Human mitochondrial DNA diversity in urban Cape Town

Master thesis in molecular anthropology

Vidar Kaspersen

Table of contents

Summary

This work arose from a collaborative project between the University of Western Cape and the University of Oslo, funded by the Research Council of Norway. The original project was set up to identify informative Y-chromosomal markers for use in forensic casework in South Africa, after it was found that the standard Y-chromosome markers used in crime cases were not suited to the ethnically varied South African population. Appropriate markers are necessary not only to help convict criminals, but to exclude innocent people who happen to match a suspect by chance. This research became vital in a country with a long history of racial discrimination and human right abuses. Another purpose of the original project was to train South African scientists in bone DNA typing, to help identify the remains of political activists murdered during the Apartheid period.

Very little research has been carried out on the genetic composition of complex urban populations. In this master project I used mitochondrial DNA (mtDNA) to explore the present-day genetic makeup in Cape Town, South Africa, in the context of historical information. My study group consisted of a group of students from the University of Western Cape.

MtDNA is maternally inherited, with little or no recombination. This permits maternal genetic lineages to be traced back to a most recent common ancestor (MRCA). MtDNA exhibits population-specific polymorphisms. The mtDNA control region is subject to high mutation rate, which allows closely related groups to be studied. The high mutation rate can be used to infer the phylogeographical changes since our MRCA.

In this study, a high level of genetic diversity was observed. In the 48 individuals tested, 48 unique sequences were observed, representing thirty different mtDNA groups with origins in Africa, Eurasia, Southeast Asia, and the Indian subcontinent.

1 Introduction

1.1 The present-day people of South Africa

The population of South Africa is highly diverse. It includes the Nguni (consisting of the Zulu, Xhosa, Ndebele and Swazi people), the Sotho-Tswana, who include the Southern, Northern and Western Sotho (Tswana people), Tsonga and Venda. These people belong to the Bantu-speaking groups in South Africa. The Afrikaners, English, so-called Coloureds, Indians, and other peoples immigrated to South Africa from other parts of Africa, and from Europe and Asia. Most of these groups have a cultural identity of their own. In addition, a few remaining members of the indigenous Khoikhoi and the San people also survive in South Africa.

Most of the present-day population are the descendants of Bantu-speakers, of people displaced by the slave trade in the seventh and eighteenth century, of Asian immigrants brought in as contract workers, and of voluntarily migrants of the 19th century. Only 9% of the population consider themselves as White, and are mostly descendants of the Dutch and English colonisers. A small number of French Huguenots and Germans also contributed to the present-day White population.

According to the 2001 democratic census, there were 44.8 million citizens living in South Africa. Of this, 79% regarded themselves as Blacks, 9.6% as Whites, 8.9 as Coloured and 2.5% as Indian/Asian. The "Statistics South Africa" publishes midyear population estimates. By mid 2006, the total population in South Africa had increased to 47.4 million people (Statistics-South-Africa. 2006).

South Africa is divided into nine provinces: Western Cape, the Eastern Cape, KwaZulu-Natal, the Northern Cape, Free State, North West, Gauteng, Mpumalanga and Limpopo (Northern Transvaal) (Figure 1).

Figure 1: Overview of the borders and provinces of South Africa.
(Map used with permission from University of Texas library).

1.2 The prehistory of South Africa

Some of the oldest fossil evidence of anatomical modern humans (AMH) is found in South
Africa, dated back to more than 100,000 YBP (years before present) (Deacon 1992; Cavalli-
Sforza et al. 1994; Miller et al. 1999). Some studies indicate AMH dates reaching back to
200,000 YBP (Clark 1992), and in the Blombos cave near the southernmost tip of South
Africa there are cave paintings and red ochre decorations that provide evidence of
sophisticated behaviour as early as 77,000 YBP (Henshilwood et al. 2002). These ancient
peoples were possibly the direct ancestors of the present-day San people, called Bushmen by
the Whites, and of the Khoikhoi people, also known as Hottentots (Vigilant et al. 1989;
Deacon 1992), small numbers of whom still survive.

Although the size of these population groups has decreased dramatically in the last couple of centuries, there are still small groups of San people who live traditionally as hunters, and Khoikhoi people who maintain their traditional pastoralist lifestyle. The Khoikhoi and the San people appear closely related genetically, and share common morphological characters such as steatophygia and tablier, meaning protruding behinds and extended labia (Jobling et al. 2004). They have also a close linguistic relationship, talking languages with very characteristic click sounds.

Although there are similarities between the San and the Khoikhoi, the Khoikhoi and the Bantu-speaking people appear to be genetically closer than the San and the Khoikhoi (Cavalli-Sforza et al. 1994). This is probably due to the extensive trade between the northern Bantu pastoralists and the San. This trade led to the introduction of cattle and sheep to South Africa about two-thousand years ago.

The Bantu-speaking pastoralists and agriculturalists possibly originated in central-west Africa, around the area of present-day Nigeria and Cameroon. They eventually expanded southwards along the area of present-day Mozambique and Zambia, and possibly also westwards towards Angola and Nigeria (Beleza et al. 2005) (Figure 2). They established permanent and semi-permanent settlements of agriculturalists and livestock holders. This form of society could support bigger population sizes and led to the concentration of political power. The power structures were strongly kin-based, allowing families who gained wealth in the form of cattle and land to exercise power over those less wealthy, and leading eventually to the formation of powerful kingdoms.

The Bantu people were the first people in South Africa to produce and utilize iron (Palmberg 1995). Metallurgy permitted the production of trade-goods such as tools, weapons and ornaments, and the trade of these goods was an important link in the creation of structured societies.

The demography of pre-colonial South Africa reflected the way of life of the different ethnic groups. The largest group was that of the Bantu-speaking agriculturalists and herdsmen, who lived in permanent settlements with access to sufficient food throughout the year to support a relatively large population. The Khoikhoi people were nomadic pastoralists and had only temporary settlements. This made them less able than the Bantu people to maintain large populations.

The San people lived as hunters and gatherers. They had only the goods they could carry, and were forced to move constantly to have access to animals for hunting, and roots and fruits for gathering. This meant that they lived in small groups.

This was the situation encountered by the first European colonizers who set foot on South African soil.
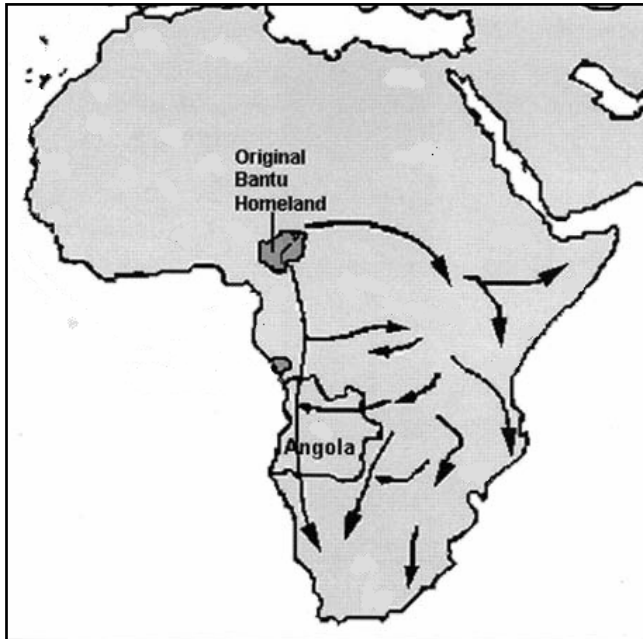
Figure 2: The eastern and western routes of the Bantu expansion from their homeland in present day Nigeria and Cameroon.
(Beleza et al. 2005)

1.3 Colonization

In 1652, the Dutch East India Company established a settlement to supply its ships with vegetables, meat, water and medical treatment (Iliffe 1995). The initial idea was to establish a supply station, and explicitly not a colony. As the need for supplies grew, the company enlarged their garden and gave concessions to some of their employees to be free settlers. This was the start of an escalating appropriation of lands already inhabited by indigenous people (Worden 1983), and marked the beginning of the colonization of South Africa.

The intensification of production led to an increased need for labour, satisfied by importing slaves. The Dutch East India Company's policy was not to enslave the indigenous people, as this was expected to create an untenable relationship with the local population. The main purpose of the settlement was to support shipping with water, food and other supplies, and to ensure this it was necessary to keep peace and avoid hostile confrontations.

Initially, most slaves were brought from the Dutch colonies in India and Indonesia, but in the 18th century Madagascar and Mozambique were the largest suppliers (Iliffe 1995). From the establishment of the Cape Colony in 1652 to the abolition of the slave trade in 1807, approximately 60,000 slaves were imported into the Colony (Ross 1999). In addition to those

brought by force, there was voluntary immigration from the Netherlands, France and Germany by people attracted by the promise of free land. However, most of these potential farmers migrated eventually to cities due to the low productivity and small size of the land they were given.

At the end of the 18th century, the English took over the Cape Colony. Around 5,000 British immigrants came to South Africa to work as small farmers. Between 1860 and 1866, about 6,000 Indian contract workers where brought to the Colony to satisfy the need for labour. The introduction of labourers from the Indian sub-continent continued to 1911. In addition, a number of Indian people came to the Cape Colony as voluntary migrants. By the 1890s, the number of Indian immigrants exceeded that of Whites (Palmberg 1995). The proportion of Whites in South Africa reached its peak in 1936, at 21% of the population. This proportion was reduced steadily during the next decades to about 10% in 1999.

1.4 Apartheid

The Apartheid era started in 1948, when the National Party gained victory and consolidated its power. In the next few years, they managed through political manoeuvring to gain two thirds of the representatives in Parliament, and succeeded in changing the constitution and establishing laws of racial segregation. The basic premise was that White people, in this case the English-speaking and the White Afrikaner population, were superior to the so-called Black (African), Coloured and Indian "races". In the course of the following decades, the non-White sections of the population basically lost all their legal rights.

A series of security laws, and the prohibition of African political parties, ensured that the government held power over the people and organizations. The government took control of the African educational system, and non-Whites were forcefully moved to often-remote homelands, according to ethnic classifications based on physical features. These homelands were isolated areas with virtually no infrastructure, schools or health care.

In 1960, Africans and Coloureds lost the right to a White representative in Parliament (Thompson 2001), and thereby the last fragment of influence over their future. This happened at the time when other African countries were gradually being released from their foreign colonizers and gained political and national sovereignty.

The domination of a White minority consisting of 10% of the population had a devastating consequence on the South African economy. To keep its sovereignty, the country had to spend huge resources on administration, the police force and the army. Political and economic pressures eventually forced a change.

1.5 The post-Apartheid era

The Apartheid policy came to an end when Frederik Willem De Clerk came to power in 1989, first as leader of the National Party and then as Prime Minister after the former radical president Pieter Willem Botha suffered a stroke. African political parties were authorized and political prisoners were released, and after intense political negotiations and struggles to avoid civil war, Black and White politicians managed to establish a non-racial interim constitution in 1993. In 1994, Nelson Mandela was sworn in as president (Thompson 2001).

The post Apartheid government, beginning with the leadership of Nelson Mandela, up to the present-day president Thabo Mbeki, has faced a tremendous challenge to deal with the economic ruin inherited from the Apartheid era. They preside over an extremely violent society with an enormous crime rate, and a murder rate five times higher than in the USA. The difference between poor and rich people is comparable to that in underdeveloped countries, and health care is poor, at least in rural areas. Over 20% of the population is infected with HIV, and half a million people die of AIDS annually, leaving about 1,100,000 orphans. This results in enormous poverty and many young people turn to crime to survive. The educational system is short of resources, both in form of money and qualified teachers, and corruption is a significant problem in every sector.

Although international economic sanctions have been lifted, and membership in the British Commonwealth re-established, it has been difficult to attract foreign investors to the South African manufacturing industry, a crucial step to achieve economic growth. Nevertheless, there have been improvements. For example, the economy is improving slowly, the murder rate has declined mostly as a result of a reduction in political murders, and education is improving. Most importantly, the people of South Africa have equal legal rights, independent of their ethnic background.

Although the Apartheid system divided people into artificially defined groups, the South African people are a complex mixture of people of many geographical origins. The people of South Africa have coexisted for a very long time, and have many diverse genetic influences, which have changed constantly over time.

1.6 Genetic affinities of human populations

The classification of various population groups into different races, as done by the Apartheid regime, is not only discriminatory, but also conceptually wrong. The genetic variation in human populations as a whole is so small that it is meaningless to try to classify people into distinct "races". In fact, genetic diversity can be higher within a group than between geographically distant groups. This was already clear in the 1960s, when classical genetic

markers, such as blood group antibodies, human leukocyte antigens (Fadhlaoui-Zid et al.), and a variety of other proteins, were used to explore human evolutionary history (Cavalli-Sforza 1994). Nevertheless, human beings exhibit diversity, and both classical markers and the more recently developed DNA markers are useful tools to help understand the origins and migrations of human populations in different parts of the world.

Mitochondrial DNA (mtDNA) is frequently used in human evolutionary studies. MtDNA polymorphisms are among the first DNA polymorphisms to be used in human evolutionary studies (Denaro et al. 1981; Johnson et al. 1983). Considerable variation in the patterns of restriction enzyme polymorphisms was observed when comparing Khoisan, Bantu and other populations.

Mitochondrial DNA is found in the mitochondria, organelles present in eukaryotic cells. Mitochondria are responsible for cellular respiration, that produces the energy needed to perform cellular processes (Cavalli-Sforza et al.1994). The mitochondrial genome consists of 16569 base pairs (bp), forming 37 genes. Of these, 13 code for proteins involved in oxidative phosphorylation, 2 code for rRNAs and 22 for tRNAs (Jobling et al. 2004).

Each animal cell contains between a few hundred to several thousand mitochondria depending on the particular tissue's energy needs. Each mitochondrion contains two to ten copies of the mitochondrial genome (Jobling et al. 2004). The mitochondria are maternally inherited, and mtDNA undergoes little or no recombination. These characteristics, combined with a high rate of observed mutation, make mtDNA a useful tool in human evolutionary studies.

Different mtDNA variants (haplotypes) are observed in human populations around the world. The haplotypes have been shown to be continent specific to a large extent, and can be used to divide human populations into groups with distinct evolutionary histories (Figure 3).

The highest degree of genetic variation is found among African populations. The large levels of mtDNA variation in Africa has been taken as evidence of the Out of Africa hypothesis, that predicts a recent origin and dispersal of anatomically modern human from Africa, spreading to the rest of the world, as illustrated in Figure 4.
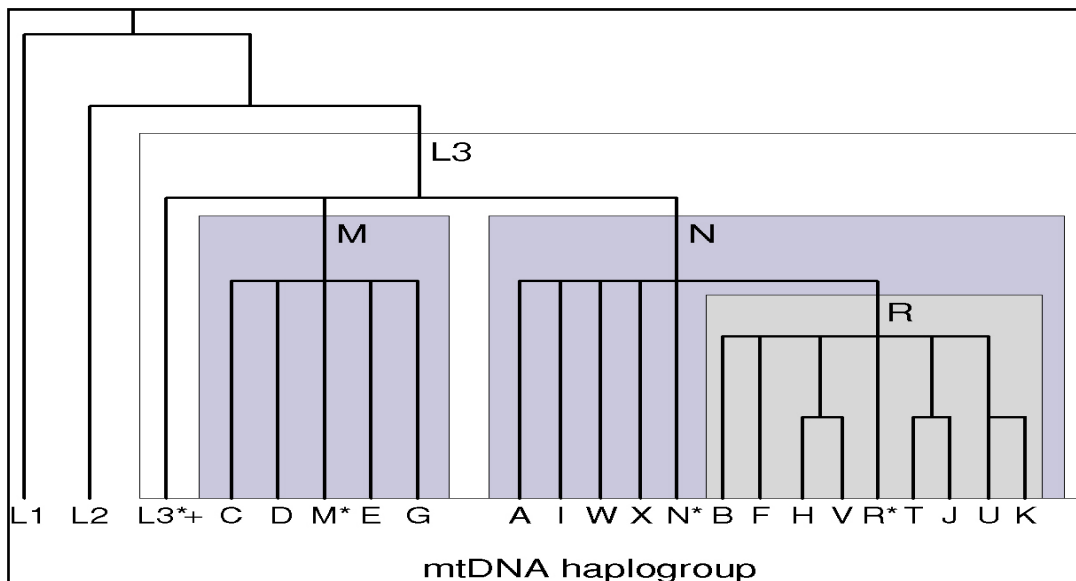
Figure 3: World mtDNA haplogroup tree. Macrohaplogroups L are African, and M and N are the out of Africa haplogroups derived from the African L3 branch.
Reproduced from Jobling et al. 2004.
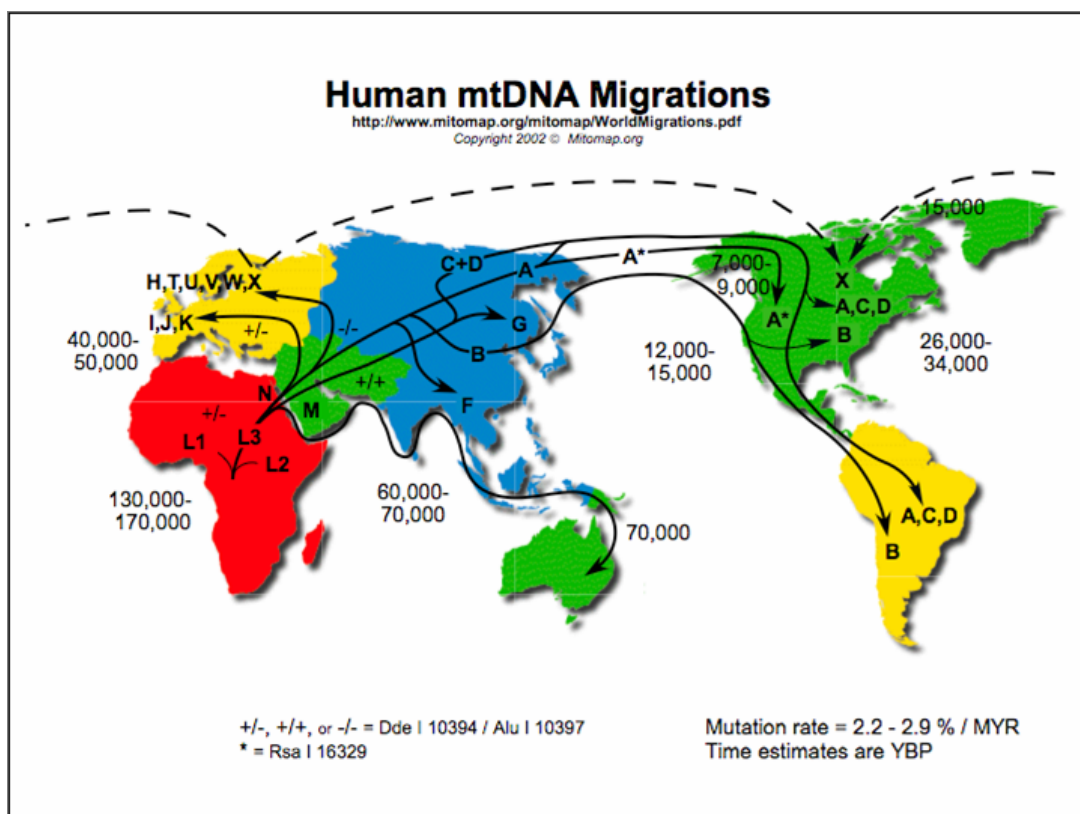
 Figure 4: Mitochondrial DNA haplogroups and migration routes. African haplogroups are denoted with the letter L. The L3 lineage migrated out of African and split into the subgroups M and N. In Europe, most mtDNA types derive from N, while in Asia there are many subgroups of M and N. (Map from http://www.mitomap.org/)

2. Materials and methods

2.1 DNA samples

The forty-eight human DNA samples used in this study where collected at the University of Western Cape, South Africa, and kindly donated by Prof. Sean Davison and Dr. Neil Leat. The DNA was from epithelial cells collected with buccal swabs, and was extracted in Cape Town using a standard protocol with Chelating Resin (Sigma). The samples were named CT1 to CT48, and diluted 1:10 to provide a template suitable for DNA analysis.

2.2 General amplification and sequencing approach

The DNA samples were analyzed using the techniques of polymerase chain reaction (PCR), restriction enzyme digestion, and DNA sequencing. The PCR amplifications were done in 50µl reactions using AmpliTaq Gold Taq polymerase (Applied Biosystems). A master-mix was prepared for the number of samples processed at each run, plus one as a negative PCR blank. In brief, 48 µl of the master mix were added to each PCR tube, and 2µl of the DNA to a final volume of 50µl. Each individual reaction contained 5µl 10x PCR buffer, 5µl 25mM $MgCl_2$, 0.5µl 10µM dNTPs, 0.8µl 5U/µl Taq polymerase, 1.5µl 10µM forward and reverse primer, 33.7µl dH_2O, and 2µl DNA . The PCRs were ran as follows: Initial denaturation at 94^0C for 5min, followed by 30 cycles of denaturation at 94^0C for 30s, primer annealing at 55^0C for 30s, elongation at 72^0C for 30s, a final step elongation at 72^0C for 7min. The amplifications were done in a Gene Amp PCR System 9700 Thermocycler.

2.3 MtDNA hypervariable region 1

All the primers used for the amplification and sequencing reactions are listed in Table 1, which shows the mtDNA locus in question, the primer sequences and the originator of the respective primers.

The hypervariable region1 (HVRI) of the mitochondrial D-loop was amplified twice. The initial PCR experiments were performed at the Department of Biology, University of Oslo, using primers D1 (L15996) (forward), and B1 (H16401) (reverse). This generated a segment of 445 bp from position 15975 to 16420 of the mtDNA reference sequence (Anderson et al. 1981). The HVRI segment was amplified again at the Nevrologisk Forskningslaboratorie Haukeland Sykehus in Bergen, with primers D1F (forward) and D1R (reverse) that yielded a 667 bp fragment spanning from mtDNA position 15879 to 16545 (667bp).

2.4 MtDNA hypervariable region 2

A fragment of 446 bp of the mtDNA hypervariable region 2 (HVRII), from position 16495 to 389 of the reference sequence was amplified using primers D2F (forward) and D2R (reverse).

2.5 COII / tRNA^{Lys} intergenic region / Region V

A fragment of 120bp spanning from 8196 to 8316 of the reference sequence was amplified with primers A (forward) and B (reverse) as described previously (Wrischnik et al. 1987). The presence or absence of the 9-bp deletion was verified by agarose gel electrophoresis as described below.

A 690 bp fragment from position 7930 to 8620 of the mtDNA reference sequence (Anderson et al. 1981) and spanning the COII / tRNA^{Lys} intergenic region was amplified using primers 13F (forward) and 13R (reverse). The presence or absence of the 9-bp deletion was verified by sequencing with primers M13F (forward) and M13R (reverse).

These amplified fragments contain a tandem repeat of the 9bp sequence CCCCCTCA in the mtDNA reference sequence. One of these repeats is frequently deleted in individuals of Asia and sub-Saharan Africa, and the deletion is fixed (present at a frequency near or approaching 100%) in Polynesians.

The PCR products of the mtDNA region V fragment (120bp) were ran on a 4% gel containing 1% agarose and 3% NuSieve as described previously (Hagelberg et al. 1993). An aliquot of 10µl DNA and 3µl loading dye (Promega) was used in each case, together with 2µl 100bp DNA ladder (Promega) as a size reference.

2.6 Gel electrophoresis

To check the yield and the size of the HVRI, HVRII and the 690bp COII / tRNA^{Lys} intergenic region PCR products, 3µl of the PCR products were run on 1% agarose gels, together with 1µl 6x Promega Load Dye. The gels contained 50µl EtBr (0.25mg/ml) and were made in 1XTBE buffer. The gels were run at 80V for 45min, together with a 100bp DNA ladder (Promega). After running the gels, the PCR products were visualized under UV light.

2.7 Cleaning of PCR products

The PCR products amplified from HVRI, HVRII and the 690bp COII / tRNA^{Lys} intergenic region were cleaned prior to sequencing using Millipore Montage^{TM} PCR centrifugal Filter Devices. The cleaned products were run on a 1% agarose gel as described above (section 2.6).

2.8 DNA sequencing

The sequencing reactions were performed in a Gene Amp PCR System 9700 Thermocycler, in 10µl reactions, using the "BigDye Terminator v3.1 Cycle Sequencing Kit" from Applied Biosystems. Individual reactions contained 1µl 5x sequencing buffer, 1µl Big-Dye 3:1 enzyme, 3µl 2.5µM primer, 2µl DNA, and 3µl dH_2O. The primers used for sequencing were M13F and M13R (Table 1). The sequencing reaction cycle was as follows: Initial denaturation at 96^0C for 5min, followed by 25 cycles of denaturation at 96^0C for 10s, primer annealing at 50^0C for 5s and elongation at 60^0C for 4min. Ten µl dH_2O was added to each tube to a final volume of 20µl before sequencing.

The sequencing was carried out at the Høyteknologi senteret i Bergen. The sequence reaction products were cleaned using the Montage Sequencing Cleanup Kit from Millipore, with a BIOMEK 2000 pipetting robot. The sequencing reactions were done in an Applied Biosystems 3730xI DNA Analyzer.

2.9 Analysis of sequence data

The HVRI and HVRII sequences were to be assigned to known mtDNA haplogroups. This is a laborious process that requires detailed manipulation of the sequencing files. The initial processing of the sequences was performed with the Staden software package (Bonfield 2004).

The first step involved the program Pregap4. In this program, the AB1 trace files were imported and renamed to standardise the nomenclature for further treatment using other programs. The main purpose of this step was to check the quality of the sequences and sort out ambiguities. Forward and reverse sequences were compared to each other.

The second step was carried out with the program Gap4 from the Staden software package. The reverse sequences were reverse complemented and manually aligned with the forward sequences. They were aligned with the revised Cambridge reference sequence (rCRS) (Andrews et al. 1999). The sequences were manually scanned for errors and ambiguities and corrected if a reliable notation was possible. The corrections were done by comparison with the complementary strand and the trace files. Discrepancies from the rCRS were marked as mutations and/or ambiguities according to universal practice. Ambiguous single stranded ends were trimmed.

The Staden files were exported to a Pearl script created by Ian Frame (personal communication). The files were renamed and converted to FASTA format. The files were concatenated to a single file including the part of the reference sequence that covered the area of the sequences of interest.

The file created in the Pearl script was aligned using the ClustalW alignment tool and opened in the program MacClade. Here, the sequences were processed to a format with that permitted the visualization of the sites variable with respect to the reference sequence.

To assign the individual sequences to haplogroups, sequences were compared to a database created by Ian Frame, and to previously published data.

2.10 Restriction enzyme analysis

RFLP (Restriction Fragment Length Polymorphism) analysis was performed to distinguish between the two major mtDNA haplogroups M and N, which are derived from the African macrohaplogroup L3 (Hill et al. 2007). A mtDNA fragment of 654 bp spanning from 9842 to 10496 of the reference sequence, contains the cutting sites 5'-AG^CT-3' for restriction enzyme AluI, and 5'-C^TNAG-3' for DdeI. The fragments were amplified using primers 16F (forward) and 16R (reverse), shown in Table1

Digestion of the 654bp fragment with DdeI was performed by mixing 8µl PCR product, 1µl Buffer D10x (Promega) and 1µl DdeI 10u/µl restriction enzyme (Promega) and the reaction was incubated at 37^0 for two hours. To terminate the digestion process, the reactions were incubated at 60^0C for 10min.

All the 10µl of digested PCR products were ran on a 2% NuSieve gel in a 1x TBE buffer together with a pGEM DNA Marker from Promega.

Digestion of the 654bp fragment with AluI restriction enzyme was performed with 8µl PCR product, 1µl Buffer 10xL (TaKaRa), 1µl AluI 8u/µl restriction enzyme (TaKaRa), and performed according to the same protocol as for DdeI.

Locus	Fragment size	Name	Forward primer (5'- 3')	Reverse primer (5'- 3')	Ref.
HVRI	667bp	D1F/D2R	AATGGGCCTGTTCCTTGTAG	AACGTGTGGGCTATTTAGGC	1
HVRI	445bp	D1/B1	CTCCACCATTAGCACCCAAAGC	TGATTTCACGGAGGATGGTG	2
HVRII	446bp	D2F/D2R	CGACATCTGGTTCCTACTTC	CTGGTTAGGCTGGTGTTAGG	1
COII/tRNA^{Lys}	690bp	13F/13R	ATTATTCCTAGAACCAGGCG	TGATGAGATATTTGGAGGTGG	1
mtDNA region V	120bp	A / B	ACAGTTTCATGCCCATCGTC	ATGCTAAGTTAGCTTTACAG	3
RFLP fragment	654bp	16F/16R	ACTTCACGTCATTATTGGCTC	AGTGAGATGGTAAATGCTAG	1
Sequencing primers		M13F/M13R	TGTAAAACGACGGCCAGT	CAGGAAACAGCTATGACC	1

Table1: Primers used for PCR and sequencing reactions

[1] L.A. Bindoff (personal communication)
[2] Vigilant et al.1989
[3] Wrischnik et al. 1987

3. Results

3.1 Summary of results

All 48 Cape Town individuals were typed successfully by RFLP analysis with the restriction enzymes AluI and DdeI to distinguish between mtDNA macrohaplogroups M and N. All 48 individuals were also typed for the intergenic 9 base pair (bp) deletion between the genes for COII and tRNA^{Lys}. Inside this intergenic area there is a tandem repeated motif of 9 bp. The deletion of one of these motifs is one of the features of haplogroup B.

All but one of the samples could be amplified successfully for the first hypervariable segment of the mtDNA noncoding region (HVRI), and 45 for the second hypervariable region (HVRII). Thus, all 48 samples yielded sequence information for at least one HVR segment.

Assignment of individuals to mtDNA haplogroups was done by comparison of HVRI and HVRII sequences to a database established by Ian Frame. Sequences not found in the database were manually compared to previously published data.

A high degree of variation was observed. Forty six different haplotypes were observed among the 48 individuals, which were assigned 30 different haplogroups originating in African, Eurasia, and South and Southeast Asia.

3.2 Hypervariable region I and II

The PCR amplification and sequencing of the first hypervariable region were performed twice. The first PCR and sequencing experiments were done at the Department of Biology, University of Oslo. The PCR and sequencing was repeated at Nevrologisk Forskningslaboratorie, Haukeland Sykehus in Bergen to improve the quality and reliability of the results.

Of the total 48 DNA samples, 47 yielded clear, unambiguous HVRI sequences and were assigned to haplogroups. The samples processed in Oslo yielded 445bp of double stranded sequence, from 15975 to 16420 of the mtDNA reference sequence (Anderson et al. 1981), and the samples amplified in Bergen yielded 667bp of double stranded sequences, from position 15879 to 16545.

One sample (CT31) failed to give a clear readable HVRI sequence. However, it gave an HVRII sequence identical to CT21 except for one variable site, and therefore probably belonged to haplogroup L1a1a. Nevertheless, as it is not common practice to assign haplogroups based solely on HVRII sequences, CT31 was not included in the final results.

Two pairs of samples yielded identical HVRI sequences, CT28 and CT46, and CT29 and CT48, respectively. However, these 4 individuals all had different HVRII sequences. Thus, based on HVRI and HVRII together, it is possible to state that each person in the study

group represented a unique mtDNA sequence, meaning a haplotype diversity of 1 (maximum).

The PCR amplification of HVRII gave 45 different sequences of double stranded DNA of typically 446bp in length, between nucleotides 16495 and 389 relative to the reference sequence. Sample CT7, CT8, and CT12 produced ambiguous HVRII sequences, but gave clear and unambiguous HVRI sequences which could be used for haplogroup assignment.

The HVRII sequences were not informative enough by themselves to assign the individuals to haplogroups. The HVRII sequences were therefore used exclusively to confirm haplogroups assigned by the HVRI sequences. This confirmation by HVRII was in cases where the HVRI sequences were not informative enough on their own, and were sometimes crucial to make an accurate assignment.

One sequence (CT44) carried a heteroplasmy at position 16189 in HVRI (Figure 5). Heteroplasmy is a condition where two variants of mitochondria are present in a cell. Some of the mitochondria can carry a specific mutation while others in the same cell do not carry the mutation.
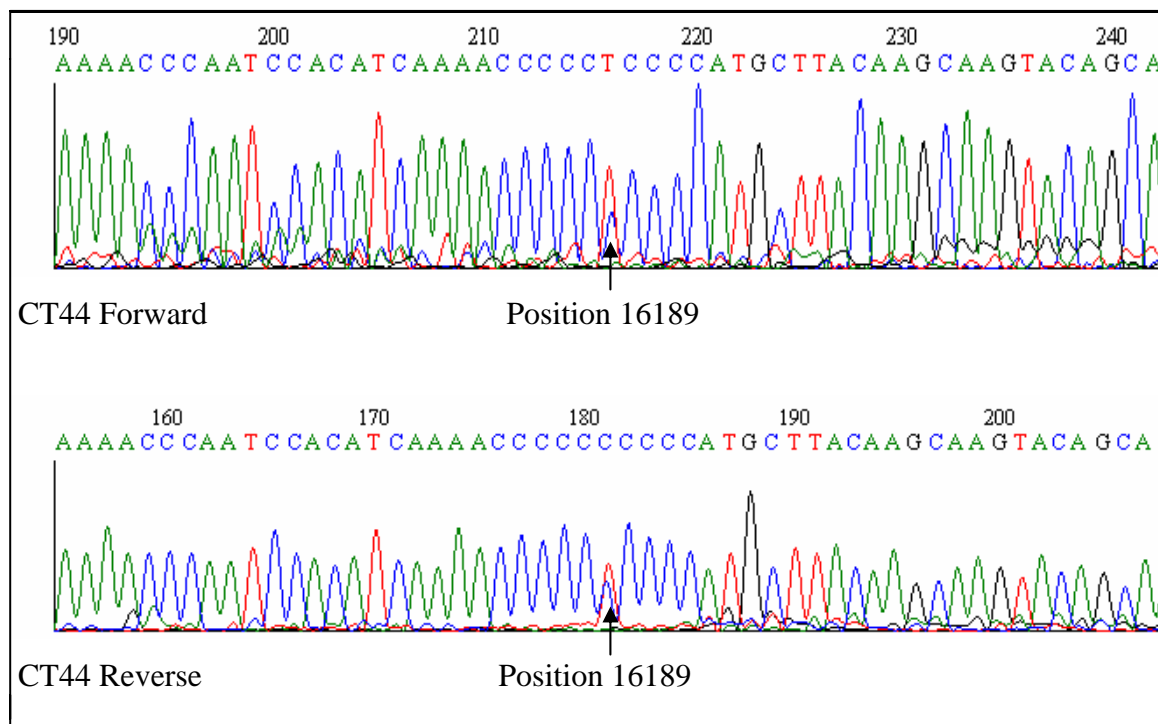
Figure 5: Heteroplasmy at position 16189 of the first mtDNA hypervariable segment (HVRI) in sample CT44. The reverse strand is reverse complemented.

3.3 RFLP analysis

The amplification of the fragments used for RFLP analysis yielded double stranded DNA sequences of 654bp for all the 48 samples, from mtDNA position 9842 to 10496 of the reference sequence.

Nine individuals carried the AluI 10400 C to G restriction site witch defines macrohaplogroup M, and seven individuals carried the DdeI 10398 A to G restriction site defining macrohaplogroup N (Hurles et al. 2005). Table 2 shows the individuals that belong to M (AluI+) and N (DdeI+).

| Enzyme	Sample numbers								
AluI +	CT10	CT12	CT15	CT27	CT28	CT34	CT36	CT45	CT46
DdeI +	CT11	CT19	CT33	CT37	CT38	CT40	CT42		

Table 2: Nine individuals had the AluI 10400 C to G restriction site characteristic of mtDNA macrohaplogroup M, and seven individuals carried the DdeI 10398 A to G restriction site defining the macrohaplogroup N. All others fitted into African L macrohaplogroups.

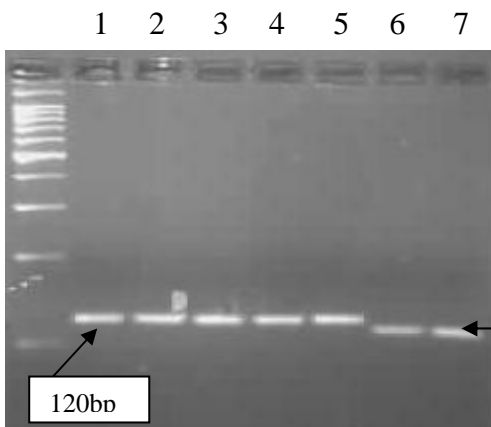## 3.4 COII / tRNA^{Lys} intergenic 9bp deletion

The 48 samples were first amplified using primers specific for a 690bp fragment between mtDNA position 7930 and 8620 of the reference sequence. All but one (CT9) were amplified successfully. This fragment contains a tandem repeat of the 9bp sequence CCCCCTCA in a noncoding area between the mtDNA genes for cytochrome oxidase II and lysine transfer RNA. A deletion of one of these 9bp repeats was originally thought to be diagnostic for haplogroup B. This deletion was first discovered in Asian populations and later at low frequencies in sub-Saharan African populations (Soodyall et al. 1996; Rosa et al. 2004).

The presence or absence of the deletion was verified by sequencing. Two of the samples, CT33 and CT40 carried the 9bp deletion. The HVR sequence of CT33 showed it had the full Polynesian motif defining haplogroup B4a, characterized by the 9bp deletion and base transitions at position 16189 (T to C), 16217 (T to C), 16247 (A to G) and 16361 (C to T) (Hagelberg et al. 1999) (Fig 6).

The 9bp deletion site was reamplified with primers that specified a120bp segment from nucleotide 8197 to 8216 to confirm the absence or presence of the deletion by gel electrophoresis (Figure 7). In this case, successful amplification was achieved for sample CT9, that had earlier failed to yield sequence. CT9 carried two copies of the 9bp motif (Figure 8).
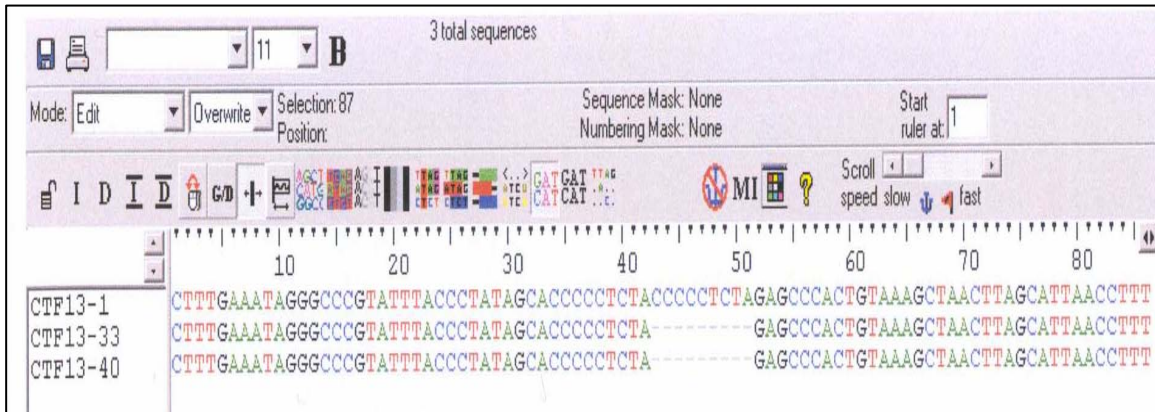
Figure 6: Nine base pair deletion in samples CT33 and CT40, compared with CT1 who had the 9bp tandem repeat.

Figure 7: The 120bp COII / tRNA^{Lys} intergenic fragment with the 9bp tandem repeat was visualized by gel-electrophoresis. Lanes 6 and 7 show the deletion in samples CT33 and CT40, while lanes 1-5 are the undeleted individuals CT10, 11, 12, 13, and 14.

120bp

111bp
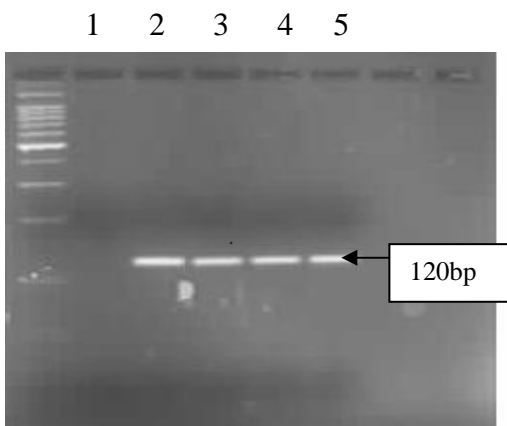
Figure 8: The 120bp COII / tRNA^{Lys} intergenic region with the 9bp tandem repeat. Lane 1 is the bank control and lanes 2-5 are individuals CT9, 10, 11, and 12.

120bp

3.5 Haplogroup assignment

Forty-five unique HVRI sequences were assigned to 30 different mtDNA haplogroups, which were subgroups of 9 different haplogroup lineages representing all three principal macrohaplogroups L, M and N, as shown in Figure 9. Figure 10 shows the world mtDNA tree with the arrows indicating the haplogroups detected in the present study, and their respective frequency in the study group.

African L lineage exhibited the highest frequency by far in the study group, as 32 (68%) individuals were from this macrohaplogroup. The individuals of L represented several L sub-groups, namely L0 (n=6,13%), L1 (n=9, 19%), L2 (n=6, 13%) and L3 (n=11, 23%). Macrohaplogroup M was represented in this study exclusively by Asian haplotypes, and counted for 15% of the group, consisting of 1 person of M*, 4 of D and 2 of E. Eight individuals (17%) belonged to macrohaplogroup N, of which 2 people were B, 1 person F and 1 person R, while 1 person was of the common European mtDNA type H and 3 people belonged to K.

Four individuals (who in all four cases defined themselves as "Coloured") in the study group had White maternal ancestors (Table 4), but belonged to the African haplotypes L0a, L0d and L1d. CT2 had a maternal grandmother who defined herself as White, CT29 had a maternal grandmother and mother defining themselves as White, CT16 had maternal grandmother defining herself as White, and CT13 had a maternal grandmother defining herself as White.

Four individuals were assigned to African haplogroup L2a1b. All four had the 10400 C to G AluI restriction site diagnostic for macrohaplogroup M. The inconsistent RFLP result in the coding region is due to homoplasy (parallel mutations), something that is extremely common in mtDNA, but the individuals clearly belonged to haplogroup L2a1b according to diagnostic variable sites in HVRI and HVRII.

Table 3: MtDNA polymorphisms in 47 individuals from Cape Town. Variable HVRI and HVRII sites, RFLPs with restriction enzyme AluI and DdeI, and presence or absence of the 9bp deletion. A letter after a variable site indicates a transversion to the base indicated. Variable sites are given by position -16000, except positions below 16000, given by the full number.

Sample	Haplogroup	Ref.	DdeI 10398	AluI 10400	9bpdel	HVRI Variable sites +16000	HVRII Variable sites
CT02	L0a	6	-	-	-	15941, 129, 187, 189, 230, 234, 243, 266A, 311	73, 146, 199, 247, 309.1, 315.1
CT07	L0a	6	-	-	-	129, 187, 189, 230, 234, 243, 266A, 294, 311	No sequence
CT29	L0a	6	-	-	-	129, 187, 189, 209, 230, 234, 243, 266A, 311	16491, 16493, 16519, 73, 146, 153, 195, 199, 247, 315.1, 391, 393
CT30	L0a	6	-	-	-	129, 187, 189, 230, 234, 243, 266A, 311	16491, 16519, 16524, 73, 146, 195, 199, 247, 309.1, 315.1,
CT48	L0a	6	-	-	-	129, 187, 189, 209, 230, 234, 243, 266A, 311	16493, 16519, 73, 146, 153, 195, 199, 247, 315.1, 391, 393, 394
CT16	L0d	9	-	-	-	129, 187, 189, 223, 230, 239, 243, 294, 311, 325	73, 146, 152, 195, 247, 309.1, 315.1
CT21	L1a1a	4	-	-	-	129, 148, 168, 172, 187, 188G, 189, 223, 230, 278, 293, 311, 320	93, 95, 185, 189, 236, 247, 263, 309.1, 315.1
CT24	L1c1	4	-	-	-	017, 129, 163, 187, 189, 209, 223, 278, 293, 294, 311, 360	16519, 73, 151, 152, 182, 186, 189, 195, 247, 263, 315.1, 316
CT32	L1c3	4	-	-	-	15978, 129, 182C, 183C, 189, 215, 223, 278, 294, 311, 360	16519, 73, 151, 152, 182, 186, 189, 247, 263, 309.1, 315.1, 309.2, 316
CT04	L1d	4	-	-	-	129, 187, 189, 212, 223, 230, 243, 311, 390	73, 146, 152, 195, 198, 247, 309.1, 315.1,
CT13	L1d	4	-	-	-	187, 189, 223, 230, 243, 266, 274, 290, 300, 311	73, 146, 150, 195, 247, 315.1, 316, 391, 393
CT17	L1d	8	-	-	-	129, 187, 189, 218, 223, 239, 243, 294, 311	73, 146, 152, 195, 247, 315.1
CT43	L1d	4	-	-	-	187, 189, 223, 230, 243, 274, 278, 290, 300, 311	16519, 73, 146, 150, 195, 247, 309.1, 315.1, 316, 391, 393, 394
CT05	L1d1	8	-	-	-	129, 187, 189, 218, 223, 239, 243, 294, 311	16519, 73, 146, 152, 195, 247, 315.1
CT14	L1d1	4	-	-	-	129, 187, 189, 223, 230, 239, 243, 294, 311	73, 146, 152, 195, 247, 315.1

CT47	L2a1	4	-	-	-	189, 223, 278, 294, 309, 390	16491, 16493, 16519, 73, 146, 152, 196, 263, 309.1, 309.2, 315.1
CT23	L2a1a	8	-	-	-	223, 278, 286, 294, 309, 390	16519, 73, 146, 152, 195, 263, 309.1, 315.1, 391
CT10	L2a1b	4	-	+	-	182C, 183C, 189, 223, 278, 290, 294, 309	73, 146, 152, 195, 263, 315.1, 391, 393
CT15	L2a1b	4	-	+	-	182C, 183C, 189, 223, 278, 290, 294, 300, 309,390	73, 146, 152, 195, 263, 309.1, 315.1
CT28	L2a1b	4	-	+	-	182C, 183C, 189, 223, 278, 290, 294, 309, 390	73, 146, 152, 195, 263, 309.1, 309.2, 315.1
CT46	L2a1b	4	-	+	-	182C, 183C, 189, 223, 278, 290, 294, 309, 390	73, 146, 152, 195, 199, 263, 309.1, 309.2, 315.1
CT20	L3d	8	-	-	-	124, 215, 223, 319	16493, 73, 150, 152, 263, 309.1, 315.1, 391
CT01	L3d1	4	-	-	-	124, 223, 319	73, 150, 152, 263, 309.1, 315.1, 391, 393
CT25	L3d1	2	-	-	-	026Y,124, ,215, 223, 319	16493, 73, 150, 152, 263, 309.1, 315.1, 391
CT03	L3e1	4	-	-	-	223, 327	73, 150, 189, 200, 214, 263, 315.1
CT26	L3e1	3	-	-	-	176, 223, 234, 287, 291, 327	16493, 73, 150, 152, 189, 200, 263, 315.1, 391, 393
CT09	L3e1a	4	-	-	-	185, 223, 311, 327	73, 150, 185, 189, 263, 315.1, 391
CT35	L3e1a	4	-	-	-	15942, 185, 223, 311, 327	73, 150, 185, 189, 263, 315.1, 391, 393
CT44	L3e1b	8	-	-	-	026Y, 223, 239, 384R	73, 150, 185, 189, 263, 309.1, 315.1, 391
CT06	L3e2b	4	-	-	-	172, 183C, 189, 223, 320	73, 150, 195, 263, 315.1, 391
CT08	L3e3	4/8	-	-	-	223, 265T	No sequence
CT18	L3f	1/4	-	-	-	129, 209, 223, 292, 295, 311	73, 189, 200, 263, 272, 309.1, 315.1, 391
CT36	D	4	-	+	-	182C, 183C, 189, 192, 223, 278, 290, 294, 309, 390	16491, 16493, 73, 146, 152, 195, 263, 315.1, 391, 393, 394
CT39	D4	4	-	-	-	189, 223, 256, 274, 294, 309, 344, 390	16519, 73, 146, 152, 195, 263, 309.1, 315.1

CT41	D4	4	-	-	-	129, 187, 189, 218, 223, 239, 243, 294, 311, 354	16491, 16493, 16519, 73, 146, 152, 195, 247, 315.1, 391, 393, 294, 395
CT27	D5	5	-	+	-	189, 223, 311, 362	73, 152, 263, 309.1, 315.1
CT34	E1a	4	-	+	-	223, 291, 262, 390	16519, 173, 263, 309.1, 315.1, 391
CT45	E1a	4	-	+	-	223, 263, 291, 362, 390	16491, 16493, 16519, 73, 263,309.1 , 315.1, 391,393, 394, 395
CT12	M46	5	-	+	-	172, 173, 223, 278, 311	No sequence
CT33	B4a	4	+	-	+	182C, 183C, 189, 217, 247, 261	16519, 73, 146, 263, 309.1, 309.2, 315.1
CT40	B4b	4	+	-	+	093, 136, 183C, 189, 217	16519, 73, 207, 263, 309.1, 315.1, 391, 393, 394, 395
CT38	F1b	4	+	-	-	093, 189, 220C, 265, 298, 311, 362	16510, 73, 150, 152, 263, 315.1, 391, 393, 394
CT11	H7	4	+	-	-	15996N, 019Y, 021B,189, 384R	16519, 73, 195, 263, 315.1,
CT22	R1	4	-	-	-	266, 297, 304, 311, 355, 356	16524, 73, 152, 263, 309.1, 315.1, 391
CT19	K	7	+	-	-	167, 261, 256	73, 152, 195, 263, 315.1, 391
CT42	K	7	+	-	-	111, 144, 223, 261, 311,	73, 228, 234, 263, 315.1, 315.2, 391, 393, 394
CT37	K2a	4	+	-	-	224, 311	16491, 16493, 16519, 73, 146, 152, 263, 315.1, 391

[1] Allard et al. 2005
[2] Abu-Amero et al. 2007
[3] Beleza et al. 2005
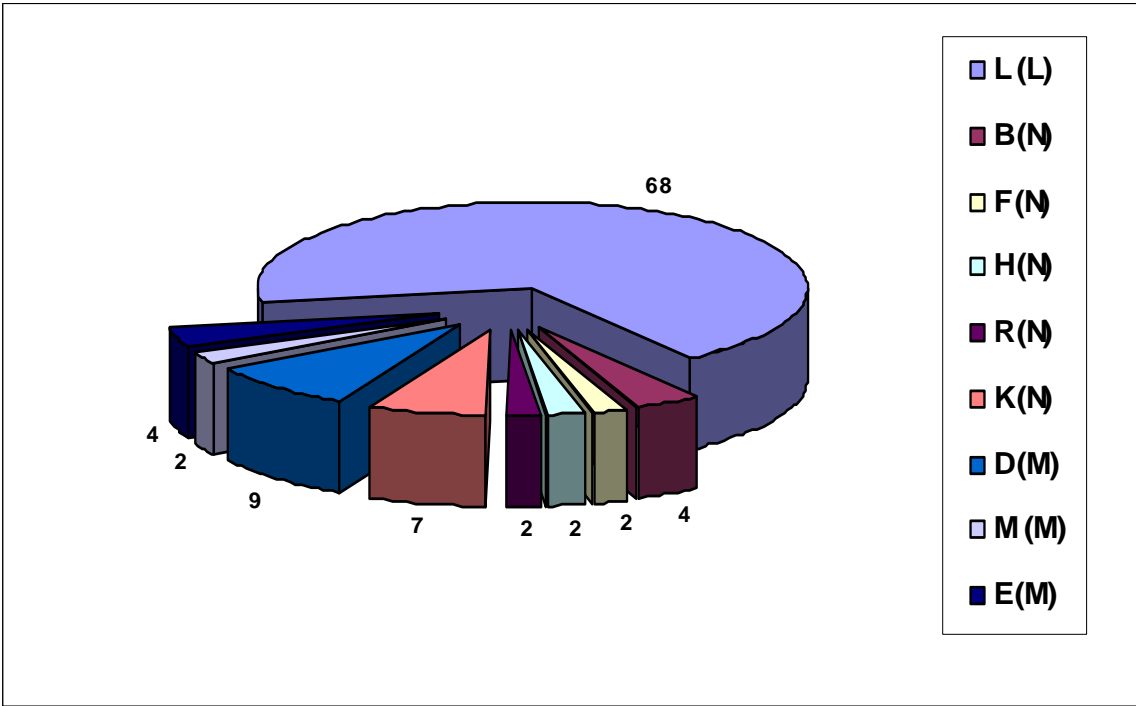[4] Ian Frame (personal communication)
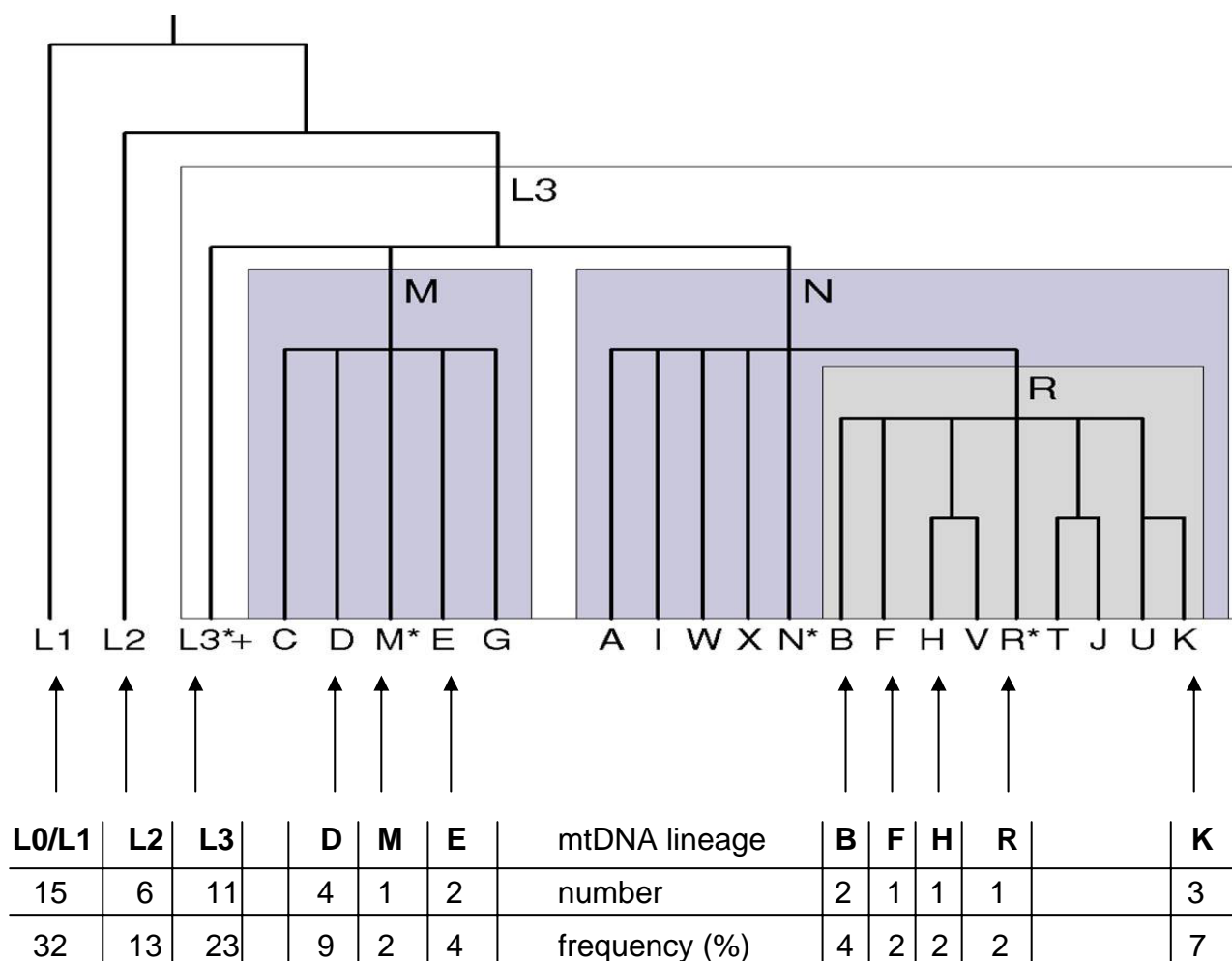[5] Hill et al. 2007
[6] Kivisild et al 2004
[7] Richards et al. 2000
[8] Salas et al. 2002
[9] Salas et al. 2004

Figure 9: Distribution of haplogroup lineages in percent among the individuals of the study group. Macrohaplogroups in parenthesis.

L0/L1	L2	L3		D	M	E	mtDNA lineage	B	F	H	R		K
15	6	11		4	1	2	number	2	1	1	1		3
32	13	23		9	2	4	frequency (%)	4	2	2	2		7

Figure 10: World tree of the mtDNA macrohaplogroups and principal mtDNA haplogroups from Jobling et al. 2004. In the Cape Town samples, there were representatives of macrohaplogroups L, M, N, and the subgroups of R. The number of individuals and frequencies of each study group are listed under each principal mtDNA haplogroup.

Table 4. List of the individuals included in this study, with their place of birth, and their own and their close family's ethnic affiliation, according to their own definition.

Individuals	Province	Ethnic affiliation				
		Mothers Mother	Mother	Own	Father	Fathers Father
CT1	Tanzania	Black	Black	Black	Black	Black
CT2	West Cape	White	Coloured	Coloured	Coloured	White
CT3	West Cape	Coloured	Coloured	Coloured	Coloured	Coloured
CT4	West Cape	Coloured	Coloured	Coloured	Coloured	Coloured
CT5	Eastern Cape	Black	Black	Black	Black	Black
CT6	Zambia	Black	Black	Black	Black	Black
CT7	West Cape	Coloured	Coloured	Coloured	Coloured	Coloured
CT8	West Cape	Coloured	Coloured	Coloured	Coloured	Coloured
CT9	West Cape	Coloured	Coloured	Coloured	Coloured	Coloured
CT10	Eastern Cape	Black	Black	Black	Black	Black
CT11	West Cape	Coloured	Coloured	Coloured	Coloured	Coloured
CT12	West Cape	Coloured	Coloured	Coloured	Coloured	Coloured
CT13	West Cape	White	Coloured	Coloured	Coloured	Coloured
CT14	West Cape	Coloured	Coloured	Coloured	Coloured	Coloured
CT15	Kwazulu Natal	Coloured	Coloured	Coloured	Coloured	Coloured
CT16	West Cape	White	Coloured	Coloured	Coloured	White
CT17	West Cape	Coloured	Coloured	Coloured	Coloured	Coloured
CT18	Namibia	Black	Black	Black	Black	Black
CT19	West Cape	Coloured	Coloured	Coloured	Coloured	Coloured
CT20	West Cape	?	Coloured	Coloured	Coloured	Coloured
CT21	Limpopo	Black	Black	Black	Black	Black
CT22	West Cape	Coloured	Coloured	Coloured	Black	
CT23	West Cape	Coloured	Coloured	Coloured	Coloured	Coloured
CT24	West Cape	Coloured	Coloured	Coloured	Coloured	Coloured
CT25	West Cape	Coloured	Coloured	Coloured	White	White
CT26	Malawi	Black	Black	Black	Black	Black
CT27	West Cape	Coloured	Coloured	Coloured	Coloured	Coloured
CT28	West Cape	Coloured	Coloured	Coloured	Coloured	Coloured
CT29	West Cape	White	White	Coloured	Coloured	Coloured
CT30	Eastern Cape	Black	Black	Black	Black	Black
CT31	West Cape	Black	Black	Black	Black	Black
CT32	West Cape	Coloured	Coloured	Coloured	Coloured	White
CT33	West Cape	Coloured	Coloured	Coloured	Coloured	Coloured
CT34	Saudi Arabia	Coloured	Coloured	Coloured	Coloured	Coloured
CT35	West Cape	Coloured	Coloured	Coloured	Coloured	Coloured
CT36	West Cape	Coloured	Coloured	Coloured	Coloured	Coloured
CT37	West Cape	Coloured	Coloured	Coloured	Coloured	Coloured
CT38	West Cape	Coloured	Coloured	Coloured	Coloured	Asian Indi
CT39	West Cape	Coloured	Coloured	Coloured	Coloured	Coloured
CT40	West Cape	Coloured	Coloured	Coloured	Coloured	Coloured
CT41	West Cape	Coloured	Coloured	Coloured	Coloured	Coloured
CT42	West Cape	White	White	White	White	White
CT43	West Cape	Cape Malay	Cape Malay	Cape Malay	Cape Malay	Cape Malay
CT44	Eastern Cape	Black	Black	Black	Black	Black
CT45	West Cape	Asian Indian	Asian Indian	Asian Indian	Asian Indian	Asian Indian
CT46	Botswana	Black	Black	Black	Black	Black
CT47	West Cape	Coloured	Coloured	Coloured	Coloured	Coloured
CT48	West Cape	Coloured	Coloured	Coloured	Coloured	Coloured

4. Discussion

The vast majority of molecular anthropology studies worldwide have focused on so-called indigenous populations. During the 18th and 19th centuries, scholars classified human groups using physical traits and cranial measurements. In the 20th century, many studies were done on human evolution and migrations, and blood groups and other classical genetic markers were widely used to study the origins of indigenous groups. From the 1980s onwards, molecular genetic markers have been used to explore the genetic affinities of populations. In 1994, Cavalli-Sforza, Paolo Menozzi and Alberto Piazza published "The History and Geography of Human Genes" which is a huge compilation of data generated in the second part of the 20th century (Cavalli-Sforza et al. 1994), including a wealth of data on South Africa. More recently, several research groups have published data on African and South African so-called indigenous populations (Hitzeroth 1984; Vigilant et al. 1989; Deacon 1992; Klein 2001; Salas et al. 2002; Beleza et al. 2005; Ely et al. 2006). However, with the exception perhaps of an analysis of uniparental and biparental markers in an urban population from Buenos Aires (Marignac et al. 2004), little attempt has been made to understand the genetic composition of urban populations.

In the present study, about two thirds of the urban Cape Town group was found to belong to various subgroups of the major African macrohaplogroup L. To define the geographical affinity of the various African L haplotypes, except L0, extensive reference was made to the 2002 study by Antonio Salas and colleagues (Salas et al. 2002). L0 has been described recently, or, more precisely, split from L1 where it was formerly included. The L0 branch of the mitochondrial tree was discussed by Spencer Wells in a recent (24 May 2007) presentation on the worldwide Genographic Project. Wells explained that L0 represents the deepest lineage of the human mitochondrial tree in Africa and defines the line leading to Mitochondrial Eve, the coalescent of the mtDNAs of all people living today (S. Wells and others, unpublished observations).

Six of the 48 individuals were found to belong to L0. They were CT2, CT7, CT29, CT30 and CT48, assigned to haplogroup L0a, and CT16 assigned to haplogroup L0d. These mtDNA types were previously observed at high frequencies in Mozambique and Kenya, eastern Africa, the place of origin of all humans (Rosa et al. 2004). If L0 represents the oldest mtDNA lineages in the world, it makes sense that they have survived at comparatively high frequencies in the southernmost part of Africa, where they were gradually "pushed" by expanding waves of Bantu agriculturalists and other later migrants. The L1a type, observed in one individual, CT21 (L1a1a), is the most common mtDNA type in southeast Africa, probably introduced by the Eastern Bantu migration. According to Salas and colleagues

(Salas et al. 2002), the lineage might have originated in the western parts of Central Africa, close to the place of origin of Bantu languages, or along the western route of the Bantu expansion.

Two individuals belonged to the L1c type, CT24 (L1c1) and CT32 (L1c3). The suggested origin of this haplotype is the region of Central Africa toward the Atlantic coast, along the western route of the Bantu migration (Salas et al. 2002; Beleza et al. 2005). Six individuals belong to haplogroup L1d, namely CT4, CT13, CT17 and CT43 (L1d), and CT5 and CT14 (L1d1). L1d is present at frequencies over 50% in the Khoisan people of southern Africa. This haplotype is among the oldest in Africa, and is probably indicative of the ancestors of the early San. This haplotype is also found in 5% of the Bantu-speakers (Salas et al. 2002).

The L2a lineage carried by CT47 (L2a1), CT23 (L2a1a) and CT10, CT15, CT28, and CT46 (L2a1b) is thought to have originated in western Africa. It is also a Bantu lineage, and it is the most widespread and common mtDNA haplotype in Africa, with 25% of the indigenous Africans belonging to this haplotype (Salas et al. 2002).

The L3 haplogroup reaches its highest frequency in East Africa, and probably is of East African origin. Eleven individuals were assigned to this lineage. L3d was carried by CT20 (L3d), and CT1 and CT25 (L3d1). L3d is a branch of L3 branch that expanded from East to West Africa, and is also found at significant frequencies in south-eastern parts of Africa (Salas et al. 2002). Seven individuals were assigned to haplotype L3e (CT3 and CT26 (L3e1), CT9 and CT35 (L3e1a), CT44 (L3e1b), CT6 (L3e2b) and CT8 (L3e3)). This haplotype is the most frequent and ancient of the L3 branches. It appears in 30% of sub-Saharan L3 haplotypes, and is thought to have originated in Central Africa (Salas et al. 2002). L3f is a rare haplotype, found only in one individual of our group, CT8. This haplotype originated in East Africa, were it is found almost exclusively today, with the exception of a limited radiation to south-eastern Africa (Salas et al. 2002).

Concerning the out of Africa macrohaplogroup M, the individuals in this study represented haplogroups M46, E1a, D, D4 and D5. The M46 haplogroup was found in one individual, CT12. M46 is found in Island Southeast Asia (ISEA), the Malay Peninsula and South China. It has also been observed in Burma (Hill et al. 2007). Two individuals carried haplogroup E1a, CT34 and CT45. Like M46, E1a is common in ISEA and among Taiwanese aboriginals, but seems rare in China (Hill et al. 2007). Haplogroup D and D4 were carried by three individuals, namely CT36 (D), CT39 and CT41 (D4). These lines are present in Han Chinese and Koreans (Thangaraj et al. 2003; Allard et al. 2004). One individual, CT27,

carried haplotype D5. This haplogroup is found at low frequency in ISEA. In Sulawesi, D5 is present at frequencies above 10% and the haplotype may have originated in mainland China (Hill et al. 2007)

The N haplogroups in this study include B4a, B4b, F1b, H7, K, K2a and R1. Haplogroup B4 is common in Southeast Asia. One individual (CT33) belonged to haplogroup B4a (CT33). B4a is common in ISEA is found at high frequencies in Taiwanese aboriginals and Filipinos. This haplogroup includes the full "Polynesian motif" carried by CT33, which is present at extremely high frequencies in Polynesia, reaching fixation in some islands (Hill et al. 2007). Interestingly, the Polynesian motif is found at frequencies around 20% in Malagasy (Soodyall et al. 1995). The occurrence of the "Polynesian motif" in urban Cape Town is a fascinating observation, and might be the result of the transportation of Malagasy slaves into the Cape Colony during the 18th century. Conversely, the "Polynesian motif" might have been brought into Cape Town in more recent times by a migrant labourer from the Indonesian Archipelago, where the Polynesian motif is observed at low frequencies.

CT40 belongs to haplogroup B4b which is most common in China and Thailand (Metspalu et al. 2004). One individual carried F1b (CT38), a type quite common in China and India (Metspalu et al. 2004). The H7 lineage was observed in one individual (CT11) and is a typically European linage, abundant also in Pakistan and Iran (Metspalu et al. 2004; Wiesbauer et al. 2006). Two individuals (CT19) and (CT42) belonged to the K haplogroup that is widespread in eastern and western Europe (Feder et al. 2007). Haplogroup R, carried by one individual, CT22 (R1), is found at significant frequencies in India, Pakistan and Bangladesh (Metspalu et al. 2004).

This study, albeit on one single genetic locus, mtDNA, demonstrates the very elevated levels of genetic variation among the people living in urban Cape Town today. Of the 47 individuals that could be assigned, 30 different haplogroups were observed. More than two thirds of the individuals of the study had African maternal lineages, while a quarter of the African haplotypes belonged to the deep African lineages L0 and L1. About a quarter of the sample could be traced to Asian source populations, including India, China, Korea, Taiwan and Island Southeast Asia, while less than a tenth of the group carried haplotypes typical of European or Eurasian populations.

The genetic diversity observed in our group of Cape Town students correlated well with the history of the province. The Bantu expansion caused a stream of peoples to migrate into the southern part of the African continent (Beleza et al. 2005). Part of the migration was directly from Central West Africa along a southward route into South Africa. Another route

was from West Africa along an easterly route, and then south into South Africa (Salas et al. 2002).

The East African lineage L3 represented about one quarter of the African mtDNA types in this study. Although a very large number of slaves were brought from East Africa into South Africa (Iliffe 1995), the majority of the slaves were men, who did not contribute their DNA to the gene pool of the extant population (Iliffe 1995; Ross 1999). Thus, because of this skewed representation of genders, the effective population size of the people displaced from East Africa was comparatively small, and would be expected to be less well represented in the present urban population than the ancient indigenous African lineages and the Bantu lineages. It is evident that historical events have blurred the genetic picture, and any conclusions drawn from this study must be tempered both by the small size of the sample, and by the fact that the sample was not random, but was a self-selected group of young people who reached university entrance, in a formerly "Coloured" University of Western Cape, and do not represent the demography of Cape Town as a whole.

Nevertheless, a possible explanation for the high frequency of L3 may be that the Bantu migrants acquired the L3 haplotype along the eastern migration route into South Africa. L3d, found in three individuals, has an East African origin, but has spread to West Africa where the source of the Bantu expansion is situated. This could also explain the comparatively low frequency of the L2 Bantu lineage, even though there are very many Bantu-speakers in South Africa (Cavalli-Sforza et al. 1994).

The high number of ancient African L0 and L1 types (15 individuals) may seem odd at first glance, as in pre-colonial times the indigenous San (Bushmen) and Khoikhoi (Hottentots) were exceeded greatly by the Bantu-speakers (Iliffe 1995). Curiously, four individuals with such ancient mtDNA types had mothers and/or maternal grandmothers who defined themselves as White (CT2, CT7, CT13, CT16 and CT29). The San people were used extensively as domestic servants by the White settlers. The predominantly male settlers undoubtedly fathered children with their female servants, and after several generations African mtDNA lineages would become widespread among the White population.

Nearly a quarter of the group had Asian maternal ancestors. Slaves were brought into South Africa from the Dutch colonies in India and Indonesia. Madagascar were also an important source of slaves in the 18th century (Iliffe 1995). Madagascar was originally settled by Malay traders (linguistic evidence links Malagasy languages with Borneo) and, as mentioned earlier, haplogroup B4a with the full "Polynesian motif" is present in the Indonesian archipelago. It is therefore possible that the person with the "Polynesian motif"

(CT33) is the descendant of a slave brought from Madagascar, where this haplotype is present at a frequency of about 20% (Soodyall et al. 1995; Hurles et al. 2005). After the abolition of the overseas slave trade in 1807, contract workers from Asia became the new source of labour. In addition, voluntary migrants from India and China came to find work in South Africa (Ross 1999).

The individuals with European haplogroups are probably the descendents of European colonizers or later settlers, but may also be the descendants of more recent migrants into South Africa.

Four individuals of the African lineage L2a1b carried the AluI (10400AG) restriction site diagnostic for the M lineage that split from the East African L3 branch. This mutation has not been described earlier in this haplogroup, and may therefore be a phenomenon linked to the dynamics of an urban population. Homoplasies in mtDNA are usually explained by parallel mutations. However, although mtDNA recombination happens at very low frequencies, if at all, it is legitimate to suggest it as an explanation for such apparent parallel mutations. In any event, it would be worthwhile to investigate such incidents on a broad basis (Hagelberg 2003)

Concerning future studies, it would be important to include data from Y chromosomal markers. Y chromosome data are available to a growing degree for the male population in Cape Town through forensic casework, and Y STR markers have already been described in certain population groups, but not in the same group as in the present study (Leat et al. 2004; Leat et al. 2007). Y chromosome biallelic markers would provide information on the origins of the paternal lineages of the subjects of this study, and it is likely that the proportion of haplotypes of different origins would be quite different, reflecting differential patterns of male and female migration. Together with information on the maternal lineages, a study of paternal lineages would provide a fuller picture of the genetic makeup of the urban population.

As we can see from the results, the mtDNA variation in urban Cape Town is huge. A much more extensive study would be required to include with any degree of statistical significance a large proportion of the mtDNA lineages present in the Cape Town population. The people of Cape Town are highly diverse, and the results presented here are further proof of the stupidity of the racial classifications of the former Apartheid regime. People who define themselves as White may have African mtDNA types, and the converse is true also. The policy of discrimination based on skin colour was motivated by a minority eager to retain political power and economical and social privileges was not only unjust, but also scientifically meaningless.

Figures 11 and 12 illustrate the lack of correlation between the geographical origin the mtDNA haplogroups observed in this study, and the population affiliation that the subjects themselves adhered to. Despite their own affiliation, the students had maternal lineages that originated in many diverse parts of the world. Although the sample size was very small and did not begin to mine the potential mtDNA variation in urban Cape Town, the results are a tantalising glance into an unknown area of molecular anthropology. Future studies should be performed, and will undoubtedly contribute important knowledge to the dynamics of human populations.
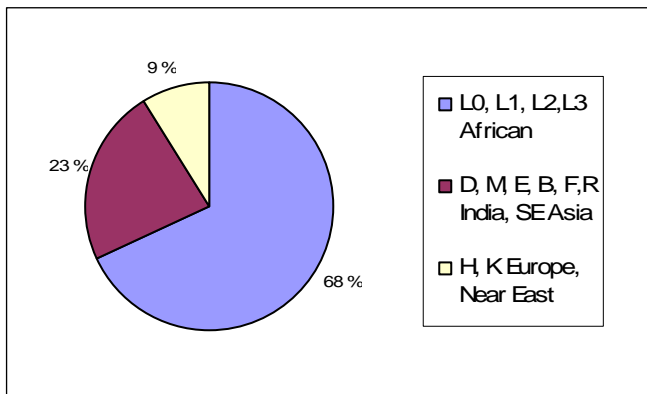
Figure 11: Distribution of the geographical origin of the haplogroups of the individuals in this study.
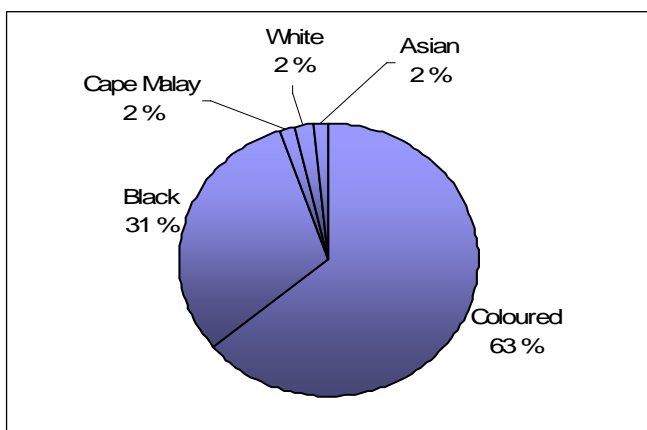
Figure 12: Breakdown of the ethnical affiliation (as defined by themselves) of the 47 individuals of this study who could be assigned to mtDNA haplogroups.

5. Acknowledgements

6. References

Allard, M. W., M. R. Wilson, K. L. Monson and B. Budowle (2004). "Control region sequences for East Asian individuals in the scientific working group on DNA analysis methods forensic mtDNA data set." Leg Med (Tokyo) **6**(1): 11-24.

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

Andrews, R. M., I. Kubacka, P. F. Chinnery, R. N. Lightowlers, D. M. Turnbull, et al. (1999). "Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA." Nat Genet **23**(2): 147-147.

Beleza, S., L. Gusmao, A. Amorim, A. Carracedo and A. Salas (2005). "The genetic legacy of western Bantu migrations." Hum Genet **117**(4): 366-375.

Bonfield, J. (2004). Staden Vs. 1.6.0. Rodger Staden, James Bonfield, Kathryn Beal, Mark Jordan and Y. Cheng. Cambridge, http://staden.sourceforge.net/.

Cavalli-Sforza, L. L., Paolo Menozzi, Alberto Piazza (1994). The history and geography of human genes. New Jersey, Princeton University Press

Clark, J. D. (1992). "African and Asian perspectives on the origins of modern humans." Philos T Roy Soc B **337**(1280): 201-215.

Deacon, H. J. (1992). "Southern Africa and modern human origins." Philos T Roy Soc B **337**(1280): 177-183.

Denaro, M., H. Blanc, M. J. Johnson, K. H. Chen, E. Wilmsen, et al. (1981). "Ethnic variation in Hpa-I endonuclease cleavage patterns of human mitochondrial-DNA." P Natl Acad Sci-Biol **78**(9): 5768-5772.

Ely, B., J. L. Wilson, F. Jackson and B. A. Jackson (2006). "African-American mitochondrial DNAs often match mtDNAs found in multiple African ethnic groups." BMC Biology **4**(1): 34-48.

Fadhlaoui-Zid, K., S. Plaza, F. Calafell, M. Ben Amor, D. Comas, et al. (2004). "Mitochondrial DNA heterogeneity in Tunisian Berbers." Ann Hum Genet **68**(3): 222-233.

Feder, J., O. Ovadia, B. Glaser and D. Mishmar (2007). "Ashkenazi Jewish mtDNA haplogroup distribution varies among distinct subpopulations: lessons of population substructure in a closed group." Eur J Hum Genet **15**(4): 498-500.

Hagelberg, E. (2003). "Recombination or mutation rate heterogeneity? Implications for Mitochondrial Eve." Trends in Genetics **19**(2): 84-90.

Hagelberg, E. and J. B. Clegg (1993). "Genetic polymorphisms in prehistoric Pacific Islanders determined by analysis of ancient bone DNA." P Roy Soc Lond B Bio **252**(1334): 163-170.

Hagelberg, E., M. Kayser, M. Nagy, L. Roewer, H. Zimdahl, et al. (1999). "Molecular genetic evidence for the human settlement of the Pacific: analysis of mitochondrial DNA, Y-chromosome and HLA markers." Philos T Roy Soc B **354**(1379): 141-152.

Henshilwood, C. S., F. d'Errico, R. Yates, Z. Jacobs, C. Tribolo, et al. (2002). "Emergence of modern human behavior: Middle stone age engravings from South Africa." Science **295**(5558): 1278-1280.

Hill, C., P. Soares, M. Mormina, V. Macaulay, D. Clarke, et al. (2007). "A mitochondrial stratigraphy for island southeast Asia." Am J Hum Genet **80**(1): 29-43.

Hitzeroth, H. W. (1984). "Anthropometric interrelationship of the Southern African Negro, Khoi and San populations .2." Homo **34**(3-4): 194-227.

Hurles, M. E., B. C. Sykes, M. A. Jobling and P. Forster (2005). "The dual origin of the Malagasy in Island Southeast Asia and East Africa: Evidence from maternal and paternal lineages." Am J Hum Genet **76**(5): 894-901.

Iliffe, J. (1995). *Africans the history of a continent*. Cambridge, Cambridge University Press.

Jobling, M. A., Hurls., Matthew., Tyler-Smith, Chris. (2004). Human evolutionary genetics. New York, Garland Science.

Johnson, M. J., D. C. Wallace, S. D. Ferris, M. C. Rattazzi and L. L. Cavallisforza (1983). "Radiation of human mitochondria DNA types analyzed by restriction endonuclease cleavage patterns." J Mol Evol **19**(3-4): 255-271.

Klein, R. G. (2001). "Southern Africa and modern human origins." J Anthropol Res **57**(1): 1-16.

Leat, N., M. Benjeddou and S. Davison (2004). "Nine-locus Y-chromosome STR profiling of Caucasian and Xhosa populations from Cape Town, South Africa." Forensic Sci Int **144**(1): 73-75.

Leat, N., L. Ehrenreich, M. Benjeddou, K. Cloete and S. Davison (2007). "Properties of novel and widely studied Y-STR loci in three South African populations." Forensic Sci Int **168**(2-3): 154-161.

Marignac, V. L. M., B. Bertoni, E. J. Parra and N. O. Bianchi (2004). "Characterization of admixture in an urban sample from Buenos Aires, Argentina, using uniparentally and biparentally inherited genetic markers." Hum Biol **76**(4): 543-557.

Metspalu, M., T. Kivisild, E. Metspalu, J. Parik, G. Hudjashov, et al. (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 Genetics **5**(1): 26-51.

Miller, G. H., P. B. Beaumont, H. J. Deacon, A. S. Brooks, P. E. Hare, et al. (1999). "Earliest modern humans in southern Africa dated by isoleucine epimerization in ostrich eggshell." Quaternary Sci Rev **18**(13): 1537-1548.

Palmberg, M. S., Per. (1995). <u>Sydafrika en regnbågsnation føds</u>. Uppsala, Nordiska Afrikainstituten och forfattarna.

Rosa, A., A. Brehm, T. Kivisild, E. Metspalu and R. Villems (2004). "MtDNA profile of West Africa Guineans: towards a better understanding of the Senegambia region." <u>Ann Hum Genet</u> **68**(4): 340-352.

Ross, R. (1999). <u>A concise history of South Africa</u>. Cambridge, Cambridge University Press.

Salas, A., M. Richards, T. De la Fe, M. V. Lareu, B. Sobrino, et al. (2002). "The making of the African mtDNA landscape." <u>Am J Hum Genet</u> **71**(5): 1082-1111.

Soodyall, H., T. Jenkins and M. Stoneking (1995). "'Polynesian' mtDNA in the Malagasy." <u>Nat Genet</u> **10**(4): 377-378.

Soodyall, H., L. Vigilant, A. V. Hill, M. Stoneking and T. Jenkins (1996). "mtDNA control-region sequence variation suggests multiple independent origins of an "Asian-specific" 9-bp deletion in sub-Saharan Africans." <u>Am J Hum Genet</u> **58**(3): 595-608.

Statistics-South-Africa. (2006). Mid-year population estimates, South Africa 2006, Statistics South Africa, Private Bag X44, Pretoria, South Africa, 0001**:** 11.

Thangaraj, K., L. Singh, A. G. Reddy, V. R. Rao, S. C. Sehgal, et al. (2003). "Genetic affinities of the Andaman Islanders, a vanishing human population." <u>Curr Biol</u> **13**(2): 86-93.

Thompson, L. (2001). <u>A history of South Africa</u>. London, Yale Nota Bene.

Vigilant, L., R. Pennington, H. Harpending, T. D. Kocher and A. C. Wilson (1989). "Mitochondrial-DNA sequences in single hairs from a Southern African population." <u>P Natl Acad Sci USA</u> **86**(23): 9350-9354.

Wiesbauer, M., D. Meierhofer, J. A. Mayr, W. Sperl, B. Paulweber, et al. (2006). "Multiplex primer extension analysis for rapid detection of major European mitochondrial haplogroups." <u>Electrophoresis</u> **27**(19): 3864-3868.

Worden, N. (1983). <u>Slavery in Dutch South Africa</u>. Cambridge, Cambridge University Press.

Wrischnik, L. A., R. G. Higuchi, M. Stoneking, H. A. Erlich, N. Arnheim, et al. (1987). "Length mutations in human mitochondrial-DNA - direct sequencing of enzymatically amplified DNA." <u>Nucleic Acids Res</u> **15**(2): 529-542.

7. Appendix

7.1 Reagents

Buffer 5x TBE Stock Solution:

In one litre:

54.0g Trizma base
       Tris (hydroxymethyl)-aminomethan
       $H_2NC (CH_2OH)_3$
       M=121.14g/mol
       1.08387.2500
       Lot K34682787523
       MERCK

27.5g Boric Acid
       H_3Bo_3
       M=61.83g/mol
       1.00165.1000
       Lot A646565
       MERCK

20.0ml 0.5M EDTA pH8
       $C_{10}H_{14}N_2Na_2O_8x2H_2O$
       1.08418.1000
       Lot K34033918503
       MERCK

1.0l dH_2O

pGEM DNA Ladder:
Cat No G174A20198501
50µl
Promega

100bp DNA Ladder:
Cat No G210A1429071
Promega

Montage^{TM} PCR centrifugal filter devices:
Kit: Lot R3DN02183
Millipore

1% Agarose gel in one litre:

1litre 1xTBE Buffer => 0.2l 5xTBE/L dH$_2$O from stock solution

10g agarose
SeaKem^R LE Agarose
Cat No: 50004
Size: 500g
Lot no: AG2510

0.25mg/ml EtBr

NuSieve 3:1 agarose:
For analysis of nucleic acid $\leq$ 1kb
Cat No: 50090
Size 125g
Lot no: AG5183

EtBr 1% solution in Water:
Lot no: 76562-30
MERCK

Chelating Resin:
(Imidodiacetic acid)
C-7901
Lot: 107H0335
Sigma

Loading dye:
Cat No G190A20319001
Blue/Orange 6x loading dye
Promega

AluI:
Cat Nr 1004A
Lot No K4001AA
500u
TaKaRa

Buffer 10xL for AluI:
Lot No A1020
TaKaRa

DdeI:
Code R629B 18374529
1000u
Promega

Buffer D10x for DdeI:
Code R004A 20327901
Promega

All PCR where performed with
AmpliTaq Gold
Gene Amp 10x PCR Buffer
MgCl Solution
From Applied Biosystems
Part nr: N808-0167
Lot nr: F04613