure 3. 18 Shows our prediction table used as an algorithm to assign haplogroups.

A mathematical algorithm that confirms the screened samples, PCA analysis and predicted haplogroups as an additional confirmation test was designed. The algorithm is based on the theoretical melting temperatures for each haplogroup and predicts expected haplogroups, the formulas is shown in Chapter 2 (Table 2.4).

Figure 3.18 shows the prediction table that can be used as an additional confirmation test, after the visual analysis and automated assignment has been done. The confirmed results shown in Multiplex 1 is shown in Figure 3.18. Sample 202, shown with the red arrow is confirmed to be derived for the L1 haplogroup as seen

in Figure 3.10. Samples 204 and 205 shown with the blue arrows are confirmed to be derived for the L0 haplogroup as seen in Figure 3.11.

The samples 206, 207, 208 and 210 shown with purple arrows was assigned as ancestral to haplogroups L1, L3 and L0 in Multiplex 1 (Figure 3.12). This assignment with the prediction table as shown in Figure 3.18 confirms the controls and the screened samples for Multiplex 1.

3.6 HRM screening to identify haplogroups

KAPA™ HRM FAST PCR kit was used for fast, high-throughput post PCR analysis of the sequence variants. The KAPA™ HRM FAST ready master mix contained a novel DNA polymerase, a buffer and the EvaGreen® saturating dye which is optimized for maximum discrimination power between the sequence variants (Kapabiosystems, 2016).

Five hundred Limpopo samples were genotyped on the Rotor-Gene Q (QIAGEN) real time thermocycler for High Resolution Melting analysis. A total of 95 individuals derived for the L3 haplogroup which gives a 19% ratio of individuals was confirmed. Only 56 individuals derived for the L1 haplogroup which gives a percentage of 11%. A total of 249 individuals were derived for the L0 haplogroup, making up a 50%. Only 100 samples were L0a (Khoesan) which are ancestral for all three SNPs making up 20% of samples screened with Multiplex 1.

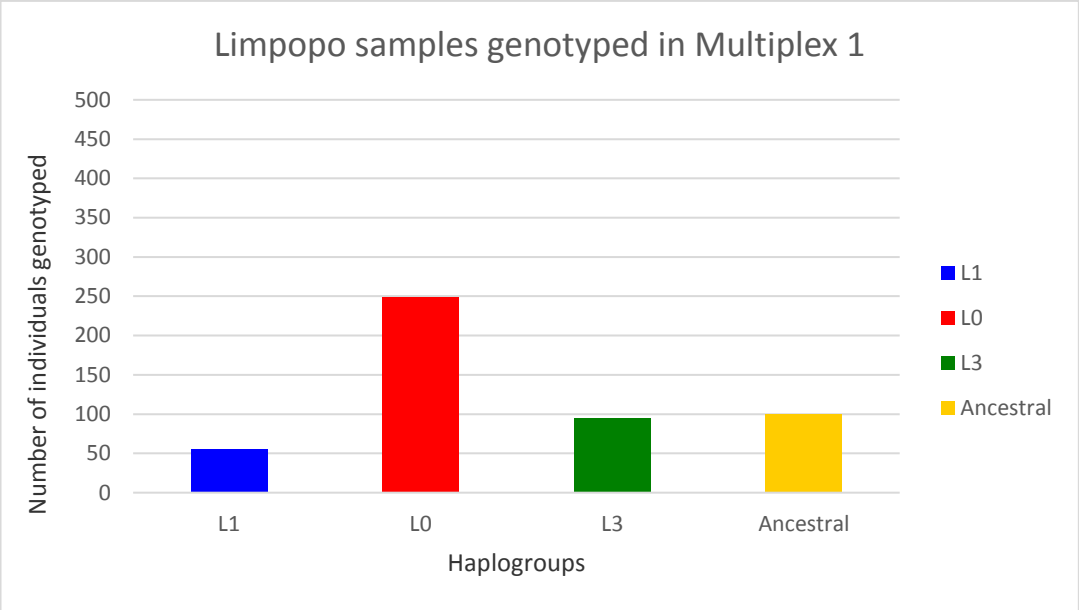


Figure 3. 19 Bar chart presents the number of samples genotyped, comparing the number of Limpopo individuals assigned to a specific haplogroup in Multiplex 1.

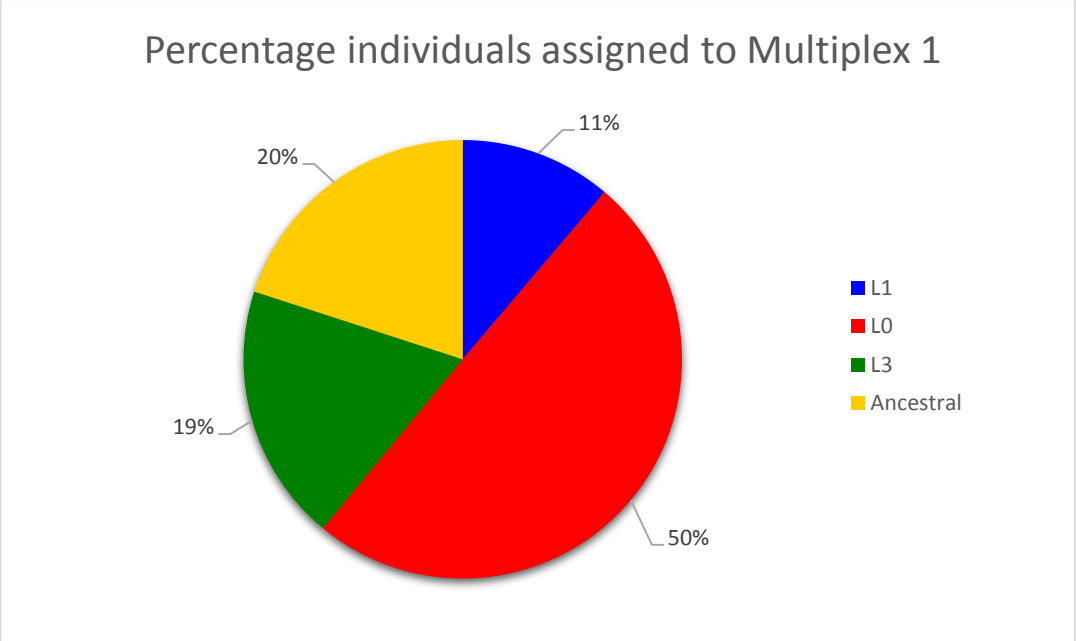


Figure 3. 20 Pie chart presents the percentage of individuals assigned to haplogroups L0, L1 and L3 and individuals who are ancestral for these three haplogroups.

Figure 3.19 and Figure 3.20 shows the categorical data of 500 Limpopo samples genotyped. The assigned individuals are more likely to be derived for the L0

haplogroup based on the results presented by the bar- and pie chart, it shows that most of the samples have African (Khoesan) descent.

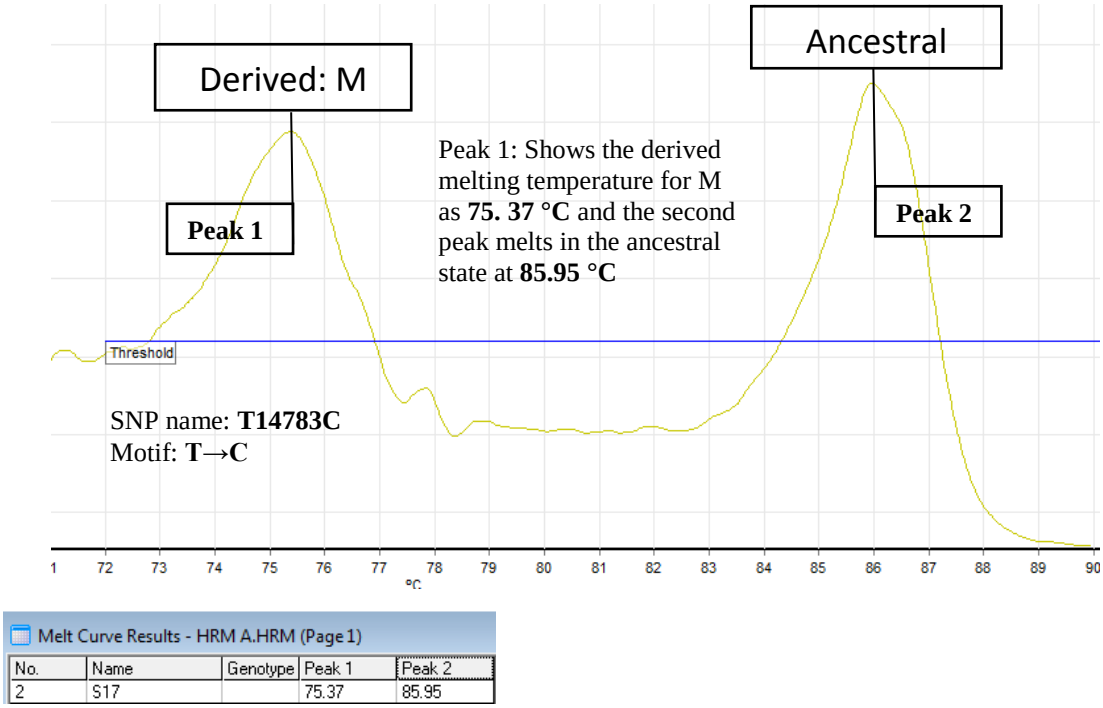
This would be expected because of the contribution of the hunter-gatherers and admixture that occurred between Khoesan and Bantu population expansions and growth. The hunter-gathers and others have contributed to the genetic adaptions of an agricultural lifestyle (Schuster *et al*., 2010).

3.7 HRM ancestral testing for Multiplex 2

As discussed in Chapter 2, Multiplex 1 was designed to assign individuals to a specific haplogroup and used as the first screening. This is the first multiplex to differentiate whether individuals have African (L3*) or out of Africa (M and N) descent. The multiplex is structured in a hierarchal arrangement. The first multiplex looks at the main branches of the mtDNA phylogenetic tree. Thus, the results of Multiplex 1 will determine Multiplex 2.

The 95 individuals from Multiplex 1, derived for haplogroup L3 (non-African descent or African descent) was then screened with Multiplex 2. Multiplex 2 consists of two informative SNPS under investigation, M: T14783C, N: G8701A (Refer to table 2.1), which gives an indication of non-African (M and N) or African (L3*) descent. The L3* branch would refer to (L2, L4, L5 or L6).

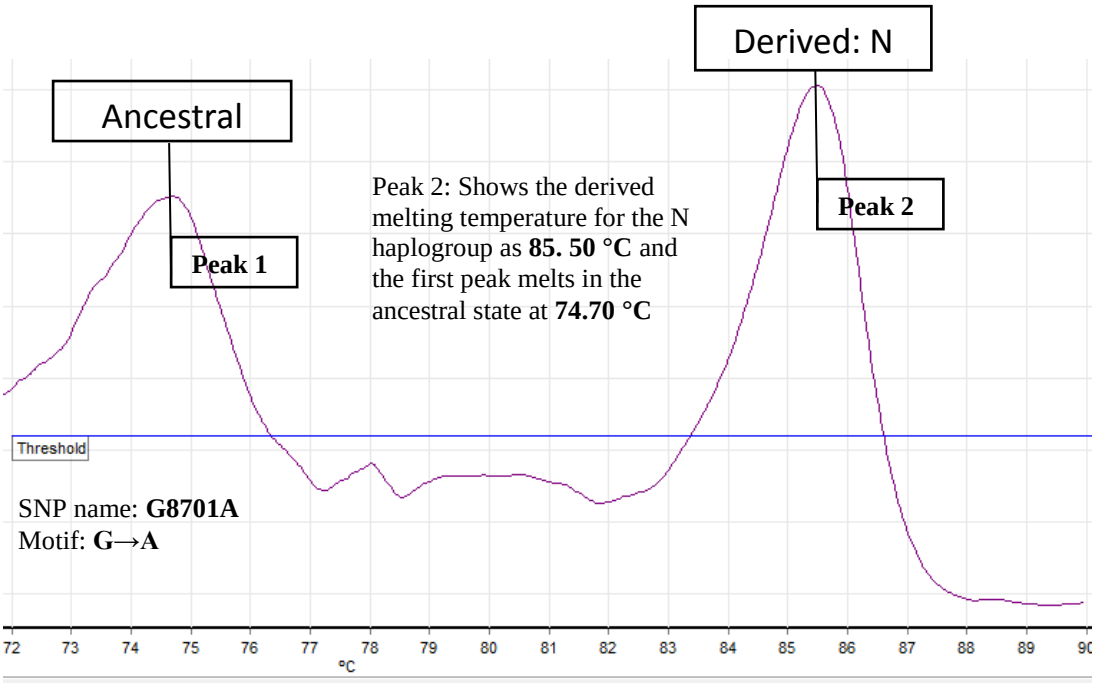
Each SNP under investigation at a specific site was tested, this would establish whether the SNP under investigation is derived or ancestral within a given state for a specific haplogroup, the same principle applies as mentioned in the previous section with Multiplex 1. As stated in Chapter 2 (Table 2.1), the theoretical temperatures for each SNP is presented and used as reference to confirm the melting temperatures for Multiplex 2. The melting temperature difference between every amplicon would be classified as the ancestral or derived state.



Melt Curve Results - HRM A.HRM (Page 1)

No.	Name	Genotype	Peak 1	Peak 2
2	S17		75.37	85.95

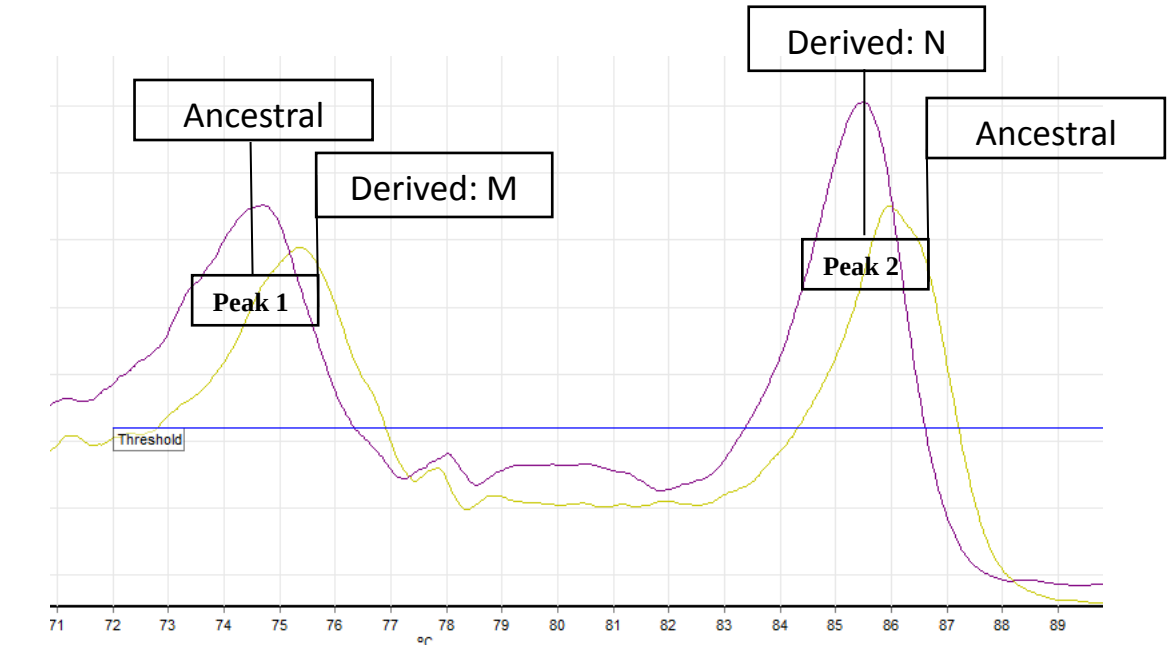
Figure 3. 21 Multiplex 2 is represented by control sample S17 (yellow) derived for the M haplogroup.



Melt Curve Results - HRM A.HRM (Page 1)

No.	Name	Genotype	Peak 1	Peak 2
4	Pet		74.70	85.50

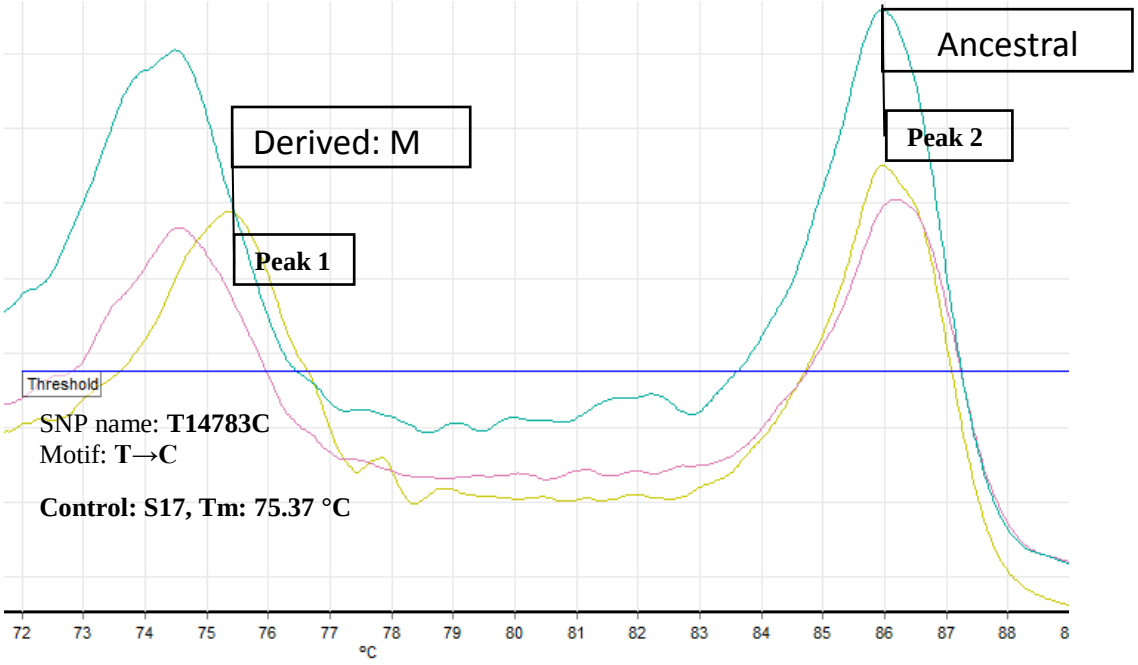
Figure 3. 22 Multiplex 2 is represented by control sample LM-P (purple) derived for the N haplogroup



Melt Curve Results - HRM A.HRM (Page 1)

No.	Name	Genotype	Peak 1	Peak 2
2	S17		75.37	85.95
4	Pet		74.70	85.50

Figure 3. 23 Multiplex 2 is presented by the two control samples S17 (yellow), and LM-P (purple) of SNPS which are derived for haplogroup M and N.



Melt Curve Results - HRM A.HRM (Page 1)

No.	Name	Genotype	Peak 1	Peak 2
2	S17		75.37	85.95
19	77		74.57	86.17
55	754		74.50	85.95

Figure 3. 24 Multiplex 2 is represented by sample S17 (yellow) derived for haplogroup M, screened against sample 77 and 754.

Table 3. 5 The theoretical and melting temperatures of the control S17 sample and the screened samples 77 and 754.

	Derived	**Ancestral**
Theoretical temperature	75°C	74 °C
Melting temperature (S17)	75.37 °C	85.95 °C
Melting temperature (77)	74.57 °C	86.17 °C
Melting temperature (754)	74.50 °C	85.95 °C

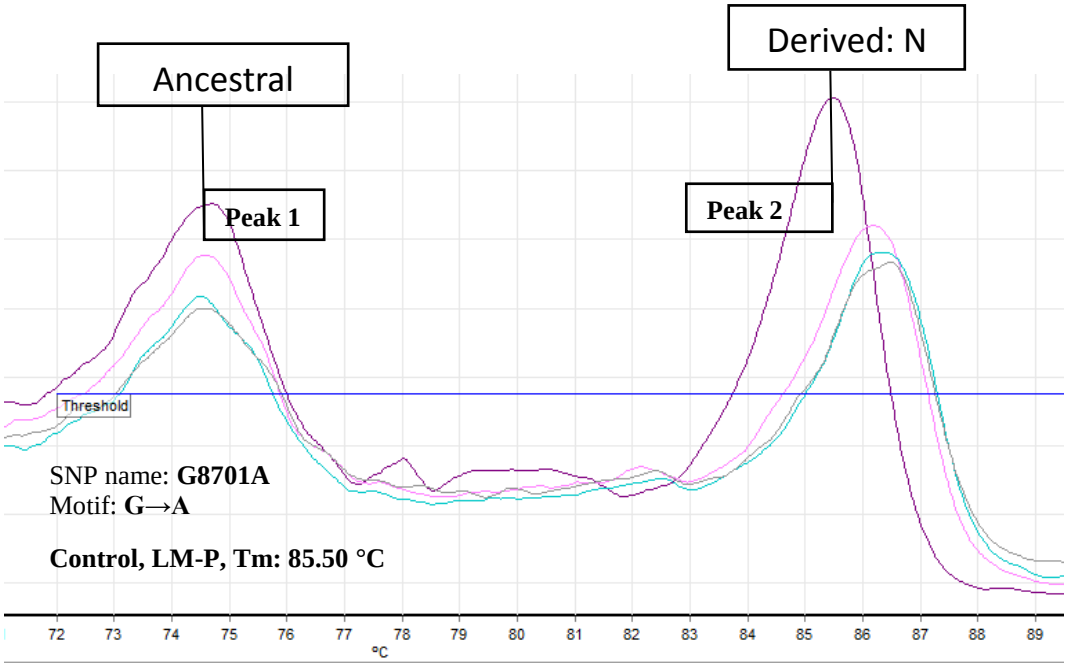
Figure 3.24 shows three DNA melting profiles, where the yellow melting curve indicates the sequenced control with the derived state for the M SNP at position T14783C. The yellow curve has a Tm of 75.37°C. The pink melting curve shows the screened sample 77, does not melt within the same temperature ranges as the S17 control, with a Tm of 74.57°C. The green melting curve shows the screened sample 754 which not melt within the same temperature ranges as the S17 control, with a Tm of 74.50°C.

The primers were designed in a way to obtain amplicons from the same set of primers with a 0.5 ˚C difference in melting temperatures between alternative amplicons. This meant that within a 0.5°C temperature range the SNP would be classified as ancestral or derived state. As seen above (Table 3.5) the difference in melting temperatures between the control sample and the screened samples was more than the 0.5°C Tm range for classification of the derived state.

The difference in temperature for every screened sample compared to the control is 0.55°C and 0.5°C as seen in Table 3.5 the theoretical temperature for this SNP melts

within the same temperate range as the control sample with a temperature difference of -0.37 °C. However, the screened samples (77 and 754) cannot be assigned to haplogroup M as the Tm's are not in the same Tm range of the S17 control and the theoretical temperature of the SNP.

If the difference in temperature is higher or equal to the 0.5°C Tm range, we cannot assign the screened sample to that specific haplogroup. In figure 3.23 the individual would more likely be ancestral for the M haplogroup and not derived, as seen the melting temperature of the screened samples are closer to the ancestral theoretical Tm for the SNP.



Melt Curve Results - HRM A.HRM (Page 1)

No.	Name	Genotype	Peak 1	Peak 2
4	Pet		74.70	85.50
5	78		74.58	86.18
12	136		74.50	86.32
30	658		74.57	86.47

Figure 3. 25 Multiplex 2 is represented by sample LM-P (purple) derived for haplogroup N, screened against sample 78, 136 and 658.

Table 3. 6 The theoretical and melting temperatures of the control LM-P sample and the screened samples 78, 136 and 658.

	Derived	Ancestral
Theoretical temperature	85 °C	86.7 °C
Melting temperature (LM-P)	85.50 °C	74.70 °C
Melting temperature (78)	86.18 °C	74.58 °C
Melting temperature (136)	86.32 °C	74.50 °C
Melting temperature (658)	86.47 °C	75.57 °C

Figure 3.25 shows four DNA melting profiles, where the purple melting curve indicates the sequenced control with the derived state for the N-SNP at position G8701A. The purple curve has a Tm of 85.50 °C. The pink melting curve shows the screened sample 78, does not melt within the same temperature range as the LM-P control, with a Tm of 86.18 °C. The blue melting curve shows the screened sample 136, does not melt within the same temperature range as the LM-P control, with a Tm of 86.32 °C. The grey melting curve shows the screened sample 658 which, does not melt within the same temperature range as the LM-P control, with a Tm of 86.47 °C.

The difference in temperature for every screened sample compared to the control is 0.68 °C, 0.82 °C and 0.97 °C as seen in Table 3.6 the theoretical temperature for this SNP melts within the same temperate range as the control sample with a temperature difference of 0.5 °C. However, the screened samples (78, 136 and 658) cannot be assigned to haplogroup N as the Tm's are not in the same Tm range of the LM-P control and the theoretical temperature of the SNP.

The primers were designed in a way to obtain amplicons from the same set of primers with a 0.5 ˚C difference in melting temperatures between alternative amplicons. This meant that within a 0.5°C temperature range the SNP would be classified as ancestral or derived state. As seen above (Table 3.6) the difference in melting temperatures between the control sample and the screened samples were more than the (0.5°C) standard temperature criteria for classification of the derived state.

If the difference in temperature is higher or equal to the 0.5°C standard, we cannot assign the screened sample to that specific haplogroup. In figure 3.25 the LM-P control does however melt higher than the expected theoretical temperature with a 0.5°C difference. However, the samples all do not melt within the Tm range of the control and the theoretical Tm of the SNP. The individuals would more likely be ancestral for the N haplogroup and not derived, as seen the melting temperature of the screened samples are closer to the ancestral theoretical Tm of the control sample.

3.8 Screen cluster analysis for Multiplex 2

The 95 individuals showed to be ancestral for haplogroup M and N as seen in the above-mentioned genotyping results for Multiplex 2. PCA clustering: The PCA clustering was used as an automated method to confirm the HRM results and as a second method to infer individuals to their ancestral haplogroups, thus confirmed screened and control samples and identified the haplogroups shown in clusters. The same screened samples shown in the HRM results above was used in PCA analysis for Multiplex 2.

3.8.1 Multiplex 2: Haplogroup clusters: Supervised mode

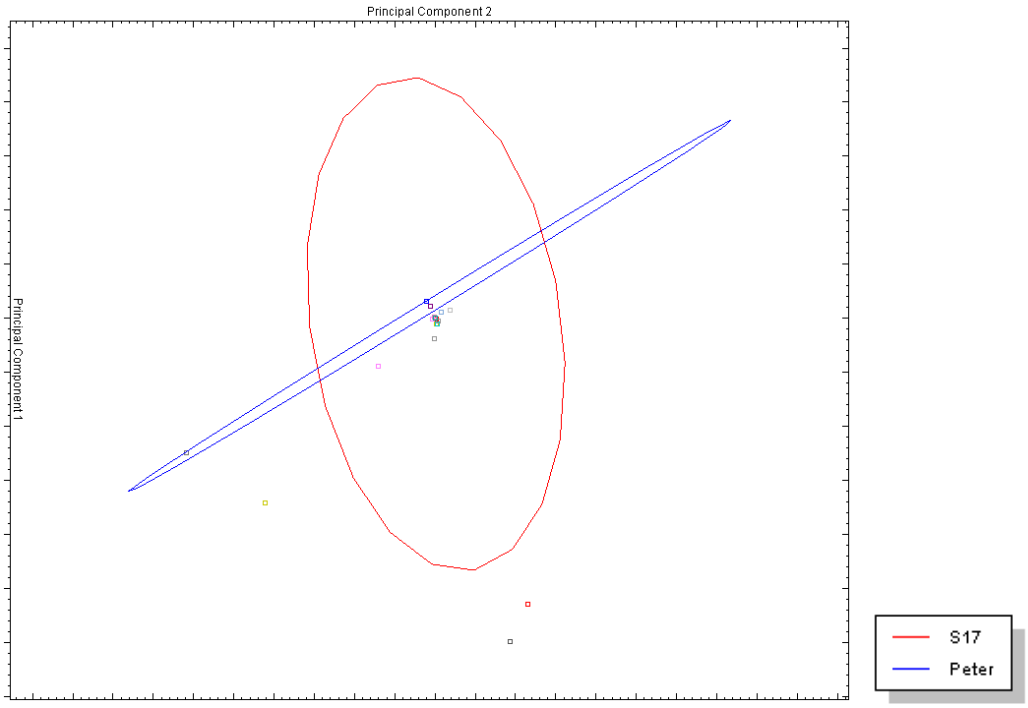


Figure 3. 26 Supervised cluster analysis for Multiplex 2. The clusters are separated in red (S17) haplogroup M, blue (LM-P) haplogroup N according to the key.

Figure 3.26 shows the supervised clustering of samples genotyped with Multiplex 2. With this run majority of the samples fall outside of the (LM-P) cluster and are clustered with the S17 cluster. The supervised mode does not coincide to the visual analysis. Three samples are shown to cluster with control (LM-P) M haplogroup which confirms the controls for Multiplex 2. Three samples are shown as outliers.

3.8.2 Multiplex 2: Haplogroup clusters: Unsupervised mode

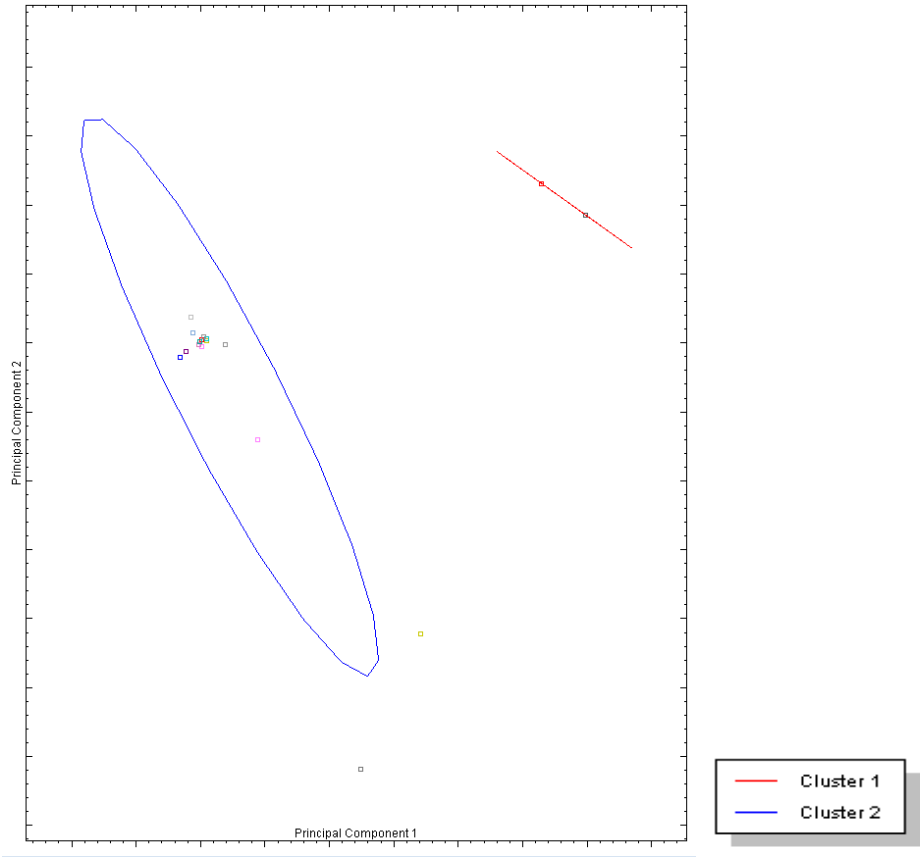


Figure 3. 27 Shows the unsupervised cluster analysis for Multiplex 2. The clusters are separated as red (Cluster 1), blue (Cluster 2).

Figure 3.27 shows the unsupervised clustering for Multiplex 2. The unsupervised clusters majority of the samples in Cluster 2 (blue). This would coincide with the visual results for Multiplex 2, all the individuals screened with Multiplex 2 found to be ancestral for Haplogroup M and N. This confirms the screened samples to have African ancestry as they do not fall within haplogroups of out-of-Africa. Two samples fall outside of both clusters, which indicates that these samples were too low to detect.

The two samples that are clustered in Cluster 1, coincide with the control samples that would be clustered as non-African descent. The unsupervised mode has shown to confirm the visual results for Multiplex 2. In summary PCA can be used as an automated method to confirm the visual analysis of the individuals assigned to

haplogroups with HRM. Based on these results, it can be shown that majority of the screened samples from Limpopo have African descent, comprising of Southern Bantu-, Khoesan contribution to the population structure and ancestry of these individuals.

3.9 HRM screening to identify haplogroups

Of the 95 samples derived for the L3 haplogroup, 87 individuals were found to be ancestral for both M and N, which meant that 91.57% of the screened samples have African descent falling in L3* branches. Further screening would need to be done to establish which African haplogroups (L2, L5, L6 or L4) the 87 individuals can be assigned to.

None of the individuals screened with Multiplex 2, showed to be derived for haplogroup M and N. Eight samples couldn't be assigned as the melting curves were too low to detect. Multiplex 2 confirms the results found in Multiplex 1, which validates that at least 78.4% of our 500 Limpopo samples have African descent as we expected with Khoesan and Southern Bantu contribution.

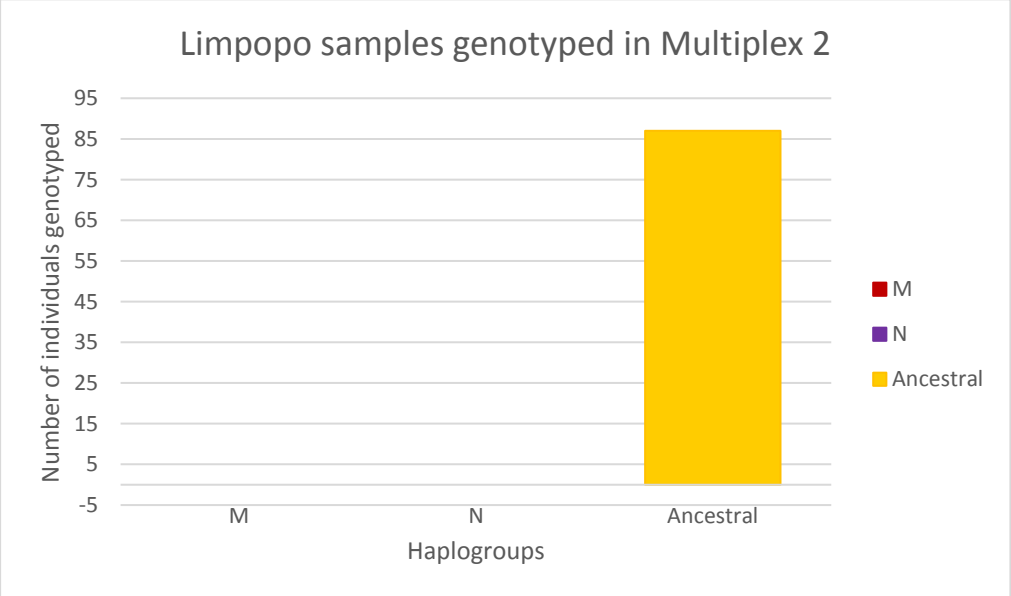


Figure 3. 28 Bar chart presents the number of samples genotyped, comparing the number of Limpopo individuals assigned to a specific haplogroup in Multiplex 2.

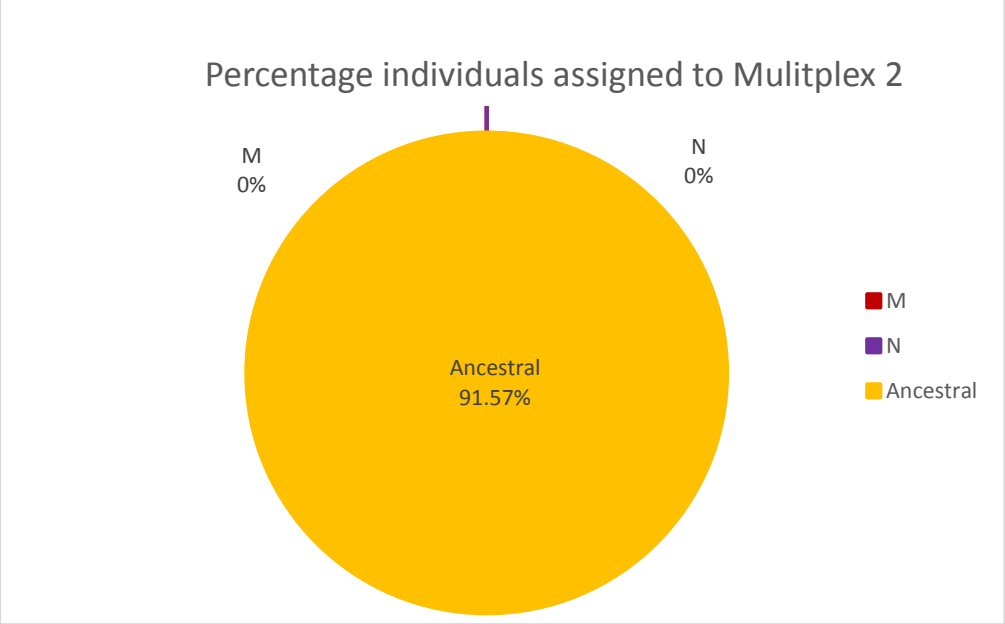


Figure 3. 29 Pie chart presents the percentage individuals assigned as ancestral for haplogroup M and N in Multiplex 2.

Figure 3.28 and Figure 3.29 shows the categorical data of 95 Limpopo samples genotyped with Multiplex 2 as a second screening. The assigned individuals have African ancestry as shown by the results presented in the bar- and pie chart.

This would be expected because of the contribution of the hunter-gatherers and admixture that occurred between Khoesan and Bantu population expansions and growth. The Bantu migrations have contributed to majority of the African history (Rosa and Brehm, 2011), because the hunter-gatherers had to transition to an agricultural lifestyle and iron-melting (Newman, 1995, Phillipson, 1993). Population growth and expansion occurred due to spatial expansion following the migration southwards (Rosa and Brehm, 2011).

The genetic evidence, linguistic and archaeological evidence support the theory that Bantu-speakers spread gradually, and gene flow and strong interactions occurred with other populations on route (Ehret, 2001, Tishkoff *et al*., 2007, Castri *et al*., 2009). The arrival of Bantu-speaking individuals within Southern regions of Africa, could have led to isolation or admixture with pastoralists and hunter-gatherers (Patin *et al*., 2014).

This would have contributed to the population structure and composition of the populations in Limpopo, because of the gene flow and migration southwards towards Limpopo regions. The in-house multiplex system shows a L0 (Khoesan) and L1 (Bantu) contribution to the genetic pool of Limpopo populations and the other individuals who are not identified by this genotyping system.

Research has shown the Bantu expansion and interactions between inhabitants has affected other regions in Southern Africa, specifically where Bantu-speaking groups were present (Patin *et al*., 2014). The interactions between residents and inhabitants were different in different regions, which has resulted in the admixture and different degrees of introgression (Pickrell *et al*., 2012, 2014).

The genetic differences may thus guide future studies to whether population replacement, rather than cultural exchange occurred. This has driven the expansion and growth of agriculture in the Southern regions of Africa (Berniell-Lee *et al*., 2009).

Individuals who were ancestral for (M and N), were thus derived for L3*, this meant these individuals have (L2, L4, L5 or L6) ancestry, north and eastern African origin. Another possibility of individuals being ancestral for M and N, could be because of reversal mutations, mutating back to the ancestral state. These individuals are possibly assigned to haplogroups somewhere else on the mtDNA phylogenetic tree.

3.10 Comparison of assignment methods used in conjunction with our multiplex system.

Table 3. 7 Comparison of the haplogroup assignment methods

Haplogroup	Manual assignment (HRM)	Automated assignment (PCA)	Algorithm (Prediction table)
Hg L0	249	201	49
Hg L1	56	46	34
Hg L3	95	89	48
Hg M	0	0	0
Hg N	0	0	0
Ancestral	100	87	49
Not assigned	0	14	11

In Summary it is shown as seen in Table 3.7 based on visual/manual, automated assignment in conjunction with an algorithm used in a prediction table, users are able successfully assign the screened samples of individuals residing in Limpopo to their ancestral haplogroups.

Showing that the screening/assigning methods and multiplex system to be a quick, inexpensive, time saving and reliable method for haplogroup assignment. Thus, the simplicity of the hierarchical diagnostic multiplex system has shown HRM technology as a choice for screening and assigning individuals to haplogroups. This method can attract new users, as it is a multipurpose analytical tool used to analyse nucleic acids in forensic applications.

3.11 Advantages and disadvantages of assignment methods

HRM: The results have shown that high resolution melting analysis is a reproducible, simple, inexpensive, accurate, easy to use and fast screening method to visually assign individuals to haplogroups. HRM requires no big changes in the laboratory and does not require specific skills which makes laboratory work easy (Vossen *et al.*, 2009). It is affordable with the only expensive element being the acquisition of the instrument, thus significant cost savings are achieved.

It is a high-sensitivity and non-destructive method that does not require post-PCR separation (Vossen *et al*., 2009). High resolution melting analysis is advantageous because of the low contamination risk, high analytical sensitivity and high speed (Wittwer *et al.,* 2009, Vossen *et al*., 2009).

The only disadvantage about HRM is that multiplexing would require more time, but the results have shown that multiplexing can be achieved with two or three SNPs by temperature and colour differences. The success of the multiplex system is however dependent on the time frame and high sample quality and the sensitivity of the assay (Vossen *et al*., 2009).

PCA: It was shown that PCA can be used as a quick and easy method to automatically assign individuals to haplogroups. This clustering method allows users to do an automated assignment, known as a principle component analysis (PCA) to confirm the screened and control samples and identify the haplogroups shown in clusters.

This method is an alternative to Bayesian clustering, which has been applied in several SNP studies of human population structure (Patterson *et al*., 2006). The advantages of PCA include, a quick method that can be performed on desktop computers, facilitates the identification of subset markers to differentiate between populations (Paschou *et al.,* 2007). In some instances, it was reported to outperform Bayesian methods for inferring population structure (Reeves and Richards, 2009).

The only drawback is that PCA is sensitive to missing data and sampling effects, specifically, in populations with continuous distributions (Novembre *et al*., 2008), this would limit inference about the historical and demographic processes (Helyar *et al*., 2011).

ALGORITHM:

A mathematical algorithm was designed that confirms the screened samples and assigned haplogroups. The algorithm is based on the theoretical melting temperatures for each haplogroup and predicts expected haplogroups.

The advantages of using the algorithm is that it can be used as a validation and confirmation test of haplogroup assignment. This method is easy, quick and effective to use.

Thus, the only drawback is that the formula is consistent within a set melting temperature range, if samples fall outside of the Tm range, you are not able to confirm the predicted haplogroups with this equation.

3.11 Why study SNPs and mtDNA in population genetics

SNPs have the characteristics to place historical demography and speciation studies within a molecular framework. This is seen in work done with mtDNA (Brumfiled *et al*., 2003). Studies have shown SNPS to be more favourable than microsatellites (Rosenberg *et al*., 2003) due to it being highly informative having a very high information content for population structure analysis (Paschou *et al*., 2007, Helyar *et al*., 2011).

Although microsatellites display greater allelic diversity per locus, the individual SNPs can segregate strongly among populations (Freamo *et al*., 2011, Karlsson *et al*., 2011). SNPs are the best markers to use for ancestral testing because of the slow mutation rate. Their simplicity, abundance and modelling make them powerful tools in multiple bi-parentally inherited, to infer population histories (Brumfield *et al*., 2003). SNPs are ideal due to its low error rates, the potential for high-throughput genotyping and downstream analysis, being high-informative markers for population analysis (Helyar *et al*., 2011).

In population genetics it important to determine the ancestral origin of these individuals so that we can contribute to the knowledge of Southern African ancestry. We can thus better understand our heritage and travels as human species (Helgason *et al*., 2003)

In forensic genetic applications using SNPs to infer population history and ancestral origin has become significant, this study allows us to study population groups by investigating their genetic markers and the application of the results allows researchers to do downstream analyses.

Chapter 4

Conclusion and future work

4.1 Conclusion

The main conclusion deduced from the results are that the screening/assigning methods and multiplex system can be used as a quick, inexpensive, time saving and reliable method for haplogroup assignment. Thus, the simplicity of the hierarchical diagnostic multiplex system has shown HRM technology as a choice for screening and assigning individuals to haplogroups. This method can attract new users, as it is a multipurpose analytical tool used to analyse nucleic acids in forensic applications.

Majority of the individuals screened with Multiplex 1, were assigned to L0 haplogroup. Based on the genotyping results most of the samples have African (Khoesan) descent. The study found 95 individuals to be derived for the L3 haplogroup which gives a 19% ratio of individuals residing in Limpopo. Only 56 individuals derived for the L1 haplogroup were found, which gives a percentage of 11%. A total of 249 individuals were derived for the L0 haplogroup, making up a 50% of the 500 genotyped samples. Only 100 samples were ancestral for all three SNPs (L1, L0 and L3) making up 20% of samples having Khoesan ancestry.

In Multiplex 2 the assigned individuals shown to have African ancestry. Of the 95 samples derived for the L3 haplogroup, the study found 87 individuals to be ancestral for both M and N, which meant that 91.57% of the screened samples have African descent falling in L3* branch. Further screening would need to be done to establish which African haplogroups (L2, L5, L6 or L4) the 87 individuals can be assigned to as having northern eastern African origin. This would be expected because of the contribution of the hunter-gatherers and admixture that occurred between Khoesan and Southern Bantu population expansions and growth. Bantu individuals are present in Limpopo representing (Sepedi, Sotho, Tswana, and Tsonga speaking individuals).

4.2 Future work

Future work would include additional sub-screening with multiplexes of individuals who could not be assigned to Multiplex 1 and 2. To have a better understanding of the population composition and structure of Southern Bantu populations in Limpopo as each SNP found within the coding region of the mitochondrial genome would be investigated. This would allow researchers to then understand the admixture between hunter-gathers and Bantus who adopted an agricultural lifestyle through gene flow and migration.

NGS is a powerful tool and the future, it would be ideal to sequence the entire mitochondrial genome, to identify mutations that have not been previously identified or well-studied in South Africa. NGS is the future of SNP genotyping, which produces longer reads and large data sets with more information, this would contribute to the knowledge and literature of Southern African ancestry. The only drawback of NGS is that it is expensive and labour intensive and would require a lot of time for optimisation and analysis.

Random sample collection of individuals across regions in South Africa would be interesting, to then use Analysis of Molecular Variance (AMOVA) and compare the genetic diversity between and among Limpopo populations and other Southern African populations. The only weak points for this study was that the diagnostic multiplex system required labour intensive optimization and that 10-20 ng/µl of DNA with a 260/280 purity was needed for sequential genotyping necessary from one multiplex to another.

The strong points of the study was that the hierarchical diagnostic multiplex system has shown to be quick, easy and effective for ancestral testing. The system is reliable and has shown reproducible results that were validated (Ristow, 2017), the results were consistent with other projects done in the laboratory. Lastly this study shows that researchers can manually, automatically and by using an algorithm successfully assign Limpopo individuals to ancestral groups.

References

Alves-Silva et al. (2000). The ancestry of Brazilian mtDNA lineages. *Am J Hum Genet* 3, 775

Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijn, M.H., Coulson, A.R., Drouin J. (1981). Sequence and organization of the human mitochondrial genome. *Nature* 290(5806), 457-465.

Anderson, S., Ellmar, M., Jorgensen, T.H., Palmé, A. (2009). Quantitative genetic effects of bottlenecks: experimental evidence from a wild plant species, *Nigella degenii. J. Hered,* 101, 298-307.

Andrews, R.M., Kubacka, I., Chinnery, P.F., Lightowlers, R.N., Turnbull, D.M., Howell, N. (1999). Reanalysis and revision of Cambridge reference sequence for human mitochondrial DNA. *Nature Genetics* 23(2), 147.

Andrews, C.A. (2010). Natural Selection, Genetic Drift, and Gene flow DO Not Act in Isolation in Natural Populations. *Nature Education Knowledge* 3, 10, 5.

Armitage, S. J., Jasim, S.A., Marks, A.E., Parker, A.G., Usik, V.I., Uerpmann, H. (2011). The southern route 'Out of Africa' evidence for an early expansion of modern humans into Arabia. *Science*, 331, 453-456.

Auton, A., Bryc, K., Boyko, A.R., Lohmueller, K.E., Novembre, J. et al. (2009). Global distribution of genomic diversity underscores rich complex history of continental human populations. *Genome Res*, 19, 795-803.

Bandelt, H.J., Parson, W. (2004). Sources of errors in mitochondrial DNA datasets and evaluation of the mtDNA database "D-Loop-Base". *Rechtmedizin* 14, 251-255.

Bandelt, H.J., Kivisild, T., Parik, J., Villems, R., Bravi, C., Yao, Y.G.(Ed.). (2006). *Lab-specific mutation process. In Human Mitochondrial DNA and the Evolution of Homo sapiens.* Berlin: Springer-Verlag.

Bandelt, H.J., Alves-Silva, J., Guimaráes, P.E., Santos, M.S., Brehem, A., Pereira, L. et al. (2001). Phylogeography of the human mitochondrial haplogroup L3e: a snapshot of African prehistory and Atlantic slave trade. *Ann. Hum. Genet*, 65, 549-563.

Barbieri, C. et al. (2013). Ancient substructure in early mtDNA lineages of Southern Africa. *American Journal for Human Genetics* 92, 285-292.

Barbieri, C. et al. (2014). Unraveling the complex maternal history of Southern African Khoesan populations. *American Journal of Physical Anthropology*, 153, 435-448.

Barker, J.S.F., Frydenberg, J., González, J., Davies, H.I., Ruiz, A., Sørensen, J.G. et al. (2009). Bottlenecks, population differentiation and apparent selection at microsatellite loci in Australian *Drosophila buzzatti*. *Heredity,* 102, 389-401.

Barnard, A. (1988). Kingship, language and production: a conjectural history of Khoesan social structure, Africa. *Journal of International African Institute*. 58, 29-50.

Barnard, A. (1992). *Hunters and herders of Southern Africa.* Cambridge University Press: Cambridge.

Behar, D.M., Villems, R., Soodyall, H., Blue-Smith, J., Pereira, L., Metspalu, E. (2008). The dawn of human matrilineal diversity. *American Journal of Human Genetics* 82(5), 1130-1140.

Beleza, S., Gusmao, L., Amorim, A., Carracedo, A., Salas, A. (2005). The genetic legacy of western Bantu migrations. *Hum. Genet.*, 117, 366-375.

Berniell-Lee, G., Calafell, F., Bosch, E., Heyer, E., Sica, L., Mouguiama-Daouda, P. (2008). Genetic and demographic implications of the Bantu expansion: insights from human paternal lineages. *Molecular Biology Evolution* 26, 1581-1589. [Pubmed: 19369595].

Berniell-Lee, G., Calafell, F., Bosch, E., Heyer, E., Sica, L. et al. (2009). Genetic and demographic implications of the Bantu expansion: insights from human paternal lineages. *Mol Biol Evol* 26, 1581-1589.

Bergstrom, D.E. (2001). Encyclopedia of Genetics. Science Direct

Blench, R. (Ed.). (1993). Recent developments in African language classification and their implications for prehistory. *In the archaeology of Africa: Food, metals and towns.* London.

Blench, R. (Ed.). (2006). *Archaeology, Language, and the African Past*. New York: Rowman & Littlefield Publishers, Inc.

Blows, M.W., Hoffmann, A.A. (2005). A reassessment of genetic limits to evolutionary change. *Ecology*, 86, 1371-1384.

Böeseken, A.J. (1977). Slaves and Free Blacks at the Cape 1658-1700. Cape Town, Tafelberg,

Brandstätter, A., Parsons, T.J., Parson, W. (2003). Rapid screening of mtDNA coding region SNPs for the identification of west European Caucasian haplogroups. *International Journal of Legal Medicine* 117, 291-298.

Brandstätter, A., Peterson, C.T., Irwin, J.A., Mpoke, S., Koech, D.K. et al. (2004). Mitochondrial DNA control region sequences from Nairobi (Kenya): inferring phylogenetic parameters for the establishment of a forensic database. *Int. J. Legal. Med*, 118, 294-306.

Brauer, G. (2008). The origin of Modern Anatomy: By Speciation or Intraspecific Evolution. *Evolutionary Anthropology* 17, 22-37.

Brumfield, R.T., Beerli, P., Nickerson, D.A., Edwards, S.V. (2003). Single nucleotide polymorphisms (SNPs) as markers in phylogeography. *Trends in Ecology and Evolution*, 18, 249-256.

Buckingham, L. (2012). Chromosomal structure and chromosomal mutation. In: Buckingham, L. (Ed.). Molecular fundamentals methods and clinical applications. Philadelphia: F.A. Davis Company

Butzer, K.W. (1981). Rise and Fall of Axum, Ethiopia: A geo-archaeological interpretation. *Am. Antiq.* 46, 471-495.

Burrows, A.M., Ristow, P.G., D'Amato, M.E. (2017). Preservation of DNA from saliva samples in suboptimal conditions. *Forensic Science International: Genetics* 6. 80-81.

Callaway, E. (2017). Oldest *Homo sapiens* fossil claim rewrites our species' history. [Online]. Retrieved July 7, 2019: http://www.nature.com/news/

Campbell, M.C., Tishkoff, S.A. (2008). African genetic diversity: implications for human demographic history, modern human origins, and complex disease mapping. *Annu Rev Genomics Hum Genet* 9, 403-433. [Pubmed: 18593304].

Campbell, M.C., Tishkoff, S.A. (2010). The evolution of Human Genetic and Phenotypic variation in Africa. *Curr Biol*, 20(4), 166-173.

Cann, R.L., Stoneking, M., Wilson, A.C. (1987). Mitochondrial DNA and human evolution. *Nature* 325, 31-36.

Castri, L., Tofanelli, S., Garagnani, P., Bini, C., Fosella, X., Pelotti, S. et al. (2009). mtDNA variability in two Bantu-speaking populations (Shona and Hutu) from Eastern Africa: implications for peopling and migration patterns in sub-Saharan Africa. *Am. J. Phys. Anthropol*, 140, 302-311.

Cavalli-Sforza L.L. (1994). *The history and geography of human genes.* Princeton University Press, Princeton, NJ.

Cerný, V., Fernandes, V., Costa, M.D., Hajek, M., Mulligan, C.J., Pereira, L. (2009). Migration of Chadic speaking pastoralists within Africa based on population structure of Chad Basin and phylogeography of mitochondrial L3f haplogroup. *BMC Evol. Biol*, 9, 63.

Coelho, M., Sequeira, F., Luiselli, D., Beleza, S., Rocha, J. (2009). On the edge of Bantu expansions: mtDNA, Y chromosome lactase persistence genetic variation in southwestern Angola. *BMC Evol. Biol*, 9, 80.

Compton, J.S. (2011). Pleistocene sea-level fluctuations and human evolution on the southern coastal plain of South Africa. *Quat Sci Rev* 30, 506-527.

Conrad D.F., Jakobsson M., Coop G., Wen X., Wall J.D., Rosenberg N.A. et al. (2006). A worldwide survey of haplotype variation and linkage disequilibrium in the human genome. *Nat Genet* 38, 1251-1260. [Pubmed: 17057719]

Cox, M.P., Woerner, A.E., Wall, J.D., Hammer, M.F. (2008). Intergenic DNA sequences from the human X chromosome reveal high rates of global gene flow. *BMC Genet* 9, 76. [Pubmed: 19038041]

Chan, E.K.F., Timmerman, A., Baldi, B.F., Moore, A.E., Lyons, R.J., Lee, S.S. et al. (2019). Human origins in a Southern African palaeo-wetland and first migrations. *Nature*, 575, 185-189.

Chen, Y.S., Torroni, A., Excoffier, L., Santachiara-Benerecetti, A.S., Wallace, D.C. (1995). Analysis of mtDNA variation in African populations reveals the most ancient of all human continent-specific haplogroups. *American Journal for Human Genetics* 57, 133-149.

Chen, Y.S., Olckers, A., Schurr, T.G., Kogelnik, A.M., Huoponen, K., Wallace, D.C. (2000). MtDNA variation in the South African Kung and Khwe-and their genetic relationships to other African populations. *American Journal for Human Genetics* 66, 1362-1383.

Daly, A.K. (2010). Pharmacogenetics and human genetic polymorphisms. *The Biochemical Journal*, 429, 435-449.

Dobrowolski, S.F., Wittwer, C.T. (2011). High resolution melt profiling. In: Rapley R., Harbron S., eds. Molecular analysis and genome discovery. Hoboken, N.J: Wiley, 81-113.

Destro-Bisol, G., Coia, V., Boschi, I., Verginelli, F., Caglia, A., Pascali, V. (2004). The analysis of variation of mtDNA hypervariable region 1 suggests that Eastern and Western Pygmies diverged before the Bantu expansion. *Am Nature,* 163, 212-226.

Destro-Bisol, G., Donati, F., Coia, V., Boschi, I., Verginelli, F., Caglia, A. (2004). Variation of female and male lineages in sub-Saharan populations: the importance of sociocultural factors. *Mol. Biol. Evol*, 21, 1673-1682.

De Filippo, C., Bostoen, K., Stoneking, M., Pakendorf, B. (2012). Bringing together linguistic and genetic evidence to test the Bantu expansion. Proc. R. Soc. B 279, 3256-3263.

Dwight, Z., Palais, R., Wittwer, C.T. (2011). uMELT: prediction of high-resolution melting curves and dynamic melting profiles of PCR products in a rich web application. *Bioinformatics* 27, 1019-20.

Ehret, C., Posnansky, M. (1982). *The archaeological and linguistic reconstruction of African history.* University of California Press.

Ehret, C. (Ed.). (1993). Nilo-Saharans and the Saharo-Sudanese Neolithic. *In the archaeology of Africa: Food, metals and towns.* London, 104-125.

Ehret, C. (1995). *Reconstructing proto-Afroasiatic: Vowels, tone, consonants and vocabulary.* University of California Press.

Ehret, C. (1998). *An African classical age: Eastern and Southern Africa in world history, 1000 B.C. to A.D. 400.* University Press of Virginia.

Ehret, C. (Ed.). (2000). *African languages: an introduction in Language and history*. Cambridge University Press, 272-297.

Ehret, C. (2001). Bantu expansions: Re-envisioning a central problem of early African history. *International Journal of African History Studies*, 34, 5-41.

Ehret, C. (2002). *The civilizations of Africa: A history to 1800.* University Press of Virginia.

Fendt L., Huber G., Röck, A.W., Zimmermann B., Bodner M., Delport R. et al. (2012). Mitochondrial DNA control region data from indigenous Angolan Khoe-San lineages. *Forensic Science International: Genetics*, 6, 662-663.

Feuk, L., Carson, A.R., Scherer, S.W. (2006). Structural variation in the human genome. *Nature reviews. Genetics*, 7(2), 85-97.

Foster, P. (2004). Ice Ages and the mitochondrial DNA chronology of human dispersals. *Biological Science* 359, 255-264.

Frazer, K.A., Murray, S.S., Schork, N.J., Topol, E.J. (2009). Human genetic variation and its contribution to complex traits. *Nat Rev Genet* 10, 241-251. [Pubmed: 19293820]

Fraemo, H., O'Reilly, P., Berg, P., Lien, S., Boulding, E. (2011). Outlier SNPs show more genetic structure between two Bay of Fundy metapopulations of Atlantic salmon than do neutral SNPs. *Molecular Ecology Resources*, 11, 243-256.

Gabriel, W., Lynch, M., Burger, R. (1993). Mullers ratchet and mutational meltdowns. *Evolution*, 47, 1744-1757.

Garrigan, D., Kingan, S.B., Pilkington, M.M., Wilder, J.A., Cox, M.P., Soodyall, H. et al. (2007). Inferring human population sizes, divergence times and rates of gene flow from mitochondrial, X and Y chromosome resequencing data. *Genetics* 177, 2195-2207. [Pubmed: 18073427]

Galtier, N., Nabholz, B., Glémin, S., Hurst, G.D. (2009). Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Molecular ecology* 18, 4541-4550.

Gillespie, R.G., Roderick, G.K. (2002). Arthropods on islands: colonization, speciation and conservation. *Annu. Rev. Entomol*, 47, 595-632.

Goldstein, D.B., Chiki, L. (2002). Human migrations and population structure: What we know and why it matters. *Annual Review of Genomics and Human Genetics*, 3, 129-152.

Gomez, F., Hirbo, J., Tishkoff, S.A. (2014). Genetic variation and adaption in Africa: Implications for Human Evolution and Disease. *Cold Spring Harb Perspectives in Biology*, 6, 1-21.

Gonder, M.K., Mortensen, H.M., Reed, F.A., de Sousa, A., Tishkoff, S.A. (2007). Whole-mtDNA genome sequence analysis of ancient African lineages. *Mol Biol Evol* 24, 757-768.

Greenberg, J.H. (1963). The languages of Africa. *International Journey of Am. Linquist.* 29, 1.

Greenberg, J.H. (1972). Linguistic evidence concerning Bantu origins. *Journal of African History* 13, 189-216.

Gugliotta, G. (2008, July). The Great Human Migration: Why humans left their African homeland 80,000 years ago to colonize the world. *Smithsonian magazine*.

Güldemann, T., Vossen, R. (2000). Khoesan: In Heine, B and Nurse, D (Eds.). *African languages: An Introduction*. (pp.99-122). Cambridge University Press. Cambridge

Güldemann, T., Stoneking, M. (2008). A historical appraisal of clicks: a linguistic and genetic population perspective. *Ann. Rev. Anthropol* 37, 93-109.

Gundry, C.N., Vandersteen, J.G., Reed, G.H., Pryor, R.J., Chen, J., Wittwer, C.T. (2003). Amplicon melting analysis with labelled primers: a close-tube method for differentiating homozygotes and heterozygotes. *Clin Chem* 49, 396-406.

Guthery, S.L., Salisbury, B.A., Pungliya, M.S., Stephens, J.C., Bamshad, M. (2007). The structure of common genetic variation in United States populations. *Am J Hum Genet* 81, 1221-1231. [Pubmed: 17999361]

Guthrie, M. (1948). *The classification of the Bantu languages.* London, UK: Oxford University Press for the International African Institute.

Haga, S.B., Venter, J.C. (2003). FDA races in wrong direction. *Science*, 301: 466

Hall, T.A. (1999). *BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT*. (pp. 95-98). Oxford University Press. USA.

Heerwaarden, B., Kellermann, V., Schiffer, M., Blacket, M., Sgrò, C.M., Hoffmann, A.A. (2009). Testing evolutionary hypotheses about species borders: patterns of genetic variation towards the southern borders of two rainforest *Drosophila* and a related habitat generalist. *Proc. Biol. Sci.* 276, 1517-1526.

Helyar, S.J., Hemmer-Hansen, J., Bekkevold, D., Taylor, M.I., Ogden, R. et al. (2011). Applications of SNPs for population genetics nonmodel organisms: new opportunities and challenges. *Molecular ecology Resources*, 11, 123-136.

Helgason et al. (2003). A Reassessment of Genetic Diversity in Icelanders: Strong Evidence from Multiple Loci for Relative Homogeneity Caused by Genetic Drift. *Annals of Human Genetics*, 67, 281-297.

Henn, B.M., Gignoux, C.R., Jobin, M., Granka, J.M., Macpherson, J.M., Kidd, J.M. et al. (2011). *Hunter-gatherer genomic diversity suggests a Southern African origin for modern humans.* [Online]. Available: http://www.pnas.org/cgi/doi/10.1073/pnas.1017511108

Henshilwood, C.S. (1996). A revised chronology for pastoralism in southernmost Africa: new evidence of sheep at c. 2000 bp from Blombos Cave, South Africa. *Antiquity*

Hermann, M.G., Durtschi, J.D., Wittwer, C.T. (2007). Expanded instrument comparison of amplicon DNA melting analysis for mutation scanning and genotyping. *Clin Chem* 53, 1544-8.

High-Resolution Melt Analysis for Mutation Scanning Prior to Sequencing. [Appliedbiosystems] [Online]. Retrieved September 10, 2018: http://www.lifetechnologies.com.html.

Hinds, D.A., Stuve, L.L., Nilsen, G.B., Halperin, E., Eskin, E., Ballinger, D.G., Frazer, K.A., Cox, D.R. (2005). Whole genome patterns of common DNA variation in three human populations. *Science* 307. 1072-1079. [Pubmed: 15718463]

Hoffmann, A.A., Hallas, R.J., Dean, J.A., Schiffer, M. (2003). Low potential for climatic stress adaption in a rainforest *Drosophila* species. *Science,* 301, 100-102.

Holden, C.J. (2002). Bantu language trees reflect the spread of farming across sub-Saharan Africa: a maximum-parsimony analysis. *Proc. R. Soc. Lond. B.* 269, 793-799.

Howell, N., Elson, J.L., Howell, C., Turnbull, D.M. (2007). Relative rates of evolution in the coding and control regions of African mtDNAs. *Mol Biol Evol* 24(10), 2213-2221.

Hublin, J. *et al*. (2017). New fossils from Jebel Irhoud, Moroco and the pan-African origin of Homo Sapiens. *Nature* 546, 289-292.

Hughes, A.L., Welch, R., Puri, V., Matthews, C., Haque, K., Chanock, S.J., Yeager, M. (2008). Genome-wide SNP typing reveals signatures of population history. *Genomics* 92, 1-8. [Pubmed: 18485661]

Huffman, T.N. (1982). Archaeology and the ethnohistory of the African Iron Age. *Annu. Rev. Anthropol*. 11, 133-150.

Huffman, T.N. (1992). Southern Africa to the south of Zambesi. In M.E. Fasi, and I. Hrbek (Eds.). *General history of Africa Volume 3: Africa from the 7th to the 11th Century* (pp. 664-680). California: University of California Press.

Ingman, M., Kaessmann, H., Pääbo, S., Gyllensten, U. (2000). Mitochondrial genome variation and the origin of modern humans. *Nature* 408, 708-713.

Jakobsson, M., Scholz, S.W., Scheet, P., Gibbs, J.R., VanLiere, J.M., Fung, H.C., Szpiech Z.A., Degnan J.H., Wang K., Guerreiro R. (2008). Genotype, haplotype and copy-number variation in worldwide human populations. *Nature* 451, 998-1003. [Pubmed: 18288195]

Jobling, M.A., Tyler-Smith, C. (2003). The human Y chromosome: an evolutionary marker comes of age. *Nat Rev Genet* 4(8), 598-612.

Johnston, H.H. (1913). A survey of the ethnography of Africa: and the former racial and tribal migrations of that continent. *J. R. Anthropol. Inst.* XLIII, 391-392.

Jolly, P. (1994). Strangers to brother: Interaction between southeastern San and southern Nguni/Sotho communities. M.A thesis, University of Cape Town.

Kanderian, S., Jiang, L., Knight, I. (2015). Automated classification and cluster visualization of genotypes derived from high resolution melt curves. *PLoS one*, 10(11): e0143295.

Kapa Biosystems (2016) 'Technical Data Sheet KAPA Hyper Prep Kit', (January), pp. 1–16. Available at: https://www.sigmaaldrich.com/content/dam/sigma-aldrich/docs/Roche/Datasheet/1/hrmfastkbdat.pdf.

Karlsson, S., Moen, T., Lien, S., Glover, K., Hindar, K. (2011). Generic genetic differences between farmed and wild Atlantic salmon identified from a 7K SNP-chip. *Molecular Ecology Resources*, 11, 236-242.

Kawecki, T.J. (2008). Adaption to marginal habitats. *Annu. Rev. Ecol. Evol. Syst*, 39, 321-342.

Kayser, M. (2007). Uni-parental markers in human identity testing including forensic DNA analysis. *Biotechniques* 43(6), S16-S21.

Kayser, M. (2017). Forensic use of Y-chromosome DNA: a general overview. Human genetics 136(5), 621-635.

Kellermann, V.M., Van Heerwaarden, B., Hoffmann, A.A., Sgrò, C.M. (2006). Very low additive genetic variance and evolutionary potential in multiple populations of two rainforest *Drosophila* species. *Evolution*, 60, 1104-1108.

Kim, H.L., Ratan, A., Perry, G.H., Montenegro, A., Miller W., *et al*. (2014). Khoesan hunter-gatherers have been the largest population throughout most of modern human demographic history, *Nature Communications*, 5, 5692.

Kinahan, J. (2011). From the beginning: the archaeological evidence. In a History of Namibia: From the Beginning to 1990, M. Wallace (pp. 15-43). London: Hurst and Company

Kivisild, T., Metspalu, M., Bandelt, H.J., Richards, M., Villems, R. (2006). The world mtDNA phylogeny. In H.J. Bandelt, V. Macaulay & M. Richards (Eds.). *Human mitochondrial DNA and the evolution of Homo sapiens*. (pp. 149-179). Berlin (Germany): Springer-Verlag.

Kivisild, T., Shen, P., Wall D.P., *et al*. (2006). The role of selection in the evolution of human mitochondrial genomes. *Genetics*, 172, 373-387.

Kloss-Brandstätter, A., Pacher, D., Schönherr, S., Weissensteiner, H., Binna, R., Specht, G., Kronenberg, F. (2011). Haplogrep: a fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups. *Hum. Mutat.* 32, 25-32.

Knight, A., Underhill, P.A., Mortensen, H.M., Zhivotovsky, L.A., Lin, A.A., Henn, B.M., Louis, D., Ruhlen M., Mountain, J.L. (2003). African Y chromosome and mtDNA divergence provides insight into the history of click languages. *Curr. Biol.* 13,464-473.

Kong et al. (2003). Mitochondrial DNA sequence polymorphisms of five ethnic populations from northern China. *Human Genetics*, 113, 391-405.

König, C. (2008). Khoesan languages. *Lang. Linguist. Compass* 2, 996-1012.

Lander, F., Russell, T. (2018). The archaeological evidence for the appearance of pastoralism and farming in Southern Africa. *PLoS One*, 13, e0198941.

Lemire, B (2005). Mitochondrial genetics. *Wormbook*, ed. The *C.elegans* Research Community. University of Alberta Edmonton, Canada.

Li, J.Z., *et al*. (2008). Worldwide human relationships inferred from genome-wide patterns of variation. *Science* 319, 1100-1104.

Li, M., Liu, Z., Yue, S., Zhou, L., Zhang, Q., Cheng, S., Li, R.W., Smith, P.N., Lu, S. (2009). The status of KRAS mutations in patients with non-small cell lung cancers from mainland China. *Oncol Rep* 22, 1013-20.

Li, M., Zhang, Q., Liu, L., Liu, Z., Zhou, L.Z., Yue, S., Xiong, H., Feng, L., Lu, S. (2011). The different clinical significance of EGFR mutations in exon 19 and 21 in non-small cell lung cancer patients of China. *Neoplasma* 58, 74-81.

Li, H., Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature*, 475, 493-496.

Li, M., Zhou, L., Palais, R.A., Wittwer, C.T. (2014). Genotyping accuracy of high-resolution DNA melting instruments. *Clinical Chemistry* 60, 864-72.

Liew, M., Pryor, R., Palais, R., Meadows, C., Erali, M., Lyon, E., Witter C. (2004). Genotyping of singe-nucleotide polymorphisms by high-resolution melting of small amplicons. *Clin Chem* 50, 1156-64.

Liu, H., Prugnolle, F., Manica, A., Balloux, F. (2006). An explicit genetic model of worldwide human-settlement history. *Am J Hum Genet* 79, 230-237. [Pubmed: 16826514]

Loewe, L., William, G.H. (2010). The population genetics of mutations: good, bad and indifferent. *Philosophical transactions of the royal society*, 365, 1153-1167.

Lombard, M., Schlebusch, C., Soodyall, H. (2013). Bridging disciplines to better elucidate the evolution of early *Homo sapiens* in Southern Africa. *S Afr J Sci* 109, 27-34.

Long, J.C., Li, J., Healy, M.E. (2009). Human DNA sequences: more variation and less race. *Am J Phys Anthropol* 139, 23-34 [Pubmed: 19226648]

Macaulay, V., Hill, C., Achilli, A., Rengo, C., Clarke, D., Meehan, W., Blackburn, J., Semino, O., Scozzari, R., Cruciani, F. (2005). Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Science* 308(5724), 1034-1036.

Maca-Meyer, N., Gonzalez, A.M., Larruga, J.M., Flores, C., Cabrera, V.M. (2001). Major genomic mitochondrial lineages delineate early human expansions. *BMC Genetics* 2, 13.

Maji, S., Krithika, S., Vasulu, T.S. (2009). Phylogeographic distribution of mitochondrial DNA Macrohaplogroup M in India. Indian Academy of Sciences: Review article, *Journal Genet* 88, 1-21.

Manica, A., Amos, W., Balloux, F., Hanihara, T. (2007). The effect of ancient population bottlenecks on human phenotypic variation. *Nature* 448,346-348. [Pubmed: 17637668]1

Marean et al. (2010). The stratigraphy of the Middle Stone Age sediments at Pinnacle Point Cave 13B (Mossel Bay, Western Cape Province, South Africa). *J Hum Evol*, 59, 3-4.

Marks, S.J. et al. (2014). Static and moving frontiers: the genetic landscape of Southern African Bantu-speaking populations. *Mol. Biol. Evol*. 32, 29-43.

Matute, D.R. (2013). The role of founder effects on the evolution of reproductive isolation. *Journal of Evolutionary Biology*, 26, 2299-2311.

McKenna, A. (2018). *Pedi* [Online]. Retrieved 8 November 2019: https://www.britannica.com/topic/pedi

McKenna, A. (2019). *Venda* [Online]. Retrieved 8 November 2019: https://www.britannica.com/place/Venda-former-republic-Africa

Medrano, J., Aasen, E., Sharrow, L. (1990). DNA Extraction from Nucleated Red Blood Cells. *Biotechnique*s, 8, 43.

Mellars, P. (2006). Why did modern human populations disperse from Africa ca 60,000 years ago? A new model. *Proc Natl Acad Sci* 103, 9381-9386. [Pubmed: 16772383]

Mellars, P. (2006). Going east: New genetic and archaeological perspectives on the modern human colonization of Eurasia. *Science* 313,796-800.

Metspalu, M. (2004). Most of the extant mtDNA boundaries in south and southwest Asia were likely shaped during the initial settlement of Eurasia by anatomically modern humans. *BMC Genet* 5, 26-50.

Mishmar, D., Ruiz-Pesimi, E., Golik, P., Macaulay, V., Clark, A.G., Hosseini, S., Brandon, M., Easley, K., Chen, E., Brown, M.D. (2003). Natural selection shaped regional mtDNA variation in humans. *Proc Natl Acad Sci* 100(1), 171-176.

Mitchell, P.J. (2002). *The archaeology of Southern Africa*. Cambridge University Press, Cambridge.

Mitchell, P., Plug, I., Bailey, G. (2008). Bringing the Kalahari debate to the mountains: domestic livestock and other evidence of late first millennium AD hunter-gatherer/farmer interaction from highland Lesotho. Before farming [Online version], 1-22.

Mitchell, P. (2009). Hunter-gatherers and farmers: some implications of 2000 years of interaction in the Maloti-Drakensberg region of Southern Africa. In: Ikeya, K, Ogawa, H, Mitchell, P.J. (ed.*). Interactions between hunter-gatherers and farmers: from prehistory to present*. (pp.15-46). National Museum of Ethnology. Japan.

Mitchell, P. (2010). Genetics and Southern African prehistory: an archaeological view. *Journal of Anthropology Sciences,* 88, 5-8.

Montgomery, J.L., Sanford, L.N., Wittwer, C.T. (2010). High-resolution DNA melting analysis in clinical research and diagnostics. *Expert Rev Mol Diagn* 30, 219-40.

Morin, P.A., Luikart, G., Wayne, R.K. (2004). SNPs in ecology, evolution and conservation. *Trends in Ecology and Evolution*, 19, 208-216.

Morris, A.G., Heinze, A., Chan, E.K.F., Smith, A.B., Hayes, V.M. (2014). First Ancient Mitochondrial Human Genome from a prepastoralist Southern Africa.

Mountain, A. (2003). *The first people of the Cape*. South Africa. David Philip Claremont.

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., Elrich, H. (1987). Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb Symp Quant Biol*, 1: 263-73.

Neiman, M., Taylor, D.R. (2009). The causes of mutation accumulation in mitochondrial genomes. *Proceedings of the Royal Society B*: *Biological Sciences*, 1-13.

Nielsen, R., Akey, J.M., Jakobsson, M., Pritchard, J.K., Tishkoff, S. *et al*. (2017). Tracing the people of the world through genomics. *Nature* 546, 289-292.

Novembre, J., Johnson, T., Bryc, K., Kutalik, Z., Boyko, A.R. et al. (2008). Genes mirror geography within Europe. *Nature*, 456, 98-101.

Nurse, D. (1997). The contributions of linguistic to the study of history in Africa. *Journal of African History* 38, 359-391.

O'Dushlaine, C.T. et al. (2005). Tandem repeat copy number variation in protein-coding regions of human genes. *Genome Biology*, 6, 6-8.

Oppenheimer, S. (2012). Out-of-Africa, the peopling of continents and islands: tracing uniparental gene trees across the map. *Phil. Trans. R. Soc. B* 367, 770-784.

Oppenheimer, S. (2003). Out of Eden: the peopling of the world. London, UK: Constable.

Oppenheimer, S. (2002). The first exodus: journey out of Africa. *Georg. Mag.* 74, 32-35.

Pakendorf, B., Stoneking, M. (2005). Mitochondrial DNA and human evolution. *Annu Rev Genomics Hum Genet* 6, 165-183.

Pakendorf, B., Bostoen, K., and de Filippo, C. (2011). Molecular prospectives on the Bantu expansion: a synthesis. *Language Dynamics and change* 1, 50-88.

Palais, R., Wittwer, C.T. (2009). Mathematical algorithms for high-resolution DNA-melting analysis. *Methods Enzymol* 454, 323-43.

Parson, W., Bandelt, H.J. (2007). Extended guidelines for mtDNA typing of population data in Forensic Science, *Forensic Science International*, 1, 13-19.

Paschou, P., Ziv, E., Burchard, E.G. *et al*. (2007). PCA-correlated SNPs for structure identification in worldwide human populations. *PLoS Genetics*, 3, 1672-1686.

Patin, E. et al. (2014). The impact of agricultural emergence on the genetic history of African rainforest hunter-gatherers and agriculturalists. *Nat Commun* 5, 3163.

Patterson, N., Price, A., Reich, D. (2006). Population structure and eigenanalysis. *PLoS Genetics*, 2, e190.

Pauls, E.P. (2017). *Tsonga* [Online]. Retrieved 8 November 2019: http://www.britannica.com/topic/Tsonga

Pereira, L., Macaulay, V., Torroni, A., Scozzari, R., Prata, M.J., Amorim, A. (2001). Prehistoric and historic traces in the mtDNA of Mozambique: insights into the Bantu expansions and the slave trade. *Ann Human Genetics*, 65, 439-458.

Pereira, L., Richards, M., Goios, A., Alonso, A., Albarrán, C., Garcia, O. et al. (2005). Evaluating the forensic informativeness of mtDNA haplogroup H sub-typing on a Eurasian scale. *Forensic Science International* 159, 43-50.

Petersen, D.C. *et al*., (2013). Complex Patterns of Genomic Admixture within Southern Africa. *Plos Genetics*, 9, 1-18.

Phillipson, D. (1993). *African Archaeology*, Cambridge University Press, Cambridge.

Philipson, D.W. (2005). *African Archaeology*. Cambridge: Cambridge University Press.

Phillipson, L. (2009). Lithic artefacts as a source of cultural, social and economic information: The evidence from Askum, Ethiopia. *African Archaeology*, 26, 45-58.

Pickrell, J.K. et al. (2012). The genetic prehistory of Southern Africa. *Nature communications*, 3, 1143.

Pickrell, J.K. et al. (2014). Ancient west Eurasian ancestry in southern and eastern Africa. *Proc Natl Acad Sci*, 111, 2632-2637.

Pleurdeau, D. et al. (2012). Of sheep and men: earliest direct evidence of caprine domestication in Southern Africa at Leopard Cave (Erongo, Namibia). *PLoS One*, 7, e40340.

Prugnolle, R., Manica, A., Balloux, R. (2005). Geography predicts neutral genetic diversity of human populations. *Curr Biol*, 15, 159-160.

QIAGEN (2009) 'For use with Rotor-Gene cyclers Sample & Assay Technologies QIAGEN Sample and Assay Technologies', (September). Available at: https://www.qiagen.com/us/resources/download.aspx?id=af33be05-14c6-4ac3...%0A.

Quintana-Murci, L., Semino, O., Bandelt, H.J., Passarino, G., McElreavey, K., Santachiara-Benerecetti, A.S. (1999). Genetic evidence of an early exit of Homo sapiens from Africa through eastern Africa. *Nature Genetics* 23, 437-441. [Pubmed: 10581031]

Quintana-Murci, L., Quach, H., Harmant, C., Luca F., Massonnet, B., Patin, E. et al. (2008). Maternal traces of deep common ancestry and asymmetric gene flow between Pygmy hunter gatherers and Bantu-speaking farmers. *Proc. Natl. Acad. Sci.* 105, 1596-1601.

Quintana-Murci, L., Harmant, C., Quach, H., Balanovsky, O., Zaporozhchenko, V., Bormans, C. et al. (2010). Strong maternal Khoesan contribution to the South African coloured population: a case of gender-biased admixture, *American Journal of Human Genetics* 86, 611-620.

Quintáns, B., Álvarez-Iglesias, V., Salas, A., Phillips C., Lareu M.V., Carracedo Á. (2004). Typing of mitochondrial DNA coding region SNPs of forensic and anthropological interest using SNaPshot minisequencing. *Forensic Science International* 140, 251-257.

Ramachandran, S., Deshpande, O., Roseman, C.C., Rosenberg, N.A., Feldman, M.W., Cavalli-Sforza, L.L. (2005). Support from the relationship of genetic and

geographic distance in human populations for a serial founder effect originating in Africa. *Proc Natl Acad Sci* 102, 15942-15947. [Pubmed: 16243969]

Ray, N., Currat, M., Berthier, P., Excoffier, L. (2005). Recovering the geographic origin of early modern humans by realistic and spatially explicit simulations. *Genome Res* 15, 1161-1167.

Reed, F.A., Tishkoff, S.A. African human diversity, origins and migrations. (2006). *Curr Opin Genet Dev* 16, 597-605. [Pubmed: 17056248]

Reid, A., Sadr, K., Hanson-James, N. (1998). Herding traditions. In Ditswa MMung: P. Lane, A. Reid, and A. Segobye (Eds.), The Archaeology of Botswana (pp. 81-100). Gaborone: Pula Press and the Botswana Society.

Reja, V. *et al.* (2010) 'ScreenClust: Advanced statistical software for supervised and unsupervised high resolution melting (HRM) analyses, *Methods*. Elsevier Inc., 50(4), pp. S10–S14. doi: 10.1016/j.ymeth.2010.02.006

Relethford, J.H., Harding R.M. (2001). *Population Genetics of Modern Human Evolution* [Online]. Retrieved May 20, 2019: http://www.els.net.html.

Relethford, J.H. (2003). Reflections of our past: *How human history is revealed in our genes.* Boulder, Colorado: Westview Press.

Restriction fragment length polymorphisms. [i.j.] [online]. Retrieved November 11, 2019: http://www.informatics.jax.org/silver/chapter/8-2.shtml

Ririe, K.M., Rasmussen, R.P., Wittwer C.T. (1997). Product differentiation by analysis of DNA melting curves during the polymerase chain reaction. *Analytical Biochemistry* 245, 154-60.

Ristow, P (2017) Comprehensive analyses of the genetic variation between forensic markers in South Africa: Mitogenomes, Autosomal STRs, and Retrotransposons. Published Doctor of Philosophiae thesis. Western Cape: University of the Western Cape.

Rito, T., Vieira, D., Silva, M., Conde-Sousa, E., Pereira, L., Mellars, P. (2019). A dispersal of *Homo Sapiens* from southern to eastern Africa immediately preceded the out-of-Africa migration. *Scientific Reports*, 9, 1-10.

Rocha, J. (2010). Bantu-Khoesan interactions at the edge of the Bantu expansion: insights from southern Angola. *Journal of Anthropology Sciences,* 88, 5-8.

Rodriquez-Murillo, L., Salem, R.M. (2013). Insertion/Deletion Polymorphism. *Encyclopedia of Behavioural Medicine, 1*-3.

Robbins, L.H. et al. (2005). The advent of herding in South Africa: early AMS dates on domestic livestock from the Kalahari Desert. *Curr Anthropol,* 46, 671-677.

Rosa, A., Brehm, A. (2011). African human mtDNA phylogeography at-a-glance. *Journal of Anthropological Sciences*, 89, 1-34.

Rosa, A., Brehm, A., Kivisild, T., Metspalu, E., Villems, R. (2004). mtDNA profile of West Africa Guineans: towards a better understanding of the Senegambia region. *Ann. Hum. Genet*, 68, 340-352.

Rosenberg, N.A., Li, L.M., Ward, R., Pritchard, J.K. (2003). In formativeness of Genetic markers for Inference of Ancestry. Am J Hum Genet, 73(6): 1402–1422.

Rusk, N. (2009). Cheap third-generation sequencing. *Nature Methods*, 6, 244-245.

Salas, A., Richards, M., De la Fe, T., Lareu, M.V., Sobrino, B., Sanchez-Diz, P., Macaulay V., Carracedo A. (2002). The making of the African mtDNA landscape. *American Journal of Human Genetics*, 71, 1082-1111.

Salas, A., Richards, M., Lareu, M.V., Scozzari, R., Coppa, A., Torroni, A. et al. (2004). The African diaspora: mitochondrial DNA and the Atlantic slave trade. *American Journal of Human Genetics*, 74, 454-465.

Scheffler, I.E. (1999). *Mitochondria*. New York: Wiley-Liss.

Schlebusch, C.M., Lombard, M., Soodyall, H. (2013). MtDNA control region variation affirms diversity and deep-structure in populations from Southern Africa. *BMC Evol Biol*, 13, 56.

Schlebusch, C.M. (2012). Genomic variation in seven Khoe-San groups reveals adaption and complex African history. *Science*, 338, 374-379.

Schlebusch, C.M., Malmström, H., Günther, T., Sjödin, P., Coutinho, A. *et al*. (2017). Southern African ancient genomes estimate modern human divergence to 350,000 to 260,000 years ago. *Science*, 358, 652-655.

Schuster, S.C., Miller, W., Ratan, A., Tomsho, L.P., Giardine, B. *et al.* (2010). Complete Khoesan and Bantu genomes from Southern Africa. *Nature*, 463, 943-947.

Sealy, J.C., Yates, R. (1994). The chronology of the introduction of pastoralism ton the Cape, South Africa. *Antiquity*, 68, 58-67.

Sharp, J., Douglas, S. (1998). Prisoners of their reputation? The veterans of the Bushmen battalions in South Africa. In P.Skotnes (Ed.), *Miscat: Negotiating the Presence of the Bushmen* (pp.323-330). Cape Town: University of Cape Town.

Shi, W., Ayub, A., Vermeulen, M., Shao, R.G., Zuniga, S. *et al.* (2010). A worldwide survey of human male demographic history based on Y-SNP and Y-STR data from the HGDP-CEPH populations. *Mol. Biol. Evol.* 27, 385-393.

Stanyon, R., Sazzini, M., Luiselli, D. (2009). Timing the first human migration into eastern Asia. *J Bio*, 8-18. [Pubmed: 19232059]

Stats SA (2019). *Mid-year population estimates*. Statistical release. P0302. Pretoria

Stevanovitch, A., Gilles, A., Bouzaid, E., Kefi, R., Paris, F., Gayraud, R.P. et al. (2004). Mitocondrial DNA sequence diversity in a sedentary population from Egypt. *Ann Human Genetics*, 68, 23-39.

Stone, A., Bornhorst, J. (Ed.). (2012). *An Introduction to personalized medicine*. School of Medicine. Little Rock: AR.

Soares, P.A., Ermini, L., Thomson, N., Mormina, M., Rito, T., Röhl, A. et al. (2009). Correcting for purifying selection: an improved human mitochondrial molecular clock, *Am. J. Hum. Genet*, 84, 740-759.

Soares, P.A. et al. (2016). Resolving the ancestry of Austronesian-speaking populations. *Human Genetics*, 135, 309-326.

Soodyall, H., Vigilant, L., Hill, A.V., Stoneking, M., Jenkins, T. (1996). mtDNA control-region sequence variation suggests multiple independent origins of an "Asian-specific" 9-bp deletion in sub-Saharan Africans. *Am. J. Hum*. Genet, 58, 595-608.

Southern, E.M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *Journal of Molecular Biology*, 98, 503-508.

Smith, A.B. (1992). Origins and spread of pastoralism in Africa. *Annu Rev Anthropol*, 21, 125-141.

Smith, A.B. (2006). Excavations at Kasteelberg and the origins of the Khoekhoen in the Western Cape, South Africa. *Oxford BAR International Series*, 1537.

Smith, M., Pascal, C., Grauvogel, Z., Habicht, C., Seeb, L. (2011). Multiplex preamplification PCR and microsatellites validation allows accurate single nucleotide polymorphism (SNP) genotyping of historical fish scales. *Molecular Ecology Resources*, 11, 257-266.

Snapshot Multiplexing System for SNP Genotyping. [Online]. Retrieved November 15, 2019: http://www.cd-genomics.com/SNaPshot.html

Sutton, J. (1973). The settlement of East Africa. In Zamani and B.A. Ogot (Eds.), (pp. 70-97). Longmans: Nairobi.

Teama, S. (2018). DNA Polymorphisms: DNA-Based Molecular Markers and Their Application in Medicine. In Teama, S (eds.). Egypt: Ain Shams University, Molecular Genetics Unit.

Torroni, A., Achilli, A., Macaulay, V., Richards, M., Bandelt, H.J. (2006). Harvesting the fruit of the human mtDNA tree. *Trends Genet* 22(6), 339-345.

Torroni, A., Sukernik, R.I., Schuur, T.G., Starikorskaya, Y.B., Cabell, M.F., Crawford, M.H., Comuzzie, A.G., Wallace, D.C. (1993). MtDNA variation of aboriginal Siberians reveals distinct genetic affinities with Native Americans. *Am J. Hum. Genet*, 53, 591-608.

Tishkoff, S.A., Dietzch, E., Speed, W., Pakstis, A.J., Kidd, J.R., Cheung, K. et al. (1996). Global patterns of linkage disequilibrium at the *CD4* locus and modern human origins. *Science* 271, 1380-1387. [Pubmed: 8596909]

Tishkoff, S.A., Gonder, M.K., Henn, B.M., Mortensen, H., Knight, A., Gignoux, C. et al. (2007). History of click-speaking populations of Africa inferred from mtDNA and Y chromosome genetic variation. *Mol. Biol. Evol.* 24, 2180-2195.

Tishkoff, S.A., Reed, F.A., Friedlaender, F., Ehret, C., Ranciaro, A., Froment, A. et al. (2009). The genetic structure and history of Africans and African American, *Science* 324, 1035-1044. [Pubmed: 19407144]

Umetsu, K., Tanaka, M., Yuasa, I., Saitou, N., Takeyasu, I., Fuku, N. (2001). Multiplex amplified product-length polymorphism analysis for rapid detection of human mitochondrial DNA variations. *Electrophoresis* 22, 3533-3538.

Umetsu, K., Yuasa, I. (2005). Recent progress in mitochondrial DNA analysis, *Legal Medicine*, 259-262.

Umetsu, K., Tanaka, M., Yuasa, I., Saitou, N., Adachi, N., Miyoshi, A. et al. (year) Multiplex amplified product-length polymorphism analysis of 36 mitochondrial single nucleotide polymorphisms for haplogrouping of East Asian populations, *Electrophoresis*, 26, 91-98.

Underhill, P.A., Kivisild, T. (2007). Use of Y chromosome and mitochondrial DNA population structure in tracing human migrations. *Annu Rev Genet* 41,539-564.

Underhill, P., Passarino, G., Lin, A., Shen, P., Lahr, M., Foley, R. (2001). The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. *Ann. Hum. Genet.* 65, 43-62.

Vallone, P.M., Just, R.S., Coble, M.D., Butler, J.M., Parsons, T.J. (2004). A multiplex allele-specific primer extension assay for forensically informative SNPs distributed throughout the mitochondrial genome, *International Journal of Legal Medicine* 118, 114-157.

Vandersteen, J.G., Bayrak-Toydemir, P., Palais, R.A., Wittwer, C.T. (2007). Identifying common genetic variants by high-resolution melting. *Clinical Chemistry* 53, 1191-8.

Vansina, J. (1995). New linguistic evidence and the Bantu expansion. *Journal of African History* 36, 173-195.

Van Beek, G.W. (1967). Monuments of Axum in the light of South Arabian archaeology. *J Am Orient Sac* 87, 113-122.

Van Oven, M., Kayser, M. (2008). Updated Comprehensive Phylogenetic Tree of Global Human Mitochondrial DNA Variation. *Mutation in brief*, 29, E386-E394.

Van Oven, M., Kayser, M. (2009). Updated Comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum. Mut.* 30, E386-E394.

Van Oven, M. (2015). PhyloTree Build 17: Growing the human mitochondrial DNA tree. *Forensic Science International: Genetics Supplement Series* 5, e392-e394.

Van der Ross, R.E. (2005). Up from Slavery: Slaves at the Cape, their origins, treatments and contribution. Ampersand and the University of the Western Cape: Cape Town.

Vigilant, L., Pennington, R., Harpending, H., Kocher, T.D., Wilson, A.C. (1989). Mitochondrial DNA sequences in single hairs from a Southern African population. *Proc Natl. Acad. Sci.* 86, 9350-9354.

Vigilant, L., Stoneking, M., Harpending, H., Hawkes, K., Wilson, A.C. (1991). African populations and the evolution of human mitochondrial DNA. *Science* 253, 1503-1507.

Vinnicombe, P. (2009). Basotho oral knowledge: the last Bushmen inhabitants of the Mashai district, Lesotho. In P.J. Mitchell & B.W. Smith (Eds.), *The eland's people: New perspectives on the rock art of the Maloti-Drakensberg Bushmen. Essays in memory of Patricia Vinnicombe* (pp.165-191). Johannesburg: Witwatersrand University Press.

Vogal, J.O. (Ed.). (1994). *In Encyclopedia of precolonial Africa*. Walnut Creek, CA: Alta-Mira Press.

Vossen, R.H., Aten, E., Roos, A., den Dunnen, J.T. (2009). High-resolution melting analysis (HRMA): more than just sequence variant screening. *Hum Mutat* 30, 860.

Wall, J.D., Cox, M.P., Mendez, F.L., Woerner, A., Severson, T., Hammer, M.F. (2008). A novel DNA sequence database for analysing human demographic history. *Genome Res* 18, 1354-1361. [Pubmed: 18493019]

Watson, E., Forster, P., Richards, M., Bandelt, H.J. (1997). Mitochondrial footprints of human expansions in Africa. *The American Journal of Human Genetics* 61,691-704.

What is DNA? [NIH] [Online]. Retrieved November 15, 2019: https://ghr.nlm.nih.gov/primer/basics/dna.

White, T., Asfaw, B., DeGusta, D., Gilbert, H., Richards, G., Suwa, G. et al. (2003). Pleistocene Homo sapiens from Middle Awash, *Nature*, 423, 742–747.

Wittwer, C.T., Reed, G.H., Gundry, C.N., Vandersteen, J.G., Pryor, R.J. (2003). High-resolution genotyping by amplicon melting analysis using LC Green. *Clinical Chemistry* 49, 853-60.

Wittwer, C.T. (2009). High-resolution DNA melting analysis: advancements and limitations. *Human Mutations* 30, 857-9.

Wittwer, C.T., Rasmussen, R.P., Ririe, K.M. (2010). Rapid PCR and melting analysis. In S.A. Bustin (Ed.). *The PCR revolution: basic technologies and applications.* (pp. 48-69). New York: Cambridge University Press.

Yakovchuk, P., Protozanova, E., Frank-Kamenetskii, M.D. (2006). Base-stacking and base-pairing contributions into thermal stability of the DNA helix. *Nucleic Acids Research* 34.2, 564-574.

Ye, F., Samuels, D.C., Clark, T., Guo, Y. (2014). High-throughput sequencing in mitochondrial DNA research. *Mitochondrion* 17, 157-163.

Zuccarelli, G., Alechine, E., Caputo, M., Bobillo, C., Corach, D., Sala, A. (2011). Rapid screening for Native American mitochondrial and Y-chromosome haplogroups detection in routine DNA analysis. *Forensic Science International: Genetics* 639, 1-4.

Zhao, Z., Yu, N., Fu, Y.X., Li, W.H. (2006). Nucleotide variation and haplotype diversity in a 10-kb noncoding region in three continental human populations. *Genetics*, 174:399-409. [Pubmed: 16783003]

Zhou, L., Palais, R.A., Ye, F., Chen, J., Montgomery, J.L., Wittwer, C.T. (2013). Symmetric snapback primers for scanning and genotyping of the cystic fibrosis transmembrane conductance regulator gene. *Clin Chem* 59, 1052-61.

Supplementary data

A table with the Nano-drop results for the extracted DNA samples, including the haplogroup assignment from the melting curve analysis.

Lab code	**Concentration**	**Purity**	**Expected haplogroup**
FDL_H1_S1_160_0041	65.3	1.79	Ancestral for L1,L0&L3
FDL_H1_S1_160_0044	3.1	2.08	Ancestral for L1,L0&L3
FDL_H1_S1_160_0045	7.8	1.54	Ancestral for L1,L0&L3
FDL_H1_S1_160_0046	2.2	1.87	L3
FDL_H1_S1_160_0047	14.7	1.72	L0
FDL_H1_S1_160_0048	40.1	1.8	Ancestral for L1,L0&L3
FDL_H1_S1_160_0049	143.7	1.79	L0
FDL_H1_S1_160_0050	5.5	2.33	L0
FDL_H1_S1_160_0052	8.6	1.19	L3
FDL_H1_S1_160_0053	10.9	1.29	Ancestral for L1,L0&L3
FDL_H1_S1_160_0054	5.7	1.04	L3
FDL_H1_S1_160_0055	3.2	1.11	L0
FDL_H1_S1_160_0056	2.8	0.9	Ancestral for L1,L0&L3
FDL_H1_S1_160_0057	8.3	0.92	L0
FDL_H1_S1_160_0058	12.8	1.09	Ancestral for L1,L0&L3
FDL_H1_S1_160_0059	82.8	1.75	L0
FDL_H1_S1_160_0060	1.8	1.19	Ancestral for L1,L0&L3
FDL_H1_S1_160_0065	12	1.57	Ancestral for L1,L0&L3
FDL_H1_S1_160_0067	23.1	1.73	L1
FDL_H1_S1_160_0068	279.8	1.81	Ancestral for L1,L0&L3
FDL_H1_S1_160_0069	90.5	1.83	Ancestral for L1,L0&L3
FDL_H1_S1_160_0070	45.7	1.46	Ancestral for L1,L0&L3
FDL_H1_S1_160_0071	6	1.7	Ancestral for L1,L0&L3
FDL_H1_S1_160_0072	30.4	1.82	L3
FDL_H1_S1_160_0073	122.8	1.81	Ancestral for L1,L0&L3
FDL_H1_S1_160_0074	119.4	1.86	L0
FDL_H1_S1_160_0075	6.4	1.76	Ancestral for L1,L0&L3
FDL_H1_S1_160_0076	64.2	1.9	Ancestral for L1,L0&L3
FDL_H1_S1_160_0077	53.5	1.81	L3
FDL_H1_S1_160_0078	39.3	1.92	L3

FDL_H1_S1_160_0079	133.8	1.82	L0
FDL_H1_S1_160_0081	19.8	1.56	L1
FDL_H1_S1_160_0082	3.6	1.1	L0
FDL_H1_S1_160_0084	1.5	1.1	L1
FDL_H1_S1_160_0085	6.8	1.45	L0
FDL_H1_S1_160_0086	5.1	1.29	Ancestral for L1,L0&L3
FDL_H1_S1_160_0087	9.7	1.1	Ancestral for L1,L0&L3
FDL_H1_S1_160_0088	2.8	1.66	L1
FDL_H1_S1_160_0089	3.4	1.02	Ancestral for L1,L0&L3
FDL_H1_S1_160_0090	3.1	1.35	Ancestral for L1,L0&L3
FDL_H1_S1_160_0093	4.9	1.24	Ancestral for L1,L0&L3
FDL_H1_S1_160_0094	20.6	1.21	L3
FDL_H1_S1_160_0095	4.9	1.82	Ancestral for L1,L0&L3
FDL_H1_S1_160_0096	6	1.08	L0
FDL_H1_S1_160_0097	10.3	1.34	L1
FDL_H1_S1_160_0099	2.6	1.23	Ancestral for L1,L0&L3
FDL_H1_S1_160_0100	18.3	1.67	L3
FDL_H1_S1_160_0102	93.3	1.87	Ancestral for L1,L0&L3
FDL_H1_S1_160_0103	11.5	1.85	L0
FDL_H1_S1_160_0105	169.5	1.82	Ancestral for L1,L0&L3
FDL_H1_S1_160_0121	19.1	1.57	L0
FDL_H1_S1_160_0122	28.3	1.76	L0
FDL_H1_S1_160_0123	21.1	1.81	L0
FDL_H1_S1_160_0124	94.1	1.78	L3
FDL_H1_S1_160_0126	9.1	1.66	L3
FDL_H1_S1_160_0127	16.5	1.6	L3
FDL_H1_S1_160_0128	8.4	1.66	Ancestral for L1,L0&L3
FDL_H1_S1_160_0129	2.3	1.47	Ancestral for L1,L0&L3
FDL_H1_S1_160_0130	6.4	1.49	Ancestral for L1,L0&L3
FDL_H1_S1_160_0131	109.4	1.76	Ancestral for L1,L0&L3
FDL_H1_S1_160_0132	21	1.56	L0
FDL_H1_S1_160_0133	21.4	1.74	L0
FDL_H1_S1_160_0134	42.3	1.7	L3
FDL_H1_S1_160_0136	68.4	1.71	L3
FDL_H1_S1_160_0137	10.2	1.44	Ancestral for L1,L0&L3
FDL_H1_S1_160_0138	11.6	1.34	Ancestral for L1,L0&L3
FDL_H1_S1_160_0139	6	1.21	L0
FDL_H1_S1_160_0140	10.5	1.62	Ancestral for L1,L0&L3
FDL_H1_S1_160_0141	47.1	1.67	Ancestral for L1,L0&L3
FDL_H1_S1_160_0142	149.1	1.73	Sequence
FDL_H1_S1_160_0143	34.2	1.81	L0
FDL_H1_S1_160_0145	5.1	1.74	Ancestral for L1,L0&L3
FDL_H1_S1_160_0148	7.3	2.12	Ancestral for L1,L0&L3

FDL_H1_S1_160_0149	30.9	1.49	Ancestral for L1,L0&L3
FDL_H1_S1_160_0150	27.6	1.39	Ancestral for L1,L0&L3
FDL_H1_S1_160_0151	1,4	1,17	L0
FDL_H1_S1_160_0152	40	1,88	L0
FDL_H1_S1_160_0153	7,5	1,74	Ancestral for L1,L0&L3
FDL_H1_S1_160_0154	6,1	1,58	L3
FDL_H1_S1_160_0155	77,7	1,79	L0
FDL_H1_S1_160_0156	13,2	1,8	Ancestral for L1,L0&L3
FDL_H1_S1_160_0157	103,9	1,77	Ancestral for L1,L0&L3
FDL_H1_S1_160_0172	83	1.64	Ancestral for L1,L0&L3
FDL_H1_S1_160_0173	56.3	1.7	L3
FDL_H1_S1_160_0175	49	1.7	Ancestral for L1,L0&L3
FDL_H1_S1_160_0176	2.8	1.28	Ancestral for L1,L0&L3
FDL_H1_S1_160_0177	152.1	1.77	L0
FDL_H1_S1_160_0180	5.9	1.28	Ancestral for L1,L0&L3
FDL_H1_S1_160_0181	39,7	1,63	L1
FDL_H1_S1_160_0182	92,4	1,74	Ancestral for L1,L0&L3
FDL_H1_S1_160_0183	41	1,79	Ancestral for L1,L0&L3
FDL_H1_S1_160_0184	11,6	1,72	Ancestral for L1,L0&L3
FDL_H1_S1_160_0185	6	1,63	L1
FDL_H1_S1_160_0186	47,3	1,8	L0
FDL_H1_S1_160_0187	198	1,85	L0
FDL_H1_S1_160_0188	10,6	1,71	Ancestral for L1,L0&L3
FDL_H1_S1_160_0189	177,1	1,86	Ancestral for L1,L0&L3
FDL_H1_S1_160_0201	61.5	1.93	L3
FDL_H1_S1_160_0202	14.2	1.73	L1
FDL_H1_S1_160_0203	3.4	2.16	L0
FDL_H1_S1_160_0204	14.8	2.03	L0
FDL_H1_S1_160_0205	27	1.9	L0
FDL_H1_S1_160_0206	28.9	1.8	Ancestral for L1,L0&L3
FDL_H1_S1_160_0207	32.8	1.81	Ancestral for L1,L0&L3
FDL_H1_S1_160_0208	30.8	1.36	Ancestral for L1,L0&L3
FDL_H1_S1_160_0209	2	1.67	L3
FDL_H1_S1_160_0210	40.9	1.77	Ancestral for L1,L0&L3
FDL_H1_S1_160_0211	69.4	1.85	L3
FDL_H1_S1_160_0212	81	1.57	Ancestral for L1,L0&L3
FDL_H1_S1_160_0213	90.7	1.79	L3
FDL_H1_S1_160_0214	28.9	1.8	Ancestral for L1,L0&L3
FDL_H1_S1_160_0215	285.2	1.74	L0
FDL_H1_S1_160_0216	10.1	1.55	Ancestral for L1,L0&L3
FDL_H1_S1_160_0217	10.9	1.7	Ancestral for L1,L0&L3
FDL_H1_S1_160_0218	11.7	1.97	L3
FDL_H1_S1_160_0219	57.4	1.82	L3

FDL_H1_S1_160_0220	29	1.87	Ancestral for L1,L0&L3
FDL_H1_S1_160_0221	95.1	1.92	L1
FDL_H1_S1_160_0222	70.4	1.89	L3
FDL_H1_S1_160_0223	6.3	1.42	Ancestral for L1,L0&L3
FDL_H1_S1_160_0224	88.2	1.9	L3
FDL_H1_S1_160_0225	2.8	1.69	Ancestral for L1,L0&L3
FDL_H1_S1_160_0226	179	1.92	L1
FDL_H1_S1_160_0227	10.6	1.77	Ancestral for L1,L0&L3
FDL_H1_S1_160_0228	305.1	1.89	L3
FDL_H1_S1_160_0229	7.4	1.7	Ancestral for L1,L0&L3
FDL_H1_S1_160_0230	10.2	1.72	Ancestral for L1,L0&L3
FDL_H1_S1_160_0231	105.3	1.88	L0
FDL_H1_S1_160_0232	49.9	1.91	Ancestral for L1,L0&L3
FDL_H1_S1_160_0233	16.9	2.03	L3
FDL_H1_S1_160_0234	26.2	1.84	Ancestral for L1,L0&L3
FDL_H1_S1_160_0235	696.2	1.87	L3
FDL_H1_S1_160_0236	19.5	1.96	L1
FDL_H1_S1_160_0237	84.1	1.91	L3
FDL_H1_S1_160_0238	37.8	1.83	L3
FDL_H1_S1_160_0239	28.7	1.96	Ancestral for L1,L0&L3
FDL_H1_S1_160_0240	87.8	1.86	Ancestral for L1,L0&L3
FDL_H1_S1_160_0241	21.6	1.79	Ancestral for L1,L0&L3
FDL_H1_S1_160_0242	18.8	1.75	L3
FDL_H1_S1_160_0243	5.9	1.74	Ancestral for L1,L0&L3
FDL_H1_S1_160_0244	139.4	1.71	Ancestral for L1,L0&L3
FDL_H1_S1_160_0245	176.5	1.81	Ancestral for L1,L0&L3
FDL_H1_S1_160_0246	65.1	1.72	Ancestral for L1,L0&L3
FDL_H1_S1_160_0247	10.7	1.79	L3
FDL_H1_S1_160_0248	3.1	1.53	Ancestral for L1,L0&L3
FDL_H1_S1_160_0249	45.9	1.79	Ancestral for L1,L0&L3
FDL_H1_S1_160_0250	71.7	1.84	Ancestral for L1,L0&L3
FDL_H1_S1_160_0251	145.2	1.82	L0
FDL_H1_S1_160_0252	63.7	1.89	L1
FDL_H1_S1_160_0253	25.4	1.95	Ancestral for L1,L0&L3
FDL_H1_S1_160_0254	8.6	1.62	Ancestral for L1,L0&L3
FDL_H1_S1_160_0255	4.8	1.86	L3
FDL_H1_S1_160_0256	14.6	1.95	L3
FDL_H1_S1_160_0257	43	1.92	Ancestral for L1,L0&L3
FDL_H1_S1_160_0258	2.5	1.93	L0
FDL_H1_S1_160_0259	16.9	1.93	Ancestral for L1,L0&L3
FDL_H1_S1_160_0260	9.9	1.84	L0
FDL_H1_S1_160_0261	1.9	2.28	L3
FDL_H1_S1_160_0262	63.8	1.91	L3

FDL_H1_S1_160_0263	86.1	1.91	L3
FDL_H1_S1_160_0264	92.8	1.9	L3
FDL_H1_S1_160_0265	79.5	1.88	Ancestral for L1,L0&L3
FDL_H1_S1_160_0266	288.8	1.87	L3
FDL_H1_S1_160_0267	50.3	1.89	L3
FDL_H1_S1_160_0268	48.3	1.91	Ancestral for L1,L0&L3
FDL_H1_S1_160_0269	82.8	1.92	Ancestral for L1,L0&L3
FDL_H1_S1_160_0270	2.1	1.86	Ancestral for L1,L0&L3
FDL_H1_S1_160_0271	59.8	1.8	L3
FDL_H1_S1_160_0272	193	1.85	Ancestral for L1,L0 & L3
FDL_H1_S1_160_0273	101.6	1.87	Ancestral for L1,L0 & L3
FDL_H1_S1_160_0274	30.3	1.83	Rare sample/sequence
FDL_H1_S1_160_0275	33.3	1.92	L3
FDL_H1_S1_160_0276	23.4	1.78	L0
FDL_H1_S1_160_0277	19	1.86	L0
FDL_H1_S1_160_0278	39.8	1.81	Ancestral for L1,L0&L3
FDL_H1_S1_160_0279	17.9	2.01	Rare sample /sequence
FDL_H1_S1_160_0280	14.2	1.71	Ancestral for L1,L0&L3
FDL_H1_S1_160_0281	1.3	1.4	Ancestral for L1,L0&L3
FDL_H1_S1_160_0282	4.4	2	Ancestral for L1,L0&L3
FDL_H1_S1_160_0283	18.5	1.93	Ancestral for L1,L0&L3
FDL_H1_S1_160_0284	272.6	1.86	L0
FDL_H1_S1_160_0285	30	1.73	L1
FDL_H1_S1_160_0286	133.1	1.77	Ancestral for L1,L0&L3
FDL_H1_S1_160_0287	21.2	1.97	Ancestral for L1,L0&L3
FDL_H1_S1_160_0288	204.9	1.85	Ancestral for L1,L0&L3
FDL_H1_S1_160_0289	23.4	1.86	L1
FDL_H1_S1_160_0290	14.6	1.95	Ancestral for L1,L0&L3
FDL_H1_S1_160_0291	146.8	1.8	L3
FDL_H1_S1_160_0292	20.3	1.96	L3
FDL_H1_S1_160_0293	32.2	1.87	L3
FDL_H1_S1_160_0294	187.3	1.68	Ancestral for L1,L0&L3
FDL_H1_S1_160_0295	54.2	1.92	Ancestral for L1,L0&L3
FDL_H1_S1_160_0296	115.7	1.87	L0
FDL_H1_S1_160_0297	93.9	1.9	Ancestral for L1,L0&L3
FDL_H1_S1_160_0299	26.9	1.85	L3
FDL_H1_S1_160_0300	124.9	1.85	L3
FDL_H1_S1_160_0310	30.1	1.76	Ancestral for L1,L0 &L3
FDL_H1_S1_160_0312	67.3	1.62	Ancestral for L1,L0&L3
FDL_H1_S1_160_0314	63.1	1.59	Sequence
FDL_H1_S1_160_0315	19.2	1.49	L3
FDL_H1_S1_160_0316	14.4	1.12	L1
FDL_H1_S1_160_0317	5.4	1.75	Ancestral for L1,L0 &L3

FDL_H1_S1_160_0318	44.9	1.77	Ancestral for L1,L0 &L3
FDL_H1_S1_160_0319	88.7	1.69	Ancestral for L1,L0 &L3
FDL_H1_S1_160_0320	19.2	1.46	Ancestral for L1,L0 &L3
FDL_H1_S1_160_0501	57,7	1,93	L0
FDL_H1_S1_160_0502	67,1	1,93	L0
FDL_H1_S1_160_0503	30,6	1,9	L1
FDL_H1_S1_160_0504	63,2	1,79	Ancestral for L1,L0&L3
FDL_H1_S1_160_0505	26,3	1,9	L0
FDL_H1_S1_160_0506	35,3	1,9	L0
FDL_H1_S1_160_0507	27,9	2,07	L0
FDL_H1_S1_160_0508	70,5	1,87	L0
FDL_H1_S1_160_0509	306,9	1,87	L0
FDL_H1_S1_160_0510	82,1	1,85	Ancestral for L1,L0&L3
FDL_H1_S1_160_0511	19,6	1,79	L1
FDL_H1_S1_160_0512	39,1	1,79	L0
FDL_H1_S1_160_0513	16,7	1,8	Ancestral for L1,L0&L3
FDL_H1_S1_160_0514	110,2	1,86	L1
FDL_H1_S1_160_0515	100,3	1,79	L3
FDL_H1_S1_160_0516	208,1	1,87	L0
FDL_H1_S1_160_0517	17,3	1,76	L0
FDL_H1_S1_160_0518	12,8	1,73	L0
FDL_H1_S1_160_0519	4,8	1,57	L1
FDL_H1_S1_160_0520	57,4	1,87	L1
FDL_H1_S1_160_0521	257,5	1,82	L0
FDL_H1_S1_160_0522	34	1,73	L0
FDL_H1_S1_160_0523	171,5	1,82	L0
FDL_H1_S1_160_0524	29,4	1,91	Ancestral for L1,L0&L3
FDL_H1_S1_160_0525	44,8	1,78	Ancestral for L1,L0&L3
FDL_H1_S1_160_0526	184,1	1,87	L3
FDL_H1_S1_160_0527	171	1,78	L1
FDL_H1_S1_160_0528	101,8	1,79	Ancestral for L1,L0&L3
FDL_H1_S1_160_0529	113,6	1,89	L3
FDL_H1_S1_160_0530	43,8	1,72	L3
FDL_H1_S1_160_0531	5.3	2.73	Ancestral for L1,L0&L3
FDL_H1_S1_160_0532	41.9	1.96	Ancestral for L1,L0&L3
FDL_H1_S1_160_0533	15.7	1.94	L3
FDL_H1_S1_160_0534	131.7	1.86	L3
FDL_H1_S1_160_0535	69.1	1.71	L1
FDL_H1_S1_160_0536	19.6	1.79	L0
FDL_H1_S1_160_0537	147.9	1.84	L3
FDL_H1_S1_160_0538	14.1	2.05	Ancestral for L1,L0&L3
FDL_H1_S1_160_0539	24.9	1.94	Ancestral for L1,L0&L3
FDL_H1_S1_160_0540	9.3	2.07	L3

FDL_H1_S1_160_0541	47,5	1,92	L0
FDL_H1_S1_160_0542	65,5	1,8	Ancestral for L1,L0&L3
FDL_H1_S1_160_0543	63,2	1,85	Ancestral for L1,L0&L3
FDL_H1_S1_160_0544	63	1,8	Ancestral for L1,L0&L3
FDL_H1_S1_160_0545	223,5	1,83	Ancestral for L1,L0&L3
FDL_H1_S1_160_0546	189,1	1,85	L0
FDL_H1_S1_160_0547	146,7	1,88	L0
FDL_H1_S1_160_0548	19,6	1,8	L1
FDL_H1_S1_160_0549	62,5	1,82	L0
FDL_H1_S1_160_0550	60,2	1,82	L0
FDL_H1_S1_160_0551	20,4	1,7	L0
FDL_H1_S1_160_0552	84,8	1,69	L0
FDL_H1_S1_160_0553	51,3	1,87	Ancestral for L1,L0&L3
FDL_H1_S1_160_0554	106,3	1,83	L0
FDL_H1_S1_160_0555	58	1,86	L1
FDL_H1_S1_160_0556	48,5	1,83	Ancestral for L1,L0&L3
FDL_H1_S1_160_0557	23	1,87	Ancestral for L1,L0&L3
FDL_H1_S1_160_0558	74,2	1,83	Ancestral for L1,L0&L3
FDL_H1_S1_160_0559	30,1	1,73	L0
FDL_H1_S1_160_0560	95	1,83	L3
FDL_H1_S1_160_0561	30,6	1,9	L0
FDL_H1_S1_160_0562	58,2	1,89	Ancestral for L1,L0&L3
FDL_H1_S1_160_0563	119,7	1,9	L0
FDL_H1_S1_160_0564	38,2	1,9	L3
FDL_H1_S1_160_0565	28,4	1,98	L3
FDL_H1_S1_160_0566	101,1	1,8	L0
FDL_H1_S1_160_0567	46,2	1,86	L1
FDL_H1_S1_160_0568	34,6	1,95	L1
FDL_H1_S1_160_0569	289,3	1,91	L0
FDL_H1_S1_160_0570	25,9	1,77	L3
FDL_H1_S1_160_0571	17,9	2,02	L3
FDL_H1_S1_160_0572	203	1,9	L3
FDL_H1_S1_160_0573	9	1,73	L0
FDL_H1_S1_160_0574	66,6	1,87	L0
FDL_H1_S1_160_0575	79,8	1,85	L0
FDL_H1_S1_160_0576	8,7	1,85	L0
FDL_H1_S1_160_0577	55,8	1,79	L3
FDL_H1_S1_160_0578	44,4	1,77	L1
FDL_H1_S1_160_0579	107	1,78	L0
FDL_H1_S1_160_0580	7,1	1,65	L0
FDL_H1_S1_160_0581	23,7	1,86	L3
FDL_H1_S1_160_0582	12,1	2,03	L3
FDL_H1_S1_160_0583	67	1,91	L3

FDL_H1_S1_160_0584	177,6	1,83	L3
FDL_H1_S1_160_0585	26,7	1,94	L0
FDL_H1_S1_160_0586	28,6	1,96	L0
FDL_H1_S1_160_0588	158,8	1,77	L0
FDL_H1_S1_160_0589	16,7	1,86	L0
FDL_H1_S1_160_0590	30,3	1,85	L3
FDL_H1_S1_160_0591	82.2	1.9	L0
FDL_H1_S1_160_0592	38.5	1.92	Ancestral for L1,L0&L3
FDL_H1_S1_160_0593	147.1	1.82	Ancestral for L1,L0&L3
FDL_H1_S1_160_0594	22.3	1.95	Ancestral for L1,L0&L3
FDL_H1_S1_160_0595	7.7	2.06	Ancestral for L1,L0&L3
FDL_H1_S1_160_0596	84.8	1.88	Ancestral for L1,L0&L3
FDL_H1_S1_160_0597	29.7	1.86	Ancestral for L1,L0&L3
FDL_H1_S1_160_0598	102.5	1.9	L0
FDL_H1_S1_160_0599	4.8	1.97	L1
FDL_H1_S1_160_0600	18.9	1.59	L0
FDL_H1_S1_160_0601	6,8	1,32	L0
FDL_H1_S1_160_0602	6,2	1,38	L3
FDL_H1_S1_160_0603	13,5	2,3	L1
FDL_H1_S1_160_0604	10,7	1,51	L1
FDL_H1_S1_160_0605	100,6	1,8	L0
FDL_H1_S1_160_0606	65,8	1,96	L0
FDL_H1_S1_160_0607	144,5	2	L0
FDL_H1_S1_160_0608	42,4	1,95	L0
FDL_H1_S1_160_0609	28	1,91	L0
FDL_H1_S1_160_0610	135,8	1,87	L0
FDL_H1_S1_160_0611	20,1	1,87	Ancestral for L1,L0&L3
FDL_H1_S1_160_0612	23,3	1,84	Ancestral for L1,L0&L3
FDL_H1_S1_160_0613	21,8	1,83	L1
FDL_H1_S1_160_0614	46,1	1,86	L3
FDL_H1_S1_160_0615	20,8	1,7	Ancestral for L1,L0&L3
FDL_H1_S1_160_0616	23,6	1,79	L0
FDL_H1_S1_160_0617	9,5	1,93	L3
FDL_H1_S1_160_0618	23,7	1,77	L0
FDL_H1_S1_160_0619	30,7	1,9	L0
FDL_H1_S1_160_0620	17,5	1,95	L0
FDL_H1_S1_160_0621	145,4	1,83	Ancestral for L1,L0&L3
FDL_H1_S1_160_0622	38,2	1,92	Ancestral for L1,L0&L3
FDL_H1_S1_160_0623	66,7	1,87	L3
FDL_H1_S1_160_0624	24,8	1,8	L3
FDL_H1_S1_160_0625	32,1	2,28	L3
FDL_H1_S1_160_0626	54,8	1,78	L0
FDL_H1_S1_160_0627	44,7	1,65	Ancestral for L1,L0&L3

FDL_H1_S1_160_0628	134,5	1,81	Ancestral for L1,L0&L3
FDL_H1_S1_160_0629	70,7	1,85	L1
FDL_H1_S1_160_0630	87,8	1,82	L0
FDL_H1_S1_160_0631	31,5	1,8	Ancestral for L1,L0&L3
FDL_H1_S1_160_0632	7,7	1,88	Ancestral for L1,L0&L3
FDL_H1_S1_160_0633	44,3	1,84	Ancestral for L1,L0&L3
FDL_H1_S1_160_0634	145,5	1,86	Ancestral for L1,L0& L3
FDL_H1_S1_160_0635	132	1,88	L0
FDL_H1_S1_160_0637	13,3	1,67	L1
FDL_H1_S1_160_0638	52,6	1,71	L1
FDL_H1_S1_160_0639	88,1	1,86	L0
FDL_H1_S1_160_0640	46,3	1,73	Ancestral for L1,L0 &L3
FDL_H1_S1_160_0641	54	1,94	L3
FDL_H1_S1_160_0642	65	1,93	L1
FDL_H1_S1_160_0643	121,4	1,84	L3
FDL_H1_S1_160_0644	14	2,53	L3
FDL_H1_S1_160_0645	39,7	1,99	Ancestral for L1,L0&L3
FDL_H1_S1_160_0646	42,3	2,01	Ancestral for L1,L0&L3
FDL_H1_S1_160_0647	28,2	2,13	L0
FDL_H1_S1_160_0648	6,5	2,32	Ancestral for L1,L0&L3
FDL_H1_S1_160_0649	58	1,95	Ancestral for L1,L0 &L3
FDL_H1_S1_160_0650	5	1.74	L3
FDL_H1_S1_160_0651	93,7	1,85	Ancestral for L1,L0&L3
FDL_H1_S1_160_0652	57,9	1,83	L3
FDL_H1_S1_160_0653	51,1	1,78	L3
FDL_H1_S1_160_0654	227,8	1,81	L0
FDL_H1_S1_160_0655	47	1,84	L0
FDL_H1_S1_160_0656	118,4	1,87	L0
FDL_H1_S1_160_0657	47,1	1,96	Ancestral for L1,L0&L3
FDL_H1_S1_160_0658	57,6	1,83	L3
FDL_H1_S1_160_0659	127,7	1,85	L1
FDL_H1_S1_160_0660	363,1	1,88	L0
FDL_H1_S1_160_0661	82	1,78	L3
FDL_H1_S1_160_0662	57,3	1,76	L0
FDL_H1_S1_160_0663	125	1,8	L0
FDL_H1_S1_160_0664	195,8	1,82	Ancestral for L1,L0&L3
FDL_H1_S1_160_0665	20,1	1,61	Ancestral for L1,L0&L3
FDL_H1_S1_160_0666	39,9	1,67	L3
FDL_H1_S1_160_0667	100,5	1,94	L3
FDL_H1_S1_160_0668	10,9	1,31	L3
FDL_H1_S1_160_0669	225,7	1,84	L1
FDL_H1_S1_160_0670	65,2	1,74	L1
FDL_H1_S1_160_0671	132,1	1,86	L0

FDL_H1_S1_160_0672	15,6	1,88	L0
FDL_H1_S1_160_0673	7,3	1,97	L0
FDL_H1_S1_160_0674	51,5	1,88	L3
FDL_H1_S1_160_0675	102,5	1,89	L3
FDL_H1_S1_160_0676	37,6	1,74	L0
FDL_H1_S1_160_0677	15,4	2,1	Ancestral for L1,L0&L3
FDL_H1_S1_160_0679	52,1	1,87	L0
FDL_H1_S1_160_0680	58,2	1,84	L1
FDL_H1_S1_160_0681	25	1,8	L3
FDL_H1_S1_160_0682	63,6	1,88	L0
FDL_H1_S1_160_0683	39,5	1,79	Ancestral for L1,L0&L3
FDL_H1_S1_160_0684	79,6	1,85	Ancestral for L1,L0&L3
FDL_H1_S1_160_0685	33,4	1,87	Ancestral for L1,L0&L3
FDL_H1_S1_160_0686	40,2	1,86	L3
FDL_H1_S1_160_0687	303,7	1,78	L1
FDL_H1_S1_160_0688	40,5	1,86	L0
FDL_H1_S1_160_0689	64,4	1,74	L0
FDL_H1_S1_160_0690	43,3	1,83	L1
FDL_H1_S1_160_0691	33,9	1,88	L1
FDL_H1_S1_160_0692	184,2	1,83	L0
FDL_H1_S1_160_0693	11,2	1,97	L0
FDL_H1_S1_160_0694	40,4	1,88	Ancestral for L1,L0&L3
FDL_H1_S1_160_0695	40,5	1,85	L0
FDL_H1_S1_160_0696	14,7	1,87	Ancestral for L1,L0&L3
FDL_H1_S1_160_0697	110,7	184	Ancestral for L1,L0&L3
FDL_H1_S1_160_0698	17,5	1,86	L0
FDL_H1_S1_160_0699	23,1	1,8	L0
FDL_H1_S1_160_0700	18,1	1,89	L0
FDL_H1_S1_160_0701	151,7	1,95	L3
FDL_H1_S1_160_0702	47,3	1,78	Ancestral for L1,L0&L3
FDL_H1_S1_160_0703	26,3	1,85	Ancestral for L1,L0&L3
FDL_H1_S1_160_0705	24,2	1,68	Ancestral for L1,L0&L3
FDL_H1_S1_160_0706	16,1	1,86	Ancestral for L1,L0&L3
FDL_H1_S1_160_0707	21,5	1,7	L3
FDL_H1_S1_160_0708	231	1,83	Ancestral for L1,L0&L3
FDL_H1_S1_160_0709	134,3	1,86	Ancestral for L1,L0&L3
FDL_H1_S1_160_0710	63,5	1,78	L3
FDL_H1_S1_160_0711	51,1	1,7	L3
FDL_H1_S1_160_0712	3	1,35	L3
FDL_H1_S1_160_0713	3,7	1,17	L0
FDL_H1_S1_160_0714	31,9	1,72	Ancestral for L1,L0&L3
FDL_H1_S1_160_0715	82,5	1,75	L1
FDL_H1_S1_160_0717	33,9	1,77	Ancestral for L1,L0&L3

FDL_H1_S1_160_0718	36,6	1,61	Ancestral for L1,L0& L3
FDL_H1_S1_160_0719	36,2	1,74	L3
FDL_H1_S1_160_0720	147,2	1,8	Ancestral for L1,L0& L3
FDL_H1_S1_160_0721	190,1	1,96	L0
FDL_H1_S1_160_0722	21,3	1,74	Ancestral for L1,L0& L3
FDL_H1_S1_160_0723	0,9	1,78	Ancestral for L1,L0& L3
FDL_H1_S1_160_0724	32,7	1,89	Ancestral for L1,L0& L3
FDL_H1_S1_160_0725	113,5	1,86	L1
FDL_H1_S1_160_0726	140,1	1,84	L0
FDL_H1_S1_160_0727	23,6	1,78	Sequence
FDL_H1_S1_160_0728	153,5	1,9	L3
FDL_H1_S1_160_0729	0,7	1,35	L3
FDL_H1_S1_160_0730	21,1	1,84	Ancestral for L1,L0& L3
FDL_H1_S1_160_0731	38,4	1,85	Ancestral for L1,L0& L3
FDL_H1_S1_160_0732	98,9	1,87	Ancestral for L1,L0&L3
FDL_H1_S1_160_0733	59,8	1,85	L0
FDL_H1_S1_160_0734	41,8	1,77	Ancestral for L1,L0& L3
FDL_H1_S1_160_0736	8,3	1,59	Ancestral for L1,L0& L3
FDL_H1_S1_160_0737	27,5	1,72	Ancestral for L1,L0& L3
FDL_H1_S1_160_0738	89	1,81	L0
FDL_H1_S1_160_0739	66,2	1,9	Ancestral for L1,L0& L3
FDL_H1_S1_160_0740	26,7	1,84	L1
FDL_H1_S1_160_0741	46,6	1,79	L1
FDL_H1_S1_160_0742	27,8	1,57	Ancestral for L1,L0&L3
FDL_H1_S1_160_0743	167,1	1,8	Ancestral for L1,L0& L3
FDL_H1_S1_160_0744	88,4	1,79	Ancestral for L1,L0& L3
FDL_H1_S1_160_0745	31,9	1,56	L1
FDL_H1_S1_160_0746	41,6	1,75	L0
FDL_H1_S1_160_0747	13,2	1,77	Ancestral for L1,L0& L3
FDL_H1_S1_160_0748	72,1	1,75	L1
FDL_H1_S1_160_0749	67	1,76	Ancestral for L1,L0& L3
FDL_H1_S1_160_0750	15,2	1,69	Ancestral for L1,L0&L3
FDL_H1_S1_160_0751	23	1,8	Ancestral for L1,L0& L3
FDL_H1_S1_160_0752	23,9	1,88	Ancestral for L1,L0&L3
FDL_H1_S1_160_0753	157,6	1,87	Ancestral for L1,L0& L3
FDL_H1_S1_160_0754	138,4	1,77	L3
FDL_H1_S1_160_0755	77,8	1,7	Ancestral for L1,L0&L3
FDL_H1_S1_160_0756	10,1	1,86	Ancestral for L1,L0& L3
FDL_H1_S1_160_0757	12,9	1,82	L3
FDL_H1_S1_160_0758	53,7	1,87	Ancestral for L1,L0&L3
FDL_H1_S1_160_0759	65,6	1,73	Ancestral for L1,L0&L3

Table with the melting temperatures for samples screened with Multiplex 1.

Theoretical melting temperatures	**Temperature range**		
Primer site	**D**	**A**	
L3	73.5	72	
L1	75.5	74.5	
L0	81	80.5	
Melting temperatures for Multiplex 1			
Sample name	**Peak 1 (°C)**	**Peak 2(°C)**	**Peak 3(°C)**
936	71.95	72.27	75.1
312	72.27	74.18	77.97
318	72.47	74.85	80.67
320	72.27	74	80.65
274	72.07	75.28	81.42
202	72.6	75.62	80.57
204	72.47	74	81.35
205	72.45	74.65	81.45
206	72.5	74.67	77.65
207	72.6	74.8	77.65
208	72.4	74.57	80.45
210	72.4	74.62	80.52
203	71.72	74.03	80.93
249	72.08	74.33	80.2
252	71.98	74.92	80.25
262	73.55	74.11	80.27
274	72.4	75.52	81.35
314	72.18	75.3	81.22
826	72.32	74.52	80.42
821	72.08	75.4	80.3
219	73.77	74.08	80.17
263	73.04	74.19	80.15
283	71.85	74.15	80.28
286	71.85	74.3	80.25
288	71.37	71.67	73.95

290	71.95	74.35	77.7
285	71.93	75.25	80.1
279	72.97	75.3	80.33
289	71.98	72.38	75.6
519	72.1	75.9	78.63
503	72.42	75.48	80.57
264	73.1	74.1	80.47
300	73.15	74.47	80.47
256	73.92	74.35	82.95
511	72.3	75.65	80.53
514	72.5	75.75	80.6
568	72.95	75.65	80.85
572	74.92	75.22	81
581	74.8	81.02	
584	74.7	77.35	80.8
567	73.15	76.4	81.27
571	74.48	75.15	80.87
821	73.35	75.35	80.65
279	73.62	75.78	80.97
314	72.82	76.12	78.6
631	71.9	74.35	80.35
632	71.85	72.18	74.12
633	71.97	74.55	80.45
635	71.95	74.47	81.33
643	73.47	77.17	80.57
645	72.4	74.67	80.7
646	72.1	74.5	80.52
648	72.35	74.6	80.55
653	73.67		
654	72.37	74.67	81.6
658	73.52	80.53	
664	71.97	74.52	80.48
665	72.35	74.58	80.53
669	71.93	75.28	80.47
672	72.2	74.65	80.6
677	71.9	74.48	80.45
679	72.65	74.75	81.72
680	72.07	75.45	80.42
681	73.47	80.53	
682	72.32	74.6	81.48
683	72.05	74.52	80.5
684	71.98	74.5	80.57
685	72.2	74.55	80.62

686	73.55	80.6	
694	71.87	74.75	80.72
697	71.93	74.52	80.52
698	72.03	74.42	81.48
699	72.3	74.52	81.18
701	73.5	80.3	
703	72.35	74.6	80.52
706	71.81	74.63	80.6
707	73.53	80.45	
675	73.86	74.87	80.4
564	74.7	80.75	80.73
708	72.8	73.55	74.85
636	73	75.08	81.05
710	73.57	71.57	74.8
711	74.7	80.8	
712	74.17	74.61	80.36
713	72.85	74.85	81.45
714	73	74.98	80.9
715	72.68	75.91	75.9
717	72.85	74.83	80.7
718	72.97	74.95	80.97
720	72.85	73.55	74.87
721	72.93	74.9	81.9
722	72.9	74.93	80.75
723	72.3	75.35	80.32
724	72.85	73.58	74.9
725	72.55	75.7	80.55
726	72.88	75.75	81.68
729	74	74.7	80.33
730	73.05	75.07	81.02
733	72.9	74.92	81.52
679	72.35	74.75	81.62
694	72.2	74.37	80.43
695	72.35	74.6	81.38
700	72.23	74.55	81.41
709	72.05	72.25	74.35
719	74.3	80.53	
727	73.07	75.47	80.53
731	72.65	74.83	80.32
734	72.03	74.4	80.4
736	72.23	74.57	80.53
738	71.55	72.25	81.5
739	72.18	74.53	80.45

740	71.95	75.33	80.2
743	72.25	74.38	80.35
744	72.4	74.8	80.57
746	72.63	74.83	81.7
747	72.15	74.37	80.35
748	71.93	75.25	80.14
749	72.32	74.6	80.48
751	72.28	74.55	80.45
753	72.2	74.3	79.1
754	74.12	80.25	
756	72.33	74.63	80.43
757	73.52	80.38	
679	72.3	73.15	81.43
694	71.95	72.4	74.5
695	71.97	74.37	81.17
700	72.25	74.4	81.3
719	74.35	80.47	
727	72.9	75.25	80.37
728	74.35	79.8	
731	71.05	73.23	75.28
734	72.23	74.35	80.3
736	72.05	74.23	79.9
738	72.25	74.35	81.12
739	72.25	74.58	80.5
740	71.75	75.02	79.93
743	73.03	75.22	81.05
744	73.1	75.25	81.05
747	72.32	74.55	80.5
748	71.68	75.05	79.77
749	72.3	74.52	80.45
751	71.88	74.2	79.98
753	71.9	74.28	78.97
754	74.35	80.27	
757	73.6	74.1	80.15
757	73.6	74.1	80.15
218	74.15	80.1	
225	71.9	74.42	80.15
231	72.23	74.35	81.2
232	72.17	74.47	80.45
233	74.4	80.33	
235	73.02	74.35	80.2
237	74.05	80.2	
238	73.07	74.25	80.18

239	71.92	74.37	80.35
247	74.15	80.27	
248	71.5	73.98	79.98
255	74.4	80.5	
266	73.04	74.25	80.37
267	74.27	80.4	
268	72.25	74.55	80.57
269	72	74.42	80.4
270	71.42	73.92	79.85
277	71.85	74.35	81.23
280	72.3	74.22	80.45
281	71.68	74.15	80.03
696	71.8	74.35	80.35
676	71.9	74.35	80.35
705	71.8	74.35	80.37
737	71.85	74.35	80.42
741	71.75	75.15	80.35
742	71.75	74.3	80.32
745	71.6	74.92	80.15
750	71.8	74.33	80.17
755	71.97	74.43	80.45
758	71.8	74.22	80.15
759	71.83	74.32	80.4
650	73.85	80.05	
41	72.1	74.55	80.55
277	72.35	74.57	81.03
280	71.87	74.27	79.85
281	71.7	74.08	79.85
696	71.75	74.18	80.12
702	72.12	74.47	80.48
728	72.27	74.2	80.3
737	71.7	75.02	77.62
742	71.72	75.07	80.14
755	71.72	74.18	80.17
758	73.54	75.55	81.37
650	71.95	74.33	76.55
41	72.35	74.7	80.4
218	74.05	80.1	
225	71.9	74.38	80.12
231	72.42	74.65	81.5
232	71.88	74.28	80.25
233	74.4	80.27	
235	74.45	80.05	

237	73.8	79.85	
238	73.41	73.92	79.9
239	71.72	74.1	80.03
247	73.78	79.9	
248	71.52	73.83	79.9
255	75	81	
266	74.05	80.05	
267	74.02	80.03	
268	72.2	74.47	80.25
269	72.4	74.5	80.35
270	71.57	73.9	79.8
54	74.05	80.38	
58	72.35	74.67	80.48
81	72.15	75.4	80.5
84	72.68	75.88	80.7
85	72.18	74.5	81.35
86	71.85	74.27	80.32
94	74.3	80.45	
95	72.23	74.52	80.45
96	72.37	74.62	81.4
99	72.35	74.62	80.53
100	74.55	80.58	
121	72.05	72.17	74.4
123	71.85	74.32	81.25
128	71.95	74.43	80.22
129	72.05	74.5	80.33
130	72.03	74.55	80.33
131	72.03	74.5	80.33
132	72.55	74.7	81.62
54	74.28	80.23	
58	72.35	74.57	80.32
81	72.13	75.3	80.4
84	72.9	76	80.85
85	71.95	74.27	81.15
86	71.68	74.1	79.92
91	72.05	74.33	79.9
94	74.07	79.95	
95	71.97	74.42	80.52
96	72.25	74.65	81.48
98	72.33	74.63	80.55
99	74.3	80.45	
100	71.98	74.3	81.15
122	72.12	74.3	81.1

127	72.02	74.37	79.92
128	72.32	74.57	80.45
129	72.42	73.43	74.65
130	71.85	72.13	74.37
131	72.33	74.62	81.5
70	71.65	72.45	74.65
79	72.13	74.4	81.22
87	72.47	74.67	80.4
88	72.3	75.4	80.47
89	72.45	74.65	80.83
90	72.43	74.63	80.45
93	72.18	74.5	80.27
97	71.83	75.25	80.22
122	72.28	74.55	81.37
124	73.15	74.35	80.28
126	74.32	80.35	
133	72.4	74.62	81.4
134	73.43	74.05	74.4
136	74.48	80.4	80.68
138	72.32	74.62	80.45
139	72.17	74.47	81.27
141	72.4	74.65	80.45
142	72.68	75.2	81.07
143	72.45	74.65	81.65
144	72.28	74.47	81.45
145	72.43	74.62	80.48
146	72.33	74.65	81.5
156	72.35	74.55	80.45
148	72.23	74.38	80.2
149	72.33	74.65	80.42
150	72.4	74.63	80.47
151	72.37	74.53	81.2
152	71.75	72.3	74.48
153	72.25	74.4	80.25
746	71.9	74.32	81.17
236	72.35	75.5	80.47
752	72.3	74.48	80.32
70	72.22	74.47	80.65
79	71.75	74.22	81.03
87	71.78	72	74.37
88	71.98	75.2	80.12
89	72.65	74.82	80.68
90	72.4	74.8	80.5

97	71.58	75.15	79.9
122	71.73	74.22	81
126	73.98	79.87	
133	71.65	74.15	81
134	73.23	74.25	79.93
138	72.5	74.85	80.55
141	71.83	74.15	80.43
142	72.85	75.43	81.41
143	72.23	74.55	81.51
155	71.7	73.95	81.02
145	72	74.38	80.37
146	72.12	74.53	81.35
236	71.97	75.25	80.4
752	71.87	72.85	80.48
77	73.07	74.19	80.33
137	72.03	74.45	80.32
140	71.98	74.35	80.27
142	71.52	73.95	79.97
156	71.97	74.33	80.37
157	72.2	74.6	80.6
160	73.7	74.6	80.67
172	71.92	74.33	80.37
173	73.09	74.43	80.33
175	71.88	74.35	80.2
176	72.02	74.47	80.23
177	71.7	74.3	81.99
180	72.23	74.55	80.4
181	71.88	75.25	80.45
182	71.85	74.23	80.3
183	71.7	74.17	80.22
184	71.58	74.03	80.2
185	71.65	74.95	80.18
186	71.85	74.22	81.3
187	72.18	74.6	81.42
188	71.9	74.18	80.2
189	71.8	74.18	80.18
102	72.32	74.75	80.57
103	71.9	74.3	81.32
105	72.15	74.5	80.43
77	73.07	74.04	80.25
137	71.85	74.35	80.37
140	71.72	74.28	80.28
142	71.57	73.98	76.57

154	73.08	74.08	79.79
156	71.53	73.95	79.87
157	71.85	74.37	80.43
160	73.07	74.7	80.59
172	71.68	73.95	79.83
173	73.07	74.7	80.59
175	71.65	74.1	80.32
176	71.8	74.12	79.82
177	71.75	74.08	81.86
180	71.88	74.28	80.05
181	71.57	75.2	80.13
182	71.5	73.87	79.85
183	71.45	73.82	79.9
184	71.35	73.78	79.78
185	71.35	74.72	79.78
186	71.45	73.83	81.04
187	72.28	74.55	81.38
188	71.58	73.87	80.05
189	71.45	73.85	79.87
102	72.17	74.57	80.42
103	71.53	73.87	81.12
105	71.85	74.3	80.35
89	72.2	74.8	80.8
102	71.62	74.25	80.57
143	71.53	74.22	81.48
160	73.02	74.28	80.55
185	71.67	75.33	80.52
177	72.02	74.52	82.36
289	71.82	75.42	80.73
279	73	75.4	80.65
694	71.93	74.58	80.7
726	71.65	74.28	81.47
706	71.78	74.57	80.58
743	71.72	74.55	80.62
744	71.97	74.57	80.7
218	70.03	74.15	80.1
225	71.9	74.42	80.15
231	72.23	74.35	81.2
232	72.17	74.47	80.45
233	74.4	80.33	
234	72	74.4	
235	72.82	74.35	80.2
236	72.77	74.23	

237	70.07	74.05	80.2
238	70.15	72.75	74.25
239	71.92	74.37	80.35
247	70.02	74.15	80.27
248	71.5	73.98	79.98
255	70.08	70.35	74.4
266	72.95	74.25	80.37
267	74.27	80.4	
268	72.25	74.55	80.57
269	72	74.42	80.4
270	71.42	73.92	79.85
277	71.85	74.35	81.23
280	72.3	74.22	80.45
281	71.68	74.15	80.03
696	71.8	74.35	80.35
676	71.9	74.35	80.35
702	72.42	74.65	
705	71.8	74.35	80.37
728	74.35		
737	71.85	74.35	80.42
741	71.75	75.15	80.35
742	71.75	74.3	80.32
745	71.6	74.92	80.15
750	71.8	74.33	80.17
752	73	75.07	
755	71.97	74.43	80.45
758	71.8	74.22	80.15
759	71.83	74.32	80.4
760	72.35	73.3	74.58
650	73.85	80.05	
41	72.1	74.55	80.55
42	72.55	74.83	
54	74.05	80.38	
58	72.35	74.67	80.48
61	74.53		
81	72.15	75.4	80.5
84	72.68	75.88	80.7
85	72.18	74.5	81.35
86	71.85	74.27	80.32
91	72.05	74.43	
94	74.3	80.45	
95	72.23	74.52	80.45
96	72.37	74.62	81.4

99	72.35	74.62	80.53
736	74.55	80.58	
121	72.05	72.17	74.4
123	71.85	74.32	81.25
124	72.42	74.3	
125	74.72		
128	71.95	74.43	80.22
129	72.05	74.5	80.33
130	72.03	74.55	80.33
131	72.03	74.5	80.33
132	72.55	74.7	81.62
70	71.65	72.45	74.65
79	72.13	74.4	81.22
87	72.47	74.67	80.4
88	72.3	75.4	80.47
89	72.45	74.65	80.83
90	72.43	74.63	80.45
93	72.18	74.5	80.27
97	71.83	75.25	80.22
122	72.28	74.55	81.37
124	73.15	74.35	80.28
125	74.68	80.9	
126	74.32	80.35	
127	74.5		
133	72.4	74.62	81.4
134	73.43	74.05	74.4
135	74.3		
136	74.48	80.4	80.68
138	72.32	74.62	80.45
139	72.17	74.47	81.27
141	72.4	74.65	80.45
142	72.68	74.95	81.07
143	72.45	74.65	81.65
144	72.28	74.47	81.45
145	72.43	74.62	80.48
146	72.33	74.65	81.5
156	72.35	74.55	80.45
148	72.23	74.38	80.2
149	72.33	74.65	80.42
150	72.4	74.63	80.47
151	72.37	74.53	81.2
152	71.75	72.3	74.48
153	72.25	74.4	80.25

154	74.68		
746	71.9	74.32	81.17
236	72.35	75.5	80.47
587	72.25	74.57	
52	72.3	74.48	80.32

Table with melting temperatures of screened samples with Multiplex 2

Sample	**Peak 1**	**Peak 2**
MR14.003	74.47	74.6
S17 (Control M)	75.37	85.95
Eug (Control N)	73	74.67
Pet (Control N)	74.7	85.5
78	74.58	86.18
54	72.27	74.8
94	72.5	74.73
100	74.7	86.47
124	74.63	86.37
126	72.45	74.6
136	74.5	86.32
77	74.57	86.17
583	72.1	74.35
560	74.65	86.53
617	74.27	74.77
623	74.65	86.35
707	74.25	86.55
625	74.75	86.45
658	74.57	86.47
235	74.52	86.05
247	74.68	86.45
255	74.43	85.95
266	74.73	86.55
267	74.95	86.57

650	74.45	86
754	74.5	85.95
757	74.67	86
46	74.65	86.07
52	74.67	86.45
315	73.82	74.43
271	72.35	73.5
275	73.57	85.72
201	73.75	85.7
209	73.93	85.68
213	73.53	74.13
228	74.15	85.9
242	74.25	85.87
262	73.72	84.72
219	73.53	73.87
263	73.65	85.82
293	77.78	
13	73.97	85.9
18	74.1	85.88
222	73.63	85.8
533	73.87	85.83
534	73.85	85.9
537	74	85.92
264	74.5	86.6
300	73.58	74.25
256	74.37	86
515	73.65	85.87
526	74.05	85.95
529	74.1	85.95
Peter	74.2	85.45
D13-003	75.52	86.02