FACULTADE DE MEDICINA
E ODONTOLOXÍA
DEPARTAMENTO DE ANATOMÍA
PATOLÓXICA E CIENCIAS FORENSES
INSTITUTO DE MEDICINA LEGAL

Tesis doctoral

Genetic ancestry of the Bolivian population

Tanja Maria Heinz

Santiago de Compostela

2015

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

El Doctor D. Antonio Salas Ellacuriaga, Profesor Titular en la Universidad de Santiago de Compostela,

CERTIFICA:

Que la presente memoria que lleva por título "**Genetic ancestry of the Bolivian population**" realizada por **Tanja Maria Heinz** ha sido realizada bajo mi dirección, considerándola en condiciones para optar al Grado de Doctor y autorizándola para su presentación ante el Tribunal correspondiente.

Y para que así conste, firmamos la presente en Santiago de Compostela, a 10 de febrero de 2015.

Prof. Dr. Antonio Salas Ellacuriaga

Tanja Heinz

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Acknowledegement

This thesis never would have been possible without all the people that I have met during my time here in Galicia. I am very grateful to have had the opportunity to write my PhD thesis abroad here in Spain in Santiago de Compostela.

I first want to thank Antonio for being such a patient, enthusiastic, and dedicated supervisor of my thesis. I am very grateful that I could work together with you in an interesting project in the last three years. I learned a lot from you, from your advices and conversations. Thank you!

I also want to thank Ángel Carracedo, Maviki, and Chris for providing me such a nice environment here in the lab in Santiago to write my thesis. For the success of a thesis it is especially important to be at a place where people treat each other with warmth and respect.

During the three years here in the lab I have met many people from different places and cultures. It was a pleasure for me to meet all of you.

Thank you to Vanesa for your helping instructions in the lab and your contributions to the projects I was working on. I was also lucky that I had colleagues like Ana Mosquera, Ana Freire and Fonde as they always lent a helping hand for me in the lab when I asked them.

I also want to thank all the people that have shared time and room with me during the three years. Jens, it was always a pleasure for me to talk with you to share thoughts about life and writing a Phd thesis abroad. Danke! Thank you to Jorge, Carla, and Jacobo, I very enjoyed working together with you in projects. I learned a lot from you. Thank you to Niaz, María de la Puente, Olalla, Raquel, Meli, Simona, Laura, Danel, Miguel, Luis, Francesca, Yarimar, Carolina, Anna Barbaro, Jeremy, Fabián, Stefanie, Fabio, Cristina, Ilker, Tolga, and Özlem to share either most or some part of my thesis here in the lab together in Santiago.

All of you contributed somehow to my thesis as surrounded by good, respectful, and kind people is important for the success of a thesis.

Thank you to Martin Richards and Maria Pala for sharing your experience about mitochondrial DNA. I had a great time in Huddersfield and I learned a lot from you.

I also want to thank the people of EUROTAST that I could be a part of the whole project. Thank you to Hannes, Tom, Temi, Colleen, Cecilie for organizing and providing us such great training opportunities and symposia. The EUROTAST meetings enriched my thesis in a very important way. Thank you to all my fellows Anu, Cesar, Petr, Jessy, Erna, Hákon, Marcela, Winston, Ryan, Sarah, Judy, and Carlos.

Owing to Patricia, who brought all the samples from Bolivia to Santiago, I could do this thesis. Thank you! I wish we could have met during the time I was here in Santiago.

Thank you to my whole family and all my friends, who always have supported me during the time working and writing on my thesis. Thank you!

Funding

This work has been funded in part by a fellowship of the Marie Curie Actions Initial Training Network (ITN), EUROTAST, granted by the European Union under the Seventh Framework Programme, and coordinated by the University of Copenhagen, Denmark (Grant agreement no.:290344/21.09.2011).

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

La jlor di liri by Juan Angola Maconde (versión afroboliviano)

Tudu lus día salgu a sentá
a la loma di aquí pue ta.
Y sempre mi quedu a mirá
a la mosa di qui dirá

Cuchuqui cho disti andis toy
li jondio cun liriu quisoy.
Cumu alza y lu quepecha
samona pue lui inborracha.

Al sol li pidu apurau
consejo di enamorau,
y al liriu li pidu amor
pa conquistá esa jlor.

Cuasquier jlor qui ta jawirau
in sudor di cuerpu cansau,
chajchura jay pensamentu
y chajaya intendimientu.

Sol dici toy inamorau,
lirio dici jay, soy morau.
Mumas pur su dircio pietu
luz rejuicilia completu.

Qui dirá la linda mosa
qui creció sempre hermosa.
Cuandu lus jlor di la loma
li digan adios paloma.

La flor de lirio by Juan Angola Maconde (version castellano)

Todos los días salgo a sentarme
A la loma de aquí está.
Y siempre me quedo a mirar
A la moza de que dirá.

Tranquilo cho desde donde estoy
Le lanzo con lirio quisoy.
Como lo levanta y lo envuelve
Su perfume pues lo emborracha.

Al sol le pido apurado
Consejo de enamorado,
Y, al lirio le pido amor
Para conquistar a esa flor.

Cualquier flor que está remojada
En sudor de cuerpo cansado,
Riega jay pensamiento
Y adormece el entendimiento.

Sol dice estoy enamorado,
El lirio dice jay, soy morado.
Nomás por su silueta negra
la luz relampaguea completa.

Qué dirá la linda moza
Que creció siempre hermosa.
Cuando las flores de la loma
Le digan adiós paloma.

Table of Contents

1 Abbreviations 17

2 Terminology 21

3 Introduction 23

3.1 Overview of the history of the Transatlantic Slave Trade (TAST) 23

3.1.1 Regions of enslavement in Africa 23

3.1.2 Middle Passage 35

3.1.3 Slave work in the Americas 36

3.1.4 End of the TAST 38

3.2 Interdisciplinary approach 39

3.3 Methods to test for genetic ancestry 40

3.3.1 Overview of methods 40

3.3.2 Uniparental markers 41

3.3.3 Autosomal markers 46

3.3.4 Limitations of DNA testing 49

3.4 Investigating the TAST with genetics 53

3.4.1 Application of genetics 53

3.4.2 The TAST in North America 54

3.4.3 The TAST on the Caribbean islands 57

3.4.4 The TAST in South America 60

3.4.5 The TAST in Central America 64

3.4.6 Genetic ancestry in medical research 65

3.5 Mitochondrial DNA 66

3.5.1 Localization 66

3.5.2 Organizationof the mtDNA genome 68
3.5.3 Differences between the nuclear and mitochondrial genome 70
3.5.4 Numts 70
3.5.5 High mutation rate 71
3.5.6 Maternal inheritance 72
3.5.7 Lack of recombination 73
3.5.8 Nomenclature 74
3.5.9 Human population history 77
3.5.9.1 Overview 77
3.5.9.2 African mtDNA lineages 78
3.5.9.3 American mtDNA lineages 92
3.5.9.4 European mtDNA lineages 93
3.6 Y chromosome 94
3.6.1 Organisation of the Y chromosome 94
3.6.2 Y chromosome polymorphisms 96
3.6.2.1 Y-SNPs 96
3.6.2.2 Y-STRs 106
3.6.3 Paternal inheritance 109
3.6.4 Lack of recombination 110
3.6.5 Human population history 110
3.7 Autosomal DNA 112
3.7.1 Genome-wide SNP genotyping 112
3.7.2 Insertion-deletion polymorphisms (Indels) 115
3.8 Study Population 117
3.8.1 Bolivia 117
3.8.2 History of the TAST in Bolivia 118

3.8.3 Genetic analyses in Bolivia 120
4 Aims 123
5 Materials and Methods 125
5.1 Population samples 125
5.2 Mitochondrial DNA analyses 126
5.2.1 Whole mtDNA genome sequencing 126
5.2.1.1 PCR amplification 127
5.2.1.2 Polyacrylamide Gel Electrophoresis (PAGE) 131
5.2.1.3 Purification of PCR product 132
5.2.1.4 Sequencing 132
5.2.1.5 Purification of sequencing product 134
5.2.1.6 Capillary Electrophoresis (CE) 135
5.2.1.7 Haplogroup classification 135
5.2.2 MtDNA control region sequencing 136
5.2.3 Statistical analyses 136
5.3 Y chromosome 138
5.3.1 Sample preparation 138
5.3.2 Y-STR genotyping 138
5.3.2.1 PCR of Y-STRs 138
5.3.2.2 Capillary Electrophoresis (CE) 139
5.3.3 Y-SNP genotyping 139
5.3.3.1 PCR of Y-SNPs 139
5.3.3.2 Post-PCR purification 141
5.3.3.3 Minisequencing 141
5.3.3.4 Post-extension purification 142
5.3.3.5 Capillary Electrophoresis (CE) 142

5.3.4 Statistical analyses 142
5.4 Autosomal DNA 143
5.4.1 Indels 143
5.4.1.1 AIM-Indelplex 143
5.4.1.2 PCR 146
5.4.1.3 Capillary Electrophoresis (CE) 146
5.4.1.4 Statistical analyses 147
5.4.2 SNPs 148
5.4.2.1 Genotyping 148
5.4.2.2 Statistical analyses 149
6 Results 151
6.1 Mitochondrial DNA 151
6.1.1 MtDNA analysis of Bolivian populations 151
6.1.1.1 Mitochondrial DNA diversity 151
6.1.1.2 Phylogeography of control region mtDNAs 155
6.1.1.3 Phylogenetic tree of whole mtDNA genomes 159
6.1.1.4 AMOVA analysis 162
6.1.2 MtDNA analysis of the Yungas populations 163
6.1.2.1 Mitochondrial DNA diversity 163
6.1.2.2 Phylogeography of control region mtDNAs 163
6.1.2.3 Phylogenetic tree of whole mtDNA genomes 165
6.1.3 Admixture analysis with control region haplotypes 172
6.2 Y Chromosome 173
6.2.1 Y-STR molecular diversity 173
6.2.2 Y-STR AMOVA analysis 175
6.2.3 Y-SNP molecular diversity 177

6.2.4 Y-SNP haplogroup diversity 177
6.2.5 Principal Component Analysis with Y-STRs and Y-SNPs 178
6.2.6 Admixture analysis of Bolvia 180
6.2.7 Geographic origin of Y chromosomes 180
6.3 Autosomal DNA 182
6.3.1 Indel genotyping 182
6.3.1.1 Continental ancestry in the Bolivian population 182
6.3.1.2 Continental ancestry in the Yungas population 187
6.3.2 SNP genotyping 189
6.3.2.1 Analysis of IBS and Principle component analysis 189
6.3.2.2 Admixture analysis of the African component in 'Afro-Bolivians' 193
6.3.2.3 The 'Native American' component in 'Afro-Bolivians' 197
6.4 Gender-bias between the Y chromosome, mtDNA, and AIMs 199
6.4.1 Gender bias in the Bolivian population 199
6.4.2 Gender bias in the Yungas population 201
7 Discussion 205
7.1 Continental ancestry in the Bolivian population 205
7.1.1 Continental ancestry based on mtDNA 205
7.1.2 Continental ancestry based on the Y chromosome 206
7.1.3 Continental ancestry based on autosomes 208
7.2 Geographic origin of 'Afro-Bolivians' 208
7.3 Gender bias in the Bolivian population 213
8 Conclusions 215
9 Resumen 217
10 Literature 227
11 Publications 257

12 Appendix 259
12.1 Samples for mitochondrial phylogenetic analyses 259
12.2 Population data set for whole genome SNP typing 263
12.3 List of Tables 265
12.4 List of Figures 267

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

1 Abbreviations

A:	adenine
AIM:	Ancestry Informative Marker
AMOVA:	Analysis of Molecular Variance
bp:	base pair
C:	cytosine
CE:	Capillary Electrophoresis
CEU:	Utah residents with Northern and Western European ancestry
CHB:	Han Chinese in Beijing, China
CHS:	Southern Han Chinese
CLM:	Colombians from Medellin, Colombia
dATP:	desoxyadenosintriphosphate
dCTP:	desoxycytidintriphosphate
dGTP:	desoxyguanosin-triphosphate
DNA:	deoxyribonucleic acid
dTTP:	desoxythimidintriphosphate
G:	guanine
GBR:	British in England and Scottland
GWAS:	Genome-Wide Association Studies
HVS:	Hypervariable Region of the mtDNA
IBS:	Identity-by-state
Indel:	Insertion Deletion polymorphism
JPT:	Japanese in Tokyo, Japan

Abbreviations

kb:	kilobase
KYA:	thousand years ago
LWK:	Luhya in Webuye, Kenya
n:	number
N_e:	effective population size
NGS:	Next-Generation Sequencing
Mb:	mega base
MSY:	male specific region of the Y chromosome
mtDNA:	mitochondrial DNA
MXL:	Mexican ancestry from Los Angeles, USA
ng:	nanogram
NRY/NRPY:	non recombining region of the Y chromosome
PAGE:	Polyacrylamide Gel Electrophoresis
PAR:	pseudoautosomic region
PCA:	Principal Component Analysis
PCR:	Polymerase Chain Reaction
PUR:	Puerto Ricans from Puerto Rico
rpm:	revolutions per minute
SNP:	Single Nucleotide Polymorphism
SRY:	Sex-determining region protein
STR:	Short Tandem Repeat
T:	thymine
TAST:	Transatlantic Slave Trade
TMRCA:	Time of Most Recent Common Ancestor
TSI:	Toscani in Italia

µl:	microliter
VNTR:	Variable Number Tandem Repeat
YRI:	Yoruba in Ibadan, Nigeria
1000G:	1000 Genomes Project

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

2 Terminology

The following thesis will explore the genetic ancestry and geographic origin of the Bolivian population. Bolivia or Plurinational State of Bolivia as it is officially known is a multiethnic country located in the center of the South American continent. According to the official governmental website, Bolivia has 36 official languages which are reflected in the multiethnic character of the country. Among the Bolivian population persons of African descendants, or 'Afro-Bolivians', respectively, are recognized finally as a constituent ethnic group by the Bolivian government. The term 'Afro-Bolivia' will be used throughout this thesis as it was formerly used by Juan Angola Maconde, author of the article "Los afrodescendientes bolivianos", published in 2008 in *The Journal of Latin American and Caribbean Anthropology*. The terms African descendant and 'Afro-Bolivian' or other 'Afro-American', respectively, are here used interchangeable. In this thesis the 'Afro-Bolivian' community Tocaña will be the main representative group for all these communities of Bolivia whose history is directly linked to the TAST. Threfore, the community Tocaña will be used here in order to investigate more specifically the African ancestry and geographic origin; however, the African ancestry is present in other localities of Bolivia. The aim of this thesis is to contribute to new evidences about the history of the Transatlantic Slave Trade.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

3 Introduction

3.1 Overview of the history of the Transatlantic Slave Trade (TAST)

3.1.1 Regions of enslavement in Africa

The beginning of the history of the Transatlantic Slave Trade can be traced back to the year of 1492 when Christopher Columbus reached in search of India the New World. His explorations and discoveries of the Americas not only initiated the colonization of the New World by European settlers searching for wealth, but also a period of forced migrations of people from the African continent to the Americas. Christopher Columbus' explorations changed considerably the course of the global history and relations among people aroundthe world.

Throughout the course of the Transatlantic Slave Trade some 12.5 million African enslaved people had to embark on ships in ports along the African coast, but only about 10.7 million people reached at the end the Americas. During the peak of the Transatlantic Slave Trade, in the second half of the 18th century, some 60,000 enslaved African people on average reached each year the New World [1]. According to the Transatlantic Slave Trade Database most African enslaved people embarked on ships destined for the Americas between 1751 and 1800 (Figure 1).

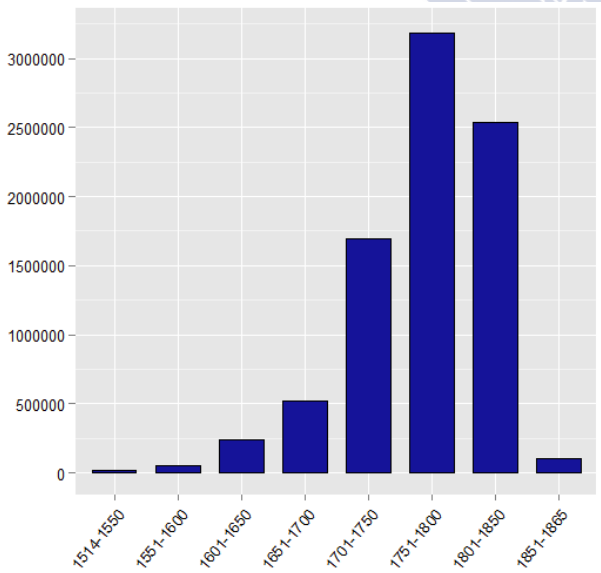


Figure 1: Total number of enslaved African people that embarked in Africa between 1514 and 1866. Statistical presentation is based on information from The Transatlantic Slave Trade Database [1].

The high forced migration of people from Africa can be attributed partially to the labor situation both in Europe, and inthe Americas, respectively. Initially, European colonizers tried to force the indigenous populations of the Americas to work, because labor demand in Europe was high, and thus, wages for European workers in the Americas would have been too costly. Later, due to economic, political, and religious reasons, the first colonizers of the Americas, Portugal and Spain, restricted the exploitation of the indigenous populations. Furthermore, diseases brought in by Europeans to America were particularly epidemic among the indigenous populations [2]. Therefore, the first European colonizers were looking for other possibilities to fill the gaps that the indigenous populations left in mines, plantations, and farms.

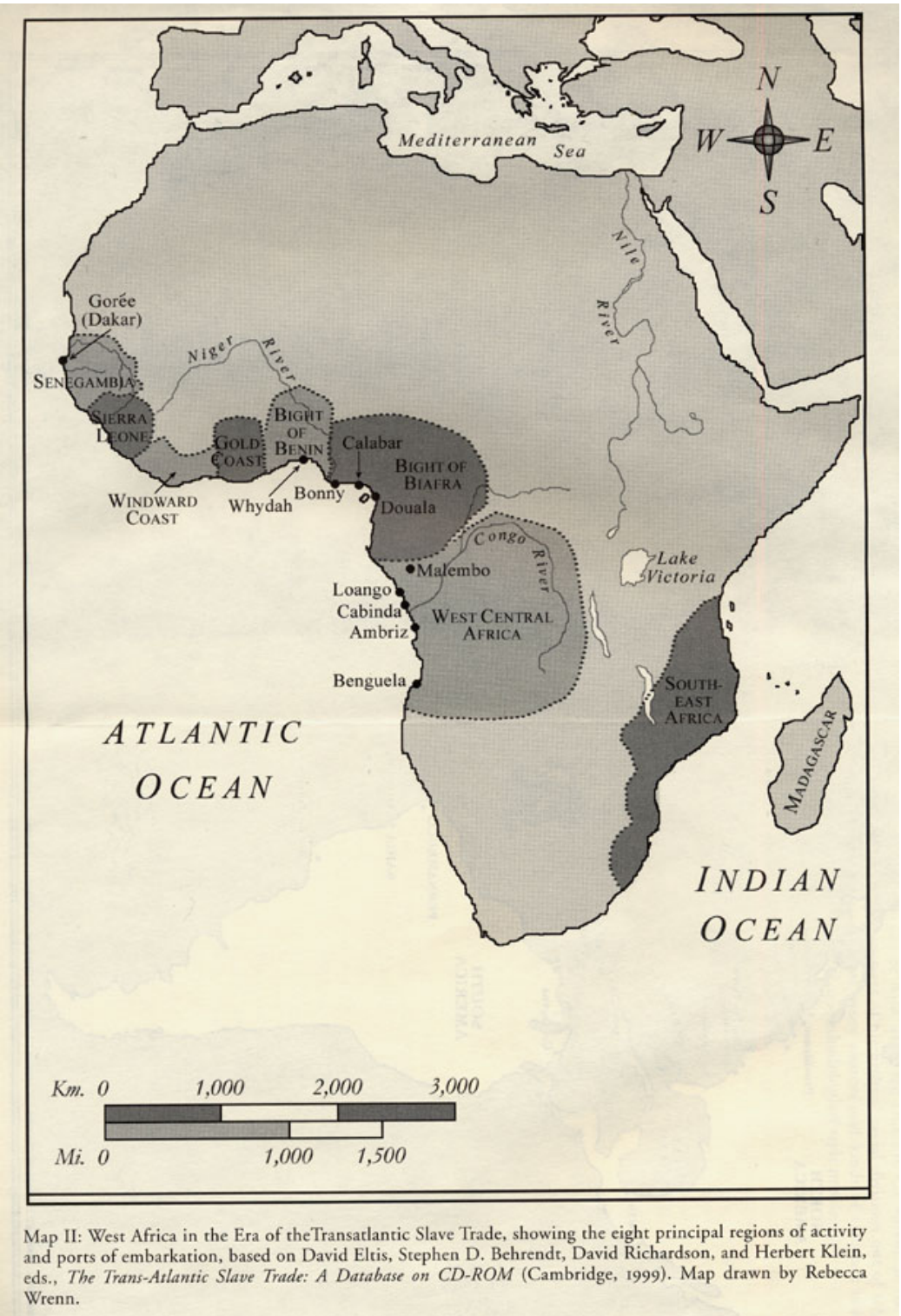
During the same time period of the conquest of the Americas, that is, in the 15th century, the first Europeans also landed on the sub-Saharan African coast. The first Europeans in Africa were the Portuguese that reached the Senegal River in the mid-century. The Portuguese were at first primarily interested in gold and not in slaves [3]. Thus, the first colonizer that transported African enslaved to the Americas people was not the Portuguese, but the Spaniards. The Spanish empire, that maintained good trading relationships with other European empires, was the first that had the capital to import African enslaved people to the Americas. Therefore, the Spaniards began to transport slaves from Africa to the Americas at the beginning of the 16th century which meant that the first African enslaved people were primarily destined to Spanish America at the beginning of the TAST. Portugal, in contrast, began transporting African slaves to the Americas later between the middle and the end of the 16th century. However, before they sent enslaved people to the Americas, Portugal collected first experiences with African enslaved people on their Atlantic islands like São Tomé and Príncipe. The Portuguese also established at the beginning of the TAST access to African labor markets, especially contacts with the Kingdom of Congo which will later provide a considerable number of enslaved people to the colonizer [2].

In the mid 17th century the colonization of the Americas was further extended by an increasing participation of northern Europeans. However, colonization of the Americas by northern Europeans differed from the colonization by Spain and Portugal. The labor situation in northern European countries was worse when compared with Portugal and Spain. This meant that wages for European workers in the Americas were low when the northern Europeans entered the TAST. Consequently, in contrast to the Iberians, the northern Europeans relied principally on indentured contracts at the beginning of their participation in the TAST and not on indigenous workers. As many poor people could not afford their passage to the Americas they sold their workforce to American employers in

indentured contracts. The reason why the northern Europeans had chosen this way of colonizing the Americas was, firstly, because the northern Europeans were not able to exploit a great number of indigenous workers to establish their colonies. Secondly, because northern Europeans did not find valuable resources in their new colonies, they could not trade for enslaved African people. Therefore, completing indentured contracts was the method of choice for the establishment of northern European colonies in the Americas at the beginning of the TAST. At the end of the 17th century economical and labor situation eventually improved in northern European countries which brought them to turn also to African enslaved people as workforce in their colonies [2]. In summary, the labor situation in the Americas and in Europe, but also, the opening up of the African Slave Trade on the West African Coast initiated an intense forced migration of enslaved African people to the New World.

Major regions of slave embarkations in Africa were Senegambia and offshore Atlantic, Sierra Leone, Windward Coast, Gold Coast, Bight of Benin, Bight of Biafra and Gulf of Guinea islands, West-Central Africa and St. Helena, and Southeast Africa and Indian Ocean islands (Figure 2) [1]. Some regions provided more enslaved people than others and the importance of regions in providing enslaved African people to the colonizers varied in the course of the Transatlantic Slave Trade.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



Map II: West Africa in the Era of theTransatlantic Slave Trade, showing the eight principal regions of activity and ports of embarkation, based on David Eltis, Stephen D. Behrendt, David Richardson, and Herbert Klein, eds., *The Trans-Atlantic Slave Trade: A Database on CD-ROM* (Cambridge, 1999). Map drawn by Rebecca Wrenn.

Figure 2: Major Slaving Regions of West Africa, Image Reference ELTIS2, as shown on www.slaveryimages.org, compiled by Jerome Handler and Michael Tuite, and sponsored by the Virginia Foundation for the Humanities and the University of Virginia Library.

Senegambia and offshore Atlantic became the first major region that provided African enslaved people to ships heading to the Americas (Figure 3). Before the Europeans colonized the Americas, African enslaved people from Senegambia were sent principally to the Iberian Peninsula. When the Spaniards began to transport enslaved African people to the Americas, Senegambia was their first choice. In America, more precisely, in Spanish America, most slaves during the first two centuries of the Transatlantic Slave Trademainly originated from Senegambia [4]

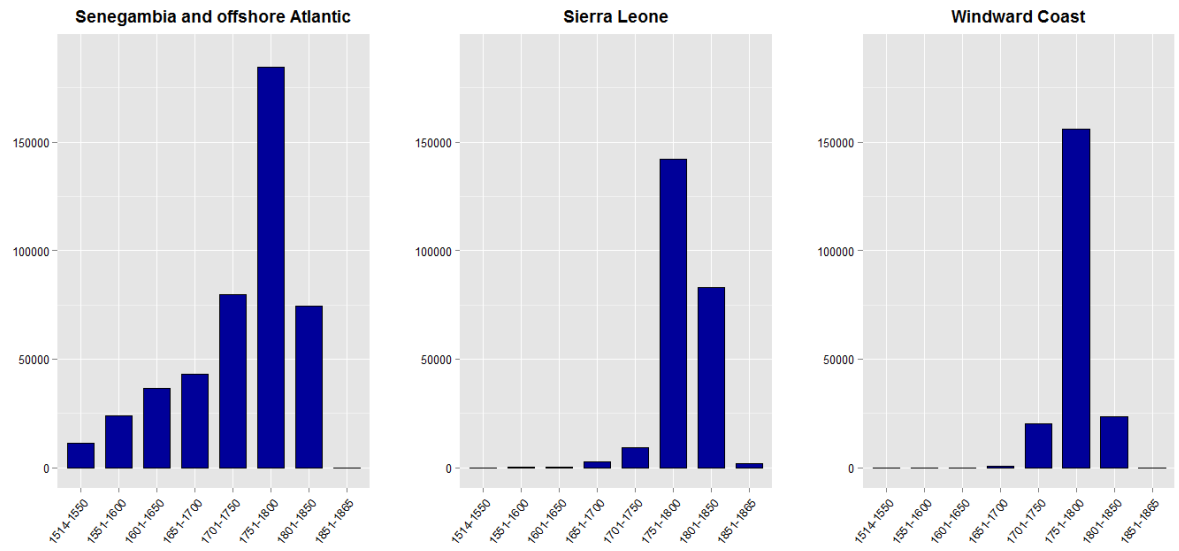


Figure 3: Number of enslaved African people that embarked in West Africa between 1514 and 1866. Statistical presentation is based on information from The Transatlantic Slave Trade Database [1].

Major destinations of Senegambians at the beginning of the TAST were the Caribbean, Mexico, and Peru. In Peru, for examples, 55% of African enslaved people documented between 1560 and 1650 came from Senegambia and only 33% from West-Central Africa. African enslaved people in Peru that arrived in Cartagenas de India, Colombia, principally came from Senegambia whereas African enslaved people from West-Central Africa mainly disembarked in Río de la Plata to further migrate to other regions, including Upper Peru (nowadays Bolivia). Due to their skills in rice and indigo production clusters of African enslaved people from Senegambia formed also across Southern North American colonies like the Carolinas, Georgia, and Louisiana during the 18th century (see chapter 3.4.2). Accordingly, number of slave embarkation in Senegambia and offshore Atlantic increased in the course of this century although estimates of enslaved African people from Senegambia and offshore Atlantic are probably in general underestimated [4].

According to the Transatlantic Slave Trade Database, most enslaved people from Senegambia between 1514 to 1866 were destined for the Caribbean (43.3%), followed by Brazil (25.9%), Mainland North America (14.8%), and Spanish American Mainland (13.5%) [1].

The regions around Sierra Leone and the Windward Coast became like Senegambia and offshore Atlantic increasingly important during the second half of the 18th century (Figure 3). The Windward Coast is widely understood as comprising modern Liberia and Ivory Coast. However, during the TAST, merchants usedthe term Windward Coast more imprecisely meaning someone coming from between the Bight of Benin and Senegambia [4]. The Windward Coast became increasingly important during the second half of the 18th century because of the presence of Dutch traders [5]. Enslaved people from the Grain Coast, including Sierra Leone, were highly demanded in rice-growing regions of the Americas [4], and thus, number of embarkations from this region as well increased especially in the second half of the 18th century.

Most enslaved African people from Sierra Leone (78.4%) and the Windward Coast (88.0%) were estimated to have arrive in the Caribbean between 1514 and 1866 [1]. As estimates of slave embarkations show, Sierra Leone and the Windward Coast did not provide as many slaves as did Senegambia and offshore Atlantic. However, Senegambia and offshore Atlantic became not the leader in numbers of enslaved people, neither.

Similar to western Africa, number of African enslaved people that embarked further central, that is, at the Gold Coast, Bight of Benin, and Bight of Biafra and Gulf of Guinea islands peaked in the 18th century (Figure 4). Most enslaved African people that were forced to migrate to the Americas from the Gold Coast during the 18th century left via trading castles like Elmina, Cape Coast, or Anomabu [3]. Before 1650, the Gold Coast was not a significant region in providing slaves to European traders that left for the Americas as enslaved people were mostly needed at the Gold Coast itself for the production of gold. Most voyages from Africa to America to that time left from Senegambia and West-Central Africa [4]. The reason behind this was a rather unreliable supply of enslaved people from the Gold Coast. Only at the beginning of the 18th century when political conflicts arose, which resulted in several wars at the Gold Coast, a steady flow of enslaved African people was guaranteed [3]. At the Gold Coast European traders were confined to coastal regions, and thus, had not direct influence on the enslavement of people but built, as mentioned above, trading castles along the African coast. Links to the Muslim trade existed in this region, and therefore, commercial connections to the interior were strongly established [3].

Most enslaved people from the Gold Coast that were destined to leave for the Americas were documented in the Caribbean (87.4%), especially in Jamaica [1].

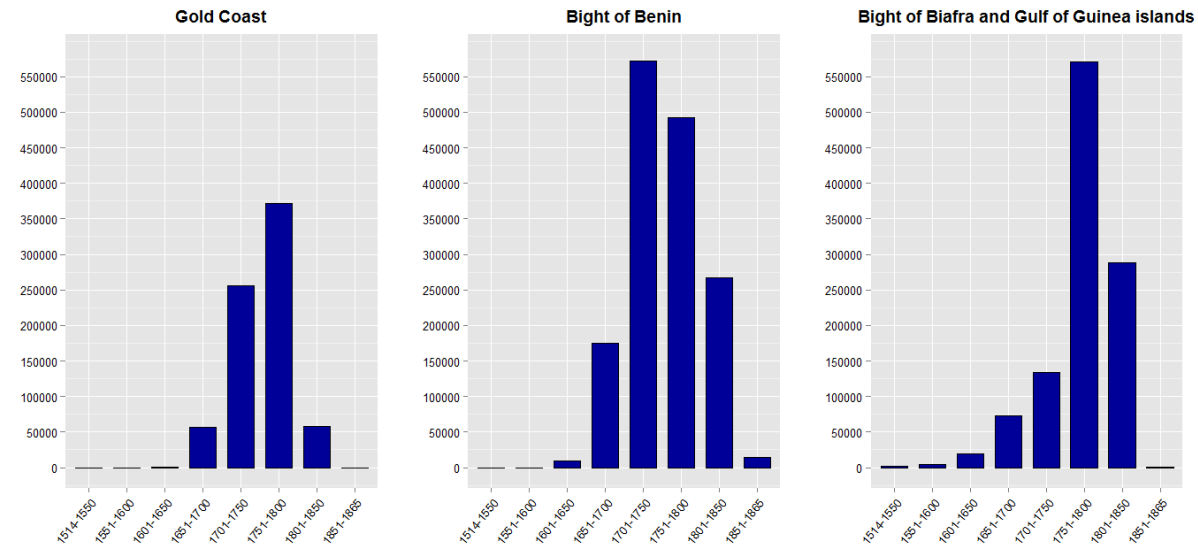


Figure 4: Number of enslaved African people that embarked in Central Africa between 1514 and 1866. Statistical presentation is based on information from The Transatlantic Slave Trade Database [1].

A restricted access to the interior slave trade and a pronounced strong Muslim influence was also found at the Bight of Benin [3]. However, when compared to the Gold Coast the Bight of Benin provided even a greater number of enslaved African people to the Americas, not only in the course of the 18th century, but throughout the centuries (Figure 4). The Bight of Benin, or the "Slave Coast", respectively, became heavily involved in the TAST during 1681 and 1730 when it even outnumbered West-Central Africa with respect to the number of enslaved people (see below). A higher number of enslaved people could be attributed in part to war activities in the interior of the Bight of Benin [3]. After the collapse of the Oyo empire at the beginning of the 19th century and extended Yoruba wars enslavements of Yoruba people rised significantly. During that time period almost half of all enslaved African people that left at ports of the Bight of Benin were Yoruba, numbering nearly one million individuals. Most Yoruba were forced to migrate to Cuba, Bahía (Brazil), and Saint Domingue between 1780 and 1850 [6].

According to the Transatlantic Slave Trade Database most enslaved people between 1514 and 1866 were forced to migrate from the Bight of Benin to Brazil (56.8%) followed by the Caribbean (38.3%) [1].

The Bight of Biafra and Gulf of Guinea islands reached its peak in slave embarkations at the end of the 18th century (Figure 4), especially in the 1780s and 1790s [3]. Due to the environmental conditions found at the Bight of Biafra, European traders did not establish trading castles on the coast like they did at the Gold Coast and the Bight of Benin. Furthermore, European traders did also not encounter strong centralized empires at the Bight of Biafra that confined the Europeans to the coast. The region was politically segmented, and thus, no empire in the scale of those found at the Gold Coast or at the Bight of Benin was established. This meant that most enslaved people came not from wars, but from small-scale raids and kidnapping. Additonally, Muslim influence at the Bight of Biafra was only small [3].

Most enslaved people that left the Bight of Biafra and Gulf of Guinea islands between 1514 and 1866 entered British vessels and came to the Caribbean (78.5%), many to Jamaica, and only a fewer percentage to Brazil (11.4%) [1].

About 40% of all African enslaved people that embarked in ports at the African coast on ships with direction to the Americas came from West-Central Africa. Between 1514 and 1866 more than 3.3 million enslaved people left Africa on the West-Central coast, but less than 3 millions eventually arrived in the Americas [1]. Number of slave embarkation reached its peak in the first half of the 19th century (Figure 5), but slave embarkation was highest during almost the whole period of the TAST when compared to other regions, except during the end of the 17th and the beginning of the 18th century when West-Central Africa was outnumbered by the Bight of Benin. Angola with its principal base Luanda in West-Central Africa was mainly dominated by the Portuguese. Only for a short period of time in the 1640s the Dutch expelled the Portuguese from their base which the Portuguese used as a point of departure for raids into the interior. The region further north of Luanda, that is, at Cabinda and Loango, north of the Congo River, was principally occupied by Dutch and English traders [3].

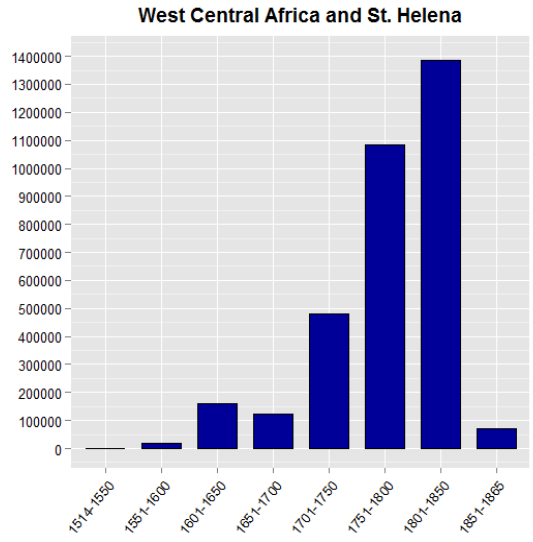


Figure 5: Number of enslaved African people that embarked in West-Central Africa between 1514 and 1866. Statistical presentation is based on information from The Transatlantic Slave Trade Database [1].

In comparison to other regions that provided a great number of enslaved people, West-Central Africa's politics was comparatively strongly influenced by Europeans. Similar to the Bight of Biafra, Muslim influence in West-Central Africa was minimal, or absent, respectively. In other parts of Africa, Europeans were confined to the coast, but in West-Central Africa the Portuguese established extended trading routes far into the interior [3, 4]. When the Portuguese landed in West-Central Africa they gradually established trading relations with the kingdom of Kongo. At first the Portuguese merchants did not trade for enslaved people, but in the 1490s, after settling the Atlantic island of São Tomé, the European trader began to ship enslaved people to their newly established sugar plantations [4]. Later, in the mid-16th century, the Portuguese moved further south to the port of Luanda where they penetrated deeply into the interior in order to capture enslaved people that arose as by-products of wars. The high number of African enslaved people in this region, made Luanda finally to the most important slave-providing region for both Portuguese, but also Brazilian traders, respectively. Trading relations between Luanda and Brazil eventually grew very strong. During 1580 and 1640 when the Portuguese crown was merged with the Spanish crown, many enslaved people that left Luanda did not only reach Portuguese colonies, but also Spanish America [2, 4].

In contrast to Luanda, which was principally dominated by the Portuguese, northern Europeans like the Dutch, English, and French, took the region at the Loanga Coast, north of the Congo River, under their control. Unlike the Portuguese at Luanda, Europeans could not establish trading posts at Loango, but they had to trade from their vessels. Trading of enslaved African people at the Loanga Coast developed later, that is, at the end of the 17th century. Given the close geographical location, slave trading networks at both Luanda and Loango, respectively, overlapped. Nonetheless, African enslaved people that embarked at Loango principally were Kikongo speakers and were forced to migrate to the United States, North America, or to the Caribbean whereas enslaved people that departed from Luanda mainly were Kimbundu speakers and sent to either Brazil or to the southeast coast of Spanish America, respectively [4].

According to the TAST Database, enslaved Africans that had to leave West Central Africa and St. Helena between 1514 and 1866 arrived principally in Brazil (63.7%), followed by the Caribbean (27.8%) [1].

Southeast Africa became increasingly important in providing slaves at the end of the TAST (Figure 6) whereby Mozambique was the principal region from where enslaved people embarked on ships [4]. During Dutch occupation of Luanda in west central Africa in the 1640s, the Portuguese turned to Southeast Africa in order to provide their colonies in the Americas with slaves. Number of enslavements increased in the second half of the 18th century when French merchants bought slaves for the Mascarene Islands, but some also for their Caribbean colonies, especially for St. Domingue/Haiti. However, after the slave revolt on Haiti in 1791, the slave trade in Mozambique became principally dominated by both Portuguese, and Brazilians, respectively, but extent of control of the trade in Mozambique by Portuguese was weak when compared to West-Central Africa [4].The reason why Mozambique became an increasingly important source of enslaved people at the end of the TAST might be due to the anti-slave trade patrols that were less effective in the Indian Ocean. Before 1842 anti-slave trade activity was not present below the equator.

According to the TAST Database, slaves that left southeast Africa and Indian Ocean islands between 1514 and 1866 had to principally migrate to Brazil (73.2%), and a fewer percentage to the Caribbean (18.0%) [1].

Some African enslaved people could not be specified to a major region within Africa (Figure 6).

At the end, the majority of enslaved African people that embarked at ports from Senegambia along the coast to southeast Africa ended in the Caribbean (51.5%), followed by Brazil (38.9%) (Figure 7 and 8).

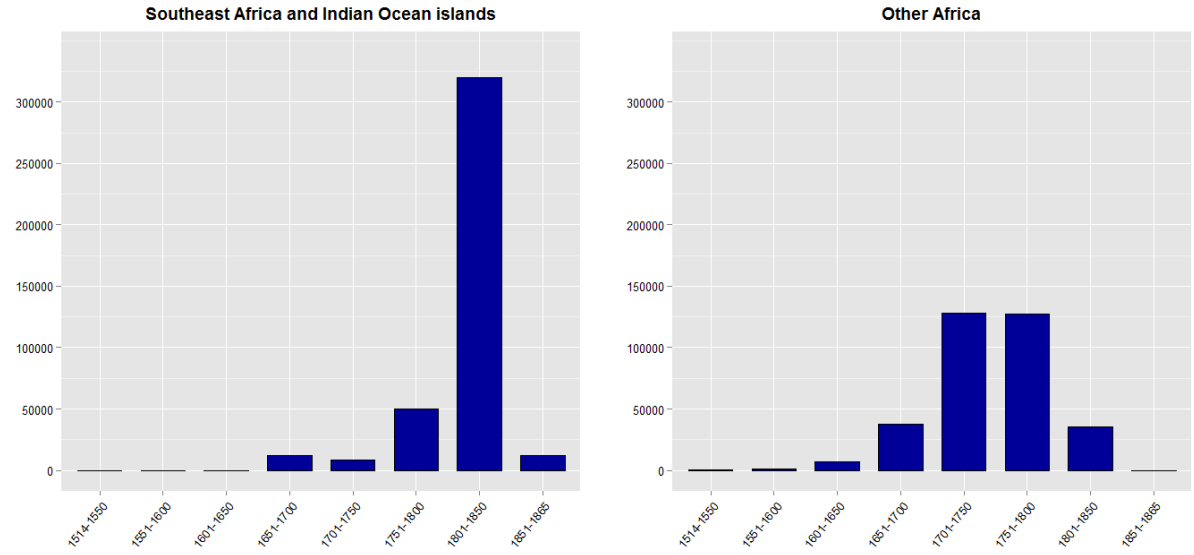


Figure 6: Number of enslaved African people that embarked in Southeast and other parts of Africa between 1514 and 1866. Statistical presentation is based on information from The Transatlantic Slave Trade Database [1].

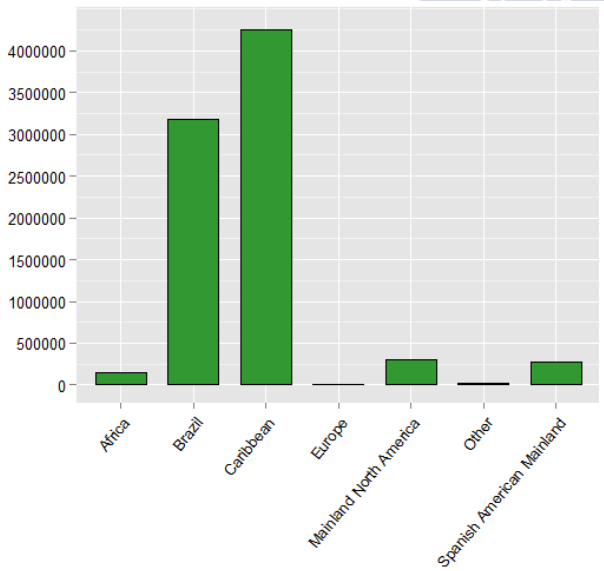


Figure 7: Number of African enslaved people that disembarked in the Americas (and Africa) from 1514 to 1866. Statistical presentation is based on information from The Transatlantic Slave Trade Database [1].

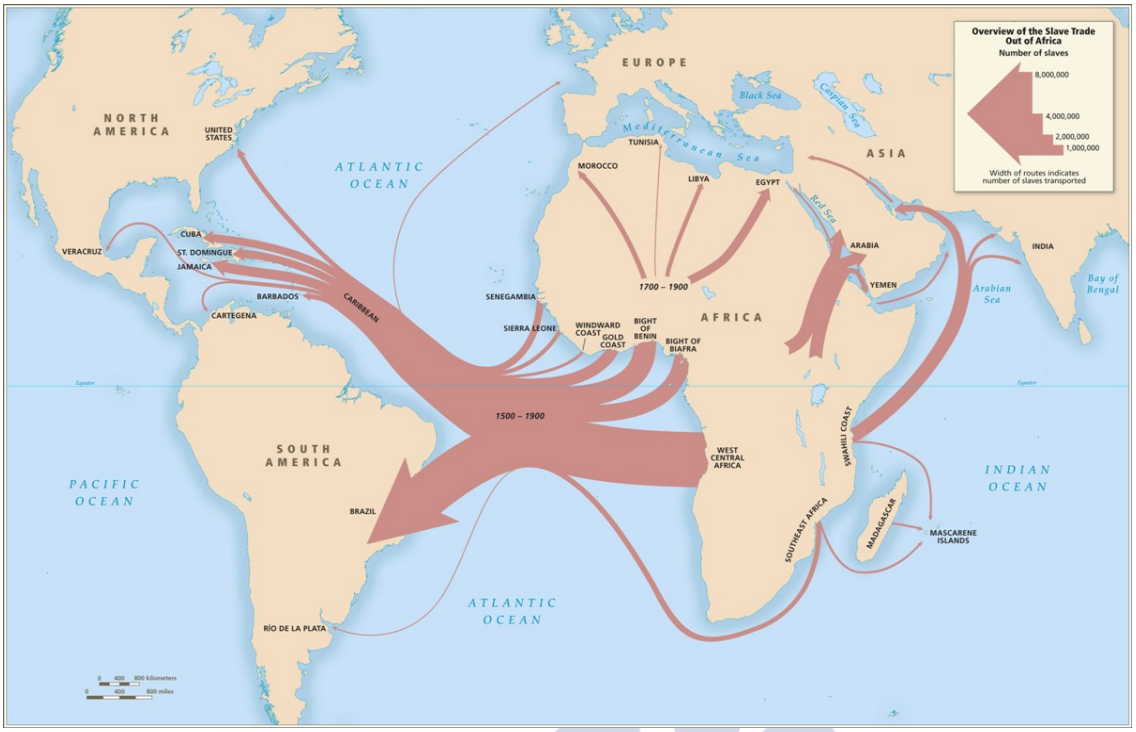


Figure 8: Overview of the Slave Trade out of Africa, 1500-1900 [7].

When the first Europeans arrived in Africa, slavery had already existed throughout Africa and other parts of the world like in Europe and Asia. However, when compared to other slave trades like the Trans-Saharan or Red Sea Slave Trade, the TAST stands out considerably regarding numbers of enslaved people [3]. The forced displacement of an incredibly high number of people significantly had impacted African's demography. More than 60% of African enslaved people were adult males, although about 20% were children (including both genders) [1]. Planters in America were principally interested in prime-age males as workers, but many times, depending on the market, planters had to leave their preferences to purchase women, children, or aged slaves [2].

3.1.2 Middle Passage

The journey that the African enslaved people had to undertake is known as the "Middle Passage". Depending on the region, a journey from Africa to the Americas took on average about 60 days, whereas voyages from West-Central Africa or Senegambia were shorter than voyages from the Gold Coast or Southeastern Africa [1]. However, most African enslaved people usually had to wait several months from capture onwards until boarding the ship. Mortality rate during the Middle Passage was highest at the beginning of the TAST and declined on ships of all nations towards the 19th century. Average mortality rate came to 12.4% whereas the lowest rates were achieved on ships originating from Congo and Angola. This can be possibly attributed in part to the fact that voyages from West-Central Africa to the Americas were among the shortest ones. However, other local factors like diseases and political or ecological crises that influenced food availabilitymight have been another important influential circumstance why the mortality rate of a region was higher than another one as poor nutrition could further weaken enslaved people on their journey to the Americas [2].

At the height of the TAST the European colonizers carried between 350 and 450 African enslaved people per voyage to the Americas with ships ranging from 150 to 250 tonnage [2]. Men, women, and children were usually placed below the deck (Figure 9).

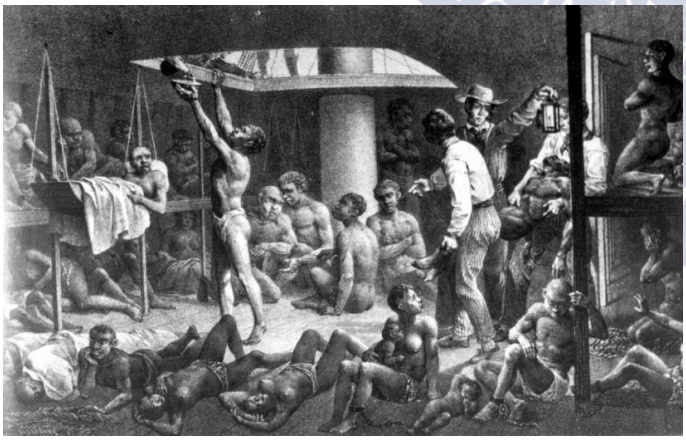


Figure 9: Enslaved Africans in Hold of Slave Ship, Image Reference E019, as shown on www.slaveryimages.org, compiled by Jerome Handler and Michael Tuite, and sponsored by the Virginia Foundation for the Humanities and the University of Virginia Library.

In general, tonnage of slave vessels increased during the course of the TAST concurrently with declining slaves-per-ton ratios and a higher number of African enslaved people on the ships. The

Europeans were certainly encouraged in reducing mortality rates of enslaved people, for example, by providing more space to each slave. However, the reason behind that was not a humanitarian one but rather an interest for good profit [2]

The TAST is known for having operated as a triangular trade, that is, a trade beginning in Europe to bring European goods to Africa, followed by the transportation of African enslaved people to the America, and finally carrying sugar back to Europe. However, in reality ships, depending on the cargo, usually operated only ona specific route. For example, usually large vessels moved from America to Europe as they carried a high amount of American products. A further discrepancy to a triangular trade was that many slavers from Brazil only travelled between Brazil and Africa by leaving out Europe [2].

After the African enslaved people had arrived the Middle Passage, many of them had to further migrate long distances in order to reach their final place were they were forced to work for the European colonizers.

3.1.3 Slave work in the Americas

European colonizers often showed preferences for slaves that came from a specific region in Africa, sometimes associated with where a specific skill is widely spread. For example, the colonizers of Colombia favored experienced gold miners from the Gold Coast, whereas the Spaniards in Peru highly prized enslaved African people from Senegambia [4]. In contrast, planters from Jamaica purchased enslaved people with different African origins [8] although price and availability of slaves often determined purchases. Especially African enslaved people originating from West-Central Africa were available in great numbers, and thus, were sold at low prices to all places [4].

In the Americas African enslaved people were forced to work in a wide range of fields. Some were sent on sugar, rice, tobacco, coffee, cocoa, or cotton plantations while others had to give their workfoce to the colonizers in gold, diamond, or silver mines. Some African enslaved people also worked in the construction industry or as domestic servants. In some cases they also had to perform skilled labour or worked on farms. However, the task African enslaved people had to carry out depended to a great extent also on the region in the Americas [2].

For example, sugar plantations based on African slave labour were expecially common in Brazil and on the Caribbean islands. Major zone of sugar production in Brazil included Bahia and

Pernambuco. Sugar production in the Caribbean began on smaller islands like Barbados within the British and Martinique within the French empire, but moved to larger estates, that is, to Jamaica and Saint Domingue later during TAST. Other important islands in the Caribbean, where sugar productionwas important and maintained by forced labour of African people, included Santo Domingo, Martinique, Cuba (Figure 10), and Puerto Rico among other islands [2].

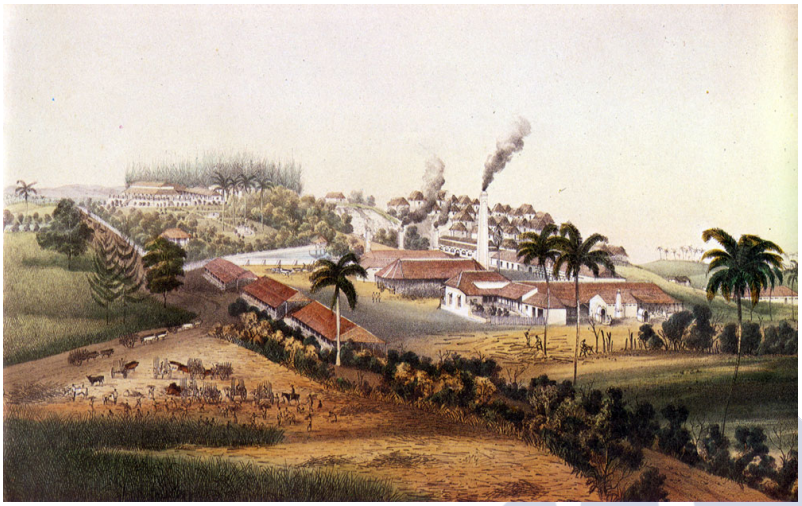


Figure 10: Sugar plantation, San José de la Angosta, Cuba, 1857, Image Reference Cantero1, as shown on www.slaveryimages.org, compiled by Jerome Handler and Michael Tuite, and sponsored by the Virginia Foundation for the Humanities and the University of Virginia Library.

One of the major coffee production centers were Cuba and Saint Domingue, but also Brazil again. Cocoa plantations were found in Venezula where African enslaved people replaced completely Indian laborers by the end of the 18th century.Tobacco production based on forced African slave labour could be found especially in the Chesapeake Bay (Virginia, Maryland, and parts of North Carolina) in the United States [2]. In the neighbouring region, that is, in South Carolina and Georgia, rice production was the main production crop [9, 10] although rice, but also tobaccoproduction in the Chesapeake Baywere replaced later by cotton production. Even new regions within the United States (Alabama, Mississippi, Louisiana) were settled to produce more cotton which was again maintained with slave labour. But in contrast to other regions like Brazil or the Caribbean, the slave population in the United States grew substantially internally which resulted in a lower number of slave imports from Africa. Therefore, it was common within the United States not to import more and more slaves but to relocate them depending on were slave work was needed [2].

Gold mining developed especially in the Choco region in nowadays northeastern Colombia. Gold, but also diamond mining based on an African slave workforce took place also in Minas Gerais, Goias, Bahia, and Mato Grosso in Brazil although fewer African people were used for this kind of work in the country when compared to the sugar production industry [2]. When rich silver deposits were found in Upper Peru (nowadays Bolivia) many African enslaved people were forced to migrate to Potosí in order to work for the colonizers in the royal mint [11].

In cities of Spanish mainland African enslaved people were also well represented in the construction industry where they were involved both in unskilled, but also skilled work. Skilled African enslaved people were especially found not only in construction, but also in metalworking and clothing [2]. African enslaved people were forced to work as domestic servants, for example, around Potosí in Upper Peru [11]. Another work area, however, small when compared to plantation slave labour, was the grazing and cattle industry found, for example, in Southern Brazil [2].

Nonetheless, the main task African enslaved people had to perform depended also partially on the period when the new slaves arrived at a region. For example, Brazil was principally a major sugar producer, but after the collapse of Saint Domingue, coffee production increased substantially, and thus, even exceeded sugar exports resulting in many African enslaved people found on coffee plantations [2].

3.1.4 End of the TAST

The TAST ended ultimately in the year 1867 although as early as 1830 the transportation of African enslaved people to the Americas was illegal in all parts of America, whereby in some regions the TAST was prohibited even before [2]. Nonetheless, after 1830 still more than 700,000 African enslaved people embarked on ships heading for the Americas [1].

Efforts to abolish the TAST arose in the British colonies, and thus, the first colonies in which the slave trade was prohibited were the British ones in 1808. After 1815 only the Portuguese and the Spaniards continued actively the slave trade, although the Portuguese prohibited slave trading north to the equator meaning that until 1830 mainly Brazil and Spanish colonies continued to import African enslaved people [2]. The number of African enslaved people that disembarked in the Americas towards the end of the TAST was especially high for Brazil, Cuba, and Puerto Rico [1].

In Brazil slave trading became illegal after 1830, but only after 1850 because ofthe British interventions the importation of African enslaved people stopped. Slave trading in Cuba was even earlier illegal, that is, after 1820, but the slave trade to Cuba was the last one to be terminated during the TAST. The slave trade terminated after the British and the U.S. allied against Spain who signed an anti-slave trade act in fear of losing its colony Cuba. Thus, the last ships that brought African enslaved people to the Americas arrived in 1967 in Cuba [2].

3.2 Interdisciplinary approach

About 150 years have passed since the end of the TAST in 1867 and there are still many gaps to fill about how the trade operated in the course of its history.Different approaches have been used to disentangle this complex history ranging from history, anthropology, social sciences [3, 4], toarchaeology [12–15]. However, the development of new technology has opened up new possibilities not onlyin the natural sciences. It can also be used to investigate past historical events, including the TAST. Only in the last few years some research has been published in which the TAST from a genetic point of view has been investigated [16–19, 39].

If historical documents lack or African cultural practices became lost or mixed in the Americas, genetics can provide further information of one's biological ancestors and geographic origin in Africa taking into consideration that the TAST was responsible for kidnapping millions of people from their homes to force them to adapt into a different environment on a different continent.African enslaved people encountered in the Americas not only their European enslavers, but also Native Americans that had already lived there before the Europeans arrived. Genetics here can provide information about where one's individual ancestor came from in Africa. Tracing back one's African ancestors is especially popular in the United States [21].

Genetics research can be used not only on an individual, but also on a population level like it is done here in this thesis. It has to be emphasized here, however, that genetics research will notreplace traditional areas of TAST research. On the contrary, genetics can be understood rather as being complementary to other research fields. In this thesis genetics will be used in order to contribute to the complex history of the TAST, more precisely, to the history of the TAST in Bolivia.

3.3 Methods to test for genetic ancestry

3.3.1 Overview of methods

Genetic ancestry testing can be conducted with different types of genetic material from the human genome. The human genome is the whole genetic material of a person and comprises genetic information that is stored both as DNA sequence on chromosomes (Figure 11) in the cell nuclei, but also on mitochondrial DNA in mitochondria (Figure 24, chapter 3.5.2). The nuclear genome comprises 23 pairs of linear chromosomes that range from ~50 to 250 Mb in size.In total, the nuclear diploid human genome, that is, one haploid set of chromosomes from the father and one from the mother consists of 6 billion base pairs. Each autosomal pair of chromosomes, that ischromosomes 1 to 22, contains the same genetic information for traits. That meansthat a gene at the same position on the maternal chromosome and paternal chromosome determines the same trait (e.g. hair color) and variations of these gene pairs (alleles) result in differences in the manifestation of traits, or phenotypes, respectively (e.g., black, brown, red, or blond hair). The chromosome pair 23 represents the sex-determining chromosomes. A female has two X chromosomes whereas males have one X and one Y chromosome. However, Y chromosomes are passed on to the next generations only by the father, and moreover, only to his sons and not to his daughters. A daughter only gets the X chromosome from her father [22].

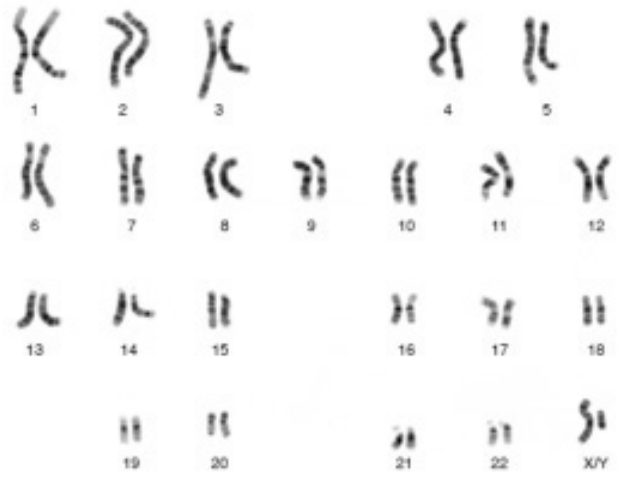


Figure 11: 23 pairs of chromosomes of the human nuclear genome. Shown is a male with one X and one Y chromosome. http://www.genome.gov/Glossary/index.cfm?id=114.

The mitochondrial genome, in contrast,is circular and reaches only ~16,5 kb (Figure 24, chapter 3.5.2). Both males and females have mitochondrial DNA, however, compared with the nuclear genome, the mitochondrial genome is like the Y chromosome inherited not by both parents, but in this case solely by the mother.

The human nuclear genome comprises in total about 20.000 genes, including the Y chromosome with 27 genes, whereas the mitochondrial genome is made up of only 37 genes. Genes contain instructions for proteins encoded as DNA sequence. DNA sequences are composed of the four bases adenine, cytosine, guanine, and thymine to form the respective nucleotides, that is, dATP (Desoxyadenosintriphosphate), dCTP (Desoxycytidintriphosphate), dGTP (Desoxyguanosin-triphosphate), and dTTP (Desoxythimidintriphosphate). Both DNA sequences of the nuclear and mitochondrial genome form double strands made of two complementary single-stranded DNA molecules, whereas adenine (A) pairs with guanine (G) and thymine (T) with cytosine (C). Before cell division the genetic material has to be doubled in order to distribute equally DNA, and thus, genetic information from the mother cell to to each daughter cell. The process of producing identical DNA molecules, which is known as DNA replication, can result in mistakes where, for example, a G instead of a T is incorporated in the newly created DNA double strand. These mistakes, called mutations, are useful for genetic ancestry testing, as they result in variations between individuals and populations [22].

Genetic ancestry tests can be divided, firstly, into analyses that look for the maternal (mtDNA) and/or paternal lineage (Y chromosomes), and secondly, into analyses that uses ancestry informative markers (AIMs) on autosomal DNA (1-22 chromosomes). Uniparental markers (mtDNA, Y chromosome) are powerful tools to investigate human evolution and geographic origin, but also toconduct genealogical research. Autosomal markers are the method of choice to investigate genetic ancestry, but alsobiological geographic origin.

3.3.2 Uniparental markers

In contrast to autosomal markers, uniparental markers make up solely about 1%of the human genome given that the Y chromosome consists of about 60 Mb and the mtDNA as mentioned above of about 16,5 kb. Uniparental markers can be used to trace back either the maternal or the paternal lineage, as mtDNA is transmitted only by females (maternal lineage) and Y chromosomes only by males (paternal lineage) (See Figure 12 and Figure 13). The difference between mitochondrial DNA

and the Y chromosome is that the former can be transmitted from the mother to both daughters and sons, respectively, and the latter from father only to sons but not to daughters. However, the maternal lineage terminates when the mtDNA is transmitted to a son [22].

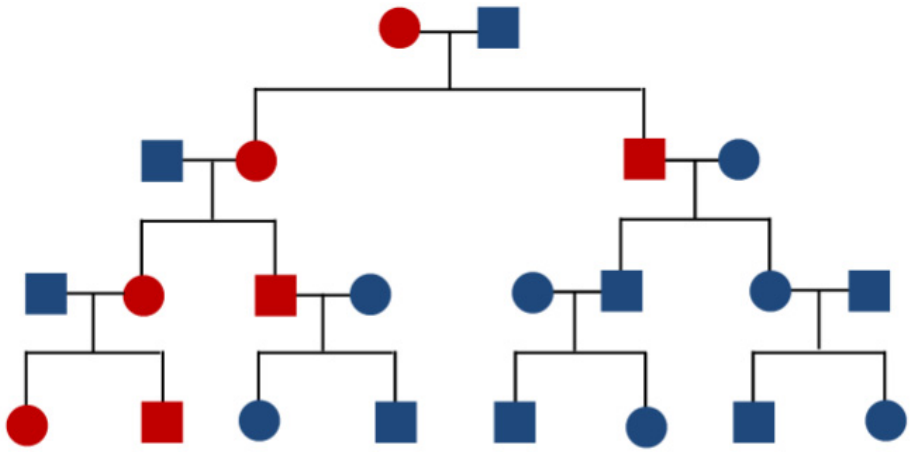


Figure 12: Matrilineal inheritance of mtDNA. Females are shown in circles, males in squares. Red indicates the path of the mitochondrial DNA.

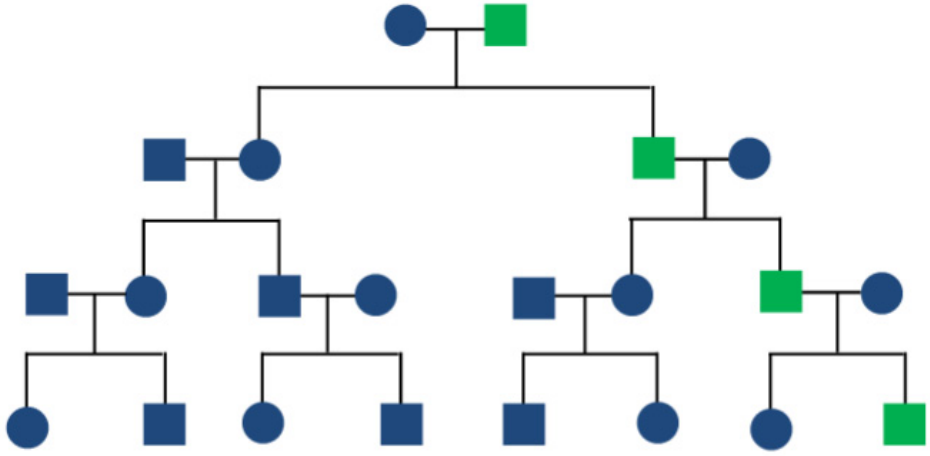


Figure 13: Paternal inheritance of the Y chromosome. Females are shown in circles, males in squares. Green indicates the path of the Y chromosome.

Both markers are useful to investigate human evolution like past migrations because they lack recombination. Due to the lack of recombination there exist a single phylogeny for both the mtDNA and the Y chromosome. Mitochondrial DNA does not recombine, because the mitochondria from the father in the sperm apparently do not survive after the egg was fertilized although

recombinationwas reported in previous, but few studies [23, 24]. Most parts of the Y chromosome (>90%) neither recombine. These parts are called the nonrecombining portion of the Y chromosome (NRPY or NRY) (see chapter 3.6.4) [22].

As mtDNA and Y chromosomes have a lower effective population size (N_e) than autosomal DNA, both uniparental markers are more sensitive to the effects of genetic drift. Genetic drift means a random fluctuation in allele frequency within the gene pool of a population. Genetic drift is more pronounced in smaller populations meaning that genetic drift in smaller populations rather give rise to the fixation or the elimination of an allele. Thus, genetic drift can cause a migrating population separating from another population to become genetically distinct from its original population faster when the populations are smaller. In contrast to the census population size that gives the present number of individuals in a population to a given time, the effective population size allows to compare genetic drift experienced by different populations. A lower effective population size indicates a greater genetic drift [25].A lower effective population size means that the number of individuals that contribute genes to the next generation is correspondingly smaller. The effective population size in mtDNA and the Y chromosome is lower when compared to autosomes, because four copies for each autosome (a diploid set of chromosome from each parent during recombination) contrasts to only one copy of Y chromosome and one copy of mtDNA that are transmitted to the next generation. Thus, as the effective population size (Ne) in both markers is lower than in autosomal DNA, the mtDNA and the Y chromosome experience a greater genetic drift and are well suited tools to detect geographic population structure [22].

Another reason why uniparental markers are useful tools is that both the mtDNA and the Y chromosome share a higher mutation rate when compared to autosomes. The mutation rate for the mitochondrial DNA can reach up to 5×10^{-6} per bp per generation in the control region [22] although discrepancy of values differ in literature [26–29]. A value of 3×10^{-8}per bp per generation was estimated for the Y chromosome [30]. In contrast, autosomes show a mutation rate of about 10^{-8} per bp per generation [31].

A high mutation rate combined with a low effective population size and high genetic drift, respectively, facilitates the detection of genetic variation between populations. Due to migration of human populations, groups diversify over time and over geographical distance, and thus, like it is the case for the mtDNA can be classified in groups to detect human migration patterns (Figure 14).

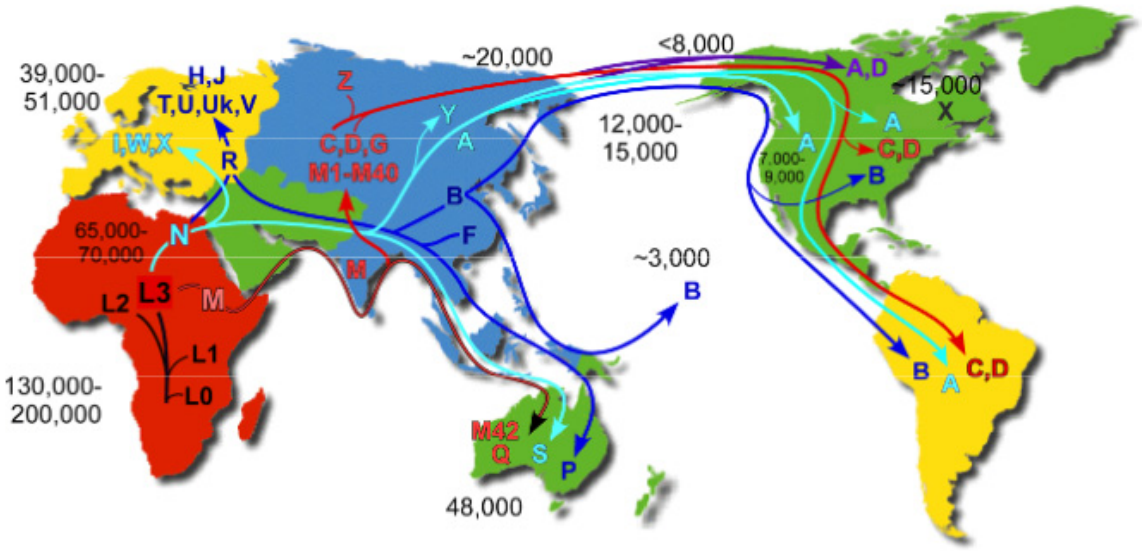


Figure 14: Global human mtDNA migrations, http://www.mitomap.org/ [32].

An example of how a uniparental marker can be used to infer geographic origin is shown in Figure 15 in which different African ethnic groups (numbered circles) have several haplotypes of the Y chromosome haplogroup E in common (big Network circles). Less frequent haplotypes (smallerNetwork circles) are obviously specific to certain groups although some groups, for example, the Yoruba and the Urhobo in Nigeria, or the Kru and Kissi in the region around Sierra Leone also share less frequent paternal lineages [33]. Based on these observations a person like, for example, from the Americas that has ancestors due to the TAST in Africa can trace back one's biological ancestor on the African continent given that this person shows a less frequent haplotype which is found in only a specific region in Africa.

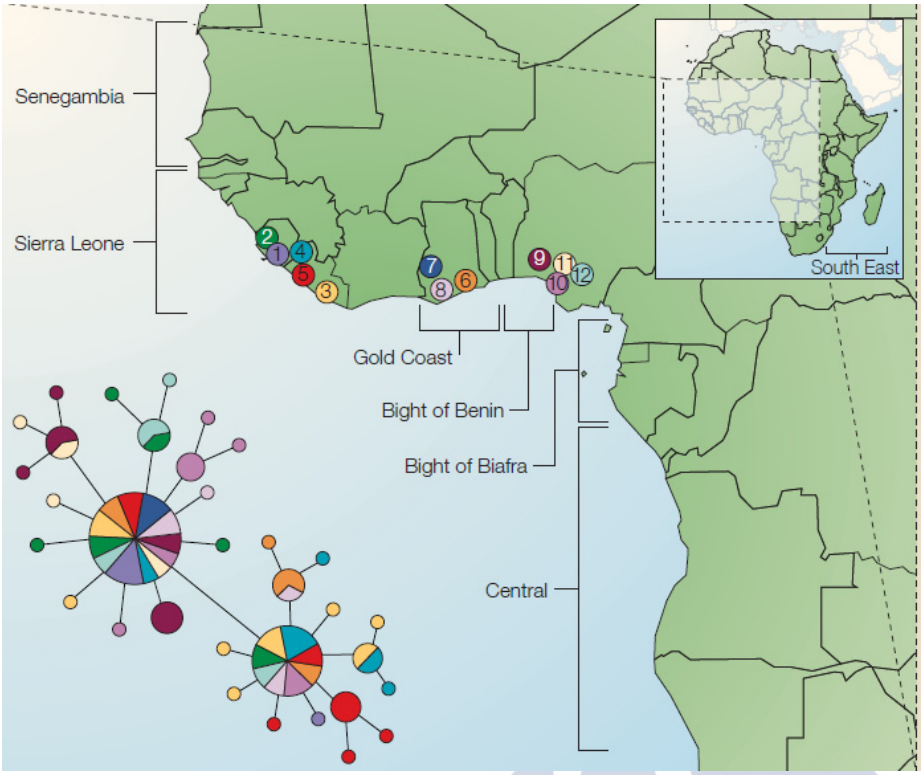


Figure 15: Inferring the genetic ancestry based on the Y chromosome. 1: Mende, 2: Temna, 3: Kru, 4: Kissi, 5: Basso, 6: Ga, 7: Akan, 8: Fante, 9: Yoruba, 10: Itsekiri, 11: Urhobo, 12: Edo, published in [33].

Both mtDNA and Y chromosomes can also be used in genealogical research. For example, a Y chromosomal DNA study revealed that Thomas Jefferson, the third president of the United States, apparently have fatheredone son of one of his slaves [34]. Foster and colleagues [34] found that a descendent of Eston, son of Jefferson's former slave Sally Hemings, shares Y haplogroup T with four descendents of Field Jefferson, Thomas Jefferson's uncle.

Y chromosomes are typically investigated for microsatellites, SNPs, and Alu insertion/deletions (see chapter 3.6). Mitochondrial DNA, in contrast, is particularly investigated for the HVS-I and/or HVS-II region, but also for SNPs distributed on the coding regions (see Chapter 3.5).

A major disadvantage of mitochondrial DNA and Y chromosomal studies is that, firstly, they provide information only about one line of descent although an individual does not have solely one ancestor per generation, but many that contributed to the genetic composition. Secondly, mtDNA and Y chromosomes are comparatively small when compared to autosomal DNA. Thus, autosomes can provide comparatively more genetic material for investigation than uniparental markers.

Thirdly, another drawback is that common and widely distributed lineages can provide information only about broad geographic origin. Only if a lineage is less common it is possible to infer more precisely about one's geographic and biological origin [33].

3.3.3 Autosomal markers

Uniparental markers can effectively geographically differentiate populations because of their smaller effective population size that result in strong genetic drift, and thus, genetically distinct groups. However, both markers consider only one out of many possible ancestors. In contrast to the mtDNA and Y chromosomes, autosomal DNA is passed down to the next generation from both parents.Before being passed on to the next generation DNA is reshuffled during meiosis. Each pair of chromosomes (one chromosome from the mother and one from the father) assorts randomly and independently before being segregated through cell division. The reduction from diploidy (2n=46 chromosomes) to haploidy (n=23 chromosomes) means that there are 2^{23} possible combinations for a set of haploid chromosomes to assort. Furthermore, each pair of chromosomes is also exposed to recombination, that means, the exchange of DNA segments (crossing over) leading to segments on the DNA that are different from the grand-parental segments (Figure 16).

Figure 16: Grandparental genome inheritance, http://www.dnainheritance.kahikatea.net/autosomal.html.

From generation to generation the segments become more and more broken down. This means that fewer DNA segments are shared with ancestors that lived a long time ago when compared, for example, to parents, grand-parents or great-grand-parents. Thus, the nature of the inheritance of the genome results in a mosaic of DNA segments of different ancestral origin. Furthermore, the

continuous migration of human populations (willingly or unwillingly) results inevitably in genetic admixture, and thus, in several ancestors from different continents. For example, the period of the Transatlantic Slave Trade when African enslaved people were forced to migrate to the Americas led to genetic admixture among three continental populations, that is, among African enslaved people, European colonizers, and the indigenous populations of the New World. In this context, the mitochondrial DNA or the Y chromosome might be able to trace back the maternal or the paternal lineage and its biological geographic origin to Africa, but both markers cannot provide further information about European, African, or Native American ancestry in the remaining genome.

Therefore, it is necessary to apply genetic markers on the autosomal chromosomes in order to estimate ancestry contributions. A prerequisite is, however, that genetic markers are able to distinguish populations genetically according to geographic origin. Useful genetic markers in this regard are Ancestry Informative Markers (AIMs) as they show different allele frequency (variation of genes) between different regions. AIMs can be continent-specific to differentiate genetically between Europe, Asia, Africa or America. Some AIM panels, however, are evenable to distinguish among populations within a continent like it was the case in Europeans [35].

Typical autosomal markers as AIMs are SNPs (Single Nucleotide Polymorphisms) or Indels (Insertion Deletion Polymorphisms). The use of microsatellites, in contrast, as predictor of population origin is limited [36]. Especially SNPs become more and more popular in the scientific community particularly in medical research [37, 38], but also in the study of human population history [39].

SNPs occur when DNA is not correctly replicated in the germ line and passed on to the next generation. Offspring should theoretically share more SNPs with their parents than with more distantly related relatives. In order to conduct a genetic ancestry test, DNA fragments have to be compared with each other (schematically shown in Figure 17). A specific DNA fragment (blue one), for examples, can be characterized by the combination of different SNPs (Haplotype 1, Haplotype 2, Haplotype 3). Theoretically, the haplotype of a child (Haplotype 2) should show more similarity to the haplotype of its mother (Haplotype 1) than to a distantly related individual (Haplotype 3).

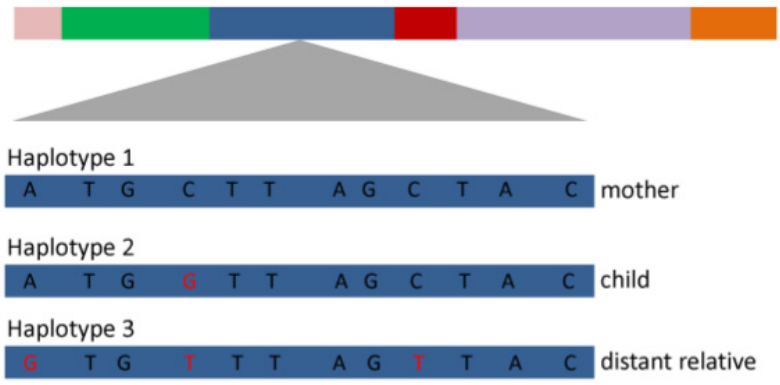


Figure 17: Different haplotypes of one DNA fragment.

Comparisons of DNA fragments can be conducted on many different regions on the genome. Due to the availability of new technology this can be done with SNP chips. Genomewide SNP chips were used, for example, to distinguish insular from mainland Caribbean populations [39]. Novembre and colleagues [35] could show that genomewide SNPs are able to distinguish between European populations.AncestryDNATM could increase genetic resolution for West Africa with SNP chips, and thus, separate West African populations into six population groups, that is, into Senegal, Mali, Ivory Coast/Ghana, Benin/Togo, Nigeria, and Cameroon/Congo (Figure 18).

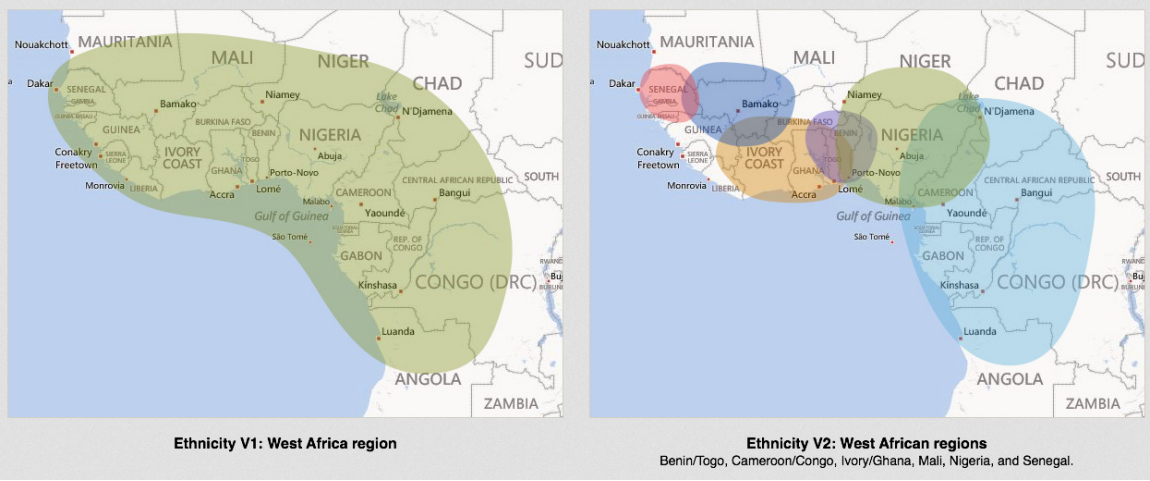


Figure 18: Ethnicity of West Africa region presented on the Smithsonian Institute's symposium: The African Diaspora. Source: AncestryDNA™, http://www.ancestry.com/.

While autosomal DNA provides information about one's different genetic ancestors, both mtDNA and Y chromosome are able to reveal maternal and paternal ancestry. Eventually, both autosomal and uniparental markers are popular complementary tools which are applied in genetic ancestry testing.

3.3.4 Limitations of DNA testing

When DNA testing companies like, for example, 23andme began to offer direct-to-consumer genetic tests, interest in testing one's DNA became increasingly popular. Su and colleagues [40] revealed in their study about the motivation of buying a genetic test that there are five principal reasons why individuals decide to test their DNA, and that is, health, curiosity and fascination, genealogy, contributing to research, and recreation.

The rise of new advanced technology in geneticshas not only extended the possibilities to test one's individual genome, but also brought forward scientific research opportunities, andfurthermore, induced novel approaches with respect to biological application. Next-Generation Sequencing (NGS) has changed considerably biological research by means of reducing reaction volume, and furthermore, increasingsimultaneously the number of sequencing reactions [41]. Especially the costs might have played a key role why Next-Generation Sequencing became popular in the scientific community and why genetic ancestry testing companies emerged on the market and started to offer direct-to-consumer genetic tests. According to The National Human Genome Research Institute,the cost per Megabase (a million bases) of DNA sequence and the cost per genome in September of 2001 came to $5.292 and $95.263.072, respectively. In contrast, the cost per Megabase of DNA sequence reached only $0.045 and the cost per genome $4.008 in January 2014. Especially the change from "first generation" sequencing platforms, that is, Sanger sequencing with capillary-based instruments (October 2007: $397 cost per Megabase of DNA sequence, $7.147.571 cost per genome) to "next generation" sequencing platforms (January 2008: $102 cost per Megabase of DNA sequence, $3.063.820 cost per genome) led to a decrease in costs [42].

The decreasing costs and the concurrent possibility of massively parallel sequencing not only led arise DNA testing companies, but also opened up new research areas like, for example, "metagenomics", which allows to investigate the microbal and viral content of the human body [43], or Genome-Wide Association Studies (GWAS) which is applied to detect genes that are

associated with diseases in order to evaluate risk like it was done, e.g., for coronary artery diseases [37]. The new technology in genomics became alsoincreasingly important in forensic sciences, that is, in human identification, as larger multiplexes might exhibit similar or even greater power to discriminate genetically individuals [44].

Some genetic testing companies offer their customers information about their risk for diseases. Incorporating new technology in research areas like personal genomics or medical research, however, poses new challenges that have to be resolved first. One of its challenges is to bioinformatically process, and subsequently, interpret the great amount of data obtained by the new instruments [45]. Furthermore, most Next Generation Sequencing technology in medical research so far was applied to investigate gene variations in populations of European descent meaning that a genetic association of a gene with a disease in the European population might not exist in a population of Africa [46]. Moreover, someone might not develop a disease despite carrying a risk gene for the respective disease. Thus, the validity of one's predisposition to a disease by now should be treated with caution as research on this topic is still in its infancy. It is therefore especially necessary that consumers of genetic ancestry tests receive comprehensive information about their results [47] although research on gene-associated diseases indeed might prove to provide reliable information in the future about one's predisposition for a disease. At least, an increased positive attitudefrom 2002 to 2010 regarding genetic tests was found in a population in the Netherlands when considering the future benefits and use of DNA test [48].

An important component of Genome-wide association studies is the previous estimation of genetic ancestry. However, autosomal markers sometimes have not the power to distinguish genetically populations from different geographic regions. That is, depending on the population under study a set of autosomal markers might be more or less useful in distinguishing genetically populations. For example, a set of 128 autosomal markers (Ancestry Informative Single-Nucleotide Polymorphisms) was able to distinguish ancestry in African and Mexican Americas [49], but was less useful in determining admixture in populations of Eurasia [50] meaning that the outcome of a genetic analysis depends both on the population under study and the genetic markers that were selected.

Consumers of genetic ancestry tests should be especially cautious when the company is providing ancestral links to famous historical characters like Cleopatra, Napoleon, or the Queen of Sheba. It might be possible that there are indeed ancestral links, but usually they are so general, and thus, on the personal level meaningless. The reason behind this is the rapid accumulation of ancestors. An

individual has theoretically after 200 generations 2^{200} (~$1.6*10^{60}$) genetic ancestors. However, in practice there never had existed a population that size 200 generations ago. After 20 generations the same individual would have theoretically 2^{20} (~1,000,000) ancestors, but in reality, again someone would not have this number of ancestors as not all genealogical ancestors contribute genetically to one's genome [51]. Ralph and Coop [51] have shown in their study that many of the ancestors within Europe are the same and shared among people meaning that even individuals from geographically distant populations within the continentshare the same set of ancestors that lived about 1,000 years ago whereas individuals living in neighboring populations even share 2-12 genetic ancestors over the last 1,500 years.This means, that going back far enough many people can exhibit ancestral links to a famous historical person.

Uniparental markers are powerful tools to trace back one's geographic paternal and maternal origin because of the lack of recombination, high mutation rate, and the small effective population size which make the markers geographically more distinctive [33]. However, mitochondrial DNA gives information only about the maternal and the Y chromosome only about the paternal ancestor.Furthermore, haplogroups of mtDNA and Y chromosomes are rarely unique to one population, and thus, often provide information only about broad geographic origin [52]. For example, Ely and colleagues [53] found in their study that <10% of African Americans could be traced back to one single geographic region in Africa. Thus, the authors concluded that only few people will be able to trace their maternal ancestor to a particular region or ethnic group in Africa. It is also possible that only a small fraction of African mtDNA haplotypes have been detected so far which might explain why 40% of African Americans do not match with African mtDNAs [53]. Another aspect that has to be taken into consideration is that ethnic groups that existed 200 or 400 years ago might differ from ethnic groups that nowadays exist. Therefore, conclusions about one's ethnic originin Africa have to be treated with caution when only the mtDNA or the Y chromosome was analyzed.

The most comprehensive approach in investigating one's ancestors is the combination of all three genetic system markers, because, for example, autosomal DNA might reveal a principal European ancestry, but the same individual might still exhibit an African Y and/or mtDNA haplotype [33]. Especially in individuals of recently admixed populations like in the Americas due to the Transatlantic Salve Trade, the combination of different genetic markers can reveal unexpected patterns of admixture as males and females from different continental populations contributed unequally to the gene pool in the Americas.

Tracing one's genetic ancestry back to a geographical region underlies continuous improvements. For example, a new method, called Geographic Population Structure (GPS) was able to place 83% individuals from different parts of the world to their country of origin, and even a quarter of about 200 Sardinians to their village [54]. Moreover, researchers also began to develop methods that allow for determining the time of admixture from a specific geographic area [55]. For example, in the context of the Transatlantic Slave Trade, Moreno-Estrada and colleagues [39] found two migration pulses from two geographic different regions in Africa into the Caribbean.

However, since the study of genetic variation among human populations, discussions about 'racial' categorizations in human genetics became the center of many interdisciplinary scientific disputes. For example, scholars from various research fields ranging from the humanities, social sciences, life science, medicine, and law engaged in an intensely interdisciplinary dialogue at the Stanford University to work out which race and ethnicity categories can be used in human genetics [56]. In order to evade 'racial' categorizationsby the scientists themselves, many life science researchers have classified their study populations according to "self reported ethnicity" like it was done in a genetic study in Colombian populations [57], but also in Hispanics and African Americans of the U.S. [58].

Concerns about DNA testing especially appear, because information about one's ancestry components might lead to a change in one's notion of identity. As all people, from the genetics point of view, are admixed, Baylis [59] has illustratedelaborately that identity is not found in the genes, but in the society that constructs racial identity. DNA does not imply racial identity, but concerns about DNA testing are still explicable as categorizations in human genetics might be misunderstood, misinterpreted and misused as it can potentially provoke people to act like racists. Therefore, dialogue across disciplines in academia is necessary to find a common agreement how to categorize human populations in the life science, also because controversies about how to label groups of individuals still remain [56]. However, dialogue within academia is probably not enough, and thus, has to be extended to the people outside academia where it should be explained not in jargon, but in plain language to avoid racism.

3.4 Investigating the TAST with genetics

3.4.1 Application of genetics

During the Transatlantic Slave Trade more than 12 million Africans were kidnapped and forced to migrate across the Atlantic to America in order to work on plantations, in households, mines or elsewhere. The Trans-Atlantic Slave Trade database [1] is a useful source to investigate the Transatlantic Slave Trade between 1501 and 1866. The database provides useful information, for example, about where a voyage began and ended, but also about the number of enslaved people thatembarked at a specific port before they were confined on vessels and shipped involuntarily to the New World. However, the disembarkation point was many times not the final destination for many enslaved people from Africa. For example, enslaved African people that were destined to work in the royal mint in Potosí, Upper Peru (nowadays Bolivia), not only arrived in El Callao, Peru, but also in ports located in Brazil, Colombia, Panamá and Argentina, and thus, had to migrate long distances before reaching the place where they finally had towork [11]. It was not uncommon that the disembarkation point was not the final destination for African enslaved people. For instance, the Caribbean island Jamaica was known for being a transitional point for many enslaved African people [8]. Sometimes African or African descendant enslaved people were relocated later during the course of the TAST. For example, the first settlers of the colony South Carolina that came principally from the Caribbean islands, especially from Barbados, brought many slaves with them from their former plantations in the Caribbean [60].The Chesapeake Bay region, as well in North America, became the final destination for many African enslaved people from the Bight of Biafra, especially for Igbo (Ibo) people. However, as birth rate in Igbo women was very high, many American-born children from Igbo women that were born as slaves were relocated from Maryland and Virginia at the Chesapeake Bay to adjacent regions [18, 61].

In this context, the research area of genetics can provide further useful information about the history of the TAST. This is especially so if historical records are poorly preserved or absent like, for example, it is the case in Afro-Bolivians that migrated in the course of the TAST from Potosí in the South of Bolivia further north to the Yungas [62]. Furthermore, cultural characteristics from different African ethnic groups intermingled in the Americas, and thus, making it difficult to tell African origin. For example, African refugees that escaped from slavery like Noir Marrons have a widespread African origin ranging from Senegambia to Angola resulting in different African ethnic groups that met in the Americas. However, despite interethnic mixture and also intense cultural

exchange with Europeans and Amerindians, Noir Marrons apparently highly conserved their genetic African ancestry by showing most genetic similarities to the Bight of Benin [63].

It has to be emphasized, however, that genetics does not make other research fields like historyredundant, but it can complement the study of the Transatlantic Slave Trade in a very important way. Genetics can help disentangle the complex history of the TAST, provoke and give rise to new questions, but also provide information about the biological impact on African American populations. Studies about genetics in the context of the Transatlantic Slave Trade have proven to be a reliable tool in confirming and backing up historical records. For example, Moreno-Estrada and colleagues [39] found two major migration pulses from Africa into the Caribbean based on genetics. The first migration pulse was detected at around 1550 and the second at the peak of the Transatlantic Slave Trade at the end of the 18th century. Coinciding with historical records, the authors could also show that African enslaved people from the first migration pulse principally came from the Senegambian region, whereas during the second pulse most enslaved people originated from West Central Africa. In a study by Stefflova and colleagues [19] it was shown that African populations contributed not equally to different populations in the America as the genetic contribution of West/West-Central Africa was detected to be highest in the Caribbean, less in the United States and Colombia, and lowest in Brazil where Southwest/West-Central Africa was more prevalent.

As the following sections will show, investigations about the Transatlantic Slave Trade from a genetics point of view is principally focused, firstly, on the genetic origin of different African American populations (chapters 3.4.2 – 3.4.5), and secondly, on the biological impact in medical research (chapter 3.4.6).

3.4.2 The TAST in North America

The TAST in North America was mainly concentrated in the region around the Chesapeake Bay (Virginia, Maryland, and parts of North Carolina) and in the neighbouring region in South Carolina and Georgia.At the beginning, in the Chesapeake Bay many African enslaved people were involved in tobacco [2] and in South Carolina in rice production [9, 10] before being replaced by cotton as the main crop [2].

As in South Carolina rice was the main crop at the beginning, merchants apparently preferentially chose slaves that originated from the Grain Coast (modern Senegal, Gambia, Guinea, Sierra Leone, and west Liberia)due to their rice cultivation skills [9] (Figure 19) although available historical records could not provide convincing information that the proportion of enslaved African people from rice-cultivating countries indeed contributed most to the African enslaved community in South Carolina [10].

Figure 19: Rice planting in South Carolina during the early 1890s, Image Reference Ralph02, as shown on www.slaveryimages.org, compiled by Jerome Handler and Michael Tuite, and sponsored by the Virginia Foundation for the Humanities and the University of Virginia Library.

Consistent with historical records in a study about the genetic ancestry in an African American population of South Carolina it was shown that Y chromosomes of males from South Carolina are genetically most similar to males from the Grain Coast whereas African American males from Washington D.C. were genetically closer to males from the Bight of Biafra [64]. Benn Torres and colleagues [64] observed also a lower European genetic ancestry on the Y chromosome in the South Carolina population (32.3%) than in the African American population from Washington D.C (36.6%) which is probably due to a more pronounced social and/or physical isolation in South Carolina.

With respect to the maternal lineage, Salas and colleagues [16] observed a principal West African ancestry (>55%) on the maternal line in African Americans from the United States followed by West-Central and southwestern Africa (<41%). In contrast, genetic contribution from North Africa, East Africa, southern Africa, but also from southeastern Africa to the mtDNA gene pool in African Americans of the United States was observed to be marginally [16, 17]. A principal West and West-Central African genetic contribution on the mtDNA in the United States was found, for instance,

inPhiladelphia whereas mitrochondrial African ancestry apparently mostly came from the Bight of Biafra and Benin [19].

Uniparental markers indicate a rather directional mating with males coming from Europe and females originating from Africa which is also supported by elevated levels of African ancestry on the X chromosome in an African American population [65]. In self-declared African Americans mitochondrial DNAswere principally of African ancestry (92.7%), but African ancestry in the Y chromosome reached a lower level (69.5%) and was followed by European ancestry (29.7%) indicating gender-biased mating with some males from Europe and females principally from Africa [58].

European autosomal ancestry in general highly varies among different populations of African Americans in the U.S. [65]. European ancestry in African Americans in Charleston, South Carolina seems to reach low levels (11.6%), but was in general observed to be higher, but not consistent among African Americans in the northern states of the U.S. [66]. An especially low level of European autosomal ancestry of 3.5% was found in the relative isolated African American population, the Gullah Sea Islanders of Coastal South Carolina [67].

Depending on the population under study in the United States, autosomal ancestry estimates vary between populations. Hispanics of the San Luis Valley in Colorado, for example, show a high European autosomal admixture (62.7%) followed by Native American admixture (34.1%). The low African ancestry (3.2%) is partly either the result of gene flow from African Americans within the U.S., or was probably already present in the people that migrated from Mexico to the North, as Mexico was heavily affected by the Transatlantic Slave Trade at the beginning of the colonization of the Americas [68]. Likewise, a low African autosomal ancestry (11.6%) could be observed in self-declared U.S. Hispanics from different parts of the United States that exhibit principally a high European ancestry (61.2%) followed by Native American ancestry (14.9%). In comparison, a higher sub-Saharan African autosomal ancestry (86.2%) was found in self-declared African Americans from the U.S. [58].

As will be seen in the following chapters, African, European, and Native American ancestries differ among various regions within the Americas. African ancestry is especially prevalent in the Caribbean, whereas Native American ancestry is particularly common in Mexico and Guatemala. European ancestry, in contrast, is most pronounced in South America [69].

3.4.3 The TAST on the Caribbean islands

During the Transatlantic Slave Trade from 1514 to 1866 more than 4.2 million African enslaved people reached the Caribbean islands which contrast with about 0.3 million enslaved people from Africa that disembarked on Mainland North America [1].

Similar to African Americans in the United States, descendants of formerly enslaved African people on the Caribbean islands exhibit a similar elevated African genetic ancestry. African descendants of Jamaica and Barbados were observed to reach a West African autosomal ancestry of 89.6% and 84.4%, respectively. European ancestry was slightly higher in the population of Jamaica (12.4%) than in Barbados (10.2%) [70] although European ancestry on Jamaica could also be lower (6.8%) [66]. Native American ancestry came to only 0.2% in Barbados and 3.2% in Jamaica [70].

The high African Ancestry on the British colonies of Jamaica and Barbados can be explained by the history of both islands. Barbados experienced a pronounced transformation regarding population composition after introducing sugar on the island around 1645 [2]. In fact, about 375,000 enslaved African people disembarked on the island during the TAST [1]. Jamaica (Figure 20), in contrast, became increasingly important during the early 18th century when the colonizers were looking for greater sugar plantations, and thus, replacing Barbados as the major sugar production colony within the British empire [2]. In total, about 930,000 enslaved African people reached Jamaica during the TAST [1]. In addition to it, Jamaica became a major market for chattel slavery providing enslaved Africans from different regions in Africa not only for the Jamaican sugar plantations, but also for Spanish America [8].

Figure 20: Sugar cane harvest, Jamaica, 1820s, Image Reference: NW0055, as shown on www.slaveryimages.org, compiled by Jerome Handler and Michael Tuite, and sponsored by the Virginia Foundation for the Humanities and the University of Virginia Library.

In contrast to the autosomal low European ancestry, European admixture in Jamaican men on the Y chromosome reached up to 41.1% whereas Jamaican men that belong to an African-specific Y chromosome haplogroup were observed to be genetically most similar to men from the Bight of Biafra [64].

In comparison, European ancestry on the mtDNA was minimal. In a study about the mitochondrial DNA in a Jamaican population it could be shown that 97.5% belonged to a sub-Saharan haplogroup. It was also observed that African maternal mtDNA contribution on individuals on thisCaribbean island could be mostly attributed to the Gold Coast, and surprisingly, only marginally to the Bight of Biafra and West-Central Africa [71]. Deason and colleagues [71] hypothesized that the highest genetic influx most probably came from the Bight of Biafra and West-Central Africa, as both regions heavily contributed to slave deportations at the end of the Transatlantic Slave Trade on Jamaica. As fecundity in the Jamaican population was low and mortality rate was high, the Gold Coast was expected to be underrepresented as this region was heavily involved in the TAST before the Bight of Biafra and West-central Africa became the major source of enslaved people [71].Therefore, the study about the genetics provokes further questions about the abilities of acclimatization and acculturation of different African groups in the Americas. Furthermore, it is still unknown to which extent the conquest contributed to the reduction of the indigenous population, mirrored in the extremely low levels of Native American ancestry in the Caribbean [70].

After Jamaica with almost one million of slave disembarkations, Cuba (>700,000) and Saint-Domingue (<700,000) represented the second and third major destination for enslaved people from Africa between 1514 and 1866 in the Caribbean. Less slave arrivals were documented, for example, in Puerto Rico (<17,000). The Spanish West Indies, including Puerto Rico, became the destination for many enslaved African people after the middle of the 18th century [1]. An increasing number of slave arrivals on Puerto Rico can be attributed to the British that captured Cuba during the 1760s which led the Spanish Crown in 1789 finally to open up a free trade of slaves for all nations to Spanish America. Puerto Rico was the second major Spanish colony after Cuba, but in comparison to other Caribbean islands, Puerto Rico was a relatively small colony where free wage laborers became increasingly involved in sugar production. However, despite a relatively small number of slaves, the concentration of slaves that worked in sugar was particularly high on Puerto Rico [2]. According to the Transatlantic Slave Trade Voyages Database, most enslaved people that arrived in Puerto Rico between 1514 and 1866 embarked at the Bight of Biafra and Gulf of Guinea islands (39.6%) followed by West Central Africa and St. Helena (32.4%) [1].

In comparison to Jamaica and Barbados, the traces that left the Transatlantic Slave Trade in populations of Puerto Rico are less visible from the genetic point of view. That is, African autosomal ancestry on average reached only 21.2% [72] which is about 4 times lower when compared to Jamaica and Barbados [70]. Native American ancestry was higher on Puerto Rico and averaged 15.2%. The principal ancestry component on the island was European which came to 63.7% [72]. Individuals from Puerto Rico and other Caribbean islands (Cuba, Haiti, Dominican), but also Colombia, Honduras, and Mexico apparently share a Latino-specific component in their European ancestry which differ from northern and southern Europeans, but shared partly with Spanish and Portuguese individuals [39].

Autosomal ancestries were observed to vary among individuals and regions on Puerto Rico, especially with respect to the African and European ancestry. For example, African ancestry exceeds the average value in the East of Puerto Rico where individuals live close to historical sugar mills (31.8%). The highest African ancestry (47.8%) correspondingly was in one municipality in the eastern part of the island (Loíza). European ancestry in the East of Puerto Rico (54.7%), in contrast, did not reach the average value of the whole island [72].

In a mitochondrial DNA study in Puerto Ricans the African genetic component was genetically closest to West Africans with probably genetic contributions from African populations that lie between the Cape Verde and São Tomé. However, mitochondrial DNA lineages were not principally of African origin (27.2%), but of Native American ancestry (61.3%). West Eurasian ancestry in Puerto Rican mtDNAs was even lower (11.5%). Similar to autosomal continental ancestry, African mtDNA ancestry was highest in the east where most of the sugar mills were established whereas Native American mtDNA ancestry was highest in the western part of Puerto Rico [73].

The TAST also left its biological traces in a population on New Providence, Bahamas. Simms and colleagues [74] found a principal genetic contribution from Equatorial Guinea and Angola which is supported by the historical fact that most enslaved African people that arrived in the Bahamas came from West Central Africa and St. Helena [1]. Studies like these and those mentioned above contribute further important information about the history of the TAST and the geographic origin of descents of formerly African enslaved people.

3.4.4 The TAST in South America

Most African enslaved people that were forced to South America were sent to Brazil [1]. Ancestry components in Brazil differ to those found in North America and the Caribbean.

African autosomal ancestry in Brazil seems to reach only a low level, that is, 14.3% and Native American ancestry apparently cameeven to only 8.5%, but still higher when compared to Jamaica and Barbados. European ancestry in Brazil accounts for 77.1%, whereas European ancestry obviously is higher in the South than in the North [75] which was found in several studies [76–78]. A predominant European ancestry was found in the states of São Paulo (European ancestry 79%, African ancestry 14%, Amerindian ancestry 7%) [79] and Rio de Janeiro (>55% European ancestry followed by Amerindian and African ancestry) [80]. Pena and colleagues [76] suggested that the reason for a predominant European ancestry in Brazil might be attributed to the government which stimulated at the end of the TAST during the 19th and 20th century the immigration of 6 million people from Europe to Brazil.However, before the Europeans dominated immigration to Brazil, the country, in specific the Northeastern part of the country, Pernambuco and Bahia, experienced a massive slave importation from the African continent. This region in Brazil developed even before Jamaica and Barbadosto one of the most important sugar production zones during the TAST (Figure 21) which resulted in a massive enslavement of African people [2].

Figure 21: Sugar mill, Brazil, 1816, Image Reference: KOSTER2, as shown on www.slaveryimages.org, compiled by Jerome Handler and Michael Tuite, and sponsored by the Virginia Foundation for the Humanities and the University of Virginia Library.

As enslavement of the local American Indian population in Brazil did not work out for the Portuguese, African enslaved people became increasingly the primary source of enforced labor, especially after the smallpox epidemic at the peak of Indian slavery in 1560. Therefore, slave labor in the Brazilian colony predominantly consisted of African enslaved people beginning in 1630 [2]. Eventually, more than 3.3 million enslaved African people between 1514 and 1866 had to leave their home in Africa to work in Brazil, mainly on sugar plantations, for the profit-seeking colonizers [1].

Silva and colleagues [81] found in their study, which is based on mitochondrial DNA, an ethnic sex difference in the forced migration to the Americas during the TAST. More specifically, they suggested,referring to their genetic data,that there was a male prevalent forced migration from different regions in West Africa to Brazil [81]. The majority of enslaved African people (66.2%) that disembarked in Brazil were male, and correspondingly, the percentage of males that embarked in Africa and arrived in Brazil ranged from 57.7% (Bight of Biafra and Gulf of Guinea islands) to 86% (Gold Coast) [1]. West-Central Africa became the most important source of enslaved people for Brazil during the TAST. More than 2 million enslaved people embarked in West-Central Africa to be destined to Brazil. The majority of individuals again, that is 65%, were male [1].

In a genetic study in two populations of Rio de Janeiro and Porto Alegre, Brazil, a principal West-Central and southeastern origin based on mitochondrial DNA was found. According to the mtDNA, 69% and 82% of the maternal lineages found in Rio de Janeiro and Porto Alegre, respectively, could be attributed to West-Central and southeastern Africa whereas Y chromosomes principally were of European origin [82]. African-specific haplogroups of the Y chromosome that could be found in the Brazilian populations were mainly of sub-Saharan origin [83]. Genetics data indicate a pronounced directional mating in Brazil with males being principally from Europe (Portugal) and females mainly of Amerindian or African origin [77, 78, 84, 85].

Both uniparental markers (mtDNA and Y chromosome) provided evidence that European ancestry in Brazil principally can be attributed to European males that migrated to this country. Genetic studies have shown that despite more male than female Africans arrived in Brazil, African women principally contributed to the genetic African ancestry in the Americas [83]. It has to be mentioned here, however, that Brazilian populations are highly heterogeneous. For example, one population in Rio Grande do Sol, Brazil, that was classified as “white” was genetically almost completely of

European ancestry, whereas another population living in the same state exhibited elevated Native American (36%) and African (16%) ancestries in their maternal lineages [86].

No other region in South America like Brazil was affected by the Transatlantic Slave Trade to that great extent. Beside Brazil, Colombia was another important region where the Europeans settled during colonization and left its traces.

Colombia shows a high genetic autosomal interpopulation variability [69]. Individuals in most regions in Colombia seem to have a principal European and Native American ancestry [87], nonetheless, autosomal ancestry can vary considerable across Colombia and among geographically relatively close regions [88]. An elevated African ancestry was observed in individuals in the North from Antioquia (88.5-88.8%) [89], Bolivar (43.8%) and Magdalena (28.2%) [88], in the Northwest in individuals from Chocó (53.8% [87], 68.1% [88]), and in the West in individuals from Mulaló (46.6%) [87], Cauca (23.5%) and Valle de Cauca (21.5%) [88].An increased African ancestry is most probably the result of slave importations from Africa during the course of the TAST.

In Colombia, the port of Cartagena de Indias in the North received an increasing number of African enslaved people towards the end of the 16th century resulting in >100.000 arrivals [1]. Although Cartagena de Indias was for many enslaved African people a transitional point, Colombia itself was early in demand of experienced gold miners (Figure 22). However, people of African ancestry did not exist only in the gold mines in Colombia, but also in the surrounding mountainous regions, becausemany African people escaped from slave work in the rich gold-mining areas to form powerful runaway communities close to Cartagena de Indias which began to emerge as early as the 1530s [4].

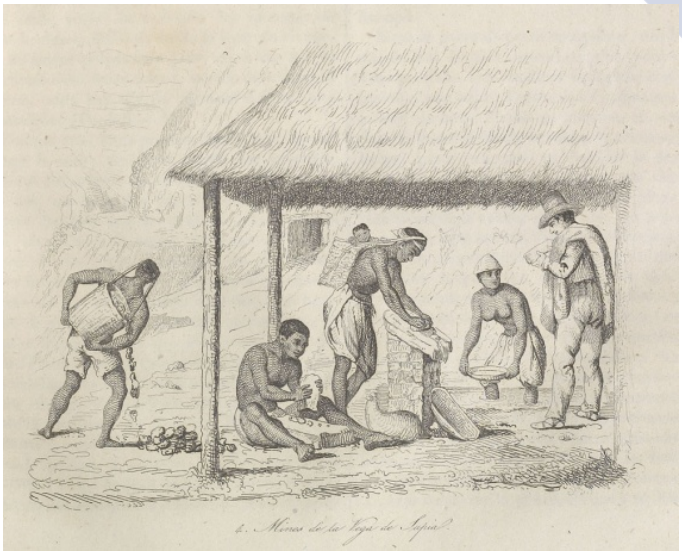


Figure 22: Gold production Colombia, 1826, Image Reference: NW0312b, as shown on www.slaveryimages.org, compiled by Jerome Handler and Michael Tuite, and sponsored by the Virginia Foundation for the Humanities and the University of Virginia Library.

Most enslaved African people that disembarked at Cartagena de Indias came from West-Central Africa and St. Helena (45.2%) and Senegambia and offshore Atlantic (38.9%) [1].

In the mtDNA a principal West/West-Central, southwest, but also southeast African component was found in self-declared mulattos and Afro-Colombians [57]. Likewise, in a mitochondrial DNA study with individuals from the department of Chocó in Colombia the authors could reveal a major sub-Saharan African genetic component (~84%) and only a minor Native Central/South American one(~16%) [90]. However, mitochondrial DNA in Colombia seems to be principally of Native American origin and African-specific mtDNA haplogroups in Colombia seem to be mostly confined to the Caribbean coast and the Pacific region like, for example, San Basilio del Palenque in the North orQuibdó [91, 92] and Chocó in the West [88]. Y chromosome ancestry, in contrast, was observed to be mostly non-Native American.That is, European ancestry in Y chromosomes apparently is distributed across all populations in Colombia, whereas African ancestry in the Y chromosome is like the mitochondrial DNA mainly detected in West and Northwest Colombia [88].The example of Colombia illustrates that, depending on the population under study, ancestry estimates can highly vary in populations of South America.

African ancestry was also highly preserved in other South American populations like, e.g., in Noir Marron communities in French Guiana which are the descendents of African refugees from Dutch plantations [63, 93].

In other parts of South America genetic contribution of African people was detected to be in general low, like it was the case in Argentina [94] although Río de la Plata received more than 63.000 slave arrivals from Africa between 1514 and 1866 [1]. In Argentina European ancestry seems to be predominant in the population, but is followed by Native American ancestry [94, 95]. European ancestry was found to be even more prevalent on the Y chromosome of Argentinans whereas mitochondrial DNA was observed to be principally of Native American ancestry [94]. Corach and colleagues [94] argued that a lack of African ancestry, despite of many slave arrivals from Africa, might be the result of a social barrier that prevailed in Argentina which prevented admixture of European or mixed European-Native American ancestry, respectively, with African ancestry.

3.4.5 The TAST in Central America

Likewise Argentina African ancestry in individuals of Mexico seems to be only marginal (typically less than 5%) with European and Native American ancestry as the principal ancestry components [96]. Similar was found in another study [97]. The low African ancestry in the Mexican population can be explained by the history of the TAST. After the Spaniards began transporting enslaved people from the African continent to the Americas, Mexico became one of the major destinations at the beginning of the TAST. Before 1650 the Spanish colonizers imported between 250,000 and 300,000 African enslaved people to their Spanish colonies, that is, mainly to Peru and Mexico. However, in contrast to Peru, African slavery was less important in Mexico due to a predominant Indian and mestizo population. As free Indian wage laborers grew in importance in the labor force, enslavement of African people decreased early again during the TAST in Mexico, that is, after 1650 [2]. Hence, a predominant Native American ancestry was found on the maternal lineage (about 90%) in a population of Mexico City, but a principal European ancestry on the paternal lineage (about 60%) followed by Native American ancestry (about 40%) [97].

However, not all populations of Central America show a principal Native American ancestry. For instance, in Garífunas of Honduras, also known as “Black Caribs”, Native American ancestry on the maternal lineage reaches lower levels (~16%) whereas African ancestry predominates (~84%) [90].

The varying estimates of African, European, and Native American ancestry in American populations is a consequence of, at least in part, recent admixture during the Transatlantic Slave Trade between indigenous populations in America, the European colonizers, and enslaved people from the African continent. Depending on the history of the respective population under study ancestry estimates can vary. Information about ancestry estimates is especially useful in disease association studies, like case-control studies, as population stratification, that is, a systematic difference of allele frequencies between populations due to differences in ancestry, can mislead associations of genes with diseases [98]. The next chapter is about how ancestry estimates can be linked to diseases in order to assess risk factors for different admixed populations.

3.4.6 Genetic ancestry in medical research

The Transatlantic Slave Trade led to the encounter of three continental populations: Africa, Europe, and Native America. As seen above, depending on the history of a region in the Americas, estimates of continental ancestry can vary between populations. In individuals of self-declared African Americans from the United States, for example, African ancestry was detected to be the principle ancestry estimate whereas African ancestry was shown to be low in populations of Mexico. In contrast, individuals of Mexico, but also Hispanics of the United States were seen to reach high levels of Native American ancestry. A low level of a Native American genetic component, in comparison, was found in Caribbean populations like Jamaica and Barbados. European ancestry was observed to be low in African Americans in Charleston and in Gullah Sea Islanders, South Carolina, but was the principal genetic component in many Brazilian populations.

Physical characters like, e.g., skin colors, are rather unreliable methods to infer genetic ancestry. Especially in Brazil, for instance, due to high genetic variances on the individual level, skin color was proven to be a poor predictor of genetic African ancestry [77, 99, 100].

Information of genetic ancestry is,due to population stratification,especially necessary to know in medical research like in GWAS (Genome-wideAssociation Studies) or case-control studies, respectively [101]. In a GWAS a high number of SNPs are involved to find an association with a disease. For example, risk disease genes where already foundfor schizophrenia [102], coronary artery disease [37], type 2 diabetes [103], Crohn's disease [104], Parkinson [105], and breast cancer [38]. A lack of information of genetic ancestry can confound associations of diseases with genes, because even within a small geographic area a high population differentiation can exist like itwasrecently foundamong indigenous populations in Mexico [96].

Information about genetic ancestry was shown to be crucial in predicting the risk for a disease. For example, African admixture was shown to increase risk for hypertension as African ancestry in African American and Hispanic American women was positively associated with hypertension [106] although environmental factors certainly also play an important role in the cause of hypertension [107]. An elevated African and Native American ancestry was also found to be associated with obesity, that is, with a high body mass index (BMI) in African American women [108] and in Native American people [109], respectively. Native American ancestry is possibly also connected with an increased risk of developing type 2 diabetes [97]. Ancestry estimates mightbe associated with a disease in one population but not in another one. Individuals with higher levels of

European ancestry in Mexican Americans, for examples, exhibited more severe asthma which contrast with Puerto Ricans, as in the latter asthma severity was not related to any ancestry [110].

An increased risk of developing breast cancer was observed to be associated with elevated levels of European ancestry like it was found among US Latinas and Mexican women [111]. However, a high incidence of aggressive breast cancer apparently also occurs among African American women of the Chesapeake Bay. As mentioned above, the Chesapeake Bay is historically connected to the Bight of Biafra.Jackson [18] used this historical information of the TAST to explain breast cancer etiology in African American women of the Chesapeake Bay. Apparently there indeed existgenetical ancestral links between women of the Chesapeake Bay and women of the Bight of Biafra that contribute to the onset of breast cancer. Knowledge about the sources of polymorphisms that are involved in genetic-related disease susceptibilities can help to understand and treat a disease [18].

Nonetheless, essential for an informative and thoroughly conducted medical genetic study is the integration of a broad ensemble of populations. Most medical genomic studies wereso far rather focused on populations of European descent [46]. Therefore, studies like it was conducted by the African Genome Variation Project are necessary for a better understanding of medical genetics in sub-Saharan Africa [112]. The Transatlantic Slave Trade was responsible for the displacement of millions of people from different African ethnic groups to various regions in the Americas. Understanding the genetic contribution of ethnic groups to a specific region might, as suggested in Jackson [18], advance medical research and treatment in a very important way, that is, in this context, both for African American and African populations.

3.5 Mitochondrial DNA

3.5.1 Localization

The mitochondrial DNA is a circular double-stranded genome located in cell organelles, called the mitochondria (Figure 23). These cell organelles can be found in the cytoplasma of cells. Two cell membranes encompass the mitocondrion to compart the interior section of the mitochondrion from the exterior located cell cytoplasma [113]. According to the endosymbiontic theory, the double membrane of mitochondria can be ascribed to formerly free-living bacterial cells that were taken up by larger cells to live in endosymbiosis. In the course of millions of years of evolution, the ingested

cells and the host cell became more and more dependent to each other. Organelles like mitochondria lost their autonomy and became more specialized resulting in an inability to live without the host cell, and vice versa. The endosymbiontic theory is a hypothesis, but research on the origin of mitochondria and other cell organelles further advanced in the past starting with Sagan [114] who formulated cytological, biochemical, and paleontological evidence for the bacterial origin of organelles, including mitochondria. In a recent study it could be found that mitochondria probably evolved from a planktonic marine alphaproteobacteria [115]. Besides the double membrane, further evidence for a bacterial origin is the presence of DNA (mtDNA), which differs from the nuclear genome in size and gene content (see chapter 3.5.3). The mitochondrion is able to produce some of its own proteins for its functioning although most of its proteins are encoded in the cell nucleus [116]. Replication and mitochondrial division is directed within the mitochondria [117].

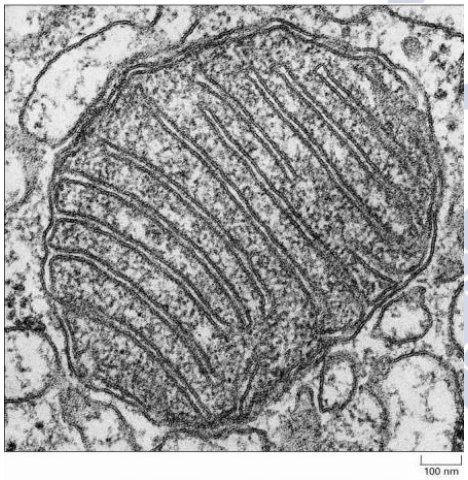


Figure 23: Mitochondrion, [118].

The principal role of mitochondria is the cellular energy metabolism. The energy metabolism comprises the fatty acid ß oxidation, the urea cycle, and the respiratory chain in order to produce adenosine triphosphate, ATP [113].

The number of mitochondria within a cell varies widely. Liver, nerve or muscle cells might contain hundreds or thousands of mitochondria in their cells, each harbouring 2-10 copies of mtDNAmolecules. In contrast, red blood cells (erythrocytes) do not contain any mitochondria at all. Oocytes contain around 100,000 mitochondria, each harbouring one mtDNA molecule, whereas sperms contain only 50-75 mitochondria [22, 119].

Depending on the type of cell 1000 to 100,000 copies of mtDNA can be present in one human cell. If mtDNA copies are identical homoplasmy is present. A mixture of mtDNA copies is calledheteroplasmy [113].

3.5.2 Organizationof the mtDNA genome

The human mitochondrial genome is a circular and double-stranded DNA molecule found in various numbers of copies in mitochondria. In contrast to the diploid nuclear genome with about 6,000,000,000 base pairs, the mtDNA reaches a size of only 16569 base pairs (Figure 24).

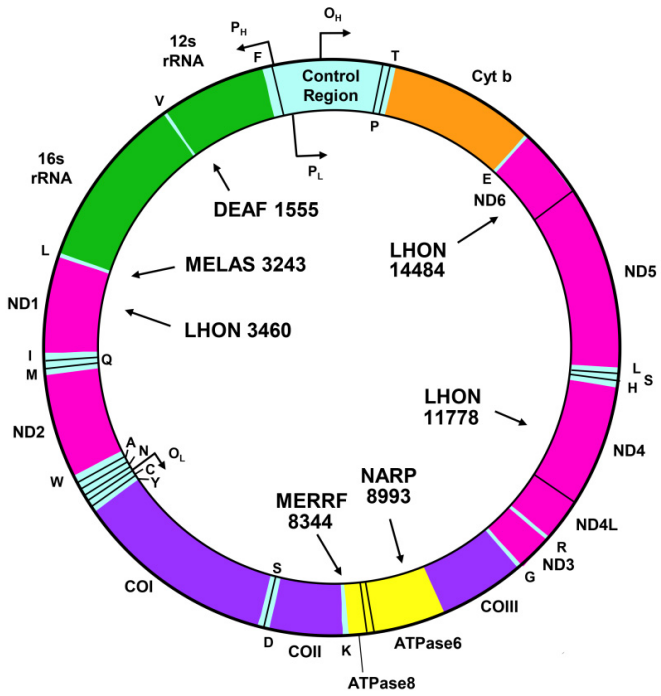


Figure 24: The mitochondrial Genome, http://www.mitomap.org/MITOMAP [32].

The mitochondrial DNA contains 37 genes that encode for proteins of the respiratory chain, ribosomal RNA (rRNA), and transfer RNA (tRNA). The 13 proteins of the respiratory chain encompass ND1-ND6 (pink), Cytb (orange), COI-COIII (violet), ATPase6 and ATPase8 (yellow). The 2 ribosomal RNA genes are 12S and 16S RNA (green), and the 22 transfer RNA genes are distributed across the whole genome. The transfer RNA genes are indicated by the amino acid they carry: F (Phenylalanine, from position 577 to 647), V (Valine, 1,602-1,670), L (Leucine, 3,230-3,304 and 12,266-12,336), I (Isoleucine, 4,263-4,331), Q (Glutamine, 4,329-4,400), M (Methionine,

4,402-4,469), W (Tryptophan, 5,512-5,579), A (Alanine, 5,587-5,655), N (Asparagine, 5,657-5,729), C (Cysteine, 5,761-5,826), Y (Tyrosine, 5,826-5,891), S (Serine, 7,446-7,514), D (Aspartic acid, 7,518-7,585), K (Lysine, 8,295-8,364), G (Glycine, 9,991-10,058), R (Arginine, 10,405-10,469), H (Histidine, 12,138-12,206), S (Serine, 12,207-12,265), E (Glutamic acid, 14,674-14,742), T (Threonine, 15,888-15,953), P (Proline, 15,956-16,023). Most genes are located in the outer strand, the 'heavey' strand (H-strand)which is rich in Guanine bases (G) whereas the inner strand with fewer genes represents the 'light' strand (L-strand) and is rich in Cytosine bases. Proteins, rRNAs and tRNAs encoded by the H-strand are placed outside the circle in Figure 24. Proteins and tRNAs encoded by the L-strand are placed inside the circle. The origin of replication for the H-Strand (OH) and transcription for both the H (PH)- and L-strand (PL) can be found in the control region, except the origin of replication for the L-strand (OL).

Mitochondrial inherited diseases are LHON (Leber hereditary optic neuropathy), NARP(Neurogenic muscle weakness, ataxia, and retinitis pigmentosa), MERRF (Myoclonus epilepsy and ragged red muscle fibers), DEAF (Maternally inherited DEAFness or aminoglycoside-induced DEAFness), MELAS (Mitochondrial encephalomyopathy, lactic acidosis, stroke-like episodes).

The mitochondrial control region or D-loop (displacement loop) extends from position 16024 to 576. This region of the mtDNA does not encode for rRNA, tRNA or proteins of the respiratory chain, but seems to be responsible for mtDNA genome replication and transcription [113]. It is a highly variable region, and thus, is also referred to as hypervariable sequence (HVS) which can be divided into HVS-I, HVS-II, and HVS-III. Start and end positions of the three sub-regions are not consistent in different research fields. In forensics HVS-I ranges from 16024 to 16365, HVS-II from 73-340, and HVS-III from 438-576. In contrast, population geneticists label HVS-I from 16024 to 16400, HVS-II from 44 to 340, and HVS-III from 438 to 576 [113]. In this thesis, HVS-I refers to the DNA fragment ranging from 16024 to 16569 and HVS-II encompass the region between position 1 and 576.

The first whole mitochondrial DNA sequence was published in 1981 and named Cambridge Reference Sequence (CRS). The mtDNA sequence was obtained from a placental mtDNA sample of one single European individual [120]. The original mtDNA sample was resequenced and reanalyzed 18 years later by Andrews and colleagues [121]. The researchers found discrepancies in the sequence of the original mtDNA molecule assuming probably due to sequencing artifacts, or

contamination with bovine and HeLa samples. Andrews and colleagues [121] found 10 substitutionerrors and only one, instead of two, cytosine residue at position 3106 and 3107. The authors suggested to keep the original, but corrected sequence as reference and name it Revised Cambridge Reference Sequence (rCRS). The numbering of the rCRS is based on the L-strand of the mitochondrial DNA.

3.5.3 Differences between the nuclear and mitochondrial genome

As mentioned above, the size of the mtDNA reaches only about 16,5 kb, wheras the nuclear genome is made up of about 6,000,000,000 base pairs. Number of chromosomes for a diploid cell is 46, but a cell can contain several thousand copies of mtDNA molecules. The number of genes that are encoded by the nuclear genome lies between ~20,000 and ~30,000. The mitochondrial DNA, in contrast, encodes for only 37 genes. Gene density is higher in mtDNA than in the nuclear genome (1 per 450 bp vs. ~1 per 40,000 bp). Introns are absent in the mtDNA, but frequently found in most genes of the nuclear genome. Correspondingly, percentage of coding DNA is 93% in the former and 3% in the latter genome. Furthermore, the genetic codes for both genomes do not completely match. Chromosomes are stabilized by nucleosome-associated histone and non-histone proteins whereas the mitochondrial genome is not joint with histones, but with several proteins like, e.g., TFAM, that form nucleotides. The mtDNA is exclusively inherited maternally while the autosomes and the X chromosome follow Mendelian and the Y chromosome paternal inheritance. During transcription all genes on both L- and H-strand of the mtDNA are transcriped as a large polycistron whereas in the nuclear genome most genes are transcriped individually [122]. Replication for the mtDNA probably follows a strand-displacement model [123] although a strand-coupled mechanism like in the nuclear genome might be possible. MtDNA uses DNA polymerase γ and the nuclear genome DNA polymerase α and δ. Recombination is most probably absent in the mitochondrial genome, but occurs in the nuclear genome during the prophase of the meiosis when homologous chromosomes pair with each other [122].

3.5.4 Numts

Numts, or nuclear inserts of mtDNA, are mitochondrial DNA fragments that are transferred and incorporated in the nuclear genome. It is still unknown how numts become part of the genome in

the nucleus. However, numts could be of interest as outgroup for the human mtDNA phylogeny, and thus, replace chimpanzee and bonobos mtDNAs that are commonly used as outgroups, because numts evolve considerably more slowly in the nucleus than they would evolve in the mitochondrion [113].

3.5.5 High mutation rate

One feature of the mitochondrial DNA is its high mutation rate in comparison to the nuclear DNA. This is a key characteristic that makes the mtDNA a popular tool to investigate the past of human history. The mutation rate of the mitochondrial DNA is 10-fold higher than the average mutation rate in the nuclear DNA [22]. However, mutation rate within the mitochondrial genome is not consistently uniform. The majority of mutational hotspots are located in the HVS-I and HVS-II region, and thus, exhibit the highest mutation rate when compared to other parts of the mtDNA [26]. Soares and colleagues [26] proposed a substitution rate for the whole mitochondrial DNA of 1.665×10^{-8}(+/- $1.479 * 10^{-9}$) base substitution per nucleotide per year. Positions with high mutation rates are shown in Figure 25.

The mitochondrion is responsible for the oxidative phosphorylation, and thus, contains many mutagenic oxygen free radicals which might be one of the reasons why the mutation rate in the mtDNA is comparatively high. Furthermore, the process of mtDNA replication requires the genome to be longer single-stranded when compared to the nuclear chromosomes which makes the mtDNA more vulnerable. The control regionis sometimes named D-loop (displacement loop) as the D-loop is a structure of the L- and H-strand that becomesdisplaced by a third strand that replicates at one of the strands.Mutational events especially occur in the control region. In addition to it, the mtDNA genome is not packaged with histones which might also result in a greater vulnerability to mutations [22].

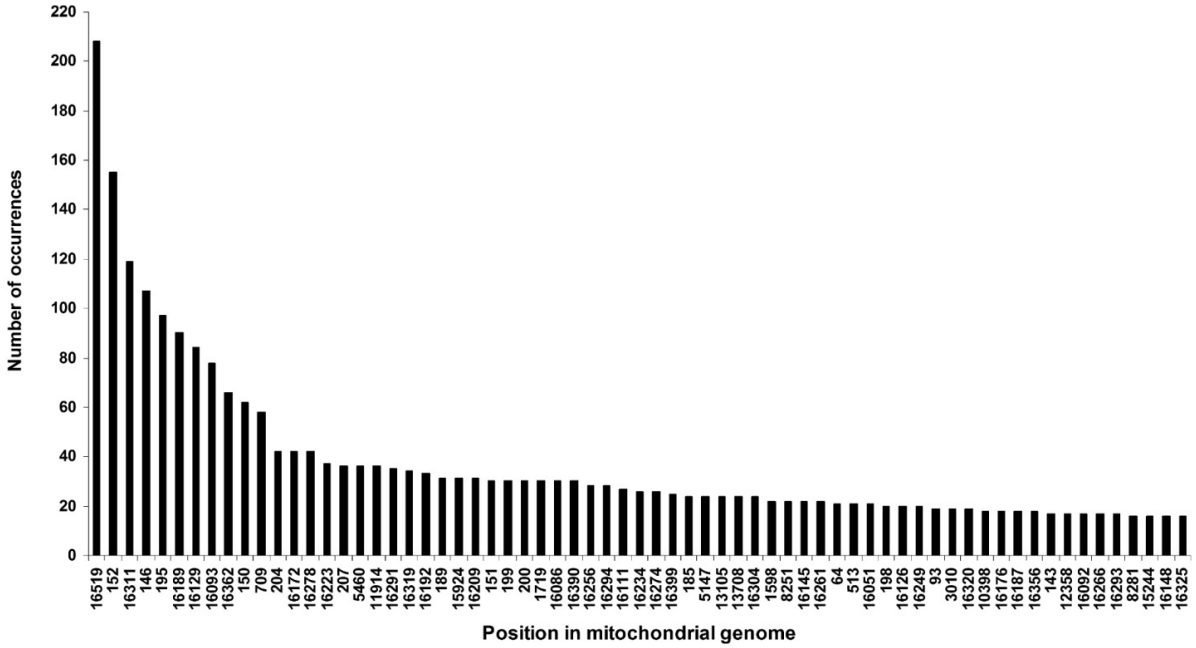


Figure 25: Hotspots in the mitochondrial DNA genome. Shown are mutations that appear >25 times in the global phylogenetic tree Phylotree. Published in Soares and colleagues [26].

3.5.6 Maternal inheritance

One of the key characteristics of the mitochondrial DNA is its obvious maternal inheritance [124]. Therefore, the mtDNA allows constructing the maternal lineage of humans to investigate past migration history of human populations (Figure 14, chapter 3.3.2).

However, when the mitochondria are passed on from the mother to the next generation, only a small number of mtDNA molecules will be in fact transferred to her offspring. This phenomenon is called the mitochondrial "genetic bottleneck" which apparently occurs during postnatal folliculogenesis [125]. The mtDNA "genetic bottleneck" is the result of an unequal segregation of different mtDNA genotypes during cell division, known as cytoplasmic segregation [126]. The reduced number of mtDNA molecules will eventually increase again in the offspring. If the mother is heteroplasmic, she can pass on different genotypes to her offspring. Homoplasy is present when all mtDNA copies are identical, and thus, the mother can transmit only one type of mtDNA to the next generation. In contrast, heteroplasmy means that two or more mtDNA genotypes coexist in one cell whereby the amount of mutant mtDNA can vary between cells [122]. If the mother is heteroplasmic, than she can pass on either only the original mtDNA genoptype, or only the mutant mtDNA, or both,

respectively, given that the mutation occurred in one of her egg cells. Mutations that occurred in other cells of the mother cannot be passed on to the next generation. This feature of the mtDNA during transmission can lead to an accumulation of mtDNA genotypes in offspring that were previously rare in the mother. The reason for the accumulation of a previously rare mtDNA genotype is that the mtDNA "genetic bottleneck" eventually result in a rapid and random genetic drift that leads to the loss of most newly arisen mutations, but also, to a rapid fixation of some mtDNA genotypes [113]. This means that a deleterious mutant mtDNA molecule in the mother can become common in her offspring, and thus result in an mtDNA-associated disease with varying degree of severity [122]. Nonetheless, the threshold effect is crucial for the degree of severity of an mtDNA-associated disease, that means, the proportion of mutant mtDNA must surpass a threshold to result in a defect of proteins in the respiratory chains [113, 127].

The majority of mutant mtDNA molecules probably remain undetected due to low concentrations. Up to the present day, a frequency of >10% of a mutant mtDNA genotype is required to be detected with Sanger sequencing [22, 128]. In a study of a West European population it was found that the frequency of heteroplasmic individuals is higher than 61% [129].

The mitochondrial DNA is most probably exclusively inherited down the maternal line although 'leakage' of paternal mtDNA might occur [24]. However, up to now paternal 'leakage' seems to be rather rare, and thus, the traditional view of maternal mtDNA transmission keeps widely accepted in the scientific community [113].

3.5.7 Lack of recombination

A prerequesite of recombination of mtDNA molecules is that the genome is passed on to the next generation by both parents. However, as paternal inheritance is rather unlikely [113] maternal and paternal mtDNAs cannot recombine although the molecular machinery that is necessary for recombination is available within mitochondria [22]. Previous studies apparently found evidence for mitochondrial DNA recombination [23, 130], however, controversy remains within the scientific community as most evidences of recombination so far could be attributed to errors in sequencing or databases [22].

3.5.8 Nomenclature

The nomenclature of the mitochondrial DNA is based on the revised Cambridge Reference Sequence, rCRS [121].Variations from the rCRS accumulate sequentially along the maternal lineage leading to different mtDNA haplotypes. A mitochondrial DNA haplotype is a combination of polymorphisms that differ from the rCRS and are inherited together from the mother to her offspring. Thus, similar mitochondrial haplotypes share a set of common mutations and can be traced back to a common maternal ancestor. Similar mitochondrial DNA haplotypes are clustered into monophyletic clades, that is, haplogroups [131]. This means that two individuals that share a mtDNA haplogroup can differ in their mtDNA haplotype through one or more mutations that are not associated with the haplogroup.The five samples shown in Figure 26 all belong to haplogroup L2a1c4a1 although haplotypes among all samples differ, because individuals have mutations that are only present in their haplotype. For example, DQ304951has a mutation (9497) that is only present in this individual. The other haplotypes, in contrast, exhibit at least twoor more mutations that are not associated with L2a1c4a1.

Back mutations in the mitochondrial genome, which are mostly unstable positions in the control region, are nucleotide changes back from the derived state to the ancestral state again. Back mutations are usually indicated with an @ or an !. Recurrent mutations are highly unstable mutations or uncertain up to the present and shown in brackets as it was done in Phylotree [132]. A nucleotide change is shown with the respective nucleotide (A, G, C, T) only in the case of atransversion (A-T, A-C, G-T, G-C mutations). Transitions (A-G or T-C mutations) are shown only with the respective position where the mutation occurred. Insertions are usually indicated with the number of insertions plus the respective nucleotide (e.g., 960.1C). Deletions are shown usually with the minuscule d (e.g., 8281-8289d).

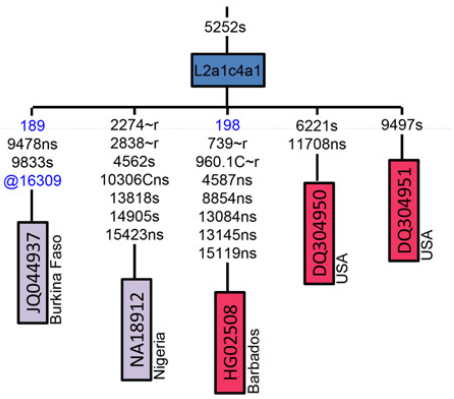


Figure 26: Mitochondrial DNA haplogroup L2a1c4a1 and its associated haplotypes. Mutations shown in blue belong to the control region and mutations in black to the coding region; s means synonymous and ns non-synonymous mutation. American samples are shown in pink, African sample in grey.

Each haplogroup is characterized by specific mutations and summarized in a phylogenetic tree of global human mitochondrial DNA variation accessible on http://www.phylotree.org [132]. For example, all haplogroups of L3 share the mutations 769, 1018, and 16311 (Table 1) which are also present in the haplogroups M or N, respectively,as both haplogroups are sub-branches of L3 (Figure 28, chapter 3.5.9.1).

As mentioned above, due to the lower effective population size of mtDNA, which makes the mitochondrial genome more sensitive to genetic drift, the mitochondrial molecule is suitable for detecting geographic population structure based on haplotypes or haplogroups, respectively. For example, haplogroups of the L-branch are typically distributed in Africa whereas haplogroup H are relatively common in Europe. The revised Campridge Reference Sequence belongs to haplogroup H2a2a1. Haplogroups A, B, C, and D are relatively widely distributed in Asia and America (Figure 27).

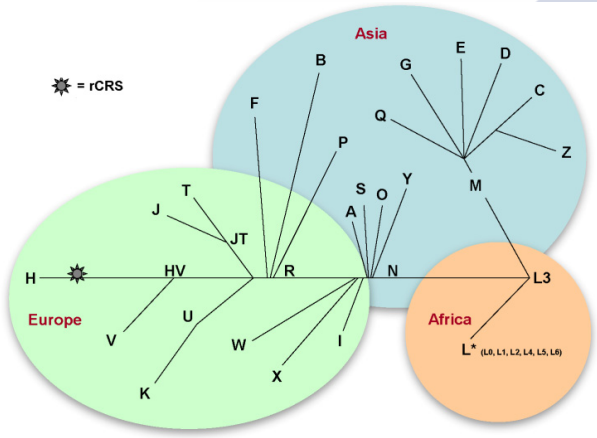


Figure 27: Simplified global mtDNA lineages distribution,http://www.mitomap.org/MITOMAP [32].

Table 1: Polymorphisms of haplogroups. Back mutations are indicated with an @, insertions as .X, deletions with a d, and transversions with the respective letter of the nucleotide after the position of the mutation.

Hg	Polymophisms	Population
L0	263 1048 3516A 5442 6185 9042 9347 10589 12007 12720	Africa
L1	3666 7055 7389 13789 14178 14560	Africa
L2	146 150 152 2416 8206 9221 10115 13590 @16311 16390	Africa
L3	769 1018 16311	Africa
L4	195 5460 16362	Africa
L5	459.1C 3423 7972 12432 12950 16148 16166	Africa
L6	146 152 185C 709 770 961 1461 4964 5267 6002 6284 9332 10978	Africa
	11116 11743 12771 13710 14791 14959 15244 15289 15499 16048 16224	
M	489 10400 14783 15043	Asia
C	3552A 9545 11914 13263 14318 16327	Asia
Z	6752 9090 15784 16185 16260	Asia
D	5178A 16362	Asia
E	3027 3705 7598 13626 16390	Asia
G	709 4833 5108 16362	Asia
Q	4117 5843 8790 12940 16129 16241	Asia
N	8701 9540 10398 10873 15301	Asia
A	235 663 1736 4248 4824 8794 16290 16319	Asia
S	8404	Asia
O	6755 9140 16213	Asia
Y	8392 10398 14178 14693 16126 @16223 16231	Asia
F	249d 6392 10310	Asia
B	8281-8289d	Asia
P	15607	Asia
R	12705 16223	Europe
JT	11251 15452A 16126	Europe
J	295 489 10398 12612 13708 16069	Europe
T	709 1888 4917 8697 10463 13368 14905 15607 15928 16294	Europe
HV	14766	Europe
H	2706 7028	Europe
V	4580	Europe
U	11467 12308 12372	Europe
K	10550 11299 14798 16224 16311	Europe
W	195 204 207 1243 3505 5460 8251 8994 11947 15884C 16292	Europe
X	6221 6371 13966 14470 16189 16278	Europe
I	10034 16129	Europe

3.5.9 Human population history

3.5.9.1 Overview

The mitochondrial DNA is only a tiny molecule when compared to the nuclear genome, but this small genome has given insightful information in the past about how humans spread across the globe. Cann and colleagues [133] were the first that could show that all present-day mtDNAs are rooted in Africa, and thus, concluded that all humans today descended from one specific female ancestor who lived in Africa about 200,000 years ago.Estimates about the Time of Most Recent Common Ancestor (TMRCA) on the matrilineal lineage, that is, the estimated time that has passed since all modern females shared a common female ancestor, vary from study to study. The TMRCA of maternal mtDNA ranges from 140,000-210,000 [134], over 171,500 +/- 50,000 [135], 192,400 (151,600-233,600) [26], and 194,300 +/- 32,550 [136], to 204,800 (116,800-295,700) [137]. All estimates about TMRCA are based on the molecular clock. The molecular clock postulates that the rate of sequence evolution is roughly constant, and thus, the time of divergence between sequences after they split from a common ancestor can be accurately obtained [22, 138].

Time estimates of branching events provide useful information about human dispersals across the globe. In terms of mitochondrial DNA, the control region DNA has proven to be an informative tool to investigate the past. However, whole mtDNA sequences might be more advantageous when the aim of the study is to investigate historical events that date back less generations. For example, Cerezo and colleagues [139] used whole mtDNA sequences and found out that about 65% of African L lineages came to Europe rather during recent migrations, probably during the time window of the Transatlantic Slave Trade.

The distribution of human mitochondrial DNA haplogroups and its time estimates are powerful tools to investigate patterns of human migration. The African-specific macro-haplogroup L of the mtDNA tree is the root of all modern mtDNAs as it represents the most ancestral maternal lineages. The L haplogroup can be divided into L0 and its sister branch L1’2’3’4’5’6 (Figure 28). The root of this branch is seen as the deepest root of the mtDNA tree. Haplogroup L3 diverges into haplogroups M and N which give rise to all non-African haplogroups [132].

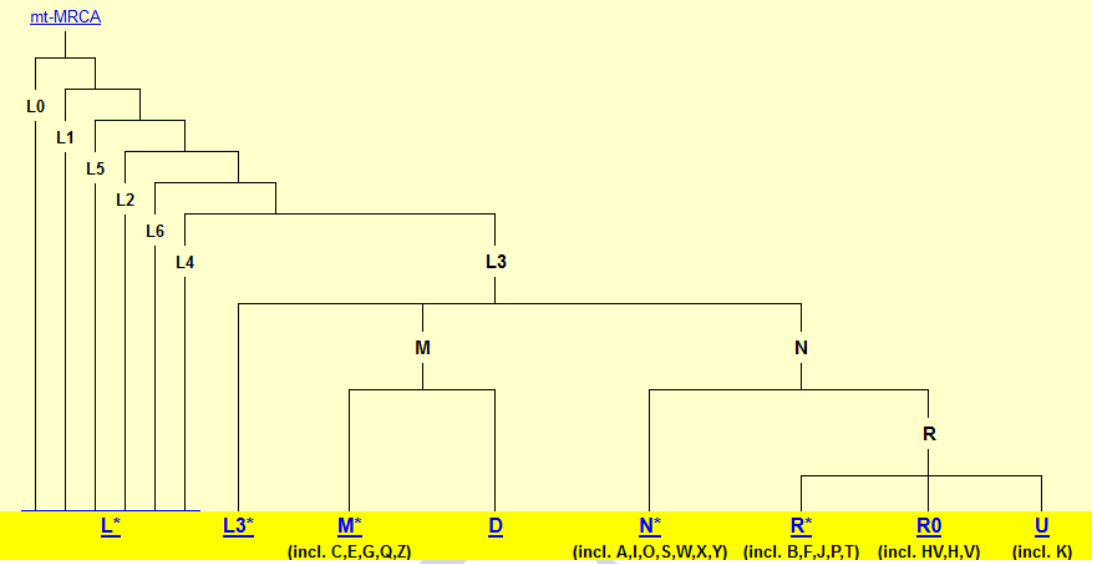


Figure 28: Overview of the global mtDNA phylogenetic tree [132].

3.5.9.2 African mtDNA lineages

L0 haplogroup

Haplogroup L0can be divided into five sub-branches, that is to say, in L0a, L0b, L0f, L0k, and L0d[132] and is estimated to have coalesced around 146,400 (± 25,100) [136], 149,700 (112,200-188,000) [26], 153,971 (± 11,905) [134], 158,700 (88,600-231,500) [137] or 168,247 (135,942-201,037) (unpublished data). Coalescence estimates for the sub-haplogroups are shown in Table 2.

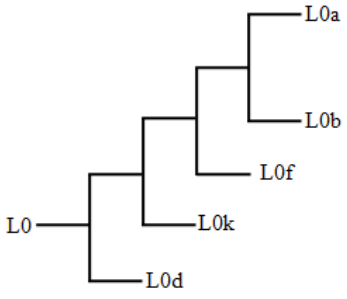


Figure 29: Clade of haplogroup L0 [132].

Table 2: Age estimates for L0 sub-haplogroups.

Publications	L0a	L0b	L0f	L0k	L0d
Salas et al. 2002 [140]	40,350 ± 16,250	-	-	-	49,600 ± 13,450
Gonder et al. 2007 [136]	54,600 ± 5,700	-	94,900 ± 9,400	70,900 ± 19,700	106,000 ± 20,200
Behar et al. 2008 [134]	53,176 ± 7,143	-	88,098 ± 9,524	39,683 ± 8,730	101,589 ± 10,318
Soares et al. 2009 [26]	44,700 (28,900-61,500)	-	108,000 (77,900-139,200)	11,200 (3,300-19,500)	-
Schuster et al. 2010 [137]	61,100 (29,200-95,100)	-	93,800 (46,900-139,700)	78,600 (37,200-122,000)	107,200 (56,800-159,900)
Barbieri et al. 2013 [141]	-	-	-	40,000 (28,000-53,000)	95,000 (79,000-121,000)
Publication in preparation (2015)*	62,256 (46,185-80,997)	69,354 (47,559-92,154)	92,877 (74,291-111,967)	41,654 (26,601-57,452)	100,128 (74,037-127,124)

* not publishedto date, calculated with all previously published whole mtDNA genomes found in literature

L0d is the most divergent haplogroup of L0. Initial studies of L0d found the highest frequency of haplogroup L0d in Southern African Khoisan (or Khoe-San) speakers (Figure 30A) [140, 142]. A principal distribution of L0d in Southern Africa was observed again in a later study in populations of South Africa, but this time also in Namibia, and BotswanawherebyBarbieri and colleagues [141] found the highest concentration of L0d in Khoisan speaking populations of central Botswana, but also in South African populations that have Khoisan ancestors [141]. Khoisans were observed to have contributed significantly to the maternal ancestry of South African Coloreds [143]. L0d was also found in lower frequencies in southeastern [141, 143, 144], southwestern [141], and eastern [141, 142, 145] Africa. Very low levels of L0d are present in South African "Whites" and "Indians", and South-western Bantu[143]. Beyond L0d's principal distribution, it was found in the Congoin one Bantu-speaking individual [146], and furthermore, in one individual of Kuwait [134]. L0d1 is common in all Khoisan populations whereas L0d2 apparently is geographically restricted to mainly Khoisan foragers and Bantu speakers of Botswana, but also to pastoralists of Namibia [141].

Like L0d, L0k was observed to be a haplogroup of the Khoisan speaking populations [134, 140]. However, in contrast to L0d, L0k seems to be more restricted, with its highest frequency in one Khoisan foragers population of Namibia, followed by other foragers of Namibia and Botswana of the same language group and a Bantu-speaking populations of southwestern Zambia (Figure 30B) [141]. The sub-lineages of L0k, L0k1b and L0k2, are distributed almost exclusively in

populationsof Bantu speakers in Zambia and are absent in Khoisan speaking populations [141]. A very low level of L0k was detected in Southeastern Bantu speakers [143] although L0k was absent in Southeastern Bantu speakersin another study [140]. Apart from that, L0k, was found in one individual from Yemen [134].

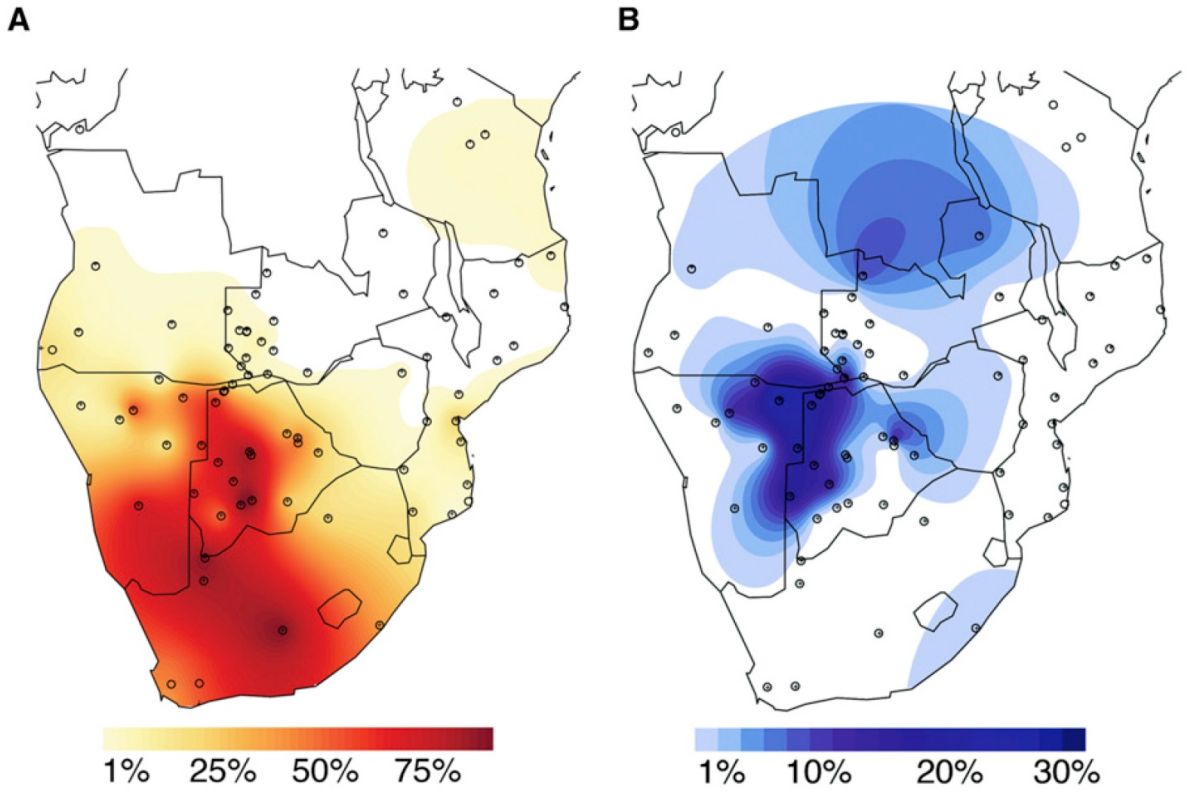


Figure 30: Spatial distribution of L0d and L0k in Africa. Haplogroup L0d (A) and Haplogroup L0k (B) are shown with a different scale. Published in Barbieri and colleagues [141].

L0f seems to be restricted to East Africa where it occurs in low frequency [136, 140] like, for example, in Ethiopia [134], Uganda [147], and Kenya [148]. However, L0f was also found in two individuals of South Africa and in one sample of the Near East in Oman [134]. Interestingly, one individual from the USAwas also seen to belong to haplogroup L0f [134], given the fact, that most enslaved African people left the continent from West and West-Central, and southeastern African ports to the Americas during the TAST [1].

Likewise L0f, L0b seems to occur so far exclusivelyin East Africa as the haplogroup was detected in one sample from Ethiopia [134] and in three samples from Kenya [148]. Due to its low frequency coalescence age was not calculated in previous publication.

In contrast to the other L0 sub-haplogroups, L0a is more widely distributed on the African continent. Salas and colleagues [140] found a principal distribution of L0a in East, Central and southeastern Africa, whereas the haplogroup was almost absent in North, West and southern Africa. L0a, or a subset of L0a mtDNAs, respectively, probably have been brought from East to southeastern Africa through the Bantu expansion [140, 144]. In East Africa L0a could be found in Ethiopia [134] and Kenya [134, 148, 149]. In southeastern Africa L0a was observed in Mozambique [134, 140] and Zambia [150] and in Central Africa in pygmies of the Central African Republic [151] and Congo [135, 146, 151], but also in Burkina Faso [152],Chad [134], and Nigeria [148]. Apart from its main distribution, L0a was also found sparsely in South Africa, including Bantu speakers [28, 134], in West Africa like one individual from Guinea-Bissau [134] and two from Sierra Leone [148], and in two individuals from North Africa, one from Algeria [134] and one from Morocco [153]. Outside Africa L0a appeared in the Near East like Egypt [134, 154], Saudi-Arabia, Oman, Yemen, Israel, and Iran [134]. In the Americas L0a was detected in individuals from North America, that is, the United States [134, 148, 155], but also in South America in one individual from Colombia [148]. In the Caribbean L0a was observed in Barbados, Puerto Rico [148] and the Dominican Republic [156].

In an African sample set of 1726 whole mtDNA genomes, 604 (35%) individuals belonged to haplogroup L0.

L1 haplogroup

Haplogroup L1can be divided into the branches L1b and L1c (Figure 31) [132]. TMRCA for L1 lies around 140,400 (± 32,900) [136], 140,600 (107,600-174,300) [26], 142,007 (110,135-174,599) (unpublished data) and 146,828 (± 11,905) [134] (Table 3).

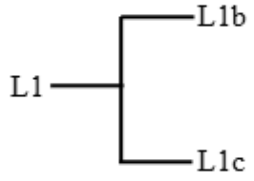


Figure 31: Clade of haplogroup L1 [132].

Table 3: Age estimates for L1 sub-haplogroups.

Publications	L1b	L1c
Salas et al. 2002 [140]	30,550 ± 16,250	59,650 ± 11,800
Gonder et al. 2007 [136]	15,000 ± 26,700	95,600 ± 23,900
Behar et al. 2008 [134]	29,366 ± 7,937	102,383 ± 7,937
Soares et al. 2009 [26]	9,700 (5,200-14,300)	85,400 (63,500-107,900)
Publication in preparation (2015)*	39,158 (25,879-53,040)	90,015 (73,797-106,632)

* not published to date, calculated with all previously published whole mtDNA genomes found in literature

Haplogroup L1b exhibits its highest concentration in West Africa, but can be found also to a lesser extent in Central and North Africa. Frequency of L1b seems to be lower in eastern, southeastern, and southern Africa [140]. In West Africa, L1b could be found in Guinea-Bissau [134], Guinea [157], Gambia, Sierra Leone [148], and in the Mandenka of Senegal [152]. In Central Africa L1b appearedat a medium-high frequency level in Mande- and Gur-speaking populations of Burkina Faso [152], but is distributed also in Yoruba [148, 152], and other ethnic populations of Nigeria [135, 148], in Chad [134], andin pygmies of Gabon [151]. In northern Africa, L1b was observed in Mauritania [153] and Tunisia [134, 158]. In East Africa, individuals exhibited occasionally haplogroup L1b like in Ethiopia [134] and Kenya [148] and in southeastern Africa in Zambia [150]. In southern Africa, in contrast, L1b seems to be almost absent [28]. Apart from Africa, L1b occurred sparselyin the Near Eastern countries of Egypt, Israel, Jordan, and Syria [134], in Europe like Portugal [134], Spain, Italy [148], and Poland [159], and even in Russia, Asia [160].

However, L1b is a common haplogroup seen in African Americans [16, 17, 140, 148, 155], and furthermore, appears in Caribbean populations like Barbados and Puerto Rico [148]. African Americans seem to share their L1b mtDNA sequences mostly with individuals from West Africa [17]. This is consistent with the history of the Transatlantic Slave Trade, as the greatest number of African enslaved people were forced to migrate to the Americas from West and Central Africa [1].

L1c seems to be principally distributed in Central Africa and appears in low frequency in West and Southeastern Africa [140]. L1c is a typical haplogroup of pygmies [135, 146, 151, 161]. However, two sub-lineages of L1c (L1c2a and L1c2b) are apparently characteristic not for pygmies, but for

Bantu speakers in Zambia [150].L1c occurred in West Africa in Guinea-Bissau [134], Gambia and Sierra Leone [148], in central Africa in some samples of Nigeria [135, 148], and in one sample of Burkina Faso [152], and in southeastern Africa in Mozambique [134]. Apart from its main distribution L1c was observed occasionally in East Africa in Kenya [135, 148], in North Africa in Morocco [134] and Western Sahara (sample from FamilyTreeDNA), and in southern Africa in South Africa [134]. Outside Africa L1c appeared in Portugal, Syria [134] and Egypt [154]. In the Americas L1c was observedin North America, USA [134, 148, 149], but most L1c haplotypes of African Americans apparently do not match with L1c haplotypes from continental Africa [140]. Most probably due to the TAST L1c is also distributed in South America like Colombia [148], and in the Caribbean like Barbados, Puerto Rico [148], and Dominican Republic [156].

Out of 1726 African sample 261 (15.1%) could be ascribed to haplogroup L1.

L2 haplogroup

Haplogroup L2 (Figure 32) encompasses the haplogroups L2a, L2b, L2c, L2d, and L2e [132] and has a TMRCA that seems to lie between 70,100 (± 15,300) [140], 89,300 (68,100-111,100) [26], 94,500 (± 4,500) [136], 104,764 (± 8,730) [134], and 119,026 (95,604-142,994) (unpublished data) (Table 4).

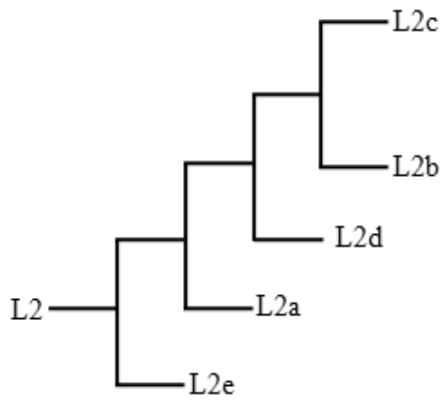


Figure 32: Clade of haplogroup L2 [132].

Table 4: Age estimates for L2 sub-haplogroups.

Publications	L2c	L2b	L2d	L2a	L2e
Salas et al. 2002 [140]	27,500 ± 7,250	31,600 ± 11,200	121,900 ± 34,200	55,150 ± 19,350	-
Behar et al. 2008 [134]	23,810 ± 3,175	26,985 ± 3,968	23,810 ± 7,143	46,033 ± 7,937	46,826 ± 7,143
Soares et al. 2009 [26]	25,200 (13,400-37,700)	14,300 (7,000-21,900)	-	48,300 (31,100-66,300)	-
Publication in preparation (2015)*	22,782 (18,857-26,775)	30,405 (21,662-39,448)	19,132 (12,833-25,624)	97,908 (76,283-120,170)	45,532 (32,752-58,809)

* not published to date, calculated with all previously published whole mtDNA genomes found in literature

L2e is apparently mostly distributed in central Africa like in Nigeria [148], but also in West Africa like in Sierra Leone, Gambia [148], and Guinea-Bissau [134]. However, it was also found in one individual of North Africa from Tunisia [158] and in low frequency in African Americans [162].

L2a is the most common and mostly distributed haplogroup of L2 and seems to be principally distributed in Central and West Africa, that is, especially in Burkina Faso [152], Nigeria [135, 148, 149, 152] and pygmies [146, 151], but also in Chad [134] and Benin [163] in Central Africa and in West Africa especially in Sierra Leone and Gambia [148], but alsoin Guinea-Bissau [134] and Senegal [152]. However, L2a seems to occur often as well in East Africa like in Kenya [134, 146, 148, 149] and Ethiopia [134] and was also observed in southeastern Africa like in Zambia [150] and Mozambique [134]. L2a appeared also in North Africa like in Tunisia [134, 158] and Morocco [134] and in southern Africa like in South Africa [28, 134]. In the Near East L2a was observed in Egypt [134, 154], Israel [134, 146], Yemen, Oman, Jordan, Kuwait, Syria, and Libya [134], in Europe in Spain [164], Portugal [134], and Slovakia [160], and in Asia in Pakistan [134]. L2a is considered as the most widely distributed haplogroup within Africa [17] and apparently is also the most common sub-Saharan African haplogroup in the Americas [17, 162]. In North America L2a is especially distributed among African Americans [148, 155, 162]. In South America L2a was observed in Colombia [90, 148] and in Peru [148]. In the Caribbean L2a could be found especially in Barbados, but was also seen in one individual from Puerto Rico [148] and one individual of the Bahamas (sample from FamilyTree DNA).

In comparison to L2a, L2d, but also L2b and L2c are less common haplogroups of L2 and are possibly principally confined to West and West Central Africa [17, 140]. In Central Africa L2d was

observed in Burkina Faso [152] and Nigeria [148], in West Africa in Gambia and Sierra Leone [148] and in North Africa in one individual from Algeria [134]. Apart from Africa, L2d was infrequently found in the Yemen [134], and again probably due to the TAST in the Caribbean like Barbados [148] and Dominican Republic [156], and in South America [146]. However, L2d in general seems to occur rather rarely [165].

Like L2d, L2b seems to be largely confined to West and Central Africa. It was observed in Nigeria [148], Burkina Faso [152], and pygmies [151] in Central Africa and in Gambia, Sierra Leone [148], and Guinea-Bissau [134] in West Africa. In other parts of Africa L2b appearseven more infrequently, although it was seen in North Africa in Tunisia [158], in East Africa in Kenya [148], in southeastern Africa in Zambia [150] and Mozambique [134], and in southern Africa in South Africa [28, 134]. L2b can be also foundsparsely in Egypt, Saudi-Arabia, and Israel in the Near East [134] and in North America in African Americans of the U.S. [148, 155] and in the Caribbean in Barbados and Puerto Rico [148]. However, as L2b, but also L2c, are less frequent than L2a in Africa [140], correspondingly due to the TAST, L2b and L2c are less found in the Americas [162].

Like mentioned above L2c has a rather West and Central African distribution [140]. In West Africa L2c was observed especially in Gambia and Sierra Leone [148], but also in Senegal [152] and Guinea-Bissau [134] and in Central Africa in Burkina Faso [152] and Nigeria [148, 149]. L2c was sparsely found in North Africa in Morocco [134] and Mauritania [153], in southern Africa in South Africa [28], and in southeastern Africa in Zambia [150] and Mozambique [134]. In Europe L2c was more infrequently seen, however, it appeared in The Netherlands and in the Near East in Lebanon [134]. In the Americas L2c was detected in African Americans from the U.S. [148, 155] and in the Caribbean in Barbados and Puerto Rico [148].

In 1726 African samples 333 (19.2%) belonged to haplogroup L2.

L3 haplogroup

Haplogroup L3 can be divided into L3a, L3b, L3c, L3d, L3e, L3f, L3h, L3i, L3k,and L3x (Figure 33), and additionally, give rise to the non-African haplogroups M and N [132]. Time of divergence for L3 ranges between 61,300 (± 11,650) [140], 70,783 (58,384-83,491) (unpublished data) and 76,192 (± 4,762) [134] (Table 5). L3 probably originated in East Africa as most sub-clades of L3 are principally distributed in East Africa [156, 166].

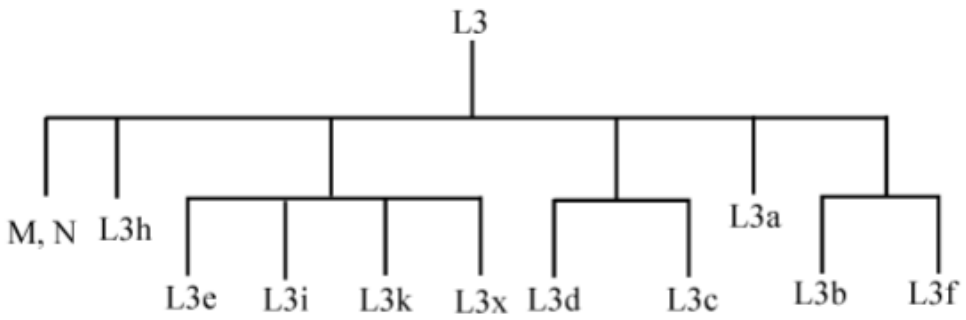


Figure 33: Clade of haplogroup L3 [132].

Table 5: Age estimates for L3 sub-haplogroups.

Publications	L3h	L3e	L3i	L3k	L3x
Salas et al. 2002 [140]	-	49,250 ± 11,750	-	-	-
Behar et al. 2008 [134]	70,636 ± 5,556	44,445 ± 5,556	45,239 ± 9,524	38,096 ± 7,143	38,096 ± 5,556
Soares et al. 2009 [26]	66,700 (52,000-81,800)	39,000 (27,200-51,200)	34,800 (19,900-50,500)	-	36,100 (23,300-49,600)
Publication in preparation (2015)*	67,156 (55,005-79,622)	44,296 (33,715-55,222)	51,113 (37,310-65,447)	34,723 (23,679-46,215)	39,071 (27,026-51,610)
Publications	**L3d**	**L3c**	**L3a**	**L3b**	**L3f**
Salas et al. 2002 [140]	30,250 ± 8,450	-	-	21,600 ± 6,850	36,600 ± 12,800
Behar et al. 2008 [134]	38,096 ± 5,556	11,905 ± 5,556	65,081 ± 9,524	28,572 ± 4,762	60,319 ± 6,349
Soares et al. 2009 [26]	31,000 (19,200-43,300)	-	-	16,400 (9,800-23,200)	53,200 (38,600-68,300)
Publication in preparation (2015)*	31,221 (25,033-37,556)	9,863 (2,342-17,720)	56,138 (41,848-70,951)	25,741 (18,896-32,785)	56,214 (42,789-70,100)

* not published to date, calculated with all previously published whole mtDNA genomes found in literature

Among all L3 haplotypes L3e is the most widespread and frequent haplogroup of L3 [140]. L3e can be further divided into L3e1, L3e2, L3e3, L3e4, and L3e5. As all five sub-branches are distributed in Central Africa L3e probably originated in Central Africa [166].

L3e1 is estimated to have diverged between 15,873 (± 2,381) [134], 21,949 (17,046-26,961) (unpublished data) and 32,150 (± 11,450) [140] and is probably of (West) Central African origin

[140]. The haplogroup was observed in Central Africa like in Burkina Faso [152], Nigeria [148, 149, 152], Chad [134, 166], Cameroon [135], and in pygmies [151] and was the most frequent haplogroup within L3e in Angola [167]. L3e1 seems to be rare in West Africa [140] where it was seen, for example, in Gambia [134]. The haplogroup L3e1 appeared frequently in southeastern African Bantu speakers [140] like in Zambia [150] and Mozambique [134]. L3e1 probably reached East Africa through the eastern Bantu migration [140] where it can be found in Kenya [134, 148, 149] and Somalia [166]. Despite its main distribution in central and southeastern Africa, L3e1 was also found in North Africa like in Tunisia and in southern Africa like in South Africa [134]. In the Near East L3e1 was observed in Israel, Jordan, Syria, Oman, and Saudi-Arabia [134].

Coalescence age for L3e2 came to 26,985 (± 7,143) [134], 27,554 (15,392-40,341) (unpublished data) and 37,400 (± 18,350) [140]. Like L3e1, L3e2 is common among southeastern African Bantu speakers, but is apparently more frequent in Central and West Africa [140]. In central Africa L3e2 appeared in Nigeria [148, 149, 152], Burkina Faso [152], Chad [166], Congo [135], and pygmies [151], in West Africa in Guinea-Bissau [134], Gambia, and Sierra Leone [148] and in southeastern Africa in Zambia [150]. Apart from its main distribution L3e2 was also seen in Kenya in East Africa [134, 148] and in Morocco in North Africa [168]. In the Near East L3e2 appeared in Egypt [134, 154], Oman, and Yemen [134]. Its sub-clade L3e2b, which is found primarily in West Africa, was possibly carried to Angola during the Bantu migrations [167].

Time of divergence for L3e3 was estimated to reach around 14,150 (± 4,500) or 16,746 (9,683-24,065) (unpublished data) [140]. It is distributed in West Africa in Sierra Leone and Gambia [148], in Central Africa in Nigeria [148, 152], Burkina Faso [152], Chad [134, 166], Cameroon [135], and pygmies [151], in southeastern Africa in Mozambique [134], and in east Africa in Kenya [148] and Somalia [166]. In the Near East L3e3 was detected so far in Israel and Oman [134].

Coalescence age for L3e4 was estimated to come to around 20,314 (10,319-30,790) (unpublished data) or 24,200 (± 10,400) [140]. It was found in West Africa in Gambia, Sierra Leone [148], Guinea-Bissau [134], and Senegal [152], in Central Africa in Nigeria [148], and in southeastern Africa in Mozambique [134]. The haplogroup was also observed in the Lebanon in the Near East [134].

Time estimate of L3e5 was calculated to be 14,227 (8,379-20,257) (unpublished data). L3e5 is mostly restricted to northwestern Africa where it was detected, for example, in Tunisia [158], but

occurred so far also in central Africa like in Chad [26] and Nigeria [156] and in east Africa in Sudan [166]. L3e5 was also detected in Egypt and Libya in the Near East [134].

L3e was observed to be the most commonly found haplogroup of clade L3 in African Americans [162], and furthermore, seems to occur proportionally more often in South America due to the pronounced importation of enslaved people from West-Central and southeastern Africa [17]. Indeed, L3e was detected to be the most frequent sub-Saharan haplogroup in a population of Brazilians [84]. Haplogroup L3e1 was observed in the U.S. [148, 155], in Colombia [148, 149], in Barbados, and Puerto Rico [148] whereas L3e1 haplotypes in the Americas are apparently most similar to southeastern African haplotypes. Similar was found in Brazil, as haplotypes within L3e1 from Brazil match highly with haplotypes from Mozambique [17]. Haplogroup L3e2, particularly L3e2b, is widely distributed in African Americans [140, 148, 155], but was detected also in Barbados, Puerto Rico [148], Dominican Republic [156], and in Colombia [148]. L3e3 was so far detected in African Americans of the U.S. [148, 155] and in Peru [148], and the less frequent haplogroup L3e4 in the Americas was seen in the U.S., Barbados, and Colombia [17, 148].

L3i, L3k, and L3x are sister-clades to L3e. L3i and L3x are almost exclusively found in East Africa [166]. L3i was detected in East Africa in Ethiopia [156, 166] and Kenya [148] and L3x in Ethiopia [134, 156, 166], Kenya [148], Somalia, and Sudan [166], but also in North Africa in Algeria [134]. Haplogroup L3k is rare and probably principally found in North Africa [166] where it was detected in Libya and Tunisia [134] and apart from that also in Chad [166] and Gambia [148]. In the Near East L3i was observed in Kuwait and Oman and L3x in Israel and Jordan [134]. In the Americas L3i was detected in Gujarati Indians in the U.S., L3k in Barbados and the U.S. [148], and L3x in African Americans of the U.S. [162].

L3a and L3h seem to be distributedprincipally in East Africa [166]. Indeed, L3a can be found in Ethiopia [134, 156, 166], Somalia [166], and Kenya [148]. L3h, in contrast, was detected not only in East Africa like in Ethiopia [134, 156, 166], Somalia, Sudan [166], Kenya [148], and Burundi [135], but also in North Africa like Tunisia [134] and Morocco [168], in West Africa like in Guinea-Bissau [134], Gambia, and Sierra Leone [148] and in Central Africa like in Chad [134, 166]. Outside of Africa, L3h was detected in the Yemen [134, 166], in the Lebanon, and in Egypt in the Near East [134], and unlike L3a, L3h was so far also sparsely detected in African Americans [148, 162].

The haplogroups L3b and L3f form one sister-clade within the macro-haplogroup L3. L3b and L3f, together with L3e and L3d are the most common haplogroups of L3 in Central and West Africa [166]. L3b was seen in Central Africa in Nigeria [148], Chad [134, 166], and Burkina Faso [152] and in West Africa in Sierra Leone [148], Gambia [134, 148], Guinea-Bissau [134], and Senegal [152]. Like L3b, L3f was observed in Central Africa in Nigeria [148, 152, 156], Chad [134, 166], and Burkina Faso [152] and unlike L3b in West Africa only in Guinea-Bissau [134] and Sierra Leone [148]. Both haplogroupsspread into North, East and Southern Africa via migrations within Africawhereas L3b is most probably of West and Central African origin and L3f of East African origin [166]. L3b was observed in North Africa in Tunisia [134, 158], Morocco [168], and Mauritania [153], in East Africa in Kenya [134, 148, 149], Somalia, and Sudan [166], in southeastern Africa in Zambia [150] and Mozambique [134], and in southern Africa in South Africa [28]. In contrast, L3f was found in North Africa in Morocco [168], in East Africa in Ethiopia [134, 156, 166], Kenya [134, 149], Sudan, and Somalia [166], in southeastern Africa in Zambia [150] and Mozambique [134] and in southern Africa in South Africa [134]. Both haplogroups were detected in the Near East, that is, both haplogroups were detected in Egypt [134, 154], Israel, Syria, Yemen, and Jordan whereas L3b appeared also in Libya and Kuwait and L3f in Lebanon, and Oman [134]. Furthermore, L3b was also observed in Asia, that is, in Russia [160] and in Southern Europe in Cyprus [134]. As L3b and L3f are common in West and Central Africa, both haplogroups appear in the Americas most possibly as a result of the TAST. In the Americas L3b is distributed in African Americans of the U.S. [134, 148, 155, 162], but also in Colombia in South America [148], and in the Dominican Republic [156], Barbados and Puerto Rico in the Caribbean [148]. Like L3b, L3f was observed in African Americans from the U.S. [148, 155] and in the Caribbean in Barbados, but in South America in Peru [148].

L3c and L3d build a sister-clade within haplogroup L3. L3c is an extremely rare haplogroup which was found so far in Ethiopia in East Africa [156] and in Israel in the Near East [134]. In contrast, L3d together with L3b, L3e, and L3f represent the most common L3 lineages in West and Central Africa [166]. In West Africa L3d was observed in Gambia, Sierra Leone [148], Senegal [152], and Guinea-Bissau [134] and in Central Africa in Nigeria [146, 148, 152], Chad [134, 135, 166], Burkina Faso [152], and pygmies [151]. L3d most probably originated in West and Central Africa from where it spread into East Africa [166]. The haplogroup can be found in Kenya [134, 148], Somalia, and Sudan [166], but was observed as well in North Africa like in Tunisia [134, 158] and Morocco [168], in southeast Africa in Zambia [150] and Mozambique [134], and in southern Africa

in South Africa [28, 134]. L3d was detected also in the Near East like Jordan [153], Yemen, Egypt, and Syria [134], in Europe like Poland [160] and Spain [148], and in Asia in Pakistan [148]. L3d is due to its principal West African distributioncommon in African Americans [17, 140, 148, 162] and appeared as well in the Caribbean like in Barbados, Puerto Rico [148] and in the Dominican Republic [156].

Out of the previously mentioned African whole mtDNA genome data set 484 (28%) individualsbelonged to haplogroup L3.

L4 haplogroup

Haplogroup L4 can be divided into the two clades L4a and L4b (Figure 34). Time of divergence for L4 is estimated to reach around 85,508 (68,590-102,891) (unpublished data) or 95,240 (± 7,143) [134].

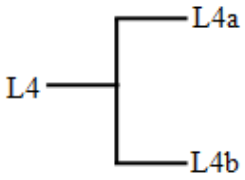


Figure 34: Clade of haplogroup L4 [132].

Age estimates for L4a and L4b probably lie around 46,585 (33,327-60,369) and 83,556 (66,697-100,891) (unpublished data). L4 seems to occur mainly in East Africa where it was observed in Tanzania [142], Kenya [148, 149], and Ethiopia [134, 156]. However, L4 was also detected so far in North Africa in Tunisia [158], in Central Africa in Nigeria [148], Burkina Faso [152], and in pygmies [151], in West Africa in Gambia and Sierra Leone [148], and in southern Africa in South Africa [134]. Outside of Africa, L4 was observed in the Near East in Yemen, Israel, Saudi-Arabia, and Syria and in Asia in Pakistan [134]. Probably due to its principal East African distribution, L4 is less frequent in the Americas than other macro-haplogroups (L0, L1, L2, L3). However, L4 was observed in African Americans of the U.S. [162], but also in the Caribbean in Barbados [148].

Haplogroup L5

Haplogroup L5 is split into two sister-clades, that is, into L5a and L5b (Figure 35). Coalescence age of L5 was estimated to lie between 82,950 (± 24,900) [140], 95,900 (± 26,200) [169], 120,200 (114,200-126,200) [26], 129,400 (± 22,100) [136], 138,098 (± 11,905) [134], 150,964 (107,286-195,842) (unpublished data).

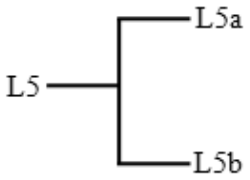


Figure 35: Clade of haplogroup L5 [132].

L5a, formerly known as L1e [170], whose coalescence age is estimated to lie around 33,604 (21,579-46,173) (unpublished data) or 37,302 (± 7,143) [134] is principally restricted to East Africa. L5b, which diverged between 31,747 (± 8,730) [134] and 74,997 (51,802-99,236) (unpublished data) years ago, apparently has a more southern distribution [140, 169] as it was observed in East Africa in Kenya [148, 149], Ethiopia [134, 169], and Tanzania [136]. In other parts of Africa L5 was seen in Central Africa in pygmies [151] and in Chad [134] and in southeastern Africa in Zambia [150] and Mozambique [134]. Outside of Africa L5 appeared in Egypt and Kuwait in the Near East, in Pakistan in Asia [134], and even in the Caribbean in the Dominican Republic [156].

L6 Haplogroup

Haplogroup L6, which can be divided into L6a and L6b, is estimated to have diverged between 21,474 (-6,671-53,791) (unpublished data), 22,223 (± 5,556) [134] and 38,900 (± 25.600) [169] years ago (Figure 36).

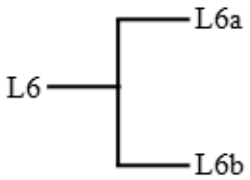


Figure 36: Clade of haplogroup L6 [132].

L6 was the most frequent haplogroup in a population from the Yemen [169], but was so far also detected in Egypt, Saudi-Arabia, and Israel in the Near East [134]. L6 possibly originated in East Africa as L6 can be found in Ethiopia [156, 169] and L6's sister haplogroups L2, L3, and L4 are diverse and frequently distributed in East Africa [169].

3.5.9.3 American mtDNA lineages

The basal haplogroups A, B, C, and D are haplogroups that are frequently found in Asia and America whereas mitochondrial DNA genomes from the Americas mostly belong to A2, B2, C1, and D1 [171]. The ancestral haplogroup of A is N, for B it is haplogroup R, and for C and D haplogroup M [132]. The haplogroups A, C, and D most likely came to the Americas from northern Siberia via the Bering land bridge whereas haplogroup B probably arrived later as it is absent in Siberia and only infrequently seen in northern North America, and furthermore, less diverse than the three previous haplogroups [28].

Within haplogroup A the pan-American haplogroup A2 is distributed across both continents whereas the distribution of A2a and A2b seems to be restricted to North America [172, 173]. B2 of haplogroup B is like A2 found both in North and South America [172, 174]. The three sub-haplogroups C1b, C1c, and C1d (including C1d1) of haplogroup C are pan-American [172, 174, 175] whereas C4c appears only in North America [174, 176]. D1 and D4h3a of haplogroup D are also distributed across both North and South America. In contrast, D2a, D3 and D4e1 are found only in North America [174, 175]. The haplogroups mentioned above (A2, A2a, A2b, B2, C1b, C1c, C1d, C1d1, C4c, D1, D2a1 [formerly D2a], D3, D4e1c [formerly D4e1], D4h3a), but also X2a and X2g belong to the 16 maternal founding lineages in the Americas [177, 178]. X2a and X2g are both confined to North America [175, 179] (Figure 37).

Whole mitochondrial DNA genome sequencing has improved considerably the resolution of American-specific mtDNA lineages in order to detect clades that appeared from the founding haplogroups in the Americas (A, B, C, D) during and after human arrival. The geographic distribution of clades and its ages can give insightful information about the colonization process of early humans in the Americas [177]. Achilli and colleagues [178] found that the first wave of peopleing the Americas following the Pacific route left the greatest genetic mark on the continents. However, within-continent migrations mirrowed in the geographic distribution and age estimates of X2a and C4c changed the Amerindian maternal genetic pool in northern North America. It was also

observed by means of A2a, migrations back to Asia and to the circumpolar regions of nowadays Canada (Figure 37) [178].

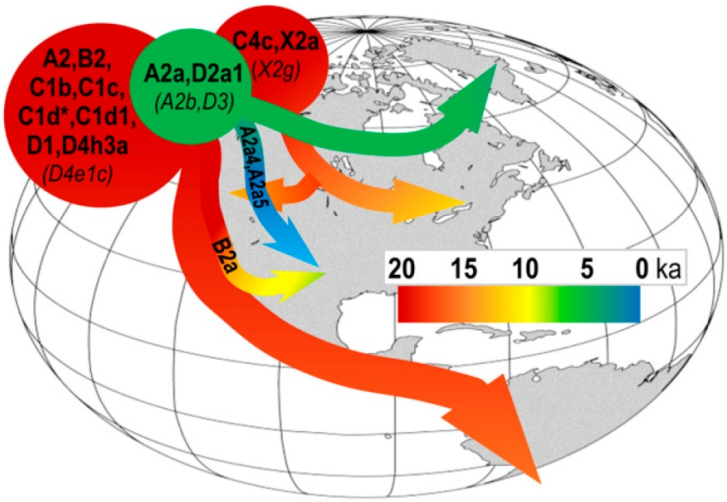


Figure 37: Three major migratory events based on 16 maternal founding lineages. Founder lineages in parenthesis and in italics are not sufficiently analyzed. One migration event based on A2a4 and A2a5 and another based on B2a are additionally shown. Published in Achilli and colleagues [178].

3.5.9.4 European mtDNA lineages

The Revised Cambridge Reference Sequence (rCRS) for the mitochondrial genome belongs to a typical European-specific haplogroup, that is, H2a2a1. Haplogroup H together with the haplogroups I, J, N1b, T, U, V, W, and X make up >98% of all mtDNAs found in Europe.All of them arederived from the macro-haplogroup N [28].

Haplogroup H is mostly found in Europe (40-50%), followed by the Near Eastand the Caucasus region (~20%) and the Gulf (<10%) (Figure 38). The largest sub-haplogroup is H1 which peaks on the Iberian Peninsula and decreases further North and East. H3 is the second largest sub-haplogroup of H.

Haplogroup H arrived in Europe from the Near East about ~20,000-25,000 years ago [180]. At first view haplogroup H seems to be distributed uniformly across Europe. However, Loogväli and colleagues [181] observed a complex structure of sub-haplogroups that can be useful for phylogenetical analyses.

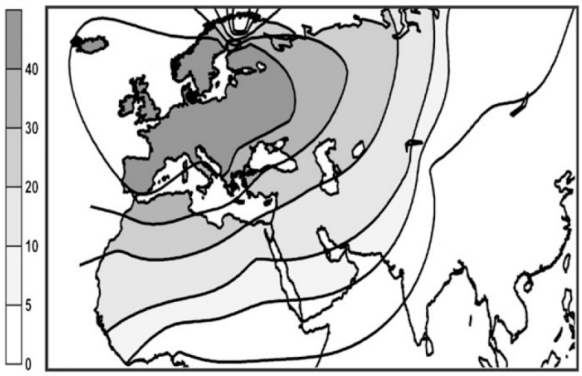


Figure 38: Spatial frequency distribution of haplogroup H. Published in Loogväli and colleagues [181].

Both haplogroup H and haplogroup V derive from HV [132]. Haplogroup V is especially common in the Saami of Scandinavia [182]. Haplogroup J and T derive from JT [132] whereas J was observed, for example, in Polish Roma [183] and haplogroup T appeared in Sephardic Jews [184]. The haplogroups I and W were like J, for example, found in Polish Roma [183]. N1b is especially present in Southwest Asia [185]. Haplogroup U appears both in the European and Near Eastern population, respectively, whereas the haplogroup seems to have a similar age in both regions [186].

3.6 Y chromosome

3.6.1 Organisation of the Y chromosome

The Y chromosome and the X chromosome represent the sex chromosomes in humans. The presence of the Y chromosome determines the gender of a person due to the production of the SRY protein (Sex-determining region protein), which is required for male sex determination [187].

The Y chromosome is a molecule of about 60 Mb in size and the MSY (male-specific region of the Y chromosome, see below) can be divided into euchromatin (transcriptionally active part, extended conformation) and heterochromatin (transcriptionally inert part, highly condensend conformation) region (Figure 39) [22].

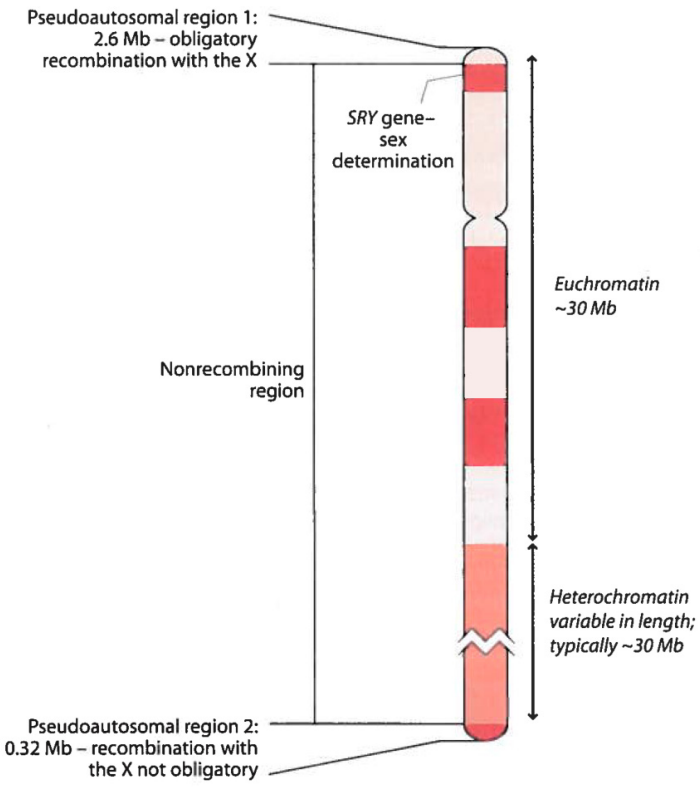


Figure 39: The Y chromosome. Shown is a G-banded Y chromosome (stained with Giemsa), Published in Jobling and colleagues [22].

The euchromatic part of the Y chromosome is approximately 23 Mb in size whereas about 8 Mb are located on the short arm and about 14.5 Mb on the long arm. The euchromatin can be divided into three classes: X-transposed, X-degenerate, and ampliconic (Figure 40). The X-transposed region (3.4 Mb) shows 99% identity to the X chromosome, is distributed in two blocks on the proximal (towards the centromer of a chromosomal arm) Y chromosome, and contains two coding genes. The X-degenerate region (8.6 Mb), in contrast, represents a pseudogene that is homologous to X-linked genes and distributed in 8 blocks on the proximal and distal (towards the telomere of a chromosomal arm) Y chromosome and which encodes for 16 genes. The apmpliconic region (10.2 Mb) appears in 7 blocks on both distal and proximal regions and comprises 60 coding genes. The heterochromatic part of the Y chromosome consists of at least six sequence species in the centromeric region and at the distal and proximal ends [188].

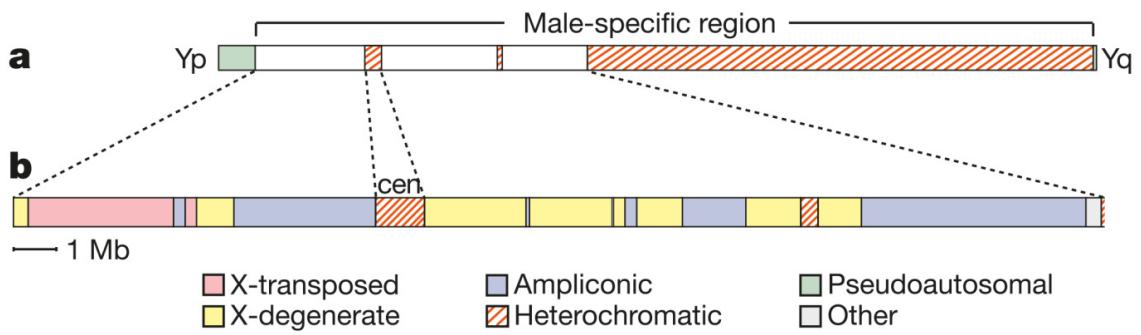


Figure 40: Distribution of euchromatin and heterochromatin in the MSY region of the Y chromosome. a) Representation of the whole Y chromosome including the heterochromatic, euchromatic, and pseudoautosomal regions. b) 23-Mb MSY region. Published in Skaletsky and colleagues [188].

The Y chromosome is like the mtDNA haploid, but is passed on to the next generation only from the father to his son(s) (Figure 13, chapter 3.3.2). Other characteristiccs of the Y chromosome that is shared with the mtDNA is the lack of recombination (except PAR1 and PAR2, see chapter 3.6.4), high mutation rate and the smaller effective population size. The smaller effective population size results in higher genetic drift, and thus, to a lower diversity in the Y chromosome and to a highly geographically informative marker [22].

3.6.2 Y chromosome polymorphisms

Polymorphisms that can be found on the Y chromosome comprise hundreds of microsatellites or STRs and Y-STRs, respectively, thousands of SNPs, SINEs (Alu repeats, short repeat sequences of high copy number), and LINEs (long repeat sequences) [22]. Mutation rate for the whole Y chromosome lies around 3 x 10^{-8} per bp per generation [30]. In this thesis Y-SNPs and Y-STRs are used for analysis, and thus, described in the following two sub-chapters.

3.6.2.1 Y-SNPs

Y-SNPs are Single Nucleotide Polymorphisms on the Y chromosome and are especially useful in genetic population studies [189], genealogical research [190, 191], or forensics [192]. Mutation rate of Y-SNPs is relatively low, and thus, are suitable markers to identify paternal lineages that can be traced back in time for thousands of years [193].

First attempts to standardize nomenclature of Y-SNPs were iniciated by The Y Chromosome Consortium (YCC) in 2002. The Y Chromosome Consortium described a single most parsimoniuos tree based on 245 mutational events which gave rise to 153 NRY haplogroups [193]. The phylogenetic tree of Y chromosomal lineages was later updated in 2003 by Jobling and Tyler-Smith (Figure 41) [194]. The main global distribution of Y-SNP haplogroups is shown in Figure 42 where the frequency of the 18 main haplogroups A to R among populations on different continents is shown (indicated by colour).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

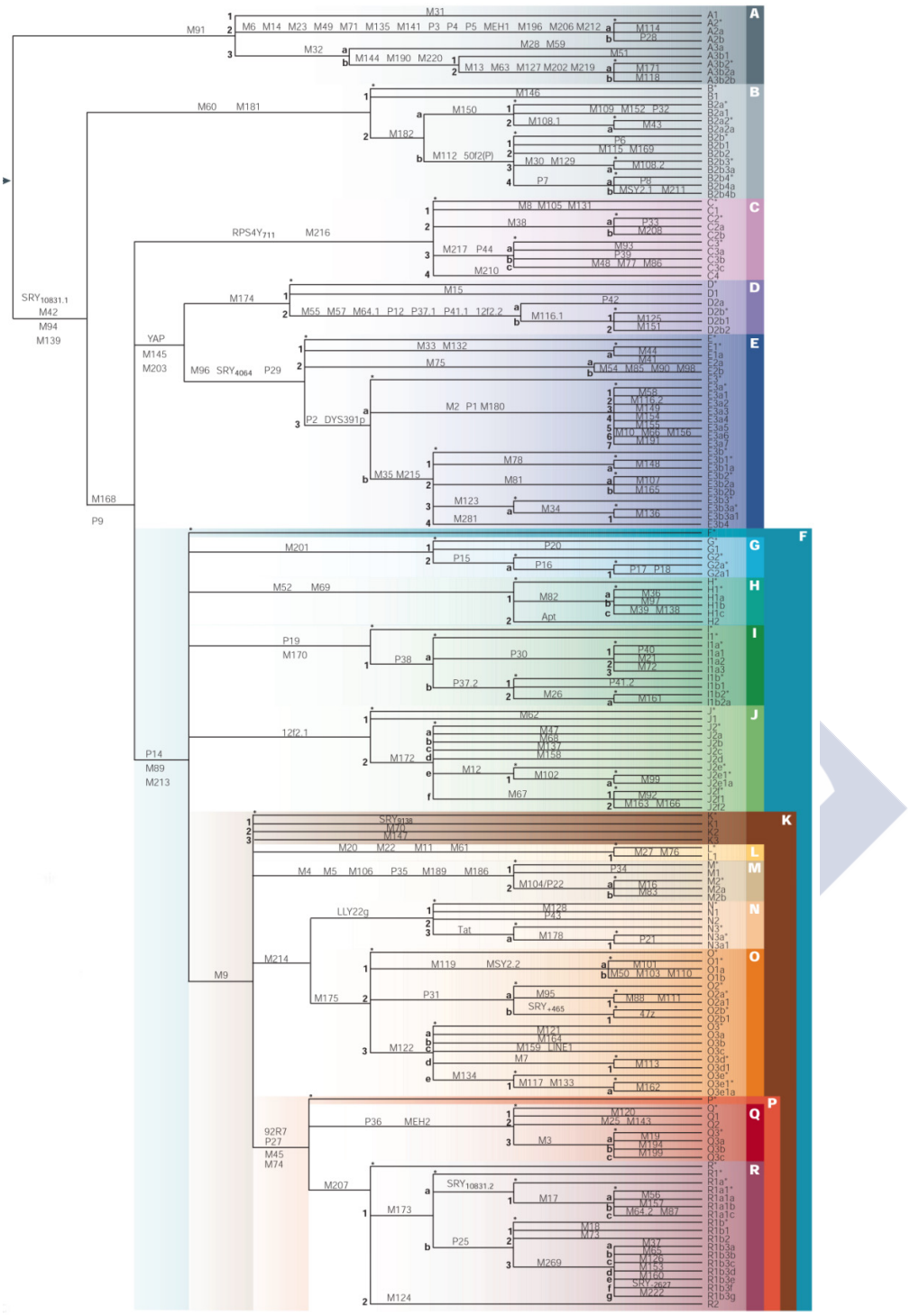


Figure 41: The phylogenetic tree of binary chromosomal haplogroups A-R (YCC2003 Tree). Clades and clade names are coloured according to global distribution (see Figure 42). An asterisk indicates a paragroup that comprises lineages that were not defined by 2003. The letter of a SNP indicates the research team and the number means the order in which the SNP was defined. Published in Jobling an dTyler Smith [194].

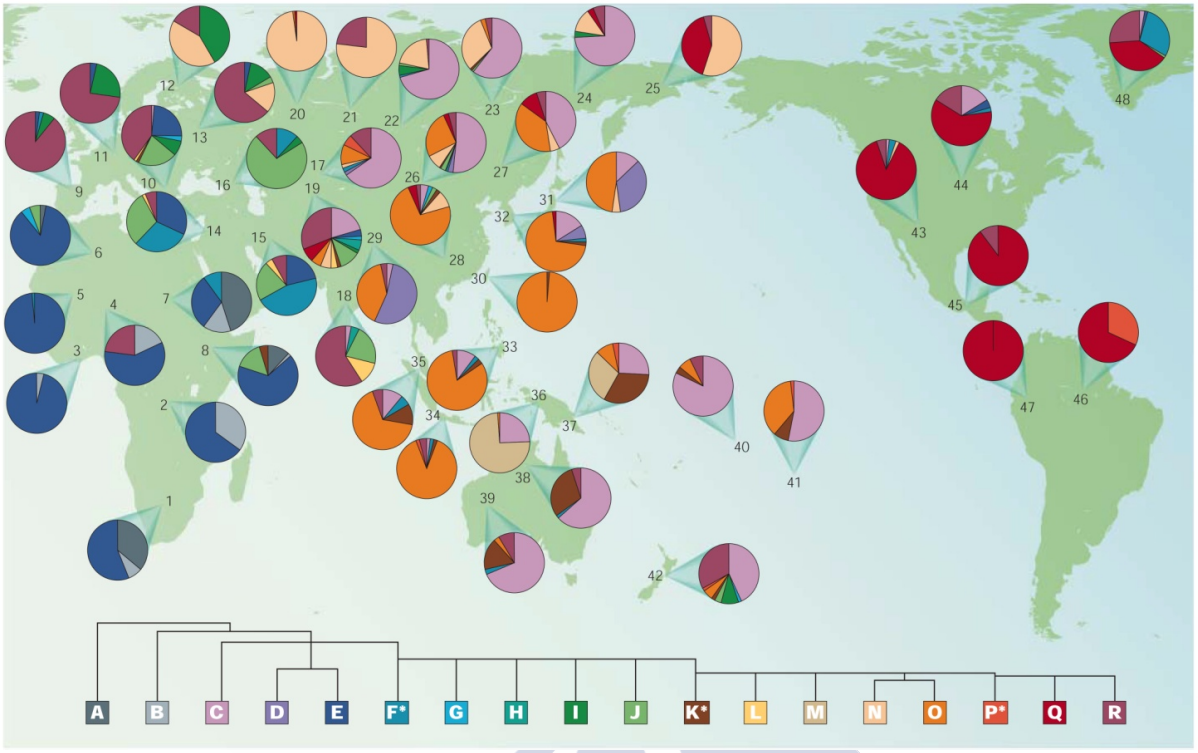


Figure 42: Global distribution of binary haplogroups. Published in Jobling and Tyler-Smith [194]. Populations are numbered as follows: 1, !Kung; 2, Biaka Pygmies;3, Bamileke; 4, Fali; 5, Senegalese; 6, Berbers; 7, Ethiopians; 8, Sudanese; 9, Basques; 10, Greeks; 11,Polish; 12, Saami; 13, Russians; 14, Lebanese; 15, Iranians; 16, Kazbegi (Georgia); 17, Kazaks; 18, Punjabis;19, Uzbeks; 20, Forest Nentsi; 21, Khants; 22, Eastern Evenks; 23, Buryats; 24, Evens; 25, Eskimos;26, Mongolians; 27, Evenks; 28, Northern Han7; 29, Tibetans; 30, Taiwanese; 31, Japanese; 32, Koreans;33, Filipinos; 34, Javanese; 35, Malaysians; 36, West New Guineans (highlands); 37, Papua New Guineans (coast);38, Australians (Arnhem); 39, Australians (Sandy Desert); 40, Cook Islanders; 41, Tahitians; 42, Maori; 43, Navajos;44, Cheyenne; 45, Mixtecs; 46, Makiritare; 47, Cayapa; 48, Greenland Inuit. Published in Jobling and Tyler-Smith [194].

The last update of the Y chromosome phylogenetic tree was done by Karafet and colleagues (Figure 43) [195]. The new revised Y chromosome tree is now described by 311 distinct haplogroups based on 600 binary markers. The new major haplogroups (S and T) were included to the haplogroups A-R in the new phylogenetic tree.

In this thesis the nomenclature described by Karafet and colleagues [195] will be used as this is the official nomenclature of the Y-STR Haplotype Reference Database (http://yhrd.org/) used in forensic genetics although the Y Chromosome Consortium (www.isogg.org/tree) updates annually the Y chromosome phylogenetic tree.

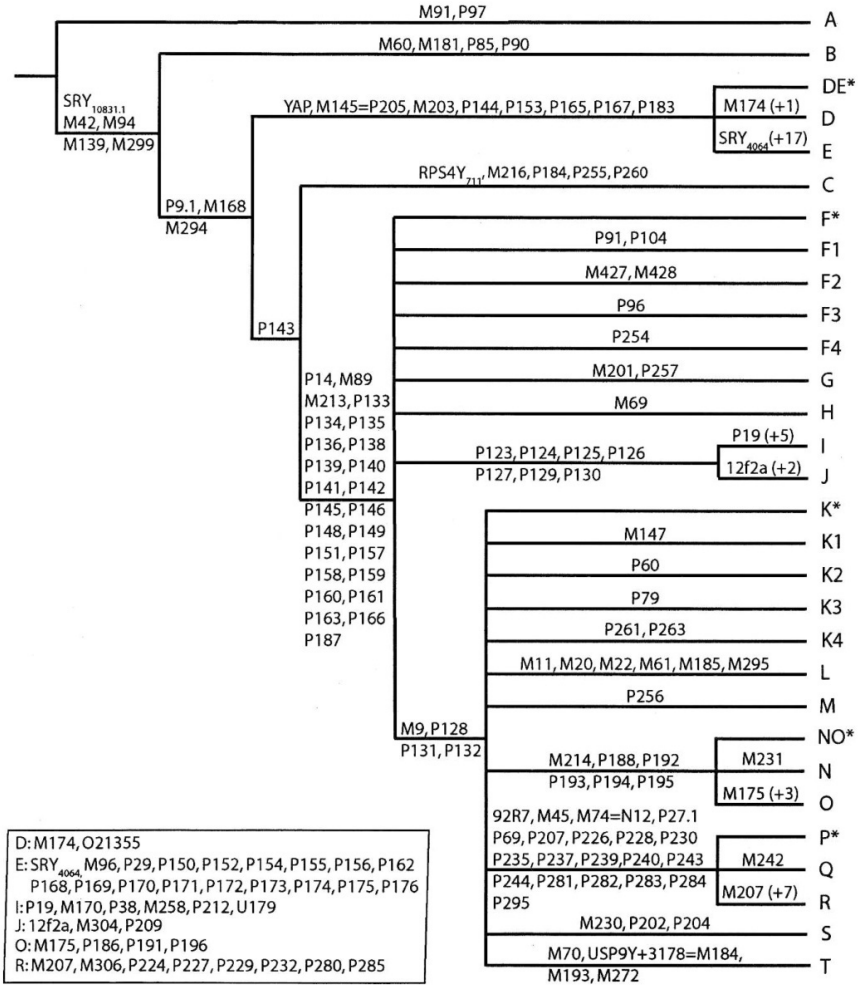


Figure 43: Updated summary of the Y chromosome phylogenetic tree. Mutation names are indicated on the branches. Haplogroup names are shown at the tips of the tree. Published in Karafet and colleagues [195].

As mentioned above the Y chromosome is a highly geographically informative marker and there exists a strong correlation between Y-SNPs and geographic distribution of the various haplogroups in different regions around the globe.

The haplogroups A and B are the two most ancient Y chromosomal lineages and are almost exclusively restricted to sub-Saharan Africa [196, 197]. Haplogroup A was observed in Khoisan speaking people and East African populations like Ethiopians and Sudaneses [198], and scatteredly as well on the Arabian Peninsula [199]. Similarly, haplogroup B appeared in Khoisan speaking people and in Ethiopia and Sudan [198], but also especially in pygmies [200].

Haplogroup C can be found in southern and eastern Asia (Figure 44) whereby branches of this lineage arose also in the Americas [201, 202], central Asia [203, 204], Europe [205], and Australia [206], but also in Polynesia and Melanesia [207]. Thus, this clade was found in populations like, for example, Japan [208, 209], China [210], Pakistan [211], and New Guinea [212].

Haplogroup D is primarly observed in Centraland Southeast Asia(Figure 44) like in China [213], but also in Japan [208]. The haplogroup is frequently seen in Tibet [214] and appeared as well in Mongolia [215].

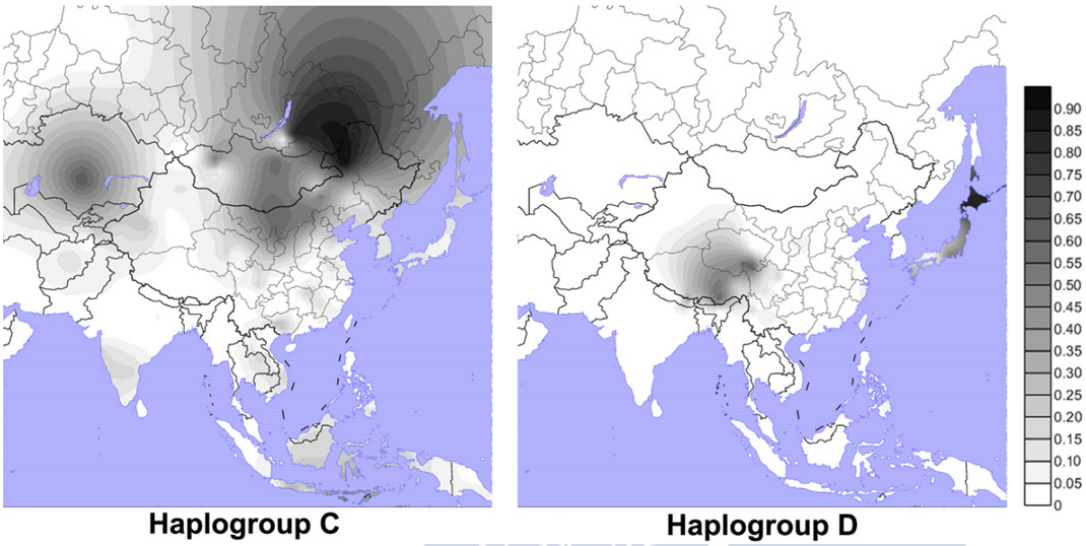


Figure 44: Distribution of haplogroup C and D. Published in [216].

Haplogroup E comprises a wide range of Y chromosome sub-haplogroups [217] and probably arose in the northern hemisphere of Africa due to the high concentration and diversity of E sub-haplogroups found in this area [215]. Haplogroup E1b1 is the most common Y chromosome haplogroup in Africa where it has its widest geographic distribution. Itcan be divided into twomajor sub-clades, that is, E1b1a and E1b1b. E1b1ais the most frequent haplogroup in sub-Saharan Africa and originated possibly in West and Central Africa where it has its highest concentration. Sub-lineages of E1b1a (former E1b1c) almost exclusively appear in East Africa. In contrast, E1b1b is especially common from southern Europe to northern and eastern Africa where it probably arose [218]. Another major sub-clade of haplogroup E is E2 which probably originated in Central Africa from where it spread in low frequency across Africa [215]. Due to the TAST sub-haplogroups of E can be frequently found in the Americas like, for example, on the Caribbean islands of Jamaica

(JAM), Haiti (HAT) and the Bahamas (Exuma/EXU, Long Island/LIS, New Provence/NWP, Eleuthera/ELU, Abaco/ABA, Grand Bahama/GRB) [219] (Figure 45) or the United States [220].

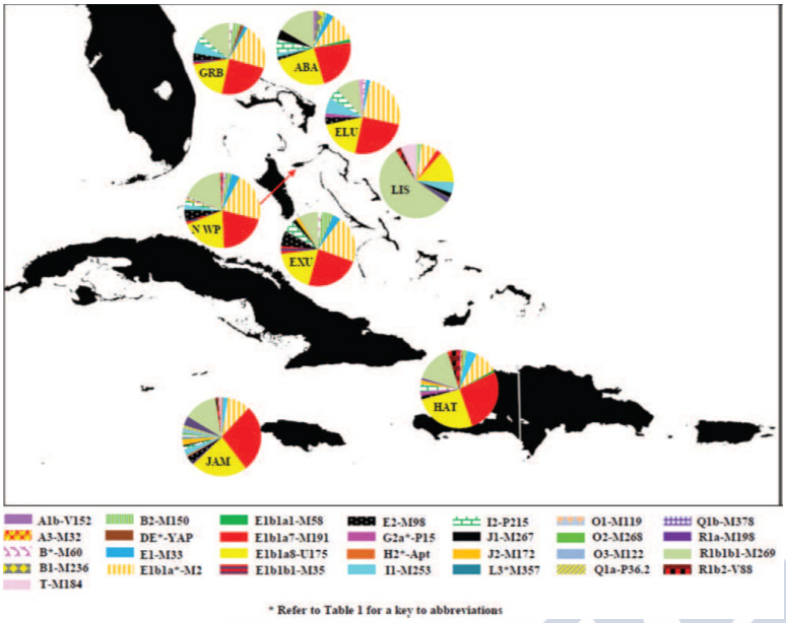


Figure 45: Y chromosome haplogroup distribution in Haiti, Jamaica and the Bahamas. Published in [219].

Haplogroup F is the parent haplogroup of all lineages from G through T whereby its major clades are represented by the haplogroups G, H, I, J, and K [215]. Haplogroup F appeared so far primarly in India [221].

Haplogroup G is the first branch of haplogroup F [222] and haplogroups like G2a2b2a1 and G2a2b1 are frequently seen in Europe and the eastern Mediterranean region, respectively. Haplogroup G2a2a is found in southwest and southern Asia [215].

Haplogroup H has not been studied comprehensively so far. However, members of haplogroup H were observed in India [223] or in some Balkan countries [205].

Haplogroup I is a European haplogroup where it has its main distribution (Figure 46A) and where it is representedin about one fifth of the European population. This haplogroup is highly infrequent outside of Europe. Its main sub-haplogroups are I1 and I2 [215]. About 95% sub-haplogroups belong either to I1a, I1b, or I1c and the remaining to the paragroup I*. Sub-clade I1a is widely distributed in Scandinavia whereas the distribution of I1b* range from the eastern Adriatic region to eastern Europe. I1c has a wide range in Europe, but peaks in Northern Europe [224].

A) B)

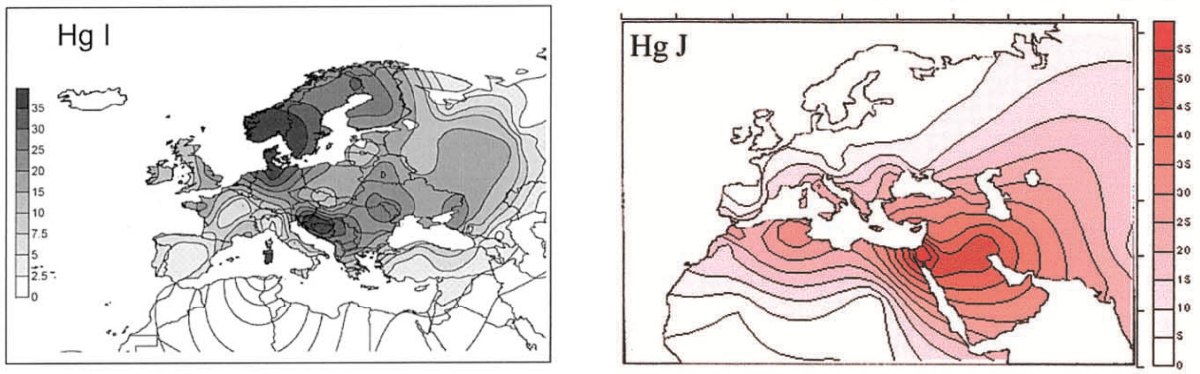


Figure 46: Distribution of haplogroup I and J. Published in [224] and [225].

Haplogroup J originated in the Near East where it is most frequently distributed (Figure 46B) and spread from there into North Africa, Europe, Central Asia, Pakistan and India [215]. Eight sub-lineages of haplogroup J were observed to be informative in a study about the peopling of southern and central-eastern Europe from the Levant [226]. The distribution of haplogroup J in Europe can be traced back in part to the Neolithic expansion [215].

Haplogroup K is the parental haplogroup for the haplogroups L through T [215]. It can be divided into four major sub-lineages, namely into K2a-d. K2b is the largest sub-lineage and further divided into K2b1 and K2b2. K2b2 comprises haplogroup P, and thus, its sub-clades Q and R which are widely distributed in Europe, Central Asia, and the Americas [227]. Sub-haplogroups of K itself were infrequently found in Africa, Eurasia, Australia, and in the South Pacific [215].

Haplogroup L is spread primarly across southern, central, and western Asia. The haplogroup is virtually absent in Europe, but infrequently observed in the Middle East and in the southern Caucasus [223]. The haplogroup was found in populations like Pakistan and India [228].

Haplogroup M peaks in Papua New Guinea [215]. On West New Guinea the majority of males, both in the lowlands and highlands, belonged to haplogroup M [229]. This haplogroup together with haplogroup K are probably markers of the expansion of Trans-New Guinea speakers [212].

Haplogroup N is widely spread throughout Northern Eurasia [215] (Figure 47). Its major sub-clades comprise N1, N2, and N3. The paragroup N* has a broad range but occurs only in low frequencies on the Fiji islands, in Borneo, Campodia, southern China, Japan, or southern Siberia. N3 is the most frequent haplogroup of the three major sub-clades and belongs to the most common haplogroup in northern Eurasia. Haplogroup N2 is most frequent in north-western populations of Siberia and N1 occurs in low frequency in some central Asian populations and Korea [230].

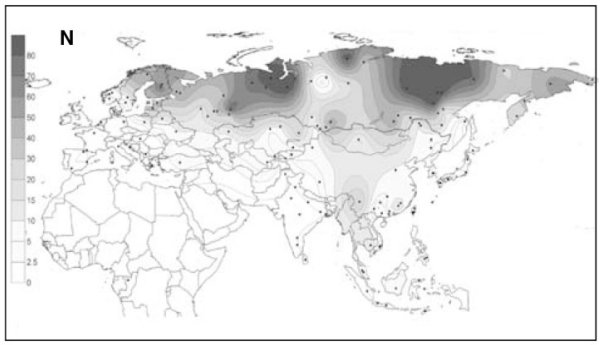


Figure 47: Distribution of haplogroup N. Published in [230].

Haplogroup O is an East Asian-specific haplogroup (Figure 48A and 48B). Three major sub-lineages, that is, O1, O2, and O3, form this haplogroup. About 60% of all Y chromosome haplogroups in East Asian populations belong to one of these sub-branches. Thus, for example, three quarter of the Chinese population and more than half of the Japanese population belong to haplogroup O. O3 is the most predominant haplogroup. O1 is common along the southeast coast of China and in Taiwan aborigines. O2 was observed in Han Chinese [216].

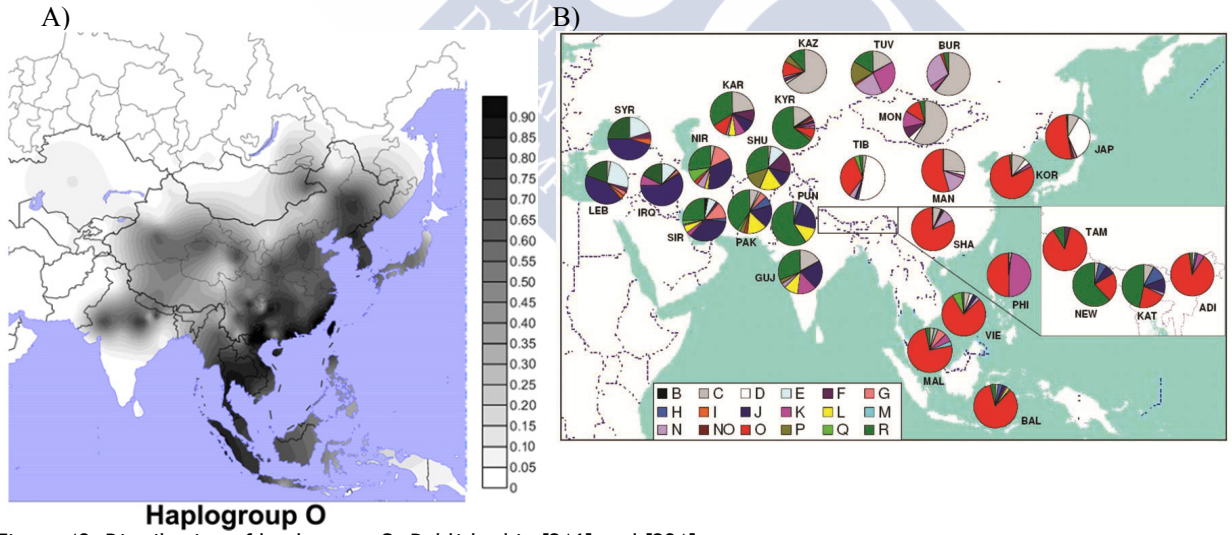


Figure 48: Distribution of haplogroup O. Published in [216] and [231]

Haplogroup P is mostly represented by the two sub-clades Q and R which are the predominant haplogroups in the Americas and in Europe, respectively [215]. In North America haplogroup Q seems to be most frequent in the Southwest, but also in Mesoamerica [202] (Figure 49). In South America Q was the predominant haplogroup, for example, in the Ecuadorian [232] or the Bolivian [233] population whereby most Native American males in South America belong to the Q1a3a sub-

haplogroup [195, 234]. Sub-lineages of Q1a3a are informative haplogroups to study population history like it was done in an Andean population where a new sub-lineages of Q1a3a was detected [235].

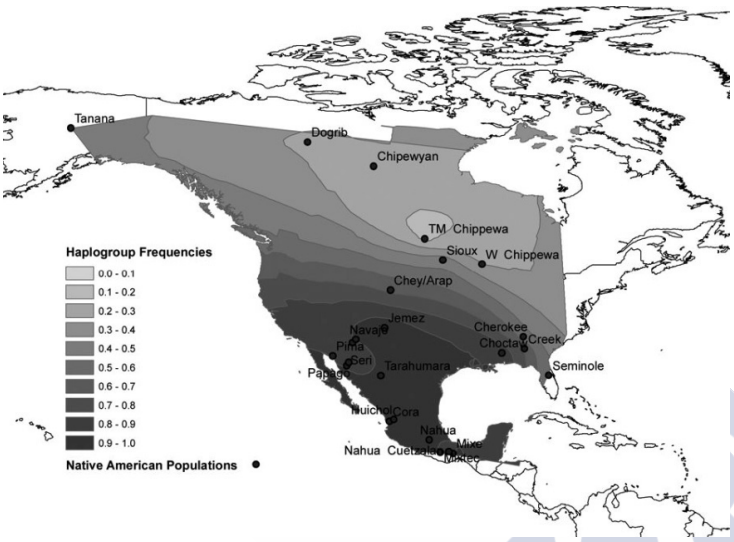


Figure 49: Distribution of haplogroup Q in North America and Mesoamerica. Published in [202].

Haplogroup R is the predominant haplogroup observed in Europe and divided into R1 and R2 whereby most individuals of haplogroup R belong to R1 sub-lineages. The main sub-lineages of R1 are R1a and R1b [215]. R1a is a widely distributed haplogroup. Its spatial distribution is apparently divided into two major groups, that is, one group primarly found in Europe and another one principally spread across Central and South Asia. R1a probably originated in present-day Iran [236]. R1b probably arose in West Asia, but in Europe its frequency increase to the Northwest of Europe [237]. Due to the TAST haplogroups of R are found in the Americas like, for example, in Brazil [85].

The haplogroups S and T were, as mentioned above, described by Karafet and colleagues [195]. Haplogroup S is mainly distributed in the highlands of Papua New Guinea, but also found in Indonesia and Melanesia. Haplogroup T appears at low frequencies throughout Europe and sparsely in the Middle East, North and East Africa [215].

3.6.2.2 Y-STRs

STRs (Short Tandem Repeats), which are also known as microsatellites, are repeating sequences consisting of usually 1-7 bp units. STRs are made up of sequences (e.g., CA, TATT) that are repeated a variable number of times. Therefore, STRs are also known as Variable Number of Tandem Repeats (VNTR). The most suitable STRs have a copy number of about 10-30. As the number of repeat units between individuals can vary, STRs are suitable markers to distinguish between individuals. Each individual has two alleles at one loci whereby homozygotic individuals have two alleles of the same length and heterozygotes alleles of different lengths. While the number of repeats is variable between individuals, the flanking regions where PCR primers bind are constant (Figure 50). Microsatellite mutations can be attributed to replication slippage, that is, a misalignment between the newly synthesized DNA strand and the template strand during replication which leads to a gain or loss of a repeat unit [22].

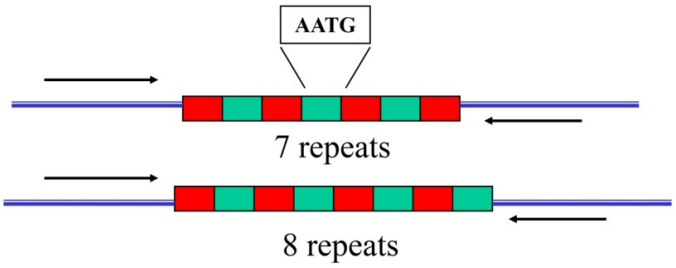


Figure 50: Short Tandem Repeats (STRs). Published on the Short Tandem Repeat DNA Internet Data Base (http://www.cstl.nist.gov/biotech/strbase/index.htm).

STRs that are only found on the Y chromosome in males are known as Y-STRs. Y-STRsare informative markers used in population and evolutionary genetics studies [238], genealogical research [239], or forensic investigations [240, 241]. The number of Y-STRs that are described in the literature and available for research is continuously growing. For example, in 2004 Kayser and colleagues [242] described 166 new Y-STRs, whereby 139 out of them are polymorphic. Due to the increasing number of Y chromosomal markers, including Y-STRs, internet-accessable databases continue to describe new Y-STR profiles [243].

The widely accepted nomenclature for Y-STRs is provided by NIST (National Institute of Standards and Technology) STRBase (http://www.cstl.nist.gov/biotech/strbase/). However, the DNA Commission of the International Society of Forensic Genetics (ISFG) regularly publishes

guidelines and recommendations about the use of DNA polymorphisms, including Y-STRs in forensics [244].

The high polymorphic character of Y-STRs can be attributed at least in part to the fact that mutations at Y-STR loci are relatively frequent which are highly appreciated in microevolutionary studies, but not in forensics [245]. In an extensive collaborative work in 2005, Gusmão and colleagues [245] analyzed 17 Y-STR markers in 3,026 father/son pairs to find an overall mutation rate of 1.998×10^{-3}across all 17 loci. A similar value, that is, 2.80×10^{-3} for 15 Y-STRs was found by Kayser and colleagues [246]. In a more recent publication, Ballantyne and colleagues [247] found an average mutation value of 3.35×10^{-3} for 186 Y-STRs.

The first Y-STR marker (DYS19) was described in 1992 by Roewer and colleagues [248]. However, as the number of Y-STRs is continuously growing since then, the scientific community was interested in finding an agreement on the Y-STR markers used to compare individuals. Thus, the European forensic community decided in 1997 to use a core set of Y-STRs, known as "minimal haplotype", to identify individuals [243]. This "minimum haplotype" included the Y-STRs DYS19, DYS389I/II, DYS390, DYS391,DYS392, DYS393, and DYS385 a/b. Seven haplotypes are unilocal except DYS385a/b which detects variation at two loci simultaneously [249]. As the flanking regions of DYS385a and DYS385b have identical sequences,one primer pair amplifies the two loci simultaneously. Thus, bothloci cannot be unambiguously distinguished by using a convential PCR [244]. By means of a nested PCR approach, however, both DYS385a and DYS385b can be amplified separately [250] (Figure 51).

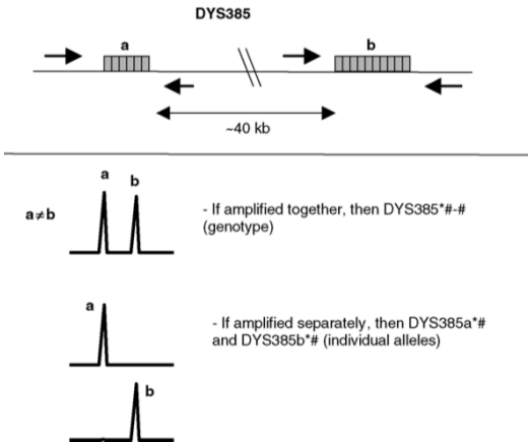


Figure 51: Illustration of amplification of the Y-STR marker DYS385. Published in [244].

The positions of the “minimum haplotype” Y-STRs on the Y chromosome are shown in Figure 52. In some studies the “minimum haplotype” set consists of 7 Y-STRs excluding DYS385a/b like it was done in [251].

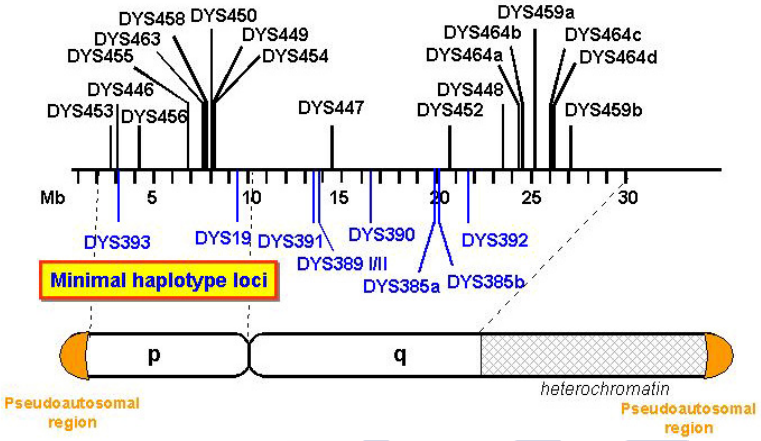


Figure 52: Y STR Positions along the Y chromosome.Published on the Short Tandem Repeat DNA Internet Data Base (http://www.cstl.nist.gov/biotech/strbase/ystrpos1.htm).

These 9 Y-STRs (if the DYS385 is counted separately) have been used to generate more than 135.000 9-locus minmal profiles on the Y-chromosomal Short Tandem Repeat Haplotype Reference Database YHRD (http://yhrd.org/, last update on Nov 10, 2014) represented by 917 populations in 127 countries whereas the largest databases (>7,000 haplotypes) include China, USA, Germany, Brazil, and Poland. The YHRD website pools Y-STR haplotype data for the forensic community [249, 252].

However, due to continous research new multiplex PCRs and commercial kits that include Y-STRs beyond the “minimum haplotype” was conducted. A multiplex PCR allows the simultaneous amplification of more than one Y-STR [243]. Commercial kits like, for example, the AmpFlSTR®YFiler™ kit from Applied Biosystems or the PowerPlex® Y from Promega Corporation allow the amplification of 17, or 12, respectively, Y-STRs in one PCR. In this thesis the commercial kit AmpFlSTR®YFiler™ was used in addition to the “minimum haplotype”.

The AmpFlSTR®YFiler™ kit comprises, as mentioned above, 17 Y-STRs including DYS19, DYS389I, DYS389II, DYS390, DYS391,DYS392, DYS393, and DYS385 a/b from the “minimum haplotype”, but alsoDYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYS635, and GATA H4. PCR multiplex amplifications with the AmpFlSTR®YFilerTM kit can be conducted

with low levels of male DNA which is important in criminal investigations [253] and allows a four-colour detection of the 17 Y-STRs (Figure 53).

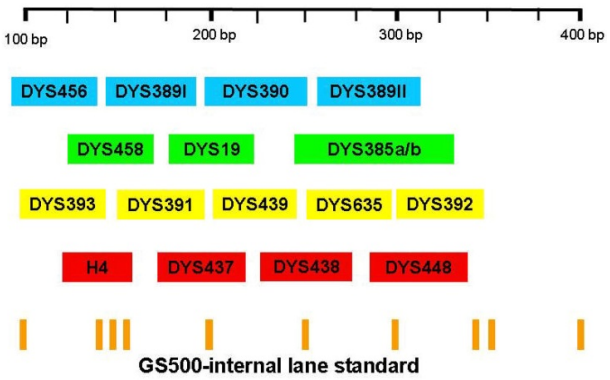


Figure 53: Schematic illustration of the Y-STRs from the AmpFlSTR® YFiler™ kit. Shown are the fluorescent dye label colors and the relative PCR product size. Published on the Short Tandem Repeat DNA Internet Data Base (http://www.cstl.nist.gov/biotech/strbase/kits/Yfiler.htm).

According to the latest version of the YHRD database (Nov 10, 2014) more than 96.000 Y-STR haplotypes for the PowerPlex Y and more than 84.000 for the YFiler were described.

3.6.3 Paternal inheritance

The Y chromosome is like the mtDNA subject to uniparental inheritance where most parts of the Y chromosome does not recombine (see below). This means that a male shares the same Y chromosome with his father, paternal grandfather, paternal great-grandfather and so on.

Mutations in the Y chromosome that occur in the germ line can be passed on to the male next generation, and thus, leading to genetic differentiation. Genetic differentiation for the Y chromosome is in general greater than for the mtDNA. The reason might be based on cultural practices. That means, in a patrilocal society in which males remain at their birthplaces and females leave their home, Y chromosome differentiation is greater than in societies in which matrilocality is widespread. Matrilocality is the opposite of patrilocality in which females stay at their birthplace and males migrate [22]. This phenomenon was observed in three matrilocal and three patrilocal groups in northern Thailand supporting the role of sex-specific migrations in shaping genetic variation and the fact that migration decreases population differentiation [254]. Patrilocality

occursin approximately 70% of the human population [255] which might explain why genetic differentiation of the Y chromosome in general is greater than for the mtDNA [22].

3.6.4 Lack of recombination

The Y chromosome in general lacks recombination or in most of its length (>90%), respectively, and is known as the nonrecombining portion of the Y (NRPY or NRY) or the male-specific region of the Y chromosome (MSY). However, a small region at both tips of the Y chromosome undergoes recombination with the X chromosome during male meiosis. These small fractions of the Y chromosome show sequence similarity with the X chromosome and are known as pseudoautosomal becasuse DNA from this part can be inherited from both mother and father. The pseudoautosomal region 1 (PAR1) is located at the end of the short arm of the Y chromosome and of approximately 2.6 Mb in size. At the long arm of the Y chromosome the pseudoautosomal region (PAR2) make up about 0.32 Mb. The PAR2 region evolved recently and does not play an important role in the segration of chromosomes whereas PAR1 is ancient dating back about 300 MYA and is target of recombination in any male meiosis [22].

Because of these features the X chromosome recombines during male meiosis only at the pseudoautosomal regions, but during female meiosis along the entire chromosome. Later the dose of X-chromosomal genes is balanced via X-inactivation. X-linked diseases in females can be masked by the second X chromosome while in males no second copy can compensate the disease effect of a mutation [22].

3.6.5 Human population history

The male Y chromosome is like the mtDNA a useful marker to understand population origin and past human migration events as this uniparental marker is passed on to the next generation only by males (see chapter 3.3.2 and 3.6.3). At high molecular resolution, the Y chromosome can provide comprehensive information about the migration out-of-Africa and further population differentiation on other continents [256].

For example, in a population of North Indians, including three higher caste groups (Brahmins, Chaturvedis, and Bhargavas) and two Muslim groups (Shia and Sunni), it was shown that a substantial part of the paternal gene pool can be traced back to Indo-European speakers of Central

Asia (or west Eurasia) as the lineages R1a1* and R2 were most represented in North Indians. Thus, it was suggested that Y chromosome lineages in Northern India are most probably of Central Asian origin. However, some Y chromosome lineages (J2*, G*, L*) of North Indians can be traced back in part as well to the Middle East whereby one lineange (E1b1b1) indicates an African/Middle Eastern paternal ancestry in Shia Muslims as this haplogroup was observed only in this population. The third Y chromosome lineages observed in the five groups are of indigenous Indian origin (H1*, F*, O*, C*). The information of the three paternal Y-lineages observed in North Indian populations can be used to infer migration patterns into India (Figure 54).

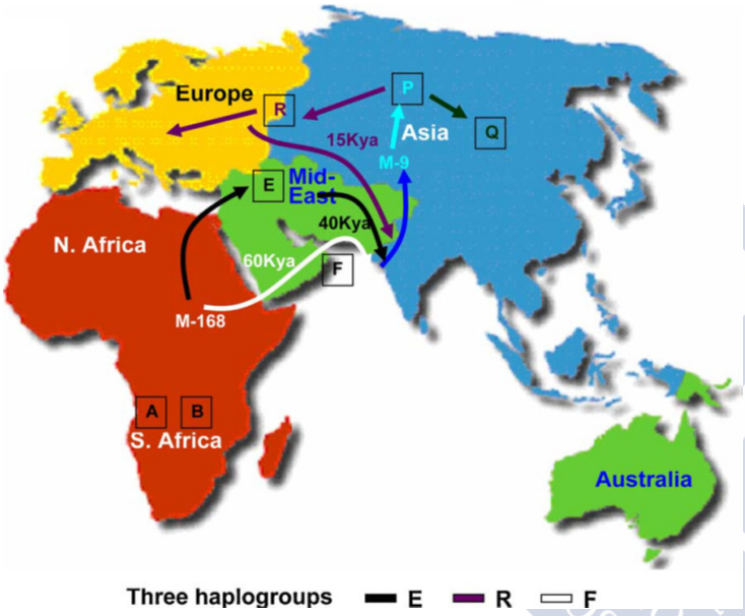


Figure 54: Migration routes based on paternal lineages in North Indians. Published in [221].

Similar research was conducted in East Asian populations. In a study in East Asian populations it was shown that the four most common haplogroups (C, D, O, N) are most probably of Southeast Asian origin supporting migrations of humans from Africa to East Asia via Southeast Asia [216].

The Y chromosome has been as well a popular marker to investigate the peopling of the Americas as it was done, for example, in extant Native American populations like Na Dené and South Amerinds [201]. Based on these populations, Bortolini and colleagues [201] support two migration events from southern/central Siberia to the Americas with the second migration wave being restricted to North America. After the initial peopleing of the Americas, however, population expansions and migrations further shaped the genetic pool of the first Native

Americans.Consequently, in Eskimoan and Athapaskan speaking people of Canada, for example, it was observed that both populations are distinct from each other which can be attributed to population expansions that occurred early after the first arrivals of humans in the Americas [257]. However, as the arrival of Europeans in the Americas has significantly changed the paternal gene pool in North America, conclusions about prehistoric population migrations should be treated with caution when using only the Y chromosome [202].

3.7 Autosomal DNA

3.7.1 Genome-wide SNP genotyping

By genotyping of SNPs in the whole genome genetic variation between individuals within and between populations can be measured. Genome-wide SNP genotyping has been developed to a popular method in various fields in the last few years due to new, more sophisticated, and cheaper technology giving rise to Next Generation Sequencing (NGS). NGS has promoted especially the usage of SNPs, for example, in ancestry testing, including direct-to-consumer genomics tests [258], but also gave rise to large-scale genomics project like the 1000 Genomes project [148]. Genome-wide SNP genotyping is especially popular in case-control studies of diseases where ancestry estimates are important information as population stratification can lead to false positives [259, 260]. Human population studies [261] and forensic genetics [262] represent more research fields in which SNP genotyping was used in the past.

A great amount of SNPs can be analyzed with DNA microarrays or DNA chips, respectively. The Affymetrix Human SNP Array 5.0 chip can genotype more than 500,000 SNPs and was used, for example, in a study about schizophrenia [263]. The Ifinium HumanCore BeadChip from Ilumina is able to genotype more than 250,000 SNPs and was applied in a study about panic and social anxiety disorders [264]. Manufacturers of DNA microarrays comprise besides companies like Affymetrix (www.affymetrix.com) and Illumina (www.illumina.com) also GE Healthcare Life Sciences (www.gelifesciences.com), or Applied Biosystems (www.lifetechnologies.com).

In the present thesis the Axiom® Genome-Wide Human Origins 1 Arrayfrom Affymetrix was used to estimate genome ancestry. The array comprises genomic markers (SNPs) of 11 different modern human populations, but also from archaic hominids (Neanderthals, Denisovans). About 629,000 SNPs can be typed with this array. Affymetrix also developed specific arrays for target populations

like, for example, for East Asians including JPT and CHB (Axiom® Genome-Wide ASI 1 Array Plate) or for Africans including West and East African populations (Axiom® Genome-Wide PanAFR Array Plate Set). The former array was used in a Genome-wide association study (GWAS) in the Japanese population to study susceptibility loci for Primary Biliary Cirrhosis [265]. Risk loci (SNPs) are commonly represented in Manhattan plots as it was done in the Japanese population (Figure 55).

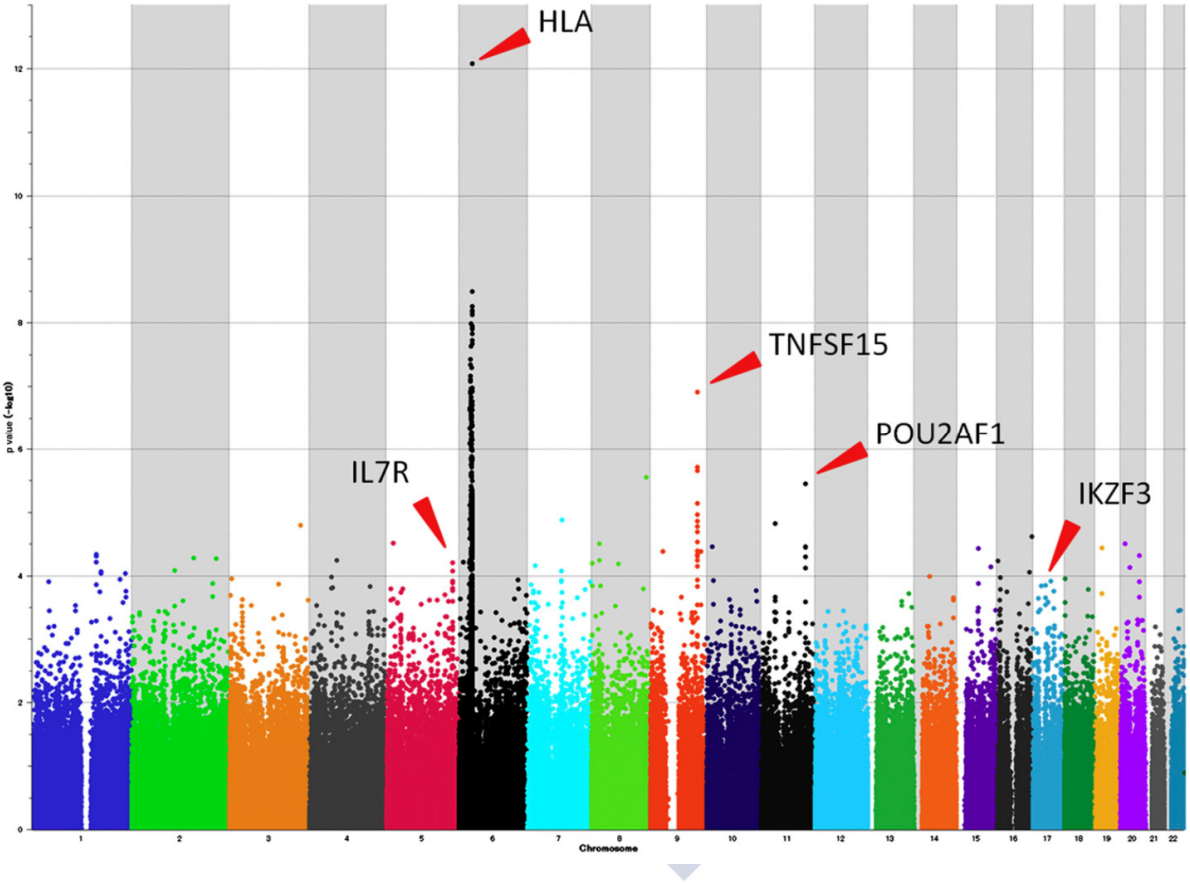


Figure 55: GWAS study of Primary Biliary Cirrhosis in a Japanese population. Susceptibilty loci are indicated with a red arrow. Published in Nakamura and colleagues [265].

In the study about susceptibility loci for Primary Biliary Cirrhosis an East Asian-specific gene chip from Affymetrix was used. As mentioned above, an association of a risk gene to a disease might be counfounded because of population stratification, or more precisely, because of allele frequency differences between populations due to different ancestry. SNPs that show different allele frequencies in different populations are named Ancestry Informative Markers (AIMs). Pardo-Seco

and colleagues [266] found that the ability to estimate genome ancestry depends to a great extent on the number of SNPs being used whereas the SNP informativeness plays not the primarily role.

The efficiency of a SNP array can be evaluated with Principal Component Analysis (PCA) and/or admixture analyses. For example, Novembre and colleagues [35] were able to detect with the Affymetrix GeneChip Human Mapping 500K Array Set genetic differentiation corresponding to geographic distance (Figure 56).

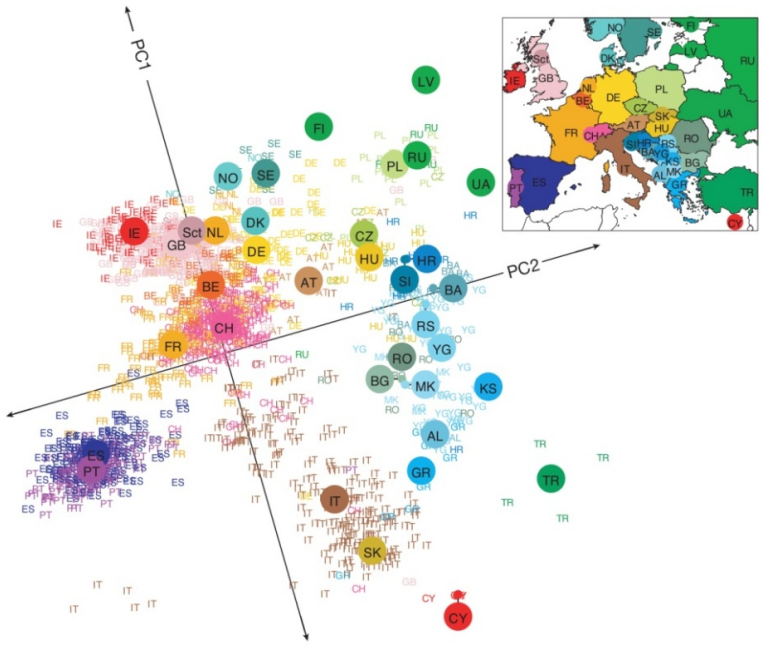


Figure 56: Principal Component Analysis of the Europen population. Published in Novembre and colleagues [35].

Despite fast advances with new technology there are still limitations in the use of whole-genome SNP genotyping for genetic ancestry testing, but also for GWAS studies. For example, it is not possible with a genetic ancestry test for an African American to trace back one's ancestry to an African village that existed during the time of the TAST [267]. With respect to GWAS studies, NGS has changed medical research in a very important way, but GWAS studies are still in itsinfancy. Janssens and colleagues [268] found out that the scientific evidence for most gene-disease associations are still insufficient although they attribute potential to this research field as genetic information could be effectively applied in the future in disease prevention.

In a recent study by Moreno-Estrada and colleagues [39] whole genome SNP genotyping was used to investigate the TAST. The researchers found two major migration pulses from Africa into the

Caribbean, the first one beginning around 1550 and the second one at the peak of the TAST towards the end of the 18th century. By means of genomic segment length analysis the authors observed that the first migration pulse came from the Senegambia region and the second pulse from West-Central Africa. The separate geographic origin of the two migration pulses to the Caribbean during the TAST is based on the fact that chromosomes that were contributed earlier in American populations were subjected to more recombination events, and thus, leading to shorter African ancestry fragments. Longer African ancestry fragments, in contrast, underwent less recombination as they arrived later [39].

3.7.2 Insertion-deletion polymorphisms (Indels)

Insertion-deletion polymorphisms are length polymorphisms that can be straight-forwardly genotyped by fragment size differentiation [269]. Indels are Ancestry Informative Markers (AIMs). AIMs are suitable tools to investigate population differentiations as they show high allele frequency divergence between populations of different ancestry or geographic distance (see above Figure 56). Most available AIM panels consist ofSNPs including, for example, the 128 In4 (four population informativeness) AIM-SNP set developed by Kosoy and colleagues [49], but also the SNP gene chips described in the previous chapter. Microsatellites and indels (Insertion and Deletions) likewise have proven to be suitable markers to distinguish between populations of different ancestry or geographic origin, respectively. For example, CoAim, which is based on 36 microsatellites, was shown to efficiently apportion continental admixture and discern European/Middle Eastern, African, East Asian, Native American, and Oceanean populations [270].

However, Indels appear to be especially useful as AIMs because of their ease-of-use and time and cost effectiveness which minimizes errors. Indels can be genotyped in one multiplex PCR and subsequently separated by a Capillary Electrophoresis to analyse each sample in one electropherogram (Figure 57) like it was done by Pereira and colleagues [269]. As it was shown by the researchers these indels are able to effectively differentiate continental ancestry (Europe, Africa, America, East Asia)[269]. Similar was done by Santos and colleagues [271] who developed a 48 Insertion-Deletion (INSEL) Ancestry Informative Marker (AIM) panel which effectively estimates sub-Saharan African, European, and Native American ancestry proportions like, for example, in admixed Brazilians [271].

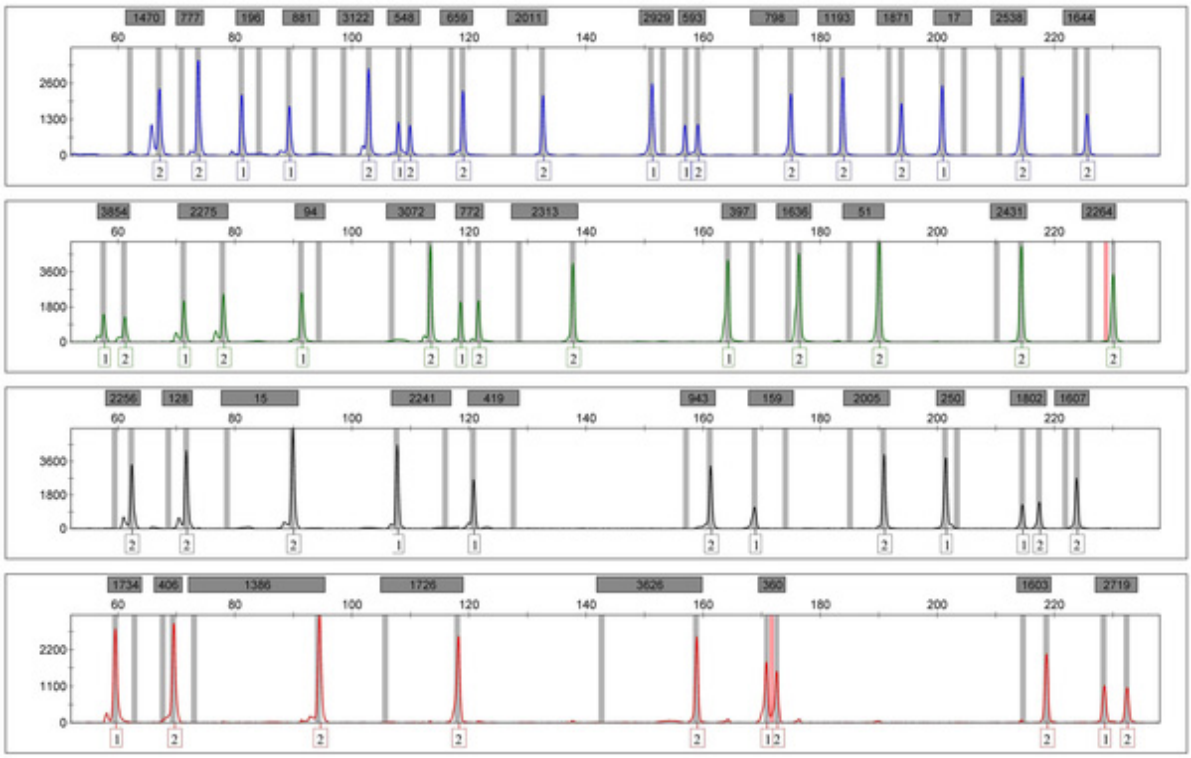


Figure 57: Electropherogram for the sample HGDP-CEPH 0452 with the 46 AIM-INDEL panel developed by Pereira and colleagues [269]. Indels are highlighted in grey, MID number is not shown. Published in [269].

As mentioned above, in practice AIMs are of great importance in Genome-Wide Association Studies (GWAS) because population stratification can lead to false positive variant-disease associations, and thus, a disease which was found to be associated with a variant might be the result of its ancestral origin and not be in general the disease-causing variant.

Individual ancestry based on indels are frequently estimated with the software STRUCTURE [272] or ADMIXTURE [273]. Moreno-Estrado and colleagues [39], for example, used the latter software to estimate individual ancestry of Latin American populations (Cuba, Haiti, Dominican, Puerto Rico, Colombia, Honduras, Mexico) (Figure 58). The reasearchers used three continental reference populations, that is, populations from Africa, Europe, and Native America. Assuming K=3 (number of K population),all Latin American population exhibited European, Native American, and African admixture whereby the range of European and African ancestry seems to be wide (especially in Cuba) and Native American ancestry was most prevalent in Mexico. However, at a value of K=8, Moreno-Estrada and colleagues [39] could detect an European-specific component (black) in the Latin American populations which were only observed in Europe in Mediterranean populations.

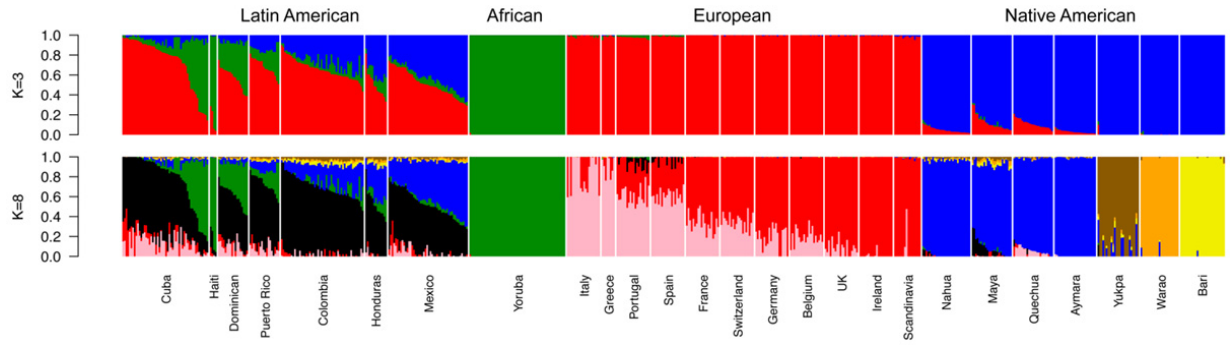


Figure 58: Admixture in latin American populations. Published in [39].

3.8 Study Population

3.8.1 Bolivia

Subject of the present thesis it the Bolivian population. Bolivia is located in the center of South America bordered by Brazil in the North and Northeast, Paraguay and Argentina in the South, Chile in the Southwest, and Peru in the Northwest and therefore does not have immediate access to the sea. It has a total area of 1,083,581 square kilometers and represents the 28th largest country in the world [274]. The main capital of Bolivia is Sucre in the department of Chuquisaca, but the government resides in the city of La Paz within the same-named department. Two mountain ranges cross Bolivia from North to South dividing the country into three climatically different regions: Andeans, Sub-Andeans, and Llanos (Figure 59). Bolivia can be further divided into nine departments, that is, Pando, Beni, and Santa Cruz in the Llanos, Cochabamba, Chuquisaca, and Tarija, in the Sub-Andeans, and La Paz, Oruro, and Potosí in the Andeans. The country is inhabited by about 10.6 million people (July 2014) who principally speak Castellano [274]. However, all 36 indigenous languages like, e.g., Aymara, Quechua, and Guaraní are like Castellano as well official languages within the plurinational state of Bolivia (Instituto Nacional de Estadística de Bolivia, INE, http://www.ine.gob.bo/).

Figure 59: Map of Bolivia.

3.8.2 History of the TAST in Bolivia

During the colonial period Europeans not only established themselves in the Americas, they also brought in millions of African enslaved people, principally to the Caribbean and Brazil, but also to Spanish America, including Bolivia. The onset of the TAST in Bolivia can be traced back to the discovery of rich silver deposits in the mountainous Andean region around nowadays Potosí. First African enslaved people in Potosí were documented beginning in 1549 [11] where they principally had to work in the royal mint (Casa de la Moneda), which operated until 1773 in Potosí, but also as domestic servants [62]. African enslaved people probably did not work in the surrounding silver mines, which was principally done by indigenous workers [11] (as shown in Figure 60 in Peru) although the contrary was reported as well elsewhere [275]. However, the royal mint contributed to a great extent to the economic boom of the Spanish colony, but African enslaved people, who were working and living under very harsh conditions and excluded from the increasing economic advancement, suffered high mortality rates [11].

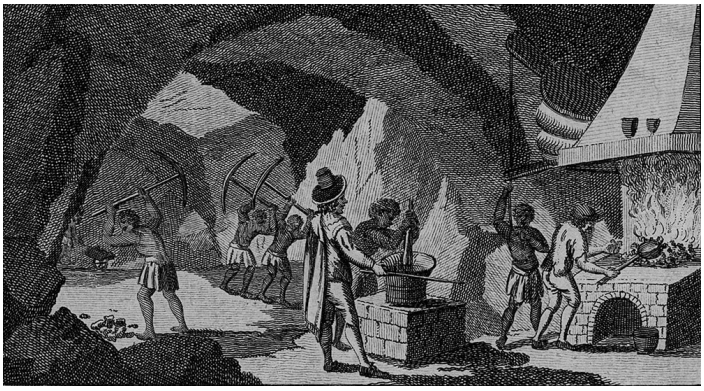


Figure 60: Working in silver mines, Peru, 1789, Image Reference: JCB_35592-6, as shown on www.slaveryimages.org, compiled by Jerome Handler and Michael Tuite, and sponsored by the Virginia Foundation for the Humanities and the University of Virginia Library.

In the course of the 18th and 19th century descendants of the first enslaved Africans in Potosí were migrating towards the North from the highland mining regions to the tropical Yunga valleys. Here, the Spanish colonists built up a so called "hacienda system" which was maintained by forced slave labor and commonly ruled by absentee estate holders. Enslaved people, 'African descendants' together with local Native Americans, was mainly working on coca, but sometimes also on coffee and sugar plantations [62, 276].

After the transition from colony to republic in 1825 and the first Bolivian constitution, new slavery was abolished and enslaved people could purchase their freedom for a high price. However, protests by estate holders hindered complete abolition of slavery, and thus, situation for 'African descendant' people in the Yunga valleys of being forced to work did not change immediately after the colonial period. Only in 1952 after the agrarian reform the "hacienda system", and thus, slavery was abolished in Bolivia [11, 62, 276]. Hence, African descendant people were free from "hacienda labor". Nonetheless, the majority of them remained in the Yungas valleys where they occupied former estates. Therefore, African descendant communitiescan be found in the Yungas, that is, in the provinces of the Nor (North) and the Sud (South) Yungas in the department of La Paz. The Bolivian Yungas region is a tropical and subtropical moist forest eco-region with poor accessibility from main cities. Here, 'Afro-Bolivian' communities remained mostly isolated after the agrarian reform. Today, there are several 'Afro-Bolivian' communities in the Nor Yungas (i.e. Tocaña, Mururata, Chijchipa, Dorado Chico, Coscoma, and Khala Khala) and in the Sur Yungas (Chicaloma) [62].

In the last twenty years 'Afro-descendant' people have increasingly migrated out from the Yungas and the category 'Afro-Bolivian'was included in the national census for the first time in 2007 [277]. Not only in Bolivia itself [278] but also within scholars of the TAST, 'Afro-Bolivian' people have been neglected or not mentioned at all in the first instance, possibly due to their low number when compared to other American countries and regions with a more evident African genetic heritage (e.g. Caribbean, Brazil, Colombia) [57, 81, 90, 279].

Information about the African geographic origin of 'Afro-Bolivians' is so far mainly based on assumptions. For example, a Congolese and Angolan origin can be supported by the presence of the African surnames such as Angola or Maconde, which apparently come from Congo [276]. Furthermore, when the Portuguese and Spanish crown were merged during 1580 and 1640 and many slaves were sold to the Portuguese during the wars in the hinterland of Luanda, Angola, many African people were forced to migrate to the Americas not only to Brazil, but also to Spanish America [2]. However, ancestors of nowadays 'Afrobolivian' communities apparently not have originated exclusively from West-Central Africa. According to historical documents most African enslaved people (about 55%) between 1560 and 1650 in Peru came from Senegambia that entered South America via ports like Cartagena de Indias in Colombia, Mexico or the Caribbean [4]. Some of the enslaved people that reached Lima, Peru, were further sold to Upper Peru (Bolivia), but others also to Chile, or Quito [2].

3.8.3 Genetic analyses in Bolivia

Previous genetic studies in Bolivia were so far primary focused on Native American populations, including especially the main Andean linguistic groups Aymara and Quechua [281–283], but also Amerindians from La Paz [284] and populations from the lowlands of Bolivia like the Ignaciano and Trinitario (Moxo), Movima and Yuracare in Beni [285], or Chimane and Moseten from Piedmont which is located again in the department of Beni [286].

Analyses of the mitochondrial genome in previous studies have shown so far that haplogroups are mainly of Native American ancestry. For example, in the Llanos de Moxo C (50%) was the predominant haplogroup followed by B (24.1%), A (18.5%), and D (5.6%) [285]. Haplogroup C (84%) was almost uniquely observed in Ayoreo Indians of Bolivia and Paraguay with haplogroup D (16%) accounting for the remaining prevailing haplogroups. Haplogroup A and B, however, were absent in the Ayoreo Indians [283]. In contrast, haplogroup B4 (57.27%) was most frequent in La

Paz with C1 (19.09%) at the second, A2 (10%) at the third, and D1 (2.73%) and D4h3 (1.82%) at the fourth and fith position [284]. Haplogroup B2 seems to be most prevalent in Aymaras (81%) and Quechua (61%). In Aymaras the most frequent haplogroup after B2 was A2 (7%) followed by C1 (6%) and D1 (5%) and in Quechuas C1 (19%), A2 (15%), and D1 (4%) [281]. In Native American groups from Piedmont most individuals exhibited firstly haplogroup B (69.5%), and secondly, A (19.6%), D (6.5%), and C (2.2%) [286]. Mitochondrial DNA diversity was observed to be high in the Amerindian populations of the Llanos de Moxo [285], in a population of La Paz [284], in Aymara- and Quechua-speaking groups [281], and in Amerindians of the Piedmont where groups were also observed to exhibit high inter-group variability [286]. In contrast, Ayoreo Indians of Bolivia and Paraguay seem to have a low mtDNA diversity, and furthermore, exhibit distinct mtDNAs when compared to other Indian groups which might be according to the authors the result of a strong founder effect and/or random loss of variability in the mtDNA [283]. The ethnic group Uros also seems to be genetically distinct. It was observed that Uros populations from Bolivia and Peru are heterogenous groups with mtDNA and Y chromosome lineages that are derived from different ancestors than those from other Andean populations [287].

With respect to the Y chromosome the main proportion of haplogroups so far was observed to be of Native American ancestry like, e.g, in Aymara- (97%) and Quechua-speaking (78%) people. Aymara individuals exhibited either Q1a3a (89%) or Q1a3* (11%) whereas in Quechuas only Q1a3a (100%) was found [281]. In a Bolivian population, including individuals from the departments La Paz and Chuquisaca, Native American Y chromosomes were predominant (61%), especially Q1a3a1a*-M3. However, European ancestry on the Y chromosome reached also a comparatively high value (39%). In La Paz (70.2%) more Native American haplogroups were found than in Chuquisaca (50%), and thus, European ancestry was higher in the latter population. Sub-Saharan haplogroups were not found in La Paz and Chuquisaca [233]. Similar was found in another population of Bolivia with individuals from different locations within the country. That is, most individuals (61%) again had a Native American ancestry (Q1a3a1) on the Y chromosome and the other part (39%) exhibited European haplogroups (R1b, J2, G) [288]. In a survey in Andean populations (Peru, Colombia) for new SNP markers on the Y chromosome a new sub-lineage (Q1a3a4) based on one new single Y-SNP (SA01) was found. The derived state at this SNP was again found in 16 individuals out of a sample set with individuals of haplogroup Q1a3a (n=1754) from various South American populations (Peru, Bolivia, Equador, Brazil), including 15 samples from Peruvian indigenous communities and 1 from Bolivia [235]. In a population from Beni itcould

be observed on 17 Y-STR loci that the genetic differentiation with Peruvians is strong [289]. Apparently, populations from the high altitudes (La Paz) in Bolivia are genetically less diverse (HD=92%) than populations from low altitudes (Alto Beni, HD=99%) based on 11 Y-STRs [290].

In the same population of [288] some of the individuals were also observed with a substantial European admixture (22-47%) on autosomal SNPs whereas African admixture again was observed to be minimal or absent [288]. In a study with other autosomal markers, that is, with autosomal and X chromosome Alu insertions, it was revealed that the two main linguistic groups Aymara and Quechua are genetically similar, but differed from two Quechua-speaking population from Peru suggesting an expansion of the Quechua language in Bolivia without any important demographic genetic contributions according to the authors [291].

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

4 Aims

This thesis is conducted within the Marie Curie Initial Training Network (ITN) EUROTAST which is an interdisciplinary network of historians, social scientists, archeologists, and geneticists to uncover the history and the contemporary legacy of the Transatlantic Slave Trade. Projects within the EUROTAST network are focused on investigating the genetic African geographic origin of American populations, finding out about health, nutrition, and diseases of slaves during enslavement, documenting the material legacy of slavery, understanding the memorialization of the TAST, and investigatingthe impact of genetics on identiy.

The primary aim of the present thesis is to unravel the history of the TAST in the Bolivian population with genetics. The majority of the Bolivian population has indigenous origins, but due to the TAST African descendant people in Bolivia exists, primary in the tropical valleyes of the Yungas. The Yungas valleys are geographically very isolated and the number of African descendants or 'Afro-Bolivians', respectively, is low when compared to other regions in the Americas that were as well affected by the TAST. Studying the Bolivian population is especially of importance and interest therefore, because they have been widely neglected so far not only by geneticists, but also by historians.

The little knowledge about the TAST in Bolivia can be attributed as well in part to the lack of historical documents or if excerpts are available, many of them are incomplete. In this context genetics can be an important complementary research area that has the capability to provide new perspectives on the history of the TAST.

The principal goals of the present thesis is

(i) to compare continental ancestry of the population from the Yungas with the remaining Bolivian population based on uniparental and biparental markers,

(ii) to find out the genetic geographic origin of African descendant 'Afro-Bolivians' of the Yungas valley,

(iii) to evaluate the paternal and maternal genetic contribution from Europe, Africa, and America in Bolivia.

Aims

The mitochondrial DNA and the Y chromosome are highly geographically informative markers due to their small effective population size which makes them sensitive to genetic drift. Based on these characteristics, differences between populations are large. With respect to the mtDNA, both control region (HVS-I and HVS-II) and whole mitochondrial genomes will be analyzed to infer geographic origin of Bolivians. The sequencing of the whole mtDNA genomes of individuals from the 'Afro-Bolivian' community of Tocaña will be used to infer African genetic origin. During the TAST most enslaved people that embarked on ships in Africa came from West-Central Africa (present-day Angola and Congo). Other important regions of slave embarkations in Africa included Senegambia, Sierra Leone, Windward Coast, Gold Coast, Bight of Benin, Bight of Biafra, and Southeast Africa. The aim of the thesis is to find out the geographic region in Africa that contributed most to'Afro-Bolivian' individuals, here to individuals from the community of Tocaña. With respect to the Y chromosome, both Y-STRs and Y-SNPs will be investigated in Bolivians to analyze genetic contribution by males from America, Europe, and Africa to the whole population.

As with both uniparental markers only one lineage (either maternal or paternal) is represented, continental ancestry will be additionally investigated with autosomal Indels (Insertion Deletion polymorphisms) andSNPs. Genome-wide SNP genotyping will be conducted to provide further inferences about the geographic origin of African descendant people from Bolivia. Continental ancestry based on Indels will provide additional information about population structure within Bolivia. As the Yungas are historically most connected to the TAST, a comparison of African ancestry between different populations within the country will reveal the genetic impact of the TAST in different populations of Bolivia.

At last,results of uniparental and autosomal markers will be compared. The colonization of the Americas and TAST was almost exclusively directed by European males. Traders at the African coast mostly preferred males in their prime age, but because trading also depended on the availability of slaves, both males and females were forced to migrate to the Americas. In the Americas populations of three different continental origins met: America, Europe, and Africa. By means of both uniparental and autosomal markers the final aim of the thesis is to look for a gender bias in the Bolivian population.

5 Materials and Methods

5.1 Population samples

In the present thesis a total of 822 samples from the Bolivian population were included in the genetic analyses whereby 102 samples came from the department of Beni, 103 from Cochabamba, 89 from Chuquisaca, 90 from Pando, 83 from Santa Cruz, and 355 from La Paz (Table 6). Out of all samples from La Paz 19 individuals originated from the 'Afro-Bolivian' community of Tocaña which is situated in the Yungas valley. Individuals from the Yungas valleys, including the community of Tocaña, are listed as Sub-Andeans as the Yungas valley are steep regions and the transition from the Andeans to the Llanos cannot be perfectly assigned to departments.

The Bolivian samples were collected from saliva of in- and outpatients in clinical centres via cotton swabs with the authorization of the director of each respective facility and with the consent of each patient. DNA extraction was done with a kit from QIAGEN following the protocol "DNA Purification from Buccal Swabs-Spin".

All 822 samples were investigated for the mitochondrial control region (HVSI, HVSII) whereby 16 individuals were selected to be sequenced on the whole mtDNA genome including 8 samples from the 'Afro-Bolivian' community Tocaña. With respect to the Y chromosome 226 individuals were analyzed for Y-STRs and Y-SNPs. Out of all samples 703 individuals were used for indel genotyping and 2 for genotyping with the Axiom® Genome-Wide Human Origins 1 Arrayfrom Affymetrix.

Table 6: Summary of Bolivian population samples for autosomal genotyping (indel, SNP), mtDNA and Y chromosome analyses.

		Autosomal DNA		mtDNA		Y
Department	Total	Indel	SNP	HVS1/HVSII	whole	chromosome
Beni	**102**	77	-	102	2	41
Cochabamba	**103**	37	-	103	1	28
Chuquisaca	**89**	73	-	89	-	-
Pando	**90**	89	-	90	2	22
Santa Cruz	**83**	82	-	83	3	26
La Paz	**355**	345	2	355	8	109
Total	**822**	**703**	**2**	**822**	**16**	**226**

The analyses are separated between individuals from the Yungas and the remaining population. All samples from the Yungas came from La Paz, whereas 97 originate from the Nor Yungas and 8 from the Sud Yungas. Among the Nor Yungas 19 individuals are from the 'Afro-Bolivian' community Tocaña.

Furthermore, 327 reference samples from the Human Genome Diversity Cell Line Panel (HGCP-CEPH), represented by African (n=105), American (n=64), and European (n=158) individuals, were included for analyzing the indel genotypes [269, 292]. For the analysis of the African geographic origin of the 8 whole mtDNA genome sequences, 2233 previously published mtDNA sequences from individuals of known geographic origin that exhibited the African-specific L-haplogroup were used for sequence comparision and creating phylogenetic trees. Samples that were used for the phylogenetic trees are shown in the Appendix (chapter 12.1). Furthermore, a population set of n=1795 samples for the whole genome SNP genotyping was used for the statistical analyses (Appendix, chapter 12.2).

5.2 Mitochondrial DNA analyses

5.2.1 Whole mtDNA genome sequencing

Sequencing the whole mitochondrial genome consisted of six steps and explained in detail in the following sections:

1: PCR amplification

2: Polyacrylamide Gel Electrophoresis

3: Purification of PCR product

4: Sequencing

5: Purification of sequencing product

6: Capillary Electrophoresis

5.2.1.1 PCR amplification

PCR amplifications were performed in the thermocyclers GeneAmp® PCR System 2700, 9700 or in the 2720 Thermal Cycler.

PCR Reagents	Volume
Kit Qiagen	4µl
Primer F	1µl
Primer R	1µl
H_2O	2µl
DNA	2µl
Final Volume	**10µl**

The PCR was conduction with the following conditions.

Temperature	Time	Cycles
95°C	15min	1x
94°C	30sec	
58°C	90sec	35x
72°C	90sec	
72°C	10min	1x
8°C	∞	

Each sample was amplified with the primers 1-11 which were developed by Torroni and colleagues [165] (Table 7). The length of fragments that were produced by the primers ranged between 1740 and 1980 bp. Samples with lower quality, and thus, PCR or sequencing failure were amplified with primers that yielded smaller fragments. The primers that produced shorter fragments were developed by Kivisild and colleagues [27] and resulted in fragments of lengths between 200 and 601 basepairs (Table 8).

Table 7: PCR primers for whole mtDNA Sequencing (developed by Torroni and colleagues [165]).

PCR Primer	basepairs	5' np	3'np	Sequence	MTemp. (°C)
1	1,845	14897	14918	CTAGCCATGCACTACTCACCAG	59.96
		155	134	AATAGGATGAGGCAGGAATCAA	59.93
2	1,759	16488	16509	CTGTATCCGACATCTGGTTCCT	59.93
		1677	1656	GTTTAGCTCAGAGCGGTCAAGT	60.08
3	1,832	1404	1425	ACTTAAGGGTCGAAGGTGGATT	60.23
		3235	3214	CTTAACAAACCCTGTTCTTGGG	59.9
4	1,784	2900	2921	CAATAACTTGACCAACGGAACA	59.9
		4683	4662	TTAGAAGGATTATGGATGCGGT	59.83
5	1,771	4381	4402	ACCTATCACACCCCATCCTAAA	59.59
		6151	6130	ACTAGTCAGTTGCCAAAGCCTC	59.95
6	1,747	5871	5892	GCTTCACTCAGCCATTTTACCT	59.79
		7617	7596	TCTTGTAGACCTACTTGCGCTG	59.72
7	1,980	7239	7260	GCATACACCACATGAAACATCC	60.13
		9218	9197	TTGGTGGGTCATTATGTGTTGT	60.02
8	1,740	8910	8931	CTTACCACAAGGCACACCTACA	60.09
		10649	10628	AGGCACAATATTGGCTAAGAGG	59.65
9	1,769	10457	10478	TCATATTTACCAAATGCCCCTC	60.04
		12225	12204	AGTTCTTGTGAGCTTTCTCGGT	59.57
10	1,816	12014	12035	CTCACCCACCACATTAACAACA	60.7
		13829	13808	AGTCCTAGGAAAGTGACAGCGA	60.44
11	1,873	13477	13498	GCAGGAATACCTTTCCTCACAG	60.13
		15349	15328	GTGCAAGAATAGGAGGTGGAGT	59.64

Table 8: PCR/SEQ primers for whole mtDNA Sequencing (developed by Kivisild and colleagues [27]).

PCR/SEQ Primer	basepairs	5' np	3'np	Sequence
1F	509	412	435	GGTATGCACTTTTAACAGTCACC
1R		921	901	ACTTGGGTTAATCGTGTGACC
2F	532	870	888	CAGGGTTGGTCAATTTCGT
2R		1402	1380	TCATAAGGGCTATCGTAGTTTTC
3F	521	1348	1366	GTGGCAAGAAATGGGCTAC
3R		1869	1845	TCTAGTTAATTCATTATGCAGAAGG
4F	522	1814	1838	AAGCATAATATAGCAAGGACTAACC
4R		2336	2317	ATGCGGAGGAGAATGTTTTC
5F	495	2238	2262	AACATATAACTGAACTCCTCACACC
5R		2733	2713	CCATAGGGTCTTCTCGTCTTG
6F	482	2665	2686	TTTAACCAGTGAAATTGACCTG
6R		3147	3125	CCTTATTTCTCTTGTCCTTTCGT
7F	399	3096	3119	TTTCTATCTACCTTCAAATTCCTC
7R		3495	3477	GGGTTTTAGGGGCTCTTTG
8F	446	3442	3460	CTACAACCCTTCGCTGACG
8R		3888	3867	GTTGGTCTCTGCTAGTGTGGAG
9F	445	3850	3873	GCCATAATATGATTTATCTCCACA
9R		4295	4268	TATCAAAGTAACTCTTTTATCAGACAT
10F	453	4223	4244	CCATACCCATTACAATCTCCAG
10R		4676	4658	GATTATGGATGCGGTTGCT
11F	477	4613	4631	AAACCCTCGTTCCACAGAA
11R		5090	5064	AATAGTTAAATTAAGAATGGTTATGTT
12F	427	5032	5055	GATGAATAATAGCAGTTCTACCGT
12R		5459	5439	GATGAGTGTGGGGAGGAATGG
13F	467	5396	5419	CAACGTAAAAATAAAATGACAGTT
13R		5863	5841	TGTAAATCTAAAGACAGGGGTTA
14F	486	5805	5825	ATATGAAAATCACCTCGGAGC
14R		6291	6270	GGTAGACTGTTCAACCTGTTCC
15F	441	6239	6258	GCTCGCATCTGCTATAGTGG
15R		6680	6657	AGTTACAATATGGGAGATTATTCC
16F	477	6473	6495	CACAGCAGTCCTACTTCTCCTAT
16R		6950	6928	GGTGAAAAGAAAGATGAATCCTA
17F	492	6809	6829	AGCATATTTCACCTCCGCTAC
17R		7301	7276	TATTACTGCTGTTAGAGAAATGAATG
18F	469	7231	7248	ACCCCGATGCATACACCA
18R		7700	7678	GGACTAGGAAGCAGATAAGGAAA
19F	505	7599	7619	CGCAAGTAGGTCTACAAGACG
19R		8104	8082	ATCTGTTTTTAAGCCTAATGTGG
20F	462	7931	7956	CTAATCTTCAACTCCTACATACTTCC
20R		8393	8371	GCCATACGGTAGTATTTAGTTGG
21F	436	8332	8355	AAGATTAAGAGAACCAACACCTCT

PCR/SEQ Primer	basepairs	5' np	3'np	Sequence
21R		8768	8746	GTGGCAATAAAAATGATTAAGGA
22F	519	8707	8729	ACAACACTAAAGGACGAACCTG
22R		9226	9208	GCATGTGATTGGTGGGTCA
23F	528	9098	9123	TCATCTTCACAATTCTAATTCTACTG
23R		9626	9605	TGATGCGAGTAATACGGATGTG
24F	547	9308	9329	ATTTCACTTCCACTCCATAACG
24R		9855	9834	TAGTGAGGAAAGTTGAGCCAAT
25F	528	9751	9770	TCGAGTCTCCCTTCACCATT
25R		10279	10259	GGTAAAAGGAGGGCAATTTCT
26F	487	10175	10193	CGGCTTCGACCCTATATCC
26R		10662	10640	CTAGTATGGCAATAGGCACAATA
27F	452	10600	10622	CTACTCTCATAACCCTCAACACC
27R		11052	11028	ATAGAGAGGTAGAGTTTTTTTCGTG
28F	543	10997	11014	AGCCAACGCCACTTATCC
28R		11540	11520	AGGGGTAGGCTATGTGTTTTG
29F	534	11453	11475	GCCGCAGTACTCTTAAAACTAGG
29R		11987	11965	TGTAGAGGGAGTATAGGGCTGTG
30F	563	11947	11970	AGGACTCAACATACTAGTCACAGC
30R		12510	12487	GTCTAGGCACATGAATATTGTTGT
31F	441	12395	12414	TTACCACCCTCGTTAACCCT
31R		12836	12818	GCTGTGTTGGCATCTGCTC
32F	489	12783	12806	TATATCCTTCTTGCTCATCAGTTG
32R		13272	13250	GAGTCCTAGTTGACTTGAAGTGG
33F	503	13122	13139	CTTCCACCCCCTAGCAGA
33R		13625	13603	GTGAGAAGAATTATTCGAGTGCT
34F	224	13510	13530	AAAGACCACATCATCGAAACC
34R		13734	13715	AAATCCTGCGAATAGGCTTC
35F	328	13627	13648	CTAACAGGTCAACCTCGCTTCC
35R		13955	13936	AGATAGGGGATTGTGCGGTG
36F	200	13835	13855	CAGCCCTAGACCTCAACTACC
36R		14035	14010	AAATTGTTTTAGGTAATAGCTTTTCT
37F	430	13940	13959	GCACAATCCCCTATCTAGGC
37R		14370	14352	TTCGTCCTGTGGGTGAAAG
38F	521	14288	14311	CCTTCATAAATTATTCAGCTTCCT
38R		14809	14789	GAGGTCGATGAATGAGTGGTT
39F	505	14701	14723	CAATGATATGAAAAACCATCGTT
39R		15206	15185	GGATGGCGGATAGTAAGTTTGT
40F	490	15048	15071	GCCTATATTACGGATCATTTCTCT
40R		15538	15519	GGTGTTTAAGGGGTTGGCTA
41F	601	15439	15463	CTTACTTCTCTTCCTTCTCTCCTTA
41R		16040	16022	GCTTCCCCATGAAAGAACA

5.2.1.2 Polyacrylamide Gel Electrophoresis (PAGE)

The aim of conducting a Polyacrylamide Gel Electrophoresis was to check if the PCR reaction was successfully performed. In the case of a PCR failure the respective PCR product was not further used in the following steps.

At first, a glass plate with a fotosensible film (Gel-fix®/Serva, Heidelberg, Germany) and a glass plate with notches and cleaned with Repel-Silane (Pharmacia Biotech) were placed together with clamps and properly vertically fixed. The polyacrylamide gel was prepared with the following reagents:

Reagent	Volume
T_9C_5	4.4ml
H_2O	225µl
Glycerol	350µl
APS	25µl
10µl	TEMED

T_9C_5		Tris-HCl	APS (ammonium persulfate)
Acrylamide	8.55g	22,715g in a total volume	1g ammonium persulphate
Bisacrylamide	0.45g	of 500ml H2O	10ml H_2O
Tris-HCl	100ml	HCl up to a pH of 8.8	

The polyacrylamide gel was left to polymerize between the two glass plates for 20 to 30 minutes. After polymerization the fotosensible film with the gel was removed and the wells on the gel loaded with 1.5µl PCR product. The gel was then placed in a gel apparatus and connected to the device via paper strips which were soaked in Tris-Glycin.

Tris-Glycin	
7.6g	Tris
500ml	H_2O
~1.2g	Glycin up to a pH of 8.8
	(bromphenol blue)

As the electrophoresis runs from the minus to the plus pole the paper strips on the minus pol were soaked in Tris-Glycin which contained bromphenol blue. This color marker is used to indicate the process of the electrophoresis. The gel was run with a voltage of maximum 15V.

After electrophoresis the gel on the film was placed in a tray together with ethanol (10%) and agitated for a minimum of 10 minutes. After fixing the DNA to the gel, ethanol was removed and replaced by nitric acid (1%) for 3to 5 minutes. After washing with destilled water, silver nitrate ($AgNO_3$) was placed for 20-25 minutes on the gel to stain the DNA bands. Following another washing step with destilled water, DNA bands were finally made visible with sodium carbonate (7.5g sodium carbonate in 250ml destilled water) mixed with 100µl formaldehyde in 3 washing steps. Finally, the stained DNA bands were fixed with acetic acid (5%).

5.2.1.3 Purification of PCR product

If the result on the Polyacrylamide Gel was positive, PCR products were purified via physical filtration. Each sample was filled up with distilled water to a volume of 100µl in 96-well MulitScreen® filter plates for High Throughput Separations (Millipore). Subsequently, the platewas placed on a vacuum purification device (Millipore) for about 10 minutes until the whole water passed through the filter. After filtration 28µl distilled water was added to the purified PCR products and left for about 5-10 minutes before placing each sample into its original tube.

5.2.1.4 Sequencing

The thermocyclers GeneAmp® PCR System 2700, 9700 or the 2720 Thermal Cycler were used for sequencing the PCR production. Sequencing was undertaken with the following reagents and conditions:

PCR Reagents	Volume
Buffer Big Dye	2µl
Kit Big Dye	0.5µl
Primer	1µl
H_2O	1.5µl
Purified PCR product	5µl
Final Volume	**10µl**

Temperature	Time	Cycles
96°C	3min	1x
96°C	30sec	25x
50°C	15sec	
60°C	4min	
12°C	∞	

Each sample was firstly sequenced with the sequencing primers which were developed by Torroni and colleagues [165] (Table 9, SEQ Primer 1a to 11c). In case of a sequencing failure samples were amplified via PCR and subsequently sequenced with primers that were created by Kivisild and colleagues [27] (Table 8, Primer 1F/1R to 41F/41R). Sequencing reagents and conditions were the same for both sets of primers.

Table 9: Sequencing Primer for Whole mtDNA Sequencing. PCR primers from 1 to 11 and SEQ primers from 1a to 11c were developed by Torroni and colleagues [165]. PCR and SEQ primers from 1F/1R to 41F/41R were developed by Kivsild and colleagues [27].

PCR Primer	SEQ Primer		PCR/SEQ Primer F		PCR/SEQ Primer R	
1	1a	14948F			39R	15185R
14897F	1b	15564F	40F	15071F	40R	15519R
155R	1c	131R	41F	15463F	41R	16022R
2	2a	16522F				
16488F	2b	584F	1F	435F	1R	901R
1677R	2c	1060F	2F	888F	2R	1380R
3	3a	1445F	3F	1366F	3R	1845R
1404F	3b	2047F	4F	1838F	4R	2317R
3235R	3c	2509F	5F	2262F	5R	2713R
			6F	2686F	6R	3125R
4	4a	3085F	7F	3119F	7R	3477R
2900F	4b	3598F	8F	3460F	8R	3867R
4683R	4c	4010F	9F	3873F	9R	4268R
			10F	4244F	10R	4658R
5	5a	4410F	11F	4631F	11R	5064R
4381F	5b	5014F	12F	5055F	12R	5439R
6151R	5c	5544F	13F	5419F	13R	5841R
6			14F	5825F		
5871F	6a	6041F			14R	6270R
7617R	6b	6600F	15F	6258F	15R	6657R
			16F	6495F	16R	6928R
			17F	6829F	17R	7276R
7	7a	7336F	18F	7248F	18R	7678R
7239F	7b	7937F	19F	7619F	19R	8082R
9218R	7c	8459F	20F	7956F	20R	8371R
			21F	8355F	21R	8746R
8	8a	8975F	22F	8729F	22R	9208R

PCR Primer	SEQ Primer		PCR/SEQ Primer F		PCR/SEQ Primer R	
8910F	8b	9589F	23F	9123F	23R	9605R
10649R	8c	10147F	24F	9329F	24R	9834R
			25F	9770F	25R	10259R
			26F	10193F	26R	10640R
9	9a	10498F	27F	10622F	27R	11028R
10457F	9b	11081F	28F	11014F	28R	11520R
12225R	9c	11644F	29F	11475F	29R	11965R
			30F	11970F		
10	10a	12114F			30R	12487R
12014F	10b	12600F	31F	12414F	31R	12818R
13829R	10c	13134F	32F	12806F	32R	13250R
			33F	13139F	33R	13603R
11	11a	13568F	34F	13530F	34R	13715R
13477F	11b	14103F	35F	13648F	35R	13936R
15349R	11c	14603F	36F	13855F	36R	14010R
			37F	13959F	37R	14352R
			38F	14311F	38R	14789R
			39F	14723F		

5.2.1.5 Purification of sequencing product

For the purification of the sequencing product, firstly, the following reagents were added to 10µl of each sequencing product of a 96-well plate:

Reagents	Volume
125mM EDTA	1µl
3M NaOAc	1µl
Ethanol (100%)	25µl

The plate was closed with Adhesive PCR Seal (4titude) and agitated to homogenize the mix of sequencing product and reagents. The plate was left for 15minutes at room temperature. After that the plate was placed on a centrifuge (Centrifuge 5804, Eppendorf) at 3700rpm for 45 minutes. The whole plate was then inverted on paper and centrifuged for 1 minuate at 200rpm. The next step was to add 70µl of ethanol (70%) to each sample and centrifuge at 3700rpm for 15 minutes. The whole plate was inverted again and centrifuge for 1 minuate at 200rpm. The plate was left until ethanol was completely evaporated.

5.2.1.6 Capillary Electrophoresis (CE)

After purification each sample was filled with 10µl Hi-Di™ Formamide and placed in an ABI3730xl sequencer for Capillary Electrophoresis (CE). The results of the CE were analyzed with the software SeqScape v2.1. For analysis all nucleotide positions, which differed from the revised Cambridge Reference Sequence [121], were noted and used for further analyses.

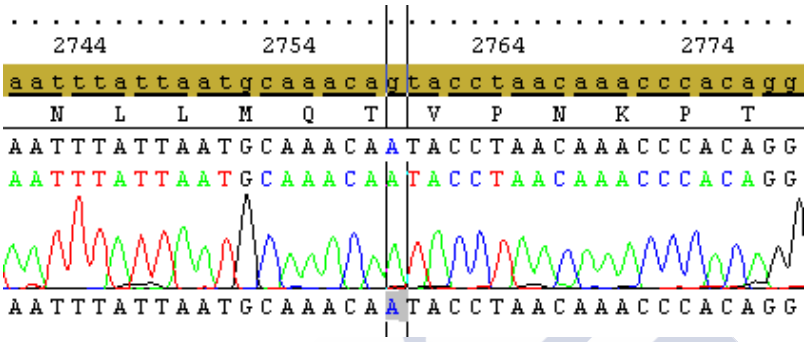


Figure 61: Analyzing whole mtDNA sequence with SeqScape v2.1. Shown is Tocaña295 who has a mutationat position 2758 and exhibits A instead of a G in the reference sequence.

5.2.1.7 Haplogroup classification

Mitochondrial DNA sequences were classified by using the online software HaploGrep which is based on Phylotree build 16 [293]. Phylotree, the phylogenetic tree of global human mitochondrial DNA variations, was last updated in February 19, 2014 [132]. HaploGrep was used principally to determine mtDNA haplogroups of samples. However, the online software also provides information about the quality of haplogroup assignment to a sample indicated either by red, yellow, or green, where green indicates the most reliable assignment of haplogroups. According to quality, HaploGrep lists alternative haplogroups assignments into ranks. Moreover, the programme shows all the expected mutations that belong to a haplogroup and found in a sample sequence. As based on Phylotree, 309.1C(C), 315.1C, 523-524d (522-523d), 16182C, 16183C, 16193.1C(C) and 16519 were not used for haplogroup classification and shown as hotspots in HaploGrep. Local private mutations of a sample included mutations that are not associated with the corresponding haplogroup, but detected in the sample. Global private mutations comprise mutations that so far had not been seen in Phylotree.

5.2.2 MtDNA control region sequencing

The mitochondrial control region was sequenced following the same procedure as explained for the whole mtDNA genome (5.2.1).

The primers, which were used for the PCR and for sequencing, are shown in Table 10whereas for sequencing only 15997L and 16555L were used.

Table 10: Primers for PCR and sequencing the control region.

Primer	Control Region	Sequence	Reference
15997L	HSVI	5'-CACCATTAGCACCCAAAGCT-3'	Ward and colleagues [294]
017H	HSVI	5'-CCCGTGAGTGGTTAATAGGGT-3'	Álvarez-Iglesias and colleagues [295]
16555L	HSVII	5'-CCCACACGTTCCCCTTAAAT-3	Álvarez-Iglesias and colleagues [295]
611H	HSVII	5'CAGTGTATTGCTTTGAGGAGG-3'	Cerezo Fernández [296]

Quality control of mtDNA sequences was conducted using EMPOP (http://empop.org/), haplogrep (http://haplogrep.uibk.ac.at), and manual inspection.

5.2.3 Statistical analyses

The software DnaSP v.5 was used to calculate haplotype (H) and nucleotide (p) diversity values, but also mean number of pairwise differences (M), number of different haplotypes, and number segregating sites [297]. Arlequin 3.5.12 was used to compute AMOVA (Analysis of Molecular Variance) and the significance of the covariance components associated with different levels of genetic structure were tested on haplotype frequencies applying a non-parametric permutation procedure. The programme was also used to calculate population pairwise FST values, between/within population average nucleotide pairwise differences, and Nei's inter-population distances [298, 307].

The sequence range 16090 to 16365 was used to compute diversity indices, phylogeographic inferences, and inter-population comparisons as this sequence range is most commonly reported in the literature. Sequence variations around 16189, which were usually associated with length heteroplasmy like, e.g., 16182C or 16183C, were ignored for the analyses.

The R software (http://cran.r-project.org/) was used to conduct Fisher's exact test and Pearson's chi-quare test [299]. The significance nominal value was set to α=0.05.

Maximum parsimony trees for the 8 Bolivian complete mtDNA genomes of haplogroup B2 and A2 were built together with 10 complete mtDNA genomes from the literature. For each cluster of samples (B2o, A2ah, B2b, B2b2) the time of the most recent common ancestor (TMRCA) was calculated by computing the averaged distance (r) of all haplotypes in a clade to the respective root haplotype. Heuristic estimates of the standard error (s) were calculated from an estimate of the genealogy, as done bySaillard and colleagues [300]. Hotspot mutations like 16182C, 16183C, and 16519 were excluded from the calculations. The mutational distances were converted into years based on the corrected evolutionary rate proposed by Soares and colleagues [26].

Maximum parsimony trees for the 8 Bolivian complete mtDNA genomes that are classified into African-specific L haplogroups were built with the software mtPhyl v4.015 (http://eltsov.org). All phylogenetic trees were subsequently adapted to the most recent version of PhyloTree build 16. The 8 complete mtDNA sequences were analyzed with n=2233 previously published whole genomes. Sequences that clustered in one phylogenetic tree (L0a, L1c3, L3d1) together with the 8 Bolivian genomes finally comprised n=152 samples (Appendix, chapter 12.1).

The software paml 4.8 (Phylogenetic Analysis by Maximum Likelihood) [301] was used to calculate Maximum Likelihood Age estimates. The analysis is based on the phylogenetic trees gained by mtPhyl and adapted to the most recent version of Phylotree. Here the HKY85 mutation model was used, and thus, it was assumed unequal base frequencies ($\pi_A \neq \pi_C \neq \pi_G \neq \pi_T$) and a different substitution rate ($\alpha \neq \beta$) for transitions and transversions. For the analysis all indels were removed except the 9-bp deletion. 16519 was removed in all sequences and 3107 replaced by a C. The sequences were divided into three partitions: Gene 1 (length=16022), Gene 2 (length=196), and Gene 3 (length=350) corresponding to HVS-I, HVS-II, and the remaining sequence. Branch Length, Standard Error, mutation rates and lengths for the three genes were used to calculate ρ and σ for each branch. Age estimates were extrapolated by using the corrected mutation rate (corrected molecular clock) developed by Soares and colleagues [26]. Results are shown in the chapter 3.5.9.2 (Introduction) and publication in preparation.

The map for the location and distribution of the African mitogenomes used for the phylogeographic analysis was conducted with SAGA v. 2.1.1 (http://www.saga-gis.org/).

5.3 Y chromosome

5.3.1 Sample preparation

Extracted DNA was quantified on a NanoDrop® ND-1000 Spectrophotometer, and normalized to a 0.1ng/µl concentration for Y-STR typing and to 1-5ng/µl for Y-SNP typing.

5.3.2 Y-STR genotyping

5.3.2.1 PCR of Y-STRs

The following PCR and all other PCR and sequencing reactions below were carried out in a GeneAmp® PCR System 9700thermocycler (Applied Biosystems).

DNA samples for Y-STR typing were amplified using the AmpFlSTR®YFiler™ kit (Applied Biosystems, USA), which includes 17 Y-STRs (DYS19, DYS385 a/b, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYS635, GATA H4)by following the manufacturer's instructions. The final reactionvolume was changed to 7µl.

PCR Reagents	Volume
Reaction mix	3µl
Primer set	1.9µl
Amplitaq Gold®(Life Technology)	0.4µl
H_2O	1µl
DNA	0.7µl
Final Volume	**7µl**

Temperature	Time	Cycles
95˚C	11min	1x
94˚C	60sec	30x
61˚C	60sec	
72˚C	60sec	
60˚C	80min	1x

5.3.2.2 Capillary Electrophoresis (CE)

Amplified products were separated in a capillary electrophoresis with an ABI PRISM 3130xl (Applied Biosystems) by using the polymer POP4®. For the capillary electrophoresis 1.5µl PCR product were mixed with 0.15µl size standard LIZ 500. Results were analyzed with the software Genemapper ID 3.2.1 (Applied Biosystems, Foster City, USA).

5.3.3 Y-SNP genotyping

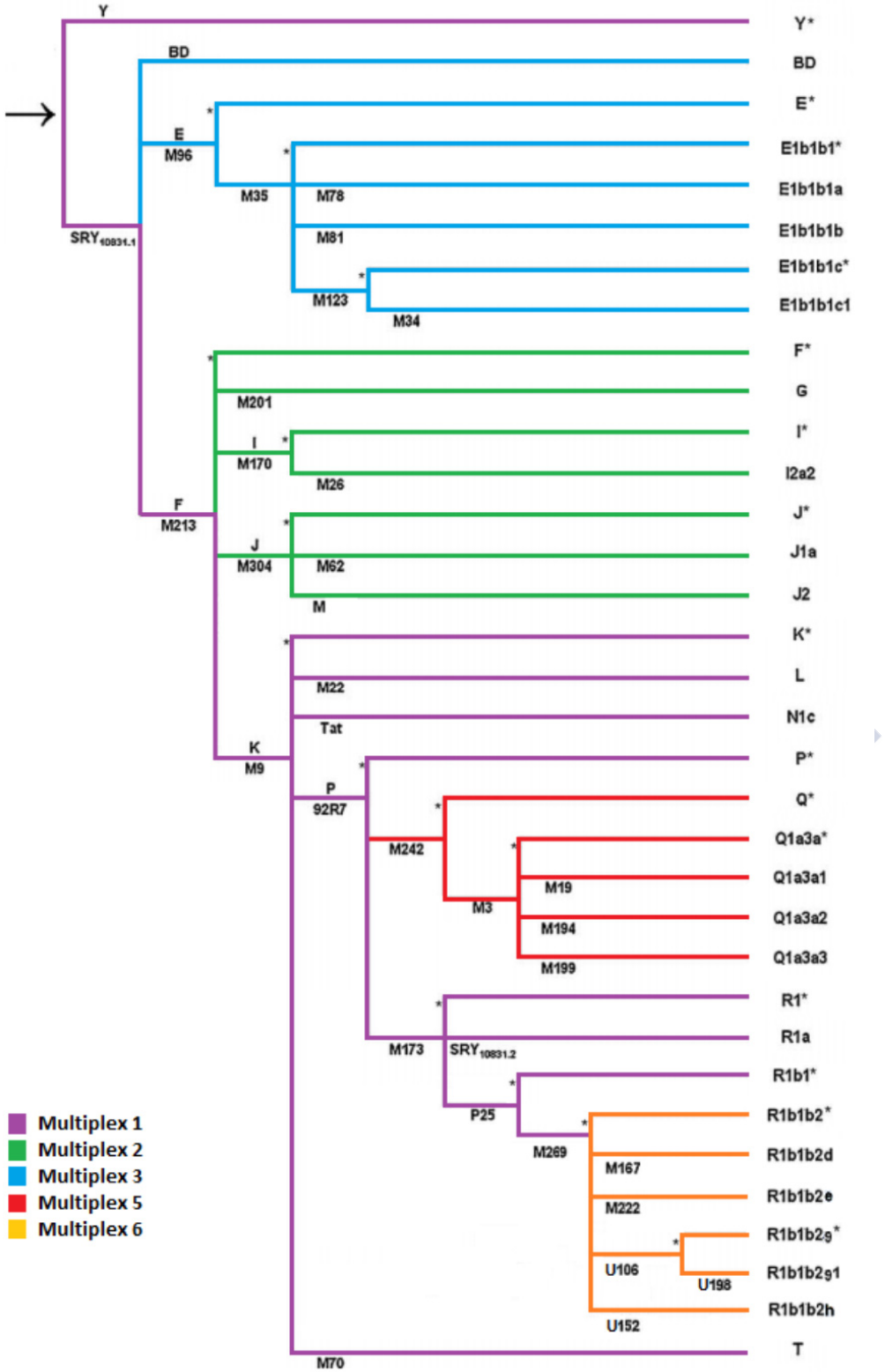
5.3.3.1 PCR of Y-SNPs

For Y-SNP haplogroup typing, five different PCR multiplex reactions were conductedfollowing [302, 303] in order to genotype 32 Y-SNPs. By means of the five multiplex reactions 34 haplogroups that are present in Europea, Africa, Middle East, and America can be detected (Figure 62). All samples where genotyped for multiplex 1. Depending on the observed SNPs of each individual, samples were further genotyped with one more multiplex which detects either SNPs that are specific for the Middle East (multiplex 2), Africa (multiplex 3), America (multiplex 5), or Europe (multiplex 6), respectively. PCR multiplex primers are described in [304] which were published in [302, 303, 305, 306].

PCR Reagents	Volume
PCR master mix Qiagen (Hilden, Germany)	5µl
Primer mix	1µl
DNA (5-10ng/µl)	1µl
H_2O	3µl
Final Volume	**10µl**

Temperature	Time	Cycles
95°C	15min	1x
94°C	30sec	35x
58°C	90sec	
72°C	90sec	
72°C	10min	1x

Figure 62: Phylogenetic tree of Y chromosome haplogroups. Based on [195] and published in [304].



5.3.3.2 Post-PCR purification

PCR products were purified with the enzyme Ilustra™ EXOSTAR™ (GE Healthcare) in order to remove all not incorporated dNTPs and primers. The enzyme has two functions. Firstly, it works as Exonuclease I in order to degrade primers, and secondly, as Alkaline phosphataseto dephosphorylate dNTPs. The following reagents and conditions were used:

Reagents	Volume
PCR product	1µl
EXOSTAR™	0.5µl

Temperature	Time
37°C	15min
85°C	15min

5.3.3.3 Minisequencing

For the minisequencing reaction a SNaPShot® (AB) kit was used. During the minisequencing reaction an unlabelled primer aligns one base before the polymorphism of interst where a fluorescent-labelled dideoxynucleotide (ddNTP) can be incorporated by the AmpliTaq® DNA FS polymerase. After the incorporation of one ddNTP, the reaction is terminated. The fluorophore were dR110 (blue, ddGTP), dR6G (green, ddATP), dTAMRA™ (black, ddCTP), and dROX™ (ddUTP). Mix of probes are listed in [304] and were published in [302, 303, 305, 306].

Reagents	Volume
Ready Reaction Mix	2µl
Mix of probes/primers	1.5µl
Purified PCR product	~1.5µl
Final Volume	**~5µl**

Temperature	Time	Cycles
96°C	10sec	25x
50°C	5sec	
60°C	30sec	

5.3.3.4 Post-extension purification

In a post-extenstion purification all not incorporated ddNTPs are removed. The purification was conducted with the enzyme SAP (Shrimp Alkaline Phosphatase) which dephosphorylates the 5'-end of the ddNTPs to inactivate them in order to avoid interferences during the capillary electrophoresis.

Reagents	Volume
SNaPshot product	5-10µl
USB® SAP (Affymetrix, Inc.)	1µl

Temperature	Time
37°C	60min
85°C	15min

5.3.3.5 Capillary Electrophoresis (CE)

The fluorescent-labelled fragments were separated by means of a capillary electrophoresis with ABI PRISM 3130xl® Genetic Analyzer (AB) by using the polymer POP4®. For the separation 0.3µl size standard GeneScan-LIZ120™ and 9.5µl Formamide HiDi™ were added to 2µl of the purified sample. Results were analyzed with the Genemapper™ ID 3.2.1 software.

5.3.4 Statistical analyses

Both minimum haplotype, which is based on 9 STR-markers (DYS19, DYS385 a/b, DYS389 I, DYS389 II, DYS390, DYS391, DYS392, DYS393), and the haplotype for all 17 Yfiler markers (DYS19, DYS385 a/b, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYS635, GATA H4) were determined. The software Arlequin v3.5.1.2 was used to estimate haplotype frequency and 'gene' diversity [298, 307]. However, DYS389II repeats were substracted from DYS389I repeats in order to calculate the real diversity indices. Statistical significance was estimated by using a permutation procedure in Arlequin (10,000 permutations).

Information about Y-STR intermediate alleles and their population distribution was obtained by researching the YHRD website (http://yhrd.org/). Haplotype discrimination capacity (DC) was

calculated by dividing thenumber of different haplotypes found in each department by the sample size.

In order to determine genetic structure and percentage of variance between the five departments and three regions of Bolivia, an AMOVA (Analysis of Molecular Variance) for Y-STRs data was carried out for which the R_{ST} genetic distances option was used.

Inter-population relationships between Bolivians and other South American populations werevisualizied by conducting a PCA (Principal Component Analysis) with the R software [299] using *cmdscale* (stat library).

Populations, which were used for the analyses included sample from Africa like Fon [308, 309], Fulbe [310], São Tomé e Príncipe [311], Bakaka, Ewondo, Ngoumba, Makina and Nzebi, Tiv and Igala [312], Angola [313], Bamelike [308], and Democratic Republic of the Congo [314], and from Native America like Nicaragua [315], Aymara, Quechua [281], La Paz, and Chuquisaca from Bolivia [233], Apalai [316], Diaguitas, Kolla-Jujuy, Kolla-Salta, and Mapuche [303], and from East Asia like Bali (Indonesia) [317], Philippines, Taiwan, She, Yao, and Vietnam [318], Han [319], but also from Europe like Cantabria, León [320], England, Wales, Scotland [321], Galicia [306],Germany [322], and Portugal [323].

Finally, admixture analysis was carried out using SPSS v17.0 among the different Y-SNPs frequency data populations used by Simms and colleagues [219]. In brief, this analysis employs a least-squaresmethod [324] that compares major Y-SNP haplogroups, grouped in parental sets from African, European, East Asian, South Asian, and Native American populations, and calculates anadmixture proportion in a given population related to these parental groups.

5.4 Autosomal DNA

5.4.1 Indels

5.4.1.1 AIM-Indelplex

In this thesis a set of 46 AIMs are used to investigate ancestral genetic components of the Bolivian population. The ancestry informative markers were developed by Pereira and colleagues [269] and are able to distinguish continental ancestry, that is, between African, European, East Asian, Native American, and even Oceanian populations. A list of all markers is shown in Table 11 and Table 12.

Table 11: AIM-INDEL multiplex developed by Pereira and colleagues [269].

MID	rs number	Chrom.	Position (bp)	Alleles described in dbSNP
MID-1470	rs2307666	11	64729920	-/GTTAC
MID-777	rs1610863	16	6551830	-/GAA
MID-196	rs16635	6	99789775	-/CAT
MID-881	rs1610965	5	79746093	-/ACTT
MID-3122	rs35451359	18	45110983	-/ATCT
MID-548	rs140837	6	3708909	-/CT
MID-659	rs1160893	2	224794577	-/CT
MID-2011	rs2308203	2	109401291	-/CTAGA
MID-2929	rs33974167	8	87813725	-/TA
MID-593	rs1160852	6	137345857	-/TT
MID-798	rs1610884	5	56122323	-/GGGAAA
MID-1193	rs2067280	5	89818959	-/AT
MID-1871	rs2308067	7	127291541	-/TT
MID-17	rs4183	3	3192524	-/TAAC
MID-2538	rs3054057	15	86010538	-/AACA
MID-1644	rs2307840	1	36099090	-/GT
MID-3854	rs60612424	6	84017514	-/TCTA
MID-2275	rs3033053	14	42554496	-/TCAGCAG
MID-94	rs16384	22	42045009	-/AAC
MID-3072	rs34611875	18	67623917	-/GCCCCCA
MID-772	rs1610859	5	128317275	-/TAG
MID-2313	rs3045215	1	234740917	-/ATTATAACT
MID-397	rs25621	6	139858158	-/TTCT
MID-1636	rs2307832	1	55590789	-/AA
MID-51	rs16343	4	17635560	-/TTTAT
MID-2431	rs3031979	8	73501951	-/ATTG
MID-2264	rs34122827	13	63778778	-/AAGT
MID-2256	rs133052	22	41042364	-/CAT
MID-128	rs6490	12	108127168	-/ATT
MID-15	rs4181	2	42577803	-/AAATACACAC
MID-2241	rs3030826	6	67176774	-/GTCCAATA
MID-419	rs140708	6	170720016	-/AATGGCA
MID-943	rs1611026	5	82545545	-/TGAT
MID-159	rs16438	20	25278470	-/CCCCA
MID-2005	rs2308161	10	69800909	-/AACAAT
MID-250	rs16687	7	83887882	-/CA
MID-1802	rs2307998	5	7814345	-/GGA
MID-1607	rs2307803	3	108981031	-/TG
MID-1734	rs2307930	6	84476378	-/CCAT
MID-406	rs25630	6	14734341	-/AG
MID-1386	rs2307582	1	247768775	-/AAACTATTCATTTTTCACCCT
MID-1726	rs2307922	1	39896964	-/CAAGAACTATAAT/CACTATCTATTAT
MID-3626	rs11267926	15	45526069	-/AATATAATTTCTCCA
MID-360	rs25584	12	112145217	-/AA
MID-1603	rs2307799	5	70828427	-/TTGT
MID-2719	rs34541393	20	30701405	-/AACT

Table 12: PCR primer sequences for AIM-Indelplex developed by Pereira and colleagues [269]. Lowercase italic letters of unlabeld primers are nucleotide tails that were added to the primers.

MID	Labeled Primer	Unlabeld Primer
MID-1470	6FAM-GAGTCTGACCCTTCATAAGC	gCCATGGTGATATTACGTCCC
MID-777	6FAM-TGGAAGACACGTCCTAAGAG	gTATTCCTCCAGGCTCTTTGC
MID-196	6FAM-CCAAGTTCTAGCCATATGGA	gtttcttGACTATCTTCTCTGACCATC
MID-881	6FAM-TTGGCTCCCTATGATAATCC	gtttcTTGTGTTCCCAAAGTTCTCC
MID-3122	6FAM-TCACAAGTCCGGAATACCAG	gAGTTATGGGATGGGAAGGAG
MID-548	6FAM-AGTCAGGACTGAAGAAACCC	gtttCAGTAAAACAAGAGCCCGTG
MID-659	6FAM-CACTGCATCAGACTGACTTC	gtttcttGGCTGCTTTGCTTTGAATTG
MID-2011	6FAM-TGAGAAACTAGGAGCTCTGG	gtTTCCTAAGAGCCACTGACAT
MID-2929	6FAM-TGTGATGTGGATAGGCAAGG	gAGGCTCCATTTGTTAAGAGG
MID-593	6FAM-TGCTCACTTTAGTGGAGACC	gtTTGCGTTTAGGTCCCTTCTG
MID-798	6FAM-ACGACAGTGTTCACAAAGAG	gCTGTTGTCTGACCTGTGAAG
MID-1193	6FAM-GCTGGGTAGTTTTTTCCTCC	gtTCCACCATCTACCTTCTATG
MID-1871	6FAM-TTGTAGTCAGAGAGTGTGCC	gAGCCTTTTCCCTAACGTCAC
MID-17	6FAM-AGAAACTGCAACCCTCCAAG	GATCCCAGACACTGAAGATG
MID-2538	6FAM-CTCGCAAAGTAGGCAAGTTC	gACACCAACAATCTTGGCACC
MID-1644	6FAM-ACACCACTGAAGATCTGACC	GGTCTAAAGTCAGTGCACAG
MID-3854	VIC-TCACCCTTATTCAGGGTTGC	GCCAGGGATTTAGTGTAGAG
MID-2275	VIC-CTACCTGACTACCACCTATG	gACCCAGCCTATCTGACTTTG
MID-94	VIC-TGGTGGCTCATGCACTTTTG	gtttACAGGGTCTCGCTATGATGC
MID-3072	VIC-AGCTTTTTCCGGCAACTCTC	gttTGGATGTGTCTGAGCTCAAC
MID-772	VIC-GTCTCRTTTTCCTGCAGTAG	gtttcttATCCTTCTGCTCACTCTACC
MID-2313	VIC-GCACACATGCAGAAATGCAG	GTTGTAACATCTGTGAGGTC
MID-397	VIC-TGGGCTTCTTCTGGGAAAAC	gCCACATTCAGGCGTTTTGTC
MID-1636	VIC-TTAGGAAGAGGTGCTATGGG	gCCTCCTTTGAAGACACACAG
MID-51	VIC-AAGATTGGAGGGAAAAGTGC	gCGTCCTCCACCTTCTTTTTC
MID-2431	VIC-AGGAGGAGCTGATAGACTTC	GCAGTGTTGCAACTGATACG
MID-2264	VIC-CTTTGGCTATCCTGTCTCAC	GTAGGAGACCACTCACATTC
MID-2256	NED-ATCGAACCGTTCCTAAGGAC	GCAAGAAAAGGAATCCAGGC
MID-128	NED-ATCAGGAGACAATCCAGCAG	gTCCAGCCATTCAGACAAAGG
MID-15	NED-GGGTTATTTGCCTCATCTCC	gtttctAGGTATTCTCTGTTCCCACG
MID-2241	NED-ACATACACGTGGAAGACTGC	gttACTGTCGACTGATCCAATAG
MID-419	NED-CAGGAAAGTATGGCCCATTC	GTCCATGTTTTCTTTGAGCATC
MID-943	NED-TCTTCCTACCCCTGTTAGTG	GACAAGATCACTAGCTTGAC
MID-159	NED-ACCAGAGCACTACAGCCTTT	gCAAGGYAGTAACAATGAGGG
MID-2005	NED-TGTAGCGGCAATATAGGCAG	GAAAGTTGTGGCTTAACTGG
MID-250	NED-ATGGAGCAGTAAAGCAGCAC	GTCACTTTGGTTTTTGCAGG
MID-1802	NED-ACGGTCAACTTTGTAGCTCC	gCCAGTTGAGAATCACTGCAC
MID-1607	NED-TGTTGCAGAAGAACTCAACC	GATAAGCACCTAACTCCCAG
MID-1734	PET-TTCGTGTTCTCACACTGTCC	GTGCATCCCATACAACTGAC
MID-406	PET-TGGCTGCTGTAGATTGTAGG	gACAAATGGACAACGGCCAAG
MID-1386	PET-AGAGGATCATGGAGACCAAC	gtttATGTTCCAAGTCAGCAGCAC
MID-1726	PET-GGTCCAAATGCACCACAATC	GCTCTGCTATTTTGGTTTGC
MID-3626	PET-TGTTGGTTCTCTCCTTTTCC	GGTGACCCCTTCTTTATCTC
MID-360	PET-AGATCAACTGCCAATCTGGG	gCTCAAGTGACCAACCCACCT
MID-1603	PET-TTACAATTTCAAGCCTCCGC	GGAGCTGTTAGTCTGAGTAG
MID-2719	PET-GTCAGGAGTCTAGAAACTTC	GGGTGATGAAATGTTCCGAA

The work procedure of indel genotyping consists of a single multiplex-PCR followed by capillary electrophoresis (CE). All 46 markers generate short fragments of less than 230 and are biallelic except MID360 and MID2264. The third sequence length variant of MID360 results from a T insertion in the short allele and is apparently specific to African individuals whereas the third allele of MID2264 arose from a T deletion in the long allele and represented seemingly specific in European individuals [269].

5.4.1.2 PCR

The 46-mulitplex PCR amplifications were performed in the thermocyclers GeneAmp® PCR System 2700, 9700 or in the 2720 Thermal Cycler. Each sample was treated with the same volumes of reagents and PCR condition.

PCR Reagents	Volume
2x Quiagen Multiplex PCR Master Mix	5µl
10x Primer Mix	1µl
DNA	0.5µl
H_2O	3.5µl
Final Volume	**10µl**

Temperature	Time	Cycles
95˚C	15min	1x
94˚C	30sec	28x
60˚C	90sec	
72˚C	60sec	
72˚C	60min	1x

5.4.1.3 Capillary Electrophoresis (CE)

Following PCR amplification 11.5µl Hi-Di Formamide (Applied Biosystems) and 0.3µl Liz-500 Size Standard (Applied Biosystems) were added to 0.8µl PCR product and separated according to size on the 3130 Genetic Analyzer (Applied Biosystems). The software GeneMapper (Applied Biosystems) was used to analyze the results.

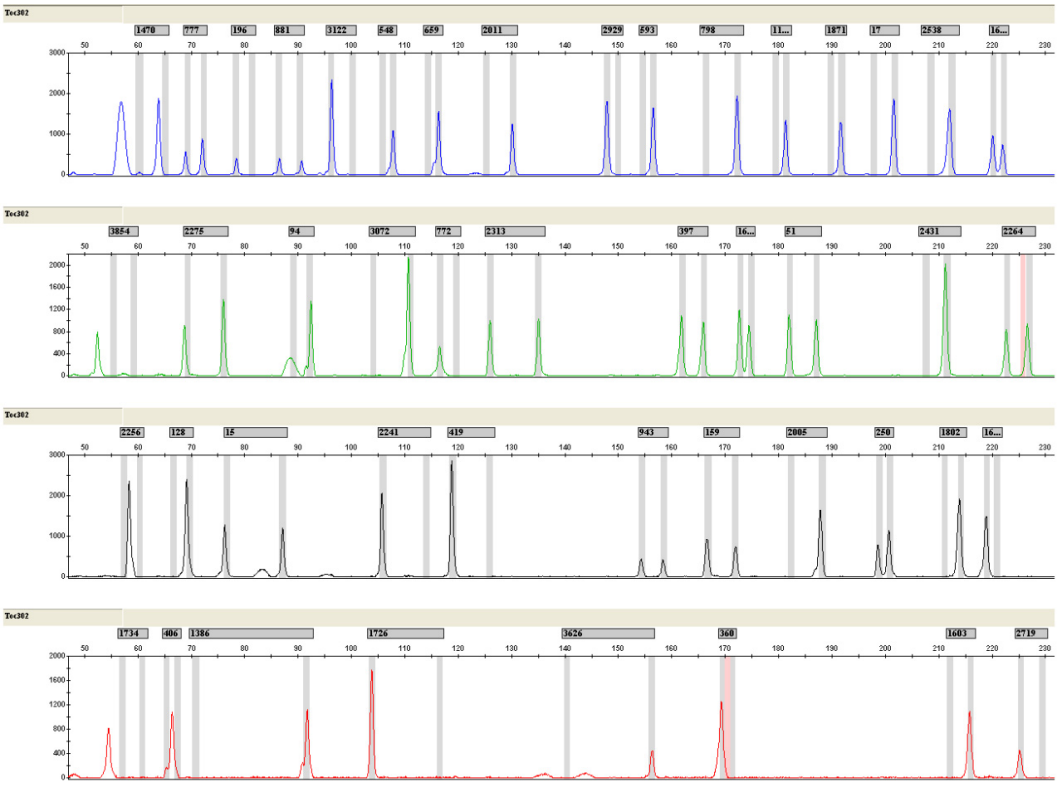


Figure 63: Length size separation of AIM-Indelplex fragments.

5.4.1.4 Statistical analyses

Population structure to infer continental ancestry of the Bolivian samples was analyzed with the software STRUCTURE 2.3.4 [272] together with CLUMPP 1.1.2 [325] and distruct 1.1 [326] for visualization, but also with the software ADMIXTURE v. 1.22 [273].

For STRUCTURE burn-in and Markov Chain Monte Carlo (MCMC) repetition were set to a length of 100,000 and the parameters Admixture Model with LOCPRIOR (use sampling location as prior) and Allele Frequency Independent were selected. The reference samples from the Human Genome Diversity Cell Line Panel, HGCP-CEPH [269, 292], were involved in the analyses to assist the clustering method of the software. Each run in STRUCTURE was repeated 5 times from K=2 to K=7. The online software Structure Harvester (http://taylor0.biology.ucla.edu/structureHarvester/) [327] was used to estimate the optimal value of K. CLUMPP was applied to prepare output files gained from Structure Havester for visualization with distruct.

The software STRUCTURE was later replaced by ADMIXTURE [273] which uses a maximum likelihood estimation of individual ancestries from multilocus genotypes. STRUCTURE was used

at the beginning to estimate individual continental ancestry in the Bolivian population. ADMIXTURE was used to estimate percentages of admixture in the Yungas, again to a continental level considering three main groups (sub-Saharan Africa, Native Americans, and Europeans) represented by 327 samples from the Human Genome Diversity Cell Line Pandel (HGDP-CEPH) which were also used in the STRUCTURE analysis.

The software R 3.1.1[299] and the package SNPassoc [328] were used to run two-dimensional Principal Component Analyses (PCA) in order to compare and evaluate the results from STRUCTURE and ADMIXTURE.

A simulation experiment has been designed in order to estimate the dependence of ancestry inferences on sample sizes. For each of the two Bolivian sets of genetic profiles, we randomly take 1000 sub-samples of variable sizes (from 5 to 70 profiles; in stepwise increments of 5 and taken without replacement). Thus, for example, we obtained 1000 sub-samples of size 5, also 1000 sub-samples of size 10, and so on until a maximum sample size of 70. For each of the sub-samples we computed ancestry proportions as indicated in the next section. Continental ancestry estimates were obtained as the mean values calculated from the 1000 sub-samples of each sample window (i.e. mean values for the 1000 sub-samples of size 5, and so on) and bootstrapping intervals were built accordingly (for each sample window). Simulations were carried out using R 3.1.1 [299]. Simulation analyses were conducted with the population from Chuquisaca (n=73) and the population from La Paz (n=105).

5.4.2 SNPs

5.4.2.1 Genotyping

For the genome-wide SNP analysis a subset of 9 samples from the Tocaña that showed high African ancestry based on the 46-AIM-INDEL panel was selected. Genotyping was conducted by using the Axiom® Genome-Wide Human Origins 1 Arrays in the Centro National de Genotipado (CEGEN) of Santiago de Compostela, Spain. As the 9 samples did not contain enough DNA and/or meet the required quality for genome-wide SNP genotyping, only 2 samples were further analyzed for 580,144 SNPs in total.

5.4.2.2 Statistical analyses

Initially, the SNP data of the two Tocañas were intersected with genome population data that were downloaded and processed from various SNP repositories (Appendix, chapter 12.2). Three main sets of data were used for the comparative analyses. At first, the SNP profile of the two 'Afro-Bolivians' were compared with data from the 1000 Genomes Project (1000G) based on 580,144 SNPs (1.7% of missing data), to determine main continental allocation. The 1000G data was retrieved from the original repository as it is described in the article by Pardo-Seco and colleagues [330]. In the second analysis, the two Tocaña individuals were compared with a large data set comprising 1.161 Africans (including 57 different populations/ethnic groups), this time, based on 25,192 SNPs (0.7% missing value). The Native American component of the two 'Afro-Bolivians' was further analyzed by using the unmasked (0.2% missing data) and masked (7.8% missing data) dataset from Reich and colleagues [331].

The software PLINK [329] was used in order to determine identiy-by-state (IBS) values from the SNP data. The two profiles were checked for close familial relationships as it was done in Pardo-Seco and colleagues [330]. Individuals showing identical profiles or close relationships were disregarded from the analysis. Furthermore, we did not detect close relationships between the two Tocaña analyzed.

Several PCAs were carried out in order to separate clusters of genetic variation in the population sets analyzed. The analyses were based on a matrix of pairwise individual IBS. The PCAs were conducted by using the function *cmdscale* (library *stats*) from R (http://www.r-project.org).

Individual admixture of the two Tocaña individuals was determined by using the software ADMIXUTRE which is based on a maximum likelihood estimation of individual ancestries from multilocus SNP data [273].

The software PCAdmix 1.0 [332] was used to assign ancestry-specific haplotypes across the genome in order to determine local ancestry. For this analysis three pre-defined ancestral groups were integrated: Yoruba (representing African ancestry), CEU (representing European ancestry), and Guahibo, Maya, and Quechua (representing Native American ancestry). Genomic segments were assigned to the three possible ancestries by following a posterior threshold of 0.8.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

6 Results

6.1 Mitochondrial DNA

6.1.1 MtDNA analysis of Bolivian populations

6.1.1.1 Mitochondrial DNA diversity

Diverstiy indices were computed by categorizing the Bolivian population in departments, rural vs. urban areas, or geographic regions (Andean, Sub-Andean, Llanos). Values of haplotype diversity among departments are more or less similar with Pando showing the highest and La Paz the lowest values (Table 13). Nucleotide diversity, in contrast, was highest in Santa Cruz and lowest in Chuquisaca. Mean number of pairwise difference was highest again in Santa Cruz, but this time lowest in La Paz. According to these observations, there seems to be no clear correlation between the populations of the departments and the molecular diversity indices, but when examining the rural versus the urban population, the rural population exhibits higher haplotype and nucleotide diversity values and a higher mean number of pairwise difference than the urban population. Among geographic regions the Llanos showed the highest molecular diversity values followed by the Sub-Andean, and subsequently, the Andean region.Therefore, the molecular diversity seems to increase longitudinally from the high mountains in the Andeans via the Sub-Andeans to the lowlands of the Llanos.

Molecular diversity values of the samples used in this study were compared with previously published data (Table 13). Some ethnic groups apparently show particularly low diversity. For example, the Ayora from Bolivia and Paraguay, and a population of Aymara- and Quechua-speakers from Beni exhibited extremely low diversity values when compared to other Bolivian populations and all Native Bolivians. The low molecular diversity among the Ayora population is most probably due to a strong founder effect as it was discussed by Dornelles and colleagues [283] whereas the low diversity values of the Aymara- and Quechua-speaking population from Beni is apparently the result of a small sample size [286] as higher molecular diversity values were observed in a larger sample size of an Aymara-speaking population [281]. The diversity values of the population from Beni in this thesis reached higher values than populations from the same department published in previous studies like the Aymara- and Quechua-speaking population, Moseten, and Chimane [286] and the Llanos populations analyzed by Bert and colleagues [285].

Table 13: Diversity indices in Bolivian mtDNAs and main American and African regions. Published in [333], Table 1, doi:10.1371/journal.pone.0058980.t001.

Populations groups	Reference	n	*k*	*S*	*H*	*π*	M
Department							
Beni	Present thesis	102	63	71	0.982±0.002	0.02470±0.00130	6.818
Chuquisaca	Present thesis	89	57	59	0.971±0.009	0.01588±0.00099	5.718
Cochabamba	Present thesis	103	71	77	0.981±0.007	0.01855±0.00086	6.677
La Paz	Present thesis	253	121	81	0.960±0.007	0.01864±0.00086	5.106
Pando	Present thesis	90	54	60	0.983±0.006	0.01977±0.00068	7.117
Santa Cruz	Present thesis	83	49	46	0.978±0.006	0.02618±0.00097	7.226
All Bolivia	Present thesis	720	306	134	0.976±0.002	0.02237±0.00041	6.130
Rural *vs.* Urban							
Rural	Present thesis	189	108	92	0.980±0.004	0.02440±0.00069	6.709
Urban	Present thesis	531	240	121	0.975±0.003	0.02225±0.00049	6.098
Regions							
Andean	Present thesis	240	116	78	0.961±0.007	0.01856±0.00088	5.086
Sub-Andean	Present thesis	204	118	90	0.973±0.006	0.02215±0.00083	6.092
Llanos	Present thesis	276	141	100	0.982±0.003	0.02458±0.00059	6.759
Native groups							
Aymara-speakers (Beni)	[286]	10	5	9	0.667±0.163	0.00499±0.00251	1.800
Moseten (Beni)	[286]	10	8	18	0.956±0.059	0.01761±0.00209	6.356
Quechua-speakers (Beni)	[286]	16	7	17	0.692±0.124	0.01094±0.00358	3.950
Chimane (Beni)	[286]	10	8	16	0.933±0.077	0.01422±0.00212	5.133
Aymara-speakers	[281]	96	39	48	0.956±0.009	0.01645±0.00142	4.523
Ayoreo	[283]	91	8	10	0.473±0.061	0.00626±0.00104	2.260
Ignaciano	[285]	15	11	23	0.933±0.054	0.01773±0.00237	6.400
Movina	[285]	12	8	12	0.894±0.078	0.00814±0.00174	2.940
Quechua-speakers	[281]	93	40	48	0.946±0.012	0.01997±0.00118	5.471
Trinitario	[285]	12	11	22	0.985±0.040	0.01884±0.00212	6.803
Yuracare	[285]	15	11	22	0.952±0.040	0.01852±0.00151	6.686
All Native		380	114	84	0.946±0.006	0.02201±0.00043	6.008

NOTE: *n* = sample size; *K* = number of different haplotypes; *S* = number of segregating sites; *H*= haplotype diversity; *π* = nucleotide diversity; *M* = average number of pairwise differences (mismatch observed mean).

Consistent with the high mtDNA diversity in the Bolivian population only few mtDNA haplotypes occur more than once (Figure 64). In Pando and La Paz singletons of haplotypes were observed four times. Few haplotypes occurred three times in all departments except in Cochabamba and some twice in all departments.

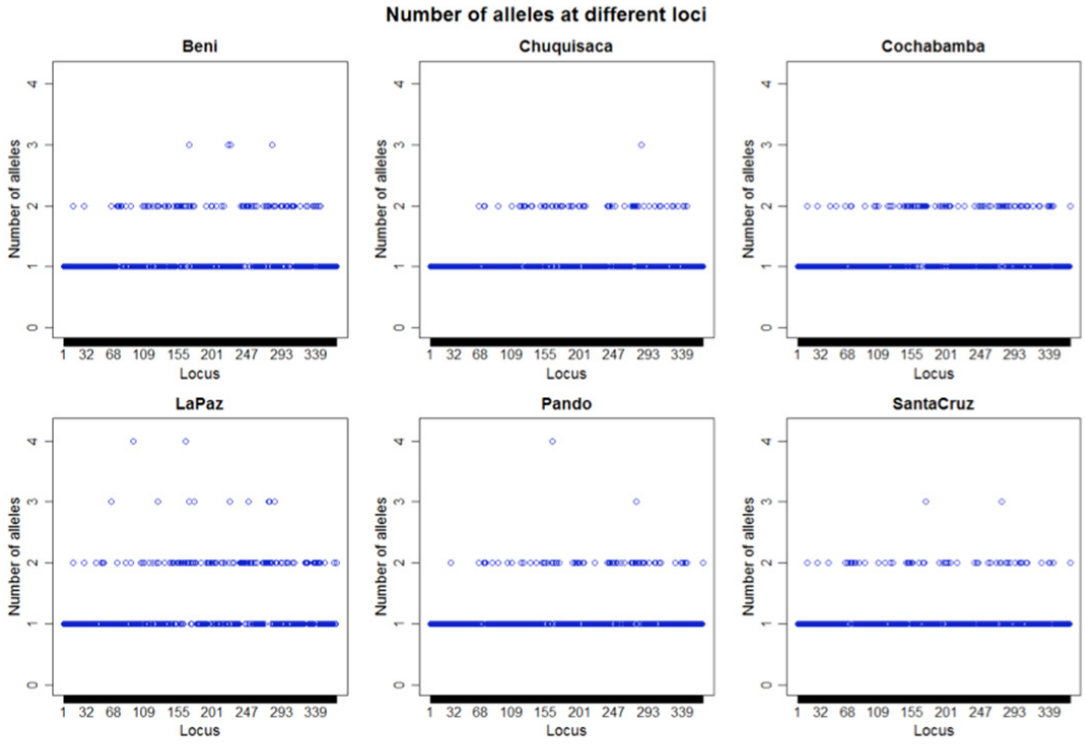


Figure 64: Number of alleles at different loci within departments. Published in [333], Figure S1.

Pairwise F_{ST} values were observed to be highest between Santa Cruz and other departments (Figure 65). Correspondingly, the Llanos, which include Santa Cruz, showed highest F_{ST} values with the two other regions, whereas the F_{ST} value between the Sub-Andean and the Andean was lowest.

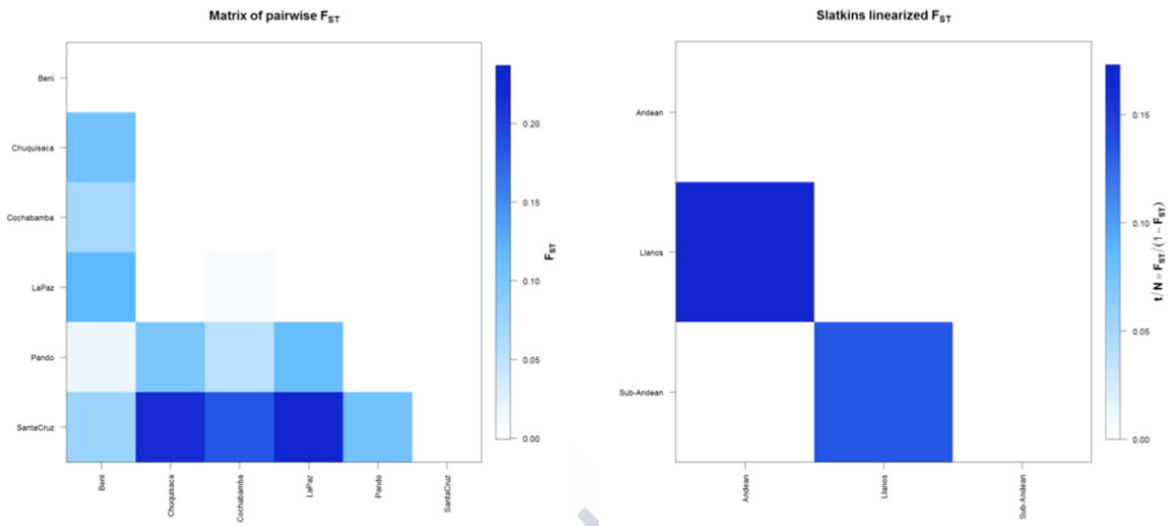


Figure 65: Pairwise F_{ST} values between departments and regions. Published in [333], Figure S3 and S4.

The expected (virtual) heterozygosities both in the main departments (Figure 66) and in the geographic regions (Figure 67), respectively, indicate the existence of substantial mtDNA diversity due to the dispersion of values.

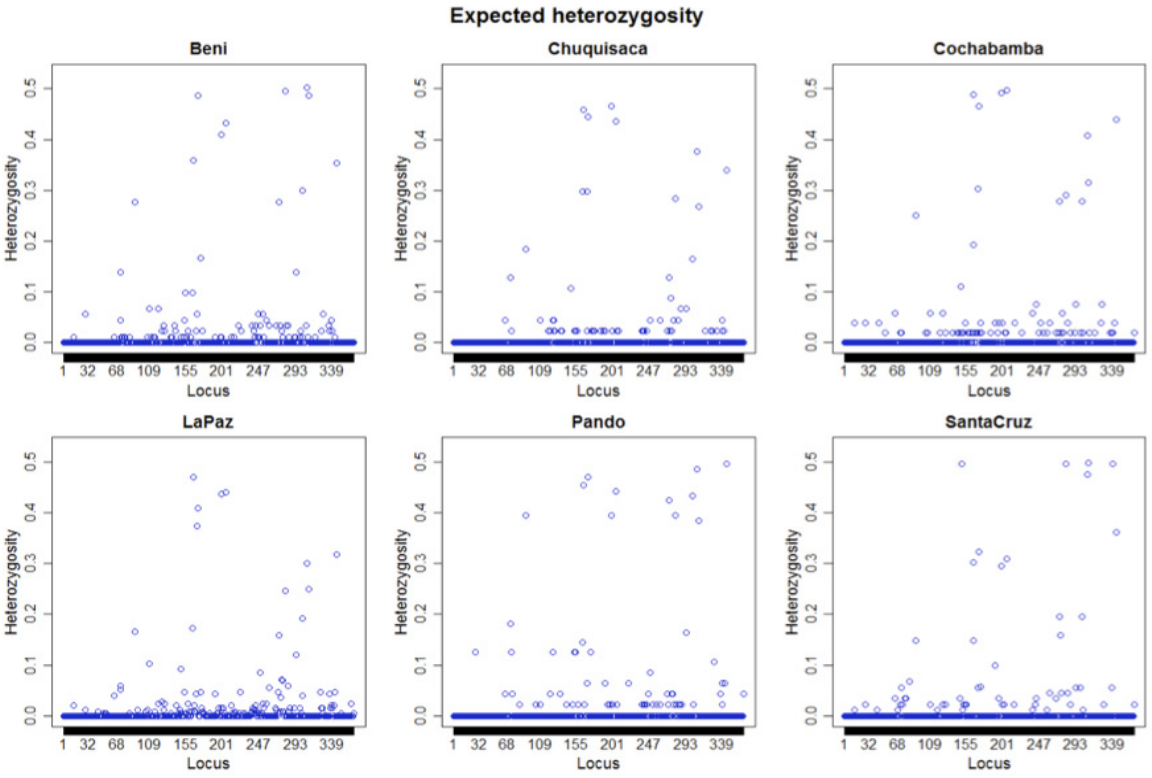


Figure 66: Expected (virtual) heterozygosity by departments. Published in [333], Figure S5.

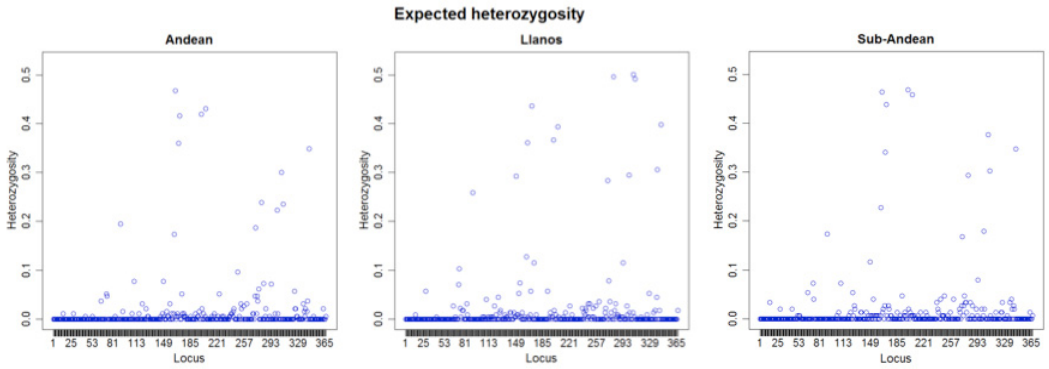


Figure 67: Expected (virtual) heterozygosity by main ecological regions. Published in [333], Figure S6.

6.1.1.2 Phylogeography of control region mtDNAs

In most individuals of the Bolivian population a Native American component predominates in the mitochondrial DNA (98.4%). Consequently, most Bolivians can be classified in one of a Native American haplogroup (A: 14.3%, B: 52.6%, C: 21.9%, D: 9.6%) (Figure 68).

However, haplogroup frequencies among Bolivian populations vary remarkably among departments. For example, in La Paz most individuals showed haplogroup B (71%) followed by C (12%), A (10%), and D (7%). Similar was found in a smaller sample size from La Paz where B was again the most frequent haplogroup [284]. Individuals of Beni, in contrast, could be firstly classified to haplogroup C (48%), and only secondly, to haplogroup B (30%). Haplogroup A (12%) and D (10%) in Beni were like in La Paz least common. The most important pattern of haplogroup distribution in Bolivia was the difference between Andean and Sub-Andean versus the Llanos regions. For instance, B is the most frequent haplogroup in the Andean (71%) and Sub-Andean (61%) region and decreases in the Llanos (32%) whereas the frequency of haplogroup C reaches its highest distribution in the Llanos (35%) followed by the Sub-Andean (18%) and the Andean (12%).

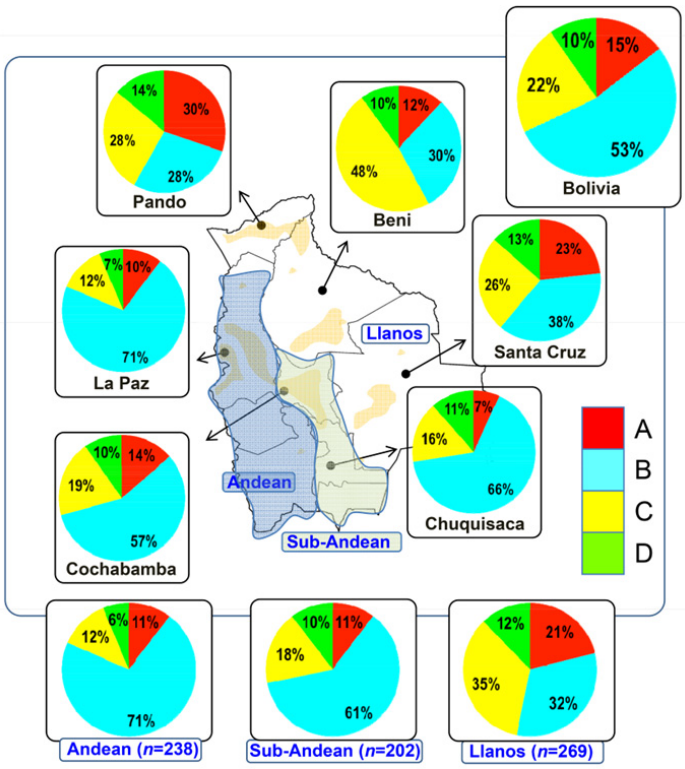


Figure 68: Map of Bolivia showing the location of the samples collected in the present study. The pie charts represent the distribution of basal Native American haplogroup frequencies in the region. Published in [333], Figure 1, doi:10.1371/journal.pone.0058980.g001.

As it was expected, Bolivian populations share more mtDNA haplotypes with South American than with North and Meso American population (Table 14), but due to the different population sizes the values have to be treated with care. However, only 68 (out of 383, 18%) haplotypes are shared with the largest population size, that is, with North America, but 187 (out of 383, 49%) with South America. Furthermore, as 48 haplotypes are shared between the three Americas most of the 68 shared haplotypes with North America are also shared with South America. Nonetheless, a high number of unique haplotypes (250 out of 383, 65%) were observed in Bolivia.

Table 14: Shared haplotypes between Bolivia and other American regions. Published in [333], Table 2, doi:10.1371/journal.pone.0058980.t002.

Population	N	SH with Bolivia	H	H_1
North America	7551	68	1875	1085
Meso America	2792	68	925	612
South America	5727	187	1587	1009
Bolivia	1127	383	383	250

Only the segment 16024 to 16385 was considered for comparison. N = sample size, SH = shared haplotypes, H = number of different haplotypes, H_1 = number of unique haplotypes per region.

There are eight individuals that most likely belong to haplogroup C1d as they all carry the characteristic transition at 16051. One mtDNA is most probable of haplogroup C1d2a as it did not show 16051, but a transition at 16209. C1d has been reported to be one of the founding Paleo-Indian mtDNA lineages due to its ancient coalescence time and the geographical distribution of the derived haplogroups or subhaplogroups from C1d in the Americas [334]. C1d did not necessarily follow a coastal route once it reached South America [175], and thus, would explain why this mtDNA lineage is more frequent in the Llanos than in the Andeans.

The five mtDNAs, which belong to D4h3a, all exhibited a transition at 16241. Three of them could be further classified into D4h3a4 as they had a transition at 16294. D4h3a has a similar coalescence time as C1d. However, D4h3a apparently followed a different route, and thus, probably spread along the Pacific coast due to its geographic distribution. In agreement, all of the five individuals of D4h3a indeed came from the Andean region. Haplogroup D4h3a4 was observed so far in Peruvians [175].

Consistent with previous studies X2a, which belongs to the founding Paleo-Indian haplogroups and is distributed in North America [175], was not found among individuals of Bolivia. Similarly, haplogroups D1g and D1j, which occur basically in populations from the Southern Cone of South America, that is in Chile and Argentina, were absent in Bolivia.

One Bolivian mtDNA of haplogroup C1b had a transition at 258, and thus, could be furtherclassified as C1b13. This mtDNA lineage probably arose in the Southern Cone and is most frequent in Paleo-Mapuche people of Chile and Argentina and only sporadically found outside of this region [177]. Saint Pierre and colleagues [177] proposed that South-American-specific sub-haplogroups like C1b13 might have originated locally later after the arrival of humans, and

consequently, represent informative markers to investigate migrations across the Andeans. Thus, C1b13 could have arrived in Bolivia from the South.

Some Bolivian mtDNAs are especially interesting, because they occur not only in Bolivia itself but also in neighbouring areas. For example, the motif 16189 16217 16290 was found six times within haplogroup B4 in different regions of Bolivia, that is, in Cochabamba, La Paz, Pando, and Chuquisaca, but also appeared in four Andean Peruvian Quechua individuals [335]. The motif 16173 16192 16223 16298 16325 16327 16346 was observed twelve times in Bolivian individuals from the Llanos that were classified to haplogroup C. However, this motif was also found in an individual from La Paz [284] and in one from the Chaco region of Northern Argentina [336].

The African-specific L haplogroups only accounted for 1% of all mtDNAs within the Bolivian population. The proportion of African ancestry on the mtDNA in Bolivians was very low when compared to other South American populations that were more influenced by the TAST [17, 57, 90]. Seven individuals were observed to belong to a sub-Saharan L haplogroup. One individual (#BNI19, see Table S1 of [333]) was classified to haplogroup L0a1b2, which is spread all over sub-Saharan Africa and the Americas, although it carries a distinctive transition (within L0a1) at position 16271 that is not associated with the haplogroup. The haplogroups L1b (#BNI75) and L1c1a1 (#LaPaz467) might have come from West-Central Africa due to the Transatlantic Slave Tradeas these haplogroups are widely distributed in this region of Africa (see chapter 3.5.9.2). L1c3b1a (#CBBA195) and L3f1b4a (#Cobija595) have been observed so far in Cabinda, Angola and Mozambique [140, 144, 313, 337], but also in Gabon [161]. L2a1h (#LP6) was found so far in Kenya [134] and L3e2b (#Cobija625) in Burkina Faso [152], Guinea-Bissau [134], and in pygmies [151]. As expected [16, 17], the most likely origin of the Bolivian mtDNAs might be from West/West-Central, but also from Southeast Africa.

The European ancestry in the mtDNA of Bolivians was observed to be less than 1%. The three mtDNA sequences could be classifierd as H1af (#STACRUZ159), K (#Cobija612), and X1 (#Cobija572). An Asian origin of 16104 16233 (X1) cannot be ruled out, because there are two exact haplotype matches with individuals from China [338], but no match with European ones (compared with an in-house database of 26700 HVS-I profiles).

6.1.1.3 Phylogenetic tree of whole mtDNA genomes

Two mtDNAs of haplogroup A2, which showed the tandem variants T16097C A16098G, and six sequences, which carried the transition G16145A and belonged to haplogroup B2, were chosen for phylogenetic tree analysis based on whole mtDNA genomes.

The motif T16097C A16098G is rarely found in mtDNA databases. Within A2 this motif alone without any other control or coding region variants defines a novel sub-branch of the Native American phylogeny, that is, A2ah (Figure 69). A2ah most likely originated in Bolivia or in the surrounding areas due to its very restrictive local occurrence in Bolivia. Furthermore, both individuals of this haplogroup show variability within this clade, that is, one individual (#4=BNI86) shows a transition at T4634C and the other individual (#5=STACRUZ221) mutations at A12373G, C16260T, and C16286T. In public databases there was only one whole mtDNA sequence (JQ702082) [339] of haplogroup A2 that had the transition A16098G. However, the sequence lacked T16097C, and thus, is probably not of haplogroup A2ah. Furthermore, in comparison to the two Bolivian samples, the sequence lacked C64T, T16097C, and T16362C, but carried mutations at T466C, T14025C, T15670C, and G16106A. Nonetheless, haplogroup A2ah probably originated around 5,200 thousand years ago, kya (95% CI: 0.1-10.5). In a literature survey only nine mtDNA sequences matched the motif of haplogroup A2ah (based on the control region). Geographic origin is not known for four samples, but two HVS-I mtDNAs that carry the A2ah motif were found in 'Hispanic' individuals of North America in the SWGDAM database [340] and three A2ah mtDNAs were found in South America whereby two of them came from Brazil [84] and one was observed in the Toba from Gran Chaco (North Argentina) [341]. In Bolivia, A2ah was found four times (0.5%) (including the two whole genome mtDNAs), but exclusively in individuals from the Llanos (three in Santa Cruz and one in the Beni department). The motif T16097C A16098G of A2ah was also found in Mainland Scottland (Western Islands, Isla of Sky), and two other individuals in [342], however, none of the sequences carried the variations that characterize A2.

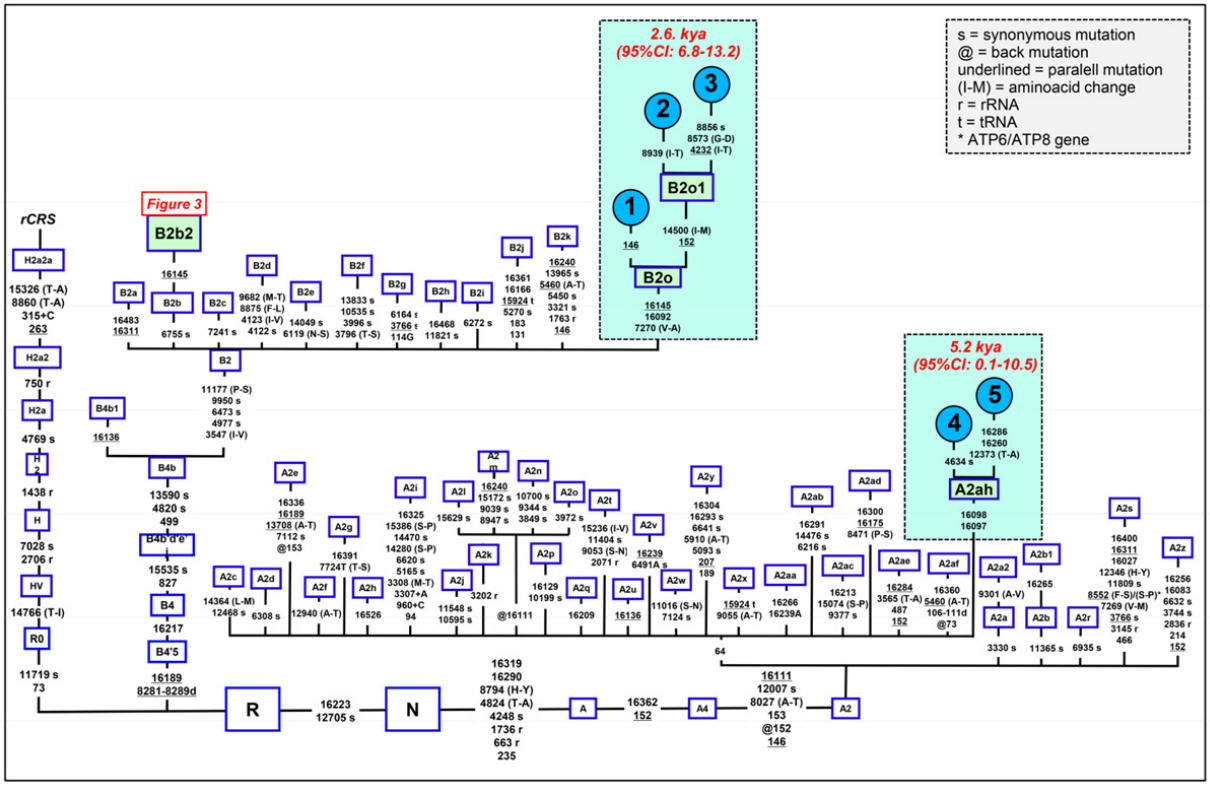


Figure 69: Maximum parsimony tree of the main branches characterizing haplogroups A2 and B2, indicating the new branches generated in the present thesis, namely, A2ah and B2o. Published in [333], Figure 2, doi:10.1371/journal.pone.0058980.g002.

All of the six individuals of haplogroup B2 carry a transition at G16145A. However, three of them (#1=COBIJA577, #2=Cobija604, #3=STACRUZ261) had non-synonymous transitions at T7270C and T16092C but lacked G6755A (Figure 69) which was observed in the other three individuals (#6=BNI90, #7=STACRUZ258, #8=CBBA110) (Figure 70). As G16145A apparently has a greater mutation rate than G6755A (22 versus 2 mutational hits inSoares and colleagues [26], 24 versus 5 in Phylotree [132]), the phylogeny was resolved like it is shown in Figure 69 and Figure 70. Therefore, the motif T7270C T16092C G16145A define the haplogroup B2o and T152CA14500G haplogroup B2o1 (see [333], according to phylotree Build 16: B2o: T16092C; B2o1: T7270CG16145A; B2o1a: T152CA14500G [132]). The three individuals of B2o show variability among their sequences. A transition at T146C was found in the first sequence (#1=Cobija577), another transition at T8939C in the second sequence (#2=Cobija604), and three more transition at T4232C, G8573A, and G8856A in the third sequence (#3=STACRUZ261). So far, no whole mtDNA genome of haplogroup B2o was observed in public databases. Nonetheless, B2o (based on

T7270C T16092C G16145A, equivalent to B2o1) probably originated around 2.6 kya (95% CI: 6.8-13.2). MtDNA control region sequences that carry T7270C, T16092C, and G16145A were found in only two individuals in public databases, that is, in one individual of San Martin de Pangoa in Perú, which is a small town in the Eastern Andeans and inhabited by Quechua and Nmatsiguenga people [343] and in one individual of Guam, an island located in the western Pacific Ocean and of U.S. territory. The mtDNA sequence of Guam, however, exhibits the motif T16189C T16217C A16247G C16261T, which is principally common in Micronesia, Australia, etc. [344, 345] and probably forms a different clade within B4a of the worldwide phylogeny. In Bolivia, B2o was observed seven times (1%) whereby most of them came from the Llanos (five from Pando and one from Santa Cruz) and only one individual came from the Sub-Andean region La Paz. Consequently, the B2o lineage most probably originated around the Andean or nearby regions.

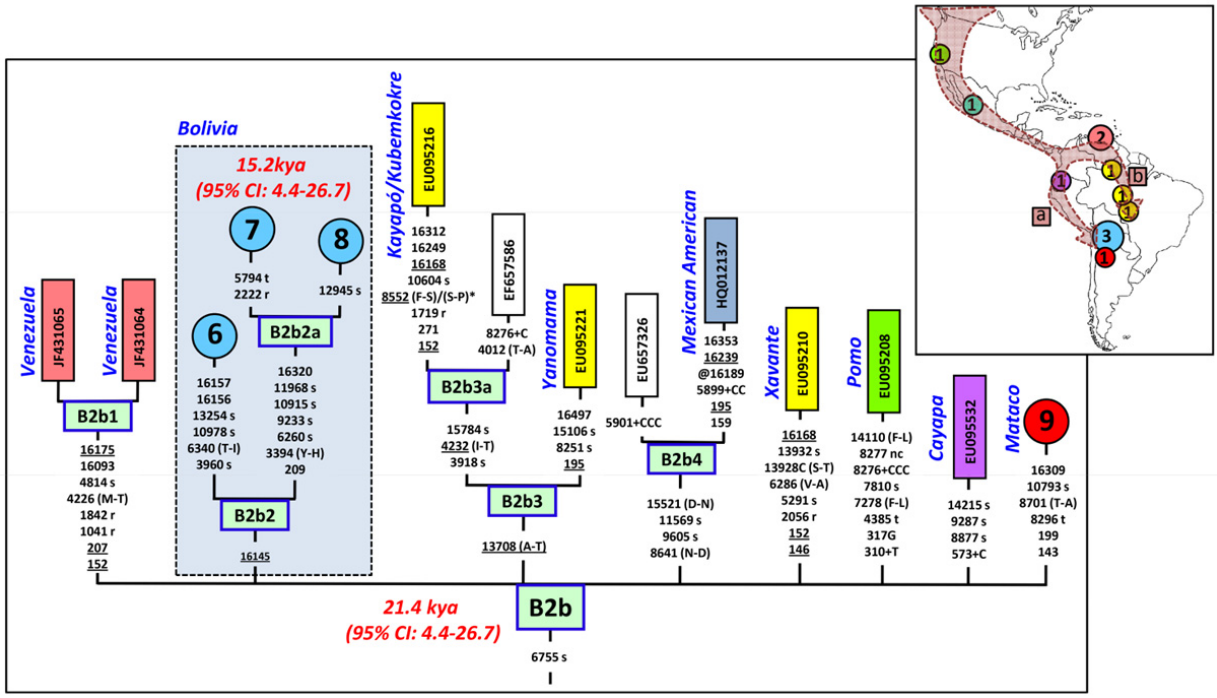


Figure 70: Maximum parsimony tree of haplogroup B2b. Published in [333], Figure 3, doi:10.1371/journal.pone.0058980.g003.

The three Bolivian whole mtDNA sequences that share a transition at G6755A could be classified as B2b as it was previously described [172, 179]. However, the internal phylogeny of B2b was not resolved in more detail so far. By searching databases the transition was found in ten other mtDNAs of the same lineage. All three Bolivian samples that belong to B2b share a transition at

G16145Awhich is now named B2b2 (Figure 70). Two of the individuals (#7=STACRUZ258, #8=CBBA110) further share transitions at T209C, T3394C, G6260A, T9233C, T10915C, A11968G, and C16320T, and thus, form a novel sub-branch, that is, B2b2a. The phylogeny of B2b comprises 11 more mtDNA sequences in addition to the three Bolivian samples: one Pomo (North Carolina, USA) [346], one Mexican-American [174], two Venezuelans from Pueblo Llano [347], one Yanomama (Brazil, Venezuela) [346], one Cayapa (Equador) [172], one Kayapó/Kubemkokre (South Amazonia, Brazil) [346], one Xavante (Brazil) [346], two from North Americans [339], and one whole mtDNA genome of a Mataco Native from North Argentina (#9=Mco-10) from the in-house DNA databank. In comparison to B2o, B2b appeared earlier and is estimated to have been originated around 21.4 kya (95% CI: 4.4-26.7). B2b2 arose probably around 15.2 kya (95% CI: 4.4-26.7). Based on the HVS-l control region B2b was observed 14 times in the HVS-I database. In Bolivia 7 individuals were found to belong to haplogroup B2b, that is, one Andean (La Paz), four Sub-Andean (Cochabamba, La Paz, and Chuquisaca), and two from the Llanos (Beni and Santa Cruz).

6.1.1.4 AMOVA analysis

An AMOVA was conducted by comparing molecular variance between departments, regions, provinces, and between rural vs. urban populations (Table 15). Irrespective of population sub-division, within-population variation exceeds substantially among-population variance ranging between 93.15 in populations of different regions to 99.85 in urban vs. rural populations. The highest values of among-population variance were reached when dividing the populations in regions and departments (6.85 and 6.56, respectively). Categorizing the Bolivian population into provinces does not increase among-population variance. Among rural vs. urban populations a minimal variance could be detected (0.15).

Table 15: AMOVA computed based on haplotype pairwise differences of Bolivian populations (significant tests: 20,022 permutations, adjusted P-value<0.0000). Published in [333], Table 3, doi:10.1371/journal.pone.0058980.t003.

	Within populations	**Among populations**
Rural vs. Urban	99.85	0.15
Andean vs. Sub-Andeanvs. Llanos	93.15	6.85
Departments	93.44	6.56
Provinces	93.60	6.40

6.1.2 MtDNA analysis of the Yungas populations

6.1.2.1 Mitochondrial DNA diversity

Haplotype diversity seems to be lower in the Yungas when compared to the remaining Bolivian population (Table 16) possibly due to the isolation of the Yungas valley. However, Tocaña seems to contribute slightly more to haplotype diversity of the Yungas than the residual population. Nucleotide diversity is highest in the Tocaña within the Yungas which can be attributed to the African-specific mtDNAs in this 'Afro-Bolivian' community. Nucleotide diversity decreases substantially in the Yungas when Tocaña is not included in the calculations. Consequently, average number of pairwise nucleotide differences again is highest in the Tocaña as both measures are directly correlated.

Table 16: Mitochondiral DNA diversity of the Yungas including Tocaña. The indices were computed by considering the HVS-I region from position 16090 to 16365. Submitted to PLos ONE.

Population	*N*	*k*	*S*	*H*	*π*	*M*
Yungas	105	52	59	0.943±0.016	0.02364±0.00168	6.500
Yungas without Tocaña	86	44	54	0.920±0.024	0.01895±0.00161	5.212
Tocaña	19	11	26	0.924±0.037	0.02990±0.00230	8.222
All Bolivians	720	306	134	0.976±0.002	0.02237±0.00041	6.130

N = sample size; k = number of haplotypes; S = number of segregating size; H = Haplotype diversity; π = nucleotide diversity; M = average number of pairwise differences (mismatch observed mean).

6.1.2.2 Phylogeography of control region mtDNAs

About 81% of all haplogroups found in the Yungas are of Native American origin while 18% can be attributed to typical sub-Saharan haplogroups. Only 1% is of European origin. However, Native American and sub-Saharan ancestry on the mtDNA are unequally distributed among the Nor and Sud Yungas. That is, in the Sud Yungas only Native American haplogroups could be observed with C1 being the most frequent haplogroup (50%) followed by B2/B4 (37%), and D1 (13%) (Figure 71). In the Nor Yungas, in contrast, the Native American comonent in the mtDNA reached 79% with B2/B4 as the most numerous haplogroup among the individuals (61%), and C1 as the second (13%), A as the third (3%), and D1/D4 as the fourth (2%). One haplotype among D in the Nor Yungas belonged to haplogroup D4h3a, a clade that is principally found in South America along the

Pacific coast [175]. Only one haplogroup of the Nor Yungas seems to be of European origin (haplogroup HV0).

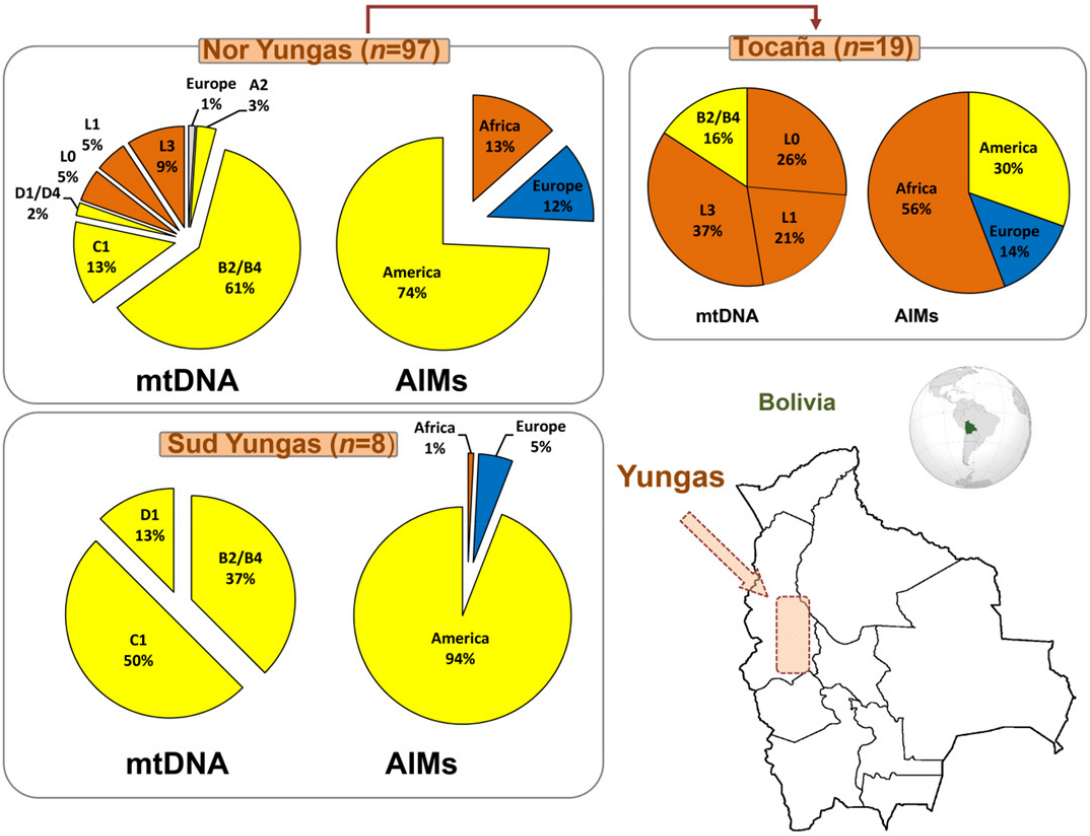


Figure 71: Proportions of continental ancestries on the mtDNA compared with autosomal AIMs (see chapter 6.3.1.2). Percentages of the Nor Yungas sum to 99% (instead of 100%) due to rounding. Submitted to PLoS ONE.

The most common Native American motif found among the individuals of the Yungas is the HVS-I motif of B4 (T16189C-T16217C) whereas 24% of all haplotypes furthermore showed a transition at C16188T. This motif is also very common in the remaining Bolivian population (see chapter 6.1.1.2)and was previously observed in Bolivians inanother study [286] and in other Andean locations like the Uros (Peru) [335], Jujuy (Argentina) [348], and Coyas (Argentina) [295, 348]. As the transition at C16188T is relatively stable (compare Soares and colleagues [26] and Phylotree Build 16 [132]), this motif might represent a haplogroup, or lineage, respectively, that has evolved locally in the Andeans. Furthermore, due to its spatial distribution in the Andeans, this haplogroup could signal trans-Andean movements.

Sub-Saharan L-haplogroups were absent in the Sud Yungas, but 19% of the haplotypes in the Nor Yungas could be classified to different L-haplogroups like L0 (5%), L1 (5%), and L3 (9%). The majority of the L-haplogroups, however, were observed in the community of Tocaña (84% of all individuals from Tocaña). The most common L-haplogroup in the Yungas was L1c3b1a (26% of all L-haplogroups).

European ancestry on the mtDNA in the Yungas was even lower. Only in one individual of the Nor Yungas a European-specific haplogroup was observed.

The lower haplotype diversity in the Yungas in general when compared to the residual Bolivian population can be explained by the fact that singleton haplotypes are less common in the Yungas. Consequently, many haplotypes occur at least twice in the Yungas not only with respect to the Native American haplotypes, but also to those of recent African ancestry. The lower haplotype diversity could be the result of the relatively isolated Yungas valleys within Bolivia.

6.1.2.3 Phylogenetic tree of whole mtDNA genomes

The 8 individuals from Tocaña that were analyzed for the whole mtDNA genome belonged to L0a1b2 (#Toc289, #Toc290), L0a2a2a (#Toc305), L1c3b1a (#Toc293, #Toc295, #Toc304), L3d1a1a (#Toc299), and L3d1b3 (#Toc294). According to the phylogenetic trees, geographic origin cannot be allocated to one single location in Africa although some tentative geographic inferences can be done.

The sample set of the mitogenomes for the phylogeographic analysis comprised 2235 samples that belonged to an L-haplogroup. Most samples were from African individuals (n=1726). The distribution of all African samples used for the analysis is shown in Figure 72. Eventually 152 samples were used for the analysis.

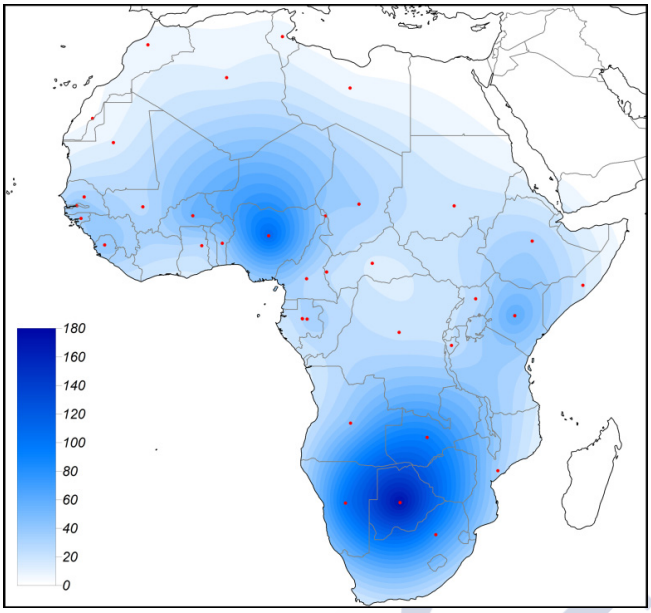


Figure 72: Map showing the location and sample size of the African mitogenomes used for phylogeographic analysis. Submitted to PLos ONE.

The two Tocañas #Toc289 and #Toc290 belonged to haplogroup L0a1b2 like three individuals from the USA and one from Saudi-Arabia that could be further classifiedto haplogroup L0a1b2a (Figure 73). Both mitogenomes of the two Tocañas built a sub-branch with the three mitogenomes from the USA which is based on @152. However, two individuals from the USA have on top the variants T7711C and G9804A, and thus, determine a new sub-branch (L0a1b2b) as more than one mutation, that is not a hot-spot [26] are shared between at least two individuals. The two 'Afro-Bolivians', in contrast, share the back mutation @16093 with one of the other USA samples. According to the most immediate ancestral node, that is L0a1b, it is not possible to allocate the members to a specific region in Africa. Nonetheless, America seems to be well represented in L0a1b (~37%) as do East Africa and the Middle East. In the next immediate ancestral node L0a1 East Africa and Middle East again is frequently seen. Some branches are also well represented in South Africa (L0a1b1a1) and the Middle East (L0a1d) while in other branches singleton sequences from Southeastern Africa (Mozambique) and North Africa (Morocco) appear. Central Africa (Chad) also seems to occur infrequently within L0a1.

Individual #Toc305 belongs to haplogroup L0a2a2a within a sub-branch that is determined by the reversion @95C although this mutation is highly unstable and uncertain based on current available data (see L0a'b) (Figure 73). Four samples from Southeastern Africa (Zambia, Mozambique), three

samples from the Middle East (Oman, Yemen, Saudi-Arabia), three from the USA, two from East Africa (Kenya), and one from the Dominican Republic as well match L0a2a2a whereas three of the Southeastern samples (Zambia) form a sub-branch (based on mutation T10361C and T16093C). The 'Afro-American' (36%) component was most represented within L0a2a2a, followed by Southeastern Africa (29%). The sub-Saharan African region that is most represented within the next immediate ancestral node, that is, L0a2a, is again Southeast Africa followed by East Africa and the Middle East.

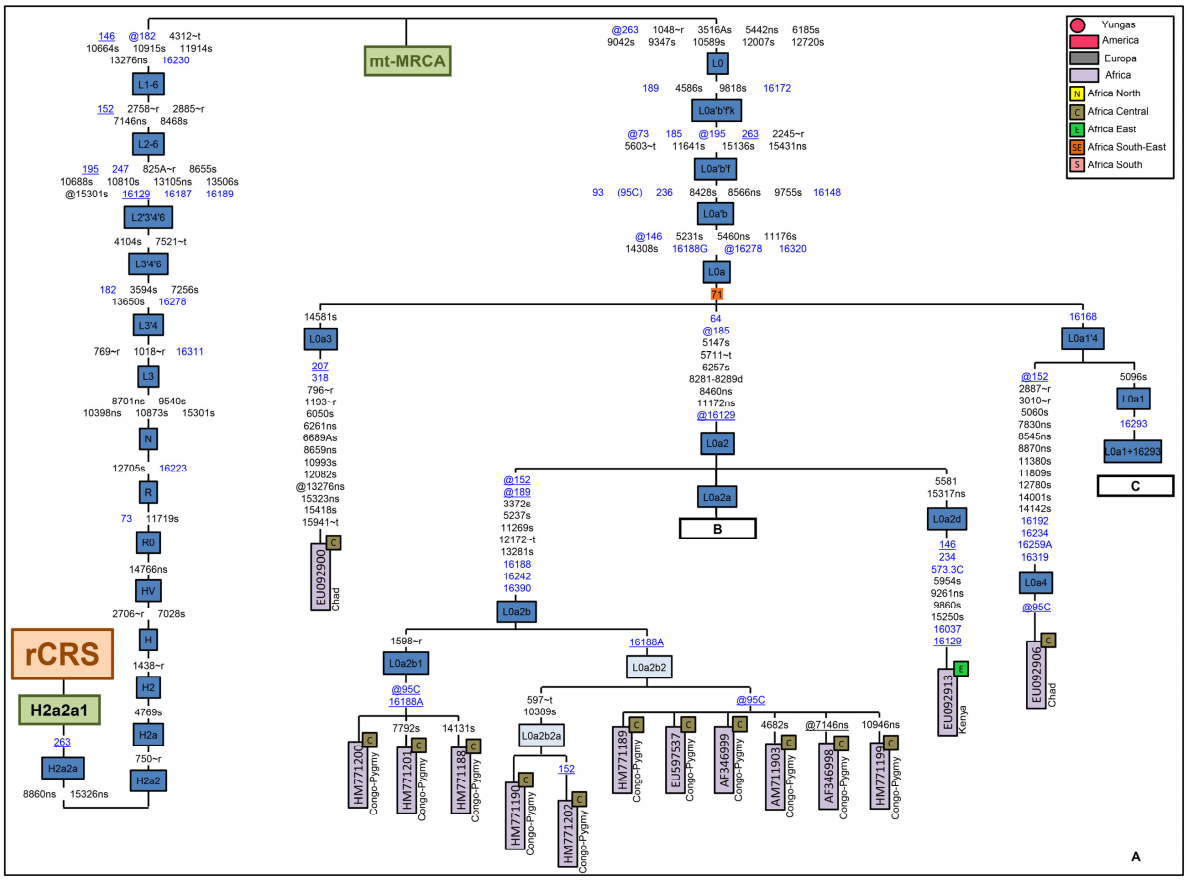


Figura 73 A: Phylogenetic tree L0a. Submitted to PLoS ONE.

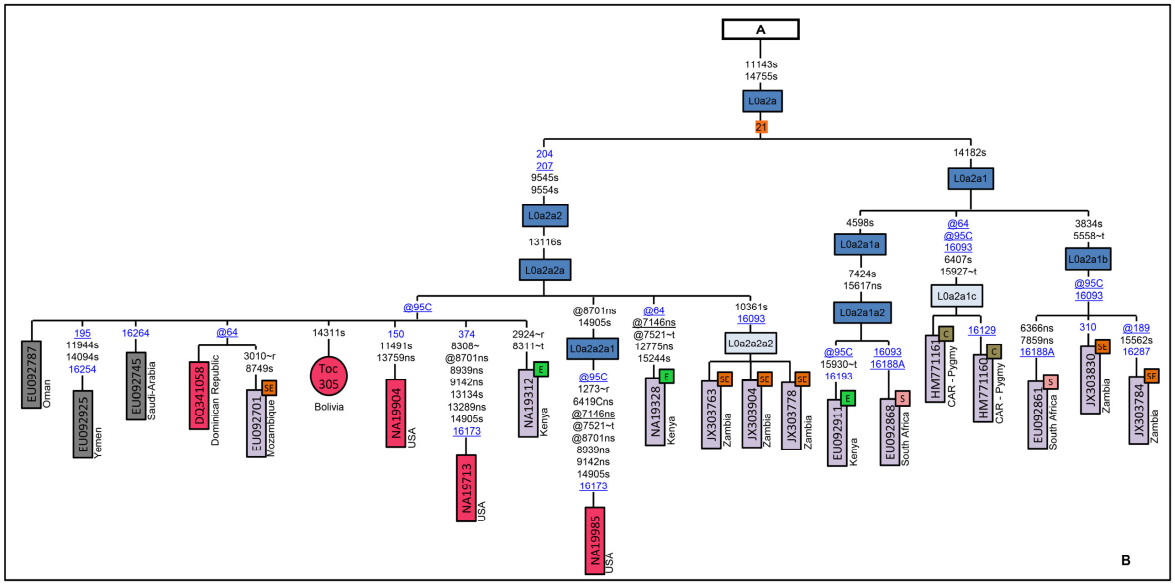


Figura 73 B: Phylogenetic tree L0a. Submitted to PLoS ONE.

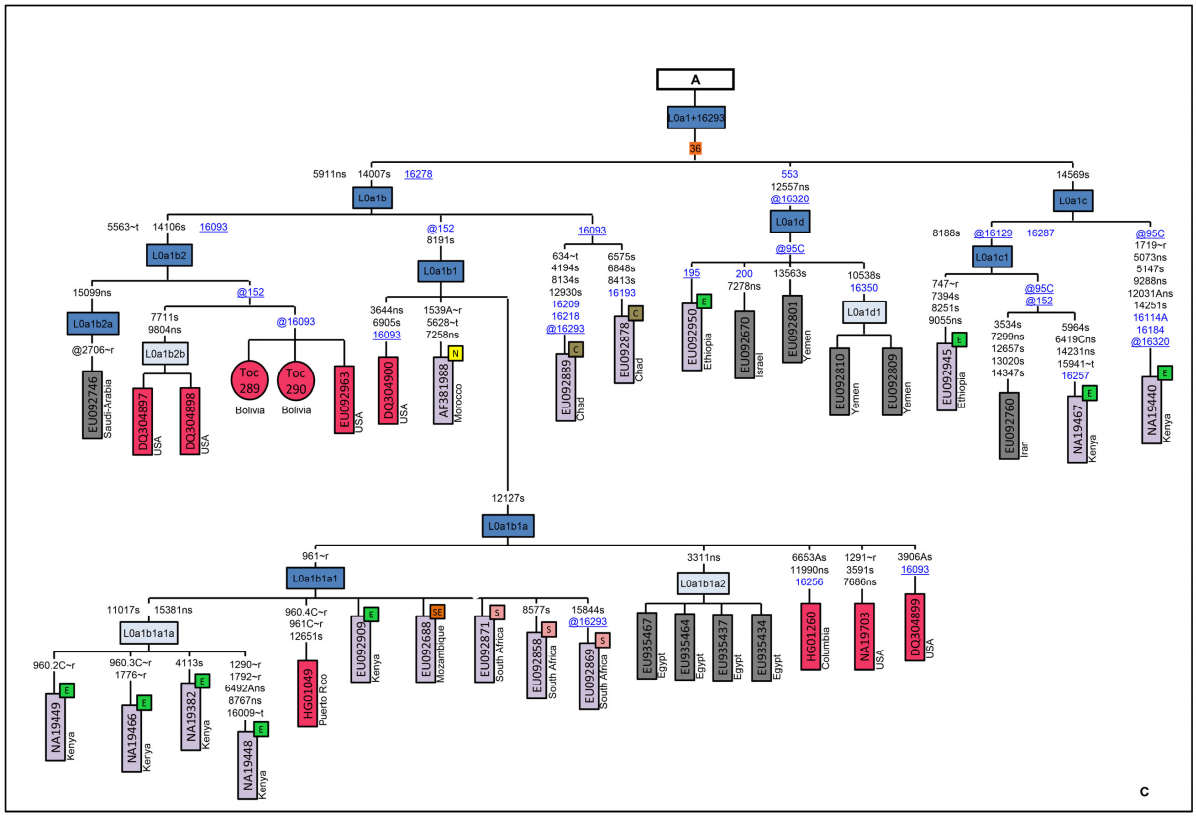


Figura 73 C: Phylogenetic tree L0a. Submitted to PLoS ONE.

The three ‘Afro-Bolivians’ #Toc293, #Toc295, and #Toc304 match L1c3b1a together with one mitogenome each from Kenya, Zambia, Morocco, a Bantu-speaking Fang, and Pygmy (Figure 74). Apart from the individuals from Tocaña, only one more sample from the Americas, that is, from Puerto Rico, appears in this clade. Nonetheless, the individuals from America are still well represented within L1c3b1a (44%). The next immediate ancestral node L1c3b1 comprises one mitogenome from each Colombia and North Africa (Western Sahara), plus a Southeastern (Mozambique) and Central African (Nigeria, Yoruba) within L1c3b1b.

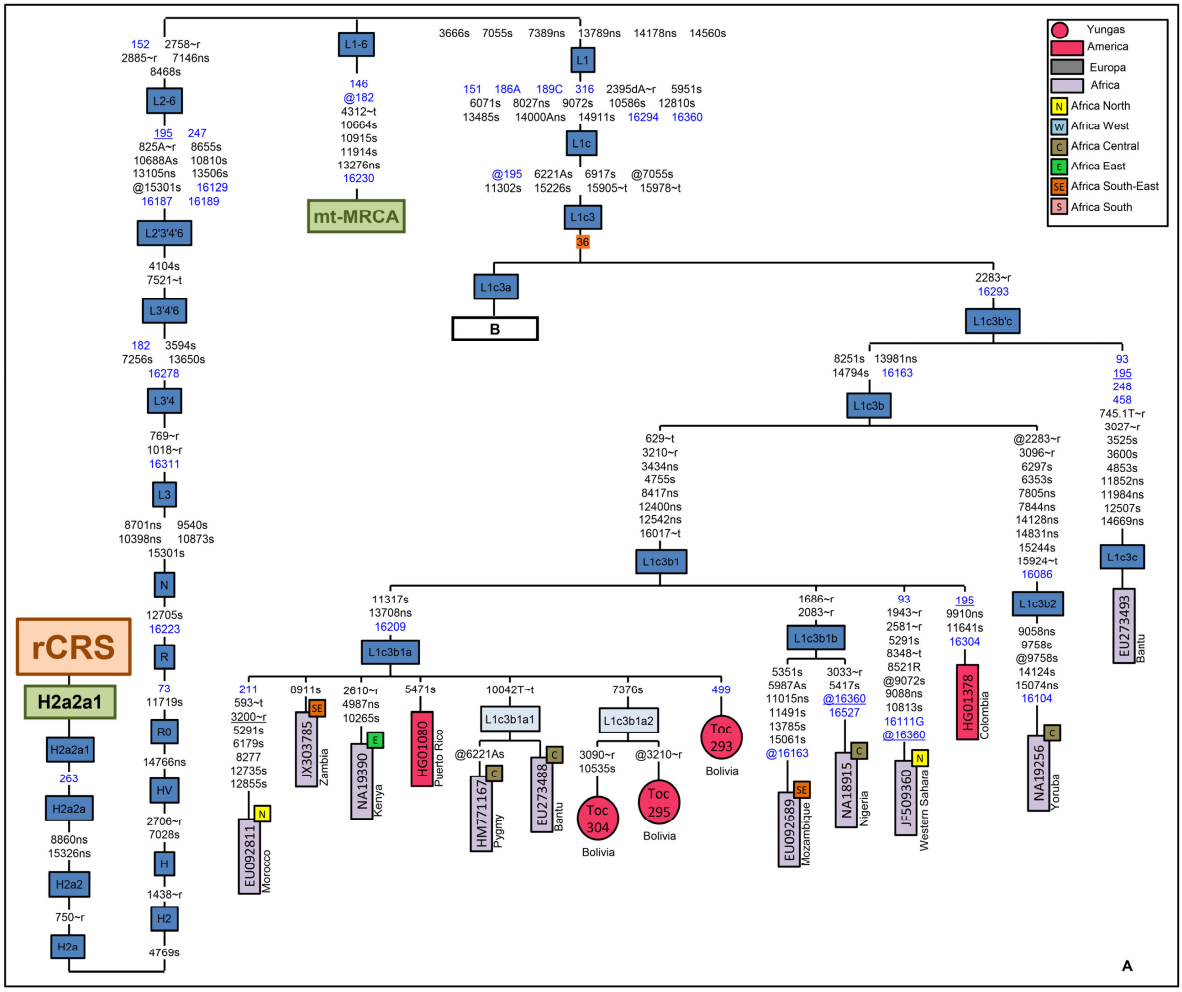


Figure 74 A: Phylogenetic tree L1c3.Submitted to PLos ONE.

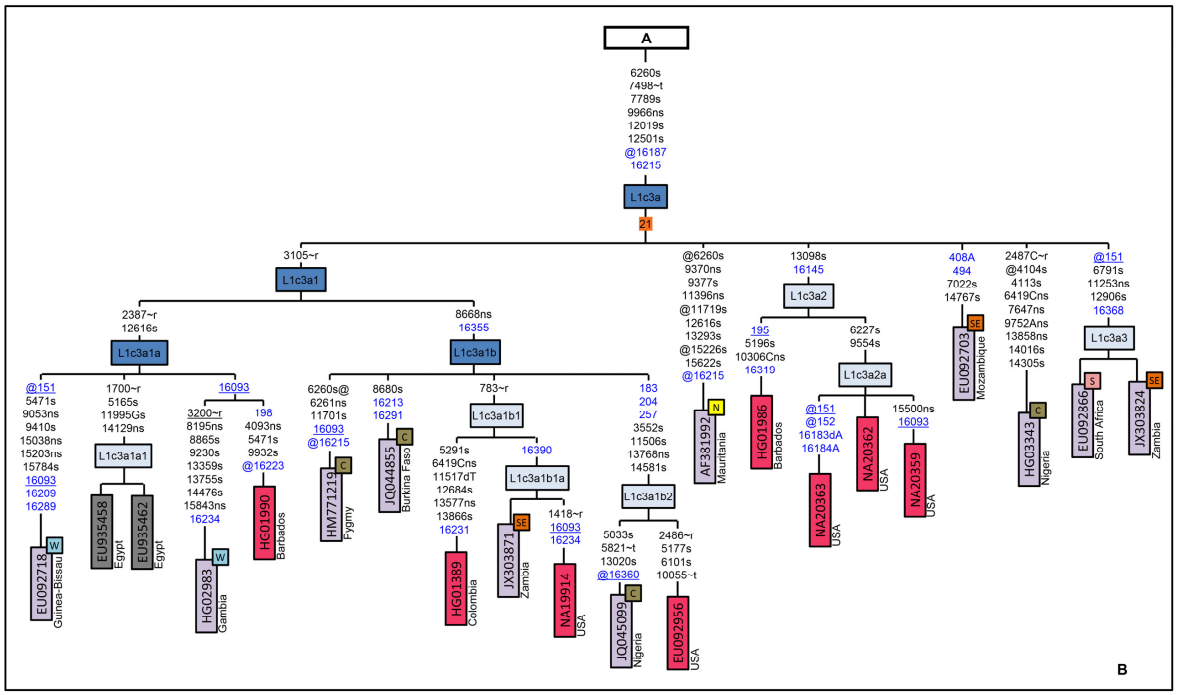


Figure 74 B: Phylogenetic tree L1c3.Submitted to PLos ONE.

The ‘Afro-Bolivian’ #Toc299 matches haplogroup L3d1a1a, a clade in which East Africa is well represented with two samples from Kenya and one from Somalia. Central Africa (Nigeria/Chad, Pygmy) is represented by two samples (Figure 75). Southeastern Africa (Mozambique), Middle East (Yemen), and even one individual from Pakistan appear each once. The latter individual forms together with one Kenyan the sub-haplogroup L3d1a1a1. The next immediate ancestral node L3d1a1 encompasses samples from West (Gambia), Central (Bukina Faso, Nigeria), East (Somalia), and South Africa.

Individual #Toc294 can be found within L3d1b3 but builds a new sub-branch (L3d1b3b), which is based on the transversion A10694T, together with one individual from Barbados. L3d1b3 further comprises two samples from Central Africa (Burkina Faso), and each one from West (Sierra Leone), and Southeastern Africa (Zambia), but also USA. West Africa (36%), including Sierra Leone, Gambia, and Guinea-Bissau, is most represented within L3d1b and followed by the American samples (32%) from Barbados, USA, Dominican Republic, Puerto Rico, and Bolivia.

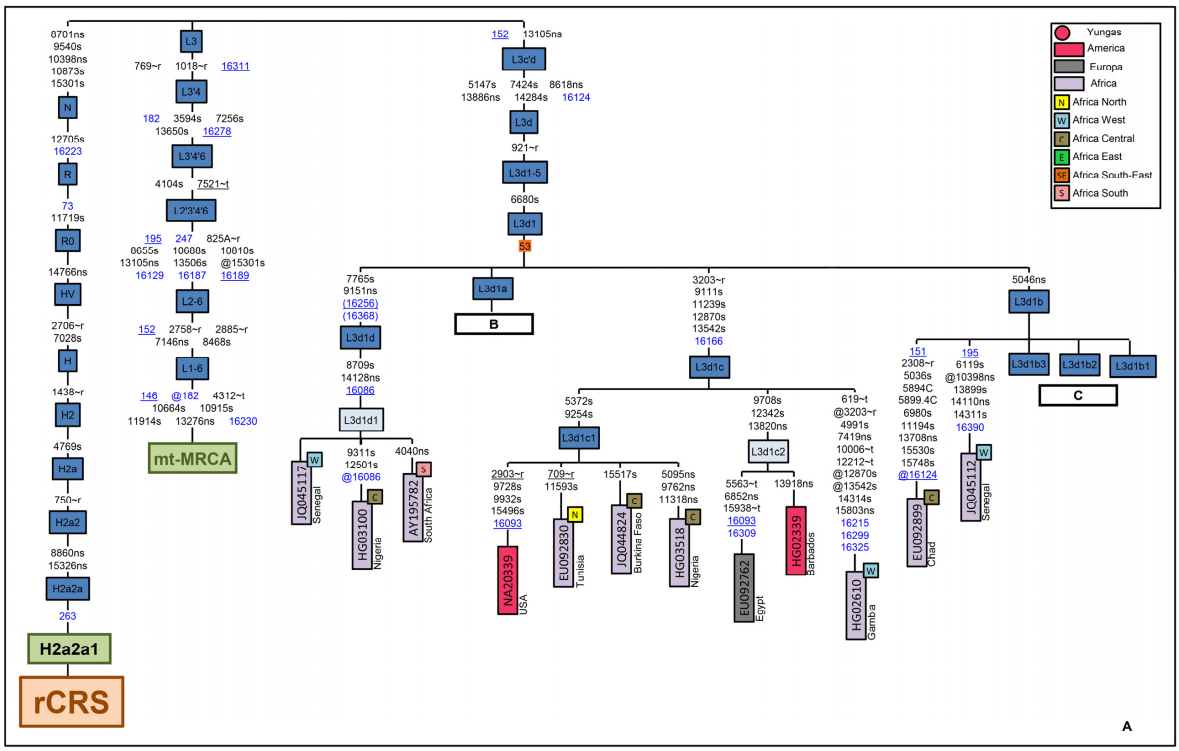


Figure 75 A: Phylogenetic tree L3d1. Submitted to PLos ONE.

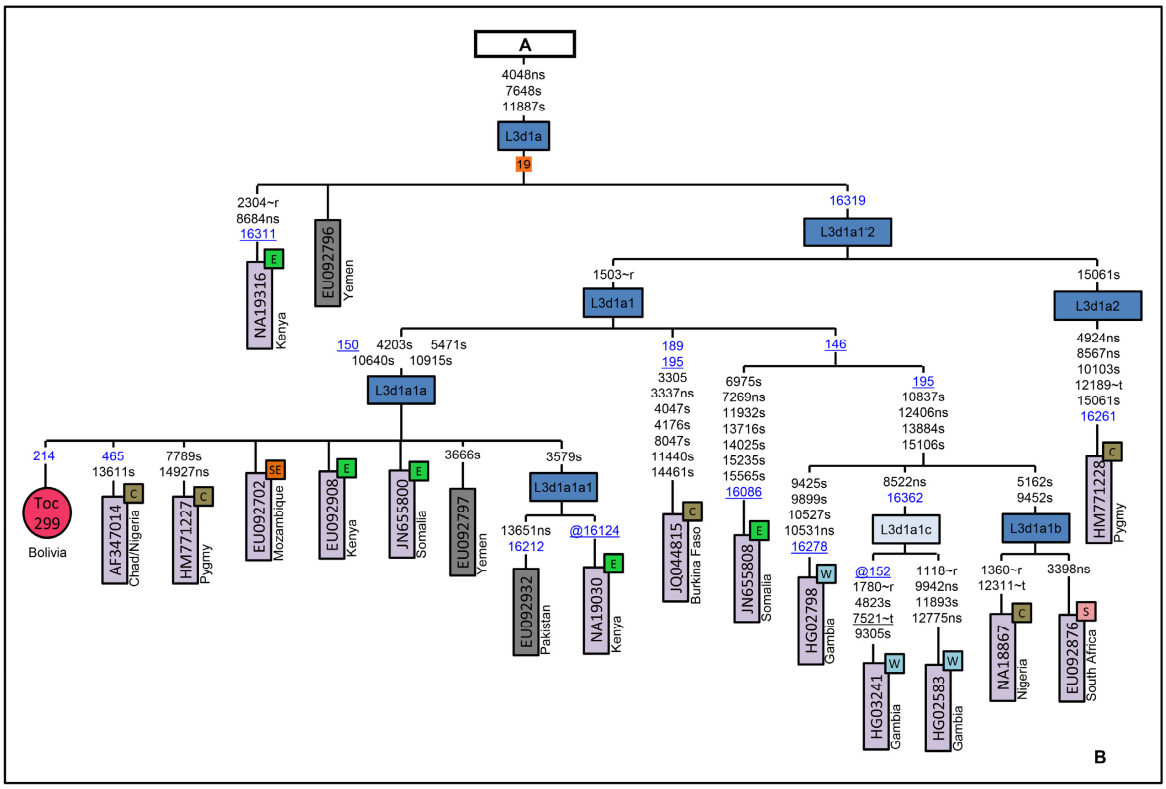


Figure 75 B: Phylogenetic tree L3d1. Submitted to PLos ONE.

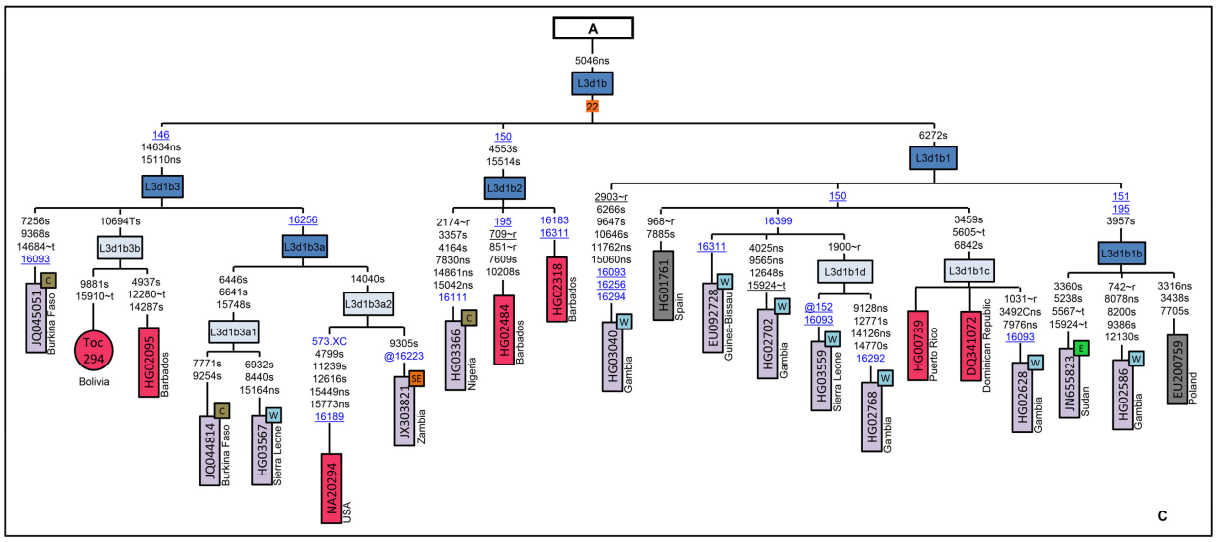


Figure 75 C: Phylogenetic tree L3d1. Submitted to PLos ONE.

6.1.3 Admixture analysis with control region haplotypes

In total 27 mtDNAs (17 different haplotypes) were observed in Bolivia that belong to the African-specific L haplogroup. These haplotypes were used to calculate proportions of admixture from different sub-continental African regions in an admixture analysis (Table 17).

Table 17: Admixture proportions, P_0, P_1, and P_2 (and 95% C.I.) from sub-continental regions in Africa in mtDNA haplotypes in Bolivia. Submitted to PLos ONE.

African region	P_0	95% C.I. (P_0)	P_1	95% C.I. (P_1)	P_2	95% C.I. (P_2)
North	0.0189	0.0180-0.0198	0.0344	0.0332-0.0356	0.0514	0.0499-0.0528
West-Central	0.1238	0.1217-0.1260	0.1521	0.1498-0.1545	0.2045	0.2019-0.2071
Southwest	0.5063	0.5030-0.5095	0.3617	0.3586-0.3649	0.3314	0.3283-0.3345
South	0.0054	0.0049-0.0059	0.0859	0.0841-0.0877	0.0986	0.0966-0.1005
Southeast	0.2607	0.2579-0.2636	0.2544	0.2515-0.2572	0.2058	0.2032-0.2085
East	0.0849	0.0831-0.0867	0.1114	0.1094-0.1135	0.1083	0.1063-0.1104

In contrast to other American populations like Brazil and Colombia where admixture can be mostly attributed to West-Central Africa [16, 17, 53, 84, 90, 349], the Bolivian mtDNAs of haplogroup L are firstly of Southwestern (mainly represented by Angola and Cabinda) and secondly of Southeastern (mainly represented by Mozambique) ancestry. However, admixture estimates differ

depending on the admixture proportions. That means, based on exact matches (P0) the admixture proportion in Bolivians came principally from Southwest (51%) followed by the Southeast (26%), but changes to 36% and 25%, respectively, when one mutation-step difference between haplotypes is allowed (P1). West-Central Africa, in contrast, lies only at the third position with 12% for P0 and 15% for P1. For East Africa, however, admixture proportions came up to 8% and 11%, respectively.

6.2 Y Chromosome

6.2.1 Y-STR molecular diversity

A total of 202 Yfiler and 128 minimum haplotypes were found in the Bolivian population(including the Yungas) (Table 18). Among the Yfiler haplotypes 15 different haplotypes were shared by 2 and 2 different haplotypes by 4 individuals. With respect to the minimum haplotype, 1 profile was similar in 22, 17, and 5, 2 profiles in 7, 3 profiles in 4, 9 profiles in 3, and finally 15 profiles in 2 samples. In Pando 17, in Beni 32, in Santa Cruz 25, in Cochabamba 24 and La Paz 60 different minimum haplotypes were observed whereby 27 haplotypes out of these haplotypes were shared between departments. The most frequentminimal haplotype was 13-14-17-23-10-16-14 and was observed in La Paz (16.8%), Santa Cruz (7.7%), and Cochabamba (7.1%).

Haplotype and gene diversity, but also mean number of pairwise differences were higher for Yfiler than for minimum haplotypes which is especially important in forensic applications. Consequently, the discrimination capacity is only 0.57 for the minimum, but 0.91 for the Yfiler haplotypes.

Among the departments Santa Cruz (Llanos) showed the highest diversity values for the minimumhaplotype. With respect to the Yfiler haplotypes Santa Cruz again exhibited the highest haplotype diversity, but in this case together with Cochabamba (Sub-Andean). Gene diversity and average number of pairwise differences for the Yfiler haplotypes, in contrast, was highest in the department of Beni (Llanos).

When eco-geographical regions are considered, the Llanos exhibited the highest diversity values for the minimum haplotype. Referring to the Yfiler haplotype the Llanos showed the highest values for gene diversity and average number of pairwise differences, but haplotype diversity, in contrast, was highest in the Sub-Andean region.

The intermediate allele 14.2 at the DYS19 locus was observed in one individual from Santa Cruz, but also at least in four individuals from different populations in the Americas

(http://www.yhrd.org). Three individuals of Cochambamba were observed with an intermediate allele at DYS458. Two individuals showed 17.2 (within haplogroup J*) and one individual 20.2 (haplogroup R1b1*, here refered to as R1b1(xM269) meaning that haplogroups found via Multiplex 6 are not included). A null allele at DYS439 (haplogroup Q*) was found in one individual from Cochabamba, which also has been observered previously in at least nine populations including Native America, Eurasia, East Asia, and 'African Americans' (http://www.yhrd.org).

Table 18: Y-STR Molecular diversity in Bolivian populationsamples. Published in [350], Table 1.

Population	N	K	DC	UH	HD	h	M
Yfiler							
Pando	21	20	0.95	20	0.9952 ± 0.0165	0.610417 ± 0.325057	9.766667 ± 4.659243
Beni	41	38	0.93	36	0.9963 ± 0.0063	0.652287 ± 0.337224	10.436585 ± 4.857652
Santa Cruz	26	26	1.00	23	1.0000 ± 0.0107	0.649808 ± 0.341165	10.396923 ± 4.899982
Cochabamba	28	28	1.00	27	1.0000 ± 0.0095	0.610946 ± 0.321076	9.775132 ± 4.614161
La Paz	107	93	0.87	87	0.9965 ± 0.0021	0.599442 ± 0.306762	9.591077 ± 4.431930
Llanos	88	82	0.93	81	0.9984 ± 0.0020	0.642682 ± 0.328070	10.282915 ± 4.737945
Andean	107	93	0.47	87	0.9965 ± 0.0021	0.593480 ± 0.303917	9.495680 ± 4.390827
Sub-Andean	28	28	2.32	27	1.0000 ± 0.0095	0.604993 ± 0.318155	9.679894 ± 4.572186
All Bolivians	223	202	0.91	196	0.9989 ± 0.0006	0.624785 ± 0.317280	9.996566 ± 4.588061
Minimum haplotype							
Pando	21	17	0.81	10	0.9810 ± 0.0197	0.576190 ± 0.334459	4.033333 ± 2.097376
Beni	41	32	0.78	21	0.9890 ± 0.0074	0.598955 ± 0.337344	4.192683 ± 2.125982
Santa Cruz	26	25	0.96	16	0.9969 ± 0.0117	0.640879 ± 0.363256	4.486154 ± 2.282552
Cochabamba	28	24	0.86	14	0.9815 ± 0.0181	0.561224 ± 0.322740	3.928571 ± 2.029159
La Paz	107	60	0.56	45	0.9552 ± 0.0120	0.582890 ± 0.324388	4.080233 ± 2.050375
Llanos	88	66	0.75	52	0.9916 ± 0.0034	0.610651 ± 0.338452	4.274556 ± 2.138451
Andean	107	60	0.56	45	0.9552 ± 0.0120	0.582890 ± 0.324388	4.080233 ± 2.050375
Sub-Andean	28	24	0.86	14	0.9815 ± 0.0181	0.561224 ± 0.322740	3.928571 ± 2.029159
All Bolivians	223	128	0.57	128	0.9806 ± 0.0043	0.594895 ± 0.328544	4.164263 ± 2.078542

N = sample size; *K* = number of different haplotypes; *DC*= discrimination capacity;*UH*= number of unique haplotypes; *HD*= Haplotype diversity; *h*= gene diversity over loci; M = average number of pairwise differences.

6.2.2 Y-STR AMOVA analysis

An AMOVAanalysis was conducted by considering both the Yfiler, but also the minimum haplotype (Table 19). The AMOVA in Bolivians was carried out by grouping the populations, firstly, into the three eco-geographical regions, and secondly into departments, and thirdly into departments again but in this case three other Bolivian datasets, which were previously reported in the literature, were included in the analysis. The Y-haplotypes from the literature came from Gayà-Vidal and colleagues (Aymara: n = 59; Quechua: n = 55) [281], Tirado and colleagues (Beni: n = 87) [351], and Vullo and colleagues (Beni: n = 87, Chuquisaca: n = 38, n = 57) [233]. AMOVA results for the Bolivian populations was compared with other Y-haplotypes from the literature which were published in Romero and colleagues (Colombia: n = 305) [352], Toscanini and colleagues (Argentina, Collas: n = 29; Toba: n = 49) [353], Roby and colleagues (Chile, Santiago de Chile: n = 196; Iquique: n = 196) [354], Cainé and colleagues [355], and Schwengber and colleagues (South Brazil: Rio Grande do Sul and Santa Catarina, n = 364) [356].

Table 19: AMOVA analysis in Bolivian and other South American populations. Published in [350], Table 2.

Population	*Pop*	*n*	*17 Y-STRs*		*7 Y-STRs*	
			Among	Within	Among	Within
Bolivia (regions)	3	223	3.75	96.25	5.04	94.96
Bolivia (departments)	5	223	5.51	94.49	4.94	95.06
Bolivia (including literature)	8	424	11.15	88.85	9.74	90.26
South America (all)	14	1563	14.80	85.20	19.27	80.73
America minus Colombia	13	1258	18.67	81.33	21.98	78.02
America minus Argentina	12	1485	14.09	85.91	18.63	81.37
America minus Chile	12	1171	16.93	83.07	23.25	76.75
America minus Brazil	13	1199	16.19	83.81	19.77	80.23
America minus Bolivia	6	1139	4.86	95.14	7.28	92.72

Pop = number of population datasets; n = total sample size.

Most variance was observed within populations. Among population variance was slightly higher for the minimum than for the Yfiler haplotype, except when the Bolivian population was grouped into departments and populations from the literature were included. This observation has been already described in [251], and it could be due to the fact that Yfiler profiles increase the amount of

singleton haplotypes in all the sample populations, and therefore, raise haplotype diversity to the maximum theoretical values. All P-values were significant.

Among population variance came to 5.51% for the Yfiler and 4.94% for the minimum haplotype when populations are grouped into departments. However, values rise to 11.15% and 9.74%, respectively, when the three additional Bolivian profiles from the literature were included into the AMOVA.

Values for among population variance were slightly lower when populations were grouped into regions and not into departments (Yfiler). Among population variance reached 3.75% for the Yfiler and 5.04% for the minimum haplotype when populations are divided into the main geographical regions, and thus, are slightly lower than those values obtained when the Bolivian population is clustered into departments.

Non-Bolivian South American profiles (n=1437) were collected from the literature (see above) [233, 281, 289, 303, 351–360] in order to compare among and within population variance with other American populations. Among population variance for the Yfiler and minimum haplotypes reacheda higher value in South America (all) with 14.60% and 19.21%. When Colombia was removed from the analysis among population variance increased to 18.67% which represents the highest value for the Yfiler haplotypes. With respect to the minimum haplotype the highest value of among population variance of 23.25% was reached when Chile was removed from the AMOVA. However, when the Bolivians were removed from the analysis values decreased significantly to 4.86% for the Yfiler and 7.28% for the minimum haplotype. Thus, Bolivian haplotypes apparently account significantly for the stratification of the sample.

6.2.3 Y-SNP molecular diversity

With respect to the Y-SNP markers, Santa Cruz exhibited the highest diversity values (Table 20). Consequently, the diversity values were highest in the Llanos (including Santa Curz).

Table 20: Y-SNP Molecular diversity in Bolivian population samples. Published in [350], Table 1.

Population	N	HD	h	M
Y-SNPs				
Pando	21	0.6857 ± 0.0892	0.099471 ± 0.061342	2.685714 ± 1.482775
Beni	41	0.6585 ± 0.0479	0.093586 ± 0.057017	2.526829 ± 1.385439
Santa Cruz	26	0.7908 ± 0.0674	0.122963 ± 0.072505	3.320000 ± 1.756627
Cochabamba	28	0.7328 ± 0.0689	0.105232 ± 0.063484	2.841270 ± 1.538839
La Paz	107	0.6928 ± 0.0288	0.072298 ± 0.045729	1.952037 ± 1.114639
Llanos	88	0.7965 ± 0.0293	0.102048 ± 0.057705	3.673720 ± 1.874895
Andean	107	0.7161 ± 0.0327	0.067576 ± 0.040865	2.432728 ± 1.328156
Sub-Andean	28	0.7354 ± 0.0701	0.076940 ± 0.046623	2.769841 ± 1.506807

N = sample size; *HD*= Haplotype diversity; *h*= gene diversity over loci; M = average number of pairwise differences.

6.2.4 Y-SNP haplogroup diversity

About 94% of all samples were either of European or Native American ancestry (Figure 76). That is, 65.4% were of European origin comprising R1b1-P25 (including R1b1b2-M269), I-M170 (including I2a2-M26), G-M201, and J-M304 (including J1a-M62 and J2-M172) and 28.3% of Native American origin like P and Q (including Q1a3a-M3). About 6% of the haplotypes were African-specific encompassing BR-SRY10831.1 and E-M96 (including E1b1b1-M35, E1b1b1a-M78, and E1b1b1b-M81). Most African-specific haplogroups were found in the Llanos and the Sub-Andean regions. However, in the 'Afro-Bolivian' community of Tocaña (Sub-Andean) 26.6% haplotypes were of African ancestry.

The European component in the Andean and Sub-Andean region represents 75% and 86%, respectively, of all haplogroups whereas in the Llanos the European component reached only 46%. The frequency of R1b1-P25 reached 32% in the Llanos (R1b1b2-M269), but 71% both in the Sub-Andean and Andean region (R1b1b2* and R1b1*). However, R1b1(xM269) was not found in the Llanos, but came to 46% in the Sub-Andeans and Andeans. In contrast, most Native American

haplogroups were found in the Llanos. That is 34% of Q1a3a*-M3, 7% of Q-M242, and 6% of P-92R7.

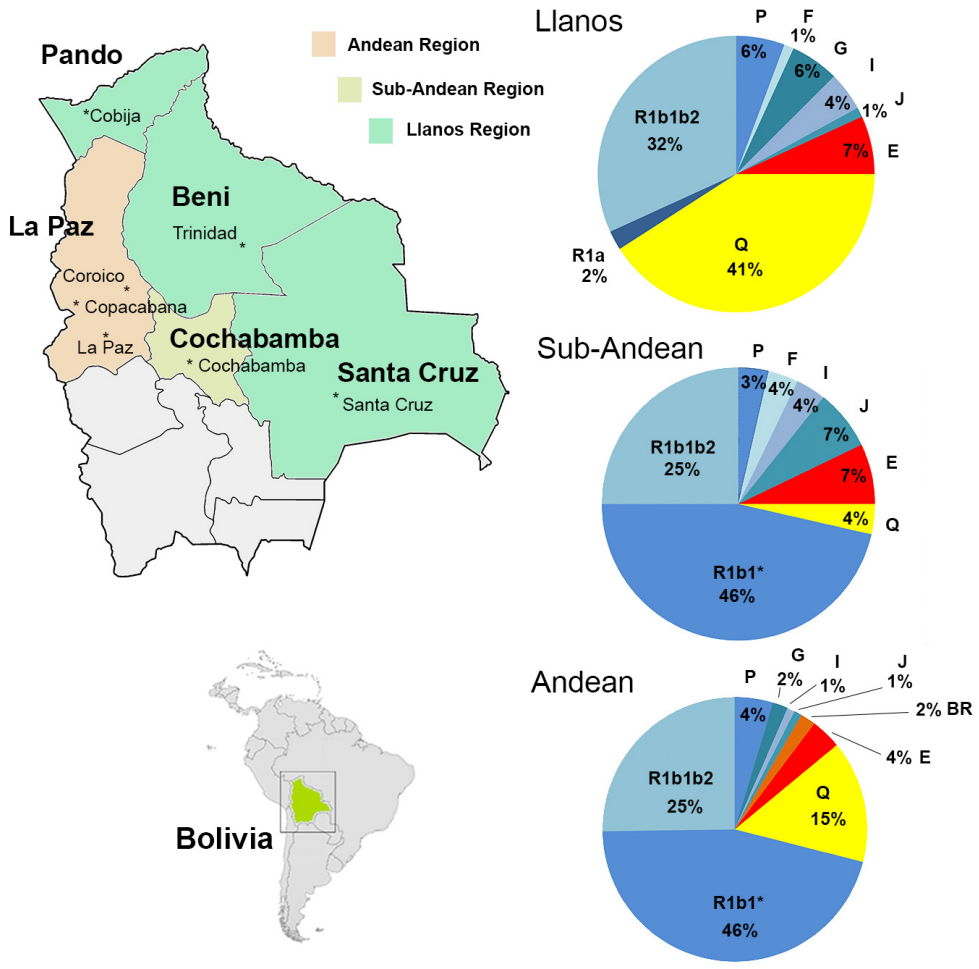


Figure 76: Y chromosome haplogroup frequencies among the Bolivian population. Sample locations are indicated as asterisks, published in [350], Figure Figure 1.

6.2.5 Principal Component Analysis with Y-STRs and Y-SNPs

Both PC1 and PC2 exhibited separate clusters (Africa, East Asia, and Europe/America) whereby variability for PC1 reached a value of 42.7% and PC2 a value of 32.4% (Figure 77). Native American populations in the PC1 and PC2 plot show a close affinity to the European and not to the East Asian sample. However, the proximity of both populations in the PC1 and PC2 can be

attributed to the high European ancestry of the Native American populations that were considered for the analysis.

In the PC1 and PC3 plot it can be observed that the Andean and Sub-Andean samples appear closer to the European cluster due to their high European ancestry in the Y chromosome and the Llanos populations affiliated closer to the Native American samples which can be attributed here to the lower European, but the higher Native American ancestry on the Y chromosome. The PC3 explained 20.4% of variance. Other populations from Bolivia from the literature, like Aymara and Quechua, rather affiliate to the Native American pole.

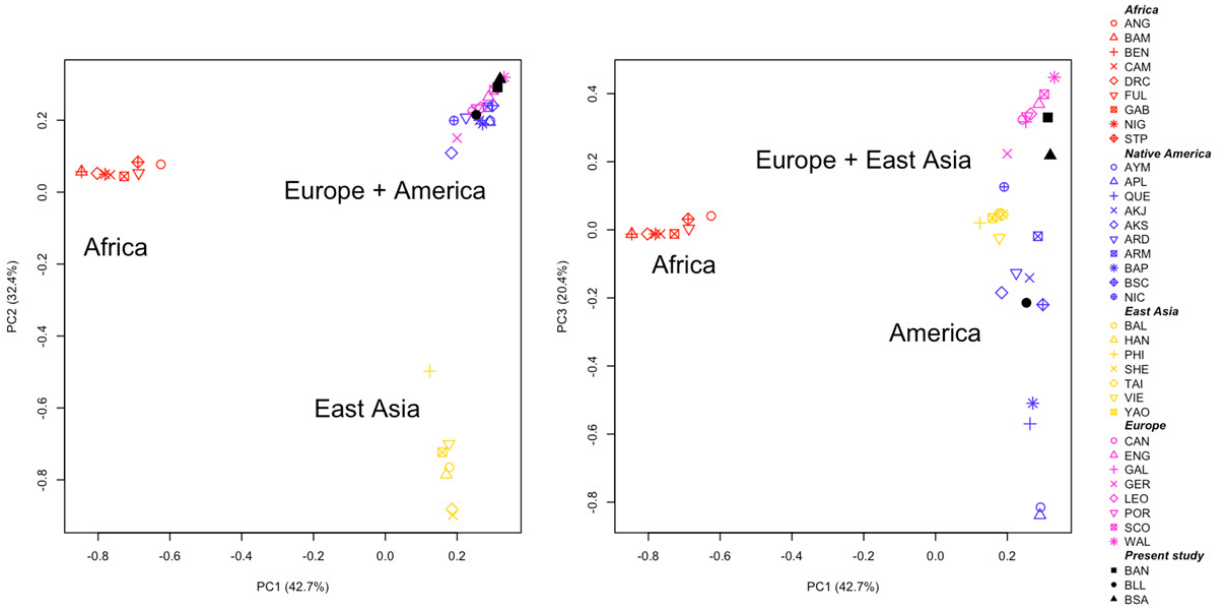


Figure 77: Principal Component Analysis (PCA) of Y-STRs and Y-SNPs. Populations are: Fon (BEN) [308, 309], Fulbe (FUL) [310], São Tomé e Príncipe (STP) [311], Bakaka, Ewondo, Ngoumba (CAM), Makina and Nzebi (GAB), Tiv and Igala (NIG) [312], Angola (ANG) [313], Bamelike (BAM) [308], Democratic Republic of the Congo (DRC) [314], Native America = Nicaragua (NIC) [315], Bolivian - Aymara (AYM) and Bolivian - Quechua (QUE) [281], Andean - La Paz (BAP), and Sub-Andean - Chuquisaca (BSC) [233], Apalai (APL) [316], Diaguitas (ARD), Kolla - Jujuy (AKJ), Kolla - Salta (AKS), and Mapuche (ARM) [303], East Asia = Bali (Indonesia) (BAL) [317], Philippines (PHI), Taiwan (TAI), She (SHE), Yao (YAO), and Vietnam (VIE) [318], Han (HAN) [319], Europe = Cantabria (CAN), León (LEO) [320], England (ENG), Wales (WAL), Scotland (SCO) [321], Galicia (GAL) [306], Germany (GER) [322], Portugal (POR) [323],Andean Region (BAN), Sub-Andean Region (BSA), and Llanos (BLL).

6.2.6 Admixture analysis of Bolvia

Continental ancestry proportions in Bolivians among the Llanos, Sub-Andeans, and Andeanswas determined with an admixture analysis which was based on worldwide main major Y-SNP haplogroups found in Africa, Europe, Native America, East Asia, and South Asia [219] (Table 21).

Table 21: Admixture proportions among eco-geographical regions in Bolivia. Published in [350].

Parental Groups	Llanos	Sub-Andean	Andean
African	0.000 ± 0.258	0.000 ± 0.870	0.000 ± 0.927
European	0.194 ± 0.236	0.945 ± 0.797	0.757 ± 0.849
Native American	0.799 ± 0.054	0.055 ± 0.182	0.243 ± 0.194
East Asian	0.000 ± 0.035	0.000 ± 0.119	0.000 ± 0.127
South Asian	0.000 ± 0.071	0.000 ± 0.241	0.000 ± 0.256

In the Llanos the largest ancestry proportion is Native American (80%) followed by European (19%). In contrast, European admixture was highest in the Sub-Andean (95%) and Andean regions (76%) whereby Native American ancestry reached lower values (6% and 24%, respectively). Thus, although African haplogroups were observed in about 6% of the Bolivian population, an African compononent could not be detected with the admixture analysis. An East and South Asian component could be neither found in Bolivians.

6.2.7 Geographic origin of Y chromosomes

According to the Y chromosome Reference Database (YHRD, http://yhrd.org) some Y chromosome profiles found in Bolivia have a clear geographic distribution in Southeast Africa like, e.g., #Tocaña291 (Figure 78) while others have profiles that are more widely distributed like the one of #Tocaña295 (Figure 79). Some profiles are only found in the Americas like it was the case for #Tocaña294 (Figure 80).

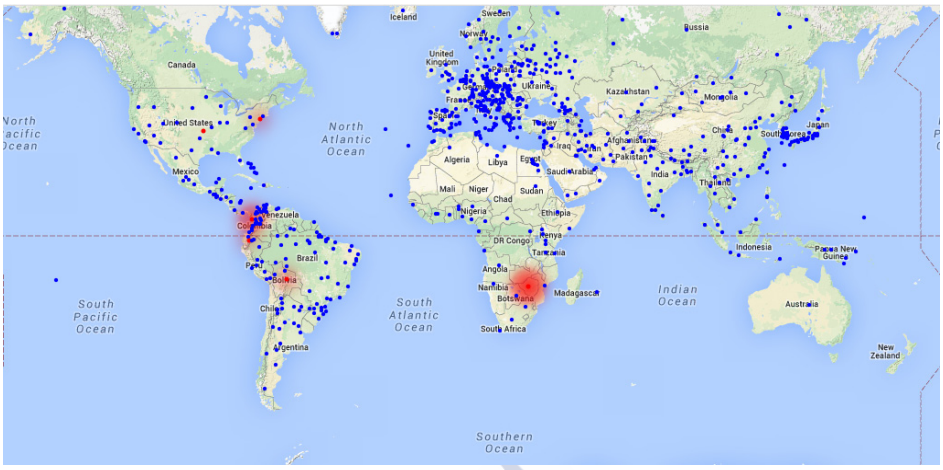


Figure 78: Distribution of the Y chromosome profile of Tocaña291 belonging to haplogroup E*. Submitted to PLos ONE.

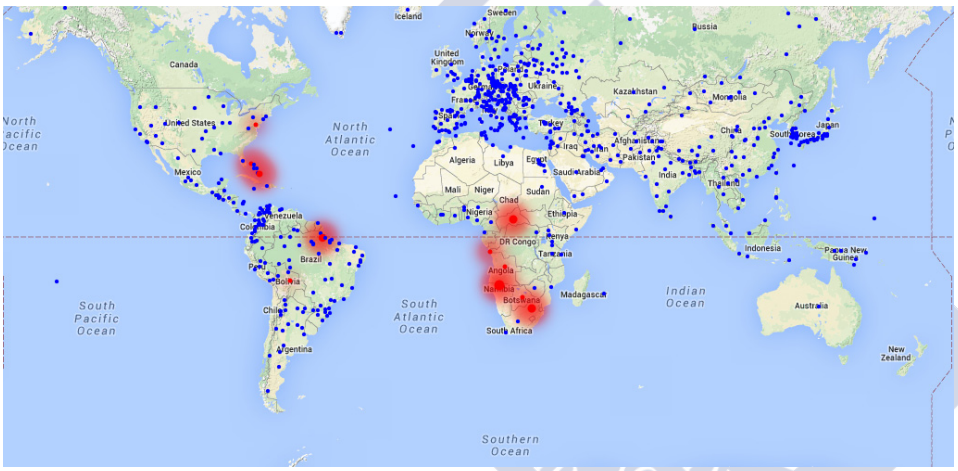


Figure 79: Distribution of the Y chromosome profile of Tocaña295 belonging to haplogroup BR. Submitted to PLos ONE.

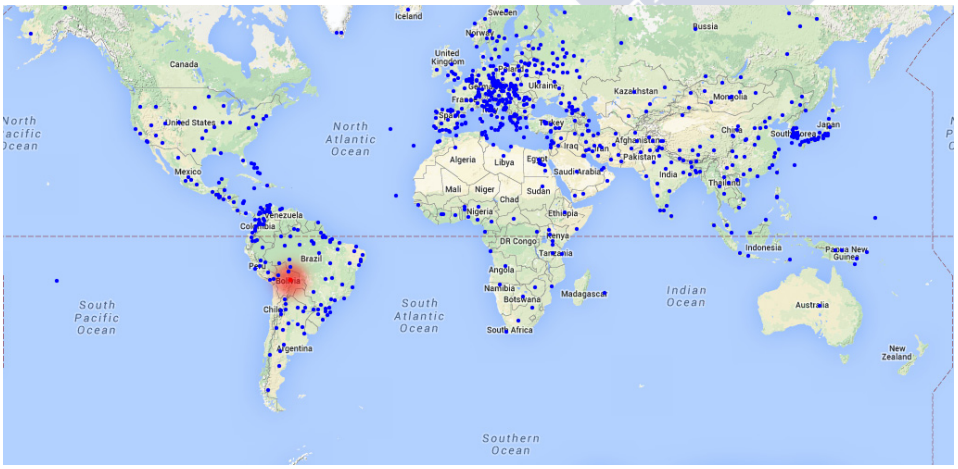


Figure 80: Distribution of the Y chromosome profile of Tocaña294 belonging to haplogroup E1b1b1*. Submitted to PLos ONE.

6.3 Autosomal DNA

6.3.1 Indel genotyping

6.3.1.1 Continental ancestry in the Bolivian population

In summary, continental ancestry in the Bolivian population is mainly Native American (73.3%) followed by European (25.8%). African ancestry (0.9%), in contrast, was observed to be minimal (Figure 81). When analyzing continental ancestry by department, La Paz (81%) exhibited the highest Native American, but the lowest European contribution (18.7%). In comparison, Santa Cruzwas observed to have the lowest Native American (57.8%), but the highest European ancestry (39.4%). In all departments African ancestry was lowwith Pando showing the highest value (3%).Among ecological regions the Native American genetic contribution was highest in the Andeans (80.5%) and decreasing from the West via the Sub-Andean (71.9%) to the East in the Llanos (67%). In contrast, European ancestry was lowest in the Andeans (19.2%) rising in the Sub-Andean (27.9%) and is highest in the Llanos (31.2%). The highest African ancestry was observed in the Llanos (1.9%). Native American ancestry was almost similar in rural (77.2%) and in urban (72.1%) populations. Likewise, European (22.2%, 26.9%), and African ancestry (0.6, 0.9%) differed only slightly.

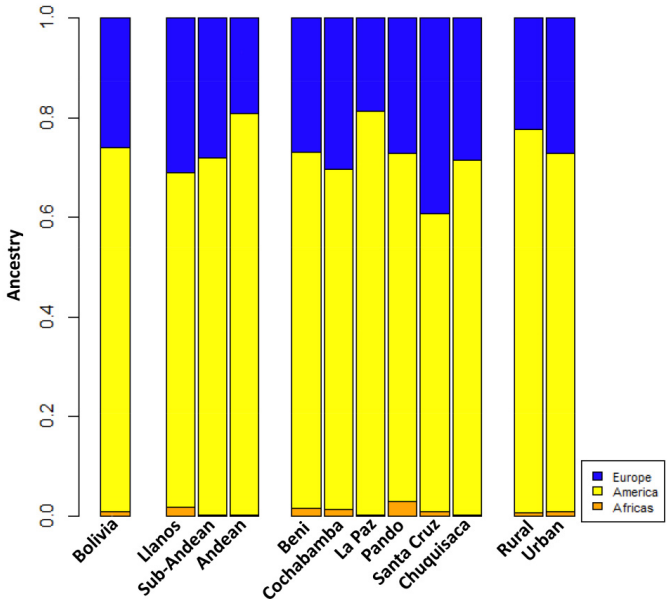


Figure 81: Average continental ancestry of the Bolivian population based on 46 AIMs. Adapted from Figure 4 published in [333], doi:10.1371/journal.pone.0058980.g004, East Asia was removed as reference population. Based on STRUCTURE results.

Continental ancestry for each individual of the respective population is shown in STRUCTURE graphs (Figure 82, 83, 84). The optimal value of K given by Structure Harvester [327] was K=3, and thus, reflecting the three continental regions Africa, Europe, and Native America.The graphs clearly demonstrate that the Native American ancestry is prevalent in the Bolivian population with Santa Cruz exhibiting the lowest Native American, but the greatest European genetic component when compared to other departments. Likewise, a decreasing Native American and an increasing European genetic component from the Andeans via the Sub-Andeans to the Llanos can be clearly seen. In all departments, there are a few individuals with European ancestry values of >50% or slightly below that level, especially in Santa Cruz (n=22) and Pando (n=6). The highest value (69%) was found in an individual from Pando. African ancestry, in contrast, was low with the highest values found in La Paz (23.6%) and Pando (16.9%).

The Bolivian populations seem to exhibit asimilar genetic sub-structure. At K=5 all Bolivian departments show a genetic component (green) which is shared with the Native American reference group. At K=6 the Native American populations seem to share this genetic component with the European population. The African reference population, in contrast, apparently does not share a genetic component (pink) which can be observed from K=4 to K=6.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

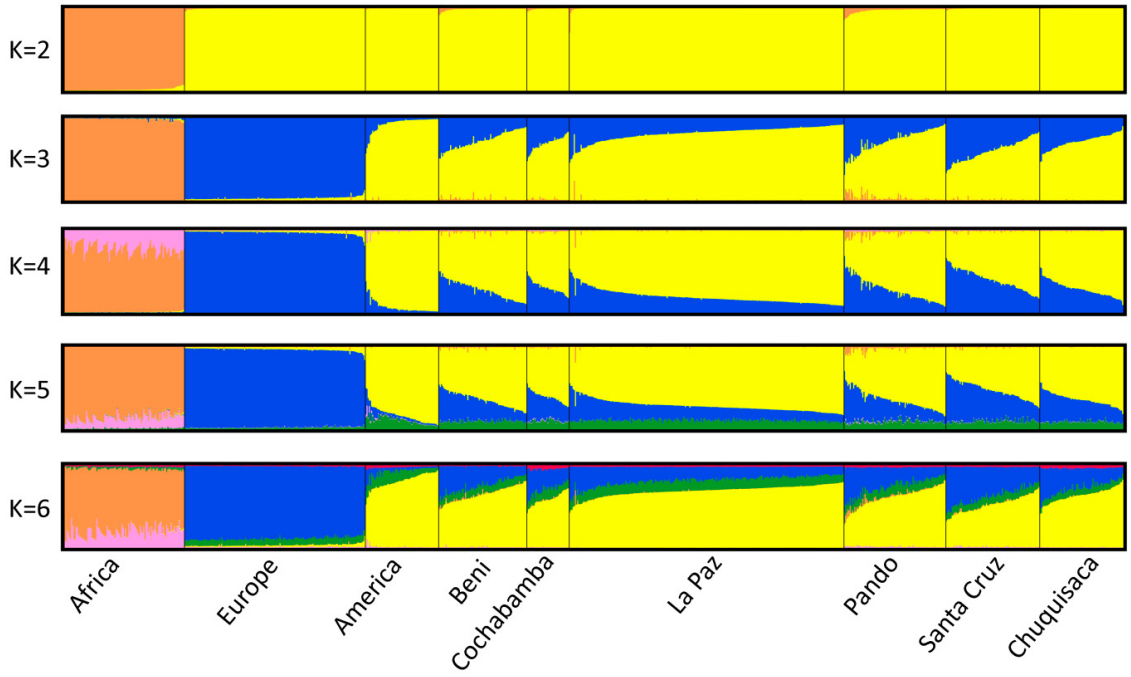


Figure 82: Structure analysis of the Bolivian population among departments. Adapted from Figure S7 published in [333], doi:10.1371/journal.pone.0058980.g005, East Asia was removed as reference population. Based on STRUCTURE results.

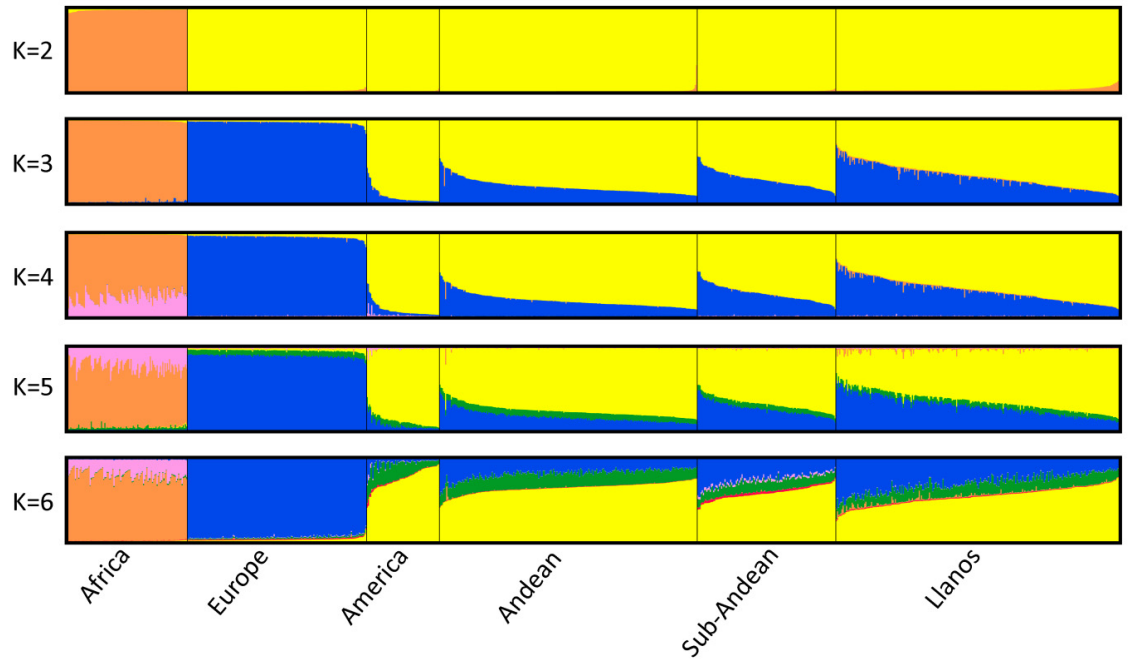


Figure 83: Structure analysis of the Bolivian population among regions. Adapted from S8. published in [333], doi:10.1371/journal.pone.0058980.g005, East Asia was removed as reference population. Based on STRUCTURE results.

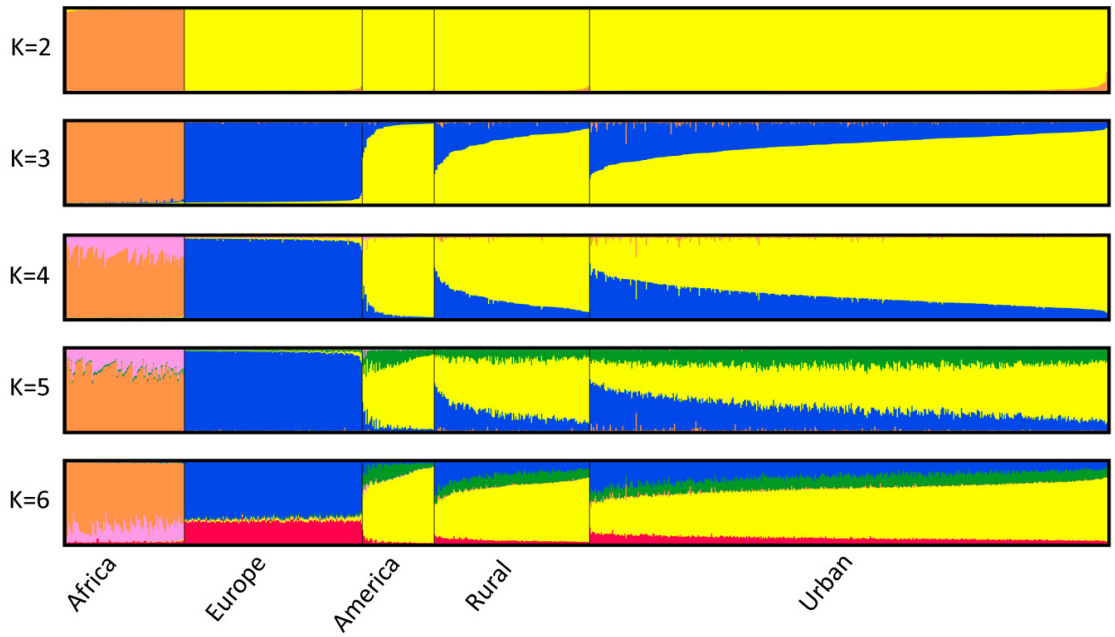


Figure 84: Structure analysis of the Bolivian population among rural and urban regions. Adapted from S9 published in [333], doi:10.1371/journal.pone.0058980.g005, East Asia was removed as reference population. Based on STRUCTURE results.

In the PCA analysis the Bolivianpopulationclearly groups with the Native American reference group with only some samples that show some closer affiliations to Europe (Figure 85A). Only one individual, that is, the individual from La Paz with a relative high African ancestry (23.6%) when compared to other Bolivians shows some affinity to the African cluster. The first principal component (PC1=25.84%) separates Africa from Native America and Europe and the second (PC2=11.75%) separates Europe from Native America. It can be observed that the Andean populations are more tightly grouped into clusters whereas the Llanos populations are more dispersed (Figure 85B). The Sub-Andean populations occupy an intermediate position. This dispersion can be observed again in the standard deviation (SD) of the main ancestry genetic component in Bolivians, that is, among Native American ancestry. The mean of Native American ancestry in the Andeans reached 80.5% with a standard deviation of 8.5%. The Sub-Andeans had a mean of 71.8% and a SD of 9.9%. In contrast, the populations in the Llanos were observed with a Native American ancestry mean of 67% and a SD of 14.1%. This pattern is again observed when considering the departments. Santa Cruz and Pando (from the Llanos) seem to highly disperse (Figure 85C). No clear pattern can be observed when rural populations are compared against urban populations (Figure 85D).

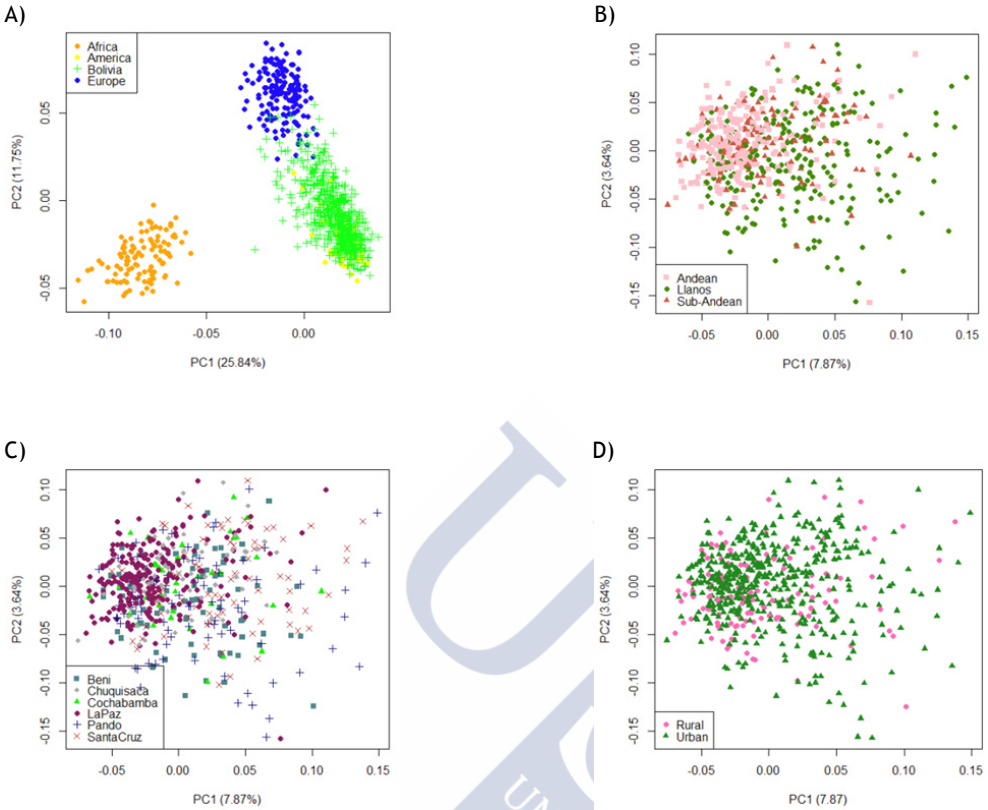


Figure 85: PCA of Bolivian profiles with 46 AIM indel markers. Based on Figure 6 published in [333], doi:10.1371/journal.pone.0058980.g006, East Asia was removed as reference population from figure A.

The outcome of estimating ancestries in populations depends on many decisions like the panel of SNPs, but also the number of SNPs included, the reference samples, and number of samples. Twelve of the Bolivian samples were previously genotyped with the LACE panel [259]. By comparison, Native American ancestry with the 46-AIM panel was lower, and correspondingly, led to an increased European ancestry (on average the difference is about 15% between the 46-AIM and the LACE panel). Therefore, European ancestry might be slightly overestimated with the 46-AIM panel in the Bolivian population. The mean of the European ancestry in the whole Bolivian population reached 24.6%. In a previous study with a 324 Native American-specific AIM panel European ancestry was observed to be around 12% [288].

The influence of sample size on the estimation of population ancestry was evaluated with a simulation analysis (Figure 86). When using a sample size between 30 and 70 the mean value for the ancestry proportion varies to a very low degree. For example, Native American ancestry in La Paz (n=105) reached 80.1%, 83.1% and 85.5% for a sample size of 5, 30, and 70. The respective

values for Chuquisaca (n=73) came to 72.3%, 74.6%, and 76.3%. As shown in Figure 82, the mean values in the simulations approach the values which were obtained with the original sample size. As it was expected, the bootstrap intervals decrease as the sample sizes increase. Above a sample size of 30, the bootstrap intervals are already small. Bootstrap intervals do not overlap between continental ancestries even for sample size five.

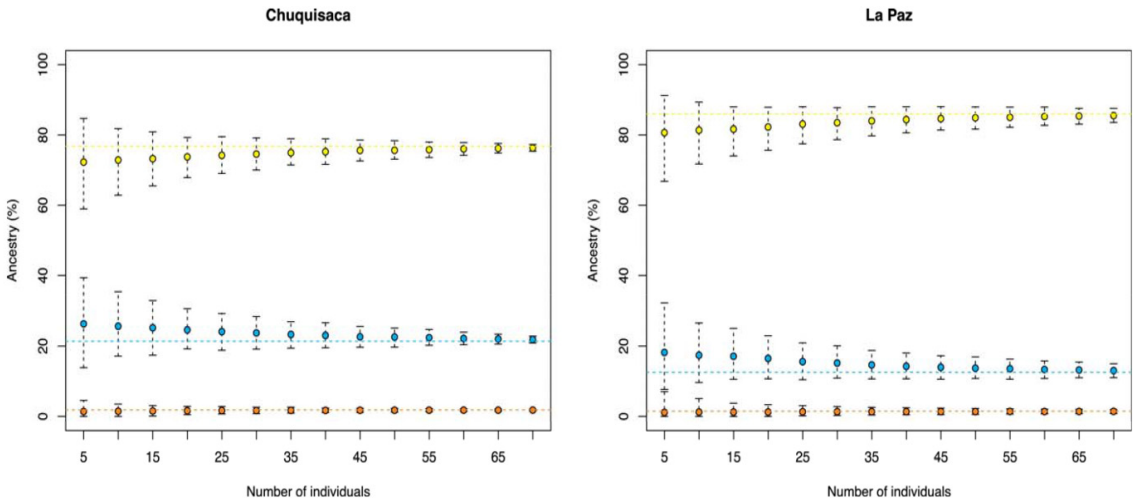


Figure 86: Ancestry values obtained for La Paz and Chuquisaca by way of simulations. Vertical dotted bars indicate bootstrap intervals obtained from the computation of ancestry on 1000 simulations (each sample size); the central dots indicate the mean values. Published in [361].

6.3.1.2 Continental ancestry in the Yungas population

Individual ancestry is shown in ADMIXTURE barplots (Figure 87). Overall, Native American ancestry came to 76%, European to 12%, and African as well to 12%. However, these values vary substantially in the Nor and Sud Yungas, as the Nor Yungas show high African ancestry which is especially concentrated in the community of Tocaña. In the Nor Yungas African ancestry reached 13%, but fall to 3% when Tocaña is excluded from the estimate. African ancestry in Tocaña in fact came to 56%. The highst value of African ancestry (78%) again can be found in Tocaña. In the Sud Yungas African ancestry is virtually absent (<1%).

European ancestry was higher in the Nor Yungas (12% with and without Tocaña) than in the Sud Yungas (5%). The maximum value for European ancestry in the Nor Yungas (61%) was more than four times higher than Sud Yungas (14%).

In contrast, Native American ancestry was higher in the Sud Yungas (94%) than in the Nor Yungas (74%). The Native American component in the Nor Yungas was highly variable ranging from 1% to 100%. In the Sud Yungas the minimum value came to 86% and the maxmimum one to 100%.

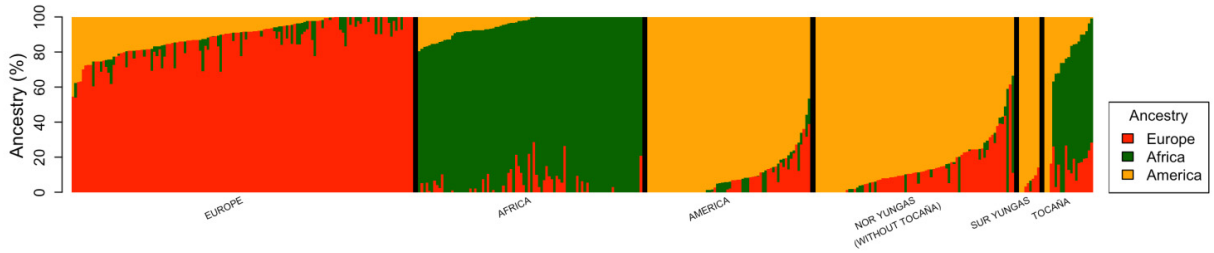


Figure 87: Individual ancestry in the Nor and Sud Yungas. Based on ADMIXTURE software. Submitted to PLos ONE.

The PCA analysis clearly separates the three main ancestral components Europe, Africa, and Native America (Figure 88). The PCA1 explains 14% of the variation and separates Europen and Native America from Africa whereas PC2 accounts for 9% of the variation and separates Europe from the other two populations. Most individuals from the Yungas can be found in the Native American cluster, however, some of them rather approach the European cluster. A few individuals can be found close to the African cluster. The majority of the Tocañas affiliate closely with the African population, except three individuals which affiliate most with Native America which corresponds with the ADMIXTURE results that revealed three individuals of Tocaña with African ancestry values ranging from 0.8% to 1.2%.

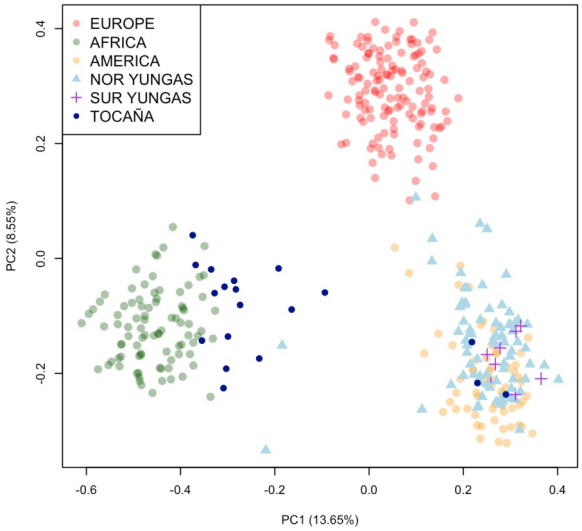


Figure 88: PCA of the Yungas population. Submitted to PLos ONE.

6.3.2 SNP genotyping

6.3.2.1 Analysis of IBS and Principle component analysis

The genetic distance based on IBS between the Tocaña individuals and each African population sample set was calculated to examine the genetic proximity of the two 'Afro-Bolivians' to different sub-Saharan African groups. In a first analysis the two individuals were compared against a population set from 1000G and in a second one against a large dataset of African populations (including populations that were affected by the TAST). According to the first analysis with the 1000G samples, the two 'Afro-Bolivians' showed the highest IBS values with populations from Africa (represented by West-Central and East Africa) and the lowest values with non-African samples (Figure 89A). In the second analysis with the large dataset of African populations, the highest IBS values could be observed for the two Tocañs with Yoruba (Nigeria) and other populations from West-Central, but also with the Luhya from Kenya in East Africa (Figure 89B).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

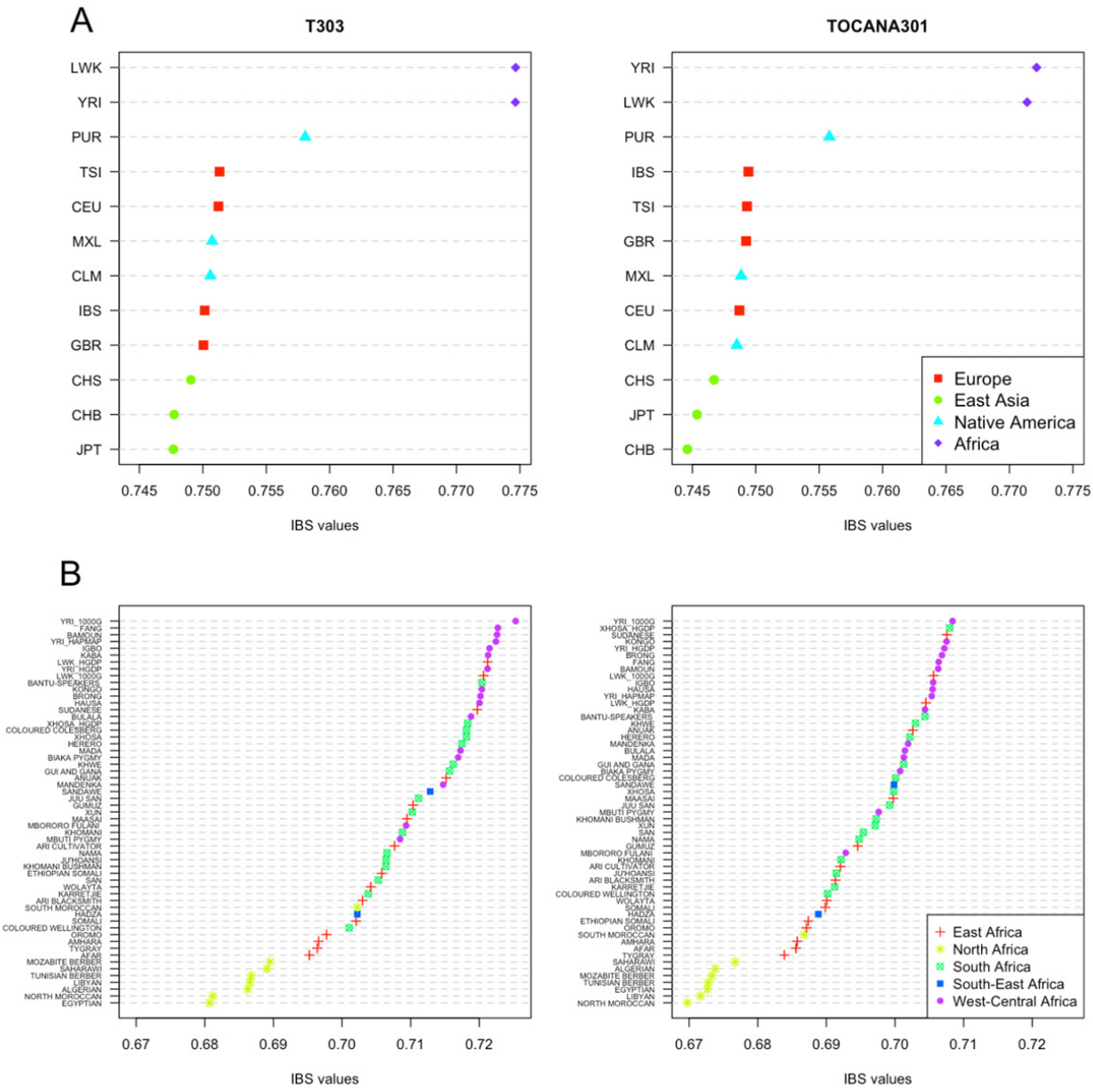


Figure 89: Average IBS values between the two 'Afro-Bolivians' and samples from 1000G (A) and a large dataset of Africans (B). Submitted to Heredity.

A PCA of the two 'Afro-Bolivians' against populations from The 1000 Genomes Project was done in order to determine main continental group affiliation (Figure 90). PC1 clearly separates sub-Saharans (LWK, YRI), East Asians (JPT, CHB, CHS), and Europeans (TSI, CEU, GBR) in three clusters. The remaining populations (CLM, MXL, PUR) occupy inter-medial positions by afilliating closest to the European cluster but dispersing into the direction of the African and the East Asian

pole which is in good agreement with their admixed ancestry history. For example, the most proximal cluster of Puerto Ricans is the European one, but show a main projection towards the African pole indicating African admixture. The two individuals from Tocaña are close to the sub-Saharan cluster with respect to both the PC1 and the PC2, respectively.

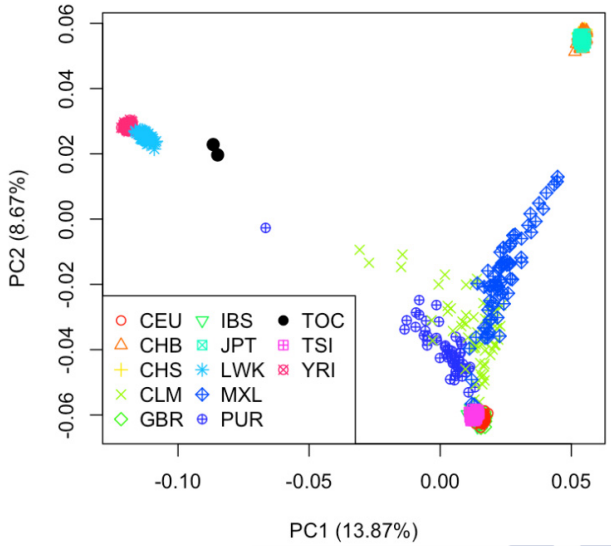


Figure 90: Principal Component Analysis of Tocañas against 1000 Genomes population. Submitted to Heredity.

In a second PCA analysis the two 'Afro-Bolivians' were investigated for their African intra-continental affiliation by using the large dataset of African populations (Figure 91). The PC1 and PC2 components separted the populations in a triangular pattern with South Africans at one pole, North Africans and Europeans at the second one, and West-Central Africas at the third pole. The South African pole, however, is higly dispersed indicating admixture with other African groups. East African populations are distributed between the clusters represented by West-Central and North Africa. Southeast Africans are close to the West-Central African cluster, but disperse slightly both to the South African pole and to the East African samples. The two Tocaña individuals are clearly located in the West-Central pole although they also show proximity to East and Southeast Africans.

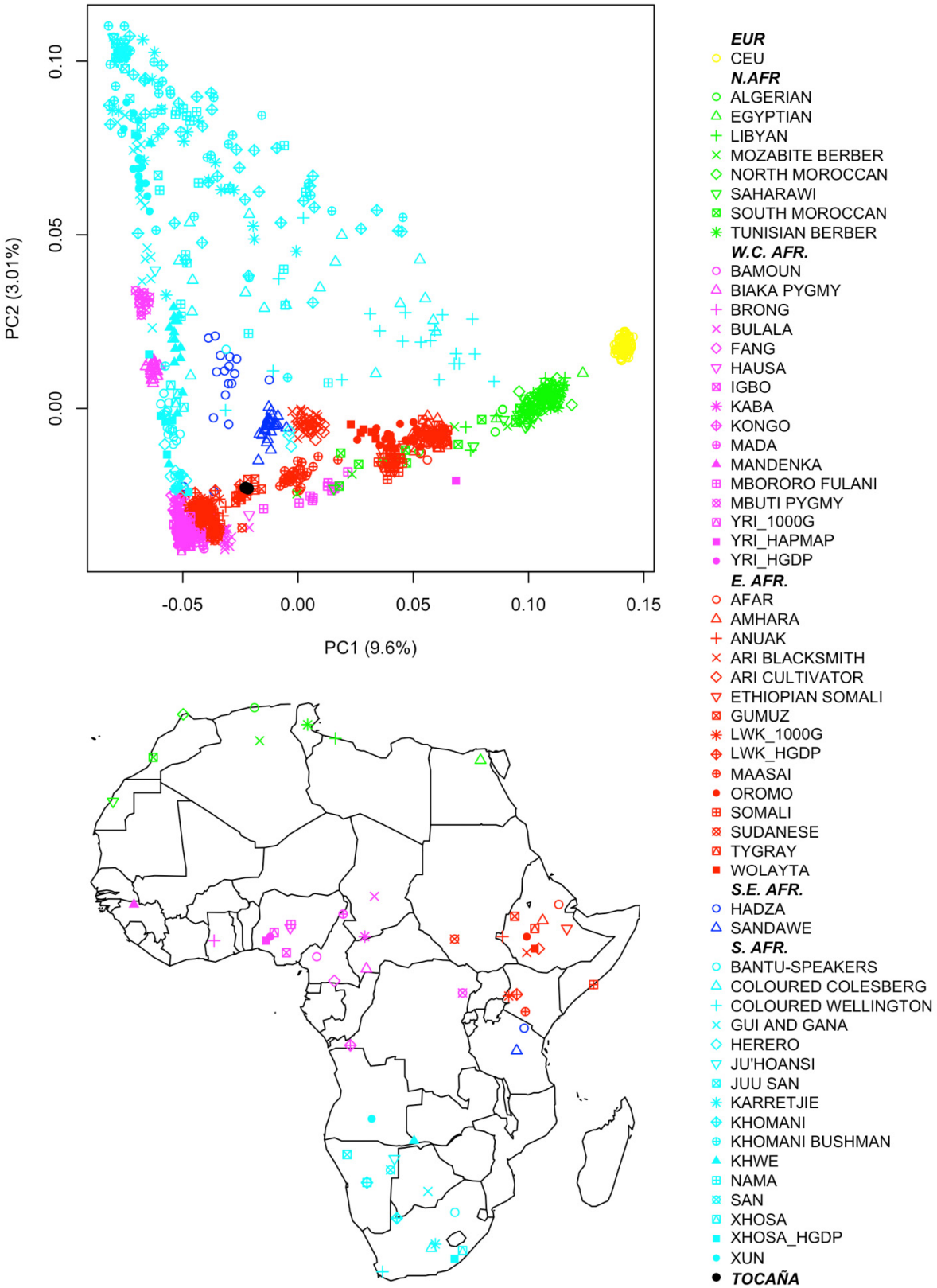


Figure 91: PCA analysis of the two 'Afro-Bolivians' in a large dataset of African groups. Submitted to Heredity.

In two PCA analyses with only West-Central or East Africans, the two Tocaña individuals were further explored for their genetic proximity with groups within both samples (Figure 92). In the PCA plot with the West-Central African population, the Biaka and Mbuti pygmies are clearly separated from the other groups (Figure 92A). Likewise, the Mbororo Fulami group clearly disperses from the other West-Central African groups. The two 'Afro-Bolivians' appear to be closest to the Mada (from Western Cameroon on the border to Nigeria), the Mandenka (Senegal), and the Bulala (Chad). In the PCA plot with the East African population, the two Tocaña individuals are most proximal to the Luhya from Kenya (Figure 92B).

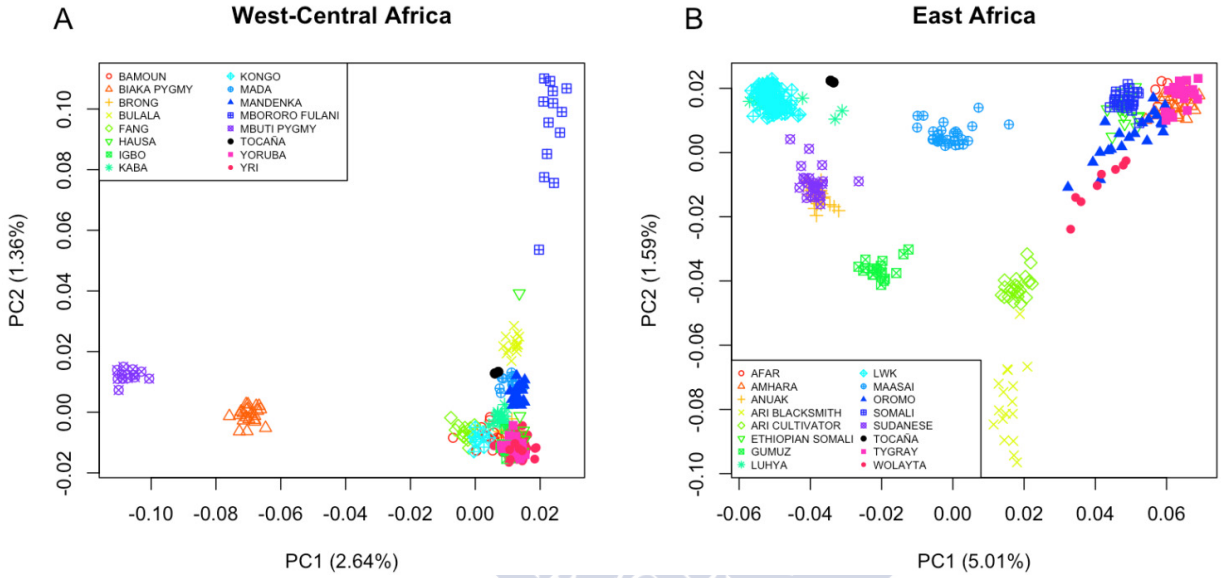


Figure 92: PCA analyses of the two 'Afro-Bolivians' with West African (A) and East African (B) samples. Submitted to Heredity.

6.3.2.2 Admixture analysis of the African component in 'Afro-Bolivians'

The genomic architecture of the two Tocaña individuals was analyzed with the software PCAdmix (Figure 93). Consistent with the previous PCA analysis, the two 'Afro-Bolivians' carry a predominant African component in their genome. The European and Native American components appear to be highly dispersed across the genome. Percentages of admixture in the two 'Afro-Bolivians' from PCAdmix were: 78.6% (#T303) and 79.0% (#Tocana301) African ancestry; 3.2% (#T303) and 2.7% (#Tocana301) European ancestry, and 18.1% Native American ancestry in both individuals.

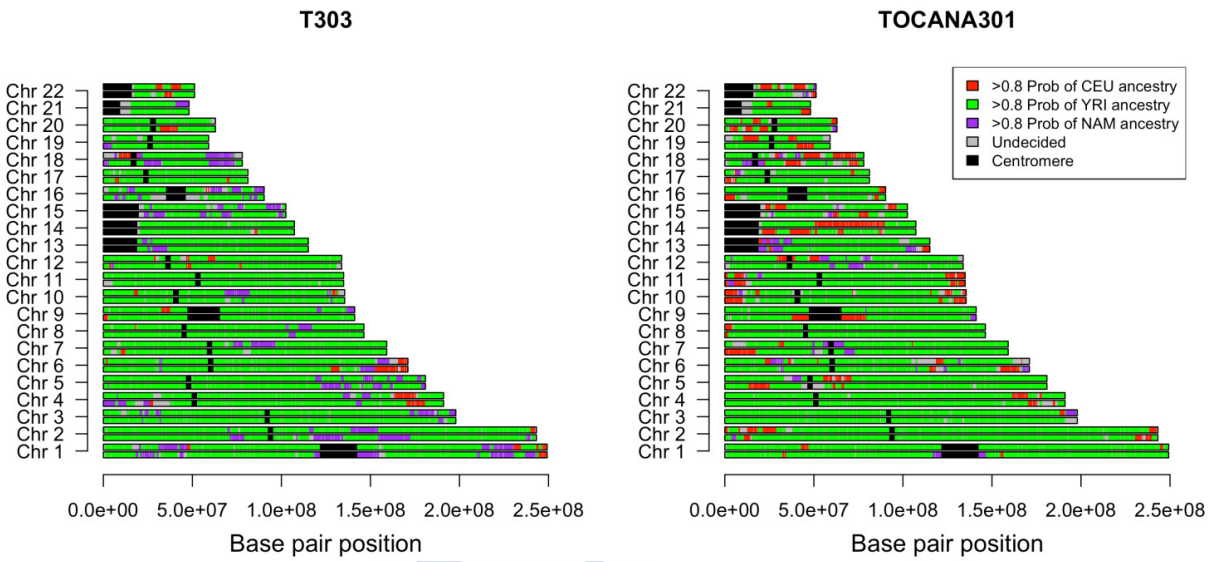


Figure 93: Genomic architecture and main continental ancestries of the two 'Afro-Bolivians' computed with PCAdmix. Submitted to Heredity.

In an ADMIXTURE analysis with the 1000G samples, the two 'Afro-Bolivians' again showed a principal African ancestry (Figure 94A). For an optimal value of K=5 the 'Afro-Bolivian' individual #T303 shows an African ancestry value of 81.4% and individual #Tocana301 a value of 79.1% (Table 22). Native American and European ancestry for #T303 came to 14% and 2.8% and for #Tocana301 to 6.4% and 9.3%. Percentages of admixture from both PCAdmix and ADMIXTURE were comparable although different reference datasets and algorithms were used to infer ancestry.

The ADMIXTURE analysis was also conducted with the large dataset of African populations (Figure 94B). At an optimal value of K=12 both individuals show one predominant component (42-49%) that is also prevalent in West-Central Africa (56% on average) especially in Yoruba (Nigeria, 77%), Hausa (Cameroon, 71%), Brong (Ghana, 79%), and Mandenka (Senegal, 77%) (Table 23). The second most predominant component in the two 'Afro-Bolivians' (23-25%) is most prevalent in Luhya from Kenya (46-55%) which is consistent with the previous PCA analysis that indicated a genetic proximity of the Tocaña individuals with the Luhya. This genetic component, however, is also shared with Southeast Africans and some South Africans whereby the highest value was reached (37%) by the Sandawe (Southeast Africa). The genetic component shared with Europeans came to 9-13% and the one shared with East Asieans (as subrogate population from Native Americans) to 6-9% in the two Tocaña individuals.

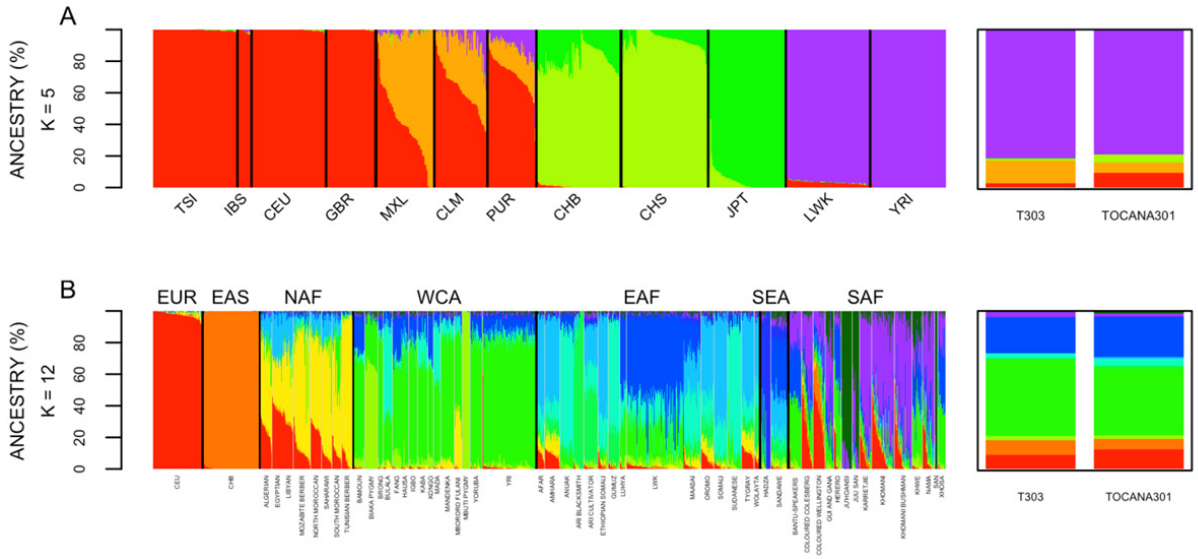


Figure 94: Individual ancestry of the two 'Afro-Bolivians'. Individual ancestry was computed by using (A) the 1000G samples only as reference populations and (B) a wide dataset of African populations together with Native Americans, Eurpeans (CEU), and East Asians (CHB). Based on ADMIXTURE software. Submitted to Heredity.

Table 22: Admixture analysis for K=5 and using the populations from The 1000 Genomes Project as population reference datasets. Cells are gradually collored from white (0%) to red (100%).

Population group	Continent	1	2	3	4	5
LWK	Africa	3.1	0.1	96.5	0.2	0.1
YRI	Africa	0	0	100	0	0
CLM	America	61.9	31.7	6.2	0.2	0
MXL	America	38.9	58.9	1.5	0.3	0.3
PUR	America	71.6	16.5	11.7	0.1	0.1
CHB	East Asia	0.4	0.2	0	19	80.4
CHS	East Asia	0	0	0	6.3	93.6
JPT	East Asia	0	0	0	96.2	3.8
CEU	Europe	99.7	0.1	0	0.1	0.1
GBR	Europe	99.8	0.1	0	0	0
IBS	Europe	98.8	0.3	0.7	0.1	0.1
TSI	Europe	99.7	0	0	0.1	0.1
T303	America	2.8	14.2	81.4	0.2	1.3
TOCANA301	America	9.3	6.4	79.1	0	5.2

Table 23: Admixture analysis considering different African datasets, CEU, and CHB and for the optimal K = 12. Cells are gradually collored from white (0%) to red (100%). EA=East Asia, E Asia=East Asia, NA=North Africa, SA=South Africa, SEA=Southeast Africa, WCA=West-Central Africa.

Population group	Region	1	2	3	4	5	6	7	8	9	10	11	12
Afar	EA	4.4	0.4	0.9	15.1	0.9	8.9	58.7	0.7	1.1	0.8	1.1	7.1
Amhara	EA	8.6	1	0.8	11.8	1.7	8.9	52.6	1.1	0.9	0.8	0.4	11.4
Anuak	EA	4.6	0.7	8	75	1.5	0	0.1	1.8	1	0.2	7.1	0
Ari Blacksmith	EA	89.9	0.7	0.4	2	0.7	0.3	3.9	0.7	0.4	0.2	0.2	0.6
Ari Cultivator	EA	45.2	3	0.9	16.4	3.7	0.4	25	3.2	1.4	0.7	0	0.1
Esomali	EA	4.2	0.6	0.5	20.5	0.9	2	63.1	1.5	1.2	1.1	2.9	1.6
Gumuz	EA	21.3	1.2	0	69.6	3.2	0.7	0.6	1	0.9	0.9	0.2	0.4
Luhya	EA	2.2	1	46.1	14.4	1.1	1.1	2.9	4	2.1	0.5	24.1	0.6
LWK	EA	1.6	1	55.2	13.4	1.1	0.7	1.5	3.7	1.2	0.5	19.2	0.8
Maasai	EA	5.1	2.2	20	32.2	3.4	1.2	24.6	1.1	1.4	0.9	3.5	4.5
Oromo	EA	13.1	1.3	1.3	16.5	2	6.4	46.5	1.6	1.2	0.7	0.3	9.2
Somali	EA	2.1	0.5	1.4	21.2	0.9	1.2	66.4	1.4	0.6	0.7	2	1.6
Sudanese	EA	0.9	1.1	6.1	77.8	0.9	0	0.7	2.3	0.5	0.2	9.6	0.1
Tygray	EA	6.7	0.8	0.9	11.6	1	10	53.9	1	0.9	0.8	1	11.6
Wolayta	EA	23.2	2.1	0.2	16.7	1.7	5.4	40.7	1.7	1.1	1.4	1.4	4.2
CHB	E Asia	0	0	0	0	0	0.2	0	0	0	99.7	0	0
CEU	Europe	0.1	0	0	0.1	0.1	97.5	0.3	0	0	0.6	0	1.2
Algeria	NA	0.8	0.9	1.1	1.3	0.7	23.5	13.3	0.9	0.4	1.1	5.9	50.1
Egypt	NA	0	0.3	1.6	0.3	0.1	37.3	28.5	0.3	0.2	1.5	0.8	29
Libya	NA	0.5	0.3	1.6	1.7	0.2	28.4	20.4	1	0.4	1.5	2.4	41.6
Morocco	NA	0.9	0.4	1	0.3	0.6	25.6	10.8	0.5	0.6	0.6	2.2	56.4
Mozabite	NA	0.8	0.5	1.7	3.9	0.5	11.3	3.7	1	0.6	0.7	5.6	69.7
Sahara Occidental	NA	0.3	0.3	1.9	1	1	7.5	13.9	0.6	0.4	0.7	6.3	66
South Morocco	NA	1.1	0.8	1.2	3	1	10.2	9	1.5	0.6	0.5	18.9	52.2
Tunisia	NA	1.5	0.4	0.2	0.9	0.5	4.8	0.2	0.3	0.4	1.2	0	89.4
Bantu-Speakers	SA	0.8	21.8	33.2	0.7	0.5	0.7	0.1	2.8	0.7	0.3	38.3	0.1
Coloured Colesberg'	SA	0.4	35.4	12.8	0.9	0.4	23.3	0.7	1.7	1.9	5.8	14.8	1.7
Coloured Wellington'	SA	0.8	18.1	9.6	1.6	1	33.6	3.4	1.4	0.6	17.6	10.6	1.7
Guighanakgal	SA	0.2	46.9	14.2	0.2	1	0	0.1	2.1	19.4	0.1	15.9	0
Herero	SA	0.6	7.3	36.5	1	0.7	6.2	0.1	4.6	2.4	0.3	39.8	0.3
Juhoansi	SA	0.1	8.3	0.9	0.3	0.1	0	0	0.2	88.9	0	1.3	0
Juu San	SA	0.2	21.2	2.6	0.6	0.1	0	0.5	1	72.2	0	1.6	0
Karretjie	SA	0.5	68.2	6	0.3	0.5	7.3	0.5	1.5	4.3	2.9	7.8	0.3
Khomani	SA	1	63.7	3.5	0.9	0.6	12.8	3.8	0.9	4.3	2.8	5	0.7
Khomani Bushman	SA	0.4	70.5	4.9	0.5	0.7	7.6	2.5	0.7	4.2	2.3	4.9	0.8
Khwe	SA	1.5	15.4	23.9	1.1	1.2	0.5	2.8	5.6	19.1	0.2	28.1	0.4

Nama	SA	1.8	57.8	8.6	1.6	1	8.5	5.8	0.7	3.6	0.7	8.5	1.3
San	SA	0	8	0	0	0.3	0	0	0.2	91.5	0	0	0
Xhosa	SA	0.4	20.1	36.4	1	0.7	0.3	0.3	2.1	1.9	0.3	36.1	0.3
Xun	SA	0.6	16.7	9.1	1.4	0.8	0.1	0.4	2.3	58.8	0.1	9.6	0.1
Hadza	SEA	1.2	1.1	11.6	3.1	72.6	0.1	2.5	1.5	1.4	0	4.7	0.2
Sandawe	SEA	9.7	4.6	36.9	5.8	6.9	0.3	23.8	4.1	5	0.7	1.3	0.9
Bamoun	WCA	1.1	1.1	21.3	4.1	1.5	0.3	0.3	6.2	1.7	0.5	61.9	0.1
Biaka Pygmy	WCA	0.4	1.6	0.9	0	0.3	0	0.1	65.4	1.9	0	29.4	0
Brong	WCA	1.7	1.3	5.9	3.8	2.2	0.2	0.8	2.1	2.2	0.4	78.6	0.7
Bulala	WCA	0	0.5	6.2	55.5	0.3	0.2	4.5	0.1	0.3	0.2	28.8	3.4
Fang	WCA	1.6	1.4	25.8	3.4	0.8	0.2	0.1	10.1	2.1	0.3	54	0.3
Hausa	WCA	1.3	1	12.6	7.7	1.1	0.2	1	1.7	0.9	0.4	70.5	1.5
Igbo	WCA	1.8	1.2	13.1	4	0.9	0.5	0.3	2.8	1.6	0.3	73	0.4
Kaba	WCA	0.6	1.5	11.1	19.2	0.5	0.3	0.2	3.7	1.1	0.3	60.5	1
Kongo	WCA	0.5	0.5	28.6	3.7	1	0	0.3	7.4	2.1	0.3	55.5	0.1
Mada	WCA	0.2	0.8	6.6	36.6	0.2	0.2	1.2	1.7	0.3	0.3	50.1	1.7
Mandenka	WCA	1.3	1.5	1	8.7	1.8	0.1	0.6	1.4	1.9	0.2	77.1	4.4
Mbororo Fulani	WCA	0.6	0.9	1.5	5.8	1.3	0.2	4.4	0.5	0.8	0.8	50.2	32.8
Mbuty Pygmy	WCA	0	0	0	0.3	0.1	0	0	99.6	0	0	0	0
Yoruba	WCA	1.8	1.3	10.5	3.5	1.4	0.1	0.6	1.7	1	0.4	77.3	0.3
YRI	WCA	1.4	1	9.5	3.3	1.6	1	0.8	1.7	1.2	0.3	77.9	0.4
T303	Tocaña	0.7	3.8	23.1	2.9	0	9	0	2.8	0	9.2	48.5	0
TOCANA301	Tocaña	1.7	1.8	25	4.2	0.5	12.6	1.3	2.5	1.7	6.4	42.3	0

6.3.2.3 The 'Native American' component in 'Afro-Bolivians'

The Native American genetic component in the two 'Afro-Bolivians' is only minor when compared to the African component and further complicated to evaluate by the fact that, firstly, the Native American reference sample is admixed to different extents [331], and secondly, the complex admixture of the two 'Afro-Bolivians'. The Native American component was analyzed by using the unmasked (whole reference dataset) and the masked (filtering out the non-Native American component of the individuals in the reference populations) data of the Native American groups analyzed by Reich and colleagues [331]. This admixture analysis clearly reveals the presence of a unique Native American genetic component in the two Tocaña individuals (Figure 95). However, the Native American component of the two 'Afro-Bolivians' is not shared with the Surui from Mato Grosso do Sul, Brazil, which is located close to the Llanos of Bolivia.

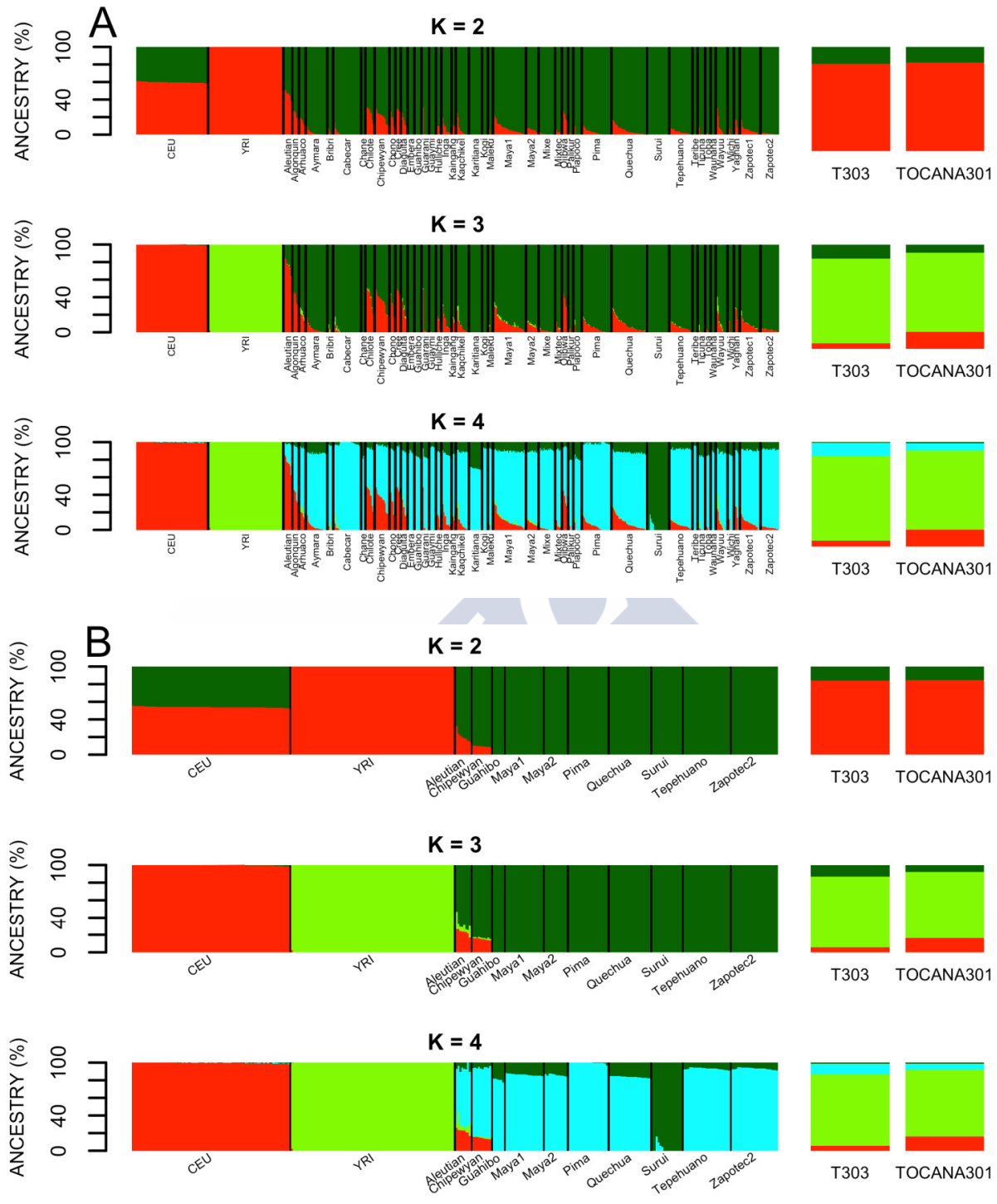


Figure 95: Admixture analysis of the two 'Afro-Bolivians' with an unmasked (A) and masked (B) Native American dataset. Based on ADMIXTURE software. Submitted to Heredity.

6.4 Gender-bias between the Y chromosome, mtDNA, and AIMs

6.4.1 Gender bias in the Bolivian population

Only samples for which information about Y chromosome, mtDNA, and AIMs were available were included in the following analyses. Out of the 226 individuals that were genotyped for the Y chromosome, 152 were also analyzed for the mitochondrial hypervariable segment (HVS-I) and the 46-AIM-Indels.

The Native American component for the AIMs predominates in the Bolivian populations (94.7%) and is followed by European ancestry (5.3%) (Figure 96, Table 24). European ancestry in the Bolivian population can be completely attributed to the Llanos (9.6%) as European ancestry was not detected in the other two regions. African ancestry was completely absent.

Native American ancestry was slightly higher for the mtDNA (98.7%) than for the AIMs, whereby a small amount of African ancestry (1.3%) was detected on the mtDNA. One African-specific mtDNA haplogroup was observed both in the Llanos and in the Sub-Andeans. However, as sample number of the Llanos (n=83) was higher than for the Sub-Andeans (n=9), African ancestry was higher in the latter (11.1%) than in the former (1.2%). European ancestry was completely absent in Bolivia on the mtDNA.

The Y chromosome, in contrast, exhibits principally a Europen ancestry in the Bolivian population (62.5%) followed by Native America (35.5%) and Africa (2%). Whereas Native American ancestry was highest in the Llanos (47%), European ancestry (88.9%) predominates in the Sub-Andean region.

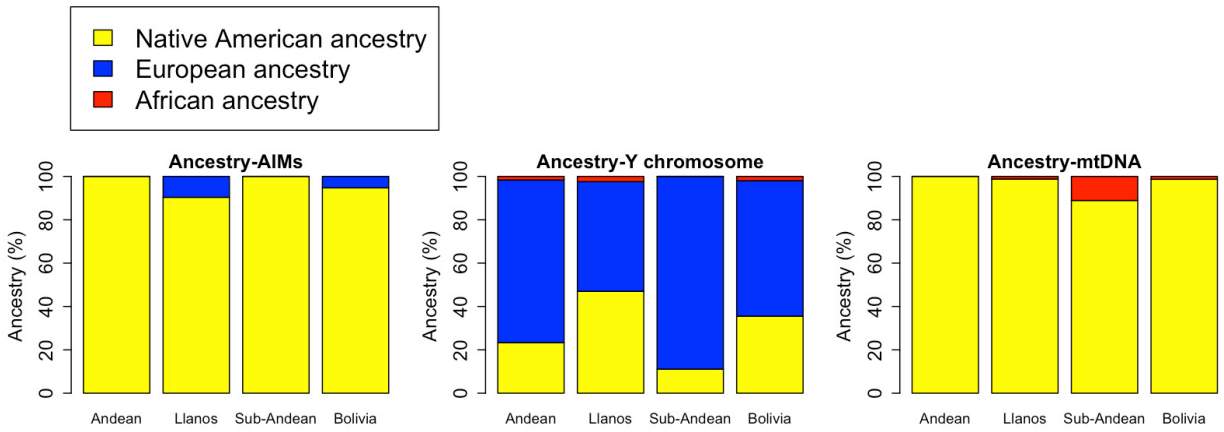


Figure 96: Ancestry proportions based on AIM-Indels, mtDNA, and Ychromosome of only males in Bolivia including Yungas (here as Sub-andeans). Published in [350] whereby estimates for the Y chromosome are taken from [350] and for the mtDNA and AIMs from [333]. Ancestry-AIMs based on STRUCTURE software.

Table 24: Ancestry estimates for mtDNA, Y chromosome, and AIMs in the Bolivian population.

Marker	*Location*	*African*	*Native American*	*European*	*n*	*Publication*
mtDNA	Bolivia	2 (1.3%)	150 (98.7%)	0 (0%)	152	[333]
	Andean	0 (0%)	60 (100%)	0 (0%)	60	[333]
	Sub-Andean	1 (11.1%)	8 (88.9%)	0 (0%)	9	[333]
	Llanos	1 (1.2%)	82 (98.8%)	0 (0%)	83	[333]
Y chromosome	Bolivia	3 (2%)	54 (35.5%)	95 (62.5%)	152	[350]
	Andean	1 (1.7%)	14 (23.3%)	45 (75%)	60	[350]
	Sub-Andean	0 (0%)	1 (11.1%)	8 (88.9%)	9	[350]
	Llanos	2 (2.4%)	39 (47%)	42 (50.6%)	83	[350]
AIMs	Bolivia	0	0.947	0.053	152	[333]
	Andean	0	1.0	0	60	[333]
	Sub-Andean	0	1.0	0	9	[333]
	Llanos	0	0.904	0.096	83	[333]

Additionally, 6 individuals were previously typed for the ancestry panel LACE [259], and thus, compared to Y chromosome and mtDNA ancestry, but also to the 46-AIM-Indels (Table 25).

Table 25: Gender-bias between Y chromosome, mtDNA and AIMs of 6 individuals from La Paz(Andeans). 46-Aim-Indels and mtDNA taken from [333], 446-AIM LACE from [259]. Published in [350], Table 4. Na=Native American, A=African, E=European.

Region	*Sample*	*Y-Chr*	*mtDNA*	*LACE 446-AIM panel (%)*			*AIM-Indels panel (%)*		
				NA	A	E	NA	A	E
Andean	337COROICO	R1b1(xM269)	A2	100	0	0	90	0	10
Andean	422COPACABANA	R1b1(xM269)	B4	100	0	0	82	0	18
Andean	409COPACABANA	R1b1(xM269)	B4	100	0	0	90	0	10
Andean	433COPACABANA	R1b1(xM269)	D4h3a4	100	0	0	75	0	25
Andean	414COPACABANA	R1b1(xM269)	B4	100	0	0	78	0	22
Andean	418COPACABANA	R1b1(xM269)	B4	98	0	2	80	0	20

According to the LACE panel all individuals are virtually completly of Native American ancestry (0.0099 ± 0.0082). In contrast, the AIM-Indels indicates a less Native American (0.8250 ± 0.0625), but also some European ancestry (0.1800 ± 0.0678), however, with a large standard deviation. An elevated European ancestry estimate for the AIM-Indels does not necessarily reflect European descent as this could be due to the smaller number of AIMs used for the analysis [266]. As the LACE panel encompasses more AIMs than the 46-AIM-Indels, the LACE panel is expected to perform more precisely in estimating ancestry. The Y chromosomes all belong to European-specific haplogroups (R1b1(xM269), while the the mtDNA haplogroups all show typical Native American ancestry.

6.4.2 Gender bias in the Yungas population

In the previous chapter gender bias was analyzedwith all Bolivian males who were genotyped at the Y chromosome, mtDNA, and AIMs (excluding the Yungas). In this chapter gender bias is further analyzed within the Yungas valleys.

Both mtDNA and autosomal AIMs indicate a principal Native American ancestry in Bolivia (98.6%, and 77.7%, respectively) while in the Y chromosome European ancestry is predominant

(65.5%). The African ancestry in the Bolivian population in general was observed to be low (Figure 96, Figure 97, Table 26).

In the Yungas valleys a strong gender bias was observed, similar to the one found in the remaining Bolivian population. However, a siginificant difference to the rest of the Bolivian population is the increased introgression of African ancestry in the Yungas valleys, especially in the Nor Yungas in which the community of Tocaña is situated. The African ancestry in the Nor Yungas on the autosomal AIMs came to 13.4%, for the mtDNA to 19.6%, and for the Y chromosome to 8.5%. The elevated African ancestry in the Nor Yungas can be mainly attributed to Tocaña as in this community African ancestry reached 55.8%, 84.2%, and 44.4% for the AIMs, mtDNA, and Y chromosome. An elevated European ancestry in the Yungas is like in the remaining population found on the Y chromosome. European ancestry was higher in the Yungas (76.6%) than in the remaining Bolivian population (65.5%).

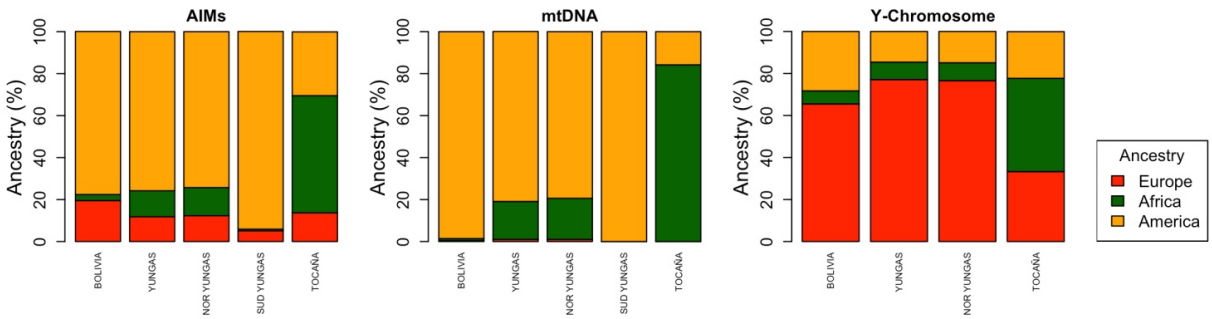


Figure 97: Ancestry proportions based on AIM-Indels, mtDNA, and Y chromosome in the Yungas compared to Bolivia. Estimates for Bolivia are taken from [333]. AIMs based on ADMIXTURE. Submitted to PLos ONE.

Table 26: Ancestry estimates for mtDNA, Y chromosome, and AIMs in the Yungas.

Marker	*Location*	*African*	*Native American*	*European*	*n*	*Publication*
mtDNA	Bolivia	7 (1%)	709 (98.6%)	3 (0.4%)	719	[333]
	Yungas	19 (18%)	85 (81%)	1 (1%)	105	Submitted to PLos ONE
	Nor Yungas	19 (19.6%)	77 (79.4%)	1 (1%)	97	Submitted to PLos ONE
	Sud Yungas	0 (0%)	8 (100%)	0 (0%)	8	Submitted to PLos ONE
	Tocaña	16 (84.2%)	3 (15.8%)	0 (0%)	19	Submitted to PLos ONE
Y chromosome	Bolivia	14 (6.2%)	64 (28.3%)	148 (65.5%)	226	[350]
	Yungas	4 (8.3%)	7 (14.6%)	37 (77.1%)	48	[350]
	Nor Yungas	4 (8.5%)	7 (14.9%)	36 (76.6%)	47	[350]
	Sud Yungas	0 (0%)	0 (0%)	1 (100%)	1	[350]
	Tocaña	4 (44.4%)	2 (22.2%)	3 (33.3%)	9	[350]
AIMs	Bolivia	0.029	0.777	0.195	598	Submitted to PLos ONE
	Yungas	0.124	0.758	0.118	105	Submitted to PLos ONE
	Nor Yungas	0.134	0.743	0.123	97	Submitted to PLos ONE
	Sud Yungas	0.008	0.942	0.051	8	Submitted to PLos ONE
	Tocaña	0.558	0.304	0.137	19	Submitted to PLos ONE

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

7 Discussion

7.1 Continental ancestry in the Bolivian population

7.1.1 Continental ancestry based on mtDNA

Most part of the Bolivian population clearly shows a principal Native American origin on the mtDNA (98.4%) with haplogroup B being the most prevalent lineage (52.6%) followed by C (21.9%), A (14.3%), and D (9.6%). African (1%) and European (<1%) ancestry, in contrast, was only minimal in the Bolivian population. In the Yungas valleys, however, African ancestry came to a value of 18% which is principally concentrated in the community of Tocaña (16 out of 19 individuals, 84%). In the Sud Yungas African-specific mtDNA lineages are absent whereas in the Nor Yungas 19% of the haplogroups could be assigned to either L0 (5%), L1 (5%), or L3 (9%), respectively. The other part of the Yungas (81%) could be assigned again to a typical Native American haplogroup. Only one individual (1%) belonged to a European-specific mtDNA haplogroup in the Yungas.

The high African ancestry on the mtDNA most probably can be attributed to the history of the Yungas valleys. The first enslaved people from Africa reached at first Potosí early during the TAST, but during the 18th and 19th century African descendant people from the high cold mountainous regions in the South of Bolivia migrated towards the North into the tropical valleys of the Yungas where the Spanish colonists established a 'hacienda system' which was based on slave labour. In the Yungas valleys both African descendant and indigenous people worked for a mostly absentee estate holder. After abolition of slavery in Bolivia, many African descendant people remained in the Yungas valleys. Hence, African ancestry on the mitochondrial is higher in the Yungas when compared to other regions of Bolivia.

In other populations of Bolivia the Native American component is most prevalent whereby the haplogroup composition varies remarkably among populations. In most departments (La Paz, Chuquisaca, Cochabamba, Santa Cruz) haplogroup B was the predominant haplogroup, but in others haplogroup C (Beni) or A (Pando) were the most frequent Native American lineage. Correspondingly, haplogroup B is prevailing in the Andeans and the Sub-Andeans, but not in the Llanos where haplogroup C exceeds the other haplogroups. The major differentiation of variability within Bolivia occurs between the eco-geographical region and departments.

Some mtDNA lineages found in Bolivia are also present in neighbouring areas like in Andean Peruvian Quechuas [335] which might be the results of the common demographic history during the Inca's Empire period as discussed in Aymaras and Quechuas by Gayà-Vidal and colleagues [281].

By sequencing the whole genome of Bolivian individuals, three new Native American lineages could be detected: A2ah, B2o, and B2o1 (now B2o, B2o1, and B2o1a according to Phylotree 16). All three sub-lineages seem to be rare in Bolivia (ranging from 0.5% to 1%) and are only infrequently found on the continent. However, all three clades show a remarkably diversity within Bolivia indicating that they probably evolved locally after their arrival from other neighbouring regions.

In the Nor Yungas the Native American component on the mtDNA reached 79% withB2/B4 as the most prevalent haplogroup followed by C1, A2, and D1/D4. The Sud Yungas, in contrast, is represented only by Native American-specific haplogroups with C1 being the most prevalent haplogroup, B2/B4 the second, and D1 the third one.This pattern of haplogroup distribution in the Yungas is similar to the one observed in the remaining population. However, it has to be noted here that the proportions of haplogroup composition in the Sud Yungas is based on only 8 individuals which contrasts with 97 individuals from the Nor Yungas.

The African-specific L haplogroups found in the Yungas belonged principally to haplogroup L3 followed by L0 and L1. In the remaining population only 1% mtDNAs were of an African-specific haplogroup. In other South American populations African ancestry on the mtDNA reached higher values[283, 285, 286]. The lower African ancestry might be due to the fact that the TAST was not that active in Bolivia than in other regions. Indeed, Bolivia had no immediate access to the sea and many slaves had to migrate long distances in order to reach the place where they were forced to work.

7.1.2 Continental ancestry based on the Y chromosome

European ancestry on the Y chromosome in the whole Bolivian population (including males from the Yungas) reached 62.5%, and thus, is comparable to other South and Central American populations like Alagoas, Northeastern Brazil (95%) [362], 'Caucasian-Mestizos' of Colombia (49%) [363], El Salvador [364], 'Nicaraguan-Mestizos' (64.4%) [315], and Córdoba, central Argentina (97%) [365]. The Native American ancestry came to only 35.5% in Bolivia. The African

ancestry with 6% is even lower. The high European ancestry on the Y chromosome can be mostly explained by the fact that the colonization of the Americas was ruled and directed by principally European males.

The most pronounced differencein Bolivia can be observed within the eco-geographical region the Llanos versus the Andeans and Sub-Andeans. European ancestry on the Y chromosome in the Llanos reached only 46% whereas in the Andeans and Sub-Andeans European ancestry came to 75% and 86%, respectively. Correspondingly, Native American ancestry was highest in the Llanos (47%), lower in the Andeans (19%) and lowest in the Sub-Andeans (7%). The higher European ancestry in the Andeans and Sub-Andeans and the higher Native American ancestry in the Llanos were also observed in an admixture analysis. The highest value of European ancestry in the Sub-Andeans might be probably due to the 'hacienda system'. The Europeans arrived at first primary in the region around Potosí but established a 'hacienda system' in the Yungas which belongs to the Sub-Andeans.

Surprisingly, most African ancestry on the the Y chromosome was observed not only in the Andeans (including the Yungas), but also in the Llanos (each 7%). African ancestry in the Andeans was only slightly lower (6%). The Sub-Andean region comprises the Yungas, but the African ancestry is not significantly higher and comparable to the estimates found in the two other eco-geographical regions.

However, the high European ancestry in the Sub-Andean (including the Yungas) region possibly reflects the fact that the Europeans were most active in this region. The low African ancestry values on the Y chromosome might possibly reflect the high mortality rate of African males during slavery or the arrival of more African descendant women than men in the Yungas.

Nonetheless, as the 'hacienda system' was mainly ruled by European absentee estate holder one would expect a pronounced African and Native American ancestry in the region where it was established. Indeed, this was observed in the Yungas especially in the community of Tocaña. African ancestry on the Y chromosome came to 44%, whereas European and Native American ancestry reached only 33% and 22%. As mentioned above African ancestry on the mtDNA in Tocaña reached 84%, and thus, is almost twice as much as on the Y chromosome. In this context, it is unclear why the African ancestry on the mtDNA is higher. One reason could be that more females than males worked on the estates as slaves or it could be possible that the mortality rate of

males as slaves was especially high. According to the TAST database about 64% of all enslaved people that embarked in Africa were male [1].

7.1.3 Continental ancestry based on autosomes

As it was seen for the mtDNA and Y chromosome, Native American ancestry in the Bolivian population (without Yungas) is again predominant. Native American ancestry (73.3%) is almost three times higher than the European one (25.8%). African ancestry was less than 1%. Whereas in La Paz Native American ancestry was especially high, in Santa Cruz European ancestry was most prevalent. Correspondingly, Native American ancestry was highest in the Andeans and lowest in the Llanos.

In the Yungas, in contrast, Native American ancestry was again predominant (76%), but African ancestry reached 12% which contrasts 1% in the remaining Bolivian population. African ancestry is especially concentrated in Tocaña as in the Nor Yungas African ancestry came to 13% but falls to 3% when Tocaña is excluded from the analysis. In Tocaña African ancestry was most prevalent (56%).

In the comparison of the Yungas valleys with the remaining Bolivian population, it can be clearly seen that the genetic heritage that the TAST left in Bolivia is unequally distributed in the country and concentrated in the region where slavery was common. Information about continental ancestry on uni- and biparental markers is of relevance in an historical context because it can be concluded that European ancestry on the autosomes in the Bolivian population came principally from European males whereas African ancestry on the autosomes came from both genders but rather more from African women than from African men.

7.2 Geographic origin of 'Afro-Bolivians'

Based on the eight mitogenomes, geographic origin of 'Afro-Bolivians' could not be clearly resolved although some tentative geographic inferences can be done.

Apparently no historical documentation is available about the African origins of modern-day 'Afro-Bolivians'. First enslaved African people were documented as early as 1549 in Potosí [11] indicating that the arrivals of enslaved people to Bolivia began early during the TAST. In Africa the

TAST began in the Senegambia region. Therefore, as most slave deportations until the first decades of the 17th century were mainly destined to Spanish colonies, especially to Mexico, but also Lima (Peru) where slaves were further sold to Upper Peru (Bolivia), Chile or Quito [2], Senegambia might have been a possible African origin of nowadays 'Afro-Bolivians'. Senegambia became gradually less important after the beginning of the 17th century as other regions, especially West-Central Africa, became more and more important for the European traders [1]. Indeed, some sources indicate that the Congo and Angola could have contributed slaves to Bolivia [62] which is further supported by the apparently Congolese origin of the surnames Angola and Maconde in 'Afro-Bolivian' communities [276]. Furthermore, it was also documented in Charcas (Bolivia) that ethnicities arriving in Upper Peru between 1650 and 1710 came apparently from West and Central Africa (including principally females), but also from Upper Guinea [4]. Out of all disembarkations in Spanish Mainland America, 57% can be attributed to West-Central Africa and 19.4% to Senegambia [1].

Enslaved Africans who embarked in Senegambia and arrived in Upper Peru principally entered the Americas in Cartagena de Indias (Colombia), whereas those embarked from West-Central Africa mainly had to first pass Río de la Plata (present day Uruguay / Argentina). African enslaved people arriving at Río de la Plata, predominantly from Angola, had to either remain in the region where they disembarked or were further transshipped to Upper or Lower Peru, but also to Chile [4]. As Upper Peru had no immediate access to the sea, other principal ports of arrival were Recife, Salvador-Bahía and Río de Janeiro in Brazil, and Panamá and El Callao in Peru [11, 278].

However, mitogenomes observed in 'Afro-Bolivians' in this thesis cannot be traced back to one single geographic location in Africa. No sample from the Senegambia or an adjacent region (except one individual from Sierra Leone belonged to L3d1b3 like #Toc294), were observed in the same haplogroup as the 'Afro-Bolivians'. A West or West-Central genetic impact on 'Afro-Bolivian' communities cannot be excluded in Bolivia as, for example, the 'Afro-Bolivian' individuals #Toc293, #Toc295, and #Toc304 form the sub-branch L1c3b1a, including one Pygmy and one Bantu individual. Sub-branch L3d1a1a with #Toc299 also comprises one individual from Nigeria/Chad (Yoruba), plus one Pygmy. A sub-branch, L3d1b3, with #Toc294 includes two individuals from Burkina Faso and one from Sierra Leone. Nonetheless, 'Afro-Bolivians' share haplogroups not exclusively with West-Central African samples.

The data indicate that the genetic impact on 'Afro-Bolivians' came rather from East or Southeast Africa, but surprisingly also from the Middle East. Individual #Toc305 belongs to the same haplogroup (L0a2a2a) as three individuals from Oman, Yemen, and Saudi-Arabia, together with one individual from Mozambique, three from Zambia, and two from Kenya. Likewise, 'Afro-Bolivian' #Toc299 can be found in the same haplogroup (L3d1a1a) with two samples from Kenya and one from Somalia, and one sample from the Middle East, namely from Yemen. The individuals #Toc293, #Toc295, and #Toc304 also form a sub-branch (L1c3b1a) together with one sample from Kenya and one from Zambia.

The Analyses of admixture based on a large dataset of African mtDNA control region sequences are very reassuring with these phylogeographic inferences. Thus, although the main contribution to 'Afro-Bolivians' comes from the West-Central and Southwest (all together >50-60%), a significant proportion of African ancestry comes from Southeast (>20%) and from East Africa (>8%).

Southeast Africa contributed an increasing number of enslaved people especially towards the end of the TAST. Slave embarkation in Southeast Africa began between 1661 and 1665, but developed to the second largest slave provider after West-Central Africa between 1816 and 1820 after the British abolished slave activities in their English colonies in 1808 [1, 2]. Most African enslaved people in Southeastern Africa, that is, a number of 112.800 embarked on vessels between 1821 and 1830. The principal destinations of vessels from Southeastern Africa was Brazil (70.6%), followed by the Carribean (19.9%) [1]. The reason behind the increasing number of African enslaved people in Southeastern Africa towards the end of the TASTwas the British effort to force abolition of slavery, but which became noticele on the Atlantic coast only after 1840 and even later in East Africa [2].

The slave trade in East Africa was more complex than the TAST on the Western and West-Central African region. Enslaved people in East Africa were not only destined for the TAST, but also to the Trans-Saharan trade and to the trade across the Red Sea and Indian Ocean. The slave trade in East Africa can be divided, firstly, into a trade that led to northern states such as Arabia, Persia, and India, secondly, a trade to the Americas and the Mascarenes, and thirdly, a coastal trade. The coastal trade along the East African coast might explain the Middle Eastern origin of 'Afro-Bolivians'. An estimate of about 400,000 slaves in the 18th century in the East African slave trade accounts for the movement of about two thirds who were destined for traditional markets in the Muslim world and the rest who was either sent to the Mascarene Islands or to the New World. About 1,650,000 enslaved people were involved in this trade in the 19th century. Half of them

remained in East Africa and the other half were forced to migrate to the Americas or to other places in the Indian Ocean [3]. Most African enslaved people from Eastern and Southeastern Africa left the continent for the Americas on ships in Mozambique. This region was particularly important between 1836 and 1840 with around 35.000 slaves [1]. Mozambique became an especially important source of African slaves for Brazilians and the Portuguese [4] and provided African enslaved people as early as 1661 [1] although other sources reported African enslaved people in Bolivia from Mozambique already in 1535 [275].

At around the same time in Bolivia slavery was not yet abolished in the country. The emancipation from the Spanish colonists took place in 1825, but did not abolish slavery in Bolivia although there were attempts to realize it. Slavery was officially abolished in Bolivia only after the agrarian reform in 1952 [11, 62]. Therefore, due to the late ending of the TAST in Southeastern Africa and the late abolition of slavery in Bolivia, some enslaved African people might have continued to arrive in the country towards the end of the TAST. Consequently, a genetic impact from East and Southeastern Africa on 'Afro-Bolivians' might seem reasonable. Brazil, Cuba, and Puerto Rico were the principal destination for African enslaved people at the end of the TAST [2]. As Bolivia had no immediate access to the sea, principal ports of slave arrivals in Bolivia included ports in Brazil like Recife, Salvador-Bahía and Río de Janeiro[11, 278]. Thus, enslaved African people who reached Bolivia at the end of the TAST might have entered the Americas via Brazil. However, it is unclear if new African enslaved people indeed arrived towards the end of the TAST in Bolivia.

A genetic input from Near Eastern countries in 'Afro-Bolivians' is rather unclear. It might indirectly reflect a connection between the TAST and the Trans-Saharan Slave Trade which was heavily influenced by the Muslim trade [3]. Most Muslim influence on the TAST was in the Senegambia region although the Bight of Benin, the Gold Coast, and the Bight of Biafra as well had profound connections to the Muslim slave trade. In contast, in the Bight of Biafra the Muslim influence was minimal and in West-Central Africa literally absent. Nonetheless, apparently up to 40% of all slaves that had to leave the African continent for America between 1500 and 1800 can be attributed to the Muslim trade although indiscriminate enslavement of free Muslims to the New World or across the Sahara displeased many Muslims. Muslim traders were an important agent during the TAST as they played a pivotal role in establishing a link between the far interior and the states that exported slaves. For example, on the Bight of Benin or 'Slave Coast' the states were more centralized, and thus, Europeans were confined to the coast for trade and did not raid for slaves which contrasts West-Central Africa where the states were more fragmented and Europeans

could invade to the interior. There were also six major routes that linked sub-Saharan Africa to North and East Africa like, for example, the Muslim trade routes from ancient Ghana to Morocco [3]. However, based on the genetic and historical information available so far, only vague conclusions can be made about a connection between the Muslim trade and 'Afro-Bolivian' communities.

The results from whole genome SNP typing support the results gained from the 'Afro-Bolivian' mitogenomes because the two 'Afro-Bolivian' genomes seem to be a mosaic of different ancestries. The main ancestry proportion is sub-Saharan (83-84%) and the remaining European and Native American (16-17%).

West-Central Africa appears to be most represented in 'Afro-Bolivians' (42-49%) with Yoruba (Nigeria) being the most prominent population that contributed to the two individuals of Tocaña whereby other populations from the Atlantic coast (like Ghana, Senegal, Nigeria) also might have contributed to the genomes of 'Afro-Bolivians'.

However, according to the statistical analysis East (or Southeast) Africa also seem to contribute significantly to the 'Afro-Bolivian' genomes accounting for about 22-25%. Based on the admixture analysis, the Luhya from Kenya, a Bantu ethnic group, make up by far the largest contribution among different East African populations to the two 'Afro-Bolivian' individuals. The Luhya is one of the largest ethnic tribes in Kenya that live in the western region of the country. As the Luhya is a Bantu group, they probably migrated together with other Bantus from Central and West Africa during the Bantu expansion towards Southeast Africa [140]. Furthermore, the Luhya also live in nowadays Uganda and Tanzania which represent a territory that served as an important passageway during the Bantu expansion between West and Southeast Africa. It is also important to note that the two populations from Southeast Africa in the datasetused in this thesis, Hadza and Sandawe, share an important percentage of ancestry with the Luhya, that is, 12% and 37%, respectively. As East Africa did not directly contribute enslaved people to the TAST [1], there might be two possible explanation why East Africans and 'Afro-Bolivians' share genetic components. On the one hand it might be possible that there is a lack of historical documentation about the connection between East Africa and the TAST. On the other hand the East African component in the two individuals of Tocaña is the one appearing in Southeast Africans. As mentioned elsewhere in this thesis, Southeast Africa provided a significant number of enslaved people to the Americas towards the end of the TAST. The second hypothesis seems to be more likely because African enslaved people from

Mozambique were indeed documented in Bolivia [275]. Furthermore, as Bolivia had no immediate access to the sea principal ports of arrivals were located in other parts of South America, including Brazil [11, 278]. Brazil still received a significant number of enslaved people towards the end of the TAST, and that especially from Southeastern Africa [1].

The results from the mitogenomes and whole genome SNP genotyping are further supported by the Y chromosome which show profiles that are also distributed in Southeastern Africa.

7.3 Gender bias in the Bolivian population

In the Bolivian population a strong gender-bias was observed. In general the Native American component predominates in the Bolivian populations, butan important European component was observed on the Y chromosome. That is, 62.5% of all Y chromosome lineages were of European ancestry in the Bolivian populations (including the Yungas) whereby the European component was most prevalent in the Sub-Andean region. A predominant European ancestry on the Y chromosome (77%) was also observed in the Yungas population. However, European ancestry on the Y chromosome in Tocaña was reduced severely due to a high introgression of Y chromosomes of African ancestry.

The mtDNAs of Bolivians (excluding the Yungas) were mainly of Native American origin. But due to the history of the Yungas, an increased introgression of African ancestry could be found on the mtDNA in the Bolivian population mainly in the Yungas. African ancestry was absent on the mtDNA in the Sud Yungas, but not in the Nor Yungas as the community of Tocaña contributed most of the increased African ancestry on the mtDNA in this population. In Tocaña African ancestry on the mtDNA reached 84%, on the Y chromosome 44%, and on the autosomes 56%.

In the Bolivian population it could be revealed that African ancestry is higher for the mtDNA than for the Y chromosome. This finding might be the result of a higher mortality rate of African males during slavery and/or the arrival of more females than males of African descent in the Yungas. Unfortunately there are apparently no historical documents available on this aspect of the TAST. It is fact that the mortality rate of African enslaved people was especially high at the beginning of the TAST in Potosí where enslaved people had to work under harsh conditions [11]. However, no information is available about the mortality rate of enslaved people in the Yungas. Possibly the mortality rate was lower in the Yungas valleys than in the high altitude and cold mountainous

Andeans. It is also not known if more African descendant females than males arrived in the Yungas which also might explain the higher African ancestry on the mtDNA. It was documented in Charcas (Bolivia) that African enslaved people arriving in Upper Peru between 1650 and 1710 came also from West and Central Africa including principally females [4].

However, according to the TAST database between ~63-65% male African enslaved people disemparked in Spanish American mainland. A similar pattern is documented for Brazil (~62-79% disembarked males), where many African enslaved people, which were forced to migrate to nowadays Bolivia, entered during the TAST.

The introgression of European Y chromosomes in Tocaña is lower when compared to the Yungas (including Tocaña) or to the whole Bolivian population. The 'hacienda system' was mainly ruled by absentee estate holder. Maybe because of the common absence of European males in the 'Afrobolivian' communities, African ancestry on the Y chromosome could become more prevalent when compared to other regions in Bolivia.

A sex biased mating in the Americas involving principally European men and Native American or African women was observed already in other South American populations. A prevalent Native American ancestry on the mtDNA was observed, for example, in Uruguay [367] and Belém of the Amazon region [368] whereas African-specific mtDNAs were frequently found in African Brazilians from São Paulo [369]. According to self-reported ethnicitiy in Colombia, 'mulatos' and'Afro-Colombians' carried mainly an African-specific mtDNA while 'mestizos' rather belong to a Native American mtDNA haplogroup [57]. A genetic impact on the Y chromosome by European males was detected, for example, in the Caribbean [370] and in Brazil [85].

8 Conclusions

1- The principal genetic component on the autosomes in the Bolivian population is of Native American ancestry. However, in the Yungas a high introgression of African ancestry was observed especially in the community of Tocaña.

2- On the mtDNA a similar pattern was observed. Most Bolivians carry a Native American mtDNA haplogroup but in the Yungas again a higher proportion of African ancestry wasobserved and this was especially high in the community of Tocaña. Most individuals of Tocaña belonged to an African-specific L-haplogroup; L(xM,N).

3- On the Y chromosome a high introgression of European haplogroups was observed in the Bolivian population, which is consistent with historical records, as the TAST was primary influenced in the Americas by European males. African ancestry on the Y chromosome was high again in the Yungas, especially in Tocaña, but proportions of African ancestry on the Y chromosome was less than on the mtDNA due to an elevated introgression of European-specific haplogroups in Tocaña.

4- Geographic orgin of 'Afro-Bolivians' in Africa could not be resolved unambiguously:

   a. According to the phylogenetic analyses of mitogenomes of recent African origin, some of them can be traced back both to West and West-Central Africa, while others also to East and Southeast Africa. 'Afro-Bolivian' mitogenomes did not cluster with samples fromexclusively one geographic region.
   b. Whole genome SNP genotyping of two individuals from Tocaña again showed a mosaic of ancestry components. West Africa with Yoruba was the main genetic ancestry in the two 'Afro-Bolivians', but a high proportion of East and Southeast African ancestry was again observed on the autosomes.

c. Some profiles of the Y chromosome found in 'Afro-Bolivians' were not only exclusively found in West or West-Central Africa, but some were also principally distributed in Southeastern Africa.

d. Due to the results gained by uni- and biparental markers a Southeastern genetic influence cannot be excluded, although most African enslaved people in the Americas came from West and West-Central African regions. The Southeastern genetic component observed in Bolivia could be explained if we take into account that Southeastern Africa became an important source of African enslaved people towards the end of the TAST. Most enslaved people during that time period entered the Americas through Brazil; in fact, this countrycould be an important entry for African enslaved people that arrived inBolivia (as Bolivia has no direct access to the sea). Therefore, as slavery was abolished officially only in the 20th century in Bolivia, some African enslaved people might have arrived late during the TAST *via* Brazil.

5- A significant gender bias was observed in the Bolivian population. Most Bolivians are of Native American ancestry on the autosomes and the mtDNA but show an increasedproportion of European-specific Y chromosomes, which can be attributed most probably to the fact that the colonization of the Americas was ruled by European males. In the Yungas, especially in the community of Tocaña, European ancestry on the Y chromosome was decreased due to the elevated proportion of African ancestry. It might be possible that the higher African ancestry on the mtDNA than on the Y chromosome can be the result of a higher mortality rate of African enslaved males or maybe because more African enslaved females than males arrived in the Yungas during the TAST.

9 Resumen

Antecedentes

Aproximadamente 10 millones de africanos fueron forzados a abandonar su tierra natal para trabajar como esclavos en las Américas durante el período del tráfico de esclavos Transatlántico. En la segunda parte del siglo XVIII, durante el punto culminante del tráfico de esclavos, unos 60.000 africanos llegaban a los puertos del Nuevo Mundo cada año. La mayor parte de ellos tubo como destino las islas del Caribe (aproximadamente 52%) y Brasil (aproximadamente 39%), pero una parte de ellos viajó a otros lugares como el Alto Perú (hoy en día Bolivia). En total, un 3,3% de los africanos llegaron a la América española durante el tráfico de esclavos.

Los primeros esclavos llevados a Bolivia están documentados en Potosí (sur de Bolivia) en el año 1549. Cuando los colonizadores llegaron al Alto Perú, se encontraron con mucha plata en el entorno de Potosí. Para la explotación de este recurso explotaron inicialmente a la población indígena. La mortalidad era extremadamente alta debido a las duras condiciones de trabajo en las minas y los colonizadores tuvieron que encontrar una manera de reemplazar la falta de trabajadores. Además mucha población indígena sufría enfermedades llevadas allí por los colonizadores.

Por otro lado, enviar gente de Europa a las Américas no era económico durante ese período porque a su vez, la demanda de trabajadores en Europa, sobre todo en la Península Ibérica, era muy elevada. Al mismo tiempo, los europeos empezaron a establecerse también en el continente africano. Los portugueses fueron los primeros en llegar, pero al principio estaban interesados fundamentalmente en el oro y solo marginalmente en el tráfico de los esclavos. Los portugueses forzaron a los primeros esclavos llevados a Europa a trabajar como servidores domésticos y a trabajar en las plantaciones de azúcar del Atlántico. No obstante, la combinación de la conquista de las Américas con la situación laboral en Europa y la llegada de los europeos a África, estimuló la migración forzada de los esclavos africanos.

Los españoles fueron los primeros que tuvieron el capital necesario para mandar a los esclavos africanos a las Américas, al principio sobre todo a Perú y México. El número de los esclavos africanos aumentó en particular cuando los colonizadores españoles encontraron plata en el entrono de Potosí, Bolivia. Aquí, los esclavos africanos tuvieron que trabajar principalmente en la Casa de la Moneda o como servidores domésticos y quizás también en las minas junto con la gente

indígena. Las condiciones laborales eran duras en la Casa de Moneda, y la mortalidad entre la población esclava era muy alta.

En los siglos XVIII y XIX, los descendientes de los primeros esclavos africanos llegados a Bolivia migraron hacia al norte del país, a los valles tropicales los Yungas. En esta región, los colonizadores se establecieron en lo que se conocía como 'sistemas de hacienda',un método de explotación de la tierra que era controlado por colonizadores generalmente ausentes y mantenido por esclavos descendientes de África y esclavos indígenas.

Los primeros esfuerzos para abolir la esclavitud en Bolivia en 1825 fallaron, por eso tan solo después de la reforma agrícola en 1952 los descendientes africanos en los Yungas no tuvieron que trabajar más para un dueño. La mayor parte de los descendientes africanos se quedaron en los Yungas y ocuparon las propiedades que sus dueños habían abandonado. Este es el motivo por el que hoy en día se observan descendientes africanos en la región de los Yungas. Esta región está formada por valles tropicales muy aislados dónde la población 'Afro-Boliviana' vive en determinadas comunidades, como por ejemplo en Tocaña. Debido al fuerte aislamiento geográfico de esta región geográfica, la población 'Afro-Boliviana' se ha mantenido en el anonimato hasta tiempos muy recientes; incluso dentro de Bolivia, esta población no formaba parte de los últimos censos 'étnicos' del país.

El interés del estudio genético de la población Boliviana

Durante los últimos años, la genética se ha venido interesando por el estudio del tráfico Transatlántico y la reconstrucción de los patrones de migración y mestizaje. Por su aislamiento y también por su número reducido actual, la historia del tráfico de esclavos en Bolivia no se ha estudiado con el mismo interés que otras zonas de América, y por lo tanto aun es muy desconocida. Además, en los documentos históricos la información también es muy limitada.

En este proyecto de tesis, el objetivo fundamental ha sido el desvelar desde la genética aspectos nuevos de todo este proceso de esclavitud. Para los análisis genéticos llevados a cabo en esta tesis se usaron marcadores uniparentales (cromosoma Y, ADN mitocondrial) y biparentales (marcadores informativos de ancestralidad o AIMs y polimorfismos de nucleótido simple o SNPs).

Los objetivos fundamentales de la tesis son los siguientes:

(i) Comparar las diferentes ancestralidades existentes en la población de Bolivia, centrándonos posteriormente en la población de los Yungas (donde asumimos que existe

el mayor componente 'Afro-Boliviano' del país). Para ello se analizarán diversos marcadores de uniparentales y biparentales,.

(ii) Investigar sobre el origen genético y geográfico en África de los patrones genéticos observados en la población 'Afro-Boliviana'.

(iii) Evaluar la contribución genética de Europa, África y América a la población de Bolivia. Usando marcadores genéticos que se heredan exclusivamente por vía paterna o materna nos permitirá además estudiar los sesgos de género que pudieron afectar a la población Boliviana en los distintos procesos de migración y mestizaje.

Los marcadores uniparentales como el ADN mitocondrial y el cromosoma Y pueden aportar informaciónmuy útil sobre el origen geográfico de los linajes (filogeografía). El poder informativo de estos marcadores viene dado fundamentalmente porque el tamaño efectivo de la población es especialmente bajo en ambos marcadores; esto hace que sean más sensibles a los efectos estocásticos de la deriva genética. La desventaja de los marcadores uniparentales es que solo pueden aportar información relacionada con único antecesor. Así, debido a los diferentes patrones de herencia, el ADN mitocondrial informa solamente del linaje maternal, mientras que el cromosoma Y solo proporciona información sobre el linaje paternal.

Los marcadores autosómicos típicos son los polimorfismos de nucleótido simple o SNPs (*Single Nucleotide Polymorphism*). Existen determinados SNPs que tienen unas características diferenciales y que se conocen como marcadores informativos de ascendencia o AIMs (*Ancestry Informative Markers*). Estos marcadores presentan la particularidad de que sus frecuencias son muy diferentes en los distintos grupos ancestrales, por lo que el estudio de un pequeño numero de los mismos permite la estimación de los distintos niveles de ancestralidad en el genoma de las poblaciones estudiadas. El estudio de AIMs en un número reducido (46 marcadores de *Indels* o polimorfismos de inserción y deleción) nos ha permitido el estudio de la ascendencia continental en el genoma de la población Boliviana. También en este proyecto de tesis, se llevó a cabo el análisis de polimorfismos de nucleótido simple a lo largo de todo genoma humano (*genome-wide)* con el objetivo de profundizar aun más en el estudio del el origen ancestral y geográfico de la población 'Afro-Boliviana'. Los marcadores autosómicos, al contrario que los uniparentales, proporcionan información sobre más antecesores. Estos marcadores son transmitidos a la siguiente generación a través de los dos progenitores. Por otro lado, hay que tener en cuenta que durante la meiosis tiene lugar la recombinación (intercambio de fragmentos de ADN entre parejas de cromosomas homólogos) en el ADN autosómico de ambos progenitores. Este hecho, transportado a través de las

múltiples generaciones a las que se remonta el genoma de cada individuo, hace que la reconstrucción coalescente de los fragmentos de ADN que se han heredado de un progenitor lejano sea cada vez más difícil; y por lo tanto, el rastro de su herencia se va 'degradando' con el paso de las generaciones. Este es el motivo por el que cualquier descendiente comparte más fragmentos de ADN con sus padres que con sus antepasados varias generaciones atrás.

La combinación de los marcadores uniparentales y biparentales en un mismo estudio es importante para el estudio integral de poblaciones que habitan el continente americano debido a los patrones de mestizaje tan particulares que experimentó este continente, en especial, la contribución de tres ancestralidades tan contrastadas genéticamente como son la nativo americana, la africana, y la europea. Y estos hechos han tenido lugar durante el periodo del tráfico de esclavos, donde la llegada de los Europeos y la población Africana esclavizada se fue mezclando con la población nativa de cada región. Como ya se sugirió anteriormente, hay que tener en cuenta que el linaje paternal y maternal no representan siempre la composición entera de la ascendencia. Por ejemplo, un individuo puede tener mayoritariamente una antecedencia europea pero sin embargo el ADN mitocondrial y/o el cromosoma Y pueden ser de origen África. Ese es el motivo por el que es importante completar los estudios de marcadores uniparentales con el estudio de marcadores que se hereden de ambos progenitores.

Patrones de variabilidad genómica en la población boliviana

Como resultado del presente proyecto de tesis, hemos podido observar que en Bolivia hay diferencias importantes en cuanto a la ancestralidad estimada en el ADN mitocondrial, el cromosoma Y o en el ADN autosómico.

La mayor parte del ADN mitocondrial en la población boliviana (salvo en los Yungas) es de origen indígena (98.4%). El haplogrupo B es el linaje más frecuente en Bolivia (52.6%) seguido por el haplogrupo C (21.9%), el haplogrupo A (14.3%), y el haplogrupo D (9.6%). La ascendencia africana (1%) y europea (<1%) es mínima en la mayor parte del país. En Bolivia, la ancestralidad mitocondrial indígena es mayoritaria (excepto en la población de los Yungas) pero la distribución geográfica de los distintos haplogrupos es diferente en términos de frecuencias. Por ejemplo, en los departamentosde La Paz, Chuquisaca, Cochabamba, y Santa Cruz, el haplogrupo B es el más frecuente, en Beni es el haplogrupo C, mientras que en Pando el haplogrupo A es el más común.

Además, el haplogrupo B es el más prevalente en las regiones andinas y sub-andinas, mientras que el haplogrupo C es el más común en la zona de los Llanos bolivianos.

En las valles de los Yungas la composición de los haplogrupos es diferente. La ascendencia africana en la población de los Nor Yungas en el linaje maternal asciende hasta el 19%. Así mismo, esta ancestralidad africana se concentra en la comunidad de Tocaña (84% en total). Los haplogrupos mitocondriales específicos africanos observados en los Yungas pertenecen a los linajes L0 (5%), L1 (5%) y L3 (9%). La mayor parte de la población residual en los Nor Yungas tiene principalmente origen nativo americano (79%). Los haplogrupos B2/B4 son los más frecuentes seguidos por el C1, A2 y D1/D4. En los Sud Yungas todos los individuos son de ancestralidad nativa en el linaje maternal. Los linajes más frecuentes en esta región son por este orden, el haplogrupo C1, B2/B4 y D1.

La secuenciación del genoma entero de ADN mitocondrial de ocho individuos con haplogrupos de origen indígena resultó en el descubrimiento de tres haplogrupos nuevos en la filogenia nativo americana, todavía no referenciados en la literatura: A2ah, B2o y B2o1. Estos tres sub-linajes son sin embargo poco frecuentes en Bolivia (0,5%-1%) y aparentemente raros en todo el continente americano.

El análisis de mitogenomas de origen Africano observados en Bolivia ha sido también de ayuda para hacer determinadas inferencias demográficas. Los ocho mitogenomas 'Afro-Bolivianos' pertenecen a los haplogrupos L0a1b2, L0a2a2a, L1c3b1a, L3d1a1a y L3d1b3. El origen geográfico preciso de los ocho mitogenomas en el continente africano todavía no está muy claro, pero se han hecho ciertas descubrimientos novedosos al respecto. La naturaleza filogeográfica de los mitogenomas cobra interés cuando las inferencias genéticas se hacen a la luz de la documentación histórica. El tráfico de los esclavos empezó muy temprano en Bolivia porque los primeros esclavos en el Alto Perú fueron documentados a partir de 1549. Se sabe además que el tráfico de los esclavos empezó en la región africana de Senegambia. Más tarde otras regiones, sobre todo África oeste-central, fueron cobrando más y más importancia para los colonizadores. Por lo tanto, la mayor parte de los esclavos llevados a América tienen su origen geográfico en el oeste-central y oeste de África. Por otro lado, especular sobre el origen geográfico de los genomas mitocondriales 'Afro-Bolivianos' en el oeste u oeste-central de África parece tener sentido ya que algunos de los haplotipos observados se agrupan perfectamente junto a muestras de esas regiones de África. Pero, algunos haplotipos Africanos observados en Bolivia tienen también afinidad genética por el este y

el sudeste de África. En este contexto, hay que destacar que la región del sudeste, sobre todo Mozambique, fue una importante fuente de esclavos sobre todo al final del siglo XIX. Esta época coincide con el final del período del tráfico de los esclavos en el Atlántico. Por lo tanto, este hecho podría explicar porque algunas muestras del sudeste de África forman haplogrupos junto a los 'Afro-Bolivianos'. Sorprendentemente, existen individuos muestreados en el Oriente Medio que también se encuentran compartiendo los mismos haplogrupos Africanos observados en Bolivia. La conexión de 'afro-boliviana' con el Oriente Medio es más bien confusa. Podría indicar la existencia de una relación indirecta del tráfico de los esclavos en el Atlántico con el tráfico de los esclavos dentro de África que fue especialmente dirigido por la población musulmana. Algunos historiadores sostienen que el 40% de los esclavos en las Américas se pudo atribuir al tráfico de los esclavos de los musulmanes. Una gran parte de esta influencia pudo estar canalizada a través del sudeste de África, fuertemente influida por el Este africano, donde la conexión con el oriente medio está mejor documentada.

En cuanto al cromosoma Y, el 62.5% de los linajes en la población boliviana (incluyendo a los Yungas) fueron de origen europeo. Esto no es de extrañar si tenemos en cuenta que este patrón está bastante generalizado en otras localizaciones de América, como Brasil, Colombia o Argentina. La ancestralidadde cromosoma Y indígena en Bolivia fue del 35.5%, y la africana del 6%. La ancestralidad europea alta es el resultado de una presencia elevada de varones en el proceso de colonización europea de las Américas. La diferencia más grande en Bolivia en cuanto al cromosoma Y se detecta entre los Llanos y las regiones andinas y sub-andinas. En los Llanos la antecedencia europea fue solo del 46% pero en las regiones andinas y sub-andinas ascenció al 75% y 86%, respectivamente. Por otro lado, la ascendencia nativo americana fue más alta en los Llanos (47%) y más baja en las regiones andinas (19%) y sub-andinas (7%). De nuevo, en la región de los Yungas, la ascendencia africana fue elevada. En la comunidad de Tocaña por ejemplo alcanzó el 44% de los linajes de cromosoma Y. El 33% de los linajes fueron de origen europeo, y el restante 22% de origen Nativo Americano.

El estudio basado en el análisis de AIMs indica que la ancestralidad Nativo Americana (73,3%) fue de nuevo la más alta en la población de Bolivia (sin tener en cuenta las poblaciones de los Yungas). La ancestralidad Nativo Americana fue casi tres veces más alta que la europea (25,8%) y la africana fue menor al 1%. No existe una distribución geográficamente homogénea en la población boliviana en cuanto a los patrones de ancestralidad. La ascendencia indígena fue sobre todo predominante en el departamento La Paz, mientras que en Santa Cruz, la ascendencia europea

fue especialmente alta. La ascendencia indígena fue en particular alta en la región andina y bajo en los Llanos.Por el contrario, en los Yungas la ancestralidad nativo americana inferida a través del análisis de AIMs fue también predominante (76%), pero la africana alcanzó un valor de 12% en los Yungas (comparar con 1% observado en el resto de la población de Bolivia). La ascendencia africana está concentrada sobre todo en la comunidad de Tocaña, localizada en Nor Yungas. De hecho, en Nor Yungas la ascendencia Africana alcanza el 13%, pero cae al 3% cuando los individuos de los Tocaña se excluyen del análisis. En Tocaña la antecedencia africana fue la más importante (56%).

Finalmente, en este estudio se llevó a cabo el análisis de polimorfismos de nucleótido simple (SNPs) a lo largo de todo genoma (estudio pan-genómico) en individuos seleccionados pertenecientes a la población de Tocaña, y para los cuales ya se conocía una ancestralidad 'Afro-Boliviana' mayoritaria, inferida esta a través del estudio de AIMs. El análisis de ancestralidad indicó que el componente principal de los 'afro-bolivianos' es sub-sahariano (83-84%), y que el componente más dominante de África se puede atribuir al oeste-central (42-49%). Curiosamente, y en consistencia con las inferencia hechas en los marcadores uniparentales, un porcentaje importante del genoma de los 'Afro-Bolivianos se puede atribuir también el este y sudeste de África. Este hecho se corresponde muy bien con los resultados de los mitogenomas los cuales también indican la existencia de un componente filogeográfico proveniente del este/sudeste de África. Una de las poblaciones del este de África que tiene una similitud especial con los Afro-Bolivianos es población de los Luhya; un grupo Bantú que habita los territorios del Sur del Kenia. Esta gran similitud se explica por la estrecha relación genética existente entre estas poblaciones Bantús provenientes del Este de África y el sudeste Africano (ej. Mozambique), esta última región a su vez en estrecha vinculación con el tráfico de esclavos, especialmente en el periodo final.

Sesgo de género existente en la población Boliviana

El desequilibrio de ancestralidades existente en la población boliviana e inferido a través del estudio de marcadores binarios indica que el impacto de las migraciones y los mestizajes tuvo un fuerte sesgo de género. Los datos de cromosoma Y indican un fuerte impacto del hombre europeo en la población indígena, cruzándose prioritariamente con mujeres de ancestralidad indígena (y también africana). En Bolivia se puede observar los resultados de esa estructura social del pasado. En la población general de Bolivia la ancestralidad nativo americana fue predominante en los autosomas y en el ADN mitocondrial. Pero la mayor parte de los cromosomas Y fueron de origen europeo. El

porcentaje europeo en el cromosoma Y fue más alto en la región sub-andina y más bajo en los Llanos. El componente europeo fue también importante en los Yungas pero en la comunidad de Tocaña la mayor parte de los hombres tenía origen africano en el cromosoma Y. En cuanto al ADN mitocondrial las linajes maternales pertenecieron mayormente a haplogrupos específico de América indígena en la población entera de Bolivia. Solo en la comunidad de Tocaña los linajes de ADN mitocondrial fueron principalmente de África. Pero según los valores de la ancestralidad observada en el ADN mitocondrial y el cromosoma Y, el porcentaje de haplogrupos específicos de África fue más alto para el ADN mitocondrial que para el cromosoma Y.

Está documentado que fueron más hombres que mujeres los que fueron forzados a dejar África durante el tráfico de los esclavos en el Atlántico. Los resultados genéticos parecen indicar por lo tanto que la mortalidad pudo haber sido mayor para los hombres que para las mujeres. Otra posibilidad podría ser que fueron más mujeres que hombres de descendientes africanos los que finalmente migraron a los Yungas. Nada hay bien documentado al respecto en la literatura sobre la historia de Bolivia. El hecho que la antecedencia europea en el cromosoma Y en la comunidad de Tocaña es más bajo que en la población residual puede reflejar la manera de dirigir las haciendas durante la esclavitud. Muchos colonizadores estaban presente en las haciendas y dirigieron las haciendas a la distancia.

Conclusiones

Tal y como fue investigado en este proyecto de tesis, la genética es una ciencia que puede aportar aspectos novedosos sobre la historia de los esclavos en las Américas. Mirar a la historia del tráfico de los esclavos en el Atlántico desde una perspectiva diferente a la histórica es importante porque puede arrojar nuevos datos sobre como pudo haber sido el proceso de esclavitud, mestizaje, etc. Pero además, estos estudios ayudan a promover más la investigación en otros campos sumamente importantes como son la historia o las ciencias sociales en uno de los episodios más dramáticos de la historia reciente de las poblaciones humanas.

La genética tiene también una dimensión social importante. Así, esta disciplina puede ser una herramienta útil para conocer algo más sobre nuestros orígenes biológicos. En el contexto del tráfico de esclavos este tema cobra un valor especial, debido a que fueron muchos millones de personas las que fueron forzados a dejar sus tierras de origen y a perder para siempre la memoria de su pasado. Por tal motivo, es comprensible poner pruebas de ADN a disposición de personas individuales para que cada una indague en la medida de sus posibilidades sobre su pasado

genealógico. Pero al mismo tiempo, y para evitar un uso erróneo de este potencial es importante informar correctamente a los consumidores sobre los límites de esas pruebas.

El desarrollo moderno de la tecnología en el campo de genética y genómica, y la reducción de los costes de las pruebas genéticas, ha elevado en gran medida el número de personas que desea realizar estas pruebas a través de compañías privadas; un tipo de prueba que se ha venido popularizando en el mercado con el nombre de *direct-to-consumer-test*. Estas pruebas tienen un ámbito de interés más amplio, y cubren también la genética clínica.

Una de las mayores limitaciones de este tipo de estudios es el conocimiento limitado de las poblaciones de referencia, y el hecho de que las poblaciones existentes hoy en día no necesariamente reflejan los patrones de variabilidad existen en el pasado. Otra limitación importante de este tipo de estudios es que la mayor parte de los clientes solicitan pruebas basadas en el análisis de marcadores uniparentales. A este efecto es importante recordar que estos marcadores solo informan de uno de nuestros ancestros de entre un total de 2^n (siendo n el número de generaciones) ancestros teóricos que existen en nuestro pasado ancestral genealógico. Además, en grupos poblacionales cercanos, todos los individuos de la población comparten ancestros en tiempos generacionales cercanos, por lo que la información aportada por el cromosoma Y o el ADN mitocondrial puede ser totalmente irrelevante en cuanto al deseo del cliente de conocer algo más de su pasado.

En este proyecto de tesis tenemos claros ejemplos que apuntan a que reconstruir el origen geográfico preciso de un perfil de cromosoma Y o de ADN mitocondrial (aun cuando este se basa en el análisis de todo el mitogenoma) es imposible. Y aun en el caso de que se pudiera mejorar en el futuro, con el incremento de las bases de datos de referencia, estos marcadores solamente informarían de una parte muy pequeña de nuestro genoma global, y por lo tanto de nuestro pasado biológico.

Es convenientehacer especial hincapié en dos ideas finales. Primero, una prueba genética no va a poder solucionar todos los problemas sobre la identidad biológica de una persona. Segundo, la genética no puede reemplazar otros campos de la ciencia, como son la historia o las ciencias sociales. La genética tiene su ámbito de aplicación y muchas limitaciones. La genética sin embargo tiene un gran potencial para ayudar a reconstruir la historia de las poblaciones humanas, y en el caso de este proyecto de tesis, descifrar la compleja y dramática historia del tráfico de esclavos en el Atlántico y sus implicaciones en la población actual de Bolivia.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

10 Literature

1. Eltis, D. (2008). A Brief Overview of the Trans-Atlantic Slave Trade. Voyag. Trans-Atl. Slave Trade Database. Available at: http://www.slavevoyages.org/.

2. Klein, H. S. (2010). The Atlantic Slave Trade (Cambridge University Press).

3. Lovejoy, P. E. (2000). Transformations in Slavery: A History of Slavery in Africa 2nd ed. (Cambridge University Press).

4. Hall, G. M. (2005). Slavery and African Ethnicities in the Americas: Restoring the Links (University of North Carolina Press).

5. Vos, J. (2010). The Slave Trade from the Windward Coast: The Case of the Dutch, 1740–1805. Afr. Econ. Hist. *38*, 29–51.

6. Falola, T., and Childs, M. D. (2005). The Yoruba Diaspora in the Atlantic World (Bloomington and Indianapolis: Indiana University Press).

7. Eltis, D., and Richardson, D. (2010). Atlas of the Transatlantic Slave Trade (Yale University Press).

8. Burnard, T., and Morgan, K. (2001). The dynamics of the slave market and slave purchasing patterns in Jamaica, 1655-1788. William Mary Q. *58*, 205–228.

9. Littlefield, D. C. (1981). Rice and Slaves: Ethnicity and the Slave Trade in Colonial South Carolina (Baton Rouge: Louisiana State University Press).

10. Morgan, P. D. (1982). Review of Rice and Slaves: Ethnicity and the Slave Trade in Colonial South Carolina by Daniel C. Littlefield. William Mary Q. *39*, 709–712.

11. Davies, C. E. B. (2008). Encyclopedia of the African Diaspora: Origins, Experiences, and Culture: Origins, Experiences, and Culture (ABC-CLIO).

12. Handler, J. S., and Corruccini, R. S. (1983). Plantation Slave Life in Barbados: A Physical Anthropological Analysis. J. Interdiscip. Hist. *14*, 65–90.

13. Handler, J. S., Corruccini, R. S., and Mutaw, R. J. (1982). Tooth Mutilation in the Caribbean: Evidence from a Slave Burial Population in Barbados. J. Hum. Evol. *11*, 297–313.

14. Samford, P. (1996). The Archaeology of African-American Slavery and Material Culture. William Mary Q. *53*, 87–114.

15. Schroeder, H., O'Connell, T. C., Evans, J. A., Shuler, K. A., and Hedges, R. E. M. (2009). Trans-Atlantic slavery: Isotopic evidence for forced migration to Barbados. Am. J. Phys. Anthropol. *139*, 547–557.

16. Salas, A., Carracedo, A., Richards, M., and Macaulay, V. (2005). Charting the ancestry of African Americans. Am. J. Hum. Genet. *77*, 676–680.

17. Salas, A., Richards, M., Lareu, M.-V., Scozzari, R., Coppa, A., Torroni, A., Macaulay, V., and Carracedo, A. (2004). The African diaspora: mitochondrial DNA and the Atlantic slave trade. Am. J. Hum. Genet. *74*, 454–465.

18. Jackson, F. L. C. (2008). Ancestral links of Chesapeake Bay region African Americans to specific Bight of Bonny (West Africa) microethnic groups and increased frequency of aggressive breast cancer in both regions. Am. J. Hum. Biol. Off. J. Hum. Biol. Counc. *20*, 165–173.

19. Stefflova, K., Dulik, M. C., Barnholtz-Sloan, J. S., Pai, A. A., Walker, A. H., and Rebbeck, T. R. (2011). Dissecting the Within-Africa Ancestry of Populations of African Descent in the Americas. PLoS ONE *6*, e14495.

20. Schlebusch, C. M., Skoglund, P., Sjödin, P., Gattepaille, L. M., Hernandez, D., Jay, F., Li, S., Jongh, M. D., Singleton, A., Blum, M. G. B., et al. (2012). Genomic Variation in Seven Khoe-San Groups Reveals Adaptation and Complex African History. Science *338*, 374-379.

21. Fehler, B. O. (2011). (Re)constructing Roots: Genetics and the "Return" of African Americans to Ghana. Mobilities *6*, 585–600.

22. Jobling, M. A., Hollox, E., Hurles, M., Tyler-Smith, C., and Kivisild, T. (2013). Human Evolutionary Genetics (Taylor & Francis Limited).

23. Kraytsberg, Y., Schwartz, M., Brown, T. A., Ebralidse, K., Kunz, W. S., Clayton, D. A., Vissing, J., and Khrapko, K. (2004). Recombination of Human Mitochondrial DNA. Science *304*, 981–981.

24. Schwartz, M., and Vissing, J. (2002). Paternal Inheritance of Mitochondrial DNA. N. Engl. J. Med. *347*, 576–580.

25. Charlesworth, B. (2009). Effective population size and patterns of molecular evolution and variation. Nat. Rev. Genet. *10*, 195–205.

26. Soares, P., Ermini, L., Thomson, N., Mormina, M., Rito, T., Röhl, A., Salas, A., Oppenheimer, S., Macaulay, V., and Richards, M. B. (2009). Correcting for purifying selection: an improved human mitochondrial molecular clock. Am. J. Hum. Genet. *84*, 740–759.

27. Kivisild, T., Shen, P., Wall, D. P., Do, B., Sung, R., Davis, K., Passarino, G., Underhill, P. A., Scharfe, C., Torroni, A., et al. (2006). The role of selection in the evolution of human mitochondrial genomes. Genetics *172*, 373–387.

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

29. Loogväli, E.-L., Kivisild, T., Margus, T., and Villems, R. (2009). Explaining the Imperfection of the Molecular Clock of Hominid Mitochondria. PLoS ONE *4*, e8260.

30. Xue, Y., Wang, Q., Long, Q., Ng, B. L., Swerdlow, H., Burton, J., Skuce, C., Taylor, R., Abdellah, Z., Zhao, Y., et al. (2009). Human Y Chromosome Base-Substitution Mutation Rate Measured by Direct Sequencing in a Deep-Rooting Pedigree. Curr. Biol. *19*, 1453–1457.

31. The 1000 Genomes Project Consortium (2010). A map of human genome variation from population-scale sequencing. Nature *467*, 1061–1073.

32. MITOMAP: A Human Mitochondrial Genome Database (2013). Available at: http://www.mitomap.org.

33. Shriver, M. D., and Kittles, R. A. (2004). Opinion: Genetic ancestry and the search for personalized genetic histories. Nat. Rev. Genet. *5*, 611–618.

34. Foster, E. A., Jobling, M. A., Taylor, P. G., Donnelly, P., Knijff, P. de, Mieremet, R., Zerjal, T., and Tyler-Smith, C. (1998). Jefferson fathered slave's last child. Nature *396*, 27–28.

35. Novembre, J., Johnson, T., Bryc, K., Kutalik, Z., Boyko, A. R., Auton, A., Indap, A., King, K. S., Bergmann, S., Nelson, M. R., et al. (2008). Genes mirror geography within Europe. Nature *456*, 98–101.

36. Phillips, C., Fernandez-Formoso, L., Garcia-Magariños, M., Porras, L., Tvedebrink, T., Amigo, J., Fondevila, M., Gomez-Tato, A., Alvarez-Dios, J., Freire-Aradas, A., et al. (2011). Analysis of global variability in 15 established and 5 new European Standard Set (ESS) STRs using the CEPH human genome diversity panel. Forensic Sci. Int. Genet. *5*, 155–169.

37. Lee, J.-Y., Lee, B.-S., Shin, D.-J., Woo Park, K., Shin, Y.-A., Joong Kim, K., Heo, L., Young Lee, J., Kyoung Kim, Y., Jin Kim, Y., et al. (2013). A genome-wide association study of a coronary artery disease risk variant. J. Hum. Genet. *58*, 120–126.

38. Garcia-Closas, M., Couch, F. J., Lindstrom, S., Michailidou, K., Schmidt, M. K., Brook, M. N., Orr, N., Rhie, S. K., Riboli, E., Feigelson, H. S., et al. (2013). Genome-wide association studies identify four ER negative-specific breast cancer risk loci. Nat. Genet. *45*, 392–398, 398e1–2.

39. Moreno-Estrada, A., Gravel, S., Zakharia, F., McCauley, J. L., Byrnes, J. K., Gignoux, C. R., Ortiz-Tello, P. A., Martínez, R. J., Hedges, D. J., Morris, R. W., et al. (2013). Reconstructing the Population Genetic History of the Caribbean. PLoS Genet *9*, e1003925.

40. Su, Y., Howard, H. C., and Borry, P. (2011). Users' motivations to purchase direct-to-consumer genome-wide testing: an exploratory study of personal stories. J. Community Genet. *2*, 135–146.

41. Schuster, S. C. (2008). Next-generation sequencing transforms today's biology. Nat. Methods *5*, 16–18.

42. Wetterstrand, K. A. (2014). DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP) Available at: www.genome.gov/sequencingcosts.

43. Mardis, E. R. (2011). A decade/'s perspective on DNA sequencing technology. Nature *470*, 198–203.

44. Parson, W., Wootton, S., and Schade, L. The Future of Forensics Has Arrived: The Application of Next Generation Sequencing. Forensic Magazine. Available at: http://www.forensicmag.com.

45. Yngvadottir, B., MacArthur, D. G., Jin, H., and Tyler-Smith, C. (2009). The promise and reality of personal genomics. Genome Biol. *10*, 237.

46. Bustamante, C. D., Vega, F. M. D. L., and Burchard, E. G. (2011). Genomics for the world. Nature *475*, 163–165.

47. Hunter, D. J., Khoury, M. J., and Drazen, J. M. (2008). Letting the genome out of the bottle--will we get our wish? N. Engl. J. Med. *358*, 105–107.

48. Henneman, L., Vermeulen, E., van El, C. G., Claassen, L., Timmermans, D. R. M., and Cornel, M. C. (2013). Public attitudes towards genetic testing revisited: comparing opinions between 2002 and 2010. Eur. J. Hum. Genet. *21*, 793–799.

49. Kosoy, R., Nassir, R., Tian, C., White, P. A., Butler, L. M., Silva, G., Kittles, R., Alarcon-Riquelme, M. E., Gregersen, P. K., Belmont, J. W., et al. (2009). Ancestry informative marker sets for determining continental origin and admixture proportions in common populations in America. Hum. Mutat. *30*, 69–78.

50. Kidd, J. R., Friedlaender, F. R., Speed, W. C., Pakstis, A. J., Vega, F. M. D. L., and Kidd, K. K. (2011). Analyses of a set of 128 ancestry informative single-nucleotide polymorphisms in a global set of 119 population samples. Investig. Genet. *2*, 1.

51. Ralph, P., and Coop, G. (2013). The Geography of Recent Genetic Ancestry across Europe. PLoS Biol *11*, e1001555.

52. Callaway, E. (2012). Ancestry testing goes for pinpoint accuracy. Nature *486*, 17–17.

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

54. Elhaik, E., Tatarinova, T., Chebotarev, D., Piras, I. S., Maria Calò, C., De Montis, A., Atzori, M., Marini, M., Tofanelli, S., Francalacci, P., et al. (2014). Geographic population structure analysis of worldwide human populations infers their biogeographical origins. Nat. Commun. *5*.

55. Gravel, S. (2012). Population Genetics Models of Local Ancestry. Genetics, *191*, 607-619.

56. Lee, S. S., Mountain, J., Koenig, B., Altman, R., Brown, M., Camarillo, A., Cavalli-Sforza, L., Cho, M., Eberhardt, J., Feldman, M., et al. (2008). The ethics of characterizing difference: guiding principles on using racial categories in human genetics. Genome Biol. *9*, 404.

57. Salas, A., Acosta, A., Alvarez-Iglesias, V., Cerezo, M., Phillips, C., Lareu, M. V., and Carracedo, A. (2008). The mtDNA ancestry of admixed Colombian populations. Am. J. Hum. Biol. Off. J. Hum. Biol. Counc. *20*, 584–591.

58. Lao, O., Vallone, P. M., Coble, M. D., Diegoli, T. M., van Oven, M., van der Gaag, K. J., Pijpe, J., de Knijff, P., and Kayser, M. (2010). Evaluating Self-declared Ancestry of U.S. Americans with Autosomal, Y-chromosomal and Mitochondrial DNA. Hum. Mutat. *31*, E1875–E1893.

59. Baylis, F. (2003). Black as Me: Narrative Identity. Dev. World Bioeth. *3*, 142–150.

60. Joyner, C. (1985). Down by the Riverside: A South Carolina Slave Community (University of Illinois Press).

61. Gomez, M. A. (1998). Exchanging Our Country Marks : The Transformation of African Identities in the Colonial and Antebellum South: The Transformation of African Identities in the Colonial and Antebellum South (Univ of North Carolina Press).

62. Lipski, J. M. (2006). Afro-Bolivian language today: the oldest surviving Afro-Hispanic speech community. Afro-Hisp. Rev. *25*, 179.

63. Brucato, N., Cassar, O., Tonasso, L., Tortevoye, P., Migot-Nabias, F., Plancoulaine, S., Guitard, E., Larrouy, G., Gessain, A., and Dugoujon, J.-M. (2010). The imprint of the Slave Trade in an African American population: mitochondrial DNA, Y chromosome and HTLV-1 analysis in the Noir Marron of French Guiana. BMC Evol. Biol. *10*, 314.

64. Benn Torres, J., Doura, M. B., Keita, S. O. Y., and Kittles, R. A. (2012). Y Chromosome Lineages in Men of West African Descent. PLoS ONE *7*, e29687.

65. Bryc, K., Auton, A., Nelson, M. R., Oksenberg, J. R., Hauser, S. L., Williams, S., Froment, A., Bodo, J.-M., Wambebe, C., Tishkoff, S. A., et al. (2010). Genome-wide patterns of population structure and admixture in West Africans and African Americans. Proc. Natl. Acad. Sci. U. S. A. *107*, 786–791.

66. Parra, E. J., Marcini, A., Akey, J., Martinson, J., Batzer, M. A., Cooper, R., Forrester, T., Allison, D. B., Deka, R., Ferrell, R. E., et al. (1998). Estimating African American admixture proportions by use of population-specific alleles. Am. J. Hum. Genet. *63*, 1839–1851.

67. Parra, E. J., Kittles, R. A., Argyropoulos, G., Pfaff, C. L., Hiester, K., Bonilla, C., Sylvester, N., Parrish-Gause, D., Garvey, W. T., Jin, L., et al. (2001). Ancestral proportions and admixture dynamics in geographically defined African Americans living in South Carolina. Am. J. Phys. Anthropol. *114*, 18–29.

68. Bonilla, C., Parra, E. J., Pfaff, C. L., Dios, S., Marshall, J. A., Hamman, R. F., Ferrell, R. E., Hoggart, C. L., McKeigue, P. M., and Shriver, M. D. (2004). Admixture in the Hispanics of the San Luis Valley, Colorado, and its implications for complex trait gene mapping. Ann. Hum. Genet. *68*, 139–153.

69. Salzano, F. M., and Sans, M. (2014). Interethnic admixture and the evolution of Latin American populations. Genet. Mol. Biol. *37*, 151–170.

70. Benn Torres, J., Bonilla, C., Robbins, C. M., Waterman, L., Moses, T. Y., Hernandez, W., Santos, E. R., Bennett, F., Aiken, W., Tullock, T., et al. (2008). Admixture and Population Stratification in African Caribbean Populations. Ann. Hum. Genet. *72*, 90–98.

71. Deason, M. L., Salas, A., Newman, S. P., Macaulay, V. A., Morrison, E. Y. st A., and Pitsiladis, Y. P. (2012). Interdisciplinary approach to the demography of Jamaica. BMC Evol. Biol. *12*, 24.

72. Via, M., Gignoux, C. R., Roth, L. A., Fejerman, L., Galanter, J., Choudhry, S., Toro-Labrador, G., Viera-Vera, J., Oleksyk, T. K., Beckman, K., et al. (2011). History Shaped the Geographic Distribution of Genomic Admixture on the Island of Puerto Rico. PLoS ONE *6*, e16513.

73. Martínez-Cruzado, J. C., Toro-Labrador, G., Viera-Vera, J., Rivera-Vega, M. Y., Startek, J., Latorre-Esteves, M., Román-Colón, A., Rivera-Torres, R., Navarro-Millán, I. Y., Gómez-Sánchez, E., et al. (2005). Reconstructing the population history of Puerto Rico by means of mtDNA phylogeographic analysis. Am. J. Phys. Anthropol. *128*, 131–155.

74. Simms, T. M., Garcia, C., Mirabal, S., McCartney, Q., and Herrera, R. J. (2008). The genetic legacy of the Transatlantic Slave Trade in the island of New Providence. Forensic Sci. Int. Genet. *2*, 310–317.

75. Lins, T. C., Vieira, R. G., Abreu, B. S., Grattapaglia, D., and Pereira, R. W. (2010). Genetic composition of Brazilian population samples based on a set of twenty-eight ancestry informative SNPs. Am. J. Hum. Biol. Off. J. Hum. Biol. Counc. *22*, 187–192.

76. Pena, S. D. J., Di Pietro, G., Fuchshuber-Moraes, M., Genro, J. P., Hutz, M. H., Kehdy, F. de S. G., Kohlrausch, F., Magno, L. A. V., Montenegro, R. C., Moraes, M. O., et al. (2011). The Genomic Ancestry of Individuals from Different Geographical Regions of Brazil Is More Uniform Than Expected. PLoS ONE *6*, e17063.

77. Parra, F. C., Amado, R. C., Lambertucci, J. R., Rocha, J., Antunes, C. M., and Pena, S. D. J. (2003). Color and genomic ancestry in Brazilians. Proc. Natl. Acad. Sci. U. S. A. *100*, 177–182.

78. Callegari-Jacques, S. M., Grattapaglia, D., Salzano, F. M., Salamoni, S. P., Crossetti, S. G., Ferreira, M. E., and Hutz, M. H. (2003). Historical genetics: spatiotemporal analysis of the formation of the Brazilian population. Am. J. Hum. Biol. Off. J. Hum. Biol. Counc. *15*, 824–834.

79. Ferreira, L. B., Mendes-Junior, C. T., Wiezel, C. E. V., Luizon, M. R., and Simões, A. L. (2006). Genomic ancestry of a sample population from the state of São Paulo, Brazil. Am. J. Hum. Biol. Off. J. Hum. Biol. Counc. *18*, 702–705.

80. Manta, F. S. N., Pereira, R., Caiafa, A., Silva, D. A., Gusmão, L., and Carvalho, E. F. (2012). Analysis of genetic ancestry in the admixed Brazilian population from Rio de Janeiro using 46 autosomal ancestry-informative indel markers. Ann. Hum. Biol. *40*, 94–98.

81. Silva, W. A., Bortolini, M. C., Schneider, M. P. C., Marrero, A., Elion, J., Krishnamoorthy, R., and Zago, M. A. (2006). MtDNA haplogroup analysis of black Brazilian and sub-Saharan populations: implications for the Atlantic slave trade. Hum. Biol. *78*, 29–41.

82. Hünemeier, T., Carvalho, C., Marrero, A. R., Salzano, F. M., Pena, S. D. J., and Bortolini, M. C. (2007). Niger-Congo speaking populations and the formation of the Brazilian gene pool: mtDNA and Y-chromosome data. Am. J. Phys. Anthropol. *133*, 854–867.

83. Abe-Sandes, K., Silva, W. A., Jr, and Zago, M. A. (2004). Heterogeneity of the Y chromosome in Afro-Brazilian populations. Hum. Biol. *76*, 77–86.

84. Alves-Silva, J., da Silva Santos, M., Guimarães, P. E. M., Ferreira, A. C. S., Bandelt, H.-J., Pena, S. D. J., and Prado, V. F. (2000). The Ancestry of Brazilian mtDNA Lineages. Am. J. Hum. Genet. *67*, 444–461.

85. Carvalho-Silva, D. R., Santos, F. R., Rocha, J., and Pena, S. D. J. (2001). The Phylogeography of Brazilian Y-Chromosome Lineages. Am. J. Hum. Genet. *68*, 281–286.

86. Marrero, A. R., Das Neves Leite, F. P., De Almeida Carvalho, B., Peres, L. M., Kommers, T. C., Da Cruz, I. M., Salzano, F. M., Ruiz-Linares, A., Da Silva Júnior, W. A., and Bortolini, M. C. (2005). Heterogeneity of the genome ancestry of individuals classified as White in the state of Rio Grande do Sul, Brazil. Am. J. Hum. Biol. Off. J. Hum. Biol. Counc. *17*, 496–506.

87. Ibarra, A., Freire-Aradas, A., Martínez, M., Fondevila, M., Burgos, G., Camacho, M., Ostos, H., Suarez, Z., Carracedo, A., Santos, S., et al. (2014). Comparison of the genetic background of different Colombian populations using the SNPforID 52plex identification panel. Int. J. Legal Med. *128*, 19–25.

88. Rojas, W., Parra, M. V., Campo, O., Caro, M. A., Lopera, J. G., Arias, W., Duque, C., Naranjo, A., García, J., Vergara, C., et al. (2010). Genetic make up and structure of Colombian populations by means of uniparental and biparental DNA markers. Am. J. Phys. Anthropol. *143*, 13–20.

89. Gómez-Pérez, L., Alfonso-Sánchez, M. A., Pérez-Miranda, A. M., García-Obregón, S., Builes, J. J., Bravo, M. L., De Pancorbo, M. M., and Peña, J. A. (2010). Genetic admixture estimates by Alu elements in Afro-Colombian and Mestizo populations from Antioquia, Colombia. Ann. Hum. Biol. *37*, 488–500.

90. Salas, A., Richards, M., Lareu, M.-V., Sobrino, B., Silva, S., Matamoros, M., Macaulay, V., and Carracedo, Á. (2005). Shipwrecks and founder effects: Divergent demographic histories reflected in Caribbean mtDNA. Am. J. Phys. Anthropol. *128*, 855–860.

91. Rodas, C., Gelvez, N., and Keyeux, G. (2003). Mitochondrial DNA studies show asymmetrical Amerindian admixture in Afro-Colombian and Mestizo populations. Hum. Biol. *75*, 13–30.

92. Yunis, J. J., and Yunis, E. J. (2013). Mitochondrial DNA (mtDNA) haplogroups in 1526 unrelated individuals from 11 Departments of Colombia. Genet. Mol. Biol. *36*, 329–335.

93. Brucato, N., Tortevoye, P., Plancoulaine, S., Guitard, E., Sanchez-Mazas, A., Larrouy, G., Gessain, A., and Dugoujon, J.-M. (2009). The genetic diversity of three peculiar populations descending from the slave trade: Gm study of Noir Marron from French Guiana. C. R. Biol. *332*, 917–926.

94. Corach, D., Lao, O., Bobillo, C., Van Der Gaag, K., Zuniga, S., Vermeulen, M., Van Duijn, K., Goedbloed, M., Vallone, P. M., Parson, W., et al. (2010). Inferring Continental Ancestry of Argentineans from Autosomal, Y-Chromosomal and Mitochondrial DNA. Ann. Hum. Genet. *74*, 65–76.

95. Seldin, M. F., Tian, C., Shigeta, R., Scherbarth, H. R., Silva, G., Belmont, J. W., Kittles, R., Gamron, S., Allevi, A., Palatnik, S. A., et al. (2007). Argentine population genetic structure: Large variance in Amerindian contribution. Am. J. Phys. Anthropol. *132*, 455–462.

96. Moreno-Estrada, A., Gignoux, C. R., Fernández-López, J. C., Zakharia, F., Sikora, M., Contreras, A. V., Acuña-Alonzo, V., Sandoval, K., Eng, C., Romero-Hidalgo, S., et al. (2014). The genetics of Mexico recapitulates Native American substructure and affects biomedical traits. Science *344*, 1280–1285.

97. Martinez-Marignac, V. L., Valladares, A., Cameron, E., Chan, A., Perera, A., Globus-Goldberg, R., Wacher, N., Kumate, J., McKeigue, P., O'Donnell, D., et al. (2007). Admixture in Mexico City: implications for admixture mapping of type 2 diabetes genetic risk factors. Hum. Genet. *120*, 807–819.

98. Freedman, M. L., Reich, D., Penney, K. L., McDonald, G. J., Mignault, A. A., Patterson, N., Gabriel, S. B., Topol, E. J., Smoller, J. W., Pato, C. N., et al. (2004). Assessing the impact of population stratification on genetic association studies. Nat. Genet. *36*, 388–393.

99. Lins, T. C., Vieira, R. G., Abreu, B. S., Gentil, P., Moreno-Lima, R., Oliveira, R. J., and Pereira, R. W. (2011). Genetic heterogeneity of self-reported ancestry groups in an admixed Brazilian population. J. Epidemiol. Jpn. Epidemiol. Assoc. *21*, 240–245.

100. Pimenta, J. R., Zuccherato, L. W., Debes, A. A., Maselli, L., Soares, R. P., Moura-Neto, R. S., Rocha, J., Bydlowski, S. P., and Pena, S. D. J. (2006). Color and genomic ancestry in Brazilians: a study with forensic microsatellites. Hum. Hered. *62*, 190–195.

101. Choudhry, S., Coyle, N. E., Tang, H., Salari, K., Lind, D., Clark, S. L., Tsai, H.-J., Naqvi, M., Phong, A., Ung, N., et al. (2006). Population stratification confounds genetic association studies among Latinos. Hum. Genet. *118*, 652–664.

102. Ripke, S., O'Dushlaine, C., Chambert, K., Moran, J. L., Kähler, A. K., Akterin, S., Bergen, S. E., Collins, A. L., Crowley, J. J., Fromer, M., et al. (2013). Genome-wide association analysis identifies 13 new risk loci for schizophrenia. Nat. Genet. *45*, 1150–1159.

103. Wheeler, E., and Barroso, I. (2011). Genome-wide association studies and type 2 diabetes. Brief. Funct. Genomics *10*, 52–60.

104. Lee, J. C., and Parkes, M. (2011). Genome-wide association studies and Crohn's disease. Brief. Funct. Genomics *10*, 71–76.

105. Simón-Sánchez, J., Schulte, C., Bras, J. M., Sharma, M., Gibbs, J. R., Berg, D., Paisan-Ruiz, C., Lichtner, P., Scholz, S. W., Hernandez, D. G., et al. (2009). Genome-wide association study reveals genetic risk underlying Parkinson's disease. Nat. Genet. *41*, 1308–1312.

106. Kosoy, R., Qi, L., Nassir, R., Garcia, L., Allison, M., Shigeta, R., Robbins, J., and Seldin, M. F. (2012). Relationship between hypertension and admixture in post-menopausal African American and Hispanic American women. J. Hum. Hypertens. *26*, 365–373.

107. Tang, H., Jorgenson, E., Gadde, M., Kardia, S. L. R., Rao, D. C., Zhu, X., Schork, N. J., Hanis, C. L., and Risch, N. (2006). Racial admixture and its impact on BMI and blood pressure in African and Mexican Americans. Hum. Genet. *119*, 624–633.

108. Fernández, J. R., Shriver, M. D., Beasley, T. M., Rafla-Demetrious, N., Parra, E., Albu, J., Nicklas, B., Ryan, A. S., McKeigue, P. M., Hoggart, C. L., et al. (2003). Association of African genetic admixture with resting metabolic rate and obesity among women. Obes. Res. *11*, 904–911.

109. Williams, R. C., Long, J. C., Hanson, R. L., Sievers, M. L., and Knowler, W. C. (2000). Individual estimates of European genetic admixture associated with lower body-mass index, plasma glucose, and prevalence of type 2 diabetes in Pima Indians. Am. J. Hum. Genet. *66*, 527–538.

110. Salari, K., Choudhry, S., Tang, H., Naqvi, M., Lind, D., Avila, P. C., Coyle, N. E., Ung, N., Nazario, S., Casal, J., et al. (2005). Genetic admixture and asthma-related phenotypes in Mexican American and Puerto Rican asthmatics. Genet. Epidemiol. *29*, 76–86.

111. Fejerman, L., Chen, G. K., Eng, C., Huntsman, S., Hu, D., Williams, A., Pasaniuc, B., John, E. M., Via, M., Gignoux, C., et al. (2012). Admixture mapping identifies a locus on 6q25 associated with breast cancer risk in US Latinas. Hum. Mol. Genet. *21*, 1907–1917.

112. Gurdasani, D., Carstensen, T., Tekola-Ayele, F., Pagani, L., Tachmazidou, I., Hatzikotoulas, K., Karthikeyan, S., Iles, L., Pollard, M. O., Choudhury, A., et al. (2015). The African Genome Variation Project shapes medical genetics in Africa. Nature *517*, 327–332.

113. Bandelt, H.-J., Macaulay, V., and Richards, M. (2006). Human Mitochondrial DNA and the Evolution of Homo sapiens.

114. Sagan, L. (1967). On the origin of mitosing cells. J. Theor. Biol. *14*, 255–274.

115. Thrash, J. C., Boyd, A., Huggett, M. J., Grote, J., Carini, P., Yoder, R. J., Robbertse, B., Spatafora, J. W., Rappé, M. S., and Giovannoni, S. J. (2011). Phylogenomic evidence for a common ancestor of mitochondria and the SAR11 clade. Sci. Rep. *1*.

116. Fox, T. D. (2012). Mitochondrial Protein Synthesis, Import, and Assembly. Genetics *192*, 1203–1234.

117. O'Connor, C., Adams, J. U., and Fairman, J. (2010). Essentials of Cell Biology (Cambridge: MA: NPG Education).

118. Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. The Mitochondrion. In Molecular Biology of the Cell (New York: Garland Science).

119. Zhang, Z.-W., Cheng, J., Xu, F., Chen, Y.-E., Du, J.-B., Yuan, M., Zhu, F., Xu, X.-C., and Yuan, S. (2011). Red blood cell extrudes nucleus and mitochondria against oxidative stress. IUBMB Life *63*, 560–565.

120. Anderson, S., Bankier, A. T., Barrell, B. G., de Bruijn, M. H., Coulson, A. R., Drouin, J., Eperon, I. C., Nierlich, D. P., Roe, B. A., Sanger, F., et al. (1981). Sequence and organization of the human mitochondrial genome. Nature *290*, 457–465.

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

122. Taylor, R. W., and Turnbull, D. M. (2005). Mitochondrial DNA mutations in human disease. Nat. Rev. Genet. *6*, 389–402.

123. Brown, T. A., Cecconi, C., Tkachuk, A. N., Bustamante, C., and Clayton, D. A. (2005). Replication of mitochondrial DNA occurs by strand displacement with alternative light-strand origins, not via a strand-coupled mechanism. Genes Dev. *19*, 2466–2476.

124. Giles, R. E., Blanc, H., Cann, H. M., and Wallace, D. C. (1980). Maternal inheritance of human mitochondrial DNA. Proc. Natl. Acad. Sci. U. S. A. *77*, 6715–6719.

125. Wai, T., Teoli, D., and Shoubridge, E. A. (2008). The mitochondrial DNA genetic bottleneck results from replication of a subpopulation of genomes. Nat. Genet. *40*, 1484–1488.

126. Griffiths, A. J. F., Wessler, S. R., Lewontin, R. C., Gelbart, W. M., Suzuki, D. T., and Miller, J. H. (2005). An Introduction to Genetic Analysis 8th ed. (New York: W. H. Freeman & Co).

127. Schon, E. A., Bonilla, E., and DiMauro, S. (1997). Mitochondrial DNA Mutations and Pathogenesis. J. Bioenerg. Biomembr. *29*, 131–149.

128. Li, M., Schönberg, A., Schaefer, M., Schroeder, R., Nasidze, I., and Stoneking, M. (2010). Detecting heteroplasmy from high-throughput sequencing of complete human mitochondrial DNA genomes. Am. J. Hum. Genet. *87*, 237–249.

129. Ramos, A., Santos, C., Mateiu, L., Gonzalez, M. del M., Alvarez, L., Azevedo, L., Amorim, A., and Aluja, M. P. (2013). Frequency and Pattern of Heteroplasmy in the Complete Human Mitochondrial Genome. PLoS ONE *8*, e74636.

130. Ladoukakis, E. D., and Eyre-Walker, A. (2004). Evolutionary genetics: Direct evidence of recombination in human mitochondrial DNA. Heredity *93*, 321–321.

131. Ruiz-Pesini, E., Lott, M. T., Procaccio, V., Poole, J. C., Brandon, M. C., Mishmar, D., Yi, C., Kreuziger, J., Baldi, P., and Wallace, D. C. (2007). An enhanced MITOMAP with a global mtDNA mutational phylogeny. Nucleic Acids Res. *35*, D823–D828.

132. Van Oven, M., and Kayser, M. (2009). Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Hum. Mutat. *30*, E386–394.

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

134. Behar, D. M., Villems, R., Soodyall, H., Blue-Smith, J., Pereira, L., Metspalu, E., Scozzari, R., Makkan, H., Tzur, S., Comas, D., et al. (2008). The Dawn of Human Matrilineal Diversity. Am. J. Hum. Genet. *82*, 1130–1140.

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

136. Gonder, M. K., Mortensen, H. M., Reed, F. A., De Sousa, A., and Tishkoff, S. A. (2007). Whole-mtDNA Genome Sequence Analysis of Ancient African Lineages. Mol. Biol. Evol. *24*, 757–768.

137. Schuster, S. C., Miller, W., Ratan, A., Tomsho, L. P., Giardine, B., Kasson, L. R., Harris, R. S., Petersen, D. C., Zhao, F., Qi, J., et al. (2010). Complete Khoisan and Bantu genomes from southern Africa. Nature *463*, 943–947.

138. Wilson, A. C., Cann, R. L., Carr, S. M., George, M., Gyllensten, U. B., Helm-Bychowski, K. M., Higuchi, R. G., Palumbi, S. R., Prager, E. M., Sage, R. D., et al. (1985). Mitochondrial DNA and two perspectives on evolutionary genetics. Biol. J. Linn. Soc. *26*, 375–400.

139. Cerezo, M., Achilli, A., Olivieri, A., Perego, U. A., Gómez-Carballa, A., Brisighelli, F., Lancioni, H., Woodward, S. R., López-Soto, M., Carracedo, Á., et al. (2012). Reconstructing Ancient Mitochondrial DNA Links Between Africa and Europe. Genome Res. *22*, 821–826.

140. Salas, A., Richards, M., De la Fe, T., Lareu, M.-V., Sobrino, B., Sánchez-Diz, P., Macaulay, V., and Carracedo, Á. (2002). The Making of the African mtDNA Landscape. Am. J. Hum. Genet. *71*, 1082.

141. Barbieri, C., Vicente, M., Rocha, J., Mpoloka, S. W., Stoneking, M., and Pakendorf, B. (2013). Ancient Substructure in Early mtDNA Lineages of Southern Africa. Am. J. Hum. Genet. *92*, 285–292.

142. Tishkoff, S. A., Gonder, M. K., Henn, B. M., Mortensen, H., Knight, A., Gignoux, C., Fernandopulle, N., Lema, G., Nyambo, T. B., Ramakrishnan, U., et al. (2007). History of Click-Speaking Populations of Africa Inferred from mtDNA and Y Chromosome Genetic Variation. Mol. Biol. Evol. *24*, 2180–2195.

143. Schlebusch, C. M., Naidoo, T., and Soodyall, H. (2009). SNaPshot minisequencing to resolve mitochondrial macro-haplogroups found in Africa. ELECTROPHORESIS *30*, 3657–3664.

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

145. Watson, E., Forster, P., Richards, M., and Bandelt, H. J. (1997). Mitochondrial footprints of human expansions in Africa. Am. J. Hum. Genet. *61*, 691–704.

146. Hartmann, A., Thieme, M., Nanduri, L. K., Stempfl, T., Moehle, C., Kivisild, T., and Oefner, P. J. (2009). Validation of microarray-based resequencing of 93 worldwide mitochondrial genomes. Hum. Mutat. *30*, 115–122.

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

148. The 1000 Genomes Project Consortium (2012). An integrated map of genetic variation from 1,092 human genomes. Nature *491*, 56–65.

149. Zheng, H.-X., Yan, S., Qin, Z.-D., and Jin, L. (2012). MtDNA analysis of global populations support that major population expansions began before Neolithic Time. Sci. Rep. *2*.

150. Barbieri, C., Butthof, A., Bostoen, K., and Pakendorf, B. (2013). Genetic perspectives on the origin of clicks in Bantu languages from southwestern Zambia. Eur. J. Hum. Genet. *21*, 430–436.

151. Bandelt, H.-J., Achilli, A., Kong, Q.-P., Salas, A., Lutz-Bonengel, S., Sun, C., Zhang, Y.-P., Torroni, A., and Yao, Y.-G. (2005). Low "penetrance" of phylogenetic knowledge in mitochondrial disease studies. Biochem. Biophys. Res. Commun. *333*, 122–130.

152. Barbieri, C., Whitten, M., Beyer, K., Schreiber, H., Li, M., and Pakendorf, B. (2012). Contrasting maternal and paternal histories in the linguistic context of Burkina Faso. Mol. Biol. Evol. *29*, 1213–1223.

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

154. Kujanová, M., Pereira, L., Fernandes, V., Pereira, J. B., and Cerný, V. (2009). Near eastern neolithic genetic input in a small oasis of the Egyptian Western Desert. Am. J. Phys. Anthropol. *140*, 336–346.

155. Just, R. S., Diegoli, T. M., Saunier, J. L., Irwin, J. A., and Parsons, T. J. (2008). Complete mitochondrial genome sequences for 265 African American and U.S. "Hispanic" individuals. Forensic Sci. Int. Genet. *2*, e45–48.

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

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

158. Costa, M. D., Cherni, L., Fernandes, V., Freitas, F., Ammar El Gaaied, A. B., and Pereira, L. (2009). Data from complete mtDNA sequencing of Tunisian centenarians: testing haplogroup association and the "golden mean" to longevity. Mech. Ageing Dev. *130*, 222–226.

159. Mielnik-Sikorska, M., Daca, P., Malyarchuk, B., Derenko, M., Skonieczna, K., Perkova, M., Dobosz, T., and Grzybowski, T. (2013). The History of Slavs Inferred from Complete Mitochondrial Genome Sequences. PLoS ONE *8*, e54360.

160. Malyarchuk, B. A., Derenko, M., Perkova, M., Grzybowski, T., Vanecek, T., and Lazur, J. (2008). Reconstructing the phylogeny of African mitochondrial DNA lineages in Slavs. Eur. J. Hum. Genet. EJHG *16*, 1091–1096.

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

162. Johnson, D. C., Shrestha, S., Wiener, H. W., Makowsky, R., Kurundkar, A., Wilson, C. M., and Aissani, B. (2013). Mitochondrial DNA diversity in the African American population. Mitochondrial DNA, 1–7.

163. Achilli, A., Iommarini, L., Olivieri, A., Pala, M., Hooshiar Kashani, B., Reynier, P., La Morgia, C., Valentino, M. L., Liguori, R., Pizza, F., et al. (2012). Rare Primary Mitochondrial DNA Mutations and Probable Synergistic Variants in Leber's Hereditary Optic Neuropathy. PLoS ONE *7*, e42242.

164. Gómez-Carballa, A., Cerezo, M., Balboa, E., Heredia, C., Castro-Feijóo, L., Rica, I., Barreiro, J., Eirís, J., Cabanas, P., Martínez-Soto, I., et al. (2011). Evolutionary analyses of entire genomes do not support the association of mtDNA mutations with Ras/MAPK pathway syndromes. PloS One *6*, e18348.

165. Torroni, A., Rengo, C., Guida, V., Cruciani, F., Sellitto, D., Coppa, A., Calderon, F. L., Simionati, B., Valle, G., Richards, M., et al. (2001). Do the Four Clades of the mtDNA Haplogroup L2 Evolve at Different Rates? Am. J. Hum. Genet. *69*, 1348–1356.

166. Soares, P., Alshamali, F., Pereira, J. B., Fernandes, V., Silva, N. M., Afonso, C., Costa, M. D., Musilova, E., Macaulay, V., Richards, M. B., et al. (2011). The expansion of mtDNA haplogroup L3 within and out of Africa. Mol. Biol. Evol., msr245.

167. Plaza, S., Salas, A., Calafell, F., Corte-Real, F., Bertranpetit, J., Carracedo, A., and Comas, D. (2004). Insights into the western Bantu dispersal: mtDNA lineage analysis in Angola. Hum. Genet. *115*, 439–447.

168. Harich, N., Costa, M. D., Fernandes, V., Kandil, M., Pereira, J. B., Silva, N. M., and Pereira, L. (2010). The trans-Saharan slave trade - clues from interpolation analyses and high-resolution characterization of mitochondrial DNA lineages. BMC Evol. Biol. *10*, 138.

169. Kivisild, T., Reidla, M., Metspalu, E., Rosa, A., Brehm, A., Pennarun, E., Parik, J., Geberhiwot, T., Usanga, E., and Villems, R. (2004). Ethiopian mitochondrial DNA heritage: tracking gene flow across and around the gate of tears. Am. J. Hum. Genet. *75*, 752–770.

170. Shen, P., Lavi, T., Kivisild, T., Chou, V., Sengun, D., Gefel, D., Shpirer, I., Woolf, E., Hillel, J., Feldman, M. W., et al. (2004). Reconstruction of patrilineages and matrilineages of Samaritans and other Israeli populations from Y-chromosome and mitochondrial DNA sequence variation. Hum. Mutat. *24*, 248–260.

171. Bandelt, H.-J., Herrnstadt, C., Yao, Y.-G., Kong, Q.-P., Kivisild, T., Rengo, C., Scozzari, R., Richards, M., Villems, R., Macaulay, V., et al. (2003). Identification of Native American founder mtDNAs through the analysis of complete mtDNA sequences: some caveats. Ann. Hum. Genet. *67*, 512–524.

172. Tamm, E., Kivisild, T., Reidla, M., Metspalu, M., Smith, D. G., Mulligan, C. J., Bravi, C. M., Rickards, O., Martinez-Labarga, C., Khusnutdinova, E. K., et al. (2007). Beringian Standstill and Spread of Native American Founders. PLoS ONE *2*, e829.

173. Helgason, A., Pálsson, G., Pedersen, H. S., Angulalik, E., Gunnarsdóttir, E. D., Yngvadóttir, B., and Stefánsson, K. (2006). mtDNA variation in Inuit populations of Greenland and Canada: migration history and population structure. Am. J. Phys. Anthropol. *130*, 123–134.

174. Kumar, S., Bellis, C., Zlojutro, M., Melton, P. E., Blangero, J., and Curran, J. E. (2011). Large scale mitochondrial sequencing in Mexican Americans suggests a reappraisal of Native American origins. BMC Evol. Biol. *11*, 293.

175. Perego, U. A., Achilli, A., Angerhofer, N., Accetturo, M., Pala, M., Olivieri, A., Hooshiar Kashani, B., Ritchie, K. H., Scozzari, R., Kong, Q.-P., et al. (2009). Distinctive Paleo-Indian migration routes from Beringia marked by two rare mtDNA haplogroups. Curr. Biol. CB *19*, 1–8.

176. Hooshiar Kashani, B., Perego, U. A., Olivieri, A., Angerhofer, N., Gandini, F., Carossa, V., Lancioni, H., Semino, O., Woodward, S. R., Achilli, A., et al. (2012). Mitochondrial haplogroup C4c: a rare lineage entering America through the ice-free corridor? Am. J. Phys. Anthropol. *147*, 35–39.

177. De Saint Pierre, M., Gandini, F., Perego, U. A., Bodner, M., Gómez-Carballa, A., Corach, D., Angerhofer, N., Woodward, S. R., Semino, O., Salas, A., et al. (2012). Arrival of Paleo-Indians to the Southern Cone of South America: New Clues from Mitogenomes. PLoS ONE *7*, e51311.

178. Achilli, A., Perego, U. A., Lancioni, H., Olivieri, A., Gandini, F., Kashani, B. H., Battaglia, V., Grugni, V., Angerhofer, N., Rogers, M. P., et al. (2013). Reconciling migration models to the Americas with the variation of North American native mitogenomes. Proc. Natl. Acad. Sci. *110*, 14308-14313.

179. Achilli, A., Perego, U. A., Bravi, C. M., Coble, M. D., Kong, Q.-P., Woodward, S. R., Salas, A., Torroni, A., and Bandelt, H.-J. (2008). The Phylogeny of the Four Pan-American MtDNA Haplogroups: Implications for Evolutionary and Disease Studies. PLoS ONE *3*, e1764.

180. Pereira, L., Richards, M., Goios, A., Alonso, A., Albarrán, C., Garcia, O., Behar, D. M., Gölge, M., Hatina, J., Al-Gazali, L., et al. (2005). High-resolution mtDNA evidence for the late-glacial resettlement of Europe from an Iberian refugium. Genome Res. *15*, 19–24.

181. Loogväli, E.-L., Roostalu, U., Malyarchuk, B. A., Derenko, M. V., Kivisild, T., Metspalu, E., Tambets, K., Reidla, M., Tolk, H.-V., Parik, J., et al. (2004). Disuniting uniformity: a pied cladistic canvas of mtDNA haplogroup H in Eurasia. Mol. Biol. Evol. *21*, 2012–2021.

182. Tambets, K., Rootsi, S., Kivisild, T., Help, H., Serk, P., Loogväli, E.-L., Tolk, H.-V., Reidla, M., Metspalu, E., Pliss, L., et al. (2004). The western and eastern roots of the Saami--the story of genetic "outliers" told by mitochondrial DNA and Y chromosomes. Am. J. Hum. Genet. *74*, 661–682.

183. Malyarchuk, B. A., Grzybowski, T., Derenko, M. V., Czarny, J., and Miścicka-Sliwka, D. (2006). Mitochondrial DNA diversity in the Polish Roma. Ann. Hum. Genet. *70*, 195–206.

184. Bedford, F. L. (2012). Sephardic signature in haplogroup T mitochondrial DNA. Eur. J. Hum. Genet. *20*, 441–448.

185. Fernandes, V., Alshamali, F., Alves, M., Costa, M. D., Pereira, J. B., Silva, N. M., Cherni, L., Harich, N., Cerny, V., Soares, P., et al. (2012). The Arabian Cradle: Mitochondrial Relicts of the First Steps along the Southern Route out of Africa. Am. J. Hum. Genet. *90*, 347–355.

186. Richards, M., Macaulay, V., Hickey, E., Vega, E., Sykes, B., Guida, V., Rengo, C., Sellitto, D., Cruciani, F., Kivisild, T., et al. (2000). Tracing European Founder Lineages in the Near Eastern mtDNA Pool. Am. J. Hum. Genet. *67*, 1251–1276.

187. Berta, P., Hawkins, J. R., Sinclair, A. H., Taylor, A., Griffiths, B. L., Goodfellow, P. N., and Fellous, M. (1990). Genetic evidence equating SRY and the testis-determining factor. Nature *348*, 448–450.

188. Skaletsky, H., Kuroda-Kawaguchi, T., Minx, P. J., Cordum, H. S., Hillier, L., Brown, L. G., Repping, S., Pyntikova, T., Ali, J., Bieri, T., et al. (2003). The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature *423*, 825–837.

189. Ramana, G. V., Su, B., Jin, L., Singh, L., Wang, N., Underhill, P., and Chakraborty, R. (2001). Y-chromosome SNP haplotypes suggest evidence of gene flow among caste, tribe, and the migrant Siddi populations of Andhra Pradesh, South India. Eur. J. Hum. Genet. *9*, 695–700.

190. Larmuseau, M. H. D., Vanoverbeke, J., Gielis, G., Vanderheyden, N., Larmuseau, H. F. M., and Decorte, R. (2012). In the name of the migrant father—Analysis of surname origins identifies genetic admixture events undetectable from genealogical records. Heredity *109*, 90–95.

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

192. Lessig, R., Zoledziewska, M., Fahr, K., Edelmann, J., Kostrzewa, M., Dobosz, T., and Kleemann, W. J. (2005). Y-SNP-genotyping – a new approach in forensic analysis. Forensic Sci. Int. *154*, 128–136.

193. The Y chromosome Consortium. (2002). A Nomenclature System for the Tree of Human Y-Chromosomal Binary Haplogroups. Genome Res. *12*, 339–348.

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

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

196. Batini, C., Ferri, G., Destro-Bisol, G., Brisighelli, F., Luiselli, D., Sánchez-Diz, P., Rocha, J., Simonson, T., Brehm, A., Montano, V., 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., msr089.

197. Cruciani, F., Trombetta, B., Massaia, A., Destro-Bisol, G., Sellitto, D., and Scozzari, R. (2011). A Revised Root for the Human Y Chromosomal Phylogenetic Tree: The Origin of Patrilineal Diversity in Africa. Am. J. Hum. Genet. *88*, 814–818.

198. Semino, O., Santachiara-Benerecetti, A. S., Falaschi, F., Cavalli-Sforza, L. L., and Underhill, P. A. (2002). Ethiopians and Khoisan share the deepest clades of the human Y-chromosome phylogeny. Am. J. Hum. Genet. *70*, 265–268.

199. Cruciani, F., Santolamazza, P., Shen, P., Macaulay, V., Moral, P., Olckers, A., Modiano, D., Holmes, S., Destro-Bisol, G., Coia, V., 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*, 1197–1214.

200. Gomes, V., Sánchez-Diz, P., Amorim, A., Carracedo, Á., and Gusmão, L. (2010). Digging deeper into East African human Y chromosome lineages. Hum. Genet. *127*, 603–613.

201. Bortolini, M.-C., Salzano, F. M., Thomas, M. G., Stuart, S., Nasanen, S. P. K., Bau, C. H. D., Hutz, M. H., Layrisse, Z., Petzl-Erler, M. L., Tsuneto, L. T., et al. (2003). Y-Chromosome Evidence for Differing Ancient Demographic Histories in the Americas. Am. J. Hum. Genet. *73*, 524–539.

202. Malhi, R. S., Gonzalez-Oliver, A., Schroeder, K. B., Kemp, B. M., Greenberg, J. A., Dobrowski, S. Z., Smith, D. G., Resendez, A., Karafet, T., Hammer, M., et al. (2008). Distribution of Y chromosomes among Native North Americans: A study of Athapaskan population history. Am. J. Phys. Anthropol. *137*, 412–424.

203. Di Cristofaro, J., Pennarun, E., Mazières, S., Myres, N. M., Lin, A. A., Temori, S. A., Metspalu, M., Metspalu, E., Witzel, M., King, R. J., et al. (2013). Afghan Hindu Kush: Where Eurasian Sub-Continent Gene Flows Converge. PLoS ONE *8*, e76748.

204. Derenko, M., Malyarchuk, B., Denisova, G. A., Wozniak, M., Dambueva, I., Dorzhu, C., Luzina, F., Miścicka-Śliwka, D., and Zakharov, I. (2006). Contrasting patterns of Y-chromosome variation in South Siberian populations from Baikal and Altai-Sayan regions. Hum. Genet. *118*, 591–604.

205. Bosch, E., Calafell, F., González-Neira, A., Flaiz, C., Mateu, E., Scheil, H.-G., Huckenbeck, W., Efremovska, L., Mikerezi, I., Xirotiris, N., et al. (2006). Paternal and maternal lineages in the Balkans show a homogeneous landscape over linguistic barriers, except for the isolated Aromuns. Ann. Hum. Genet. *70*, 459–487.

206. Hudjashov, G., Kivisild, T., Underhill, P. A., Endicott, P., Sanchez, J. J., Lin, A. A., Shen, P., Oefner, P., Renfrew, C., Villems, R., et al. (2007). Revealing the prehistoric settlement of Australia by Y chromosome and mtDNA analysis. Proc. Natl. Acad. Sci. *104*, 8726–8730.

207. Kayser, M., Brauer, S., Cordaux, R., Casto, A., Lao, O., Zhivotovsky, L. A., Moyse-Faurie, C., Rutledge, R. B., Schiefenhoevel, W., Gil, D., et al. (2006). Melanesian and Asian Origins of Polynesians: mtDNA and Y Chromosome Gradients Across the Pacific. Mol. Biol. Evol. *23*, 2234–2244.

208. Naitoh, S., Kasahara-Nonaka, I., Minaguchi, K., and Nambiar, P. (2013). Assignment of Y-chromosomal SNPs found in Japanese population to Y-chromosomal haplogroup tree. J. Hum. Genet. *58*, 195–201.

209. Hammer, M. F., Karafet, T. M., Park, H., Omoto, K., Harihara, S., Stoneking, M., and Horai, S. (2005). Dual origins of the Japanese: common ground for hunter-gatherer and farmer Y chromosomes. J. Hum. Genet. *51*, 47–58.

210. Shou, W.-H., Qiao, E.-F., Wei, C.-Y., Dong, Y.-L., Tan, S.-J., Shi, H., Tang, W.-R., and Xiao, C.-J. (2010). Y-chromosome distributions among populations in Northwest China identify significant contribution from Central Asian pastoralists and lesser influence of western Eurasians. J. Hum. Genet. *55*, 314–322.

211. Mohyuddin, A., Ayub, Q., Underhill, P. A., Tyler-Smith, C., and Mehdi, S. Q. (2006). Detection of novel Y SNPs provides further insights into Y chromosomal variation in Pakistan. J. Hum. Genet. *51*, 375–378.

212. Mona, S., Tommaseo-Ponzetta, M., Brauer, S., Sudoyo, H., Marzuki, S., and Kayser, M. (2007). Patterns of Y-Chromosome Diversity Intersect with the Trans-New Guinea Hypothesis. Mol. Biol. Evol. *24*, 2546–2555.

213. Deng, W., Shi, B., He, X., Zhang, Z., Xu, J., Li, B., Yang, J., Ling, L., Dai, C., Qiang, B., et al. (2004). Evolution and migration history of the Chinese population inferred from Chinese Y-chromosome evidence. J. Hum. Genet. *49*, 339–348.

214. Shi, H., Zhong, H., Peng, Y., Dong, Y.-L., Qi, X.-B., Zhang, F., Liu, L.-F., Tan, S.-J., Ma, R. Z., Xiao, C.-J., et al. (2008). Y chromosome evidence of earliest modern human settlement in East Asia and multiple origins of Tibetan and Japanese populations. BMC Biol. *6*, 45.

215. International Society of Genetic Genealogy (2015). Y-DNA Haplogroup Tree 2014 (Version: 10.01). Available at: http://www.isogg.org/tree/.

216. Wang, C.-C., and Li, H. (2013). Inferring human history in East Asia from Y chromosomes. Investig. Genet. *4*, 11.

217. Cruciani, F., Fratta, R. L., Trombetta, B., Santolamazza, P., Sellitto, D., Colomb, E. B., Dugoujon, J.-M., Crivellaro, F., Benincasa, T., Pascone, R., et al. (2007). Tracing Past Human Male Movements in Northern/Eastern Africa and Western Eurasia: New Clues from Y-Chromosomal Haplogroups E-M78 and J-M12. Mol. Biol. Evol. *24*, 1300–1311.

218. Trombetta, B., Cruciani, F., Sellitto, D., and Scozzari, R. (2011). A New Topology of the Human Y Chromosome Haplogroup E1b1 (E-P2) Revealed through the Use of Newly Characterized Binary Polymorphisms. PLoS ONE *6*, e16073.

219. Simms, T. M., Wright, M. R., Hernandez, M., Perez, O. A., Ramirez, E. C., Martinez, E., and Herrera, R. J. (2012). Y-chromosomal diversity in Haiti and Jamaica: contrasting levels of sex-biased gene flow. Am. J. Phys. Anthropol. *148*, 618–631.

220. Hammer, M. F., Chamberlain, V. F., Kearney, V. F., Stover, D., Zhang, G., Karafet, T., Walsh, B., and Redd, A. J. (2006). Population structure of Y chromosome SNP haplogroups in the United States and forensic implications for constructing Y chromosome STR databases. Forensic Sci. Int. *164*, 45–55.

221. Zhao, Z., Khan, F., Borkar, M., Herrera, R., and Agrawal, S. (2009). Presence of three different paternal lineages among North Indians: a study of 560 Y chromosomes. Ann. Hum. Biol. *36*, 46–59.

222. Poznik, G. D., Henn, B. M., Yee, M.-C., Sliwerska, E., Euskirchen, G. M., Lin, A. A., Snyder, M., Quintana-Murci, L., Kidd, J. M., Underhill, P. A., et al. (2013). Sequencing Y Chromosomes Resolves Discrepancy in Time to Common Ancestor of Males Versus Females. Science *341*, 562–565.

223. Kivisild, T., Rootsi, S., Metspalu, M., Mastana, S., Kaldma, K., Parik, J., Metspalu, E., Adojaan, M., Tolk, H.-V., Stepanov, V., et al. (2003). The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations. Am. J. Hum. Genet. *72*, 313–332.

224. Rootsi, S., Magri, C., Kivisild, T., Benuzzi, G., Help, H., Bermisheva, M., Kutuev, I., Barać, L., Perićić, M., Balanovsky, O., et al. (2004). Phylogeography of Y-chromosome haplogroup I reveals distinct domains of prehistoric gene flow in europe. Am. J. Hum. Genet. *75*, 128–137.

225. Semino, O., Magri, C., Benuzzi, G., Lin, A. A., Al-Zahery, N., Battaglia, V., Maccioni, L., Triantaphyllidis, C., Shen, P., Oefner, P. J., et al. (2004). Origin, Diffusion, and Differentiation of Y-Chromosome Haplogroups E and J: Inferences on the Neolithization of Europe and Later Migratory Events in the Mediterranean Area. Am. J. Hum. Genet. *74*, 1023–1034.

226. Di Giacomo, F., Luca, F., Popa, L. O., Akar, N., Anagnou, N., Banyko, J., Brdicka, R., Barbujani, G., Papola, F., Ciavarella, G., et al. (2004). Y chromosomal haplogroup J as a signature of the post-neolithic colonization of Europe. Hum. Genet. *115*, 357–371.

227. Karafet, T. M., Mendez, F. L., Sudoyo, H., Lansing, J. S., and Hammer, M. F. (2014). Improved phylogenetic resolution and rapid diversification of Y-chromosome haplogroup K-M526 in Southeast Asia. Eur. J. Hum. Genet.

228. Sengupta, S., Zhivotovsky, L. A., King, R., Mehdi, S. Q., Edmonds, C. A., Chow, C.-E. T., Lin, A. A., Mitra, M., Sil, S. K., Ramesh, A., et al. (2006). Polarity and Temporality of High-Resolution Y-Chromosome Distributions in India Identify Both Indigenous and Exogenous Expansions and Reveal Minor Genetic Influence of Central Asian Pastoralists. Am. J. Hum. Genet. *78*, 202–221.

229. Kayser, M., Brauer, S., Weiss, G., Schiefenhövel, W., Underhill, P., Shen, P., Oefner, P., Tommaseo-Ponzetta, M., and Stoneking, M. (2003). Reduced Y-Chromosome, but Not Mitochondrial DNA, Diversity in Human Populations from West New Guinea. Am. J. Hum. Genet. *72*, 281–302.

230. Rootsi, S., Zhivotovsky, L. A., Baldovič, M., Kayser, M., Kutuev, I. A., Khusainova, R., Bermisheva, M. A., Gubina, M., Fedorova, S. A., Ilumäe, A.-M., et al. (2006). A counter-clockwise northern route of the Y-chromosome haplogroup N from Southeast Asia towards Europe. Eur. J. Hum. Genet. *15*, 204–211.

231. Gayden, T., Cadenas, A. M., Regueiro, M., Singh, N. B., Zhivotovsky, L. A., Underhill, P. A., Cavalli-Sforza, L. L., and Herrera, R. J. (2007). The Himalayas as a Directional Barrier to Gene Flow. Am. J. Hum. Genet. *80*, 884–894.

232. Geppert, M., Baeta, M., Núñez, C., Martínez-Jarreta, B., Zweynert, S., Cruz, O. W. V., González-Andrade, F., González-Solorzano, J., Nagy, M., and Roewer, L. (2011). Hierarchical Y-SNP assay to study the hidden diversity and phylogenetic relationship of native populations in South America. Forensic Sci. Int. Genet. *5*, 100–104.

233. Vullo, C., Gomes, V., Romanini, C., Oliveira, A. M., Rocabado, O., Aquino, J., Amorim, A., and Gusmão, L. (2014). Association between Y haplogroups and autosomal AIMs reveals intra-population substructure in Bolivian populations. Int. J. Legal Med.

234. Bailliet, G., Muzzio, M., Ramallo, V., Jurado Medina, L. S., L., E., J., and M., C. (2012). Pre-Columbian Male Ancestors for the American Continent, Molecular Y-Chromosome Insight. In Genetic Diversity in Microorganisms, M. Caliskan, ed. (InTech).

235. Jota, M. S., Lacerda, D. R., Sandoval, J. R., Vieira, P. P. R., Santos-Lopes, S. S., Bisso-Machado, R., Paixão-Cortes, V. R., Revollo, S., Paz-Y-Miño, C., Fujita, R., et al. (2011). A new subhaplogroup of native American Y-Chromosomes from the Andes. Am. J. Phys. Anthropol. *146*, 553–559.

236. Underhill, P. A., Poznik, G. D., Rootsi, S., Järve, M., Lin, A. A., Wang, J., Passarelli, B., Kanbar, J., Myres, N. M., King, R. J., et al. (2015). The phylogenetic and geographic structure of Y-chromosome haplogroup R1a. Eur. J. Hum. Genet. *23*, 124–131.

237. Myres, N. M., Rootsi, S., Lin, A. A., Järve, M., King, R. J., Kutuev, I., Cabrera, V. M., Khusnutdinova, E. K., Pshenichnov, A., Yunusbayev, B., et al. (2011). A major Y-chromosome haplogroup R1b Holocene era founder effect in Central and Western Europe. Eur. J. Hum. Genet. *19*, 95–101.

238. Stumpf, M. P. H., and Goldstein, D. B. (2001). Genealogical and Evolutionary Inference with the Human Y Chromosome. Science *291*, 1738–1742.

239. Jobling, M. A. (2001). In the name of the father: surnames and genetics. Trends Genet. TIG *17*, 353–357.

240. Kayser, M., Caglià, A., Corach, D., Fretwell, N., Gehrig, C., Graziosi, G., Heidorn, F., Herrmann, S., Herzog, B., Hidding, M., et al. (1997). Evaluation of Y-chromosomal STRs: a multicenter study. Int. J. Legal Med. *110*, 125–133, 141–149.

241. Jobling, M. A., Pandya, A., and Tyler-Smith, C. (1997). The Y chromosome in forensic analysis and paternity testing. Int. J. Legal Med. *110*, 118–124.

242. Kayser, M., Kittler, R., Erler, A., Hedman, M., Lee, A. C., Mohyuddin, A., Mehdi, S. Q., Rosser, Z., Stoneking, M., Jobling, M. A., et al. (2004). A Comprehensive Survey of Human Y-Chromosomal Microsatellites. Am. J. Hum. Genet. *74*, 1183–1197.

243. Butler, J. M. (2003). Recent Developments in Y-Short Tandem Repeat and Y-Single Nucleotide Polymorphism Analysis. Forensic Sci. Rev. *15*.

244. Gusmão, L., Butler, J. M., Carracedo, A., Gill, P., Kayser, M., Mayr, W. R., Morling, N., Prinz, M., Roewer, L., Tyler-Smith, C., et al. (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. Int. J. Legal Med. *120*, 191–200.

245. Gusmão, L., Sánchez-Diz, P., Calafell, F., Martín, P., Alonso, C. a., Álvarez-Fernández, F., Alves, C., Borjas-Fajardo, L., Bozzo, W. r., Bravo, M. l., et al. (2005). Mutation rates at Y chromosome specific microsatellites. Hum. Mutat. *26*, 520–528.

246. Kayser, M., Roewer, L., Hedman, M., Henke, L., Henke, J., Brauer, S., Krüger, C., Krawczak, M., Nagy, M., Dobosz, T., et al. (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*, 1580–1588.

247. Ballantyne, K. N., Goedbloed, M., Fang, R., Schaap, O., Lao, O., Wollstein, A., Choi, Y., van Duijn, K., Vermeulen, M., Brauer, S., et al. (2010). Mutability of Y-Chromosomal Microsatellites: Rates, Characteristics, Molecular Bases, and Forensic Implications. Am. J. Hum. Genet. *87*, 341–353.

248. Roewer, L., and Epplen, J. T. (1992). Rapid and sensitive typing of forensic stains by PCR amplification of polymorphic simple repeat sequences in case work. Forensic Sci. Int. *53*, 163–171.

249. Roewer, L., Krawczak, M., Willuweit, S., Nagy, M., Alves, C., Amorim, A., Anslinger, K., Augustin, C., Betz, A., Bosch, E., et al. (2001). Online reference database of European Y-chromosomal short tandem repeat (STR) haplotypes. Forensic Sci. Int. *118*, 106–113.

250. Kittler, R., Erler, A., Brauer, S., Stoneking, M., and Kayser, M. (2003). Apparent intrachromosomal exchange on the human Y chromosome explained by population history. Eur. J. Hum. Genet. *11*, 304–314.

251. Brisighelli, F., Blanco-Verea, A., Boschi, I., Garagnani, P., Pascali, V. L., Carracedo, A., Capelli, C., and Salas, A. (2012). Patterns of Y-STR variation in Italy. Forensic Sci. Int. Genet. *6*, 834–839.

252. Willuweit, S., Roewer, L., and International Forensic Y Chromosome User Group (2007). Y chromosome haplotype reference database (YHRD): update. Forensic Sci. Int. Genet. *1*, 83–87.

253. Mulero, J. J., Chang, C. W., Calandro, L. M., Green, R. L., Li, Y., Johnson, C. L., and Hennessy, L. K. (2006). Development and validation of the AmpFlSTR Yfiler PCR amplification kit: a male specific, single amplification 17 Y-STR multiplex system. J. Forensic Sci. *51*, 64–75.

254. Oota, H., Settheetham-Ishida, W., Tiwawech, D., Ishida, T., and Stoneking, M. (2001). Human mtDNA and Y-chromosome variation is correlated with matrilocal versus patrilocal residence. Nat. Genet. *29*, 20–21.

255. Burton, M. L., Moore, C. C., Whiting, J. W., Romney, A. K., Aberle, D. F., Barcelo, J. A., Dow, M. M., Guyer, J. I., Kronenfeld, D. B., Levy, J. E., et al. (1996). Regions Based on Social Structure: A Reconsideration. Curr. Anthropol. *37*, 87–123.

256. Underhill, P. A., and Kivisild, T. (2007). Use of Y Chromosome and Mitochondrial DNA Population Structure in Tracing Human Migrations. Annu. Rev. Genet. *41*, 539–564.

257. Dulik, M. C., Zhadanov, S. I., Osipova, L. P., Askapuli, A., Gau, L., Gokcumen, O., Rubinstein, S., and Schurr, T. G. (2012). Mitochondrial DNA and Y Chromosome Variation Provides Evidence for a Recent Common Ancestry between Native Americans and Indigenous Altaians. Am. J. Hum. Genet. *90*, 229–246.

258. Goldsmith, L., Jackson, L., O'Connor, A., and Skirton, H. (2012). Direct-to-consumer genomic testing: systematic review of the literature on user perspectives. Eur. J. Hum. Genet. *20*, 811–816.

259. Galanter, J. M., Fernandez-Lopez, J. C., Gignoux, C. R., Barnholtz-Sloan, J., Fernandez-Rozadilla, C., Via, M., Hidalgo-Miranda, A., Contreras, A. V., Figueroa, L. U., Raska, P., et al. (2012). Development of a Panel of Genome-Wide Ancestry Informative Markers to Study Admixture Throughout the Americas. PLoS Genet *8*, e1002554.

260. Montana, G., and Pritchard, J. K. (2004). Statistical Tests for Admixture Mapping with Case-Control and Cases-Only Data. Am. J. Hum. Genet. *75*, 771–789.

261. Lao, O., van Duijn, K., Kersbergen, P., de Knijff, P., and Kayser, M. (2006). Proportioning whole-genome single-nucleotide-polymorphism diversity for the identification of geographic population structure and genetic ancestry. Am. J. Hum. Genet. *78*, 680–690.

262. Phillips, C., Salas, A., Sánchez, J. J., Fondevila, M., Gómez-Tato, A., Álvarez-Dios, J., Calaza, M., de Cal, M. C., Ballard, D., Lareu, M. V., et al. (2007). Inferring ancestral origin using a single multiplex assay of ancestry-informative marker SNPs. Forensic Sci. Int. Genet. *1*, 273–280.

263. Kurotaki, N., Tasaki, S., Mishima, H., Ono, S., Imamura, A., Kikuchi, T., Nishida, N., Tokunaga, K., Yoshiura, K., and Ozawa, H. (2011). Identification of novel schizophrenia loci by homozygosity mapping using DNA microarray analysis. PloS One *6*, e20589.

264. Jensen, K. P., Stein, M. B., Kranzler, H. R., Yang, B. Z., Farrer, L. A., and Gelernter, J. (2014). The α-endomannosidase gene (MANEA) is associated with panic disorder and social anxiety disorder. Transl. Psychiatry *4*, e353.

265. Nakamura, M., Nishida, N., Kawashima, M., Aiba, Y., Tanaka, A., Yasunami, M., Nakamura, H., Komori, A., Nakamuta, M., Zeniya, M., et al. (2012). Genome-wide Association Study Identifies TNFSF15 and POU2AF1 as Susceptibility Loci for Primary Biliary Cirrhosis in the Japanese Population. Am. J. Hum. Genet. *91*, 721–728.

266. Pardo-Seco, J., Martinón-Torres, F., and Salas, A. (2014). Evaluating the accuracy of AIM panels at quantifying genome ancestry. BMC Genomics *15*, 543.

267. Bandelt, H.-J., Yao, Y.-G., Richards, M. B., and Salas, A. (2008). The brave new era of human genetic testing. BioEssays *30*, 1246–1251.

268. Janssens, A. C. J. W., Gwinn, M., Bradley, L. A., Oostra, B. A., van Duijn, C. M., and Khoury, M. J. (2008). A critical appraisal of the scientific basis of commercial genomic profiles used to assess health risks and personalize health interventions. Am. J. Hum. Genet. *82*, 593–599.

269. Pereira, R., Phillips, C., Pinto, N., Santos, C., Santos, S. E. B. dos, Amorim, A., Carracedo, Á., and Gusmão, L. (2012). Straightforward Inference of Ancestry and Admixture Proportions through Ancestry-Informative Insertion Deletion Multiplexing. PLoS ONE *7*, e29684.

270. Londin, E. R., Keller, M. A., Maista, C., Smith, G., Mamounas, L. A., Zhang, R., Madore, S. J., Gwinn, K., and Corriveau, R. A. (2010). CoAIMs: A Cost-Effective Panel of Ancestry Informative Markers for Determining Continental Origins. PLoS ONE *5*, e13443.

271. Santos, N. P. C., Ribeiro-Rodrigues, E. M., Ribeiro-dos-Santos, Â. K. C., Pereira, R., Gusmão, L., Amorim, A., Guerreiro, J. F., Zago, M. A., Matte, C., Hutz, M. H., et al. (2010). Assessing individual interethnic admixture and population substructure using a 48–insertion-deletion (INSEL) ancestry-informative marker (AIM) panel. Hum. Mutat. *31*, 184–190.

272. Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. Genetics *155*, 945–959.

273. Alexander, D. H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome Res. *19*, 1655–1664.

274. The World Factbook 2013-14 (2013). (Washington, D.C.: Central Intelligence Agency). Available at: https://www.cia.gov/library/publications/the-world-factbook/index.html.

275. Rodríguez, R. J. (2001). The Afro populations of America's southern cone: organization, development, and culture in Argentina, Bolivia, Paraguay, and Uruguay. In African roots/American cultures - Africa in the creation of the Americas (Lanham, MD: Rowman & Littlefield Publishers, Inc.).

276. Lipski, J. M. (2006). Afro-Bolivian Spanish and Helvécia Portuguese: semi-creole parallels. Papia *16*, 96–116.

277. Busdiecker, S. (2009). Where Blackness Resides: Afro-Bolivians and the Spatializing and Racializing of the African Diaspora. Radic. Hist. Rev. *2009*, 105–116.

278. Maconde, J. A. (2007). Los afrodescendientes bolivianos. J. Lat. Am. Caribb. Anthropol. *12*, 246–253.

279. Mendizabal, I., Sandoval, K., Berniell-Lee, G., Calafell, F., Salas, A., Martinez-Fuentes, A., and Comas, D. (2008). Genetic origin, admixture, and asymmetry in maternal and paternal human lineages in Cuba. BMC Evol. Biol. *8*, 213.

280. Henn, B. M., Gignoux, C. R., Jobin, M., Granka, J. M., Macpherson, J. M., Kidd, J. M., Rodríguez-Botigué, L., Ramachandran, S., Hon, L., Brisbin, A., et al. (2011). Hunter-gatherer genomic diversity suggests a southern African origin for modern humans. Proc. Natl. Acad. Sci. *108*, 5154–5162.

281. Gayà-Vidal, M., Moral, P., Saenz-Ruales, N., Gerbault, P., Tonasso, L., Villena, M., Vasquez, R., Bravi, C. M., and Dugoujon, J.-M. (2011). mtDNA and Y-chromosome diversity in Aymaras and Quechuas from Bolivia: Different stories and special genetic traits of the Andean Altiplano populations. Am. J. Phys. Anthropol. *145*, 215–230.

282. Bert, F., Corella, A., Gené, M., Pérez-Pérez, A., and Turbón, D. (2001). Major mitochondrial DNA haplotype heterogeneity in highland and lowland Amerindian populations from Bolivia. Hum. Biol. *73*, 1–16.

283. Dornelles, C. L., Battilana, J., Fagundes, N. J. R., Freitas, L. B., Bonatto, S. L., and Salzano, F. M. (2004). Mitochondrial DNA and Alu insertions in a genetically peculiar population: the Ayoreo Indians of Bolivia and Paraguay. Am. J. Hum. Biol. Off. J. Hum. Biol. Counc. *16*, 479–488.

284. Afonso Costa, H., Carvalho, M., Lopes, V., Balsa, F., Bento, A. M., Serra, A., Andrade, L., Anjos, M. J., Vide, M. C., Pantoja, S., et al. (2010). Mitochondrial DNA sequence analysis of a native Bolivian population. J. Forensic Leg. Med. *17*, 247–253.

285. Bert, F., Corella, A., Gené, M., Pérez-Pérez, A., and Turbón, D. (2004). Mitochondrial DNA diversity in the Llanos de Moxos: Moxo, Movima and Yuracare Amerindian populations from Bolivia lowlands. Ann. Hum. Biol. *31*, 9–28.

286. Corella, A., Bert, F., Pérez-Pérez, A., Gené, M., and Turbón, D. (2007). Mitochondrial DNA diversity of the Amerindian populations living in the Andean Piedmont of Bolivia: Chimane, Moseten, Aymara and Quechua. Ann. Hum. Biol. *34*, 34–55.

287. Sandoval, J. R., Lacerda, D. R., Jota, M. S. A., Salazar-Granara, A., Vieira, P. P. R., Acosta, O., Cuellar, C., Revollo, S., Fujita, R., Santos, F. R., et al. (2013). The Genetic History of Indigenous Populations of the Peruvian and Bolivian Altiplano: The Legacy of the Uros. PLoS ONE *8*, e73006.

288. Watkins, W., Xing, J., Huff, C., Witherspoon, D., Zhang, Y., Perego, U., Woodward, S., and Jorde, L. (2012). Genetic analysis of ancestry, admixture and selection in Bolivian and Totonac populations of the New World. BMC Genet. *13*, 39.

289. López-Parra, A. M., Tirado, M., Baeza, C., Bert, F., Corella, A., Pérez-Pérez, A., Gamba, C., Fernández, E., and Arroyo-Pardo, E. (2008). Genetic structure of the population of Beni department (North Bolivia). Forensic Sci. Int. Genet. Suppl. Ser. *1*, 348–349.

290. Lee, J., Ewis, A. A., Hurles, M. E., Kashiwazaki, H., Shinka, T., and Nakahori, Y. (2007). Y chromosomal STRs haplotypes in two populations from Bolivia. Leg. Med. Tokyo Jpn. *9*, 43–47.

291. Gayà-Vidal, M., Dugoujon, J.-M., Esteban, E., Athanasiadis, G., Rodríguez, A., Villena, M., Vasquez, R., and Moral, P. (2010). Autosomal and X chromosome Alu insertions in Bolivian Aymaras and Quechuas: two languages and one genetic pool. Am. J. Hum. Biol. Off. J. Hum. Biol. Counc. *22*, 154–162.

292. Cann, H. M., Toma, C. de, Cazes, L., Legrand, M.-F., Morel, V., Piouffre, L., Bodmer, J., Bodmer, W. F., Bonne-Tamir, B., Cambon-Thomsen, A., et al. (2002). A Human Genome Diversity Cell Line Panel. Science *296*, 261–262.

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

294. Ward, R. H., Frazier, B. L., Dew-Jager, K., and Pääbo, S. (1991). Extensive mitochondrial diversity within a single Amerindian tribe. Proc. Natl. Acad. Sci. U. S. A. *88*, 8720–8724.

295. Alvarez-Iglesias, V., Jaime, J. C., Carracedo, A., and Salas, A. (2007). Coding region mitochondrial DNA SNPs: targeting East Asian and Native American haplogroups. Forensic Sci. Int. Genet. *1*, 44–55.

296. Cerezo Fernández, M. (2012). Human mitochondrial DNA variability: multidisciplinary applications in the fields of forensic, medical and population genetics. Phd thesis.

297. Librado, P., and Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics *25*, 1451–1452.

298. Excoffier, L., and Lischer, H. E. L. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol. Ecol. Resour. *10*, 564–567.

299. The R Core Team (2014). R: A Language and Environment for Statistical Computing (Vienna, Austria: R Foundation for Statistical Computing). Available at: http://www.R-project.org.

300. Saillard, J., Forster, P., Lynnerup, N., Bandelt, H. J., and Nørby, S. (2000). mtDNA variation among Greenland Eskimos: the edge of the Beringian expansion. Am. J. Hum. Genet. *67*, 718–726.

301. Yang, Z. (2007). PAML 4: Phylogenetic Analysis by Maximum Likelihood. Mol. Biol. Evol. *24*, 1586–1591.

302. Ramos-Luis, E., Blanco-Verea, A., Brión, M., Van Huffel, V., Carracedo, A., and Sánchez-Diz, P. (2009). Phylogeography of French male lineages. Forensic Sci. Int. Genet. Suppl. Ser. *2*, 439–441.

303. Blanco-Verea, A., Jaime, J. C., Brión, M., and Carracedo, A. (2010). Y-chromosome lineages in native South American population. Forensic Sci. Int. Genet. *4*, 187–193.

304. Cárdenas Paredes, J. M. (2014). Análisis de linajes uniparentales en población Boliviana.

305. Whitfield, L. S., Sulston, J. E., and Goodfellow, P. N. (1995). Sequence variation of the human Y chromosome. Nature *378*, 379–380.

306. Brion, M., Quintans, B., Zarrabeitia, M., Gonzalez-Neira, A., Salas, A., Lareu, V., Tyler-Smith, C., and Carracedo, A. (2004). Micro-geographical differentiation in Northern Iberia revealed by Y-chromosomal DNA analysis. Gene *329*, 17–25.

307. Excoffier, L., Laval, G., and Schneider, S. (2007). Arlequin (version 3.0): An integrated software package for population genetics data analysis. Evol. Bioinforma. Online *1*, 47–50.

308. Luis, J. R., Rowold, D. J., Regueiro, M., Caeiro, B., Cinnioglu, C., Roseman, C., Underhill, P. A., Cavalli-Sforza, L. L., and Herrera, R. J. (2004). The Levant versus the Horn of Africa: Evidence for Bidirectional Corridors of Human Migrations. Am. J. Hum. Genet. *74*, 532–544.

309. Simms, T. M., Martinez, E., Herrera, K. J., Wright, M. R., Perez, O. A., Hernandez, M., Ramirez, E. C., McCartney, Q., and Herrera, R. J. (2011). Paternal lineages signal distinct genetic contributions from British Loyalists and continental Africans among different Bahamian islands. Am. J. Phys. Anthropol. *146*, 594–608.

310. Rosa, A., Ornelas, C., Jobling, M. A., Brehm, A., and Villems, R. (2007). Y-chromosomal diversity in the population of Guinea-Bissau: a multiethnic perspective. BMC Evol. Biol. *7*, 124.

311. Gonçalves, R., Spínola, H., and Brehm, A. (2007). Y-chromosome lineages in São Tomé e Príncipe islands: evidence of European influence. Am. J. Hum. Biol. Off. J. Hum. Biol. Counc. *19*, 422–428.

312. Montano, V., Ferri, G., Marcari, V., Batini, C., Anyaele, O., Destro-Bisol, G., and Comas, D. (2011). The Bantu expansion revisited: a new analysis of Y chromosome variation in Central Western Africa. Mol. Ecol. *20*, 2693–2708.

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

314. Filippo, C. de, Barbieri, C., Whitten, M., Mpoloka, S. W., Gunnarsdóttir, E. D., Bostoen, K., Nyambe, T., Beyer, K., Schreiber, H., Knijff, P. de, et al. (2010). Y-chromosomal variation in Sub-Saharan Africa: insights into the history of Niger-Congo groups. Mol. Biol. Evol., msq312.

315. Núñez, C., Geppert, M., Baeta, M., Roewer, L., and Martínez-Jarreta, B. (2012). Y chromosome haplogroup diversity in a Mestizo population of Nicaragua. Forensic Sci. Int. Genet. *6*, e192–195.

316. Mazieres, S., Callegari-Jacques, S. M., Crossetti, S. G., Dugoujon, J.-M., Larrouy, G., Bois, E., Crubezy, E., Hutz, M. H., and Salzano, F. M. (2011). French Guiana Amerindian demographic history as revealed by autosomal and Y-chromosome STRs. Ann. Hum. Biol. *38*, 76–83.

317. Karafet, T. M., Lansing, J. S., Redd, A. J., Reznikova, S., Watkins, J. C., Surata, S. P. K., Arthawiguna, W. A., Mayer, L., Bamshad, M., Jorde, L. B., et al. (2005). Balinese Y-chromosome perspective on the peopling of Indonesia: genetic contributions from pre-neolithic hunter-gatherers, Austronesian farmers, and Indian traders. Hum. Biol. *77*, 93–114.

318. Karafet, T. M., Hallmark, B., Cox, M. P., Sudoyo, H., Downey, S., Lansing, J. S., and Hammer, M. F. (2010). Major East–West Division Underlies Y Chromosome Stratification across Indonesia. Mol. Biol. Evol. *27*, 1833–1844.

319. Zhong, H., Shi, H., Qi, X.-B., Duan, Z.-Y., Tan, P.-P., Jin, L., Su, B., and Ma, R. Z. (2010). Extended Y-chromosome investigation suggests post-Glacial migrations of modern humans into East Asia via the northern route. Mol. Biol. Evol., msq247.

320. Flores, C., Maca-Meyer, N., González, A. M., Oefner, P. J., Shen, P., Pérez, J. A., Rojas, A., Larruga, J. M., and Underhill, P. A. (2004). Reduced genetic structure of the Iberian peninsula revealed by Y-chromosome analysis: implications for population demography. Eur. J. Hum. Genet. EJHG *12*, 855–863.

321. Campbell, K. D. (2007). Geographic Patterns of Haplogro up R1b in the British Isles. J. Genet. Geneal. *3*, 1–13.

322. Kayser, M., Lao, O., Anslinger, K., Augustin, C., Bargel, G., Edelmann, J., Elias, S., Heinrich, M., Henke, J., Henke, L., et al. (2005). Significant genetic differentiation between Poland and Germany follows present-day political borders, as revealed by Y-chromosome analysis. Hum. Genet. *117*, 428–443.

323. Beleza, S., Gusmão, L., Lopes, A., Alves, C., Gomes, I., Giouzeli, M., Calafell, F., Carracedo, A., and Amorim, A. (2006). Micro-phylogeographic and demographic history of Portuguese male lineages. Ann. Hum. Genet. *70*, 181–194.

324. Long, J. C., Williams, R. C., McAuley, J. E., Medis, R., Partel, R., Tregellas, W. M., South, S. F., Rea, A. E., McCormick, S. B., and Iwaniec, U. (1991). Genetic variation in Arizona Mexican Americans: estimation and interpretation of admixture proportions. Am. J. Phys. Anthropol. *84*, 141–157.

325. Jakobsson, M., and Rosenberg, N. A. (2007). CLUMPP: A Cluster Matching and Permutation Program for Dealing with Label Switching and Multimodality in Analysis of Population Structure. Bioinformatics *23*, 1801–1806.

326. Rosenberg, N. A. (2003). distruct: a program for the graphical display of population structure. Mol. Ecol. Notes *4*, 137–138.

327. Earl, D., and von Holdt, B. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv. Genet. Resour. *4*, 359–361.

328. González, J. R., Armengol, L., Solé, X., Guinó, E., Mercader, J. M., Estivill, X., and Moreno, V. (2007). SNPassoc: an R package to perform whole genome association studies. Bioinforma. Oxf. Engl. *23*, 644–645.

329. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., et al. (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. Am. J. Hum. Genet. *81*, 559–575.

330. Pardo-Seco, J., Gómez-Carballa, A., Amigo, J., Martinón-Torres, F., and Salas, A. (2014). A Genome-Wide Study of Modern-Day Tuscans: Revisiting Herodotus's Theory on the Origin of the Etruscans. PLoS ONE *9*, e105920.

331. Reich, D., Patterson, N., Campbell, D., Tandon, A., Mazieres, S., Ray, N., Parra, M. V., Rojas, W., Duque, C., Mesa, N., et al. (2012). Reconstructing Native American population history. Nature *488*, 370–374.

332. Henn, B. M., Botigué, L. R., Gravel, S., Wang, W., Brisbin, A., Byrnes, J. K., Fadhlaoui-Zid, K., Zalloua, P. A., Moreno-Estrada, A., Bertranpetit, J., et al. (2012). Genomic Ancestry of North Africans Supports Back-to-Africa Migrations. PLoS Genet *8*, e1002397.

333. Taboada-Echalar, P., Álvarez-Iglesias, V., Heinz, T., Vidal-Bralo, L., Gómez-Carballa, A., Catelli, L., Pardo-Seco, J., Pastoriza, A., Carracedo, Á., Torres-Balanza, A., et al. (2013). The Genetic Legacy of the Pre-Colonial Period in Contemporary Bolivians. PLoS ONE *8*, e58980.

334. Perego, U. A., Angerhofer, N., Pala, M., Olivieri, A., Lancioni, H., Kashani, B. H., Carossa, V., Ekins, J. E., Gómez-Carballa, A., Huber, G., et al. (2010). The Initial Peopling of the Americas: A Growing Number of Founding Mitochondrial Genomes from Beringia. Genome Res*20*, 1174-1179.

335. Barbieri, C., Heggarty, P., Castrì, L., Luiselli, D., and Pettener, D. (2011). Mitochondrial DNA variability in the Titicaca basin: Matches and mismatches with linguistics and ethnohistory. Am. J. Hum. Biol. Off. J. Hum. Biol. Counc. *23*, 89–99.

336. Bobillo, M. C., Zimmermann, B., Sala, A., Huber, G., Röck, A., Bandelt, H.-J., Corach, D., and Parson, W. (2010). Amerindian mitochondrial DNA haplogroups predominate in the population of Argentina: towards a first nationwide forensic mitochondrial DNA sequence database. Int. J. Legal Med. *124*, 263–268.

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

338. Li, H., Cai, X., Winograd-Cort, E. R., Wen, B., Cheng, X., Qin, Z., Liu, W., Liu, Y., Pan, S., Qian, J., et al. (2007). Mitochondrial DNA diversity and population differentiation in southern East Asia. Am. J. Phys. Anthropol. *134*, 481–488.

339. Behar, D. M., van Oven, M., Rosset, S., Metspalu, M., Loogväli, E.-L., Silva, N. M., Kivisild, T., Torroni, A., and Villems, R. (2012). A "Copernican" Reassessment of the Human Mitochondrial DNA Tree from its Root. Am. J. Hum. Genet. *90*, 675–684.

340. Monson, K. L., Miller, K. W., Wilson, M. R., DiZinno, J. A., and Budowle, B. (2002). The mtDNA Population Database: An Integrated Software and Database Resource for Forensic Comparison. Forensic Sci. Commun. *4*.

341. Cabana, G. S., Merriwether, D. A., Hunley, K., and Demarchi, D. A. (2006). Is the genetic structure of Gran Chaco populations unique? Interregional perspectives on native South American mitochondrial DNA variation. Am. J. Phys. Anthropol. *131*, 108–119.

342. Behar, D. M., Rosset, S., Blue-Smith, J., Balanovsky, O., Tzur, S., Comas, D., Mitchell, R. J., Quintana-Murci, L., Tyler-Smith, C., Wells, R. S., et al. (2007). The Genographic Project Public Participation Mitochondrial DNA Database. PLoS Genet *3*, e104.

343. Fuselli, S., Tarazona-Santos, E., Dupanloup, I., Soto, A., Luiselli, D., and Pettener, D. (2003). Mitochondrial DNA diversity in South America and the genetic history of Andean highlanders. Mol. Biol. Evol. *20*, 1682–1691.

344. Murray-McIntosh, R. P., Scrimshaw, B. J., Hatfield, P. J., and Penny, D. (1998). Testing migration patterns and estimating founding population size in Polynesia by using human mtDNA sequences. Proc. Natl. Acad. Sci. *95*, 9047–9052.

345. Redd, A. J., Takezaki, N., Sherry, S. T., McGarvey, S. T., Sofro, A. S., and Stoneking, M. (1995). Evolutionary history of the COII/tRNALys intergenic 9 base pair deletion in human mitochondrial DNAs from the Pacific. Mol. Biol. Evol. *12*, 604–615.

346. Fagundes, N. J. R., Kanitz, R., Eckert, R., Valls, A. C. S., Bogo, M. R., Salzano, F. M., Smith, D. G., Silva, W. A., Jr, Zago, M. A., Ribeiro-dos-Santos, A. K., et al. (2008). Mitochondrial population genomics supports a single pre-Clovis origin with a coastal route for the peopling of the Americas. Am. J. Hum. Genet. *82*, 583–592.

347. Gómez-Carballa, A., Ignacio-Veiga, A., Álvarez-Iglesias, V., Pastoriza-Mourelle, A., Ruíz, Y., Pineda, L., Carracedo, Á., and Salas, A. (2011). A melting pot of multicontinental mtDNA lineages in admixed Venezuelans. Am. J. Phys. Anthropol. *147*, 78–87.

348. Cardoso, S., Palencia-Madrid, L., Valverde, L., Alfonso-Sánchez, M. A., Gómez-Pérez, L., Alfaro, E., Bravi, C. M., Dipierri, J. E., Peña, J. A., and de Pancorbo, M. M. (2013). Mitochondrial DNA control region data reveal high prevalence of Native American lineages in Jujuy province, NW Argentina. Forensic Sci. Int. Genet. *7*, e52–55.

349. Li, J. Z., Absher, D. M., Tang, H., Southwick, A. M., Casto, A. M., Ramachandran, S., Cann, H. M., Barsh, G. S., Feldman, M., Cavalli-Sforza, L. L., et al. (2008). Worldwide human relationships inferred from genome-wide patterns of variation. Science *319*, 1100–1104.

350. Cárdenas, J. M., Heinz, T., Pardo-Seco, J., Álvarez-Iglesias, V., Taboada-Echalar, P., Sánchez-Diz, P., Carracedo, Á., and Salas, A. The multiethnic ancestry of Bolivians as revealed by the analysis of Y-chromosome markers. Forensic Sci. Int. Genet*14*, 210-218.

351. Tirado, M., López-Parra, A. M., Baeza, C., Bert, F., Corella, A., Pérez-Pérez, A., Turbón, D., and Arroyo-Pardo, E. (2009). Y-chromosome haplotypes defined by 17 STRs included in AmpFlSTR Yfiler PCR Amplification Kit in a multi ethnical population from El Beni Department (North Bolivia). Leg. Med. *11*, 101–103.

352. Romero, R. E., Briceño, I., Lizarazo, R. del P., Willuweit, S., Roewer, L., and Gómez, A. (2008). A Colombian Caribbean population study of 16 Y-chromosome STR loci. Forensic Sci. Int. Genet. *2*, e5–8.

353. Toscanini, U., Gusmão, L., Berardi, G., Amorim, A., Carracedo, A., Salas, A., and Raimondi, E. (2006). Genetic variability of 17 Y chromosome STRs in two Native American populations from Argentina. Int. Congr. Ser. *1288*, 154–155.

354. Roby, R. K., Gonzalez, S. D., Phillips, N. R., Planz, J. V., Thomas, J. L., Astudillo, J. A. P., Ge, J., Morales, E. A., Eisenberg, A. J., Chakraborty, R., et al. (2009). Autosomal STR allele frequencies and Y-STR and mtDNA haplotypes in Chilean sample populations. Forensic Sci. Int. Genet. Suppl. Ser. *2*, 532–533.

355. Cainé, L. M., de Pancorbo, M. M., and Pinheiro, F. (2010). Y-chromosomal STR haplotype diversity in males from Santa Catarina, Brazil. J. Forensic Leg. Med. *17*, 92–95.

356. Schwengber, S. P., Kommers, T., Matte, C. H. F., Raimann, P. E., Carvalho, B. A., Leite, F. P. N., Medeiros, M. A., Souza, L. F., Castro, C. S., Chassot, F. G. C., et al. (2009). Population data of 17 Y-STR loci from Rio Grande do Sul state (South Brazil). Forensic Sci. Int. Genet. *4*, e31–33.

357. González-Andrade, F., Roewer, L., Willuweit, S., Sánchez, D., and Martínez-Jarreta, B. (2009). Y-STR variation among ethnic groups from Ecuador: Mestizos, Kichwas, Afro-Ecuadorians and Waoranis. Forensic Sci. Int. Genet. *3*, e83–91.

358. Iannacone, G. C., Tito, R. Y., Lopez, P. W., Medina, M. E., and Lizarraga, B. (2005). Y-chromosomal haplotypes for the PowerPlex Y for twelve STRs in a Peruvian population sample. J. Forensic Sci. *50*, 239–242.

359. Borjas, L., Bernal, L. P., Chiurillo, M. A., Tovar, F., Zabala, W., Lander, N., and Ramírez, J. L. (2008). Usefulness of 12 Y-STRs for forensic genetics evaluation in two populations from Venezuela. Leg. Med. Tokyo Jpn. *10*, 107–112.

360. Marino, M., and Furfuro, S. (2010). Genetic population data of 12 Y-chromosome STRs loci in Mendoza population (Argentina). Forensic Sci. Int. Genet. *4*, e89–93.

361. Heinz, T., Alvarez-Iglesias, V., Pardo-Seco, J., Taboada-Echalar, P., Gómez-Carballa, A., Torres-Balanza, A., Rocabado, O., Carracedo, A., Vullo, C., and Salas, A. (2013). Ancestry analysis reveals a predominant Native American component with moderate European admixture in Bolivians. Forensic Sci. Int. Genet. *7*, 537–542.

362. De Azevedo, D. A., da Silva, L. A. F., Gusmão, L., and de Carvalho, E. F. (2009). Analysis of Y chromosome SNPs in Alagoas, Northeastern Brazil. Forensic Sci. Int. Genet. Suppl. Ser. *2*, 421–422.

363. Acosta, M. A., Blanco-Verea, A., Lareu, M. V., Brion, M., and Carracedo, A. (2009). The genetic male component of two South-Western Colombian populations. Forensic Sci. Int. Genet. *3*, e59–e61.

364. Lovo-Gómez, J., Blanco-Verea, A., Lareu, M. V., Brión, M., and Carracedo, A. (2007). The genetic male legacy from El Salvador. Forensic Sci. Int. *171*, 198–203.

365. Salas, A., Jaime, J. C., Alvarez-Iglesias, V., and Carracedo, A. (2008). Gender bias in the multiethnic genetic composition of central Argentina. J. Hum. Genet. *53*, 662–674.

366. Pagani, L., Kivisild, T., Tarekegn, A., Ekong, R., Plaster, C., Gallego Romero, I., Ayub, Q., Mehdi, S. Q., Thomas, M. G., Luiselli, D., et al. (2012). Ethiopian Genetic Diversity Reveals Linguistic Stratification and Complex Influences on the Ethiopian Gene Pool. Am. J. Hum. Genet. *91*, 83–96.

367. Bonilla, C., Bertoni, B., González, S., Cardoso, H., Brum-Zorrilla, N., and Sans, M. (2004). Substantial Native American female contribution to the population of Tacuarembó, Uruguay, reveals past episodes of sex-biased gene flow. Am. J. Hum. Biol. Off. J. Hum. Biol. Counc. *16*, 289–297.

368. Batista dos Santos, S. E., Rodrigues, J. D., Ribeiro-dos-Santos, A. K., and Zago, M. A. (1999). Differential contribution of indigenous men and women to the formation of an urban population in the Amazon region as revealed by mtDNA and Y-DNA. Am. J. Phys. Anthropol. *109*, 175–180.

369. Gonçalves, V. F., Carvalho, C. M. B., Bortolini, M. C., Bydlowski, S. P., and Pena, S. D. J. (2008). The phylogeography of African Brazilians. Hum. Hered. *65*, 23–32.

370. Benn Torres, J., Kittles, R. A., and Stone, A. C. (2007). Mitochondrial and Y chromosome diversity in the English-speaking Caribbean. Ann. Hum. Genet. *71*, 782–790.

371. Arnason, U., Gullberg, A., Janke, A., and Kullberg, M. (2007). Mitogenomic analyses of caniform relationships. Mol. Phylogenet. Evol. *45*, 863–874.

11 Publications

The results shown in the present thesis are part of the following publications:

Taboada-Echalar, P, Álvarez-Iglesias, V, **Heinz, T**, Vidal-Bralo, L, Gómez-Carballa, A, Catelli, L, Pardo-Seco, J, Pastoriza, A, Carracedo, Á, Torres-Balanza, A, Rocabado, O, Vullo, C, Salas, A (2013). "The Genetic Legacy of the Pre-Colonial Period in Contemporary Bolivians." *PLoS ONE* 8 (3): e58980. doi:10.1371/journal.pone.0058980.

Heinz, T, Álvarez-Iglesias, V, Pardo-Seco, J, Taboada-Echalar, P, Gómez-Carballa, A, Torres-Balanza, A, Rocabado, O, Carracedo, Á, Vullo, C, Salas, A (2013). "Ancestry AnalysisReveals a Predominant Native American Component with Moderate European Admixture inBolivians".*Forensic Science International. Genetics* 7 (5): 537–42. doi:10.1016/j.fsigen.2013.05.012.

Cárdenas, JM, **Heinz, T**, Pardo-Seco, J, Álvarez-Iglesias, V, Taboada-Echalar, P, Sánchez-Diz, P, Carracedo, Á, Salas, A (2014). "The multiethnic ancestry of Bolivians as revealed by the analysis of Y-chromosome markers".*Forensic Science International Genetics* 14: 210-218. http://dx.doi.org/10.1016/j.fsigen.2014.10.023

Heinz, T, Cárdenas, JM, Álvarez-Iglesias, V, Pardo-Seco, J, Gómez-Carballa, A, Santos, C, Taboada-Echalar, P, Salas, A (2015). "The genomic legacy of the Transatlantic Slave Trade in the Yungas valley of Bolivia". Submitted to *PLoS ONE*

Pardo-Seco, J, **Heinz, T**, Taboada-Echalar, P, Martinón-Torres, F, Salas, A (2015). "Mapping the genomic mosaic of two 'Afro-Bolivians' from the Yungas". Submitted to *Heredity*

Heinz, T, Pala, M, Richards, M, Salas, A (2015). "An update of the Mitochondrial Phylogenetic Tree". In preparation.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

12 Appendix

12.1 Samples for mitochondrial phylogenetic analyses

SampleID	Haplogroup	Geographic Area	Country	Reference
EU273488	L1c3b1a	Africa	unknown	[161]
EU273493	L1c3c	Africa	unknown	[161]
JQ044815	L3d1a1	Africa Central	Burkina Faso	[152]
JQ044814	L3d1b3a	Africa Central	Burkina Faso	[152]
JQ044824	L3d1c1	Africa Central	Burkina Faso	[152]
JQ044855	L1c3a1b	Africa Central	Burkina Faso	[152]
JQ045051	L3d1b3	Africa Central	Burkina Faso	[152]
HM771160	L0a2a1	Africa Central	Central African Republic	[151]
HM771161	L0a2a1	Africa Central	Central African Republic	[151]
EU092878	L0a1b	Africa Central	Chad	[134]
EU092889	L0a1b	Africa Central	Chad	[134]
EU092900	L0a3	Africa Central	Chad	[134]
EU092906	L0a4	Africa Central	Chad	[134]
EU092899	L3d1b	Africa Central	Chad	[134]
AF347014	L3d1a1a	Africa Central	Chad/Nigeria	[135]
AF346998	L0a2b	Africa Central	Congo	[135]
AF346999	L0a2b	Africa Central	Congo	[135]
AM711903	L0a2b	Africa Central	Congo	[371]
EU597537	L0a2b	Africa Central	Congo	[146]
HM771189	L0a2b	Africa Central	Congo	[151]
HM771190	L0a2b	Africa Central	Congo	[151]
HM771199	L0a2b	Africa Central	Congo	[151]
HM771202	L0a2b	Africa Central	Congo	[151]
HM771188	L0a2b1	Africa Central	Congo	[151]
HM771200	L0a2b1	Africa Central	Congo	[151]
HM771201	L0a2b1	Africa Central	Congo	[151]
HG03343	L1c3a	Africa Central	Nigeria	1000 Genomes GATK
HG03366	L3d1b2	Africa Central	Nigeria	1000 Genomes GATK
HG03518	L3d1c1	Africa Central	Nigeria	1000 Genomes GATK
HG03100	L3d1d	Africa Central	Nigeria	1000 Genomes GATK
JQ045099	L1c3a1b	Africa Central	Nigeria	[152]
NA18915	L1c3b1b	Africa Central	Nigeria	1000 Genomes GATK
NA19256	L1c3b2	Africa Central	Nigeria	1000 Genomes GATK
NA18867	L3d1a1	Africa Central	Nigeria	1000 Genomes GATK

Appendix

SampleID	Haplogroup	Geographic Area	Country	Reference
HM771219	L1c3a1b	Africa Central	unknown	[151]
HM771167	L1c3b1a	Africa Central	unknown	[151]
HM771227	L3d1a1a	Africa Central	unknown	[151]
HM771228	L3d1a2	Africa Central	unknown	[151]
EU092945	L0a1c1	Africa East	Ethiopia	[134]
EU092950	L0a1d	Africa East	Ethiopia	[134]
NA19440	L0a1+16293	Africa East	Kenya	1000 Genomes GATK
NA19382	L0a1b1a1	Africa East	Kenya	[149]
NA19448	L0a1b1a1	Africa East	Kenya	1000 Genomes GATK
NA19449	L0a1b1a1	Africa East	Kenya	1000 Genomes GATK
NA19466	L0a1b1a1	Africa East	Kenya	1000 Genomes GATK
NA19467	L0a1c1	Africa East	Kenya	1000 Genomes GATK
NA19312	L0a2a2a	Africa East	Kenya	1000 Genomes GATK
NA19328	L0a2a2a	Africa East	Kenya	1000 Genomes GATK
NA19390	L1c3b1a	Africa East	Kenya	1000 Genomes GATK
NA19316	L3d1a	Africa East	Kenya	1000 Genomes GATK
NA19030	L3d1a1a1	Africa East	Kenya	1000 Genomes GATK
EU092909	L0a1b1a1	Africa East	Kenya	[134]
EU092911	L0a2a1a2	Africa East	Kenya	[134]
EU092913	L0a2d	Africa East	Kenya	[134]
EU092908	L3d1a1a	Africa East	Kenya	[134]
JN655808	L3d1a1	Africa East	Somalia	[166]
JN655800	L3d1a1a	Africa East	Somalia	[166]
JN655823	L3d1b1b	Africa East	Sudan	[166]
AF381992	L1c3a	Africa North	Mauritania	[153]
EU092811	L1c3b1a	Africa North	Morocco	[134]
AF381988	L0a1b1	Africa North	Morocco	[153]
EU092830	L3d1c1	Africa North	Tunisia	[134]
JF509360	L1c3b1	Africa North	Western Sahara	FamilyTree DNA
EU092866	L1c3a	Africa South	South Africa	[134]
EU092858	L0a1b1a1	Africa South	South Africa	[134]
EU092869	L0a1b1a1	Africa South	South Africa	[134]
EU092871	L0a1b1a1	Africa South	South Africa	[134]
EU092868	L0a2a1a2	Africa South	South Africa	[134]
EU092861	L0a2a1b	Africa South	South Africa	[134]
EU092876	L3d1a1	Africa South	South Africa	[134]
AY195782	L3d1d	Africa South	South Africa	[28]
EU092703	L1c3a	Africa Southeast	Mozambique	[134]
EU092702	L3d1a1a	Africa Southeast	Mozambique	[134]
EU092689	L1c3b1b	Africa Southeast	Mozambique	[134]
EU092688	L0a1b1a1	Africa Southeast	Mozambique	[134]

SampleID	Haplogroup	Geographic Area	Country	Reference
EU092701	L0a2a2a	Africa Southeast	Mozambique	[134]
JX303784	L0a2a1b	Africa Southeast	Zambia	[150]
JX303830	L0a2a1b	Africa Southeast	Zambia	[150]
JX303763	L0a2a2a	Africa Southeast	Zambia	[150]
JX303778	L0a2a2a	Africa Southeast	Zambia	[150]
JX303904	L0a2a2a	Africa Southeast	Zambia	[150]
JX303824	L1c3a	Africa Southeast	Zambia	[150]
JX303871	L1c3a1b	Africa Southeast	Zambia	[150]
JX303785	L1c3b1a	Africa Southeast	Zambia	[150]
JX303821	L3d1b3a	Africa Southeast	Zambia	[150]
HG02983	L1c3a1a	Africa West	Gambia	1000 Genomes GATK
HG02583	L3d1a1	Africa West	Gambia	1000 Genomes GATK
HG02798	L3d1a1	Africa West	Gambia	1000 Genomes GATK
HG03241	L3d1a1	Africa West	Gambia	1000 Genomes GATK
HG02628	L3d1b1	Africa West	Gambia	1000 Genomes GATK
HG02702	L3d1b1	Africa West	Gambia	1000 Genomes GATK
HG02768	L3d1b1	Africa West	Gambia	1000 Genomes GATK
HG03040	L3d1b1	Africa West	Gambia	1000 Genomes GATK
HG02586	L3d1b1b	Africa West	Gambia	1000 Genomes GATK
HG02610	L3d1c	Africa West	Gambia	1000 Genomes GATK
EU092718	L1c3a1a	Africa West	Guinea-Bissau	[134]
EU092728	L3d1b1	Africa West	Guinea-Bissau	[134]
JQ045112	L3d1b	Africa West	Senegal	[152]
JQ045117	L3d1d	Africa West	Senegal	[152]
HG03559	L3d1b1	Africa West	Sierra Leone	1000 Genomes GATK
HG03567	L3d1b3a	Africa West	Sierra Leone	1000 Genomes GATK
DQ304900	L0a1b1	America North	USA	[155]
DQ304899	L0a1b1a	America North	USA	[155]
NA19703	L0a1b1a	America North	USA	1000 Genomes GATK
DQ304897	L0a1b2	America North	USA	[155]
DQ304898	L0a1b2	America North	USA	[155]
EU092963	L0a1b2	America North	USA	[134]
NA19713	L0a2a2a	America North	USA	1000 Genomes GATK
NA19904	L0a2a2a	America North	USA	1000 Genomes GATK
NA19985	L0a2a2a	America North	USA	1000 Genomes GATK
NA20359	L1c3a	America North	USA	1000 Genomes GATK
NA20362	L1c3a	America North	USA	1000 Genomes GATK
NA20363	L1c3a	America North	USA	[149]
EU092956	L1c3a1b	America North	USA	[134]
NA19914	L1c3a1b	America North	USA	1000 Genomes GATK
NA20294	L3d1b3a	America North	USA	1000 Genomes GATK
NA20339	L3d1c1	America North	USA	1000 Genomes GATK

Appendix

SampleID	Haplogroup	Geographic Area	Country	Reference
HG01260	L0a1b1a	America South	Colombia	1000 Genomes GATK
HG01389	L1c3a1b	America South	Colombia	1000 Genomes GATK
HG01378	L1c3b1	America South	Colombia	1000 Genomes GATK
EU092932	L3d1a1a1	Asia	Pakistan	[134]
HG01986	L1c3a	Caribbean	Barbados	1000 Genomes GATK
HG01990	L1c3a1a	Caribbean	Barbados	1000 Genomes GATK
HG02318	L3d1b2	Caribbean	Barbados	1000 Genomes GATK
HG02484	L3d1b2	Caribbean	Barbados	1000 Genomes GATK
HG02095	L3d1b3	Caribbean	Barbados	1000 Genomes GATK
HG02339	L3d1c	Caribbean	Barbados	1000 Genomes GATK
DQ341058	L0a2a2a	Caribbean	Dominican Republic	[156]
DQ341072	L3d1b1	Caribbean	Dominican Republic	[156]
HG01049	L0a1b1a1	Caribbean	Puerto Rico	1000 Genomes GATK
HG01080	L1c3b1a	Caribbean	Puerto Rico	1000 Genomes GATK
HG00739	L3d1b1	Caribbean	Puerto Rico	1000 Genomes GATK
EU200759	L3d1b1b	Europe East	Poland	[160]
HG01761	L3d1b1	Europe West	Spain	1000 Genomes GATK
EU935434	L0a1b1a	Near East	Egypt	[154]
EU935437	L0a1b1a	Near East	Egypt	[154]
EU935464	L0a1b1a	Near East	Egypt	[154]
EU935467	L0a1b1a	Near East	Egypt	[154]
EU935458	L1c3a1a	Near East	Egypt	[154]
EU935462	L1c3a1a	Near East	Egypt	[154]
EU092762	L3d1c	Near East	Egypt	[134]
EU092760	L0a1c1	Near East	Iran	[134]
EU092670	L0a1d	Near East	Israel	[134]
EU092787	L0a2a2a	Near East	Oman	[134]
EU092746	L0a1b2a	Near East	Saudi-Arabia	[134]
EU092745	L0a2a2a	Near East	Saudi-Arabia	[134]
EU092801	L0a1d	Near East	Yemen	[134]
EU092809	L0a1d	Near East	Yemen	[134]
EU092810	L0a1d	Near East	Yemen	[134]
EU092925	L0a2a2a	Near East	Yemen	[134]
EU092796	L3d1a	Near East	Yemen	[134]
EU092797	L3d1a1a	Near East	Yemen	[134]

12.2 Population data set for whole genome SNP typing

Ethnic group	Region	Country	Sample size	Reference
Afar	EA	Ethiopia	12	[366]
Amhara	EA	Ethiopia	26	[366]
Anuak	EA	Ethiopia	23	[366]
Ari Blacksmith	EA	Ethiopia	17	[366]
Ari Cultivator	EA	Ethiopia	24	[366]
Ethiopian Somali	EA	Ethiopia	17	[366]
Gumuz	EA	Ethiopia	19	[366]
Luhya	EA	Kenya	11	[349]
LWK	EA	Kenya	100	[148]
Maasai	EA	Kenya	30	Hapmap
Oromo	EA	Ethiopia	21	[366]
Somali	EA	Somalia	23	[366]
Sudanese	EA	South Sudan	24	[366]
Tygray	EA	Ethiopia	21	[366]
Wolayta	EA	Ethiopia	8	[366]
Algerians	NA	Algeria	19	[280]
Egyptians	NA	Egypt	19	[280]
Libyans	NA	Libya	17	[280]
North Moroccans	NA	Morocco	18	[280]
Mozabite Berber	NA	Algeria	29	[349]
Sahrawi	NA	West Sahara	18	[280]
South Moroccans	NA	Morocco	16	[280]
Tunisian Berber	NA	Tunisia	18	[280]
Coloured'	SA	Colesberg; SA	20	[20]
Coloured'	SA	Wellington, SA	20	[20]
Gui and Gana	SA	Botswana	15	[20]
Ju'hoansi	SA	Namibia	18	[20]
Juu San	SA	Namibia	12	[137]
Karrettjie	SA	SA	20	[20]
Khomani	SA	SA	39	[20]
Khwe	SA	Angola	17	[20]
Nama	SA	Namibia	20	[20]
San	SA	Namibia	5	[349]
Khomani Bushman	SA	SA	31	[280]
Bantu-speakers	SA	SA	20	[20]
Herero	SA	Bantu	12	[20]
Xhosa	SA	SA	5	[65]
Xhosa	SA	Migrants	8	[349]
Xun	SA	Angola	19	[20]
Hadza	SEA	Tanzania	17	[280]
Sandawe	SEA	Tanzania	28	[280]
Bamoun	WCA	Cameroon	18	[65]
Biaka Pygmy	WCA	Democratic Republic of Congo	22	[349]
Brong	WCA	Ghana	8	[65]
Bulala	WCA	Chad	15	[65]
Fang	WCA	Cameroon	15	[65]

Appendix

Ethnic group	Region	Country	Sample size	Reference
Hausa	WCA	Cameroon	12	[65]
Igbo	WCA	Nigeria	15	[349]
Kaba	WCA	Chad	17	[65]
Kongo	WCA	Democratic Republic of Congo	9	[65]
Mada	WCA	Cameroon	12	[65]
Mandenka	WCA	Senegal	22	[349]
Mbororo Fulani	WCA	Nigeria	13	[65]
Mbuti Pygmy	WCA	Democratic Republic of Congo	13	[349]
Yoruba	WCA	Nigeria	21	[349]
YRI	WCA	Ibadan, Nigeria	88	[148]
Yoruba	WCA	Ibadan, Nigeria	5	Hapmap
MXL	Central America	Mexico	66	[148]
PUR	Central America	Puerto Rico	55	[148]
CEU	North America	Utah Residents (CEPH)	95	[148]
Tocaña	South America	Bolivia	2	
CLM	South America	Colombia	60	[148]
CHB	East Asia	China	97	[148]
CHD	East Asia	China	100	[148]
CHS	East Asia	China		
JPT	East Asia	Japan	89	[148]
GBR	-	England and Scotland	56	[148]
Iberian in Spain	-	Spain	14	[148]
TSI	-	Italy	98	[148]

Samples from Hapmap: http://hapmap.ncbi.nlm.nih.gov, WCA=West-Central Africa, SEA=Southeast Africa, EA=East Africa, SA=South Africa, NA=North Africa.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

12.3 List of Tables

Table 1: Polymorphisms of haplogroups.

Table 2: Age estimates for L0 sub-haplogroups.

Table 3: Age estimates for L1 sub-haplogroups.

Table 4: Age estimates for L2 sub-haplogroups.

Table 5: Age estimates for L3 sub-haplogroups.

Table 6: Summary of Bolivian population samples for autosomal genotyping (indel, SNP), mtDNA and Y chromosome analyses.

Table 7: PCR primers for whole mtDNA Sequencing (developed by Torroni and colleagues [165]).

Table 8: PCR/SEQ primers for whole mtDNA Sequencing (developed by Kivisild and colleagues [27]).

Table 9: Sequencing Primer for Whole mtDNA Sequencing.

Table 10: Primers for PCR and sequencing the control region.

Table 11: AIM-INDEL multiplex developed by Pereira and colleagues[269].

Table 12: PCR primer sequences for AIM-Indelplex developed by Pereira and colleagues[269].

Table 13: Diversity indices in Bolivian mtDNAs and main American and African regions.

Table 14: Shared haplotypes between Bolivia and other American regions.

Table 15:AMOVA computed based on haplotype pairwise differences of Bolivian populations.

Table 16: Mitochondiral DNA diversity of the Yungas including Tocaña.

Table 17: Admixture proportions, P_0, P_1, and P_2 (and 95% C.I.) from sub-continental regions in Africa in mtDNA haplotypes in Bolivia.

Table 18: Y-STR Molecular diversity in Bolivian population samples.

Table 19: AMOVA analysis in Bolivian and other South American populations.

Table 20: Y-SNP Molecular diversity in Bolivian population samples.

Table 21: Admixture proportions among eco-geographical regions in Bolivia.

Table 22: Admixture analysis for K=5 and using the populations from The 1000 Genomes Project as population reference datasets.

Table 23: Admixture analysis considering different African datasets, CEU, and CHB and for the optimal $K = 12$.

Table 24: Ancestry estimates for mtDNA, Y chromosome, and AIMs in the Bolivian population.

Table 25: Gender-bias between Y-Chromosome, mtDNA and AIMs of 6 individuals from La Paz (Andeans).

Table 26: Ancestry estimates for mtDNA, Y chromosome, and AIMs in the Yungas.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

12.4 List of Figures

Figure 1: Total number of enslaved African people that embarked in Africa between 1514 and 1866.

Figure 2: Major Slaving Regions of West Africa.

Figure 3: Number of enslaved African people that embarked in West Africa between 1514 and 1866.

Figure 4: Number of enslaved African people that embarked in Central Africa between 1514 and 1866.

Figure 5: Number of enslaved African people that embarked in West-Central Africa between 1514 and 1866.

Figure 6: Number of enslaved African people that embarked in Southeast and other parts of Africa between 1514 and 1866.

Figure 7: Number of African enslaved people that disembarked in the Americas (and Africa) from 1514 to 1866.

Figure 8: Overview of the Slave Trade out of Africa, 1500-1900.

Figure 9: Enslaved Africans in Hold of Slave Ship.

Figure 10: Sugar plantation, San José de la Angosta, Cuba, 1857.

Figure 11: 23 pairs of chromosomes of the human nuclear genome.

Figure 12: Matrilineal inheritance of mtDNA.

Figure 13: Paternal inheritance of the Y chromosome.

Figure 14: Global human mtDNA migrations.

Figure 15: Inferring the genetic ancestry based on the Y chromosome.

Figure 16: Grandparental genome inheritance.

Figure 17: Different haplotypes of one DNA fragment.

Figure 18: Ethnicity of West Africa region presented on the Smithsonian Institute's symposium: The African Diaspora.

Figure 19: Rice planting in South Carolina during the early 1890s.

Figure 20: Sugar cane harvest, Jamaica, 1820s.

Figure 21: Sugar mill, Brazil, 1816.

Figure 22: Gold production Colombia, 1826.

Figure 23: Mitochondrion.

Figure 24: The mitochondrial Genome.

Figure 25: Hotspots in the mitochondrial DNA genome.

Figure 26: Mitochondrial DNA haplogroup L2a1c4a1 and its associated haplotypes.

Figure 27: Simplified global mtDNA lineages distribution.

Figure 28: Overview of the global mtDNA phylogenetic tree.

Figure 29: Clade of haplogroup L0.

Figure 30: Spatial distribution of L0d and L0k in Africa. Haplogroup L0d (A) and Haplogroup L0k (B) are shown with a different scale.

Figure 31: Clade of haplogroup L1.

Figure 32: Clade of haplogroup L2.

Figure 33: Clade of haplogroup L3.

Figure 34: Clade of haplogroup L4.

Figure 35: Clade of haplogroup L5.

Figure 36: Clade of haplogroup L6.

Figure 37: Three major migratory events based on 16 maternal founding lineages.

Figure 38: Spatial frequency distribution of haplogroup H.

Figure 39: The Y chromosome.

Figure 40: Distribution of Euchromatin and Heterochromatin in the MSY region of the Y chromosome.

Figure 41: The phylogenetic tree of binary chromosomal haplogroups A-R (YCC2003 Tree).

Figure 42: Global distribution of binary haplogroups.

Figure 43: Updated summary of the Y chromosome phylogenetic tree.

Figure 44: Distribution of haplogroup C and D.

Figure 45: Y chromosome haplogroup distribution in Haiti, Jamaica and the Bahamas.

Figure 46: Distribution of haplogroup I and J.

Figure 47: Distribution of haplogroup N.

Figure 48: Distribution of haplogroup O.

Figure 49: Distribution of haplogroup Q in North America and Mesoamerica.

Figure 50: Short Tandem Repeats (STRs).

Figure 51: Illustration of amplification of the Y-STR marker DYS385.

Figure 52: Y STR Positions along the Y chromosome.

Figure 53: Schematic illustration of the Y-STRs from the AmpFlSTR® YFiler™ kit.

Figure 54: Migration routes based on paternal lineages in North Indians.

Figure 55: GWAS study of Primary Biliary Cirrhosis in a Japanese population.

Figure 56: Principal Component Analysis of the Europen population.

Figure 57: Electropherogram for the sample HGDP-CEPH 0452 with the 46 AIM-INDEL panel developed by Pereira and colleagues[269].

Figure 58: Admixture in latin American populations.

Figure 59: Map of Bolivia.

Figure 60: Working in silver mines, Peru, 1789.

Figure 61: Analyzing whole mtDNA sequence with SeqScape v2.1.

Figure 62: Phylogenetic tree of Y chromosome haplogroups.

Figure 63: Length size separation of AIM-Indelplex fragments.

Figure 64: Number of alleles at different loci within departments.

Figure 65: Pairwise F_{ST} values between departments and regions.

Figure 66: Expected (virtual) heterozygosity by departments.

Figure 67: Expected (virtual) heterozygosity by main ecological regions.

Figure 68: Map of Bolivia showing the location of the samples collected in the present study.

Figure 69: Maximum parsimony tree of the main branches characterizing haplogroups A2 and B2, indicating the new branches generated in the present thesis, namely, A2ah and B2o.

Figure 70: Maximum parsimony tree of haplogroup B2b.

Figure 71: Proportions of continental ancestries on the mtDNA compared with autosomal AIMs.

Figure 72: Map showing the location and sample size of the African mitogenomes used for phylogeographic analysis.

Figura 73: Phylogenetic tree L0a.

Figure 74: Phylogenetic tree L1c3.

Figure 75: Phylogenetic tree L3d1.

Figure 76: Y chromosome haplogroup frequencies among the Bolivian population.

Figure 77: Principal Component Analysis (PCA) of Y-STRs and Y-SNPs.

Figure 78: Distribution of the Y chromosome profile of Tocaña291 belonging to haplogroup E*.

Figure 79: Distribution of the Y chromosome profile of Tocaña295 belonging to haplogroup BR.

Figure 80: Distribution of the Y chromosome profile of Tocaña294 belonging to haplogroup E1b1b1*.

Figure 81: Average continental ancestry of the Bolivian population based on 46 AIMs.

Figure 82: Structure analysis of the Bolivian population among departments.

Figure 83: Structure analysis of the Bolivian population among regions.

Figure 84: Structure analysis of the Bolivian population among rural and urban regions.

Figure 85: PCA of Bolivian profiles with 46 AIM indel markers.

Figure 86: Ancestry values obtained for La Paz and Chuquisaca by way of simulations.

Figure 87: Individual ancestry in the Nor and Sud Yungas.

Figure 88: PCA of the Yungas population.

Figure 89: Average IBS values between the two 'Afro-Bolivians' and samples from 1000G (A) and a large dataset of Africans (B).

Figure 90: Principal Component Analysis of Tocañas against 1000 Genomes population.

Figure 91: PCA analysis of the two 'Afro-Bolivians' in a large dataset of African groups.

Figure 92: PCA analyses of the two 'Afro-Bolivians' with West African (A) and East African (B) samples.

Figure 93: Genomic architecture and main continental ancestries of the two 'Afro-Bolivians' computed with PCAdmix.

Figure 94: Individual ancestry of the two 'Afro-Bolivians'.

Figure 95: Admixture analysis of the two 'Afro-Bolivians' with an unmasked (A) and masked (B) Native American dataset.

Figure 96: Ancestry proportions based on AIM-Indels, mtDNA, and Y chromosome of only males in Bolivia including Yungas (here as Sub-andeans).

Figure 97: Ancestry proportions based on AIM-Indels, mtDNA, and Y chromosome in the Yungas compared to Bolivia.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA