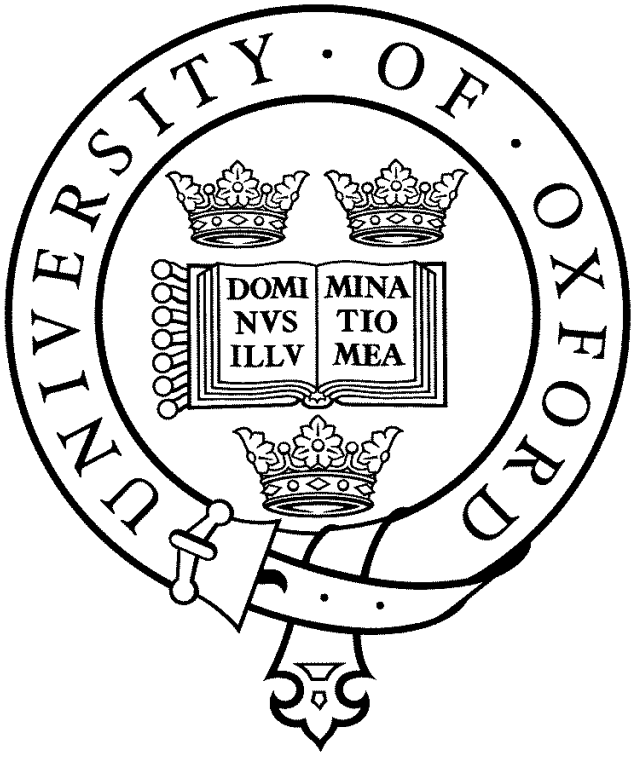
The Peopling of Southern Africa: a Genetics Approach



A Thesis submitted for the Degree of Doctor of Philosophy

Sarah Marks

Balliol College, University of Oxford

Trinity 2012

“It always seems impossible until it’s done”

Nelson Mandela

To Rachael, Simon and Katie

Declaration

This thesis is structured as a set of independent chapters in the style of scientific journal articles with a general methods chapter and final conclusion linking them, as approved by the Director of Graduate Studies in the Department of Zoology, University of Oxford. Two chapters, 4 and 5, will form the basis for two articles which are in preparation. Part of chapter 3 is the basis for the following paper, which is published in Molecular Ecology:

Marks, S. J., et al (2012) 'Migration distance rather than migration rate explains genetic diversity in human patrilocal groups.' *Molecular Ecology* 21(20) 4958-4969.

All of the work presented in this thesis is my own, and any research that I discuss of others is referenced accordingly. I include a Contributors section at the end of the thesis indicating those who have helped with sampling, provided samples and helped with laboratory work. These individuals are also indicated in the main text where appropriate.

The Peopling of Southern Africa: a Genetics Approach

Sarah Marks, Balliol College, University of Oxford

A thesis submitted for the degree of Doctor of Philosophy, Trinity 2012

Abstract

Human populations in Africa have high levels of genetic, cultural and linguistic diversity. Despite this, only a small proportion of African populations have been studied from a genetics perspective. There is a particular dearth of information for Southern Africa, even though this region is one of the few places where hunter-gatherer, pastoralist and farmer populations remain, and where interaction between these groups can be studied. This thesis analyses novel populations from Lesotho, Namibia and South Africa, for both maternal and paternal markers, in order to provide a more accurate understanding of the history of this region. Analysis of biodemographic data for these populations highlights signatures of specific cultural traits. Interestingly, contrary to expectations, results indicate that in the patrilocal Basotho there are no differences in within and between area mtDNA and NRY variation, despite observations of a higher female than male migration rate. Females move preferentially at shorter distances than males, minimising the impact of the higher female migration rate. Further analysis of the Southern African samples indicates that the genetic composition of these populations is different to previously studied populations from Sub-Saharan Africa. Notably, there is a significantly higher maternal hunter-gatherer component, potentially as a result of an archeologically defined static frontier which existed along the Maloti/Drakensberg escarpment. Analysis of Namibian samples provides additional information about the history of populations with different lifestyles in the region, with support for a link between Southern Africa pastoralists and East Africa, while other Namibian populations appear genetically different from previously studied populations. Overall, this work demonstrates the diversity of populations in Southern Africa, improves understanding of the history of this region, and also emphasises the value of having access to geographically and ethnically well-defined samples.

Acknowledgements

There are many people without whom this thesis could not have been completed. Firstly I would like to thank my supervisor, Dr Cristian Capelli, to whom I am extremely grateful for all his advice and support. I would also like to thank The Boise Fund, who have provided monetary support, not only in the form of my Studentship, but also significant contributions to sampling and genotyping costs. The Wenner Gren Foundation also provided funds for collection of samples. Thank you to my examiners for taking the time to read this work.

None of the work in this thesis could have been carried out without the DNA samples I had access to, and for this I owe a great deal of thanks to the people who provided their saliva during sampling trips to Lesotho and Namibia. Additionally I thank all those who helped with the collection of samples, and the organisation of these trips, and also those who provided DNA samples from other populations.

I would also like to thank the other members of the Capelli group: George for help with R scripts amongst other things, and his entertaining conversations, and Francesca for much lab related advice and for preventing me being outnumbered by the men! I have also benefitted from help (and numerous cups of coffee) from the many visiting students and researchers: Chiara, Conrado, Francesco, Gijsbert, Hila, Jeremy, Marco, Maria, Nick.

Throughout my DPhil, Oxford University Water Polo Club has provided a great distraction and stress relief, both in and out of the water. I thank all the members for providing such entertainment, despite the many bruises, concussion and the fact that I have smelt of chlorine for the past four years. Thank you also to Helen and Penny, for putting up with my inability to

stay in touch, but for providing drink (mostly tea), cheese, mushrooms and much fun when we do manage to meet up.

I owe my family huge thanks for all their support. My Grandfather, Ron Davis, for providing a poor student with meals and pub trips. My sister, Katie, for replying to all my messages, and for (mostly) being on my side. Marvin and the boys, for looking after the sofa. My father for his proof reading and taxi services, for his help in Namibia despite illness, and for providing much entertainment without meaning to. In particular, a delirious trip back from Namibia, where he attempted to pay for flights at Frankfurt Airport in Namibian Dollars. And finally, I thank my mother for her 'super-freak' exercise regime, and the never-ending supply of caffeine, which I still cannot compete with.

Contents

Abstract 9

Acknowledgements 11

Contents 13

List of Figures 15

List of Tables 17

Abbreviations 21

CHAPTER 1: Introduction 23

- Introduction 25
- Genetic Markers 26
- Understanding human history in Africa 39
- Southern African populations: migrations and current distributions 41
  - 'Khoisan' 41
  - Bantu-speaking agriculturalists 50
  - Frontiers 53
- Cultural impact on genetic diversity 59
- Populations to study 62
  - Lesotho 62
  - Namibia 65
- Aims 69

CHAPTER 2: Materials and Methods 71

- Samples 71
  - Lesotho 71
  - Namibia 72
  - South Africa 74
  - Botswana 74
- Laboratory Methods 75
  - Extractions 75
  - mtDNA 76
  - Y chromosome 79
- Chapter-specific analyses 85
  - Chapter 3: Cultural impact on genetic diversity 85
  - Chapter 4: Frontiers and Southern African Bantu-speaking populations 91

Chapter 5: Hunter-gatherers, pastoralists and farmers: migrations and interactions in Namibian populations 96
CHAPTER 3: Cultural impact on genetic diversity 101
Introduction 101
Populations studied 105
Namibia 105
Lesotho 109
Results 111
Namibia 111
Lesotho 119
Discussion 134
CHAPTER 4: Frontiers and Southern African Bantu-speaking populations 141
Introduction 141
Samples 144
Results 146
Discussion 177
CHAPTER 5: Hunter-gatherers, pastoralists and farmers: migrations and interactions in Namibian populations 183
Introduction 183
Pastoralism 183
Who are the Damara? 185
Namibian Bantu populations 188
Results 190
Discussion 231
CHAPTER 6: Conclusion 241
Appendix 1 247
Participant Questionnaire 247
Appendix 2 246
Supplementary Table 1 248
Supplementary Table 2 250
Supplementary Table 3 252
Supplementary Figure 1 255
Contributors 257
Bibliography 261

List of Figures

Figure 1.1: Regions of Africa 23
Figure 1.2: mtDNA phylogeny 33
Figure 1.3: The Y chromosome parsimony tree 35
Figure 1.4: Location of Southern African Ju, Tuu and Khoe-speaking populations 45
Figure 2.1: Sampling locations in Lesotho 72
Figure 2.2: Sampling locations in Namibia 73
Figure 2.3: SNPs genotyped in the initial multiplex 84
Figure 2.4: SNPs genotyped in the E1 and E2 multiplexes 84
Figure 3.1: Summary of the number of migrants for the Basotho data 132
Figure 4.1: Agriculturalist limit mid-second millennium AD 143
Figure 4.2: mtDNA haplogroup frequencies for the Lesotho, Xhosa and Zulu populations 147
Figure 4.3: NRY haplogroup frequencies for the Lesotho, Xhosa and Zulu populations 154
Figure 4.4: Wilcoxon tests between regions 163
Figure 4.5: Wilcoxon tests within regions for mtDNA vs. Y chromosome 164
Figure 4.6: L0d/L0k (mtDNA) and A3b1/B2b (NRY) haplogroup frequencies in Southernmost African populations 165
Figure 4.7: Mismatch distributions for all mtDNA haplogroups for population groups 170
Figure 4.8: Mismatch distributions for the mtDNA haplogroup L0d for population groups 170
Figure 4.9: Mismatch distributions for the mtDNA haplogroup L3e for population groups 171
Figure 4.10: Mismatch distributions for the mtDNA haplogroup L2a1 for population groups 171
Figure 4.11: Phylogenetic network of mtDNA L0d haplotypes in southernmost African populations 173
Figure 4.12: Phylogenetic network of mtDNA L3e haplotypes in southernmost African populations 174
Figure 4.13: Phylogenetic network of NRY E haplotypes in southernmost Africa 175
Figure 4.14: Phylogenetic network of NRY A3b1* haplotypes in southernmost African populations 176
Figure 5.1: Diagram depicting theories 1 and 2 of the possible routes for a Damara migration into Namibia 187
Figure 5.2: mtDNA haplogroup frequencies in the Namibian populations 194
Figure 5.3: NRY haplogroup frequencies in the Namibian populations 195

Figure 5.4: NRY E sub-haplogroup frequencies in the Namibian populations..... 196
Figure 5.5: Principal component analysis plot of mtDNA haplogroup frequencies for Sub-Saharan African populations 205
Figure 5.6: Principal component analysis plot of mtDNA haplogroup frequencies for Southern African and East African populations 205
Figure 5.7: Principal component analysis plot of mtDNA haplogroup frequencies for South East and South West African populations 206
Figure 5.8: Principal component analysis plot of NRY haplogroup frequencies for Sub-Saharan African populations 206
Figure 5.9: Principal component analysis plot of NRY haplogroup frequencies for Southern African populations 207
Figure 5.10: Principal component analysis plot of NRY haplogroup frequencies, including E sub-haplogroups for South East and South West African populations..... 207
Figure 5.11: mtDNA lineage sharing between individuals in the six Namibian populations and regions in Sub-Saharan Africa 211
Figure 5.12: mtDNA lineage sharing between haplotypes in the six Namibian populations and regions in Sub-Saharan Africa 212
Figure 5.13: NRY lineage sharing between individuals in the six Namibian populations and regions in Sub-Saharan Africa 216
Figure 5.14: NRY lineage sharing between haplotypes in the six Namibian populations and regions in Sub-Saharan Africa 217
Figure 5.15: Phylogenetic networks for the NRY A3b1 haplotypes – 15STRs..... 219
Figure 5.16: Phylogenetic network for NRY haplogroup E1b1b1g-M293..... 228
Supplementary Figure 1: Map showing locations of populations used in this work 255

List of Tables

Table 1.1: DNA that can be analysed 26
Table 1.2: Classification of Southern African click-languages 44
Table 2.1: Samples requiring HVS-II analysis 78
Table 2.2: Primers used in initial SNP multiplex 81
Table 2.3: Primers used in E1 SNP multiplex 82
Table 2.4: Primers used in E2 SNP multiplex 83
Table 2.5: Trio combinations 86
Table 2.6: Matrix between population geographic distances (in km) for the Basotho populations 88
Table 3.1: Owambo tribes 106
Table 3.2: Nama tribes 107
Table 3.3: Damara sub-groups 108
Table 3.4: Clans of Lesotho 110
Table 3.5: Summary of ethnographic information available for Namibia 111
Table 3.6: Population pairwise F_{ST} values and (p values) among Namibian ethnic groups 113
Table 3.7: Frequency of female parent-male parent pairs with matching or different ethnic groups, for the Namibian populations 115
Table 3.8: Frequency of trio combinations for clan affiliation in Namibian ethnic groups 116
Table 3.9: Frequency of trio combinations for places of birth in Namibian ethnic groups 118
Table 3.10: Number of migrants in the Namibian ethnic groups 119
Table 3.11: Population pairwise genetic distance (F_{ST}) & (p value) by ethnic group, Lesotho . 120
Table 3.12: Frequency of female parent-male parent pairs in Lesotho 121
Table 3.13: Frequency (%) of clans present in the Basotho samples 122
Table 3.14: Frequency of different trio combinations for clan affiliation in the Basotho 123
Table 3.15: Sample size and gene diversity (with 95% confidence intervals) for mtDNA and Y STRs for four Basotho clans 124
Table 3.16: Population pairwise F_{ST} value and (p value) for mtDNA for four Basotho clans 124
Table 3.17: Population pairwise F_{ST} value and (p value) for NRY for four Basotho clans 125
Table 3.18: NRY and mtDNA genetic differentiation for the Basotho clans 125
Table 3.19: Frequency of different trio combinations for sub-clan affiliation in the Basotho . 125
Table 3.20: Frequency of different location trio combinations in the Basotho 127

Table 3.21: Number of migrants in the Basotho 127
Table 3.22: NRY and mtDNA genetic differentiation for the Basotho samples, and other populations from published works with various residential norms 131
Table 3.23: Absolute values of the slope of the regression line (1/Nb) for the Basotho data, calculated using F_{ST} and F'_{ST} values 132
Table 3.24: F_{ST} and Nm ratios in Lesotho villages for NRY and mtDNA 133
Table 4.1: Frequencies of mtDNA haplogroups across Southern African populations 148
Table 4.2: Frequencies of NRY haplogroups across Southern African populations..... 151
Table 4.3: Percentages of NRY E sub-haplogroups for samples genotyped in this work and those from previous studies..... 152
Table 4.4: Population pairwise F_{ST} values for mtDNA haplogroup frequencies and haplotypes, comparing the Lesotho, Xhosa and Zulu populations with other Southern African populations from previous studies..... 156
Table 4.5: Population pairwise F_{ST} values for NRY haplogroup frequencies, comparing the Lesotho, Xhosa and Zulu populations with other Southern African populations from previous studies 159
Table 4.6: Population pairwise F_{ST} values for NRY haplotypes, comparing the Lesotho, Xhosa and Zulu populations with other Southern African populations from previous studies 160
Table 4.7: Summary statistics for mtDNA for the SE Bantu-Click (Southernmost) and SE Bantu (South-East) populations..... 168
Table 5.1: Number of individuals per mtDNA haplogroup in the Namibian populations 193
Table 5.2: Number of individuals per NRY haplogroup in the Namibian populations 193
Table 5.3: Gene diversity for the Namibian populations..... 198
Table 5.4: Population pairwise F_{ST} values for mtDNA haplotypes, comparing the Owambo and Ju|'hoansi populations with other Southern African and East African populations from previous studies 200
Table 5.5: Population pairwise F_{ST} values for NRY haplotypes, comparing the Namibia populations with Southern African and East African populations..... 202
Table 5.6: mtDNA haplotype sharing between the Namibian populations and other regions of Sub-Saharan Africa 208
Table 5.7: NRY haplotype sharing (based on 10STRs) between the Namibian populations and other regions of Sub-Saharan Africa 214
Table 5.8: Haplotype sharing for NRY haplogroups A2 and A3b1, 15 STRs..... 220
Table 5.9: Haplotype sharing for NRY haplogroup A3b1, 10 STRs..... 221

Table 5.10: A3b1 gene diversity 223
Table 5.11: NRY E-M293 haplotype sharing, 10 STRs 225
Table 5.12: NRY E-M293 haplotype sharing, 8 STRs 226
Table 5.13: E-M293 gene diversity 229
Table 5.14: STR variance for E-M293 230
Supplementary Table 1: Populations used in analyses in Chapter 4 248
Supplementary Table 2: Populations used in mtDNA comparisons in Chapter 5 250
Supplementary Table 3: Populations used in NRY comparisons in Chapter 5 252

Abbreviations

AD	Anno Domini
AMOVA	Analysis of Molecular Variance
BC	Before Christ
BP	Before Present
bp	Base pairs
DNA	Deoxyribonucleic Acid
Hgs	Haplogroups
HVS-I	Hypervariable segment one of the mitochondrial DNA control region
HVS-II	Hypervariable segment two of the mitochondrial DNA control region
IBD	Isolation By Distance
kb	Kilo bases
km	Kilometre
kya	Thousand years ago
Mb	Mega bases
MSA	Middle Stone Age
mtDNA	Mitochondrial deoxyribonucleic acid
N_e	Effective population size
NRY	Non-Recombining section of the Y chromosome
PCA	Principal Components Analysis
RFLP	Restriction Fragment Length Polymorphism
SAK	South African 'Khoisan'
SNP	Single Nucleotide Polymorphism
SSD	Sum of Squared Deviation
STR	Short Tandem Repeat, or microsatellite

TMRCA	Time to Most Recent Common Ancestor
ya	Years ago
YSTR	Y chromosome Short Tandem Repeat

CHAPTER 1: Introduction

Note

Geographic definitions

The majority of this work will focus on events that have occurred in the southern half of the African continental landmass. As past population movements did not conform to modern-day country boundaries, it will sometimes be necessary to use terms other than country names to describe locations on the continent. When reference is made to 'Sub-Saharan Africa', this includes all countries and part-countries that are below the Sahara desert; it will not include areas in the Sahel zone. 'Southern Africa' is harder to define, as no obvious geographical boundary exists which is not controversial in one way or another. Here, the term will be used as in Nurse et al. (1985), and will include: the whole of South Africa, Botswana, Lesotho and Swaziland; the whole of Namibia except for Eastern Caprivi; and any of Mozambique which lies below the Limpopo River. See Figure 1.1.

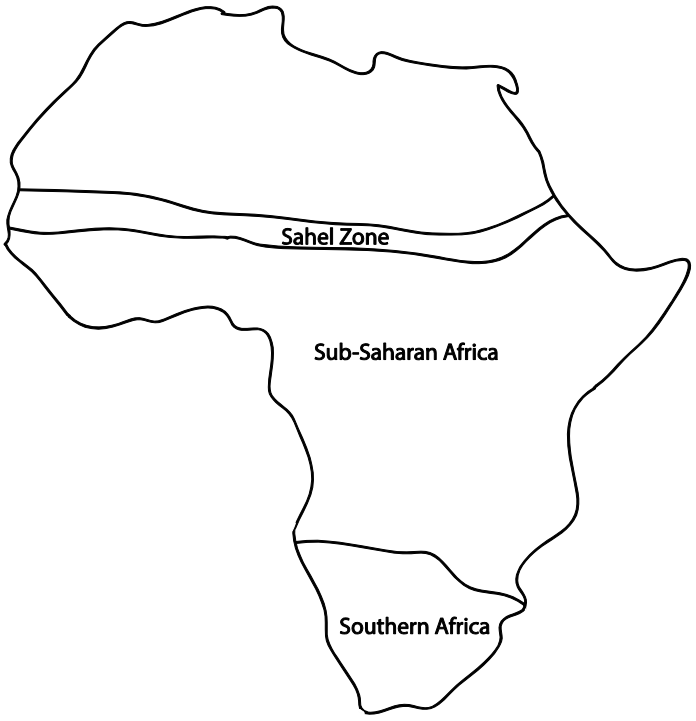


Figure 1.1: Regions of Africa.

Use of ‘Khoisan’

The use of ‘Khoisan’ as a grouping of Southern African hunter-gatherers and herders is discussed in more detail in the Introduction. It is acknowledged that the term ‘Khoisan’ groups together populations that are linguistically and biologically distinct. Populations in this work will be referred to as ‘hunter-gatherer’ or ‘pastoralist’ or more-specifically by language group as an alternative to ‘Khoisan’ where possible. However, if the term ‘Khoisan’ is used, it is used in quotation marks, and is meant as a loose grouping of all hunter-gatherers and pastoralists from the region, indicating populations speaking languages that have previously been classified as ‘Khoisan’. It is not meant to suggest that these populations are closely related, nor does it have any political meaning.

Introduction

Around eight to five million years ago the human and chimpanzee lineages diverged. What happened after this split, and indeed before it, is of great interest to modern biologists; not only because it is a history that involves each and every one of us, but also because of the benefits that understanding human evolution and genetic variation between individuals and populations are thought to have. By studying patterns of genetic structure it is possible to gain a crucial understanding of the genetic basis of human disease, and why some people respond differently to pharmaceutical treatment.

The study of human evolution, although perhaps not recognisable in today's sense, has been considered as far back as the 5th and 6th centuries BC, when Plato and Aristotle were interested in the origins of humans and humanity (Wood 2005). In the 1st century BC, the Roman philosopher Lucretius proposed an evolutionary hypothesis on the creation and extinction of life forms (Lucretius and W.H.D. Rouse 1992). This work went on to influence many evolutionary theorists in Europe, including Lamarck, and arguably Darwin. Following the collapse of the Roman Empire, ideas from Genesis became more favoured, however, and it was not until the Enlightenment, with people such as Francis Bacon and Andreas Vesalius that ideas began to change again in favour of a more scientific approach (Wood 2005).

In 1863, Thomas Henry Huxley wrote a review of the differences between modern humans, chimpanzees and gorillas (Huxley 1863). This work marked the beginning of modern studies of human evolutionary genetics, as it concluded that anatomical differences are less marked between modern humans, chimpanzees and gorillas, than they are between chimpanzees, gorillas and the orangutan. This prompted Darwin to propose that the ancestors of modern humans were more likely to be from Africa than elsewhere (Darwin 1871). Since these

publications, much work involving archaeology, linguistics, and more recently genetics, has been done on understanding the origins of modern humans, and how we have migrated and interacted with each other since.

Genetic Markers

Since the start of the 20th century numerous studies have used genetic data to examine the human past. These studies have played a significant role in shaping current views about events in human history, and have helped guide interpretation of evidence from other disciplines, such as archaeology and linguistics.

Human genome structure

The human genome is billions of bp in length (approximately three billion) (Venter et al. 2001), meaning there are vast amounts of DNA that can be studied and compared between individuals and populations (Table 1.1).

	Genomic compartment			
Feature	**Autosomes**	**X chromosomes**	**NRY**	**mtDNA**
Location	Nuclear	Nuclear	Nuclear	Cytoplasmic
Inheritance	Bi-parental	Bi-parental	Uni-parental	Uni-parental
Ploidy	Diploid	Haploid-diploid	Haploid	Haploid
Relative N_e	4	3	1	1
Recombination rate	Variable	Variable	Zero	Zero
Mutation rate	Low*	Low*	Low*	High

Table 1.1: DNA that can be analysed.
Taken from Garrigan and Hammer (2006). *This refers to a SNP mutation rate: STR mutation rates in the autosomes, X chromosome and NRY are high.

Autosomes and X chromosome

The majority of our DNA is made up of the autosomes, and along with the X chromosome, these are not uni-parentally inherited, and do undergo recombination, meaning that they have many evolutionarily independent loci (Hammer et al. 2008). The fact that the autosomes represent the majority of the genome, means that there are a large number of polymorphisms to study (Pakendorf et al. 2011).The presence of recombination might be important in studying TMRCA and effective population size (N_e) as it might reduce the impact of selection.

X chromosomes are present in two copies in females and one in males, and autosomes are present as two copies in both sexes. This means that the N_e of the autosomes and X chromosome are different to one another, and also different to those of the mtDNA and Y chromosome (Table 1.1). It has been suggested that a larger N_e, like that of the autosomes and X chromosome, will result in a deeper ancestry than that of the haploid regions, meaning that markers in these regions can be used to study events that might have occurred before the TMRCA of the mtDNA and Y chromosome (Garrigan and Hammer 2006). Sex biased population processes can also be detected by comparing levels of genetic variation between the X chromosome and the autosomes. Previous studies comparing the X chromosome and the autosomes have shown that the effective population size of males and females might differ, but have produced conflicting results as to which had the larger size (Hammer et al. 2008; Keinan et al. 2009). Large-scale autosomal data can also be used for investigating human demographic events, for example relating to shared ancestry in populations (Rosenberg et al. 2002; Hellenthal et al. 2008; Li et al. 2008; Tishkoff et al. 2009; Henn et al. 2011). These studies are, however, costly and can require complicated analysis. As a result, to date the majority of research on human population genetics has concentrated on haploid markers such as the mtDNA and NRY in the first instance. This has allowed an initial understanding of the history of a population or region, before progression on to larger autosomal studies.

mtDNA & NRY

mtDNA and the NRY are uni-parentally inherited, haploid, and probably do not undergo recombination (Table 1.1). This lack of recombination allows haplotypes to pass between generations affected only by mutation; making their history easier to interpret. mtDNA and the NRY also have a four times reduction in N_e compared to the autosomes, which leads to increased genetic drift in the absence of selection, and can lead to large differences between populations. This property means that mtDNA and the NRY are useful tools for studying demographic processes, as population geographic structure is often present. It is important to consider both male (NRY) and female (mtDNA)-associated markers in order to take into account that the different sexes might have made different contributions to population structure in the past. Differences in patterns of diversity of the mtDNA and NRY might demonstrate sex-biased processes (Pilkington et al. 2008).

mtDNA

mtDNA is located in the mitochondria of cells, not the nucleus, and is a circular molecule with a length of approximately 16.5 kb (Anderson et al. 1981). It is now generally accepted that the mitochondrion was originally an endosymbiotic bacterium that was engulfed by proto-eukaryotic cells about 1.5 billion years ago (Margulis 1981). Human mtDNA has been fully sequenced (Anderson et al. 1981; Andrews et al. 1999), and there are only 13 polypeptides coded for in the mtDNA, all of which are involved in oxidative phosphorylation (Wallace et al. 1999). The region of the mtDNA that has been studied the most in human population genetics is the control region (approximately 1200 bp long), which is made up of two hypervariable regions, or segments, (HVS-I and HVS-II) which are each approximately 350 bp in length (Tishkoff and Verrelli 2003).

Mitochondria are present at levels of hundreds to thousands of copies in cells, at different frequency in different cell types. The high copy number of mitochondria can be problematic, however, as it can mean that different types of mtDNA might be present within an individual – this is heteroplasmy. Levels of heteroplasmy are, however, low and if heteroplasmy is present in an individual (it occurs in about 14% of the population), the second mtDNA type is probably present only at very low frequency (about 1%) within that individual (Pakendorf and Stoneking 2005).

For many years the maternal inheritance of mtDNA was accepted as definitive, however, more recently a study found paternal mtDNA present in an individual (Schwartz and Vissing 2002), suggesting that maybe a small amount of paternal inheritance might occur (Jobling et al. 2004; Pakendorf and Stoneking 2005). However, it is known that mitochondria from sperm are destroyed in the oocyte (Manfredi et al. 1997; Shitara et al. 1998), along with paternal mtDNA (Sutovsky et al. 1999, 2000), and anyway, there are thousands more mitochondria present in the oocyte compared to the sperm (Jobling et al. 2004), making it likely that if any paternal mitochondria did survive, they would be extremely diluted. In fact, many other studies have shown no indication of paternal mtDNA (Giles et al. 1980; Jazin et al. 1998; Howell et al. 2003), and it is therefore likely that the observed case is as a result of a breakdown in the process that normally destroys the paternal mitochondria (Pakendorf and Stoneking 2005)

mtDNA's lack of recombination has also recently been called into question. A number of studies reported evidence of recombination in human mtDNA (Awadalla et al. 1999; Eyre-Walker et al. 1999; Hagelberg et al. 1999; Kivisild et al. 2000). However, these studies have been shown to be inaccurate; using incorrect data, or incorrect statistical methods (Pakendorf and Stoneking 2005).

mtDNA has a high base substitutional rate; about 10 times higher than the average rate in nuclear DNA (Brown et al. 1979), being estimated at 0.017 x 10^{-6} substitutions per site per year for the mtDNA excluding the control region (Ingman et al. 2000). Consequently, there are a large number of different mtDNA sequences in the human population, although parallel mutations and reversions must be taken into consideration at fast-mutating sites. It is known that the mutation rate in the hypervariable regions is higher, however, an overall consensus as to the exact rate does not exist. Numerous attempts have been made to study it as it is important in understanding population history. Studies have used analysis of pedigrees to try and calculate a rate, estimating 0.0 – 1.46 x 10^{-6} substitutions/site/year (Bendall et al. 1996; Parsons et al. 1997; Howell et al. 2003; Pakendorf and Stoneking 2005), while phylogenetic studies have estimated 0.075 – 0.165 x 10^{-6} substitutions/site/year (Ward et al. 1991; Stoneking et al. 1992; Hasegawa et al. 1993; Tamura and Nei 1993). Some of the differences in mutation rates might be as a result of hot spots in the control region; some sites mutating at a much higher rate than others (Pakendorf and Stoneking 2005). In fact, different sites can have different, site specific, mutation rates. A commonly used mutation rate has been that of Mishmar et al (2003), at 1.26 x 10^{-8} substitutions/site/year. Soares et al (2009) suggest a rate for the whole mtDNA of 1.665 x 10^{-8} substitutions/site/year, taking into account selection and saturation.

NRY

Y chromosomes are involved in sex determination, are haploid, and are transmitted from father to son (Jobling and Tyler-Smith 2003). Some sections of the Y chromosome undergo recombination with the X chromosome; these are pseudoautosomal regions. The majority of the Y chromosome, however, does not undergo recombination (about 90%); this portion is referred to as the Non-Recombining section, the NRY. It accounts for approximately 2% of the human genome and is around 60 Mb in length. This lack of recombination has made the NRY

important in studies of human population genetics. The Y chromosome does not contain genes important for human survival that are not shared with the X chromosome, and in fact, contains very few genes at all; the non-pseudoautosomal region codes for only 27 proteins (Skaletsky et al. 2003).

The average mutation rate of the NRY is lower than that for the mtDNA, however, it is higher than other nuclear DNA (Jobling et al. 2004). In fact, the NRY has a high number of different types of repeat polymorphisms. This is because of the Y chromosome being passed through the male germline, which is more mutagenic than the female due to the high number of meiosis that occur in the continuous production of sperm. However, although this causes high between species divergence, there is relatively low within species diversity due to genetic drift, due to the small N_e of the NRY.

Polymorphisms

Classical markers

The first human genetic polymorphism to be defined and used for anthropological purposes was the ABO blood group system: ABO groups are carried at different frequencies in different populations (Hirschfeld and Hirschfeld 1919). Further work by Emile Zuckerkandl on peptides (Zuckerkandl et al. 1960), Morris Goodman on albumin (Goodman 1963) and Vince Sarich and Allan Wilson on variations in protein structure (Sarich and Wilson 1966) supported the conclusions of Huxley and Darwin; that modern humans and African apes are very closely related (Huxley 1863; Darwin 1871).

SNPs

Single nucleotide polymorphisms (SNPs) are mutations that occur when a single nucleotide in the genome sequence is altered. This can be a substitution, or a deletion or insertion (indels). A substitution is known as a transition when a pyrimidine base is substituted for another pyrimidine base (e.g. C for T), or a purine base for another purine base (e.g. A for G); and a transversion when a pyrimidine is exchanged for a purine or vice versa. Transitions occur twice as often as transversions, and indels are considerably less frequent (Jobling et al. 2004). For a mutation to count as a SNP it must occur in at least 1% of the population. Occurring approximately once every 1000 bases in the HapMap database, SNPs are the most common form of genetic variation between individuals, and millions of human SNPs have been discovered (The International HapMap Consortium 2007). However, SNPs have a relatively low mutation rate; approximately 10^{-8} mutations/generation (Kayser et al. 2000). Comparison of sequences between individuals at known, or even previously unknown, SNPs provides a measurement of genetic variation which can be compared across populations. SNP data can be problematic, however, particularly with Y chromosomal analyses, due to ascertainment bias (Jobling and Tyler-Smith 2003). Ascertainment bias results from methodology which favours certain markers or samples; if markers are chosen because they are highly variable in one population, when analysed in another region they might show reduced variation. This does not necessarily mean that these other populations have lower genetic diversity, but perhaps just that they are variable at different markers.

Over time mutations accumulate sequentially in the DNA, for example in mtDNA or NRY lineages, resulting in different mtDNA or NRY in different populations, in different parts of the world. Studies have been conducted in order to analyse these differences and to try and understand more about human history. A pioneering study was carried out on mtDNA in 1987 by Rebecca Cann, Mark Stoneking, and Allan Wilson (Cann et al. 1987). By comparing mtDNA

from 147 modern humans across the world, they showed that the tree constructed from this data has a deep African branch, and that the Sub-Saharan African branch contained the most variation (Cann et al. 1987). The time depth for mtDNA in Africa is between 100,000 – 200,000 years (Kittles and Keita 1999). Recent studies have usually involved sequencing of the control region or amplification of the complete mtDNA molecule. This has allowed the identification of SNPs in the mtDNA, which has made possible the elucidation of a human mtDNA phylogeny, which is classified into haplogroup clades, named using alphabetical letters (Figure 1.2). The different haplogroups have different distributions, with varying frequencies across populations depending upon where they originate. Analysis of which haplogroups are present in populations and their frequencies therefore provides information about a population's history. Nomenclature used for mtDNA haplogroups in this work will follow that from the most recent PhyloTree build (Build 14) (van Oven and Kayser 2009).

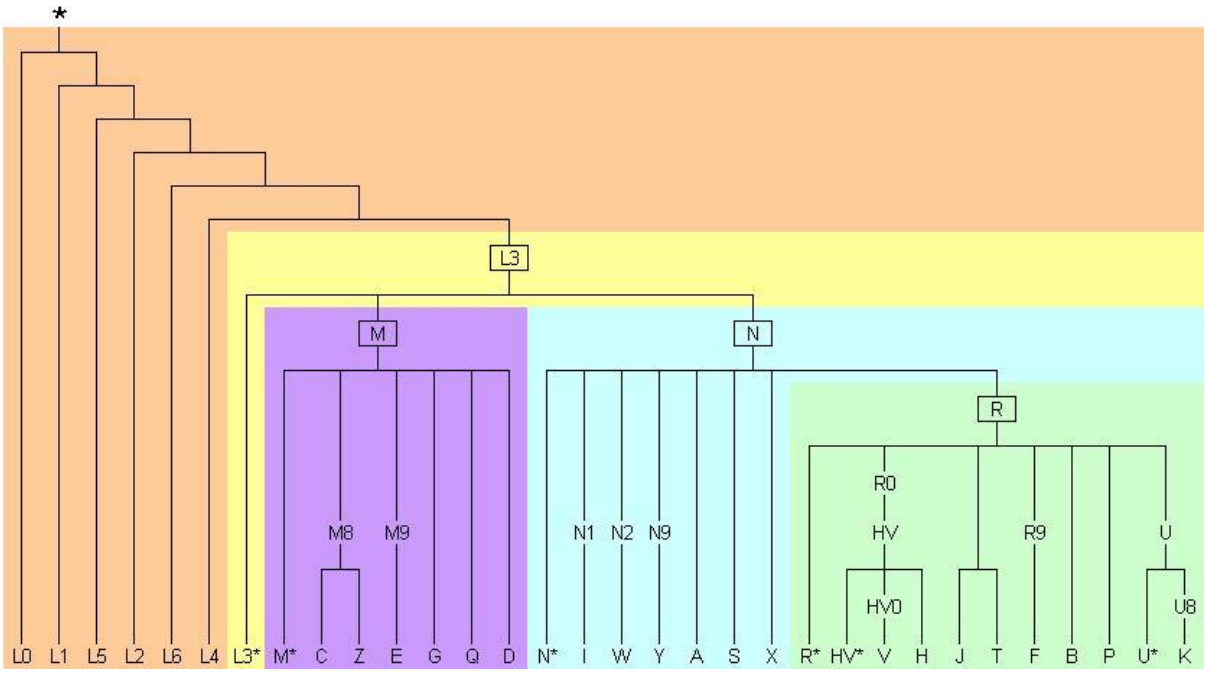


Figure 1.2: mtDNA phylogeny.
From van Oven and Kayser (2009). The star at the top of the diagram indicates the root of the tree, and thus the most recent common ancestor of females. Stars after haplogroups (letters) indicate that this includes all other descendant lineages not shown of that particular clade.

The major mtDNA haplogroups in Sub-Saharan Africa are the 'L-type' haplogroups, which are considered to be at the root of the human evolutionary tree (Salas et al. 2002) and are Africa-specific. L0 contains the most recent common ancestor of human mtDNA (Salas et al. 2002). L3 gave rise to the haplogroups that contain all variation outside of Africa (M, N and R).

The complete sequence of the euchromatic region of the NRY was published in 2003 (Skaletsky et al. 2003), but studies using NRY polymorphisms had been carried out before this. The first polymorphic marker on the NRY was discovered in 1985 (Casanova et al. 1985), and by the early 21st century over 200 SNPs had been found (Shen et al. 2000; Underhill et al. 2000; Hammer et al. 2001). In fact, the number of NRY SNPs now described is around the 600 mark (Karafet et al. 2008), and increasing numbers are becoming available. In particular, more SNPs will become available as a result of whole genome sequencing projects such as the 1000 Genomes project (The 1000 Genomes Project Consortium 2010). SNPs have been used to define Y chromosome lineages (Underhill et al. 2001; Jobling and Tyler-Smith 2003; Karafet et al. 2008), and an NRY phylogeny (Figure 1.3) has been constructed using these SNP markers (The Y Chromosome Consortium 2002; Karafet et al. 2008). Nomenclature used for NRY SNP markers discussed in this work will follow the nomenclature of the Y Chromosome Consortium (The Y Chromosome Consortium 2002; Karafet et al. 2008).

Some NRY haplogroups are specific to certain geographic regions or populations, a characteristic which can help understand population migrations and admixture. This usefulness is increased by the fact that the low N_e of the NRY, and the subsequent genetic drift, can lead to large differences between populations. Haplogroups A and B are the deepest branches of the tree, originating from, and predominantly found in, Africa (Underhill et al. 2001; Chiaroni et al. 2009). Haplogroup A is typically found in indigenous 'Khoisan' populations, and haplogroup B (B2b in particular) in African Pygmies in Central Africa (Batini et

al. 2011). Haplogroup E is also found mostly in African populations. Other haplogroups on the tree are associated with populations involved in or descendant from those that migrated out of Africa, and have geographically distinct distributions (Jobling et al. 2004). This supports the Out of Africa hypothesis for the origin of modern humans.

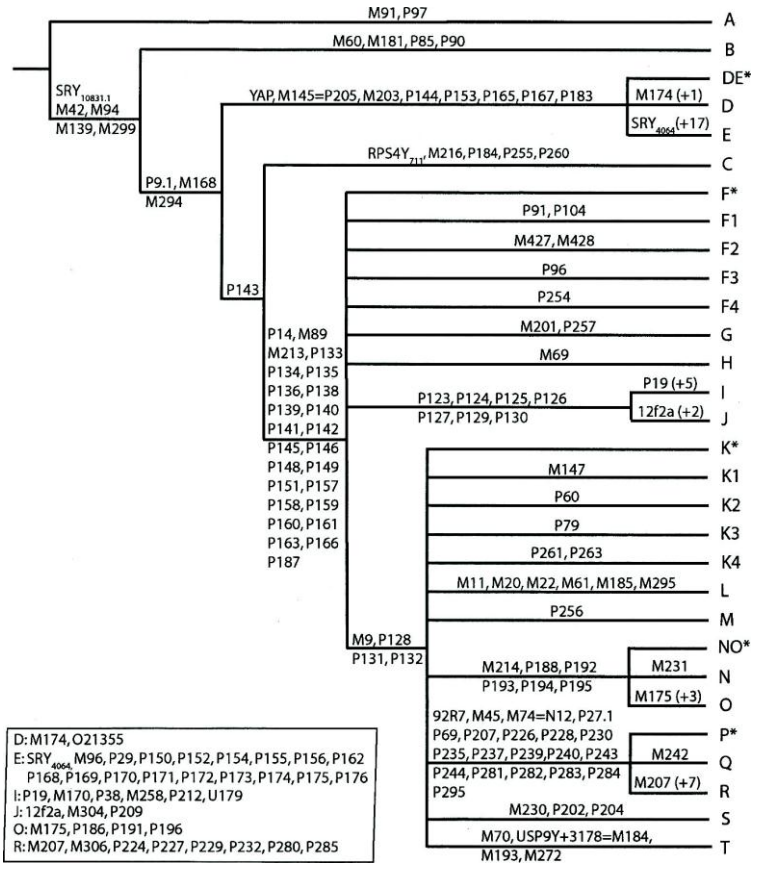


Figure 1.3: The Y chromosome parsimony tree.
From Karafet et al. (2008). Haplogroup names are at the tips of the tree, and mutation names on the branches.

STRs

Short Tandem Repeats (STRs), or microsatellites, are repeats of sequence with repeat units of 1 – 6 base pairs (bp) in length, and copy numbers of about 7 – 30. STRs are used in studies of evolutionary genetics as their copy number is often variable. The majority of mutations are characterised by an increase or decrease of only a single repeat unit (Brinkmann et al. 1998; Xu

et al. 2000). STRs have mutation rates that are higher than those for SNPs, meaning that more recent evolutionary events can be studied. STR variation can be used to estimate the time to most recent common ancestor (TMRCA). There are, however, problems with dating using STRs, including (according to Jobling and Tyler-Smith 2003): the mutation rate is not know precisely; mutation rates differ between STR alleles; there is disagreement over what should be used as a generation time (anywhere between 20 – 35 years have been used in different studies); variation does not accumulate at a steady rate. Despite these issues, mutation rates for STRs have been estimated. This is usually done by pedigree analysis (Brinkmann et al. 1998; Xu et al. 2000); examination of pedigrees separated by a large number of generations (Heyer et al. 1997); or by detection of mutations in sperm DNA molecules (Holtkemper et al. 2001), due to mutation being maintained and/or arising in these populations.

STRs in the NRY (YSTRs) have been analysed, providing more detailed information about populations, and used to date lineages. As mentioned above, however, this dating using YSTRs is controversial, as calculation of a mutation rate for the NRY is not agreed upon, and the molecular characteristics of the different STRs might have different age estimates (Busby et al. 2012). A number of studies have proposed mutation rates, however, these should be used with caution, due to considerable disagreements in rates between them. For example, although Gusmao et al. (2005) (mutation rate of 1.998×10^{-3}), Kayser et al. (2000) (mutation rate of 2.8×10^{-3}), and Dupuy et al. (2004) (mutation rate of 2.3×10^{-3}) report fairly similar results, a study using phylogenetic networks and both YSTRs and SNPs, gives a mutation rate of 2.6×10^{-4} mutations/20 years (Forster et al. 2000). This is considerably different. Zhivotovsky et al. (2004) report an average mutation rate per repetition per STR of 6.9×10^{-4}/25 years, again using STRs and SNPs.

Despite no concordance over mutation rates, dating using YSTRs is still often used to calculate TMRCAs for lineages and for events such as the TMRCA of modern humans in Africa. Estimates for the TMRCA of modern humans using YSTRs have been much more recent than those for mtDNA. For example, dating using 8 STRs has suggested a TMRCA of 46 (16 – 126) kya (Pritchard et al. 1999). Dating using SNPs has also produced relatively recent estimates, e.g. 59 (40 – 140) kya (Thomson et al. 2000). This N_e for the NRY could be lower due to selection, greater variance in male reproductive success, greater variance in mtDNA mutation rates, or differences in generation times used (Jobling and Tyler-Smith 2003; Garrigan and Hammer 2006). It has also been shown that not all STRs are equally useful in calculating TMRCAs, and this might affect dates produced (Busby et al. 2012). In fact a more recent study by Cruciani et al. (2011), which uses sequencing of a large region of the NRY, produces an older estimate of about 140 kya.

Summary

A vast number of genetic studies carried out using the above markers and regions support, or are consistent with, the proposed African origin for modern humans (Cann 1988; Vigilant et al. 1991; Nei 1995; Tishkoff et al. 1996; Quintana-Murci et al. 1999; Ingman et al. 2000; Salas et al. 2002), demonstrating that genetic diversity in Africa is extremely high; more so than that around the rest of the world. Recent work using autosomal markers (Rosenberg et al. 2002; Hellenthal et al. 2008; Li et al. 2008; Tishkoff et al. 2009; Henn et al. 2011) has supported results from mtDNA and Y chromosome studies, showing that there is vast genetic diversity within Africa. It is likely that a population bottleneck occurred as populations migrated out of Africa, thus causing a reduction in genetic diversity in non-African populations relative to Africans (Campbell and Tishkoff 2010). Genetic estimates for the origin of *Homo sapiens* also tally with paleontological estimates, falling between 100 – 200 kya for mtDNA estimates (Cann

et al. 1987; Ingman et al. 2000; Salas et al. 2002; Behar et al. 2008), and 200 – 65 kya for NRY estimates (Scozzari et al. 1999; Underhill et al. 2000; Cruciani et al. 2011).

Relevant haplogroup distributions for both the mtDNA and NRY are discussed in detail later in the Introduction in sections relating to populations present in Southern Africa, and also in Chapters 4 and 5.

Understanding human history in Africa

In addition to the evidence showing that Africa displays the most genetic diversity, archaeological, palaeontological and linguistic evidence also supports the idea that modern humans originated in Africa about 200 kya (Stringer 2003; White et al. 2003; Campbell and Tishkoff 2010). Africa is home to over 2000 language groups; nearly a third of the languages in the world (Lewis and SIL International 2009), and is also home to the earliest known modern human remains (Stringer 2003). Cultural evidence also supports this view, with effectively the only signs of behavioural modernity during the Middle Stone Age (MSA) (about 250 – 40 kya) being in Africa (Mitchell 2002; Scheinfeldt et al. 2010). As the place where modern humans originated, and where contemporary populations display the highest genetic diversity, studying African populations is essential for understanding our past.

Where in Africa?

Despite all the evidence supporting an African origin for modern humans, there is still debate as to where exactly in Africa *Homo sapiens* originated. Fossil remains from East Africa, dated to around 195 – 150 kya, seem to show the earliest evidence of traits associated with anatomically modern humans (Stringer 2003; White et al. 2003), and it has been suggested that populations from other continents display genetic diversity that is a subset of that in Eastern Africa (Ramachandran et al. 2005; Gonder et al. 2007). It has been argued, however, that the presence of modern morphological traits in other areas of Africa before 100 kya cannot be completely ruled out (Reed and Tishkoff 2006). It is also possible that results are biased towards Eastern Africa, as in regions such as Central and West Africa very limited archaeological research has been conducted with a view to answering this question, and in regions outside East Africa specimens tend to be less well preserved. Reed and Tishkoff (2006) even suggest that a multiregional model in which *Homo sapiens* evolved from *Homo erectus*

concurrently in multiple locations, might not be as unlikely within Africa as it is for Out of Africa. Recent archaeological research has in fact shown marine resource use and pigment use at Pinnacle Point on the south coast of South Africa occurring earlier than has been discovered anywhere else, at 165 kya (Marean et al. 2007), and a 100,000 year old ochre-processing workshop has also been found at Blombos Cave on the southern Cape coast of South Africa (Henshilwood et al. 2011). This evidence of modern human behaviour in South Africa at a much earlier date than previously expected, not only increases evidence for early modern humans in Southern Africa, but also adds sway to the argument for modern human behaviour and anatomy evolving at similar times. Additionally, climatic evidence suggests that until about 60,000 – 70,000 years ago Southern Africa was more hospitable than Eastern Africa (Scholz et al. 2007), and Southern Africa might therefore have been a refuge for modern humans (125 – 60 kya) (Compton 2011). Further, a recent linguistic study supports a South-Western African origin of language (Atkinson 2011).

Genetic studies have also added support for a Southern African origin of *Homo sapiens*. A study from 2009 proposed an origin in Southern Africa based on data from over 1000 microsatellites in 121 African populations (also 4 African-American populations, and 60 non-African populations) (Tishkoff et al. 2009). However, the number of Southern African samples in this study was small. A study of over 580,000 SNPs in several hunter gatherer populations suggests that populations of 'Khoisan' are among the most diverse human populations (Henn et al. 2011), and that the patterns observed support an origin in Southern Africa, and not Eastern Africa. As well as potentially playing a role in the origins of modern humans, Southern Africa is one of the few places in Africa where hunter gatherer, pastoralist and agriculturist population remain. Despite this, relatively few genetic studies have concentrated on Southern Africa, especially when compared to other parts of the world.

However, even recent genetic studies indicating an important role for Southern Africa are still controversial, as according to Hublin and Klein (2011), the populations that now reside in Southern Africa might not have actually lived there 60,000 years ago. The place of origin of a population does not always equate to their place of residence, and searching for the place of origins of a population is difficult; instead it might be possible only to detect signatures of genetic lineages that have deep branches in phylogenies.

Southern African populations: migrations and current distributions

Africa is a large continent spanning the equator, meaning that the topography and climate across different regions of the continent varies greatly. Within Southern Africa alone, most of the different climates and topographies seen across the rest of the continent are present (Nurse et al. 1985). Even within relatively small regions, high mountain ranges can cause the temperature to vary greatly over small distances. This enormous variation in conditions is a large factor in the distribution of human populations discussed below. Agriculturalists were, for example, restricted by the climatic conditions that their crops could withstand, and so were prevented from settling in certain areas. Soil conditions are also extremely variable in Southern Africa; a vital factor for agricultural practices.

'Khoisan'

Southern African indigenous populations, often grouped together and referred to as 'Khoisan' or 'Khoesan', may have split from other modern humans from about 90,000 years ago (Behar et al. 2008). 'Khoisan' is a contentious grouping (Mitchell 2010), however, and in fact the Khoi (or Khoe) and the San (or Bushmen or hunter-gatherers) are separate groups (Barnard 1992; Mitchell 2002); with the term 'San' being ambiguous in itself (Guldemann and Stoneking 2008).

'San' often refers to populations of hunter-gatherers in Southern Africa, but is inaccurate for genetic purposes as it includes groups that speak different languages from different language families (Guldemann and Stoneking 2008). 'Khoe' refers to a linguistic family in Southern Africa, but groups together culturally and biologically distinct populations. The groups have often been linked together due to similarities in language and culture: 'Khoisan' languages include languages with click consonants. The grouping together of these in the 'Khoisan' language family is, however, made more debatable due to the group also containing indigenous populations from East Africa (the Sandawe and Hadza, both from Tanzania). The language family has high diversity within subclades, and high divergence among the subclades (Nurse 1997), thus grouping languages together that might be diverse enough to be counted as separate entities. It has therefore been suggested that the languages spoken in South Africa should be separated into a different branch (Southern African 'Khoisan', 'SAK') from the more divergent Sandawe and Hadza (Heine and Nurse 2000), especially as the SAK languages are found considerably further away geographically.

Genetic evidence supports the idea that the Hadza and Sandawe have more commonality with each other than either has to the SAK (Tishkoff et al. 2007a; Scheinfeldt et al. 2010). Some linguists have argued, however, that Sandawe and some SAK languages might actually share more linguistic similarity than either does to Hadza (Ambrose 1982; Guldemann and Vossen 2000; Scheinfeldt et al. 2010). Mitchell (2010) points out, however, that language groupings must be treated with caution, and cannot necessarily be used to infer ethnicity; for example, whether or not languages have spread when groups come into contact is not often known. This is also the case for the spread of the Bantu-languages (discussed below). Thus, without further evidence linking them together, it has been suggested that it is even best to count the various SAK families as separate groups (Guldemann and Stoneking 2008), referred to as: Tuu, Ju and Khoe (Guldemann and Stoneking 2008); or Southern, Northern and Central 'Khoisan' (or

'Khoesan') (Blench 2006); or !Ui-Taa, Ju and Khoe (Crawhall 2006) (Table 1.2). Tuu and Ju speakers are hunter-gatherers, while Khoe-speakers are both herders and hunter-gatherers (or were until recently). See Figure 1.4 for a map of the distribution of the Southern African Ju, Tuu and Khoe-speaking populations.

Until relatively recently, Southern Africa was occupied only by indigenous hunter-gatherer populations (Barnard 1992). By around the end of the first millennium BC, however, there is evidence of domesticated animals (sheep) in Southern Africa; with sheep bones at Blombos Cave, South-Eastern Western Cape Province dated to the first century AD, and bones at Spoegrivier (Namaqualand) dating to the first century BC (Henshilwood 1996; Vogel et al. 1997; Mitchell and Whitelaw 2005). Ancient DNA evidence also supports this dating, with analysis of cattle and sheep bones suggesting that sheep and cattle could both be found in northern Southern Africa around the first millennium BC (Horsburgh 2008; Mitchell 2010). Pleurdeau et al. (2012) record the oldest dates for sheep in southern Africa, supporting ideas that they were present at the end of the first millennium BC.

In the mid-1980s the general view was that these animals belonged to herding groups, and had been brought into the region by migrating Khoekhoen pastoralists around 2000 years ago from the Zambezi region, who had in turn acquired the animals from Bantu-speakers in the region (Elphick 1985; Smith 1990b; Fauvelle-Aymar and Sadr 2008). New evidence, however, means that today there is much less of a consensus as to the origins and owners of these animals, and the debate is still on-going (Mitchell 2010). Generally it seems to be agreed that the domesticated animals must have come from further north as none are indigenous to Southern Africa (Mitchell 2002), but beyond this, there is much disagreement.

Family		Language(s)			
Ju-≠hoan	≠Hoa				≠Hoã
	Ju		Northwest		!'O!Xũ, !Xũ
			Southeast		Ju|'hoã, ≠Au||ei
Tuu	Taa-Lower Nossob	Taa	West		*|Nu||en*, West !Xõ
			East		East !Xõ, *Kakia*, *N|amani*, N|u
		Lower Nossob			*|'Auni, |Haasi*
	!Ui				*!Gã!ne, ||Ku||e, ||Kx'au, ||Ng, 'Seroa', 'Sonqua', |Xam, ||Xegwi*
Khoe-Kwadi	Kwadi				*Kwadi*
	Khoe	Khoekhoe	North		*Eini*, Hai||om, ≠Aakhoe, Nama-Dama
			South		*Cape Khoekhoe, !Ora, Xiri*
		Kalahari	East	Shua	Cara, Danisi, *Deti,* Ts'ixa, |Xaise
				Tshwa	Cua, Kua, Tsua
			West	G||ana	G|ana, G|wi, ≠Haba (including N|haints'e, Qabekhoe, Ts'aokhoe)
				Kxoe	||Ani, Buga, G|anda, Khwe
				Nharo	Nharo

Table 1.2: Classification of Southern African click languages.
From Mitchell (2010). Extinct languages are marked by italics.

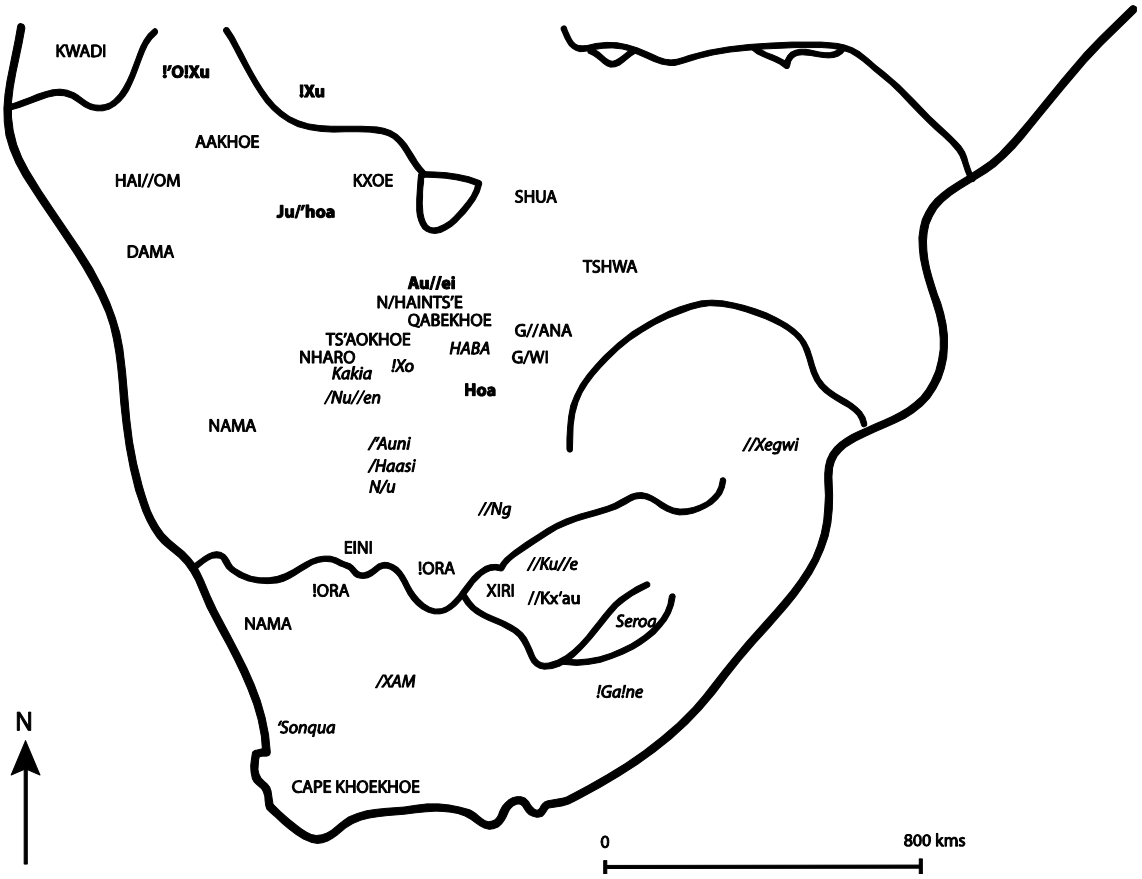


Figure 1.4: Location of Southern African Ju, Tuu and Khoe-speaking populations from Mitchell (2010).

Although it is undeniable that there are Khoekhoe-speaking herders in Southern Africa today, the idea that the Khoekhoen introduced the first livestock into the region is controversial. Proponents of the migrating Khoekhoen theory argue that linguistic evidence is consistent with the dates of the sheep bones, and also ceramic evidence, and suggest the Khoekhoen arrived at the Cape in around the first century AD (Ehret 1982a; Smith 1990b; Ehret 2008). But this idea has been challenged by those who have suggested that livestock were actually introduced into Southern Africa as a result of hunter-gatherers taking up herding prior to the Iron-Age (Denbow and Campbell 1986; Sadr 1998, 2003). It has been noted that evidence of sheep in Southern Africa predates the arrival of Bantu-speaking farmers and herders in Zambia, thus

questioning the idea that the Khoekhoen could have acquired livestock in this region before migrating southwards (Mitchell and Whitelaw 2005; Sadr and Sampson 2006; Sadr 2008). Sadr goes as far to say that there is 'no convincing archaeological evidence' for the Khoekhoen having introduced the first livestock to the area (Sadr 2008). Others are also unsure of the association between the Khoekhoen and the first livestock in the area (Deacon 1984; Klein 1986; Kinahan 2001).

It is also suggested that lugged ware pottery that has been associated with a Khoekhoen migration might actually indicate a movement at a much later date – in the second millennium CE (Sadr 1998, 2008). There is also other evidence that supports a later arrival for the Khoekhoen at the Cape. For example, Khoekhoe has a low linguistic diversity despite being spoken in a vast area; the Cape Khoekhoen are reported to have had a low population density in the 17th century; and there were suggestions that not all suitable areas had been occupied by then and that the expansion might not have had time to finish when Europeans arrived (Elphick 1977; Ehret 1982a; Fauvelle-Aymar 2008). It should be noted, however, that it has been argued that it would be a hard process for the hunter-gatherers already present to have taken up food production. This might have been due to cultural factors such as their idea of sharing in which private ownership of food was not accepted, while their social structure would have had to undergo fundamental changes in order to allow for acceptance of herds as status symbols rather than just as a food source (Smith 1990a; Barnard 2008). Such transitions have, however, been made before (Mitchell 2002); for example, it has been suggested that some hunter-gatherers in Namibia adopted pastoralism (Kinahan 1996, 2001).

Whether or not the Khoekhoen introduced the first livestock into the area, there is still much interest as to where these herders originated. As mentioned above, linguistically the Khoe languages, of which Khoekhoe is a member, are included in the 'Khoisan' language group. This

was not always the case, however, as Khoekhoe was originally thought to be a Hamitic language (Schapera 1930). It has since been suggested that proto-Khoe and proto-Khoekhoe developed further south, perhaps in the northern Botswana, Zambia, Zimbabwe area (Bleek 1929; Westphal 1963), and then interacted with Bantu-speaking Iron-age farmers from East Africa (Elphick 1977; Smith 1992). Some argue, however, that there is not obvious evidence for a Khoekhoe-Bantu link (Fauvelle-Aymar 2008), and there have been suggestions of contact with pre-Bantu Eastern Sahelian groups (Ehret 1998) or Cushitic speakers for the source of the Khoekhoen livestock (Blench 2009), whilst some still suggest a slightly more northern origin in East Africa, or at least links with the area (Guldemann 2008; Smith 2008). Evidence that has been proposed for an East African origin includes some of the methods of herding used by the Khoekhoen, which might have been learnt from East African groups. The Namibian Nama and the Cape Khoekhoen use war oxen and vaginal insufflation (a technique used in cows reluctant to let down milk), rare practices which are used in only a very small number of other populations in Africa, most of which are in East Africa (Fauvelle-Aymar 2008). This evidence also supports discounting the Bantu groups as the source for the Khoekhoen's livestock, as Southern-African Bantu speakers do not use these techniques (Fauvelle-Aymar 2008). Heine and Konig (2008), on the other hand, propose that the current distribution of Khoe languages implies that proto-Khoekhoe developed in the west of the Kalahari, between Botswana and Namibia, and not in the Zambezi Basin.

No current theory about the arrival dates and origins of the Khoekhoen is certain. It has been suggested that some of the problem with finding archaeological evidence to support a Khoekhoen migration which tallies with the first sheep is that they left little evidence and were invisible due to frequent movements (Smith 2005; Smith 2008). Or, perhaps they were invisible to some extent because they interacted with the hunter-gatherers already in the region, losing some of their own distinguishing cultural features (Fauvelle-Aymar 2008; Guldemann 2008).

Sadr (2008) suggests that some of the problems with the conflicting theories might be to do with data from the various disciplines (archaeology, history, linguistics) not being comparable. It will be interesting to see whether genetic studies can help address these questions, and help understand how the various populations grouped in the 'Khoisan' category are related, if at all.

Genetic studies on extant populations of Khoe-speaking herders and Ju and Tuu hunter-gatherers have attempted to help understand the history of, and relationships between, the groups. Initial studies used blood groups, protein, red cell enzymes and HLA markers, and in addition to demonstrating the diversity in these populations, they indicated a genetic structure between San groups and Khoe groups (the Nama), with different types found in each group (Jenkins et al. 1970; Jenkins et al. 1971; Steinberg et al. 1975). Studies also showed that despite the differences between the Khoe and San groups, they are more closely related to each other than they are to Bantu-speaking groups (Jenkins 1986). Use of mtDNA restriction fragment length polymorphisms (RFLPs) also demonstrated differences between San and Khoe populations (Soodyall and Jenkins 1992). More recent studies have identified 'Khoisan' associated markers as being L0d and L0k for mtDNA – the oldest branches on the tree (Salas et al. 2002), and A(xA3b2) (M91 for A, and M13 for A3b2), B2b (M112) and E1b1b1g (M293) for the NRY (Underhill et al. 2000; Cruciani et al. 2002; Semino et al. 2002; Henn et al. 2008; Soodyall et al. 2008; Batini et al. 2011). These markers, the deepest known clades for mtDNA and the NRY, are found at the highest frequencies in 'Khoisan' populations of Southern Africa, and are thought to show Southern African click-speaking populations as being the most genetically diverged (Chen et al. 2000; Knight et al. 2003; Tishkoff et al. 2007a; Behar et al. 2008). Analysis of the Khoe-speaking Nama pastoralists has shown that they are indeed more closely related to Ju-speaking foragers than to other African populations (Guldemann and Stoneking 2008). Other evidence, however, shows different Khoe-speaking groups (Damara and Khwe) as being genetically closer to other Africans than to each other (Guldemann and

Stoneking 2008). Additionally, a study has suggested that NRY haplotype sharing between populations from Southern Africa and Eastern Africa supports the idea of an eastern origin of the Khoe-Kwadi pastoralist migration (Henn et al. 2008). Guldemann's (2008) proposal is that the Khoe-Kwadi speaking pastoralists arrived in the region later than the Ju and Tuu groups, and interactions occurring between these three groups as the pastoralists migrated might explain the genetic and linguistic patterns seen today.

Despite this apparent importance of 'Khoisan' populations in the history of the continent, relatively few studies have been carried out on these populations, and those that have use small numbers of samples. There is also often no distinction made between the Khoe herders and the Ju/Tuu hunter-gatherers, and there has been little work carried out on the separation between the Ju and Tuu themselves, and so little is known about the origins of the Southern African forager populations. Additionally, the majority of the SAK groups still present in the region have been affected by migrations within Southern Africa, and it is likely that interactions have taken place between the various groups. It has been pointed out that present-day Khoekhoen populations might not be representative of the original Khoe-speaking migrators due to interactions with the Tuu/Ju populations (Fauvelle-Aymar 2008). Khoe and Tuu/Ju associated markers might also be identifiable in many of the contemporary populations present in Southern Africa that they have interacted with, not just those that still practice a hunter-gatherer/pastoralist lifestyle. Studying the distribution of these markers across different populations in Southern Africa is expected to help identify any admixture that occurred, both between the herders and the hunter-gatherers, and also between both of these groups and the Bantu-speakers who arrived later.

Bantu-speaking agriculturalists

Continental African languages can be split into four major phyla: Niger-Congo (including the Bantu family), Nilo-Saharan, Afroasiatic and 'Khoisan' (as discussed above) (Greenberg 1966). The Bantu languages are a large group of over 400 related languages that are spoken across Africa south of 5°N (Holden 2002); the majority of people living in Sub-Saharan Africa today speak Bantu languages (Nurse and Philippson 2003). The 'Bantu-expansion' is believed to have started around 5,000 years ago, with movement from North-Western Cameroon/Southern Nigeria to throughout Sub-Saharan Africa (Newman 1995). Understanding the current distribution of the Bantu languages, and those who speak them, has engendered a large amount of research by linguists, archaeologists and geneticists.

Archaeological evidence supports the idea of a 5000 – 4000 BP start date for the Bantu dispersal, as it corresponds with the appearance of Neolithic technologies in the Grassfields region of Cameroon (Lavachery 2001; Pakendorf et al. 2011). Sites further south in Cameroon, Gabon, the Congo, Central African Republic and Democratic Republic of Congo have produced finds from slightly younger dates, ranging from the second millennium BC to the end of the last centuries BC, with sites further south and east producing increasingly more recent finds (Pakendorf et al. 2011). Archaeological evidence of ceramics from the Chifumbaze complex (Phillipson 2005) suggests the presence of farming communities in Southern Africa from about the third century AD, although no further south than about 33.5°S (Mitchell 2004).

The spread of Bantu languages was probably not just a single event, but involved both western and eastern expansions (Ehret 1998, 2001), although it is possible that these were not independent of each other (Newman 1995). In fact, there is not agreement about the exact routes the expansion took, and two main possibilities exist (de Filippo et al. 2012): an 'early split' whereby the Eastern branch of Bantu languages is a primary branch of Proto-Bantu,

arriving in East Africa about 4000 ya as the result of a migration from Cameroon, north of the rainforest, towards the Great Lakes (Newman 1995; Holden 2002); or a 'late split' with East Bantu languages only splitting from West Bantu languages about 2000 ya, after the Bantu languages had already spread south of the rainforest region (Ehret 2001; Rexova et al. 2006). Whatever the exact route of dispersal, it is arguably one of the largest and quickest movements of modern humans (Berniell-Lee et al. 2009), involving cultural, linguistic and technological transmission, along with gene transfer (Diamond and Bellwood 2003).

In addition to the widespread distribution of Bantu languages being explained by a migration of iron-using agropastoralist populations (speaking Bantu languages) (Diamond and Bellwood 2003; Phillipson 2005), there has also been the suggestion that language shift and the spread of culture might have occurred without the actual movement of people (Nichols 1997). A similar debate has occurred over the European spread of agriculture (Currat and Excoffier 2005). Ammerman and Cavalli-Sforza (1973) propose a demographically driven 'wave of advance' of agriculturalist genes over short distances into Europe, with a corresponding increase in population density. The general idea is that if farming spread via culture, then there should be a large amount of Palaeolithic (hunter-gatherer) lineages still present. Conversely, if the spread of farming involved the movement of people, there should be more Neolithic lineages in the gene pool. Despite this theory, genetic evidence has not produced any decisive results for Europe (Currat and Excoffier 2005).

This theory might also have a role to play when studying the spread of agriculture in Sub-Saharan Africa. If the 'Bantu expansion' was indeed a movement of people, it might have potentially led to the isolation of, or admixture with, neighbouring hunter gatherer and other pastoralist populations, and genetic evidence of gene flow among any groups that came into contact might be evident. In fact, there is evidence from genetic studies of a Bantu-impact on

populations across Sub-Saharan Africa, although it seems that the extent to which the Bantu populations interacted with the indigenous populations varies across regions. Principal coordinates analysis of African populations suggests that Bantu populations from different areas of Africa cluster together, with evidence of gene flow between 'Khoisan' and Bantu populations (Cavalli-Sforza et al. 1994; Tishkoff et al. 2009). As with the 'Khoisan' lineages mentioned above, mtDNA and NRY studies have also identified Bantu-associated markers and haplogroups. These have been identified as L0a, L1c1b, L1c2, L1c3, L1e, L2a1, L3b, L3d and L3e (Bandelt et al. 1995; Chen et al. 1995; Watson et al. 1997; Alves-Silva et al. 2000; Pereira et al. 2001; Salas et al. 2002; Plaza et al. 2004; Batini et al. 2007; Quintana-Murci et al. 2008; Berniell-Lee et al. 2009; de Filippo et al. 2010) for mtDNA and B2a and E (more specifically, E1b1a7a and E1b1a8) for the NRY (Scozzari et al. 1999; Underhill et al. 2000; Underhill et al. 2001; de Filippo et al. 2011). Although E1b1a7a and E1b1a8 are not uniquely found in Bantu-speaking populations, they are still informative about the history of this group (Pakendorf et al. 2011).

Bantu-speaking populations that have been analysed from Angola (Plaza et al. 2004; Coelho et al. 2009), Zambia (de Filippo et al. 2010), Zimbabwe (Castri et al. 2009) and Mozambique (Pereira et al. 2001; Pereira et al. 2002; Salas et al. 2002) demonstrate that the majority of lineages present in current populations are 'Bantu' associated, and there are very low frequencies (if present at all) of the 'Khoisan' associated markers. This adds support to the idea that the 'Bantu-expansion' was a movement of people, not just of their culture. A recent study looking at the link between language and genetics supports the view that the Bantu expansion was that of both language and people (de Filippo et al. 2012), although, a study of autosomal SNPs has recently reignited the debate by suggesting that language shift might be responsible for the spread of Bantu languages to Mozambique (Sikora et al. 2011).

Analysis of language history in Sub-Saharan Africa also adds support to the 'wave' model: it has been argued that the spread of languages is less likely to happen without the spread of the people, as unlike other cultural traditions, it is hard to acquire a language as an adult (Bellwood 2001). For the Bantu languages to have spread so significantly over Sub-Saharan Africa, and to have replaced the majority of the 'Khoisan' languages previously spoken (Lewis and SIL International 2009), it suggests that there was a spread of Bantu-speakers and not just their language and agriculture. Study of specific words in Bantu languages supports this idea; for example, analysis of the geographical distribution of loan words for cattle, sheep and cereal cultivation (Jobling et al. 2004). Linguistic studies have also led to the suggestion that the Bantu languages in Africa are very closely related, meaning that the depth of their historical divergence is probably limited (Nurse et al. 1985). Trees of Bantu languages are also consistent with archaeological evidence for the spread of agriculture (Holden 2002).

Frontiers

The small number of studies that have been carried out on Southern African Bantu-speaking populations, however, suggest that populations from this region might have considerably higher levels of the 'Khoisan' markers. For example, L0d/L0d has been found at 29% in a South East Bantu population (Schlebusch et al. 2009), and at 25% and 50% in Xhosa and Zulu populations respectively (Soodyall (1993) reported in (Mitchell 2010). These frequencies suggest clear differences between the different regions of Sub-Saharan Africa. In order to understand the reasons for these possible differences in introgression and the effect of the Bantu expansion we must look at the manner of their arrival.

An important process which might help to explain and predict the structure of genetic variation across Sub-Saharan Africa is a frontier model. Frontier dynamics were first proposed in the late 19th and early 20th centuries by researchers studying North America (Alexander

1977), and were defined as 'the temporary boundary of an expanding society at the edge of substantially free lands' (Turner 1962). This concept can be applied to most areas of the world. Alexander (1977) proposes that several stages exist in the development of the 'frontier': both 'moving' and 'static' frontiers. 'Moving frontiers' arise during the initial phase of agricultural arrival, where farmers continue spreading until usable land is exhausted, natural boundaries are reached, or their crops and animals cannot tolerate the conditions any further. There are two phases of a 'moving frontier' (Alexander 1984). First, small groups from the farming communities move into uninhabited, or hunter-gatherer inhabited, regions and use some of its resources. These groups do not cause problems with the local hunter-gatherers, and may even be welcomed by them (Alexander 1984). Secondly, the farmers begin to 'subdue the wilderness', as opposed to only exploiting some aspects of it (Ammerman and Cavalli-Sforza 1973). This results in denser group sizes, and is more likely to disrupt the hunter-gatherer societies. Once the limit of the area usable by the farmers is reached, however, changes occur in their social organisation, often resulting in different relationships with the nearby hunter-gatherers (Bohannan and Plog 1967). A 'static frontier' is created. While a static frontier exists, long term interactions might occur between farmers and hunter-gatherers (Alexander 1984).

With regards to Sub-Saharan Africa, the Bantu-speaking agriculturalists had reached modern day Angola, Zambia, Mozambique and Zimbabwe by the mid second millennium CE. As discussed above, genetics-wise, the current populations in these regions have typically Bantu genetic characteristics with little presence (if any) of lineages from the indigenous Southern-African 'Khoisan' peoples (Pereira et al. 2001; Pereira et al. 2002; Salas et al. 2002; Plaza et al. 2004; Coelho et al. 2009; de Filippo et al. 2010; de Filippo et al. 2011). A frontier existed across this region as part of the spread of the sorghum/millet/bovid/caprid complex with the Bantu-speaking farmers (Alexander 1984). These animals and crops are not indigenous to Southern Africa (Harlan et al. 1976), and a moving frontier must have existed while they were

introduced. This advance was rapid, as not only were there few human or natural barriers in the way, the poor condition of the region's soils meant that sites could not be intensively cultivated, and thus occupied for long, before moving on was necessary (Newman 1995). Archaeological evidence supports the idea of a rapid expansion along the frontier in this region, with large agricultural settlements being built in inland woodlands within centuries of the first villages appearing along the coastlines (Hall 1987). Some of the earliest structures built by farmers in Southern Africa have been located in the village of Mabveni on the Zimbabwe Plateau (Hall 1987). There is also archaeological evidence that domesticated animals were present in some more northern parts of South Africa from the 1st millennium BC (Beaumont and Vogel 1984), and that farming of sorghum and millet was also present in the Transvaal by the 1st millennium AD (Mason 1981).

This expansion of farmers through the majority of Southern Africa might have impacted negatively on the indigenous hunter-gatherers due to over-grazing and a reduction of land available. The speed of the expansion may have also left little time for relationships to develop between the farmers and the hunter-gatherers, and in some cases, such as in Zimbabwe, the hunter-gatherers seem to have disappeared quickly (Walker and Thorp 1997). In the Magaliesberg region of South Africa, for example, farming sites have little evidence of trade with the hunter gatherers (few ostrich egg shell beads or stone tools), and hunter-gatherer shelters similarly have little evidence of farming technologies (Wadley 1996; Mitchell 2002).

By the middle of the 1st millennium AD, however, unsatisfactory growth conditions for sorghum and millet caused the moving frontier to reach its limits, with a static frontier being created (Harlan et al. 1976; Maggs 1980; Mitchell et al. 2008). The farmers were unable to move any further south than the borders of Lesotho and surrounding South Africa and the spread of farmers, and the associated Iron Age technologies in this region were therefore slow

(Hall 1987). Hunter-gatherer groups were still present over large areas of South Africa between 700 and 1500 AD, and in some areas into the 1800s, resulting in a static frontier between the settling farmers and these indigenous peoples (Mitchell 2009).

The static frontier might have resulted in increased interactions between the farmers and the hunter-gatherers, and Alexander (1984) suggests that hybrid societies were formed with characteristics from both ways of life. Hunter-gatherers were not necessarily subservient to the farmers, and close, mutually beneficial relationships might have developed (Mazel 1989). In some cases hunter-gatherers worked for the farmers, and farmers often used hunter-gatherer techniques; for example, stone tools have been found at numerous early farming villages (Hall 1987), as have ostrich egg shell beads (Mitchell 2002). There is also evidence of Bantu groups making use of San rainmakers in order to try and improve growing conditions (Prins 2009), thus providing the opportunity for hunter-gatherer individuals to integrate into farming communities. It has also been noted that some Bantu rituals appear to have a San source (Mitchell 2009), for example some Nguni groups borrowed trance dance from the San (Hammond-Tooke 1998), and also adopted divinatory animals (Hammond-Tooke 1999). Archaeological evidence in South Africa, extending as far as the Lesotho border, demonstrates farmer-associated ceramics and iron remains in hunter-gatherer settlements (Carter and Vogel 1974; Mazel 1986; Mitchell and Charles 2000). This iron might have been intentionally produced for trade with the hunter-gatherers (Mitchell 2002). Rock art created by hunter-gatherers in the Caledon Valley shows cattle, and in other locations sheep and other Bantu-associated symbols appear (Mitchell 2002), demonstrating further the close-proximity of the two groups. Furthermore, the interpretation of rock art in Lesotho, although controversial (Mitchell 2009) has been suggested to indicate that Bushmen in this region interacted with the Bantu-farmers to such an extent that they incorporated some of the Bantu groups religious beliefs and practices into their own (Jolly 2005). Interactions such as these on the ‘static

frontier' might have allowed hunter-gatherer populations to survive for longer, and for increased admixture to occur. This in turn might have resulted in a larger number of hunter-gatherer genetic markers being be present in the current populations of areas such as Lesotho and the southern regions of South Africa than are present in previously studied Sub-Saharan African populations.

Evidence of interaction can also be seen by the level of click-sounds present in the languages of Bantu-speaking populations to the south of the frontier. The languages spoken in Lesotho and South Africa belong to the South-East Bantu language group (Holden 2002), yet unlike languages from the same group in Mozambique, Zambia and Zimbabwe, they also contain a large number of click sounds, typical of 'Khoisan' languages (Herbert 1990; Guldemann and Stoneking 2008). It has been reported that 15% of the Xhosa language is derived from 'Khoisan', with Zulu and Sotho also containing high levels (Lewis and SIL International 2009).

The static frontier lasted for at least 1000 years (Alexander 1984), with a large part of the interior of South Africa not being settled by farmers until at least 900 to 1400 AD. The static frontier was only completely disrupted after about 1600 AD when Sotho/Tswana-speaking groups expanded across the Vaal, although a number of Nguni-speaking groups may have arrived slightly earlier. Later, the introduction of new European plants (such as maize) and animals further disrupted the equilibrium. The advances in yield and durability of growing maize had consequences on the population sizes of the farmers; agriculturalist numbers increased pushing them further into marginal land, gradually impinging on the hunter-gatherers more and more (Mitchell 2002). Today there are very few hunter-gatherer groups remaining in Southern Africa.

Together, the wave model (Ammerman and Cavalli-Sforza 1973) (see above) and the frontier hypothesis, might explain the results in the small number of studies that have already been conducted in Southern Africa (Soodyall 1993 in (Mitchell 2010), (Schlebusch et al. 2009)). It has been suggested that genetic drift is important at the front of range expansions (Edmonds et al. 2004; Klopfstein et al. 2006). Mutations that occur at the edge of the range expansion have been seen to 'surf' on the 'wave of advance', resulting in an increase in frequency and distribution (Klopfstein et al. 2006). This can produce a cline of frequency of a particular allele, increasing along the direction of the migration (Barbujani et al. 1995). It has been proposed that 'surfing' like this occurred during the spread of agriculture across Europe, resulting in some mutations reaching very high frequencies. It does not seem unreasonable to suggest that this process might also have occurred in Sub-Saharan Africa.

The interactions that occurred on the static frontier between farmers and hunter-gatherers in southern South Africa might therefore have resulted in hunter-gatherer markers entering the farming populations. These markers would then have been on the edge of the expansion once the farmers began moving again after the frontier dissolved. As surfing has been shown to be favoured in small, fast-growing populations (Excoffier and Ray 2008), it seems possible here. In fact, simulations carried out for European populations suggest that only a very small hunter-gatherer contribution is needed to create a current gene pool which traces the majority of its genes back to the hunter-gatherers (Currat and Excoffier 2005). It is predicted therefore, that levels of hunter-gatherer markers will be higher in the Lesotho and South African populations than in those from other Southern African populations where there was not enough interaction between the two groups before the range expansion began. Currat and Exoffier's (2005) simulations with European data show that to achieve hunter-gatherer contributions to current populations as low as they are in other Southern African countries such as Zimbabwe and Southern Mozambique (Pereira et al. 2001; Salas et al. 2002; Castri et al. 2009), there was

virtually no genetic contact between the expanding farmers and the hunter-gatherers. This would support the frontier model discussed above. The majority of the work on wave models has so far been on new mutations (Excoffier and Ray 2008; Balaresque et al. 2010), however, this does not mean that mutations already present cannot end up on the expanding wave.

Cultural impact on genetic diversity

In addition to the impact of migrations, there are various other factors that might also affect the distribution of, and resulting genetic diversity in, populations. In particular, cultural practices are thought to have a significant impact.

Gender-biased gene flow

Differential histories of males and females might have an impact on the genetic structure of a population which is detectable today. For example, it has been suggested that farmer-hunter-gatherer interactions were mostly sex-biased (Pakendorf et al. 2011). A well-known example of this can be seen in studies of Central Africa, and the hunter-gatherer populations of Pygmies that live in the equatorial rainforests. Analysis of the Pygmy and Bantu-speaking farmer populations that live in the region has demonstrated that sex-biased gene flow occurs (Cavalli-Sforza 1986; Destro-Bisol et al. 2004a; Destro-Bisol 2005; Wood et al. 2005). Pygmy maternal lineages are present in the farming populations of Central Africa, however, it is much rarer to find farmer maternal lineages in the Pygmy populations, or Pygmy paternal lineages in the farmer populations (Destro-Bisol et al. 2004b). This is thought to be a consequence of cultural taboos which stigmatise Pygmy males as being 'lowest on the social scale' (Cavalli-Sforza 1986), while Pygmy females are accepted into Bantu communities due to their low dowries and believed high fertility (Destro-Bisol 2005). Bantu male lineages might enter the hunter-

gatherer populations as a result of marriage breakdown, with male children returning with the mother to the hunter-gatherer population (Cavalli-Sforza 1986). Certain circumstances might, however, have allowed for male hunter-gatherers to have integrated into the farming societies; for example the use of San rainmakers by Southern African Bantu populations (Prins 2009). It is known from previous studies, that Sotho, Tswana, Xhosa, and Zulu populations from this region do actually display some hunter-gatherer associated NRY haplogroups (Wood et al. 2005), suggesting that some male hunter-gatherer to farmer gene flow must have occurred. The possibility that male and female members of populations might have had differing interactions with other groups makes it extremely important to study both the maternal (mtDNA) and paternal (NRY) genetic markers in order to get a full understanding of a population's history.

Marriage customs and descent rules

The occurrence of residence norms within an ethnic group can also have an impact on the genetic variation seen between males and females (Jobling and Tyler-Smith 2003). Patrilocality (or virilocality) involves a couple living in the group of the male, while matrilocality is where a couple live with the group of the female partner. These practices prevent random mating, and affect the distribution of genetic variation across populations. Patrilocality is usually given as the explanation for high mtDNA and low NRY diversity within groups, and greater genetic differences across populations for the NRY than for mtDNA (Oota et al. 2001). A number of studies have been conducted on worldwide populations in order to try and understand this process, and to see if any differences between male and female patterns are discernible (Seielstad et al. 1998; Oota et al. 2001; Wilder et al. 2004; Kumar et al. 2006; Langergraber et al. 2007). Unfortunately, to date, contradictory results have been observed.

Another custom that might impact the genetic variation seen in populations today is inheritance of clan affiliation between parents and offspring. Populations can be organised into 'descent groups', in which they are divided into smaller groups known as tribes or clans, which may have a common ancestor. In some populations, offspring belong to the clan of both parents (cognatic, 39%) (Heyer et al. 2012). In others, inheritance is unilineal: patrilineal societies pass clan affiliation between a father and his offspring (45%), whilst in matrilineal societies inheritance is from the mother (12%) (Heyer et al. 2012). Patrilineal societies are approximately twice as common as matrilineal ones (Burton et al. 1996). This practice has the potential to have an impact on genetic variation in a population; it would be expected that in a patrilineal society there should be correspondence between NRY diversity and clan-association. A number of previous studies have found differing results for NRY and mtDNA variation in patrilineal groups, with reduced NRY diversity, indicating that the different inheritance patterns might be causing this (Chaix et al. 2007; Segurel et al. 2008; Heyer et al. 2009).

It has been seen that social organisation and cultural practices can influence genetic diversity in human populations, and it is therefore very important to take this into consideration in order to properly understand genetic variation observed in a region. Without taking information such as this into account it is impossible to get a true understanding of the history of a population. For example, individuals grouped together based on their clan affiliation might not display expected patterns of variation: if mtDNA variation is tested in a patrilineal group, results will be impacted by the fact that the maternal line is likely to be comprised of a mixture of clans, whereas the paternal line should all remain the same.

Populations to study

As previously mentioned, genetic studies have been carried out on a number of Sub-Saharan African populations, however, there is a distinct lack of information about southern-most Africa, a region that is archaeologically, geographically, climatically and ethnographically different to the regions already genetically analysed (Nurse et al. 1985; Mitchell 2002). Populations that have been previously studied are also often not well characterised, and in some cases even incorrectly classified (Mitchell 2010), making it difficult to draw confident conclusions from them. For example, samples described as 'South African Kung and Khwe' in several papers (Chen et al. 2000; Salas et al. 2002) are actually Angolan (Mitchell 2010), being part of a group that moved to Schmidtsdrift, Northern Cape Province, South Africa in 1990. Here, populations with well characterised ethnic and geographical information will be studied in two different regions of Southern Africa in order to provide information about the history of an area which is presently understudied.

Lesotho

The Kingdom of Lesotho is a landlocked country, surrounded entirely by the Republic of South Africa. Lesotho is a relatively small country, with an area of 30,355 km^2, and is the only independent state in the world to lie entirely above 1000 metres (BBC 2011a). As a result of this altitude, Lesotho has a relatively cool climate compared to other Southern African countries. Lesotho emerged as a single nation under King Moshoeshoe I in the 1800s (Gill 1993), and today the majority of people in Lesotho belong to the Basotho ethnic group, speaking Southern Sotho, or Sesotho, a South-East Bantu language in the Sotho-Tswana group (Lewis and SIL International 2009). A small number of people also belong to Nguni groups; Ndebele, Phuti, Thembu, Xhosa and Zulu (Lewis and SIL International 2009). Lesotho's

population is approximately 2.2 million (United Nations Department of Economic and Social Affairs 2011).

Modern day Lesotho and the adjacent Maloti-Drakensberg region of South Africa were among the last places in Africa south of the Limpopo to be settled by the Bantu expansion. In Lesotho, this settlement began in the mid-seventeenth century and was only completed at the end of the nineteenth century (Mitchell 1992; Gill 1993). Before the Bantu arrival, Lesotho was inhabited by San populations, known as 'Mountain Bushmen' (How 1962). Today, none of these Mountain Bushmen survive, and the predominant language spoken in the country is Southern Sotho, with no 'Khoisan' languages being spoken (Lewis and SIL International 2009). After the Nguni languages which occur in Southern Africa, the next closest relatives of Sotho/Tswana seem to occur in East Africa (Mitchell 2002).

As discussed above, when the Bantu-speaking farmers reached the Maloti-Drakensberg region and the borders of Lesotho, a static frontier was created acting as a barrier against the farming migrations (Alexander 1984). Despite this late settlement by the Bantu-speakers, a considerable amount of archaeological evidence has been amassed from Lesotho supporting the idea that the Mountain Bushmen communities were not isolated from the nearby farmers, and might even have been interacting long before the Bantu-speakers actually settled in the region (Mitchell et al. 2008). For example, excavations at a site at Likoaeng in highland Lesotho uncovered remains of domesticated cattle dating from the ninth century AD (Mitchell et al. 2008). This is approximately 1000 years before the Bantu agriculturalists actually settled in the region. Historical accounts even report that the first King of Lesotho, Moshoeshoe, had a Mountain Bushman wife (Gill 1993), adding weight to the argument that it was not taboo for a Bushman women to marry into the Bantu tribes. There is evidence that in many cases hunter-gatherers incorporated Bantu-associated technologies into their lifestyles, such as the use of

iron tools (Mitchell 2009). Sotho oral traditions also describe trade with the hunter-gatherers, with exchange of skins, ostrich feathers and ostrich eggshell beads for tobacco and grains (Maggs 1976; Jolly 1996).

By the seventeenth century, however, the equilibrium changed, and unrest caused by competition over grazing lands, which were becoming increasingly in short supply, resulted in large population displacement (Newman 1995). In the Natal region of South Africa, the Zulu expansion caused other Bantu groups to flee; some along the coast towards Thembu and Xhosa groups, and others into the Drakensberg Mountains and thus into modern day Lesotho. By this time as well, the Sotho farming techniques had caused increased stress on the land available for the hunter-gatherers, and their adoption of new crops allowed them to begin the move into the mountains. However, even when the Bantu-speakers eventually made the move into modern day Lesotho, they did not immediately displace the Mountain Bushmen populations. For example, it is known that when missionaries arrived in Lesotho in the 1830s, San languages were still spoken in the Southern Highveld (Gill 1992), and that in some instances, they might even have been dominant over the Sotho settlers for a while (Ellenberger 1912). Oral history research in the northern Drakensberg region has even indicated that San might have been present in the region as recently as the 1930s (G Warrick, personal communication).

So, although there are no longer any people present in Lesotho who self-identify as San, archaeological, linguistic and historical evidence suggest that there were significant interactions between these Mountain Bushmen and the Bantu farmers (Barnard 1992; Mitchell 1992; Gill 1993), and that genetic exchanges might also have occurred. Yet, despite the possibility of San lineages still being detectable in Lesotho populations, no previous genetic work has been carried out on the population genetic history of Lesotho. This work will

therefore study both maternal and paternal markers from populations from Lesotho (Basotho, Ndebele and Thembu groups) and South Africa, and compare them to those previously available (Chen et al. 2000; Pereira et al. 2001; Salas et al. 2002; Knight et al. 2003; Plaza et al. 2004; Tishkoff et al. 2007a; Castri et al. 2009; Coelho et al. 2009; de Filippo et al. 2010; Pereira et al. 2010; Quintana-Murci et al. 2010; de Filippo et al. 2011), in an attempt to help understand the impact of the Bantu expansion on populations in Southern Africa.

Namibia

The Republic of Namibia is located on the southwest coast of Africa, and is bordered by Angola, Botswana, South Africa and Zambia. The country has an area of 824,292 km^2 (BBC 2011b), and with the current population estimated to be 2.3 million (United Nations Department of Economic and Social Affairs 2011), it is one of the most sparsely populated countries in the world (United Nations Department of Economic and Social Affairs 2010). There are over 11 indigenous languages spoken in Namibia, and also several international languages (Government of the Republic of Namibia 2011). These languages not only span three major linguistic groups, but the populations speaking them also display a number of different lifestyles, making Namibia one of the few places where hunter-gatherers, pastoralists and farmers coexist. Despite this, little data is available from Namibian populations in the literature, making comparisons within Namibian ethnic groups, and between populations from across Southern Africa difficult. By studying the people living in Namibia today, we can try to understand the history of Namibia, and also that of the surrounding region.

To date, the archaeological research on Namibia's past is sparse and thus inconclusive. Evidence such as a femur from an archaic hominid dated to the Middle or Late Pleistocene (Grine et al. 1995; Trinkaus et al. 1999) does, however, suggest that Namibia has probably been continuously occupied by hominids for about two million years (Sandelowsky 1983). The

earliest *Homo sapiens* occupants of Namibia were hunter-gatherers, and there is a range of archaeological evidence supporting this, and indicating where they were present. Evidence of modern humans in a Middle Stone Age (MSA) context has been found at Apollo 11 Cave in southwest Namibia dating to between 40,000-20,000 BP, while rock art from this cave has also been dated to between 25,000 and 27,000 years (Wendt 1972, 1976). More recent occupation, from the Later Stone Age, is also found at Apollo 11, along with Pockenbank 1 (also in southwest Namibia), with ostrich eggshells and beads found at both sites, and bone tools at Pockenbank 1 (Wendt 1972, 1976). Human occupation across Namibia has, however, not been evenly distributed. For example, it seems that people were absent from southern Namibia between 5100 and 2300 BP (Mitchell 2002). In other regions, such as northern Namibia and the Desert Biome, settlement seems to have been sporadic, perhaps due to weather conditions affecting resource availability (Parkington and Hall 1987; Mitchell 2002). Sites in the Erongo Mountains, western Namibia, date from about 6000 BP (Wadley 1979), and sites at Falls Rock and Snake Rock Shelters seem to have been intermittently occupied from 6500 BP (Vogel and Marais 1971; Mitchell 2002).

As discussed above, by around 2000 BP, evidence suggests that there were now also people practising pastoralism, or at least owning livestock, in Southern Africa. Namibia, especially the northern part, is often considered important in trying to understand the introduction of livestock to the Southern Africa region (Smith 1992). However, archaeological evidence from Namibia is limited (Albrecht et al. 2001). Evidence so far dates pottery finds to around 1980 +/- 50 BP and sheep bones to 1800 BP at Geduld (south west of Outjo) (Smith et al. 1995). Additionally, pottery at Oruwanje 95/1 rock shelter (north west of country) has been dated to 2000 BP, and livestock might have appeared at the same time (Albrecht et al. 2001). Sheep hairs dating from 1550 +/- 50 BP have also been discovered at Mirabib, western Namibia (Sandelowsky et al. 1979). These dates are consistent with those for the Cape region, South

Africa (see above). As with the rest of the region, however, there is still a debate as to how this livestock arrived in Namibia, and who brought it. It is possible that the pre-existing hunter-gatherers and the new pastoralists coexisted at several locations, and sheep bones at Big Elephant Shelter (a hunter-gatherer site) seem to support this theory (Wadley 1979).

Soon after the arrival of the Khoekhoe-speakers in Namibia, a further wave of migrants arrived in the country; this time Bantu-speaking agriculturalists. Ceramic evidence indicates the presence of Bantu-speakers in northernmost Namibia from about the first millennium AD (Mitchell 2002), while evidence also indicates the presence of an 'Early Iron Age' at Kapako on the Kavango River from about 1000 AD (Richter 2002-2003). Further west and south, however, there seems to be no real indication of Iron Age artefacts, and thus Bantu-speakers, before 1000 AD (Mitchell 2002). It is possible, however, that some of the current Bantu-speaking populations did not arrive in Namibia until about the 16th century AD (Nurse et al. 1985; Malan 1995). The exact dates and original locations of these Bantu groups is not very clear, but today there are numerous Bantu-speaking populations in Namibia, making up a large proportion of the country's population: e.g. Herero, Himba, Owambo.

In addition to these 'Khoisan' and Bantu populations, in relatively recent times there has also been a significant European impact on Namibian history. The first Europeans known to have arrived in Namibia were Portuguese explorers who landed in the 1480s, but significant impact only began in the late 18th century (Government of the Republic of Namibia 2011). The number of Europeans in Namibia, and also the Cape region of South Africa, gradually increased, resulting in displacement of the indigenous populations. For example, in the latter half of the 18th century/start of the 19th century increasing numbers of Oorlam moved into Namibia from the Cape. This, in turn, caused tension with the existing populations, and over the 19th century there was a great deal of fighting within Namibia; between Oorlams, Herero and Nama

('Khoisan'-speaking pastoralists) groups. In 1884 Germany took control of the region, ruling it for 30 years as the colony of German South-West Africa. German settlers were given large amounts of land, and as a result conflicts developed with the indigenous populations. Between 1904 – 1907 the Herero and Nama rebelled against the German occupiers, but despite initial successes, the German forces prevailed and retaliated brutally, killing huge numbers. Reports suggest that by 1907 Herero numbers had been depleted from 80,000 to just 15,300 (Bridgman and Worley 2004; Zimmerer 2008), and that the Nama lost perhaps 40 – 60% of their population (Christie 2005; Cooper 2007; Schaller 2011). During WWI, Britain encouraged South Africa to take Namibia from the Germans. This still left the indigenous Africans marginalised, however, and restricted to 'native areas'; Bushmanland and Hereroland in the Kalahari, Damaraland and Kaokoland bordering on the Namib. It was not until 1990 that Namibia gained independence, after increasing international pressure and violent resistance.

In this work, six Namibian populations are discussed: three Bantu-speaking (Herero, Himba, Owambo) and three 'Khoisan'-speaking (Damara, Ju|'hoansi, Nama).

Aims

There is currently a lack of knowledge about the genetic history of Southern Africa, with few populations analysed to date. Often this small number of populations and samples are accepted as being representative of diversity across the whole of Africa. In many cases maternal and paternal markers are not analysed in the same populations, potentially leading to biased conclusions with regards to the populations' history. This work will analyse previously unstudied populations from Lesotho and Namibia in order to increase the coverage of populations in Southern Africa, and to try and improve the understanding of the history of the region. These new samples are characterised ethnically and geographically, providing important information which is not often available for the other populations in the literature. The new samples will be compared to samples previously available in order to address a number of questions.

The impact of culture on genetic variation

The Lesotho and Namibian samples are among the few datasets that have been collected with good biodemographic information, meaning that they have the potential to provide important information on cultural aspects of population history. This information will be used to detect cultural practices present in these samples, such as patrilocality, and analyse any impact on genetic diversity.

Gene flow between hunter-gatherer and farmer populations in Southern Africa

The hypothesis that in the Lesotho region populations will have increased levels of hunter-gatherer associated haplogroups as the result of a frontier restricting the Bantu-expansion is tested. Sex-biased gene flow is also expected, with higher frequencies of female lineages present.

History of pastoralism in Southern Africa

The origins of pastoralism in Southern Africa will be examined using the Namibian pastoralist population, the Nama. Analyses will be conducted to try and understand any evidence linking the Nama, and thus possibly the origins of pastoralism, to other regions of Sub-Saharan Africa.

Namibian population history

Understanding of the origins of other Namibian populations is sparse. Analysis of the Damara, a population with unknown ancestry, will aim to provide more information with regards to their origins and interactions with other populations. Analysis of samples from Namibian Bantu-speaking populations will be used to address questions relating to migrations in the region which have had a significant impact on the current distribution of genetic diversity seen today.

CHAPTER 2: Materials and Methods

Samples

Lesotho

An initial trip to Lesotho was organised for March/April 2009 with the help of Chiara Batini and Cristian Capelli, and advice from Peter Mitchell, Charles Arthur and Brian Stewart. Organisation involved arranging car hire, accommodation, permits, flights etc. Permission was obtained before travelling from The Oxford Tropical Research Ethics Committee (OxTREC) and The Lesotho Ministry of Tourism, Environment and Culture. Whilst in Lesotho meetings were held with The Lesotho Ministry of Tourism, Environment and Culture and The Lesotho Ministry of Health and Social Welfare in order to gain additional permissions for sampling. Permission was also sought from the Lesotho Ministry of Local Government. Unfortunately, permission was not obtained before the end of the trip and no samples were collected. Permission was finally obtained in August 2009 from the three Ministries. A second sampling trip was then carried out (organised by Cristian Capelli and Chiara Batini, with help from Peter Mitchell), and samples were collected in October/November 2009 by Cristian Capelli, Chiara Batini and George Busby with the help of Charles Arthur. Help (translation and logistics) was also provided by Njapeli Matlanyane and Nthontsi Qokolo.

Saliva samples were finally collected from 377 unrelated males, using Oragene® DNA sample collection kits (DNA Genotek, Inc., Ottowa, Ontario, Canada). Participants were healthy adults and appropriate informed consent was obtained. Samples were collected from six different locations within the country (Figure 2.1), including three locations from the Lesotho Lowlands

(LOW-1, LOW-2, LOW-3), and three from the Highlands (HIGH-1, HIGH-2, HIGH-3), in order to investigate whether these regions have been affected differently by the country's history. In addition, ethnic and linguistic information was collected about the donors, their parents and their grandparents, where known (See Appendix 1). In total samples were collected from individuals from seven ethnolinguistic groups: 283 Basotho, 55 Ndebele, 24 Thembu, 8 Xhosa, 2 Phuti, 2 Swazi and 1 Zulu. Two individuals declined to answer the questions.

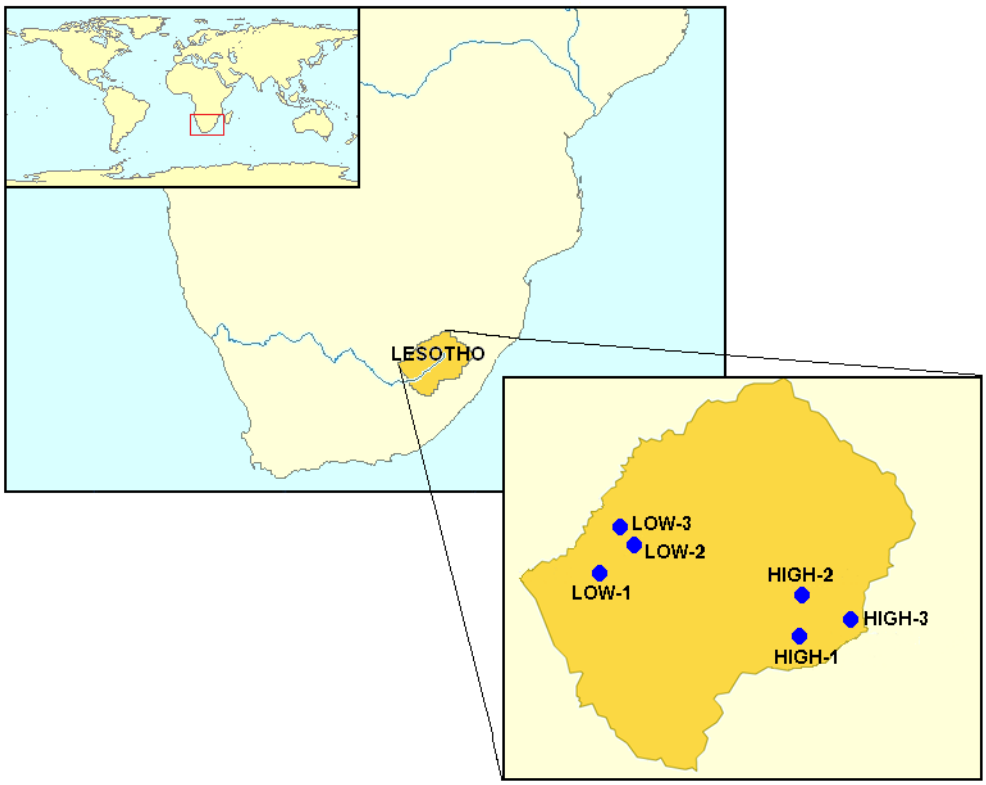


Figure 2.1: Sampling locations in Lesotho.
Sampling was conducted at six regions in Lesotho: three in the Highlands, and three in the Lowlands.

Namibia

During Spring/Summer 2010 organisation was carried out for a sampling trip in Namibia. Dr Ockie Oosthuizen and Erica Oosthuizen helped organise the trip. Permission was obtained from OxTREC and from The Namibian Ministry of Health prior to travelling. Accommodation, flights, car hire and the sampling locations and route were organised with the help of the

Oosthuizens'. Sampling was carried out in September/October 2010 with the help of Cristian Capelli, Ockie and Erica Oosthuizen and Simon Marks. Assistance was provided by numerous translators within Namibia.

Saliva samples and ethno-linguistic information (See Appendix 1) were collected from Damara (72), Herero (45), Himba (50), Ju Hoansi (58), Nama (54) and Owambo (81) individuals within Namibia (Figure 2.2). Participants were unrelated, healthy adult males and appropriate informed consent was obtained. Samples were again collected using Oragene® DNA sample collection kits (DNA Genotek, Inc., Ottowa, Ontario, Canada).

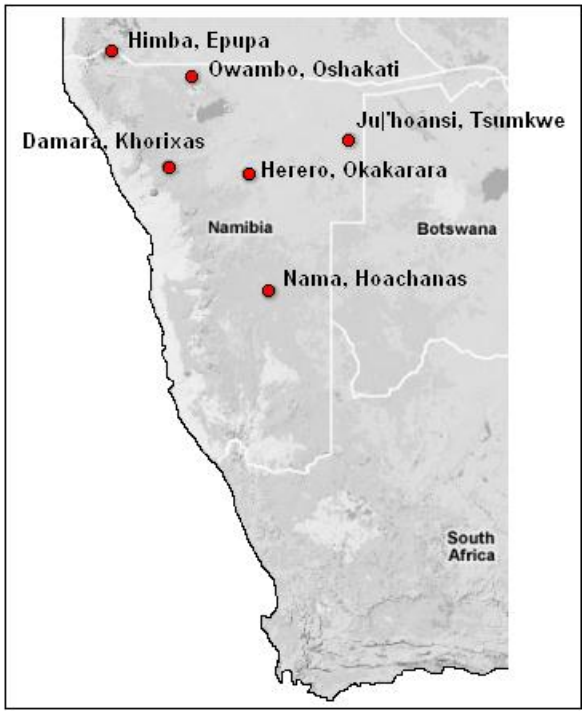


Figure 2.2: Sampling locations in Namibia.
Sampling was conducted at six regions in Namibia: Epupa (Himba population), Hoachanas (Nama population), Khorixas (Damara population), Okakarara (Herero population), Oshakati (Owambo population), Tsumkwe (Ju|'hoansi population).

South Africa

DNA samples from 58 Xhosa and 54 Zulu populations were provided by Maria Eugenia D'Amato and Sean Davison from The Forensic DNA Lab, Department of Biotechnology, University of the Western Cape in Cape Town, South Africa (Leat et al. 2004; Leat et al. 2007). These samples were collected in the Cape Town region, and when included in tables with other Xhosa and Zulu samples from previous works, will be referred to as 'Xhosa Cape Town' and 'Zulu Cape Town' in order to distinguish them from the other populations.

Botswana

177 serum samples from a Bantu population and 183 from a Bushman population were provided by Peter Ebbesen from the Institute of Health Science at the University of Aalborg, Denmark (Ebbesen et al. 1989). Information available for these samples with regards to sampling methods and ethnic classification is unclear. Samples from these populations will be referred to as 'Mixed-Bantu' and 'Mixed-Bushmen' in this work.

Laboratory Methods

See Contributors section at the end of this work for information about others who assisted with laboratory work.

Extractions

Lesotho and Namibia saliva samples

Samples collected using Oragene® DNA sample collection kits (DNA Genotek, Inc., Canada) were extracted using the manufacturer's protocol. Extractions were checked with agarose gel electrophoresis. Initially a small number of extractions failed, but on re-extraction the majority were successful. The majority of samples produced large, bright bands on the gel, however, some did appear to be degraded slightly. Although the sex of these samples was known, once extracted they were tested using the amelogenin test (Salido et al. 1992), as a way of ascertaining what concentration would be sufficient for further analyses (3μl Qiagen reaction mix, 1μl primer mix and 1μl DNA, amplified at: 94°C for 10 minutes; 20 cycles of 94°C for 1 minute, 68°C for 2 minutes reduced by -0.5°C per cycle and 72°C for 2 minutes; followed by 15 cycles of 94°C for 1 minute, 58°C for 2 minutes and 72°C for 2 minutes; and a final extension step of 72°C for 10 minutes. With the primer mix made up of 2.5μl of each of 100μM of the forward and reverse primers and 95μl water). mtDNA and YSNP and YSTR analyses were then carried out on these samples.

Botswana 'Mixed Bantu' and 'Mixed Bushmen' serum samples

174 'Mixed Bushmen' and 175 'Mixed Bantu' serum samples from Botswana (Ebbesen et al. 1989) were extracted using the Qiagen QIAmp DNA Blood Mini Kit. These samples were supposed to be buffy coat only, however, the majority were red in colour (as opposed to the

whitish colour you would expect from buffy coat), and so perhaps also contained some erythrocytes. Extractions were checked with agarose gel electrophoresis, and the majority seemed to have bright bands, suggesting high concentrations. The sex of these samples was not previously known, so they were then also checked using the amelogenin test (see above). 73 of the 'Mixed Bushmen' samples and 52 of the 'Mixed Bantu' samples were found to be male. Analyses were carried out on the 73 + 52 male samples, some of which are available in Batini et al. (2011).

Whole Genome Amplification of Xhosa

65 Xhosa samples collected in Cape Town, South Africa were amplified with the Illustra GenomiPhi V2 DNA Amplification Kit.

mtDNA

For the mtDNA analysis, Hypervariable Segment I (HVS-I) was amplified, with variable positions being determined from positions 16,040 – 16,519 (in analyses with other populations from the literature, positions 16,040 – 16,383 were used). In the initial amplification stage, 1:100 dilutions of the Lesotho, Namibian and Botswana stock extractions were used, with primers L15997 (forward) (5'-CACCATTAGCACCCAAAGCT-3') and H00017 (reverse) (5'-CCCGTGAGTGGTTAATAGGGT-3'). PCR reactions were carried out using 1μl of the diluted DNA, 4μl of Qiagen multiplex PCR master mix (Qiagen), 4μl H_2O, 0.5μl primer L15997 (10μM) and 0.5μl primer H00017 (10μM), and amplified with thermocycling conditions of denaturation at 95°C for 15 minutes, 35 cycles of 94°C for 30 seconds, 58°C for 90 seconds and 72°C for 90 seconds, and a final extension step of 72°C for 10 minutes. The PCR products from this amplification were checked on a 2% agarose gel with 20μl of Sybr® Safe DNA Stain, using 3μl of the PCR product and 3μl loading buffer per well. Products were then purified using polyethylene glycol (PEG) purification and eluted with 30μl H_2O. Sequencing was performed

with the forward primer, using BigDye Terminator Chemistry v1.1 (Applied Biosystems) according to the manufacturer's protocol. A Big Dye master mix of 2.75µl H_2O, 1µl 5 x buffer and 0.25µl Big Dye was added to each well, along with 4µl of 2µM primer and 2µl of the PEG-purified PCR product. Sequencing was carried out using 30 cycles of 96°C for 10 seconds, 50°C for 5 seconds and 60°C for 2 minutes. This was followed by shrimp alkaline phosphate (SAP) purification by adding 1µl SAP to each well and amplifying at 37°C for 1 hour followed by 60°C for 15 minutes. Further purification was carried out using ethanol-sodium acetate. Purified products were sequenced on an ABI 3730x1 Genetic Analyser (Applied Biosystems). Xhosa and Zulu samples were of poorer quality and 1:50 dilutions of the DNA were used initially. A number of samples did not work at this concentration, and dilutions of 1:20 and 1:10 were used instead.

Variable positions in the sequences were determined by manual inspection of electropherograms, and comparison with the revised Cambridge Reference Sequence (rCRS) (Anderson et al. 1981; Andrews et al. 1999), using Geneious Pro 4.8.3 (Biomatters Ltd.) for alignments. For sequences with a poly-C stretch between positions 16184 and 16193, or with other ambiguities, sequencing was repeated with the reverse primer (H00017), or the primer H16401 (5'-TGATTTCACGGAGGATGGTG-3'). Each sequence was assigned to a haplogroup (Hg) using information from previous studies of human mtDNA phylogeny, including the Phylotree website (Behar et al. 2008; van Oven and Kayser 2009).

Haplogroup assignment was verified using the HaploGrep website (Kloss-Brandstatter et al. 2011), which is based on Phylotree (van Oven and Kayser 2009). HaploGrep identifies sequences that have poor quality (less than 90%) matches with previously published haplotypes and thus might have been incorrectly sequenced. For sequences that were ambiguous on HaploGrep electropherograms were re-checked and samples re-sequenced for

HVS-I to ensure that there was not an error with sequencing. If this process resulted in the same haplotype, thus ruling out a sequencing error, Hypervariable Segment II (HVS-II) was additionally amplified in order to try and obtain a more accurate haplogroup assignment. Amplification and sequencing were carried out as for HVS-I, but using forward primer 16555L (5'-CCCACACGTTCCCCTTAAAT-3') and reverse primer 599H (5'-TTGAGGAGGTAAGCTACATA-3') for the initial amplification reaction, and additionally primer 408H (5'-CTGTTAAAAGTGCATACCGCCA-3') for sequencing (Cerezo et al. 2009).

Population		**Samples needing HVS-II analysis**	**Number resolved by HVS-II analysis**	**Number still giving ambiguous results**
Lesotho		92	31	61
Xhosa		9	4	5
Zulu		13	5	8
Namibia	Damara	5	0	5
	Herero	3	1	2
	Himba	5	3	2
	Ju|'Hoansi	11	0	11
	Nama	4	1	3
	Owambo	16	5	11

Table 2.1: Samples requiring HVS-II analysis.

New haplotypes (combining both the HVS-I and HVS-II mutations) were checked again using HaploGrep. A number of samples now presented more convincing results with good quality indicated in HaploGrep, however, there were still some samples that returned poor quality results (see Table 2.1). Although HaploGrep reported problems with these samples, in the majority of cases these samples had the necessary defining mutations for a specific haplogroup and so it was in fact possible to assign them to a haplogroup, but they could not be defined to a higher resolution at the sub-haplogroup level. This is probably because the sequences exhibited additional new mutations that have not been previously reported, meaning that the program could not successfully match them with previous sequences. For example, a number of samples were defined by HaploGrep as L0d1'2, as they had the necessary mutations for L0d

(16129, 16187, 16189, 16223, 16230, 16243, 16311), and also a reversion at position 16278 to place them in the L0d1'2 branch, but not the other required polymorphisms to distinguish between either L0d1 or L0d2. Additionally these samples had a number of extra mutations that the program calls 'local private mutations' which prevented matches with the database. These samples were therefore included as L0d* in analyses.

Y Chromosome

SNPs

NRY analysis was carried out to determine the paternal ancestry of the samples. All samples were genotyped for 11 single nucleotide polymorphisms (SNPs) in one multiplex reaction. The SNPs genotyped were: A-P248, A2-M206, A2-P247, A3b1*-M51, B-M181, B2a-M150, B2b-M112, E-M96, F-R-M89, J-M304, J1-M267 (Figure 2.3; Table 2.2). PCR reactions were conducted using 1μl of 1:50 dilutions of extracted samples, 2.5μl Qiagen multiplex PCR master mix (Qiagen), 1.5μl H_2O and 1μl primer mix (made up of 2μl of the 100μM forward and reverse primer for each SNP, made up to a volume of 100μl with H_2O). Thermal cycling conditions involved a denaturation step of 95°C for 15 minutes, 32 cycles of 94°C for 30 seconds, 58°C for 90 seconds and 72°C for 1 minute, and a final extension step of 72°C for 10 minutes. Amplified products were purified using 1μl ExoSAP-IT® (Affymetrix) and 1.5μl of the PCR product at 37°C for 15 minutes and 80°C for 15 minutes. 2μl of the purified product was mini-sequenced with 2μl of SNaPshot® Mulitplex Kit (Applied Biosystems), 1μl H_2O and 1μl mini-sequencing primer mix (made up of 2μl of the 100μM mini-sequencing primer for each SNP, made up to a volume of 100μl with H_2O). Thermal cycling conditions were 25 cycles of 96°C for 10 seconds, 50°C for 5 seconds and 60°C for 30 seconds. These products were purified using 1μl SAP at 37°C for 1 hour and 75°C for 15 minutes. Products were electrophoresed on an ABI 3130 Genetic Analyser (Applied Biosystems). Results were analysed with GeneMapper software (v. 3.7; Applied Biosystems). Haplogroups were assigned according to Karafet et al. (2008).

A number of the samples identified as E-M96 were further genotyped using an extra multiplex with ten SNPs (E1-P147, E1b-P177, E1b1a-M2, E1b1a7-M191, E1b1a7a-U174, E1b1a8-U175, E1b1a8a-U209, E1b1b1-M35, E1b1b1a-M78, E1b1b1g-M293) (Figure 2.4; Table 2.3), and some then with a third multiplex with three SNPs (E2-M75, E2b-M90, E2b1-M85) (Figure 2.4; Table 2.4). Protocols were as above, apart from different volumes of primers in the primer mixes were used (Tables 2.3 and 2.4).

Y-STRs

A subset of samples were also genotyped for 17 short tandem repeats (Y-STRs). The Y-STRs used (DYS19, DYS385, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYS635, Y GATA H4) were from the AmpFℓSTR Yfiler Amplification Kit (Applied Biosystems). PCR was carried out using 1:50 dilutions of the extracted DNA, with 2.94μl Yfiler PCR reaction mix, 1.6μl Yfiler primer set, 0.256μl AmpliTaq Gold® DNA polymerase (Applied Biosystems) and 2.204μl H_2O, with thermocycling conditions of denaturation at 95°C for 11 minutes, 30 cycles of 94°C for 1 minute, 61°C for 1 minute and 72°C for 1 minute, and a final extension step of 60°C for 1 hour and 20 minutes. Products were electrophoresed on an ABI 3130 Genetic Analyser (Applied Biosystems), and fragment lengths converted to repeat numbers using allelic ladders and analysed using GeneMapper software (v. 3.7; Applied Biosystems). DYS389II was recorded as DYS389II-DYS389I(Redd et al. 2002; Gusmao et al. 2006), and Y GATA H4 as Y GATA H4 + 9(Butler et al. 2008). The DYS385 locus consists of a duplicated tetra-nucleotide STR region and was omitted from analyses. When comparing the samples genotyped in this work with those from published works, the number of Y-STRs used for comparison had to be reduced to 10 (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439) in order to ensure that they had been typed in all of the populations.

SNP		PCR Primers	Minisequencing Primers	Primer designed by
A2-P248	F	TTGTCCCTGTGTTTCCATTTC	$(GACT)_3$GTTCTTCCACACAAAATACTGAA	GF
	R	CAGGCATCAGCTAGATTGTGTT		
A2-M206	F	ACTGTGGTTGGAAAGCACTAA	GAGACTGACTGACTGACTGACTGACTATAATGCTGTAAGCCAAAATTAA	GF
	R	TGCCTGTTTTGTCTGTGACTC		
A2-P247	F	TTGTCCCTGTGTTTCCATTTC	GACTATGAACAAGTACAGGAGAAATGT	GF
	R	CAGGCATCAGCTAGATTGTGTT		
A3b1*-M51	F	GAGCCTCTATCTCCTGAAGC	GACTGACTGACTGACTGACTGACTTATCTCCTGAAGCAGAGTAGACACA	GF
	R	CCTAGTCCCTTTGCTCCATTC		
B-M181	F	GCTAGCAAAGTTGGCTTGGG	GACTGACTGACTGACTGGACAACTTGATCATCTTTTTGA	(Onofri et al. 2006)
	R	GCACACTAGCTATAAGCAAAAGAAAT		
B2a-M150	F	GCAGTGGAGATGAAGTGAGAC	TTTTTTTTGCCCACACACACAGATAGAAGT	CB
	R	CCTACTTTCCCCCTCTTCTG		
B2b-M112	F	AAAAGCAAAAGAGAACTGCCTC	TTTTTTTTTTTTGAGGTGAGATAAAAACAAAGCAGT	CB
	R	TTCAATTCTTGTCTGTTGCAGAA		
E-M96	F	TTCTCCATATCTGTGTAAAGGCAAGT	GACTGACTGACTGACTGACTGACTGTAACTTGGAAAACAGGTCTCTCATAATA	(Onofri et al. 2006)
	R	CCATAGGTTTTTAATATTATACCTGAGTG		
F-R-M89	F	CTGCTCAGCTTCCTGGATTCA	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTCAACTCAGGCAAAGTGAGAGAT	(Onofri et al. 2006)
	R	CACTTTGGGTCCAGGATCACC		
J1-M267	F	CCAAGTCTGGATAGCGGATT	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTCCACACAAAATACTGAACGT	(Onofri et al. 2006)
	R	GCAGGCATCAGCTAGATTGT		
J-M304	F	TTGTAACAAACAGTATGTGGGATTTT	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTACCAAAATATCACCAGTTGT	(Onofri et al. 2006)
	R	AAATAAACAAATAACTTTCAAAACGTC		

Table 2.2: Primers used in initial SNP multiplex.

Primers designed by: GF = Gianmarco Ferri; CB = Chiara Batini. F = forward primer, R = reverse primer.

SNP		PCR Primers	Volume (in µl) of 100µM primer used in the 50µl primer mix	Minisequencing Primers	Volume (in µl) of 100µM primer used in the 50µl minisequencing primer mix
E1-P147	Fw	AAGGACTGGGCAAGTAATGCT	1.25	gactgactgactgactgactgactgactgactgactgactgactCCAAAAGTTCTCTTCCTGG	1.5
	Rev	CCTCCTAATTTCCCAGCAGAC	1.25		
E1b-P177	Fw	TGCCTCAGCCTCCCAAAGTGC	1.5	gactgactgactgactgactgactgactgactgactgactgactGAAAAAATACTCAGATTCC	1.5
	Rev	CCCTTCCACTCAGGGTCTACT	1.5		
E1b1a-M2	Fw	GGCTCCCCTGTTTAAAAATGTA	2.25	gacgactgactgactgactgactCTTTATCCTCCACAGATCTCA	2.25
	Rev	TACAGCTCCCCCTTTATCCTC	2.25		
E1b1a7-M191	Fw	TGGAGTGTTTATCAGAGCTTAATG	2	ACACACCAAAATATCTCATATTTTCAT	2
	Rev	CCCAGACACACCAAAATATCTC	2		
E1b1a7a-U174	Fw	TCCCTGCAGTGAAATAGTTTTG	1	gactgactgactgactgactgactgactgactTCACTGTCTCTGTTCTTCACAG	1
	Rev	TGCTTCACTGTCTCTGTTCTTCA	1		
E1b1a8-U175	Fw	CTGGTCACACTAAGGCACCAT	0.5	gactgactgactgactgactgactgactgactGGCCACAGGTGCTAATGAAACC	0.5
	Rev	TGTGCATCTAATGACCAGGAG	0.5		
E1b1a8a-U209	Fw	CCACAGGAATGCAAAAGATGT	1	gactgactgactgactgactgactgactGAAGACTGCAAGTTAAAATCA	1
	Rev	ACCTGCAGCATTAAATGGAAG	1		
E1b1b1-M35	Fw	AACTGAGAGGGCATGGTCCC	1.5	gactgactgactgactgactCGGAGTCTCTGCCTGTGTC	1.5
	Rev	TGAACAACTAATCCATGCAGACTTTC	1.5		
E1b1b1a-M78	Fw	TCGACATGAACACAAATTGATACACTT	1	tTTTTGAAATATTTGGAAGGGC	1
	Rev	CCTTAAATCTAAATATTGGAAGCTTACCAT	1		
E1b1b1g-M293	Fw	TGAAAGAATGAAAAATCTGGGAAT	1.25	gactgactgactgactgactgactgactgactgactgactAGAGATTGATCGGTGCATA	1.5
	Rev	CTATTTCCCCCAATCCAAAAC	1.25		

Table 2.3: Primers used in E1 SNP multiplex.
Multiplex designed by Gianmarco Ferri.

SNP		PCR Primers	Volume (in µl) of 100µM primer used in the 50µl primer mix	Minisequencing Primers	Volume (in µl) of 100µM primer used in the 50µl minisequencing primer mix
E2-M75	Fw	GACTTGTCAAAAGCCAAAACAG	2	gactgactgactgactCAATTATCAAACCACATCC	2
	Rev	TTGAACAGAGGCATTTGTGAAT	2		
E2b-M90	Fw	TCTCAGGATTTCTGGATGGTG	1	GATGGTGTTCTGTAACT	1
	Rev	AAAGGAACAAAGCCAAAGCTC	1		
E2b1-M85	Fw	TGGTAACTACCAAAAGTTGAATGG	2.25	gacgactTTCTATTAAGTGTAGTTTTGTTAG	2.25
	Rev	TTAATGCTCACGCTTGTGTTC	2.25		

Table 2.4: Primers used in E2 SNP multiplex.
Multiplex designed by Gianmarco Ferri.

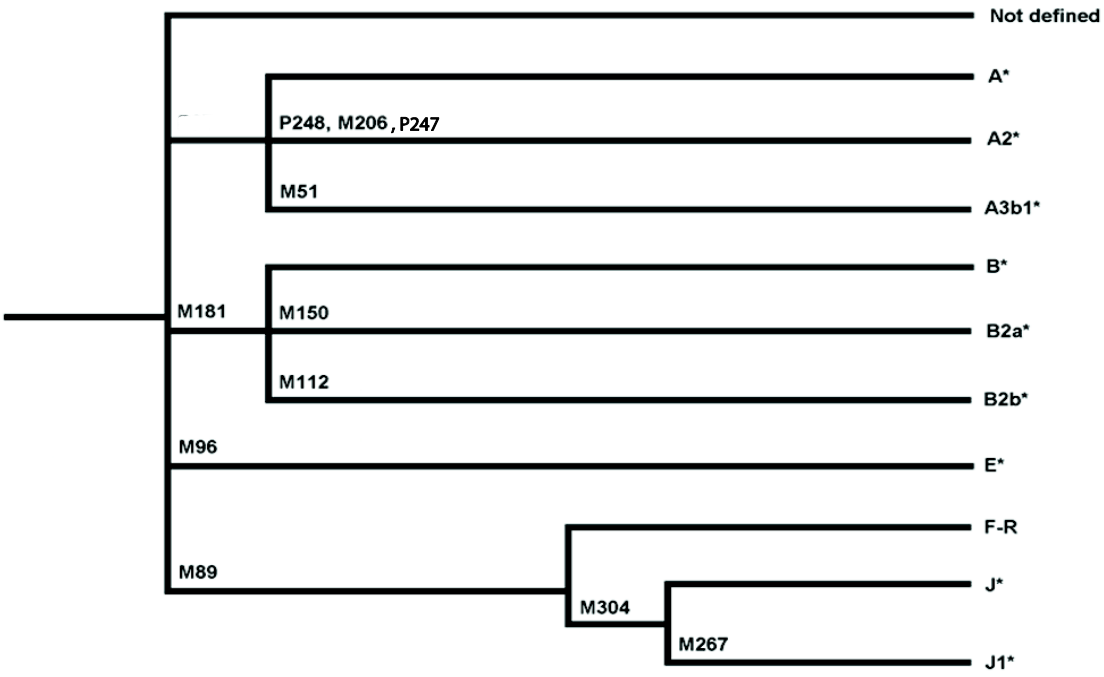


Figure 2.3: SNPs genotyped in the initial multiplex.

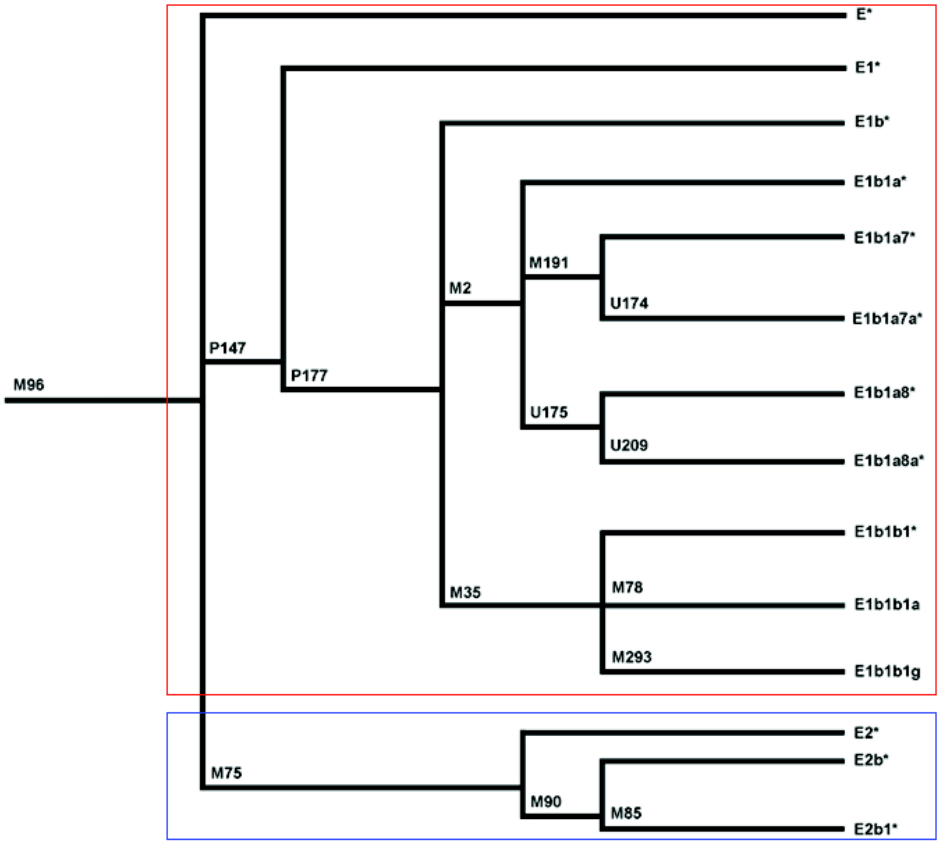


Figure 2.4: SNPs genotyped in the E1 and E2 multiplexes.
The red box indicates multiplex E1, and the blue box multiplex E2.

Chapter-specific analyses

Chapter 3: Cultural impact on genetic diversity

Unless otherwise stated the same methods were used for both the Lesotho and Namibian samples.

Within Lesotho and within Namibia comparisons

In order to detect any significant differences between groups within Lesotho and Namibia, population pairwise F_{ST} values (Reynolds et al. 1983; Slatkin 1995) were calculated using Arlequin 3.5 (Excoffier and Lischer 2010) based on the number of different alleles. For Lesotho, analyses were carried out between: ethnic groups with over 20 individuals (Basotho, Ndebele, Thembu); the six sampling areas; and the three Highland sampling areas vs. the three Lowland sampling areas. For Namibia, analyses were carried out between the six populations. Both mtDNA and Y chromosome data was used, with both haplogroup frequencies and haplotypes taken into account.

Sampling questionnaires

In addition to saliva samples, ethnic and linguistic information was also collected for each donor in both Lesotho and Namibia. See Appendix 1 for a copy of the questionnaire used. Questionnaires were explained to participants using translators where necessary. Information was collected with regards to ethnic group, clan, sub-clan, places of birth and languages spoken for the donors, their parents and their grandparents, where known. Information about current residence and age were also collected for the donors themselves. Data from the

questionnaires was stored in spread sheet format. Lesotho data was inputted by Chiara Batini, George Busby and Hila Levy. Namibian information was inputted by myself.

Analysis of questionnaire information

In order to ensure that comparisons within ethnic groups were fair, for both the Lesotho and Namibian bio-demographic information, only samples with all four grandparents belonging to the same ethnic group as the individual sample were used. This reduced the number of samples available (see Chapter 3).

Female-parent and Male-parent pairs, and trio combinations

Within the ethnic groups information was compared for clan affiliation and place of birth. Offspring – Female Parent – Male Parent (O – F – M) trios were used for comparisons, including parent-grandparent trios, thus increasing the amount of data available. Table 2.5 shows the possible trio combinations available. Female parent-Male parent pairs were used to determine any cross-ethnic group associations that had occurred.

Code	In full	Explanation
O = F = M	Offspring = Female Parent = Male Parent	All same clan/location
O = F ≠ M	Offspring = Female Parent ≠ Male Parent	Offspring and Female Parent match, Male Parent different
O = M ≠ F	Offspring = Male Parent ≠ Female Parent	Offspring and Male Parent match, Female Parent different
O ≠ F ≠ M	Offspring ≠ Female Parent ≠ Male Parent	All different clan/location
O ≠ F = M	Offspring ≠ Female Parent = Male Parent	Female and Male Parents match, but Offspring different

Table 2.5: Trio combinations.
O = Offspring, F = Female parent, M = Male parent.

For both the clan and location analyses, only trios with information for all three individuals were used. Trios were discounted where information was missing or ambiguous. With regards to the location analyses, individuals born outside of the country (Lesotho or Namibia) were discounted as often accurate information was not provided for birthplace. Some

questionnaires only named other countries or regions of countries e.g. Angola, or Free State, Republic of South Africa.

Microsoft Excel was used to identify matches or mismatches between birth place or clan for individuals in the trios, and also for pairs of parents. The frequencies of the different permutations (Table 2.5) were then calculated. The Basotho samples from Lesotho also had information for sub-clans for some of the samples. Again, Microsoft Excel was used to identify matches or mismatches within trios.

Testing results

Comparisons of absolute numbers of migrants, or numbers of different trio combinations for the clan analyses, were conducted using a Pearson chi squared test in the JMP 8.0 statistical package (SAS Institute Inc 2009), with the expected values estimated assuming an equal number of males and females.

Fishers exact tests were also carried out in Arlequin 3.5 (Excoffier and Lischer 2010) to test whether the frequencies obtained for the various trio combinations were different. For example, whether the location trio frequencies were the same when taking into account all trios available, or only those available with coordinate information.

Location coordinates and geographic distances – Lesotho

For Lesotho it was possible to carry out more in-depth analysis on the Basotho samples, taking into account the exact location of the place of birth reported by the participants for themselves, their parents and their grandparents, in order to calculate distances between the individuals within each trio. Geographic coordinates for these locations were obtained from a list tabulated in the “Gazetteer for Basutoland” in 1950 by R. S. Webb (Webb 1950). Out of

nearly 300 different locations reported in the questionnaires, it was possible to obtain unambiguous coordinate information for 137 locations.

The distances in kilometres between the locations were calculated using the spherical law of cosines, which estimates the shortest distance over the earth's surface between two points on the basis of a spherical earth. The distances in kilometres between the locations were calculated using the following equation:

Distance=ACOS(SIN(lat1)*SIN(lat2)+COS(lat1)*COS(lat2)*COS(lon2–lon1))*6371

Where SIN, COS and ACOS are the sine, cosine and arcsine functions respectively, lat and lon are the latitude and longitude of the different geographical coordinates in radians, and 6371 is the mean earth's radius in kilometres. In Lesotho the average distance across all of the six sampled regions was 86.8 km. The matrices of the between populations geographic distances in Lesotho are reported in Table 2.6.

	HIGH-1	**HIGH-2**	**HIGH-3**	**LOW-1**	**LOW-2**	**LOW-3**
HIGH-1	0					
HIGH-2	24.07	0				
HIGH-3	31.95	32.56	0			
LOW-1	111.9	104.2	136.3	0		
LOW-2	124.7	121.5	152.2	26.79	0	
LOW-3	125.4	116.3	148.7	14.47	30.49	0

Table 2.6: Matrix of between population geographic distances (in km) for the Basotho populations.

Averages were calculated for distances between places of birth for all offspring-female parent-male parent trios for which there was location information for all of the individuals (also including trios involving grandparents) (Table 2.5).

Genetic analyses – Lesotho

For the purpose of genetic analysis, the Basotho samples were grouped into the six sampling regions. Sample numbers for the mtDNA were (number of samples per region in brackets): HIGH-1 (23), HIGH-2 (27), HIGH-3 (17), LOW-1 (37), LOW-2 (36) and LOW-3 (48). Seven samples were removed from the NRY genetic analyses due to ambiguous results, resulting in groupings of: HIGH-1 (21), HIGH-2 (27), HIGH-3 (17), LOW-1 (37), LOW-2 (34) and LOW-3 (45).

Genetic differentiation and corresponding 95% confidence intervals were calculated in Analyses of Molecular Variance (AMOVAs) (Excoffier et al. 1992) using Arlequin 3.5 (Excoffier and Lischer 2010). Several different methods were used: Φ_{ST} calculated for number of different alleles for both mtDNA and NRY haplotypes (Excoffier et al. 1992), frequency of different haplotypes for both mtDNA and NRY (Excoffier et al. 1992), and the Tamura and Nei model (gamma = 0.22) (Kumar et al. 1993; Tamura and Nei 1993) for mtDNA; and R_{ST} (Slatkin 1995) for NRY. A standardized Φ_{ST} value (Φ'_{ST}) was also calculated for the number of different alleles (Hedrick 2005; Meirmans 2006; Langergraber et al. 2007), in order to take into account the differing mutation rates of mtDNA and the NRY. Confidence intervals were calculated by bootstrapping genetic differentiation values in a locus-by-locus AMOVA, and those for the standardized values were calculated by dividing the lower value produced for the Φ_{ST} calculations by the upper value for the transformed data, and the upper Φ_{ST} value by the lower transformed value (Langergraber et al. 2007). Nm values (N = effective population size; m = migration rate) to estimate the number of migrants between populations were calculated for Φ_{ST} values obtained from calculations based on the number of different alleles (Wright 1943), modified as described in Meirmans and Hedrick (2011). 95% confidence intervals for Nm values were estimated by combining the values produced from the Φ_{ST} and Φ'_{ST} calculations (Langergraber et al. 2007). The genetic distances and the Nm values were similarly calculated for populations from previous studies in order to make comparisons with the Lesotho results:

six patrilocal (Oota et al. 2001; Tommaseo-Ponzetta et al. 2002; Kayser et al. 2003; Kumar et al. 2006; Marks et al. 2012), and two matrilocal groups (Oota et al. 2001; Kumar et al. 2006).

Analyses were additionally conducted using an Isolation by Distance model (IBD) approach, with the summary statistic considered being the neighbourhood size Nb, the product of the dispersal rate and the population density. In a two-dimensional space Nb is equal to $4\pi D\sigma^2$, with the reciprocal being calculated as the slope of the regression line when $F_{ST}/(1-F_{ST})$ between each pair of populations is plotted against their geographic distance (or its log value) (Rousset 1997). Here both population pairwise F_{ST} values (Reynolds et al. 1983; Slatkin 1995) and standardised pairwise F'_{ST} values were used, and were estimated in Arlequin 3.5 (Excoffier and Lischer 2010) based on the number of different alleles. Geographic distances between populations were estimated as described above: between the centre points of the sampling areas for the Lesotho data. Regression lines were estimated using the JMP 8.0 statistical package (SAS Institute Inc 2009). The differences between the slopes (effectively representing the NRY and mtDNA 1/ Nb estimates) were evaluated using the standard errors of the slopes, as also provided by the JMP software.

Population pairwise F_{ST} values, and genetic differentiation and standardised genetic differentiation values were also calculated in the same way for four Basotho clans.

Number of migrants

The number of migrants was calculated as the number of parents whose place of birth was different from that of the offspring, again considering only trios where all three members provided birthplace information and, in the case of Lesotho, where geographic coordinates for the locations were recovered. The total number of trios used in Namibia was: 74 Herero, 67 Himba, 193 Owambo, 93 Damara, 165 Ju|'hoansi, 95 Nama. The total number of trios used in

Lesotho was 494 and 150 for geographic distance analysis. The number of migrants for male parents and female parents were estimated by considering each gender separately. In Lesotho, the average dispersal distance was calculated as the average of all the parent-offspring distances, including parent-offspring pairs where both individuals were born in the same location. The average migration distance was estimated considering only pairs when the parent and offspring were born in different places. Distances were calculated considering males and females both together and separately.

Chapter 4: Frontiers and Southern African Bantu-speaking populations

The Lesotho, Xhosa and Zulu mtDNA and NRY genotyping results were used for analyses in this chapter. Haplotypes from the Botswana samples were also used in network analyses but not in other analyses due to the uncertainty of the sample provenance. For analysis purposes the ethnic information of the maternal grandmother was used for mtDNA analyses, and the ethnic information of the paternal grandfather for the Y chromosome analyses. Taking into account samples grouped in such a way, the following Lesotho samples were used: mtDNA – 287 Basotho, 33 Ndebele, 24 Thembu; NRY SNPs – 267 Basotho, 52 Ndebele, 21 Thembu, and of these 181 Basotho, 50 Ndebele and 21 Thembu were typed for YSTRs with Yfiler.

Haplogroup and haplotype composition

For both mtDNA and NRY results, haplogroup frequencies were calculated, and pie charts produced to indicate the frequencies of different haplogroups across the different populations. For both the mtDNA and NRY, the number of different haplotypes were calculated using Arlequin 3.5 (Excoffier and Lischer 2010), and haplotype diversity values were also calculated (Nei's diversity) (Nei 1987).

Population pairwise F_{ST} values (Reynolds et al. 1983; Slatkin 1995) based on both haplogroup frequencies and haplotypes were calculated for comparisons between the three main ethnic groups in Lesotho (Basotho, Ndebele and Thembu), the Xhosa and Zulu samples, and also to previously published Bantu-speaking populations from the literature (see Supplementary Table 1 and Supplementary Figure 1, Appendix 2 for populations used and references).

Botswana and Zambia populations from de Filippo et al. (2011) were not included in haplogroup comparisons as this paper does not differentiate between the B haplogroups B2a and B2b, making it harder to compare them with the other data available. These samples were, however, included in haplotype comparisons, and were counted as separate populations if they had over 15 individuals; otherwise they were grouped together and called 'Mixed Bantu West' as in the original paper. A number of the populations used also had less samples available for haplotype analysis than for haplogroup frequencies, possibly as these samples did not work when genotyped for STRs.

Wilcoxon tests

Wilcoxon tests were carried out using the JMP 8.0 statistical package (SAS Institute Inc 2009) to determine whether the mean frequencies of hunter-gatherer related lineages differed between regions. The Lesotho and South African (Xhosa and Zulu) samples were grouped together as South-East Bantu-click speaking populations from Southernmost Africa. This group was compared to populations from the neighbouring region of South East Africa (Mozambique, Zambia and Zimbabwe) (Pereira et al. 2001; Salas et al. 2002; Castri et al. 2009; de Filippo et al. 2010; de Filippo et al. 2011).

Hunter-gatherer related Hgs present in Southernmost Africa before the arrival of the Bantu groups are generally accepted as being represented by L0d/L0k for mtDNA and A/B2b for the Y

chromosome (Chen et al. 2000; Pereira et al. 2001; Pereira et al. 2002; Salas et al. 2002; Behar et al. 2008; Coelho et al. 2009; de Filippo et al. 2010). For the other regions under consideration here, however, less is known about the populations present before the arrival of the Bantu-speakers. In order to avoid possible underestimates related to currently unidentified pre-Bantu types, a conservative approach was followed and all the non-Bantu specific lineages present in the analysed populations were considered as being potentially introgressed local types (see Table 4.1, Chapter 4, for the haplogroups excluded). This most probably represents an overestimate of the frequency of hunter-gatherer types present in the regions, however, this seemed the best way to avoid underestimating the levels of introgression in the South-East region. Non-African haplogroups, identified as belonging to European-associated haplogroups I, R1a and R1b1b2, were also removed. Further overestimation is also likely as a result of using samples from de Filippo et al. (2011): these populations were genotyped for haplogroup B, but are not separated into B2a and B2b. It therefore was not possible to determine if any B2b was present in these populations, so in order to remain conservative all samples listed as haplogroup B from this paper were classed as possible non-Bantu ones.

Wilcoxon tests were also used to detect any differences between mtDNA and NRY hunter-gatherer type frequencies within regions. When carrying out this analysis solely on the Southernmost Africa it was possible to focus on the identified hunter-gatherer types for this region (L0d/L0k and A3b1*/B2b), producing a more accurate result for comparison of the introgression between mtDNA and the NRY in this region.

Summary statistics

Various summary statistics were calculated for mtDNA in Arlequin 3.5 (Excoffier and Lischer 2010) for the Lesotho, Xhosa and Zulu samples, and for the other populations available in South-East Africa. Summary statistics calculated were: number of haplotypes, gene diversity

(Nei 1987), nucleotide diversity (Tajima 1983; Nei 1987), mean number of pairwise differences (Tajima 1983, 1993), Tajima's D (Tajima 1989a, b, 1993), Fu's F_S (Fu 1997) and the Chakraborty test of population amalgamation (Chakraborty 1990). The R_2 statistic (Ramos-Onsins and Rozas 2002) was calculated in DnaSP v 5 (Librado and Rozas 2009). For Tajima's D test, theta values (calculated based on the mean number of pairwise differences and the number of segregating sites) are expected to produce the same values as each other under equilibrium, while differences in the values might indicate population growth (Tajima 1989b, a, 1993). Fu's F_S uses pairwise differences to calculate the probability of observing a certain number of alleles in a population, with departures from the expected perhaps indicating population expansion (Fu 1997). The R_2 statistic is calculated by taking into account the 'number of singleton mutations and the average number of nucleotide differences' (Ramos-Onsins and Rozas 2002). The Chakraborty test detects population amalgamation using observed homozygosity, by calculating the probability of observing a certain number of alleles in a neutral, stationary population (Chakraborty 1990). As a number of the tests (Tajima's D, Fu's F_S, R_2) are thought to be sensitive to changes in population size (Ramos-Onsins and Rozas 2002), comparison of these tests across the regions (Southernmost and South-East) might therefore produce demonstrable differences, indicating distinction in the speed and size of the Bantu-migration in the regions. A previous study suggested that Fu's F_S test is the best test for detecting population growth for large sample sizes, with the R_2 test being suitable for smaller sample sizes (Ramos-Onsins and Rozas 2002).

Summary statistics were also calculated for the separate Lesotho sampling locations for the Basotho samples: HIGH-1, HIGH-2, HIGH-3, LOW-1, LOW-2 and LOW-3. In order to confirm that any results were not influenced by sample size, the summary statistics were additionally calculated for a random sub-sample from these areas (size 20) to make the size of the Basotho samples more comparable with the Mozambique populations available.

Mismatch distributions

Mismatch distribution analysis was also carried out on the mtDNA haplotype data using Arlequin 3.5 (Excoffier and Lischer 2010), and the distributions of the observed number of differences between pairs of haplotypes were plotted. Analyses were carried out on all samples (all haplogroups), and three specific haplogroups separately: L0d (hunter-gatherer) and L3e (Bantu) were analysed as these contained a large number of samples. Bantu-associated haplogroup L2a1 was additionally analysed to provide comparison with L3e, although not all groups could be compared for this due to not having enough haplotypes: Arlequin reported 'the variance of the mismatch distribution is too small', and so parameters could not be estimated. The sum of squared deviation (SSD) and its corresponding p value is reported, along with Harpending's raggedness index (H_r) and its p value for each distribution (Li 1977; Harpending 1994; Rogers 1995; Schneider and Excoffier 1999; Excoffier 2004). The results can be used to understand what demographic history the populations displaying the analysed lineages may have undergone. The SSD demonstrates the goodness of fit to the null model (expansion), so SSD values that are not significant ($p > 0.05$) suggest that the population has undergone a recent, rapid expansion. The shape of the distribution is important, with smooth, bell-shaped histograms (Harpending raggedness index < 0.03) generally suggesting rapid population growth, and ragged distributions suggesting constant population size (Slatkin and Hudson 1991; Rogers and Harpending 1992). Populations used for comparison are recorded in Supplementary Table 1 and Supplementary Figure 1, Appendix 2, excluding the Botswana (Mixed-'Khoisan' and Mixed-Bantu) samples genotyped in this work.

Networks

Phylogenetic networks were constructed for both mtDNA and NRY haplotype data, using Network 4.5.1.6 (Bandelt et al. 1999) (http://www.fluxus-engineering.com/sharenet.htm) in order to compare haplotypes across the various populations. Initially the median joining

algorithm was used (Bandelt et al. 1999), and networks were then post-processed with the Steiner maximum parsimony algorithm (Polzin and Daneshmand 2003). For the NRY data the Reduced Median algorithm was initially used, prior to the median joining one. For the mtDNA, the highly mutable poly-cytosine region in the HVS-I sequences (bp 16182-16193) (Salas et al. 2002) was removed to obtain results unaffected by unstable loci, and a transversion/transition ratio of 3:1 was used (Rosset et al. 2008). Weights for mutable sites were determined according to the frequency of mutations in the population at that site relative to the total number of mutations in the population (Coelho et al. 2009; Batini et al. 2011).

Haplotypes were compared within haplogroups. For the mtDNA Hgs L0d/L0k (hunter-gatherer) and L3e (Bantu) were compared among ten Southernmost African Bantu-speaking and 'Khoisan' populations from this work and from the literature (Chen et al. 2000; Tishkoff et al. 2007a). NRY chromosome haplotypes for the haplogroups A3b1* (hunter-gatherer) and E (Bantu) were compared across nine Southernmost African Bantu-speaking and 'Khoisan' populations. These Bantu-associated Hgs were chosen for comparison due to the large number of samples available within them.

Chapter 5: Hunter-gatherers, pastoralists and farmers: migrations and interactions in Namibian populations

The Namibian mtDNA and NRY genotyping results were used for analyses in this chapter, and were compared to samples in previous studies, and also the Lesotho, Xhosa, Zulu and Botswana samples discussed in Chapter 4. For analysis purposes the ethnic information of the maternal grandmother was used for mtDNA analyses, and the ethnic information of the

paternal grandfather for the NRY analyses. Taking into account samples grouped in such a way, the following samples were available: mtDNA – 66 Damara, 44 Herero, 51 Himba, 58 Ju|'hoansi, 57 Nama, 79 Owambo; NRY SNPs – 54 Damara, 42 Herero, 50 Himba, 58 Ju|'hoansi, 54 Nama, 87 Owambo. 55 Damara, 39 Herero, 21 Himba, 53 Ju|'hoansi, 53 Nama, 50 Owambo were typed for YSTRs with Yfiler.

Comparison populations

Supplementary Table 2, Supplementary Table 3 and Supplementary Figure 1 (Appendix 2) list the populations used for comparison in this chapter, and indicate which of the comparisons they were used in. Populations from Southern Africa only were used to calculate gene diversity averages and for F_{ST} comparisons, whilst populations from across the whole of Sub-Saharan Africa were used in PCA and haplotype sharing analyses.

Haplogroup and haplotype composition, genetic diversity

For both mtDNA and NRY results, haplogroup frequencies were calculated and pie charts produced to indicate the frequencies of different haplogroups across the populations. The number of different haplotypes and Nei's gene (or haplotype) diversity (Nei 1987) were calculated using Arlequin 3.5 (Excoffier and Lischer 2010). Gene diversity values were calculated for both haplogroup frequencies and haplotypes, as not all populations available in the literature had haplotype data available. NRY haplotype values were calculated for 15 STRs, but in order to allow comparison with populations from previous studies, values were also calculated for 10 STRs. Additionally mtDNA diversity values were calculated for a subset of Owambo samples to check that the larger number of individuals available for the Owambo were not skewing the results in any way.

The mean number of pairwise differences (Tajima 1983, 1993) was calculated for mtDNA haplotypes using Arlequin 3.5 (Excoffier and Lischer 2010), and the Mean Number of Mutational Steps for NRY haplotypes, using the R Statistical Package (R Development Core Team 2011). Additionally, the Weighted Intralineage Mean Pairwise difference (WIMP) was calculated (Hurles et al. 2002), measuring the mean within-haplogroup diversity and giving a better reflection of diversity in populations that might have been through a bottleneck.

Average gene diversity, MNPD and WIMP values for Southern African Bantu populations and SAK populations were calculated using populations indicated in Supplementary Table 2 and Supplementary Table 3 (Appendix 2). Values were calculated for each population separately, and then an overall average calculated. Standard deviations given for the means were calculated as the standard deviation of the separate values for all the populations. Some populations were excluded from haplogroup frequency analyses due to not having the same SNPs genotyped.

F_{ST} comparisons – with other Sub-Saharan African populations

Population pairwise F_{ST} values (Reynolds et al. 1983; Slatkin 1995) based on haplotypes were calculated for comparisons between the Namibian populations and previously published populations (Supplementary Table 2 and Supplementary Table 3, Appendix 2). mtDNA haplotype comparisons are reported in Davies (2011) for four of the populations, so only the Owambo and Ju|'hoansi are reported here. NRY haplogroup frequency comparison is used in Davies (2011), however, there is a lot of non-significance with this, probably due to the low resolution of the genotyping in previous studies, so here haplotype comparison has been used.

Principal component analysis

Principal component analysis (PCA) was performed using POPSTR (Harpending 2005) based on the haplogroup frequencies of populations for both mtDNA and NRY in order to demonstrate the relationship of the Namibia populations to other populations. Supplementary Tables 2 and 3 (Appendix 2) indicate the populations used in these comparisons. Initial analyses for both the mtDNA and NRY included all of the populations, and gradually populations were removed from the analyses.

Networks and haplotype sharing

Haplotype sharing between the Namibian populations and Sub-Saharan African populations from the literature (in Supplementary Table 2 and Supplementary Table 3, Appendix 2) was conducted for mtDNA and NRY using Arlequin 3.5 (Excoffier and Lischer 2010). For the NRY, 10 STRs were used for analysis (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439) in order to allow comparison between the maximum number of populations. Charts were made for each Namibian population separately indicating the percentage of individuals and also the percentage of haplotypes that are shared with the other regions of Sub-Saharan Africa. Only haplotypes or individuals that were shared with at least one other region were included in these charts. The number of unique haplotypes per population is listed separately.

Haplotype sharing was analysed in greater detail for NRY haplogroup A, with separate haplotype sharing tables created including additional populations. Additional samples from Batini et al. (2011) were included that were not included in the previous analyses as only data for specific haplogroups was available for these populations. Analyses were conducted for both 15 STRs (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYS635, Y GATA H4) and 10 STRs (as for analysis of all

haplogroups) to allow comparison with different populations. A phylogenetic network was constructed using the 15 STRs in the same way as for the Lesotho data (see above). Nei's gene diversity (Nei 1987) values were also calculated for various groupings of the A3b1 samples, in order to compare diversity across the region. Groupings were: all A3b1 samples available; the Nama separately; all other A3b1 (excluding the Nama); the SAK (excluding the Nama); Southern African Bantu; and the SAC from Quintana-Murci et al. (2010). Analyses were carried out for 15 STRs where possible, and also 10 STRs where data was not available for all 15. Calculations were also carried out removing the multistep locus DYS448 as the large allele differences caused by the multistep nature of this locus might have been obscuring other patterns.

Haplotype sharing was also analysed for NRY haplogroup E-M293, with separate haplotype sharing tables again produced. Here 10 STRs (as above) were used where possible, as these were available for the majority of the populations. A number of samples from Henn et al. (2008) only had 8 STRs available that overlapped with the data genotyped here (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS439), so haplotype sharing was additionally carried out on just 8 STRs. A phylogenetic network was constructed for E-M293 using the 8 STR dataset, and methods as described in Henn et al. (Henn et al. 2008): no weighting was used. Again, Nei's gene diversity (Nei 1987) values were calculated, using both the 10 STR and 8 STR datasets for maximum comparison, and various groupings for the Southern African populations. Calculations were also carried out removing the multistep DYS389I locus.

The mean STR variance was also calculated within the A3b1 and E-M293 haplogroups. The R statistical package (R Development Core Team 2011) was used, calculating the average of the STR repeat size variances over all STR loci using 1000 bootstrap replicates.

CHAPTER 3: Cultural impact on genetic diversity

Introduction

Cultural practices can have an impact on genetic variation across human populations. Some practices might cause selection to occur, favouring certain alleles, whilst others might cause changes in variation as a result of genetic drift. A well-studied example of a cultural practice affecting genetic variation is associated with diet (Laland et al. 2010). Lactase persistence allows adult digestion of lactose in milk and other dairy products, however this is not the ancestral state, with humans typically losing the ability to digest lactose in childhood (Swallow 2003). Lactose-tolerance is, however, present at high frequency in certain parts of the world (Ingram et al. 2009). Several SNPs have been identified as being associated with lactase persistence; different ones in Europe and Africa, suggesting convergent evolution (Tishkoff et al. 2007b). It is believed that the practice of farming, and the increased availability of fresh milk this offered, provided the selection pressure for these SNPs to increase in frequency (Holden and Mace 1997). Other cultural practices, such as residential norms (patrilocality etc.) and marriage customs (polygyny etc.) are known to affect the distribution of genetic variation, but are not subject to selection. Knowledge of these practices and the history of regions should be taken into account when trying to understand patterns of genetic variation.

Marriage residence patterns

Patrilocality, where a couple live in the group of the male, is present in approximately 70% of human cultures (Murdock 1981), and is not unique among hominins (Copeland et al. 2011;

Lalueza-Fox et al. 2011). The opposite system, matrilocality, in which the male joins the group of the female, is much less prevalent, but is found across the world (Jones 2011). More recently neolocal residence has become more common in developed counties, with couples living separately from both individuals' parents (Ember 1967). Matrilocality and patrilocality prevent random mating, and so are expected to affect the distribution of genetic variation across human groups.

It is possible to predict how matrilocality and patrilocality might affect between and within population variation for genetic markers that have gender-biased inheritance: the paternally inherited non-recombining part of the Y chromosome (NRY) and maternally inherited mitochondrial DNA (mtDNA). Patrilocal groups are expected to show higher female than male among group migration as the result of preferential female mobility, which should result in a larger decrease in between group differentiation for mtDNA than NRY, and an increase in within group diversity for mtDNA, but not NRY. Matrilocal groups should demonstrate opposite patterns: higher male among group migration, and lower between group differentiation for the NRY.

Patterns of mtDNA and NRY have previously been analysed in some populations in order to see if there are any noticeable differences between male and female demographic patterns. Contradictory results have, however, been observed (Seielstad et al. 1998; Oota et al. 2001; Wilder et al. 2004; Kumar et al. 2006; Langergraber et al. 2007). In their seminal paper, Seielstad et al (1998) proposed that the observed differences in NRY and mtDNA differentiation among worldwide populations were mostly the result of patrilocality, with a female migration rate calculated from mtDNA being eight times higher than that of males calculated from the NRY data (Seielstad et al. 1998; Stoneking 1998). A later investigation, however, aimed at reducing the impact of different mutation rates between these two genetic

systems failed to identify significant differences in genetic differentiation between worldwide populations (Wilder et al. 2004). In this work, both NRY and mtDNA were genotyped on the same panel of individuals, with results suggesting that patrilocality might have an effect only at a local level (Wilder et al. 2004). Expected patterns were observed for NRY but not mtDNA data in Central Asian patrilocal pastoralist and bilocal farmer populations (Chaix et al. 2007), while comparisons of the distribution of genetic variation in matrilocal and patrilocal groups was as predicted in highland populations of Thailand (Oota et al. 2001) but not in those resident in India (Kumar et al. 2006). Recent investigations of how NRY and mtDNA diversity are shaped among common chimpanzee groups show that none of the human patrilocal groups investigated so far appear to apply this cultural norm as strongly as *P. paniscus and P. troglodytes* (Eriksson et al. 2006; Langergraber et al. 2007).

Two main explanations have been proposed for these conflicting results. Firstly, a failure to take into account the differences in the mutation rate of the loci used to define NRY and mtDNA haplotypes may influence the estimation of the fraction of variation among groups (Wilder et al. 2004; Meirmans and Hedrick 2011). It has been suggested that the use of highly variable markers would increase the amount of within group variation and similarly reduce the maximum fraction of variation that it is possible to detect across groups (Meirmans and Hedrick 2011). Secondly, the analysis of culturally heterogeneous and geographically separated groups (Oota et al. 2001; Tommaseo-Ponzetta et al. 2002; Kayser et al. 2003; Destro-Bisol et al. 2004b; Langergraber et al. 2007) makes it difficult to disentangle the specific contribution of patrilocality on the observed pattern of genetic diversity (Kumar et al. 2006).

Clan affiliation

In addition to these residence norms, there are other cultural traits and customs that might be gender-biased, thus differentially affecting genetic variation of mtDNA and the NRY. Many

populations are comprised of 'descent groups', such as clans. Organisation like this can affect marriage patterns, with exogamous societies marrying outside the social group, and endogamous societies within the group. In exogamous groups, in which cross-clan marriage has occurred, the offspring can either inherit clan affiliation from both parents (cognatic), the father (patrilineal), or the mother (matrilineal). Inheritance traits of populations are important to take into consideration as they might have an effect on the genetic data. For example, when individuals sampled from a patrilineal society identify themselves as a member of a particular clan, this is therefore an indication of their paternal history, and does not represent their maternal history, which may be different. Without checking a sample's history further, it might not be possible to make any inferences about clan and maternal genetic markers. It is therefore important to take into account any cultural practices when defining population sets for genetic analysis.

Previous studies have demonstrated that there are in fact differences in NRY and mtDNA diversity in patrilineal groups, with reduced NRY diversity found in Turkic populations (Chaix et al. 2007). This is as a result of females migrating between descent groups, while the males remain within the same group. Studies of the paternal NRY have also shown that individuals within descent groups share a common ancestor at certain levels of organisation, although not always at all levels; for example in the Turkic populations clans appear to share a common ancestor, but tribes do not (Chaix et al. 2004; Heyer et al. 2012). In fact, studies investigating surnames and genetic diversity are perhaps equivalent to this, as surnames are commonly inherited in a similar fashion to clans in a patrilineal society (Jobling 2001). Results from these also indicate correspondence between surname inheritance and patterns of genetic diversity observed in the NRY of surname groups (McEvoy and Bradley 2006; King and Jobling 2009b, a).

In order to make sure accurate analyses are carried out, it is therefore important to collect and analyse as much information about the samples and populations as possible. Unfortunately, a large number of samples that have been collected in the past do not have sufficient ethnographic data to make sensible comparisons possible. For example, for many populations the exact locations, accurate ethnic information, and clan information are not provided. In the case of some studies, it is even ambiguous as to which population samples belong to, and which country they are from, let alone more precise geographic information being available (Mitchell 2010).

Populations studied

Biodemographic data was analysed from the Lesotho and Namibian datasets with the aim of identifying different cultural practices that should be taken into account when making conclusions with genetic data. The datasets available here provide a linguistically and ethnographically homogeneous set of samples with precise geographic localisation.

Namibia

Six Namibian populations are discussed here: three Bantu-speaking (Herero, Himba, Owambo) and three 'Khoisan'-speaking (Damara, Ju|'hoansi, Nama).

Owambo

The Owambo are a south-west Bantu speaking group that occupy Owamboland in northern Namibia, and also some parts of southern Angola. Within Namibia there are a number of Owambo tribes (Table 3.1).

No.	Name	Frequency
1	Kwanyama	36.56%
2	Ndonga	28.66%
3	Kwambi	11.84%
4	Ngandjera	7.74%
5	Mbalanhu	7.38%
6	Kwaluudhi	5.03%
7 & 8	Eunda and Nkolonkadhi	2.79%

Table 3.1: Owambo tribes.
Adapted from (Malan 1995).

Although often referred to as speaking a general 'OshiWambo' language, the different Owambo tribes actually speak different dialects which should be classified separately (Lewis and SIL International 2009). A number of the questionnaire answers obtained from the Owambo sampled do list 'OshiWambo' as their language, however, it should be possible to deduce the dialect spoken from the tribe listed. The Owambo are a matrilineal group, and residence is generally neolocal, although it may have originally been matrilocal, and today an increasing number are actually practising patrilocality (Malan 1995). Large numbers of men may often be absent due to migratory labour (Malan 1995).

Herero and Himba

The Herero and Himba, originally part of the same population, probably did not enter Namibia until the middle of the 16th century (Malan 1995), occupying the northern Kaokoveld region, having migrated from Angola. The Herero suffered massive population losses during the various conflicts of the 19th and 20th centuries. Both populations practise double descent reckoning, with a mixture of patrilineal and matrilineal ideas. Residence patterns fall under the patrilineal principles, and so the Herero and Himba are patrilocal (Malan 1995).

Nama

The Nama are Khoe-speaking pastoralists that entered Namibia via a northward migration from South Africa. They originated in the Cape region, dividing into two groups: the Great

Nama who moved into Namibia pre-European arrival, and the Little Nama (Oorlam) who moved northward in around the 19th century (Barnard 1992). The Nama are patrilocal, and their tribes patrilineal (Barnard 1992), with 14 recognised today (Table 3.2) (Malan 1995).

No.	Proper name	Alternative name	Afrikaans name
1	Gai-‖khaun	‖Khauben	Rooi nasie
2	ǂ‖Aonin	!Naranin, Hurinin	(Southern) Topnaars
3	!Gomén		(Northern) Topnaars
4	!Gami-ǂnûn		Bondelswarts
5	!Khara-khoin	!Khara-gai-khoin	Fransmanne; Kopers
6	‖Haboben		Velskoendraers
7	‖Õ-gain		Groot Doden
8	‖Khau-|gõan	ǂKhari-‖khaun	Swartboois
9	Kharo-!oan	Tsain	Keetmanshopers
10	!Aman	!Amaîn	Bethaniërs
11	|Hõa-|aran	‖Aixa-|ain; Toroxa-‖ain	Afrikaners
12	Gai-|khauan		Lamberts; Amraals
13	|Hai-|khauan	ǂKhari-|khauan	Bersebaers
14	|Khobesin	!Uri-gamn	Witboois

Table 3.2: Nama tribes.
From (Malan 1995).

Damara

The Damara speak a dialect of Nama (Lewis and SIL International 2009) and were hunter-gatherers until recently (Barnard 1992), but they have since obtained livestock off the Bantu groups that migrated into Namibia (Malan 1995). How the Damara are related to the other populations in Namibia is not well understood. Over the years the Damara have been controlled by both the Nama and Herero, perhaps explaining the shared language with the Nama (Barnard 1992).

Traditionally the Damara have practised patrilocality, however, more recently people have tended to move to areas of employment. The Damara can also be divided into various sub-groups, although these do not have the expected leadership structure to justify being called tribes (Malan 1995) (Table 3.3).

No.	Name	Known as	Comments
1	|Gobanin	Sand dune people	Originally from sand dunes near Rehoboth, Seeis & Hoachanas
2	|Khomá-daman or |Khománin	Khomas Hochland Damara	
3	Hago-daman	Damara of Hakos Mountains	
4	Tsoa-xau-daman	Swakop Damara	Originally from Swakop River area
5	!Kuise-daman	Kuiseb Damara	
6	!Oe-ǂgân	Erongo Damara	From the Erongo Mountains
7	Dãore-daman or Dãorén	Brandenberg Damara	
8	|Gaiö-daman	Parasis Damara	From Parasis Mountains between Outjo and Otjiwarongo
9	ǁHuruben		From the Namib & semi-Namib area between the Unjab and Huab Rivers
10	ǂAo-daman		From Outjo, Kamanjab & Khorixas
11	Ao-guwun		From Sesfontein & !Nani-|ous in the southern Kaokoland

Table 3.3: Damara sub-groups.
Adapted from (Malan 1995).

Ju'|hoansi

The Ju'|hoansi are part of the Ju (or Northern) 'Khoisan' family (Lewis and SIL International 2009), and are also known as !Kung, being classed in Central !Kung (Barnard 1992). They are traditionally hunter-gatherers, and bands generally consist of about 25 individuals, often made up of closely related people (Barnard 1992). The stereotypical image of hunter-gatherer groupings constantly moving around has been questioned, and it is possible that some !Kung individuals might stay at a waterhole throughout the year, living a more stationary existence (Barnard 1992).

Another stereotype, that of !Kung groups only recently being disturbed by other societies, is also likely to be incorrect, with the !Kung and Khoe groups in the Kalahari having had considerable contact in the past (Barnard 1988). The !Kung have also had contact with Bantu-speaking groups for centuries, including the Owambo and Herero, as they traded and shared land with each other (Gordon 1984; Barnard 1988; Barnard 1992). Disturbance has increased

in more recent years, however, and in 1970 a Bushman 'homeland' was established by the South African government. In 1978 South Africa began recruiting Ju'|hoansi individuals as soldiers (Lee and Hurlich 1982), setting up a base at Tsumkwe. In the 1990s, after the end of apartheid and South African control of Namibia, the Ju'|hoansi continued to be marginalised by other Namibian groups who disliked their helping of the South Africans during 1970s and 1980s. The result of the numerous disturbances is that many Ju'|hoansi are no longer hunter-gatherers, as the population in this area can no longer be supported by traditional hunter-gatherer methods (Barnard 1992). More recently, international aid organisations have attempted to help the Ju'|hoansi either resist or adapt to new social and cultural systems introduced by other Namibian ethnic groups, and which are not always compatible with their traditional lifestyles.

Lesotho

Lesotho is comprised of several different ethnic groups, with people belonging to various tribes within these (Table 3.4). The largest group, the Basotho, are not a homogenous group, being an amalgamation of people from a number of different ethnic groups, brought together under Moshoeshoe I during the formation of the Kingdom of Lesotho (Gill 1993). The Basotho are patrilineal and practise patrilocality (Murdock 1981), and according to previous ethnographic surveys, the largest Sotho clan is the Koena (Sheddick 1953). The Nguni are also patrilocal and, although they have a different history, are well integrated with the Sotho, and some might even regard themselves as Basotho (Sheddick 1953).

The samples collected in Lesotho for this work are from seven different ethnolinguistic groups, representing both Sotho and Nguni populations. The Basotho, Ndebele (Nguni) and Thembu (Nguni) were present at large enough frequencies to be used as separate groups in analyses.

Sotho	Koena Group	Koena (various) – crocodile	
		Hlakoana – crocodile	
		Khoakhoa – crocodile	
		Fokeng – antelope	
	Khatla Group	Khatla (various) – baboon	
		Kholokoe – wild cat/baboon	
		Sia – qoabi, another species of wild cat	
		Tlokoa – *qoabi*	
	Taung Group	Taung (various)	
		Hlalele	
		Taung of Moletsane – lion	
		Kubung – hippopotamus	
Nguni	Original settlers	Phetla	Dinake (MaTzetza)
			Polane
		Phuthi	
	Later Nguni immigrants	Natal Nguni	Boshoane's Tebele (Ndebele)
			Hlakoene
			Hlubi
		Cape Nguni	Thembu (various)
			Vova's Vundle

Table 3.4: Clans of Lesotho.
From Sheddick (1953).

Results

Namibia

Ethnographic information was collected along with saliva samples in Namibia in order to try and increase our understanding of the history of the various populations. Full information is not available for all individuals as, unsurprisingly perhaps, not all individuals were aware of details for all of their grandparents.

For F_{ST} tests below, a sample's ethnic group was defined by taking into account the ethnic group of the female parent and maternal grandmother for the mtDNA analyses, and the male parent and paternal grandfather for the NRY analyses. For comparisons of clan affiliation and birthplace location, individuals were used that had all four grandparents recorded as the same ethnic group. Calculations were carried out comparing birthplace and clan affiliation within Offspring-Female Parent-Male Parent trios (O, F, M). Table 3.5 shows a summary of the data available for the various Namibian populations.

	Damara	**Herero**	**Himba**	**Ju|'hoansi**	**Nama**	**Owambo**
Total number of samples	69	45	50	58	54	80
Total number with 4 GP same*	40	29	47	57	36	77
Total number of trios**	120	87	141	171	108	231
Number of trios with locations	93	74	67	165	95	193
Number of trios with clans	92	X	X	X	98	213

Table 3.5: Summary of ethnographic information available for Namibia.
*Sampled individuals with all four grandparents from the same ethnic group.
**Number of trios available from the samples with all four grandparents from the same ethnic group.
Clan information was not available for the Herero and Himba samples, and for the Ju|'hoansi samples only a small number of individuals reported clan information meaning there was not enough to make any comparisons.

For a number of the populations there is a noticeable reduction in the number of samples available if using only individuals that have all four grandparents with the same ethnic group

(Table 3.5). Some of this is as a result of cross-group marriages (see below), however, some of the reduction is due to missing information. The Himba and Ju|'hoansi have only a small reduction in sample size (three Himba and one Ju|'hoansi), all of which is a consequence of missing information. For the Damara, Nama and Herero, however, a large number of the samples had to be discounted due to a lack of information (14, 15 and 6 respectively).

For the groups where there is a large reduction in the sample size if using only individuals with four grandparents from that ethnic group (Damara, Herero and Nama), results were also recorded for trios that did not just include those from 'four grandparent' individuals (i.e. all members of the trio were always the same ethnic group, but a relative might not have been). Fishers exact tests were carried out to test whether the frequencies obtained for each category using the two classifications were different. All tests came back as not significant for location trios (Damara: p = 0.88 +/- 0.0041, n = 204; Herero: p = 0.99 +/- 0.0008, n = 162; Nama: p = 0.97 +/- 0.0025, n = 209) and for clan trios (Damara: p = 0.35 +/- 0.0048, n = 208; Nama: p = 1.00 +/- 0, n = 216). All further comparisons were carried out with the 'four grandparent' trios for continuity across all the ethnic groups.

Comparisons between ethnic groups

F_{ST} comparing ethnic groups

Population pairwise F_{ST} tests were conducted between the Namibian ethnic groups in order to determine whether there are any differences between the six populations. Values are reported in Davies (Davies 2011) for mtDNA and NRY haplogroup composition for four of the six groups. Values based on mtDNA haplotypes show all the Namibian populations to be significantly different from each other (Table 3.6a). When comparisons are carried out using mtDNA haplogroup frequencies (Table 3.6b), the Ju|'hoansi and Nama are not significantly different, possibly because both populations have high frequencies of the 'Khoisan'-associated

haplogroup L0d (see Chapter 4), however, all other comparisons are significantly different. The NRY comparisons for the Ju|'hoansi and the Nama also indicate that they are significantly different from the other Namibian populations; for both haplogroup frequencies (Table 3.6c) and haplotypes (Table 3.6d). The other four populations are, however, not significantly different for some of the comparisons (Table 3.6c and d). For example, the Himba and Herero, are not significantly different from one another (for both NRY haplogroup frequencies and haplotypes); perhaps not surprising considering they are believed to have originated from the same population (Malan 1995).

a. mtDNA haplotypes

	Damara	Herero	Himba	Ju|'hoansi	Nama	Owambo
Damara	x					
Herero	**0.0374*** **(0.0090+-0.0091)**	x				
Himba	**0.12*** **(0.00+-0.00)**	**0.053*** **(0.018+-0.012)**	x			
Ju|'hoansi	**0.35*** **(0.00+-0.00)**	**0.40*** **(0.00+-0.00)**	**0.34*** **(0.00+-0.00)**	x		
Nama	**0.25*** **(0.00+-0.00)**	**0.31*** **(0.00+-0.00)**	**0.27*** **(0.00+-0.00)**	**0.045*** **(0.00+-0.00)**	x	
Owambo	**0.14*** **(0.00+-0.00)**	**0.16*** **(0.00+-0.00)**	**0.11*** **(0.00+-0.00)**	**0.19*** **(0.00+-0.00)**	**0.14*** **(0.00+-0.00)**	x

b. mtDNA haplogroups

	Damara	Herero	Himba	Ju|'hoansi	Nama	Owambo
Damara	x					
Herero	**0.036*** **(0.025+-0.0056)**	x				
Himba	**0.14** **(0.00+-0.00)**	**0.061*** **(0.0088+-0.0029)**	x			
Ju|'hoansi	**0.50*** **(0.00+-0.00)**	**0.54*** **(0.00+-0.00)**	**0.46*** **(0.00+-0.00)**	x		
Nama	**0.39*** **(0.00+-0.00)**	**0.42*** **(0.00+-0.00)**	**0.36*** **(0.00+-0.00)**	0.036 (0.055+-0.0071)	x	
Owambo	**0.18*** **(0.00+-0.00)**	**0.17*** **(0.00+-0.00)**	**0.082*** **(0.00+-0.00)**	**0.35*** **(0.00+-0.00)**	**0.27*** **(0.00+-0.00)**	x

c. NRY haplogroups

	Damara	Herero	Himba	Ju|'hoansi	Nama	Owambo
Damara	x					
Herero	0 (0.46+-0.015)	x				
Himba	-0.0027 (0.0059+-0.0022)	0.014 (0.13+-0.0089)	x			
Ju|'hoansi	**0.25*** **(0.00+-0.00)**	**0.21*** **(0.00+-0.00)**	**0.26*** **(0.00+-0.00)**	x		
Nama	**0.16*** **(0.00+-0.00)**	**0.17*** **(0.00+-0.00)**	**0.22*** **(0.00+-0.00)**	**0.25*** **(0.00+-0.00)**	x	
Owambo	**0.048*** **(0.0029+-0.0016)**	0.21 (0.066+-0.0084)	-0.0098 (0.82+-0.012)	**0.26*** **(0.00+-0.00)**	**0.21*** **(0.00+-0.00)**	x

d. NRY haplotypes

	Damara	Herero	Himba	Ju|'hoansi	Nama	Owambo
Damara	x					
Herero	0.0087 (0.13+-0.0034)	x				
Himba	**0.058*** **(0.0011+-0.0003)**	0.0072 (0.25+-0.0039)	x			
Ju|'hoansi	**0.21*** **(0.00+-0.00)**	**0.14*** **(0.00+-0.00)**	**0.13*** **(0.00+-0.00)**	x		
Nama	**0.14*** **(0.00+-0.00)**	**0.10*** **(0.00010+-0.0001)**	**0.12*** **(0.00+-0.00)**	**0.17*** **(0.00+-0.00)**	x	
Owambo	**0.017*** **(0.026+-0.0015)**	**0.015*** **(0.046+-0.0019)**	0.017 (0.12+-0.0032)	**0.16*** **(0.00+-0.00)**	**0.14*** **(0.00+-0.00)**	x

Table 3.6: Population pairwise F_{ST} values and (p values) among Namibian ethnic groups. For: a. mtDNA haplotypes; b. mtDNA haplogroup frequencies; c. NRY haplogroup frequencies; d. NRY haplotypes.
* = significant at the 0.05 significance level.

Between group associations

In order to assess interactions between the different Namibian populations, the ethnic affiliations of female parent-male parent pairs were analysed: this provided information on the frequency of cross-group marriages (Table 3.7). Female parent-male parent pairs were used where ethnic information was available for both individuals. The Himba and Ju|'hoansi have no female parent-male parent pairs belonging to different ethnic groups. These populations have been isolated until relatively recently (Malan 1995), which might explain reduced

interactions with other groups, and account for all marriages taking place within the same population. The majority of the Nama and Owambo individuals have parents of the same ethnicity (Nama 97.8%; Owambo 98.8%). Two of the Owambo pairs with a mismatch have one parent with European ancestry. There are also a further seven individuals with European ancestry across the other populations, reflecting the European aspect of Namibian history. The Damara and Herero, on the other hand, have a larger number of pairs with mismatches (Damara 11.7%; Herero 11.6%), although the number with parents of the same ethnic group is still statistically significant (Damara: 151 vs. 20, $\chi^2 = 100.36$, $p < 0.0001$; Herero: 107 vs. 14, $\chi^2 = 71.48$, $p < 0.0001$). Arguably, these two groups have had the most inter-group contact as a result of conflicts in the past two centuries, and the reported subservience of the Damara to the Nama and Herero (Malan 1995). This might therefore explain some of the mixed ancestry. In both cases, the majority of the mismatches show the male parent as having a different ethnic group to that which the individual associates with (Damara: 9.4% male parent different vs. 2.3% female parent different; Herero: 9.1% male parent different vs. 2.5% female parent different). These differences are statistically significant for the actual values (Damara: 16 vs. 4, $\chi^2 = 7.2$, $p = 0.0073$; Herero: 11 vs. 3, $\chi^2 = 4.57$, $p = 0.033$).

Ethnic group pairs	**Damara**	**Herero**	**Himba**	**Ju|'hoansi**	**Nama**	**Owambo**
Number of female parent-male parent pairs	171	121	147	171	134	240
F = M (%)	88.3	88.4	100	100	97.8	98.8
Female parent different (%)	2.3	2.5	0	0	0.7	0.8
Male parent different (%)	9.4	9.1	0	0	1.5	0.4

Table 3.7: Frequency of female parent-male parent pairs with matching or different ethnic groups, for the Namibian populations.
Pairs included where ethnic group information was available for both parents. F = M: Female parent and Male parent both from same ethnic group.

Since Namibian Independence, an increasing number of individuals across all ethnic groups have migrated to rapidly expanding cities and towns looking for jobs (Tvedten 2004). Sampling

in areas such as Windhoek, rather than in the historical homelands for each ethnic group, as was carried out here, might therefore detect increased inter-group marriages.

Clans

Clan affiliation was compared across trios within ethnic groups in order to see if any interesting patterns emerged with regards to inter-clan associations (Table 3.8). In any instances where cross-clan marriages do take place, we might expect matrilineal populations to show a larger number of matches between the individual and the maternal parent, and patrilineal populations to show a larger number of matches between the individual and the paternal parent.

Clan affiliation trios	**Herero**	**Himba**	**Owambo**	**Damara**	**Ju|'hoansi**	**Nama**
Total trios available*	X	X	213	92	X	98
All the same clan (%)	X	X	89.67	85.87	X	88.78
Not all the same clan (%)	X	X	10.33	14.13	X	11.22
All same vs. not all same**	X	X	**134.09 (<0.0001)**	**47.35 (<0.0001)**	X	**58.94 (<0.0001**)
Of the trios that are not all the same clan:						
O = F ≠ M (%)	X	X	5.16	4.35	X	1.02
O = M ≠ F (%)	X	X	2.35	3.26	X	6.12
O = F ≠ M vs. O = M ≠ F***	X	X	2.25 (0.13)	0.14 (0.71)	X	3.57 (0.059)

Table 3.8: Frequency of trio combinations for clan affiliation in Namibian ethnic groups. *Trios with four grandparents from the same ethnic group, and with clan information for all individuals in the trio. **Chi squared test on actual values for 'All same clan' vs. 'Not all same clan' trios (p value in brackets). ***Chi squared test on actual values for 'male parent different' vs. 'female parent different' trios (p value in brackets).

Results indicate that for all three of the populations where sufficient clan information was available (Damara, Nama and Owambo), the majority of trios consisted of individuals from the same clan. O = F = M trios were significantly more frequent than trios where the clans were not all the same (Damara: 79 vs. 13, $\chi^2 = 47.35$, $p < 0.0001$, $n = 92$; Nama: 87 vs. 11, $\chi^2 = 58.94$, $p < 0.0001$, $n = 98$; Owambo: 191 vs. 22, $\chi^2 = 134.09$, $p < 0.0001$, $n = 213$). This suggests that the majority of individuals in these groups marry within the same clan. As each group was

sampled in a relatively small region, however, this might have affected the information collected. Different clans traditionally occupy different regions (Malan 1995), so sampling in just one region is likely to sample mostly from one clan. In fact, for the Owambo, the most frequent clan (the Kwanyama) are present at over 60% across all generations in the questionnaires. This is the clan that Malan (1995) indicates as the most frequent, but at a lower frequency of 36.56% across the whole country. The next highest clans, the Kwambi and Ndonga are present at 15% or less. For the Nama, the Gai||Khaun clan is present at 92.59% in the individuals sampled, dropping to 59.26% for the paternal grandfather, but only because the number of unknowns increases. The other Nama clans reported in the questionnaires all remain below 5%. For both the Owambo and the Nama, the clans with the highest frequency (Kwanyama and Gai||Khaun, respectively) are significantly more frequent than all the other clans added together, across all generations covered by the questionnaires (Owambo sampled individuals: $\chi^2 = 15.51$, $p < 0.0001$, $n = 80$; Nama sampled individuals: $\chi^2 = 41.68$, $p < 0.0001$, $n = 53$). It is perhaps not surprising, therefore, that the O = F = M trios are the most frequent, seeing as the sampled areas seem to have high frequencies of a particular clan. The Damara results are less clear cut, with the most frequent clan, the Ao Dama, not being significantly higher than the sum of all the other clans across all the generations. The Ao Dama are, however, significantly more frequent than the next most frequent clan (the Oe-gan) across all the generations (sampled individuals: $\chi^2 = 15.36$, $p < 0.0001$, $n = 57$).

Locations

Places of birth were also compared across the Namibian trios in order to identify any evidence of a particular residential practice. Again, it might be expected that matrilocal populations will have a larger number of matches between the individual's place of birth and the female parent's place of birth, while patrilocal populations should show a larger number of matches between the individual and the male parent. Table 3.9 displays the results for each group.

The majority of trios for the Damara, Himba, Ju|'Hoansi, Nama and Owambo have all three individuals with the same place of birth (O = F = M). For the Ju|'hoansi 95.67% of trios are O = F = M, a frequency which is significantly higher than all other trios put together (χ^2 = 138.19, p < 0.0001, n = 165). In fact, the majority of Ju|'hoansi trios (154 out of 165) are all from the same location (Tsumkwe), indicating that the Ju|'hoansi appear to stay in the same location. For the Damara, Nama and Owambo, the number of O = F = M trios is significantly higher than the number of trios in the next most frequent category (Damara: χ^2 = 5.79, p = 0.016, n = 56; Nama: χ^2 = 11.66, p = 0.0006, n = 58; Owambo: χ^2 = 10.29, p = 0.0013, n = 119). The O = F = M Himba trios are not significantly more frequent than any other category, however (χ^2 = 2.38, p = 0.12, n = 42). The Herero are the only group that do not have O = F = M as the highest category. Their highest category is O ≠ F ≠ M (50%), with all three individuals in the trio being born at a different location. These trios are significantly more frequent than the next most frequent category (χ^2 = 7.41, p = 0.0065, n = 54).

Location trios	**Herero**	**Himba**	**Owambo**	**Damara**	**Ju|'hoansi**	**Nama**
Total trios available*	74	67	193	93	165	95
O = F = M (%)	8.11	38.81	39.90	39.78	95.76	44.21
O = F ≠ M (%)	22.97	11.94	13.47	16.13	1.21	14.74
O = M ≠ F (%)	9.46	5.97	5.18	5.38	0.61	7.37
O ≠ M ≠ F (%)	50.00	23.88	21.76	18.28	0.00	16.84
O ≠ M = F (%)	9.46	19.40	19.69	20.43	2.42	16.84

Table 3.9: Frequency of trio combinations for places of birth in Namibian ethnic groups.
*Trios with four grandparents from the same ethnic group, and with location information for all individuals in the trio. **Chi squared test on actual values for 'male parent different' vs. 'female parent different' trios (p value in brackets).

Using the birthplace of the offspring as a proxy for the family place of residency, it was possible to calculate the number of parents that had migrated. Results indicate that all populations have a higher percentage of male migrants than female migrants, although none of these values are significant (Table 3.10). These results are unexpected as the Damara and Herero are traditionally patrilocal populations, and so it would be expected that more females

would migrate than males. Although thought to be generally neolocal, the Owambo are a matrilineal society, so increased male migration here might be more easily explained. In fact, the second most frequent trio category for the Owambo is O ≠ F ≠ M, at 21.76%, indicating that a large number do indeed report a different birthplace location for all individuals.

	Herero	Himba	Owambo	Damara	Ju|'hoansi	Nama
Trios*	74	67	193	93	165	95
Parents**	148	134	386	186	330	186
Total number of migrants~	112	70	196	92	11	85
Total number of female migrants	51 (45.5%)	33 (47.1%)	90 (45.9%)	41 (44.6%)	5 (45.5%)	39 (45.9%)
Total number of male migrants	61 (54.5%)	37 (52.9%)	106 (54.1%)	51 (55.4%)	6 (54.5%)	46 (54.1%)
Male vs. female migrants#	0.89 (0.34)	0.23 (0.63)	1.31 (0.25)	1.09 (0.30)	0.091 (0.76)	0.58 (0.45)

Table 3.10: Number of migrants in the Namibian ethnic groups.
Percentages for the male and female migrants are percentage of total migrant.
*Trios with four grandparents from the same ethnic group, and with location information for all individuals in the trio. **Total number of parents. ~Total number of those parents that migrated, i.e. have a different location listed to the offspring. #Chi squared test on actual values for male migrants vs. female migrants (p value in brackets).

Lesotho

The total Lesotho sample included individuals from seven ethnolinguistic groupings: 283 Basotho, 55 Ndebele, 24 Thembu, 8 Xhosa, 2 Phuti, 2 Swazi and 1 Zulu. Two individuals declined to answer the questions. 188 Basotho individuals reported all four grandparents belonging to the Basotho population: 121 from the Lowlands (37 LOW-1, 36 LOW-2, 48 LOW-3) and 67 from the Highlands (23 HIGH-1, 27 HIGH-2, 17 HIGH-3), with the average across the six regions being 31, the average in the Lowlands 40, and the average in the Highlands 22. F_{ST} analyses were carried out on ethnic groups with > 20 individuals, classifying individuals using the ethnic group of the female parent and maternal grandmother for the mtDNA, and the male parent and paternal grandfather for the NRY. Trio analyses of clan affiliation and birth place were conducted on various subsets of the 188 Basotho samples with all four grandparents

Basotho, as this was the only Lesotho ethnic group with sufficient data. Due to the size of the Basotho dataset, and the fact that the samples were collected in a variety of locations, it was also possible to carry out genetic analyses to support the ethnographic data.

Comparisons between ethnic groups

F_{ST} comparing ethnic groups

Population pairwise F_{ST} values between ethnic groups with over 20 individuals (Basotho, Ndebele, Thembu) produced no significant results for mtDNA when using both haplogroup frequency and haplotypes analysis (Table 3.11a and Table 3.11b). NRY haplogroup frequencies also produced no significant results (Table 3.11c), however, all three ethnic groups were significantly different when comparing NRY haplotypes (Table 3.11d).

a.	Basotho	Ndebele	Thembu
Basotho	x		
Ndebele	-0.0064 (0.65+-0.0048)	x	
Thembu	0.013 (0.17+-0.0043)	-0.0030 (0.45+-0.0051)	x

b.	Basotho	Ndebele	Thembu
Basotho	x		
Ndebele	-0.002 (0.50+-0.005)	x	
Thembu	0.0060 (0.23+-0.0042)	-0.0058 (0.57+-0.0056)	x

c.	Basotho	Ndebele	Thembu
Basotho	x		
Ndebele	0.0064 (0.19+-0.0036)	x	
Thembu	0.021 (0.15+-0.0031)	-0.016 (0.76+-0.0043)	x

d.	Basotho	Ndebele	Thembu
Basotho	x		
Ndebele	**0.026 (0.00098+-0.001)**	x	
Thembu	**0.041 (0.0068+-0.0023)**	**0.060 (0.00098+-0.001)**	x

Table 3.11: Population pairwise genetic distance (F_{ST}) & (p value) by ethnic group, Lesotho.
a - mtDNA haplogroups; b - mtDNA haplotypes; c - NRY haplogroups; d - NRY haplotypes.

Between group associations

Although the majority of those sampled in Lesotho are Basotho, a number do belong to different ethnic groups. Comparisons were carried out using female parent-male parent pairs to determine the frequency of associations between individuals of different ethnic groups.

Across the whole dataset, 1084 pairs had ethnic group information for both members (Table 3.12). Of these the majority, 748 (69%), were unions between two Sotho individuals. There were 198 (18.2%) unions between one Sotho individual and one non-Sotho individual, of which 86 were a non-Sotho female and 112 (10.3%) a non-Sotho male. Unions between two non-Sotho individuals accounted for 12.73% (138) of the pairings.

Pair type	**Lesotho**
Number of female parent-male parent pairs	1084
Sotho-Sotho (%)	748 (69%)
Male Sotho-Female non-Sotho (%)	86 (7.93%)
Female Sotho-Male non-Sotho (%)	112 (10.3%)
Both parents non-Sotho (%)	138 (12.73%)

Table 3.12: Frequency of female parent-male parent pairs in Lesotho.

The occurrence of cross-group marriages in the Sotho is significantly less frequent than Sotho-Sotho associations (198 cross-group vs. 748 Sotho-Sotho; $\chi^2 = 319.77$, $p < 0.0001$). There is not a significant difference between the number of non-Sotho female parents and the number of non-Sotho male parents (86 female vs. 112 male; $\chi^2 = 3.41$, $p = 0.065$).

Clans

According to the questionnaires, all 188 Basotho samples have a clan listed. There are 13 different Basotho clans recorded, along with a small number that could not be verified as Basotho. Three clans appear more frequently than the others, accounting for 75% of the sampled individuals between them: BaKoena (29.8%), BaFokeng (26.1%) and BaHlakoana (19.2%) (Table 3.13). These clans have been reported as being the most frequent in Lesotho (Sheddick 1953), and so this is not surprising. The other clans are only reported by a small number of individuals. This pattern is repeated across all generations in the questionnaires (both parents, and all four grandparents), with the only difference being that the Mosia clan has a slightly increased frequency for the female parent and maternal grandfather. This

supports reports that the Basotho are patrilineal (Sheddick 1953) as Mosia clan affiliation would have been passed down from the maternal grandfather to his daughter (the sample's mother), but this affiliation would not have been passed on by the female parent to her children. Some individuals did not know information concerning some of their relatives. Interestingly, it is the maternal grandmother and paternal grandmother with by far the largest amount of missing information (37.2% and 26.6% respectively, compared to less than 5% for the other relatives) (Table 3.13). Sampled individuals knew considerably more about clans of male relatives than female relatives (123 missing information for female relatives vs. 8 for male relatives, $\chi^2 = 121.37$, $p < 0.0001$, $n = 131$).

Clan	Offspring	Female parent	Male parent	G-mother (Female parent)	G-father (Female parent)	G-mother (Male parent)	G-father (Male parent)
Lekholokoe	2.1	2.1	2.1	0.5	2.1	2.7	2.1
BaFokeng	26.1	20.7	26.1	17	20.7	19.1	26.6
BaHlakoana	19.2	14.9	19.2	9.6	14.4	15.4	18.6
BaKhatla	2.1	3.2	2.1	0.5	3.2	1.6	2.1
BaKobung	1.6	1.1	1.6	2.7	1.1	0.5	1.6
BaKoena	29.8	31.4	29.8	16.5	31.9	20.2	29.9
BaNareng	0	1.1	0	0.5	1.1	0	0
BaPhering	0.5	0	0.5	0	0	0	0.5
Mosia	2.7	8	2.7	3.2	7.4	2.1	2.7
BaTaung	7.4	6.4	7.4	5.3	6.9	6.4	7.4
BaTlokoa	2.1	1.6	2.1	2.7	1.6	1.1	2.1
BaTloung	3.2	3.7	3.2	0.5	3.2	3.2	3.2
BaTsoeneng	1.6	2.1	1.6	1.1	1.6	1.1	1.6
Other	1.6	2.1	1.6	2.7	0.5	0	1.6
Missing info	0	1.6	0	37.2	4.3	26.6	0

Table 3.13: Frequency (%) of clans present in the Basotho samples.
G-mother = Grandmother; G-father = Grandfather

Analysis of inheritance of clan affiliation was carried out using Offspring-Female Parent-Male Parent trios where clan information was available for all three individuals (423 trios) (Table 3.14). Of these, 42.8% (181/423) were O = F = M, with all individuals belonging to the same clan. The largest category was O = M ≠ F trios, with 54.8% (232/423) showing a match between the offspring and male parent, but not the female parent. Contrastingly, only 1.2% (5/423)

showed a match between the offspring and the female parent, but not the male parent (O = F ≠ M). This difference is significant (χ^2 = 217.42, p < 0.0001, n = 237), indicating patrilinearity is being practiced. In fact, the O = M ≠ F trios were significantly more frequent than the O = F = M trios, (χ^2 = 6.30, p = 0.012, n = 413). A small number of offspring matched neither parent (two trios, 0.47%, were all different, and three, 0.71%, had parents matching but the offspring different). It seems unlikely that the offspring would not have inherited their clan from either parent, so these results might be explained by errors in completion of the questionnaires.

Clan affiliation trios	**Basotho**
Total trios available*	423
All the same clan	181 (42.8%)
Not all the same clan (%)	242 (57.2%)
All same vs. not all same**	**8.8 (0.003)**
Of the trios that are not all the same clan:	
O = F ≠ M (%)	232 (54.8%)
O = M ≠ F (%)	5 (1.2%)
O = F ≠ M vs. O = M ≠ F***	**217.42 (< 0.0001)**

Table 3.14: Frequency of different trio combinations for clan affiliation in the Basotho.
*Trios with four grandparents from the Basotho ethnic group, and with clan information for all individuals in the trio. **Chi squared test on actual values for 'All same clan' vs. 'Not all same clan' trios (p value in brackets). ***Chi squared test on actual values for 'male parent different' vs. 'female parent different' trios (p value in brackets).

Genetic analyses were carried out within and between clans using gene diversity and pairwise F_{ST} tests for the mtDNA and NRY. There were insufficient individuals with all relatives from the same clan, so analyses were instead carried out taking into account only the clan that the individual identified themselves as. Seeing as this does not take into account that both parents might not be from the same clan, it will not necessarily identify differences between the clans, but it might indicate if there is evidence of patrilinearity. In a patrilineal group, it is expected that mtDNA gene diversity is higher than that for the NRY, while the NRY F_{ST} results would be more different across the clans than the mtDNA. Using this method, there were four clans with

sufficient individuals for analysis: BaKoena (56), BaFokeng (49), BaHlakoana (36) and BaTaung (14). Gene diversity results (Table 3.15) are not as expected, with NRY gene diversity being higher or overlapping with mtDNA gene diversity. Results from the mtDNA indicate that none of the clans were significantly different from one another (Table 3.16). For the NRY, however, the BaFokeng were significantly different to the BaHlakoana and the BaKoena (Table 3.17). Using an AMOVA analysis to compare genetic differentiation for all four clans, the NRY F_{ST} value is significantly higher than the mtDNA value (Table 3.18). When corrected values are used (F'_{ST}) (Langergraber et al. 2007), the NRY value remains significantly higher than the mtDNA value. Analyses were also conducted removing the BaFokeng (as they are significantly different in NRY analyses: Table 3.17). Here the NRY and mtDNA F_{ST} values are no longer significantly different (Table 3.18), with the NRY slightly lower. When corrected values are used (F'_{ST}), the NRY value is higher than the mtDNA one, although not significantly so.

Clan	mtDNA		15 YSTRs	
	n	Gene Diversity	n	Gene Diversity
BaFokeng	49	0.9847 (0.9823;0.9870)	48	0.992 (0.9899;0.9940)
BaHlakoana	36	0.9698 (0.9647;0.9749)	35	0.9916 (0.9884;0.9948)
BaKoena	56	0.9597 (0.9566;0.9628)	53	0.9884 (0.9865;0.9903)
BaTaung	14	0.978 (0.9599;0.9961)	14	0.989 (0.9725;1.0054)

Table 3.15: Sample size and gene diversity (with 95% confidence intervals) for mtDNA and Y STRs for four Basotho clans.

	BaFokeng	BaHlakoana	BaKoena	BaTaung
BaFokeng	0 0			
BaHlakoana	-0.01537 (0.98+-0.0034)	0 0		
BaKoena	-0.00248 (0.51+-0.015)	-0.00642 (0.65+-0.016)	0 0	
BaTaung	0.00438 (0.33+-0.013)	-0.0004 (0.39+-0.014)	-0.00792 (0.56+-0.013)	0 0

Table 3.16: Population pairwise F_{ST} value and (p value) for mtDNA for four Basotho clans.

	BaFokeng	BaHlakoana	BaKoena	BaTaung
BaFokeng	0 0			
BaHlakoana	**0.06644** **(0.00+/-0.00)**	0 0		
BaKoena	**0.06654** **(0.00+/-0.00)**	-0.00473 (0.63+/-0.018)	0 0	
BaTaung	0.02148 (0.13+/-0.012)	-0.01222 (0.67+/-0.016)	-0.01305 (0.73+/-0.013)	0 0

Table 3.17: Population pairwise F_{ST} value and (p value) for NRY for four Basotho clans.

	NRY F_{ST}	**mtDNA F_{ST}**	**NRY F'_{ST}**	**mtDNA F'_{ST}**
Basotho 4 clans	0.037* (0.021; 0.053)	-0.0059 (-0.012; 0.0012)	0.071* (0.034; 0.12)	-0.0067* (-0.013; 0.0014)
Basotho 3 clans	-0.0079 (-0.017; 0.001)	-0.0059 (-0.014; 0.0038)	-0.014* (-0.026; 0.0022)	-0.0076* (-0.016; 0.0057)

Table 3.18: NRY and mtDNA genetic differentiation for the Basotho clans.
The BaFokeng were removed for the 3 clan analysis. F_{ST} and Standardised F_{ST} (F'_{ST}) values, (all with 95% confidence intervals). * = F_{ST} values significant at the 0.05 significance level.

Sub-Clan affiliation trios	**Basotho**
Total trios available*	56
All the same sub-clan	28 (50%)
Not all the same sub-clan (%)	28 (50%)
All same vs. not all same**	0 (1.00)
Of the trios that are not all the same sub-clan:	
O = F ≠ M (%)	2 (4%)
O = M ≠ F (%)	23 (41%)
O = F ≠ M vs. O = M ≠ F***	**17.64** **(< 0.0001)**

Table 3.19: Frequency of different trio combinations for sub-clan affiliation in the Basotho.
*Trios with four grandparents from the Basotho ethnic group, and with clan information for all individuals in the trio. **Chi squared test on actual values for 'All same clan' vs. 'Not all same sub-clan' trios (p value in brackets). ***Chi squared test on actual values for 'male parent different' vs. 'female parent different' trios (p value in brackets).

A number of the Basotho samples also reported information on sub-clans. Of the trios with clan information for all members, sub-clan information was additionally available for 425 people. The clans with the highest frequency also have the highest number of different sub-clans listed. Generally, most sub-clans seem to only have a small number of individuals listed (one or two), but there are some that are more common than the others: the BaKoena sub-

clans BaKoteli (72) and Nkopane (42) are the most frequent. Only 56 trios were available where sub-clan information was listed for all members (Table 3.19). Of these, 50% (28/56) are O = F = M trios with all individuals belonging to the same sub-clan. The majority of the remaining trios, 41% (23/56), are O = M ≠ F where the offspring matches the male parent but not the female parent, again confirming the Basotho as a patrilineal people. There were insufficient samples with sub clan information to carry out genetic analyses.

Locations

Questionnaires were also analysed in order to understand patrilocality within the Basotho.

Geographic distances

The average distance across all of the six sampled regions in Lesotho was 86.8 km, with averages of 29 km and 24 km within the Highlands and Lowlands respectively, and an average of 27 km for within regions, taking into account both the Highlands and Lowlands. The matrices of the between populations geographic distances in Lesotho are reported in Table 2.6 in Chapter 2.

Male and female migration

A total of 494 trios were available with place of birth information for all members. The frequencies of the different types of trios are recorded in Table 3.20. Over a third (35.2%) of the offspring-parent trios showed a match between the male parent and the offspring's place of birth but not to the female parent's birthplace (O = M ≠ F), while significantly less (9.72%, $\chi^2 = 15.36$, $p < 10^{-4}$, n = 222) showed the reverse pattern (O = F ≠ M). The total number of migrants was also calculated, taking the birthplace of the offspring as a proxy for the family location of residency (Table 3.21). In total 988 parents were available for analysis from the 494 trios. The overall number of migrants (female and male combined) was 450. Of these 288 were

female and 162 male, a significant difference (χ^2 = 35.28, p < 0.0001, n = 450). This suggests a substantial difference in female vs. male migration, in agreement with ethnographic evidence that reports a patrilocal residential norm for these populations (Murdock 1981).

Comparisons were also made taking into account distances (in km) that individuals migrated by using exact geographic locations of places of birth. Geographic information was not available for all locations, so the number of trios available was reduced to 150. The frequency of the various categories for trios with known geographic (coordinate) locations is reported in Table 3.20. The total number of migrants was calculated (Table 3.21), and results were again significant when only trios with all three members with available geographic localisation were considered (150 trios; 55 female and 30 male migrants; χ^2 = 7.35, p = 0.0067, n = 85).

Trios	Frequency (all trios)	Frequency (trios with geographic info)
Total number of trios	494	150
O = F = M	32%	53.3%
O ≠ F ≠ M	17%	6.7%
O = F ≠ M	9.7%	10%
O = M ≠ F	35.2%	26.7%
O ≠ F = M	6.1%	3.3%

Table 3.20: Frequency of different location trio combinations in the Basotho.

	Frequency (all)	Frequency (with geographic info)
Trios*	494	150
Parents**	988	300
Total number of migrants~	450	85
Total number of female migrants	288 (58.3%)	55 (36.7%)
Total number of male migrants	162 (32.8%)	30 (20%)
Male vs. female migrants#	35.28 (< 0.0001)	7.35 (0.0067)

Table 3.21: Number of migrants in the Basotho.
Percentages for the male and female migrants are percentage of total migrants. *Trios with four grandparents from the same ethnic group, and with location information for all individuals in the trio. **Total number of parents. ~Total number of those parents that migrated, i.e. have a different location listed to the offspring. #Chi squared test on actual values for male migrants vs. female migrants (p value in brackets).

Genetic differentiation

This observed higher female than male migration rate might be expected to cause a differential distribution of genetic diversity. In order to test this, between population variation was investigated using the NRY and mtDNA. Contrary to expectations for patrilocal groups, no significant differences in between population differentiation were observed when comparing the two genetic systems (Table 3.22). These results from the Basotho were compared to worldwide populations characterised by specific residential norms (Table 3.22). Assuming similar male and female effective population sizes, N_m and N_f, differences in the observed ratios should be related to different gender-related migration rates, m_m and m_f. Five populations showed no differences in mtDNA and NRY differentiation among groups: the Lesotho and Spanish sets, two Indian patrilocal groups known to practise strict caste endogamy (Kumar et al. 2006) and one matrilocal group from Thailand (Oota et al. 2001) (Table 3.22). Notably, the other four groups comprised of populations geographically separated and/or speaking different languages, making it difficult to properly evaluate the specific impact of the residential practice on the observed pattern of variation (Oota et al. 2001; Tommaseo-Ponzetta et al. 2002; Kayser et al. 2003; Besaggio et al. 2007; Langergraber et al. 2007).

The data was also tested within an Isolation by Distance (IBD) framework using the approach described by Rousset (1997). This provides a more realistic model assuming a correlation between the geographic proximity of populations and their migration rate. However, the geographic localisation of samples has to be known. Unfortunately, detailed geographic data for populations where the impact of residential pattern has been explored is mostly missing (Oota et al. 2001; Tommaseo-Ponzetta et al. 2002; Kayser et al. 2003; Kumar et al. 2006), so only the Lesotho data could be analysed here, and the Spanish data is provided in Marks et al. (2012). For both the NRY and mtDNA data the slope of the regression line was estimated when

considering the genetic differentiation between pairs of populations (expressed as $F_{ST}/(1-F_{ST})$) and their geographic distances (or the logarithm of it). Rousset (1997) showed that the slope of this curve provides a good approximation for the reciprocal of the Nb parameter, a measure of dispersal and population density. The use of the absolute distance or the logarithm of this value is related to which model (uni or bi-directional) is to be applied to the habitat within which samples are collected (Rousset 1997). For completeness both correlations were carried out, even if the Lesotho data might better comply with a bi-dimensional model. The 1/Nb estimates for mtDNA and NRY were not different in the Lesotho populations when the standard errors were taken into consideration (Table 3.23).

Migration distances

The Lesotho data was further explored in relation to migration distances (Figure 3.1). The average dispersal range was 13 km, with the majority of parent-offspring pairs being in the 0 km distance class (72%). The number of migrants below the average dispersal range was significantly higher than those moving beyond this distance (13 km, 233 vs. 67, $\chi^2 = 91.85$, $p < 1 \times 10^{-3}$, n = 300). Similar results were observed when males and females were considered independently (Figure 3.1a). The average dispersal distance (including non-migrants) of the male parents (10.25 +/- 2.1 km) is significantly less than that of the female parents (15.53 +/- 2.4 km). If non-migrants are excluded, the average migration distance becomes 50 km in Lesotho, with male parents migrating further than female parents (male: 51.24 +/- 6.1 km vs. female: 42.36 +/- 6 km). The absolute number of males and females migrating beyond the average migration distances is not significant, while the number of females migrating at distances shorter than this was significantly larger (Figure 3.1b). Female migrants more commonly moved at distances below 50 km in Lesotho while males did not show a significant difference (Figure 3.1b).

Finally, to explore the impact of differential migration at very small distances, genetic data was compared from single Lesotho villages within sampled areas. In accordance with demographic data showing significant differences in the number of male and female migrants at short distances, the fraction of NRY F'_{ST} values found to be significantly larger than corresponding mtDNA F'_{ST} estimates was greater in comparisons between Lesotho villages separated by less than 24 km than in comparisons between groups at more than 24 km (33% vs. 22%) (Table 3.24).

Population	Residential norm	Locus	Φ_{ST}	(CI)	Φ'_{ST}	(CI)	Nm ($1-\Phi'_{ST}/\Phi_{ST}$)	(CI)	Reference
Basotho, Lesotho	Patrilocal	NRY	0.0039	(-0.0045;0.014)	0.0077	(-0.0074;0.033)	253.14	(71.96;∞)	This work
		mtDNA	0.019	(0.0069;0.031)	0.022	(0.0076;0.037)	50.62	(32.01;139.57)	
Spanish, Spain	Patrilocal	NRY	0.012	(0.0066;0.019)	0.021	(0.0097;0.039)	78.73	(51.37;146.066)	See (Marks et al. 2012) for list of populations and references
		mtDNA	0.0097	(0.0055;0.015)	0.0098	(0.0056;0.015)	101.95	(68.25;178.48)	
Dravidian tribes, India	Patrilocal	NRY	0.01	(0.00;0.06)	0.04	(0.00;0.22)	96	(16.67;∞)	(Kumar et al. 2006)
		mtDNA	0.03	(0.02;0.04)	0.03	(0.02;0.04)	32.33	(24.5;48)	
Mundari tribes, India	Patrilocal	NRY	0.04	(0.01;0.07)	0.08	(0.02;0.21)	23	(16.33;79)	(Kumar et al. 2006)
		mtDNA	0.01	(0.00;0.03)	0.02	(0.00;0.03)	98	(33.33;∞)	
Highland tribes, Western New Guinea	Patrilocal	NRY	0.57	(0.38;0.78)	0.77	(0.44;1.00)	0.4	(0;0.72)	(Tommaseo-Ponzetta et al. 2002; Kayser et al. 2003)
		mtDNA	0.2	(0.15;0.23)	0.22	(0.16;0.26)	3.9	(3.65;4.93)	
Lowland tribes, Western New Guinea	Patrilocal	NRY	0.16	(0.04;0.31)	0.3	(0.06;0.71)	4.38	(3.03;7.25)	(Tommaseo-Ponzetta et al. 2002; Kayser et al. 2003)
		mtDNA	0.03	(0.01;0.05)	0.04	(0.02;0.06)	32	(19.6;94)	
Hill tribes, Thailand	Patrilocal	NRY	0.39	(0.29;0.50)	0.72	(0.45;1.00)	0.72	(0;1.1)	(Oota et al. 2001)
		mtDNA	0.08	(0.05;0.11)	0.08	(0.05;0.13)	11.5	(8.64;17.4)	
Hill tribes, Thailand	Matrilocal	NRY	0.066	(0.02;0.11)	0.18	(0.06;0.40)	12.42	(8.55;30)	(Oota et al. 2001)
		mtDNA	0.16	(0.09;0.2)	0.18	(0.09;0.26)	5.15	(4.55;8.22)	
Khasi tribes, India	Matrilocal	NRY	0.003	(-0.0058;0.014)	0.01	(-0.01;0.05)	330	(72.30;∞)	(Kumar et al. 2006)
		mtDNA	0.029	(0.02;0.04)	0.031	(0.021;0.042)	33.41	(24.48;47.9)	

Table 3.22: NRY and mtDNA genetic differentiation for the Basotho samples, and other populations from published works with various residential norms.

Φ_{ST}, Standardised Φ_{ST} (Φ'_{ST}) and Nm values (all with 95% confidence intervals - CI) estimated as described in Materials and Methods.

Populations	Locus	Distance type	1/Nb (F_{ST})	SE	1/Nb (F'_{ST})	SE
Basotho, Lesotho	NRY	Normal	0.0000021	0.000081	0.0000078	0.00016
	mtDNA	Normal	-0.00013	0.00021	-0.00015	0.00024
	NRY	Logged	0.0014	0.0051	0.003	0.0099
	mtDNA	Logged	-0.0053	0.013	-0.0061	0.015

Table 3.23: Absolute values of the slope of the regression line (1/Nb) for the Basotho data, calculated using F_{ST} and F'_{ST} values.
Normal: distance between populations in km; Logged: logarithm of the distance in km between populations; SE: Standard error.

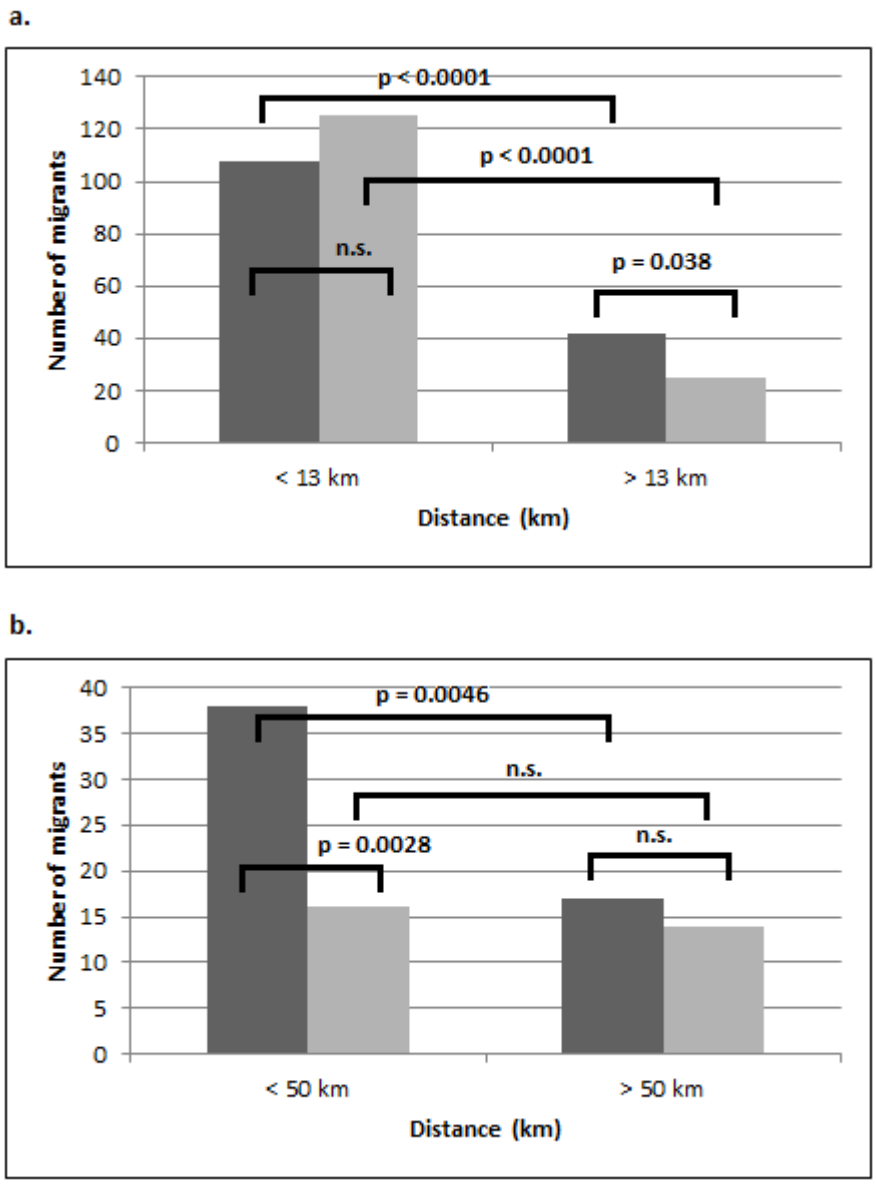


Figure 3.1: Summary of the number of migrants for the Basotho data.
Divided by female (dark grey bars) and male (light grey bars), and those migrating above and below average dispersal (a) and migration (b) distances. Significant p values are reported in the figure; n.s. = not significant at the 0.05 significance level.

Population	Distance (km)	NRY F_{ST}	mtDNA F_{ST}	NRY F'_{ST}	mtDNA F'_{ST}	Nm NRY	Nm mtDNA	male/female Nm
LOW-1a, LOW-1b	< 24 km	-0.0052 (-0.055;0.047)	0.025 (-0.028;0.085)	-0.010 (-0.090;0.11)	0.033 (-0.035;0.12)		39.04 (12.21;infinity)	
LOW-1a, LOW-1c	< 24 km	0.30 (0.21;0.39)	-0.064 (-0.097;-0.029)	0.54 (0.32;0.81)	-0.089 (-0.13;-0.043)	1.52 (0.91;1.72)		
LOW-1b, LOW-1c	< 24 km	0.21 (0.14;0.29)	-0.042 (-0.077;0.00056)	0.39 (0.22;0.60)	-0.057 (-0.099;0.00083)	2.87 (2.72;2.88)		
LOW-3a, LOW-3b	< 24 km	0.023 (-0.022;0.073)	0.21 (0.12;0.29)	0.051 (-0.041;0.19)	0.27 (0.14;0.40)	40.67 (14.32;infinity)	3.50 (2.96;5.15)	11.62 (2.78;infinity)
LOW-3a, LOW-3c	< 24 km	0.038 (-0.0093;0.083)	0.020 (-0.020;0.063)	0.079 (-0.016;0.21)	0.030 (-0.027;0.10)	24.36 (12.32;infinity)	48.47 (16.21;infinity)	0.50 (0;infinity)
LOW-3b, LOW-3c	< 24 km	-0.0066 (-0.045;0.028)	0.030 (-0.011;0.072)	-0.013 (-0.075;0.066)	0.039 (-0.013;0.10)		32.53 (13.99;infinity)	
LOW-3a, LOW-1a	> 24 km	0.13 (0.060;0.20)	0.15 (0.070;0.22)	0.27 (0.10;0.46)	0.21 (0.091;0.33)	5.57 (4.55;8.99)	5.31 (4.07;9.63)	1.05 (0.47;2.21)
LOW-3a, LOW-1b	> 24 km	0.10 (0.037;0.17)	0.019 (-0.042;0.096)	0.22 (0.066;0.42)	0.028 (-0.056;0.15)	7.59 (5.60;15.70)	50.82 (10.95;infinity)	0.15 (0;1.43)
LOW-3a, LOW-1c	> 24 km	-0.0071 (-0.090;0.086)	0.071 (0.016;0.12)	-0.014 (-0.15;0.21)	0.10 (0.021;0.18)		12.65 (8.22;52.59)	
LOW-3b, LOW-1a	> 24 km	0.011 (-0.037;0.064)	-0.048 (-0.068;-0.023)	0.022 (-0.059;0.15)	-0.063 (-0.084;-0.032)	86.80 (16.47;infinity)		
LOW-3b, LOW-1b	> 24 km	-0.023 (-0.066;0.026)	0.075 (0.0039;0.15)	-0.046 (-0.1077;0.062)	0.095 (0.0048;0.21)		12.12 (6.52;201.63)	
LOW-3b, LOW-1c	> 24 km	0.16 (0.083;0.23)	-0.04 (-0.074;-0.0029)	0.30 (0.14;0.52)	-0.053 (-0.091;-0.0042)	4.44 (3.72;5.77)		
LOW-3c, LOW-1a	> 24 km	0.017 (-0.028;0.070)	0.013 (-0.030;0.066)	0.032 (-0.044;0.15)	0.019 (-0.038;0.099)	56.22 (14.82;infinity)	73.30 (15.64;infinity)	0.77 (0;infinity)
LOW-3c, LOW-1b	> 24 km	-0.062 (-0.081;-0.041)	-0.014 (-0.067;0.048)	-0.13 (-0.14;-0.10)	-0.020 (-0.086;0.072)			
LOW-3c, LOW-1c	> 24 km	0.16 (0.059;0.25)	-0.047 (-0.077;-0.013)	0.29 (0.093;0.54)	-0.068 (-0.10;-0.020)	4.48 (3.64;7.75)		

Table 3.24: F_{ST} and Nm ratios in Lesotho villages for NRY and mtDNA.

LOW-1a, LOW-1b, LOW-1c = LOW-1 villages a, b & c, LOW-3a, LOW-3b, LOW-3c = LOW-3 villages a, b & c.

Discussion

It is known that a variety of different cultural traits, including linguistic variation, residential practices, marital customs and socio-economic factors, affect the distribution of genetic variation in human populations by influencing the amount and direction of gene flow between groups (Destro-Bisol et al. 2004b; Belle and Barbujani 2007; Hammer et al. 2008; Laland et al. 2010; Heyer et al. 2012). Although studies have been carried out previously, there is not clear information with regards to some of these cultural practices and the impact that they are likely to have.

Residential practices, such as patrilocality for example, have presented contradictory results, possibly as a consequence of other confounding factors obscuring effects that patrilocality might be having (Seielstad et al. 1998; Oota et al. 2001; Wilder et al. 2004; Besaggio et al. 2007). Studying populations that speak different languages and that are separated by relatively large distances, as with a study on Thailand Hill tribes (Oota et al. 2001; Besaggio et al. 2007) and another continental study (Wilder et al. 2004) might be problematic (Wilkins and Marlowe 2006). Previous studies also have not often taken into account the fact that the different genetic markers tested are not always easily comparable, with STRs used for NRY analysis and SNPs for mtDNA analysis producing different haplotype resolution and thus affecting the variation detected within and between groups (Meirmans and Hedrick 2011). For example, the extremely large differences observed by Seielstad *et al.* (1998) in Y chromosome vs. mtDNA variation between populations could be explained by a higher power in haplotype resolution offered by the SNPs selected for mtDNA than those genotyped on the Y chromosome as a result of a bias in marker discovery and selection (Currat and Excoffier 2005).

The way in which the samples used here from Lesotho and Namibia were collected allows for additional information to be taken into account when analysing the data. For example, a large number of the Lesotho samples were from the same language group, from the same ethnic group, and were distributed over a relatively small geographic area, thus going some way in controlling for other confounding factors. With these other factors ruled out, no differences in within and between area mtDNA and NRY variation were observed despite the confirmation by biodemographic data of a higher female than male migration rate. The approach usually applied to identify differences in migration between sexes is based on the relationship proposed by Wright (1931), $F_{ST} = 1/(4Nm +1)$, under the assumptions of the Island model: absence of spatial structure, with migration being homogenous across populations, no mutations, and populations having reached equilibrium – an unrealistic set of requirements for most natural populations (Whitlock and McCauley 1999). In order to take this into account, an Isolation by Distance model as described by Rousset (1997) was also explored. This approach also failed to detect evidence of differences in population differentiation for the NRY and mtDNA.

The information provided by the questionnaires also allowed analysis of the number of migrants, and the distances migrated. Results indicate differences between the genders, and suggest that females do indeed tend to move more often than males. Both males and females mostly migrated at short distances, below the average dispersal range (13 km, 72% and 83% of females and males respectively). However, female migrants moved at distances shorter than males as suggested by their shorter average migration distance (42 vs. 51 km) and the significantly larger number of female migrants moving below than above the average migration distance (50 km) (Figure 3.1). When the overall average migration distance is considered, significant differences in the number of male and female migrants are observed only at distances below the average migration distance

(Figure 3.1). This suggests that despite the fact that females move more often than males, they do it preferentially at distances shorter than males. Following this observation, different migration rates are expected for males and females when different distance ranges are considered. The average distance among populations analysed for the genetic data is beyond the average migration distance (87 km), where a similar effective migration rate for both genders is expected. Results showed no differences in the degree of population structure for the two genetic systems, in agreement with expectations based on migration estimates (Tables 3.22 and 3.23). It is estimated that at least twice as many females than males moved at shorter distances (below the average dispersal distance, 13 km), while a smaller number, similar to the male figure, would move at larger distances (above the average migration distance; 50 km).

Results from simulations in Marks et al. (2012) show that the between population differentiation observed for NRY and mtDNA is similar when a relatively low migration rate is considered, that is the same for both markers. When a larger female migration rate is applied, however, as in the case of geographically close populations, the mtDNA became less structured than the NRY, with a much smaller fraction of simulations showing significant between population differentiation. This implies that patrilocal residence will mostly impact geographically close groups, unsurprising when it is known that the distance between parents and offspring in human populations very rarely exceeds 50 km, and more often than not is below 10 km, close to that observed in the Lesotho dataset (Wijsman and Cavalli-Sforza 1984). Unfortunately inadequate data was available for single villages in order to test this hypothesis. Nevertheless, as predicted there is evidence of an increase in Y differentiation at shorter distances (Table 3.24).

The relationship between migration distance and number of migrants was tested further in Marks et al. (2012), using biodemographic data for a nationwide Spanish sample. Interestingly, this data also shows a higher number of female than male migrants, but no differences in mtDNA and NRY variation between samples were observed. Use of the IBD model again failed to find evidence of differences in population differentiation for the NRY and mtDNA. This suggests that the pattern observed in Lesotho could be representative of a widespread situation.

It is not sufficient, however, to concentrate on patrilocality, as other cultural factors also have an effect on genetic diversity, resulting in different patterns. In addition, inheritance of clan association, and whether a group is endogamous or exogamous might have an impact upon the genetic diversity and structure of a population. It should be noted for example, that not all marriage migration patterns will necessarily comply with the model used above. For example, strict endogamy practiced by some societies in India will lead to female movement mostly within caste boundaries, having no impact on the overall genetic variability (Kumar et al. 2006). With regards to clan affiliation, if only individual clan is taken into account this might not give a true picture of the history of the population, as it is not necessarily representative of both the male and female history. We know from our samples that in the patrilineal Basotho, for individuals that do not match the clan of both of their parents, the majority inherit their clan from the male parent (54.8%, significantly higher than those matching the female parent) (Table 3.14). The Basotho are, however, meant to be an endogamous population, choosing spouses from within the same ethnic group, and also within the same clan (Sheddick 1953). Ethnographic results suggest that, although the Basotho are mainly endogamous at ethnic group level, a large number do not marry within clan (56%) (Table 3.14), and of those that do, half marry into a different sub clan (Table 3.19). Pairwise F_{ST} results for the NRY displayed more structure than those for the mtDNA (Tables 3.16

and 3.17), suggesting some structure across the different clans, and both unstandardised and standardised AMOVA results also indicate the NRY as being more different across clans than the mtDNA when all four clans are included in analyses (Table 3.18). A previous study analysing diversity within and between clans suggests that results such as these do in fact fit with populations that are endogamous at the ethnic group level, but exogamous at the clan level (Heyer et al. 2009). Populations that are endogamous at the clan level are expected to also display differences in mtDNA between clans. It could be that the Basotho results are confused, as mentioned with regards to the Central Asian study (Heyer et al. 2009), by the fact that the Basotho are made up of a number of different tribes that combined fairly recently under Moshoeshoe I, and are therefore not a homogenous group (Sheddick 1953; Gill 1993). Heyer et al (2009) suggest that a grouping together of unrelated tribes, along with a low level of endogamy as is seen in the Basotho, might account for lower levels of mtDNA genetic differentiation among the clans than would be expected in a truly endogamous population.

The Namibian data could not be analysed at a similar genetic level to the Basotho samples due to the number of samples available for each ethnic group, clan and location. Ethnographic analysis did, however, show how important it is to collect as much biodemographic information as possible as results were not always as initially expected. Cross ethnic group comparisons (Table 3.7) and clan information (Table 3.8) suggest that the majority of individuals marry within the same group, and therefore that the populations are endogamous, as expected (Tuupainen 1970; Estermann 1976; Davies 1994). Location information, however, indicates that all Namibian populations have a higher male than female migration rate (Tables 3.9 and 3.10), although none significantly so. This is particularly interesting seeing as the Damara, Herero, Himba and Nama (Barnard 1992; Malan 1995) are reported as being patrilocal. Additionally, groups such as the Herero and Ju|'hoansi

show a different pattern of migration to the other populations, with the Herero displaying more neolocal migration, and the Ju|'hoansi a more static lifestyle, with a large number of individuals all being born in the same location (Table 3.9). This indicates that the stereotypical image of hunter-gatherer groups constantly moving location does not hold for modern groups of Ju|'hoansi. It would be interesting to analyse the age of the Ju'|hoansi individuals studied here, and if possible compare this with results from other age groups. It is possible that the pre-independence policies of the South African Defence Force, in enforcing settlement of the Ju'|hoansi in specific areas, might have resulted in the majority of individuals being born in the same locations, while analysis of more recent births might yet indicate a return to traditional practises. Results such as these emphasise that data collection is important as traditional anthropological views are not necessarily correct in every situation, and any genetic analyses conducted on these samples might have been confused by a misunderstanding of the residential norms displayed by these groups. Collection of such data might also provide information about changes in lifestyles and practices over time.

The results here emphasise the importance of more thorough sampling techniques, where biodemographic data is collected in addition to genetic information. It is necessary to understand and integrate both the cultural information that can be obtained from individuals, with genetic information in order to get a more accurate picture of the processes that have shaped human diversity (Moreau et al. 2011). The importance of this is particularly evident here, where the information collected does not always fit with expected practices. Gathering this additional information is not costly, and more care should therefore be taken to obtain careful ethnographic and geographic characterisation of the samples collected for genetic analysis (Winney et al. 2012). Of particular interest would be future work focused on the collection of a detailed set of samples

that are geographically very close, in order to verify the impact of the patrilocal norm at relatively short distances.

CHAPTER 4: Frontiers and Southern African Bantu-speaking populations

Introduction

Migrations and subsequent admixture between populations have shaped the current distribution of genetic variation. Within Sub-Saharan Africa, the expansion of Bantu-speaking iron-using agropastoralist populations (the 'Bantu expansion' (Vansina 1995)) is arguably one of the most influential demographic events to have occurred in Africa. The spread of Bantu-speaking peoples is believed to have started around 5,000 years ago, with movement from north-west Cameroon/southern Nigeria to throughout Sub-Saharan Africa (Newman 1995). Prior to this expansion, it is generally accepted that the Khoekhoen and San were the sole inhabitants of southern Africa (Ehret 1982a, b). The arrival of the Bantu-speaking farmers, therefore, potentially led to the isolation of, and/or admixture with, the neighbouring hunter-gatherer populations (Destro-Bisol et al. 2004b; Mitchell et al. 2008; Mitchell 2009).

Previous genetic studies of Sub-Saharan African populations have highlighted the presence of male and female lineages specifically associated with, and at high frequency in, Bantu speaking populations (both maternal – e.g. mtDNA haplogroups L0a, L2a, L3e, and paternal – NRY lineages E and B2a), whilst lineages specifically associated with 'Khoisan' speaking groups (maternal mtDNA haplogroups – L0d, L0k, and paternal NRY lineages – A3b1*, B2b, E1b1b1g), have been found only at very low levels in Bantu-speaking populations (Chen et al. 2000; Pereira et al. 2001; Pereira et al. 2002; Salas et al. 2002; Behar et al. 2008; Coelho et al. 2009; de Filippo et al. 2010). Work

carried out so far in Southern Africa has, however, concentrated on countries such as Angola (Plaza et al. 2004; Coelho et al. 2009), Mozambique (Pereira et al. 2001; Pereira et al. 2002; Salas et al. 2002), Zambia (de Filippo et al. 2010) and Zimbabwe (Castri et al. 2009). A small number of studies have analysed populations from further south, but these all have limited use. Quintana-Murci et al. (2010) found high frequencies of maternal 'Khoisan' lineages in the South African Coloured (SAC) population, however, this population is known to have undergone high levels of admixture with a number of other populations from outside Africa. Schlebusch et al. (2009) report high levels of L0d/L0k (29%) in 'South-eastern Bantu speakers', but do not provide the sequence data for further analysis, and a recent study includes NRY data for two Botswana populations, however, no corresponding mtDNA data is available (de Filippo et al. 2011). There is scope, therefore, for more detailed work to be carried out on southernmost Africa, with the collection of well-defined samples needed in order to answer questions relating to the peopling of the region.

Lesotho and the surrounding regions of South Africa (referred to here collectively as southernmost- Africa) were among the last places in Sub-Saharan Africa to be settled by the Bantu expansion. In Lesotho, this settlement began in the mid-seventeenth century, and was only completed at the end of the nineteenth century (Mitchell 1992; Gill 1993). Today, no San groups remain in the region, and the predominant languages spoken (Southern Sotho, Zulu, Phuthi, Xhosa) are Bantu, with no local 'Khoisan' languages present (Lewis and SIL International 2009). Although there are no longer people present in the region who self-identify as San, archaeological, linguistic and historical evidence suggest that there were significant interactions between the indigenous San and Bantu farmers (Barnard 1992; Mitchell 1992; Prins and Lewis 1992; Gill 1993; Jolly 1996). Archaeological data from Lesotho strongly supports a scenario where San communities were not isolated from the nearby farmers, suggesting they might have experienced significant

contact and exchanges long before the Bantu-speakers settled in the region (Maggs 1976; Gill 1993; Jolly 1996; Mitchell et al. 2008; Mitchell 2009). The archaeological findings from this region point towards a static frontier having existed between the Bantu farmers and indigenous hunter gatherers (Alexander 1977, 1984), not dissimilar to those in North America (Alexander 1977). This frontier extended along the Maloti/Drakensberg Mountains and into Botswana, as a result of an ecological and environmental transition between Southernmost Africa, and the neighbouring South-Eastern region (Mozambique, Zambia, Zimbabwe) (Mitchell 2009) (Figure 4.1).

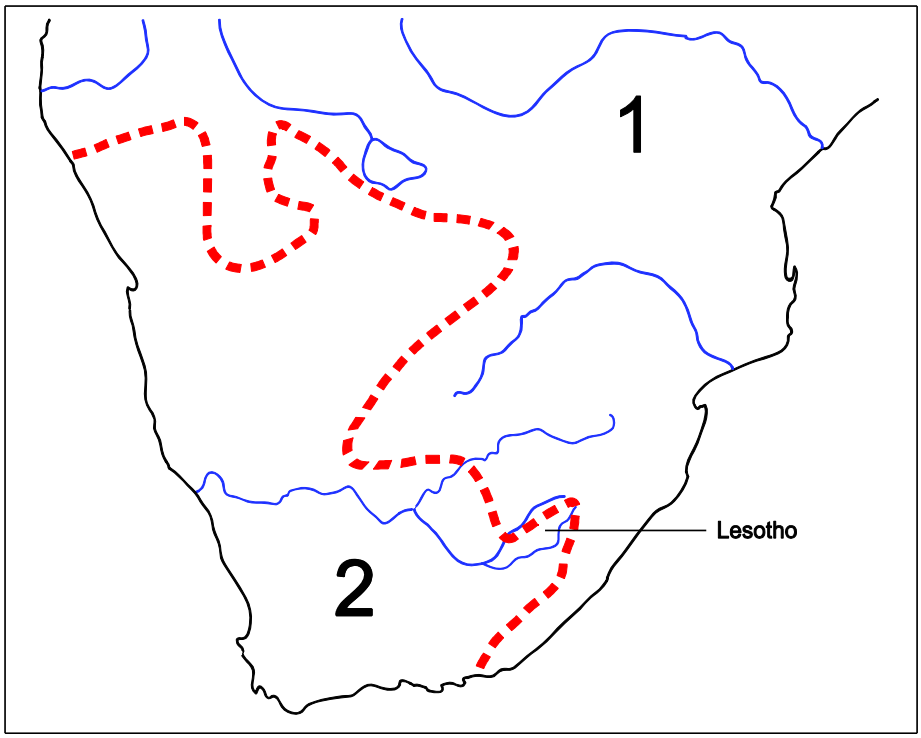


Figure 4.1: Agriculturalist limit mid-second millennium AD.
Adapted from Mitchell (2004). The dashed line indicates the approximate limit of Late Iron Age agropastoralist settlement in Southern Africa by the mid-second millennium AD. Region 1 is the area where the Bantu-speakers had settled, and Region 2 is where they had yet to reach.

As Alexander (1977, 1984) notes, static frontiers favour prolonged interaction between groups on either side of the boundary, potentially enabling increased genetic exchanges. This would make the genetic composition of populations in Southernmost Africa different to those in other areas of Southern Africa; accounting for the low frequency of hunter-gatherer lineages observed in previously studied populations of this region (Pereira et al. 2001; Pereira et al. 2002; Salas et al. 2002; Plaza et al. 2004; Castri et al. 2009; Coelho et al. 2009; de Filippo et al. 2010).

Samples

Analyses were conducted using maternal and paternal markers from ethnic groups in Lesotho with greater than 20 individuals (Basotho, Ndebele, Thembu), and populations in the surrounding region of South Africa (Xhosa and Zulu). mtDNA results for the Lowland Lesotho samples are reported in Levy (2010), and are incorporated here. The Lesotho samples were grouped according to the ethnic groups of their maternal line (using questionnaire information) for mtDNA analyses, and the ethnic groups of the paternal line for NRY analyses. Samples from Botswana were also included in some analyses. These populations fall within the borders of the archaeologically defined static frontier, and, as predicted by Alexander (Alexander 1977, 1984), are therefore expected to show genetic evidence of prolonged contact between hunter-gatherers and farmers.

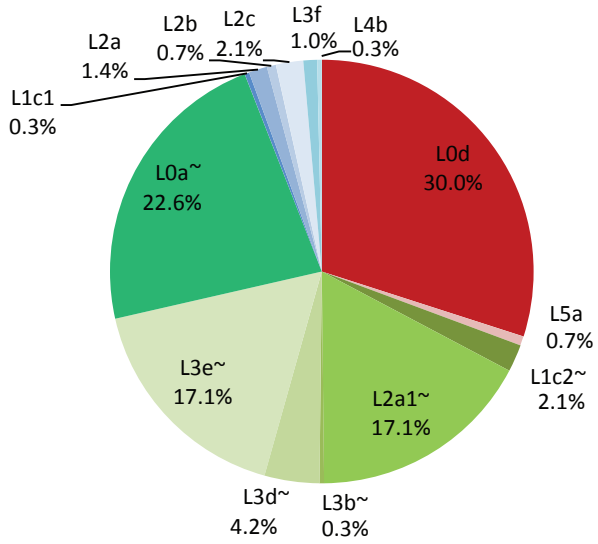
These Southernmost African samples are compared with samples from the neighbouring region of South-East Africa (Mozambique, Zambia and Zimbabwe) (Pereira et al. 2001; Salas et al. 2002; Castri et al. 2009; de Filippo et al. 2010; de Filippo et al. 2011), in an attempt to shed light on the demographic dynamics associated with the expansion of the Bantu-speaking agropastoralists. In addition to this geographical basis of grouping the populations according to where the frontier is hypothesised to be, it is also possible to divide the populations on a linguistic basis. Southern

African Bantu populations speak South-East Bantu languages (Holden 2002). A number of the languages spoken in Southern Africa, however, include a large number of 'click' sounds which are typical of 'Khoisan' languages, whereas the languages of South East Africa do not (Herbert 1990; Guldemann and Stoneking 2008). The Lesotho (speaking SeSotho) and South African populations (speaking Xhosa and Zulu) used in this work fall into the groups with high frequencies of clicks present in their languages. In fact, 15% of Xhosa is estimated to be derived from 'Khoisan' (Lewis and SIL International 2009). The populations analysed here have therefore been divided into South-East Bantu, and South-East Bantu-Click. This linguistic classification results in the same groupings as those done on a geographic basis.

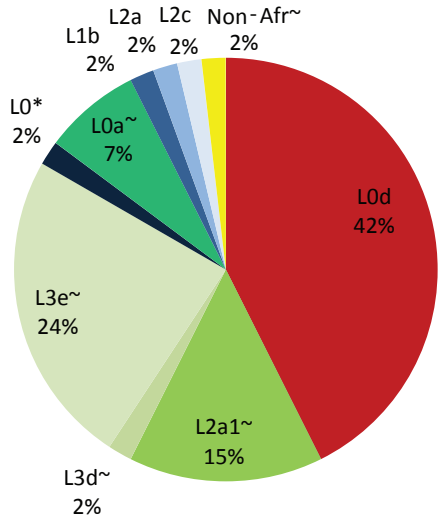
Results

Haplogroup and haplotype composition

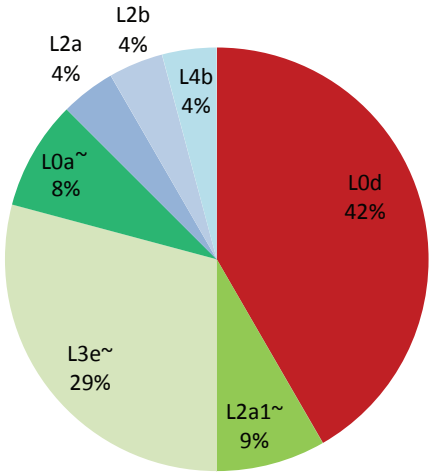
At the mtDNA haplogroup (Hg) level, all of the Lesotho samples belong to Sub-Saharan African haplogroups L(xM,N) (Figure 4.2; Table 4.1). A large proportion of these (Basotho 29.97%, Ndebele 33.33%, Thembu 41.67%, Overall 30.88%) fall within Hg L0d, a lineage associated with 'Khoisan' populations (Vigilant et al. 1991; Chen et al. 2000; Salas et al. 2002; Kivisild et al. 2004). Both the Xhosa and Zulu samples also show high frequencies of this 'Khoisan' type (42.6% and 29.6% respectively) (Figure 4.2; Table 4.1). Such large estimates of 'Khoisan' lineages in Southern African Bantu populations are not common. A recent study of the South African Coloured found L0d to be at 60% (Quintana-Murci et al. 2010), and a study of mixed "South-eastern Bantu speakers" (South African Zulu, Xhosa, Sotho and Tswana) recorded L0d levels of 29% (Schlebusch et al. 2009). Aside from this, L0d is at extremely low frequencies in populations studied elsewhere in Southern Africa, namely Mozambique, Zambia and Zimbabwe (Salas et al. 2002; Plaza et al. 2004; Coelho et al. 2009; de Filippo et al. 2010) (Table 4.1). Other haplogroups present in the Lesotho, Xhosa and Zulu populations include lineages probably introduced to southern Africa during the Bantu expansion (L0a, L1, L2, L3) (Pereira et al. 2001; Salas et al. 2002). The presence of these is to be expected, as the subject groups speak Bantu languages. The high percentage of L0d present in the populations is, however, more unusual and suggests a different maternal origin.



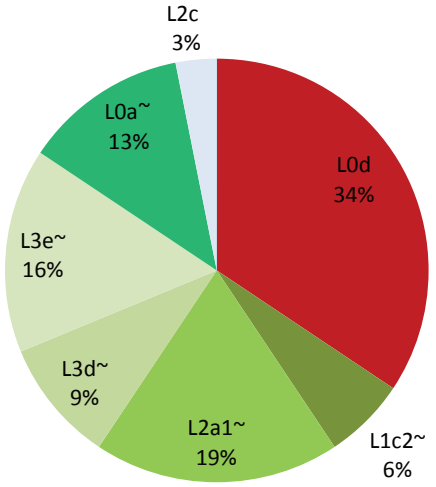
Mosotho, n = 287



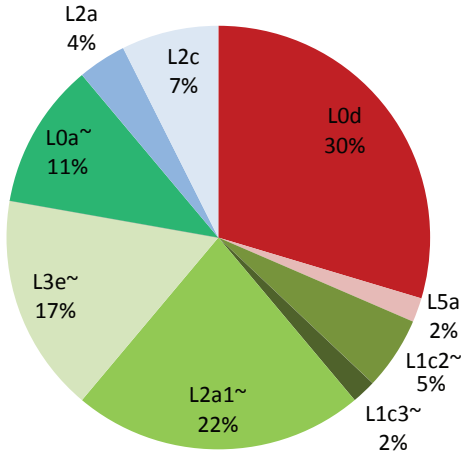
Xhosa, n = 54



Thembu, n = 24



Ndebele, n = 33



Zulu, n = 54

Figure 4.2: mtDNA haplogroup frequencies for the Lesotho, Xhosa and Zulu populations
~ = Bantu-associated haplogroups

[a]Area	[b]Country	[c]Population	N	L0*	L0a~	L0d	L0f	L0k	L1b	L1c*	L1c1	L1c1a2	L1c1b~	L1c2~	L1c3~	L2*	L2a	L2a1~	L2b	L2c	L2d	L3*	L3b~	L3d~	L3e~	L3f	L3h	L4b	L5a	Non-Afr#	Freq L0d & L0k	Freq Non-Bantu
SE	Zam	Bisa	42	0	7	0	4	0	1	0	0	0	2	3	3	0	4	10	0	0	0	0	1	1	2	2	0	2	0	0	0	0.31
SE	Moz	Chopi	27	0	4	0	0	0	1	0	0	0	0	1	1	0	0	8	1	0	0	2	0	4	3	2	0	0	0	0	0	0.22
SE	Moz	Chwabo	20	0	4	0	0	0	1	0	1	0	0	0	0	0	0	8	0	0	1	0	2	2	1	0	0	0	0	0	0	0.15
SE	Zam	Kunda	36	0	9	1	2	0	0	0	0	0	1	1	1	0	0	4	0	0	1	0	4	4	7	1	0	0	0	0	0.028	0.083
SE	Moz	Lomwe	20	0	11	0	0	0	0	1	0	0	0	0	1	0	1	4	0	0	1	0	0	0	1	0	0	0	0	0	0	0.15
SE	Moz	Makhuwa	20	0	9	0	0	0	0	0	1	0	0	1	0	0	1	3	0	0	0	0	0	1	4	0	0	0	0	0	0	0.1
SE	Moz	Makonde	19	0	3	1	0	0	0	0	1	0	0	1	0	0	0	4	0	0	0	0	2	0	7	0	0	0	0	0	0.053	0.11
SE	Moz	General	109	0	16	8	0	0	1	2	0	0	0	2	1	0	1	46	2	1	0	0	4	2	19	2	0	0	2	0	0.073	0.17
SE	Moz	Ndau	19	0	8	3	0	0	0	0	0	0	0	1	0	0	0	5	0	0	0	0	1	1	0	0	0	0	0	0	0.16	0.16
SE	Moz	Nguni	11	0	5	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	1	0	1	2	0	0	0	0	0	0	0.09
SE	Moz	Nyanja	20	0	6	0	0	0	0	0	0	0	0	0	0	0	1	8	0	0	0	0	0	3	2	0	0	0	0	0	0	0.05
SE	Moz	Nyungwe	20	0	6	0	0	0	0	1	0	0	0	0	1	0	0	4	0	0	1	0	2	2	3	0	0	0	0	0	0	0.1
SE	Moz	Ronga	21	0	4	4	0	0	0	0	0	0	0	0	0	0	0	5	1	1	0	0	0	0	4	1	0	0	0	1	0.19	0.33
SE	Moz	Sena	21	0	7	0	0	0	0	0	0	0	0	0	1	0	0	8	0	0	0	0	0	3	1	1	0	0	0	0	0	0.048
SE	Moz	Shangaan	22	0	6	1	0	0	1	0	0	0	0	2	0	0	0	6	1	0	0	0	0	2	3	0	0	0	0	0	0.045	0.14
SE	Zim	Shona	59	0	16	1	0	1	2	0	0	0	0	3	2	0	2	10	1	1	0	0	4	4	11	0	0	1	0	0	0.034	0.15
SE	Moz	Shona 2	17	0	5	0	0	0	1	0	0	0	0	0	0	0	0	7	0	1	0	0	0	0	2	1	0	0	0	0	0	0.18
SE	Moz	Tonga	20	0	3	1	0	0	0	1	0	0	0	0	1	0	0	7	1	0	0	0	0	0	6	0	0	0	0	0	0.05	0.15
SE	Moz	Tswa	19	0	2	3	0	0	0	0	0	0	0	0	0	0	1	6	0	0	0	1	0	2	2	2	0	0	0	0	0.16	0.37
SE	Moz	Yao	10	0	4	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	4	0	0	0	0	0	0	0
Mean Food Producer South East			552	0	135	23	6	1	8	5	3	0	3	15	12	0	11	156	7	4	4	4	21	32	84	12	0	3	2	1	0.044	0.15
Mean South East			552	0	135	23	6	1	8	5	3	0	3	15	12	0	11	156	7	4	4	4	21	32	84	12	0	3	2	1	0.044	0.15
Sm	RSA (Ang)*	Khwe	31	0	3	5	0	7	1	0	0	0	0	0	0	0	0	3	2	0	0	0	0	1	9	0	0	0	0	0	0.39	0.45
Sm	RSA (Ang)*	Kung	43	0	1	22	0	11	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	5	2	0	0	0	0	0	0.77	0.81

Sm	Bots/ Nam	[d]Kung	24	2	0	21	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.92	1
Sm	Bots	Mixed Bushmen	73	0	0	63	0	3	0	0	0	0	0	0	1	0	0	5	0	0	0	0	0	0	1	0	0	0	0	0	0.9	0.9
Sm	RSA (Ang)*	!Xun-Khwe	17	0	0	12	0	3	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0.89	1
Mean 'Khoisan' Southernmost Africa			188	2	4	123	0	25	1	0	0	0	0	0	1	0	0	8	6	0	0	0	0	6	12	0	0	0	0	0	0.78	0.83
Sm	Bots	Mixed Bantu	50	0	3	28	0	0	0	0	0	0	0	2	0	0	0	10	0	1	0	0	0	0	6	0	0	0	0	0	0.56	0.58
Sm	Les	Basotho	287	0	65	86	0	0	0	0	1	0	0	6	0	0	4	49	2	6	0	0	1	12	49	3	0	1	2	0	0.3	0.34
Sm	Les	Ndebele	33	0	4	11	1	0	0	0	0	0	0	2	0	0	0	6	0	1	0	0	0	3	5	0	0	0	0	0	0.33	0.39
Sm	Les	Thembu	24	0	2	10	0	0	0	0	0	0	0	0	0	0	1	2	1	0	0	0	0	0	7	0	0	1	0	0	0.42	0.54
Sm	RSA	Xhosa	54	1	4	23	0	0	1	0	0	0	0	0	0	0	1	8	0	1	0	0	0	1	13	0	0	0	0	1	0.43	0.5
Sm	RSA	Zulu	54	0	6	16	0	0	0	0	0	0	0	3	1	0	2	12	0	4	0	0	0	0	9	0	0	0	1	0	0.3	0.43
Mean Food Producer Southernmost Africa			502	1	84	174	1	0	1	0	1	0	0	13	1	0	8	87	3	13	0	0	1	16	89	3	0	2	3	1	0.35	0.46
Mean Southernmost Africa			690	3	88	297	1	25	2	0	1	0	0	13	2	0	8	95	9	13	0	0	1	22	101	3	0	2	3	1	0.47	0.65

Table 4.1: Frequencies of mtDNA haplogroups across Southern African populations.

[a] Areas: SE = South East; Sm = Southernmost. N = number of samples.

[b] Country: Ang = Angola; Bots = Botswana; Les = Lesotho; Moz = Mozambique; Nam = Namibia; RSA = South Africa; Zam = Zambia; Zim = Zimbabwe.

[c] See Supplementary Table 1 in Appendix 2 for sources of populations used.

[d] Haplogroups were re-assigned to Kung samples from Vigilant et al. (1991) using sequences available in Genbank (sequence numbers M76241-M76256), as previous papers appear to be in disagreement as to the haplogroup frequencies (Tishkoff et al. 2007a; de Filippo et al. 2010). Sequence M76255 was not included due to ambiguities in the sequence.

~ Bantu haplogroups excluded from Wilcoxon analyses (Chen et al. 1995; Watson et al. 1997; Alves-Silva et al. 2000; Bandelt et al. 2001; Pereira et al. 2001; Salas et al. 2002; Plaza et al. 2004; Batini et al. 2007; Quintana-Murci et al. 2008; Berniell-Lee et al. 2009; de Filippo et al. 2010).

European haplogroups were also excluded from Wilcoxon analyses.

* See Mitchell (2010) for a discussion on the ambiguity of these samples.

Considered together, the Lesotho, Xhosa and Zulu samples belong to 141 different mtDNA haplotypes (separately: Lesotho 124 (Basotho 105, Ndebele 25, Thembu 20), Xhosa 37, Zulu 32). Haplotype diversity (Nei's diversity) values for mtDNA show that all three populations have high genetic diversity (Lesotho = 0.9724 +/- 0.0030 (Basotho = 0.9698 +/- 0.0038; Ndebele = 0.9830 +/- 0.011; Thembu = 0.9819 +/- 0.018), Xhosa = 0.9713 +/- 0.012; Zulu = 0.9706 +/- 0.0096). Within L0d, the Lesotho samples are characterised by 49 different haplotypes (Basotho 42, Ndebele 9, Thembu 8), the Xhosa samples 18, and the Zulu samples 10. They show high genetic diversity (Nei's diversity: Lesotho overall = 0.9495 +/- 0.013 (Basotho = 0.9442 +/- 0.017; Ndebele = 0.9636 +/- 0.051; Thembu = 0.9556 +/- 0.059), Xhosa = 0.9565 +/- 0.034; Zulu = 0.8952 +/- 0.070).

As with the mtDNA analysis, 'Khoisan'-associated lineages were also observed in the paternal gene pool (Hg A3b1-M51) (Knight et al. 2003; Tishkoff et al. 2007a), at an average frequency of 9% across the entire Lesotho sample (Basotho 11%, Ndebele 3.8%, Thembu 4.3%), 3.4% for the Xhosa, and 1.9% for the Zulu (Figure 4.3; Table 4.2). The majority of the samples analysed here (Lesotho 79% (Basotho 76%, Ndebele 84.6%, Thembu 91.3%), Xhosa 86%, Zulu 90.7%) fall into Hg E-M96 – a pan-African Hg associated with the Bantu expansion, which is often of high frequency in Bantu-speaking populations (Coelho et al. 2009; de Filippo et al. 2010; Quintana-Murci et al. 2010). Another Bantu-associated Hg, B2a-M150 (Gomes et al. 2010), is also found in the Southernmost African samples, at an average frequency of 7.8% across the Lesotho, Xhosa and Zulu populations (Table 4.2).

Area[a]	Country	Population[b]	N[c]	Haplogroups							Frequency Non-Bantu[d]
				A	B				E	Other	
						*B**	*B2a*	*B2b*			
SE	Zambia-East	Bisa[3]	33	0	1	*0*	*0*	*1*	32	0	0.03
SE	Zambia-East	Kunda[3]	36	1	0	*0*	*0*	*0*	34	1	0.028
SE	Zambia-West	Fwe[4]	30	0	0	*0*	*0*	*0*	30	0	0
SE	Zambia-West	Kwamashi[4]	27	0	0	*0*	*0*	*0*	27	0	0
SE	Zambia-West	Lozi[4]	94	0	7	*7*	*N/A*	*N/A*	86	1	0
SE	Zambia-West	Luvale[4]	16	0	1	*1*	*N/A*	*N/A*	15	0	0
SE	Zambia-West	Luyana[4]	61	0	3	*3*	*N/A*	*N/A*	57	1	0
SE	Zambia-West	Mbunda[4]	49	0	1	*1*	*N/A*	*N/A*	46	2	0
SE	Zambia-West	Nkoya[4]	29	0	2	*2*	*N/A*	*N/A*	27	0	0
SE	Zambia-West	Shanjo[4]	29	0	1	*1*	*N/A*	*N/A*	28	0	0
SE	Zambia-West	Tonga[4]	30	0	3	*3*	*N/A*	*N/A*	27	0	0
SE	Zambia-West	Totela[4]	15	0	1	*1*	*N/A*	*N/A*	13	1	0
SE	Zambia-West	Mixed Bantu West[4]	100	0	4	*4*	*N/A*	*N/A*	92	0	0
SE	Zimbabwe	Shona[5]	49	0	5	*0*	*5*	*0*	44	0	0
Total Food Producer South East			118	1	6	*0*	*5*	*1*	110	1	0.017
Total South East			118	1	6	*0*	*5*	*1*	110	1	0.017
Sm	Namibia	!Kung/Sekele[11]	32	15	3	*0*	*0*	*3*	14	0	0.56
Sm	Namibia	Tsumkwe San[11]	29	19	9	*0*	*0*	*9*	1	0	0.97
Sm	RSA (Angola)*	!Kung[12]	64	23	5	*0*	*0*	*5*	36	0	0.44
Sm	RSA (Angola)*	Khwe[12]	26	3	0	*0*	*0*	*0*	23	0	0.12
Sm	RSA (Angola)*	Mixed 'Khoisan' Speakers[13]	39	17	11	*0*	*0*	*11*	11	0	0.72
Sm	Botswana	Mixed Bushmen[6,7]	66	14	36	*0*	*35*	*1*	16	0	0.23
Total 'Khoisan' Southernmost Africa			256	91	64	*0*	*35*	*29*	101	0	0.47
Sm	Botswana	Mixed Bantu[6,7]	44	1	23	*0*	*23*	*0*	20	0	0.023
Sm	Botswana	Kalanga[4]	20	0	2	*2*	*N/A*	*N/A*	17	1	0
Sm	Botswana	Tswana[4]	20	1	6	*6*	*N/A*	*N/A*	13	0	0.05
Sm	South Africa	Xhosa Cape Town[6]	57	1	4	*0*	*4*	*0*	49	3	0.018
Sm	South Africa	Zulu Cape Town[7]	54	1	3	*0*	*3*	*0*	49	1	0.019
Sm	South Africa	Xhosa[11]	80	4	4	*0*	*4*	*0*	72	0	0.05
Sm	South Africa	Sotho-Tswana[11]	28	2	5	*0*	*5*	*0*	20	1	0.071
Sm	South Africa	Zulu[11]	29	1	5	*0*	*4*	*1*	23	0	0.069
Sm	Lesotho	Basotho[7]	267	30	32	*0*	*32*	*0*	203	2	0.11
Sm	Lesotho	Ndebele[7]	52	2	5	*0*	*5*	*0*	44	1	0.038
Sm	Lesotho	Thembu[7]	23	1	1	*0*	*1*	*0*	21	0	0.043
Total Food Producer Southernmost Africa			738	46	91	*9*	*81*	*1*	582	19	0.045
Total Southernmost Africa			994	137	155	*9*	*116*	*30*	683	19	0.26

Table 4.2: Frequencies of NRY haplogroups across Southern African populations.

[a] Areas: SE = South East; Sm = Southernmost. [c] N = number of samples.

[b] See Supplementary Table 1 in Appendix 2 for sources of populations used.

Zambia populations from de Filippo et al. 2011 were classed as separate populations if they have over 15 individuals, otherwise they were grouped together as Mixed Bantu West based on the groupings in the original paper.

[d] Non-Bantu haplogroups excludes B2a, E and non-African haplogroups. Any sub-typed E samples that are E-M293 were left in the Non-Bantu group.

* See Mitchell (2010) for a discussion on the ambiguity of these samples.

Country	Population[a]	N[b]	N Haplogroup E[c]	Haplogroups											
				E*	E1b1a	E1b1a7	E1b1a7a	E1b1a8	E1b1a8a	E1b1b1	E1b1b1g	E(xE1)	E2	E2b1	E2b1a
Populations from this work															
Botswana	Mixed-Bushmen	66	**16**	0	6.25	0	12.5	6.25	56.25	0	18.75	0	0	0	0
Botswana	Mixed-Bantu	44	**20**	5	20	0	20	0	40	0	0	0	0	15	0
Lesotho	Basotho	267	**203 (40 typed)**	0	16.26	0	37.44	0	34.98	0	4.43	0	0	4.43	2.46
Lesotho	Ndebele	52	**44**	2.27	2.27	0	20.45	0	72.74	0	0	0	0	0	2.27
Lesotho	Thembu	23	**21**	0	0	0	38.09	0	47.62	0	0	0	0	0	14.29
South Africa	Xhosa	57	**49**	0	6.12	0	24.49	2.04	55.11	0	0	12.24	0	0	0
South Africa	Zulu	54	**49**	0	10.2	0	26.53	0	42.86	0	0	20.41	0	0	0
Bantu populations from previous studies															
Zambia-East	Bisa	33	**32**	0	12.5	0	37.5	46.87	N/A	0	0	0	3.13	N/A	N/A
Zambia-East	Kunda	36	**34**	0	5.88	2.95	35.29	50	N/A	0	0	0	5.88	N/A	N/A
Zambia-West	Fwe	30	**30**	0	6.67	0	16.67	73.33	N/A	0	0	0	3.33	N/A	N/A
Zambia-West	Kwamashi	27	**27**	0	11.11	0	11.11	62.96	N/A	0	0	0	14.82	N/A	N/A
Zambia-West	Lozi	94	**86**	0	10.47	0	18.6	62.79	N/A	0	0	0	8.14	N/A	N/A
Zambia-West	Luvale	16	**15**	0	6.67	0	13.33	73.33	N/A	0	0	0	6.67	N/A	N/A
Zambia-West	Luyana	61	**57**	0	8.77	0	15.79	75.44	N/A	0	0	0	0	0	0
Zambia-West	Mbunda	49	**46**	0	6.52	0	19.57	67.39	N/A	0	0	0	6.52	N/A	N/A
Zambia-West	Nkoya	29	**27**	0	3.7	0	22.22	59.27	N/A	3.7	N/A	0	11.11	N/A	N/A
Zambia-West	Shanjo	29	**28**	0	17.86	0	21.43	60.71	N/A	0	0	0	0	0	0
Zambia-West	Tonga	30	**27**	0	7.41	0	33.33	59.26	N/A	0	0	0	0	0	0
Zambia-West	Totela	15	**13**	0	7.69	0	53.85	38.46	N/A	0	0	0	0	0	0
Zambia-West	Mixed Bantu West	100	**92**	0	7.61	0	31.52	51.09	N/A	2.17	0	0	7.61	N/A	N/A
Zimbabwe	Shona	49	**44**	0	56.82	N/A	N/A	N/A	N/A	40.91	N/A	0	2.27	N/A	N/A
Botswana	Kalanga	20	**17**	0	5.88	0	52.94	35.3	N/A	5.88	N/A	0	0	0	0
Botswana	Tswana	20	**13**	0	0	0	15.38	53.85	N/A	7.69	N/A	0	23.08	N/A	N/A

South Africa	Sotho-Tswana	28	**20**	5	50	10	N/A	N/A	N/A	30	N/A	0	5	N/A	N/A
South Africa	Xhosa	80	**72**	4.17	37.5	5.55	N/A	N/A	N/A	22.22	N/A	0	30.56	N/A	N/A
South Africa	Zulu	28	**23**	4.35	43.48	0	N/A	N/A	N/A	26.09	N/A	0	26.08	N/A	N/A
'Khoisan' populations from previous studies															
South Africa (Angola)*	!Kung	64	**36**	0	41.67	19.44	N/A	N/A	N/A	27.78	N/A	0	11.11	N/A	N/A
Namibia	!Kung/Sekele	32	**14**	0	42.86	28.57	N/A	N/A	N/A	28.57	N/A	0	0	0	0
Namibia	Tsumkwe San	29	**1**	0	0	100	N/A	N/A	N/A	0	0	0	0	0	0
South Africa (Angola)*	Khwe	26	**23**	0	60.87	34.78	N/A	N/A	N/A	0	0	0	4.35	N/A	N/A
South Africa (Angola)*	Mixed 'Khoisan' Speakers	39	**11**	0	63.64	36.36	N/A	N/A	N/A	0	0	0	0	0	0

Table 4.3: Percentages of NRY E sub-haplogroups for samples genotyped in this work and those from previous studies.

Percentages are calculated as the total number of E samples with each sub-haplogroup. Only a subset of Basotho samples were genotyped for additional E SNPs, so frequencies displayed are relative frequencies.

[a] Populations were obtained from references as in Supplementary Table 1, Appendix 2. Some studies did not genotype all of the SNPs, but further information was available from de Filippo et al. (2011), where they inferred the frequency of sub-lineages from STR haplotypes using Linear Discriminant Analysis.

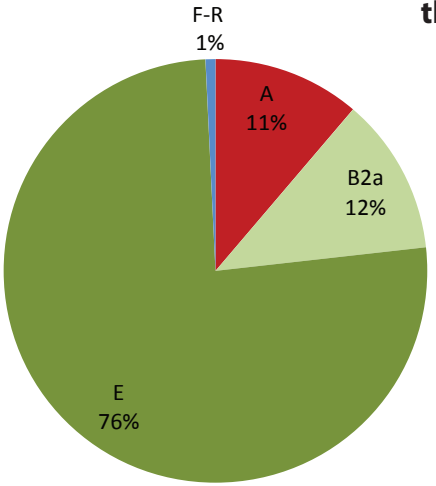
[b] N = Total number of samples.

[c] N haplogroup E = Total number of samples with E haplogroup.

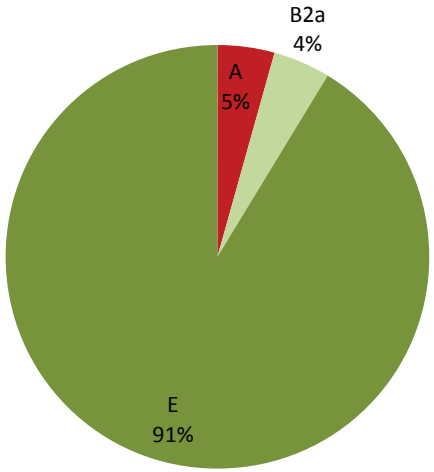
[d] N/A = Information not available – these SNPs were not genotyped in the previous study.

* See Mitchell (2010) for a discussion on the ambiguity of these samples.

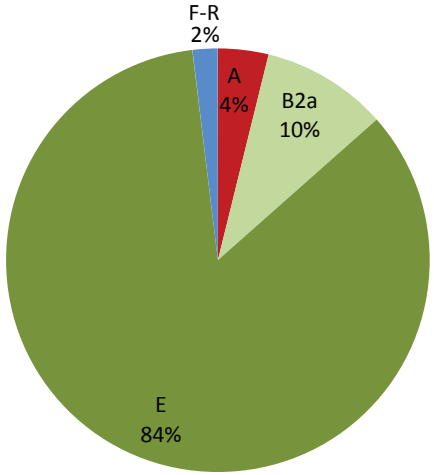
Figure 4.3: NRY haplogroup frequencies for the Lesotho, Xhosa and Zulu populations



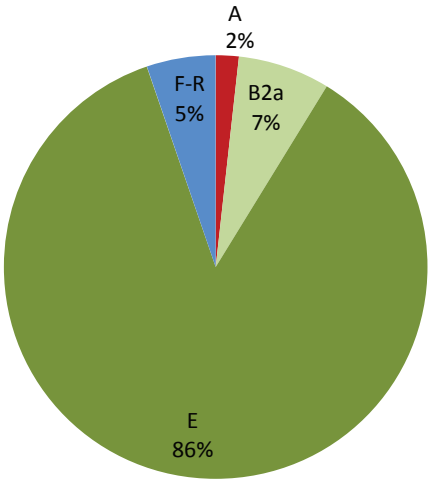
Mosotho, n = 267



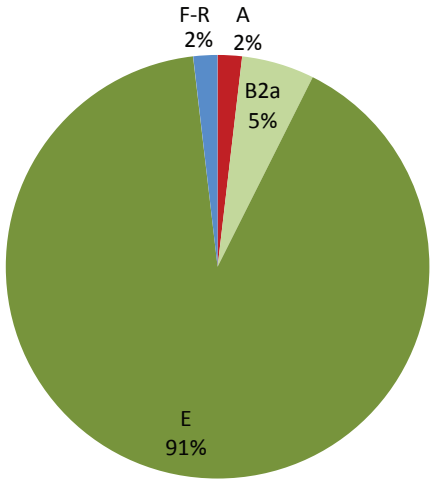
Thembu, n = 23



Ndebele, n = 52



Xhosa, n = 57



Zulu, n = 54

Of the subset of samples that were genotyped with the E multiplexes, only a small number (2 Basotho, 3 Botswana Mixed-Bushmen) had the pastoralist-associated haplogroup E1b1b1g (M293) (Henn et al. 2008). Table 4.3 shows the frequencies for the Botswana, Lesotho, Xhosa and Zulu samples when the E results are taken into account. The frequencies are relative frequencies for the Lesotho populations as only a subset of the samples were genotyped with the E multiplexes. The majority of the E samples from the Lesotho, Xhosa and Zulu samples have sub-haplogroups associated with Bantu-speakers: E1b1a7a-U174 and E1b1a8a-U209 (de Filippo et al. 2011) (Basotho: E1b1a7a 37.4%, E1b1a8a 35%; Ndebele: E1b1a7a 20.5%, E1b1a8a 72.7%; Thembu: E1b1a7a 38.1%, E1b1a8a 47.6%; Xhosa: E1b1a7a 24.5%, E1b1a8a 55.1%; Zulu: E1b1a7a 26.5%, E1b1a8a 42.9%). Frequencies are as a percentage of the total number of E samples with this particular haplogroup. Results from previous studies are also shown in the table (Table 4.3), however, it is difficult to compare frequencies of the various E types across Southern Africa as previous studies have not genotyped samples to the same resolution. It is possible, however, to observe that the majority of E haplogroups from other Bantu-speaking populations belong to Bantu-associated ones (E1b1a7a and E1b1a8), but as other studies do not genotype SNP E1b1a8a-U209 it is not possible to make direct comparisons. Samples from previous studies in Southern Africa were also not analysed for E1b1b1g-M293. A number of samples are listed as E1b1b1, however, it is not possible to determine whether these would be E1b1b1g or not.

For the NRY, the Lesotho samples analysed belong to 162 different haplotypes (Basotho 123, Ndebele 38, Thembu 15), the Xhosa 45 different haplotypes, and the Zulu 46 different haplotypes. Haplotype diversity values are high for all the groups (Nei's diversity: Lesotho overall = 0.9932 +/- 0.0013 (Basotho = 0.9924 +/- 0.0020; Ndebele = 0.9845 +/- 0.0086; Thembu = 0.9381 +/- 0.0403), Xhosa = 0.99 +/- 0.0056; Zulu = 0.9937 +/- 0.0063).

[a]Population	[b]Country	mtDNA Haplogroup F_{ST}					mtDNA Haplotype F_{ST}				
		Basotho	Ndebele	Thembu	Xhosa	Zulu	Basotho	Ndebele	Thembu	Xhosa	Zulu
Bisa	Zam	**0.086*** **(0.00+-0.00)**	**0.074*** **(0.00099+-0.0003)**	**0.15*** **(0.00+-0.00)**	**0.15*** **(0.00+-0.00)**	**0.056*** **(0.0012+-0.0004)**	**0.066*** **(0.00+-0.00)**	**0.056*** **(0.00099+-0.0003)**	**0.090*** **(0.0002+-0.0001)**	**0.10*** **(0.00+-0.00)**	**0.040*** **(0.0047+-0.0007)**
Kunda	Zam	**0.047*** **(0.0014+-0.0003)**	**0.051*** **(0.0078+-0.0009)**	**0.094*** **(0.0011+-0.0004)**	**0.11*** **(0.00+-0.00)**	**0.060*** **(0.00099+-0.0003)**	**0.022*** **(0.020+-0.0013)**	0.015 (0.11+-0.0029)	**0.039*** **(0.025+-0.0016)**	**0.052*** **(0.0020+-0.0004)**	**0.024*** **(0.043+-0.0022)**
Shona	Zim	**0.046*** **(0.00+-0.00)**	**0.052*** **(0.0028+-0.0005)**	**0.10*** **(0.0001+-0.0001)**	**0.11*** **(0.00+-0.00)**	**0.049*** **(0.00099+-0.0003)**	**0.028*** **(0.0016+-0.0004)**	**0.028*** **(0.023+-0.0016)**	**0.052*** **(0.0046+-0.0006)**	**0.061*** **(0.00+-0.00)**	**0.032*** **(0.0079+-0.0008)**
Chopi	Moz	**0.062*** **(0.0022+-0.0005)**	**0.047*** **(0.021+-0.0014)**	**0.12*** **(0.00069+-0.0003)**	**0.12*** **(0.00+-0.00)**	**0.045*** **(0.016+-0.0014)**	**0.056*** **(0.0012+-0.0003)**	**0.041*** **(0.024+-0.0016)**	**0.081*** **(0.0042+-0.0007)**	**0.073*** **(0.0024+-0.0005)**	**0.027*** **(0.049+-0.0023)**
Chwabo	Moz	**0.075*** **(0.0034+-0.0005)**	**0.065*** **(0.017+-0.0012)**	**0.15*** **(0.0004+-0.0002)**	**0.15*** **(0.0001+-0.0001)**	**0.052*** **(0.021+-0.0016)**	**0.050*** **(0.0062+-0.0008)**	0.033 (0.055+-0.0024)	**0.090*** **(0.0041+-0.0006)**	**0.086*** **(0.0025+-0.0006)**	0.027 (0.074+-0.0026)
Lomwe	Moz	**0.10*** **(0.0001+-0.0001)**	**0.15*** **(0.0001+-0.0001)**	**0.23*** **(0.0001+-0.0001)**	**0.23*** **(0.00+-0.00)**	**0.14*** **(0.0001+-0.0001)**	**0.041*** **(0.014+-0.0011)**	**0.066*** **(0.012+-0.001)**	**0.11*** **(0.0027+-0.0005)**	**0.12*** **(0.0002+-0.0001)**	**0.069*** **(0.0074+-0.0009)**
Makhuwa	Moz	**0.063*** **(0.0073+-0.0009)**	**0.093*** **(0.0046+-0.0006)**	**0.15*** **(0.00079+-0.0003)**	**0.16*** **(0.00+-0.00)**	**0.089*** **(0.0029+-0.0005)**	**0.043*** **(0.011+-0.0012)**	**0.055*** **(0.023+-0.0015)**	**0.084*** **(0.0080+-0.0009)**	**0.10*** **(0.00089+-0.0003)**	**0.057*** **(0.013+-0.0011)**
Makonde	Moz	**0.048*** **(0.021+-0.0013)**	0.048 (0.051+-0.0022)	**0.063*** **(0.043+-0.002)**	**0.080*** **(0.011+-0.0011)**	0.035 (0.081+-0.003)	**0.053*** **(0.0065+-0.0008)**	0.031 (0.066+-0.0025)	**0.063*** **(0.020+-0.0013)**	**0.066*** **(0.0075+-0.0009)**	**0.046*** **(0.023+-0.0013)**
Ndau	Moz	0.033 (0.058+-0.0026)	0.051 (0.051+-0.0023)	**0.13*** **(0.0035+-0.0006)**	**0.13*** **(0.00059+-0.0002)**	**0.055*** **(0.028+-0.0015)**	0.005 (0.27+-0.0045)	0.017 (0.15+-0.0035)	**0.050*** **(0.042+-0.0021)**	**0.057*** **(0.011+-0.0011)**	0.03 (0.065+-0.0023)
Nguni	Moz	0.048 (0.076+-0.0029)	**0.079*** **(0.037+-0.0018)**	**0.14*** **(0.011+-0.001)**	**0.15*** **(0.0034+-0.0005)**	**0.082*** **(0.022+-0.0014)**	-0.0097 (0.59+-0.0055)	0.0083 (0.30+-0.005)	0.03 (0.15+-0.0034)	0.044* (0.071+-0.0024)	0.019 (0.18+-0.0039)
Nyanja	Moz	**0.083*** **(0.0012+-0.0004)**	**0.087*** **(0.0085+-0.0008)**	**0.18*** **(0.0001+-0.0001)**	**0.17*** **(0.0001+-0.0001)**	**0.075*** **(0.0089+-0.001)**	**0.042*** **(0.014+-0.0011)**	**0.041*** **(0.042+-0.0018)**	**0.091*** **(0.0063+-0.0008)**	**0.083*** **(0.0028+-0.0005)**	**0.040*** **(0.037+-0.002)**
Nyungwe	Moz	**0.045*** **(0.019+-0.0015)**	**0.052*** **(0.031+-0.0021)**	**0.12*** **(0.0018+-0.0004)**	**0.13*** **(0.00059+-0.0002)**	**0.053*** **(0.023+-0.0015)**	0.025 (0.054+-0.002)	0.024 (0.094+-0.0029)	**0.062*** **(0.016+-0.0013)**	**0.071*** **(0.0043+-0.0006)**	0.028 (0.066+-0.0024)
Ronga	Moz	-0.01 (0.69+-0.0047)	-0.011 (0.62+-0.0048)	0.014 (0.25+-0.004)	0.022 (0.14+-0.0034)	-0.015 (0.75+-0.005)	-0.00027 (0.40+-0.0047)	-0.014 (0.78+-0.0039)	0.01 (0.24+-0.0042)	0.016 (0.14+-0.0039)	-0.0071 (0.60+-0.0043)
Sena	Moz	**0.075*** **(0.0018+-0.0004)**	**0.081*** **(0.0097+-0.001)**	**0.18*** **(0.00+-0.00)**	**0.17*** **(0.00+-0.00)**	**0.075*** **(0.0059+-0.0008)**	**0.028*** **(0.038+-0.0018)**	0.028 (0.084+-0.0029)	**0.079*** **(0.012+-0.0011)**	**0.070*** **(0.006+-0.0008)**	0.028 (0.066+-0.0026)
Shangaan	Moz	**0.029*** **(0.053+-0.0024)**	**0.031*** **(0.10+-0.0032)**	**0.097*** **(0.0058+-0.0008)**	**0.10*** **(0.00069+-0.0003)**	0.028 (0.090+-0.0028)	0.007 (0.22+-0.004)	0.0048 (0.31+-0.0045)	0.039 (0.057+-0.0024)	**0.041*** **(0.025+-0.0015)**	-0.003 (0.48+-0.005)
Shona	Moz	**0.065*** **(0.0099+-0.0009)**	**0.075*** **(0.018+-0.0013)**	**0.15*** **(0.0014+-0.0004)**	**0.15*** **(0.0002+-0.0001)**	**0.049*** **(0.044+-0.0021)**	0.029 (0.055+-0.0025)	0.021 (0.14+-0.0034)	**0.064*** **(0.022+-0.0014)**	**0.062*** **(0.014+-0.001)**	0.015 (0.17+-0.0034)
Tonga	Moz	**0.048*** **(0.020+-0.0014)**	0.042 (0.069+-0.0024)	**0.079*** **(0.026+-0.0017)**	**0.089*** **(0.0067+-0.0009)**	0.019 (0.18+-0.004)	**0.051*** **(0.0055+-0.0008)**	**0.041*** **(0.044+-0.0021)**	**0.074*** **(0.016+-0.0011)**	**0.068*** **(0.0091+-0.0009)**	0.015 (0.16+-0.0038)

Tswa	Moz	0.028 (0.073+-0.0023)	0.0047 (0.34+-0.0041)	**0.069*** **(0.036+-0.0018)**	**0.065*** **(0.025+-0.0017)**	0.0028 (0.35+-0.0045)		**0.028*** **(0.047+-0.002)**	0.013 (0.20+-0.004)	0.038 (0.070+-0.0026)	0.035 (0.053+-0.0023)	0.003 (0.33+-0.0044)
Yao	Moz	**0.065*** **(0.044+-0.002)**	**0.10*** **(0.022+-0.0012)**	**0.12*** **(0.03+-0.0019)**	**0.14*** **(0.0083+-0.001)**	**0.098*** **(0.016+-0.0012)**		0.022 (0.15+-0.0034)	0.034 (0.12+-0.0031)	0.064 (0.051+-0.0022)	**0.071*** **(0.028+-0.0017)**	0.052 (0.061+-0.0026)
General	Moz	**0.067*** **(0.00+-0.00)**	**0.058*** **(0.0023+-0.0004)**	**0.12*** **(0.0002+-0.0001)**	**0.11*** **(0.00+-0.00)**	**0.033*** **(0.0086+-0.0009)**		**0.057*** **(0.00+-0.00)**	**0.046*** **(0.0042+-0.0006)**	**0.085*** **(0.00069+-0.0003)**	**0.078*** **(0.00+-0.00)**	**0.022*** **(0.019+-0.0012)**
Mixed-Bantu	Bots	**0.050*** **(0.0004+-0.0002)**	0.02 (0.12+-0.0032)	0.015 (0.20+-0.0038)	0.011 (0.18+-0.0041)	**0.044*** **(0.017+-0.0013)**		**0.022*** **(0.0082+-0.001)**	0.0032 (0.30+-0.0044)	0.013 (0.16+-0.0037)	0.017 (0.056+-0.0024)	**0.019*** **(0.046+-0.0019)**
Basotho	Les	x	-0.0064 (0.65+-0.0048)	0.013 (0.17+-0.0043)	**0.019*** **(0.025+-0.0016)**	0.0073 (0.15+-0.0035)		X	-0.002 (0.50+-0.005)	0.006 (0.23+-0.0042)	**0.012*** **(0.037+-0.0021)**	0.0058 (0.13+-0.0037)
Ndebele	Les	-0.0064 (0.65+-0.0048)	x	-0.003 (0.45+-0.0051)	-0.002 (0.43+-0.005)	-0.013 (0.81+-0.0035)		-0.002 (0.50+-0.005)	x	-0.0058 (0.57+-0.0056)	-0.00041 (0.41+-0.0048)	-0.0041 (0.56+-0.0045)
Thembu	Les	0.013 (0.17+-0.0043)	-0.003 (0.45+-0.0051)	x	-0.025 (0.98+-0.0014)	0.016 (0.17+-0.0037)		0.006 (0.23+-0.0042)	-0.0058 (0.57+-0.0056)	x	-0.021 (0.99+-0.0011)	0.015 (0.15+-0.0037)
Xhosa	RSA	**0.019*** **(0.025+-0.0016)**	-0.002 (0.43+-0.005)	-0.025 (0.98+-0.0014)	x	0.016 (0.11+-0.0031)		**0.012*** **(0.037+-0.0021)**	-0.00041 (0.41+-0.0048)	-0.021 (0.99+-0.0011)	x	0.016 (0.067+-0.0026)
Zulu	RSA	0.0073 (0.15+-0.0035)	-0.013 (0.81+-0.0035)	0.016 (0.17+-0.0037)	0.016 (0.11+-0.0031)	x		0.0058 (0.13+-0.0037)	-0.0041 (0.56+-0.0045)	0.015 (0.15+-0.0037)	0.016 (0.067+-0.0026)	x

Table 4.4: Population pairwise F_{ST} values for mtDNA haplogroup frequencies and haplotypes, comparing the Lesotho, Xhosa and Zulu populations with other Southern African populations from previous studies.

***** = significant at the 0.05 significance level.

[a] See Supplementary Table 1, Appendix 2 for sources of populations used.

[b] Country: Bots = Botswana; Les = Lesotho; Moz = Mozambique; RSA = South Africa; Zam = Zambia; Zim = Zimbabwe.

Comparisons with other Southern African populations

Population pairwise distances (F_{ST}) based on mtDNA haplotypes and haplogroup frequencies show that the composition of the three Lesotho populations and the Xhosa and Zulu samples is unlike that of a large number of the previously published Bantu-speaking populations on the African continent (Table 4.4). Using a dataset with 21 other Bantu-speaking Southern African populations for comparison (Supplementary Table 1, Appendix 2), when analysing haplogroup frequency F_{ST}, the Basotho population is different from 17/21 of the other populations, Ndebele 15/21, Thembu 19/21, Xhosa 19/21 and Zulu 16/21. When analyses are conducted using haplotypes, again they differ from a large number of the other populations, but less than for haplogroup frequencies (Basotho: 14/21; Ndebele 8/21; Thembu 15/21; Xhosa 17/21; and Zulu 10/21) (Table 4.4).

Despite high frequencies of the Bantu NRY haplogroups, F_{ST} values from analyses of Hg frequencies show the three Lesotho populations, and the Xhosa and Zulu populations to be significantly different from some of the comparison populations used from different regions, possibly due to less non Sub-Saharan African Hgs being present (Basotho: 5/7; Ndebele 2/7; Thembu 1/7; Xhosa 2/7; and Zulu 2/7) (Table 4.5). Less NRY comparisons are significantly different than for the mtDNA haplogroup analyses, probably as a result of a lower resolution of haplogroups being used, and the majority of samples for all the populations being grouped in a large E haplogroup, not representing fully the diversity present. When NRY haplotypes were analysed virtually all comparisons were significant (Basotho: 14/16; Ndebele 14/16; Thembu 16/16; Xhosa 16/16; and Zulu 15/16) (Table 4.6).

[a]Population	[b]Country	NRY Haplogroup F_{ST}				
		Basotho	Ndebele	Thembu	Xhosa	Zulu
Bisa	Zam	**0.076*** **(0.0049+-0.0007)**	**0.043*** **(0.047+-0.0022)**	0.00071 (0.56+-0.0051)	**0.037*** **(0.047+-0.0019)**	0.0078 (0.31+-0.0047)
Kunda	Zam	**0.058*** **(0.010+-0.001)**	0.022 (0.13+-0.0033)	-0.021 (0.82+-0.0033)	0.013 (0.20+-0.0039)	-0.0077 (0.64+-0.0042)
Shona	Zim	**0.031*** **(0.034+-0.0017)**	-0.010 (0.56+-0.0051)	-0.017 (0.65+-0.0048)	-0.0025 (0.40+-0.005)	-0.012 (0.59+-0.0050)
Mixed-Bantu	Bots	**0.23*** **(0.00+-0.00)**	**0.29*** **(0.00+-0.00)**	**0.34*** **(0.0002+-0.0001)**	**0.31*** **(0.00+-0.00)**	**0.38*** **(0.00+-0.00)**
Sotho-Tswana	RSA	-0.0027 (0.39+-0.0043)	0.024 (0.14+-0.0033)	0.062 (0.072+-0.0024)	0.032 (0.090+-0.0033)	**0.088*** **(0.013+-0.0011)**
Xhosa	RSA	**0.020*** **(0.032+-0.002)**	-0.0077 (0.65+-0.0041)	-0.020 (0.83+-0.0036)	-0.0055 (0.55+-0.0048)	-0.0048 (0.51+-0.0056)
Zulu	RSA	-0.0081 (0.58+-0.0042)	-0.0083 (0.48+-0.0051)	0.018 (0.22+-0.004)	0.00054 (0.36+-0.0053)	0.034 (0.10+-0.0029)
Basotho	Les	x	0.0064 (0.20+-0.0045)	0.021 (0.14+-0.0036)	0.013 (0.096+-0.0031)	**0.036*** **(0.015+-0.0012)**
Ndebele	Les	0.0064 (0.20+-0.0045)	x	-0.016 (0.77+-0.0043)	-0.015 (0.96+-0.002)	-0.0061 (0.48+-0.0050)
Thembu	Les	0.021 (0.14+-0.0036)	-0.016 (0.77+-0.0043)	x	-0.015 (0.76+-0.0043)	-0.029 (0.99+-0.00)
Xhosa Cape Town	RSA	0.013 (0.096+-0.0031)	-0.015 (0.96+-0.0020)	-0.015 (0.76+-0.0043)	x	-0.0061 (0.55+-0.0048)
Zulu Cape Town	RSA	**0.036*** **(0.015+-0.0012)**	-0.0061 (0.48+-0.0050)	-0.029 (0.99+-0.00)	-0.0061 (0.55+-0.0048)	x

Table 4.5: Population pairwise F_{ST} values for NRY haplogroup frequencies, comparing the Lesotho, Xhosa and Zulu populations with other Southern African populations from previous studies.

***** = significant at the 0.05 significance level.

[a] See Supplementary Table 1, Appendix 2 for sources of populations used.

[b] Country: Bots = Botswana; Les = Lesotho; RSA = South Africa; Zam = Zambia; Zim = Zimbabwe.

[a]Population	[b]Country	NRY Haplotype F_{ST}				
		Basotho	Ndebele	Thembu	Xhosa	Zulu
Bisa	Zam	**0.050*** (0.00+-0.00)	**0.076*** (0.00+-0.00)	**0.13*** (0.00+-0.00)	**0.069*** (0.00+-0.00)	**0.046*** (0.0049+-0.002)
Kunda	Zam	**0.035*** (0.00098+-0.001)	**0.046*** (0.002+-0.0014)	**0.088*** (0.00098+-0.001)	**0.049*** (0.00+-0.00)	**0.032*** (0.0068+-0.0023)
Fwe	Zam	**0.10*** (0.00+-0.00)	**0.12*** (0.00+-0.00)	**0.21*** (0.00+-0.00)	**0.10*** (0.00+-0.00)	**0.086*** (0.00+-0.00)
Kwamashi	Zam	**0.090*** (0.00+-0.00)	**0.11*** (0.00+-0.00)	**0.21*** (0.00+-0.00)	**0.11*** (0.00+-0.00)	**0.084*** (0.00+-0.00)
Lozi	Zam	**0.054*** (0.00+-0.00)	**0.062*** (0.00+-0.00)	**0.13*** (0.00+-0.00)	**0.054*** (0.00+-0.00)	**0.038*** (0.00+-0.00)
Luvale	Zam	**0.099*** (0.00+-0.00)	**0.081*** (0.0039+-0.0019)	**0.28*** (0.00+-0.00)	**0.077*** (0.00098+-0.001)	**0.065*** (0.014+-0.0037)
Luyana	Zam	**0.068*** (0.00+-0.00)	**0.081*** (0.00+-0.00)	**0.17*** (0.00+-0.00)	**0.081*** (0.00+-0.00)	**0.062*** (0.00+-0.00)
Mbunda	Zam	**0.090*** (0.00+-0.00)	**0.089*** (0.00+-0.00)	**0.19*** (0.00+-0.00)	**0.096*** (0.00+-0.00)	**0.083*** (0.00+-0.00)
Nkoya	Zam	**0.086*** (0.00+-0.00)	**0.098*** (0.00+-0.00)	**0.18*** (0.00+-0.00)	**0.068*** (0.00098+-0.001)	**0.063*** (0.00+-0.00)
Shanjo	Zam	**0.058*** (0.00098+-0.001)	**0.089*** (0.00+-0.00)	**0.15*** (0.00+-0.00)	**0.086*** (0.00+-0.00)	**0.064*** (0.00+-0.00)
Tonga	Zam	**0.053*** (0.00+-0.00)	**0.057*** (0.00+-0.00)	**0.13*** (0.00+-0.00)	**0.058*** (0.00098+-0.001)	**0.046*** (0.0029+-0.0016)
Totela	Zam	0.012 (0.18+-0.012)	0.033 (0.066+-0.0097)	**0.079*** (0.012+-0.0036)	**0.040*** (0.031+-0.0076)	0.024 (0.095+-0.01)
Mixed Bantu West	Zam	**0.043*** (0.00+-0.00)	**0.057*** (0.00+-0.00)	**0.11*** (0.00+-0.00)	**0.048*** (0.00+-0.00)	**0.031*** (0.00098+-0.001)
Mixed Bantu	Bots	**0.12*** (0.00+-0.00)	**0.081*** (0.00+-0.00)	**0.17*** (0.00+-0.00)	**0.11*** (0.00+-0.00)	**0.10*** (0.00+-0.00)
Kalanga	Bots	0.016 (0.095+-0.011)	**0.067*** (0.002+-0.0014)	**0.078*** (0.0078+-0.0031)	**0.057*** (0.002+-0.0014)	**0.039*** (0.02+-0.0041)
Tswana	Bots	**0.053*** (0.00098+-0.001)	0.029 (0.051+-0.0074)	**0.11*** (0.00+-0.00)	**0.039*** (0.014+-0.0039)	**0.032*** (0.046+-0.0052)
Xhosa Cape Town	RSA	**0.039*** (0.00+-0.00)	**0.018*** (0.027+-0.006)	**0.042*** (0.018+-0.0056)	x	-0.0088 (0.90+-0.0096)
Zulu Cape Town	RSA	**0.025*** (0.00098+-0.001)	0.013 (0.097+-0.0075)	**0.056*** (0.0078+-0.0028)	-0.0088 (0.90+-0.0096)	x
Basotho	Les	x	**0.027*** (0.002+-0.0014)	**0.057*** (0.00098+-0.001)	**0.039*** (0.00+-0.00)	**0.025*** (0.00098+-0.001)
Ndebele	Les	**0.027*** (0.0020+-0.0014)	x	**0.061*** (0.002+-0.0014)	**0.018*** (0.027+-0.006)	0.013 (0.097+-0.0075)
Thembu	Les	**0.057*** (0.00098+-0.001)	**0.061*** (0.002+-0.0014)	x	**0.042*** (0.018+-0.0056)	**0.056*** (0.0078+-0.0028)

Table 4.6: Populations pairwise F_{ST} values for NRY haplotypes, comparing the Lesotho, Xhosa and Zulu populations with other Southern African populations from previous studies.

***** = significant at the 0.05 significance level.

[a] See Supplementary Table 1, Appendix 2 for sources of populations used. [b] Country: Bots = Botswana; Les = Lesotho; RSA = South Africa; Zam = Zambia; Zim = Zimbabwe

Wilcoxon tests

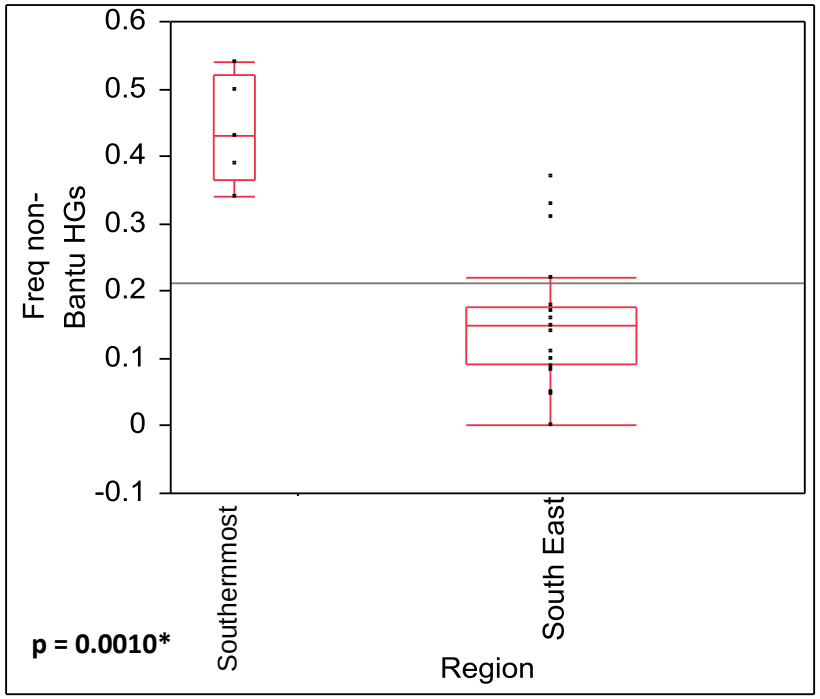
According to the frontier hypothesis, southernmost African populations beyond and around the static frontier should contain more hunter-gatherer haplogroups than those in surrounding regions as a result of increased interaction between the hunter-gatherers and Bantu-speakers on the frontier (Alexander 1984; Mitchell 2009). In order to test whether the observations of increased mtDNA hunter-gatherer Hgs in the Lesotho, Xhosa and Zulu populations are significant, the Lesotho and South African (Xhosa and Zulu) populations were grouped together as Southernmost Africa and compared to populations from the neighbouring region of South East Africa (Mozambique, Zambia and Zimbabwe) (Pereira et al. 2001; Salas et al. 2002; Castri et al. 2009; de Filippo et al. 2010).

A conservative approach was used to determine hunter-gatherer related haplogroups to be included in the analysis (see Chapter 2, Materials and Methods), as it is not certain what the hunter-gatherer haplogroups present in the South-East region were before the arrival of the Bantu-speakers. Taking this into account, all the non-Bantu specific lineages present in the analysed populations were considered as being potentially introgressed local types (see Table 4.1 for the haplogroups excluded). This most probably represents an overestimate of the frequency of hunter-gatherer types present in the regions, however, this seemed the most appropriate way of avoiding underestimating the levels of introgression in the South-East populations. Non-African types, identified as belonging to European-associated haplogroups I, R1a and R1b1b2, were also removed from the analysis. Using this method, Wilcoxon tests indicate that the level of hunter-gatherer maternal introgression in Southernmost Africa is significantly higher than South-East Africa ($Z = 3.30$, $p = 0.0010$) (Figure 4.4a).

The same comparisons were also conducted using NRY data. Again, all chromosomes not belonging to Bantu-specific lineages were considered (Hg E and B2a) as being potentially hunter-gatherer introgressed (the few known Non-Africa types were also removed from the analysis). A Wilcoxon test indicates that the level of non-Bantu Hgs in Southernmost Africa is not significantly higher than that in South-East Africa ($Z = 0.75$, $p = 0.45$) (Figure 4.4b). These comparisons include the data from de Filippo et al. (2011) which does not differentiate between Bantu-associated B2a and 'Khoisan'-associated B2b. In order to allow for the possibility that some of the B* listed in this paper was B2b, all B types were included as being possibly Non-Bantu. Even if these populations are removed, however, NRY comparisons are still not significant (Southernmost Africa vs. South East: $Z = -1.53$, $p = 0.13$).

The differences between the mtDNA and NRY results within the regions is demonstrated by further Wilcoxon tests: the mtDNA introgression of non-Bantu Hgs is significantly higher than that for the NRY (Southernmost Africa: $Z = 2.93$, $p = 0.0034$; South East: $Z = -3.96$, $p = < 0.0001$) (Figure 4.5). When carrying out this analysis solely on Southernmost Africa it was possible to focus on the identified hunter-gatherer types for this area (L0d/L0k and A3b1*/B2b), producing a more accurate result for comparison of the introgression between mtDNA and the NRY in this region. Results using just these Hgs for Southernmost Africa were in agreement with the previous comparison using all non-Bantu types: mtDNA hunter-gatherer types are present at higher frequencies than Y chromosome hunter-gatherer types in Southernmost Bantu-speaking populations (Basotho: $\chi^2 = 27.03$, $p < 0.0001$; Ndebele: $\chi^2 = 6.23$, $p = 0.013$; Thembu: $\chi^2 = 7.36$, $p = 0.0067$; Xhosa: $\chi^2 = 20.17$, $p < 0.0001$; Zulu: $\chi^2 = 13.24$, $p = 0.0003$) (Figure 4.6).

a. mtDNA



b. NRY

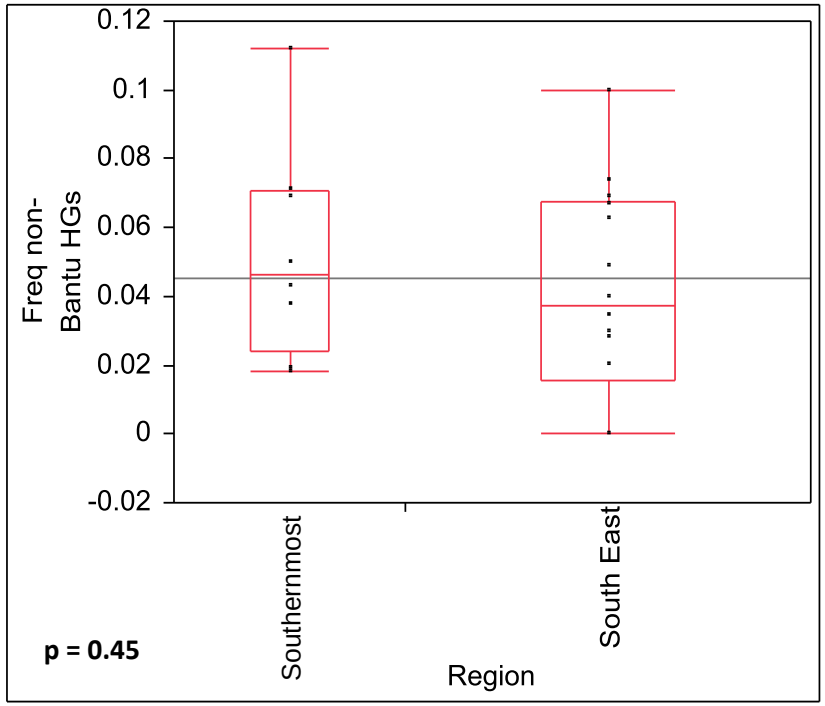
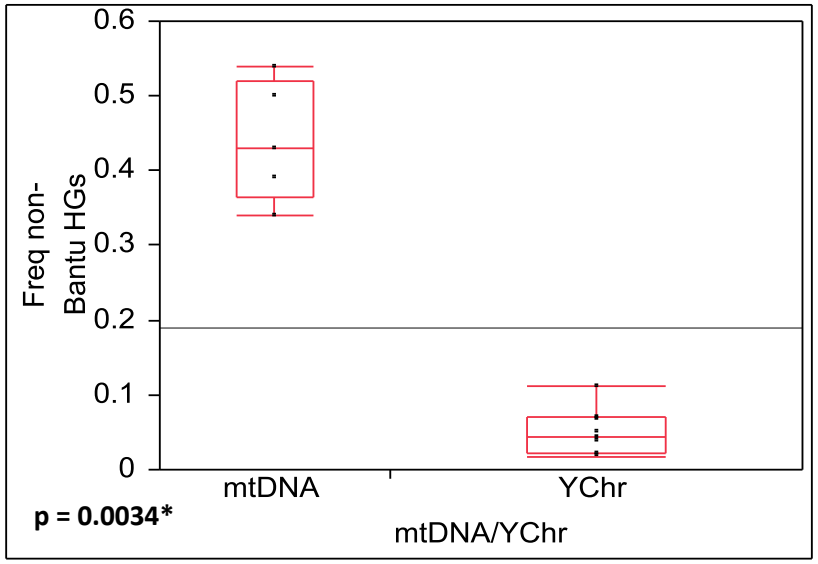


Figure 4.4: Wilcoxon tests between regions.
a – mtDNA, Southernmost vs. South East; b – NRY, Southernmost vs. South East.

a. Southernmost



b. South East

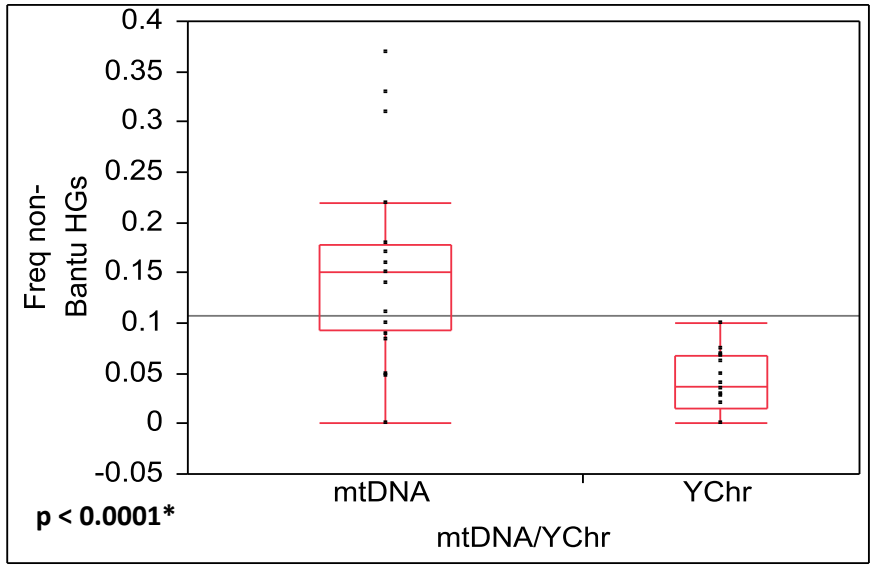


Figure 4.5: Wilcoxon tests within regions for mtDNA vs. Y chromosome.
a – Southernmost; b – South East.

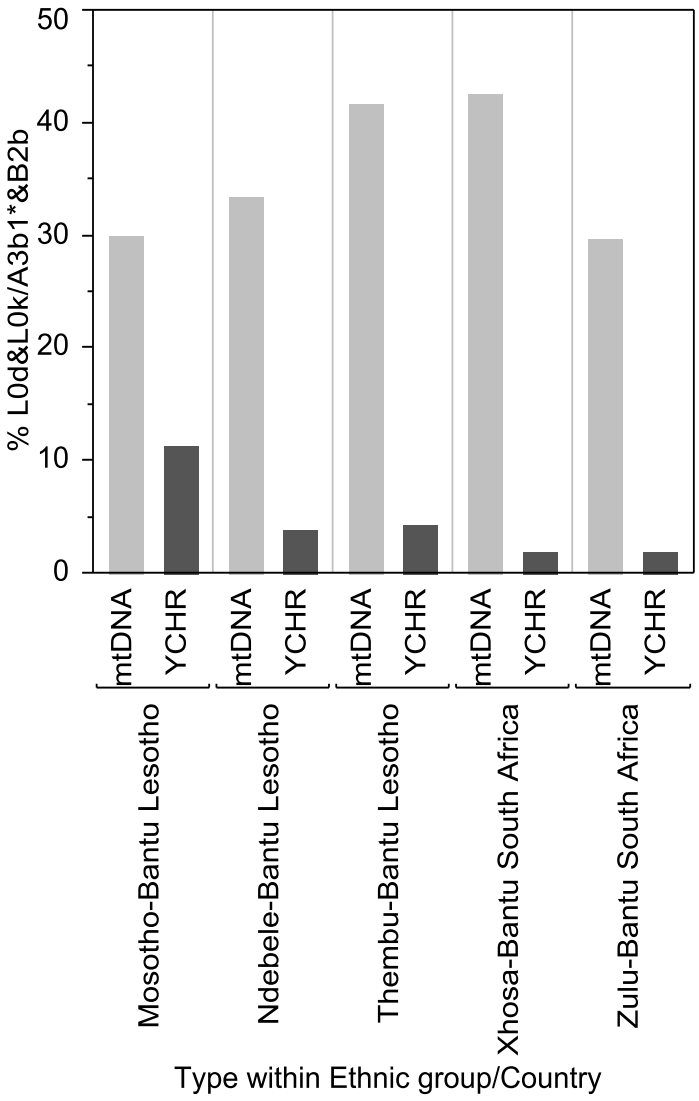


Figure 4.6: L0d/L0k (mtDNA) and A3b1/B2b (NRY) haplogroup frequencies in Southernmost African populations.

Summary statistics

Seeing as the mtDNA results are particularly interesting, various summary statistics for the mtDNA were calculated for the Lesotho, Xhosa and Zulu samples, and for the populations in the South-East region, as a comparison with populations on the other side of the frontier (Table 4.7). These summary statistics were used in order to try and determine whether any patterns emerge to help

explain the differences in frequency of non-Bantu haplogroups across the different regions/language groupings.

No populations produced a significant result for the Tajima's D test, although the majority of the populations did have a negative D value, indicating possible evidence of population expansion. All South East Bantu-click populations produced negative values, apart from LOW-3-20. Some of the Mozambique populations produced positive values, indicating a possible difference in population history, but none of the results are significant. The R_2 test also resulted in no significant results.

The Fu's F_S test, on the other hand, did produce significantly negative values for some of the populations, indicating evidence of population expansion. All of the South East Bantu-click populations produce significant negative results when the full number of samples are included, but when the sample size is reduced some of the separate sampling locations do not. A number of the South East populations are also significant, however, so again not showing a difference between the regions. Although the Chakraborty test produced a significant result for the Basotho (all), the only other significant population was the Ndau (Mozambique).

Mismatch distribution

As can be seen from the graphs and Harpending raggedness indexes (Figure 4.7), both the Southernmost and South East Bantu-speaking populations display a bell shaped histogram, with non-significant Sum of Squared Deviation (SSD) values, indicating that the null hypothesis of expansion is met (Southernmost: SSD = 0.0037, p = 0.14, H_r = 0.0089, p = 0.11; South East: SSD = 0.0029, p = 0.54, H_r = 0.0045, p = 0.76). The Southernmost 'Khoisan' populations are different, however, with a significant SSD value (SSD = 0.047, $p < 0.0001$) suggesting that the null hypothesis

of expansion is not met. When specific haplogroups are considered separately, the hunter-gatherer L0d Hg produces a bell shaped histogram for the Bantu-speaking populations of Southernmost Africa, with a non-significant SSD value indicating the null hypothesis of expansion is met (SSD = 0.0024, p = 0.65; H_r = 0.010, p = 0.75) (Figure 4.8). The Bantu-associated L3e Hg produces a slightly more ragged distribution for the Southernmost Bantu populations (H_r = 0.085, p 0.057), however, the SSD value is still not significant (SSD = 0.030, p = 0.089), suggesting that the population might have undergone an expansion (Figure 4.9). Interestingly, another Bantu-associated Hg, L2a1, does seem to produce a smooth bell-shaped distribution (SSD = 0.0074, p = 0.78; H_r = 0.018, p = 0.88) (Figure 4.10), suggesting that the raggedness in the L3e distribution might be as a result of more recent gene flow. South-East Africa produces smooth distributions for L0d and L3e, and a nearly smooth distribution for L2a1, with all SSD values not significantly different from an expansion hypothesis. When the same haplogroups were analysed using data from 'Khoisan' populations, ragged distributions were observed, although because of the small population sizes of these samples, the results should be viewed with caution.

Population[a]	Country[b]	Region[c]	Language Group[d]	Individuals (htypes)	Gene diversity (sd)	Htypes/N	Mean number of Pairwise differences	Nucleotide diversity	Tajima's D[e]	p value	Fu's F_S	P value[f]	R_2	P value	Chakraborty's Test		
															Observed homozygosity	Expected number of alleles	p (k or more alleles)
Bisa	Zam	SE	SE Bantu	42 (27)	0.96 (0.016)	0.64	9.79 (4.57)	0.028 (0.015)	-0.50	0.33	**-7.72**	**0.02**	0.09	0.32	0.04	16.72	0.31
Kunda	Zam	SE	SE Bantu	36 (25)	0.97 (0.015)	0.69	9.11 (4.29)	0.026 (0.014)	-0.65	0.30	**-8.31**	**0.01**	0.09	0.25	0.03	14.98	0.50
Shona	Zim	SE	SE Bantu	59 (39)	0.98 (0.0072)	0.66	9.16 (4.27)	0.027 (0.014)	-1.08	0.14	**-19.25**	**0.00**	0.07	0.11	0.02	18.83	0.52
Chopi	Moz	SE	SE Bantu	27 (18)	0.95 (0.025)	0.67	8.06 (3.86)	0.023 (0.013)	-0.42	0.38	-4.31	0.07	0.10	0.32	0.05	12.24	0.45
Chwabo	Moz	SE	SE Bantu	20 (15)	0.94 (0.043)	0.75	8.30 (4.01)	0.024 (0.013)	-0.31	0.40	-3.63	0.06	0.12	0.33	0.06	10.54	0.22
Lomwe	Moz	SE	SE Bantu	20 (12)	0.88 (0.065)	0.60	8.65 (4.17)	0.025 (0.014)	0.04	0.56	-0.62	0.38	0.13	0.49	0.12	10.72	0.13
Makhuwa	Moz	SE	SE Bantu	20 (12)	0.91 (0.053)	0.60	9.83 (4.70)	0.029 (0.015)	-0.01	0.54	-0.19	0.49	0.13	0.49	0.09	11.26	0.31
Makonde	Moz	SE	SE Bantu	19 (13)	0.95 (0.031)	0.68	7.10 (3.49)	0.021 (0.011)	-0.14	0.48	-2.54	0.12	0.13	0.45	0.05	9.62	0.69
Ndau	Moz	SE	SE Bantu	19 (15)	0.96 (0.036)	0.79	9.30 (4.47)	0.027 (0.015)	0.06	0.57	-3.62	0.06	0.13	0.51	0.04	23.27	**0.02**
Nguni	Moz	SE	SE Bantu	11 (11)	1 (0.039)	1.00	8.27 (4.16)	0.024 (0.014)	0.24	0.62	**-4.86**	**0.01**	0.15	0.40	0.04	10.69	0.38
Nyanja	Moz	SE	SE Bantu	20 (13)	0.94 (0.034)	0.65	7.79 (3.79)	0.023 (0.012)	0.25	0.67	-1.83	0.22	0.14	0.61	0.00	7.29	N.A.
Nyungwe	Moz	SE	SE Bantu	20 (16)	0.97 (0.025)	0.80	9.45 (4.53)	0.027 (0.015)	-0.31	0.43	-4.24	0.05	0.11	0.36	0.06	10.28	0.59
Ronga	Moz	SE	SE Bantu	21 (19)	0.99 (0.018)	0.90	9.26 (4.43)	0.027 (0.014)	-0.66	0.29	**-8.66**	**0.00**	0.10	0.17	0.03	11.09	0.64
Sena	Moz	SE	SE Bantu	21 (11)	0.89 (0.049)	0.52	7.58 (3.68)	0.022 (0.012)	0.52	0.75	-0.06	0.52	0.15	0.72	0.01	11.32	0.73
Shangaan	Moz	SE	SE Bantu	22 (17)	0.96 (0.030)	0.77	9.23 (4.41)	0.027 (0.014)	-0.39	0.37	-4.52	0.05	0.11	0.29	0.11	10.44	0.41
Shona	Moz	SE	SE Bantu	17 (15)	0.99 (0.025)	0.88	9.35 (4.52)	0.027 (0.015)	-0.55	0.30	**-5.04**	**0.02**	0.11	0.16	0.04	11.61	0.35
Tonga	Moz	SE	SE Bantu	20 (14)	0.95 (0.034)	0.70	8.79 (4.24)	0.026 (0.014)	-0.13	0.50	-2.29	0.16	0.12	0.41	0.01	10.02	0.74
Tswa	Moz	SE	SE Bantu	19 (16)	0.98 (0.027)	0.84	7.57 (3.70)	0.022 (0.012)	-0.22	0.46	**-6.24**	**0.01**	0.13	0.45	0.05	10.79	0.48
Yao	Moz	SE	SE Bantu	10 (8)	0.93 (0.077)	0.80	7.80 (3.97)	0.023 (0.013)	0.01	0.54	-0.87	0.29	0.14	0.29	0.02	9.87	0.57
General population	Moz	SE	SE Bantu	109 (51)	0.96 (0.0081)	0.47	8.78 (4.08)	0.026 (0.013)	-0.84	0.21	**-23.26**	**0.00**	0.07	0.18	0.07	6.73	0.54
Basotho (all)	Les	Sm	SE Bantu-Click	287 (103)	0.97 (0.0039)	0.36	9.14 (4.22)	0.027 (0.014)	-0.87	0.18	**-24.15**	**0.00**	0.06	0.19	0.03	32.29	**0.00**
Basotho HIGH-1 all	Les	Sm	SE Bantu-Click	23 (17)	0.97 (0.022)	0.74	9.16 (4.37)	0.027 (0.014)	-0.33	0.38	-4.05	0.06	0.11	0.32	0.03	11.87	0.64
Basotho HIGH-2 all	Les	Sm	SE Bantu-Click	27 (19)	0.97 (0.021)	0.70	9.23 (4.38)	0.027 (0.014)	-0.79	0.23	-4.56	0.05	0.09	0.18	0.03	13.00	0.57
Basotho HIGH-3 all	Les	Sm	SE Bantu-Click	17 (14)	0.98 (0.027)	0.82	9.12 (4.41)	0.027 (0.014)	-0.38	0.40	-3.63	0.07	0.11	0.24	0.02	9.93	0.76
Basotho LOW-1 all	Les	Sm	SE Bantu-Click	37 (26)	0.97 (0.014)	0.70	9.29 (4.37)	0.027 (0.014)	-0.58	0.32	**-9.02**	**0.01**	0.09	0.30	0.03	15.33	0.48
Basotho LOW-2 all	Les	Sm	SE Bantu-Click	36 (23)	0.97 (0.014)	0.64	8.50 (4.02)	0.025 (0.013)	-0.61	0.31	**-6.43**	**0.02**	0.09	0.24	0.03	14.49	0.67
Basotho LOW-3 all	Les	Sm	SE Bantu-Click	48 (29)	0.96 (0.016)	0.60	9.08 (4.25)	0.026 (0.014)	-0.29	0.44	**-9.00**	**0.02**	0.10	0.41	0.04	17.12	0.27

Basotho HIGH-1-20*	Les	Sm	SE Bantu-Click	20 (14)	0.96 (0.028)	0.70	9.06 (4.35)	0.026 (0.014)	-0.53	0.34	-2.17	0.17	0.10	0.20	0.04	14.31	0.67
Basotho HIGH-2-20	Les	Sm	SE Bantu-Click	20 (15)	0.96 (0.028)	0.75	8.82 (4.25)	0.026 (0.014)	-0.86	0.20	-3.34	0.09	0.10	0.13	0.04	14.84	0.58
Basotho HIGH-3-17	Les	Sm	SE Bantu-Click	17 (14)	0.98 (0.027)	0.82	9.12 (4.41)	0.027 (0.014)	-0.38	0.41	-3.63	0.06	0.11	0.24	0.02	14.44	0.76
Basotho LOW-1-20	Les	Sm	SE Bantu-Click	20 (16)	0.97 (0.025)	0.80	7.96 (3.86)	0.023 (0.013)	-0.28	0.43	**-5.20**	**0.02**	0.12	0.35	0.03	16.02	0.64
Basotho LOW-2-20	Les	Sm	SE Bantu-Click	20 (14)	0.96 (0.026)	0.70	8.51 (4.11)	0.025 (0.013)	-0.27	0.44	-2.43	0.14	0.12	0.39	0.04	14.84	0.77
Basotho LOW-3-20	Les	Sm	SE Bantu-Click	20 (17)	0.98 (0.025)	0.85	9.17 (4.04)	0.027 (0.014)	0.08	0.56	**-5.91**	**0.02**	0.13	0.52	0.02	16.68	0.57
Ndebele	Les	Sm	SE Bantu-Click	33 (25)	0.98 (0.011)	0.76	9.80 (4.60)	0.028 (0.015)	-0.57	0.32	**-9.10**	**0.01**	0.10	0.31	0.02	14.84	0.77
Thembu	Les	Sm	SE Bantu-Click	24 (20)	0.98 (0.018)	0.83	8.45 (4.05)	0.025 (0.013)	-0.50	0.33	**-8.43**	**0.00**	0.10	0.23	0.02	11.75	0.61
Xhosa	RSA	Sm	SE Bantu-Click	54 (37)	0.97 (0.012)	0.69	8.20 (3.86)	0.024 (0.012)	-0.59	0.33	**-20.40**	**0.00**	0.09	0.32	0.03	17.06	0.09
Zulu	RSA	Sm	SE Bantu-Click	54 (32)	0.97 (0.0096)	0.59	9.13 (4.27)	0.027 (0.014)	-0.30	0.42	**-10.80**	**0.01**	0.09	0.40	0.03	18.10	0.53

Table 4.7: Summary statistics for mtDNA for the SE Bantu-Click (Southernmost Africa and SE Bantu (South-East) populations.

* HIGH-1-20 indicates sub-sample size

[a] See Supplementary Table 1, Appendix 2 for sources of populations used.

[b] Country: Les = Lesotho; Moz = Mozambique; RSA = South Africa; Zam = Zambia; Zim = Zimbabwe.

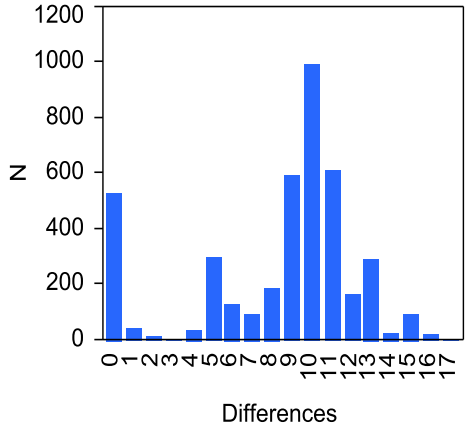
[c] Region: Sm = Southernmost; SE = South-East.

[d] Language group: populations grouped according to language spoken

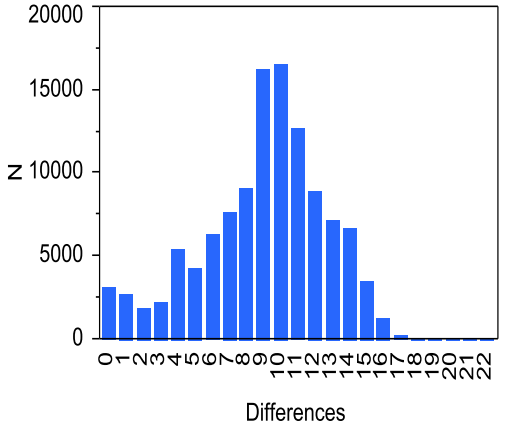
[e] Tajima's D: D > 0 indicates a population bottleneck; D < 0 indicates an expansion.

[f] Fu's F_S: significance level for this test= 0.02; negative values are evidence of a recent expansion.

Southernmost 'Khoisan' all haplogroups
(SSD = 0.047, p < 0.0001, H_r = 0.067, p<0.0001)



Southernmost Bantu all haplogroups
(SSD = 0.0037, p = 0.14, H_r = 0.0089, p = 0.11)



South-East Bantu all haplogroups
(SSD = 0.0029, p = 0.54, H_r = 0.0045, p = 0.76)

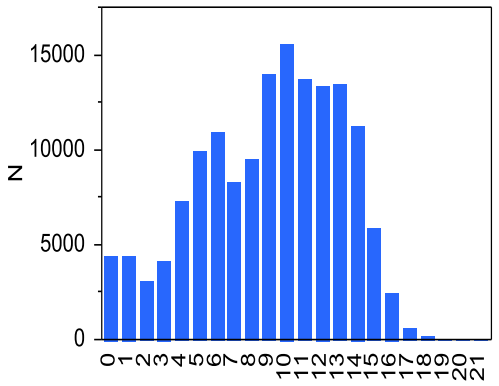
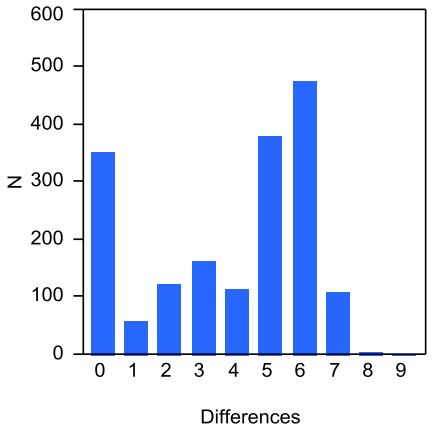
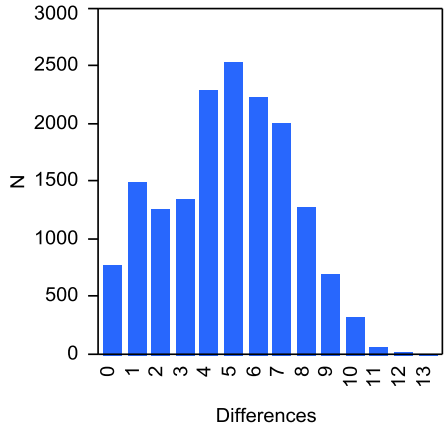


Figure 4.7: Mismatch distributions for all mtDNA haplogroups for population groups.
SSD = Sum of squared deviation; H_r = Harpending raggedness index. H_r values of < 0.03 indicate smooth, bell shaped histograms.

Southernmost 'Khoisan' L0d
(SSD = 0.056, p = 0.015, H_r = 0.10, p = 0.020)



Southernmost Bantu L0d
(SSD = 0.0024, p = 0.65, H_r = 0.010, p = 0.75)



South-East Bantu L0d
(SSD = 0.0032, p = 0.65, H_r = 0.026, p = 0.59)

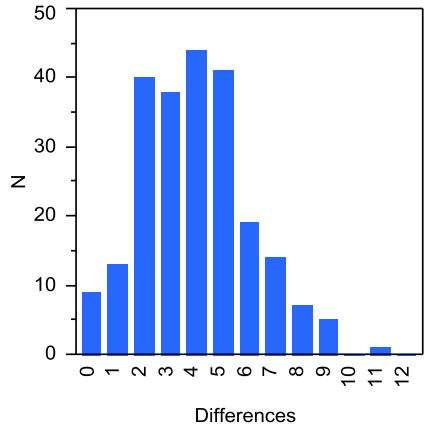
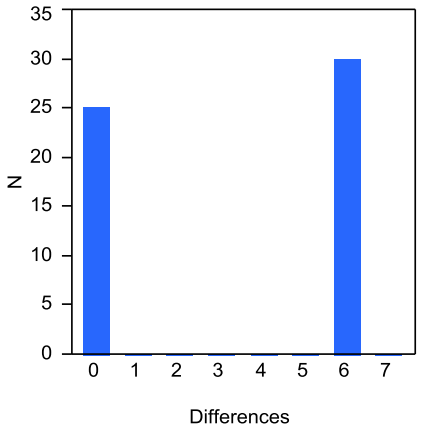
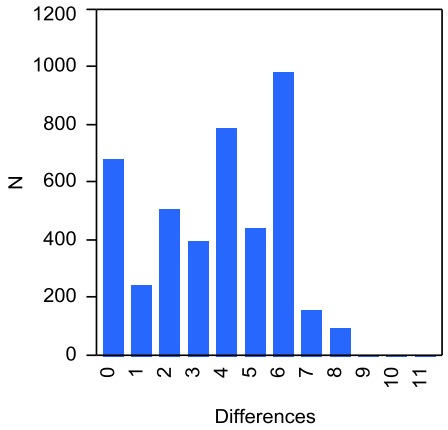


Figure 4.8: Mismatch distributions for the mtDNA haplogroup L0d for population groups.
SSD = Sum of squared deviation; H_r = Harpending raggedness index. H_r values of < 0.03 indicate smooth, bell shaped histograms.

Southernmost 'Khoisan' L3e
(SSD = 0.36, p = 0.053, H_r = 0.80, p = 0.001)



Southernmost Bantu L3e
(SSD = 0.030, p = 0.089, H_r = 0.085, p = 0.057)



South-East Bantu L3e
(SSD = 0.0025, p = 0.40, H_r = 0.021, p = 0.56)

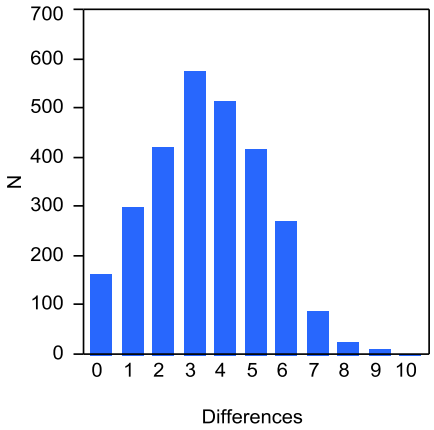
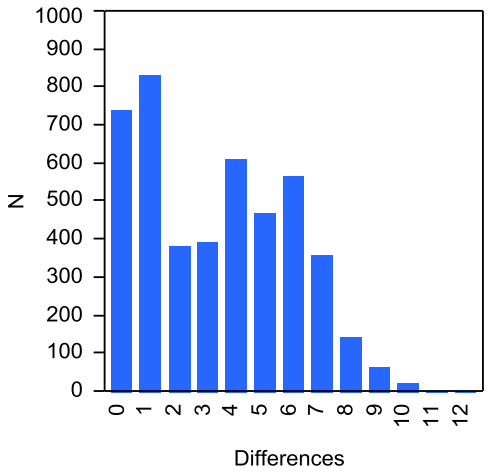


Figure 4.9: Mismatch distributions for the mtDNA haplogroup L3e for population groups.
SSD = Sum of squared deviation; H_r = Harpending raggedness index. H_r values of < 0.03 indicate smooth, bell shaped histograms.

Southernmost Bantu L2a1
(SSD = 0.0074, p = 0.78, H_r = 0.018, p = 0.88)



South-East Bantu L2a1
(SSD = 0.028, p = 0.20, H_r = 0.046, p = 0.33)

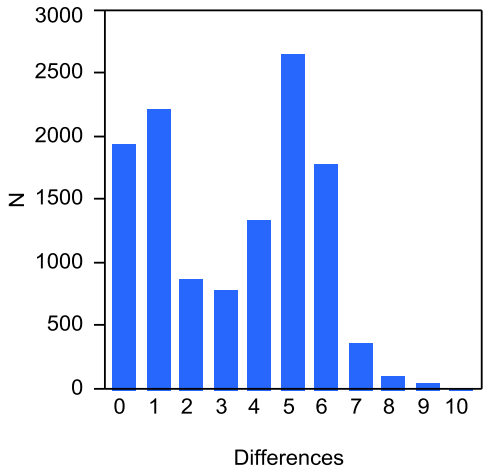


Figure 4.10: Mismatch distributions for the mtDNA haplogroup L2a1 for population groups.
SSD = Sum of squared deviation; H_r = Harpending raggedness index. H_r values of < 0.03 indicate smooth, bell shaped histograms.

Networks

In order to improve understanding of the events that played a part in creating the genetic variation seen in Southernmost Africa today, haplotypes for the mtDNA Hgs L0d/L0k (hunter-gatherer) (Figure 4.11) and L3e (Bantu) (Figure 4.12) were compared. Analysis of the mtDNA Hgs shows that a small number of haplotypes seem to be of high frequency across the whole of Southernmost Africa, being present in the Botswana, Lesotho, Xhosa and Zulu populations. This is the case for both the hunter-gatherer and Bantu-associated Hgs. Of the 69 L0d haplotypes present within the Southernmost African samples, 15 of these are shared among more than 1 Bantu population from the region (21.7%), and of the 17 L3e haplotypes present, 7 of these are shared among more than 1 Southernmost African Bantu population (41.2%). These shared haplotypes are, more often than not, the haplotypes with the highest frequency within the region.

The NRY E haplotype also shows a similar pattern (Figure 4.13), with haplotype sharing across Botswana, Lesotho, Xhosa and Zulu populations. 27 of the E haplotypes present (105) are shared. A smaller number of samples are available for analysis for A3b1* haplotypes, but the networks still demonstrate sharing across widely dispersed populations, with Lesotho and Botswana samples sharing haplotypes (Figure 4.14).

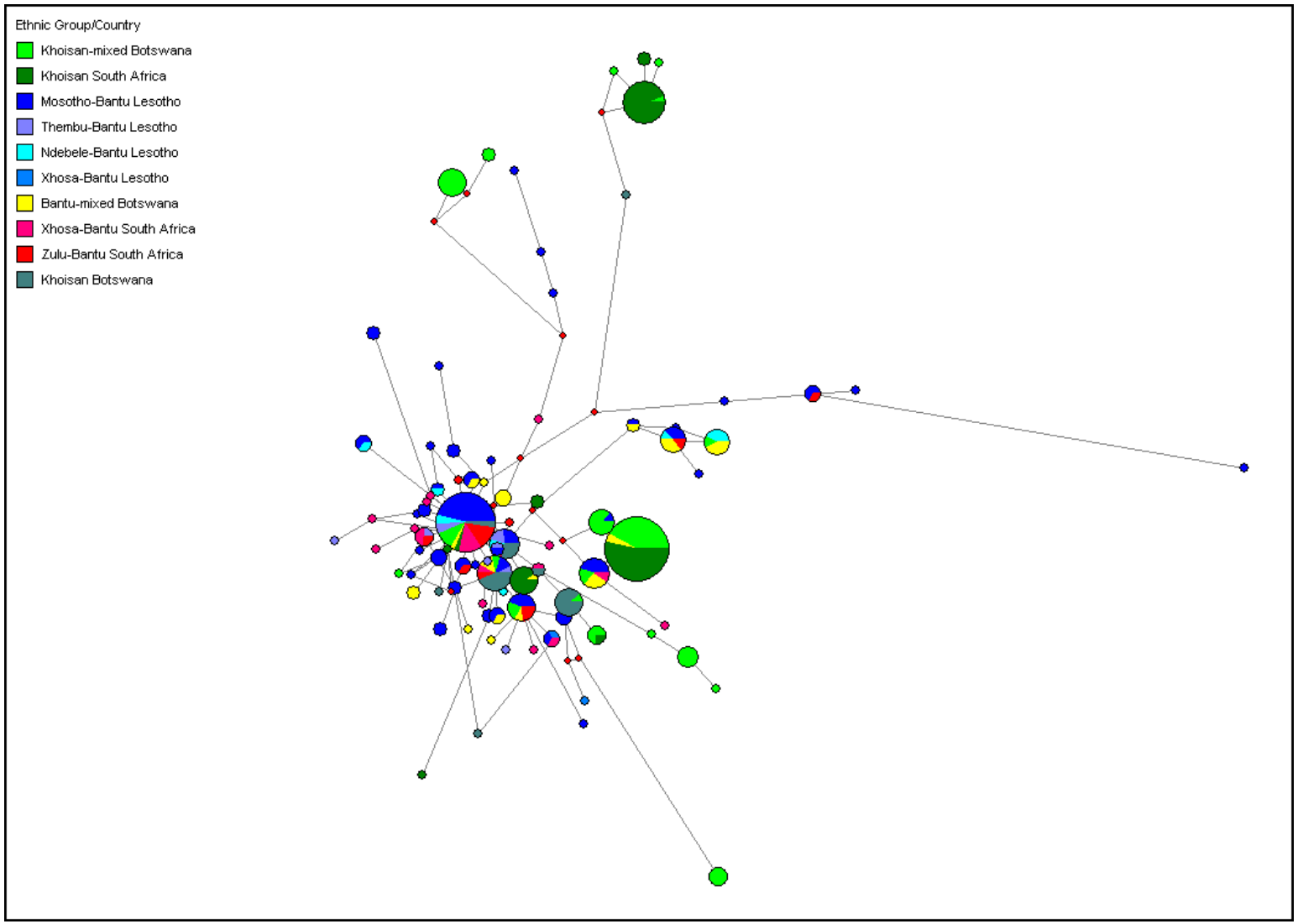


Figure 4.11: Phylogenetic network of mtDNA L0d haplotypes in southernmost African populations.

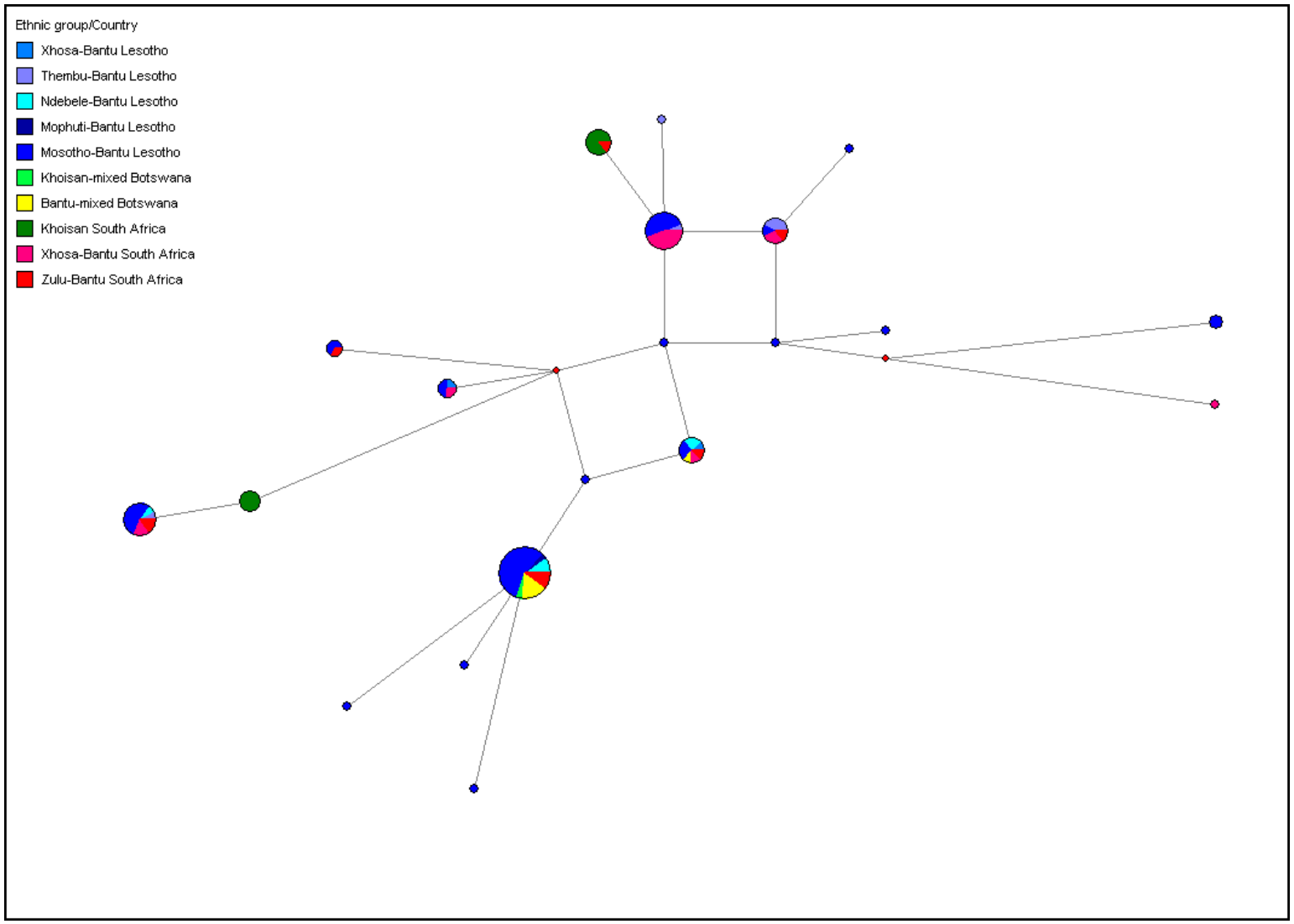


Figure 4.12: Phylogenetic network of mtDNA L3e haplotypes in southernmost African populations.

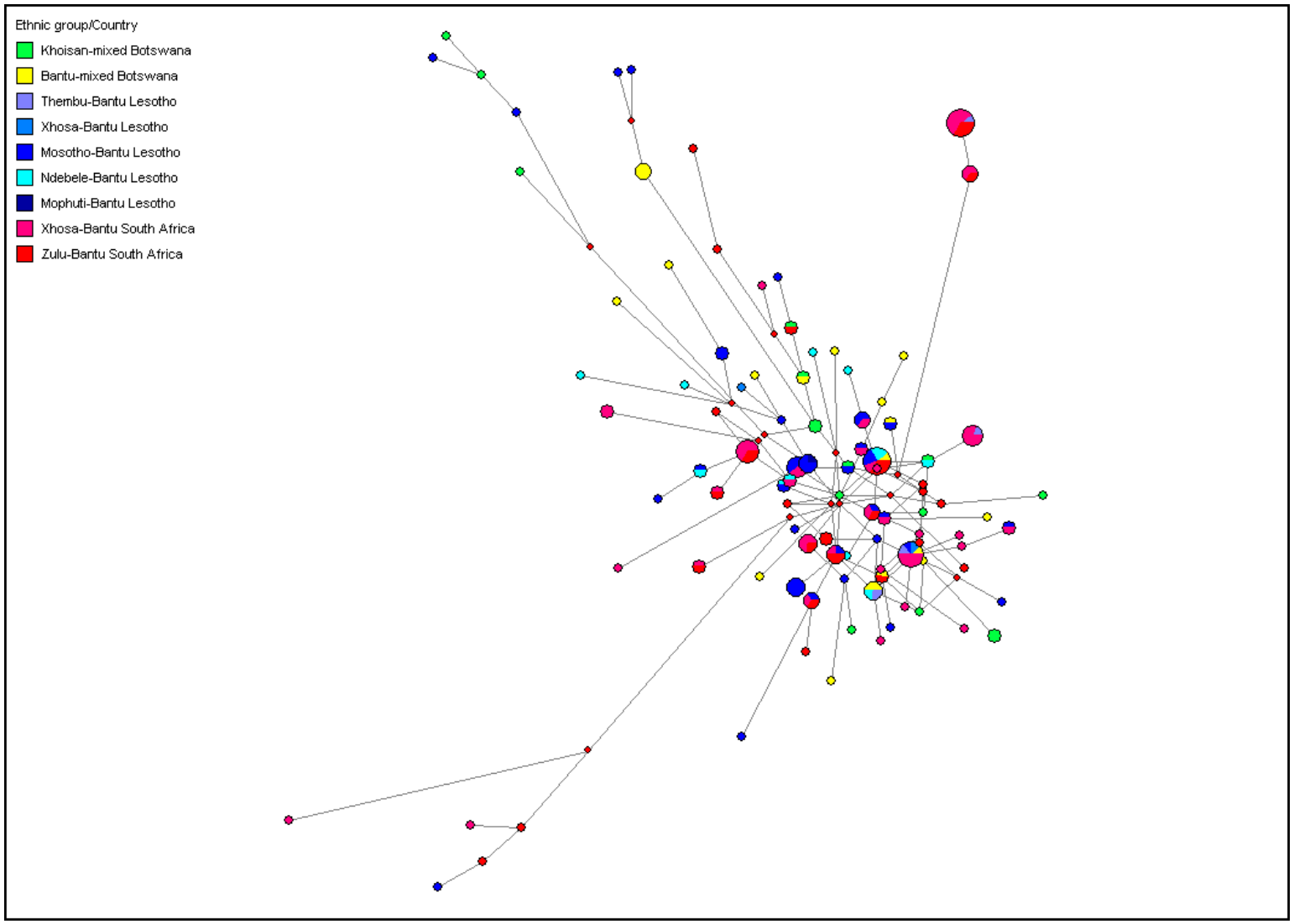


Figure 4.13: **Phylogenetic network of NRY E haplotypes in southernmost African populations.**

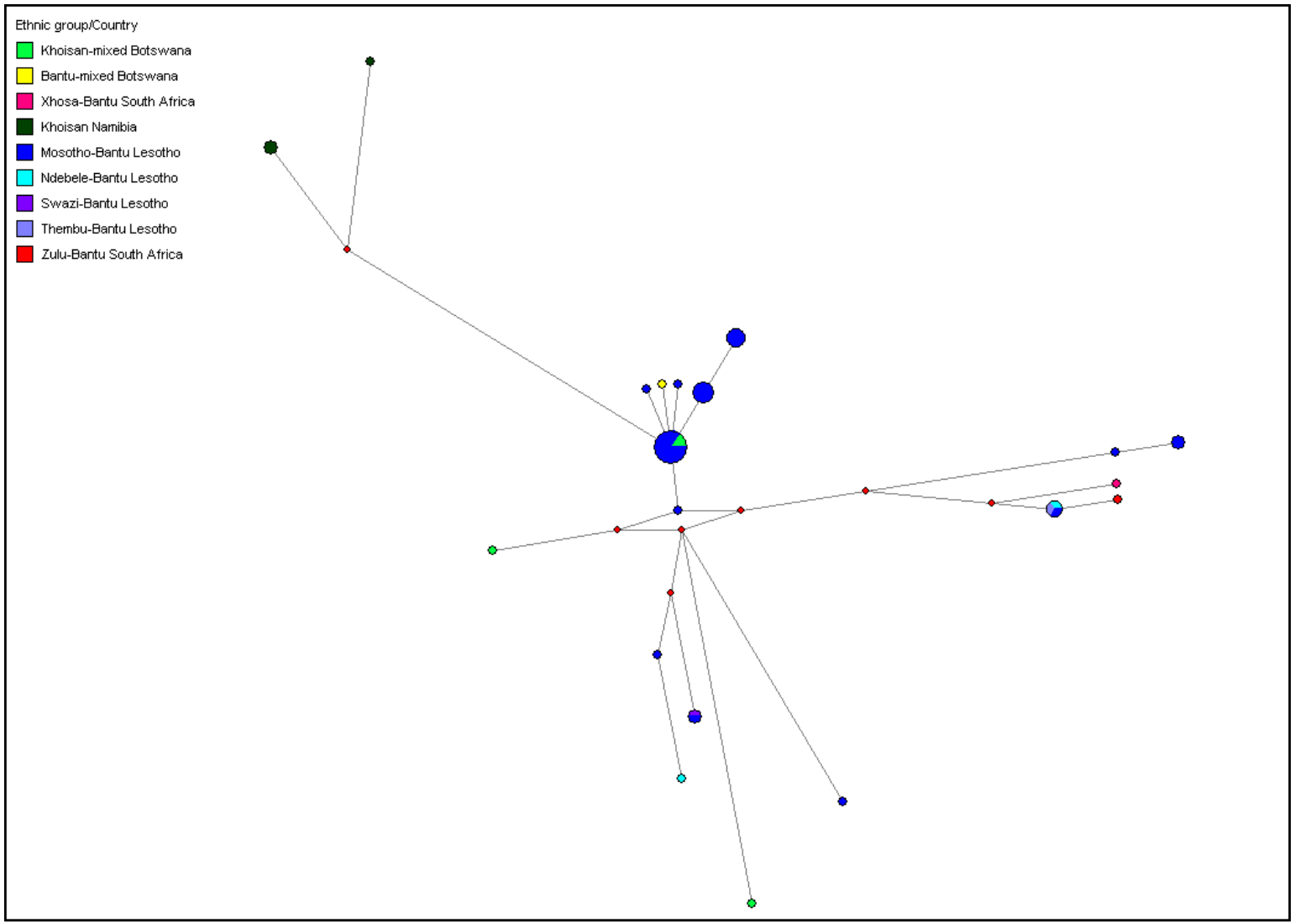


Figure 4.14: Phylogenetic network of NRY A3b1* haplotypes in southernmost African populations.

Discussion

It is known, from historical and archaeological evidence, that the Bantu-speaking farmers began migrating southwards from about 5000 years ago (Newman 1995). Movement such as this, with the settling Bantu farmers bringing in a sorghum/millet/bovid/caprid complex into Southern Africa (Figure 4.1: Region 1), has been described as a moving frontier (Alexander 1984). Archaeological evidence suggests that this advance was rapid (Hall 1987), providing little opportunity for interaction with indigenous populations. This fits with the genetic evidence available from this region, which shows populations to have typically Bantu genetic characteristics with little presence (if any) of lineages from the indigenous southern African 'Khoisan' people (Pereira et al. 2001; Pereira et al. 2002; Salas et al. 2002; Plaza et al. 2004; Castri et al. 2009; Coelho et al. 2009; de Filippo et al. 2010) (Tables 4.1 and 4.2).

By the mid-1st millennium AD the Bantu-speaking farmers became constrained by climatic and geographical conditions and thus were unable to settle any further than the edges of the Maloti/Drakensberg Mountains (Maggs 1976; Mitchell 2002; Mitchell et al. 2008): a static frontier developed (Maggs 1980) (Figure 4.1: Region 2). Hunter-gatherer populations were still present at this time across large areas of South Africa, and archaeological evidence has shown that much interaction seems to have occurred between them and the farmers on the other side of the frontier (Mitchell and Charles 2000; Mitchell 2002). For example, stone tools have been found at numerous early farming villages (Hall 1987), and farmer-associated ceramics and iron-ore remains have been found in hunter-gatherer settlements (Carter and Vogel 1974; Mazel 1986; Mitchell and Charles 2000). Oral evidence also indicates that cooperative hunting occurred between San and farmer groups in the region as recently as the 1930s (G Warrick, personal communication). It is known that the static frontier lasted for approximately 1000 years (Alexander 1984), with a large part of the interior of South Africa not being settled by farmers until at least 900 to 1400 AD and with the static frontier only being completely

disrupted after 1500 AD when new plants and animals were introduced, and Bantu-speaking groups were able to migrate further south, disrupting the equilibrium (Gill 1993; Mitchell et al. 2008).

This long-term interaction goes some way to explaining the genetic findings in the Lesotho, Xhosa and Zulu populations which fall in region 2 (Figure 4.1), Southernmost Africa, and speak South-East Bantu-Click languages. The high levels of maternal hunter-gatherer lineages found in these Bantu-speaking populations, especially when compared to populations from the surrounding regions (Figure 4.4; Table 4.1), support the idea that a static frontier allows for increased interactions. Under the 'wave of advance' model (Ammerman and Cavalli-Sforza 1973), hunter-gatherer lineages incorporated into the farming populations on the edge of an expansion, might then 'surf' to high levels once the static frontier gave way, causing particular alleles to increase in frequency along the direction of migration (Barbujani et al. 1995) – as is seen here. In region 1 on the map (Figure 4.1) there was seemingly not enough interaction between the two groups before the range expansion began (as a result of the rapid moving frontier). In fact, Currat and Excoffier's (2005) simulations from European data show that to achieve hunter-gatherer contributions to current populations as low as they are in other Sub-Saharan African countries such as Zimbabwe and southern Mozambique (Pereira et al. 2001; Salas et al. 2002; Castri et al. 2009), there must have been virtually no genetic contact between the expanding farmers and hunter-gatherers. Furthermore, modelling from Currat et al. (2008), suggests that to achieve the introgression levels that are observed in the Lesotho, Xhosa and Zulu populations (Figure 4.1; Region 2), as little as 0.5% interbreeding might be needed between the populations (this would produce introgression levels of over 50% from the local Hgs into the 'invading' populations). This idea is to be tested through simulation for further publication.

The results from both the mtDNA and NRY phylogenetic networks (Figures 4.11, 4.12, 4.13 and 4.14) produced similar results for both the hunter-gatherer and Bantu Hgs, with a large number of samples, including those from populations thousands of miles apart, sharing a small number of haplotypes. This could be because of a founder effect, with a few haplotypes being swept to high frequency during the initial Bantu arrival: it is known that rare alleles can 'surf' to high frequencies as a result of range expansions (Ammerman and Cavalli-Sforza 1973; Edmonds et al. 2004; Klopfstein et al. 2006; Excoffier and Ray 2008). Bantu haplotypes that were present in the populations on the edge of the frontier when it dissolved, would have found themselves on the edge of the wave of expansion, and might have undergone strong genetic drift, and surfed to high frequencies in the populations. This could have created the small number of very high frequency haplotypes that we see in today's populations. As the frontier broke down, and the farming populations recommenced their migration across southernmost Africa, they would have carried these haplotypes with them to the different regions of Southernmost Africa, thus explaining how populations as far apart as the Bantu-speakers in Botswana, and the Basotho in Lesotho might share the same haplotypes. Mutations that arise before or after the main wave expansion have a much smaller likelihood of surfing to high frequencies; so, the presence of the much rarer haplotypes could be explained if they were carried into the region either before or after the main wave (Klopfstein et al. 2006; Travis et al. 2007).

Support for the idea of a static frontier which was broken down only within the last 1000 years can also be found in the analysis of the mismatch distributions (Figures 4.7, 4.8, 4.9 and 4.10). The smooth bell-shaped distributions of the Southernmost Bantu-speaking populations for all the haplogroups grouped together, and additionally for two of the three haplogroups when considered separately, and the insignificant SSD value for the other, indicates a period of rapid population growth but with recent admixture with other populations. Contrastingly, the

hunter-gatherer populations from the region display distributions indicative of populations with a constant size. The mismatch distributions for South-East Africa perhaps suggest that the populations in this region have also recently mixed with other populations. Unfortunately, results for the Tajima's D, Fu's F_S, R_2 and Chakraborty tests for the mtDNA (Table 4.7) did not produce conclusive evidence for a different population history for the Southernmost populations as opposed to the South East populations.

Although the results here seem to differ from those of other populations in Sub-Saharan Africa with regards to the frequency of non-Bantu haplogroups present, they do share one characteristic; that of sex biased gene flow. The frequency of maternal hunter-gatherer haplogroups is considerably higher than that of paternal hunter-gatherer haplogroups, as shown by the Wilcoxon tests between mtDNA and Y Chromosome frequencies of non-Bantu Hgs within the regions (Figures 4.5 and 4.6; Tables 4.1 and 4.2.). This trend has previously been demonstrated in the Pygmy and farmer populations of Central Africa (Cavalli-Sforza 1986; Destro-Bisol et al. 2004a; Destro-Bisol et al. 2004b; Destro-Bisol 2005; Wood et al. 2005; Quintana-Murci et al. 2008), where socio-cultural factors appear to have influenced gene flow between the Pygmy and farmer groups, favouring marriage between female Pygmies and male farmers (Cavalli-Sforza 1986). Similar socio-cultural factors to those that existed in Central Africa might therefore have also existed in southernmost Africa, restricting male hunter-gatherers from interacting with the Bantu-farmers.

The results obtained investigating the maternal and paternal gene pool of Southern African populations suggest that population structure is present among the Bantu-speaking groups of the region. Higher maternal hunter-gatherer contributions are observed in populations in Southernmost Africa when compared to those from South East Africa, in accordance with a frontier model that has been used to explain the archaeological patterns seen.

Interestingly, it has been suggested that the San are in fact not extinct in the region, but became assimilated into the Bantu farming communities in order to survive; otherwise they faced the loss of their hunting areas, and prejudice against them (Southern African San Institute Annual Report in (Francis 2007)). Reports of recent integration between San groups and Xhosa and Sotho populations suggest farming groups made use of San traditions; for example, the use of San rainmakers to try and improve the growing conditions for their crops (Jolly 1996; Prins 2009). In other regions there is evidence of substantial interaction between the Nguni groups and San populations, where the Bantu-speakers adopted numerous San rituals (Hammond-Tooke 1998, 1999). It has been suggested that Bushmen that became integrated into these farming populations might be indistinguishable from the farmers from an archaeological and cultural perspective, as they adopted farming practices, lived in the same dwellings as the farmers and dressed like them (G Warrick, personal communication).

The high levels of 'Khoisan'-associated lineages in the Southern African populations studied here might, however, have succeeded at detecting these 'Secret San', who have a duel identity, referring to themselves as Sotho, Xhosa, Zulu etc. when asked by outsiders, but who consider themselves San amongst themselves (Prins 2009). Therefore, although the individuals used in this work identified themselves and their relatives as Sotho, Xhosa, Zulu etc., it is not known how many of these might actually be 'Secret San', but were unwilling to identify themselves as this to outsiders. Either way, any who are 'Secret San' are living integrated with the Bantu communities, speaking Bantu languages, and are probably interacting to greater and greater extents. This therefore does not remove the main outcome here; that the Bantu-speaking populations of Lesotho and South Africa studied here contain considerably more hunter-gatherer lineages, and thus underwent greater interaction with the San populations, that those north of the frontier.

CHAPTER 5: Hunter-gatherers, pastoralists and farmers: migrations and interactions in Namibian populations

Introduction

Namibia is one of the few Southern African countries where hunter-gatherer, pastoralist and farmer lifestyles still coexist. This close proximity of three different lifestyles makes Namibia a very interesting place in which to study the interactions that occur between groups with different lifestyles and/or ethnicity. Studying the population history of Namibia will also play an important role in understanding the history of Southern Africa as a whole, by providing increased coverage of samples in a currently understudied region. In fact, despite its relevance in addressing such questions, little data is available from Namibian populations in the literature, making comparisons both between ethnic groups within Namibia, and between the Namibian populations and other Southern African populations difficult.

Pastoralism

Many questions still remain about the histories of the various groups present in Namibia. One of the major questions that archaeologists and linguists have focused on with regards to Namibian population history has been the origins of pastoralism in Namibia and the wider region of Southern Africa (Fauvelle-Aymar and Sadr 2008; Mitchell 2010). Undoubtedly there are pastoralist populations present in Namibia today and have been for numerous years, with archaeological evidence of sheep bones and pottery dated to around 2000 BP (Smith et al. 1995). Evidence as to how this practice arrived in the region, and where from, is however, not

conclusive (Mitchell 2010). Using Namibian populations to try and help determine an exact origin for Southern African pastoralism might be difficult, however, as the only pastoralist Khoekhoe-speaking population in Namibia today, the Nama (Nurse et al. 1985), are known to be relatively recent arrivals in the country, having migrated northward from the Cape region in South Africa from about the 17th century (Ehret 1982a). Despite the relatively recent arrival of the Nama in Namibia, the scarcity of pastoralist populations in Southern Africa today means that analysis of the genetics of the Nama still has the potential to play an important role in answering questions concerning the history of pastoralism.

One of the major debates relating to pastoralism concerns the origins of the Southern African pastoralist populations. Three main lines of argument exist: 1. an origin in East Africa, followed by migration (Guldemann 2008); 2. an origin further south in the Zambezi region, followed by migration (Elphick 1985); 3. no migration, but a diffusion of 'techniques' instead (Sadr 1998). A large amount of the recent evidence, proposed from multiple disciplines, supports the idea of an East African origin and subsequent migration into Southern Africa. Linguistic studies of the Khoe-Kwadi suggest links with East African populations such as the Sandawe, a 'Khoisan'-speaking population (Guldemann 2008); and there is evidence of pastoralist techniques being shared between Southern African pastoralists and those in East Africa which are not seen elsewhere (Fauvelle-Aymar 2008). A genetic study has also linked a specific NRY polymorphism (E-M293) between East African populations and the Southern African Khwe and Kung populations ('Khoisan'-speaking) (Henn et al. 2008). If the theory of an East African origin is in fact correct, we would expect to see evidence in the Namibian Nama of similarity with East African populations, and possibly also the presence of the E-M293 NRY polymorphism. If either of the alternative theories are correct instead, similarity with East Africa would not be expected, but instead with South-East populations in the case of a Zambezi origin, or little sharing at all in the case of a non-migration model.

Complicating the process of understanding links between Southern African pastoralists and the rest of Africa, however, is the suggestion that it is likely that considerable contact occurred between migrating pastoralists and the Ju and Tuu hunter-gatherer groups already present in the Southern region (Guldemann 2008). This contact caused language modification (Mitchell 2010), and probably genetic exchange. As the Nama are descended from a group that migrated to southernmost Africa, it would be expected that they underwent considerable interaction with the Ju and Tuu, and thus for there to be a large San component to their genetics. Genetic signals associated with the pastoralists' area of origin might have thus been diluted. In fact, previous genetic studies do suggest that the Nama are more closely related to Ju-speaking hunter-gatherers than to other African populations (Guldemann and Stoneking 2008).

Who are the Damara?

A further question with regard to the Namibian populations is that of the origins of the Damara. Understanding the Damara's history might improve insights into the Nama and thus the arrival of pastoralism in the region, and also into general population migrations in Namibia.

The Damara speak a dialect of Nama, but do not look typically 'Khoisan' (Nurse et al. 1976), and were possibly present in Namibia before the Nama arrived (Barnard 1992). Previous genetic studies indicate that the Damara are not closely related to the Nama, and are unlikely to be closely related to any previously studied hunter-gatherer population (Nurse et al. 1976; Soodyall and Jenkins 1993; Guldemann and Stoneking 2008). Even if a close relationship with the Nama is disregarded, numerous other theories have existed with regards to Damara origins, although some are considered less plausible than others, including a suggestion that the Damara are slaves from north-east Africa brought into Southern Africa by the Khoekhoen (Vedder (1929) in (Nurse et al. 1976)). The debate over Damara origins is still wide open,

however, with previous studies even admitting that they had not 'solved the riddle of Dama origins' (Nurse et al. 1976), and suggesting that further work should be carried out.

The various current theories as to the origins of the Damara can be combined into four main hypotheses that are testable with genetic analyses:

1. **Damara as Kwadi**

As discussed above, it is possible that pastoralism entered Southern Africa by a migration of a Proto-Khoe-Kwadi group from East Africa (Guldemann 2008). It has been suggested that the Damara might have been part of this migration, and are potentially a remnant of Khoe-Kwadi, being closely related to Kwadi populations (Ehret 1982a). Once in Southern Africa, the Kwadi (and Damara) split off from the rest of the Proto-Khoe, migrating into northern Namibia (Haacke 2008) (Figure 5.1). The remainder of the Proto-Khoe migrated southwards undergoing considerable interaction with Ju and Tuu groups, but the Kwadi (and possibly Damara) did not undergo this interaction (Guldemann 2008). If this theory is correct, the Damara might be expected to show genetic similarities with Eastern pastoralist populations, but not with the San. The Kwadi are recorded as being an Angolan, cattle-herding, click-speaking population (although the clicks were not like those of 'Khoisan' languages) who resembled Herero and Kuvale populations rather than 'Khoisan' people (Rocha 2010). Unfortunately, there are no extant Kwadi-speaking populations for comparison with the Damara, as no Kwadi-speaking individuals have been recorded in Angola since the 1980s.

2. **Damara as an East African, non-Khoe-Kwadi population**

Haacke (2008) suggests that the Damara were not a Khoe-Kwadi population, but were instead a 'Negro' group that merged with the Khoe-Kwadi and underwent a language shift (Haacke 2008). He is uncertain, however, as to whether they migrated from further East into the Proto-

Khoe-Kwadi region, or whether they were already present in the Zambezi region prior to the Proto-Khoe-Kwadi arrival (Haacke 2008) (Figure 5.1). If the Damara are an East-African, or even South-East African, population it might be possible to detect genetic similarity with other East African, or South-East African, non-'Khoisan' populations.

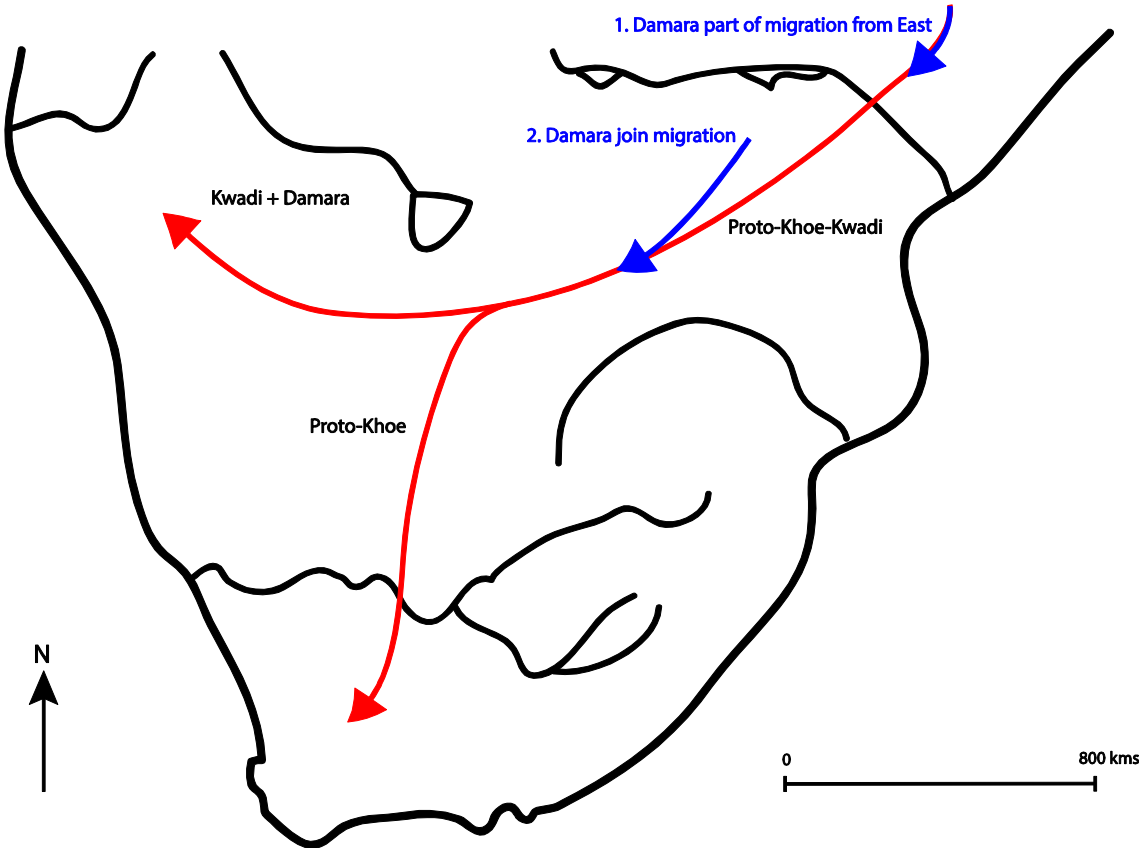


Figure 5.1: Diagram depicting theories 1 and 2 of the possible routes for a Damara migration into Namibia.
See main text for a full explanation of the theories.

3. Damara as an ancient Bantu population

It has also been suggested that the Damara might be an ancient Bantu population, among the first to have migrated into the region from Central Africa (Knussmann and Knussmann (1969/1970) in (Nurse et al. 1976)). If this is the case, it would be expected for the Damara to

appear genetically similar to Central African Bantu populations, sharing haplotypes with populations from this region.

4. Damara as a mixture of other populations

A further suggestion is that the Damara might be made up of a combination of disparate populations, perhaps joined together for convenience, that underwent a language shift to Nama as a lingua franca (Jenny (1966) in (Nurse et al. 1976)). If this is correct, the Damara might be genetically different to most populations in Sub-Saharan Africa, but would perhaps have a high diversity, being a combination of different groups. Haplotype sharing would be expected across a large number of different populations and regions.

Namibian Bantu populations

Further complicating the understanding of Namibia's history is the arrival of Bantu-speaking populations, an event that may have obscured signals of earlier migrations. The majority of people living in Namibia today belong to Bantu-speaking groups, and although the exact dates are unclear, there is evidence of Iron Age ceramics and technologies in northernmost Namibia during the first millennium AD (Mitchell 2002; Richter 2002-2003), although arrival of Bantu-speakers further south and west in Namibia might have been later. Some of the current Bantu-speaking populations might have arrived in Namibia as late as the 16th century AD (Nurse et al. 1985; Malan 1995). In addition to the lack of consensus about the history of pastoralism in the region and the origins of the Damara, there also appears to be some confusion with regards to the origins of the Bantu-speaking populations now resident in Namibia.

It is generally accepted that the Herero and Himba originated from the same source population, and arrived in Namibia in around the 16th century (Malan 1995). They probably

migrated down from Angola, and they do speak a South-West Bantu language similar to other Angolan populations (Lewis and SIL International 2009). The Owambo also speak a South-West Bantu language, closely related to Herero (Lewis and SIL International 2009), and it has been suggested that they might be related to Central African Bantu groups, having migrated south-west from Eastern Central Africa into Angola and down to Namibia in the 16th century (Malan 1995). This view is, however, contradicted by a previous genetic study which suggests that the Owambo (Ambo) might be more closely related to South-Eastern Bantu groups than to South-West ones, and that the Owambo are genetically different to the Herero (Soodyall and Jenkins 1993). With no clear evidence as to the origins of the Herero, Himba and Owambo, it will be interesting to see which other, if any, Bantu populations from Sub-Saharan Africa these Namibian populations are genetically similar too.

Although studies have been conducted previously, conclusive results are yet to be reached with regards to the history of the Namibian populations. Older studies use low resolution marker systems, or concentrate on only a small number of ethnic groups (e.g. (Nurse et al. 1976; Soodyall and Jenkins 1993). The most recent study investigates whole genomes from Namibian individuals and demonstrates the massive diversity in Southern Africans, yet it concentrates on only a very small number of samples (Schuster et al. 2010). Studies such as this show how interesting Namibian populations are with regards to Southern African genetics, yet there is scope for further genetic work to be carried out. Here, analyses will be conducted in order to address questions on the origins of pastoralism in the region (by analysing the Khoekhoe-speaking Nama), the origins and relationships of Bantu populations in Namibia (Herero, Himba and Owambo), and also to shed further light on the origins of the Damara.

Populations sampled

As part of this project six Namibian populations were sampled, with saliva and biogeographic information being obtained (See Materials and Methods, Chapter 2). Populations sampled were Damara, Herero, Himba, Ju'|hoan, Nama and Owambo (See Figure 2.2, Chapter 2).

Results

Haplogroup and haplotype composition, genetic diversity

At the haplogroup level all of the Namibian samples belong to the Sub-Saharan African mtDNA haplogroups L(xM,N) (Figure 5.2; Table 5.1). All of the Ju|'hoansi belong to haplogroups L0d or L0k, lineages known to be associated with 'Khoisan' populations (Vigilant et al. 1991; Chen et al. 2000; Salas et al. 2002; Kivisild et al. 2004). The Nama have 78.2% L0d and L0k, and 12.7% Bantu-associated haplogroups L0a, L1c1b, L1c2, L1c3, L1e, L2a1, L3b, L3d and L3e (Bandelt et al. 1995; Chen et al. 1995; Watson et al. 1997; Alves-Silva et al. 2000; Bandelt et al. 2001; Pereira et al. 2001; Salas et al. 2002; Plaza et al. 2004; Batini et al. 2007; Quintana-Murci et al. 2008; Berniell-Lee et al. 2009; de Filippo et al. 2010). By contrast, the Bantu-speaking populations, have much higher frequencies of Bantu-associated types (Herero: 68.2%; Himba: 54.9% (Davies 2011); Owambo: 36.7%), and lower frequencies of 'Khoisan'-associated types (Herero: 9.1%; Himba: 3.9% (Davies 2011); Owambo: 10.1%). Although the Damara speak a 'Khoisan' language, they display a high frequency of Bantu-associated types (80.6%) ('Khoisan'-associated types, 12%) (Davies 2011). The Owambo contain the largest number of different haplogroups.

For the NRY, although the majority of Hgs are Sub-Saharan African (A, B, E), some of the populations also contain Hgs identified as F-R in initial analyses (Figure 5.3; Table 5.2). The samples with haplogroups F-R were further genotyped by N. Davies and confirmed as belonging to European-associated haplogroups I, R1a and R1b1b2 (Davies 2011).The Ju|'hoansi and Nama contain the most 'Khoisan'-associated Hgs (A2, A3b1, B2b and E1b1b1g; Ju|'hoansi: 91.4%; Nama: 61.1%). Unlike with the mtDNA, the Ju|'hoansi do contain some non-'Khoisan'-associated Hgs (8.6%), and the Nama contain 25.9% non-'Khoisan'-associated and 13% European. The Bantu-speaking populations show high frequencies of the non-'Khoisan'-associated African Hgs (Herero: 76.2%; Himba: 96% (Davies 2011); Owambo: 97.7%), although they do contain some 'Khoisan' Hgs (Herero: 16.7%; Himba: 4% (Davies 2011); Owambo: 2.3%) and the Herero also have 7.1% European. The Damara contain minimal 'Khoisan' Hgs (3.7%), less than some Bantu-speaking populations; the majority of their Hgs are non-'Khoisan'-associated (87%), with a handful of European (9.3%) (Davies 2011).

Of the E haplogroups identified by the E multiplexes, the majority are sub-haplogroups E1b1a7a and E1b1a8a, which are found at high frequency in Bantu populations (de Filippo et al. 2011) (Figure 5.4; Table 5.2). As a percentage of the total E samples, these are (number in brackets is total number of E samples): Damara: 70.8% (48); Herero: 90.6% (32); Himba: 90.5% (42); Ju|'hoansi: 62.5% (8); Nama: 41.7% (24); Owambo: 81% (79). Only the Nama do not have E1b1a8a, the other populations have both this and E1b1a7a. The overall frequencies of the definite Bantu-associated Hgs (B2a, E1b1a7a, E1b1a8a) in the populations are therefore: Damara: 63%; Herero: 71.4%; Himba: 88%; Ju|'hoansi: 8.6%; Nama: 18.5%; Owambo: 82.8%.

Haplotype numbers, gene diversities, Mean Number of Pairwise Difference (MNPD) and Weighted Intralineage Mean Pairwise Difference (WIMP) values are reported for mtDNA haplotypes (Table 5.3a) and NRY haplotypes (Table 5. 5.3b and c) and haplogroups (Table 5.3d

and e). For the mtDNA haplotypes, the Damara, Herero and Himba have low gene diversities, and lower MNPD and WIMP values when compared to the average for Southern African Bantu populations. When comparing NRY haplotype diversity (10 STRs), however, these populations are more similar to the Southern African Bantu average, with the Herero even being slightly above the average gene diversity. In fact the Damara, Herero and Himba all have MNPD values higher than the Southern African Bantu average, and the Damara and Herero also have WIMP values higher than the average. Using haplogroup diversity, the Damara have a lower diversity than the Southern African Bantu average, but the Himba and Herero are both higher, with the Herero approximately double the other two groups. The grouping used for this is, however, large, with all of the E combined together to allow comparison with other populations. When diversity values are calculated taking into account the frequencies of the different E sub-haplogroups, the Herero are actually similar to the Damara and the Himba (using standard deviations).

Population	N	L0d	L0k	L0a~	L1c2~	L1c3~	L2a1~	L3b~	L3d~	L3e~	L0*	L1b	L1c1	L2b	L2c	L3f	L3h	L4b	L0f
Damara	67	8	0	0	11	0	0	0	41	2	0	2	0	1	0	0	1	0	1
Herero	44	3	1	0	1	0	0	0	27	2	0	0	0	0	0	9	1	0	0
Himba	51	2	0	3	0	3	0	0	17	8	0	0	1	0	0	17	0	0	0
Ju|'hoansi	55	45	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Nama	55	41	2	2	1	0	0	0	6	0	1	0	0	0	0	0	0	1	1
Owambo	79	5	3	21	5	0	7	2	9	6	0	5	4	2	2	8	0	0	0

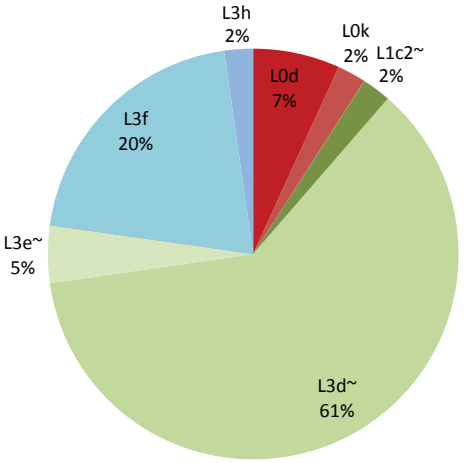
Table 5.1: Number of individuals per mtDNA haplogroup in the Namibian populations.

~ Bantu-associated haplogroups (See Table 4.1)

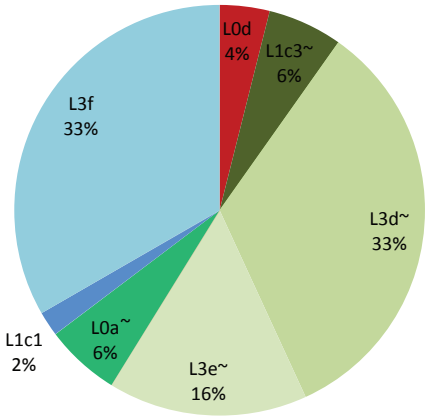
Population	N	A2	A3b1	B2a	B2b	E*	E1b1a	E1b1a7a	E1b1a8a	E1b1b1a	E1b1b1g	E(xE1)	E2b1	Non-African
Damara	54	0	1	0	0	1	8	24	10	0	1	1	3	5
Herero	42	1	2	1	3	0	2	18	11	0	1	0	0	3
Himba	50	0	0	6	2	1	3	17	21	0	0	0	0	0
Ju|'hoansi	58	13	10	0	27	0	0	1	4	0	3	0	0	0
Nama	54	0	22	0	1	0	2	10	0	0	10	0	2	7
Owambo	87	0	0	8	0	1	4	28	36	1	2	1	6	0

Table 5.2: **Number of individuals per NRY haplogroup in the Namibian populations.**

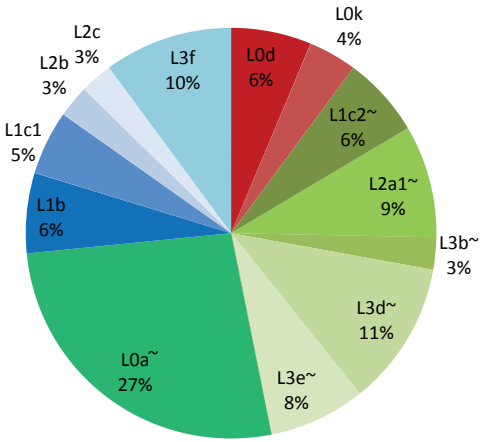
Figure 5.2: mtDNA haplogroup frequencies in the Namibian populations
~ = Bantu-associated haplogroups



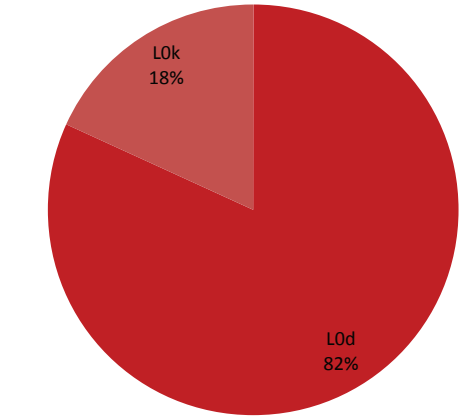
Herero, n = 44



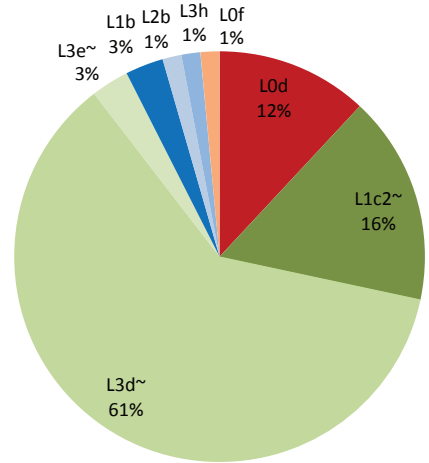
Himba, n = 51



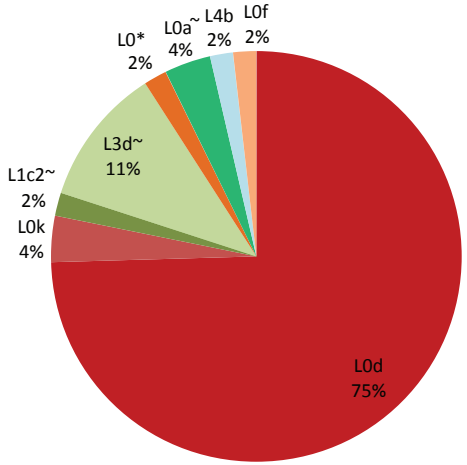
Owambo, n = 79



Ju|'hoansi, n = 55



Damara, n = 67



Nama, n = 55

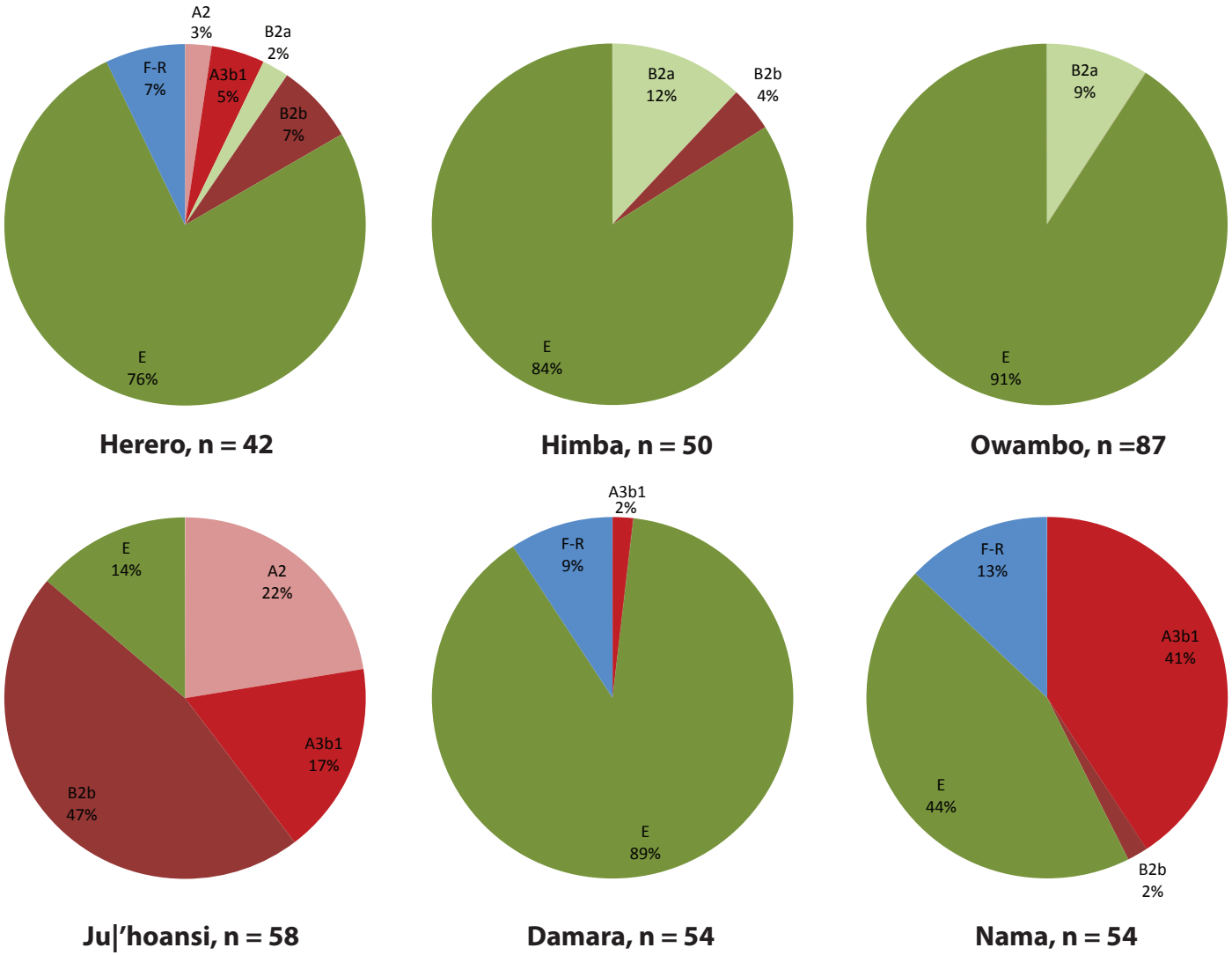
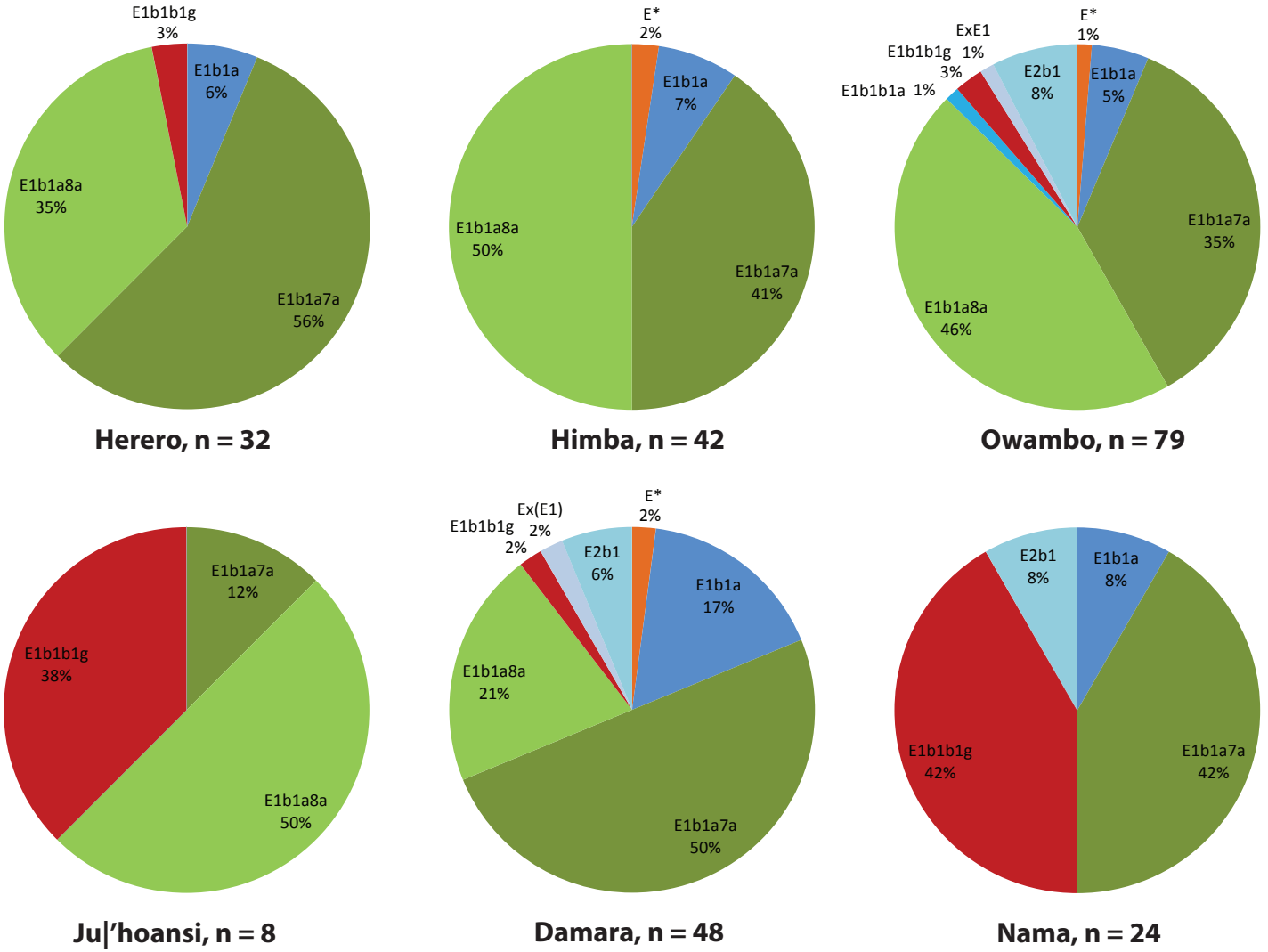


Figure 5.3: NRY haplogroup frequencies in the Namibian populations

Figure 5.4: E sub-haplogroup frequencies for the Namibian populations



Interestingly, although the Owambo have a lower mtDNA genetic diversity than the average for Southern Africa (not quite significantly different – standard deviation), they have a much higher genetic diversity than the other Namibian Bantu-speaking populations studied here. The Owambo NRY gene diversity (based on STR haplotypes) is also slightly higher than the other Namibian populations, especially with the full 15 STRs. For the Hg comparison the Owambo's gene diversity is actually lower, however, as virtually all of the Owambo samples belong to Hg E, so giving a false impression of low diversity. When comparing MNPD and WIMP values for the NRY, the Owambo are actually similar to the average for Southern Africa.

The Nama and Ju|'hoansi display diversity values that are generally lower, although not always significantly so, than the other Namibian populations for the NRY STR haplotypes, perhaps due to their low numbers of E haplotypes compared to the other populations. In fact, when WIMP values are calculated, the Nama and Ju|'hoansi have higher (not significantly) values than the other populations. When haplogroup frequencies (without E sub-haplogroups) are used these two populations are also more diverse than the other Namibian populations, as well as the 'Khoisan' average and the Southern African Bantu average. When taking into account the E sub-haplogroups, however, the Ju|'hoansi have the lowest diversity, because they do not have the range of E haplogroups that the other populations have. mtDNA-wise, the Ju|'hoansi and Nama have a higher diversity than the 'Khoisan' average, and also higher diversities than the other Namibian populations apart from the Owambo, and this is also the case for MNPD and WIMP values.

Interestingly, for both mtDNA and NRY, in all but one instance (Himba NRY), the WIMP values are lower than the MNPD values. This implies that the populations studied might comprise lineages that have low internal diversity, and are distantly related, indicating genetic bottlenecks (Hurles et al. 2002).

a. mtDNA haplotypes	N	N haplotypes	Gene diversity	SD	MNPD	SD	WIMP
Damara	66	18	0.7207	0.058	6.06	2.92	2.55
Herero	44	12	0.6765	0.071	4.78	2.38	3.9
Himba	51	13	0.80	0.035	6.05	2.93	3.89
Ju|'hoansi	58	16	0.925	0.013	6.48	3.11	6.48
Nama	57	21	0.9348	0.013	7.19	3.42	4.59
Owambo	79	41	0.9585	0.011	9.41	4.37	5.39
Owambo Subset	45	22	0.9263	0.024	9.38	4.39	
[a] *'Khoisan' (mean)*	*30.3*	*9.7*	*0.8201*	*0.060*	*7.78*	*0.98*	*4.81*
[a] *Southern Africa Bantu (mean)*	*45.6*	*26.4*	*0.9601*	*0.029*	*8.79*	*0.74*	*4.41*
b. NRY STR haplotypes (15 STRs)							
Damara	55	44	0.9872	0.0075			
Herero	39	32	0.9892	0.0085			
Himba	21	17	0.9762	0.023			
Ju|'hoansi	53	34	0.9782	0.0085			
Nama	54	35	0.963	0.016			
Owambo	50	46	0.9967	0.0048			
c. NRY STRs haplotypes (10 STRs)							
Damara	55	39	0.9744	0.011	5.84*	3.31	5.21
Herero	39	29	0.9784	0.013	7.82	4.77	5.71
Himba	21	16	0.9667	0.026	7.38	4.72	2.69
Ju|'hoansi	53	23	0.9565	0.011	12.24	6.23	7.49
Nama	54	31	0.9511	0.018	9.68	5.02	6.07
Owambo	50	35	0.9788	0.0095	5.8	3.21	4.61
[b] *Southern Africa Bantu (mean)*	*47.9*	*33*	*0.9767*	*0.016*	*5.68*	*1.01*	*4.77*
d. NRY haplogroups, using E sub-haplogroups							
Damara	54	9	0.747	0.046			
Herero	42	9	0.7491	0.051			
Himba	50	6	0.702	0.040			
Ju|'hoansi	58	6	0.7078	0.043			
Nama	54	7	0.7596	0.038			
Owambo	87	9	0.7172	0.032			
e. NRY haplogroups, without E sub-haplogroups							
Damara	54	3	0.2048	0.069			
Herero	42	5	0.4135	0.092			
Himba	50	3	0.2841	0.077			
Ju|'hoansi	58	3	0.6177	0.030			
Nama	54	4	0.631	0.033			
Owambo	87	2	0.1689	0.051			
[b] *'Khoisan' (mean)*	*38*	*2.8*	*0.51*	*0.056*			
[b] *Southern Africa Bantu (mean)*	*51*	*3.35*	*0.24*	*0.072*			

Table 5.3: Gene diversity for the Namibian populations.
a. mtDNA haplotypes; b. NRY STR haplotypes (15 STRs); c. NRY STR haplotypes (10 STRs); d. NRY haplogroups, using E sub-groups; e. NRY haplogroups, without E sub-haplogroups.
SD: Standard Deviation; MNPD: Mean Number of Pairwise Differences (*Mean Number of Stepwise Mutations for NRY); WIMP: Weighted Mean Intralineage Mean Pairwise Difference.
[a] Populations used for mtDNA haplotype comparison are listed in Supplementary Table 2, Appendix 2. [b] Populations used for NRY haplogroup and haplotype comparisons are listed in Supplementary Table 3, Appendix 2.

F_{ST} comparisons with other Sub-Saharan African populations

Population pairwise F_{ST} tests were conducted between the Namibian populations and other populations from Sub-Saharan Africa in order to better understand their histories. For the mtDNA, haplotype composition was used, and all of the Namibian populations apart from the Owambo are significantly different to all of the comparison populations. (See Davies (2011) for comparison of Herero, Himba, Nama and Damara to other populations, and Table 5.4, for the Owambo and Ju|'hoansi). The Owambo are the only Namibian population that are not significantly different to all of the other populations, suggesting they are similar to: one Angolan population, nine Mozambique populations and one Zambian population. This suggests that the Owambo, but not the Herero and Himba, have similar mtDNA to the other Bantu populations in southern Africa. All of the Namibian populations are significantly different to all the East African populations tested.

For the NRY haplotype comparisons the Ju|'hoansi and Nama are significantly different from all other populations (Table 5.5). The other four Namibian populations show some similarity to other populations from the region, although for most, the majority of comparisons are significantly different. The Damara are different to all but the Ganguela and Kuvale from Angola, and the Herero and Himba are also similar to these Angolan populations, perhaps not surprising considering that these populations are relatively close geographically and the Kuvale speak Herero (Coelho et al. 2009). The Owambo show the most similarity with other southern African populations, and are not significantly different from 13 populations, including some Angolan, and a large number of the Zambian populations. When compared with the East African populations, again the Ju|'hoansi and Nama are significantly different to all, but the other populations are not all significantly different, with the Owambo again showing most similarity.

Population	Country	Owambo	Ju|'hoansi
Ganguela	Angola	0.0044 (0.29+-0.013)	**0.24*** **(0.00+-0.00)**
Kuvale	Angola	**0.028*** **(0.0059+-0.003)**	**0.14*** **(0.00+-0.00)**
Mbundu	Angola	**0.035*** **(0.002+-0.0014)**	**0.25*** **(0.00+-0.00)**
Nyaneka-Nkhumbi	Angola	**0.019*** **(0.0078+-0.0024)**	**0.22*** **(0.00+-0.00)**
Ovimbundo	Angola	**0.031*** **(0.00+-0.00)**	**0.24*** **(0.00+-0.00)**
Chopi	Mozambique	**0.07*** **(0.00+-0.00)**	**0.31*** **(0.00+-0.00)**
Chwabo	Mozambique	**0.049*** **(0.012+-0.003)**	**0.31*** **(0.00+-0.00)**
Lomwe	Mozambique	0.02 (0.072+-0.0091)	**0.26*** **(0.00+-0.00)**
Makhuwa	Mozambique	0.01 (0.18+-0.011)	**0.26*** **(0.00+-0.00)**
Makonde	Mozambique	**0.07*** **(0.002+-0.0014)**	**0.31*** **(0.00+-0.00)**
Ndau	Mozambique	0.0044 (0.30+-0.014)	**0.19*** **(0.00+-0.00)**
Nguni	Mozambique	-0.017 (0.75+-0.016)	**0.24*** **(0.00+-0.00)**
Nyanja	Mozambique	**0.056*** **(0.0068+-0.003)**	**0.31*** **(0.00+-0.00)**
Nyungwe	Mozambique	0.01 (0.18+-0.011)	**0.28*** **(0.00+-0.00)**
Ronga	Mozambique	0.019 (0.073+-0.0077)	**0.22*** **(0.00+-0.00)**
Sena	Mozambique	**0.039*** **(0.022+-0.0056)**	**0.29*** **(0.00+-0.00)**
Shangaan	Mozambique	0.0095 (0.19+-0.013)	**0.25*** **(0.00+-0.00)**
Shona	Mozambique	0.029 (0.061+-0.0088)	**0.28*** **(0.00+-0.00)**
Tonga	Mozambique	**0.083*** **(0.00+-0.00)**	**0.31*** **(0.00+-0.00)**
Tswa	Mozambique	**0.062*** **(0.0029+-0.0016)**	**0.27*** **(0.00+-0.00)**
Yao	Mozambique	0.022 (0.16+-0.0093)	**0.30*** **(0.00+-0.00)**
General Population	Mozambique	**0.077*** **(0.00+-0.00)**	**0.28*** **(0.00+-0.00)**
Bisa	Zambia	**0.036*** **(0.0049+-0.002)**	**0.27*** **(0.00+-0.00)**
Kunda	Zambia	0.011 (0.12+-0.012)	**0.24*** **(0.00+-0.00)**
Shona	Zimbabwe	**0.015*** **(0.041+-0.0057)**	**0.24*** **(0.00+-0.00)**
Mixed 'Khoisan'	Botswana	**0.21*** **(0.00+-0.00)**	**0.065*** **(0.00+-0.00)**
Khwe	South Africa	**0.042*** **(0.0049+-0.0025)**	**0.18*** **(0.00+-0.00)**

Kung	South Africa	**0.14*** **(0.00+-0.00)**	**0.068*** **(0.002+-0.0014)**
!Xun-Khwe	South Africa	**0.17*** **(0.00+-0.00)**	**0.053*** **(0.026+-0.0057)**
Mixed Bantu	Botswana	**0.078*** **(0.00+-0.00)**	**0.11*** **(0.00+-0.00)**
Basotho	Lesotho	**0.028*** **(0.00+-0.00)**	**0.14*** **(0.00+-0.00)**
Ndebele	Lesotho	**0.031*** **(0.0059+-0.0026)**	**0.15*** **(0.00+-0.00)**
Thembu	Lesotho	**0.052*** **(0.0039+-0.0019)**	**0.13*** **(0.00+-0.00)**
Xhosa	South Africa	**0.064*** **(0.00+-0.00)**	**0.14*** **(0.00+-0.00)**
Zulu	South Africa	**0.036*** **(0.00098+-0.001)**	**0.18*** **(0.00+-0.00)**
Hadza	Tanzania	**0.20*** **(0.00+-0.00)**	**0.44*** **(0.00+-0.00)**
Datog	Tanzania	**0.042*** **(0.00098+-0.001)**	**0.24*** **(0.00+-0.00)**
Burunge	Tanzania	**0.063*** **(0.00+-0.00)**	**0.23*** **(0.00+-0.00)**
Sandawe	Tanzania	**0.14*** **(0.00+-0.00)**	**0.32*** **(0.00+-0.00)**
Sukuma	Tanzania	**0.15*** **(0.00+-0.00)**	**0.41*** **(0.00+-0.00)**
Turu	Tanzania	**0.085*** **(0.00+-0.00)**	**0.29*** **(0.00+-0.00)**

Table 5.4: Population pairwise F_{ST} values for mtDNA haplotypes, comparing the Owambo and Ju|'hoansi populations with other Southern African and East African populations from previous studies.

* = significant at the 0.05 significance level

Source: See Supplementary Table 2, Appendix 2.

Population	Country	Damara	Herero	Himba	Nama	Owambo	Ju|'hoansi
Ganguela	Angola	-0.013 (0.69+-0.018)	0.0097 (0.27+-0.011)	0.013 (0.27+-0.014)	**0.14*** **(0.00+-0.00)**	-0.00019 (0.43+-0.016)	**0.18*** **(0.00+-0.00)**
Kuvale	Angola	0.016 (0.090+-0.0089)	0.017 (0.11+-0.011)	0.06 (0.11+-0.0094)	**0.15*** **(0.00+-0.00)**	0.022 (0.056+-0.0068)	**0.19*** **(0.00+-0.00)**
Nyaneka-Nkhumbi	Angola	**0.042*** **(0.00+-0.00)**	**0.052*** **(0.00+-0.00)**	**0.042*** **(0.011+-0.0033)**	**0.2*** **(0.00+-0.00)**	**0.029*** **(0.0068+-0.0023)**	**0.23*** **(0.00+-0.00)**
Ovimbundu	Angola	**0.049*** **(0.00+-0.00)**	**0.061*** **(0.00+-0.00)**	**0.036*** **(0.016+-0.0039)**	**0.2*** **(0.00+-0.00)**	**0.027*** **(0.0059+-0.0022)**	**0.23*** **(0.00+-0.00)**
Bisa	Zambia	**0.033*** **(0.0039+-0.0019)**	**0.058*** **(0.00098+-0.001)**	**0.072*** **(0.00+-0.00)**	**0.21*** **(0.00+-0.00)**	0.0086 (0.20+-0.011)	**0.22*** **(0.00+-0.00)**
Kunda	Zambia	**0.031*** **(0.0049+-0.002)**	**0.047*** **(0.00+-0.00)**	**0.054*** **(0.0039+-0.0019)**	**0.17*** **(0.00+-0.00)**	0.0091 (0.16+-0.0091)	**0.18*** **(0.00+-0.00)**
Fwe	Zambia	**0.043*** **(0.0029+-0.0016)**	**0.055*** **(0.0020+-0.0014)**	**0.036*** **(0.035+-0.0047)**	**0.21*** **(0.00+-0.00)**	0.026 (0.052+-0.008)	**0.24*** **(0.00+-0.00)**
Kwamashi	Zambia	**0.046*** **(0.0049+-0.0025)**	**0.072*** **(0.00098+-0.001)**	**0.069*** **(0.0059+-0.0022)**	**0.21*** **(0.00+-0.00)**	**0.031*** **(0.040+-0.0069)**	**0.23*** **(0.00+-0.00)**
Lozi	Zambia	**0.034*** **(0.00+-0.00)**	**0.047*** **(0.00+-0.00)**	**0.036*** **(0.0078+-0.0024)**	**0.18*** **(0.00+-0.00)**	0.0089 (0.090+-0.0084)	**0.20*** **(0.00+-0.00)**
Luvale	Zambia	**0.052*** **(0.013+-0.0034)**	**0.036*** **(0.040+-0.0069)**	0.029 (0.12+-0.012)	**0.17*** **(0.00+-0.00)**	0.032 (0.071+-0.0068)	**0.19*** **(0.00+-0.00)**
Luyana	Zambia	**0.043*** **(0.00+-0.00)**	**0.056*** **(0.00+-0.00)**	**0.038*** **(0.017+-0.0042)**	**0.2*** **(0.00+-0.00)**	**0.02*** **(0.025+-0.0051)**	**0.22*** **(0.00+-0.00)**
Mbunda	Zambia	**0.06*** **(0.00098+-0.001)**	**0.067*** **(0.00+-0.00)**	**0.058*** **(0.0029+-0.0016)**	**0.2*** **(0.00+-0.00)**	**0.04*** **(0.00+-0.00)**	**0.23*** **(0.00+-0.00)**
Nkoya	Zambia	**0.038*** **(0.0039+-0.0019)**	**0.044*** **(0.0098+-0.0029)**	**0.037*** **(0.042+-0.0067)**	**0.18*** **(0.00+-0.00)**	0.018 (0.10+-0.012)	**0.21*** **(0.00+-0.00)**
Shanjo	Zambia	**0.037*** **(0.0088+-0.0025)**	**0.065*** **(0.00098+-0.001)**	**0.075*** **(0.0049+-0.0025)**	**0.22*** **(0.00+-0.00)**	0.016 (0.12+-0.010)	**0.22*** **(0.00+-0.00)**
Tonga	Zambia	**0.05*** **(0.00098+-0.001)**	**0.056*** **(0.0020+-0.0014)**	0.019 (0.13+-0.012)	**0.18*** **(0.00+-0.00)**	0.0097 (0.19+-0.012)	**0.19*** **(0.00+-0.00)**
Totela	Zambia	**0.037*** **(0.019+-0.0039)**	**0.046*** **(0.019+-0.0044)**	0.026 (0.14+-0.012)	**0.1*** **(0.00+-0.00)**	0.0046 (0.32+-0.015)	**0.14*** **(0.00+-0.00)**
Mixed Bantu West	Zambia	**0.021*** **(0.0049+-0.002)**	**0.038*** **(0.00+-0.00)**	**0.037*** **(0.011+-0.0033)**	**0.17*** **(0.00+-0.00)**	0.00001 (0.44+-0.015)	**0.20*** **(0.00+-0.00)**
Botswana Mixed Bantu	Botswana	**0.22*** **(0.00+-0.00)**	**0.17*** **(0.00+-0.00)**	**0.13*** **(0.00+-0.00)**	**0.18*** **(0.00+-0.00)**	**0.17*** **(0.00+-0.00)**	**0.15*** **(0.00+-0.00)**
Kalanga	Botswana	**0.05*** **(0.0039+-0.0023)**	**0.069*** **(0.00+-0.00)**	**0.067*** **(0.0039+-0.0019)**	**0.15*** **(0.00+-0.00)**	0.024 (0.067+-0.0079)	**0.16*** **(0.00+-0.00)**
Tswana	Botswana	**0.14*** **(0.00+-0.00)**	**0.099*** **(0.00+-0.00)**	**0.079*** **(0.0059+-0.0022)**	**0.16*** **(0.00+-0.00)**	**0.081*** **(0.00+-0.00)**	**0.13*** **(0.00+-0.00)**

Xhosa	South Africa	**0.08*** (0.00+-0.00)	**0.067*** (0.00+-0.00)	**0.062*** (0.00+-0.00)	**0.15*** (0.00+-0.00)	**0.043*** (0.00098+-0.001)	**0.16*** (0.00+-0.00)
Zulu	South Africa	**0.068*** (0.00+-0.00)	**0.058*** (0.00+-0.00)	**0.055*** (0.0029+-0.0016)	**0.15*** (0.00+-0.00)	**0.03*** (0.0078+-0.0028)	**0.16*** (0.00+-0.00)
Basotho	Lesotho	**0.093*** (0.00+-0.00)	**0.099*** (0.00+-0.00)	**0.086*** (0.00+-0.00)	**0.18*** (0.00+-0.00)	**0.049*** (0.00+-0.00)	**0.19*** (0.00+-0.00)
Ndebele	Lesotho	**0.13*** (0.00+-0.00)	**0.11*** (0.00+-0.00)	**0.087*** (0.00+-0.00)	**0.19*** (0.00+-0.00)	**0.064*** (0.00+-0.00)	**0.18*** (0.00+-0.00)
Thembu	Lesotho	**0.16*** (0.00+-0.00)	**0.15*** (0.00+-0.00)	**0.15*** (0.00+-0.00)	**0.23*** (0.00+-0.00)	**0.098*** (0.00+-0.00)	**0.19*** (0.00+-0.00)
Datog	Tanzania	**0.19*** (0.00+-0.00)	**0.15*** (0.00+-0.00)	**0.12*** (0.00+-0.00)	**0.13*** (0.00+-0.00)	**0.16*** (0.00+-0.00)	**0.12*** (0.00+-0.00)
Burunge	Tanzania	**0.096*** (0.00+-0.00)	**0.07*** (0.00+-0.00)	**0.056*** (0.013+-0.0031)	**0.067*** (0.0020+-0.0014)	**0.078*** (0.00+-0.00)	**0.098*** (0.00+-0.00)
Mbugwe	Tanzania	**0.068*** (0.0029+-0.0016)	**0.043*** (0.020+-0.0039)	0.035 (0.10+-0.0083)	**0.12*** (0.00+-0.00)	0.026 (0.10+-0.0096)	**0.1*** (0.00+-0.00)
Turu	Tanzania	0.025 (0.052+-0.0065)	0.024 (0.053+-0.007)	0.036 (0.058+-0.0081)	**0.089*** (0.00+-0.00)	0.021 (0.099+-0.0089)	**0.12*** (0.00+-0.00)
Wa Fiome	Tanzania	**0.46*** (0.00+-0.00)	**0.37*** (0.0059+-0.0022)	**0.37*** (0.0039+-0.0019)	**0.31*** (0.00098+-0.001)	**0.44*** (0.0029+-0.0016)	**0.26*** (0.029+-0.0052)
Hadza	Tanzania	**0.16*** (0.00+-0.00)	**0.13*** (0.00+-0.00)	**0.12*** (0.00+-0.00)	**0.14*** (0.00+-0.00)	**0.13*** (0.00+-0.00)	**0.13*** (0.00+-0.00)
Sandawe	Tanzania	**0.096*** (0.00+-0.00)	**0.068*** (0.00+-0.00)	**0.055*** (0.0039+-0.0023)	**0.096*** (0.00+-0.00)	**0.072*** (0.00+-0.00)	**0.11*** (0.00+-0.00)
Sukuma	Tanzania	**0.02*** (0.041+-0.0061)	**0.034*** (0.0059+-0.0026)	**0.035*** (0.031+-0.0043)	**0.12*** (0.00+-0.00)	0.0052 (0.26+-0.011)	**0.15*** (0.00+-0.00)

Table 5.5: Population pairwise F_{ST} values for NRY haplotypes, comparing the Namibia populations with Southern African and East African populations.

* = significant at the 0.05 significance level.

Source: See Supplementary Table 2, Appendix 2.

Principal component analysis

Principal component analysis (PCA) of mtDNA Hg frequencies for populations in Sub-Saharan Africa shows the Herero, Himba and Owambo clustering with other Southern African Bantu-speaking populations, while the Nama and Ju|'hoansi cluster with other Southern African 'Khoisan'-speaking (SAK) populations (Figure 5.5). The Damara, however, do not cluster with other 'Khoisan'-speakers, but instead with Bantu-speaking groups. If the Central African, SAK, and Southernmost Bantu-speaking populations are removed from the analysis, as they cluster separately on the plot, it becomes evident that the East African populations also cluster separately from the remaining populations (Figure 5.6). The Owambo appear to fall nearest the SW and SE populations, although not amongst them. Removal of the East African populations (Figure 5.7), results in the Owambo appearing more similar to the other Bantu populations from the SE and SW, while the Herero, Himba and Damara still fall separately.

The NRY PCA with populations from Sub Saharan Africa also shows the Damara, Herero, Himba and Owambo clustering with the Southernmost, South East and South West Bantu-speaking populations, while the Ju|'hoansi and Nama cluster with the SAK (Figure 5.8). The Eastern food producing populations and the Pygmy populations fall separately from the other populations. When only Southern African populations are included (Figure 5.9), the Nama and Ju|'hoansi populations still cluster with the SAK, although they are quite widely spread out. The other Namibian populations cluster within the large SE and SW food producers group. With these SE and SW populations it was also possible to increase the number of haplogroups used for comparison by including E sub-haplogroups in order to see if this helps with placing the Namibian populations with respect to the other populations from the region (Figure 5.10). With this analysis, the Namibian populations do not cluster with the SE and SW food producers, and instead are quite separate, although the Owambo do seem to be the closest to the SW populations.

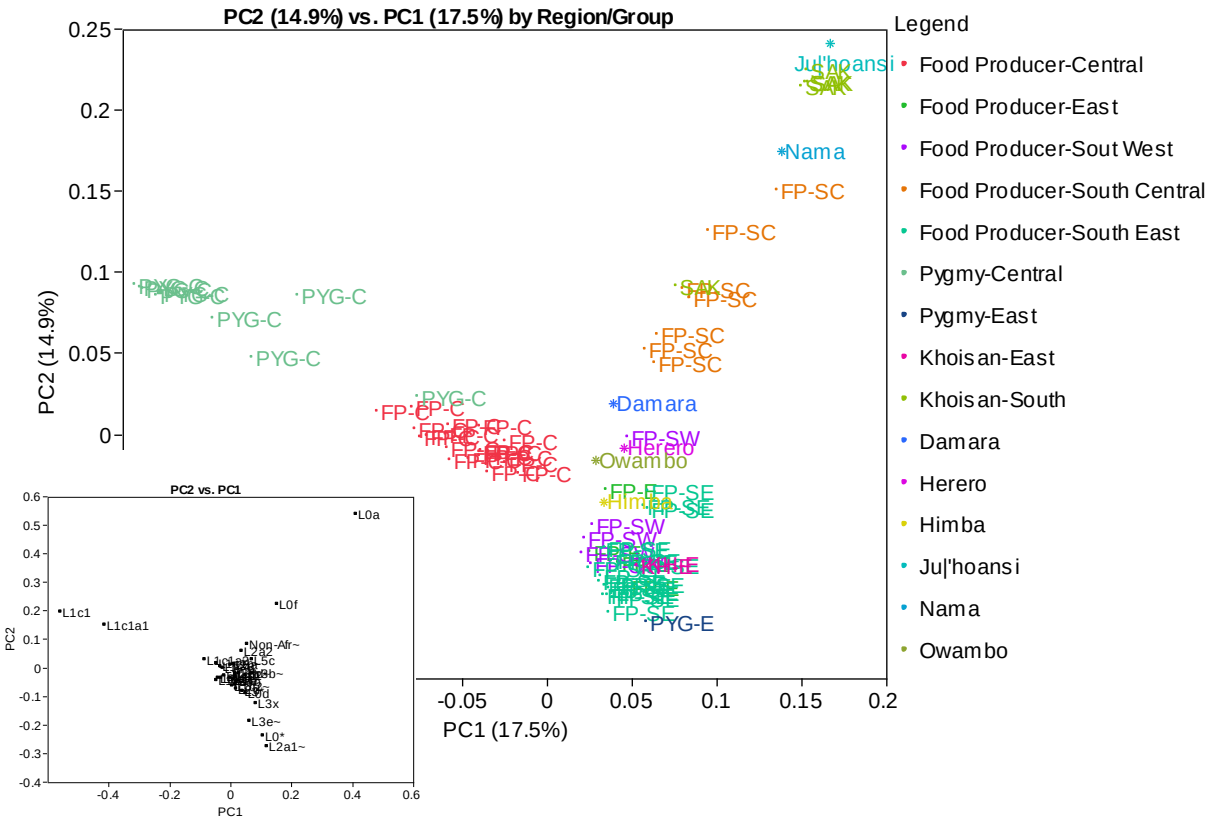


Figure 5.5: Principal component analysis plot of mtDNA haplogroup frequencies for Sub-Saharan African populations.
For populations used and abbreviations, see Supplementary Table 2, Appendix 2.

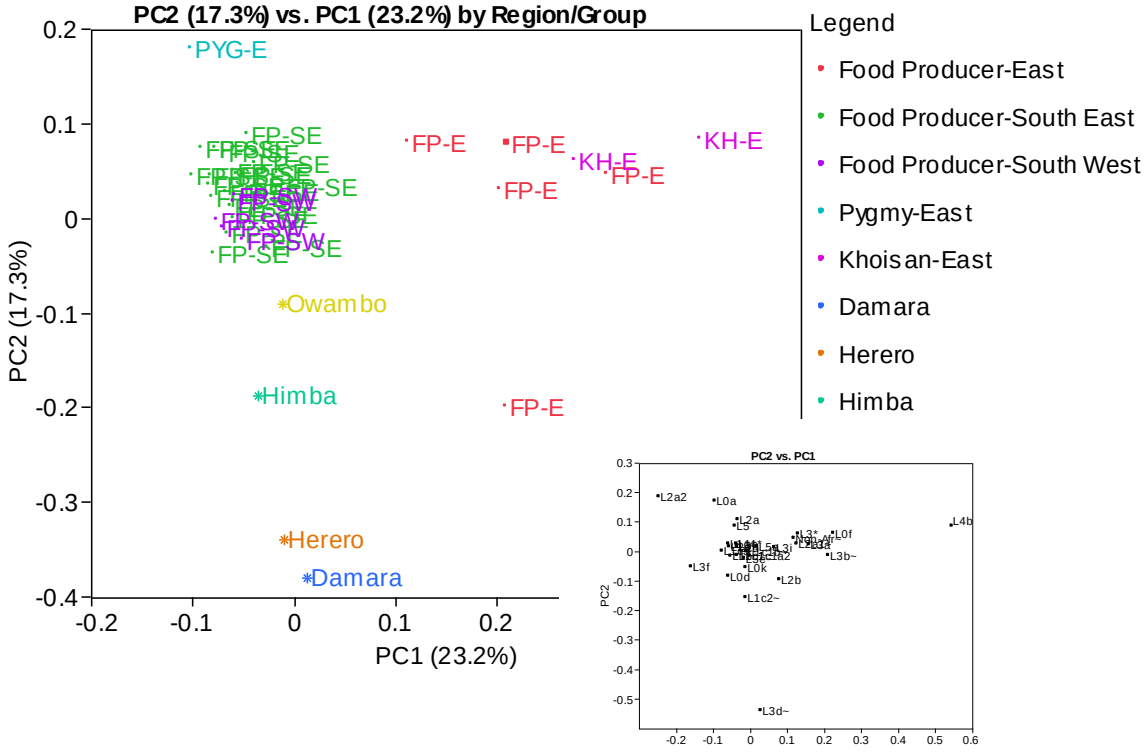


Figure 5.6: Principal component analysis plot of mtDNA haplogroup frequencies for Southern African and East African populations.
For populations used and abbreviations, see Supplementary Table 2, Appendix 2.

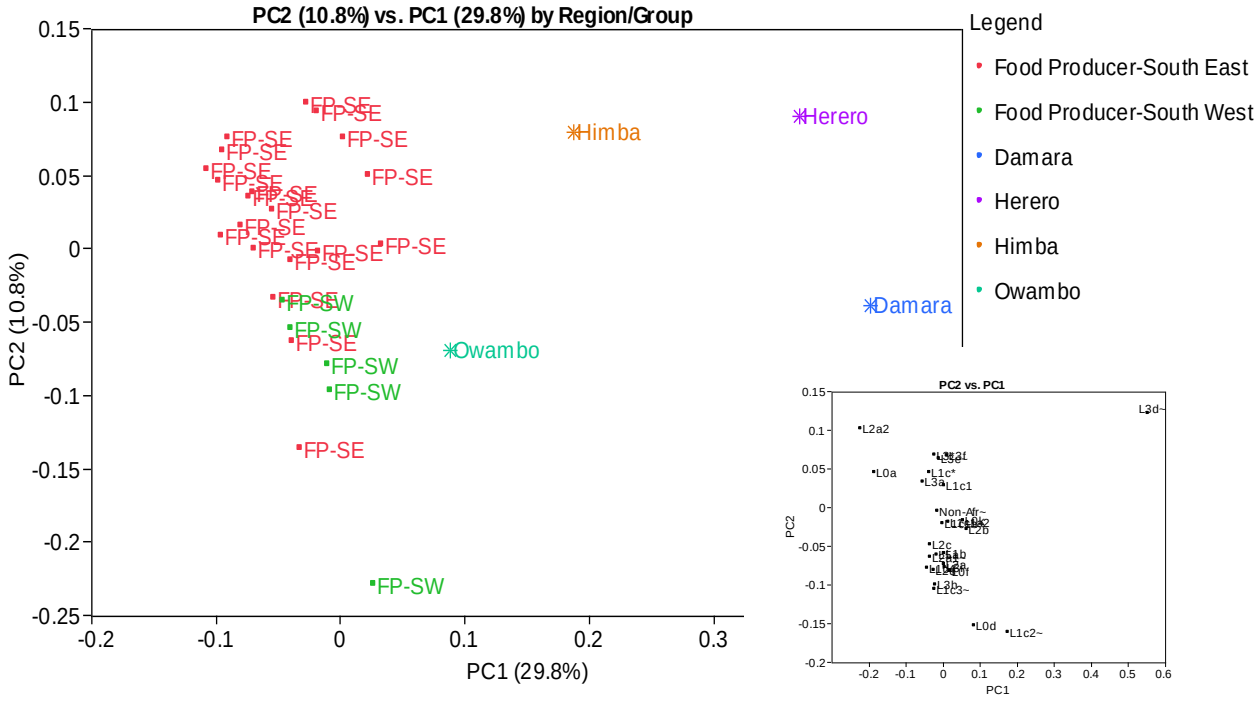


Figure 5.7: Principal component analysis plot of mtDNA haplogroup frequencies for South East and South West Africa populations.

For populations used and abbreviations, see Supplementary Table 2, Appendix 2.

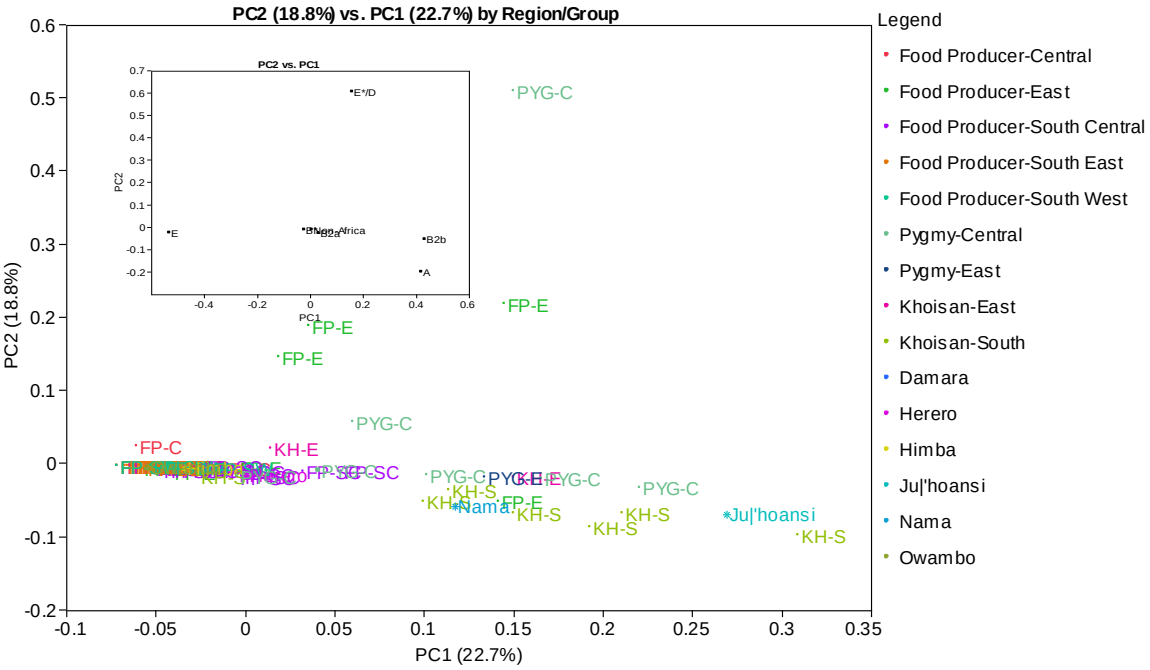


Figure 5.8: Principal component analysis plot of NRY haplogroup frequencies for Sub Saharan African populations.

For populations used and abbreviations, see Supplementary Table 3, Appendix 2.

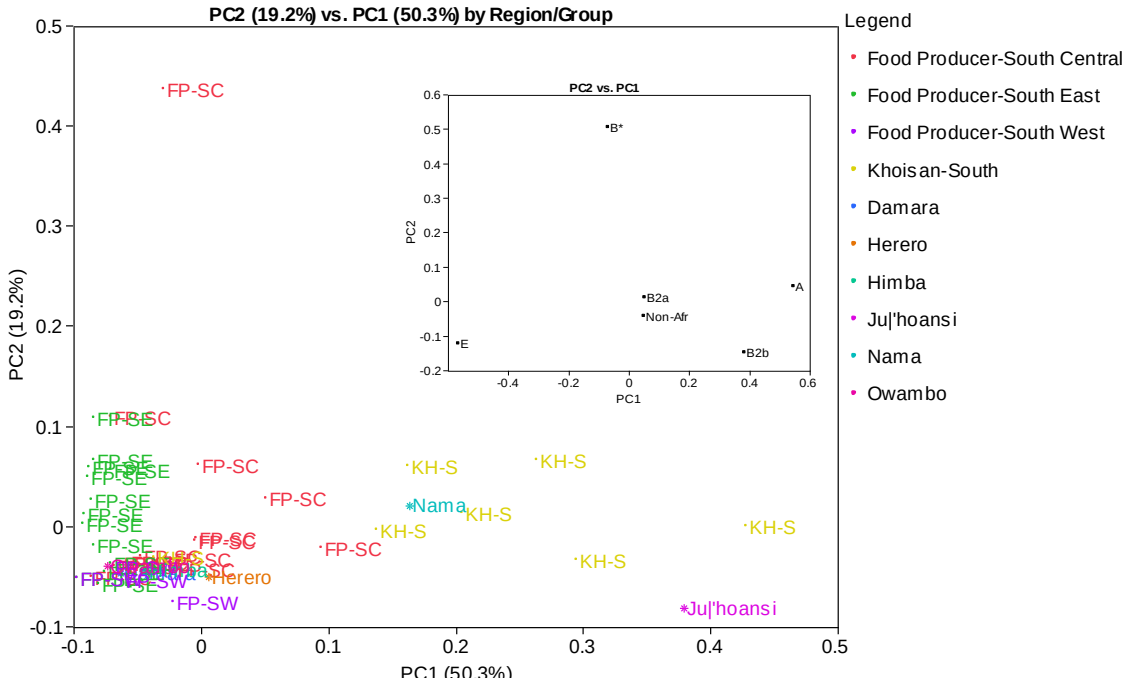


Figure 5.9: Principal component analysis plot of NRY haplogroup frequencies for Southern African populations.
For populations used and abbreviations, see Supplementary Table 3, Appendix 2.

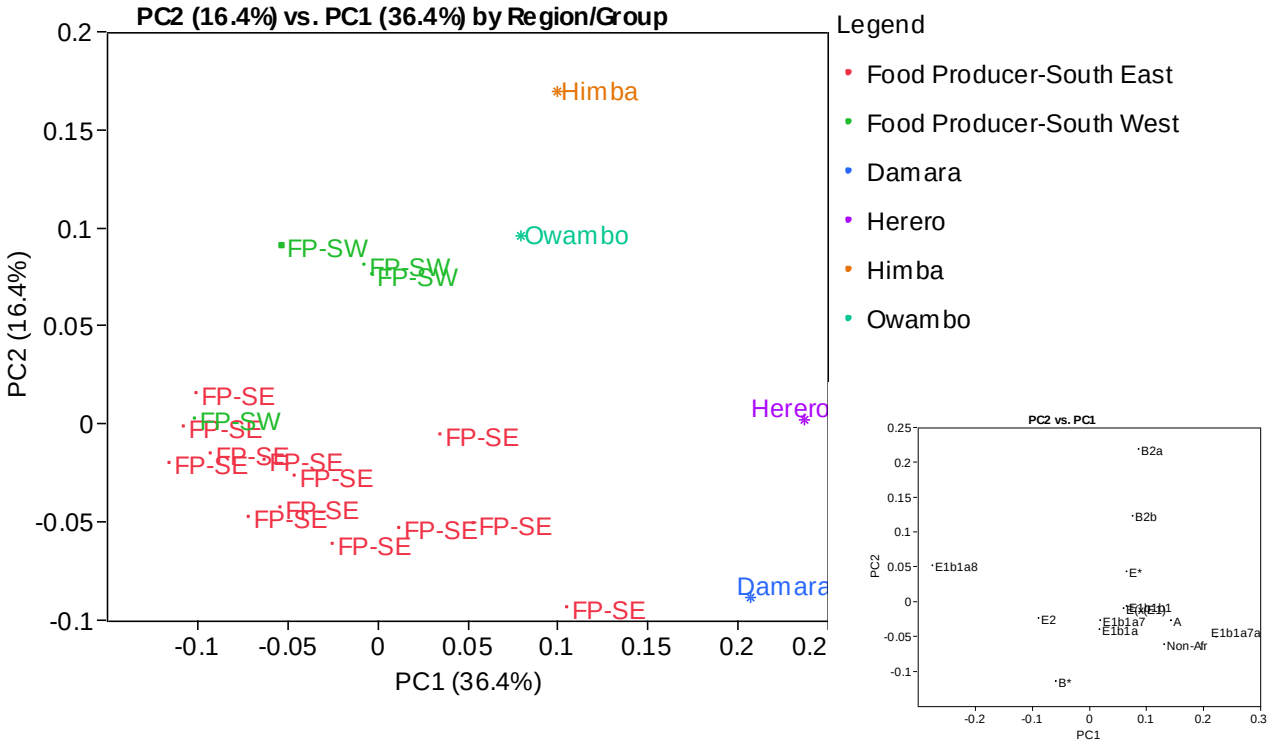


Figure 5.10: Principal component analysis plot of NRY haplogroup frequencies, including E sub-haplogroups for South East and South West African populations.
For populations used and abbreviations, see Supplementary Table 3, Appendix 2.

Haplotype Name	Haplogroup	Herero	Himba	Owambo	Damara	Nama	Ju|'hoansi		West	Central	'Khoisan' East	East	South East	Botswana Bantu	South West (Excluding Kuvale)	Kuvale	Southernmost	'Khoisan' South
mtDNA1	L0a	0	0	0	0	2	0		0	0	0	0	0	0	0	0	0	1
mtDNA2	L0a1b	0	1	0	0	0	0		0	0	0	0	0	0	0	0	0	0
mtDNA3	L0a1b	0	0	10	0	0	0		0	14	0	3	30	2	10	5	37	0
mtDNA4	L0a1b2	0	1	0	0	0	0		0	1	0	1	3	0	1	0	1	0
mtDNA5	L0a2	0	0	0	0	1	0		0	0	0	0	0	0	0	0	0	0
mtDNA6	L0a2	0	1	1	0	0	0		0	7	0	1	9	0	6	10	4	0
mtDNA7	L0a2	0	0	10	0	0	0		0	19	5	8	54	1	10	0	24	0
mtDNA8	L0d1	0	0	0	0	0	1		0	0	0	0	0	0	0	0	0	0
mtDNA9	L0d1	0	0	0	0	0	1		0	0	0	0	0	0	0	0	0	1
mtDNA10	L0d1	0	0	0	0	0	3		0	0	0	0	0	0	0	0	0	0
mtDNA11	L0d1'2	1	2	0	1	0	0		0	0	0	0	0	0	1	5	0	0
mtDNA12	L0d1'2	0	0	2	0	0	0		0	0	0	0	0	0	0	2	0	0
mtDNA13	L0d1'2	0	0	0	1	6	3		0	0	0	0	4	1	0	0	6	0
mtDNA14	L0d1'2	0	0	0	0	0	6		0	0	0	0	0	0	0	0	0	1
mtDNA15	L0d1a	0	0	0	0	1	0		0	0	0	0	0	0	0	0	0	0
mtDNA16	L0d1a	1	0	0	0	2	0		0	0	0	0	0	0	3	4	0	0
mtDNA17	L0d1a	0	0	0	0	6	0		0	0	0	0	0	0	0	0	10	2
mtDNA18	L0d1b	0	0	0	0	1	0		0	0	0	0	0	0	0	0	0	0
mtDNA19	L0d1b	0	0	0	1	1	0		0	0	0	0	0	0	0	0	0	0
mtDNA20	L0d1b	0	0	0	0	0	5		0	0	0	0	2	2	0	0	5	1
mtDNA21	L0d1b1	0	0	0	1	0	0		0	0	0	0	0	0	1	0	0	0
mtDNA22	L0d1b1	0	0	0	0	1	0		0	0	0	0	0	0	0	0	0	0
mtDNA23	L0d1b1	0	0	0	0	1	0		0	0	0	0	0	0	0	0	0	0
mtDNA24	L0d1b1	0	0	0	0	8	0		0	0	0	0	0	0	0	0	0	0
mtDNA25	L0d1b1	0	0	0	0	0	3		0	0	0	0	0	0	0	0	0	0
mtDNA26	L0d1b1	0	0	0	0	0	3		0	0	0	0	1	1	0	0	6	2
mtDNA27	L0d1c	0	0	0	1	0	0		0	0	0	0	0	0	0	0	0	0
mtDNA28	L0d1c	0	0	0	0	0	9		0	0	0	0	0	0	0	0	0	0
mtDNA29	L0d1c1	0	0	0	0	0	1		0	0	0	0	0	0	0	0	0	0
mtDNA30	L0d1c1	0	0	0	0	0	1		0	0	0	0	0	0	0	0	0	0
mtDNA31	L0d1c1	0	0	0	0	6	5		0	0	0	0	0	2	0	0	0	42
mtDNA32	L0d2a	0	0	2	0	0	0		0	0	0	0	0	0	2	0	0	0
mtDNA33	L0d2a	1	0	1	3	5	3		0	0	0	0	0	1	0	0	32	5
mtDNA34	L0d2b	0	0	0	0	0	1		0	0	0	0	0	0	0	0	0	0
mtDNA35	L0d3	0	0	0	0	1	0		0	0	0	0	0	1	0	0	1	0
mtDNA36	L0d3	0	0	0	0	3	0		0	0	0	0	1	3	0	0	5	0
mtDNA37	L0f1	0	0	0	1	1	0		0	0	0	0	0	0	0	0	0	0
mtDNA38	L0k1	0	0	1	0	0	0		0	0	0	0	0	0	0	0	0	0
mtDNA39	L0k1	0	0	1	0	0	0		0	0	0	0	0	0	0	0	0	1
mtDNA40	L0k1	1	0	1	0	0	5		0	0	0	0	0	0	0	0	0	2
mtDNA41	L0k1	0	0	0	0	2	8		0	0	0	0	0	0	1	0	0	20
mtDNA42	L1b1	0	0	1	0	0	0		3	38	0	0	2	0	2	0	1	0
mtDNA43	L1b1	0	0	2	0	0	0		0	0	0	0	0	0	0	0	0	0
mtDNA44	L1b1	0	0	2	2	0	0		0	13	0	0	4	0	10	0	0	1
mtDNA45	L1c1	0	1	1	0	0	0		0	1	0	0	0	0	1	5	0	0
mtDNA46	L1c1	0	0	2	0	0	0		0	0	0	0	0	0	0	0	0	0
mtDNA47	L1c1a2	0	0	1	0	0	0		0	0	0	0	0	0	0	0	0	0
mtDNA48	L1c2	0	0	0	1	0	0		0	0	0	0	0	0	0	0	0	0
mtDNA49	L1c2	1	0	0	8	1	0		0	1	0	0	0	0	2	0	0	0
mtDNA50	L1c2	0	0	2	0	0	0		0	0	0	0	1	0	0	0	0	0
mtDNA51	L1c2a1	0	0	1	0	0	0		0	0	0	0	0	0	0	0	0	0
mtDNA52	L1c2a1	0	0	1	0	0	0		0	2	0	0	2	1	1	1	0	0
mtDNA53	L1c2a2	0	0	1	0	0	0		0	0	0	0	0	0	1	0	0	0
mtDNA54	L1c2b	0	0	0	2	0	0		0	3	0	0	0	0	4	2	0	0
mtDNA55	L1c3a	0	1	0	0	0	0		0	0	0	0	0	0	0	0	0	0
mtDNA56	L1c3a	0	1	0	0	0	0		0	0	0	0	0	0	0	0	0	0
mtDNA57	L1c3a	0	1	0	0	0	0		0	0	0	0	3	0	0	0	0	0
mtDNA58	L2a1	0	0	1	0	0	0		0	0	0	0	0	0	0	0	0	0
mtDNA59	L2a1	0	0	1	0	0	0		0	2	0	0	0	0	0	0	0	0
mtDNA60	L2a1	0	0	1	0	0	0		0	9	0	0	0	0	0	0	0	0
mtDNA61	L2a1a2	0	0	1	0	0	0		0	1	0	0	41	0	5	0	4	0
mtDNA62	L2a1a3	0	0	2	0	0	0		0	11	0	1	0	0	2	0	0	0
mtDNA63	L2a1c1	0	0	1	0	0	0		0	2	0	0	0	0	2	0	0	0
mtDNA64	L2b	0	0	0	1	0	0		0	0	0	0	0	0	0	0	0	0
mtDNA65	L2b1a	0	0	1	0	0	0		0	0	0	0	0	0	0	0	0	0
mtDNA66	L2b1a	0	0	1	0	0	0		0	0	0	0	0	0	1	0	0	0
mtDNA67	L2c2	0	0	2	0	0	0		0	0	0	0	0	0	0	0	0	0
mtDNA68	L3b	0	0	1	0	0	0		0	0	0	0	0	0	0	0	0	0
mtDNA69	L3b1b	0	0	1	0	0	0		0	0	0	0	0	0	0	0	0	0

mtDNA70	L3d1a1	0	0	1	2	0	0		0	4	0	0	0	0	1	0	1	0
mtDNA71	L3d3a	0	0	0	1	0	0		0	0	0	0	0	0	0	0	0	0
mtDNA72	L3d3a	1	0	0	0	0	0		0	0	0	0	0	0	0	0	0	0
mtDNA73	L3d3a	2	1	0	3	0	0		0	0	0	0	1	0	1	0	0	0
mtDNA74	L3d3a	0	0	0	0	1	0		0	0	0	0	0	0	0	0	0	0
mtDNA75	L3d3a	0	0	1	0	0	0		0	0	0	0	0	0	0	0	0	0
mtDNA76	L3d3a	24	16	7	34	6	0		0	10	0	0	1	0	6	0	2	6
mtDNA77	L3e1a1	0	0	1	0	0	0		0	10	0	2	2	0	2	0	7	0
mtDNA78	L3e1a2	2	8	0	2	0	0		0	9	0	1	1	0	2	0	1	6
mtDNA79	L3e2b	0	0	1	0	0	0		0	0	0	0	0	0	0	0	0	0
mtDNA80	L3e2b	0	0	1	0	0	0		0	11	0	0	3	0	6	0	0	5
mtDNA81	L3e2b	0	0	2	0	0	0		0	29	0	3	4	0	8	0	13	0
mtDNA82	L3e3	0	0	1	0	0	0		0	23	0	3	15	0	8	0	4	0
mtDNA83	L3f1b	0	0	2	0	0	0		0	3	0	0	2	0	2	0	0	0
mtDNA84	L3f1b	0	0	2	0	0	0		1	5	0	5	0	0	0	0	0	0
mtDNA85	L3f1b1	0	0	4	0	0	0		0	12	0	2	4	0	2	0	0	0
mtDNA86	L3f1b4a	1	2	0	0	0	0		0	1	0	0	0	0	0	0	0	0
mtDNA87	L3f1b4a	8	15	0	0	0	0		0	25	0	0	3	0	11	3	3	0
mtDNA88	L3h	0	0	0	1	0	0		0	0	0	0	0	0	0	0	0	0
mtDNA89	L3h	1	0	0	0	0	0		0	0	0	0	0	0	0	0	0	0
mtDNA90	L4b2	0	0	0	0	1	0		0	0	0	0	0	0	0	0	0	0

Table 5.6: mtDNA haplotype sharing between the Namibian populations and other regions of Sub-Saharan Africa.

Only haplotypes present in the Namibian populations are included.

Red = haplotypes unique to a population; green = haplotypes unique to Namibia but shared within Namibia; blue = shared haplotypes.

See Supplementary Table 2, Appendix 2 for populations included in the regions.

Networks and haplotype sharing

Haplotypes (for all haplogroups) were compared between the Namibian populations and other populations from Sub-Saharan Africa in order to detect any sharing, and therefore potential links with different regions (Figures 5.11, 5.12, 5.13 and 5.14; Tables 5.6 and 5.7; Supplementary Tables 2 and 3, Appendix 2).

mtDNA

Of the 90 haplotypes seen across the Namibian samples (all grouped together), 39 are unique to Namibia. The rest are shared with populations from West Africa, Central Africa, East Africa or Southern Africa (Figures 5.11 and 5.12; Table 5.6). 292 of the 355 Namibian individuals have a haplotype that is shared with a population outside Namibia: Owambo 64/79; Damara 59/66; Herero 42/44; Himba 48/51; Nama 40/57; and Ju|'hoansi 39/58. This indicates that there is large amount of sharing across the whole of Sub-Saharan Africa. The haplogroups of the shared haplotypes are mixed, with both Bantu-associated and 'Khoisan'-associated lineages shared, although the 'Khoisan'-associated sharing is limited to populations in Southern Africa, as would be expected.

Southern Africa

A large number of Namibian haplotypes are shared with populations from Southern Africa: 32 with South West; 24 with South East and 20 with Southernmost Africa (Figures 5.11 and 5.12; Table 5.6). All of the Namibian populations share haplotypes with all of the southern areas. A large amount of the sharing in the Ju|'hoansi and Nama is with 'Khoisan' South populations as would be expected, with some of these haplotypes unique to these populations. The Owambo also have one unique shared haplotype with 'Khoisan' South. The Damara, Herero, Himba and Owambo populations share a large amount of haplotypes with all of the Southern populations, although the most with the South West region (Figures 5.11 and 5.12).

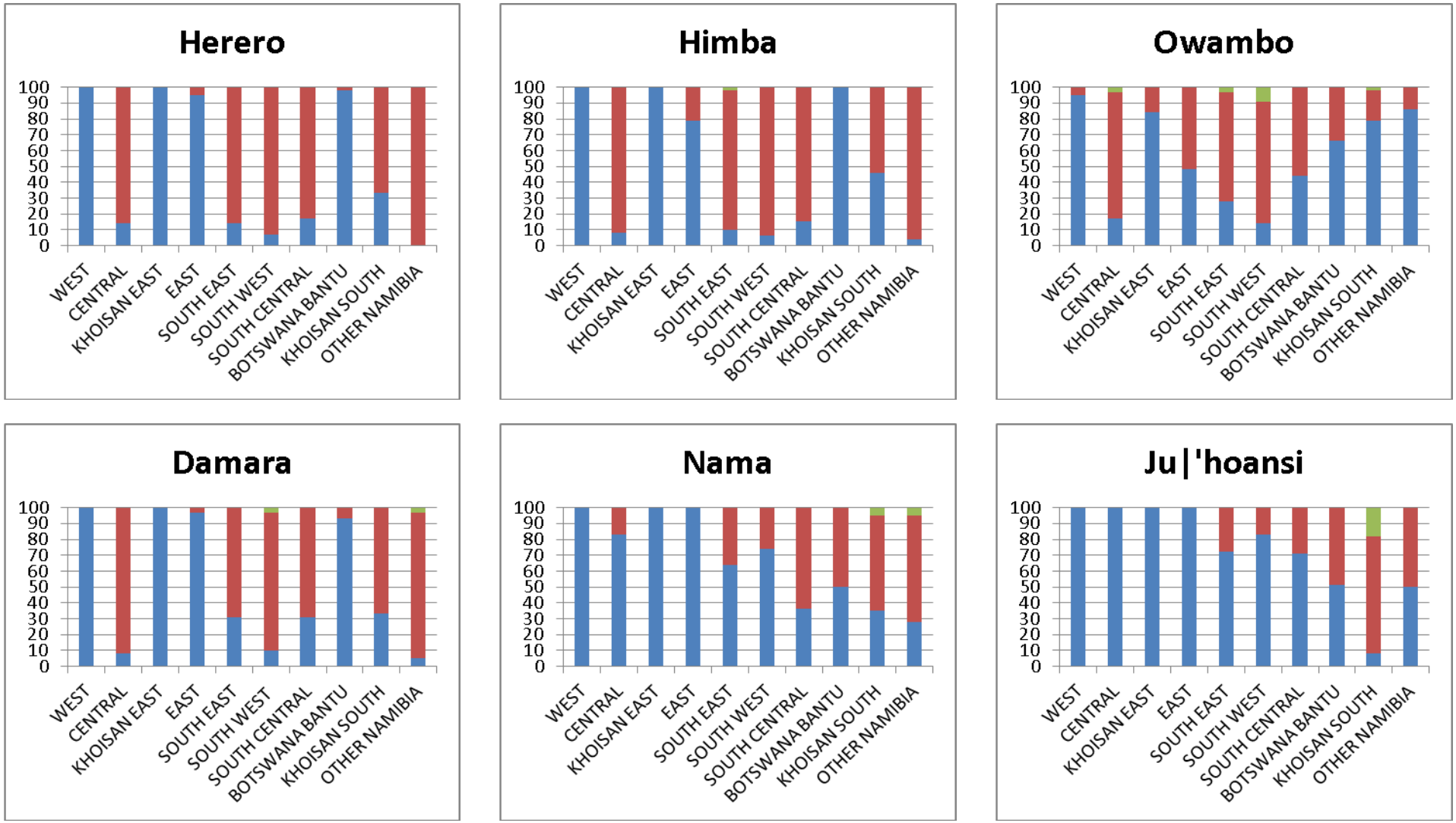


Figure 5.11: mtDNA lineage sharing between individuals in the six Namibian populations and regions in Sub-Saharan Africa.
Only mtDNAs shared between the Namibian population and at least one other region were included. Regions are the same as in Table 5.6.
Green = exclusive match; red = non-exclusive match; blue = no match.

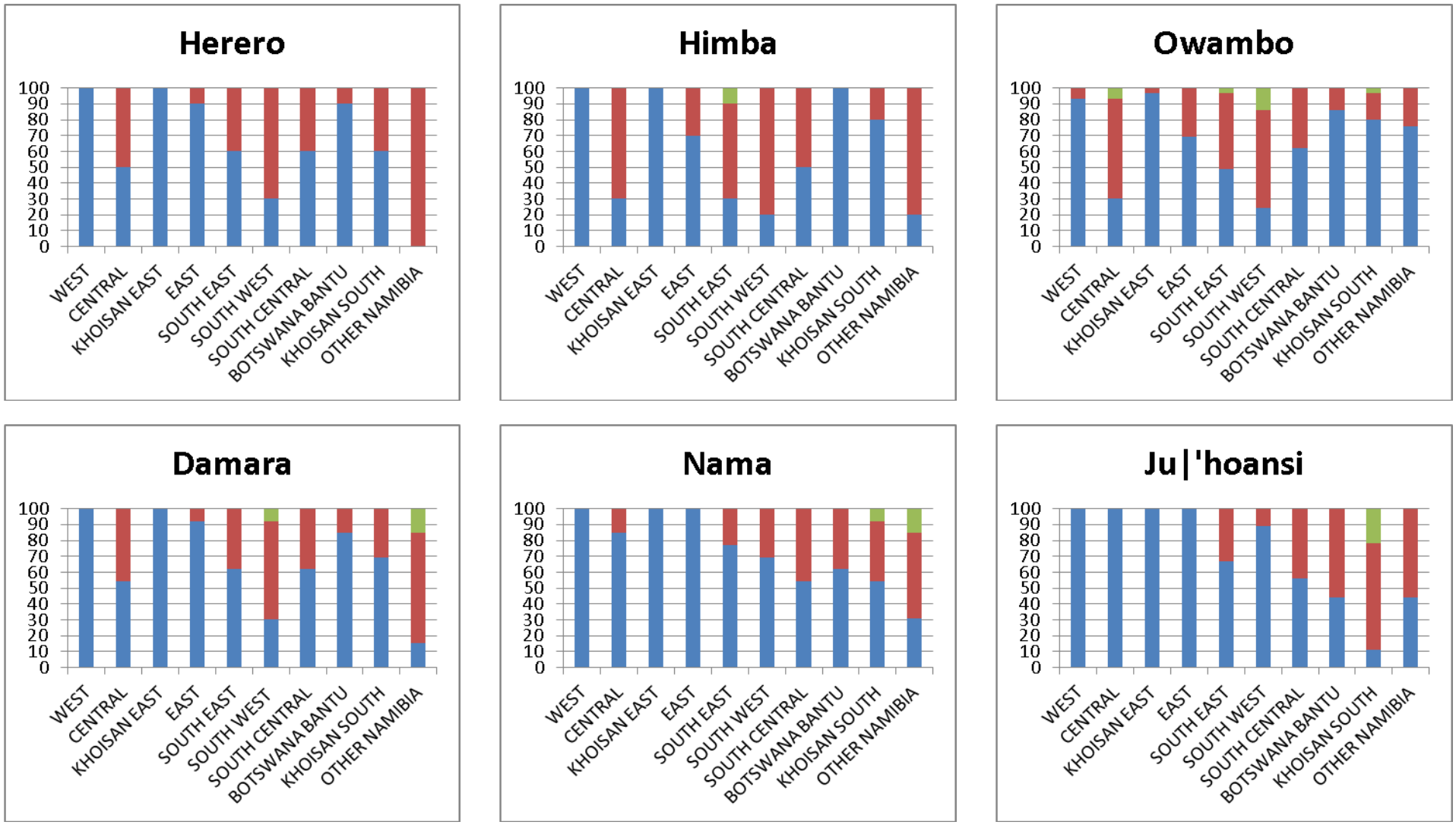


Figure 5.12: mtDNA lineage sharing between haplotypes in the six Namibian populations and regions in Sub-Saharan Africa.
Only mtDNAs shared between the Namibian population and at least one other region were included. Regions are the same as in Table 5.6.
Green = exclusive match; red = non-exclusive match; blue = no match.

East Africa

The Namibian populations share 11 haplotypes with East African populations. Of these, the majority are only found in the Owambo: eight haplotypes, and 32 individuals (Figures 5.11 and 5.12; Table 5.6). A further haplotype shared between one Owambo and one Himba individual is also found in East Africa. The remaining two haplotypes are not found in the Owambo, but are present in the Damara, Herero and Himba. These haplotypes are not, however, unique to East Africa, also being found in Central Africa, and the majority also in Southern Africa.

The Ju|'hoansi and Nama do not share any mtDNA haplotypes with East African populations. This is particularly interesting for the Nama, as pastoralism is thought to have originated in East Africa.

Central Africa

Namibian populations share 27 haplotypes with Central African populations. The majority of these haplotypes are also shared with Southern African populations; in fact only four are not (Figures 5.11 and 5.12; Table 5.6). Only six of the 27 haplotypes shared with Central African populations are not found in the Owambo. The Owambo are the only Namibian population with any unique haplotypes with Central Africa.

<u>NRY haplotype sharing</u>

For the NRY, of the 135 haplotypes found in the Namibian samples, 73 are unique to the Namibian populations (Figures 5.13 and 5.14; Table 5.7). Limited STR data was available for Southern African 'Khoisan' populations, so only Botswana Mixed 'Khoisan' samples were included as 'Khoisan'. It was, however, possible to include additional samples for separate haplogroup comparisons (see below) where whole population information was not necessary.

Haplotype Name	Haplogroup	Herero	Himba	Owambo	Damara	Ju|'hoansi	Nama		Pygmy Central	Central	'Khoisan' East	East	South-East	Botswana Bantu	South-West (Excluding Kuvale)	Kuvale	Southernmost	SAC	'Khoisan' South (Botswana)*
NRY1	A2	1	0	0	0	0	0		0	0	0	0	0	0	0	0	0	0	0
NRY2	A2	0	0	0	0	1	0		0	0	0	0	0	0	0	0	0	0	0
NRY3	A2	0	0	0	0	1	0		0	0	0	0	0	0	0	0	0	0	0
NRY4	A2	0	0	0	0	2	0		0	0	0	0	0	0	0	0	0	0	0
NRY5	A2	0	0	0	0	3	0		0	0	0	0	0	0	0	0	0	0	0
NRY6	A2	0	0	0	0	3	0		0	0	0	0	0	0	0	0	0	0	0
NRY7	A2	0	0	0	0	3	0		0	0	0	0	0	0	0	0	0	0	0
NRY8	A3b1	0	0	0	1	0	0		0	0	0	0	0	0	0	0	0	0	0
NRY9	A3b1	1	0	0	0	0	0		0	0	0	0	0	0	0	0	0	0	0
NRY10	A3b1	1	0	0	0	0	0		0	0	0	0	0	0	0	0	0	0	0
NRY11	A3b1	0	0	0	0	0	1		0	0	0	0	0	0	0	0	0	0	0
NRY12	A3b1	0	0	0	0	0	1		0	0	0	0	0	0	0	0	0	0	0
NRY13	A3b1	0	0	0	0	0	1		0	0	0	0	0	0	0	0	0	0	0
NRY14	A3b1	0	0	0	0	0	1		0	0	0	0	0	0	0	0	0	0	0
NRY15	A3b1	0	0	0	0	0	1		0	0	0	0	0	0	0	0	0	0	0
NRY16	A3b1	0	0	0	0	0	5		0	0	0	0	0	0	0	0	0	0	0
NRY17	A3b1	1	0	0	0	0	10		0	0	0	0	0	0	0	0	0	2	0
NRY18	A3b1	0	0	0	0	1	0		0	0	0	0	0	0	0	0	0	0	0
NRY19	A3b1	0	0	0	0	1	3		0	0	0	0	0	0	0	0	0	0	0
NRY20	A3b1	0	0	0	0	2	0		0	0	0	0	0	0	0	0	0	0	0
NRY21	A3b1	0	0	0	0	3	0		0	0	0	0	0	0	0	0	0	0	0
NRY22	B2a	0	1	0	0	0	0		0	0	0	0	0	0	0	0	0	0	0
NRY23	B2a	0	0	1	0	0	0		0	0	0	0	0	0	0	0	0	0	0
NRY24	B2a	0	0	1	0	0	0		0	0	0	1	0	0	0	0	0	0	0
NRY25	B2a	0	0	1	0	0	0		0	2	0	0	0	0	2	0	0	0	0
NRY26	B2a	0	0	1	0	0	0		0	0	0	0	0	0	0	0	0	0	0
NRY27	B2a	0	3	1	0	0	0		0	0	0	0	0	0	1	0	0	0	0
NRY28	B2b	2	0	0	0	0	0		0	0	0	0	0	0	0	0	0	0	0
NRY29	B2b	1	2	0	0	0	0		0	0	0	0	0	0	0	0	0	0	0
NRY30	B2b	0	0	0	0	0	1		0	0	0	0	0	0	0	0	0	0	0
NRY31	B2b	0	0	0	0	1	0		0	0	0	0	0	0	0	0	0	0	0
NRY32	B2b	0	0	0	0	1	0		0	0	0	0	0	0	0	0	0	0	0
NRY33	B2b	0	0	0	0	1	0		0	0	0	0	0	0	0	0	0	0	0
NRY34	B2b	0	0	0	0	2	0		0	0	0	0	0	1	0	0	0	0	0
NRY35	B2b	0	0	0	0	2	0		0	0	0	0	0	0	0	0	0	0	0
NRY36	B2b	0	0	0	0	3	0		0	0	0	0	0	0	0	0	0	0	0
NRY37	B2b	0	0	0	0	3	0		0	0	0	0	0	0	0	0	0	0	0
NRY38	B2b	0	0	0	0	5	0		0	0	0	0	0	0	0	0	0	0	0
NRY39	B2b	0	0	0	0	7	0		0	0	0	0	0	0	0	0	0	0	0
NRY40	E*	0	0	0	1	0	0		0	0	0	0	0	0	0	0	0	0	0
NRY41	E-M191	0	0	0	1	0	0		0	0	0	0	0	0	0	0	0	0	0
NRY42	E-M2	0	0	0	1	0	0		0	0	0	0	1	0	0	0	0	0	0
NRY43	E-M2	0	0	0	1	0	0		0	0	0	0	1	0	0	0	6	0	0
NRY44	E-M2	0	0	0	1	0	1		0	0	0	0	1	0	0	0	0	0	0
NRY45	E-M2	0	1	0	4	0	1		0	2	0	0	0	0	1	0	0	0	0
NRY46	E-M2	0	0	1	0	0	0		0	1	0	0	6	1	0	0	4	0	0
NRY47	E-M2	2	0	1	0	0	0		0	13	0	1	10	0	3	1	8	0	0
NRY48	E-M293	0	0	0	0	0	1		0	0	0	0	0	0	0	0	0	0	0
NRY49	E-M293	0	0	0	0	0	1		0	0	0	0	0	0	0	0	0	0	0
NRY50	E-M293	0	0	0	0	0	1		0	0	0	0	0	0	0	0	0	0	0
NRY51	E-M293	0	0	0	0	0	1		0	0	0	0	0	0	0	0	0	0	0
NRY52	E-M293	0	0	0	0	0	1		0	0	0	0	0	0	0	0	0	0	0
NRY53	E-M293	0	0	0	0	0	1		0	0	0	0	0	0	0	0	0	0	0
NRY54	E-M293	0	0	0	0	0	3		0	0	1	1	0	0	0	0	0	2	0
NRY55	E-M293	1	0	0	1	1	0		0	0	3	0	1	1	0	0	0	0	1
NRY56	E-M293	0	0	0	0	2	0		0	0	0	0	0	0	0	0	0	0	0
NRY57	E-M78	0	0	1	0	0	0		0	0	0	1	0	0	0	0	0	0	0
NRY58	E-M85	0	0	0	1	0	0		0	0	0	0	0	0	0	0	0	0	0
NRY59	E-M85	0	0	0	0	0	1		0	1	0	0	0	0	0	0	0	0	0
NRY60	E-M85	0	0	0	0	0	1		0	2	0	0	0	0	0	0	10	6	0
NRY61	E-M85	0	0	4	3	0	0		0	0	0	0	1	0	1	0	0	0	0
NRY62	E-U174	0	0	0	1	0	0		0	0	0	0	1	0	0	0	0	0	0
NRY63	E-U174	0	0	0	1	0	0		0	14	0	0	4	0	0	0	0	0	0
NRY64	E-U174	0	0	0	1	0	0		0	1	0	0	0	0	0	0	0	0	0
NRY65	E-U174	0	0	0	1	0	0		0	0	0	0	0	0	0	0	0	0	0
NRY66	E-U174	0	0	0	1	0	0		0	0	0	0	0	0	0	0	0	0	0
NRY67	E-U174	0	0	0	1	0	0		0	10	0	0	3	0	0	0	4	0	0
NRY68	E-U174	0	0	0	1	0	0		0	0	0	0	0	0	0	0	0	0	0
NRY69	E-U174	1	0	0	0	0	0		0	0	0	0	0	0	0	0	0	0	0

NRY70	E-U174	1	0	0	0	0	0		0	0	0	0	0	0	0	0	0	0	0
NRY71	E-U174	1	0	0	0	0	0		0	0	0	0	0	0	0	0	0	0	0
NRY72	E-U174	1	0	0	0	0	0		0	3	3	0	1	0	0	0	0	0	0
NRY73	E-U174	1	0	0	1	0	0		0	0	0	0	0	0	0	0	0	0	0
NRY74	E-U174	1	0	0	4	0	0		0	0	0	0	1	1	1	0	0	0	0
NRY75	E-U174	0	1	0	0	0	0		0	0	0	0	0	0	1	0	0	0	0
NRY76	E-U174	0	1	0	0	0	0		0	1	0	0	0	0	2	0	0	0	0
NRY77	E-U174	0	1	0	0	0	0		0	0	0	0	0	0	0	0	0	0	0
NRY78	E-U174	5	1	0	0	0	0		0	0	0	0	0	0	0	1	0	0	0
NRY79	E-U174	0	0	0	0	0	1		0	1	0	0	1	0	0	0	0	0	0
NRY80	E-U174	0	0	0	0	0	1		0	0	0	0	0	0	0	0	0	0	0
NRY81	E-U174	0	0	0	2	0	1		0	2	0	0	1	0	0	0	5	0	0
NRY82	E-U174	0	0	1	0	0	0		0	10	0	3	4	4	0	0	4	0	0
NRY83	E-U174	0	0	1	0	0	0		0	1	0	0	0	0	0	0	0	0	0
NRY84	E-U174	0	0	1	0	0	0		0	2	0	0	0	0	2	0	1	0	0
NRY85	E-U174	0	0	1	0	0	0		0	1	0	0	0	0	0	0	0	0	0
NRY86	E-U174	0	0	1	0	0	0		0	0	0	0	0	0	0	0	0	0	0
NRY87	E-U174	0	1	1	1	0	0		0	2	0	1	6	0	0	0	0	0	0
NRY88	E-U174	1	1	1	0	0	0		0	0	0	0	1	0	1	0	0	0	0
NRY89	E-U174	1	1	1	1	0	0		0	7	0	0	3	0	1	0	0	0	0
NRY90	E-U174	1	0	1	0	0	2		0	0	0	0	0	0	1	0	0	0	0
NRY91	E-U174	1	0	1	7	0	4		0	0	0	0	1	0	0	1	0	0	0
NRY92	E-U174	0	0	2	0	0	0		4	34	0	0	1	0	0	0	1	0	0
NRY93	E-U174	0	0	2	1	0	0		0	20	0	0	23	1	4	0	2	1	0
NRY94	E-U174	0	0	3	0	0	0		0	0	0	0	0	0	0	0	0	0	0
NRY95	E-U174	1	0	5	0	0	0		0	8	4	7	11	0	2	0	6	1	0
NRY96	E-U174	0	0	1	0	1	0		0	1	0	0	1	0	6	2	0	0	0
NRY97	E-U209	0	0	0	1	0	0		0	0	0	0	0	0	0	0	0	0	0
NRY98	E-U209	0	0	0	1	0	0		0	0	0	0	0	0	0	0	0	0	0
NRY99	E-U209	0	0	0	1	0	0		0	2	0	0	1	0	0	0	0	0	0
NRY100	E-U209	0	0	0	1	0	0		0	1	0	0	0	0	0	0	0	0	0
NRY101	E-U209	0	0	0	1	0	0		0	0	0	0	0	0	0	0	0	0	0
NRY102	E-U209	0	0	0	1	0	0		0	0	0	0	0	0	0	0	0	0	0
NRY103	E-U209	0	0	0	1	0	0		0	0	0	0	0	0	0	0	0	0	0
NRY104	E-U209	0	0	0	2	0	0		0	5	0	0	1	0	1	0	0	0	0
NRY105	E-U209	1	0	0	0	0	0		0	3	0	0	1	0	1	0	0	0	0
NRY106	E-U209	1	0	0	1	0	0		0	2	0	0	1	0	1	0	0	0	0
NRY107	E-U209	2	0	0	0	0	0		0	0	0	0	4	0	0	0	0	0	0
NRY108	E-U209	0	1	0	0	0	0		0	5	0	0	4	0	0	0	0	0	0
NRY109	E-U209	0	1	0	0	0	0		0	1	0	0	3	0	0	0	0	0	0
NRY110	E-U209	0	0	1	0	0	0		0	9	2	1	1	0	2	0	1	0	0
NRY111	E-U209	0	0	1	0	0	0		0	0	0	0	0	0	0	0	0	0	0
NRY112	E-U209	0	0	1	0	0	0		0	11	0	0	10	1	3	0	0	0	0
NRY113	E-U209	0	0	1	0	0	0		0	0	0	0	4	0	0	0	0	0	0
NRY114	E-U209	0	0	1	0	0	0		0	0	0	0	0	0	0	0	0	0	0
NRY115	E-U209	2	0	1	0	0	0		0	17	1	0	46	3	16	2	1	3	1
NRY116	E-U209	0	1	1	0	0	0		0	0	0	0	0	0	0	0	0	0	0
NRY117	E-U209	0	3	1	0	0	0		0	8	3	4	9	0	4	2	1	1	1
NRY118	E-U209	2	0	2	1	0	0		0	4	0	1	16	0	14	2	0	0	2
NRY119	E-U209	2	1	2	0	0	0		0	36	1	1	10	1	3	0	2	0	1
NRY120	E-U209	0	0	3	0	0	0		0	20	4	2	18	1	15	3	6	0	0
NRY121	E-U209	0	0	0	0	4	0		0	20	0	1	12	0	4	0	2	0	0
NRY122	I	0	0	0	1	0	0		0	0	0	0	0	0	0	0	0	1	0
NRY123	I	1	0	0	0	0	0		0	0	0	0	0	0	0	0	0	1	0
NRY124	I	0	0	0	0	0	2		0	0	0	0	0	0	0	0	0	3	0
NRY125	R	1	0	0	0	0	0		0	0	0	0	0	0	0	0	0	0	0
NRY126	R1	0	0	0	1	0	0		0	0	0	0	0	0	0	0	0	0	0
NRY127	R1	1	0	0	0	0	0		0	0	0	0	0	0	0	0	0	0	0
NRY128	R1	0	0	0	0	0	1		0	0	0	0	0	0	0	0	0	0	0
NRY129	R1	0	0	0	0	0	1		0	0	0	0	0	0	0	0	0	5	0
NRY130	R1	0	0	0	0	0	1		0	0	0	0	0	0	0	0	0	0	0
NRY131	R1	0	0	0	0	0	1		0	0	0	0	0	0	0	0	0	0	0
NRY132	R1	0	0	0	0	0	2		0	0	0	0	0	0	0	0	0	0	0
NRY133	R1a1	0	0	0	1	0	0		0	0	0	0	0	0	0	0	0	0	0
NRY134	R1b1	0	0	0	1	0	0		0	0	0	0	0	0	0	0	0	0	0
NRY135	R1b1b	0	0	0	1	0	0		0	0	0	0	0	0	0	0	0	0	0

Table 5.7: NRY haplotype sharing (based on 10STRs) between the Namibian populations and other regions of Sub-Saharan Africa.

Only haplotypes present in the Namibian populations are included. See Supplementary Table 3, Appendix 2 for populations included.

Red = haplotypes unique to a population; green = haplotypes unique to Namibia but shared within Namibia; blue = shared haplotypes.

* 'Khoisan' South here just includes the Mixed 'Khoisan' from Botswana, as other populations only had data available for a limited number of haplogroups.

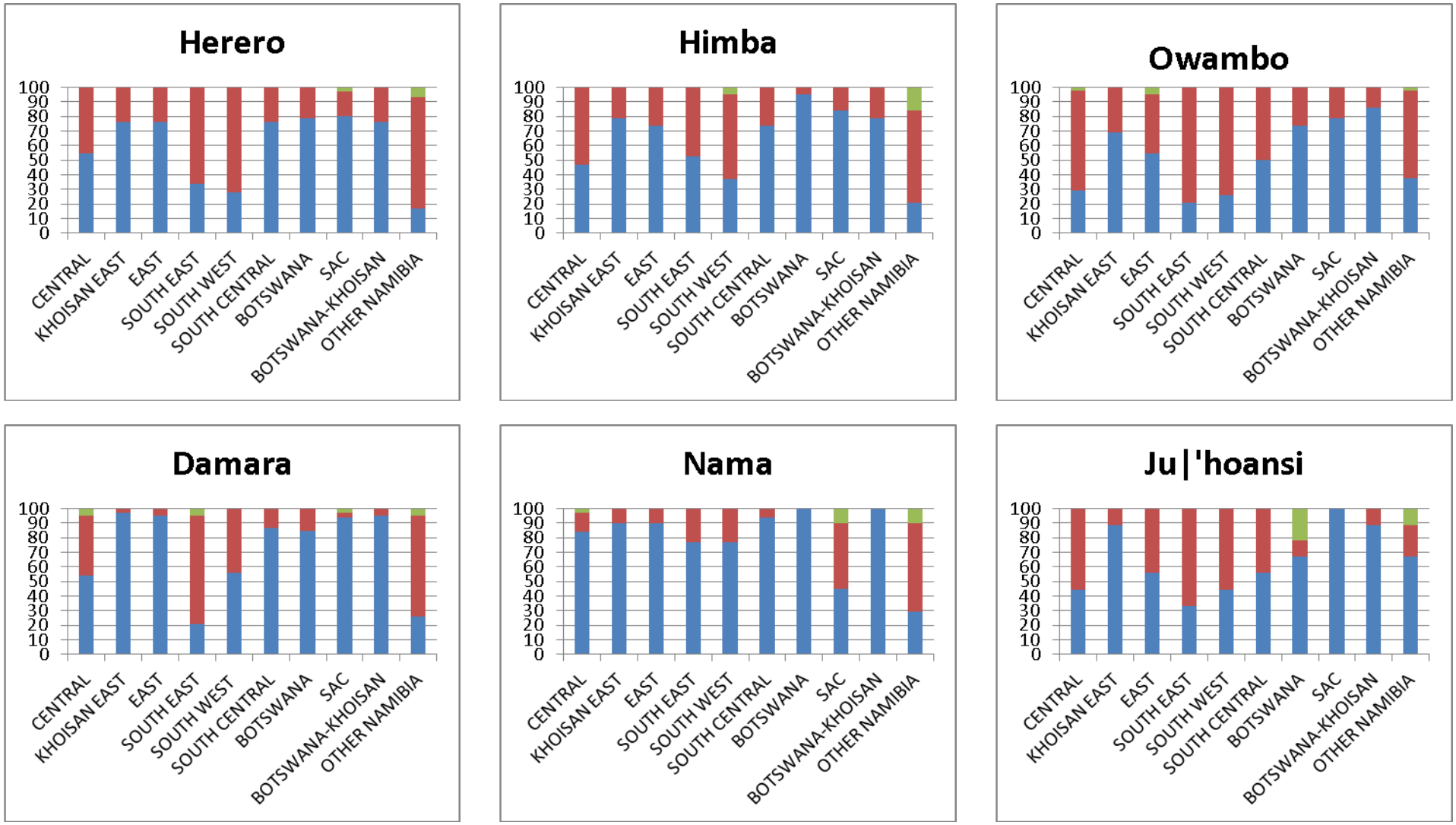


Figure 5.13: NRY lineage sharing between individuals in the six Namibian populations and regions in Sub-Saharan Africa.
Only mtDNAs shared between the Namibian population and at least one other region were included. Regions are the same as in Table 5.7.
Green = exclusive match; red = non-exclusive match; blue = no match.

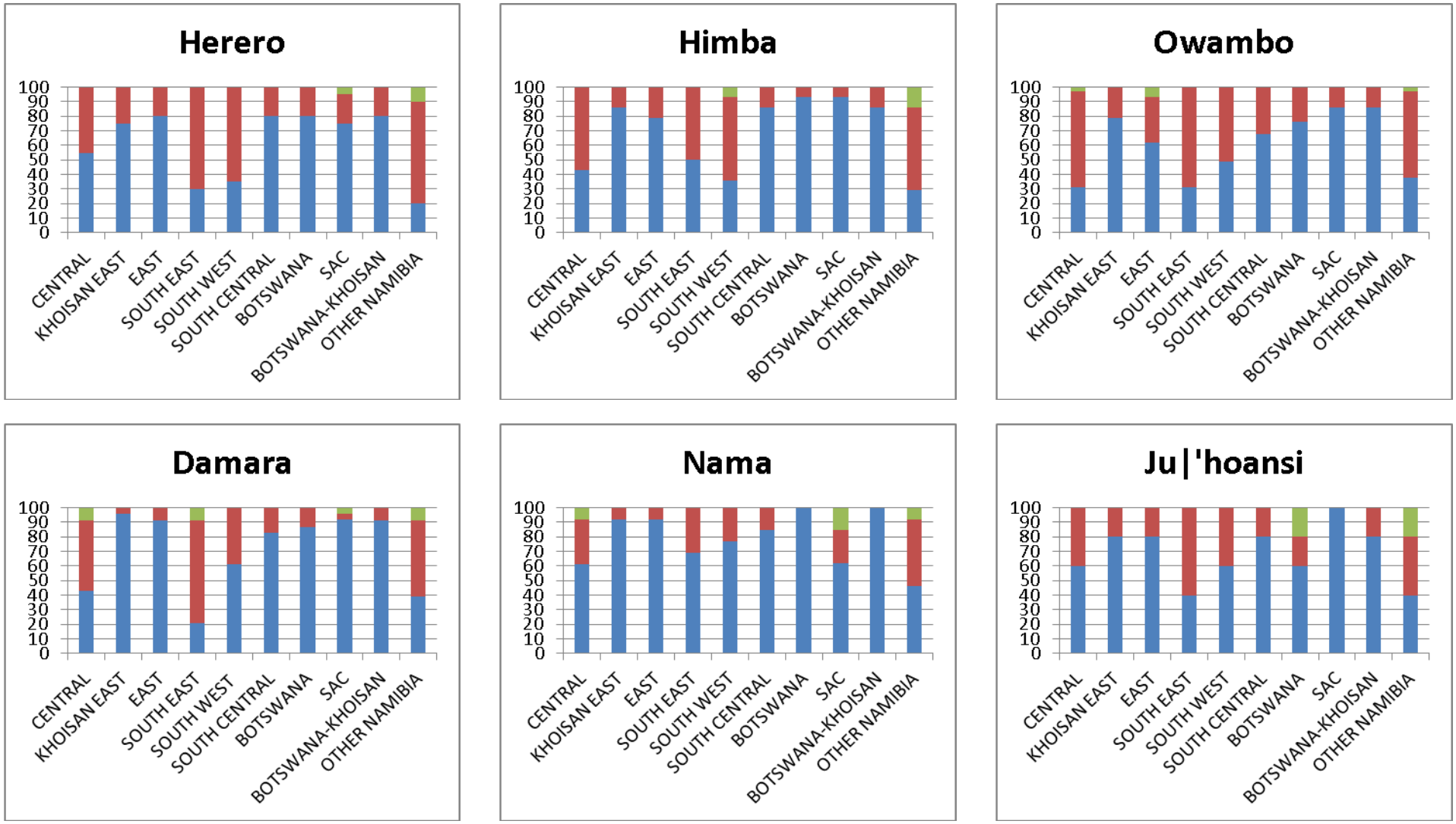


Figure 5.14: NRY lineage sharing between haplotypes in the six Namibian populations and regions in Sub-Saharan Africa.
Only mtDNAs shared between the Namibian population and at least one other region were included. Regions are the same as in Table 5.7.
Green = exclusive match; red = non-exclusive match; blue = no match.

159 of the 272 Namibian individuals analysed have a haplotype shared with populations from outside Namibia: Damara 38/55; Herero 27/39; Himba 16/21; Ju|'hoansi 8/53; Nama 28/54; Owambo 42/50 (Figures 5.13 and 5.14; Table 5.7). As with the mtDNA, this indicates large amounts of sharing with populations across the whole of Sub-Saharan Africa. The majority of this sharing appears to be with Bantu associated E haplogroups (E1b1a7a and E1b1a8a). Of the 60 haplotypes from these two haplogroups, only 19 are unique to Namibia.

Southern Africa

A large proportion of the total haplotypes found in the Namibian samples are shared with populations from Southern Africa, especially Bantu-speaking populations. The percentages (percentage of total haplotypes of that group) of the Namibian haplotypes shared with Southern African Bantu-speaking populations are: Damara 49%; Herero 59%; Himba 75%; Ju|'hoansi 17%; Nama 26% and Owambo 69% (Figures 5.13 and 5.14; Table 5.7). In contrast to the mtDNA, there appears to be more sharing (haplotypes and individuals with those haplotypes) between the Owambo and Damara and South East Africa rather than South West Africa (Figures 5.13 and 5.14). The Damara share unique haplotypes with South East Africa.

Haplotype sharing was analysed in more detail for haplogroup A, a 'Khoisan'-associated haplogroup found at high frequency in the Nama and Ju|'hoansi populations (Table 5.8). A phylogenetic network was constructed for haplogroup A3b1 (15 STRs) (Figure 5.15). For this analysis additional samples were included from Batini et al (2011). The network demonstrates a clear split between the majority of the Nama samples (all but one haplotype with three individuals) (cluster 1) and most of the other samples from the region (cluster 2). The majority of the Nama have a null allele (0 repeats) at DYS448, an allele also found in two Damara samples, three Lesotho, one Zulu (see Chapter Four) and one Xhosa (the Xhosa sample is reported as not defined in Batini et al (2011), but re-genotyping here confirmed it as null).

None of the Ju|'hoansi have this null allele, nor the majority of other samples from the region. Unfortunately, a number of previous studies do not report results for this locus, so it was not possible to include these individuals in the network, for example the SAC individuals (Quintana-Murci et al. 2010). Using the results for haplotype sharing in A3b1 for 10 STRs (Table 5.9) it is possible, however, to see that there are two SAC samples that share an exact match with 10 Nama. These Nama have the null allele for DYS448, and although it cannot be certain that these SAC also have the null allele, it is definitely a possibility.

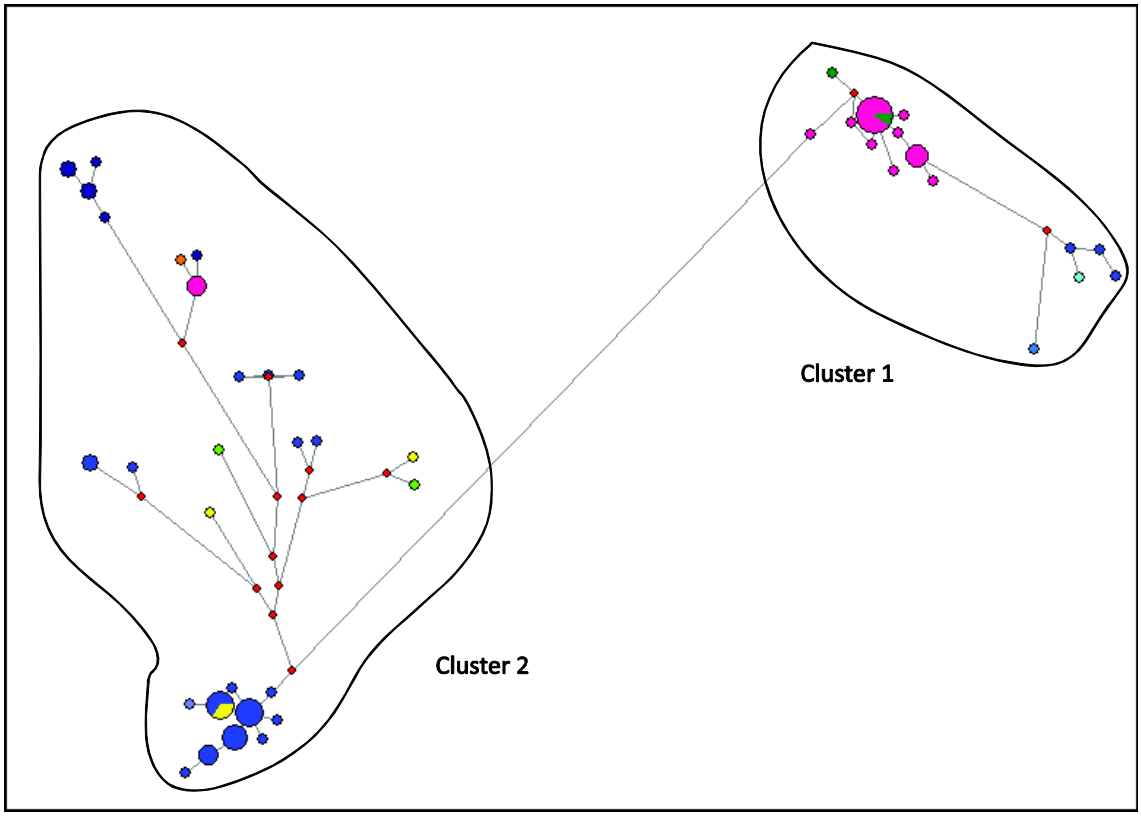


Figure 5.15: Phylogenetic networks for the NRY A3b1 haplotypes – 15 STRs.
Key: Pink = Nama; Purple = Ju|'hoansi; Dark green = Damara; Bright green = Herero; Yellow = Mixed 'Khoisan' Botswana; Orange = Namibia San (REF); Blue = South African Bantu populations (Lesotho, Xhosa, Zulu, Botswana).

Haplotype Name	Haplogroup		Damara[1]	Herero[1]	Himba[1]	Ju|'hoansi[1]	Nama[1]	Owambo[1]		Baka Pygmy[2]	Mixed 'Khoisan', Botswana[2]	Namibia San[2]		Mixed Bantu, Botswana[2]	Lesotho[1]	Xhosa[1,2]	Zulu[1]	SAC[3]	Ndau, Mozambique[2]
A15-1	A2		0	0	0	0	0	0		0	1	0		0	0	0	0	0	0
A15-2	A2		0	0	0	0	0	0		0	4	0		0	0	0	0	0	0
A15-3	A2		0	0	0	0	0	0		0	2	0		0	0	0	0	0	0
A15-4	A2		0	0	0	0	0	0		0	1	0		0	0	0	0	0	0
A15-5	A2		0	0	0	0	0	0		0	1	0		0	0	0	0	0	0
A15-6	A2		0	0	0	0	0	0		0	1	0		0	0	0	0	0	0
A15-7	A2		0	0	0	0	0	0		2	0	0		0	0	0	0	0	0
A15-8	A2		0	0	0	1	0	0		0	0	1		0	0	0	0	0	0
A15-9	A2		0	1	0	0	0	0		0	0	0		0	0	0	0	0	0
A15-10	A2		0	0	0	1	0	0		0	0	0		0	0	0	0	0	0
A15-11	A2		0	0	0	1	0	0		0	0	0		0	0	0	0	0	0
A15-12	A2		0	0	0	2	0	0		0	0	0		0	0	0	0	0	0
A15-13	A2		0	0	0	2	0	0		0	0	0		0	0	0	0	0	0
A15-14	A2		0	0	0	1	0	0		0	0	0		0	0	0	0	0	0
A15-15	A2		0	0	0	2	0	0		0	0	0		0	0	0	0	0	0
A15-16	A2		0	0	0	1	0	0		0	0	0		0	0	0	0	0	0
A15-17	A2		0	0	0	1	0	0		0	0	0		0	0	0	0	0	0
A15-18	A2		0	0	0	1	0	0		0	0	0		0	0	0	0	0	0
A15-19	A3b1		0	0	0	0	0	0		0	1	0		0	0	0	0	0	0
A15-20	A3b1		0	0	0	0	0	0		0	1	0		0	0	0	0	0	0
A15-21	A3b1		0	0	0	0	0	0		0	2	0		0	4	0	0	0	0
A15-22	A3b1		0	0	0	0	0	0		0	0	0		1	0	0	0	0	0
A15-23	A3b1		0	0	0	0	0	0		0	0	0		0	0	1	0	0	0
A15-24	A3b1		0	0	0	0	0	0		0	0	1		0	0	0	0	0	0
A15-25	A3b1		0	0	0	0	0	0		0	0	0		0	6	0	0	0	0
A15-26	A3b1		0	0	0	0	0	0		0	0	0		0	1	0	0	0	0
A15-27	A3b1		0	0	0	0	0	0		0	0	0		0	1	0	0	0	0
A15-28	A3b1		0	0	0	0	0	0		0	0	0		0	2	0	0	0	0
A15-29	A3b1		0	0	0	0	0	0		0	0	0		0	1	0	0	0	0
A15-30	A3b1		0	0	0	0	0	0		0	0	0		0	1	0	0	0	0
A15-31	A3b1		0	0	0	0	0	0		0	0	0		0	1	0	0	0	0
A15-32	A3b1		0	0	0	0	0	0		0	0	0		0	1	0	0	0	0
A15-33	A3b1		0	0	0	0	0	0		0	0	0		0	5	0	0	0	0
A15-34	A3b1		0	0	0	0	0	0		0	0	0		0	1	0	0	0	0
A15-35	A3b1		0	0	0	0	0	0		0	0	0		0	3	0	0	0	0
A15-36	A3b1		0	0	0	0	0	0		0	0	0		0	1	0	0	0	0
A15-37	A3b1		0	0	0	0	0	0		0	0	0		0	1	0	0	0	0
A15-38	A3b1		0	0	0	0	0	0		0	0	0		0	1	0	0	0	0
A15-39	A3b1		0	0	0	0	0	0		0	0	0		0	1	0	0	0	0
A15-40	A3b1		0	0	0	0	0	0		0	0	0		0	1	0	0	0	0
A15-41	A3b1		0	0	0	0	0	0		0	0	0		0	1	0	0	0	0
A15-42	A3b1		0	0	0	0	0	0		0	0	0		0	1	0	0	0	0
A15-43	A3b1		0	0	0	0	0	0		0	0	0		0	0	0	1	0	0
A15-44	A3b1		0	0	0	0	0	0		0	0	0		0	0	0	0	0	1
A15-45	A3b1		1	0	0	0	0	0		0	0	0		0	0	0	0	0	0
A15-46	A3b1		0	1	0	0	0	0		0	0	0		0	0	0	0	0	0
A15-47	A3b1		0	1	0	0	0	0		0	0	0		0	0	0	0	0	0
A15-48	A3b1		0	0	0	1	0	0		0	0	0		0	0	0	0	0	0
A15-49	A3b1		0	0	0	2	0	0		0	0	0		0	0	0	0	0	0
A15-50	A3b1		0	0	0	1	0	0		0	0	0		0	0	0	0	0	0
A15-51	A3b1		0	0	0	1	0	0		0	0	0		0	0	0	0	0	0
A15-52	A3b1		0	0	0	2	0	0		0	0	0		0	0	0	0	0	0
A15-53	A3b1		0	0	0	0	1	0		0	0	0		0	0	0	0	0	0
A15-54	A3b1		0	0	0	0	1	0		0	0	0		0	0	0	0	0	0
A15-55	A3b1		0	0	0	0	1	0		0	0	0		0	0	0	0	0	0
A15-56	A3b1		0	0	0	0	1	0		0	0	0		0	0	0	0	0	0
A15-57	A3b1		0	0	0	0	1	0		0	0	0		0	0	0	0	0	0
A15-58	A3b1		0	0	0	0	1	0		0	0	0		0	0	0	0	0	0
A15-59	A3b1		0	0	0	0	1	0		0	0	0		0	0	0	0	0	0
A15-60	A3b1		0	0	0	0	3	0		0	0	0		0	0	0	0	0	0
A15-61	A3b1		0	0	0	0	4	0		0	0	0		0	0	0	0	0	0
A15-62	A3b1		1	0	0	0	9	0		0	0	0		0	0	0	0	0	0

Table 5.8: Haplotype sharing for NRY haplogroups A2 and A3b1, 15 STRs.

See Materials and Methods (Chapter 2) for STRs included. Red = haplotypes unique to a population; green = haplotypes unique to Namibia but shared within Namibia; blue = shared haplotypes. Source: 1 (This work); 2 (Batini et al. 2011); 3 (Quintana-Murci et al. 2010).

Haplotype Name	Haplogroup		Damara[1]	Herero[1]	Himba[1]	Ju|'hoansi[1]	Nama[1]	Owambo[1]		Mixed 'Khoisan', Botswana[2]	Namibia San[2]	Tsumkwe San[3]		Mixed Bantu, Botswana[2]	Lesotho[1]	Xhosa[1,2]	Zulu[1]	SAC[4]	Ndau Mozambique[2]
A-1	A3b1		0	0	0	0	0	0		1	0	0		0	0	0	0	0	0
A-2	A3b1		0	0	0	0	0	0		1	0	0		0	0	0	0	0	0
A-3	A3b1		0	0	0	0	0	0		2	0	0		0	11	0	0	0	0
A-4	A3b1		0	0	0	0	0	0		0	0	0		1	0	0	0	0	0
A-5	A3b1		0	0	0	0	0	0		0	0	0		0	0	1	0	1	0
A-6	A3b1		0	0	0	0	0	0		0	1	1		0	0	0	0	0	0
A-7	A3b1		0	0	0	3	0	0		0	0	1		0	0	0	0	0	0
A-8	A3b1		0	0	0	0	0	0		0	0	0		0	1	0	0	0	0
A-9	A3b1		0	0	0	0	0	0		0	0	0		0	3	0	0	0	0
A-10	A3b1		0	0	0	0	0	0		0	0	0		0	2	0	0	0	0
A-11	A3b1		0	0	0	0	0	0		0	0	0		0	1	0	0	0	0
A-12	A3b1		0	0	0	0	0	0		0	0	0		0	2	0	0	0	0
A-13	A3b1		0	0	0	0	0	0		0	0	0		0	1	0	0	0	0
A-14	A3b1		0	0	0	0	0	0		0	0	0		0	5	0	0	0	0
A-15	A3b1		0	0	0	0	0	0		0	0	0		0	1	0	0	0	0
A-16	A3b1		0	0	0	0	0	0		0	0	0		0	4	0	0	0	0
A-17	A3b1		0	0	0	0	0	0		0	0	0		0	1	0	0	0	0
A-18	A3b1		0	0	0	0	0	0		0	0	0		0	1	0	0	0	0
A-19	A3b1		0	0	0	0	0	0		0	0	0		0	1	0	0	0	0
A-20	A3b1		0	0	0	0	0	0		0	0	0		0	0	0	1	0	0
A-21	A3b1		1	0	0	0	10	0		0	0	0		0	0	0	0	2	0
A-22	A3b1		0	0	0	0	0	0		0	0	0		0	0	0	0	0	1
A-23	A3b1		0	0	0	0	0	0		0	0	0		0	0	0	0	2	0
A-24	A3b1		0	0	0	0	0	0		0	0	0		0	0	0	0	2	0
A-25	A3b1		0	0	0	0	0	0		0	0	0		0	0	0	0	1	0
A-26	A3b1		0	0	0	0	0	0		0	0	0		0	0	0	0	1	0
A-27	A3b1		0	0	0	0	0	0		0	0	0		0	0	0	0	1	0
A-28	A3b1		0	0	0	0	0	0		0	0	0		0	0	0	0	2	0
A-29	A3b1		1	0	0	0	0	0		0	0	0		0	0	0	0	0	0
A-30	A3b1		0	1	0	0	0	0		0	0	0		0	0	0	0	0	0
A-31	A3b1		0	1	0	0	0	0		0	0	0		0	0	0	0	0	0
A-32	A3b1		0	0	0	1	3	0		0	0	0		0	0	0	0	0	0
A-33	A3b1		0	0	0	1	0	0		0	0	0		0	0	0	0	0	0
A-34	A3b1		0	0	0	2	0	0		0	0	0		0	0	0	0	0	0
A-35	A3b1		0	0	0	0	1	0		0	0	0		0	0	0	0	0	0
A-36	A3b1		0	0	0	0	1	0		0	0	0		0	0	0	0	0	0
A-37	A3b1		0	0	0	0	1	0		0	0	0		0	0	0	0	0	0
A-38	A3b1		0	0	0	0	5	0		0	0	0		0	0	0	0	0	0
A-39	A3b1		0	0	0	0	1	0		0	0	0		0	0	0	0	0	0
A-40	A3b1		0	0	0	0	1	0		0	0	0		0	0	0	0	0	0

Table 5.9: Haplotype sharing for NRY haplogroup A3b1, 10 STRs.
See Materials and Methods (Chapter 2) for STRs included. Red = haplotypes unique to a population; green = haplotypes unique to Namibia but shared within Namibia; blue = shared haplotypes. Source: 1 (This work); 2 (Batini et al. 2011); 3 (The Y Chromosome Consortium 2002); 4 (Quintana-Murci et al. 2010).

Nei's gene diversity values and mean STR variances were calculated for various groupings of A3b1 samples (Table 5.10). Results suggest that the Nama diversity for this haplogroup is low compared to other population groupings. When all 15 STRs are considered the Nama diversity is 0.8221 +/- 0.067, while the diversity for the Southern Bantu group is 0.9612 +/- 0.016, and other 'Khoisan' populations is 0.9545 +/- 0.047. If the multistep DYS448 locus is removed, or if

10 STRs are used in order to include more samples in the analyses, similar results are produced: the Nama diversity is the lowest (Table 5.10). With regards to the STR variances, the Nama variance is lower than that for the other population groupings, however, it is not significantly different, irrespective of which group of STRs is used (Table 5.10). The variances for all but the SAK drop considerably when the DYS448 locus is not included.

Population	15 STRs					14 STRs (removing 448)					10 STRs			
	N	Htypes	Gene diversity (sd)	Variance (CI)		N	Htypes	Gene diversity (sd)	Variance (CI)		N	Htypes	Gene diversity (sd)	Variance (CI)
All A3b1 samples	78	45	0.9667 (0.0097)	6.43 (5.35;6.97)		78	43	0.965 (0.0096)	0.84 (0.71;0.99)		92*	41	0.951 (0.0116)	0.64 (0.52;0.78)
Nama	23	10	0.8221 (0.067)	3.22 (0.065;5.95)		23	10	0.8221 (0.067)	0.39 (0.069;0.67)		23	8	0.7708 (0.071)	0.39 (0.066;0.64)
All other A3b1 (excl Nama)	55	36	0.969 (0.012)	3.47 (1.86;5.06)		55	34	0.9657 (0.012)	0.75 (0.57;0.96)		69*	35	0.9514 (0.016)	0.67 (0.54;0.85)
SAK (excl Nama)	12	9	0.9545 (0.047)	0.63 (0.27;0.81)		12	8	0.9242 (0.058)	0.64 (0.28;0.84)		14	8	0.9011 (0.058)	0.4 (0.15;0.55)
Southern Bantu	43	28	0.9612 (0.016)	3.94 (2.04;5.65)		43	27	0.9579 (0.016)	0.57 (0.38;0.85)		43	22	0.9158 (0.03)	0.51 (0.33;0.75)
Southern Bantu (excl Namibia)	38	23	0.9502 (0.02)	3.24 (0.97;5.05)		38	22	0.9459 (0.02)	0.43 (0.3;0.55)		38	17	0.8919 (0.036)	0.39 (0.26;0.48)
SAC											12	8	0.9394 (0.048)	0.68 (0.42;0.9)

Table 5.10: A3b1 gene diversity.

Analyses were carried out for 15 STRs where possible, and also 10 STRs. Calculations were also carried out removing the multi-step DYS448 locus. See Materials and Methods (Chapter 2) for STRs used.

* Extra populations were included here as more were available for 10 STRs: SAC & Namibian San

See Tables 5.8 and 5.9 for references for the populations used in groupings:

All A3b1 samples: all populations in Table 5.8 (15 STRs) or Table 5.9 (10 STRs)

All other A3b1: all populations in Table 5.8 (15 STRs) or Table 5.9 (10 STRs), excluding Nama

SAK: Ju|'hoansi; Namibian San, Tsumkwe San (10 STRs only), Mixed 'Khoisan' Botswana

Southern Bantu: Mixed Bantu Botswana; Lesotho; Xhosa; Zulu; Ndau Mozambique; Herero Namibia

SAC: South African Coloured (10 STRs only)

East Africa

Haplotype sharing between Namibia and East Africa is considerably less than that with Southern Africa (Damara 5%; Herero 14%; Himba 19%; Ju|'hoansi 4%; Nama 3% and Owambo 31%) (Figures 5.13 and 5.14; Table 5.7). Overall, 13 haplotypes are shared between Namibia and East Africa, but only two of these are not also found in Southern African populations. The Owambo display the largest amount of sharing, the majority of which (9 out of 11), are Bantu-associated haplogroups.

According to Henn et al. (2008), the E sub-haplogroup E1b1b1g-M293 is associated with pastoralists in East Africa, and provides possible evidence of a pastoralist migration between East Africa and Southern Africa due to haplotype sharing between East African samples and !Khwe in Southern Africa. There are nine E1b1b1g-M293 haplotypes (10 STRs) in the Namibian samples analysed here, six of which are unique to Namibia (note, again the lack of SAK samples available, so the uniqueness of these samples is not certain) (Table 5.11). Two of the other three haplotypes are shared with East African 'Khoisan' samples, an East African sample, a Botswana Mixed 'Khoisan' and one with a Lesotho sample. Four samples from other works that had not previously been tested for this SNP were assigned to E-M293 by haplotype inference: one Botswana sample and one Zambia sample from De Filippo et al. (2011) match haplotype M293-1 found in some of the Namibian populations (but not in the Nama); and two SAC (Quintana-Murci et al. 2010) match a Nama haplotype (M293-3) (also found in East Africa) (Table 5.11). The Nama have the highest frequency of E1b1b1g-M293 in Namibia (18.5%) (Table 5.2), supporting the idea that this is a pastoralist marker, however, this frequency is lower than that found in some East African populations (Henn et al. 2008).

Haplotype Name	Herero[1]	Damara[1]	Ju|Hoansi[1]	Nama[1]	SAC[2]	Botswana Mixed 'Khoisan'[1]	Lesotho[1]	Sotho[3]	Zambia[4]	Kalanga, Botswana[4]	Hadza[3]	Sandawe[3]	Datog[3]	Burunge[3]	Mbugwe[3]	Turu[3]	Waifome[3]	Sukuma[3]
M293-1	1	1	1	0	0	1	0	0	1	1	0	3	0	0	0	0	0	0
M293-2	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
M293-3	0	0	0	3	2	0	0	0	0	0	1	0	1	0	0	0	0	0
M293-4	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0
M293-5	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
M293-6	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
M293-7	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
M293-8	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
M293-9	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
M293-10	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0
M293-11	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0
M293-12	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
M293-13	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
M293-14	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
M293-15	0	0	0	0	0	0	0	0	0	0	3	0	1	2	0	0	0	0
M293-16	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
M293-17	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
M293-18	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0
M293-19	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
M293-20	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
M293-21	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0
M293-22	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0
M293-23	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0
M293-24	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
M293-25	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
M293-26	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
M293-27	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0
M293-28	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
M293-29	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0
M293-30	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0
M293-31	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
M293-32	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
M293-33	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
M293-34	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
M293-35	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
M293-36	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
M293-37	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
M293-38	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
M293-39	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
M293-40	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
M293-41	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
M293-42	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
M293-43	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
M293-44	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
M293-45	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
M293-46	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
M293-47	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
M293-48	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Total samples	**1**	**1**	**3**	**9**	**3**	**3**	**2**	**1**	**1**	**1**	**6**	**17**	**17**	**7**	**3**	**3**	**2**	**1**
Total haplotypes	1	1	2	7	2	3	2	1	1	1	4	10	15	6	3	3	2	1

Table 5.11: NRY E-M293 haplotype sharing, 10 STRs.

See Materials and Methods (Chapter 2) for STRs used. Source: 1 (This work); 2 (Quintana-Murci et al. 2010); 3 (Henn et al. 2008); 4 (de Filippo et al. 2011).

Haplotype Name	Herero[1]	Damara[1]	Ju|Hoansi[1]	Nama[1]	Khwe[2]	!Kung[2]	Botswana Mixed 'Khoisan'[1]	SAC[3]	Lesotho[1]	Sotho[2]	Zambia[4]	Kalanga, Botswana[4]	Hadza[2]	Sandawe[2]	Datog[2]	Burunge[2]	Mbugwe[2]	Turu[2]	Waifome[2]	Sukuma[2]	Kenyan Bantu[2]	Ethiopian[2]
M293-1	1	1	1	0	2	0	1	0	0	0	1	1	0	3	0	0	0	0	0	0	0	0
M293-2	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
M293-3	0	0	0	3	0	1	0	2	0	0	0	0	1	0	1	0	0	0	0	0	0	0
M293-4	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
M293-5	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
M293-6	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
M293-7	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
M293-8	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
M293-9	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
M293-10	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0
M293-11	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
M293-12	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
M293-13	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
M293-14	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
M293-15	0	0	0	0	0	0	0	0	0	0	0	0	3	0	1	2	0	0	0	0	0	0
M293-16	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
M293-17	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
M293-18	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0
M293-19	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
M293-20	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
M293-21	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0
M293-22	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0
M293-23	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0
M293-24	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
M293-25	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
M293-26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
M293-27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0
M293-28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
M293-29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0
M293-30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0
M293-31	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
M293-32	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
M293-33	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
M293-34	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
M293-35	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
M293-36	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
M293-37	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
M293-38	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
M293-39	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
M293-40	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
M293-41	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
M293-42	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
M293-43	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
M293-44	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
M293-45	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
M293-46	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
M293-47	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
M293-48	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
M293-49	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
M293-50	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
M293-51	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
M293-52	0	0	0	0	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
M293-53	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
M293-54	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
M293-55	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
M293-56	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
M293-57	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Total samples	**1**	**1**	**3**	**9**	**8**	**7**	**3**	**3**	**2**	**1**	**1**	**1**	**6**	**17**	**17**	**7**	**3**	**3**	**2**	**1**	**1**	**2**
Total haplotypes	1	1	2	7	3	6	3	2	2	1	1	1	4	10	15	6	3	3	2	1	1	2

Table 5.12: NRY E-M293 haplotype sharing, 8 STRs.

See Materials and Methods (Chapter 2) for STRs used.

Source: 1 (This work); 2 (Henn et al. 2008); 3 (Quintana-Murci et al. 2010); 4 (de Filippo et al. 2011).

A phylogenetic network of the E1b1b1g-M293 samples was constructed in order to better visualise the haplotype sharing (Figure 5.16). Here 8 STRs were used in order to allow comparison with the samples used in Henn et al. (2008) – allowing additional South African 'Khoisan' samples to be included (Table 5.12). In fact, even with these additional Khwe and !Kung individuals, a number of the Namibian haplotypes are still unique.

The network shows all the Namibian samples, apart from some Ju|'hoansi, to be within cluster 1 (Figure 5.16). This divide in the network is a result of the DYS389I locus: cluster 2 has an allele repeat size of 13 plus, whilst cluster 1 with the Nama samples in contains individuals derived at this locus, with a repeat size of 10. The majority of the Southern African samples reported in Henn et al. (2008) fall within the 10-repeat cluster (cluster 1), while the 13 plus allele, which Henn suggests is ancestral, is found mostly in Tanzanian populations, and not at all in the Nama (cluster 2).

Gene diversity values were calculated for the Nama for 10 STRs and 8 STRs (to allow additional comparison), and also removing the multistep DYS389-I locus (Table 5.13). Irrespective of the set of STRs used, the Nama have a higher genetic diversity than other Southern African populations (e.g. 9 STRs, Nama: 0.9167 +/- 0.092; Southern Africa all grouped together 0.75 +/- 0.11; Southern African Bantu: 0.5238 +/- 0.21; SAK: 0.7333 +/- 0.16) (Table 5.13), however, their diversity is lower than a number of the East African populations previously reported (Henn et al. 2008). Mean STR variances were also calculated (Table 5.14); the Nama variance is lower than that for the East Africa populations, significantly so for the Datog when 8 and 10 STRs are used, however, when the DYS389-I locus is removed this difference is no longer significant. In fact, for 7 and 9 STRs the Nama actually have a larger variance than some of the East Africa populations (Hadze, Sandawe). For all STR sets, the Nama have a significantly larger variance than the Southern African Bantu.

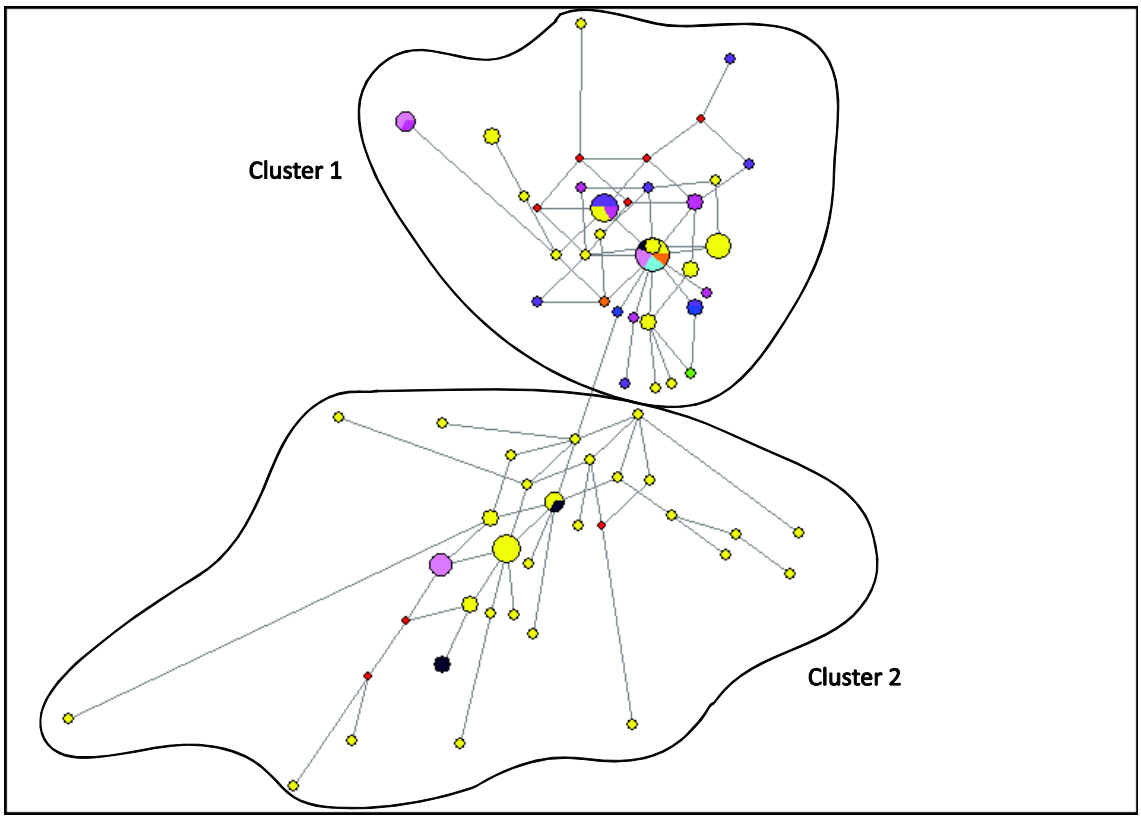


Figure 5.16: Phylogenetic network for NRY haplogroup E1b1b1g-M293.
Populations: Botswana 'Khoisan' (orange); Ju|'hoansi (dark purple); Khwe (light pink); Kung (dark pink); Nama (purple); Lesotho (dark blue); Sotho from Henn et al. (2008) (green); Namibian Bantu (light blue); other Tanzanian populations from Henn et al. (2008) (yellow).

Central Africa

All of the Namibian populations share haplotypes with populations from Central Africa, with 33 out of 38 of these being known Bantu types. The majority of these (33/38) are also shared with Southern populations (Figures 5.13 and 5.14; Table 5.7).

Population	Country	10 STRs			9 STRs (removing 389I)			8 STRs			7 STRs (removing 389I)		
		N	Htypes	Gene diversity (sd)	N	Htypes	Gene diversity (sd)	N	Htypes	Gene diversity (sd)	N	Htypes	Gene diversity (sd)
Herero	Namibia	1	1	1 (0)	1	1	1 (0)	1	1	1 (0)	1	1	1 (0)
Damara	Namibia	1	1	1 (0)	1	1	1 (0)	1	1	1 (0)	1	1	1 (0)
Nama	Namibia	9	7	0.9167 (0.092)	9	7	0.9167 (0.092)	9	7	0.9167 (0.092)	9	7	0.9167 (0.092)
Ju|'hoansi	Namibia	3	2	0.6667 (0.31)	3	2	0.6667 (0.31)	3	2	0.6667 (0.31)	3	2	0.6667 (0.31)
SAC	South Africa	3	2	0.6667 (0.31)	3	2	0.6667 (0.31)	3	2	0.6667 (0.31)	3	2	0.6667 (0.31)
Lesotho	Lesotho	2	2	1 (0.5)	2	2	1 (0.5)	2	2	1 (0.5)	2	2	1 (0.5)
Sotho	South Africa	1	1	1 (0)	1	1	1 (0)	1	1	1 (0)	1	1	1 (0)
Zambia	Zambia	1	1	1 (0)	1	1	1 (0)	1	1	1 (0)	1	1	1 (0)
Kalanga	Botswana	1	1	1 (0)	1	1	1 (0)	1	1	1 (0)	1	1	1 (0)
Botswana Mixed 'Khoisan'	Botswana	3	3	1 (0.27)	3	2	0.6667 (0.31)	3	3	1 (0.2722)	3	2	0.6667 (0.31)
Hadza	Tanzania	6	4	0.8 (0.17)	6	3	0.6 (0.22)	6	4	0.8 (0.17)	6	3	0.6 (0.22)
Sandawe	Tanzania	17	10	0.8971 (0.056)	17	8	0.875 (0.053)	17	10	0.8971 (0.056)	17	8	0.875 (0.053)
Datog	Tanzania	17	15	0.9853 (0.025)	17	13	0.9632 (0.033)	17	15	0.9853 (0.025)	17	13	0.9632 (0.033)
Burunge	Tanzania	7	6	0.9524 (0.096)	7	6	0.9524 (0.096)	7	6	0.9524 (0.096)	7	6	0.9524 (0.096)
Mbugwe	Tanzania	3	3	1 (0.27)	3	3	1 (0.27)	3	3	1 (0.27)	3	3	1 (0.27)
Turu	Tanzania	3	3	1 (0.27)	3	3	1 (0.27)	3	3	1 (0.27)	3	3	1 (0.27)
Waifome	Tanzania	2	2	1 (0.5)	2	2	1 (0.5)	2	2	1 (0.5)	2	2	1 (0.5)
Sukuma	Tanzania	1	1	1 (0)	1	1	1 (0)	1	1	1 (0)	1	1	1 (0)
Kenyan Bantu	Kenya							1	1	1 (0)	1	1	1 (0)
Ethiopia	Ethiopia							2	2	1 (0.5)	2	1	0 (0)
Khwe	Angola							8	3	0.7143 (0.12)	8	3	0.7143 (0.12)
Kung	Angola							7	6	0.9524 (0.096)	7	6	0.9524 (0.096)
Grouped													
Nama		9	7	0.9167 (0.092)	9	7	0.9167 (0.092)	9	7	0.9167 (0.092)	9	7	0.9167 (0.092)
Southern Africa		16	9	0.8583 (0.077)	16	7	0.75 (0.11)	31	15	0.9097 (0.034)	31	13	0.8731 (0.045)
Southern African Bantu		7	4	0.7143 (0.18)	7	3	0.5238 (0.21)	7	4	0.7143 (0.18)	7	3	0.5238 (0.21)
SAK		6	4	0.8667 (0.13)	6	3	0.7333 (0.16)	21	11	0.919 (0.034)	21	10	0.9 (0.039)

Table 5.13: E-M293 gene diversity.

Populations were analysed for 10 STRs where available, otherwise 8 STRs were used. Calculations were also carried out removing the multi-step DYS389I locus. See Materials and Methods (Chapter 2) for which STRs were included. See Tables 5.11 and 5.12 for references.

Groupings: Southern Africa = all Southern African populations excluding Nama; Southern African Bantu = all Southern African Bantu, excluding SAC; SAK = all Southern African 'Khoisan' populations, excluding Nama.

	N	Variance (CI)			
		10STRs	**9STRs**	**8STRs**	**7STRs**
Nama	9	0.19 (0.097;0.27)	0.21 (0.11;0.3)	0.24 (0.12;0.34)	0.27 (0.13;0.38)
Southern Africa	16/31*	0.38 (0.13;0.55)	0.12 (0.049;0.19)	0.59 (0.38;0.77)	0.27 (0.15;0.39)
South African Bantu	7	0.048 (0.014;0.076)	0.042 (0;0.063)	0.06 (0.018;0.095)	0.041 (0;0.082)
SAK	6/21*	0.58 (0.15;0.66)	0.16 (0.019;0.2)	0.73 (0.51;0.89)	0.32 (0.16;0.48)
Hadza	6	0.34 (0.017;0.63)	0.048 (0;0.063)	0.43 (0.021;0.8)	0.062 (0;0.086)
Sandawe	17	0.46 (0.2;0.62)	0.1 (0.057;0.17)	0.55 (0.24;0.77)	0.13 (0.078;0.19)
Datog	17	0.63 (0.42;0.85)	0.27 (0.13;0.47)	0.79 (0.5;1.07)	0.36 (0.17;0.62)
Burunge	7	0.72 (0.17;0.97)	0.33 (0.14;0.47)	0.85 (0.15;1.14)	0.36 (0.13;0.54)

Table 5.14: STR variance for E-M293

Calculated for grouped populations, and for East African populations with 6 or more individuals as in Table 5.13.

*Population size increased for 8 and 7 STR calculations

Discussion

Pastoralism in Namibia

Namibia is one of the few places where Khoe-speaking pastoralists remain in Sub-Saharan Africa, and analysis of samples from the Nama, a Namibian pastoralist population, was conducted in order to try and decipher events surrounding pastoralism in Southern Africa. The history of pastoralism in Southern Africa is not fully understood, and previous archaeological, linguistic and genetic studies have not reached a consensus as to the origins and migration routes of pastoralist populations, if indeed a migration was involved (Fauvelle-Aymar and Sadr 2008). Did pastoralists originate in East Africa and migrate into the south (Guldemann 2008); did they originate further south in the Zambezi region and migrate from there (Elphick 1985); or was there no actual migration of pastoralist peoples, but a diffusion of techniques instead (Sadr 1998)?

A number of recent studies, involving linguistics (Guldemann 2008), comparison of pastoral techniques (Fauvelle-Aymar 2008), and genetics (Henn et al. 2008), have favoured an East Africa origin of pastoralism, with migration carrying the practice southwards. Henn et al. (2008) report a polymorphism (E-M293) which is found at highest frequency in pastoralists in East Africa, and also in the Southern African Khwe, whilst at much lower frequencies in Southern African Bantu populations. This polymorphism has been identified as a possible marker of a pastoralist migration, likely originating in an Eastern African Southern Nilotic population, and being carried into Southern Africa (Henn et al. 2008). Analysis of the Nama demonstrates that they do in fact have a relatively high frequency of this polymorphism (18.5%) (Table 5.2), and that there is haplotype sharing within the E-M293 haplogroup with East African populations (Figure 5.16; Tables 5.11 and 5.12). The E-M293 haplotypes present in the Nama do not, however, contain the DYS389-I allele which is thought to be ancestral (13

repeats). This is not unexpected as it is known that the Nama only arrived in Namibia relatively recently after a migration from the Cape Region of South Africa (Barnard 1992), making it possible that either this was not present in the ancestral group, or that it was lost during migration. The absence of this allele in Southernmost Africa, apart from in the Khwe, supports the latter.

It has been suggested that a population such as the Nilo-Saharan speaking Datog, from Tanzania, may be the source of the polymorphism as they have the highest frequency and highest diversity of E-M293 so far reported (9 STRs: frequency = 0.43; diversity = 0.9632 +- 0.033) (Table 5.13) (Henn et al. 2008). With a frequency of 0.185 and a diversity of 0.9167 +- 0.092 (9 STRs; Table 5.13), the Nama are more diverse than all other Southern African populations (separately and grouped), excluding the !Kung, but are still less diverse (although not significantly so) than the Datog. The high frequency of the mutation in the Nama, and the fact that the level of genetic diversity for this haplogroup in the Nama is more comparable with that of the East African populations rather than the Southern African region, supports the idea of a link between the Nama and East Africa. It is possible that the mutation arose in East Africa and was carried by migration, perhaps via a pastoralist population ancestral to the Nama, into Southern Africa.

Apart from the E-M293 NRY haplotypes, however, there are no other Nama NRY haplotypes shared with East African populations: the majority of Nama haplotypes are unique to the Nama, or are shared with Southern African populations (Figures 5.13 and 5.14; Table 5.7). In fact the majority of Nama Y chromosomes are A3b1 (40.7%), a SAK associated haplogroup. The Ju|'hoansi also have a high frequency of haplogroup A (39.7%) (Table 5.2), however, the composition of this appears very different. While the Ju|'hoansi have almost equal amounts of A2 and A3b1, the Nama have exclusively A3b1, and even the A3b1 is distinct between the two

groups due to a null allele at the DYS448 locus in the Nama (Figure 5.15). This, along with the fact that there is only one NRY shared haplotype (with three Nama individuals and one Ju|'hoansi) suggests that there has been little, perhaps only recent, male gene flow between the populations.

The maternal composition of the Nama, however, actually appears to be more similar to the SAK than to any other Sub-Saharan African population, with the majority of the Nama mtDNA L0d (74.5%), an SAK associated haplogroup. The Nama cluster closest to the Ju|'hoansi and the other SAK populations in the PCA of mtDNA haplogroup frequencies (Figure 5.5), and there are shared mtDNA haplotypes between the populations (four haplotypes: 19 Nama and 19 Ju|'hoansi individuals) (Figures 5.11 and 5.12; Table 5.6). Unlike with the NRY, there seems to be no link with East Africa, with no mtDNA types shared between the Nama and any of the East African populations analysed (Figures 5.11 and 5.12; Table 5.6). The maternal correspondence between the Nama and the Ju|'hoansi is perhaps evidence of the considerable interaction that is believed to have occurred between Khoe populations as they migrated southwards, and the local Ju and Tuu populations (Guldemann 2008). In fact, the further south the Khoe-speakers migrated, the more interaction is thought to have occurred (Guldemann 2008), making it not unlikely that the Nama, a population originating from pastoralists who migrated as far south as possible, would have a considerable proportion of SAK genetics. The Nama in Namibia today are perhaps more representative of the Ju and Tuu populations that the Khoe groups interacted with, than they are of the original pastoralist populations. Interestingly, this interaction may have been gender biased, with little evidence of interaction in the NRY data of the Nama and Ju|'hoansi.

Although evidence for an East African origin is not definitive due to minimal sharing, support for other theories for the introduction of pastoralism to Southern Africa is also not clear. The

Nama do share mtDNA and NRY haplotypes with South East Africa, however, none of these are unique to the two regions, and are shared across the whole of Southern Africa, some with Central Africa too (Figures 5.11, 5.12, 5.13 and 5.14; Tables 5.6 and 5.7). In fact the majority of these haplotypes belong to haplogroups L3 (mtDNA) (Table 5.6) and E-U174 (NRY) (Table 5.7). These are both Bantu-associated haplogroups, suggesting that this sharing might be as a result of later Bantu-migrations rather than an earlier pastoralist one. Arguments for pastoralism being transferred without migration are weakened by the lack of, or very low frequencies of, E-M293 in populations in the regions linking East Africa and Namibia. Seeing as this marker is not widespread, it seems unlikely that populations from two such geographic disparate regions would share haplotypes without there having been some migration to link them.

In fact, further evidence linking the Nama to East African populations has been seen in the analysis of the lactase gene. Work carried out by Jeremy Brown for his undergraduate project (Brown 2011) indicated that the Nama have high levels of 14010*C (32%), an allele associated with lactase persistence which has previously been found only at high levels in East Africa (Tishkoff et al. 2007b). This allele is rare in other populations outside East Africa (Coelho et al. 2009; Torniainen et al. 2009), again providing evidence for some sort of migration linking pastoralists in East Africa with those in Southern Africa.

Overall, results do therefore seem to indicate evidence of past interactions with East African populations (E-M293), although this signal is not as strong as it might be. It seems possible that due to interactions between Khoe groups and Ju and Tuu populations, the Khoe-speaking Nama from southernmost Africa, appear more similar to the SAK than perhaps they originally might have been. This is certainly a possibility for the maternal line, however, the Y chromosome data is less clear: high frequencies of SAK haplogroups are present, but they do not correspond with those of the Ju|'hoansi studied here. It could of course be the case that

the populations responsible for the male A3b1 lineages in the Nama have not yet been sampled, or are perhaps extinct, and were very different to those analysed here, as the genetic diversity of populations in this area is known to be large. It has previously been suggested that a lack of correspondence between haplotypes in East and Southern Africa might be as a result of any migration being minor in scale, with only a small number of individuals moving (Henn et al. 2008).

The Damara

It has already been reported in several studies that the Damara are unlikely to be closely related to the 'Khoisan' populations of Southern Africa, with haplogroup frequencies that appear more Bantu-like rather than 'Khoisan' (Nurse et al. 1976; Soodyall and Jenkins 1993; Guldemann and Stoneking 2008). Indeed, frequencies for the Damara studied here confirm this (Tables 5.1 and 5.2), with the Damara having only 11.9% mtDNA and 3.7% NRY haplogroups associated with 'Khoisan' populations (Davies 2011). PCA plots of both the mtDNA and NRY haplogroups also demonstrate that the Damara do not cluster with SAK populations (Figures 5.5, 5.8 and 5.9). Interestingly, however, the Damara also do not appear to cluster with the Namibian Bantu populations for both mtDNA and NRY analyses (Figures 5.7 and 5.10).

Since a Damara origins shared with the 'SAK' has been discounted, there have been a number of other suggestions as to the origins of the Damara, although as yet there is no overall agreement. Here, analysis of a Damara population and comparison with other Sub-Saharan African populations has attempted to further our understanding of this conundrum, by looking at evidence for four different theories: 1. The Damara are related to the Kwadi, or Proto-Khoe-Kwadi, originating in East Africa (Ehret 1982a); 2. The Damara are an East African population, but are neither Khoe nor Kwadi (Haacke 2008); 3. The Damara are an ancient Bantu population

(Nurse et al. 1976); or 4. The Damara are made up of a mixture of different populations (Jenny (1966) in (Nurse et al. 1976)).

Theories 1 and 2 link the Damara with the Proto-Khoe-Kwadi group involved in a pastoralist migration from East Africa, either being related to this group (Ehret 1982a) (theory 1), or as a non-related East African or South-East African population that joined the group and migrated with them (Haacke 2008) (theory 2). As the Proto-Khoe migrated southwards, the Kwadi, and potentially the Damara, are believed to have branched off and migrated west into Northern Namibian and Southern Angola instead, not undergoing any of the interaction that the Khoe are believed to have had with Ju and Tuu populations (Guldemann 2008). If theory 1 is correct, the Damara might therefore be expected to be most similar to the Kwadi, but, although the Kwadi are known to have been present in Southern Angola until relatively recently (the 1950s) (Barnard 1992), unfortunately there are no longer any known extant Kwadi-speaking populations available with which to compare the Damara. To a lesser extent, it might be anticipated that the Damara would show some similarity to other Khoe-Kwadi populations. As mentioned previously, however, the Damara are definitely not genetically similar to the Nama, with very different haplogroup composition (Tables 5.1 and 5.2); significantly different population pairwise F_{ST} values (Davies 2011); and a lack of correspondence in PCA charts (Figures 5.5, 5.8 and 5.9). It is known, however, that the Nama (and other Khoe) underwent significant interaction with Ju and Tuu groups, whilst the Damara (and Kwadi) would not have (Guldemann 2008), possibly explaining such differences between the Damara and Southern Khoe populations.

Even with these differences between the Damara and other Southern Khoe populations, if the Damara were part of the Proto-Khoe-Kwadi group, then links with East African pastoralist populations that are evident in the Nama (see above) might also be found in the Damara.

Unlike the Nama, however, the Damara only have the E-M293 'pastoralist'-associated NRY marker present at a frequency of 0.0185 (one individual) (Table 5.2), and this is a haplotype found in other Namibian populations and also individuals from Zambia and Botswana (Table 5.11). Additionally, analysis of lactase persistence in the Damara indicates the 14010*C allele to be present at a frequency of only 2.6% (Brown 2011). In fact, Damara haplotype sharing with any East African population for both mtDNA and the NRY is minimal, and any types shared are also present in a number of other populations from other regions (Figures 5.11, 5.12, 5.13 and 5.14; Table 5.6 and 5.7), while PCA plots for mtDNA and NRY do not show the Damara clustering with East Africa (Figures 5.6 and 5.8). This finding also weakens theory 2, that the Damara might be related to 'Negro' populations from East Africa: there seems to be no evidence of links to East Africa, whether with pastoralist or non-pastoralist populations. Another consideration, is that if the Damara did indeed participate in a migration with the Proto-Khoe-Kwadi groups, whether related to them or not, then it would surely be expected that some degree of interaction would have occurred between the groups during this migration. As indicated, however, there is no such evidence of this interaction.

It is possible, instead, that the Damara might not have joined a Proto-Khoe-Kwadi migration until it had already reached South East Africa (theory 2), meaning instead that they might show evidence of links with South East populations. In fact the Damara do fall near the South East populations in the NRY PCA (Figure 5.10), and share a number of mtDNA and NRY haplotypes with South East populations (Figures 5.11, 5.12, 5.13 and 5.14; Tables 5.6 and 5.7); however, only three NRY haplotypes are unique to the Damara and South East populations. The rest of the haplotypes are shared across numerous other regions, making it hard to determine how sharing occurred, especially considering that the majority are Bantu-associated haplogroups which may have been spread by later migrations.

With evidence of links with East Africa minimal and those with South East Africa difficult to determine dates of contact, perhaps theory 3, which suggests that the Damara are instead an ancient Bantu population (Nurse et al. 1976), has more support. If the Damara were indeed one of the first Bantu groups to migrate from the Cameroon/Nigeria region, where the Bantu-language is believed to have originated (Newman 1995), then it would be expected that the Damara might share haplotypes, perhaps even unique ones, with Central African populations. In fact, the Damara do share 13 NRY haplotypes and 6 mtDNA haplotypes with Central populations, however, all but two of these (both NRY) are shared with numerous other populations across the region (Figures 5.11, 5.12, 5.13, 5.14; Tables 5.6 and 5.7). PCA plots also indicate that the Damara are not similar to Central food producing populations (Figures 5.5 and 5.8).

Results for the Damara are therefore still inconclusive, with no obvious relationship emerging with any populations analysed so far. Perhaps this lack of association with a particular population or region is support for theory 4, and the Damara comprise a mixture of populations, merged for convenience, adopting Nama as a lingua franca. Without an increase in the number of populations analysed it is difficult to draw firm conclusions at this stage.

Namibian Bantu populations

The majority of people in Namibia today, and in fact in Sub-Saharan Africa, speak Bantu languages. The history of the Namibian Bantu-speaking populations seems unclear, however, with linguistic classifications not necessarily matching up with previous genetic results. For example, the Owambo speak a South-West Bantu language similar to that of the Herero (Malan 1995), yet previous genetic results suggest that they may be more closely related to South East Bantu populations than they are to the Herero (Soodyall and Jenkins 1993).

Overall, the Owambo have the highest genetic diversity of any of the Namibian Bantu-speaking populations analysed here, for both mtDNA haplotypes and NRY haplotypes (Table 5.3), and although their mtDNA diversity is lower than the Southern African Bantu average, it is considerably higher than that of the Herero and Himba populations. Results here based on haplogroup frequency comparisons suggest that for both mtDNA and NRY the Owambo group slightly more towards South West populations than they do with South East (Figure 5.7 and 5.10). These PCA plots also show, however, that the Owambo definitely do not group with the other Namibian Bantu-speaking populations; they are different from the Herero and Himba. In fact, according to the PCAs, the Herero and Himba also appear to be really quite different from each other, with the Herero even grouping slightly nearer the South East populations for the NRY. The differences between the Herero and Himba could perhaps be explained as a result of the huge population bottleneck that the Herero underwent at the beginning of the 20th century during wars with Nama and European groups. When looking at F_{ST} comparisons with other Sub-Saharan African populations (Table 5.4), for the mtDNA the Owambo are the only Namibian population that is not significantly different from all other populations, with similarities to one Angolan and 10 South East populations, indicating that, although the haplogroup frequencies used in the PCAs suggest an association with South West populations, there is also correspondence with South East populations. The Owambo also have considerable similarity with other populations when considering NRY F_{ST} results: two Angolan, and 11 South East population. Again, the Owambo are the Namibian population with the most similarity to other Southern African populations.

The Owambo are also the Namibian Bantu-speaking population with the most shared haplotypes (Figures 5.11, 5.12, 5.13 and 5.14; Tables 5.6 and 5.7).The Owambo share haplotypes with all regions of Sub-Saharan Africa, but interestingly they share nine mtDNA haplotypes and eleven NRY haplotypes with East African populations. Although the majority of

these haplotypes are also shared across Central and Southern populations too, the Owambo are by far the Namibian population with the most sharing with East Africa. In fact, a number of the NRY haplotypes are unique to the Owambo and East Africa (Figures 5.13 and 5.14; Table 5.7). Even though all of the other Namibian populations share a large number of haplotypes with Central and Southern populations, they share very few with East Africa. This raises questions as to how these haplotypes, which are also found in all of the other regions, are only shared with the Owambo and not the other Namibian populations.

The Owambo are clearly different from the Herero and Himba analysed here, having very different genetic diversities, haplogroup frequencies and haplotype composition. As to the origins of the Owambo, it is slightly more complicated: although PCA results suggest an association with South West populations that speak similar languages; haplotype sharing and F_{ST} results indicate considerable similarity with South East populations. Perhaps this apparently contradictory result can be explained by the Owambo originating in eastern Central Africa, and therefore sharing a history with some of the populations that later migrated eastwards, and then a more recent migration down through Angola into Namibia.

CHAPTER 6 – Conclusion

The work in the preceding chapters has aimed to address the current paucity in genetic studies on populations from Southern Africa. This is a region in which some of the most genetically diverse populations are found (Henn et al. 2011), which might be significant in the advent of modern humans (Tishkoff et al. 2009; Atkinson 2011; Compton 2011), and where interactions between populations practising numerous lifestyles can be studied. As written records are not available for the time period in which many migratory events took place in the region, genetics, along with other fields of study such as archaeology and linguistics, can often be the only way of deciphering the past. As with all such work, it is hard to draw precise conclusions from genetic research alone, but, with the support of data from these other fields a fuller picture can begin to emerge. Despite this, few genetic studies have focused on populations south of Zimbabwe, and those that have use only a small numbers of samples (Schuster et al. 2010), low resolution markers (Nurse et al. 1976; Soodyall and Jenkins 1992) or samples that are not well defined (Mitchell 2010). The improved coverage of Southern African populations provided in this work therefore has the potential to shed new light on the history of people in the region, furthering the understanding of an important area.

The importance of well-defined samples with good geographical and ethnic information is an integral part of this thesis. It is becoming widely accepted that culture has a significant impact on a population's genetic structure (Laland et al. 2010), however, samples from previous studies are often ambiguous with regards to ethnic groups and locations (Mitchell 2010), making accurate comparisons and conclusions difficult. Here, carefully collected data provided an unprecedented opportunity to carry out more detailed analysis on the impact of cultural practices on the genetic structure of populations. In addition to analysis of ethnic groups it was

possible to analyse the impact of practices such as patrilocality and patrilinearity on populations (Chapter 3). Results indicate that in distinct ethnic groups, in two geographically distant countries (Lesotho and Namibia), cultural practices are indeed detectable in biodemographic data, and have an impact on genetic structure. It is not, however, always clear cut; some Namibian populations display unexpected patterns. For example, a number of patrilocal populations actually displayed a higher male than female migration rate. Additionally, the Basotho of Lesotho provided interesting results, with patrilocality detectable in the biodemographic data, but with the genetic data displaying inconsistencies with the expected patterns. The quality of the information collected with these samples allowed further analysis, meaning it was possible to disentangle this puzzling result. Questionnaire data allowed detection of differences in migration distance between the sexes; these were most likely obscuring the genetic patterns normally expected. Results such as this demonstrate the importance of gathering as much information as possible, so as to be able to take into account potential confounding factors. Without information about birthplaces, it would not have been possible to decipher the patterns seen here.

Another overriding factor linking together the populations studied in this thesis is that of gender-biased gene flow. This pattern has been observed in populations from previous studies, particularly in the Pygmy and farmer populations of Central Africa (Cavalli-Sforza 1986; Destro-Bisol et al. 2004b; Wood et al. 2005; Quintana-Murci et al. 2008), with female but not male hunter-gatherer lineages detected in farming populations. Results in Chapters 4 and 5 indicate that this pattern is also a feature across Southern Africa, and is clearly noticeable in the Lesotho, Namibian and South African populations studied here. The levels of female 'Khoisan' introgression are highest in the Lesotho, Xhosa and Zulu populations, although the frequencies of 'Khoisan'-types present in the Namibian Bantu populations are still at a higher level than

other Sub-Saharan Africa countries studied (Pereira et al. 2001; Pereira et al. 2002; Salas et al. 2002; Plaza et al. 2004; Castri et al. 2009; Coelho et al. 2009; de Filippo et al. 2010).

Gene flow in the opposite direction, from Bantu groups into hunter-gatherer or pastoralist groups, is often harder to analyse as a result of the small number of populations available for analysis. Results in Chapter 5 from the Namibian populations do, however, show evidence of some gene flow of Bantu haplogroups into both the Nama and Ju|'hoansi populations for the NRY, and into the Nama for the mtDNA. The Bantu-associated NRY types could be evidence of females returning to their original populations with male children after divorce from a Bantu husband, as has previously been suggested for Pygmy groups (Destro-Bisol et al. 2004b). Although, these levels are relatively low, and the close proximity of the groups today means that very recent admixture cannot be ruled out.

Results in this thesis also provide information about migratory events in the region, helping answer questions with regards to both the arrival of pastoralism and farming in Southern Africa. Chapter 4 concentrates on the manner of the Bantu-speaking farmer arrival into the region. Here, maternal 'Khoisan' introgression in the Lesotho, Xhosa and Zulu populations is particularly interesting due to its magnitude. The frequency is significantly higher than that found in populations in the neighbouring region of South East Africa (Salas et al. 2002; Castri et al. 2009; de Filippo et al. 2010). Such high levels provide support for a frontier hypothesis, whereby climatic and geographical conditions restricted the migration of farming groups, resulting in the creation of a static frontier along the Maloti/ Drakensberg Mountains (Maggs 1980; Alexander 1984; Mitchell et al. 2008). Archaeological research provides evidence of increased interactions between the farming groups to the north of the frontier, and hunter-gatherer groups to the south (Hall 1987; Mitchell and Charles 2000), and genetic results from this work now provide further evidence of significant interactions. Interestingly, the Lesotho

samples also show higher levels of male hunter-gatherer introgression than detected in previous studies (Pereira et al. 2002; de Filippo et al. 2010), although this is significantly less than the female introgression. This could also be explained by increased interactions caused by the frontier, but perhaps also by other cultural practices, such as some Bantu populations making use of San rainmakers, and thus enabling male hunter-gatherers acceptance in the farming communities (Jolly 1996; Prins 2009).

Chapter 5 focuses on understanding the histories of populations in a different region of Southern Africa: Namibia. The migration of pastoralist populations into Southern Africa is discussed with regards to the Nama population. Although results here do not solve the controversy surrounding the arrival of this practice, they do provide further important information. Evidence suggests links between pastoralists in Southern Africa and the East African region, as has previously been suggested (Henn et al. 2008), however, the signal is not as strong as might be expected. Interestingly, the Nama also show considerable levels of interaction with hunter-gatherer groups in the region. This might be obscuring genetic signatures of the region of the pastoralists' origin, but does provide important evidence supporting the theory that the pastoralist groups underwent significant interactions with resident Ju and Tuu groups when they arrived in the region (Guldemann 2008). Analyses for other Namibian populations, are unfortunately, still confusing. The Damara, a population that speak a 'Khoisan' language but share little genetic similarity with extant 'Khoisan' populations, display no obvious relationship with any populations analysed so far. The debate still remains as to whether they could be a very ancient Bantu population, or perhaps a Kwadi population that migrated into the region with the pastoralist groups. The Namibian Bantu populations are equally confusing with regards to their origins; the Owambo, for example, show some similarity with South East Bantu populations, some similarity with South West Bantu populations, but no strong correspondence with either region.

Results such as those for Namibia demonstrate the importance of further work being carried out in Southern Africa. It is impossible to get a full understanding of the diversity present and the history of the region just from studying a small number of populations. Southern Africa is a vast area, and as can be seen from the map of all the populations used in this work (Supplementary Figure 1, Appendix 2), there are still large regions that are unstudied. Samples from hunter-gatherer and pastoralist populations are particularly lacking. Even with the Nama and Ju|'hoansi samples used here, there are few good data sets of SAK populations, especially ones that have accurate information available (Mitchell 2010). Admittedly it is getting increasingly difficult to collect samples from hunter-gatherer and pastoralist populations, as these groups are becoming increasing admixed with surrounding Bantu populations, but this just emphasises the importance and urgency of collecting information about them now.

Focus should also be on ensuring that populations are analysed from both a maternal and paternal perspective, as previously this has not always been the case (Pakendorf et al. 2011). It can be seen from this work that the female and male histories of a population can be very different, and studying just one of the mtDNA or NRY is not sufficient. In fact, it is becoming increasingly feasible to analyse larger sets of data; e.g. SNP assays and whole genome sequencing, and future studies should concentrate on these where possible. Currently only a few whole genome sequences are available in Southern Africa (Schuster et al. 2010), and only a handful of studies use large autosomal datasets (Tishkoff et al. 2009; Henn et al. 2011) meaning comparisons are difficult and conclusions debateable. These studies do all demonstrate, however, the huge diversity in Southern African populations, and therefore the importance of further work in this region in order to better understand the history of humans in Sub-Saharan Africa, and also provide results informative for medicine.

Appendix 1

PARTICIPANT QUESTIONNAIRE

Investigator……..Date……..

Sampling location……Sample Code……..

PARTICIPANT

Sex:……..Date of Birth……Place of Birth…….

Participant's Ethnic group……

Participant's village…….

Participant's clan……

Participant's spoken languages…….

PARTICIPANT'S MOTHER
Place of Birth…… **Ethnic Group……**
Clan…….. **Languages Spoken…….……**

PARTICIPANT'S FATHER
Place of Birth…… **Ethnic Group……**
Clan…….. **Languages Spoken…….……**

PARTICIPANT'S GRANDMOTHER (MOTHER)
Place of Birth…… **Ethnic Group……**
Clan…….. **Languages Spoken…….……**

PARTICIPANT'S GRANDFATHER (MOTHER)
Place of Birth…… **Ethnic Group……**
Clan…….. **Languages Spoken…….……**

PARTICIPANT'S GRANDMOTHER (FATHER)
Place of Birth…… **Ethnic Group……**
Clan…….. **Languages Spoken…….……**

PARTICIPANT'S GRANDFATHER (FATHER)
Place of Birth…… **Ethnic Group……**
Clan…….. **Languages Spoken…….……**

Appendix 2

Populations from previous works used for comparison throughout this thesis.

Population	[a]Country	[b]Region	[c]Language Group	Map Code	Individuals (mtDNA)	Individuals (NRY)	Data used for								Source
							mtDNA haplogroups	NRY haplogroups	mtDNA FST	NRY FST (Haplogroup)	NRY FST (Haplotypes)	Wilcoxon tests	mtDNA Summary statistics	mtDNA mismatch	
Chopi	Moz	SE	SE BAN	MOZ-BAN	27		✓		✓			✓	✓	✓	1
Chwabo	Moz	SE	SE BAN	MOZ-BAN	20		✓		✓			✓	✓	✓	1
Lomwe	Moz	SE	SE BAN	MOZ-BAN	20		✓		✓			✓	✓	✓	1
Makhuwa	Moz	SE	SE BAN	MOZ-BAN	20		✓		✓			✓	✓	✓	1
Makonde	Moz	SE	SE BAN	MOZ-BAN	19		✓		✓			✓	✓	✓	1
Ndau	Moz	SE	SE BAN	MOZ-BAN	19		✓		✓			✓	✓	✓	1
Nguni	Moz	SE	SE BAN	MOZ-BAN	11		✓		✓			✓	✓	✓	1
Nyanja	Moz	SE	SE BAN	MOZ-BAN	20		✓		✓			✓	✓	✓	1
Nyungwe	Moz	SE	SE BAN	MOZ-BAN	20		✓		✓			✓	✓	✓	1
Ronga	Moz	SE	SE BAN	MOZ-BAN	21		✓		✓			✓	✓	✓	1
Sena	Moz	SE	SE BAN	MOZ-BAN	21		✓		✓			✓	✓	✓	1
Shangaan	Moz	SE	SE BAN	MOZ-BAN	22		✓		✓			✓	✓	✓	1
Shona	Moz	SE	SE BAN	MOZ-BAN	17		✓		✓			✓	✓	✓	1
Tonga	Moz	SE	SE BAN	MOZ-BAN	20		✓		✓			✓	✓	✓	1
Tswa	Moz	SE	SE BAN	MOZ-BAN	19		✓		✓			✓	✓	✓	1
Yao	Moz	SE	SE BAN	MOZ-BAN	10		✓		✓			✓	✓	✓	1
General pop	Moz	SE	SE BAN	MOZ-BAN	109		✓		✓			✓	✓	✓	2
Bisa	Zam	SE	SE BAN	ZAE-BAN	42	33	✓	✓	✓	✓	✓	✓	✓	✓	3
Kunda	Zam	SE	SE BAN	ZAE-BAN	36	36	✓	✓	✓	✓	✓	✓	✓	✓	3
Fwe	Zam	SE	SE BAN	ZAW-BAN		30		✓		✓	✓	✓			4
Kwamashi	Zam	SE	SE BAN	ZAW-BAN		27		✓		✓	✓	✓			4
Lozi	Zam	SE	SE BAN	ZAW-BAN		94		✓		✓	✓	✓			4
Luvale	Zam	SE	SE BAN	ZAW-BAN		16		✓		✓	✓	✓			4
Luyana	Zam	SE	SE BAN	ZAW-BAN		61		✓		✓	✓	✓			4
Mbunda	Zam	SE	SE BAN	ZAW-BAN		49		✓		✓	✓	✓			4
Nkoya	Zam	SE	SE BAN	ZAW-BAN		29		✓		✓	✓	✓			4
Shanjo	Zam	SE	SE BAN	ZAW-BAN		29		✓		✓	✓	✓			4
Tonga	Zam	SE	SE BAN	ZAW-BAN		30		✓		✓	✓	✓			4
Totela	Zam	SE	SE BAN	ZAW-BAN		15		✓		✓	✓	✓			4
Mixed Bantu West	Zam	SE	SE BAN	ZAW-BAN		100		✓		✓	✓	✓			4
Shona	Zim	SE	SE BAN	ZIM-BAN	59	49	✓	✓	✓	✓		✓	✓	✓	5
Kalanga	Bots	Sm	SE BAN-CL	BOT-BAN		20		✓		✓	✓	✓			4
Tswana	Bots	Sm	SE BAN-CL	BOT-BAN		20		✓		✓	✓	✓			4
Mixed Bantu	Bots	Sm	SE BAN-CL	BOTX-BAN	50	44	✓	✓	✓	✓	✓	✓			6,7
Basotho (all)	Les	Sm	SE BAN-CL	LES-BAN	287	267	✓	✓	✓	✓	✓	✓	✓	✓	7
Ndebele	Les	Sm	SE BAN-CL	LES-BAN	33	52	✓	✓	✓	✓	✓	✓	✓	✓	7
Thembu	Les	Sm	SE BAN-CL	LES-BAN	24	23	✓	✓	✓	✓	✓	✓	✓	✓	7
Xhosa	RSA	Sm	SE BAN-CL	XHO-BAN	54	57	✓	✓	✓	✓	✓	✓	✓	✓	6,7

Zulu	RSA	Sm	SE BAN-CL	ZUL-BAN	54	54	✓	✓	✓	✓	✓	✓	✓	✓	7
Khwe	Ang*	Sm	SAK	SAK-CHEN	31		✓							✓	8
Kung	Ang*	Sm	SAK	SAK-CHEN	43		✓							✓	8
Kung	Bots/ Nam	Sm	SAK	SAK-VIG	24		✓							✓	9
!Xun-Khwe	RSA (Ang?)*	Sm	SAK	SAK-TISH	17		✓							✓	10
!Kung/ Sekele	Nam*	Sm	SAK	NAM-KHO2		32		✓							11
Tsumkwe San	Nam*	Sm	SAK	NAM-KHO2		29		✓							11
!Kung	RSA (Ang?)*	Sm	SAK	SAK-CRU		64		✓							12
Khwe	RSA (Ang?)*	Sm	SAK	SAK-CRU		26		✓							12
Mixed 'Khoisan'	Ang/ Nam*	Sm	SAK	SAK-UND		39		✓							13
Mixed Bushmen	Bot	Sm	SAK	BOT-KHO	73	66	✓	✓						✓	6,7

Supplementary Table 1: Populations used in analyses in Chapter 4.

* See Mitchell (2010) for a discussion on the ambiguity of these samples.

[a] Ang = Angola; Bots = Botswana; Les = Lesotho; Moz = Mozambique; Nam = Namibia; RSA = South Africa; Zam = Zambia; Zim = Zimbabwe

[b] Sm = Southernmost; SE = South East

[c] SE BAN = South East Bantu; SE BAN-CL = South East Bantu-Click; SAK = Southern African 'Khoisan'

Zambia populations from de Filippo et al. (2011) with > 15 individuals were classed as separate populations, otherwise they were grouped together as Mixed Bantu West.

Sources: 1 (Salas et al. 2002); 2 (Pereira et al. 2001); 3 (de Filippo et al. 2010); 4 (de Filippo et al. 2011); 5 (Castri et al. 2009); 6 (Batini et al. 2011); 7 (This work); 8 (Chen et al. 2000); 9 (Vigilant et al. 1991); 10 (Tishkoff et al. 2007a); 11 (Wood et al. 2005); 12 (Cruciani et al. 2002); 13 (Underhill et al. 2000).

Population Country	Country	[a]Region	[b]Linguistic Affiliation	Map Code	N	Gene diversity average	F_{ST}	PCA	Haplotype sharing	Source
Tuareg	Burkina Faso	W	AFR-AS	BKF-TAU	38				✓	1
Tuareg	Mali	W	AFR-AS	MAL-TAU	21				✓	1
Tuareg	Niger	W	AFR-AS	NIG-TAU	31				✓	1
Baka	Cameroon	Pyg-C	Bantu	CAM-PYG1	30			✓		2
Baka	Cameroon	Pyg-C	Bantu	CAM-PYG1	58			✓		2
Bakola	Cameroon	Pyg-C	Bantu	CAM-PYG1	49			✓		2
Bakola	Cameroon	Pyg-C	Bantu	CAM-PYG1	88			✓		2
Tigar	Cameroon	Pyg-C	Bantu	CAM-PYG1	35			✓		2
Biaka	CAR	Pyg-C	Bantu	CAR-PYG1	56			✓		2
Bakola	Eq Guinea/ Cameroon	Pyg-C	Bantu	CAM-PYG1	30			✓		2
Babongo	Gabon	Pyg-C	Bantu	GAB-PYG1	45			✓		2
Baka	Gabon	Pyg-C	Bantu	GAB-PYG1	39			✓		2
Bakoya	Gabon	Pyg-C	Bantu	GAB-PYG1	31			✓		2
Bakaka	Cameroon	WC	Bantu	CAM-BAN1	50				✓	3
Bamileke	Cameroon	WC	Bantu	CAM-BAN1	48				✓	3
Bassa	Cameroon	WC	Bantu	CAM-BAN1	46				✓	3
Daba	Cameroon	WC	Bantu	CAM-BAN1	20				✓	3
Ewondo	Cameroon	WC	Bantu	CAM-BAN1	53				✓	3
EwondoC	Cameroon	WC	Bantu	CAM-BAN2	25			✓	✓	2
Fali	Cameroon	WC	Bantu	CAM-BAN1	41				✓	3
FangC	Cameroon	WC	Bantu	CAM-BAN2	39			✓	✓	2
Fulbe	Cameroon	WC	Bantu	CAM-BAN1	34				✓	3
Mandara	Cameroon	WC	Bantu	CAM-BAN1	37				✓	3
NgumbaC	Cameroon	WC	Bantu	CAM-BAN2	88			✓	✓	2
Podoko	Cameroon	WC	Bantu	CAM-BAN1	39				✓	3
Tali	Cameroon	WC	Bantu	CAM-BAN1	20				✓	3
Tupuri	Cameroon	WC	Bantu	CAM-BAN1	25				✓	3
Uldeme	Cameroon	WC	Bantu	CAM-BAN1	28				✓	3
Akele	Gabon	WC	Bantu	GAB-BAN1	48			✓	✓	2
Ateke	Gabon	WC	Bantu	CAM-BAN1	54			✓	✓	2
Benga	Gabon	WC	Bantu	CAM-BAN1	50			✓	✓	2
Duma	Gabon	WC	Bantu	CAM-BAN1	47			✓	✓	2
Eshira	Gabon	WC	Bantu	CAM-BAN1	40			✓	✓	2
Eviya	Gabon	WC	Bantu	CAM-BAN1	38			✓	✓	2
FangG	Gabon	WC	Bantu	CAM-BAN1	66			✓	✓	2
Galoa	Gabon	WC	Bantu	CAM-BAN1	51			✓	✓	2
Kota	Gabon	WC	Bantu	CAM-BAN1	56			✓	✓	2
Makina	Gabon	WC	Bantu	CAM-BAN1	45			✓	✓	2
Mitsogo	Gabon	WC	Bantu	CAM-BAN1	64			✓	✓	2
Ndumu	Gabon	WC	Bantu	CAM-BAN1	39			✓	✓	2
Nzebi	Gabon	WC	Bantu	CAM-BAN1	63			✓	✓	2
Obama	Gabon	WC	Bantu	CAM-BAN1	47			✓	✓	2
Orungu	Gabon	WC	Bantu	CAM-BAN1	20			✓	✓	2
Punu	Gabon	WC	Bantu	CAM-BAN1	52			✓	✓	2
Shake	Gabon	WC	Bantu	CAM-BAN1	51			✓	✓	2
Mbuti	DRC	Pyg-E	N-S	DRC-PYG	39			✓		2
Hadza	Tanzania	E	'Khoisan'	TZ-KHO	79		✓	✓	✓	16
Sandawe	Tanzania	E	'Khoisan'	TZ-KHO	82		✓	✓	✓	16
Afar	Ethiopia	E	AFR-AS	ETH-AA	16				✓	4
Amhara	Ethiopia	E	AFR-AS	ETH-AA	120				✓	4
Eritrea	Ethiopia	E	AFR-AS	ETH-AA	8				✓	4
Gurage	Ethiopia	E	AFR-AS	ETH-AA	21				✓	4
Oromo	Ethiopia	E	AFR-AS	ETH-AA	33				✓	4
Tigrai	Ethiopia	E	AFR-AS	ETH-AA	45				✓	4

Kenya	Kenya	E	Mixed	KEN-MXD	99				✓	5
Hutu	Rwanda	E	Bantu	RWA-BAN	42			✓	✓	6
Burunge	Tanzania	E	AFR-AS	TZ-AA	38		✓	✓	✓	16
Datog	Tanzania	E	N-S	TZ-NS	39		✓	✓	✓	16
Sukuma	Tanzania	E	Bantu	TZ-BAN	11		✓	✓	✓	16
Turu	Tanzania	E	Bantu	TZ-BAN	29		✓	✓	✓	16
Mixed 'Khoisan'	Botswana	Sm	SAK	BOT-KHO	73	✓	✓	✓	✓	7
Damara	Namibia	SW	SAK	NAM-KHO	66	✓	✓	✓	✓	7, 14
Ju|'hoansi	Namibia	SW	SAK	NAM-KHO	58	✓	✓	✓	✓	7
Nama	Namibia	SW	SAK	NAM-KHO	57	✓	✓	✓	✓	7
!Xun-Khwe	South Africa (Angola?)*	Sm	SAK	SAK-TISH	17	✓	✓	✓	✓	16
Khwe	South Africa (Angola)*	Sm	SAK	SAK-CHEN	31	✓	✓	✓	✓	8
Kung	South Africa (Angola)*	Sm	SAK	SAK-CHEN	43	✓	✓	✓	✓	8
Ganguela	Angola	SW	SW BAN	ANG-BAN	21	✓	✓	✓	✓	9
Kuvale	Angola	SW	SW BAN	ANG-BAN	54	✓	✓	✓	✓	9
Mbundu	Angola	SW	SW BAN	ANG-BAN2	43	✓	✓	✓	✓	10
Nyaneka-Nkhumbi	Angola	SW	SW BAN	ANG-BAN	153	✓	✓	✓	✓	9
Ovimbundo	Angola	SW	SW BAN	ANG-BAN	92	✓	✓	✓	✓	9
Herero	Namibia	SW	SW BAN	NAM-BAN	44	✓	✓	✓	✓	7
Himba	Namibia	SW	SW BAN	NAM-BAN	51	✓	✓	✓	✓	7, 14
Owambo	Namibia	SW	SW BAN	NAM-BAN	79	✓	✓	✓	✓	7
Chopi	Mozambique	SE	SE BAN	MOZ-BAN	27	✓	✓	✓	✓	11
Chwabo	Mozambique	SE	SE BAN	MOZ-BAN	20	✓	✓	✓	✓	11
Lomwe	Mozambique	SE	SE BAN	MOZ-BAN	20	✓	✓	✓	✓	11
Makhuwa	Mozambique	SE	SE BAN	MOZ-BAN	20	✓	✓	✓	✓	11
Makonde	Mozambique	SE	SE BAN	MOZ-BAN	19	✓	✓	✓	✓	11
Mozambicans	Mozambique	SE	SE BAN	MOZ-BAN	109	✓	✓	✓	✓	12
Ndau	Mozambique	SE	SE BAN	MOZ-BAN	19	✓	✓	✓	✓	11
Nguni	Mozambique	SE	SE BAN	MOZ-BAN	11	✓	✓	✓	✓	11
Nyanja	Mozambique	SE	SE BAN	MOZ-BAN	20	✓	✓	✓	✓	11
Nyungwe	Mozambique	SE	SE BAN	MOZ-BAN	20	✓	✓	✓	✓	11
Ronga	Mozambique	SE	SE BAN	MOZ-BAN	21	✓	✓	✓	✓	11
Sena	Mozambique	SE	SE BAN	MOZ-BAN	21	✓	✓	✓	✓	11
Shangaan	Mozambique	SE	SE BAN	MOZ-BAN	22	✓	✓	✓	✓	11
Shona	Mozambique	SE	SE BAN	MOZ-BAN	17	✓	✓	✓	✓	11
Tonga	Mozambique	SE	SE BAN	MOZ-BAN	20	✓	✓	✓	✓	11
Tswa	Mozambique	SE	SE BAN	MOZ-BAN	19	✓	✓	✓	✓	11
Yao	Mozambique	SE	SE BAN	MOZ-BAN	10	✓	✓	✓	✓	11
Bisa	Zambia	SE	SE BAN	MOZ-BAN	42	✓	✓	✓	✓	13
Kunda	Zambia	SE	SE BAN	MOZ-BAN	36	✓	✓	✓	✓	13
Shona	Zimbabwe	SE	SE BAN	ZIM-BAN	59	✓	✓	✓	✓	6
Mixed Bantu	Botswana	Sm	SE BAN-CL	BOTX-BAN	50	✓	✓	✓	✓	7
Basotho	Lesotho	Sm	SE BAN-CL	LES-BAN	287	✓	✓	✓	✓	7, 15
Ndebele	Lesotho	Sm	SE BAN-CL	LES-BAN	33	✓	✓	✓	✓	7, 15
Thembu	Lesotho	Sm	SE BAN-CL	LES-BAN	24	✓	✓	✓	✓	7, 15
Xhosa	South Africa	Sm	SE BAN-CL	XHO-BAN	54	✓	✓	✓	✓	7
Zulu	South Africa	Sm	SE BAN-CL	ZUL-BAN	54	✓	✓	✓	✓	7

Supplementary Table 2: Populations used in mtDNA comparisons in Chapter 5.

* See Mitchell (2010) for a discussion on the ambiguity of these samples.

[a] E = East; Pyg-C = Pygmy Central; Pyg-E = Pygmy East; Sm = Southernmost; SE = South East; SW = South West; W = West; WC = West Central

[b] AFR-AS = Afro-Asiatic; N-S = Nilo-Saharan; SE BAN = South East Bantu; SE BAN-CL = South East Bantu-Click; SAK = Southern African 'Khoisan'; SW BAN = South West Bantu

Sources: 1 (Pereira et al. 2010); 2 (Quintana-Murci et al. 2008); 3 (Coia et al. 2009); 4 (Kivisild et al. 2004); 5 (Brandstätter et al. 2004); 6 (Castri et al. 2009); 7 (This work); 8 (Chen et al. 2000); 9 (Coelho et al. 2009); 10 (Plaza et al. 2004); 11 (Salas et al. 2002); 12 (Pereira et al. 2001); 13 (de Filippo et al. 2010); 14 (Davies 2011); 15 (Levy 2010); 16 (Tishkoff et al. 2007a)

Population	Country	Region	Linguistic affiliation	Map Code	N	Gene diversity average (HG)	Gene diversity average (Htype)	F_{ST}	PCA	Haplotype sharing	Source
Ga	Ghana	W	N-CON	GHA-GA	29				✓		1
Yoruba	Nigeria	W	N-CON	NIG-YOR	13				✓	✓	2
BakaC	Cameroon	Pyg-C	Bantu	CAM-PYG2	5				✓	✓	3
Bakola Pyg	Cameroon	Pyg-C	Bantu	CAM-PYG3	33				✓		1
BakolaC	Cameroon	Pyg-C	Bantu	CAM-PYG2	22				✓	✓	3
Baka Pyg	CAR	Pyg-C	Bantu	CAR-PYG2	18				✓		1
Biaka Pyg	CAR	Pyg-C	Bantu	CAR-PYG2	51				✓		1
Mbenzele Pyg	CAR	Pyg-C	Bantu	CAR-PYG3	46				✓		4
Baka	Gabon	Pyg-C	Bantu	GAB-PYG2	33				✓	✓	3
Bakaka	Cameroon	WC	Bantu	CAM-BAN4	17				✓		1
NgumbaC	Cameroon	WC	Bantu	CAM-BAN3	24				✓	✓	3
Akele	Gabon	WC	Bantu	GAB-BAN2	50				✓	✓	3
Bekwil	Gabon	WC	Bantu	GAB-BAN2	5				✓	✓	3
Benga	Gabon	WC	Bantu	GAB-BAN2	48				✓	✓	3
Duma	Gabon	WC	Bantu	GAB-BAN2	46				✓	✓	3
Eshira	Gabon	WC	Bantu	GAB-BAN2	42				✓	✓	3
Eviya	Gabon	WC	Bantu	GAB-BAN2	24				✓	✓	3
Fang	Gabon	WC	Bantu	GAB-BAN2	60				✓	✓	3
Galoa	Gabon	WC	Bantu	GAB-BAN2	47				✓	✓	3
Kota	Gabon	WC	Bantu	GAB-BAN2	53				✓	✓	3
Makina	Gabon	WC	Bantu	GAB-BAN2	43				✓	✓	3
Ndumu	Gabon	WC	Bantu	GAB-BAN2	36				✓	✓	3
Nzebi	Gabon	WC	Bantu	GAB-BAN2	57				✓	✓	3
Obamba	Gabon	WC	Bantu	GAB-BAN2	47				✓	✓	3
Okande	Gabon	WC	Bantu	GAB-BAN2	6				✓	✓	3
Orungu	Gabon	WC	Bantu	GAB-BAN2	21				✓	✓	3
Punu	Gabon	WC	Bantu	GAB-BAN2	58				✓	✓	3
Shake	Gabon	WC	Bantu	GAB-BAN2	43				✓	✓	3
Teke	Gabon	WC	Bantu	GAB-BAN2	48				✓	✓	3
Tsogo	Gabon	WC	Bantu	GAB-BAN2	60				✓	✓	3
Mbuti Pyg	DRC	Pyg-E	Sudanic	DRC-PYG	47				✓		1
Hadza	Tanzania	E	'Khoisan'	TZ-KHO	77			✓	✓	✓	2
Sandawe	Tanzania	E	'Khoisan'	TZ-KHO	68			✓	✓	✓	2
Nanda	DRC	E	Bantu	DRC-BAN	18				✓		1
Kikuyu & Kamba	Kenya	E	Bantu	KEN-BAN	42				✓		1
Mbugwe	Tanzania	E	Bantu	TZ-BAN	15			✓		✓	2
Burunge	Tanzania	E	AFR-AS	TZ-AA	23			✓	✓	✓	2
Datoga	Tanzania	E	N-S	TZ-NS	31			✓	✓	✓	2
Sukuma	Tanzania	E	Bantu	TZ-BAN	32			✓	✓	✓	2
Turu	Tanzania	E	Bantu	TZ-BAN	20			✓	✓	✓	2
Ganda	Uganda	E	Bantu	UGA-BAN	26				✓		1
Karamoja	Uganda	E	N-S	UGA-NS	118				✓	✓	5
Mixed Bantu-Khoe	Botswana	Sm	SAK	BOT-KHO	66				✓	✓	6, 7
!Kung/Sekele	Namibia*	SW	SAK	NAM-KHO2	32	✓			✓		1
Nama	Namibia	SW	SAK	NAM-KHO2	11				✓		1
Tsumkwe San	Namibia*	SW	SAK	NAM-KHO2	29	✓			✓		1
Damara	Namibia	SW	SAK	NAM-KHO2	18				✓		1
Damara	Namibia	SW	SAK	NAM-KHO	54	✓	✓	✓	✓	✓	7, 17
Ju|'hoansi	Namibia	SW	SAK	NAM-KHO	58	✓	✓	✓	✓	✓	7
Nama	Namibia	SW	SAK	NAM-KHO	54	✓	✓	✓	✓	✓	7
!Kung	South Africa (Angola)*	Sm	SAK	SAK-CRU	64	✓			✓		8

Khwe	South Africa (Angola)*	Sm	SAK	SAK-UND	26	✓			✓		8
Mixed 'Khoisan'	South Africa (Angola)*	Sm	SAK	SAK-UND	39	✓			✓		9
Ganguela	Angola	SW	SW BAN	ANG-BAN	11	✓	✓	✓	✓	✓	10
Kuvale Herero	Angola	SW	SW BAN	ANG-BAN	26	✓	✓	✓	✓	✓	10
Nyaneka-Nkhumbi	Angola	SW	SW BAN	ANG-BAN	75	✓	✓	✓	✓	✓	10
Ovimbundu	Angola	SW	SW BAN	ANG-BAN	96	✓	✓	✓	✓	✓	10
Ambo	Namibia	SW	SW BAN	NAM-BAN2	22	✓			✓		1
Herero	Namibia	SW	SW BAN	NAM-BAN2	24	✓			✓		1
Herero	Namibia	SW	SW BAN	NAM-BAN	42	✓	✓	✓	✓	✓	7
HImba	Namibia	SW	SW BAN	NAM-BAN	50	✓	✓	✓	✓	✓	7, 17
Owambo	Namibia	SW	SW BAN	NAM-BAN	87	✓	✓	✓	✓	✓	7
Bisa	Zambia	SE	SE BAN	ZAE-BAN	33	✓	✓	✓	✓	✓	11
Fwe	Zambia	SE	SE BAN	ZAW-BAN	30		✓	✓	✓	✓	12
Kunda	Zambia	SE	SE BAN	ZAE-BAN	36	✓	✓	✓	✓	✓	11
Kwamashi	Zambia	SE	SE BAN	ZAW-BAN	27		✓	✓	✓	✓	12
Lozi	Zambia	SE	SE BAN	ZAW-BAN	94		✓	✓	✓	✓	12
Luvale	Zambia	SE	SE BAN	ZAW-BAN	16		✓	✓	✓	✓	12
Luyana	Zambia	SE	SE BAN	ZAW-BAN	61		✓	✓	✓	✓	12
Mbuna	Zambia	SE	SE BAN	ZAW-BAN	49		✓	✓	✓	✓	12
Mixed Bantu West	Zambia	SE	SE BAN	ZAW-BAN	100		✓	✓	✓	✓	12
Nkoya	Zambia	SE	SE BAN	ZAW-BAN	29		✓	✓	✓	✓	12
Shanjo	Zambia	SE	SE BAN	ZAW-BAN	29		✓	✓	✓	✓	12
Tonga	Zambia	SE	SE BAN	ZAW-BAN	30		✓	✓	✓	✓	12
Totela	Zambia	SE	SE BAN	ZAW-BAN	15		✓	✓	✓	✓	12
Shona	Zimbabwe	SE	SE BAN	ZAW-BAN	49	✓			✓		13
Kalanga	Botswana	Sm	SE BAN-CL	BOT-BAN	20		✓	✓	✓	✓	12
Mixed Bantu	Botswana	Sm	SE BAN-CL	BOTX-BAN	44			✓	✓	✓	6, 7
Tswana	Botswana	Sm	SE BAN-CL	BOT-BAN	20		✓	✓	✓	✓	12
Basotho	Lesotho	Sm	SE BAN-CL	LES-BAN	267	✓	✓	✓	✓	✓	7
Ndebele	Lesotho	Sm	SE BAN-CL	LES-BAN	52	✓	✓	✓	✓	✓	7
Thembu	Lesotho	Sm	SE BAN-CL	LES-BAN	23	✓	✓	✓	✓	✓	7
SAC	South Africa	Sm	SE BAN-CL	SA-SAC	228				✓		14
Sotho-Tswana	South Africa	Sm	SE BAN-CL	SA-BAN	28	✓			✓		1
Xhosa	South Africa	Sm	SE BAN-CL	SA-BAN	80	✓			✓		1
Xhosa (Cape Town)	South Africa	Sm	SE BAN-CL	XHO-BAN	57	✓	✓	✓	✓	✓	6, 7, 15, 16
Zulu	South Africa	Sm	SE BAN-CL	SA-BAN	29	✓			✓		1
Zulu (Cape Town)	South Africa	Sm	SE BAN-CL	ZUL-BAN	54	✓	✓	✓	✓	✓	7

Supplementary Table 3: Populations used in NRY comparisons in Chapter 5.

* See Mitchell (2010) for a discussion on the ambiguity of these samples.

[a] E = East; Pyg-C = Pygmy Central; Pyg-E = Pygmy East; Sm = Southernmost; SE = South East; SW = South West; W = West; WC = West Central

[b] AFR-AS = Afro-Asiatic; N-CON = Niger-Congo; N-S = Nilo-Saharan; SE BAN = South East Bantu; SE BAN-CL = South East Bantu-Click; SAK = Southern African 'Khoisan'; SW BAN = South West Bantu

Zambia populations from de Filippo et al. (2011) with > 15 individuals were classed as separate populations, otherwise they were grouped together as Mixed Bantu West.

Sources: 1 (Wood et al. 2005); 2 (Tishkoff et al. 2007a); 3 (Berniell-Lee et al. 2009); 4 (Coia et al. 2005); 5 (Gomes et al. 2010); 6 (Batini et al. 2011); 7 (This work); 8 (Cruciani et al. 2002); 9 (Underhill et al. 2000); 10 (Coelho et al. 2009); 11 (de Filippo et al. 2010); 12 (de Filippo et al. 2011); 13 (Castri et al. 2009); 14 (Quintana-Murci et al. 2010); 15 (Leat et al. 2004); 16 (Leat et al. 2007); 17 (Davies 2011).



Supplementary Figure 1: Map showing locations of populations used in this work. Codes for populations are given in Supplementary Tables 1, 2 and 3. Red = samples genotyped in this work; green = populations from other studies.
The Xhosa and Zulu samples used in this work (XHO-BAN and ZUL-BAN) were collected in Cape Town but have been placed on the map in the locations where these populations are originally from. The majority of the SAK samples have been placed in Angola rather than South Africa, as some studies indicate, as these samples are originally from the Angola region (see Mitchell 2010).

Contributors

The following contributed towards various aspects of the work presented here and will be included as co-authors on any papers resulting from this work.

Chapter 3: Cultural impact on genetic diversity

- **Hila Levy** (Department of Zoology, University of Oxford): Extracted 195 of the 379 Lesotho samples and genotyped these samples for mtDNA hypervariable region 1.
- **Conrado Martinez-Cadenas** (Department of Zoology, University of Oxford): Provided Spanish data for comparison with the Lesotho data.
- **Francesco Montinaro** (Institute of Legal Medicine, Catholic University, Rome, Italy): Carried out simulations.

Chapter 4: Frontiers and Southern African Bantu-speaking populations

- **Hila Levy** (Department of Zoology, University of Oxford): See above. As part of her MSc thesis Hila presented the Lesotho Lowland mtDNA results, comparing them with the Lesotho samples that I genotyped and analysed (184). Analyses presented in this chapter contain Hila's data, but all analyses have either been redone including larger datasets which I constructed, or are entirely new analyses. Additionally I include NRY data for the Lesotho samples, for which I conducted all of the genotyping (379 samples).
- **Professor Peter Mitchell** (School of Archaeology, University of Oxford & School of Geography, Archaeology and Environmental Studies, University of the Witwatersrand): Provided information with regards to organising sampling by suggesting suitable areas to sample in, and suggested relevant literature to read.

- **Francesca Brisighelli** (Department of Zoology, University of Oxford): Helped with extraction of Botswana blood samples and creating NRY SNP multiplexes used to genotype samples.
- **Francesco Montinaro** (Institute of Legal Medicine, Catholic University, Rome, Italy): Carried out simulations.

The following helped organise and conduct sampling in Lesotho:

- **Charles Arthur** (St Hugh's College, University of Oxford)
- **Dr Chiara Batini** (Department of Genetics, University of Leicester)
- **George Busby** (Department of Zoology, University of Oxford)
- **Dr Cristian Capelli** (Department of Zoology, University of Oxford)
- **Dr Brian Stewart** (McDonald Institute for Archaeological Research, University of Cambridge)

The following provided samples used in this chapter:

- **Dr Maria Eugenia D'Amato and Professor Sean Davison** (Forensic DNA Lab, University of the Western Cape, South Africa): Provided Xhosa and Zulu DNA
- **Peter Ebbesen** (Institute of Health Science, University of Aalborg, Denmark): Provided Botswana serum samples

Chapter 5: Namibia

- **Nicholas Davies** (Department of Zoology, University of Oxford): Extracted two out of six of the Namibian populations (Damara: 72; Himba: 50) and genotyped these two populations for mtDNA hypervariable region 1, Y SNPs and Y STRs. As part of his MSc thesis he presented the results for these two populations and compared them to two populations genotyped and analysed by myself (Herero: 45; Nama 54). Results

presented in Chapter 5 include these populations analysed and genotyped by Nick, but all analyses included are my own, using larger population sets and new data. I additionally include data for two further Namibian populations genotyped and analysed by myself (Ju|'hoansi: 58; Owambo: 81).

The following helped organise and conduct sampling in Lesotho:

- **Dr Cristian Capelli** (Department of Zoology, University of Oxford)
- **Dr Ockie and Erica Oosthuizen**
- **Simon Marks**

Bibliography

Albrecht, M., H. Berke, B. Eichhorn, T. Frank, R. Kuper, S. Prill, R. Vogelsang and S. Wenzel (2001). "Oruwanje 95/1: a late Holocene stratigraphy in northwestern Namibia." Cimbebasia **17**: 1-22.

Alexander, J. A. (1977). The 'frontier' concept in prehistory: the end of the moving frontier. Hunters, gatherers and first farmers beyond Europe. J. V. S. Megaw. Leicester, Leicester University Press**:** 25-40.

Alexander, J. A. (1984). Early frontiers in southern Africa. Frontiers: Southern African Archaeology Today. M. Hall, G. Avery, D. M. Avery, M. L. Wilson and A. J. B. Humphreys. Oxford, B.A.R. International Series 207**:** 12-23.

Alves-Silva, J., M. da Silva Santos, P. E. Guimaraes, A. C. Ferreira, H. J. Bandelt, S. D. Pena and V. F. Prado (2000). "The ancestry of Brazilian mtDNA lineages." Am J Hum Genet **67**(2): 444-461.

Ambrose, S. H. (1982). Archaeology and linguistic reconstructions of history in Eastern Africa. Archaeological and linguistic reconstruction of African history. C. Ehret and M. Posnansky. Berkeley (CA), University of California Press**:** 104-157.

Ammerman, A. J. and L. L. Cavalli-Sforza (1973). A population model for the diffusion of early farming in Europe. The explanation of culture change: models in prehistory: proceedings of a meeting of the Research Seminar in Archaeology and Related Subjects. C. Renfrew. London, Duckworth**:** 343-357.

Anderson, S., A. T. Bankier, B. G. Barrell, M. H. de Bruijn, A. R. Coulson, J. Drouin, I. C. Eperon, D. P. Nierlich, B. A. Roe, F. Sanger, P. H. Schreier, A. J. Smith, R. Staden and I. G. Young (1981). "Sequence and organization of the human mitochondrial genome." Nature **290**(5806): 457-465.

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

Atkinson, Q. D. (2011). "Phonemic diversity supports a serial founder effect model of language expansion from Africa." Science **332**(6027): 346-349.

Awadalla, P., A. Eyre-Walker and J. M. Smith (1999). "Linkage disequilibrium and recombination in hominid mitochondrial DNA." Science **286**(5449): 2524-2525.

Balaresque, P., G. R. Bowden, S. M. Adams, H. Y. Leung, T. E. King, Z. H. Rosser, J. Goodwin, J. P. Moisan, C. Richard, A. Millward, A. G. Demaine, G. Barbujani, C. Previdere, I. J. Wilson, et al. (2010). "A predominantly neolithic origin for European paternal lineages." PLoS Biol **8**(1): e1000285.

Bandelt, H. J., J. Alves-Silva, P. E. M. GuimarÃEs, M. S. Santos, A. Brehm, L. Pereira, A. Coppa, J. M. Larruga, C. Rengo, R. Scozzari, A. Torroni, M. J. Prata, A. Amorim, V. F. Prado, et al. (2001). "Phylogeography of the human mitochondrial haplogroup L3e: a snapshot of African prehistory and Atlantic slave trade." Annals of Human Genetics **65**(6): 549-563.

Bandelt, H. J., P. Forster and A. Rohl (1999). "Median-joining networks for inferring intraspecific phylogenies." Molecular Biology and Evolution **16**(1): 37-48.

Bandelt, H. J., P. Forster, B. C. Sykes and M. B. Richards (1995). "Mitochondrial portraits of human populations using median networks." Genetics **141**(2): 743-753.

Barbujani, G., R. R. Sokal and N. L. Oden (1995). "Indo-European origins: a computer-simulation test of five hypotheses." Am J Phys Anthropol **96**(2): 109-132.

Barnard, A. (1988). "Kinship, Language and Production: A Conjectural History of Khoisan Social Structure." Africa: Journal of the International African Institute **58**(1): 29-50.

Barnard, A. (1992). Hunters and herders of southern Africa : a comparative ethnography of the Khoisan peoples. Cambridge, Cambridge University Press.

Barnard, A. (2008). "Ethnographic analogy and the reconstruction of early Khoekhoe society." Southern African Humanities **20**: 61-75.

Batini, C., V. Coia, C. Battaggia, J. Rocha, M. M. Pilkington, G. Spedini, D. Comas, G. Destro-Bisol and F. Calafell (2007). "Phylogeography of the human mitochondrial L1c haplogroup: genetic signatures of the prehistory of Central Africa." Mol Phylogenet Evol **43**(2): 635-644.

Batini, C., G. Ferri, G. Destro-Bisol, F. Brisighelli, D. Luiselli, P. Sanchez-Diz, J. Rocha, T. Simonson, A. Brehm, V. Montano, N. E. Elwali, G. Spedini, M. Eugenia D'Amato, N. Myres, et al. (2011). "Signatures of the pre-agricultural peopling processes in sub-Saharan Africa as revealed by the phylogeography of early Y chromosome lineages." Mol Biol Evol **28**(9): 2603-2613.

BBC. (2011a). "Lesotho Profile." Retrieved July 2012, 2012, from http://www.bbc.co.uk/news/world-africa-13728325.

BBC. (2011b). "Namibia Profile." from http://www.bbc.co.uk/news/world-africa-13890723.

Beaumont, P. and J. C. Vogel (1984). Spatial patterning of the ceramic Later Stone Age in the northern Cape Province, South Africa. Frontiers: Southern African Archaeology Today. M. Hall, G. Avery, D. M. Avery, M. L. Wilson and A. J. B. Humphreys. Oxford, B.A.R. International Series 207**:** 80-95.

Behar, D. M., R. Villems, H. Soodyall, J. Blue-Smith, L. Pereira, E. Metspalu, R. Scozzari, H. Makkan, S. Tzur, D. Comas, J. Bertranpetit, L. Quintana-Murci, C. Tyler-Smith, R. S. Wells, et al. (2008). "The dawn of human matrilineal diversity." American Journal of Human Genetics **82**(5): 1130-1140.

Belle, E. M. and G. Barbujani (2007). "Worldwide analysis of multiple microsatellites: language diversity has a detectable influence on DNA diversity." Am J Phys Anthropol **133**(4): 1137-1146.

Bellwood, P. (2001). "Early Agriculturalist Population Diasporas? Farming, Languages, and Genes." Annual Review of Anthropology **30**(1): 181-207.

Bendall, K. E., V. A. Macaulay, J. R. Baker and B. C. Sykes (1996). "Heteroplasmic point mutations in the human mtDNA control region." Am J Hum Genet **59**(6): 1276-1287.

Berniell-Lee, G., F. Calafell, E. Bosch, E. Heyer, L. Sica, P. Mouguiama-Daouda, L. van der Veen, J. M. Hombert, L. Quintana-Murci and D. Comas (2009). "Genetic and demographic implications of the Bantu expansion: insights from human paternal lineages." Mol Biol Evol **26**(7): 1581-1589.

Besaggio, D., S. Fuselli, M. Srikummool, J. Kampuansai, L. Castrì, C. Tyler-Smith, M. Seielstad, D. Kangwanpong and G. Bertorelle (2007). "Genetic variation in Northern Thailand Hill Tribes: origins and relationships with social structure and linguistic differences." BMC Evol Biol **7 Suppl 2**: S12.

Bleek, D. F. (1929). Comparative vocabularies and Bushman languages. Cambridge, Cambridge University Press.

Blench, R. (2006). Archaeology, language and the African past. Walnut Creek, AltaMira Press.

Blench, R. (2009). Was there an interchange between Cushitic pastoralists and Khoisan speakers in the prehistory of Southern Africa and how can this be detected? Language Contact, Language Change and History Based on Language Sources in Africa. W. J. G. Mohlig, F. Seidel and M. Seifert. **20**.

Bohannan, P. and F. Plog, Eds. (1967). Beyond the frontier: social process and cultural change. Garden City, New York, Natural History Press, for the American Museum of Natural History.

Brandstätter, A., C. T. Peterson, J. A. Irwin, S. Mpoke, D. K. Koech, W. Parson and T. J. Parsons (2004). "Mitochondrial DNA control region sequences from Nairobi (Kenya): inferring phylogenetic parameters for the establishment of a forensic database." Int J Legal Med **118**(5): 294-306.

Bridgman, J. and L. J. Worley (2004). Genocide of the Hereros. Century of Genocide. S. Totten, W. S. Parsons and I. W. Charny. London, Routledge**:** 3-40.

Brinkmann, B., M. Klintschar, F. Neuhuber, J. Hühne and B. Rolf (1998). "Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat." Am J Hum Genet **62**(6): 1408-1415.

Brown, J. (2011). Identification of genetic variants associated with lactase persistence in populations from Namibia: evidence of a pastoralist migration from East Africa, University of Oxford.

Brown, W. M., M. George, Jr. and A. C. Wilson (1979). "Rapid evolution of animal mitochondrial DNA." Proc Natl Acad Sci U S A **76**(4): 1967-1971.

Burton, M. L., C. C. Moore, J. W. M. Whiting, A. K. Romney, D. F. Aberle, J. A. Barcelo, M. Dow, M,, J. I. Guyer, D. B. Kronenfeld, J. E. Levy and J. Linnekin (1996). "Regions Based on Social Structure." Current Anthropology **37**: 87-123.

Busby, G. B., F. Brisighelli, P. Sánchez-Diz, E. Ramos-Luis, C. Martinez-Cadenas, M. G. Thomas, D. G. Bradley, L. Gusmão, B. Winney, W. Bodmer, M. Vennemann, V. Coia, F. Scarnicci, S. Tofanelli, et al. (2012). "The peopling of Europe and the cautionary tale of Y chromosome lineage R-M269." Proc Biol Sci **279**(1730): 884-892.

Butler, J. M., M. C. Kline and A. E. Decker (2008). "Addressing Y-Chromosome Short Tandem Repeat Allele Nomenclature." Journal of Genetic Genealogy **4**(2): 125-148.

Campbell, M. C. and S. A. Tishkoff (2010). "The evolution of human genetic and phenotypic variation in Africa." Curr Biol **20**(4): R166-173.

Cann, R. L. (1988). "DNA and Human Origins." Annual Review of Anthropology **17**(1): 127-143.

Cann, R. L., M. Stoneking and A. C. Wilson (1987). "Mitochondrial DNA and human evolution." Nature **325**(6099): 31-36.

Carter, P. L. and J. C. Vogel (1974). "The Dating of Industrial Assemblages from Stratified Sites in Eastern Lesotho." Man **9**(4): 557-570.

Casanova, M., P. Leroy, C. Boucekkine, J. Weissenbach, C. Bishop, M. Fellous, M. Purrello, G. Fiori and M. Siniscalco (1985). "A human Y-linked DNA polymorphism and its potential for estimating genetic and evolutionary distance." Science **230**(4732): 1403-1406.

Castri, L., S. Tofanelli, P. Garagnani, C. Bini, X. Fosella, S. Pelotti, G. Paoli, D. Pettener and D. Luiselli (2009). "mtDNA variability in two Bantu-speaking populations (Shona and Hutu) from Eastern Africa: implications for peopling and migration patterns in sub-Saharan Africa." Am J Phys Anthropol **140**(2): 302-311.

Cavalli-Sforza, L., Ed. (1986). African Pygmies. New York, Academic Press.

Cavalli-Sforza, L., P. Menozzi and A. Piazza (1994). The History and Geography of Human Genes. Princeton, New Jersey, Princeton University Press.

Cerezo, M., H. J. Bandelt, I. Martin-Guerrero, M. Ardanaz, A. Vega, A. Carracedo, A. Garcia-Orad and A. Salas (2009). "High mitochondrial DNA stability in B-cell chronic lymphocytic leukemia." PLoS One **4**(11): e7902.

Chaix, R., F. Austerlitz, T. Khegay, S. Jacquesson, M. F. Hammer, E. Heyer and L. Quintana-Murci (2004). "The genetic or mythical ancestry of descent groups: lessons from the Y chromosome." Am J Hum Genet **75**(6): 1113-1116.

Chaix, R., L. Quintana-Murci, T. Hegay, M. F. Hammer, Z. Mobasher, F. Austerlitz and E. Heyer (2007). "From social to genetic structures in central Asia." Curr Biol **17**(1): 43-48.

Chakraborty, R. (1990). "Mitochondrial DNA polymorphism reveals hidden heterogeneity within some Asian populations." Am J Hum Genet **47**(1): 87-94.

Chen, Y. S., A. Olckers, T. G. Schurr, A. M. Kogelnik, K. Huoponen and D. C. Wallace (2000). "mtDNA variation in the South African Kung and Khwe-and their genetic relationships to other African populations." American Journal of Human Genetics **66**(4): 1362-1383.

Chen, Y. S., A. Torroni, L. Excoffier, A. S. Santachiara-Benerecetti and D. C. Wallace (1995). "Analysis of mtDNA variation in African populations reveals the most ancient of all human continent-specific haplogroups." Am J Hum Genet **57**(1): 133-149.

Chiaroni, J., P. A. Underhill and L. L. Cavalli-Sforza (2009). "Y chromosome diversity, human expansion, drift, and cultural evolution." Proc Natl Acad Sci U S A **106**(48): 20174-20179.

Christie, K. (2005). "The specter of genocide: mass murder in historical perspective, and Final solutions: mass killings and genocide in the 20th century (review)." Holocaust and Genocide Studies **19**: 317-321.

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

Coia, V., F. Brisighelli, F. Donati, V. Pascali, I. Boschi, D. Luiselli, C. Battaggia, C. Batini, L. Taglioli, F. Cruciani, G. Paoli, C. Capelli, G. Spedini and G. Destro-Bisol (2009). "A multi-perspective view of genetic variation in Cameroon." Am J Phys Anthropol **140**(3): 454-464.

Coia, V., G. Destro-Bisol, F. Verginelli, C. Battaggia, I. Boschi, F. Cruciani, G. Spedini, D. Comas and F. Calafell (2005). "Brief communication: mtDNA variation in North Cameroon: lack of Asian lineages and implications for back migration from Asia to sub-Saharan Africa." Am J Phys Anthropol **128**(3): 678-681.

Compton, J. S. (2011). "Pleistocene sea-level fluctuations and human evolution on the southern coastal plain of South Africa." Quaternary Science Reviews **30**(5-6): 506-527.

Cooper, A. (2007). Reparations for the Herero Genocide: defining the limits of international litigation. African Affairs. **106:** 113-126.

Copeland, S. R., M. Sponheimer, D. J. de Ruiter, J. A. Lee-Thorp, D. Codron, P. J. le Roux, V. Grimes and M. P. Richards (2011). "Strontium isotope evidence for landscape use by early hominins." Nature **474**(7349): 76-78.

Crawhall, N. (2006). Languages, genetics and archaeology: problems and the possibilities in Africa. The Prehistory of Africa. H. Soodyall. Johannesburg, Witwatersrand University Press**:** 109-124.

Cruciani, F., P. Santolamazza, P. Shen, V. Macaulay, P. Moral, A. Olckers, D. Modiano, S. Holmes, G. Destro-Bisol, V. Coia, D. C. Wallace, P. J. Oefner, A. Torroni, L. L. Cavalli-Sforza, et al. (2002). "A back migration from Asia to sub-Saharan Africa is supported by high-resolution analysis of human Y-chromosome haplotypes." Am J Hum Genet **70**(5): 1197-1214.

Cruciani, F., B. Trombetta, A. Massaia, G. Destro-Bisol, D. Sellitto and R. Scozzari (2011). "A revised root for the human Y chromosomal phylogenetic tree: the origin of patrilineal diversity in Africa." Am J Hum Genet **88**(6): 814-818.

Currat, M. and L. Excoffier (2005). "The effect of the Neolithic expansion on European molecular diversity." Proc Biol Sci **272**(1564): 679-688.

Currat, M., M. Ruedi, R. J. Petit and L. Excoffier (2008). "The hidden side of invasions: massive introgression by local genes." Evolution **62**(8): 1908-1920.

Darwin, C. (1871). The descent of man, and selection in relation to sex. London, John Murray.

Davies, G. (1994). The Medical Culture of the Ovambo of Southern Angola and Northern Namibia, University of Kent at Canterbury. **Doctor of Philosophy**.

Davies, N. (2011). Genetic insight into linguistic and cultural transitions during the peopling of Namibia. University of Oxford.

de Filippo, C., C. Barbieri, M. Whitten, S. W. Mpoloka, E. D. Gunnarsdóttir, K. Bostoen, T. Nyambe, K. Beyer, H. Schreiber, P. de Knijff, D. Luiselli, M. Stoneking and B. Pakendorf (2011). "Y-chromosomal variation in sub-Saharan Africa: insights into the history of Niger-Congo groups." Mol Biol Evol **28**(3): 1255-1269.

de Filippo, C., K. Bostoen, M. Stoneking and B. Pakendorf (2012). "Bringing together linguistic and genetic evidence to test the Bantu expansion." Proc Biol Sci **279**(1741): 3256-3263.
de Filippo, C., P. Heyn, L. Barham, M. Stoneking and B. Pakendorf (2010). "Genetic perspectives on forager-farmer interaction in the Luangwa valley of Zambia." Am J Phys Anthropol **141**(3): 382-394.
Deacon, J. (1984). Later Stone Age people and their descendants in southern Africa. Southern African prehistory and palaeoenvironments. R. G. Klein. Rotterdam and Boston, A. A. Balkema**:** 221-329.
Denbow, J. and A. Campbell (1986). "The early stages of food production in southern Africa and some potential linguistic connections." Sprache und Geschichte in Afrika **7**(1): 83-103.
Destro-Bisol, G. (2005). "Genetic variation and social structure: a case-study from Africa." Human Evolution **20**(2-3): 93-98.
Destro-Bisol, G., V. Coia, I. Boschi, F. Verginelli, A. Caglia, V. Pascali, G. Spedini and F. Calafell (2004a). "The analysis of variation of mtDNA hypervariable region 1 suggests that Eastern and Western Pygmies diverged before the Bantu expansion." Am Nat **163**(2): 212-226.
Destro-Bisol, G., F. Donati, V. Coia, I. Boschi, F. Verginelli, A. Caglia, S. Tofanelli, G. Spedini and C. Capelli (2004b). "Variation of female and male lineages in sub-Saharan populations: the importance of sociocultural factors." Mol Biol Evol **21**(9): 1673-1682.
Diamond, J. and P. Bellwood (2003). "Farmers and their languages: the first expansions." Science **300**(5619): 597-603.
Dupuy, B. M., M. Stenersen, T. Egeland and B. Olaisen (2004). "Y-chromosomal microsatellite mutation rates: differences in mutation rate between and within loci." Hum Mutat **23**(2): 117-124.
Ebbesen, P., C. B. Juhl, W. D. Osei and E. T. Maganu (1989). "San people (bushmen) from the central Kalahari desert found free of antibodies to HIV-1 and HIV-2." J Acquir Immune Defic Syndr **2**(4): 415-416.
Edmonds, C. A., A. S. Lillie and L. L. Cavalli-Sforza (2004). "Mutations arising in the wave front of an expanding population." Proc Natl Acad Sci U S A **101**(4): 975-979.
Ehret, C. (1982a). The first spread of food production to southern Africa. The Archaeological and Linguistic Reconstruction of African History. C. Ehret and M. Posnansky. Berkeley, University of California Press**:** 158-181.
Ehret, C. (1982b). Linguistic inferences about early Bantu history. The Archaeological and Linguistic Reconstruction of African History. C. Ehret and M. Posnansky. Berkeley, University of California Press**:** 57-65.
Ehret, C. (1998). An African Classical Age: Eastern and Southern Africa in World History, 1000 B.C. to A.D. 400. Charlottesville, University Press of Virginia.
Ehret, C. (2001). "Bantu Expansions: Re-Envisioning a Central Problem of Early African History." The International Journal of African Historical Studies **34**(1): 5-41.
Ehret, C. (2008). "The early livestock raisers of southern Africa." Southern African Humanities **20**: 7-35.
Ellenberger, D. F. (1912). History of the Basuto: Ancient and Modern. Morija, Morija Printing Works.
Elphick, R. (1977). Kraal and castle. New Haven, CT, Yale University Press.
Elphick, R. (1985). Khoikhoi and the founding of white South Africa. Johannesburg, Ravan Press.
Ember, M. (1967). "The emergence of neolocal residence." Trans N Y Acad Sci **30**(2): 291-302.
Eriksson, J., H. Siedel, D. Lukas, M. Kayser, A. Erler, C. Hashimoto, G. Hohmann, C. Boesch and L. Vigilant (2006). "Y-chromosome analysis confirms highly sex-biased dispersal and

suggests a low male effective population size in bonobos (Pan paniscus)." Mol Ecol **15**(4): 939-949.
Estermann, C. (1976). The Ethnography of Southwest Angola Vol. 1: The Non-Bantu Peoples, the Ambo Ethnic Group. London, London African Pub. Co.
Excoffier, L. (2004). "Patterns of DNA sequence diversity and genetic structure after a range expansion: lessons from the infinite-island model." Mol Ecol **13**(4): 853-864.
Excoffier, L. and H. E. L. Lischer (2010). "Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows." Molecular Ecology Resources **10**(3): 564-567.
Excoffier, L. and N. Ray (2008). "Surfing during population expansions promotes genetic revolutions and structuration." Trends Ecol Evol **23**(7): 347-351.
Excoffier, L., P. E. Smouse and J. M. Quattro (1992). "Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data." Genetics **131**(2): 479-491.
Eyre-Walker, A., N. H. Smith and J. M. Smith (1999). "How clonal are human mitochondria?" Proc Biol Sci **266**(1418): 477-483.
Fauvelle-Aymar, F. X. (2008). "Against the 'Khoisan paradigm' in the interpretation of Khoekhoe origins and history: a re-evaluation of Khoekhoe pastoral traditions." Southern African Humanities **20**: 77-92.
Fauvelle-Aymar, F. X. and K. Sadr (2008). "Trends and traps in the reconstruction of early herding societies in southern Africa." Southern African Humanities **20**: 1-6.
Forster, P., A. Röhl, P. Lünnemann, C. Brinkmann, T. Zerjal, C. Tyler-Smith and B. Brinkmann (2000). "A short tandem repeat-based phylogeny for the human Y chromosome." Am J Hum Genet **67**(1): 182-196.
Francis, M. (2007). Explorations in Ethnicity and Social Change among Zulu-speaking San Descendents of the Drakensberg Mountains, KwaZulu-Natal. Faculty of Human Sciences, University of KwaZulu-Natal. **PhD**.
Fu, Y. X. (1997). "Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection." Genetics **147**(2): 915-925.
Garrigan, D. and M. F. Hammer (2006). "Reconstructing human origins in the genomic era." Nat Rev Genet **7**(9): 669-680.
Giles, R. E., H. Blanc, H. M. Cann and D. C. Wallace (1980). "Maternal inheritance of human mitochondrial DNA." Proc Natl Acad Sci U S A **77**(11): 6715-6719.
Gill, S. J. (1992). Introductory Essay to D. Fred. Ellenberger's 'History of the Basuto: Ancient and Modern'. History of the Basuto: Ancient and Modern. Morija, Morija Printing Works**:** xvii-xxii.
Gill, S. J. (1993). A short history of Lesotho from the late stone age until the 1993 elections. Lesotho, Morija Museum & Archives.
Gomes, V., P. Sanchez-Diz, A. Amorim, A. Carracedo and L. Gusmao (2010). "Digging deeper into East African human Y chromosome lineages." Hum Genet **127**(5): 603-613.
Gonder, M. K., H. M. Mortensen, F. A. Reed, A. de Sousa and S. A. Tishkoff (2007). "Whole-mtDNA genome sequence analysis of ancient African lineages." Mol Biol Evol **24**(3): 757-768.
Goodman, M. (1963). Man's Place in the Phylogeny of the Primates as Reflected in Serum Proteins. Classification and Human Evolution. S. L. Washburn, Wenner-Gren Foundation for Anthropological Research Inc.**:** 204-234.
Gordon, R. (1984). The !Kung in the Kalahari exchange: an ethnohistorical perspective. Past and present in hunter-gatherer studies. C. Schrire. Orlando, FL, Academic Press**:** 195-224.
Government of the Republic of Namibia. (2011). "GRNNET." Retrieved Februrary 2012, from www.grnnet.gov.na.

Greenberg, J. H. (1966). The languages of Africa. Bloomington, Indiana University.
Grine, F. E., W. L. Jungers, P. V. Tobias and O. M. Pearson (1995). "Fossil Homo femur from Berg Aukas, northern Namibia." Am J Phys Anthropol **97**(2): 151-185.
Guldemann, T. (2008). "A linguist's view: Khoe-Kwadi speakers as the earliest food-producers of southern Africa." Southern African Humanities **20**(1): 93-132.
Guldemann, T. and M. Stoneking (2008). "A Historical Appraisal of Clicks: A Linguistic and Genetic Population Perspective." Annual Review of Anthropology **37**(1): 93-109.
Guldemann, T. and R. Vossen (2000). Khoisan. African Languages. B. Heine and D. Nurse. Cambridge, CUP**:** 99-122.
Gusmao, L., J. M. Butler, A. Carracedo, P. Gill, M. Kayser, W. R. Mayr, N. Morling, M. Prinz, L. Roewer, C. Tyler-Smith and P. M. Schneider (2006). "DNA Commission of the International Society of Forensic Genetics (ISFG): an update of the recommendations on the use of Y-STRs in forensic analysis." Forensic Sci Int **157**(2-3): 187-197.
Gusmão, L., P. Sánchez-Diz, F. Calafell, P. Martín, C. A. Alonso, F. Alvarez-Fernández, C. Alves, L. Borjas-Fajardo, W. R. Bozzo, M. L. Bravo, J. J. Builes, J. Capilla, M. Carvalho, C. Castillo, et al. (2005). "Mutation rates at Y chromosome specific microsatellites." Hum Mutat **26**(6): 520-528.
Haacke, W. H. G. (2008). "Linguistic hypotheses on the origins of Namibian Khoekhoe speakers." Southern African Humanities **20**: 163-177.
Hagelberg, E., N. Goldman, P. Lió, S. Whelan, W. Schiefenhövel, J. B. Clegg and D. K. Bowden (1999). "Evidence for mitochondrial DNA recombination in a human population of island Melanesia." Proc Biol Sci **266**(1418): 485-492.
Hall, M. (1987). The changing past: Farmers, kings and traders in southern Africa, 200-1860. Cape, South Africa, David Philip.
Hammer, M. F., T. M. Karafet, A. J. Redd, H. Jarjanazi, S. Santachiara-Benerecetti, H. Soodyall and S. L. Zegura (2001). "Hierarchical patterns of global human Y-chromosome diversity." Mol Biol Evol **18**(7): 1189-1203.
Hammer, M. F., F. L. Mendez, M. P. Cox, A. E. Woerner and J. D. Wall (2008). "Sex-biased evolutionary forces shape genomic patterns of human diversity." PLoS Genet **4**(9): e1000202.
Hammond-Tooke, W. D. (1998). "Selective Borrowing? The Possibility of San Shamanistic Influence on Southern Bantu Divination and Healing Practices." South African Archaeological Bulletin **53**: 9-15.
Hammond-Tooke, W. D. (1999). "Divinatory Animals: Further Evidence of San/Nguni Borrowing?" South African Archaeological Bulletin **54**: 128-132.
Harlan, J. R., J. M. J. de Wet and A. B. L. Stemler, Eds. (1976). Origins of African plant domestication. The Hague, Mouton.
Harpending, H. (2005). "POPSTR." from http://harpending.humanevo.utah.edu/popstr/.
Harpending, H. C. (1994). "Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution." Hum Biol **66**(4): 591-600.
Hasegawa, M., A. Di Rienzo, T. D. Kocher and A. C. Wilson (1993). "Toward a more accurate time scale for the human mitochondrial DNA tree." Journal of Molecular Evolution **37**(4): 347-354.
Hedrick, P. W. (2005). "A standardized genetic differentiation measure." Evolution **59**(8): 1633-1638.
Heine, B. and C. Konig (2008). "What can linguistics tell us about early Khoekhoe history?" Southern African Humanities **20**: 235-248.
Heine, B. and D. Nurse (2000). Introduction. African languages : an introduction. B. Heine and D. Nurse. Cambridge, Cambridge University Press**:** 1-10.
Hellenthal, G., A. Auton and D. Falush (2008). "Inferring human colonization history using a copying model." PLoS Genet **4**(5): e1000078.

Henn, B. M., C. Gignoux, A. A. Lin, P. J. Oefner, P. Shen, R. Scozzari, F. Cruciani, S. A. Tishkoff, J. L. Mountain and P. A. Underhill (2008). "Y-chromosomal evidence of a pastoralist migration through Tanzania to southern Africa." Proc Natl Acad Sci U S A **105**(31): 10693-10698.

Henn, B. M., C. R. Gignoux, M. Jobin, J. M. Granka, J. M. Macpherson, J. M. Kidd, L. Rodríguez-Botigué, S. Ramachandran, L. Hon, A. Brisbin, A. A. Lin, P. A. Underhill, D. Comas, K. K. Kidd, et al. (2011). "Hunter-gatherer genomic diversity suggests a southern African origin for modern humans." Proc Natl Acad Sci U S A **108**(13): 5154-5162.

Henshilwood, C. (1996). "A revised chronology for pastoralism in southernmost Africa: new evidence of sheep at c. 2000 b.p. from Blombos Cave, South Africa." Antiquity **70**(945-949).

Henshilwood, C. S., F. d'Errico, K. L. van Niekerk, Y. Coquinot, Z. Jacobs, S. E. Lauritzen, M. Menu and R. García-Moreno (2011). "A 100,000-year-old ochre-processing workshop at Blombos Cave, South Africa." Science **334**(6053): 219-222.

Herbert, R. (1990). "The sociohistory of clicks in Southern Bantu." Anthropological Linguistics **32**: 295-315.

Heyer, E., P. Balaresque, M. A. Jobling, L. Quintana-Murci, R. Chaix, L. Segurel, A. Aldashev and T. Hegay (2009). "Genetic diversity and the emergence of ethnic groups in Central Asia." BMC Genet **10**: 49.

Heyer, E., R. Chaix, S. Pavard and F. Austerlitz (2012). "Sex-specific demographic behaviours that shape human genomic variation." Mol Ecol **21**(3): 597-612.

Heyer, E., J. Puymirat, P. Dieltjes, E. Bakker and P. de Knijff (1997). "Estimating Y chromosome specific microsatellite mutation frequencies using deep rooting pedigrees." Hum Mol Genet **6**(5): 799-803.

Hirschfeld, L. and H. Hirschfeld (1919). "Serological Differences Between the Blood of Different Races: the Result of Researches on the Macedonian Front." The Lancet **194**(5016): 675-679.

Holden, C. and R. Mace (1997). "Phylogenetic analysis of the evolution of lactose digestion in adults." Hum Biol **69**(5): 605-628.

Holden, C. J. (2002). "Bantu language trees reflect the spread of farming across sub-Saharan Africa: a maximum-parsimony analysis." Proc Biol Sci **269**(1493): 793-799.

Holtkemper, U., B. Rolf, C. Hohoff, P. Forster and B. Brinkmann (2001). "Mutation rates at two human Y-chromosomal microsatellite loci using small pool PCR techniques." Hum Mol Genet **10**(6): 629-633.

Horsburgh, K. A. (2008). "Wild or domesticated? An ancient DNA approach to canid species identification in South Africa's Western Cape Province." Journal of Archaeological Science **35**(6): 1474-1480.

How, M. W. (1962). The Mountain Bushmen of Basutoland. Pretoria, Van Schaik.

Howell, N., C. B. Smejkal, D. A. Mackey, P. F. Chinnery, D. M. Turnbull and C. Herrnstadt (2003). "The pedigree rate of sequence divergence in the human mitochondrial genome: there is a difference between phylogenetic and pedigree rates." Am J Hum Genet **72**(3): 659-670.

Hublin, J. J. and R. G. Klein (2011). "Northern Africa could also have housed the source population for living humans." Proc Natl Acad Sci U S A **108**(28): E277; author reply E278.

Hurles, M. E., J. Nicholson, E. Bosch, C. Renfrew, B. C. Sykes and M. A. Jobling (2002). "Y chromosomal evidence for the origins of oceanic-speaking peoples." Genetics **160**(1): 289-303.

Huxley, T. H. (1863). 'On the relations of man to the lower animals'. Evidence as to man's place in nature. New York, D. Appleton and company**:** 72-134.

Ingman, M., H. Kaessmann, S. Paabo and U. Gyllensten (2000). "Mitochondrial genome variation and the origin of modern humans." Nature **408**(6813): 708-713.

Ingram, C. J., C. A. Mulcare, Y. Itan, M. G. Thomas and D. M. Swallow (2009). "Lactose digestion and the evolutionary genetics of lactase persistence." Hum Genet **124**(6): 579-591.

Jazin, E., H. Soodyall, P. Jalonen, E. Lindholm, M. Stoneking and U. Gyllensten (1998). "Mitochondrial mutation rate revisited: hot spots and polymorphism." Nat Genet **18**(2): 109-110.

Jenkins, T. (1986). The prehistory of the San and Khoikhoi as recorded in their blood. Contemporary Studies on Khoisan 2. R. Vossen and K. Keuthmann. Hamburg, Helmut Buske Verlag**:** 51-77.

Jenkins, T., H. C. Harpending, H. Gordon, M. M. Keraan and S. Johnston (1971). "Red-cell-enzyme polymorphisms in the Khoisan peoples of Southern Africa." American journal of human genetics **23**(5): 513-532.

Jenkins, T., A. Zoutendyk and A. G. Steinberg (1970). "Gammaglobulin groups (Gm and Inv) of various Southern African populations." American journal of physical anthropology **32**(2): 197-218.

Jobling, M. A. (2001). "In the name of the father: surnames and genetics." Trends Genet **17**(6): 353-357.

Jobling, M. A., M. E. Hurles and C. Tyler-Smith (2004). Human Evolutionary Genetics: Origins, Peoples & Disease, Garland Publishing.

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

Jolly, P. (1996). "Interaction between South-Eastern San and Southern Nguni and Sotho communities c.1400 to c.1880." Suid-Afrikaanse Historiese Joernaal-South African Historical Journal(35): 30-61.

Jolly, P. (2005). "Sharing Symbols: A Correspondence in the Ritual Dress of Black Farmers and the Southeastern San." South African Archaeological Society Goodwin Series **9**: 86-100.

Jones, D. (2011). "The matrilocal tribe: an organization of demic expansion." Hum Nat **22**(1-2): 177-200.

Karafet, T. M., F. L. Mendez, M. B. Meilerman, P. A. Underhill, S. L. Zegura and M. F. Hammer (2008). "New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree." Genome Res **18**(5): 830-838.

Kayser, M., S. Brauer, G. Weiss, W. Schiefenhovel, P. Underhill, P. Shen, P. Oefner, M. Tommaseo-Ponzetta and M. Stoneking (2003). "Reduced Y-chromosome, but not mitochondrial DNA, diversity in human populations from West New Guinea." Am J Hum Genet **72**(2): 281-302.

Kayser, M., L. Roewer, M. Hedman, L. Henke, J. Henke, S. Brauer, C. Krüger, M. Krawczak, M. Nagy, T. Dobosz, R. Szibor, P. de Knijff, M. Stoneking and A. Sajantila (2000). "Characteristics and frequency of germline mutations at microsatellite loci from the human Y chromosome, as revealed by direct observation in father/son pairs." Am J Hum Genet **66**(5): 1580-1588.

Keinan, A., J. C. Mullikin, N. Patterson and D. Reich (2009). "Accelerated genetic drift on chromosome X during the human dispersal out of Africa." Nat Genet **41**(1): 66-70.

Kinahan, J. (1996). "Alternative Views on the Acquisition of Livestock by Hunter-Gatherers in Southern Africa: A Rejoinder to Smith, Yates and Jacobson." The South African Archaeological Bulletin **51**: 106-108.

Kinahan, J. (2001). Pastoral nomads of the Namib Desert: The people history forgot. Windhoek, Namibia Archaeological Trust.

King, T. E. and M. A. Jobling (2009a). "Founders, drift, and infidelity: the relationship between Y chromosome diversity and patrilineal surnames." Mol Biol Evol **26**(5): 1093-1102.

King, T. E. and M. A. Jobling (2009b). "What's in a name? Y chromosomes, surnames and the genetic genealogy revolution." Trends Genet **25**(8): 351-360.

Kittles, R. and S. O. Y. Keita (1999). "Interpreting African Genetic Diversity." African Archaeological Review **16**(2): 87-91.

Kivisild, T., M. Reidla, E. Metspalu, A. Rosa, A. Brehm, E. Pennarun, J. Parik, T. Geberhiwot, E. Usanga and R. Villems (2004). "Ethiopian mitochondrial DNA heritage: tracking gene flow across and around the gate of tears." American Journal of Human Genetics **75**(5): 752-770.

Kivisild, T., R. Villems, L. B. Jorde, M. Bamshad, S. Kumar, P. Hedrick, T. Dowling, M. Stoneking, T. J. Parsons, J. A. Irwin, P. Awadalla, A. Eyre-Walker and J. M. Smith (2000). "Questioning evidence for recombination in human mitochondrial DNA." Science **288**(5473): 1931a.

Klein, R. G. (1986). "The Prehistory of Stone Age Herders in the Cape Province of South Africa." South African Archaeological Society Goodwin Series **5**: 5-12.

Klopfstein, S., M. Currat and L. Excoffier (2006). "The fate of mutations surfing on the wave of a range expansion." Mol Biol Evol **23**(3): 482-490.

Kloss-Brandstatter, A., D. Pacher, S. Schonherr, H. Weissensteiner, R. Binna, G. Specht and F. Kronenberg (2011). "HaploGrep: a fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups." Hum Mutat **32**(1): 25-32.

Knight, A., P. A. Underhill, H. M. Mortensen, L. A. Zhivotovsky, A. A. Lin, B. M. Henn, D. Louis, M. Ruhlen and J. L. Mountain (2003). "African Y chromosome and mtDNA divergence provides insight into the history of click languages." Curr Biol **13**(6): 464-473.

Kumar, S., K. Tamura and M. Nei (1993). MEGA, Molecular Evolutionary Genetic Analysis ver 1.0. University Park, PA 16802, The Pennsylvania State University.

Kumar, V., B. T. Langstieh, K. V. Madhavi, V. M. Naidu, H. P. Singh, S. Biswas, K. Thangaraj, L. Singh and B. M. Reddy (2006). "Global patterns in human mitochondrial DNA and Y-chromosome variation caused by spatial instability of the local cultural processes." PLoS Genet **2**(4): e53.

Laland, K. N., J. Odling-Smee and S. Myles (2010). "How culture shaped the human genome: bringing genetics and the human sciences together." Nat Rev Genet **11**(2): 137-148.

Lalueza-Fox, C., A. Rosas, A. Estalrrich, E. Gigli, P. F. Campos, A. García-Tabernero, S. García-Vargas, F. Sánchez-Quinto, O. Ramírez, S. Civit, M. Bastir, R. Huguet, D. Santamaría, M. T. Gilbert, et al. (2011). "Genetic evidence for patrilocal mating behavior among Neandertal groups." Proc Natl Acad Sci U S A **108**(1): 250-253.

Langergraber, K. E., H. Siedel, J. C. Mitani, R. W. Wrangham, V. Reynolds, K. Hunt and L. Vigilant (2007). "The genetic signature of sex-biased migration in patrilocal chimpanzees and humans." PLoS One **2**(10): e973.

Lavachery, P. (2001). "The Holocene Archaeological Sequence of Shum Laka Rock Shelter (Grassfields, Western Cameroon)." African Archaeological Review **18**(4): 213-247.

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

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

Lee, R. and S. Hurlich (1982). From foragers to fighters: South Africa's militarization of the Namibian San. Politics and history in band societies. E. Leacock and R. Lee. Cambridge, Cambridge University Press**:** 327-345.

Levy, H. (2010). The Human Genetic History of Lesotho. University of Oxford. **Msc**.

Lewis, M. P. and SIL International (2009). Ethnologue : languages of the world. Dallas, Tex., SIL International.

Li, J. Z., D. M. Absher, H. Tang, A. M. Southwick, A. M. Casto, S. Ramachandran, H. M. Cann, G. S. Barsh, M. Feldman, L. L. Cavalli-Sforza and R. M. Myers (2008). "Worldwide human relationships inferred from genome-wide patterns of variation." Science **319**(5866): 1100-1104.

Li, W. H. (1977). "Distribution of nucleotide differences between two randomly chosen cistrons in a finite population." Genetics **85**(2): 331-337.

Librado, P. and J. Rozas (2009). "DnaSP v5: a software for comprehensive analysis of DNA polymorphism data." Bioinformatics **25**(11): 1451-1452.

Lucretius and t. W.H.D. Rouse (1992). On the Nature of Things. M. F. Smith. Cambridge, MA, Harvard University Press. **5**.

Maggs, T. (1976). Iron Age Communities Of The Southern Highveld. Pietermaritzburg, Council of the Natal Museum.

Maggs, T. (1980). "The Iron Age Sequence South of the Vaal and Pongola Rivers: Some Historical Implications." The Journal of African History **21**(1): 1-15.

Malan, J. S. (1995). Peoples of Namibia. Pretoria, Rhino Publishers.

Manfredi, G., D. Thyagarajan, L. C. Papadopoulou, F. Pallotti and E. A. Schon (1997). "The fate of human sperm-derived mtDNA in somatic cells." Am J Hum Genet **61**(4): 953-960.

Marean, C. W., M. Bar-Matthews, J. Bernatchez, E. Fisher, P. Goldberg, A. I. Herries, Z. Jacobs, A. Jerardino, P. Karkanas, T. Minichillo, P. J. Nilssen, E. Thompson, I. Watts and H. M. Williams (2007). "Early human use of marine resources and pigment in South Africa during the Middle Pleistocene." Nature **449**(7164): 905-908.

Margulis, L. (1981). Symbiosis in Cell Evolution: Life and its Environment on the Early Earth. New York, W. H. Freeman & Co.

Marks, S. J., H. Levy, C. Martinez-Cadenas, F. Montinaro and C. Capelli (2012). "Migration distance rather than migration rate explains genetic diversity in human patrilocal groups." Molecular ecology **21**(20): 4958-4969.

Mason, R. J. (1981). "Early Iron Age settlement at Broederstroom 24/73, Transvaal, South Africa." South African Journal of Science **77**: 401-416.

Mazel, A. D. (1986). "Mbabane Shelter and eSinhlonhlweni Shelter: the last two thousand years of hunter-gatherer settlement in the central Thukela Basin, Natal, South Africa." Annals of the Natal Museum **27**: 389-453.

Mazel, A. D. (1989). "People making history: the last ten thousand years of hunter-gatherer communities in the Thukela Basin." Natal Museum Journal of Humanities **1**: 1-168.

McEvoy, B. and D. G. Bradley (2006). "Y-chromosomes and the extent of patrilineal ancestry in Irish surnames." Hum Genet **119**(1-2): 212-219.

Meirmans, P. G. (2006). "Using the AMOVA framework to estimate a standardized genetic differentiation measure." Evolution; international journal of organic evolution **60**(11): 2399-2402.

Meirmans, P. G. and P. W. Hedrick (2011). "Assessing population structure: F(ST) and related measures." Mol Ecol Resour **11**(1): 5-18.

Mishmar, D., E. Ruiz-Pesini, P. Golik, V. Macaulay, A. G. Clark, S. Hosseini, M. Brandon, K. Easley, E. Chen, M. D. Brown, R. I. Sukernik, A. Olckers and D. C. Wallace (2003). "Natural selection shaped regional mtDNA variation in humans." Proc Natl Acad Sci U S A **100**(1): 171-176.

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

Mitchell, P. (2004). "Some reflections on the spread of food-production in southernmost Africa." Before Farming(2004/4): Article 2.

Mitchell, P. (2009). Hunter-Gatherers and Farmers: Some Implications of 1,800 Years of Interaction in the Maloti-Drakensberg Region of Southern Africa. Interactions Between

Hunter-Gatherers and Farmers: from Prehistory to Present. K. Ikeya, H. Ogawa and P. Mitchell. Osaka, Japan, National Museum of Ethnology. **73:** 15-46.

Mitchell, P. (2010). "Genetics and southern African prehistory: an archaeological view." J Anthropol Sci **88**: 73-92.

Mitchell, P., I. Plug, G. Bailey and S. Woodborne (2008). "Bringing the Kalahari debate to the mountains: late first millennium AD hunter-gatherer/farmer interaction in highland Lesotho." Before Farming(2008/2): Article 4.

Mitchell, P. and G. Whitelaw (2005). "The archaeology of Southernmost Africa from c. 2000 BP to the early 1800s: A review of recent research." Journal of African History **46**(2): 209-241.

Mitchell, P. J. (1992). "Archaeological research in Lesotho: a review of 120 years." African Archaeological Review **10**(1): 3-34.

Mitchell, P. J. and R. L. C. Charles (2000). Later Stone Age hunter-gatherer adaptations in Lesotho. Human ecodynamics: proceedings of the conference of the Association of Environmental Archaeology. G. N. Bailey, R. L. C. Charles and N. Winder. Oxford, Oxbow Press**:** 90-99.

Moreau, C., C. Bherer, H. Vezina, M. Jomphe, D. Labuda and L. Excoffier (2011). "Deep human genealogies reveal a selective advantage to be on an expanding wave front." Science **334**(6059): 1148-1150.

Murdock, G. P. (1981). Atlas of world cultures. Pittsburgh, University of Pittsburgh Press ; London : Feffer and Simons.

Nei, M. (1987). Molecular Evolutionary Genetics. New York, NY, USA, Columbia University Press.

Nei, M. (1995). "Genetic support for the out-of-Africa theory of human evolution." Proc Natl Acad Sci U S A **92**(15): 6720-6722.

Newman, J. L. (1995). The peopling of Africa: a geographic interpretation. New Haven; London, Yale University Press.

Nichols, J. (1997). "Modelling ancient population structures and movement in linguistics." Annual Review of Anthropology **26**: 359-384.

Nurse, D. (1997). "The Contributions of Linguistics to the Study of History in Africa." The Journal of African History **38**(03): 359-391.

Nurse, D. and G. Philippson (2003). Introduction. The Bantu Languages. D. Nurse and G. Philippson. London, Routledge**:** 1-12.

Nurse, G. T., A. B. Lane and T. Jenkins (1976). "Sero-genetic studies on the Dama of South West Africa." Ann Hum Biol **3**(1): 33-50.

Nurse, G. T., S. Weiner and T. Jenkins (1985). The Peoples of Southern Africa and Their Affinities. Oxford, Clarendon Press.

Onofri, V., F. Alessandrini, C. Turchi, M. Pesaresi, L. Buscemi and A. Tagliabracci (2006). "Development of multiplex PCRs for evolutionary and forensic applications of 37 human Y chromosome SNPs." Forensic Sci Int **157**(1): 23-35.

Oota, H., W. Settheetham-Ishida, D. Tiwawech, T. Ishida and M. Stoneking (2001). "Human mtDNA and Y-chromosome variation is correlated with matrilocal versus patrilocal residence." Nat Genet **29**(1): 20-21.

Pakendorf, B., K. Bostoen and C. de Filippo (2011). "Molecular Perspectives on the Bantu Expansion: A Synthesis." Language Dynamics and Change **1**: 50-88.

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

Parkington, J. and M. Hall (1987). "Patterning in Recent Radiocarbon Dates from Southern Afirca as a Reflection of Prehistoric Settlement and Interaction." The Journal of African History **28**(1): 1-25.

Parsons, T. J., D. S. Muniec, K. Sullivan, N. Woodyatt, R. Alliston-Greiner, M. R. Wilson, D. L. Berry, K. A. Holland, V. W. Weedn, P. Gill and M. M. Holland (1997). "A high observed substitution rate in the human mitochondrial DNA control region." Nat Genet **15**(4): 363-368.

Pereira, L., V. Cerny, M. Cerezo, N. M. Silva, M. Hajek, A. Vasikova, M. Kujanova, R. Brdicka and A. Salas (2010). "Linking the sub-Saharan and West Eurasian gene pools: maternal and paternal heritage of the Tuareg nomads from the African Sahel." European Journal of Human Genetics **18**(8): 915-923.

Pereira, L., L. Gusmao, C. Alves, A. Amorim and M. J. Prata (2002). "Bantu and European Y-lineages in Sub-Saharan Africa." Ann Hum Genet **66**(Pt 5-6): 369-378.

Pereira, L., V. Macaulay, A. Torroni, R. Scozzari, M. J. Prata and A. Amorim (2001). "Prehistoric and historic traces in the mtDNA of Mozambique: insights into the Bantu expansions and the slave trade." Ann Hum Genet **65**(Pt 5): 439-458.

Phillipson, D. W. (2005). African archaeology. Cambridge, UK; New York, Cambridge University Press.

Pilkington, M. M., J. A. Wilder, F. L. Mendez, M. P. Cox, A. Woerner, T. Angui, S. Kingan, Z. Mobasher, C. Batini, G. Destro-Bisol, H. Soodyall, B. I. Strassmann and M. F. Hammer (2008). "Contrasting signatures of population growth for mitochondrial DNA and Y chromosomes among human populations in Africa." Mol Biol Evol **25**(3): 517-525.

Plaza, S., A. Salas, F. Calafell, F. Corte-Real, J. Bertranpetit, A. Carracedo and D. Comas (2004). "Insights into the western Bantu dispersal: mtDNA lineage analysis in Angola." Hum Genet **115**(5): 439-447.

Pleurdeau, D., E. Imalwa, F. Detroit, J. Lesur, A. Veldman, J. J. Bahain and E. Marais (2012). ""Of sheep and men": earliest direct evidence of caprine domestication in southern Africa at leopard cave (Erongo, Namibia)." PloS one **7**(7): e40340.

Polzin, T. and S. V. Daneshmand (2003). "On Steiner trees and minimum spanning trees in hypergraphs." Operations Research Letters **31**(1): 12-20.

Prins, F. E. (2009). "Secret San of the Drakensberg and their rock art legacy." Critical Arts: A Journal of South-North Cultural Studies **23**(2): 190 - 208.

Prins, F. E. and H. Lewis (1992). "Bushmen as Mediators in Nguni Cosmology." Ethnology **31**(2): 133-147.

Pritchard, J. K., M. T. Seielstad, A. Perez-Lezaun and M. W. Feldman (1999). "Population growth of human Y chromosomes: a study of Y chromosome microsatellites." Mol Biol Evol **16**(12): 1791-1798.

Quintana-Murci, L., C. Harmant, H. Quach, O. Balanovsky, V. Zaporozhchenko, C. Bormans, P. D. van Helden, E. G. Hoal and D. M. Behar (2010). "Strong Maternal Khoisan Contribution to the South African Coloured Population: A Case of Gender-Biased Admixture." American Journal of Human Genetics **86**(4): 611-620.

Quintana-Murci, L., H. Quach, C. Harmant, F. Luca, B. Massonnet, E. Patin, L. Sica, P. Mouguiama-Daouda, D. Comas, S. Tzur, O. Balanovsky, K. K. Kidd, J. R. Kidd, L. van der Veen, et al. (2008). "Maternal traces of deep common ancestry and asymmetric gene flow between Pygmy hunter-gatherers and Bantu-speaking farmers." Proc Natl Acad Sci U S A **105**(5): 1596-1601.

Quintana-Murci, L., O. Semino, H. J. Bandelt, G. Passarino, K. McElreavey and A. S. Santachiara-Benerecetti (1999). "Genetic evidence of an early exit of Homo sapiens sapiens from Africa through eastern Africa." Nat Genet **23**(4): 437-441.

R Development Core Team (2011). R: a language and environment for statistical computing. Vienna, Austria, R Foundation for Statistical Computing.

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

distance in human populations for a serial founder effect originating in Africa." Proc Natl Acad Sci U S A **102**(44): 15942-15947.
Ramos-Onsins, S. E. and J. Rozas (2002). "Statistical properties of new neutrality tests against population growth." Mol Biol Evol **19**(12): 2092-2100.
Redd, A. J., A. B. Agellon, V. A. Kearney, V. A. Contreras, T. Karafet, H. Park, P. de Knijff, J. M. Butler and M. F. Hammer (2002). "Forensic value of 14 novel STRs on the human Y chromosome." Forensic Science International **130**(2-3): 97-111.
Reed, F. A. and S. A. Tishkoff (2006). "African human diversity, origins and migrations." Curr Opin Genet Dev **16**(6): 597-605.
Rexova, K., Y. Bastin and D. Frynta (2006). "Cladistic analysis of Bantu languages: a new tree based on combined lexical and grammatical data." Naturwissenschaften **93**(4): 189-194.
Reynolds, J., B. S. Weir and C. C. Cockerham (1983). "Estimation of the coancestry coefficient: basis for a short-term genetic distance." Genetics **105**(3): 767-779.
Richter, J. (2002-2003). "Archaeology along the Kavango River, Namibia." Southern African Field Archaeology **11/12**: 78-104.
Rocha, J. (2010). "Bantu-Khoisan interactions at the edge of the Bantu expansions: insights from southern Angola." J Anthropol Sci **88**: 5-8.
Rogers, A. (1995). "Genetic evidence for a Pleistocene population explosion." Evolution **49**: 608-615.
Rogers, A. R. and H. Harpending (1992). "Population growth makes waves in the distribution of pairwise genetic differences." Mol Biol Evol **9**(3): 552-569.
Rosenberg, N. A., J. K. Pritchard, J. L. Weber, H. M. Cann, K. K. Kidd, L. A. Zhivotovsky and M. W. Feldman (2002). "Genetic structure of human populations." Science **298**(5602): 2381-2385.
Rosset, S., R. S. Wells, D. F. Soria-Hernanz, C. Tyler-Smith, A. K. Royyuru and D. M. Behar (2008). "Maximum-likelihood estimation of site-specific mutation rates in human mitochondrial DNA from partial phylogenetic classification." Genetics **180**(3): 1511-1524.
Rousset, F. (1997). "Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance." Genetics **145**(4): 1219-1228.
Sadr, K. (1998). "The first herders at the Cape of Good Hope." African Archaeological Review **15**: 101-132.
Sadr, K. (2003). "The Neolithic of southern Africa." Journal of African History **44**: 195-209.
Sadr, K. (2008). "Invisible herders? The archaeology of Khoekhoe pastoralists." Southern African Humanities **20**: 179-203.
Sadr, K. and C. G. Sampson (2006). "Through thick and thin: early pottery in southern Africa." Journal of African Archaeology **4**: 235-252.
Salas, A., M. Richards, T. De la Fe, M. V. Lareu, B. Sobrino, P. Sanchez-Diz, V. Macaulay and A. Carracedo (2002). "The making of the African mtDNA landscape." Am J Hum Genet **71**(5): 1082-1111.
Salido, E. C., P. H. Yen, K. Koprivnikar, L. C. Yu and L. J. Shapiro (1992). "The human enamel protein gene amelogenin is expressed from both the X and the Y chromosomes." Am J Hum Genet **50**(2): 303-316.
Sandelowsky, B. H. (1983). "Archaeology in Namibia." American Scientist **71**: 606-615.
Sandelowsky, B. H., J. H. van Rooyen and J. C. Vogel (1979). "Early Evidence for Herders in the Namib." The South African Archaeological Bulletin **34**(129): 50-51.
Sarich, V. M. and A. C. Wilson (1966). "Quantitative Immunochemistry and Evolution of Primate Albumins - Micro-Complement Fixation." Science **154**(3756): 1563-&.
SAS Institute Inc (2009). JMP Version 8.0. Cary, NC, USA.

Schaller, D. J. (2011). Genocide in colonial South West Africa: The German war against the Herero and Nama, 1904-1907. Genocide of Indigenous Peoples: a critical biographical review. S. Totten and R. K. Hitchcock. New Brunswick, NJ, Transaction Publishers**:** 37-60.

Schapera, I. (1930). The Khoisan peoples of South Africa: Bushmen and Hottentots, London; George Routledge, 1930.

Scheinfeldt, L. B., S. Soi and S. A. Tishkoff (2010). "In the Light of Evolution IV: The Human Condition Sackler Colloquium: Working toward a synthesis of archaeological, linguistic, and genetic data for inferring African population history." Proc Natl Acad Sci U S A **107**(Supplement 2): 8931-8938.

Schlebusch, C. M., T. Naidoo and H. Soodyall (2009). "SNaPshot minisequencing to resolve mitochondrial macro-haplogroups found in Africa." Electrophoresis **30**(21): 3657-3664.

Schneider, S. and L. Excoffier (1999). "Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates very among sites: Application to human mitochondrial DNA." Genetics **152**(3): 1079-1089.

Scholz, C. A., T. C. Johnson, A. S. Cohen, J. W. King, J. A. Peck, J. T. Overpeck, M. R. Talbot, E. T. Brown, L. Kalindekafe, P. Y. Amoako, R. P. Lyons, T. M. Shanahan, I. S. Castaneda, C. W. Heil, et al. (2007). "East African megadroughts between 135 and 75 thousand years ago and bearing on early-modern human origins." Proceedings of the National Academy of Sciences of the United States of America **104**(42): 16416-16421.

Schuster, S. C., W. Miller, A. Ratan, L. P. Tomsho, B. Giardine, L. R. Kasson, R. S. Harris, D. C. Petersen, F. Zhao, J. Qi, C. Alkan, J. M. Kidd, Y. Sun, D. I. Drautz, et al. (2010). "Complete Khoisan and Bantu genomes from southern Africa." Nature **463**(7283): 943-947.

Schwartz, M. and J. Vissing (2002). "Paternal inheritance of mitochondrial DNA." N Engl J Med **347**(8): 576-580.

Scozzari, R., F. Cruciani, P. Santolamazza, P. Malaspina, A. Torroni, D. Sellitto, B. Arredi, G. Destro-Bisol, G. De Stefano, O. Rickards, C. Martinez-Labarga, D. Modiano, G. Biondi, P. Moral, et al. (1999). "Combined use of biallelic and microsatellite Y-chromosome polymorphisms to infer affinities among African populations." Am J Hum Genet **65**(3): 829-846.

Segurel, L., B. Martinez-Cruz, L. Quintana-Murci, P. Balaresque, M. Georges, T. Hegay, A. Aldashev, F. Nasyrova, M. A. Jobling, E. Heyer and R. Vitalis (2008). "Sex-specific genetic structure and social organization in Central Asia: insights from a multi-locus study." PLoS genetics **4**(9): e1000200.

Seielstad, M. T., E. Minch and L. L. Cavalli-Sforza (1998). "Genetic evidence for a higher female migration rate in humans." Nat Genet **20**(3): 278-280.

Semino, O., A. S. Santachiara-Benerecetti, F. Falaschi, L. L. Cavalli-Sforza and P. A. Underhill (2002). "Ethiopians and Khoisan share the deepest clades of the human Y-chromosome phylogeny." American Journal of Human Genetics **70**(1): 265-268.

Sheddick, V. G. J. (1953). The Southern Sotho. London, International African Institute.

Shen, P., F. Wang, P. A. Underhill, C. Franco, W. H. Yang, A. Roxas, R. Sung, A. A. Lin, R. W. Hyman, D. Vollrath, R. W. Davis, L. L. Cavalli-Sforza and P. J. Oefner (2000). "Population genetic implications from sequence variation in four Y chromosome genes." Proc Natl Acad Sci U S A **97**(13): 7354-7359.

Shitara, H., J. I. Hayashi, S. Takahama, H. Kaneda and H. Yonekawa (1998). "Maternal inheritance of mouse mtDNA in interspecific hybrids: segregation of the leaked paternal mtDNA followed by the prevention of subsequent paternal leakage." Genetics **148**(2): 851-857.

Sikora, M., H. Laayouni, F. Calafell, D. Comas and J. Bertranpetit (2011). "A genomic analysis identifies a novel component in the genetic structure of sub-Saharan African populations." Eur J Hum Genet **19**(1): 84-88.

Skaletsky, H., T. Kuroda-Kawaguchi, P. J. Minx, H. S. Cordum, L. Hillier, L. G. Brown, S. Repping, T. Pyntikova, J. Ali, T. Bieri, A. Chinwalla, A. Delehaunty, K. Delehaunty, H. Du, et al. (2003). "The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes." Nature **423**(6942): 825-837.

Slatkin, M. (1995). "A measure of population subdivision based on microsatellite allele frequencies." Genetics **139**(1): 457-462.

Slatkin, M. and R. R. Hudson (1991). "Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations." Genetics **129**(2): 555-562.

Smith, A. (2008). "Pastoral origins at the Cape, South Africa: influences and arguments." Southern African Humanities **20**: 49-60.

Smith, A. B. (1990a). "On becoming herders: Khoikhoi and San ethnicity in southern Africa." African Studies **49**(2): 51-73.

Smith, A. B. (1990b). "The origins and demise of the Khoikhoi: the debate." South African Historical Journal **23**: 3-14.

Smith, A. B. (1992). Pastoralism in Africa: origins and development ecology. London, Athens, Johannesburg, Hurst & Co, Ohio University Press, Witwatersrand University Press.

Smith, A. B. (2005). African herders: emergence of pastoral traditions. Walnut Creek, CA, Alta Mira Press.

Smith, A. B., R. Yates, D. Miller, L. Jacobson and G. Evans (1995). "Excavations at Geduld and the Appearance of Early Dometic Stock in Namibia." The South African Archaeological Bulletin **50**(161): 3-20.

Soares, P., L. Ermini, N. Thomson, M. Mormina, T. Rito, A. Röhl, A. Salas, S. Oppenheimer, V. Macaulay and M. B. Richards (2009). "Correcting for purifying selection: an improved human mitochondrial molecular clock." Am J Hum Genet **84**(6): 740-759.

Soodyall, H. and T. Jenkins (1992). "Mitochondrial DNA polymorphisms in Khoisan populations from southern Africa." Ann Hum Genet **56**(Pt 4): 315-324.

Soodyall, H. and T. Jenkins (1993). "Mitochondrial DNA polymorphisms in Negroid populations from Namibia: new light on the origins of the Dama, Herero and Ambo." Ann Hum Biol **20**(5): 477-485.

Soodyall, H., H. Makkan, P. Haycock and T. Naidoo (2008). "The genetic prehistory of the Khoe and San." Southern African Humanities **20**(1): 37-48.

Steinberg, A. G., T. Jenkins, G. T. Nurse and H. C. Harpending (1975). "Gammaglobulin groups of the Khoisan peoples of Southern Africa: evidence for polymorphism for a Gm1,5,13,14,21 haplotype among the San." American journal of human genetics **27**(4): 528-542.

Stoneking, M. (1998). "Women on the move." Nature genetics **20**(3): 219-220.

Stoneking, M., S. T. Sherry, A. J. Redd and L. Vigilant (1992). "New approaches to dating suggest a recent age for the human mtDNA ancestor." Philos Trans R Soc Lond B Biol Sci **337**(1280): 167-175.

Stringer, C. (2003). "Human evolution: Out of Ethiopia." Nature **423**(6941): 692-695.

Sutovsky, P., R. D. Moreno, J. Ramalho-Santos, T. Dominko, C. Simerly and G. Schatten (1999). "Ubiquitin tag for sperm mitochondria." Nature **402**(6760): 371-372.

Sutovsky, P., R. D. Moreno, J. Ramalho-Santos, T. Dominko, C. Simerly and G. Schatten (2000). "Ubiquitinated sperm mitochondria, selective proteolysis, and the regulation of mitochondrial inheritance in mammalian embryos." Biol Reprod **63**(2): 582-590.

Swallow, D. M. (2003). "Genetics of lactase persistence and lactose intolerance." Annu Rev Genet **37**: 197-219.

Tajima, F. (1983). "Evolutionary Relationship of DNA-Sequences in Finite Populations." Genetics **105**(2): 437-460.

Tajima, F. (1989a). "The Effect of Change in Population-Size on DNA Polymorphism." Genetics **123**(3): 597-601.

Tajima, F. (1989b). "Statistical-Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism." Genetics **123**(3): 585-595.

Tajima, F. (1993). "Measurement of DNA Polymorphism." Mechanisms of Molecular Evolution: 37-59.

Tamura, K. and M. Nei (1993). "Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees." Mol Biol Evol **10**(3): 512-526.

The 1000 Genomes Project Consortium (2010). "A map of human genome variation from population-scale sequencing." Nature **467**(7319): 1061-1073.

The International HapMap Consortium (2007). "A second generation human haplotype map of over 3.1 million SNPs." Nature **449**(7164): 851-861.

The Y Chromosome Consortium (2002). "A nomenclature system for the tree of human Y-chromosomal binary haplogroups." Genome Res **12**(2): 339-348.

Thomson, R., J. K. Pritchard, P. Shen, P. J. Oefner and M. W. Feldman (2000). "Recent common ancestry of human Y chromosomes: evidence from DNA sequence data." Proc Natl Acad Sci U S A **97**(13): 7360-7365.

Tishkoff, S. A., E. Dietzsch, W. Speed, A. J. Pakstis, J. R. Kidd, K. Cheung, B. Bonne-Tamir, A. S. Santachiara-Benerecetti, P. Moral and M. Krings (1996). "Global patterns of linkage disequilibrium at the CD4 locus and modern human origins." Science **271**(5254): 1380-1387.

Tishkoff, S. A., M. K. Gonder, B. M. Henn, H. Mortensen, A. Knight, C. Gignoux, N. Fernandopulle, G. Lema, T. B. Nyambo, U. Ramakrishnan, F. A. Reed and J. L. Mountain (2007a). "History of click-speaking populations of Africa inferred from mtDNA and Y chromosome genetic variation." Mol Biol Evol **24**(10): 2180-2195.

Tishkoff, S. A., F. A. Reed, F. R. Friedlaender, C. Ehret, A. Ranciaro, A. Froment, J. B. Hirbo, A. A. Awomoyi, J. M. Bodo, O. Doumbo, M. Ibrahim, A. T. Juma, M. J. Kotze, G. Lema, et al. (2009). "The genetic structure and history of Africans and African Americans." Science **324**(5930): 1035-1044.

Tishkoff, S. A., F. A. Reed, A. Ranciaro, B. F. Voight, C. C. Babbitt, J. S. Silverman, K. Powell, H. M. Mortensen, J. B. Hirbo, M. Osman, M. Ibrahim, S. A. Omar, G. Lema, T. B. Nyambo, et al. (2007b). "Convergent adaptation of human lactase persistence in Africa and Europe." Nat Genet **39**(1): 31-40.

Tishkoff, S. A. and B. C. Verrelli (2003). "Patterns of human genetic diversity: implications for human evolutionary history and disease." Annu Rev Genomics Hum Genet **4**: 293-340.

Tommaseo-Ponzetta, M., M. Attimonelli, M. De Robertis, F. Tanzariello and C. Saccone (2002). "Mitochondrial DNA variability of West New Guinea populations." Am J Phys Anthropol **117**(1): 49-67.

Torniainen, S., M. I. Parker, V. Holmberg, E. Lahtela, C. Dandara and I. Jarvela (2009). "Screening of variants for lactase persistence/non-persistence in populations from South Africa and Ghana." BMC Genet **10**: 31.

Travis, J. M. J., T. Munkemuller, O. J. Burton, A. Best, C. Dytham and K. Johst (2007). "Deleterious Mutations Can Surf to High Densities on the Wave Front of an Expanding Population." Mol Biol Evol **24**(10): 2334-2343.

Trinkaus, E., C. B. Ruff and G. C. Conroy (1999). "The anomalous archaic Homo femur from Berg Aukas, Namibia: a biomechanical assessment." Am J Phys Anthropol **110**(3): 379-391.

Turner, F. J. (1962). The frontier in American history; with a foreword by Ray Allen Billington. New York, Henry Holt.

Tuupainen, M. (1970). Marriage in a Matrilineal Tribe: A Social Anthropological Study of Marriage in the Ondonga Tribe in Ovamboland. Helsinki.

Tvedten, I. (2004). ""A Town Is Just a Town": Poverty and Social Relations of Migration in Namibia." Canadian Journal of African Studies **38**(2): 393-423.

Underhill, P. A., G. Passarino, A. A. Lin, P. Shen, M. Mirazon Lahr, R. A. Foley, P. J. Oefner and L. L. Cavalli-Sforza (2001). "The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations." Ann Hum Genet **65**(Pt 1): 43-62.

Underhill, P. A., P. Shen, A. A. Lin, L. Jin, G. Passarino, W. H. Yang, E. Kauffman, B. Bonne-Tamir, J. Bertranpetit, P. Francalacci, M. Ibrahim, T. Jenkins, J. R. Kidd, S. Q. Mehdi, et al. (2000). "Y chromosome sequence variation and the history of human populations." Nat Genet **26**(3): 358-361.

United Nations Department of Economic and Social Affairs (2010). Demographic Yearbook 2009-2010**:** Table 3.

United Nations Department of Economic and Social Affairs (2011). World Population Prospects The 2010 Revision, Volume II: Demographic Profiles. New York.

van Oven, M. and M. Kayser (2009). "Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation." Human Mutation **30**(2): E386-E394.

Vansina, J. (1995). "New Linguistic Evidence and 'the Bantu Expansion'." The Journal of African History **36**(2): 173-195.

Venter, J. C., M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, G. G. Sutton, H. O. Smith, M. Yandell, C. A. Evans, R. A. Holt, J. D. Gocayne, P. Amanatides, R. M. Ballew, D. H. Huson, et al. (2001). "The Sequence of the Human Genome." Science **291**(5507): 1304-1351.

Vigilant, L., M. Stoneking, H. Harpending, K. Hawkes and A. C. Wilson (1991). "African populations and the evolution of human mitochondrial DNA." Science **253**(5027): 1503-1507.

Vogel, J., I. Plug and L. Webley (1997). "New dates for the introduction of sheep into South Africa: the evidence from Spoegrivier Cave in Namaqualand." South African Journal of Science **93**: 246-248.

Vogel, J. C. and M. Marais (1971). "Pretoria Radiocarbon Dates I." Radiocarbon **13**(2): 378-394.

Wadley, L. (1979). "Big Elephant Shelter: its role in the Holocene prehistory of central South West Africa." Cimbebasia (B) **3**: 1-76.

Wadley, L. (1996). "Changes in the Social Relations of Precolonial Hunter-Gatherers after Agropastoralist Contact: An Example from the Magaliesberg, South Africa." Journal of Anthropological Archaeology **15**(2): 205-217.

Walker, N. J. and C. Thorp (1997). Stone Age archaeology in Zimbabwe. Caves, monuments and texts: Zimbabwean archaeology today. G. Pwiti. Uppsala, Department of Archaeology and Ancient History, Uppsala University**:** 9-32.

Wallace, D. C., M. D. Brown and M. T. Lott (1999). "Mitochondrial DNA variation in human evolution and disease." Gene **238**(1): 211-230.

Ward, R. H., B. L. Frazier, K. Dew-Jager and S. Pääbo (1991). "Extensive mitochondrial diversity within a single Amerindian tribe." Proc Natl Acad Sci U S A **88**(19): 8720-8724.

Watson, E., P. Forster, M. Richards and H. J. Bandelt (1997). "Mitochondrial footprints of human expansions in Africa." Am J Hum Genet **61**(3): 691-704.

Webb, R. S. (1950). Gazetteer for Basutoland: the first draft of a list of names, with special reference to the 1:250,000 map G.S.G.S. No. 2567 of June 1911: (Basutoland, from a reconnaissance survey made in 1904-9 by Captain M. C. Dobson, R.F.A). South Africa, Paarl, C.P.

Wendt, W. E. (1972). "Preliminary report on an archaeological research programme in South West Africa." Cimbebasia (B) **2**: 1-61.

Wendt, W. E. (1976). "'Art Mobilier' from the Apollo Cave, South West Africa: Africa's Oldest Dated Works of Art." The South African Archaeological Bulletin **31**(121/122): 5-11.

Westphal, E. O. J. (1963). "The Linguistic Prehistory of Southern Africa: Bush, Kwadi, Hottentot, and Bantu Linguistic Relationships." Africa: Journal of the International African Institute **33**: 237-265.

White, T. D., B. Asfaw, D. DeGusta, H. Gilbert, G. D. Richards, G. Suwa and F. C. Howell (2003). "Pleistocene Homo sapiens from Middle Awash, Ethiopia." Nature **423**(6941): 742-747.

Whitlock, M. C. and D. E. McCauley (1999). "Indirect measures of gene flow and migration: FST not equal to 1/(4Nm + 1)." Heredity **82 (Pt 2)**: 117-125.

Wijsman, E. M. and L. Cavalli-Sforza (1984). "Migration and genetic population structure with special reference to humans." Annual Review of Ecology and Systematics **15**: 279-301.

Wilder, J. A., S. B. Kingan, Z. Mobasher, M. M. Pilkington and M. F. Hammer (2004). "Global patterns of human mitochondrial DNA and Y-chromosome structure are not influenced by higher migration rates of females versus males." Nat Genet **36**(10): 1122-1125.

Wilkins, J. F. and F. W. Marlowe (2006). "Sex-biased migration in humans: what should we expect from genetic data?" Bioessays **28**(3): 290-300.

Winney, B., A. Boumertit, T. Day, D. Davison, C. Echeta, I. Evseeva, K. Hutnik, S. Leslie, K. Nicodemus, E. C. Royrvik, S. Tonks, X. Yang, J. Cheshire, P. Longley, et al. (2012). "People of the British Isles: preliminary analysis of genotypes and surnames in a UK-control population." Eur J Hum Genet **20**(2): 203-210.

Wood, B. (2005). Human Evolution: A Very Short Introduction. Oxford, Oxford University Press.

Wood, E. T., D. A. Stover, C. Ehret, G. Destro-Bisol, G. Spedini, H. McLeod, L. Louie, M. Bamshad, B. I. Strassmann, H. Soodyall and M. F. Hammer (2005). "Contrasting patterns of Y chromosome and mtDNA variation in Africa: evidence for sex-biased demographic processes." Eur J Hum Genet **13**(7): 867-876.

Wright, S. (1931). "Evolution in Mendelian Populations." Genetics **16**(2): 97-159.

Wright, S. (1943). "Isolation by Distance." Genetics **28**(2): 114-138.

Xu, X., M. Peng and Z. Fang (2000). "The direction of microsatellite mutations is dependent upon allele length." Nat Genet **24**(4): 396-399.

Zhivotovsky, L. A., P. A. Underhill, C. Cinnioglu, M. Kayser, B. Morar, T. Kivisild, R. Scozzari, F. Cruciani, G. Destro-Bisol, G. Spedini, G. K. Chambers, R. J. Herrera, K. K. Yong, D. Gresham, et al. (2004). "The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time." Am J Hum Genet **74**(1): 50-61.

Zimmerer, J. (2008). Colonial genocide: the Herero and Nama War (1904-8) in German South West Africa and its significance. Historiography of Genocide. D. Stone. Basingstoke, Palgrave Macmillan**:** 323-343.

Zuckerkandl, E., R. T. Jones and L. Pauling (1960). "A Comparison of Animal Hemoglobins by Tryptic Peptide Pattern Analysis." Proc Natl Acad Sci U S A **46**(10): 1349-1360.