UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Ancient DNA analysis of the Thulamela remains: Deciphering the migratory patterns of a southern African human population

Molebogeng K Bodiba

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Ancient DNA analysis of the Thulamela remains: Deciphering the migratory patterns of a southern African human population

By

Molebogeng Keamogetswe Bodiba

Submitted in fulfilment of the requirements of the degree of

Master of Science

In the

FACULTY OF HEALTH SCIENCES

Department of Anatomy

School of Medicine

University of Pretoria

2014

Supervisor: Prof. M. Steyn

Co-supervisors: Prof. P. Bloomer

and

Prof. Dr. Rühli

© University of Pretoria

Declaration

I, Molebogeng Keamogetswe Bodiba, hereby declare that this dissertation is my own work and has not been presented by me for any degree at this or any other university/institution

Signed:………..

Date:………….

Department on Anatomy, School of Medicine, Faculty of Health Sciences, University of Pretoria

South Africa

© University of Pretoria

Abstract

Bio-archaeology is the study of biological remains found at sites of archaeological interest. It is an interdisciplinary science employing different scientific fields including physical anthropology, geography, archaeology and genetics. Genetic analysis includes ancient DNA (aDNA) studies, now a specialised field in genetics. This approach was used to analyse human skeletal material of eight individuals from various Iron Age archaeological sites in southern Africa. Included in this sample is a naturally mummified individual from Tuli, in Botswana. The context of the specimens found in the Limpopo Province (Thulamela), as well as their cultural links with the Zimbabwe Culture Complex (which includes Mapungubwe and Khami) suggests that some gene exchange might have occurred. While this is not the first aDNA study on southern African samples, it is the first aDNA study based on southern African Iron Age human individuals and also included a naturally mummified individual.

Morphometric and morphological analyses have indicated the age at death, sex and health status of the individuals, and the context in which they were found has helped in assessing their cultural affinity. Bone samples were analysed in a specialized aDNA laboratory at the Centre for Evolutionary Medicine in Switzerland. Following DNA extraction, ancestry-specific mitochondrial DNA was amplified from all samples and was compared to that of modern sub-Saharan Africans whose data were accessed from GenBank.

Some individuals show (maternal) genetic similarities to present-day Sotho/Tswana groups. The male individual from Thulamela aligns somewhat more with the groups from the west and the female with the eastern peoples. Two Early Iron Age individuals from Happy Rest presented some similarities to the Khoesan peoples. Genetic-sex determination efforts were inconclusive for all individuals.

The purpose of this study was to place the Thulamela individuals within the context of the genetic diversity in South Africa. It was noted that the introduction of genetic material from the early Sotho/Tswana was gradual in the case of Thulamela. Two other individuals from Happy Rest, who were contemporaries of each other, showed very little genetic variation and it can be said that their maternal DNA was of the same (possibly Khoesan) origin. Further resolution in haplotype assignment will be done in future. These temporally and spatially dispersed individuals can only provide a glimpse into the population interactions of the Iron Age that may have partially shaped the immense genetic diversity of present-day southern Africa.

© University of Pretoria

Abstrak

Bio-argeologie is die studie van biologiese oorskot gevind op terreine van argeologiese belang. Dit is ‘n interdissiplinêre wetenskap wat gebruik maak van verskillende wetenskaplike velde, onder andere fisiese antropologie, geografie, argeologie en genetika. Genetiese analise sluit studies van antieke DNA (aDNA) in, wat gesien word as ‘n gespesialiseerde veld in genetika. Hierdie benadering is gebruik om menslike skeletale materiaal van agt individue van verskeie Ystertydperk argeologiese terreine in suider Afrika te analiseer. Hierdie steekproef sluit ‘n natuurlik gemummifiseerde individu van Tuli, Botswana, in. Die konteks van die monsters wat in die Limpopo Provinsie (Thulamela) gevind is, asook hulle kulturele bande met die Zimbabwe Kultuur Kompleks (wat Mapungubwe en Khami insluit) dui daarop dat genetiese uitruiling moontlik plaasgevind het. Terwyl hierdie nie die eerste studie is wat fokus op aDNA van suider Afrikaanse monsters nie, is dit wel die eerste studie gebasseer op suider Afrikaanse Ystertydperk individue en sluit ook ‘n natuuurlik gemummifiseerde individu in.

Morfometriese en morfologiese analise het ‘n aanduiding gegee van ouderdom van dood, geslag en gesondheidsstatus van hierdie individue, terwyl die konteks waarin hulle gevind is gehelp het met die assessering van hulle kulturele affiniteit. Beenmonsters is geanaliseer in ‘n gespesialiseerde aDNA laborotorium by die Centre for Evolutionary Medicine in Switserland. Na DNA ekstraksie is afstamming-spesifieke mitochondriale DNA van al die monsters geamplifiseer en vergelyk met dié van moderne sub-Sahara Afrikane wie se data verkry is van GenBank.

Sommige individue vertoon (moederlike) genetiese ooreenkomste met hedendaagse Sotho/Tswana groepe. Die manlike individu van Thulamela is effens meer in lyn met groepe vanaf die weste en die vrou met oostelike mense. Twee Vroeë Ystertydperk individue van Happy Rest vertoon ooreenkomste met die Khoesan. Genetiese geslagsbepaling pogings was onbeslis vir alle individue.

Die doel van hierdie studie was om die Thulamela individue binne die konteks van die genetiese diversiteit van Suid-Afrika te plaas. Daar is waargeneem dat die bekendstelling van genetiese materiaal van die vroeë Sotho/Tswana geleidelik gebeur het in die geval van Thulamela. Twee ander individue van Happy Rest, wat tydgenote was, toon baie min genetiese variasie en daar kan gesê word dat hulle moederlike DNA van dieselfde oorsprong (moontlik Khoesan) was. Verdere resolusie in haplotipe toedeling sal in die toekoms uitgevoer word. Hierdie tydelik en ruimtelik verspreide individue gee slegs ‘n kykie in die bevolkingsinteraksies van die Ystertydperk wat gedeeltelik bygedra het tot die vorming van die geweldige genetiese diversiteit van die hedendaagse suider Afrika.

© University of Pretoria

Acknowledgements

My appreciation goes to my supervisors, Professors Maryna Steyn (Department of Anatomy, University of Pretoria), Paulette Bloomer (Department of Genetics, University of Pretoria) and Frank Rühli (Institute of Anatomy, University of Zurich). I thank Prof. Maryna Steyn for her recourcefulness and patience at the development and completion of this project. My gratitute also goes to Prof. Dr. Rühli for giving me the opportunity to work at the ZEM and collaborate with his laboratory and to the rest of his team for making my stay in Switzerland a joyful one. I thank all funding bodies (University of Pretoria, National Research Foundation, Swiss Federal Government Scholarship for Foreign Students, and SANParks/UP post-graduate fund with the assistance of Mrs Louise Euthimiou) as well as the colleagues and friends at the Department of Anatomy (University of Pretoria) and the ZEM (Zentrum für Evolutionäre Medizin/Centre for Evolutionary Medicine), University of Zurich. Many thanks are also owed to SANParks (South African National Parks authority) and SAHRA (South African Heritage Resources Agency) for granting permission for the analysis. I acknowledge Dr. Abigail Bouwman for the training in ancient DNA laboratory techniques and Mrs Gülfirde Ackgül for her technical assistance in the lab. Gratitude also goes to Prof. Paulette Bloomer and Dr. Michael Campana for the assistance and advise given on the DNA analysis. I also express thanks to Mr. Coen Nienaber and Miss Anja Meyer for their most valuable contributions with regard to comparing the DNA results to previous archaeological information. I thank Mrs Marinda Pretorious for her help with organising and designing posters of this project. Last but not least I thank my family and friends for their prayers and support.

© University of Pretoria

TABLE OF CONTENTS

Declaration ---------- ii

Abstract ---------- iii

Abstrak ---------- iv

Acknowledgements ---------- v

List of figures ---------- ix

List of tables ---------- x

Abbreviations and definitions ---------- xi

Chapter 1 ---------- 1

Introduction ---------- 1

1.2 Aim/purpose of study ---------- *3*

Chapter 2 ---------- 5

Literature review ---------- 5

2.1 Archaeological background ---------- 5

2.1.1 African population origins and Bantu-speaker migrations ---------- *5*

2.1.2 The Iron Age in Southern Africa and the Zimbabwe culture ---------- *9*

2.2 Late Iron Age archaeological remains at Thulamela ---------- 10

2.2.1 Other relevant Iron Age sites ---------- 12

2.2.1.1 Happy Rest ---------- 12

2.2.1.2 Makgope ---------- 13

2.2.1.3 Kon-Stayt ---------- 14

2.2.1.4 Nwanetsi ---------- 15

2.2.1.5 Tuli ---------- 16

2.3 Ancient DNA analyses in biological anthropology ---------- 17

© University of Pretoria

2.3.1 Sources of aDNA and relative efficiency in DNA yield ---------- *19*

2.3.2 Factors that affect the quality/ yield of aDNA ---------- *20*

2.3.2.1 Post-mortem biochemical changes in DNA ---------- 20

2.3.2.2 Environmental effects on post-mortem DNA ---------- 25

2.3.2.3 Preferable bone samples to use ---------- 25

2.3.3 Molecular techniques in aDNA research ---------- *26*

2.3.3.1 DNA markers for aDNA experiments ---------- 27

2.3.4 Amplification and sequencing of aDNA ---------- *30*

2.3.4.1 PCR ---------- 30

2.3.4.2 Sequence generation and analysis ---------- 32

2.3.4.2.1 Haplotype networks and phylogeography ---------- 35

2.4 Drawbacks of aDNA research ---------- 41

2.5 Ethical considerations ---------- 42

2.6 Conclusion ---------- 43

Chapter 3 ---------- 45

Materials and Methods ---------- 45

3.1 Samples and procedure ---------- 45

3.2 Ethical and legal clearance ---------- *46*

3.3 Methods ---------- *47*

Experimental flow diagram ---------- 48

3.3.1 Contamination control ---------- *49*

3.3.2 Sample preparation ---------- *49*

3.3.4 DNA extraction and purification ---------- *50*

3.3.5 DNA amplification, sequencing and analysis ---------- *50*

3.3.5.1 PCR amplification ---------- 50

3.3.5.2 Cloning ---------- 51

3.3.5.2.1 Vector preparation and ligation ---------- 52

3.3.5.2.2 Colony picking and PCR amplification ---------- 52

3.3.5.2.3 Sequencing and analysis ---------- 52

© University of Pretoria

Chapter 4 ---------- **54**

Results ---------- **54**

4.1.1 DNA extraction ---------- *54*

4.1.2 PCR ---------- *54*

4.1.2.1 Standard PCR ---------- 54

4.1.2.2 Real-Time PCR ---------- 58

4.1.3 Cloning ---------- *58*

4.1.3.1 Colony picking and PCR ---------- 58

4.1.4 Sequencing and network construction ---------- *60*

4.1.5 Statistical analysis ---------- *74*

Chapter 5 ---------- **75**

Discussion ---------- **75**

5.1 Sample preservation and sampling ---------- *75*

5.1.1 DNA extraction, PCR and contamination ---------- *75*

5.1.2 Contamination control, amplification and sequencing ---------- *76*

5.1.2.1 Analysis of sequences, network and haplotyping ---------- 77

5.1.2.2 Relationships between individuals ---------- 80

5.1.2.3 Comparison with other African sequences ---------- 80

5.2 Comparison to the archaeological evidence ---------- *81*

Chapter 6 ---------- **86**

Conclusions ---------- **86**

List of References ---------- **88**

© University of Pretoria

List of figures

Figure 2.1 Bantu-speaker migration patterns according to the broad definition of ceramic styles......... 8
Figure 2.2. Specific areas of aDNA damage……………………………………………………….22
Figure 2.3. Post-mortem DNA damage in fossil remains 23
Figure 2.4. Schematic representation of how jumping PCR occurs.. 24
Figure 2.5. Illumina sequencing procedure... 34
Figure 2.6. Human mtDNA phylogeny showing all known haplogroups…………………………...39
Figure 3.1. Map of the eastern part of southern Africa with the relative locations of the sites of interest.. 45
Figure 3.2. Summary of methods... 48
Figure 3.3. Photographs of some of the teeth used (and accession numbers) prior to pulverisation 49
Figure 4.1a PCR products after agarose gel electrophoresis……………………………….......56

Figure 4.1b. PCR products after agarose gel electrophoresis. 57
Figure 4.2. Gel electrophoresis image of the linear vector. 58
Figure 4.3. Gel electrophoresis images of PCR products after cloning 59
Figure 4.4a. Electropherogram of sequence generated from cloned product...................................... 61
Figure 4.4b. Electropherogram of a sequence generated directly from a PCR product....................... 61
Figure 4.5a. Representation of query sequences versus the rCRS 71
Figure 4.5b. Representation of query sequences versus the rCRS... 71
Figure 4.6. Network representing genetic distances among individuals representing each site. 72
Figure 4.7. Allele network showing the relationships to other sub-Saharan African sequences. 73

© University of Pretoria

List of tables

Table 2.1. Time-line summary of the sites relevant to this study……………………………………...9
Table 4.1a. Summary of alignments with reference sequences 62
Table 4.1b. Summary of sequences aligned with rCRS 70
Table 4.2. Haplogroup assignment for individuals representing each site 71
Table 5.1. Cultural and genetic affinities of the assessed individuals 83

© University of Pretoria

Abbreviations and definitions

The following abbreviations and definitions were used throughout this study:

DNA: Deoxyribonucleic acid

aDNA: ancient DNA

PCR: Polymerase chain reaction

EIA: Early Iron Age

MIA: Middle Iron Age

LIA: Late Iron Age

RNA: Ribonucleic acid

DNase/RNase: Enzyme that cleaves DNA/RNA (respectively)

Proteinase K: An endolytic enzyme that digests protein in the presence of detergents

Ancient DNA: DNA that is older than 100 years old (can vary)

Iron Age: The period marked by the increased use of iron technology, as well as agricultural pastoralism in southern Africa, occurring between AD 500 and 1700.

Early Iron Age: The period occurring between AD 500 and AD 1000 (approximately).

Middle Iron Age: The period occurring between AD 1000 and AD 1300 (approximately).

Late Iron Age: The period occurring between AD 1300 and AD 1700 (approximately).

Bantu: The group of Niger-Congo languages spoken by sub-Saharan Africans who descend from farmers that migrated out of their original homeland about 5000 years ago. In southern Africa, these include major groups such as the Shona, Nguni and Sotho/Tswana.

Khoesan: Includes both the Khoe (pastoralists) and the San (hunter-gatherers) peoples who are indigenous to southern Africa and whose languages are separate from the Bantu languages.

Gene genealogy: The use of DNA data to estimate the genetic relationships of individuals.

Coalescence: The probability that two alleles share a common ancestor in the previous generation.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Haplotype: A collection of alleles that are in close proximity along a chromosome or other genome and tend to be inherited together in successive generations.

Haplogroup: A collection of haplotypes that have some variants between them.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Chapter 1

Introduction

Archaeological remains are indispensable when studying population histories, aspects of culture, as well as the evolution of human cultures. They serve as physical proof of some knowledge that may have been passed down by oral teaching, or validate hypotheses developed by scientists over the years. Archaeological remains may manifest in the form of cultural artefacts, or items such as tools and pots left behind by previous complex societies; these artefacts are useful in studying the life-style and survival strategies employed by these cultural groups, but in order to obtain a closer look at populations in terms of nutrition, health and even social complexities, the skeletal remains of members of these groups are highly valuable.

A bio-archaeological study (including archaeology, skeletal and biological or molecular material) requires the cooperation of multiple disciplines; these include anatomy, physical anthropology, cultural anthropology, radiology, chemistry, and geology, to name a few. A relatively recent addition to this field of study is genetics, and with the work that has been performed on ancient DNA on skeletal and even plant remnants, genetics has become useful in the study of ancient remains of human and hominid individuals (1-3).

When DNA is available, it can reveal information about an individual's population affinity, pathogens, sex, and even the probable geographic origin and possible descendants of the individual(s) being studied. These methods of assessment have already been developed in biological anthropology, and genetics can help to substantiate such findings (1).

DNA has been used extensively in the studies of human origins, with varying success, and markers have been developed that can aid in such investigations. Examples include mitochondrial DNA markers that can help clarify maternal lineages (2, 3). Ancient DNA can help elucidate genetic relationships between ancient and contemporary populations, migration patterns, and changes in genetic diversity over time, the reconstruction of kinship systems, social structure, and mating patterns (1-3).

The discovery of an archaeological site in the Thulamela area near the Levhuvhu River and just south of the Zimbabwean border revealed, among other items such as river rocks, gold,

© University of Pretoria

beaded jewellery and clay pots, the skeletal remains of two individuals, one female and one male (UP43 and UP44 respectively), whose burials also included grave goods such as items made of gold (4). The individuals therefore seemed to be of high social status, due to the lavish artefacts found associated with their remains. Both were adults at the time of death and the second individual, UP44, appears to have undergone a secondary burial.

The archaeological evidence (such as pottery and settlement pattern) testifies to a cultural link between the Iron Age society that resided in Thulamela and the Great Zimbabwean Shona speaking peoples (4). Presently, the Thulamela area is occupied by the vha-Venda and va-Tsonga/Shangaan people and possible genetic links between these present day peoples with the Thulamela individuals can be hypothesised.

The geographical location of Thulamela, as well as the presence of the secondary burial of one individual, may lead to the question of whether its existence was a result of a break-away group of people moving into Thulamela from somewhere else and then re-burying this individual for sentimental and religious purposes. Thulamela is near the border of South Africa with Zimbabwe. It appeared that the individuals may have lived within the context of the Zimbabwe culture, as seen in the similarities in pottery and settlement pattern (4).

Thulamela is also on the path of migration of the Bantu speaking people who were ancestral to present-day Nguni and Sotho/Tswana groups of South Africa (5). However, much of the evidence found at the Thulamela site points towards a society with a Zimbabwean origin, due to similarities in architecture and positioning of the graves (6). Studying the probable population affinity of these buried individuals by obtaining a closer, molecular perspective may serve to clarify where these individuals had come from and their probable descendants.

Several groupings of ethnically diverse people exist in southern Africa. In Zimbabwe, Shona speakers make up the majority of the population, followed by the Ndebele people who settled there from South Africa under the leadership of Mzilikazi who broke away from the Zulus. The border region of South Africa and Zimbabwe is occupied by the Lemba, the Venda and the Tsonga/Shangaan people. Tshi-Venda and Xi-Tsonga are of two separate language groups. The other main groups of South African peoples are the Sotho/Tswana, which include the North Sotho (Sepedi), South Sotho and Tswana people,

© University of Pretoria

these form a single language group, as well as the Nguni, which include the Xhosa, Ndebele, Zulu and Swati people, which also form a language group (7). Other groups include the Khoesan people, who are the oldest known people to have occupied southern Africa, as well as people of Asian, European and mixed descent (7, 18).

It has been proposed that the individuals excavated from Thulamela were of a Zimbabwean origin, who may have been ancestral to the modern Shona speaking people, and were succeeded by the vha-Venda. The secondary burial may have been the result of a break-away group collecting the bones of their 'father' in order to re-bury him in their new home. Some researchers speculate that certain Zimbabwean groups had a tradition of leaving a body to decompose completely before it is buried and this could explain the secondary burial of the one individual (8).

Evidence of secondary burials in southern Africa is not common, but the practice of secondary burials in Africa, however infrequent, is not a new one (8). Two examples include pot burials from Broederstroom and the Northern Province (now known as the Limpopo Province) (9, 10). We can now only speculate as to why the secondary burial of the one individual took place.

1.2 Aim/purpose of study

The aim of this study was to attempt to extract DNA from the skeletal remains found at Thulamela and other related archaeological sites, in order to:

1. Substantiate the results obtained from the osteometric analyses with regard to sex and population affinity and

2. Compare the DNA of the individuals to each other and to that of living individuals, (using reference DNA sequences from GenBank) so as to determine the population of origin and probable descendants of these individuals, with the greater objective being to add to existing knowledge about the migration of peoples in southern Africa.

The additional individuals used in this study were from other Iron Age sites in southern Africa, namely, Tuli in Botswana (a naturally mummified individual), Nwanetsi, Makgope, Happy Rest and Kon-Stayt.

© University of Pretoria

The motivation to address these questions mainly stem from the need to add to gaps in knowledge about the history of indigenous southern African people prior to the settlement of Europeans. From a scientific standpoint, it would be interesting to see the change in genetic diversity over time within the indigenous groups of southern Africa, as the effects of migration, gene flow, population bottlenecks as well as the possibility of genetic exchange between genetically related people as a result of restrictions of marriages between members of certain groups, can be evident on the genome.

Ancestry-specific DNA markers would be useful as a first step towards understanding the genetic history of the indigenous groups of southern Africa. It should be noted that due to the small sample size, as well as the temporal and geographical diversity of the individuals under study, the results can only suggest possible genetic relationships between groups, however, answers to more complex questions can really be achieved by use of larger samples, which are difficult to obtain.

© University of Pretoria

Chapter 2

Literature review

2.1 Archaeological background

2.1.1 African population origins and Bantu-speaker migrations

The term 'Bantu' is used to group people speaking specific languages in sub-Saharan Africa. The word has many variants, with the root word '*ntu*' which means person or being, and varies from '*mutu', 'vhanu', 'vathu', 'munu', 'motho', and 'muntu'* in the different languages. Bantu languages fall under the Niger-Congo group of languages and were described by Malcolm Guthrie in 1948. The expansion of Bantu-speakers pertains to the diffusion of technology, language, culture and even genes.

Climatic changes in Africa, about 5000 years ago, caused some of the Bantu speaking farmers to migrate away from their putative land of origin in central-west Africa, across central Africa towards the east and the west and then eventually entering the southernmost parts of Africa from varying directions. These migrations also resulted in the settlement of some groups and the further migration of subsets of these groups. These migrations of farmers lead to the isolation and/or admixture with pre-existing hunter-gatherers such as the Khoesan and Pygmy people (5, 11). Archaeological remnants, such as settlement patterns and ceramic design coupled with linguistic and dating analyses, show that the general pattern of migration toward the south happened at different times, beginning along the west coast, and then along the east. It is not known for sure if these migrating groups intermingled along the way, but it is supposed that some contact occurred between these and other existing groups (5, 12). Evidence from Iron Age contexts are closely correlated to the Bantu expansion, whereas Stone Age technologies are mostly associated with the Bushmen or Khoesan, who made little use of iron technology, in southern Africa. Therefore the use of metallurgy, farming, as well as the ownership of land/permanent settlements is related to the Bantu expansion (13).

From the perspective of the southern African population demographics, which forms part of this study, it is important to place some focus on the migrations of Bantu-speaking people in this region. Prior to the dawn of molecular analyses, archaeologists and linguists have made valuable contributions towards understanding the movement and interactions of

© University of Pretoria

populations in the past, and the recent addition of molecular studies have contributed equally well to the field of bio-archaeology (3, 7, 14, 16). Archaeological studies have involved the examination of material culture, which included the analyses of ceramics among others. The decoration and design of ceramic vessels seem to be repeated among some groups over a given time and have thus been used as indicators of cultural affinity, or at least, they represent the trend that is followed by the complex society that is economically and culturally dominant.

Ceramic pattern analysis is also loosely associated with linguistic groups (14). One might raise the argument that ceramic design does not equal group affinity (15, 16), but one can at least use ceramics to determine the possible direction of group movements (14). Ceramics do not only provide the potential to place an archaeological site at a certain point in time (assisted by carbon dating of animal bone or charcoal, etc.), but the vessels may have elaborate or even typical patterns, making it possible for one to assess the possibility of interaction between people of different groups and to some extent, the cultural affinity of the people who made the ceramics. Ceramic analysis has been used to trace the migration of groups in and around southern Africa. It is widely thought that similar ceramics between groups suggest the possibility of a common or similar language, since part of the teaching process involves verbal communication (14, 17, 18).

The migration of Bantu-speakers occurred in different stages, different times and has been loosely defined into two separate streams, an eastern and a western stream (19). Huffman (1989) also suggested a third stream, called the central stream. The three migration streams are named after their corresponding ceramic styles, the details of which are beyond the scope of this discussion. These originate from the so-called Chifumbaze complex which consists of two traditions (which are also linked to linguistic and cultural groups), the Kalundu tradition or western stream and the Urewe tradition or eastern stream. In southern Africa, the two branches of Urewe are the Kwale and the Nkope branches, the latter also represents the central stream of migrations (14). The earliest East African pots, part of the Urewe tradition, are dated around the 2nd to the 5th century and the associated settlements also include iron smelting sites (20, 21). The western stream or Kalundu ceramics are found along western Tanzania, western Zimbabwe, eastern Botswana and the former Transvaal province in South Africa. For the Kwale branch of the Urewe tradition, ceramics are found along the coast from Kenya to present-day KwaZulu-Natal, as well as Zambia

© University of Pretoria

and parts of Zimbabwe and these appear around the 3rd century. The Nkope branch pots, also known as the central stream and dated to the first century, spread from Angola, western Zambia, Botswana and parts of Zimbabwe. Figure 2.1 is a summary of the migration streams into southern Africa.

Pottery from the Kalundu branch or western stream is often associated with settlements that follow the central cattle pattern or CCP. These include Sotho/Tswana and Nguni settlements. The earliest ceramics from the Kalundu tradition in southern Africa are the Gokomere type, which predate Happy Rest ceramics. The latter are followed by the Zhizo style, which were derived from the Nkope branch or central stream ceramics and appear at Schroda, an early capital that survived for approximately 100 years and was abandoned around AD 1000. At around the same time, K2 became an important capital for a short period between AD 1030-1220 and was closely followed by the rise of Mapungubwe from AD 1220-1300, near the confluence of the Shashe and Limpopo Rivers (12, 22).

The next important capital was Great Zimbabwe, which was occupied by the ancestors of the Shona. These early Shona are associated with the introduction and dominance of the Leopard's Kopje style pots in the Shashe-Limpopo which arise from the Kalundu tradition (22, 23). However, Shona is classified as an eastern Bantu language and it is therefore thought that the makers of the Kalundu tradition might have spoken an eastern Bantu language (12). The Zimbabwe culture is thought to have evolved between AD 1000 and AD 1300. When Mapungubwe was abandoned around AD 1300, some of its occupants may have moved to present-day Venda while others occupied Great Zimbabwe (22, 24). The adoption of some Shona rituals south of the Soutpansberg was evidence of the prestige and influence of Shona speakers in this region (20).

Complex societies south of the Limpopo that shared cultural similarities with Mapungubwe survived to the 14th century, which is also the same time as the establishment of the early Sotho/Tswana and Nguni, as well as the existence of Thulamela, the last of the important capitals in the Zimbabwe culture sequence. The ancestors of the Sotho/Tswana and Nguni probably arrived in southern Africa around AD 1050 and moved through the Shashe-Limpopo region somewhere between AD 1050 and 1350; the Nguni continued southwards and the Sotho/Tswana moved into the central plateau. The CCP is thus shared between Sotho/Tswana, Nguni and EIA peoples. While the CCP persists throughout the Iron Age,

© University of Pretoria

Broederstroom represents the EIA and the Nguni represent the LIA peoples (25). Sotho/Tswana and Nguni pottery are the Moloko and Blackburn branches respectively and they both derive from the Urewe tradition, or eastern stream migrations (14).

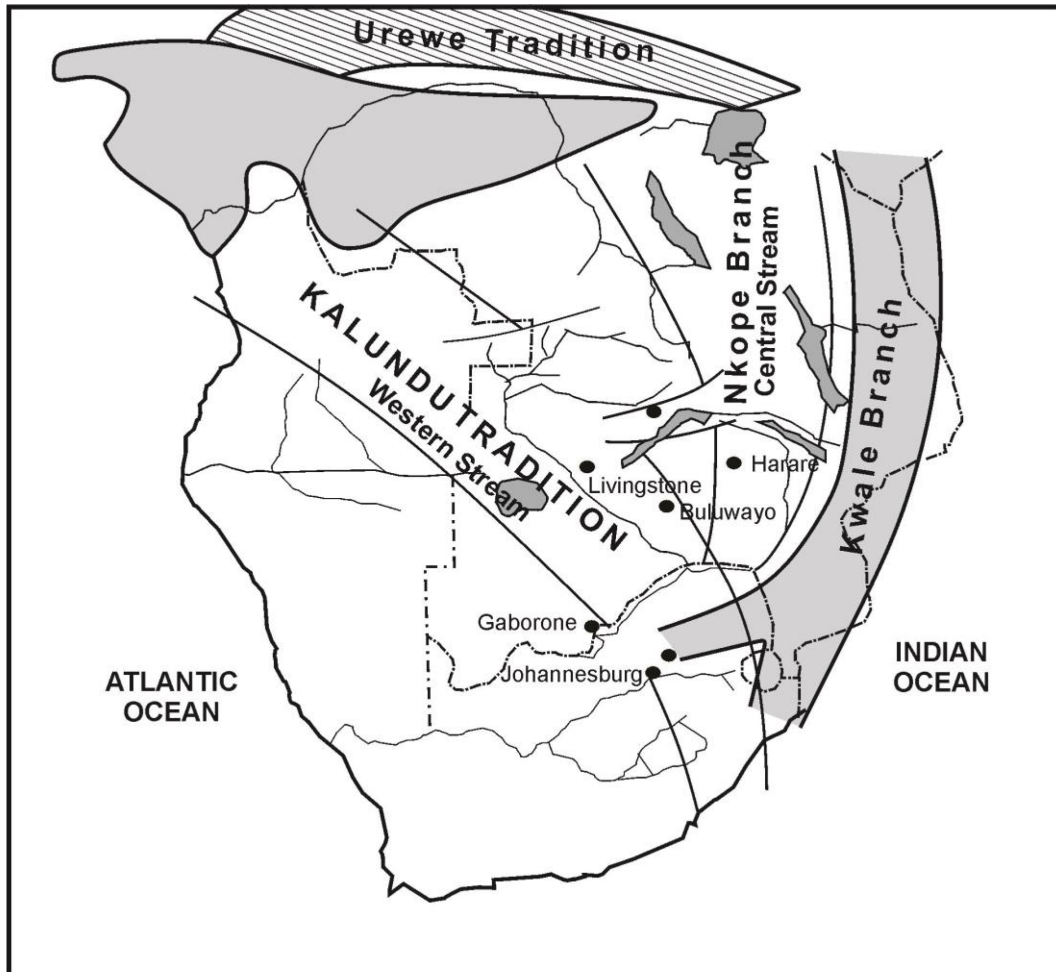


Figure 2.1. Bantu-speaker migration patterns according to the broad definition of ceramic styles (12).

Evidence from molecular studies generally supports the pattern of Bantu-speaker migrations as seen by ceramic analyses and other studies of material culture. The distribution of mitochondrial DNA (mtDNA) lineages have been associated with the migration pattern of Bantu-speakers from West Africa, through the central region and then to the south (26, 27). In addition, studies on Y-chromosome variation in sub-Saharan Africa revealed a deep rooted Bantu-speaker haplotype that originated somewhere near the

© University of Pretoria

reputed Bantu-speaker homeland and it was suggested that it was dispersed by farming populations throughout this region (28).

2.1.2 The Iron Age in Southern Africa and the Zimbabwe culture

The Iron Age can be divided into three periods: The Early Iron Age (EIA) saw the appearance of the Gokomere and Happy Rest ceramics, arising from the Kalundu tradition and the latter occurred at around AD 500; the Middle Iron Age (MIA), marks the time that Zhizo ceramics replaced Happy Rest, and the makers of Zhizo, Leokwe and Leopards's Kopje pottery developed the early Zimbabwe culture at the Shashe-Limpopo basin around the 10th century. The development of the Zimbabwe culture actually culminates with the establishment of Mapungubwe between AD 1220 and 1300 (12, 22, 29); the Late Iron Age (LIA) is marked, in part, by the introduction of the early Sotho/Tswana and Nguni into southern Africa, as well as the replacement of Khami by Thulamela as an important capital (20). Table 2.1 is a time-line summary of some of the sites that are relevant to this discussion as well as those included in this study.

Table 2.1: Time-line summary of the sites relevant to this study

Prominent state/ceramic/ event(s)	Gokomere and Happy Rest ceramics	Happy Rest individuals	Schroda	Makers of Leopard's Kopje, Leokhwe and Zhizo begin development of Zimbabwe Culture	K2	Mapungubwe, culmination of Zimbabwe Culture	Great Zimbabwe (early Shona)	Early Sotho/Tswana and Nguni	Kon-Stayt	Thulamela	Khami	Tuli	Makgope
Approximate date	AD 500	AD 683-798	AD 1000	AD 1000	AD 1030-1220	AD 1220-1300	AD 1290-1350	AD1050-1350	AD1063-1235	AD 1400-1700	AD 1450-1820	AD 1400-1500	18th cent

The Zimbabwe culture sequence may be divided into four periods, each with respect to an important capital: Mapungubwe (AD 1220 to 1300), Great Zimbabwe (AD 1290 to 1450), Khami (AD 1450 to 1820) and Thulamela in the LIA (AD 1400 to 1700) (22, 29-31). The material remains at Thulamela suggest a link between these people and that of Zimbabwe and oral tradition among the vha-Venda testifies that they originate from Zimbabwean groups. The geographical context of Thulamela and Great Zimbabwe is along the same path as the migration of the Bantu speaking peoples that began around 4000 or 5000 years ago (5, 32), and who arrived in southern Africa approximately 2000 years before present (BP) (5). The degrees of similarity between African languages provides evidence of contact between different population groups, and even indicates if some languages have a common origin. Studies have been performed in order to clarify aspects of evolutionary and demographic processes associated with the spread of Bantu languages in sub-Saharan

© University of Pretoria

Africa and to investigate whether patterns of genetic variation fit with models of population expansion that are based on linguistic and archaeological data (26). They concluded that groups that are geographically close to each other tend to have similarities between and among languages, and the same holds for the Shona and Venda language.

2.2 Late Iron Age archaeological remains at Thulamela

Thulamela, a site in Limpopo near the Levhubu River, was first described and investigated by Sydney Miller (6, 33) and this investigation revealed the graves of two individuals which were then described by Steyn and colleagues in 1998. The project continued from 1993 to 1997, and was mainly focused on the architecture and settlement layout of the site itself, as well as the artefacts found. The site provided clues pertaining to the political structure of the community, as well as the culture that was practiced there. The site was occupied in three phases between the 13th and 17th centuries and is situated on a hilltop in the northern Kruger National Park, overlooking the Levhuvhu River. The settlement pattern, consisting of a central focal point and the relative seclusion of the elite by stone-walled enclosures on a hilltop (as opposed to the peasant community at the bottom of the hill), follows what is known as the Zimbabwe pattern (30, 34).

Thulamela seems to have first been occupied in the second half of the 13th century, during which time it is not clear whether stone-walled structures existed, however, ostrich eggshell beads at the lowest layers suggest the presence of human occupation. The second occupation, dated to the 14th and mid-15th centuries, showed numerous glass beads, ivory, gold, and seashells, as well as ceramics with complex designs, and it was thus evident that there was some trading with the East Coast port settlements. Evidence of stone-walling was also noted. By the 17th century, the third and last phase of occupation, ironmongery had reached its peak and ceramics had evolved to the Khami style which derives from the western stream or Kalundu tradition. The presence of a Chinese porcelain fragment also indicates trade with countries as far as China, or at least trade with other groups that also traded with Asia.

In 1996, the excavation yielded two burials with the skeletonised remains of one male and one female individual (4). Both were buried with gold jewellery and river stones were found at or near the burial pits, suggesting that these burials may have been performed with

© University of Pretoria

some care and therefore the deceased were possibly significant members of the community. The position of these graves at the top of the hill instead of at the bottom also suggests that they were part of the elite class and this manner of burial follows that of the Zimbabwe pattern. Morphometric assessment showed that the two were adults at the time of death, ranging between the ages 45-60. Dating methods, using C-14 dating directly from bone samples, revealed that the two were not contemporaries; UP43, the female, was dated to about 340 ± 40 BP and calibrated to approximately AD 1525-1632, while UP44, the male, was dated to about 520 ± 40 BP, which was calibrated to approximately AD 1424-1437 (4). It is thus possible that these individuals are of the same population group, but different generations.

The first individual, a female, was buried in a tightly flexed position, facing eastwards and with the head pointing northward. The axial orientation may point to the direction of origin of the deceased, as suggested by van Warmelo (1940), who also mentions that burial rites depended on the status and sex of the deceased and were therefore not universal to all members of a group (35). The individual had also been adorned with gold bracelets. Except for a missing skull base, fractured zygomae and mandible, as well as fragmentary extremities, the skeleton was relatively well-preserved. Health indicators included some enamel hypoplasia and tooth-wear. The grave-pit itself was dated to about 566 ± 160 BP, corresponding to the site's first phase of occupation.

The second individual, a male, was found buried within a different stone-walled enclosure. The skeletal remains were completely disarticulated, and appeared to have undergone a secondary burial. The long bones had been deliberately arranged and piled and the skull was found face down and on top of the rest of the bones. Secondary burial practices had been previously noted among Venda chiefs (36). The preservation of these remains was relatively poor. Some osteophytes were noted that may have caused nerve impingement and head and neck pain. The third lumbar vertebrae also appeared to have been wounded but could not be clearly distinguished from post mortem damage. If the individual had been rendered incapable of ruling due to his injuries, he might have been ceremoniously relieved of his duties as noted previously in another study (37). Prior to the discovery of this grave, pieces of copper and gold foil, as well as a double iron gong and fragmented pots were found within the enclosure. A few gold beads, a hyena mandible and some lion teeth were found directly above the burial pit.

© University of Pretoria

The graves and their contents indicate that the individuals were of a high social status and have also provided clues as to the cultural affinity of these individuals, i.e. it appears as if their culture was very much like that of the Zimbabwean culture (4). The skeletal remains that were found at this site showed that the individuals were of African origin. The female was of tall stature which, after comparison with the average height of Venda women today, suggests that this individual may have been taller than most women in her community. It has been speculated that she may have had access to better food due to her royal position, or that she may have become a royal wife because of her tall stature (4).

The excavation of Thulamela was special as it also included the participation of members of the Venda and Tsonga/Shangaan communities that now reside there. After the skeletal analysis, the remains of both individuals were ceremoniously reburied by community members and only some samples were retained by the University of Pretoria. Bone samples were handled with gloves and stored in separate bags to minimise contamination with modern DNA.

2.2.1 Other relevant Iron Age sites

Samples representing the sites that are discussed in this section were selected on account of their temporal context (all were from the Iron Age) as well as on their level of preservation and availability. These sites are discussed in more detail below.

2.2.1.1 Happy Rest

This site near Schoemansdal in the Soutpansberg (Limpopo Province) was discovered during the early stages of a building project for a school. Ceramics found at this site were first described by Vaal in 1943, who likened them to Zimbabwean ceramics; Fagan later described them as similar to Gokomere ceramics (contemporaneous with phase I of Mapungubwe (38)). Happy Rest pottery has been classified to be part of the Kalundu tradition, also known as the Western stream of southern Bantu-speaker migrations and they represent some of the earliest farmer communities to come to southern Africa. It is important to note that Happy Rest pottery had been found both at the base (commoner dwelling) and top (elite homesteads) of Mapungubwe Hill (12), which might signify that

© University of Pretoria

these were important at the time and also definitely hints at the existence of some level of contact between Happy Rest and Mapungubwe people.

Further excavations at Happy Rest yielded eight graves, the contents of which were later lost. Two more graves were discovered later. The one individual was found buried in a kraal, in a vertically flexed position and the grave-pit had been lined with rocks, forming a semi-circle. This manner of burial is associated with Sotho/Tswana and Nguni individuals of high social rank (39). Broken ceramics and a lion's tooth were found associated with the body as well as an ivory bracelet on the arm. The individual was an adult male and carbon dating (using part of the femur) gave a date of 1310 ± 50 years BP. Osteometric measurements suggested some features shared with the Khoesan peoples. The skeleton was relatively well-preserved with some signs of dental wear. Indicators of tooth modification were noticed on the incisors (40). This practice has been noted in various regions in southern Africa, by both Bantu-speakers and Khoesan people, including an EIA site in Broederstroom, K2, as well as other more recent practices in the Western Cape and Namaqualand (9, 41, 42).

The second individual was an adult female, also buried in a sitting position and found to be associated with some ostrich eggshell beads, however, the grave appeared to have been disturbed. Osteometric analysis also indicated some physical features shared with the Khoesan peoples, it is thus thought that both individuals may have been Khoesan or Khoesan-related. The two Happy Rest individuals (labelled UP1 and UP2 for the male and female, respectively) are thought to have been contemporaneous with each other. Both individuals had features mostly associated with South African Black people, but some affinity to the Khoesan was also noted. The Khoesan peoples mostly utilised Stone Age technology. The archaeology of Happy Rest as well as the manner of burial of the two Happy Rest individuals is associated with Bantu speaking people, who mostly made use of iron technology. This therefore opens up interesting avenues of research regarding interactions of populations in the EIA (40).

2.2.1.2 Makgope

Makgope Hill is situated in the Bankenveld in present-day Marikana (Rustenburg district, Northwest Province). The remains of two male individuals (UP48 and UP49) were

© University of Pretoria

discovered there and both were buried in a tightly flexed, sitting position, facing south-westwards. They were also buried within two separate enclosures in the cattle byre (or kraal) and were associated with ceramics and what appeared to be the remains of sheep/goat sacrifices. The graves were dated to be from the 18th century, which was also the time that the Bakgatla and Bakwena established their chiefdoms (43). It is suggested that the cultural affinity of the Makgope Hill people was probably one of the early Batswana clans of the Bakwena Baphalane, whose capital is now at Ramokoka, east of the Pilanesberg. The Batswana are part of the eastern stream migrations or Urewe branch (and are mostly associated with Moloko ceramics).

Makgope Hill is part of a series of hills which include Malepe and Tierkop, which may have been part of Makgope's sphere of influence, based on Makgope Hill's relatively large size, the presence of granite knolls that surround the hill, a splendid men's court (kgotla), a spacious high status '*lapa*' complex, and the hill's elevation above the surrounding hills. The burial of both individuals within the kraal complex, as well as their sitting position, coincides with the manner of burial for high status individuals in Sotho/Tswana (and Nguni) communities (39). The general layout of the settlement also conforms to the central cattle pattern or CCP, which is common in Sotho/Tswana settlements. Makgope Hill flourished between AD 1716 and 1784 and was abandoned during the *difaqane* (a period of warfare and subsequent forced removal or submission of some ethnic groups by others in southern Africa) (43).

2.2.1.3 Kon-Stayt

This site is situated in the north of the Soutpansberg, and it is dated to the 13th century. This dating, well as associated ceramics and gold objects, is believed to be evidence of the sphere of influence of Mapungubwe/K2 (44). Charcoal from Kon-Stayt was actually dated to the period AD 1063-1235, which is consistent with the transitional phase between the Early and Middle Iron Ages (AD 900-1300).

The site consisted of some stone-walled enclosures, as well as remnants of copper and iron production sites. An extensive amount of glass beads were also found, as well a soapstone amulet, along with ostrich eggshell and iron beads. The presence of glass beads suggests that Kon-Stayt was part of the trade network that was under the control of Mapungubwe,

© University of Pretoria

which thus placed Kon-Stayt at some level of economic importance and also allowed for the interaction of people which resulted in similarities in culture. This is shown by the resemblances in ceramic style (45). Pottery found at the site was identified as a Venda-type ceramic; it was also dated to about AD 1700-1750. A cave-like opening, with pots containing animal bones at the mouth of the cave, was situated at the outcrop of what the excavators had named Section 1, and it was said to have been used for ritual purposes. Gold beads and copper objects were found in Section 2, including pottery associated with Mapungubwe. Mapungubwe/K2 pottery is thought to represent an early phase of the Leopard's Kopje ceramics, which ultimately arise from the Kalundu tradition, or western stream migrations (45). Undecorated pottery found at this site was thought to represent the gradual abandonment of this site, as well as the decline of Mapungubwe phase IV. Section 3 of the excavation resembles the CCP, which later changed to the Zimbabwe pattern as seen in relatively later sites such as at Mapungubwe phase III (12, 44). The culture at this site was therefore evolving from the CCP towards the Zimbabwe pattern.

The fragmentary, skeletonised remains of two individuals were found on the summit of a hill that appeared to be the main occupation site. The graves were also associated with pottery and UP77 is suspected to be a 'pot burial', however, in this case, the remains seemed to have been surrounded by pots instead of being contained within a pot. UP78 was fully articulated and found in a tightly flexed position. This individual was also associated with pottery and some black glass beads. Due to the poor preservation of the remains, UP78 might have been male but this could not be ascertained as the bones were fragmentary and the sex of UP77 could also not be ascertained as this individual was a juvenile (46). The remains of both individuals are in storage at room temperature at the University of Pretoria (Department of Anatomy bone collection). Due to the poor preservation of the bones, only those from UP77 were eligible for DNA analysis.

2.2.1.4 Nwanetsi

Nwanetsi is situated in the Kruger National Park and no extensive archaeological study has yet taken place on the site. The remains were exhumed as part of a rescue excavation. Only one potsherd was found associated with the skeletal remains of one individual, whose level of preservation was also quite poor. No other grave goods were present. No signs of occupation were found in the surrounding area at the time of the investigation. This

© University of Pretoria

skeleton was found near the Nwanetsi picnic spot, near the eastern border of the Kruger National Park. The body was on its back, although slightly turned on its left side. The upper part of the body seems to have been straight, and the lower part flexed. The remaining parts consisted of pieces of the neurocranium, which were reconstructed. All vertebrae, both scapulae and claviculae, many of the ribs, fragments of the pelvis and the complete right upper limb were preserved. The lower limbs were represented by a piece of fibula and the right foot. The individual was an African male of about 30-50 years. A single potsherd found at this site is thought to be of the Shingwedzi type, possibly made by ancestors of the present-day Venda people. Carbon dating analyses have estimated the age of the site to approximately AD 1401 (1414) 1428) (47). The remains are stored at the Department of Anatomy at the University of Pretoria, with the accession number UP98, and samples were taken for DNA analysis.

2.2.1.5 Tuli

A chance discovery at the border of Tuli Game Lodge and the Lorensa property in Botswana revealed the presence of human remains in a shallow grave at the base of a cliff (48). The site is about 20 km west of the Shashe-Limpopo confluence, where the borders of South Africa, Botswana and Zimbabwe meet. There was nothing above the remains that indicated the presence of a grave and it may have been exposed by scavenging animals. It was, however, evident that although the grave was inconspicuous, the individual was buried deliberately, placed in a tightly flexed position with the head oriented eastwards and the body was wrapped in animal skin which was secured with rope made of plant material. The remains were relatively complete, except for some phalanges and the right femur, as well as part of the skin wrapping, which may have been removed by scavengers. Hair, nails, skin and tendons were also still preserved as the warm, dry climate in the area, as well as the grave's location (away from direct sunlight) allowed for the process of natural mummification. This was the first naturally mummified human individual to be discovered in this region.

Osteometric analyses (limited as they were, so as not to damage the remains) and the presence of facial hair indicated that the individual was male and the texture of the hair indicated an individual of African origin. The presence of osteophytes and other features suggested that the individual was aged between 40 and 60 years. The remnants of what

© University of Pretoria

might have been a granary house were located just a few meters from the grave, as well as some glass beads, ostrich eggshell beads and fragmentary pots. Some potsherds were not decorated but a few could be likened to those from Khami. These arise from the western stream migrations, or the Kalundu tradition. It was thus concluded that the site might have been occupied between AD 1500 and 1800 (48).

Radiographic and CT scanning analysis of the remains has been conducted by Dr. Mosothwane, Prof. Rühli and Prof. Steyn prior to the sampling of soft and bony tissue for DNA analysis. The radiographic and sampling procedure was performed under controlled conditions with the necessary use of latex gloves and protective clothing to minimise modern DNA contamination. The samples for DNA analysis were stored in the ancient DNA laboratory at the University of Zürich, with the accession 'Bots' and the rest of the remains are kept at the University of Botswana.

2.3 Ancient DNA analyses in biological anthropology

Ancient DNA research refers to the retrieval and analysis of DNA from archaeological finds, museum specimens, fossil finds and other sources (49). The first successful sequencing of ancient DNA (aDNA) samples included specimens from a quagga (50) and an Egyptian mummy (51). These studies employed bacterial cloning methods and showed that most of the DNA retrievable in such studies is often of microbial origin. Endogenous DNA is often degraded, consisting of short sequences of low concentration (52). The difficulty in obtaining aDNA samples is based on the fact that some material is usually destroyed in order to gain access to the DNA; this has led to the reluctance of communities and museum curators in making samples available for DNA extraction.

A non-destructive method of DNA extraction from ancient bone samples has been suggested and even proven successful by Bolnick *et al.* (2012)*;* the authors were able to amplify mtDNA from 90% of the individuals tested, as well as analyze nuclear loci in 70% of the individuals (53). Because aDNA is difficult to obtain without destroying some or the entire bone sample (if a skeletal sample is used), extraction should only be conducted when necessary and with great care as the success of such an experiment is not always guaranteed. It is thus ideal that all the necessary morphometric analyses be completed prior to aDNA extraction, however, great care should be taken when doing this (discussed later).

© University of Pretoria

The aDNA research field has had a history of publications of questionable results and dubious findings that were later disputed as evidence of contaminating DNA. Ancient DNA research has now been standardised by what are called the 'nine golden criteria for ancient DNA research' highlighted by Cooper and Poinar in 2000. It has been stressed that these criteria must be strictly adhered to, as the best way to ensure that aDNA research becomes a reputable field in evolutionary biology. However, steps 6 through 9 are not deemed absolutely necessary. The criteria can be summarised as follows (54):

1) The work area should be physically isolated from other laboratories where aDNA research is not taking place, so as to minimise contamination by other sources.
2) Multiple extraction and PCR controls must be generated in order to detect low-copy number contamination. It is also suggested that positive controls should be avoided and all contaminated results be reported.
3) The older the sample, the more likely it is that the DNA extracted from it will be fragmented; therefore an amplicon of 1000 base pairs (bp) should be looked upon with caution as this is highly unusual for aDNA.
4) The results must be reproducible and PCR primers that target overlapping sites are recommended to increase the chances of detecting contaminating DNA.
5) Direct PCR products must be verified by cloning the amplicons, in order to determine the ratio of exogenous to endogenous DNA, as well as to detect damage induced changes and the presence of numts ('numts' are nuclear copies of mitochondrial DNA, and thus have the potential of obscuring phylogenetic relationships between species (55)).
6) The specimens need to be extracted and sequenced at separate laboratories in order to exclude the possibility of intra-laboratory contamination interfering with the results, especially when sequencing human DNA.
7) The biochemical preservation of the DNA must be assessed.
8) Quantification can be used to assess the possibility of contaminating DNA, especially when the amount of starting DNA is very low, and this is usually done by competitive PCR.
9) Associated remains sometimes present with similar DNA sequences, possibly due to post-mortem DNA changes, that compete with the fragmented DNA of an ancient sample. Faunal elements that were associated with the specimen of

© University of Pretoria

interest make for a good negative control when studying aDNA from an unrelated species such as a hominid.

2.3.1 Sources of aDNA and relative efficiency in DNA yield

Ancient DNA can be retrieved from a variety of samples, adding to its indispensability in ancient specimen research. If most of an individual's remains are unavailable for morphometric assessment, DNA can still provide some information. Sources of DNA include teeth, bone, tissue and others, such as hair, provided the hair root is available. Each of these sources has varying efficiency in terms of DNA that is of good enough quality to be amplified and sequenced.

Teeth and bone are believed to be some of the most reliable sources of aDNA since the degradation of DNA is hindered by its binding to hydroxyapatite in the mineral matrix (3). It has been shown that DNA quality is positively correlated with microscopic preservation of bone, but not necessarily with the age of the bone (56). The treatment of skeletal samples may differ in an attempt to retrieve DNA; shearing or powdering the bone sample is one way, and another could be drilling into the bone in order to gain access to the inner bone mass. A bone without lesions is preferable, so as to minimize contamination by external agents such as bacteria. Some scientists prefer to use small fragmentary rib samples as they are numerous in each individual and are of no significant morphological or palaeopathological importance (3).

Teeth are a good source of DNA as they allow for independent and numerous samples per individual, the results of which can be replicated. Unerupted teeth further reduce the risk of contamination, such as in teeth with dental caries. It is, however, important that the root is fully formed. Powdering the tooth yields more DNA, however, this method is highly destructive and may increase the amount of PCR inhibitors; a less destructive method would be sectioning the tooth to reach the pulp cavity, and this method also allows one to glue the tooth back together to preserve its morphology (57). Powdering the tooth is disadvantageous as it further degrades the already-compromised ancient DNA (58).

The best and least likely to be contaminated source of DNA from soft tissue is subsurface tissue. Desiccated soft tissue is known to produce better quality DNA as the desiccation

© University of Pretoria

process also shields the DNA from hydrolytic damage (3). It still remains susceptible to oxidative damage, however (59) (discussed below).

Nucleic acid extraction has also been achieved from coprolites; these also help provide a picture of the type of diet of the individual under study (60, 61). Another source of aDNA is hair, which is best when the hair root, which contains more DNA, is also available. The amount of DNA that is usable for PCR, however, is limited (62).

2.3.2 Factors that affect the quality/ yield of aDNA

One of the greatest challenges in aDNA studies pertaining to ancient human remains, is the possibility of contamination with modern human DNA, and this has led to a great deal of scepticism concerning aDNA research discoveries (52). Contamination with foreign DNA has resulted in misinterpretations, especially erroneously inferring close genetic relatedness of modern humans with some hominid samples. A famous example includes a report of the first ancient human DNA sequence of a 4 000 year old mummy (59), as well as one of a mitochondrial DNA sequence from a Mungo individual from Australia (63). In an attempt to curb the problem of contamination, certain characteristics have been found that can help distinguish modern from ancient DNA; the inevitable degradation processes that DNA undergoes post-mortem have provided markers that aid scientists in identifying modern from non-modern DNA. One sign of contamination is the production of multiple, differing sequences from cloned DNA. Most of the post-mortem DNA changes include double stranded nicks, as well as oxidative dinucleotide modifications, which contribute greatly to the difficulties incurred during enzymatic replication *in vitro* (64, 65).

2.3.2.1 Post-mortem biochemical changes in DNA

Molecular research on ancient samples has mainly been focused on DNA, however, the first effort to study ancient biomolecules was in fact performed on proteins (66). Unfortunately, the main protein in bone, collagen, consists of a repetitive primary structure, is encoded by multiple genes and is thus likely to be uninformative (67). Furthermore, ancient protein molecules are structurally heterogeneous as a consequence of post-mortem modifications (68). Analyses of amino acids can even help determine the level of preservation of a specimen by assessing the rate of amino acid racemisation. Nucleotide analyses can also verify the results obtained after DNA sequencing (69). DNA analyses

© University of Pretoria

have become common practice in the analyses of ancient biomolecules especially when proteins are degraded or uninformative (as mentioned above).

The cloning of DNA from an Egyptian mummy showed that both nuclear and mitochondrial DNA can survive for millennia (59), even though its integrity is inevitably compromised to some degree. DNA remains relatively stable in biologically active tissue, however, after death, this stability gradually decreases due to a lack of DNA repair mechanisms (70). Endogenous endonuclease activity is also known to occur after death, as well as depurination of adenine and guanine, thus compromising the DNA backbone. DNA damage therefore accumulates and nucleic acids are subjected to oxidative and hydrolytic damage; as a result, aDNA samples are often fragmented and difficult to amplify (51). Post-mortem DNA damage is often characterized by strand breaks, baseless sites, miscoding lesions and crosslinks, which then results in sequencing artefacts, as well as the preferential amplification of undamaged DNA, which is often contaminant DNA (51, 70).

Ancient DNA often contains a large number of CG-TA transitions (A-G or T-C being type 1 and C-T or G-A being type 2) due to the hydrolytic deamination of cytosine (and 5-methyl-cytosine) to uracil (65). Deamination thus occurs more rapidly on cytosine residues, converting it to uracil. If such a strand is used for PCR, deoxyuridine in the template DNA will be complementary base paired with deoxyadenosine in place of the guanine that should have been paired with the original cytosine. The deoxyuridine is thus a miscoding lesion. Similarly, adenine can be deaminated to hypoxanthine (65, 71).

The challenge in this situation comes from the ability to track the changes that are observed on a DNA sequence, i.e. is the nucleotide that we see a result of post-mortem damage that was carried on to the next strand by complementary base pairing (especially since it is impossible to know which of the two DNA strands was the parent strand and which was the daughter strand) or was it simply a mutation that occurred in life? Miscoding lesions are generated during PCR and do not hinder polymerase activity during amplification (59). However, mtDNA is known to have 'miscoding hotspots' where these miscoding lesions occur more often, and are thus not randomly distributed. More importantly, the distribution of these miscoding lesions in humans (and bovids) is similar to those occurring in ordinary evolutionary substitutions. Ancient DNA damage therefore generates sequence artefacts

© University of Pretoria

that resemble that of regular evolutionary changes (72). Other forms of DNA damage include oxidative damage that is caused by free radicals and manifests as changes in sugar residues as conversion of the pyrimidines thymine and cytosine into hydantoins. Oxidative damage, baseless sites, as well as intermolecular cross-links can block the activity of DNA polymerase during PCR (51). Efforts to minimize the negative effects of post-mortem DNA damage have been tested, and are aimed at enhancing the quality and quantity of aDNA for PCR and sequencing. One example includes the use of uracil-N-glycosylase (UNG) which removes the deamination products of cytosine; this aids in detecting the origins of sequence variation (71, 72). However, UNG has not been used extensively in aDNA research as the enzyme might destroy the little template DNA that is available (73). N-phenacylthiazolium bromide (PTB) seems to break the intermolecular cross-links that occur due to glycosylation end products, in a manner that has yet to be explained. Intermolecular crosslinks also act as PCR inhibitors (61). Figure 2.2 shows a DNA strand that is marked on the areas that are most prone to specific types of DNA damage (70). Figure 2.3 illustrates the different types of DNA damage. These modifications, such as miscoding lesions and baseless sites, result in the disruption of polymerase activity during PCR, causing ‘jumping PCR’ effects (discussed below) (52).

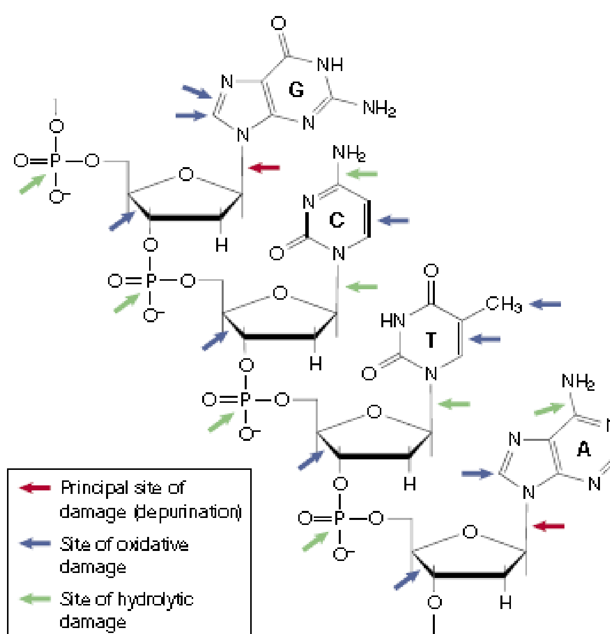


Figure 2.2. Specific areas of aDNA damage. Principle sites of damage are marked in red, sites of oxidative damage are marked in blue and the sites of hydrolytic damage in green. G, guanine; C, cytosine; T, thymine; A, adenine. (74) Modified with permission from (70).

© University of Pretoria

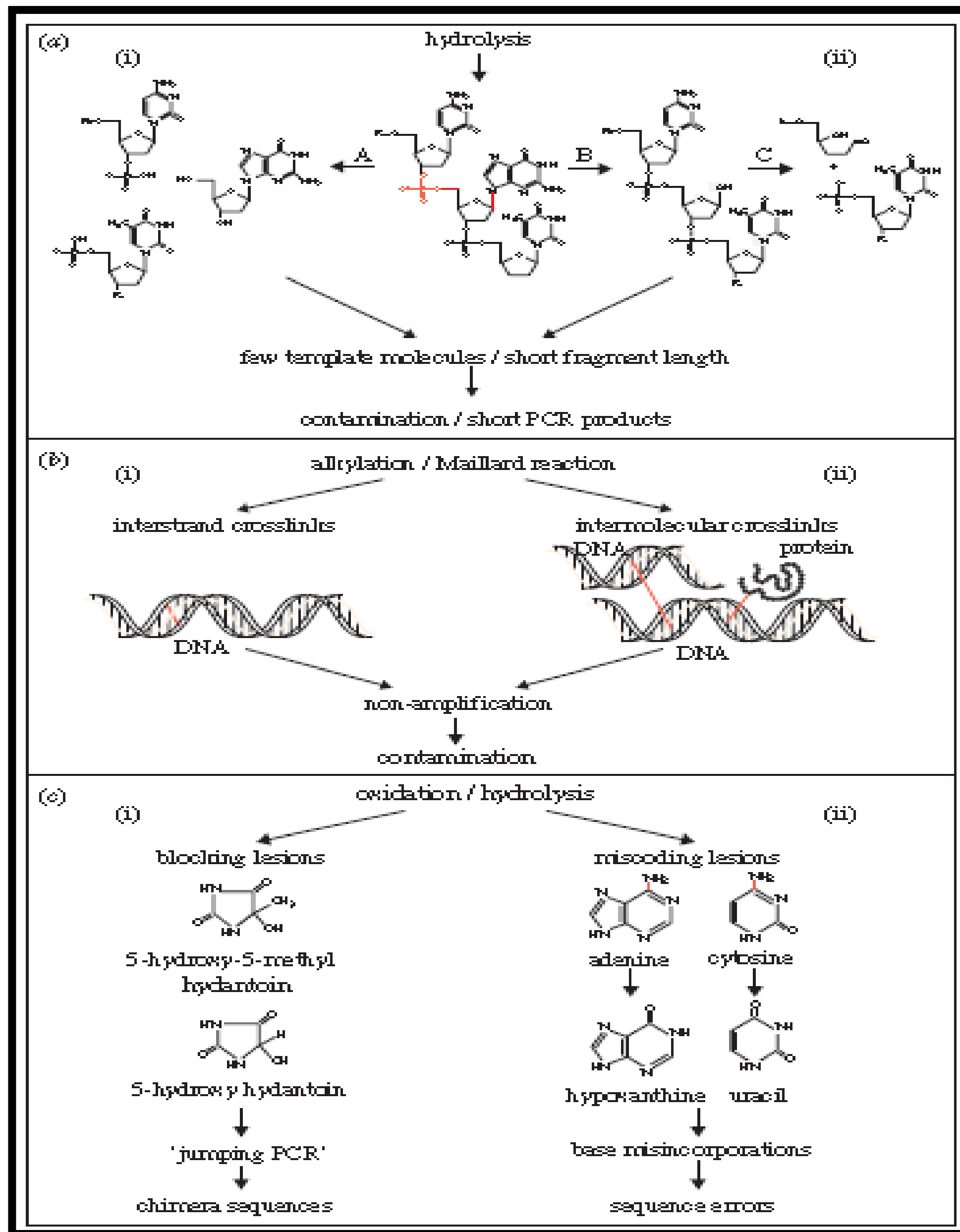


Figure 2.3. Post-mortem DNA damage in fossil remains (52). Damaged structures are marked in red: (a) single stranded nicks caused by hydrolytic damage. (i) Cleavage of the phosphodiester backbone. (A) (ii) Baseless site (AP site) caused by depurination. (B) Breakage of the sugar backbone follows, through β-elimination. (C) Strand breaks (b) Different types of crosslink formation (i) Alkylation causes inter-strand crosslinks (ii) Intermolecular crosslinks by the 'Maillard reaction'. (c) Oxidative and hydrolytic damages that result in: (i) Blocking lesions and (ii) Miscoding lesions (52).

© University of Pretoria

Other methods to help curb the post-mortem DNA damage problem is the use of high fidelity polymerase enzymes such as *Pfu* and *Taq*HiFi, as these are known to increase amplification efficiency and decrease sequence error (75). The rate of DNA damage, as well as its *modus operandi* are not well enough understood to create perfectly effective DNA repair enzymes *in vitro* (52, 76).

The phenomenon of 'jumping PCR' was first described by Pääbo *et al.* and is caused by lesions in the DNA template that cause the primer to jump from one template strand to another, usually more intact strand. The result is that a recombinant or mosaic DNA molecule, consisting of a mixture of different amplicons from different templates, is synthesized *in vitro* (51, 77). Figure 2.4 is an illustration of the manner in which jumping PCR occurs (51). The result of a jumping PCR reaction is that the authenticity of the DNA sequence is compromised, especially when analyzing autosomal nuclear DNA (which is present in two copies) of a heterozygous individual (51).

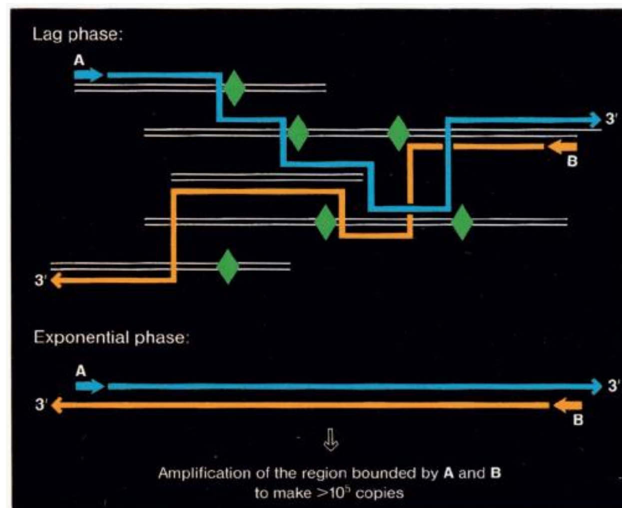


Figure 2.4. Schematic representation of how jumping PCR occurs. The two primers (A and B) use intact parts of five damaged templates resulting in the amplification of a mosaic product. In this amplification reaction, the template is so damaged that no molecules allow the DNA polymerase to continue directly from one primer site (A, blue) to the next (B, yellow); the primers will be extended during the first cycle of PCR up to points where lesions (green) or ends of fragments cause cessation of replication by the polymerase. In the subsequent cycles, the extended primers may anneal to other template molecules and then be further extended. After many cycles the two primers would have grown long enough that their 3' ends overlap and a full-length double stranded molecule is formed. This molecule will serve as a template for a conventional chain reaction (51).

© University of Pretoria

2.3.2.2 Environmental effects on post-mortem DNA

The conditions under which a sample containing DNA is stored have also been shown to affect the quality of amplifiable DNA, thus influencing the inevitable changes to the DNA that occur due to normal post mortem damage. Therefore certain conditions are able to either impede or accelerate these changes. In other words, DNA preservation is the main limiting factor of a successful aDNA research study. Hydrolytic damage of DNA is mainly caused by the presence of water which allows the growth of micro-organisms and the dissolving of the bone apatite. Microorganisms are able to metabolize organic matter in the bone, such as DNA and collagen. DNA oxidation is also accelerated by the action of UV light from the sun, and furthermore, acidic soil dissolves the calcium and phosphate in bone and this acidic property can be counteracted by alkaline environments or conditions. The ultimate place to preserve DNA is a cool, anaerobic, dry, dark, and slightly alkaline environment such as that found in caves (70, 74, 78). Cool, dry areas are also recommended for storage of samples to avoid post-excavation DNA damage and this is to be done shortly after the sample is removed from where it was found. It is also known that metal objects that contain copper may hinder microbial growth, so a piece of pottery or limestone at or near a bone sample could retard the microbial activity on the bone (79).

An ideal place of ‘storage’ for aDNA samples is under low temperature. The best quality aDNA samples have so far been retrieved from samples found in permafrost (subsurface layer below the soil that never thaws, usually in arctic and subarctic regions (74)); examples include mtDNA extracted from a mammoth of over 50 000 years old (80), and a bison of about 65 000 years old (81, 82). One of the most recent finds comes from a 4000 year old permafrost-preserved hair of an extinct Paleo-Eskimo individual. The authors were able to sequence 79% of the diploid genome and identified approximately 300 000 single nucleotide polymorphisms (SNPs) (83).

2.3.2.3 Preferable bone samples to use

When attempting to retrieve biomolecules from an archaeological sample such as bone, the sample’s condition is extremely important to ensure the success of such an experiment. A good indicator is that the bone should be compact, without openings, hard and heavy; compact bone is preferred over spongy bone which is porous and therefore open to

© University of Pretoria

contamination. Approximately 100 mg of pulverised bone sample should provide enough DNA for the analysis, however, the amount or concentration of DNA in that 100 mg sample depends on the preservation of the sample. As a control, it is recommended that the soil surrounding the individual that is being excavated should also be sampled as soil can contain compounds that inhibit the PCR reaction, such as humic acids that can bind tightly to DNA molecules (84, 85).

2.3.3 Molecular techniques in aDNA research

Demographic events such as migrations, bottlenecks and expansions have left imprints in the form of altered gene frequencies on the human genome, and because these imprints are transferred to succeeding generations, the modern human genome possesses a permanent record of our evolutionary past. Earlier studies on (human) evolution have been limited to a number of genetic polymorphisms, which were unfortunately few in number, uniform among populations and affected by natural selection; these included the use of blood groups and protein polymorphisms. The past two decades have witnessed the development of different methods that aid in deciphering population histories and provide much more information than what could previously be achieved (86, 87).

A major limiting factor to a successful aDNA experiment is the ability to extract DNA that is of good enough quality to be amplified by PCR, and as mentioned above, there are different factors that can inhibit DNA yield. It is inevitable that the specimen should be partially or wholly destroyed when attempting to access the DNA from inside the sample; the same is true for bones and teeth. It has therefore become common practice that all the necessary osteometric analyses be completed prior to DNA extraction.

The bone can be decalcified with EDTA, mechanical grinding of the bone is another method and also drilling an opening into the bone cavity is sometimes done. Some common forms of DNA extraction include the use of proteinase K/phenol-chloroform (proteinase K digests proteins in the sample), as well as silica based methods, which use guanidiniumthiocyanate (GuSCN) which can also lyse proteins and is also a chaotropic agent that allows for the binding of DNA to silica particles (88). A chaotropic agent disrupts hydrogen bonds such as those between water molecules and DNA, thus reducing the solubility of DNA, promoting the precipitation of DNA/silica particles. The advantage of the silica method is that it avoids the need to handle the toxic phenol-chloroform,

© University of Pretoria

however, silica is a PCR inhibitor and therefore washing must be done thoroughly before PCR. Although phenol/chloroform extraction methods result in a higher DNA yield, silica based methods produce better amplification success rates, as well as less PCR inhibition (89). The two above-mentioned protocols (phenol-chloroform and the silica method) are the most commonly used manners of DNA extraction, and may be further adapted.

2.3.3.1 DNA markers for aDNA experiments

As mentioned earlier, nuclear DNA (nDNA) is not always available for analysis in both forensic and ancient DNA analyses, thus making nDNA markers more difficult to retrieve by amplification. It has been shown that mtDNA is often present post-mortem. The human mtDNA molecule is double-stranded, circular, with 16 569 nucleotide pairs, and encodes the small (12S) and large (16S) ribosomal RNA, 22 transfer RNAs, as well as genes encoding 13 polypeptides that participate in the organelle's oxidative phosphorylation processes (90). The advantage of using mtDNA is that it occurs in multiple copies per cell, unlike nDNA, making mtDNA more available for retrieval and thus, mtDNA markers are more likely to be retrieved by amplification and further analysed. Mitochondrial DNA also has a high mutation rate, as well as no apparent recombination; therefore the difference between any two mitochondrial sequences represents only the mutations that have taken place since each sequence was derived from a common ancestor.

An increasing amount of data on human mtDNA variation has been accumulated in the past two decades, with greater and greater resolution (91). The other advantage of using mtDNA for a genealogical study is that since it is inherited from one parent, the mtDNA is haploid and therefore evades the complications caused by recombination processes of the diploid nDNA. The nuclear genome does, however, harbour much more historical information than does the mtDNA, but these studies have shown to be difficult. It is therefore important that one uses a combination of both mtDNA and nDNA markers wherever possible (92). A pioneering study on mtDNA variation showed a tree that depicted a separation between sub-Saharan Africans and non-African populations from about 200 000 years ago, thus supporting the recent origin of modern humans in Africa (93). In fact, African populations harbour the greatest amount of genetic diversity, especially with regard to mtDNA diversity, further substantiating the hypothesis that anatomically modern humans originated in Africa (94).

© University of Pretoria

Haplotypes are used to define groups of individuals that have similar genetic characteristics on the same location on their DNA, or in other words, the presence of similar polymorphisms on the mitochondrial genome that are identical by descent. These can help in deciphering human migration patterns, and specific haplogroups have been shown to be associated with certain ethnic groups. The definitions of haplogroups are highly dependent on phylogenetically stable regions of the DNA; however, most mutational hotspots lie within the control region of the mtDNA. As a result, much of the aDNA research (on mtDNA) has been conducted by using the well characterized and rapidly mutating hypervariable regions (HVR) of the control region, namely HVRI and HVRII. These rapidly mutating regions can only provide information with regard to recent population history (95).

High resolution methods for assessing mtDNA restriction site variation for PCR (Polymerase Chain Reaction) have been under development, and have made it possible to screen different mtDNA sequences from Europe, Asia, and America. These studies have shown that the above mentioned continents are defined by one or more continent-specific polymorphisms which make them excellent markers for determining geographical race. These mutations also appear to have arisen after the expansion of modern humans out of Africa (96, 97).

The phylogenetic analyses of these continent-specific polymorphisms thus define clusters of mtDNA haplotypes or haplogroups. For example, polymorphisms on specific regions of the mtDNA genome has allowed for the classification of Native American populations into four distinct haplogroups; A, B, C, and D, all of which are found in Asian people, which is to be expected as Native Americans descend from Asia. European populations are characterized more commonly by haplogroups H, I, J and K (90). Haplogroups were assigned by the order in which they were described, and follow an alphabetical order. This has no bearing on the chronology of their specific origin, in other words, haplogroup A is not the most common recent ancestor of all haplogroups, but instead, the first to be characterised.

African populations, known to possess the greatest amount of mtDNA genetic diversity, have a high frequency of haplogroup L, which appears to be the sister group of all continent-specific haplogroups, having arisen 100,000-130,000 years before present and

© University of Pretoria

perhaps before the expansion of *Homo sapiens* out of Africa. This haplogroup is virtually absent in non-African groups and was probably not carried out of Africa when the expansion occurred, however, isolated haplotypes that are within the L group have been observed outside Africa (94). Mitochondrial DNA has thus been a valuable tool in population genetics studies, deciphering population structure, origins and migration patterns. The Y-chromosome, which is transmitted through the paternal line, can likewise provide information on population history, but only for male lineages and apart from the recombining pseudo-autosomal region on the Y chromosome, DNA polymorphisms have also been found on this chromosome that can be used in similar ways as the mtDNA (98, 99).

Multi- and single copy DNA markers from the nuclear genome have been obtained from fossils dating back to the Pleistocene (100). The nuclear genome contains many different polymorphisms, such as SNP's (single nucleotide polymorphisms). The mutation rate of each nucleotide has been approximated to about 10^{-7} per generation, thus, unlike the rapidly mutating mitochondrial genome, nuclear DNA (nDNA) polymorphisms like these can provide information on the ancient history of a genome (101). In contrast, the more rapidly evolving nDNA markers include tandem repeats such as micro- and minisatellites that can be informative with regard to more recent population history (86). Although the hypervariable region of the mtDNA is useful for the construction of phylogenies, it cannot be reliable on its own and coding region information is therefore important, whereas the HVR is just as prone to mutations as it is to post-mortem damage (72).

Slowly evolving gene markers on the mtDNA genome have the advantage that, because they accumulate fewer mutations, they retain much of the genetic similarities in comparison and therefore provide information about the ancient history of a population. One such mtDNA marker is the cytochrome oxidase subunit II or COII, which is a functional protein coding gene and has already been used in phylogenetic studies on human origins. Because the COII is a slowly evolving gene, it can also be used to make inferences about the divergence time from a common ancestor, such as that between humans and chimpanzees (102).

Other gene markers for the determination of sex can be analysed. These include the Y-chromosome specific, sex determining region, or SRY gene which is normally expressed in

© University of Pretoria

XY males. The zinc-finger protein coding gene has also been implicated in sex determination, however, has instead been shown to be co-expressed with the SRY gene in human male brains (103). The ZF gene also presents as different variants, each of which are specific on either the X or the Y chromosome, and thus named ZFX or ZFY, respectively. The amelogenin gene presents similarly and is thus also named AMELX and AMELY. Normal XX females are thus expected to have two copies of both AMELX and ZFX in comparison to XY males. Both the ZF and AMEL genes have been used to determine sex in forensic cases (104, 105).

Genetic information can be deposited into GenBank (151) by scientists from any part of the world and the database contains huge amounts of molecular information from a plethora of organisms, including that of humans. This tool avoids the need to use living communities if the information is already available on GenBank, and this alternative therefore provides a quicker, more cost-efficient method of comparison, while still protecting the anonymity of participating individuals. Many authors have published data on sequenced regions of both the nuclear and mitochondrial genome, and GenBank is among the databases that contain these data. Mitochondrial DNA data are also well represented in these publications, as well as on databases such as GenBank, allowing for comparisons with database sequences and allowing for estimations of population affinity.

2.3.4 Amplification and sequencing of aDNA

2.3.4.1 PCR

The polymerase chain reaction, or PCR, is a method used to amplify small fragments of DNA for subsequent sequencing or other analyses. The reaction is done using two synthetic primers, matching either end of the DNA fragment that is to be amplified, the four deoxyribonucleoside triphosphates and a thermostable DNA polymerase that adds the nucleotides by complementary base pairing them to nucleotides on the parent DNA strand. Repetitive cycles of heating and cooling lead to a chain reaction that amplifies a segment of DNA exponentially (51). The advantage of PCR, apart from the fact that it is fast, is that it is an *in vitro* system that lacks repair mechanisms that sometimes occur erroneously in a bacterial cloning system. However, the polymerase used in PCR is not without errors; mis-incorporation of nucleotides occurs periodically and the rate of error varies between polymerases. If an error occurs in the initial stages of amplification, all the resulting

© University of Pretoria

molecules will have copied the mutant nucleotide/s. One way to curb this problem is by direct sequencing of multiple PCR products, or clones of PCR products, otherwise one can use a DNA polymerase with a proofreading exonuclease function (106).

Although PCR inhibition can be reduced by extracting the DNA using the silica method, it is difficult to rid the sample completely of inhibitors. Some strategies involve dilution of the DNA extract in an attempt to dilute the inhibiting elements in the sample. Another method is the addition of bovine serum albumin (BSA) that can bind to inhibitors or further digestion with proteinase K, collagenase or adding sodium hydroxide (NaOH) (2, 106). Ancient DNA is often degraded, resulting in the amplification of a mosaic product. Telenius *et al.* designed a method of increasing the quantity of amplifiable DNA, called degenerate oligonucleotide-primed PCR (or DOP-PCR) (107). The primers used for DOP-PCR are partially degenerate and anneal throughout the genome at low temperatures, allowing for general amplification. The initial amplification product is used for amplification with specific primers that target a region of interest. Although the success rate of this method is high, the chances of contamination are increased and this should therefore be employed with caution (106, 107).

An additional problem with the amplification of aDNA is that, due to its fragmented nature, primers tend to bind to each other more often than they bind to the DNA template, forming primer dimers that outcompete the target DNA and resulting in a low yield of target amplicons. This problem can be curbed by 'touchdown PCR' which systematically reduces the annealing temperature and allows for the preferential binding of primers to the target molecule. The annealing temperature is decreased by 1°C for every two amplification cycles, thus delaying the onset of non-specific primer annealing (108). Another use for touchdown PCR is when the exact sequence of the target DNA is not fully known, which would otherwise make it difficult to calculate the annealing temperature or primer for the species in question. The use of this method has also been employed when amplifying DNA from coprolites, especially when attempting to determine the animal itself and not the species of its prey. This is especially useful when the prey was a closely related species, then a highly specific touchdown PCR is necessary (108).

The polymerase chain reaction has thus far been used extensively in aDNA studies, and as mentioned above, PCR has been modified a number of times in order to improve on the

© University of Pretoria

quality of DNA for research. Recent advances in PCR technology pertaining to aDNA research is SPEX, or single primer extension PCR. This method uses a single biotinylated primer to target one strand of the aDNA template. Amplification then continues until it is stopped by the presence of a polymerase blocking lesion or fragmentation at the end of the template. SPEX therefore allows one to track the post-mortem changes in aDNA. The biotinylated DNA is then amplified by nested PCR (109). Comparison of sequence data produced from SPEX compared to standard PCR resulted in sequences of better quality from the SPEX samples. The disadvantage of SPEX is that it is laborious and time-consuming and has therefore not been widely used in aDNA research (73). Multiplex PCR is another strategy aimed at reducing the time and cost of standard PCR. It involves the use of a single reaction tube to amplify different target regions, as well as to sequence it simultaneously. This strategy has been employed by scientists to characterise the entire mtDNA genome of various species, including that of Mesolithic wild aurochs, an ancestor of domestic cattle (110).

2.3.4.2 Sequence generation and analysis

After amplification of DNA, it is often necessary to obtain a sequence of the DNA region for further analysis. Sequencing reactions are based on PCR; a single primer is used so that synthesis occurs in one direction and the strand extension can be terminated by the incorporation of a dideoxynucleotide (ddNTP). The ddNTP forms a small part of the reaction mixture (about 1%) and because it lacks the OH group that can react with the phosphate group on the adjacent nucleotide, chain extension cannot continue. When this method was first developed, the sequence of chain terminating nucleotides was read off from an electrophoresis polyacrylamide gel, and each lane on the gel represented one reaction tube containing one of the four ddNTPs. The method was first developed by Sanger and colleagues in 1977 and is commonly known as Sanger sequencing (111). The protocol has evolved to what is now known as automated Sanger sequencing, where each ddNTP is labelled with a specific dye molecule and after completion of the reaction, the sequence of dye-labelled chain terminating ddNTPs can be examined by a computer.

Recent developments have now seen the dawn of the so-called ‘next-generation sequencing’, which is roughly based on Sanger sequencing and has been used extensively in aDNA research (112). For sequencing, three widely used commercially available platforms are: 454 (Roche), Genome Analyser (Illumina/Solexa) and ABI-SOLiD (Applied

© University of Pretoria

Biosystems); these are now also being referred to a 'high-throughput', 'ultra-deep' sequencing or 'massively parallel' sequencing. What is remarkable about these technologies is that they can provide sequences at a massive scale, without the need for cloning and at a fraction of the cost of traditional sequencing (113). These platforms have become commonly used as a means of sequence production and analysis, except for the ABI-SOLID platform, which has thus far not been extensively used in aDNA studies (112).

The Roche/454 and Illumina systems use a similar rationale for producing sequences and only differ in the way that a target region is amplified as well as the sequencing chemistry, thus producing different throughputs. Sequences, also called reads, are generally produced following these three steps: (1) library preparation, (2) amplification and (3) sequencing. The library preparation involves the editing of nucleotides on both ends, as well as adding/ligating adapters on either end of the target DNA prior to PCR. The adapters consist on oligonucleotides of known sequence that are used to design complementary primers for the PCR reaction.

The Roche system amplifies the DNA in an oil emulsion (emulsion PCR or emPCR), while on the Illumina system, amplification takes place on a glass slide or flow cell, and molecules are amplified independently, forming clones or clusters. The sequencing reaction takes place separately when using the Roche protocol and it is based on pyrosequencing, while the Illumina platform performs the amplification and sequencing reaction simultaneously, thus saving time and reducing labour. Sequencing using the Illumina platform takes on a so-called 'sequencing by synthesis' (SBS) strategy that makes use of modified nucleotides. Figure 2.5 depicts the method in which sequences are generated on the Illumina platform (73). One's choice of either platform depends on the type of analysis, as well as the throughput desired.

Newer sequencing platforms are now under way, and have been referred to as the 'Third Generation Sequencing' platforms. These differ from Next Generation Sequencing methods in their sequencing chemistry and sample preparation, some even bypassing PCR amplification. The most impressive feature that is provided by these platforms is that they allow for sequencing from a single molecule, detects contamination and exogenous DNA, thus improving throughput in aDNA research (73). Orlando *et al.* used this technology to sequence the genome of a Pleistocene horse bone, and although they were able to do so

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

without the need for PCR or sequence enrichment, and were also able to identify the presence of contaminating DNA, the sequence depth was much too low (114). It is therefore suggested that Third Generation sequencing be used as a secondary venture, to help detect contamination, wherever possible, in anticipation of even better improvements. Thus far, Next Generation Sequencing avenues are used more often in aDNA research, as well as other forms of research.

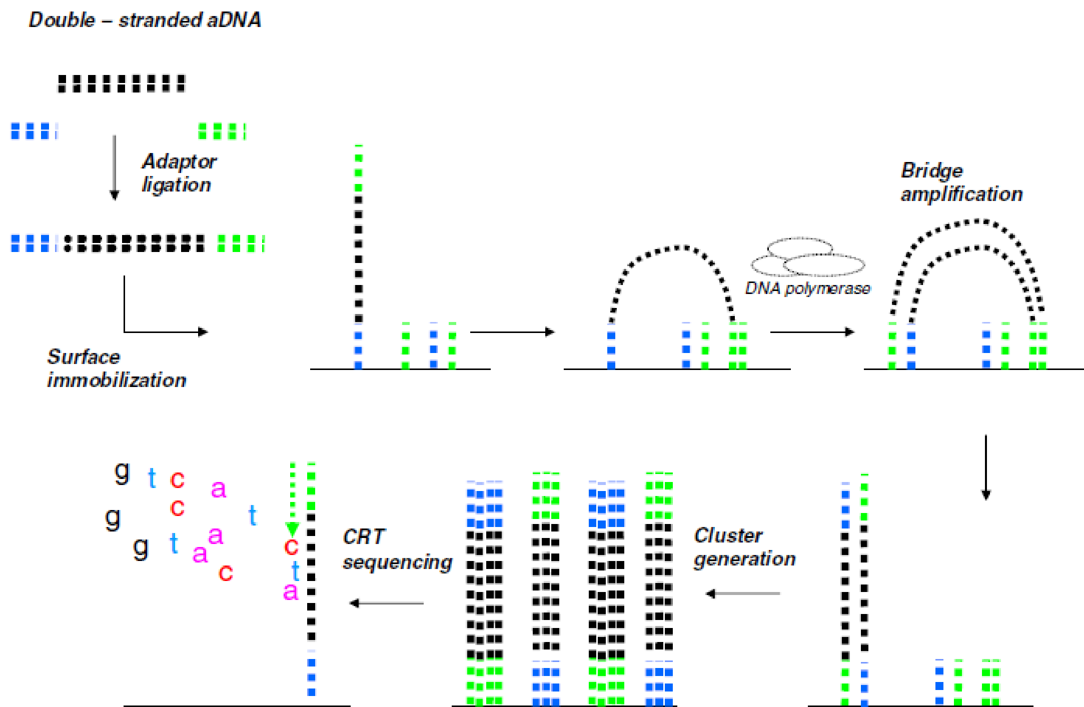


Figure 2.5. Illumina sequencing procedure.Double stranded DNA is modified on either end with blunt ended adaptors. The adaptors also provide anchorage for the DNA on a glass plate. DNA polymerase synthesises the sequences as a bridge joining the adaptors, eventually forming clusters (73).

For PCR and sequencing, one has to target a specific region of DNA in order to gain a closer look at the region, thus simplifying comparative analyses, and depending on the type of study, specific regions of the DNA are studied. To determine or verify the sex of an individual, markers on the X and Y chromosome can be sequenced, such as the sex determining region on the Y chromosome (SRY) which is only present in normal 46XY males. Other markers on the X chromosome can be targeted, and are expected to occur in two copies in females as opposed to males.

© University of Pretoria

Genetic sex verification can substantiate anthropological sex determination, but it can also diagnose XY sex reversal. Autosomal markers have been used to determine population affinity; these include SNPs as well as STRs (short tandem repeats). STR data in South African native peoples have been used to establish a correlation between linguistic and genetic similarity (115).

Due to their unilinear transmission, mitochondrial and Y-chromosome DNA data have become useful in determining ancestral relationships, and owing to their lack of significant recombination, mt-and Y-chromosome DNA can accumulate mutations along lines of descent, thus making them more valuable than autosomal DNA in phylogenetic analyses (116).

Mitochondrial DNA markers, such as the hypervariable region is useful as it does not only inform us of recent evolutionary changes (due to its high mutation rate) but mtDNA can also verify the population affinity of an individual by classing them into specific haplogroups. Once a specific region of DNA is sequenced, it can compared to others and these can be used to construct a phylogeny or a genealogy, to analyse the relationships among those genes.

2.3.4.2.1 Haplotype networks and phylogeography

In order to decipher evolutionary relationships and patterns, the use of statistics and mathematical modelling becomes increasingly important. This is because direct observations of past evolutionary events cannot be made and one must therefore use some sort of modelling of the past events.

The first step in creating a gene tree, based on a given gene region, for instance, begins in creating an alignment of all the sequences of that gene region that were obtained from the genomes of the organisms/individuals of interest. The sequences are aligned in a way that maximizes the degree of similarity between the sequences. There are therefore a myriad of alignment algorithms that aid one in doing this, including popular algorithms such as Kalign, MAFFT and Muscle. There are two methods of creating an alignment, namely global and local methods. Global methods cover the entire length of a sequence, whereas local methods cover parts of a sequence and only high scoring areas are considered. Database algorithms use the local strategy, while MSA (multiple sequence alignment) algorithms use the global strategy. One can employ both strategies if it is not known if the

© University of Pretoria

sequences at hand are of related species (117, 118). Aligned DNA sequences can also be assessed for the ratio of transitions versus transversions (R ratio) in order to estimate the evolutionary rate among related DNA sequences. Transitions occur more often than transversions and mtDNA often has higher R ratios (approximately 20) than nDNA (approximately 1). High R values are correlated with low sequence divergence and vice versa (143).

Rooted phylogeny construction usually requires the use of an outgroup, or reference species. The outgroup is usually chosen if it is a relative of the species under study, but not so closely related that it may show great similarity to the species of interest. The reference is excluded from within the taxonomic group, and is assumed to have branched out from that group sooner than the species within that taxon. The chimpanzee and Neanderthal have both been used in phylogenies involving humans or hominids; unrooted trees, however, do not require the addition of an outgroup. Phylogenies can be used to determine the genetic distance between taxa and can thus aid in deciphering the genetic relationships of different population groups.

While phylogenetic trees are good for summarising the genetic relationship between different species, they do not clearly distinguish the diversity that is present at the species level, owing to the fact that they do not account for hybridization, recombination or gene duplication; the so-called reticulate events, where ancestral genes persist and undergo recombination with ‘newer’ or mutant genes.

Phylogenetic trees are also strictly bifurcating at every node, thus assuming that each ancestor gave rise to two descendants that had no interaction with ancestral genes. Phylogenetic trees that represent different species are hierarchical because they result from reproductive isolation and fission between populations and therefore cannot account for overlapping gene pools. However, phylogenetic networks are much better at the species level as they account for all of these reticulate events, in which descendant genes coexist with ancestral genes and recombine to form reticulate relationships (119).

It has thus been suggested that traditional methods of creating these trees, such as Maximum Likelihood and Parsimony analysis make assumptions that violate the true processes that create genetic relationships within a population. And also, because individuals within a population diverge later than those between populations, the dataset

© University of Pretoria

from the within-species sample has fewer characters for the creation of a phylogenetic tree. A multi-furcating tree implies that a single ancestral gene can give rise to many descendant genes, which is often the case and is represented as loops in the network.

Genetic data, as mentioned before, possess a permanent record of population history, as well as the effects of processes such as genetic drift, selection, mate preference, etc. Scientists have worked tirelessly in using DNA markers to accurately trace population histories for a variety of reasons. The primary objective is not only to look at the genetic relatedness of individuals/species of interest, but also to discover the very origins of a population, i.e. the most recent ancestor as well as the genealogy of populations.

This quest has led to developments of methods that attempt to trace the origin of a gene as if to go 'back in time', such as the coalescent theory. It is therefore a retrospective study in population genetics and is aimed at tracing the origins of all taxa/individuals involved to a single common ancestor, using mathematical modelling that also take into account the probabilities of both forward and backward mutations at individual sites, and newer models of coalescence also account for recombination, gene flow and other processes. It is thus defined as 'the merging of ancestral lineages going back in time' (120). Coalescent theory seeks to find out the time that has elapsed since the last common ancestor (time to coalescence) and therefore provides a temporal basis on which to analyse genealogies.

The coalescent theory was independently developed by different scientists at slightly different times but most of the credit has gone to Kingman, who developed it between the years 1974 and 1982 (121-123).

The coalescent theory has had a myriad of uses, including disease gene mapping, and has helped scientists discover the possible origins of a disease gene, as well as the rate at which it is spreading and/or mutating (124, 125). The coalescent theory can be loosely explained as such: sampled lineages are viewed as picking or selecting their parents at random, as one goes back in time, and when two lineages select the same parent, their lineages coalesce. Gradually, all lineages coalesce to a single most recent common ancestor (MRCA). The rate of coalescence depends on how many lineages are sampled, as well as the size of the population under study (120).

© University of Pretoria

The haploid mtDNA and Y-chromosome DNA (the non-recombining portion) is ideal for studying coalescent processes, especially since their short coalescence times limits their utility in gene tree construction. For most phylogenies that are based on diploid DNA data, allele sequences are not separated and the heterozygous sites are considered as ‘missing’ or polymorphic. This creates complications when analysing data from closely related taxa. The network approach is therefore more efficient when analysing the genealogies of closely related individuals or groups (92).

In order to define mitochondrial haplogroups more clearly, there arose a need to have a standard reference sequence to which researchers could compare their study sequences. The Cambridge Reference Sequence of human mitochondrial DNA was first published in 1981 and was based on an individual of European descent (126).

The Cambridge Reference Sequence or CRS has thereafter been revised and corrected into a more accurate sequence (127). Positions of mutations in a sample sequence are thus numbered relative to the revised Cambridge Reference Sequence (rCRS). There are a myriad of phylogenetic tools that can aid one in assigning haplogroups to sample sequences, such as the online tool HaploGrep (153) (http://haplogrep.uibk.ac.at/). One can choose to compare samples to older or newer versions of the rCRS as they are uploaded onto HaploGrep. Mitotool (154), which is also freely available (http://www.mitotool.org/), works in a similar manner (and has other uses, such as detecting disease causing mutations) and assigns haplogroups according to phylotree (155) (www.phylotree.org).

Phylotree is an online tool with an overview of all human mtDNA haplogroups organised into a phylogenetic tree. This aids in the visualisation of haplogroups from the most recent common ancestor that lead to the African-specific haplogroups and their descendants both within and out of Africa. Both programs (HaploGrep and Mitotool) not only compare samples to the rCRS, but they also assign a haplogroup to each sample and name the positions on the sample sequence(s) that deviate from the rCRS and thus automatically detects variants.

The human mtDNA phylogeny that summarises the known haplogroups that exist around the world, is based mainly on HVRI data and wherever possible, also on whole mtDNA data. Coalescence times have been calculated using the same data (see Fig 2.5). Haplogroup L0 is thought to have been the earliest to diverge, approximately 150 kya and

© University of Pretoria

is further divided into L0a, L0d, L0f and L0k; L0 is also found almost exclusively among the Khoesan peoples. The L1’6 haplogroup diverged quite early, between 175 and 150kya, but did not start to diversify until at least 140kya, and therefore, the Khoesan specific haplogroup L0 presents with much deeper roots in comparison. The L1’6 haplogroup is the root of all other non-Khoesan mtDNA haplogroups that have spread and diversified throughout the world and also presents with many ‘shallow’ roots compared to the L0 clade (146).

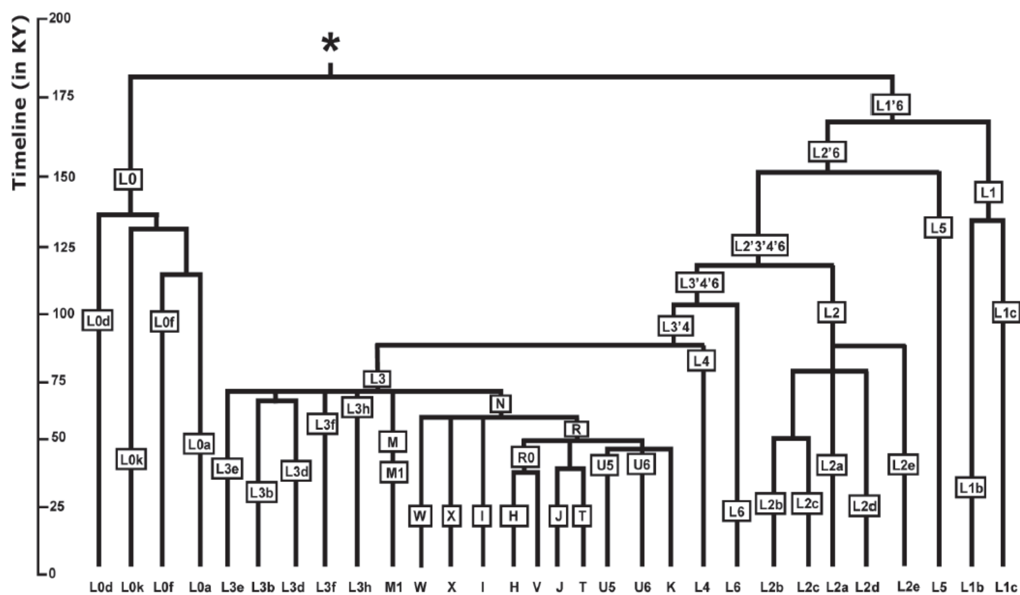


Figure 2.6: Human mtDNA phylogeny showing all known haplogroups and their coalescence times in thousands of years. The most recent common ancestor is marked with an asterisk (146).

The distribution of certain genes or alleles over a geographical area can provide clues as to the migration of individuals carrying that specific gene. This therefore brings forward the concept of phylogeography, whose use can also aid in deciphering population origins and dispersal patterns. It is the study of the past progressions or population movements that could be accountable for the observed geographic distributions of populations or individuals.

Since human populations have been characterised by haplotype, it is now possible to use a haplotype network, coupled with a phylogeography generating algorithm, to assess the distribution of these haplotypes. The nested clade analysis, or NCA, tests for the non-random distribution of haplotypes over a geographical area in one of three ways: statistical

© University of Pretoria

parsimony, likelihood networks or median joining networks (119). The statistical parsimony method first estimates the greatest number of differences between haplotypes, so that haplotypes with the fewest number of differences are placed closer together. It therefore assumes the least number of changes to have occurred between haplotypes. For the likelihood network method, nodes are treated as stochastic variables and branches represent correlations between these variables. The likelihood method is computationally cumbersome and thus not recommended for large data-sets.

Median joining networks use a parsimony criterion to combine the minimum spanning tree (a sub-graph of a connected undirected graph, organised into a tree that connects all vertices together). The median joining method is also rapid and requires the absence of recombination, which is ideal for analysing mitochondrial DNA sequences (119).

TCS, a freely available phylogenetic software (http://darwin.uvigo.es/software/tcs.html), uses the statistical parsimony method to estimate gene genealogies from sequence data (157). Geographical associations can be inferred using the program GeoDis (http://bioag.byu.edu/zoologycrandall_lab/programs.htm) which measures how widely distributed the individuals with a specific haplotype are (Dc) or how far these individuals are from each other (Dn), in other words, the higher the Dn, the longer the dispersal distance (158). The lengths of the lines joining the nodes on a haplotype network are proportional to the genetic distance between them.

GeoDis also uses a nested clade analysis to test for geographical associations by treating each sample location as a categorical variable, performing a permutation contingency test for clades at each level, and then a Chi-squared test is performed from the contingency tables in which the rows represent genetic clades and the columns are geographical locations. One can also use the geographical co-ordinates of each population and calculate the clade distance (Dc) as well as the nested clade distance (Dn) and if the haplotype tree has a root (for example, an outgroup) the oldest haplotype can be defined as the 'interior' with the younger haplotypes as the 'tips'.

Statistical significance is tested using a Monte-Carlo procedure which creates null distributions by randomising data from the contingency table for each clade and re-estimating the statistical significance for each randomised data set. One thousand permutations are recommended to make inferences at a 5% significance level (119).

© University of Pretoria

The combination of haplogroup assignments, organisation of these into a network and the calculation of coalescence, can help us to approximate the degree of relatedness between different populations, their genetic distance from each other, and the time to their most recent common ancestor.

2.4 Drawbacks of aDNA research

As mentioned above, the major setback with aDNA research is that some climatic conditions do not support the adequate preservation of DNA, thus leading to DNA degradation. Even if DNA can be extracted, amplification may be impossible. Without enough remains from an individual, i.e. if the bones are fragmented or very little tissue is present, and if there is no DNA that can be used, a small piece of bone from a single individual can be virtually useless.

Another big hurdle is contamination. Some very exciting reports on aDNA sequences have been completely discredited due to the discovery of contaminating DNA from either micro-organisms or even modern humans. Sources of contamination may be in the form of organisms that come into contact with the body after death, as well as human DNA from the researchers or anybody else that handled the remains.

Some authors have suggested ways to minimize the chances of contaminating samples with modern human DNA; these include the use of laboratories that are dedicated to aDNA research, with separate sections for each step of the protocol (i.e. one for DNA extraction, another for PCR, another for sequencing, etc.), the use of disposable laboratory clothing, as well as masks, gloves and caps, and disposable pipette-tips for example. The surface of the workbench is also treated with bleach, the lab equipment, workbench and reagents that do not contain DNA are also treated with UV radiation (128).

Although it is impossible to completely rid one's samples of contaminating DNA, it is good practice to try to minimize it wherever possible to avoid distortion of the results. Ginolhac *et al.* suggests a way to search for contaminating DNA in an ancient DNA sequence; the Perl Script mapDamage program detects typical modern DNA sequences (129). Typical sequence patterns that result from aDNA damage have been identified and can help distinguish damaged ancient sequences from modern contaminants; these include, short sequence length; an excess of cytosine to thymine (C-to-T) misincorporations at 5'

© University of Pretoria

ends of sequences, and complementary guanine to adenine (G-to-A) misincorporations at 3' termini, due to enhanced cytosine deamination in single stranded 5' overhanging ends (130, 131). While it is impossible to remove contaminating DNA from a sample, it is helpful to be able to recognise it.

2.5 Ethical considerations

Legal and ethical issues surrounding human DNA studies should always be taken into consideration, with special care to remain sensitized to social and religious beliefs that exist in different communities. Issues have been highlighted that pertain to the ownership of DNA samples, i.e. who owns the right to genetic information, who can it be transferred to, and if a disease is diagnosed from an ancestral sample, what are the ramifications for that individual's descendants? If they had never elected to know about such disease risks, what right has the researcher to release any information about the diagnosis?

The results of aDNA research has a possible impact on the legal, social and political state of affairs of living descendants, and also has the potential of offending certain beliefs that people may have about their ancestors (2, 106). Communities therefore need to be informed about any such work that could be performed on their ancestors, as well as the importance of such research. Ideally, the researcher would like to have the participation of communities and their leaders, such as in the case of the Thulamela remains. The community was fully involved in the process of retrieving, as well as the ritual re-burial of the individuals, who they believe are their ancestors (4).

Scientists have to be content with the proxy consent of the living descendants, as well as with the assumption that the decisions they make would serve the interest of the deceased (106, 132). National and institutional regulations have been put in place to secure the right of living participants in scientific research; however the debate on the rights of the dead has been a long and old one (even since Aristotle). It has been suggested that the beliefs of the deceased are to be dismissed if we do not know what those beliefs were (and if no descendants are known) (132).

In South Africa, the custodian of heritage related issues and items is SAHRA, the South African Heritage Resources Agency, which falls under the National Heritage Resources Act, No 25 of 1999. SAHRA promotes education and training to encourage public

© University of Pretoria

involvement in the discovery and understanding of historical items that can add to knowledge about much of South Africa's undocumented history. In addition, SAHRA is involved in protecting heritage sites and items, as well as the rights of communities that claim a connection to archaeological finds such as graves, human remains and land.

Overall, anthropologists and geneticists (or scientists in general) have a duty to obey a specific ethical code, one that primarily avoids offending the religious or sentimental beliefs of people and handling information in a way that will not stigmatize individuals or groups of people (for example if genetic information reveals a certain genetic disease that may be carried by living descendants). It is also imperative that informed consent be acquired from participants of such studies and individuals should be informed of the implications of their participation, as well as what may in future happen to the information they provide.

2.6 Conclusion

Ancient DNA analyses have become highly instrumental in the studies of modern human origins. The sequencing of the Neanderthal genome, for example, has provided further insight in this regard, as support for one model of modern human expansion out of Africa. The 'out of Africa' model (133), versus the 'regional continuity' model, for instance (134).

The Neanderthal genome sequence suggests that some genetic exchange between anatomically modern humans and Neanderthals did occur as some Neanderthal-specific alleles exist in modern Eurasians (134), however, the debate still continues. A resolution in this debate may help scientists to decipher more about our evolutionary past and population genetics has played a great role in clarifying long- and short-term evolutionary changes that have shaped modern populations of different species.

Population genetics studies have been performed in South Africa and have given insight into the patterns of gene flow and have shown that certain clusters of populations correspond with specific language groups, thus supporting the notion that linguistic similarities may be indicative of genetic and cultural relatedness (135). However, geographic distances do not clearly correlate with linguistic distances, which suggest that genetic and linguistic contact occurred before South African native populations became situated where they are presently (136).

© University of Pretoria

DNA studies pertaining to the genetic diversity of native and non-native South Africans have persisted over the years and have shown the effects of the country's past (and present) contact with people from other countries, so that we now know the maternal and paternal genetic contributions into certain population groups (115, 137, 138), for example, the maternal contribution of DNA from the Khoesan population in Cape Coloured people (139).

Given what is already known about the genetic substructure of Bantu-speaking people in southern Africa, there is still more to be deciphered. The purpose of this study was to clarify possible patterns of gene flow, as well as the probable interactions between groups based on combined data from ethnicity-specific markers and other lines of evidence, such as ceramics between population groups, with specific focus on the Thulamela remains.

The exchange of culture and technology is often accompanied by the exchange of genes. The cultural links between Thulamela and Zimbabwe and Venda is highly suggestive of a genetic relatedness. The information gained from this will add to the osteometric and archaeological evidence on the Thulamela remains. The information obtained should contribute to the data that have been captured on the origins of Bantu-speakers, their migration patterns, as well as how their culture may have evolved to what it is today.

© University of Pretoria

Chapter 3

Materials and Methods

3.1 Samples and procedure

After obtaining the necessary permissions, bone and/or tooth samples were collected from a total of nine individuals: one from Botswana (Tuli) and eight from South Africa (some soft tissue was also collected from the Tuli mummy). Except for the Tuli mummy, all were South African, from the north-eastern region of the country. It must be noted that apart from the Tuli mummy and Thulamela remains, none of the other samples were initially intended for DNA analysis and they were thus handled without protective clothing. The South African samples were all collected at different points in time and were stored at room temperature in the bone collection at the Department of Anatomy, University of Pretoria. Table 3.1 is a summary of the samples used in this study, the site where they were found, and their estimated age and accession numbers. Figure 3.1 is a map of the relative geographic location of the sites of interest.

Figure 3.1. Map of the eastern part of southern Africa with the relative locations of the sites of interest written in bold. (Edited from Huffman 2009).

© University of Pretoria

Table 3.1. A summary of the samples used.

(UP) accession no.	Sex	Name of site	Approx. Age of sample	Sample type	Label	Notes	Reference
UP1	Male	Happy Rest	AD 683-798	1 upper central incisor, 1 lower central incisor	1a 1b		(40)
UP2	Female	Happy Rest	AD 683-798	1 phalange, 1 vertebra	2a 2b		Unpublished
UP77	Indeterminate (juvenile)	Kon-Stayt	AD 1063-1074	1st mandibular molar, embedded	77a (1-4)		Unpublished
UP44	Male	Thulamela	AD 1421-1437	Upper left canine	44a (1-4)		(4)
UP44	Male	Thulamela	AD 1421-1437	Upper first molar	44b (1-4)		(4)
UP43	Female	Thulamela	AD 1525-1632	Left talus	43a (1-4)		(4)
Bot	Male	Tuli (Botswana)	AD 1400-1500	Mummified soft tissue and bone	B	First as part of test sample	(48)
Bot	Male	Tuli (Botswana)	AD 1400-1500	Bone	B	First as part of test sample	(48)
UP98	Male	Nwanetsi	AD 1401-1428	Vertebra	98a (1-4)	Wrapped in foil, marked 'DNA, do not touch'	Unpublished
UP98	Male	Nwanetsi	AD 1401-1428	Upper right molar	98b(1-4)		Unpublished
UP48	Male	Makgope	AD 1677-1768	Lower 3rd molar	49a (1-4)		(43)
UP49	Male	Makgope	AD 1670-1702	Upper right canine	49b (1-4)		(43)

3.2 Ethical and legal clearance

In keeping with the ethical standards that needed to be fulfilled, the remains at Thulamela were treated with care and community participation was encouraged during the excavation in 1996. Following the ceremonial reburial of the remains by the two populations that claim to have descended from the Thulamela individuals (the va-Tsonga and Vha-Venda), permission was given (also by the curators), so that parts of the skeletal remains were retained at the University of Pretoria, for DNA analysis. The South African National Parks authorities or SANParks as well as SAHRA were contacted directly in obtaining permission for this study (SAHRA permit Case ID 235, references: 9/2/269/0014, 9/2/255/0014, 9/2/269/0001). Furthermore, ethical clearance was given by the University of Pretoria Research Ethics Committee (reference number S127/2012).

© University of Pretoria

In addition, samples from other Iron Age archaeological sites were added for comparative purposes. These are the remains from two individuals excavated at Kon-Stayt, a Mapungubwe type site, as well as other sites, namely, Makgope (two individuals), Nwanetsi (one individual), Happy Rest (two individuals) and one from Tuli in Botswana, the first mummified individual to be found in this particular region. The Tuli mummy was first used as a test sample and later added to the sample collection after successful DNA extraction. It should, however, be kept in mind that no specific samples from these remains (except those from Thulamela and Botswana (Tuli)) were set apart for DNA analysis, and contamination with modern DNA may be a problem as they may not have been handled with gloves or any other protective clothing. Genetic data from anonymous living (Southern) African people were used for comparison (using a DNA database).

3.3 Methods

In summation, the DNA was extracted using the silica method which was then followed by real-time PCR amplification of the AMELX and AMELY nuclear markers, as well as standard PCR amplification of part of the SRY nuclear DNA marker and the HVRI, HVRII and the COII mitochondrial markers (53, 58). The mtDNA markers were amplified using PCR primers that target overlapping DNA regions. These regions were then sequenced by automated Sanger sequencing and analysed on an ABI analyser. Aligned sequences were then compared to the revised Cambridge Reference Sequence (rCRS), assigned to specific mtDNA haplogroups using Mitotool and then constructed into a haplotype network using the Splitstree computer algorithm (152). Pairwise comparisons of sequences were performed using MEGA6 (144). The flow diagram (Figure 3.2) is a summary of the basic steps that were followed; a more detailed description of the methods follow.

© University of Pretoria

Experimental flow diagram

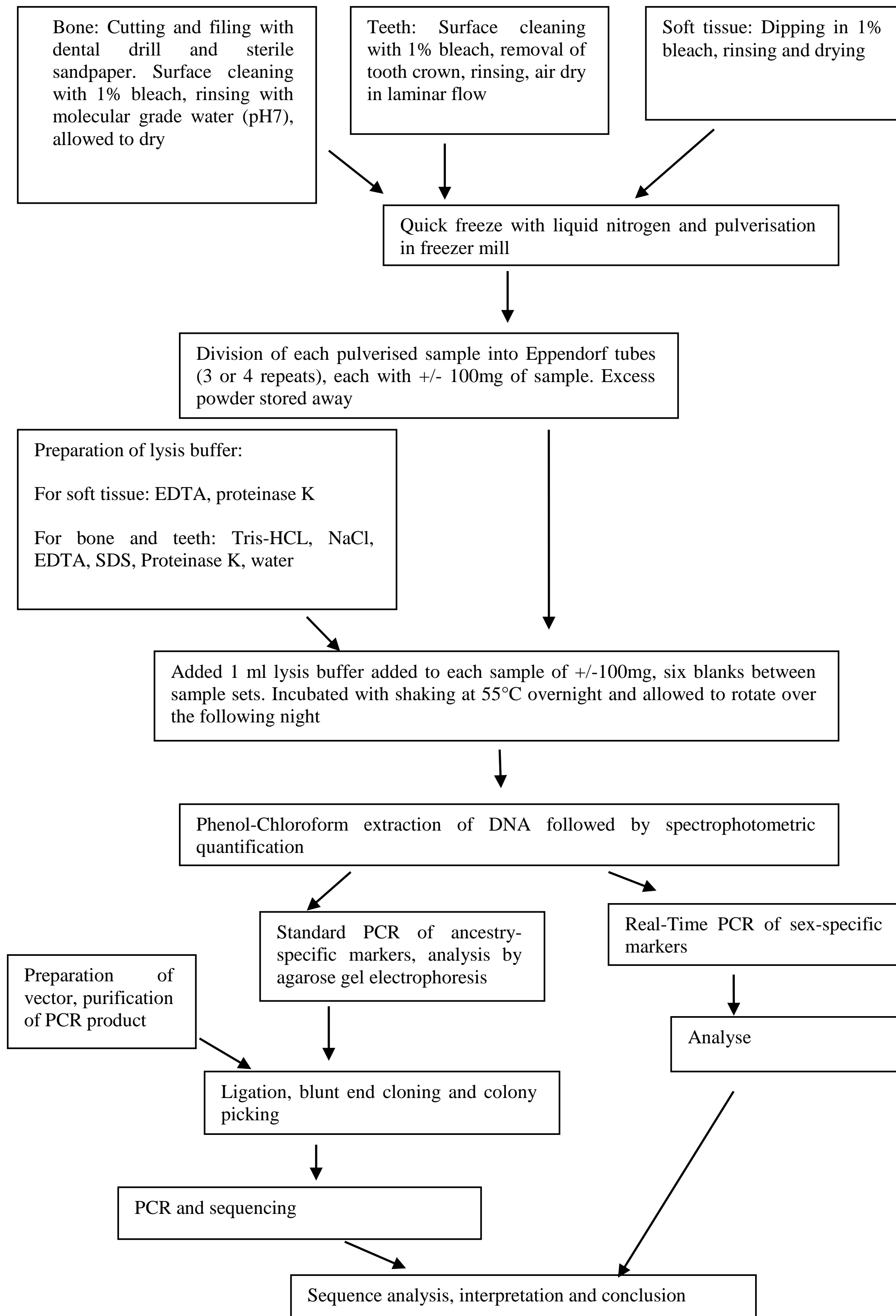


Figure 3.2. Summary of methods

© University of Pretoria

3.3.1 Contamination control

Ancient DNA extraction, amplification and sequencing were performed in a dedicated aDNA lab at the University of Zurich, Switzerland. Standard precautionary measures were followed; this included cleaning the sample surface with bleach solution, also, workbench surfaces were cleaned with bleach and UV radiation. Disposable laboratory ware was used as well as protective clothing (gloves, goggles, face masks, hair-nets and overalls). Separate rooms were used for extraction, as well as pre- and post PCR procedures and no modern DNA samples were allowed in the aDNA laboratory. To control for contamination, multiple negative controls, containing water instead of DNA sample were used for both the DNA extraction and PCR.

3.3.2 Sample preparation

Figure 3.3 shows the photographs of some of the individual teeth, prior to the process of cleaning and pulverisation for DNA extraction. Each tooth had been transported in its own separate bag, to minimize the possibility of contamination. Teeth that were still intact (i.e., not broken, especially at the root) were selected. Photography, surface cleaning with bleach solution and pulverisation also took place under sterile conditions and were followed by DNA extraction.

Figure 3.3. Photographs of some of the teeth used (and accession numbers) prior to pulverisation

© University of Pretoria

3.3.4 DNA extraction and purification

Extraction and purification of aDNA was performed according to a modified version of the method by Adachi *et al.* (140). Bone samples were dipped in 1% bleach solution, and then rinsed thoroughly in RNase-/DNase free distilled water. After allowing the bone to air dry, it was pulverised and air-dried again. The powdered sample was incubated overnight, with shaking, in a lysis buffer that included proteinase K at 55°C. DNA extraction was performed using the FAST ID DNA extraction kit (in accordance with the technical manual (72)) and with the phenol-chloroform method and the efficiency of both methods were compared. The phenol-chloroform DNA extraction method was as follows: After incubation of the sample with proteinase K and EDTA (lysis buffer), the samples were spun in a centrifuge for 5 minutes at 12 054 xg and the supernatant removed. To the rest of the sample, equal amounts (350 µl each) of phenol and chloroform were added, mixed by vortexing and then spun for 1 minute, in order to separate the excess protein in the sample from the DNA. The procedure was repeated once. Approximately 700 µl of chloroform was added to the sample, mixed and spun, and the entire supernatant removed. The sample was then added to a column tube that contains a silica mesh that binds DNA. A binding buffer was added, and after spinning, the excess binding buffer was disposed of. The sample was then treated with a wash buffer. Excess wash buffer was also removed. The column was then moved to a new, sterile Eppendorf tube and an elution buffer was added to elute the DNA from the column tube to the Eppendorf tube. This was done by adding the elution buffer (60 µl), allowing it to stay for approximately 5 minutes and then spinning in a centrifuge to allow the DNA to go through the column. The DNA extract was kept in the aDNA laboratory at 4°C.

3.3.5 DNA amplification, sequencing and analysis

3.3.5.1 PCR amplification

Following the extraction of DNA, as well as confirmation of the presence of genomic DNA by flourometric analysis (Qubit), the gene regions of interest were amplified by the polymerase chain reaction (PCR). A Hotstart Taq DNA polymerase (Qiagen) was used, to avoid non-specific amplification, because the polymerase is only active at high temperatures. The following reagents were used at the indicated concentrations, per PCR reaction: 2mM ddNTPs (Thermo-Scientific) and 10mM DNA primers (Thermo-Scientific). For the negative controls, 1µl of molecular grade water (Thermo-Scientific) was added

© University of Pretoria

instead of the DNA sample. The total volume was 25µl in each reaction. The pH of the reaction mixture was 7.3 and the buffer contained 10mM Tris-HCl (Thermo-Scientific) and 25mM $MgCl_2$ (Thermo-Scientific). A Biometra multi-block thermocycler was used for the PCR reaction. Table 3.2 is a summary of the mtDNA primers used (with forward and reverse sequences), the optimal annealing temperature (Tm) for each primer pair, as well as the expected size of the PCR product after amplification. The PCR temperature conditions were as follows: 98°C, 3 secs, (98°C, 10 secs, Tm, 20 secs, 72°C 15 secs) x 46 cycles, 72°C 5 min, 10°C, ∞.

Table 3.2. A list of the primers used, target region and expected PCR product size.

Primers used	Tm in °C	mtDNA/primer position 3'	Product size in bp	Reference
B1 B1-F CACCATGAATATTGTACGGT B1-R TTGCAGTTGATGTGTGATAG	56	HVRI 16131-16228	140	(156)
C1 C1-F AAGTACAGCAATCAACCCTC C1-R CTGTAATGTGCTATGTACGGTA	56	HVRI 16225-16325	141	(156)
D3 D3-F TACCCACCCTTAACAGTACA D3-R TATTGATTTCACGGAGGA	54	HVRI 16307-16406	136	(156)
U2e1 U2e1-F CACAGCCACTTTCCACACAG U2e1-R TCTTTGTTTTTGGGGTTTGG	63	HVRII 274-348	112	(Unpublished)
Ha Ha-F TCTGAGCCCTAGGATTCATC Ha-R TGATGGCAAATACAGCTCCT	63	COII 6938-7059	153	(156)
Hb Hb-F AGACATCGTACTACACGACACG Hb-R AAGCCTCCTATGATGGCAAA	63	COII 7013-7062	90	(Unpublished)
Amel Amel-F CCCTGGGCTCTGTAAAGAATAGTG Amel-R ATCAGAGCTTAAACTGGGAAGCTG	66	AMEL	106	(Unpublished)

3.3.5.2 Cloning

© University of Pretoria

The purpose of bacterial cloning prior to sequencing was to increase the amount of target DNA. When enough DNA is available for sequencing, each nucleotide produces an adequate signal for detection by the sequence analyser.

3.3.5.2.1 Vector preparation and ligation

PCR products were first subjected to ATP (adenosine 5'-triphosphate) and a T4 phosphorylate kinase which aids in the phosphorylation of the 5'-end of the target DNA, to create blunt ends on either side of the target DNA. Following this step, excess ATP and phosphorylate kinase were removed using the Qiaquick PCR purification kit. DNA concentration was then calculated with the Qubit assay for control purposes. A linear cloning vector (TOPO, Invitrogen) was then ligated to the target DNA with the aid of a DNA ligase (Invitrogen). Ligation was followed by transformation into *E.coli* bacterial cells, under sterile conditions. The *E.coli* cells were then plated on nutrient agar (Sigma-Aldrich) containing 5 g/l peptic digest, 3 g/l beef extract and 15 g/l agar. The antibiotic ampicillin was added for selection. Incubation was performed at 37°C overnight.

3.3.5.2.2 Colony picking and PCR amplification

After amplification of each PCR product in *E. coli* bacteria, six colonies were picked from each plate and each colony was placed in a separate PCR tube. Universal primers T3 and T7 were used to further amplify each product, after which they were visualised on an agarose gel (1% agarose in 1x Tris-acetate/EDTA buffer) and compared to a DNA marker (Thermo-Scientific) to approximate the size of the PCR products. The DNA sample was diluted 1:1 in loading buffer and electrophoresis was carried out at 80 V for 1 hour. The PCR conditions were as follows: 95°C, 10 minutes, (95°C, 30 seconds, 50°C, 30 secs, 72°C, 30 secs) x 26 cycles, 72°C, 10 min, 12°C, ∞.

3.3.5.2.3 Sequencing and analysis

Successfully cloned products were forwarded to GATC Biotech for automated Sanger sequencing. Visualisation of the DNA sequences was performed on an ABI analyser and electropherograms of the sequences were viewed and analysed using the program CLC Bio, from the CLC Bio bioinformatics software company. Sequence alignments were created for each PCR primer pair, and also, for each individual, contigs were created from the sequences of overlapping DNA regions. These alignments were compared to the revised Cambridge Reference Sequence (rCRS), through the online programme BLAST

© University of Pretoria

(Basic Alignment Search Tool) and the alignment was then assembled using the program MEGA 6 (144). Assigning each sequence to a haplogroup was done with the Mitotool program, which uses the rCRS to detect differences in the input sequences and thus assigns these to specific haplogroups. The computer algorithm, Splitstree, was used to create a network of the HVRI sequences generated. The transition: transversion ratio was also calculated using Mega 6.

© University of Pretoria

Chapter 4

Results

4.1.1 DNA extraction

DNA was successfully extracted from all samples, including the test samples from the Botswana Tuli mummy. A DNA extraction kit was used to extract DNA from the Tuli samples (both soft tissue and bone), in addition to the standard phenol-chloroform (PC) method. DNA quantity was measured using a digital flourometer (Qubit) and showed that a higher concentration of DNA was retrieved from the PC method than from the kit.

The kit produced DNA concentrations ranging from 0.269-0.300 ng/µl, whereas the PC method yielded between 4.95-10.5 ng/µl. It was therefore decided, that given the expected poor quality of aDNA, it would be more appropriate to continue with the PC method in further extractions. No DNA was detected in the control samples, as expected, indicating that no contamination occurred at this stage. The DNA samples were stored at 4°C in the aDNA laboratory.

4.1.2 PCR

PCR was used to amplify specific regions of DNA for each of the nine individuals, and again, this reaction was prepared in the aDNA laboratory. Standard PCR was used to amplify parts of the following DNA regions: HVRI, HVRII, COII and SRY. Real-Time PCR was used to amplify the Amelogenin region.

4.1.2.1 Standard PCR

The sizes of most of the amplicons were as expected. The negative controls from the extractions were also tested; none produced any amplicons, except those from the Tuli sample. The PCR results were analysed by agarose gel electrophoresis, and as can be seen in Fig. 4.1, very light bands could be observed. Also, occasionally, random amplification of non-specific regions occurred; this is also a characteristic of aDNA analysis. Most of amplicons were all of the expected size (see Table 3.1, Materials and Methods section) and were compared against a DNA low range size standard or ladder. After PCR, the samples whose concentrations ranged between 1.1 and 4.5 µg/ml (measured with the Qubit digital flourometer) were prepared for cloning.

© University of Pretoria

As can be seen on the images (Fig. 4.1a, image 2, lanes 6, 10, and 18) some PCR contamination of the negative controls from the Tuli mummy sample did occur. This is shown by the presence of PCR product in the lanes representing blank samples, where, instead of a DNA sample, water was added and thus expected to show as a blank space in the above-mentioned lanes. PCR was repeated to test for sporadic PCR contamination, and this was noted to be the case. DNA was re-extracted from the Tuli mummy for control purposes and PCR was repeated for the new extracts. Sporadic PCR contamination continued to occur.

The contaminated blank samples were forwarded for cloning and sequencing with the rest of the samples, in order to attempt to identify the source of the contamination. It is important to note that PCR contamination occurred only in association with the Tuli mummy and with none of the other samples and that at the initial stage of DNA extraction, no contamination was detected in any of the extracts.

Some primer dimer formation also occurred (such as in Figure 4.1a, image 2), due to the limited amount of template DNA. As can be seen in Figure 4.1b, gel number 1, some randomly amplified products did occur during the amplification process. This was, possibly due to the degraded nature of the template DNA, and cross amplification of environmental contaminants. While this is a common occurrence in aDNA PCR, this occurred rarely in this experiment and the majority of the images indicated that, generally, the primers used were specific, as most products were approximately the expected size. Those that differed too much from the expected size were not used further.

© University of Pretoria

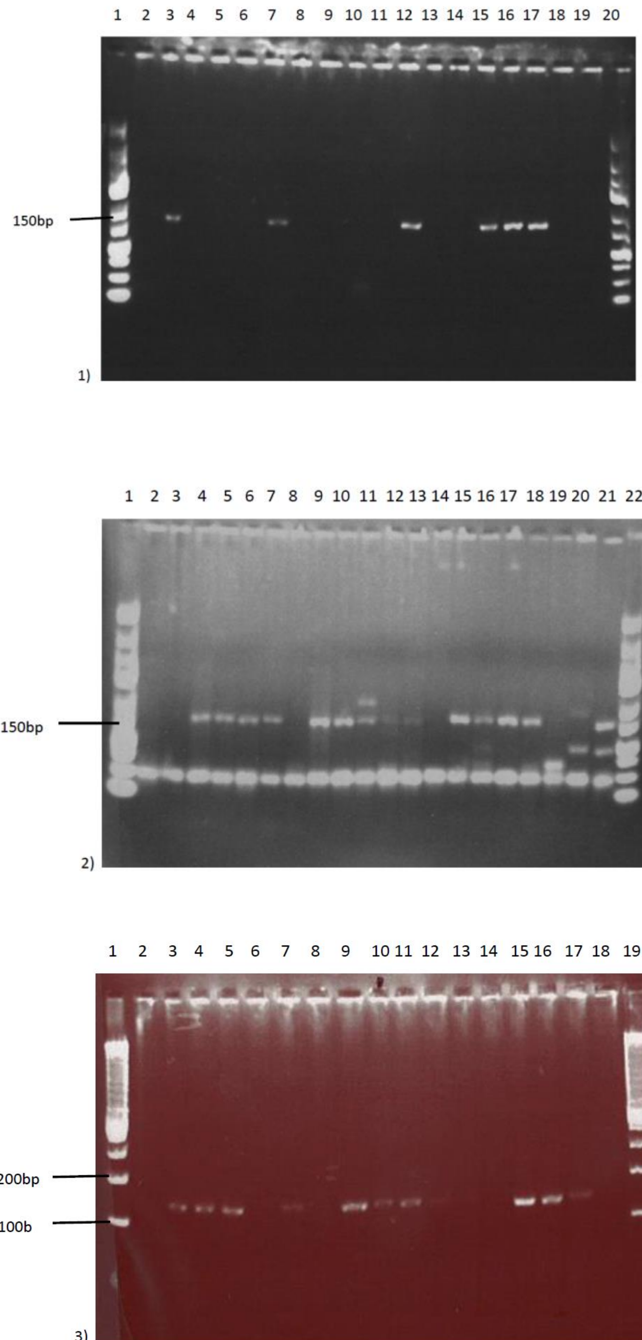
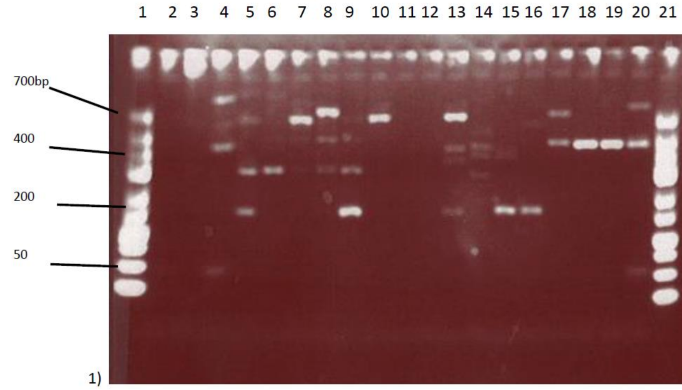
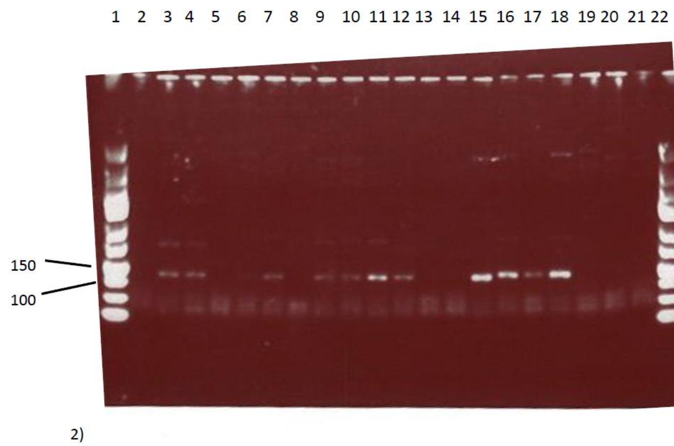


Figure 4.1a. Agarose gels after amplification of COII region using Ha primers. 2) Amplification of HVRI using C1 primers. 3) Amplification of HVRI using D3 primers. Order of samples, by well number: DNA low range (25-700bp) ladder (in image 3, a 1000bp ladder was used), negative control, F2, F4, F6, F negative, L2, L4, L6, L negative, F12, F14, F22, F negative, L12, L14, L22, L negative, DNA ladder. Note: F and L were from the Tuli mummy, the samples labelled ‘F’ were from the test extraction done according to the extraction kit while those labelled ‘L’ were extracted by the phenol-chloroform method.

© University of Pretoria





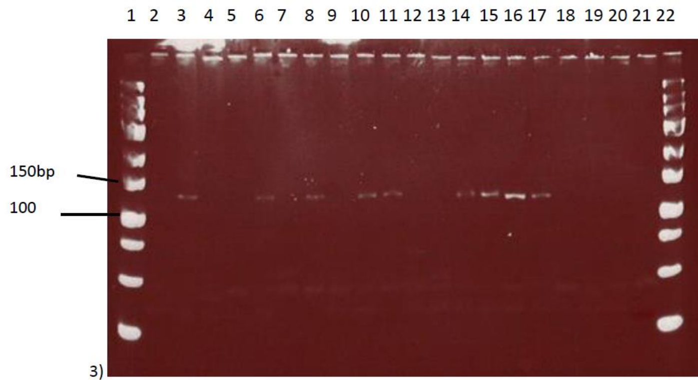


Figure 4.1b. PCR products after agarose gel electrophoresis. 1) Amplification of HVRI using B1 primers. 2) Amplification of COII using Hb primers. 3) Amplification of HVRII using U2e1 primers. Order of samples by well number: DNA low range ladder, negative control, 49b1, 43a1, 43a2, 43a3, 43a4, 98a1, 98a2, 98a3, 98a4, extraction blank 1, extraction blank 2, 98b1, 98b2, 98b3, 98b4, 77a1, 77a2, 77a3, 77a4, DNA low range ladder.

© University of Pretoria

4.1.2.2 Real-Time PCR

Real-time PCR was used for sex-specific DNA markers, using a single primer pair and two specific probes (AMELX and AMELY), in order to detect the relative copy numbers of these in each sample. For each individual, three of the four DNA extracts were used for testing and each experiment was done in triplicate. Substantial results would be if the AMELY was detected in at least two of the three samples. Only once was the AMELY detected from the individual UP77 (Kon-Stayt) and UP49 (Makgope) and the AMELX was also detected sparsely in the other samples. (Standard PCR was also performed for a 90bp region on the SRY, which was also repeated three times with different extracts, and still produced no conclusive results). It is therefore impossible to be absolutely certain as to what the genetic sex of any of these individuals were, as the amount of nuclear DNA was not enough to produce a signal (or amplification) in these experiments. For the sex of the individuals in this study, the results from the morphometric analyses that took place prior to the initiation of this project remain.

4.1.3 Cloning

4.1.3.1 Colony picking and PCR

After clones had been grown on an agar plate, single colonies were carefully selected for PCR. Six colonies were selected per plate and the products were analysed on an agarose gel. Each product was expected to be greater in size than the vector used, to indicate that an insert had indeed been ligated to the vector. The vector was analysed on an agarose gel with a DNA marker, so as to determine its size, which is approximately 1000bp (see Figure 4.2).

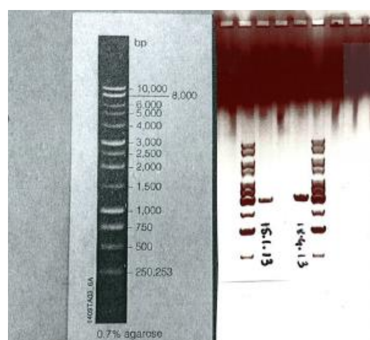


Figure 4.2. Gel electrophoresis image of the linear vector. The first and last lanes represent the DNA ladder, the 2nd and 4th lanes represent the same vector prepared on different dates (15.01.2013 and 18.04.2013). The 3rd lane is empty.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

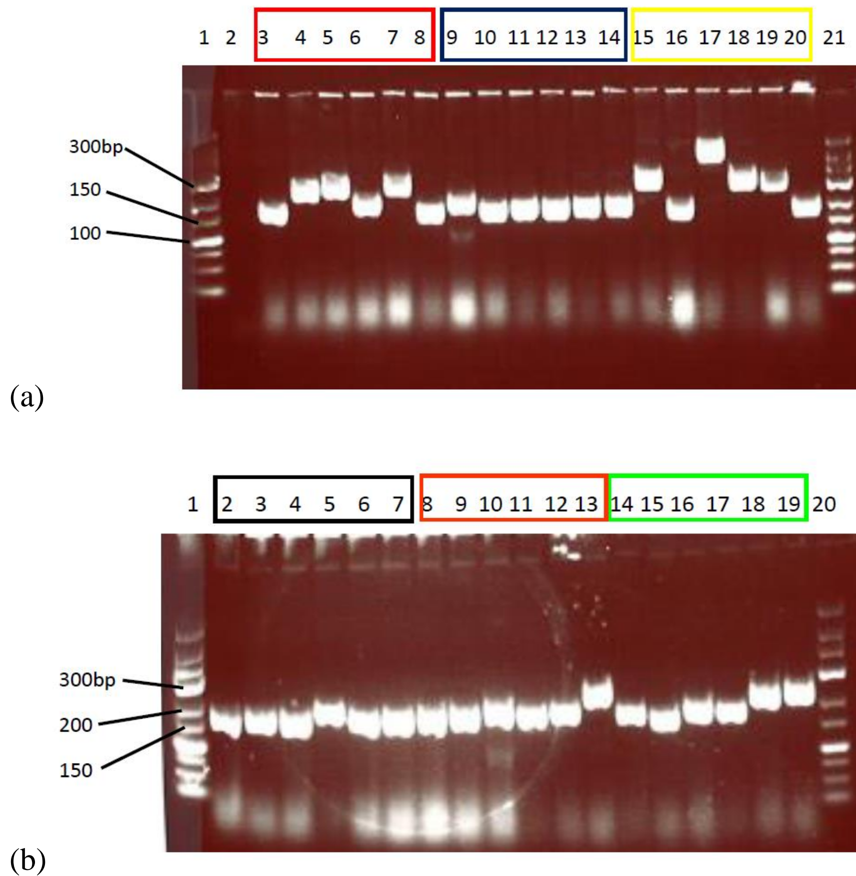


Figure 4.3. Gel electrophoresis images of PCR products after cloning. The first and last well in both images represent the DNA ladder, and well 2 of the first gel (a) represents a negative control. Samples from each plate are marked in a different coloured box; plate 1 is represented by the red box, plate 2 by the blue, and plate 3 by the yellow. In the second gel image (b), plate 4 is represented by the black box, plate 5 by the orange and plate 6 by the green. From each plate, six colonies were picked for PCR and sequencing.

The Tuli samples were the first to be cloned and sequenced. (The Tuli mummy samples were initially used to test the PC method against the FAST ID DNA extraction kit and after successful extraction, were kept in the study.) As can be seen in Figure 4.3, not all colonies produced the same sized PCR products. Figure 4.3 represents products that were cloned after inserting a 140 bp product into the ligated vector. It was thus expected that the products shown in these gels should be over 200bp in size, which was not always the case. Some PCR products were much larger than expected, such as in the image in Figure 4.3a, in wells number 3, 4, 6 and 15. This may have been a result of contamination or double inserts. Others were much smaller than expected, perhaps because the target DNA had not

© University of Pretoria

inserted into the vector. Products of a larger size were sequenced in order to find any possible contaminants.

4.1.4 Sequencing and network construction

After cloned products had been delivered for sequencing, it was noted that the majority of these had failed to incorporate the target DNA into the vector, therefore, the majority of the sequences represented only the cloning vector, and not the target DNA sequence. This was verified using the BLAST online database. Almost all of the samples that matched to the cloning vector (without target sequence) also matched to one human DNA region, namely the oxysterol 7αhydroxylase or CYP7B1, found on chromosome 8, which has been implicated in cholesterol metabolism (141). Other samples were contaminated with microbial DNA of various origins, such as *Gluconacetobacter diazotrophicus*, *Mycobacter canneti* and *Candida* species, which is normal for aDNA studies, especially since the samples had been buried in soil (or wrapped in animal skin in the case of the Tuli mummy) for a considerable amount of time, allowing plenty of access to microbial activity.

The reason for cloning prior to Sanger sequencing was to further increase the amount of target DNA after amplification by PCR, so as to produce accurate sequence signals. Ligation of the target DNA into the cloning vector was, however, problematic. When ligation occurs, after transformation, a reporter gene in the vector is expressed, for example, one that conveys resistance to an antibiotic in the growth medium. Only bacteria carrying the vector with a functional reporter gene would grow. Our experiment produced bacterial colonies, indicative of adequate transformation. A possible explanation is the presence of complementary sites on either end of the linear vector, causing the circularisation of the vector, with or without insert, and the successful expression of the resistance gene, which allowed the *E. coli* to grow.

From each plate of bacteria, six separate colonies were selected for sequencing and as mentioned before, when matched against the BLAST database, most sequences matched to the cloning vector and from each plate, nearly identical sequences were generated from each colony. Due to time constraints, it was then necessary to abandon the cloning step and instead, the remainder of the samples were sent for direct sequencing following PCR. Sequence signals generated from PCR products were weaker and included more background noise than those generated from cloned products, due to a lower amount of

© University of Pretoria

DNA in the PCR products. Figure 4.4 is an example of the electropherograms (after removal of primer or vector sequences) as viewed on CLC Bio; the first (Fig. 4.4a) is a sequence generated from a cloned product, the second, (Fig. 4.4b) from a PCR product.

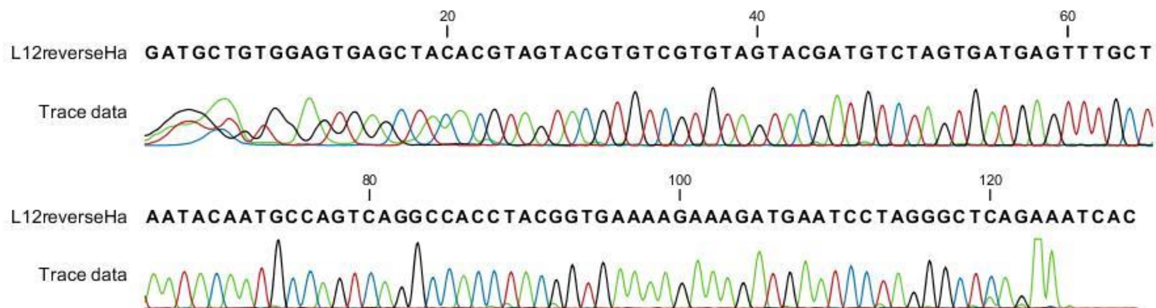


Figure 4.4a. Electropherogram of sequence generated from cloned product.

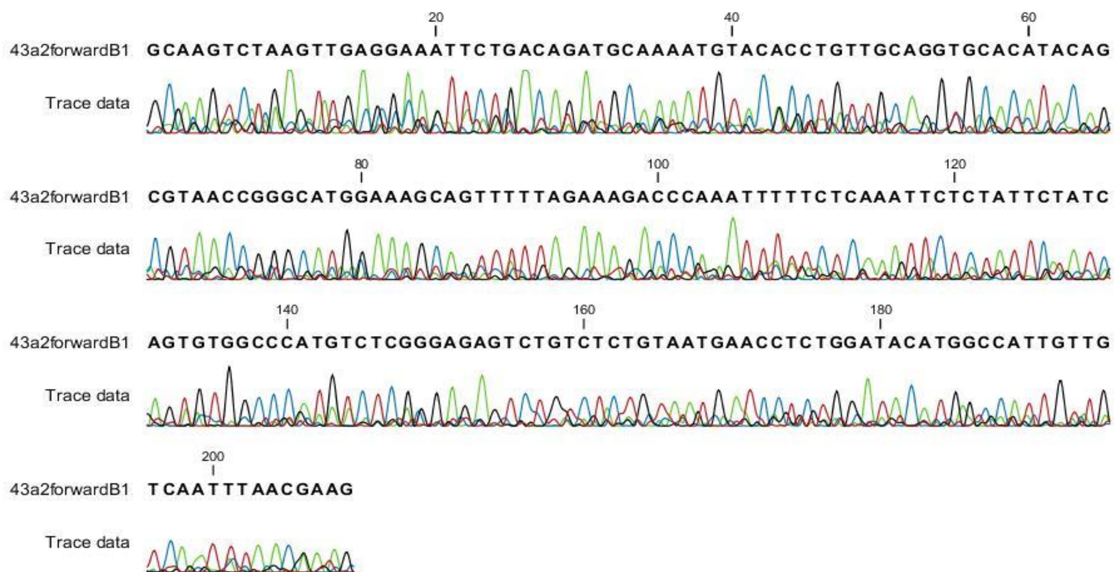


Figure 4.4b. Electropherogram of a sequence generated directly from a PCR product.

The sequences generated after cloning were trimmed in order to remove the sequence of the cloning vector, until the resulting sequence matched to a human mtDNA region on the BLAST database. Sequences generated from PCR products needed to be trimmed in order to remove the primer sequence on either end, and forward and reverse sequences of each sample was assembled to construct consensus sequences. Alignments for each gene region were constructed. Then the various sequences were assembled into contiguous sequences as overlapping primer pairs were used to amplify and sequence three smaller regions within each region. This is due to the degraded nature of aDNA, and also in order to

© University of Pretoria

increase the probability of obtaining PCR product at each amplification. Contiguous sequences were successfully assembled for much of the HVRI region, however, for the HVRII region, the same principle was employed, but none of the sequences were of good enough quality to be used for further analysis so this region was omitted from this study. The COII mtDNA region could also not be used for analysis as only the sequences from four individuals were usable, so that phylogenetic analysis of this region would not be informative.

Each sequence was verified as human mtDNA when submitted against the GenBank database. For each sequence, here referred to as a query, an alignment with one reference sequence, with which it had the highest percent similarity, is summarised in Table 4.1a and presented thereafter.

Table 4.1a. Summary of alignments with reference sequences

Individuals studied	Sequence obtained with primer set	Length of clean sequence (bp) in Blast search	GenBank reference of top hit	transition: transversion bias (R)
UP1 Happy Rest	B1/C1/D3	234	gi|656331056|gb|KJ185473.1|	0.92
UP2 Happy Rest	B1/C1/D3	141	gi|628823991|gb|KJ739542.1|	3.46
UP49 Makgope	B1/C1/D3	153	gi|300218137|gb|HM596745.1|	2.34
Bots Tuli mummy	B1/C1/D3	188	gi|380711000|gb|JF749371.1|	1.53
	Ha	86		
UP77 Stayt	Ha	78	gi|557954835|gb|KF672799.1|	2.31
UP98 Nwanetsi	C1	80	gi|481045765|gb|KC622118.1|	1.44
		46		
UP44 Thulamela	B1/C1/D3	80	gi|481045765|gb|KC622118.1|	2.07
	Ha/Hb	46		
UP43 Thulamela	B1/C1/D3	23	gi|18029545|gb|AF392099.1|	2.47

© University of Pretoria

UP1HappyRest

Homo sapiens isolate ZAM608 mitochondrion, complete genome

Sequence ID: gi|656331056|gb|KJ185473.1|Length: 16568

Score 369 bits(192) Identities 224/235(95%) Gaps 9/235(3%)

```
Query  1      TCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCA  60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  16110  TCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCA  16169

Query  61     ATCCACATCAAAACCCCCTCCCCATGCTTACAAGCAAGTACAGCAATCAACCTTCAACTA  120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  16170  ATCCACATCAAAACCCCCTCCCCATGCTTACAAGCAAGTACAGCAATCAACCTTCAACTA  16229

Query  121    TCACACATCAACTGCAA--------CCACCCCTCACCCACTAGGATACCAACAAACCTAC  172
              |||||||||||||||||        |||||||||||||||||||||||||||||||||||
Sbjct  16230  TCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTAC  16289

Query  173    CCACCCTTAACAGTACATAGTACATAAAGCCA-TTACCGTACATAGCACATTACA  226
              |||||||||||||||||||| ||||||||||| ||| ||||||||||||||||||
Sbjct  16290  CCACCCTTAACAGTACATAGCACATAAAGCCATTTATCGTACATAGCACATTACA  16344
```

UP2HappyR

Homo sapiens haplogroup H mitochondrion, complete genome

Sequence ID: gi|628823991|gb|KJ739542.1|Length: 16573

Score 266 bits(138) Identities 140/141(99%) Gaps 0/141(0%)

```
Query  1      AAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCA  60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  16210  AAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCA  16269

Query  61     CCCACTAGGATACCAACAAACCTACCCACCCTTAACAGTACATAGCACATAAAGCCATTC  120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 
Sbjct  16270  CCCACTAGGATACCAACAAACCTACCCACCCTTAACAGTACATAGCACATAAAGCCATTT  16329

Query  121    ACCGTACATAGCACATTACAG  141
              |||||||||||||||||||||
Sbjct  16330  ACCGTACATAGCACATTACAG  16350
```


© University of Pretoria

UP49Magop

Homo sapiens haplogroup L2abcd mitochondrion, complete genome

Sequence ID: gi|300218137|gb|HM596745.1|Length: 16571

Score 285 bits(148) Identities 152/154(99%) Gaps 0/154(0%)

```
Query  1     TGATGGCAAATACAGCTCCTATGGATTGGACATAGTGGAAGTGAGCTACAACGTAGTACG  60
             |||||||||||||||||||||| ||| |||||||||||||||||||||||||||||||||
Sbjct  7073  TGATGGCAAATACAGCTCCTATTGATAGGACATAGTGGAAGTGAGCTACAACGTAGTACG  7014

Query  61    TGTCGTGTAGTACGATGTCTAGTGATGAGTTTGCTAATACAATGCCAGTCAGGCCACCTA  120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  7013  TGTCGTGTAGTACGATGTCTAGTGATGAGTTTGCTAATACAATGCCAGTCAGGCCACCTA  6954

Query  121   CGGTGAAAAGAAAGATGAATCCTAGGGCTCAGAA  154
             ||||||||||||||||||||||||||||||||||
Sbjct  6953  CGGTGAAAAGAAAGATGAATCCTAGGGCTCAGAA  6920
```

Bots

Homo sapiens isolate MARBr592 control region, partial sequence; mitochondrial

Sequence ID: gi|380711000|gb|JF749371.1|Length: 1192Number of Matches: 2:

Score 325 bits(169) Identities 184/189(97%) Gaps 1/189(0%)

```
Query  98   AAAACCCAATCC-CATCAAAACCCCCCCCCCCAGGCTTACAAGCAAGTACAGCAATCAA  156
            |||||||||||| ||||||||||||||||||||| |||||||||||||||||||||||||
Sbjct  155  AAAACCCAATCCACATCAAAACCCCCCCCCCCATGCTTACAAGCAAGTACAGCAATCAA  214

Query  157  CCCTCAACTATCACACATCAACTGCAACTCCAAAGTCACCCCTCATCCATTAGGATACCA  216
            ||||||||||||||||||||||||||||||||||| ||||||||| ||| ||||||||||
Sbjct  215  CCCTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCA  274

Query  217  ACAAACCTACCCACCCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACA  276
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  275  ACAAACCTACCCACCCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACA  334

Query  277  TTACAGTCA  285
            |||||||||
Sbjct  335  TTACAGTCA  343
```


© University of Pretoria

Range 2:

Score 167 bits(87) Identities 87/87(100%) Gaps 0/87(0%)

```
Query  1      CACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGGGGTCCCT  60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  331    CACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGGGGTCCCT  390

Query  61     TGACCACCATCCTCCGTGAAATCAATA  87
              |||||||||||||||||||||||||||
Sbjct  391    TGACCACCATCCTCCGTGAAATCAATA  417
```

UP77Stayt
Homo sapiens isolate Mozambique Moz74 mitochondrion, complete genome
Sequence ID: gi|557954835|gb|KF672799.1|Length: 16567

Score 135 bits(70) Identities 76/79(96%) Gaps 0/79(0%)

```
Query  6      TGTTGGTATCCTAGAGGTTGACGGGTGGCTTTGGAGTTGCCGTTGATGTATGACAGTTGA  65
              |||||||||||||| |||||| |||||||||||||||||| |||||||||||||||||||
Sbjct  16281  TGTTGGTATCCTAGTGGTTGAGGGGTGGCTTTGGAGTTGCGGTTGATGTATGACAGTTGA  16222

Query  66     GGGTTGATTGCTGTACTTG  84
              |||||||||||||||||||
Sbjct  16221  GGGTTGATTGCTGTACTTG  16203
```

UP98Nwanet
Homo sapiens isolate KGA020 mitochondrion, complete genome
Sequence ID: gi|481045765|gb|KC622118.1|Length: 16569Number of Matches: 2

Score 91.1 bits(47) Identity 72/82(88%) Gaps 1/82(1%)

```
Query  64     AGCATGGGGGGGGGTTTTTGATGTGATATTGGGTTTTTAGGTCCTACGGGGGGTCAATTT  123
              |||||||||||||||||||||||||  |||||||||||| || |||| || |||||| | 
Sbjct  16198  AGCATGGGGGGGGGTTTTTGATGTGG-ATTGGGTTTTTATGTACTACAGGTGGTCAAGTA  16140

Query  124    TTTAGGGTACCGTACAATTTTC  145
              |||| ||||||||||||| |||
Sbjct  16139  TTTATGGTACCGTACAATATTC  16118
```


© University of Pretoria

Range 2:

Score 62.2 bits(32) Identity 42/47(89%) Gaps 0/47(0%)

```
Query  1      TTGCAGGTGATGTGTGATAGTTGAAGGTTGAATGGTGGACTTGGTTG  47
              |||||| ||||||||||||||||||||||| || || ||||| |||
Sbjct  16247  TTGCAGTTGATGTGTGATAGTTGAAGGTTGATTGCTGTACTTGCTTG  16201
```

UP44Thulam
Homo sapiens isolate KGA020 mitochondrion, complete genome
Sequence ID: gi|481045765|gb|KC622118.1|Length: 16569Number of Matches: 2

Score 91.1 bits(47) Identities 72/82(88%) Gaps 1/82(1%)

```
Query  64     AGCATGGGGGGGGTTTTTGATGTGATATTGGGTTTTTAGGTCCTACGGGGGGTCAATTT  123
              ||||||||||||||||||||||||  |||||||||||| || |||| || |||||| |
Sbjct  16198  AGCATGGGGGGGGGTTTTTGATGTGG-ATTGGGTTTTTATGTACTACAGGTGGTCAAGTA  16140

Query  124    TTTAGGGTACCGTACAATTTTC  145
              |||| |||||||||||||| |||
Sbjct  16139  TTTATGGTACCGTACAATATTC  16118
```

Range 2

Score 62.2 bits(32) Identities 42/47(89%) Gaps 0/47(0%)

```
Query  1      TTGCAGGTGATGTGTGATAGTTGAAGGTTGAATGGTGGACTTGGTTG  47
              |||||| ||||||||||||||||||||||| || || ||||| |||
Sbjct  16247  TTGCAGTTGATGTGTGATAGTTGAAGGTTGATTGCTGTACTTGCTTG  16201
```

UP43 Thulamela
Homo sapiens isolate L6 mitochondrial control region, hypervariable segment 1, partial sequence

Sequence ID: gi|18029545|gb|AF392099.1|Length: 495

Score 44.9 bits(23) Identities 23/23 (100%) Gaps 0/23(0%)

```
Query  1    CTGTAATGTGCTATGTACGGTAG  23
            |||||||||||||||||||||||
Sbjct  346  CTGTAATGTGCTATGTACGGTAG  324
```

The sequences were then each aligned against the rCRS to analyse which region each of them corresponded to, so as to verify whether the target sequence had been obtained for each individual. The alignments are presented here. Table 4.1b is a summary of where each

© University of Pretoria

sequence falls, relative to the rCRS and Figures 4.5a and 4.5b is a visual representation of query sequences against the rCRS. The vertical lines show the part of the alignment used for the construction of a network.

UP1HappyR
Sequence ID: lcl|52329Length: 226

Score 368 bits(199) Identities 224/234(96%) Gaps 9/234(3%)

```
Query  16112  CACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAA  16171
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2      CACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTACATAAAAACCCAA  61

Query  16172  TCCACATCAAAACCCCCTCCCCATGCTTACAAGCAAGTACAGCAATCAACCCTCAACTAT  16231
              ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Sbjct  62     TCCACATCAAAACCCCCTCCCCATGCTTACAAGCAAGTACAGCAATCAACCTTCAACTAT  121

Query  16232  CACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACC  16291
              ||||||||||||||||        ||||||||||||||||||||||||||||||||||||
Sbjct  122    CACACATCAACTGCAA--------CCACCCCTCACCCACTAGGATACCAACAAACCTACC  173

Query  16292  CACCCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACA  16345
              ||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Sbjct  174    CACCCTTAACAGTACATAGTACATAAAGCCA-TTACCGTACATAGCACATTACA  226
```

UP2HappyR
Sequence ID: lcl|215455Length: 144

Score 250 bits(135) Identities 139/141(99%) Gaps 0/141(0%)

```
Query  16206  AAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCA  16265
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2      AAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCA  61

Query  16266  CCCACTAGGATACCAACAAACCTACCCACCCTTAACAGTACATAGTACATAAAGCCATTT  16325
              ||||||||||||||||||||||||||||||||||||||||||||| |||||||||||||| 
Sbjct  62     CCCACTAGGATACCAACAAACCTACCCACCCTTAACAGTACATAGCACATAAAGCCATTC  121

Query  16326  ACCGTACATAGCACATTACAG  16346
              |||||||||||||||||||||
Sbjct  122    ACCGTACATAGCACATTACAG  142
```


© University of Pretoria

UP49Makgop

Sequence ID: lcl|237739Length: 155

Score 268 bits(145) Identities 151/154(98%) Gaps 0/154(0%)

```
Query  6919  TCTGAGCCCTAGGATTCATCTTTCTTTTCACCGTAGGTGGCCTGACTGGCATTGTATTAG  6978
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  154   TCTGAGCCCTAGGATTCATCTTTCTTTTCACCGTAGGTGGCCTGACTGGCATTGTATTAG  95

Query  6979  CAAACTCATCACTAGACATCGTACTACACGACACGTACTACGTTGTAGCCCACTTCCACT  7038
             ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Sbjct  94    CAAACTCATCACTAGACATCGTACTACACGACACGTACTACGTTGTAGCTCACTTCCACT  35

Query  7039  ATGTCCTATCAATAGGAGCTGTATTTGCCATCAT  7072
             |||||| ||| |||||||||||||||||||||||
Sbjct  34    ATGTCCAATCCATAGGAGCTGTATTTGCCATCAT  1
```

Bots

Sequence ID: lcl|208915Length: 286Number of Matches: 2

Range 1

Score 302 bits(163) Identities 181/189(96%) Gaps 3/189(1%)

```
Query  16163  AAAACCCAATCCACATCAAAA--CCCCCTCCCCATGCTTACAAGCAAGTACAGCAATCAA  16220
              |||||||||||| ||||||||  ||||| ||||| |||||||||||||||||||||||||
Sbjct  99     AAAACCCAATCC-CATCAAAACCCCCCCCCCCAGGCTTACAAGCAAGTACAGCAATCAA  157

Query  16221  CCCTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCA  16280
              ||||||||||||||||||||||||||||||||||| ||||||||| ||| ||||||||||
Sbjct  158    CCCTCAACTATCACACATCAACTGCAACTCCAAAGTCACCCCTCATCCATTAGGATACCA  217

Query  16281  ACAAACCTACCCACCCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACA  16340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  218    ACAAACCTACCCACCCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACA  277

Query  16341  TTACAGTCA  16349
              |||||||||
Sbjct  278    TTACAGTCA  286
```


© University of Pretoria

Range 2

Score 163 bits(88)   Identities 88/88(100%)   Gaps 0/88(0%)

```
Query  16336  GCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGGGGTCCC  16395
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1      GCACATTACAGTCAAATCCCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGGGGTCCC  60

Query  16396  TTGACCACCATCCTCCGTGAAATCAATA  16423
              ||||||||||||||||||||||||||||
Sbjct  61     TTGACCACCATCCTCCGTGAAATCAATA  88
```

UP77Stayt

Sequence ID: lcl|215695Length: 88

Score 139 bits(75)   Identities 84/88 (95%)   Gaps 1/88(1%)

```
Query  16259  CCCCTCACCCAC-TAGGATACCAACAAACCTACCCACCCTTAACAGTACATAGTACATAA  16317
              ||||||| |||| |||||||||||||||||||||||||||||||||||||||| ||||||
Sbjct  1      CCCCTCAACCACGTAGGATACCAACAAACCTACCCACCCTTAACAGTACATAGCACATAA  60

Query  16318  AGCCATTTACCGTACATAGCACATTACA  16345
              ||||||| ||||||||||||||||||||
Sbjct  61     AGCCATTCACCGTACATAGCACATTACA  88
```

UP98Nwanet

Sequence ID: lcl|199047Length: 301

Score 150 bits(81)   Identities 85/87(98%)   Gaps 0/87(0%)

```
Query  16288  TACCCACCCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGT  16347
              ||||||||||||||||||||||| |||||||| |||||||||||||||||||||||||||
Sbjct  87     TACCCACCCTTAACAGTACATAGCACATAAAGTCATTTACCGTACATAGCACATTACAGT  28

Query  16348  CAAATCCCTTCTCGTCCCCATGGATGA  16374
              |||||||||||||||||||||||||||
Sbjct  27     CAAATCCCTTCTCGTCCCCATGGATGA  1
```

UP43Thulam: No matches against rCRS

© University of Pretoria

UP44Thulam

Sequence ID: lcl|84461Length: 156

Score 278 bits(150) Identities 152/153(99%) Gaps 0/153(0%)

```
Query  6919  TCTGAGCCCTAGGATTCATCTTTCTTTTCACCGTAGGTGGCCTGACTGGCATTGTATTAG  6978
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  155   TCTGAGCCCTAGGATTCATCTTTCTTTTCACCGTAGGTGGCCTGACTGGCATTGTATTAG  96

Query  6979  CAAACTCATCACTAGACATCGTACTACACGACACGTACTACGTTGTAGCCCACTTCCACT  7038
             ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Sbjct  95    CAAACTCATCACTAGACATCGTACTACACGACACGTACTACGTTGTAGCTCACTTCCACT  36

Query  7039  ATGTCCTATCAATAGGAGCTGTATTTGCCATCA  7071
             |||||||||||||||||||||||||||||||||
Sbjct  35    ATGTCCTATCAATAGGAGCTGTATTTGCCATCA  3
```

Table 4.1b. Summary of sequences aligned with rCRS

Individual/site	Region of rCRS aligned to	Colour label
UP1 Happy Rest	16345-16112 (HVRI)	Blue
UP2 Happy Rest	16346-16206 (HVRI)	Green
UP49 Makgope	7072-6919 (COII)	Navy blue
Tuli(Bots)	16349-16163 (HVRI)	Brown
UP77 Stayt	16345-16259 (HVRI)	Red
UP98 Nwanetsi	16374-16288 (HVRI)	Purple
UP43 Thulamela	No alignment	
UP44 Thulamela	7071-6919 (COII)	Grey

© University of Pretoria

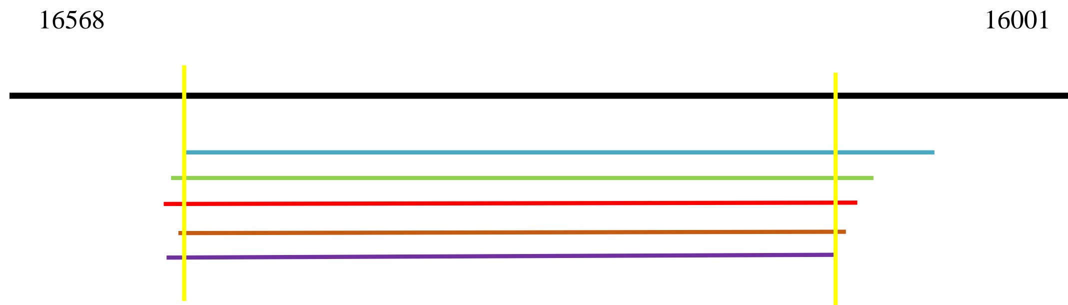


Figure 4.5a. Representation of query sequences versus the rCRS. The numbers represent the nucleotide position relative to the rCRS, represented by the black line.

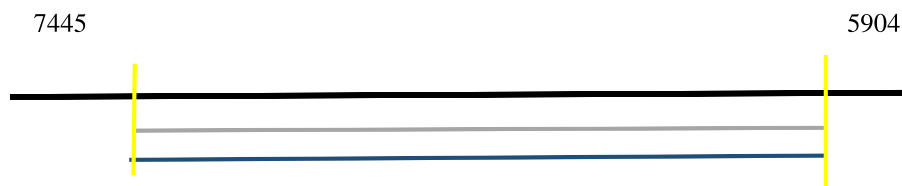


Figure 4.5b. Representation of query sequences versus the rCRS. The numbers represent the nucleotide position relative to the rCRS, represented by the black line.

Aligned sequences were compared to the control region of the revised Cambridge Reference Sequence (rCRS) mtDNA in order to analyse polymorphisms and Mitotool was used to assign each to a haplogroup. All samples were assigned the haplogroup L0, except for UP 43 and UP 49 (as HVRI sequence could not be obtained), and no variants were listed for these sequences. Missing data was noted at position 263 for each. Table 4.2 shows a summary of the individuals representing each site, the haplogroup assigned, as well as the corresponding variants.

Table 4.2. Haplogroup assignment for individuals representing each site

Name	Haplogroup	Missing	Variants
UP44Thulam	L0	L0:263	
UP77Stayt	L0	L0:263	
UP98Nwanet	L0	L0:263	
Bots	L0	L0:263	
UP1HappyR	L0	L0:263	
UP2HappyR	L0	L0:263	

© University of Pretoria

Alignments of the sequences, both with and without GenBank references sequences, were performed using Mega6. An allele network was created using aligned sequences generated from individuals from this study, and a second network also included reference sequences from GenBank, using the program Splitstree (See Figures 4.6 and 4.7). Reference sequences were selected on account of their individual matches with sequences from this study and also due to the geographical region after which they were named, with the focus being of southern African samples.

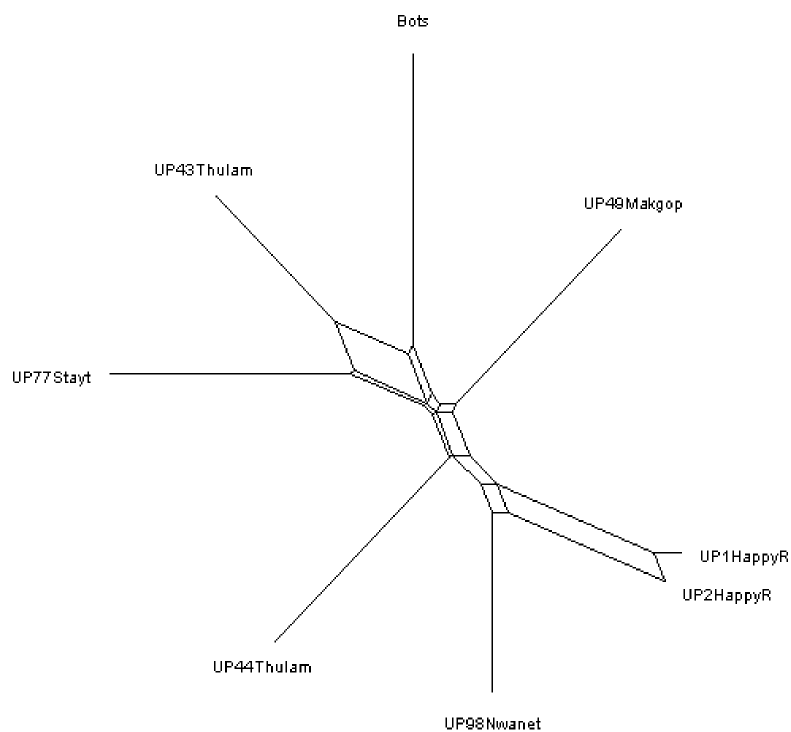


Figure 4.6. Network representing genetic distances among individuals representing each site.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

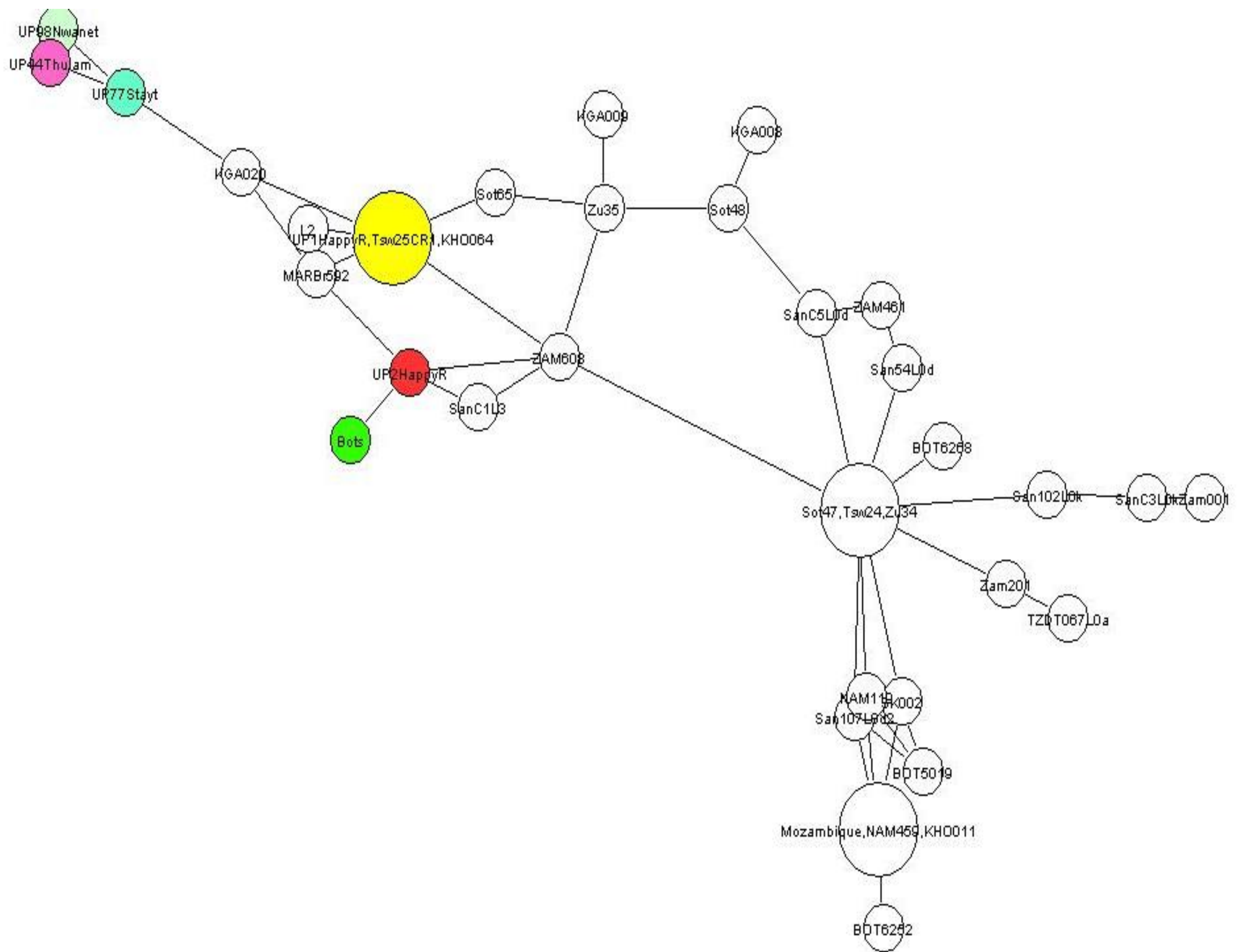


Figure 4.7. Allele network showing the relationships to other sub-Saharan African sequences. Individuals for this study are marked in different colours for visual simplicity.

© University of Pretoria

4.1.5 Statistical analysis

The number of individuals whose DNA was analysed in this study was nine in total, however, sequence data could not be generated for all of them. DNA sequences from the individual UP48 from Makgope could not be used for analysis; fortunately, adequate sequence data could be obtained from the other individual, UP49, also from Makgope. While adequate sequence data could not be retrieved from all samples, for every gene region, each archaeological site was represented by at least one individual.

The initial sample collection cannot be said to have been random, as only bones/teeth/tissue were sampled from the available Iron Age individuals whose preservation was still good, regardless of geographical origin. The geographical contexts of these individuals were mostly in the north-eastern regions of South Africa, (except for the Tuli and Makgope samples).

The alignment including sequences from this study as well as GenBank reference sequences, but excluding UP49 and UP4, was used to calculate the transition: transversion bias, R, which was done using Mega 6. The estimated transition: transversion bias (R) is 3.66. Substitution pattern and rates were estimated under the Kimura (1980) 2-parameter model, with nucleotide frequencies of A = 25.00%, T/U = 25.00%, C = 25.00%, and G = 25.00%. For estimating ML values, a tree topology was automatically computed. The maximum Log likelihood for this computation was -1079.168. The analysis involved 40 nucleotide sequences. Codon positions included were noncoding. All positions with less than 95% site coverage were eliminated. That is fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 146 positions in the final dataset. Evolutionary analyses were conducted in MEGA6 (144).

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Chapter 5

Discussion

5.1 Sample preservation and sampling

The remains were all found in a relatively good state of preservation, although it was not yet certain as to whether DNA could be successfully extracted at the commencement of the study. All remains (except the Thulamela and Tuli individuals) are still in storage at room temperature in the Department of Anatomy, University of Pretoria and only samples from these were removed for use in this study. The remainder of the Tuli mummy is kept in the Gaborone National Museum. The remains of the two individuals from Thulamela were handled with caution when they were excavated, as they were intended for DNA analysis at a later stage, and were therefore handled with gloves and placed in separate bags. The rest of their remains have since been reburied at Thulamela. The Tuli mummy was also handled in the same manner. A single vertebra from the Nwanetsi sample had been covered in foil and also kept for DNA analysis. Other samples used in this study were not originally intended for DNA analysis and were handled with bare hands, therefore contamination with modern DNA was possible.

5.1.1 DNA extraction, PCR and contamination

It was noted that the modified phenol-chloroform method yielded more DNA than with the use of a kit (see results section). We proceeded with the phenol-chloroform extraction method due to the poor and degraded state of aDNA, in order to increase the chances of amplifying enough of the target DNA. After the successful extraction of DNA from all samples, PCR amplification results suggested that the amount of DNA amplified was low, as can be seen in the images of the agarose gels (Figure 4.1) shown as light bands, despite the extended PCR reaction times. This indicated that it is possible that the DNA amplified was endogenous as contamination control and PCR conditions were stringent. Some sporadic PCR contamination did occur, in very small amounts, such as in lane 21 of image 2, figure 4.1a. PCR contamination cannot be completely avoided in aDNA analyses, as modern DNA contaminants competes with the degraded aDNA for primer binding and subsequent polymerase activity. PCR contamination occurred in two instances and was in both cases associated with the DNA extracts from the Tuli mummy. The amount of sporadic contamination was not enough to be significant. The presence of randomly

© University of Pretoria

amplified PCR products also occurred in two instances and is due to the altered nature of ancient DNA, causing some non-specific annealing of primers followed by non-specific polymerase activity. Randomly amplified products, such as those seen in figure 4.1b, image 1, were not used in downstream applications.

5.1.2 Contamination control, amplification and sequencing

After DNA extraction, it was necessary to further amplify the PCR products by bacterial cloning, in order to produce an even higher yield of DNA and produce much stronger signals for the detection of nucleotides during sequencing. Initially, clones were successfully produced, and some (from the Tuli mummy) were used for generating sequences. Clones could not be produced when using other samples and it is suspected that the linearised vector had erroneously ligated upon itself (circularised) prior to ligating to the target DNA. While this is not very likely, as linearised vectors are designed so that their 5'and 3' ends do not complement each other, perhaps the insertion of blunt ends on either side of the target DNA was inadequate, so that ligation of the target DNA to the vector was incomplete. The formation of colonies was also not indicative of successful ligation as the vector could still express antibiotic resistance genes to aid the bacteria to grow, regardless of the insertion of target DNA. After many repeated attempts, the cloning experiments were not producing colonies and due to time constraints, it was decided that sequences would be generated directly from PCR amplified products. The disadvantage of this procedure is that the PCR amplified products produce very poor nucleotide signals during the sequencing procedure, as it does not provide enough DNA for the reaction to proceed. For this reason, many of the sequences generated were rendered unusable for further analysis as they consisted of background 'noise' which can be seen as low, mixed signals on the electropherogram. Furthermore, many of these did not match to any sequences on the online alignment tool, BLAST.

Overlapping primer pairs were used for PCR and sequencing, so that short sequences could be arranged into longer contiguous sequences, wherever possible, using the CLCBio analytical tool. It is therefore important to note that the sequence from UP77 (from Kon-Stayt) was only the forward sequence generated from part of the HVRI region, as other sequences from this individual were of too poor a quality to be used. Results generated for this individual should therefore be treated with caution. Also, one individual had to be

© University of Pretoria

removed completely (UP48 from Makgope), as none of these sequences were informative. Fortunately, this site could still be represented by a different individual, UP49, from whom adequate sequencing results could be obtained. The two individuals were also estimated to have lived at approximately the same time period (UP48, AD 1677-1768 and UP49, AD 1670-1702) (43).

5.1.2.1 Analysis of sequences, network and haplotyping

When aligned against GenBank sequences, some of the samples in this study matched with human mitochondrial DNA sequences, and some even matched to African-specific sequences. A small number of sequences matched those of microbial origin, such as *Gluconacetobacter diazotrophicus, Mycobacter canneti* and *Candida* species. These were the result of contamination. All other sequences were treated with caution, because, as mentioned before, apart from the Thulamela remains, one of the Nwanetsi samples and the Tuli mummy, the other samples were initially handled without gloves, so that contamination by modern human DNA was possible. It was not possible to account for all the people who handled these remains, and contamination could have occurred at any stage of handling and analysis. The most immediate way to detect modern DNA contamination was to look out for mtDNA sequences of European origin, as most of the researchers who had handled the remains were of European descent. No European specific markers were found, but this does not completely rule out the possibility of modern DNA contamination. After viewing the sequences it was noted, however, that many of the typical patterns of aDNA were present, such as a high GC content, for example in Fig 4.4a, (due to the lack of DNA repair mechanisms after death), which attested that the sequences obtained were most likely of ancient origin and not modern DNA contamination.

After editing (removing primer or cloning vector sequences and assembling short sequences into contigs wherever possible) the sequences were verified by comparison to online sequences using BLAST. The individual sequences, along with one of their matches from the GenBank database, are presented in the results section. The query sequences presented high similarities with human mtDNA, however, these also consisted of gaps, possibly due to difficulties in obtaining sequences for those regions in spite of the fact that overlapping primers were used for PCR and sequencing. The Happy Rest individual, UP1 shared a 95% identity with a Zambian individual, with an 8 nucleotide gap from position 218 to 225 and one nucleotide missing in position 203 as well as a C-T change at positions

© University of Pretoria

192 and 204. UP2 from Happy Rest had a 99% similarity with part of an mtDNA sequence representing haplogroup H, which originated in Southwest Asia and spread to Europe some 20000 - 25000 years ago. There was a T-C transversion at position 120 of the query sequence. UP49 shared 99% similarity with an mtDNA sequence representing haplogroup L2, a typically sub-Saharan African marker, having 2 point mutations, G-T transversion and an A-T transition at positions 23 and 27, respectively. The Tuli mummy showed similarities with an mtDNA sequence of a Moroccan individual, a single gap at position 3, a T-G mutation at position 35, as well as three C-T changes at points 192, 202 and 206. UP77 from Kon-Stayt had a 96% similarity with part of an mtDNA sequence from a Mozambican individual, with a T-A mutation at position 15 and two G-C transitions at positions 22 and 41. The sequence representing UP98 from Nwanetsi shared 88% identity with one from a Kgalagadi individual, having T-G transversion at position 103, an A-C mutation at position 106, two T-G mutations at positions 114 and 128, a G-T change at position 121 and two A-T mutations at positions 123 and 142. In the second range, there is a C-G mutation at positions 44 and 35, a T-G mutation at positions 38 and 8 as well as a T-A mutation at position 32. The male from Thulamela, UP44, also shared similarities with two parts of a single sequence representing a Kgalagadi individual. There were T-G transversions at positions 7, 38, 104, 115 and 128, a G-T transversion at position 122 and an A-C transversion at position 107. C-G transitions occurred at positions 35 and 44 and a T-A and A-T transition at positions 32 and 142, respectively. A 100% 23 base-pair match occurred between UP43 from Thulamela and one representing haplogroup L6. However, a 23 base-pair match is not significant as it is too short.

In general, the alignments discussed above present short parts of sequences having similarities to other human mtDNA sequences on GenBank. The similarities therefore show that the sequences obtained were indeed human, but due to post-mortem changes and low sequence resolution, the matches were only partial and our query sequences consisted of a high number of G and C residues. The transition/transversion (R) bias calculated for each individual's sequence alignment against a reference sequence showed generally high ratios, with the exception of UP1. High R values are correlated to low sequence divergence, which is to be expected as the shared sequence identities were also high (80-100%). It is also normal that the R ratio is higher in mtDNA than in nDNA (143). However, given the length of perfectly aligned sequences, (from 23 to 234 base pairs), the

© University of Pretoria

R values calculated here are not significant in that they represent a ratio calculated from relatively short aligned sequences.

Each individual query sequence was also aligned against the rCRS, using BLAST, to assess if the correct gene region had been sequenced. UP1 from Happy Rest aligned to region 16112-16292 with a 96% identity and an 8 nucleotide gap between positions 16248-16245, another gap at position 16313 and a C-T mutation at position 16223. UP2 from Happy Rest aligned to region 16206-16346 with a 99% similarity and one C-T mutation at position 16310. UP 49 from Makgope aligned with nucleotides 6919 to 7072, which represents part of the COI region of the mtDNA; this was a 98% match consisting of a C-T mutation at position 7028, an T-A change at position 7045 and a A-C change at position 7053. The Tuli mummy's sequence matched with two parts of the rCRS, from nucleotide 16163-16349 there was a 96% match, having a two nucleotide gap at positions 16184 and 16185, a T-G mutation at position 16197, and a C-T at 16256, and 12266. The second range aligned to nucleotides 13336 through to 16423 with a 100% similarity. The sequence representing the individual from Kon-Stayt, UP77, aligned with nucleotides 16259 to 16345, having a 95% match, a gap at position 16271, a C-A mutation at position 16266 and two T-C mutations at nucleotides 16326 and 16311. UP98 from Nwanetsi aligned to nucleotides 16288 to 16374, with 98% identity and a T-C and C-T mutation at positions 16311 and 16320, respectively. UP44 from Thulamela aligned with nucleotides 6919 to 7071, also representing part of the COI gene region, with 99% identity and a C-T mutation at position 7028. UP43 could not be aligned to the rCRS, possibly due to a high number of errors in the sequence.

When creating a larger alignment for the hypervariable region (which included GenBank reference sequences), UP43 and UP49 were excluded as these did not align on the same gene region as the rest; inferences about these sequences were thus estimated based on the sequences obtained.

To assign each individual to a haplogroup, sequences were submitted in FASTA format onto Mitotool and were all assigned to the haplogroup L0, which is mostly associated with the Khoesan. The following is a list of mutations that define haplogroup L0: G263A C1048T C3516A T5442C T6185C C9042T A9347G G10589A G12007A A12720G. This does not imply that all individuals represented here were Khoesan, but this does

© University of Pretoria

suggest evidence of genetic exchange or at least shared genetic markers between Khoesan and non-Khoesan groups.

5.1.2.2 Relationships between individuals

A strictly bifurcating phylogenetic tree does not show more resolved details about the relationships between individual sequences as it suggests that one ancestor only gives rise to two descendants each time. Therefore, a network approach, which accounts for different mutational events (seen as reticulations in the network) depicts a more realistic representation of evolutionary processes. Also, the length of each branch is proportional to the estimated genetic distances. A network was thus created using the sequences generated in this study, through the computer algorithm, Splitstree, and is shown in Figure 4.6 of the results section. Beginning from the top of this network, there are fewer mutational steps between UP43, UP77 and the Tuli mummy, also, the branch leading to UP43 is relatively short, suggesting a short genetic distance. On the other hand, three different reticular events occur before the one leading to UP49, moreover, the branch leading to UP49 is long. Two more reticulation events occur before the appearance of UP44, separating this individual, quite significantly from the second Thulamela individual, UP43. The placement of UP98 is quite close to UP1 and UP2, who are also in this case placed very close to each other. This network also serves to split these individuals into two groups of closely related sequences; namely UP43, Tuli, UP49 and UP77 as well as UP44, versus UP98, UP1 and UP2.

5.1.2.3 Comparison with other African sequences

Sequences were aligned against other sequences of African origin from GenBank and the resulting network is shown in Figure 4.7 of the results section. Unfortunately, no DNA data from modern Venda and Tsonga people were accessible online at the time of this study. The general alignment which included other sub-Saharan African sequences showed a large number of variable sites with regard to sequences generated in this study, compared to the reference sequence. This is due to post-mortem changes to the DNA, the consequence of which is that the conclusions drawn from this study cannot be said to be absolute and should rather be taken as approximations. When assessing the spatial relationship of the samples from this study with reference sequences, it can be noted that there is a clear relationship between UP1 and sequences representing the Khoesan and Tswana, with UP2 being placed a few mutational steps away, and included in this cluster is a Zambian, Moroccan and L2 sample. The Tuli mummy (Botswana), is also not distant from the UP2 sample, and it can be

© University of Pretoria

suggested that this individual has some Khoesan/Tswana ancestry for the mtDNA control region.

Sequences representing UP98, UP44 and UP77 do not show a close relationship with the rest. This is possibly due to poor sequence quality and/or the lack of sequences representing other indigenous groups from southern Africa, such as the Venda and Tsonga. The spatial relationship between UP98 and UP44 in this network corresponds also to the network representing only the samples from this study. For the Makgope sample, UP49, and Thulamela, UP43, estimates are made from individual matches to other GenBank sequences, as these were not added to the general network that included other African sequences, due to the inability to obtain control region sequence from them. Both individuals show some genetic relationship to sub-Saharan people, with UP49 sharing similarities with one sequence representing haplogroup L2, present in both the Khoesan and Bantu-speaking groups and UP43 sharing some sequence identity with one representing haplogroup L6, another Bantu-speaker haplogroup. No definite conclusions can be drawn from these relationships as the sequences obtained were of poor quality.

5.2 Comparison to the archaeological evidence

Pre-colonial southern Africa was marked by the presence of hunter-gatherer and farming communities as well as the settlement of the last of the major Bantu expansions (groups continued to migrate during the colonial era). Archaeological sites are usually assessed for the cultural identity of their occupants using materials found at these sites, including the design of homesteads and the position of grave sites to name but a few. This approach is indispensable in the absence of human remains for molecular analysis. When human remains are retrieved (and the necessary precautions taken to minimise contamination), molecular analyses complements the study of material culture and can add to the value of archaeological investigations. However, due to the temporal and spatial diversity between the samples used in this study, as well as the poor quality of the DNA sequences collected, one can only suggest what the results could mean when compared to the archaeological evidence. The current study only provides a few hints into the complexity of diversity that one finds in Africa, thus, these results present only a small step in the avenue of many more research questions.

© University of Pretoria

A common way of assigning group identity or cultural affinity to previous occupants of a site was and still is to some extent, the use of pottery analysis, the details of which are beyond the scope of this study. Certain decoration patterns and/or shapes of pots are associated with different groups and these styles have come to indicate the possibility of group interaction as some patterns were thought to have been influenced by other groups. These, as well as the study of linguistics, have been used to trace possible migration patterns. It is therefore common to find that some ceramics styles overlap at different sites. The sample size plays a role as a single ceramic that could be retrieved from a site is not enough to assign cultural affinity with significant confidence.

Ceramic decorations are sometimes repeated on other objects that form part of the material culture of a group, such as that on walls and weapons etc. Thus, ceramic pattern analysis, as well as other object of material culture, can been a proxy for studying probable group identities of people who have occupied a site (12). This information can thus be supplemented by aDNA analyses of biological remains. The interchange of ceramics can be indicative of alliance formation through marriage (the family of a bride might present pots to the bride when she leaves for her husband's family) or trade (of the contents of the pot) or both (147). DNA analysis has the potential to substantiate claims of intermarriage between groups. The use of ceramic analyses to assign cultural affinities of people has been debated since there is always the possibility that some potters may have created original and even unique ceramic decorations at any given time and these may or may not have been repeated or copied by members of the same group. Also, some potters may have simply copied a style from another group, which indicates inter-communication between groups, but not necessarily the identity of a group (15).

While omitting details about specific patterns of ceramics, there are several examples of different pot-making styles at various sites and perhaps the integration of foreign women in these communities (since potters were usually women (12)). These examples are only to illustrate that there had been some level of communication between population groups: Khami style pottery, associated with the Shona, has been found north of the Soutpansberg, while Icon style pottery, associated with the Sotho/Tswana was found in the south. Pottery associated with Happy Rest had been found south of Mapungubwe Hill; Great Zimbabwe pottery was found in Thulamela. Mutamba pottery was found in the pre-walling phase at Thulamela, while the second phase, also associated with the two skeletons (UP43 and

© University of Pretoria

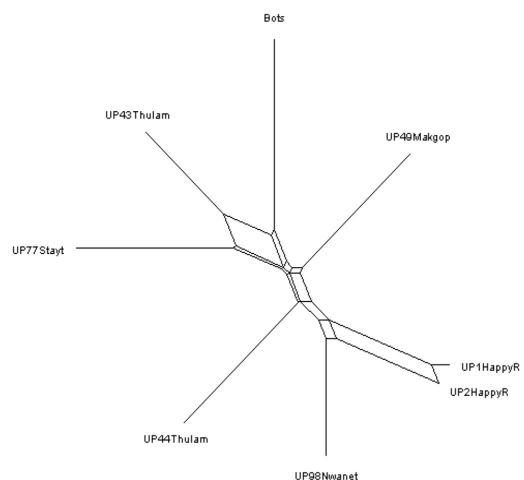
UP44), consisted of Khami style pots. Moloko pots, which are characteristic of the Sotho/Tswana, have been found at Khami and vice versa.

The Tavhatshena style pots are thought to represent a merger between Icon (Sotho/Tswana) and Khami (Shona) into Letaba, which is a Venda style pot that had been limited to the central Soutpansberg and commoner Sotho/Tswana (12). Linguistic assessment of the Venda language also shows a combination of Shona with Sotho/Tswana and some level of uniqueness, and this coincides with the hybrid style pottery of Letaba, which is suggestive of inter-group contact.

To put the results of this study into perspective, Table 5.1 compares group identity (according to pottery analysis, wherever possible, or pots that were found at certain sites), approximate time of occupation (with the estimated date of the skeletal remains in brackets) and the relationships between the tested samples according to the phylogenetic tree created earlier. A miniature copy of Figure 4.6 has been placed adjacent to Table 5.1 for easier assessment.

Table 5.1. Cultural and genetic affinities of the assessed individuals

Site/sample	Approx. date	Group Identity
Happy Rest/UP1	EIA (AD 683-798)	Happy Rest
Happy Rest/UP2	EIA (AD 683-798)	Happy Rest
Kon-Stayt/UP77	MIA (AD 1063-1074)	Mapungubwe
Thulamela/UP44	MIA (AD 1421-1437)	Khami
Thulamela/UP43	MIA (AD 1525-1632)	Khami
Nwanetsi/UP98	LIA (AD 1401-1428)	Thavhatshena
Tuli/Bots	LIA (AD 1400-1500)	Khami
Makgope/UP49	LIA (AD 1670-1702)	Icon



© University of Pretoria

Table 5.1 shows different ceramic styles at presenting at different periods and at diverse sites. The overlapping of some styles into other sites, such as the Mapungubwe and Khami pots in Kon-Stayt and Tuli, respectively, indicates that there was some interaction between members of these groups. According to this table, the ceramic evidence suggests that, with the assumption that potters were mostly women who probably modelled their pots on those of their own ancestors, the two individuals from Thulamela had maternal genetic influence from a group that was ancestral to the Shona. In contrast, the DNA analysis suggests otherwise; the network proposes that UP43, the female, has Sotho/Tswana (or Khoesan) ancestry, given the closeness of this individual to the ones from Makgope, Botswana and to some degree, Kon-Stayt (it should be noted that most southern Africans have some genetic relationship with the Khoesan peoples (138, 139)).

The other individual from Thulamela, UP44, has an affinity to those representing Nwanetsi and Happy Rest. Due to the unavailability of genetic data from the Venda and Tsonga people at the time of this analysis, clearer affinities could not be established According to the network presented earlier, UP43's proximity to the individuals from Kon-Stayt and Tuli (Botswana) suggests a link to the nearby western Sotho/Tswana, while UP44 is closer to the eastern peoples (perhaps those who are ancestral to the Tsonga/Venda, but this cannot be said for certain and is only supposed due to the geographical location of present-day Venda and Tsonga people). The female individual from Thulamela, UP43, was found buried in a manner that suggested that she was a member of the upper class, and it was thought that her relatively tall stature may indicate that she was from a different group and perhaps became a royal wife owing, in part, to her height (4). It is also possible that she brought pottery or skills in ceramic design from her place of origin.

This may suggest both the introduction of new genes as well as novel cultures or cultural practices. But perhaps ceramic styles were not completely influenced by the culture from which a new bride came, but instead, by the current or predominant trend and/or the 'indigenous' women in the community that a new bride enters into. Thus, using ceramic style analyses as a proxy for cultural affinity may still be reliable but it may not be directly indicative of the cultural affinity of foreign women that were married into a community.

The two Thulamela individuals can thus be said to have different genetic affinities; with the male being closer to EIA groups towards the east and the female being closer to the

© University of Pretoria

western peoples such as the makers of Mapungubwe and Khami ceramics. This is evidence of relationships formed through marriage. The association of the Happy Rest individuals with the Khoesan peoples confirms the archaeological evidence that shows that there may have been biological interchange with the Khoesan and Bantu-speakers in the EIA. There is also a genetic link between UP77, UP44 and the Tuli mummy. The latter two have been associated with Khami ceramics while UP77 is associated with Mapungubwe pottery. Again, this suggests that some groups were affiliating with other, perhaps more prominent or prosperous states. Three more individuals group close together; UP44 (Thulamela), UP98 (Nwanetsi) and UP49 (Makgope), and are associated with different ceramic styles; in the case of Makgope, the association with Icon ceramics is not unexpected as Makgope is known to have been occupied by the Tswana, specifically the Bakwena ba Phalane (43). The absence of foreign ceramics here could mean that this community may have been creating their own ceramics and were either not trading in ceramics with other groups or not trading at all. The ceramic style found at Nwanetsi, probably Thavhatshena type, is associated with the Venda.

The two individuals from Happy Rest showed very little (genetic) diversity between them, and incidentally, this is also true for the ceramics found at this Early Iron Age site. As mentioned before, that lack of genetic diversity between these individuals may indicate that there might have been little interaction between people from this site with other groups. Populations were smaller and more dispersed in the Early Iron Age (13, 149). Their existence is also before the settlement of the Sotho/Tswana and Nguni in this region, who arrived at the beginning of the LIA (150).

The genetic evidence does not entirely dispute the conclusions from the ceramic analyses, rather, it sheds new light on the formation of alliances between groups as well as the assignment of cultural affinity using ceramic style analysis. It is therefore imperative, that whenever possible, molecular analyses of human remains accompany the study of archaeological sites. The material evidence gives important information as to the culture or cultural affinity of the inhabitants of a community, but this has no bearing on the genetic heritage of the people practicing a particular culture. The study of material culture remains a useful tool for reckoning cultural affinity, communication and the trade of non-human goods between groups while DNA analysis can aid these findings by providing information on the patterns of gene-flow.

© University of Pretoria

Chapter 6

Conclusions

- The aim of this project was to decipher the probable genetic heritage of the two individuals from Thulamela, and those from other Iron Age sites, including a naturally mummified individual.

- An insignificant amount of PCR contamination occurred and some sequences were of poor quality. Genetic sex determination was inconclusive for all individuals.

- Haplogroup assignments showed that all individuals, that could be compared to the rCRS, were of sub-Saharan African origin, but none could be assigned to specific sub-haplogroups and this would need to be further resolved in future.

- The data can be separated into two groups of individuals who share the most similarities with each other: UP43 (Thulamela) is associated with UP49 (Makgope), UP77 (Kon-Stayt) and Bots (or the Tuli mummy), while the second group consists of UP44 (Thulamela), UP1 and UP2 (Happy Rest), as well as UP98 (Nwanetsi). The Happy Rests' affinity to the Khoesan is indicative of evolving cultural practices over time.

- The genetic evidence points to the origin of the maternal genetic heritage and for lack of Y-chromosome data for the males, the paternal heritage could not be deciphered. The origin of the Thulamela male could thus also not be ascertained.

- The results indicate a shift in genetic diversity over time, since the Iron Age and perhaps even earlier, due to differential interaction between groups.

- Short-comings from this study include to the inability to obtain autosomal DNA and poor mtDNA sequence quality. The data from mtDNA provide a useful glimpse into the migration of female lines, and thus, on marriage practices, that may have fostered alliances between populations.

© University of Pretoria

- The presence of Sotho/Tswana mtDNA in the Thulamela individuals substantiates archaeological finds (including ceramic style analysis).

- Future work will include the application of a migration-modelling computer algorithm to trace more specific migration patterns of these individuals' maternal ancestors across sub-Saharan Africa.

- Further research will also attempt to add genetic data from va-Tsonga and vha-Venda for comparison.

- The study shows the importance of DNA analysis in complementing the role of archaeological research in population history studies

© University of Pretoria

List of References

1. Mooder KP, Weber AW, Bamforth FJ, Lieverse AR, Schurr TG, Bazaliiski VI, et al. Matrilineal affinities and prehistoric Siberian mortuary practices: a case study from Neolithic Lake Baikal. J Archaeol Sci. 2005;32:619-34.

2. O'Rourke DH, Hayes MG, Carlyle SW. Spatial and temporal stability of mtDNA haplogroup frequencies in native North America. Hum Biol. 2000a;72:15-34.

3. O'Rourke DH, Hayes MG, Carlyle SW. Ancient DNA studies in physical anthropology. Annual Reviews Anthropology. 2000b;29:217-42.

4. Steyn M, Miller S, Nienaber WC, Loots M. Late Iron Age gold burials from Thulamela (Pafuri Region, Kruger National Park). S Afr Archael Bull. 1998;53(168):73-85.

5. Newman JL. The Peopling of Africa: A Geographic Interpretation. Yale University Press. 1995.

6. Miller SM. The archaeological investigation, the rebuilding of the walls, and the public presentation of the Thulamela archaeological site. Pretoria: National Parks Board, 1998.

7. Nurse GT, Weiner JS, Jenkins T. The Peoples of Southern Africa and their Affinities. Jenkins T, editor. Oxford: Claendon Press. 1985.

8. Schroeder S. Secondary disposal of the dead: Cross-cultural codes. World Cultures. 2001;12(1):77-93.

9. Van Reenen JF. The Broederstroom pot burial. J Dent Assoc S Africa. 1975;30:419-29.

10. Klapwijk M. Pot and pit burials from the north-eastern Transvaal, South Africa. S Afr Archaeol Bull. 1989;44:65-9.

11. Hewlett BS, Cavilla-Sforza LL. Cultural Transmission among Aka Pygmies. Am Anthropol. 1986;88(4):922–34.

12. Huffman TN. Handbook to the Iron Age: The Archaeology of Pre-colonial Farming Societies in Southern Africa: University of KwaZulu-Natal Press. 2007.

13. Huffman TN. The Early Iron Age and the spread of the Bantu. S Afr Archaeol Bull. 1970;25(97):3-21.

14. Huffman T. Ceramics, settlements and Late Iron Age migrations. Afr Archaeol Rev. 1989;7(1):155-82.

15. Hall M. The changing past: farmers, kings and traders in Southern Africa. David Phillip. Cape Town 1987:200-1860.

16. Vansina J. New Linguistic Evidence and 'The Bantu Expansion'. J Afr Hist. 1995;36(02):173-95.

17. Evers TM. The recognition of groups in the Iron Age of Southern Africa. Unpublished PhD thesis,University of the Witwatersrand, Johannesburg. 1988.

18. Shillington K. History of Africa: St. Martin's press New York. 1995.

19. Phillipson DW. Spread of Bantu Language. Sci Am. 1977;236(4):106-14.

20. Mitchell P. The Archaeology of Southern Africa: Cambridge University Press. 2002.

21. Phillipson DW. African Archaeology: Cambridge University Press. 1993.

© University of Pretoria

22. Pikirayi I. The Zimbabwe Culture: Origins and Decline of Southern Zambezian States: AltaMira Press, California. 2001.

23. Pikirayi I. Ceramics and group identities: Towards a social archaeology in southern African Iron Age ceramic studies. J Soc Archaeol. 2007;7(3):286-301.

24. Huffman TN. Mapungubwe and the Origins of the Zimbabwe Culture. Goodwin Series. 2000;8:14-29.

25. Hammond-Tooke W. Southern Bantu origins: light from kinship terminology. South Afr Humanit. 2004;16:71-8.

26. Castri L, Tofanelli S, Garagnani P, Bini C, Fosella X, Pelotti S, et al. 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. 2009;140(2):302-11.

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

28. Pereira L, Gusmao L, Alves C, Amorim A, Prata MJ. Bantu and European Y-lineages in Sub-Saharan Africa. Ann Hum Genet. 2002;66:369-78.

29. Loubser JHN. The ethnoarchaeology of Venda-speakers in southern Africa: Nasionale Museum; 1991.

30. Huffman TN. Mapungubwe and Great Zimbabwe: The origin and spread of social complexity in southern Africa. J Anthropol Archaeol. 2009;28(1):37-54.

31. Miller D. Smelter and Smith: Iron Age metal fabrication technology in Southern Africa. J Archaeol Sci. 2002;29:1083-131.

32. Berniell-Lee G, Calafell F, Bosch E, Heyer E, Sica L, Mouguiama-Daouda P, et al. Genetic and Demographic Implications of the Bantu Expansion: Insights from human paternal lineages. Mol Biol Evol. 2009;26(7):1581-9.

33. Miller SM. Adaptation of Traditional Building Methods. Johannesburg: University of the Witwatersrand, 1990.

34. Huffman TN. Archaeological evidence and conventional explanations of southern Bantu settlement patterns. Africa. 1986;56(03):280-98.

35. Van Warmelo NJ. The copper miners of Musina and the early history of the Zoutpansberg. Pretoria: South African Government Printer. 1940.

36. Stayt HA. The Bavenda: Cass; 1931.

37. Eloff JF, de Vaal JB. Makahane. Koedoe. 1965;8:68-74.

38. Meyer A. The Archaeological sites of Greefswald. Pretoria 1998.

39. Morris AG. A Master Catalogue: Holocene Human Skeletons from South Africa. Witwatersrand University Press; 1992.

40. Steyn M, Loots, M, Prinsloo, H. An Early Iron Age human skeleton from the Soutpansberg district, South Africa. S Afr J Sci 1994;90.

41. Van Reenen JF. Tooth mutilation amongst the peoples of Kavango and Bushmanland, South West Africa (Namibia). J Dent Assoc S Africa. 1978;33:205-18.

© University of Pretoria

42. Morris A. Dental mutilation in historic and prehistoric South Africa. Q Bull S Afr Libr. 1989;43(3):207-24.

43. Pistorius JCC, Steyn M, Nienaber WC. Two burials from Makgope - a Batswana settlement in the Bankveld. S Afr J Enthnol. 1998;21(3).

44. Prinsloo HP, Coetzee FP. Stayt: A 13th Century Iron Age Site, Soutpansberg District, Northern Province, South Africa. So Afr J Ethnol. 2001;24(3):81-7.

45. Huffman TN. Snakes and Crocodiles: Witwatersrand University Press, Johannesburg, 1996.

46. Hutten M, Nienaber, WC, Steyn M. Investigation of two 13th century Iron Age graves from Stayt, Limpopo Province. unpublished data.

47. Nienaber WC, Steyn M. Nwanetsi. Kruger National Park. unpublished data.

48. Mosothwane MN. The Tuli Mummy: A preliminary report from Northeastern Botswana. S Afr Archaeol Bull. 2011;66(194).

49. Pääbo SHP, Serre D, Jaenicke-Despr´es V, Hebler J, Rohland N, Kuch M et al. Genetic analyses from ancient DNA. Ann Rev Gen. 2004;38:645-79.

50. Higuchi R, Bowman B, Freiberger M, Ryder OA, Wilson AC. DNA sequences from the quagga, an extinct member of the horse family. Nature. 1984;312:282-4.

51. Pääbo SHP, Higuchi RG, Wilson AC. Ancient DNA and the polymerase chain reaction. J Biol Chem. 1989;264:9709-12.

52. Willerslev E, Cooper A. Ancient DNA. Proc Biol Sci. 2005;272:3-16.

53. Bolnick DA, Bonine HM, Mata-Miguez J, Kemp BM, Snow MH, Leblanc SA. Nondestructive sampling of human skeletal remains yields ancient nuclear and mitochondrial DNA. Am J Phys Anthropol. 2012;147(2):293-300.

54. Cooper A, Poinar, HN. Ancient DNA: Do it right or not at all. Science. 2000;289:1139.

55. Bensasson D, Zhang D, Hartl DL, Hewitt GM. Mitochondrial pseudogenes: Evolution's misplaced witnesses. Trends Ecol Evol. 2001;16(6):314-21.

56. Shearin NL, King EJ, O'Rourke DH. DNA preservation in Precolumbian remains from the American Southwest. Hum Evol. 1989;4:263-70.

57. Drancourt M, Aboudharam G, Signoli M, Dutour O, Raoult D. Detection of 400-year-old Yersinia pestis DNA in human dental pulp: An approach to the diagnosis of ancient septicemia. Proc Natl Acad Sci USA. 1998;95:12637-40.

58. Smith BC, Fisher DL, Weedn VW, Warnock GR, Holland MM. A systematic approach to the sampling of dental DNA. J For Sci. 1993;38:1194-209.

59. Pääbo SHP. Molecular cloning of ancient Egyptian mummy DNA. Nature.1989;314:644-5.

60. Chobe LP, Chadha MS, Banerjee K, Arankalle VA. Detection of HEV RNA in faeces by RT-PCR during the epidemics of hepatitis E in India (1976-1995). J Viral Hepatitis. 1997;4:129-33.

61. Poinar HN, Hofreiter M, Spaulding WG, Martin PS, Stankiewicz BA, Bland H, et al. Molecular coproscopy: dung and diet of the extinct ground sloth Nothrotheriops shastensis. Science. 1998;281:402-6.

© University of Pretoria

62. Allen M, Engström AS, Meyers S, Handt O, Saldeen T, et al. Mitochondrial DNA sequencing of shed hairs and saliva on robbery caps: Sensitivity and matching probabilities. J For Sci. 1998;43:453-64.

63. Adcock GJ, Dennis ES, Easteal S, Huttley GA, JermiinLS, Peacock WJ, et. Mitochondrial DNA sequences in ancient Australians: Implications for modern human origins. Proc Natl Acad Sci USA. 2001;98:537-42.

64. Thomas M, Gilbert P, Willerslev E, Hansen AJ, Barnes J, Rudbeck L, et al. Distribution Patterns of Postmortem Damage in Human Mitochondrial DNA. Am J Hum Genet. 2003;72:32-47.

65. Thomas M, Gilbert P, Hansen AJ, Willerslev E, Rudbeck L, Barnes I, et al. Characterization of Genetic Miscoding Lesions Caused by Postmortem Damage. Am J Hum Genet. 2003;72:48-61.

66. Weiner S, Lowenstam H A, Hood L. Characterisation of 80-million-year-old Mollusk Shells. Natl Acad Sei. 1976;73:2541-5.

67. Yamada Y, Awedimento E, Mudryj M, Ohkubo H, Vogeli G, Irani M, et al. The Collagen Gene: Evidence for its Evolutionary Assembly by Amplification of a DNA Segment Containing an Exon of 45bp. Cell. 1980;22:887-92.

68. Sensabaugh GF, Wilson AC, Kirk PL. Protein Stability in Preserved Biological remains I: Survival of Biologically Active Proteins in an 8-year-old Sample of Dried Blood. Int J Biochem. 1971;2:545-68.

69. Poinar HN, Höss M, Bada JL, Pääbo SHP. Amino acid racemization and the preservation of ancient DNA. Science. 1996;272:864-6.

70. Lindahl T. Instability and decay of the primary structure of DNA. Nature. 1993;362:709-715.

71. Hofreiter M, Jaenicke V, Serre S, von Haeseler A, Pääbo SHP. DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. Nucleic Acids Res. 2001a;29:4793-9.

72. Gilbert MTP, Willerslev E, Hansen AJ, Barnes, I, Rudbeck L, Lynnerup N, et al. Distribution patterns of post-mortem damage in human mitochondrial DNA. Am J Hum Genet. 2003a;72:32-47.

73. Rizzi E, Lari M, Gigli E, De Bellis G, Caramelli D. Ancient DNA studies: New perspectives on old samples. GSE. 2012;44:21.

74. Hofreiter M, Serre D, Poinar HN, Kuch M, Pääbo SHP. Ancient DNA. Nat Rev Gen. 2001;2:353-9.

75. Cooper A, Lalueza-Fox C, Anderson S, Rambaut A, Austin J, Ward R. Complete mitochondrial genome sequences of two extinct moas clarify ratite evolution. Nature. 2001b;409:704-7.

76. Di Bernardo G, Del Gaudio S, Cammarota M, Galderisi U, Cascino A, Cipollaro M. Enzymatic repair of selected cross-linked homoduplex molecules enhances nuclear gene rescue from Pompeii and Herculaneum remains. Nucleic Acids Res. 2002;30:e16.

77. Pääbo S, Irwin DM, Wilson AC. DNA damage promotes jumping between templates during enzymatic amplification. J Biol Chemi. 1990;265:4718-21.

78. Hedges EM, Millard AR. Bones and groundwater: Towards the modelling of diagenetic processes. J Archaeol Sci. 1995:155-64.

79. Bollongino R, Tresset A, Vigne J. Environment and excavation: Pre-lab impacts on ancient DNA analyses. Paleoevol. 2008;7:91-8.

80. Höss M, Pääbo SHP, Vereshchagin NK. Mammoth DNA sequences. Natue. 1994;370:333.

© University of Pretoria

81. Gibert M, Touinssi M, Reviron D, Mercier P, Boetsch G, Chiaroni J. Ancient mitochondrial DNA from hair. Curr Biol. 2004a;14:R463-R4.

82. Shapiro B. The impact of large-scale climate change on genetic diversity in large mammal populations. Science. 2004;In the press.

83. Rasmussen M, Li Y, Lindgreen S, Pedersen JS, Albrechtsen A, Moltke I, et al. Ancient human genome sequence of an extinct Palaeo-Eskimo. Nature. 2010;463(7282):757-62.

84. Hardy C, Callou C, Vigne J, Casane D, Dennebouy N, Mounolou J, Monnerot M. Rabbit mitochondrialDNAdiversity from prehistoric to modern time. J Mol Evol. 1995;40:227-37.

85. Goodyear PD, MacLaughlin-Black S, Mason IJ. A reliable method for the removal of co-purifying PCR inhibitors from ancient DNA. Biotechniques. 1994;16:232-5.

86. Jorde LB, Bamshad M, Rogers AR. Using mitochondrial and nuclear DNA markers to reconstruct human evolution. Bio Essays 1998;20:126–36.

87. Burbano HA, Hodges E, Green RE, Briggs AW, Krause J, Meyer M, et al. Targeted investigation of the Neandertal genome by array-based sequence capture. Science. 2010;328(5979):723-5.

88. Höss M, Pääbo SHP. DNA extraction from Plestocene bones by a silica-based purification method. Nucleic Acids Res. 1993;21:3913-4.

89. Cattaneo C, Craig OE, James NT, Sokol RJ. Comparison of three DNA extraction methods on bone and blood stains up to 43 years old and amplification of three different gene sequences. J For Sci. 1997;42:1126-35.

90. Torroni A, Wallace DC. Mitochondrial DNA variation in human populations and implications for detection of mtDNA mutations of pathological significance. J Bioenerg Biomembr 1994;26:261-71.

91. Batini C, Coia V, Battaggia C, Rocha J, Pilkington MM, Spedini G, et al. Phylogeography of the human mitochondrial L1c haplogroup: Genetic signatures of the prehistory of Central Africa. Mol Phylogenet Evol. 2007;43(2):635-44.

92. Sota T, Sasabe M. Utility of nuclear allele networks for the analysis of closely related species in the genus Carabus, subgenus Ohomopterus. Syst Biol. 2006;55(2):329-44.

93. Cann RL, Stoneking M, Wilson AC. Mitochondrial DNA and human evolution. Nature. 1987;325:31-6.

94. Chen YS, Torroni A, Exco YL, Santachiara-Benerecetti, S, Wallace DC. Analysis of mtDNA variation in African populations reveals the most ancient of all human continent-specifc haplogroups. Am J Hum Genet. 1995;57:133-49.

95. Brandstätter A, Parsons TJ, Parson W. Rapid screening of mtDNA coding region SNPs for the identification of west European Caucasian haplogroups. Int J Leg Med. 2003;117:291–8.

96. Ballinger SW, Schurr TG, Torroni A, Gan YY, Hodge JA, Hassan K, et al. Southeast Asian mitochondrial DNA analysis reveals genetic continuity of ancient Mongoloid migrations. Genetics. 1992;130:139-52.

97. Schurr TG, Ballinger SW, Gan YY, Hodge JA, Merriwether DA, Lawrence DN, et al. Amerindian mitochondrial DNAs have rare Asian mutations at high frequencies, suggesting they derived from four primary maternal lineages. Am J Hum Genet. 1990;46:613-23.

© University of Pretoria

98. Hammer MF, Karafet T, Rasanayagam A, Wood ET, Altheide TK, Jenkins T, et al. Out of Africa and back again: Nested cladistic analysis of human Y chromosome variation. Mol Biol Evol. 1998 Apr;15(4):427-41.

99. Underhill PA, Shen P, Lin AA, Jin L, Passarino G, Yang WH, et al. Y chromosome sequence variation and the history of human populations. Nat Genet. 2000;26(3):358-61.

100. Poinar HN, Poinar GO, Cano R J. DNA from an extinct plant. Nature. 1993;363:677.

101. Nei M. Molecular Evolutionary Genetics. New York: Columbia University Press. 1987.

102. Ruvolo M, Zehr S, von Dornum M, Pan D, Chang B, Lin J. Mitochondrial COII sequences and modern human origins. Mol Biol Evol. 1993;10(6):1115-35.

103. Mayer A, Lahr G, Swaab DF, Pilgrim C, Reisert I. The Y-chromosomal genes SRY and ZFY are transcribed in adult human brain. Neurogenetics. 1998;1(4):281-8.

104. Wilson JF, Erlandsson R. Sexing of human and other primate DNA. Biol Chem. 1998;379(10):1287-8.

105. Kim JJ, Han BG, Lee HI, Yoo HW, Lee JK. Development of SNP-based human identification system. Int J Legal Med. 2010;124(2):125-31.

106. Kaestle FA, and Horsburgh KA. Ancient DNA in Anthropology: Methods, Applications, and Ethics. Yearb Phys Anthropol. 2002;45:92-130.

107. Telenius H, Carter, NP, Bebb, CE, Nordenskjold M, Ponder BA, Tunnacliffe A. Degenerate oligonucleotide-primed PCR general amplification of target DNA by a single degenerate primer. Genomics. 1992;13:718-25.

108. Don RH, Cox PT, Wainwright BJ, Baker K, Mattick JS. Touchdown PCR to circumvent spurious priming during gene amplification. Nucleic Acids Res. 1991;19:4008.

109. Brotherton P, Endicott P, Sanchez JJ, Beaumont M, Barnett R, Austin J, et al. Novel high-resolution characterization of ancient DNA reveals C-U-type base modification events as the sole cause of post mortem miscoding lesions. Nucl Acids Res. 2007;35:5717-28.

110. Edwards CJ, Magee DA, Park SD, McGettigan PA, Lohan AJ, Murphy A, et al. A complete mitochondrial genome sequence from a mesolithic wild aurochs (Bos primigenius). PLoS One. 2010;5(2):e9255.

111. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA. 1977;74(12):5463-7.

112. Mardis ER. Next generation DNA sequencing methods. Annu Rev Genomics Hum Genet. 2008;9:387–402.

113. Marguerat S, Wilhelm, BT, Bähler J. Next-generation sequencing: applications beyond genomes. Biochem Soc Trans. 2008;36:1091-6.

114. Orlando L, Ginolhac A, Raghavan M, Vilstrup J, Rasmussen M, Magnussen K, et al. True single-molecule DNA sequencing of a pleistocene horse bone. Genome Res. 2011;21(10):1705-19.

115. Lane AB, Soodyall H, Arndt S, Ratshikhopha ME, Jonker E, Freeman C, et al. Genetic substructure in South African Bantu-speakers: evidence from autosomal DNA and Y-chromosome studies. Am J Phys Anthropol. 2002;119(2):175-85.

© University of Pretoria

116. Destro-Bisol G, Donati F, Coia V, Boschi I, Verginelli F, Caglia A, et al. Variation of female and male lineages in sub-Saharan populations: the importance of sociocultural factors. Mol Biol Evol. 2004;21(9):1673-82.

117. Pearson WR. Improved tools for biological sequence comparison. Proc Natl Acad Sci USA. 1988;85:2444-8.

118. Smith F. Identification of common molecular subsequences. J Mol Biol. 1981;147:195-7.

119. Posada D, Crandall KA. Intraspecific gene genealogies: trees grafting into networks. Trends Ecol Evol. 2001;16(1):37-45.

120. Rosenberg NA, Nordborg M. Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. Nat Rev Genet. 2002;3(5):380-90.

121. Kingman JF. Origins of the coalescent. Genetics. 2000;156(4):1461-3.

122. Hudson R. Gene genealogies and the coalescent process. Oxf Surv Evol Biol 1990;7:1-44.

123. Kingman JFC. The coalescent. Stochastic Process Appl. 1982b;13(235-248).

124. Morris AP. Coalescent methods for fine-scale disease-gene mapping. Methods Mol Biol. 2007;376:123-40.

125. Victoria M, Miagostovich MP, Ferreira MS, Vieira CB, Fioretti JM, Leite JP, et al. Bayesian coalescent inference reveals high evolutionary rates and expansion of norovirus populations. Infect Genet Evol. 2009;9(5):927-32.

126. Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J., et al. Sequence and organization of the human mitochondrial genome. Nature. 1981;9:290(5806):457-65.

127. Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nature Genetics. 1999;23(2):147.

128. Yang DY, Watt K. Contamination controls when preparing archaeological remains for ancient DNA analysis. J Archaeol Sci. 2005;32(3):331-6.

129. Ginolhac A, Rasmussen M, Gilbert MT, Willerslev E, Orlando L. mapDamage: Testing for damage patterns in ancient DNA sequences. Bioinformatics. 2011;27(15):2153-5.

130. Krause J, Briggs AW, Kircher M, Maricic T, Zwyns N, Derevianko A, et al. A complete mtDNA genome of an early modern human from Kostenki, Russia. Curr Biol. 20109;20(3):231-6.

131. Green RE, Briggs AW, Krause J, Prüfer K, Burbano HA, Siebauer M, et al. The Neandertal genome and ancient DNA authenticity. EMBO J. 2009;28:2494-502.

132. Holm S. The privacy of Tutankhamen - utilising the genetic information in stored tissue samples. Theo Med. 2001;22:437-49.

133. Klein RG. Whither the Neanderthals. Science. 2003;299:1525.

134. Green RE BA, Krause J, Marici T, Stenzel U, Kircher M. et al. A draft sequence of the Neandertal genome. Science. 2010;328:710-22.

135. Castrì L, Tofanelli S, Garagnani P, Bini C, Fosella X, Pelotti S, et al. 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 Physical Anthro. 2009;140(2):302-11.

© University of Pretoria

136. Lane AB, Soodyall H, Arndt S, Ratshikhopha ME, Jonker E, Freeman C, et al. Genetic substructure in South African Bantu-speakers: evidence from autosomal DNA and Y-chromosome studies. Am J Phys Anthro. 2002;119:175-85.

137. Soodyall H, Jenkins T. Mitochondrial DNA studies in the South African Indian population. Gene Geogr. 1992;6(3):127-37.

138. Soodyall H, Jenkins T. Mitochondrial DNA polymorphisms in Khoisan populations from southern Africa. Ann Hum Genet. 1992;56(4):315-24.

139. Quintana-Murci L, Harmant C, Quach H, Balanovsky O, Zaporozhchenko V, Bormans C, et al. Strong maternal Khoisan contribution to the South African coloured population: A case of gender-biased admixture. Am J Hum Genet. 2010;86(4):611-20.

140. Adachi N, Shinoda, K, Umetsu K, Matsumura H. Mitochondrial DNA Analysis of Jomon skeletons from the Funadomari site, Hokkaido, and its Implication for the origins of Native Americans. Am J Physical Anthro. 2009;138:255–65.

141. Stiles AR, McDonald JG, Bauman DR, Russell DW. CYP7B1: One cytochrome P450, two human genetic diseases, and multiple physiological functions. J Biol Chem. 2009;284(42):28485-9.

142. Tamura K, Nei M, Kumar S. Prospects for inferring very large phylogenies by using the neighbor-joining method. Proc Natl Acad Sci USA. 2004;101(30):11030-5.

143. Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol. 1980;16(2):111-20.

144. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol. 2011;28(10):2731-9.

145. Tamura K, Kumar S. Evolutionary distance estimation under heterogeneous substitution pattern among lineages. Mol Biol Evol. 2002;19(10):1727-36.

146. Behar DM, Villems R, Soodyall H, Blue-Smith J, Pereira L, Metspalu E, et al. The dawn of human matrilineal diversity. Am J Hum Genet. 2008;82(5):1130-40.

147. Evers T, Hammond-Tooke WD. The emergence of South African Chiefdoms: An archaeological perspective. J Afri Stud. 1986;45(1):37-41.

148. Huffman TN, Vogel JC. The chronology of Great Zimbabwe. S Afr Archaeol Bull. 1991;46(154):61-70.

149. Robertson JH, Bradley R. A new paradigm: The African Early Iron Age without Bantu migrations. Hist Afr. 2000:287-323.

150. Badenhorst S. Descent of Iron Age farmers in southern Africa during the last 2000 Years. Afr Archaeol Rev. 2010;27(2):87-106.

151. Benson DA, Cavanaugh M, Clark K, et al. GenBank. Nucleic Acids Res. 2013; 41: 36-42.

152. Hudson DH and Bryant. Application of Phylogenetic Networks in Evolutionary Studies. Mol Bio Evol. 2006;23(2): 254-267.

© University of Pretoria

153. Kloss-Brandstaetter A, Pacher D, Schoenherr S, Weissensteiner H, Binna R, Specht G and Kronenberg F. HaploGrep: A Fast and Reliable Algorithm for Automatic Classification of Mitochondrial DNA Haplogroups. 2010. http://www.haplogrep.uibk.ac.at doi: 10.1002/humu.21382

154. Fan L and Yao Y. Mitotool: A Web Server for the Analysis and Retrieval of Human Mitochondrial DNA Sequence Variations. Mitochondrion. 2011;11: 351-356.

155. van Oven M and Kayser M. Updated Comprehensive Phylogenetic Tree of Global Human Mitochondrial DNA Variation. 2009. Hum Mutat 30(2):E386-E394. http://www.phylotree.org . doi:10.1002/humu.20921

156. Dissing J, Thymann M, and Hopkinson D. Simultaneous Detection of ACP1 and GC Genotypes in Using PCR/SSCP. Ann Hum Gen. 2003; 67(1): 81-85

157. Clement M, Posada D and Crandall KA. TCS: A Computer Program to Estimate Gene Genealogies. Mol Ecol. 2000; 9(10): 1657-1660

158. Posada D, Crandall KA and Templeton AR. GeoDis: A Program for the Cladistic Nested Analysis of the Geographical Distribution of Genetic Haplotypes. Mol Ecol. 2000; 9: 487-448

© University of Pretoria