THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par *l'Université Toulouse III - Paul Sabatier*

Discipline ou spécialité : *Anthropologie génétique*

Présentée et soutenue par *FORTES LIMA César Augusto*

Le 17 Décembre *2015*

Titre: *Tracing the genetic origin of African descendants from South America*

JURY

QUINTANA-MURCI Lluis, *Directeur de recherche CNRS*
COTTIAS Myriam, *Directrice de recherche CNRS*
CHIARONI Jacques, *Professeur*
GESSAIN Antoine, *Professeur*
CALDERON Rosario, *Professeur*
DUPUY Francis, *Professeur*
DUGOUJON Jean-Michel, *Directeur de recherche émérite CNRS*

Ecole doctorale : *Biologie Santé Biotechnologie (BSB)*

Unité de recherche : *Anthropologie Moléculaire et Imagerie de Synthèse, CNRS - UMR 5288*

Directeur(s) de Thèse : *DUGOUJON Jean-Michel*

Rapporteurs *: QUINTANA-MURCI Lluis et COTTIAS Myriam*

AUTEUR : FORTES LIMA, César Augusto

TITRE: Tracing the genetic origin of African descendants from South America.

I thank Pr. Tomas Axelsson and his staff in the SNP&SEQ Technology platform in Uppsala University, Sweden, for performing the genome-wide SNP genotyping. I thank Dr. Lluis Quintana-Murci and Dr. Etienne Patin for sharing the genome-wide SNP data of Central African populations.

Likewise, I would like to acknowledge hundreds of participants from various populations in South America, West Africa and West Europe, who voluntarily donated their DNA samples and whose genetic data are reported in this study for the first time. I am very grateful to scientific partners for collecting DNA samples, such as Dr. Antonio Salas, Dr. Pedro Moral, Dr. Rosario Calderón, Dr. Luisa Pereira, Dr. Florence Migot Nabias, Dr. Gil Bellis, and Dr. André Chaventré. A very special thank you goes to Dr. Andres Ruiz-Linares and Dr. Gabriel Bedoya for recruiting the Afro-Colombians samples; Dr. Maria Cátira Bortolini and Dr. Rafael Bisso Machado for recruiting the Afro-Brazilians samples; Dr. Antoine Gessain for providing us the first Noir Marron samples; Dr. Jacques Chiaroni, Dr. Stéphane Mazieres, and Dr. Pascal Bailly for helping me to manage DNA samples from the new Noir Marron sampling.

I would also like to extend my gratitude to my dearest family, close friends, and colleagues at the *Laboratoire d'Anthropologie Moléculaire et Imagerie de Synthèse* of University Paul Sabatier for their patience and guidance during my thesis. A very special thank you goes to Dr. Nicolas Brucato for very insightful discussions regarding to Noir Marron communities and genetic analysis of human diversity.

And last but not least, I thank Dr. Jean-Michel Dugoujon, my advisor on this project, for give the chance to participate in the Marie Curie Network and his unconditional support all these years. He has been extremely patient through my trials and peculiarities, and have provided me with timely and constructive feedback and assessment throughout the length of my doctoral pursuit. I could not thank him enough for everything.

This work is dedicated to memory of distinguished Professor Georges Larrouy for telling me the unknown history of settlement in the Guianas, region that he has developed from the 1960s until nowadays. I would like to send my warmest acknowledgements to him, for his years of dedication to these populations, and with extraordinary emotion we review the genetic history of people in the Guianas and outside of the Guianas.

This dissertation is dedicated to my dearest parents, Virginia and Jose, and my best friend, Jaime, whose support and belief in me and helped me successfully complete this project.

INDEX

I. INTRODUCTION

A. *Historical background of the Trans-Atlantic Slave Trade*

A.1. *Global consequences of the Trans-Atlantic Slave Trade*

The transatlantic slave trade was one of the most extraordinary movements of people in the world history, which persisted for nearly four centuries and resulted in the forced deportation of millions of Africans to the Americas and Caribbean colonies. This historical event ranks as one of the world's worst crimes against humanity. Between around 1500 and the 1860s about 12.5 million African captives were taken from a slaving coast that stretched thousands of miles, from Senegal to Angola, and even round the Cape and on to Mozambique (Figure 1). That means around 30,000 captives a year over three and a half centuries. The captives were enslaved victims of debt, dehydration, and famine; however the largest single source of captives was violence, including warfare, state-sponsored raiding, and kidnapping. As the scale of the Atlantic slave trade grew, the circles of violence in Africa linked to transatlantic slavery intensified and widened. In some Atlantic nations slavery existed as a legal institution and socio-economic system.

Enslaved Africans were loaded on to Atlantic slave vessels; of these, around 11 million survived the transatlantic Middle Passage to reach landfall in the Americas, and many more died in the early years of captivity. Yet until quite recently the full significance of this massive enforced displacement of humanity was largely overlooked by historians. Just a generation ago, the slave trade was considered to be of interest only to maritime historians, or to those studying the histories of Africa or the Americas (Walvin, 2011).

The overriding motive that lay behind the uprooting, enslavement, and coerced long-distance transport of millions of sub-Saharan Africans was the ruthless desire for wealth of European colonizers, to find the cheapest workforce for the production and export of: precious metals, sugar, rum, rice, tobacco, cotton, coffee, indigo, and other luxury goods (Eltis & Richardson, 2010). The magnitude of the slave trade conveys at least a hint of the magnitude of human suffering. By 1820, enslaved Africans constituted around 80% of all the people who had embarked for the Americas since 1500, and mortality on the slave ships averaged at least 15%. Besides, the numerous deaths that occurred as slaves were marched from the African interior to the coast and as they waited to sail, jammed into castle prisons or on board ships (Eltis & Richardson, 2013).
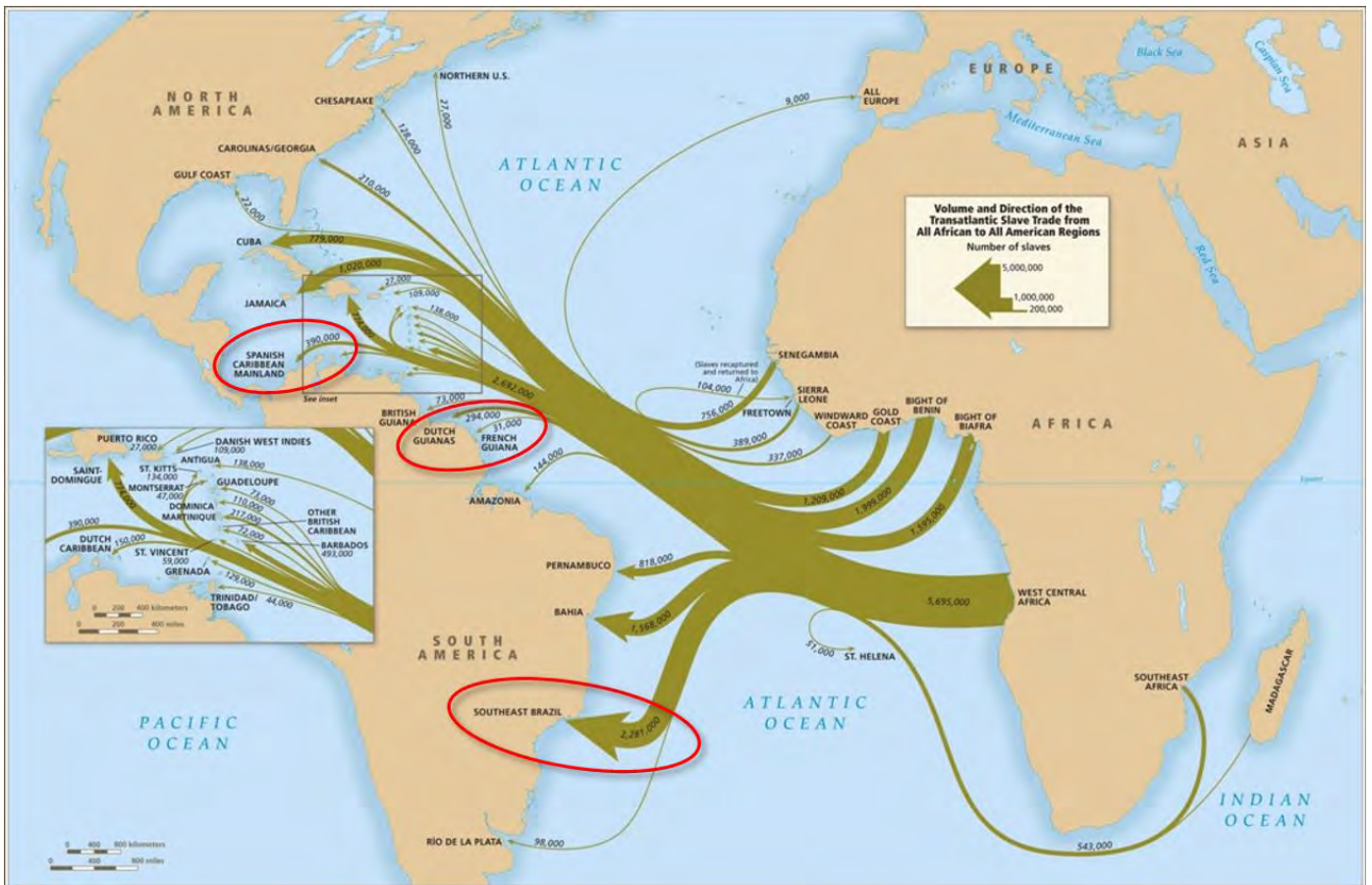
Figure 1: Overview of the transatlantic slave routes from African regions to American regions from 1501 to 1866. Red ellipses point out three different regions in South America: Spanish Caribbean Mainland, Dutch Guiana and French Guiana, and Southeast Brazil (Eltis & Richardson, 2010).

   The transatlantic trade also represented one of the most important commercial and cultural ventures in the formation of the modern world and a fundamental element in the creation of a socioeconomic world system. In very different ways, the transatlantic trade had an extraordinary impact on the American, African, and European continent. In the Americas the most obvious consequence was demographic growth and economic transformation. Until 1830, between three and four Africans crossed the Atlantic Ocean for every European, making the Americas more a demographic extension of Africa than of Europe before the 19th century. Areas that had once been only lightly populated by indigenous peoples became effectively African, they were in fact African settlements on the far side of the Atlantic (Araujo, 2014).

   The transatlantic slave trade even affected the landscape in the Americas. For instance, the labour of enslaved Africans converted Caribbean islands into major agricultural-industrial regions, and their skylines dotted with windmills and factory chimneys, as raw sugar cane was transformed into crude sugar. These colonial settlements were connected to Africa and Europe by regular arrival and departure of growing numbers of ships criss-crossing the Atlantic. They

arrived in the Americas with Africans and materials for the plantations, and departed with slave-grown produce (Walvin, 2013).

In Europe the consequences were enormous, and indeed the wider world, the rapid development of towns and ports involved in slaving was one visible effect of the trade. Profits from both the ships and the plantations flowed back to Europe to be invested in docks, quaysides, port facilities, and related industries as well as rural retreats for those who prospered. For instance Liverpool, London, and Bristol in Britain; Nantes and Bordeaux in France; Lisbon in Portugal; and Texel in Netherlands, all were transformed by their involvement in the transatlantic slave trade (Walvin, 2011). In respect to European dwellers, the slave-produced sugar, tobacco, coffee, chocolate, and other luxury stimulants not only altered the European diet by the late 1700s, had also helped to shape a costumers mentality among European workers, especially in Britain. For instance, workers became more willing to accept factory discipline in order to afford luxury stimulants, and later factory-produced cotton clothing, made possible by the cotton gin and slave labour (Brion Davis, 1999).

The slave trade was an exceptionally cruel and brutal system, both inhumane and immoral, and victims had always struggled against their fate. However, much of the European nations accepted slavery and the slave trade as legitimate and moral five centuries ago. Europeans also believed that enslaving Africans was legitimate because the institution of slavery already existed in Africa. The sense among Europeans and Americans that the slave system itself was morally flawed and irreligious (indeed that it was unchristian) was a very late development. Every European nation that had an Atlantic coastline (such as Spaniards, Portuguese, Dutch, British, French, Danes, and Swedes) participated in the transportation of slaves from Africa to the Americas during the slave-trade era. Furthermore, ideas of European racial superiority increased through the slave-trade era. For instance, in the 18th century the British Parliament passed dozens of Acts that defined Africans as a commodity (Walvin, 2013).

The first enslaved Africans who arrived into the Americas, departed from Europe rather than directly from Africa, and not until the mid-1520s did the first slave ship sail directly from Africa. Besides, African captives may have arrived in the Americas on Columbus's third voyage in 1498. The slave-trade era is generally considered to have begun un uninterrupted human traffic in 1501, when vessels crossing the Atlantic from Spain begun to carry some African captives for sale in the Greater Antilles (the largest islands of the West Indies); and ended in 1867, when the last slave ship from Africa thought to have disembarked its captives in Cuba (also in Greater Antilles) (Eltis & Richardson, 2010).

Indeed, the first enslaved Africans to arrive were probably personal servants, however with the discovery of gold in Hispaniola and the collapse of the Amerindian population in West Indies, the Spanish began to bring slave labour primarily to maintain exports. For the first century of the transatlantic slave trade, the production of precious metals (such as gold and silver) dictated the involvement of the Western Hemisphere in the slave trade. Until the 1590s, the principal source of demand for African captives was precious metals rather than sugar. For the next two centuries, both the size and the range of the transatlantic slave trade expanded dramatically; it was mainly the sugar plantations that created the demand for slave labour. In fact, within the plantation sector, sugar plantations concentrated 80% of the African captives unloaded from the crowed slave ships (Eltis & Richardson, 2010).

Sugar plantations spread widely in the three centuries after they were established in Brazil in the 1560s, and still many plans of sugar plantations have survived. For instance in Cayenne, an outpost of French sugarcane cultivation in French Guiana in the 18th century, sugar plantations required major investments in buildings and equipment as well as enslaved African labour (Figure 2). The enslaved Africans, who typically numbered 100–200 or even more, lived in rows of huts (Morgan, 2013).
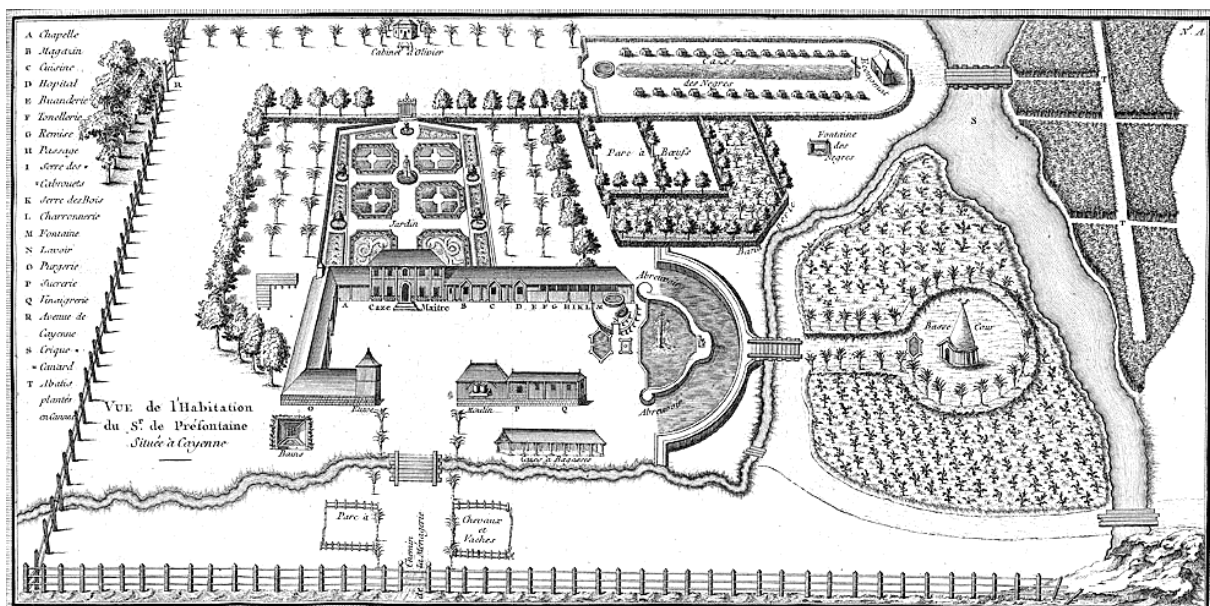
Figure 2: Plan of a Large Plantation in Cayenne (French Guiana) in 1763. Image titled; "*Vue de l'habitation du Sr. de Préfontaine, située a Cayenne*". The image consists of a detailed plan of the Préfontaine's layout, a very large plantation, which comes with the chapel, storehouse, kitchen, hospital, and sugar mill. The rather sizeable slave village (called '*cazes des negres*) is in the upper right hand corner with its rectangular houses. © John Carter Brown Library at Brown University.

Sugar-growing required a large workforce, it was hard and heavy work. Local labour was scarce in the Americas because millions pre-Colombian indigenous people had died after the colonisation of their lands by Europeans. Many were killed in battle, trying to repel the invaders. Other were worked to death, especially in mines. Ultimately, European diseases, such as measles, influenza, cholera, and smallpox wiped out large numbers on indigenous inhabitants. For instance, between 1492 and 1550 the Amerindian population of the West Indies was reduced by 90% owing to massacres. Therefore, European colonist looked to the African continent for a new supply of cheap labour. This resulted in the largest forced migration in human history (Walvin, 2011).

A breakdown of the slave-trade era into three time periods reflects the principal slave-trading nations' entry into and departure from the trade. In the first period, the Spanish and the Portuguese established the first European empires in the Americas, with large assists from northern Italian capital and maritime expertise, they were pioneered the early slave trade. The Iberian powers dominated the trade, with Portugal and Spain being united in 1580-1640 and, in commercial terms, their separation taking place in 1641. By that time, the Portuguese flag, flew over much of the transatlantic slave trade until the mid-17th century (Eltis & Richardson, 2010).

A second phase of the trade started around 1642, when the northern European nations established their own colonies in the Americas, and almost immediately began to engage in the trade, joined by traders from mainland North America. The Dutch, French, English, Danes, Swedes, and Brandenburgers (later Prussians), systematically sent vessels to Africa to obtain captives before 1700. By the early 18th century, the slave trade systems in the North and in the South Atlantic world were firmly established. The Portuguese flag no longer had much of a presence in the slave traffic to Spanish America and the Caribbean but had come to control the South Atlantic slave trade, which was conducted not from Portugal but from Brazil (Eltis & Richardson, 2010).

The second phase ended in 1808, when British and U.S. anti-slave-trade laws of 1807-1808 took effect and other northern European nations began disengaging from the trade. Disengagement ushered in the third period, which lasted through 1867, and during which abolition and suppression dramatically altered the pattern of participation in the slave trade. With the gradual withdrawal of the northern European powers, the slave trade came to be dominated once more by the Portuguese and the Spanish, operating largely from bases in Brazil and Cuba respectively, and against a background of growing abolitionist and suppression activities (Eltis & Richardson, 2010).

A.2. Abolition of the slave trade

The abolition of slavery in the Atlantic world occurred during the 19th century. However, its origins are generally recognized to be the intellectual ferment of the 18th century Enlightenment, with the political turmoil of the "Age of Revolution" and the economic transformations associated with the development of modern industrial capitalism (Drescher, 2009). The abolition movement started with small groups of Europeans and North Americans, who began to turn against the entrenched and lucrative business of slave trading, and the maritime nations of Europe and North America, led by Danish and British governments, finally began legislating to ban the slave trade (Walvin, 2011).

Although antislavery ideas circulated much more widely beginning in the 1760s, the first sustained effort to do something about slavery began in the 1780s, particularly with slave rebellions and the British campaign to end the slave trade. The Saint Domingue Revolution was a well-known slave rebellion in the French colony that broke out in 1791 and soon turned into a revolution; leading to the freedom of 500,000 enslaved Africans and to the creation of the Republic of Haiti in 1804. Consequently, the Saint Domingue Revolution closed down the biggest slave market in the Caribbean, and added a new sense of urgency to the issue in France, Great Britain, and the United States (or U.S.). As it did each of the increasingly troubling slave rebellions that erupted elsewhere in the region during this era, which dropped considerably the number of enslaved Africans (Brion Davis, 1999).

The growing popular awareness of African resistance to their suffering, both on the ships and on the plantations, persuaded more people to support the campaign against the slave system. It was becoming important that Africans hated their bondage and wanted it ended. The abolition movements in Britain, and later in the U.S., employed two powerful political images reproduced everywhere to argue the end of slavery. On the one hand, the famous image of the kneeling slave asking *"Am I not a man and a brother?"*, or *"Am I not a woman and a sister?"* (Figure 3A). This image was reproduced by abolitionists after 1787 in many different forms, such as cameos (called 'Wedgwood cameo'), medallions, and brooches. On the other hand, the other image was of the *Brookes* slave ship. It was a plan and cross-sections of the slave ship, with enslaved Africans crammed (sardine-like) head to toe, to illustrate the horror of the slave ships (Figure 3B). Both images demonstrate the inhumanity of a system that kept Africans in chains and subjected to oceanic torment in ships; and both images have survived to this day as a graphic portrayal of the slave trade and the campaign against it (Walvin, 2013).

American criticism of the slave trade became widespread in the last years of the 18th century, mainly because North America no longer needed new captives. The existing slave population was expanding quickly. The U.S. trade was mainly internal and overland, not transatlantic.

Other slave societies in the America, notably in Brazil and in the French and Spanish Caribbean, continued to demand slaves from Africa. South America still clung to plantations slavery, and expanded greatly the next century (Walvin, 2011).

A: **B:**
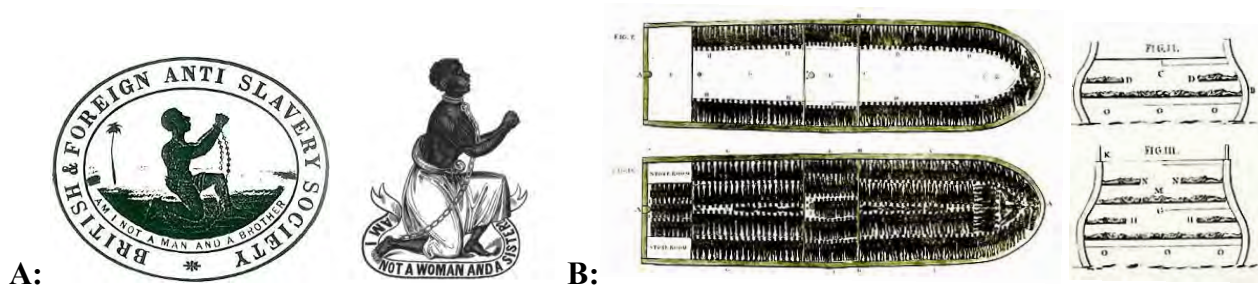
Figure 3: Representations of Wedgwood cameos (**A**) and Brooks slave ship diagram (**B**) published by the British Society for Effecting the Abolition of the Slave Trade in 1788. © British Museum.

In the first decade of the 19th century, the British and the U.S. governments abolished the slave trade in 1807, and made efforts to suppress it throughout the Atlantic world. Although, the institutions of slavery itself lived in the British colonies until 1838 and in the United States until 1865. Despite their efforts, over 2.5 million Africans were further transported across the Atlantic as slaves in the decades of the mid-19th century. Markedly, the British played a large role in suppression of human traffic, albeit the fight against the slave trade always had a strong international dimension, and many European nations took actions against the slave trade only after signing treaties agreeing to do so. Portugal was the last European nation to abolish the slave trade in 1836 (Walvin, 2011).

Slowly thereafter, slavery would be outlawed in many of the new independent Latin American nations throughout the British Empire in 1833, and thirty years later in the Dutch colonies. In 1820, the Spanish government ended the legal slave trade to Cuba. In 1848, slavery was abolished in the French colonies. Although the law was passed in 1794, the decree was never respected, not even after the French Revolution (1789-1799), and Napoleon Bonaparte restored the practice. Therefore, 54 years later, slavery became effectively illegal in metropolitan France and its colonies (Drescher, 2009).

Slavery was also gradually abolished in the Northern United States, with all Northern states passing emancipation acts between 1777 and 1804. However, for the slave owners of the Southern United States emancipation was still unthinkable. This issue was one of the prime causes of the American Civil War (1861-1865). Not until the 1860s, slavery came to a halt in the entire United States, when the American Civil War ended with the defeat of the Confederate States, and the last 40,000 slaves were freed by the final ratification of the Thirteen Amendment to the Constitution in December 1865. Finally, Brazil took actions for its own traffic in 1831,

although an illegal trade continued two decades thereafter, until 1850 when Brazil suppressed its slave trade (Campbell, 2015; Davis, 2006).

The end of the slave trade marked a dramatic turnabout, and the system, which had served the Europeans so well in their settlement and development of key areas of the Americas and brought them such material reward, was rejected. More than that, the West came to view slavery as a moral aberration. In the 19th century, because of the slave trade became illegal, slave vessels were subject to capture and condemnation. After passing their own acts of abolition, the British and the Americans embarked on a crusade to persuade, or force, everyone else to follow their example and outlaw the slave trade.

Indeed, Europeans and Americans became fierce enemies of slavery and slave trading, concerned to destroy it wherever they encountered it in Africa, India, and East Asia. Unfortunately, the history of slavery continued to hold nowadays, it is a sobering thought that there are an estimated over 27 million slaves on Earth today (Bales et al., 2009). More than at any point in history, and more than were stolen from Africa during the four centuries of the transatlantic slave trade.

A.3. Historic links between African coastal regions and the New World

Information on the sources of African captives entering the Atlantic slave trading is considerably limited, as is information on the circumstances in which they were enslaved and forced to move to the coast for the voyage to the plantations and mines of the Americas. Many captives came from places much closer to the African coast, and some captives came from several hundred miles inland and took them months to reach the African coast. They were victims of dehydration and famine before arriving at the places of embarkation. However, there is an important lack of records about the African ethnicity of slaves arriving in the Americas. For instance, the largest historical databases, such as the Transatlantic Slave Trade Database (Eltis et al., 1999), details the departure port for the majority of slaves, but not their home town or ethnicity.

Besides, political borders in the African continent were remarkably different during the slave trade era than the current political distribution of African countries. For instance, the explored West African regions by Europeans were called "*Negroland*" and "*Guinea*" (Figure 4). The extensive trade in ivory, gold, and slaves made these regions wealthy, with a considerable number of centralized African kingdoms developing in the 18th and the 19th century. Afterwards, European traders subdivided the region based on its main exports, such as the Slave Coast, the Gold Coast or the Grain Coast (Lovejoy, 2011).

Figure 4: Map of African continent in 1736, and in detail *"Negroland"* and *"Guinea"* regions established by European settlers. That is one of the finest maps of West Africa to appear in the mid-17ᵗʰ century. The coast is highly detailed with numerous notations in Latin regarding the peoples and tribes of the region. © http://www.geographicus.com/.

According to historical resources, from 1501 to 1867 enslaved Africans were embarked from eight major historical coastal regions in sub-Saharan Africa (see Figure 5): 5.7% of the captives were form Senegambia, 3.2% from Sierra Leone, 2.7% from Windward Coast, 9.6% from Gold Coast, 16.1% from Bight of Benin, 12.3% from Bight of Biafra, 46.3% from West Central Africa, and 4.1% from Southeast Africa (see Figure 6). Furthermore, the proportions of African captives arrived to the Americas differed remarkably among North, Central and specially South America (Figure 1) (Eltis & Richardson, 2010; Eltis & Richardson, 2013).

The Table 1 shows the distribution of the estimated number of enslaved Africans embarked on slave vessels in different African coastal regions during the slave-trade era (Voyages Database, 2009). West Central Africa was always the largest regional point for captives throughout most of the slave trade era (Figure 6), and much of the trade there was focused on Angola. As the transatlantic slave trade expanded after 1641; Gold Coast, Bight of Benin, Bight of Biafra, and West Central Africa slaving regions became more prominent than they had been (Figure 6). Ten ports of trade that dotted the African seaboard were embarkation points that collectively supplied almost two-thirds of the captives. All these ports were located south to the Windward Coast included: Anomabu (on the Gold Coast); Ouidah (on the Bight of Benin); Bonny and Old Calabar (on the Bight of Biafra); Luanda, Benguela, Cabinda, Malembo, and Loango (on the West Central Africa); and Quilimane (on the Southeast Africa) (Eltis & Richardson, 2010).

Table 1: Estimated number of enslaved Africans embarked on slave vessels, and leaving major African coastal regions from 1501 to 1866 (see Figure 6) (Voyages Database, 2009).

Coastal Region	Senegambia	Sierra Leone	Windward Coast	Gold Coast	Bight of Benin	Bight of Biafra	West Central Africa	Southeast Africa	Total
Period	*5.71%*	*3.17%*	*2.68%*	*9.63%*	*16.12%*	*12.31%*	*46.30%*	*4.10%*	*100.00%*
1501 - 1525	8,923	0	0	0	0	0	452	0	9,375
1526 - 1550	31,117	0	0	0	0	1,463	2,953	0	35,534
1551 - 1575	33,829	821	0	0	0	2,457	6,025	0	43,132
1576 - 1600	29,251	168	1,737	0	0	2,092	77,996	0	111,244
1601 - 1625	16,703	0	0	48	2,470	2,045	253,322	241	274,829
1626 - 1650	23,961	1,118	0	2,040	4,969	26,014	194,773	0	252,876
1651 - 1675	21,524	781	268	24,779	42,445	59,248	237,839	13,399	400,282
1676 - 1700	44,458	3,937	904	60,178	168,256	48,976	257,244	10,737	594,689
1701 - 1725	46,604	5,679	7,643	193,704	318,861	51,811	291,266	8,998	924,567
1726 - 1750	71,134	13,799	31,252	193,458	301,906	145,939	484,571	2,263	1,244,321
1751 - 1775	113,487	70,291	141,067	226,885	250,418	253,687	585,280	3,237	1,644,353
1776 - 1800	74,624	84,335	66,001	253,088	235,369	293,461	753,852	35,899	1,796,627
1801 - 1825	79,679	78,561	32,891	72,206	182,688	230,979	838,385	151,784	1,667,172
1826 - 1850	15,726	75,081	5,603	4,532	187,639	199,601	838,688	187,527	1,514,396
1851 - 1866	0	4,214	0	0	29,813	2	132,783	22,445	189,257
Total	**611,019**	**338,785**	**287,366**	**1,030,918**	**1,724,834**	**1,317,775**	**4,955,430**	**436,528**	**10,702,656**

Figure 5: Historical African coastal regions from which slaves were carried during the slave trade. The map shows the limits of the eight major coastal regions in sub-Saharan Africa during the slave trade period and their alignment with modern African nations. 'Upper Guinea' is broken down into three sub-regions: Senegambia, Sierra Leone, and Windward Coast (Eltis & Richardson, 2010).
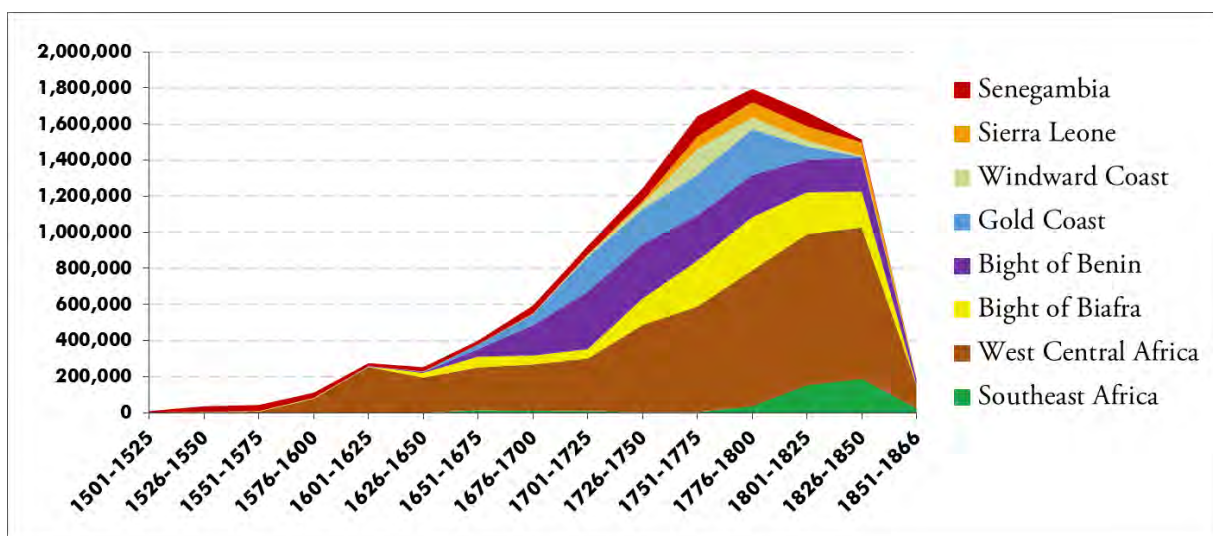
Figure 6: Estimated number of enslaved Africans disembarked on the Americas from 1501 to 1866 (see Table 1), leaving African coastal regions (Voyages Database, 2009).

In the 18th century, slave traders competed intensely with each other on the African coastal regions (Figure 7). Certain European countries dominated particular regions for a time. The Portuguese flag was usually the one seen in the major ports of West Central Africa region (Luanda and Benguela), and on the southern rivers of the Senegambia region. The English were almost as successful in excluding other nations from the Bight of Biafra region, in what is now southeast Nigeria (Lovejoy, 2011). They managed to carry off 87% of the slaves taken from this region. The French dominated slaving ports just north of the Congo River prior to the outbreak of the Saint Domingue slave rebellion (1791-1804). The Dutch had enclave markets on the Gold Coast region and at Cap Lahou (now Grand Lahou, in Ivory Coast). Moreover, traders from countries on the Baltic Sea (Finland, Sweden, and Denmark) carried off most of the captives taken from the eastern Gold Coast, the area around Fort Christiansborg (now Osu Castle, in Ghana) (Eltis & Richardson, 2010).

Senegambia, Sierra Leone, and Windward Coast slaving regions are commonly grouped into 'Upper Guinea' region (Figure 5), together these regions represent around 11.6% of the embarked enslaved Africans (see Table 1). Upper Guinea was a large area that currently includes the modern countries of Senegal, Gambia, Guinea-Bissau, Sierra Leone, Liberia and part of Ivory Coast. Senegambia slaving region includes modern Senegal, Gambia, and part of Guinea-Bissau, as well as counting offshore islands. While Sierra Leone slaving region stretched from Guinea-Bissau to west of Cape Mount (in modern western Liberia). This region includes modern Guinea, part of Guinea-Bissau, Sierra Leone, and western Liberia. Lastly, Windward Coast slaving region (also called as the Grain Coast) stretched from east of Cape Mount to Grand Lahou (in modern Ivory Coast). This region includes modern Liberia and Ivory Coast (Eltis & Richardson, 2010).

Upper Guinea was the first part of Africa to be pulled into the transatlantic slave trade, and enslaved Africans went initially to Europe. In the 16th century, the Portuguese gathered slaves from many parts of Upper Guinea to Cape Verde Islands. Uninhabited prior to 1460, the Cape Verde Islands were firstly settled by the Portuguese with the specific aim of establishing a staging post for slave ships travelling between Africa, Europe, and the Americas. The prominence of these islands lay in their strategic location at the crossroads of shipping routes. Enslaved Africans were used here as agricultural labour to produce food to sustain the small colony of European settlers and passing ships. Therefore, slavery lies at the very early origins of Cape Verde society (Lane & MacDonald, 2011).

	1501–1641		1642–1807		1808–1867	
Documented embarkations	233,000	28%	5,632,000	66%	2,057,000	64%
Estimated embarkations	601,000	72%	2,851,000	34%	1,148,000	36%
Total embarkations	834,000		8,483,000		3,205,000	
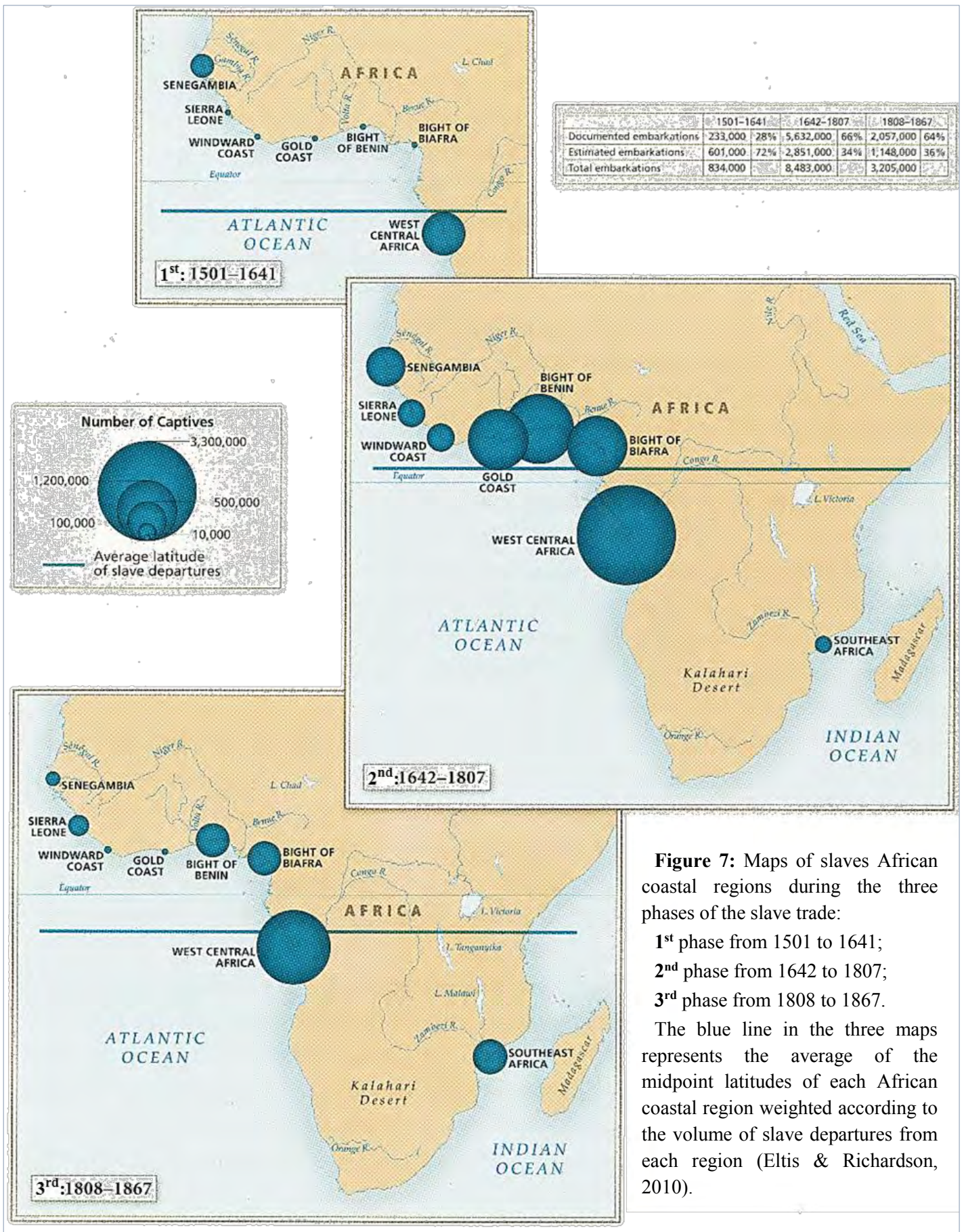
Figure 7: Maps of slaves African coastal regions during the three phases of the slave trade:

1st phase from 1501 to 1641;

2nd phase from 1642 to 1807;

3rd phase from 1808 to 1867.

The blue line in the three maps represents the average of the midpoint latitudes of each African coastal region weighted according to the volume of slave departures from each region (Eltis & Richardson, 2010).

The slave traffic from Upper Guinea grew after 1642 (Figure 7), largely because of the entry of English and French companies into the slave trade between 1660 and 1700. The English concentrated on the Gambia River, and the French on the Senegal River. Shipments of captives from Upper Guinea reached their historic peak in 1750-1780, when the Portuguese chartered companies to expand the area where enslaved Africans were transported to Brazil. Finally, the withdrawal of the British, the French, and the Dutch from slaving accelerated a general decline in transatlantic trafficking from Upper Guinea between the 1780s and the late 1820s, with the exception of Galinhas in Guinea-Bissau that continued to be a source of slaves until 1856 (Eltis & Richardson, 2010).

During the slave trade period, Europeans built castles, forts, and trading factories on some parts of the African coast to facilitate trade and confront invasions. For instance, Saint-Louis, established by the French in 1659, was a fortified trading fort commanding the entrance to the Senegal River in Senegambia. Embarkations of enslaved Africans from the fort began in the 1680s and continued through the 1820s. The French were the main carriers, and most African captives went to French America, in French Guiana and especially in St. Domingue (called Haiti after 1804) (Eltis & Richardson, 2010).

The island of Gorée (Figure 8A), which lies 3.5 Km off the coast of Senegal opposite Dakar, was the largest slave-trading centre on the Senegambia from the 15th to the 19th century. Ruled in succession by the Portuguese, Dutch, English, and French; the island was used as a 'slave warehouse' consisting of over a dozen slave houses, and each house contained between 150 and 200 enslaved Africans from different locations in Upper Guinea. For instance, the 'House of Slaves' (Figure 8B), which was built in 1776 by the Dutch, is currently the last surviving slave house in Gorée (Thiaw, 2010). Each cells of 2.60 square meters were reserved for different sex, and contained up to 15 or 20 people seated with their backs against the wall, and chained around the neck and arms. In the middle of that chain, there was a big iron ball that the captive had to carry between his two hands and two legs, which were released only once a day to satisfy their needs. The extremely poor hygienic conditions were so sickening that the first pest epidemic in the island was originated in that house in 1779 (Thiaw, 2011). Today, the island serves as a reminder of human exploitation, and as a sanctuary for reconciliation (Figure 8C).

Both the Gold Coast and the Bight of Benin were key regions in West Africa for slave traffic in the late 17th and most of the 18th century (Figure 7). Both regions supplied more than 2,750,000 of the enslaved Africans, which means more than 25.75% of the enslaved Africans embarked during the slave trade (Table 1). Ports such as Ouidah, Elmina Castle, Cape Coast, and later Lagos were among the very largest sources of captives anywhere on the West Africa coast (Eltis & Richardson, 2013).
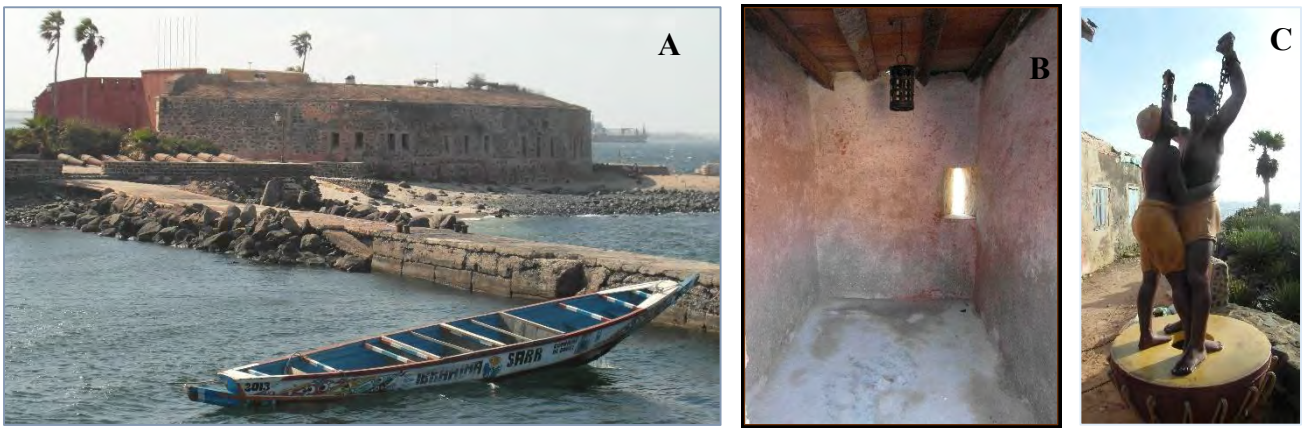
Figure 8: The island of Gorée in Senegal. **A**: Historical museum of Gorée (*'Fort d'Estrées'*), **B**: Cell in the Slave House (*'Maison des Esclaves'*) with a tiny window, and **C**: Statue of freed slaves (*'La statue de la libération de l'esclavage'*). Today the whole island was recognized by UNESCO as World Heritage Sites in 1978. © 2014 Cesar A. Fortes Lima.

The Gold Coast slaving region stretched along 400 miles (643.7 Km) coastline, from east of Axim (in Southwester Ghana) to the Volta River (Figure 5); and most of this region is in modern Ghana. European built most of their trading forts along this coastline, several of which remain today as important historic sites. For instance, Elmina Castle (Figure 9A), which was built in Ghana by the Portuguese in 1482 and captured by the Dutch in 1637, became the headquarters of Dutch trading activity on the Gold Coast. The principal destination of captives leaving Elmina shifted to Dutch Guiana and Dutch West Indies during the 18th century. Moreover, Cape Coast Castle (Figure 9B), which was built in what is now called Ghana, by the Swedes in 1653 and was seized by the British in 1664, was the headquarters of English monopoly on slave trading in West Africa. Enslaved Africans taken from Cape Coast Castle often went to Barbados and Jamaica (Eltis & Richardson, 2010).

Figure 9: Elmina Castle (**A**) and Cape Coast Castle (**B**) in Gold Coast region (modern Ghana) remain today as important historic sites, and are the Ghana's major historic attractions for tourists. Both castles are used as an exceptional testimony to one of the greatest tragedies in the history of human societies, and were recognized by UNESCO as World Heritage Sites in 1979. © 2014 Cesar A. Fortes Lima.

The Bight of Benin slaving region (also known as the Slave Coast) stretched along roughly 500 miles (804.6 Km) across the Gulf of Guinea, from near the mouth of the Volta River (Cape Saint Paul) in eastward Ghana to the Nun River, an extension of Niger River in Nigeria (Figure 5). Today this coastline is part of east Ghana, Togo, Benin, and western Nigeria. The Bight of Benin received its name from the centuries-old kingdom of Benin located in southern Nigeria (now Benin City). This is a rich region in culture and history (Law, 2004; Toyin & Childs, 2004). Prior to European colonization, major African kingdoms existed in this area, including: the Asante (from 1750 to late 1800s) in modern Ghana; Fon (during 1700s) and Dahomey kingdoms (during 1800s) in modern Togo and Benin; and the Yoruba (from 1000s to 1800s) and Benin (from 1200s to 1897) kingdoms in what is now Benin and southern Nigeria (Law, 2004; Lovejoy, 2013).

Ouidah (or *Whidah*, in modern Benin) on the Bight of Benin was the largest single embarkation point for captives shipped from West Africa by the end of the 17th century. Its emergence as a slave port dates from the late 17th century. The Kingdom of Dahomey conquered it in 1725-1727, and ruled it until the late 19th century. For the vessels leaving Ouidah, the Dutch and British Caribbean were the chief destinations, as well as the Dutch Guiana (Law, 2004). This historic site (also known as *Route de l'Esclave*) was added to the UNESCO World Heritage Tentative List in 1996.

The Bight of Biafra slaving region stretched around 370 miles (595.5 Km), from the Nun River to Cap Lopez, at the southern end of the Gulf of Guinea (Figure 5). Today this coastline belongs to modern eastern Nigeria, Cameroon, Equatorial Guinea, and northern Gabon. This slaving region also includes Bimbia Island (Cameroon) and the Gulf of Guinea islands: Príncipe, São Tomé, and Bioko. This slaving region emerged as other major source of slaves in the 1660s (Table 1). The slave traffic was largely centred on New Calabar, Bonny, and Old Calabar port (all of them in modern Nigeria). Together, those ports sent out more captives than the rest of this region combined. The majority of the captives were speakers of Igbo dialects from the multi-ethnic Niger Delta region. The gulf islands, São Tomé and Princes Islands, were also important; most of its slaves were obtained hundreds of miles south, in West Central Africa rather than the adjacent mainland. Cameroon was the smallest of the embarkation points in the Bight of Biafra, and the last to engage in the slave trade. Apart from a single voyage recorded in 1658, departures did not begin until the 1760s (Eltis & Richardson, 2010).

The West Central Africa slaving region ranged from Cap Lopez to the southern tip of Angola, which does not match with the West-Central African geographic region. The southernmost slaving port in Atlantic Africa includes modern part of Gabon, the Republic of Congo, the Democratic Republic of Congo, and Angola (Figure 5). Almost all the slaves from West Central

Africa left from ports located in modern Angola and the Democratic Republic of the Congo. The first vessels sailing for the Americas left from the Congo River region, however a decade after the founding of Luanda (in modern Angola), the new settlement come to monopolize slave traffic from this region. Luanda was established by the Portuguese in 1575, and in most years after 1620 supplied more African captives to the Americas than any other location in sub-Saharan Africa (Eltis & Richardson, 2010). Although Luanda was the principal slave port, Benguela (also in modern Angola) emerged as a major port south of the Congo River, and remained as a key embarkation point for captives from the 1720s; beating even Luanda in the 1840s. Two-thirds of the captives that embarked in Benguela (that means around 491,000 enslaved Africans) had been disembarked in Rio de Janeiro. The Portuguese, who had a continuous political presence in both Luanda and Benguela, drew many of their captives from marauding trading chiefs whose main source of enslaved Africans were Ovimbundu and Mbundu people. The Bantu-speaking Lunda Empire, in east of Luanda, was also participating to supply large number of enslaved Africans to major slaving ports in West Central Africa. Many enslaved Africans, who came from West Central Africa, were mostly called 'Congos' in the Americas; however they were in fact drawn from a wide range of ethnolinguistic groups (Eltis & Richardson, 2013).

Lastly, the Southeast Africa slaving region was the slave-trading region farthest away from the Americas. Consequently, more captives died on large voyages from that region than from anywhere else. The Southeast Africa stretched along the East Africa's coastline and included the islands of Madagascar and Zanzibar (Figure 5). This part of Africa provided enslaved Africans to the Indian Ocean and Arabian markets for many centuries. After 1781, slave departures to the Americas grew rapidly. The centre of the slave trade in this region was Mozambique, particularly in embarkation points such as Mozambique Island, Quilimane, and *Lourenço Marques* (now Maputo). By the 1830s, this region had become the largest supplier of slaves to the Americas after West Central Africa (Figure 7); probably owing to the expansion of plantations in Brazil and British-led efforts to suppress slaving activities north of the Ecuador. In this region, African captives were mostly Yao and Makua peoples (Eltis & Richardson, 2013).

Traditionally, historians have described the transatlantic slave traffic as dominated by males and adults. Adults were usually defined as those older than thirteen or fourteen years old or taller than four feet four inches (~1.33 m). Historians have attributed high ratios of males to females to the plantation owners' demand for strong labourers (Eltis & Richardson, 2013). Today, it is possible to measure variations in the gender and age of distribution of enslaved Africans transported across the African coastal regions (Figure 10). Modern research shows

that, in comparison to other long-distance migrations before 1800, the numbers of females and children in the trade were extremely high, and that important differences among African regions cannot be explained by the demands of buyers in the Americas (Morgan, 2013).
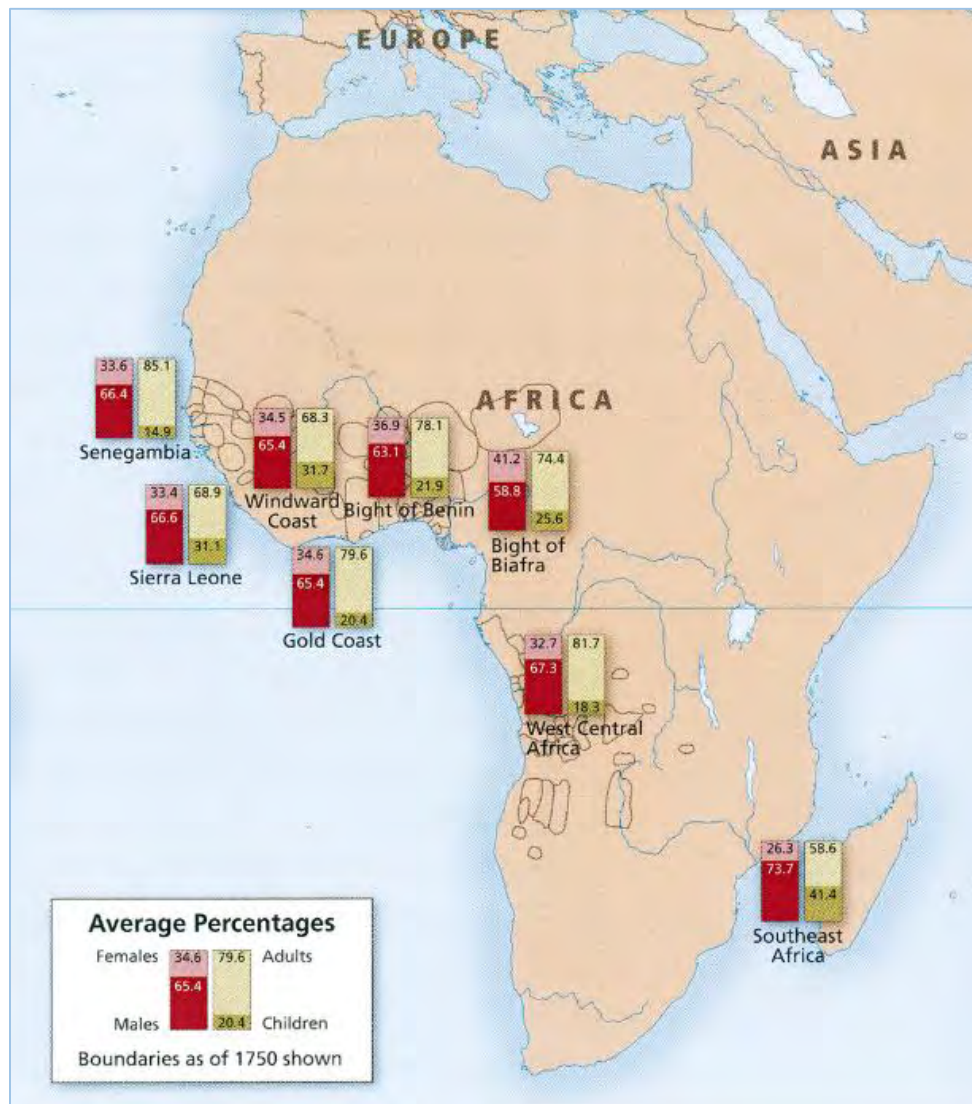
Figure 10: Map of gender and age of slaves carried from African regions to the New World from 1545 to 1864 (Eltis & Richardson, 2010).

Given the gender and age of slaves embarked on slave vessels in different African regions (Figure 10), three important patterns emerged: *i*) the male percentage of captives leaving the Bight of Biafra region was always lower than the female percentage of captives leaving other coastal regions; *ii*) adult-to-child ratios varied enormously across African coastal regions (approximately 26% of all slaves carried out to the Americas were children); and *iii*) over time, the proportion of males carried off from Africa increased as did the proportion of children, and strikingly appears across all African slaving regions at about the same time (Eltis & Richardson, 2010).

A.4. Historic links between South America and African coastal regions

The transatlantic slave trade changed dramatically the demography of Africa and also the New World. South America (43.1%) and the Caribbean (51.9%) accounted for 95% of the disembarked enslaved Africans in the Americas; and fewer than 4% disembarked in what became the United States (Figure 1). Only a very small percentage of slaves disembarked in Africa, after their ships were diverted because of slaves on-board rebelled. However, around two million enslaved Africans died during the Middle Passage (Eltis & Richardson, 2010). Indeed, the Middle Passage was one of the most dreadful and traumatic experience for African captives. They were herded naked onto small overcrowded vessels, and dispatched to far-off lands from which no return was possible (Eltis & Richardson, 2013).

To highlight the extent of this movement, the Table 2 shows the estimated number of enslaved Africans arriving at the main Spanish, French, Dutch, and Portuguese colonies in South America in different numbers and at different periods of the slave trade. This historic data was obtained from 4,349 well-documented slave vessels, which embarked over 1,640,000 enslaved Africans from 1549 to 1856 (Voyages Database, 2009). In general, slave vessels that completed the longest passages experienced also the highest shipboard mortality. Hence, slave vessels arrived to Cartagena present the highest mortality percentage during voyage (23.1%), than vessels arrived to Guiana (16.3%) or Southeast Brazil (8.8%) (Table 2). The male percentage (71.2%) and the percentage of adults (94.4%) were higher in Rio de Janeiro than other regions in South America (Table 2). This fact points out the high demands of labour workers in plantations and mines in Brazilian. In the Spanish American mainland, the 85% of the enslaved Africans arrived from 1576 to 1650, which is remarkably earlier than in French and Dutch Guiana, and in Southeast Brazil. Because of the Spanish Crown had forbidden the enslavement of Amerindians at the beginning of the 17th century; and at that time the enslavement of Africans enabled a rapid develop of Cartagena (Navarrete, 2011). Later, the slave traffic with African regions ended up in Cartagena in 1788, which is also remarkably earlier than in other slave regions in South America (Table 2).

From 1604 to 1815 about 400,000 enslaved Africans arrived to work in plantations in Guiana region, between Brazil and Venezuela, mainly ruled for Dutch settlers, but also for English, French and Portuguese, and Jewish settlers (Eltis & Richardson, 2010). On the one hand, Surinam (known now as the Republic of Suriname), was the foremost settlement for the Dutch in Guiana, and the centre of: *i*) the rapid growth of sugar cultivation before 1770, and *ii*) the inflows of African labour needed to sustain it. There are estimated around 295,000 captives that arrived in Surinam (Figure 11A), albeit there are only 221,430 documented disembarkations (Table 3). The principal sources of enslaved Africans were West Central Africa (30.4%), Gold

Coast (23.2%), and Bight of Benin (16.0%) (Figure 12A); and there are not any historic records from Southeast Africa (Voyages Database, 2009). Elmina Castel in Gold Coast was the main port with 11.4% of the captives (that means 25,147 enslaved Africans) (Table 3). On the other hand, French Guiana supported a modest and erratic inflow of African captives in its main port, Cayenne, now capital city. African captives arriving in Cayenne came from most African coastal regions, with the exception of Gold Coast (see Figure 11B) (Eltis & Richardson, 2010). Interestingly, the contribution of different African coastal regions was homogeneous and with similar periodicity, mainly with Senegambia (26.7%), Bight of Benin (22.5%), Bight of Biafra (19.2%), and West Central Africa (22.5%) (Figure 12B). Ouidah in Bight of Benin was the most prominent outfitting port with 13.2% of the captives, and Saint Louis in Senegambia was another important port with 9.4% of the captives (Table 3).

Table 2: Estimated number of embarked (Emb.) and disembarked (Dis.) enslaved Africans arriving at main Spanish, Dutch, French, and Portuguese ports in South America from 1549 to 1866. There are also indicated: the total of voyages; the percentage of slaves embarked who died during the Middle Passage; the percentage of male and female enslaved Africans; and the percentage of children and adults (Voyages Database, 2009).

Region	Spanish America		Dutch Guiana		French Guiana		Southeast Brazil	
Main port	Cartagena		Surinam		Cayenne		Rio de Janeiro	
Period	Emb.	Dis.	Emb.	Dis.	Emb.	Dis.	Emb.	Dis.
1549 to 1550	224	166	-	-	-	-	-	-
1551 to 1575	756	594	-	-	-	-	-	-
1576 to 1600	44,646	32,058	-	-	-	-	334	287
1601 to 1625	56,820	40,364	-	-	-	-	-	-
1626 to 1650	25,546	18,212	-	-	-	-	-	-
1651 to 1675	5,470	4,276	8,398	6,546	1,327	1,000	382	331
1676 to 1700	6,990	5,511	23,496	20,274	1,283	973	3,231	2,831
1701 to 1725	7,969	6,104	23,467	19,738	1,634	1,286	29,537	26,090
1726 to 1750	-	-	58,059	49,420	2,651	2,365	96,229	84,440
1751 to 1775	422	344	90,971	80,987	4,219	3,250	69,081	62,853
1776 to 1800	379	284	26,582	23,031	5,997	5,268	136,439	125,429
1801 to 1825	-	-	23,893	21,434	5,495	4,828	547,828	493,470
1826 to 1850	-	-	-	-	2,014	1,706	332,635	303,818
1851 to 1856	-	-	-	-	-	-	1,217	1,021
Total	**149,222**	**107,913**	**254,866**	**221,430**	**24,620**	**20,676**	**1,216,913**	**1,100,570**
Total of voyages	**598**		**765**		**100**		**2,886**	
% died during voyage	*23.10*		*14.30*		*18.20*		*8.80*	
% male / female	*66.70*	*33.30*	*60.60*	*39.40*	*51.00*	*49.00*	*71.20*	*28.80*
% children / adults	*14.80*	*85.20*	*18.60*	*81.40*	*31.50*	*68.50*	*5.60*	*94.40*

In the Portuguese colony of Brazil, the organization slave vessels' voyages was initially cantered on Lisbon, however in the latter 16th century migrated to Brazil, which emerged as the principal centre in the Portuguese-speaking world for the dispatch of slaving voyages from Africa (Klein, 2002). African enslaved people increased and became the primary source of enforced labour, especially after the most severe outbreak of smallpox in Brazil (1665-1666) (Childs, 2007). As Brazil became Europe's leading source of sugar in the late 17th century, also Recife, Salvador de Bahia, and Rio de Janeiro emerged as the leading markets in the Americas for slaves brought directly from Africa (Figure 13A). These centres generated large demands for slave labour within the hinterland of major ports such as Rio de Janeiro, and also generated some of the trade goods dispatched to Africa to purchase slaves. These three major ports were also transit points for captive labourers sold on to the gold producing regions of Minas Gerais and Goiás (Campbell, 2015). After 1760, transatlantic slave vessels were addressed to the Amazonian ports of Pará and Maranhão (Figure 13B), in response to the rapid growth of demand for cotton in North Brazil (Eltis & Richardson, 2010).

Rio de Janeiro was the most active and important slaving port in Southeast Brazil (Figure 14), and also was the home port for the slaves vessels that brought African captives to this region (Eltis & Richardson, 2010). West Central Africa was the major source of slaves for all the Americas, and in South America it was particularly dominant. In Rio de Janeiro, 83.7% of the disembarked enslaved Africans came from West Central Africa region (Figure 15 and 16). They were brought to Rio de Janeiro mainly originated from two important regions in Portuguese Angola: Luanda (42.3%) and Benguela (18.1%) (Table 4). Both major regions are populated by very distinct African ethnolinguistic groups, and are inhabited by speakers of languages belonging most of them to the Niger-Congo linguistic subphylum. This subphylum comprises the largest Bantu branch and includes about 500 languages, which are spoken in virtually all of Central, East, and South Africa, except for the areas occupied by the Khoisan-speaking groups (Paul et al., 2015). Therefore, African captives of Bantu origin from South of the Congo and Angola were overwhelmingly forced to move to Brazil during the slave era (see Figure 15).
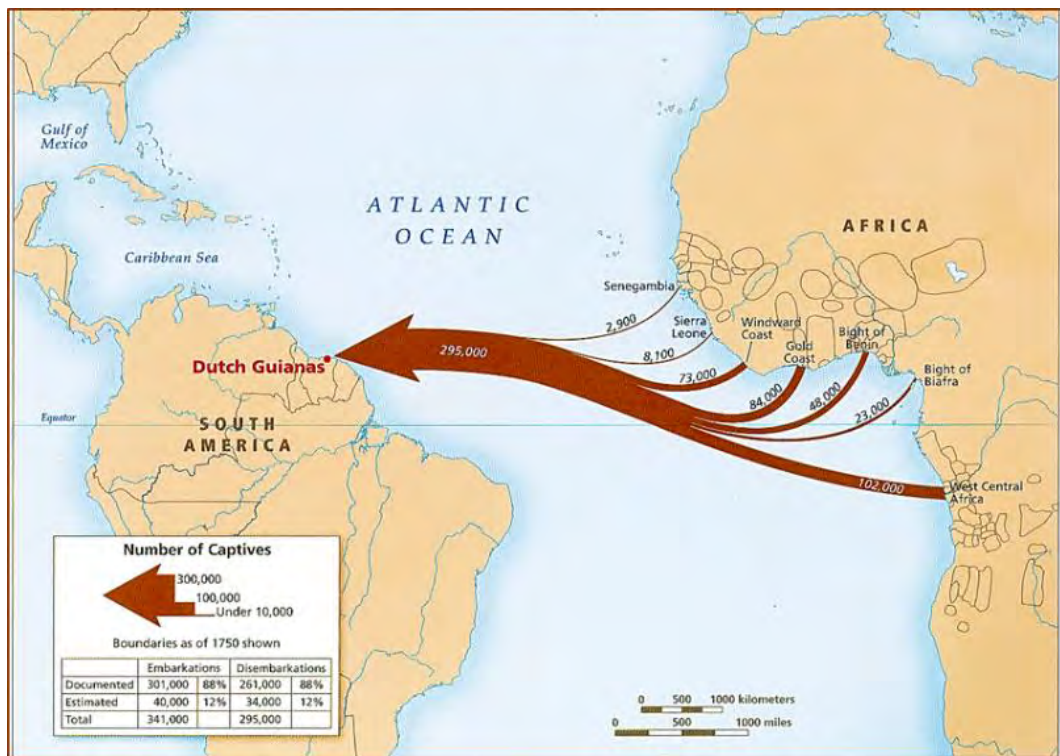
A:
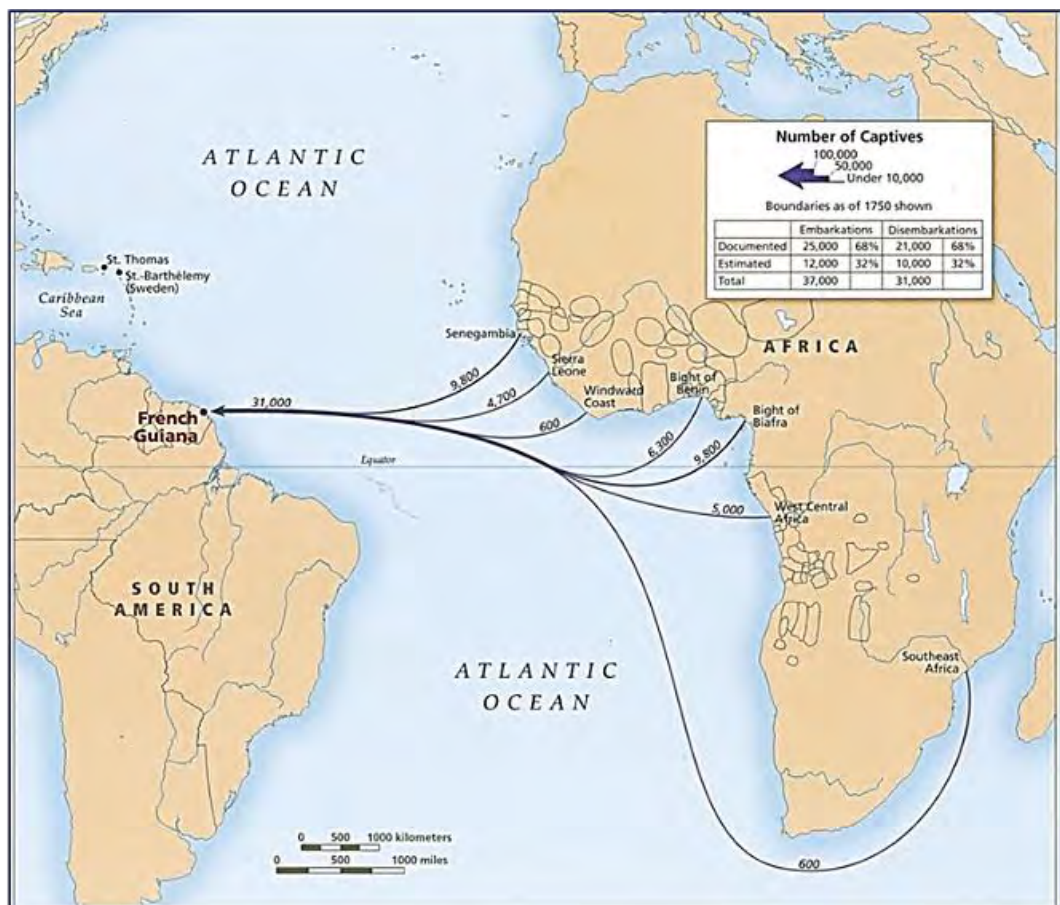
B:

Figure 11: Map of African coastal origins of slaves carried to Dutch Guiana (**A**) and French Guiana (**B**) from 1526 to 1867. The boundaries of historical African political units and ethnolinguistic groups are shown along the African coastal regions (Eltis & Richardson, 2010).
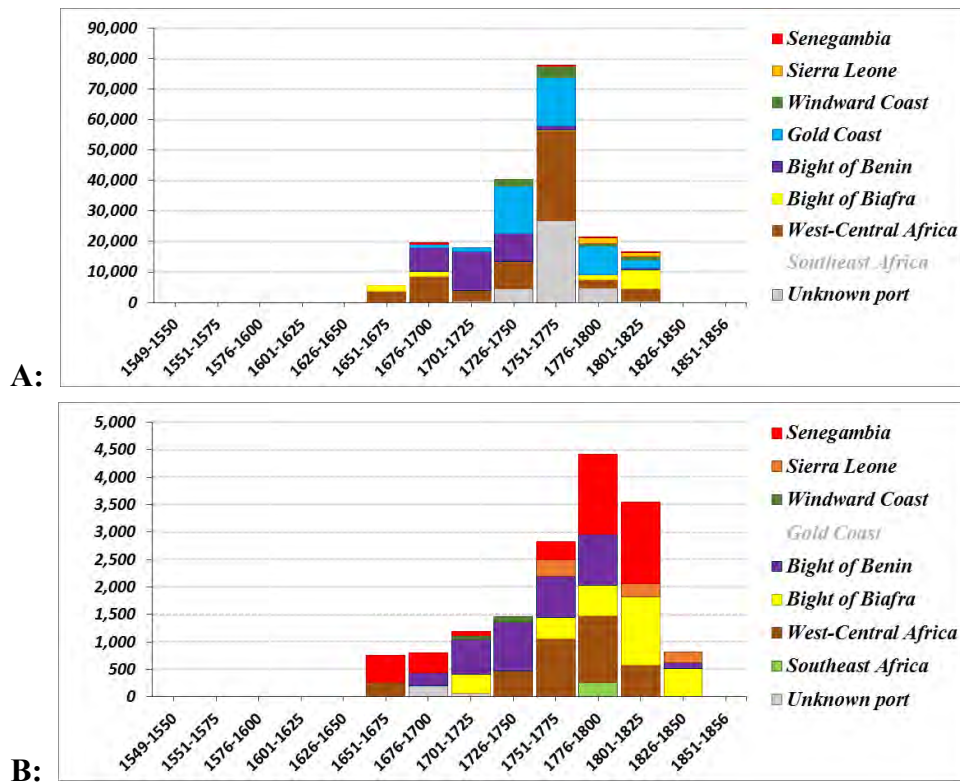
Figure 12: Estimated number of enslaved Africans disembarked on Surinam (**A**) and Cayenne (**B**), leaving African coastal regions from 1549 to 1856. There are no records of enslaved Africans arriving from Southeast Africa to Surinam, or form Gold Coast to Cayenne (Voyages Database, 2009).

Table 3: Estimated number and percentage of: 221,430 enslaved Africans disembarked on the Dutch port of Surinam in Dutch Guiana; and 20,676 enslaved Africans disembarked on the French port of Cayenne in French Guiana, from major African ports from 1664 to 1825 (Voyages Database, 2009).

Surinam	Senegambia	Gold Coast	Bight of Benin	Bight of Biafra		West Central Africa	
Period / Port	Saint-Louis	Elmina	Ouidah	Bonny	Calabar	Malembo	Loango
1664 - 1675					243	391	1,457
1676 - 1700					1,406	730	424
1701 - 1725		1,032	11,720				455
1726 - 1750		10,985				1,029	802
1751 - 1775		5,495	690			9,275	589
1776 - 1800	100	7,635		343	644		349
1801 - 1825				1,462	2,385		
Total	**100**	**25,147**	**12,410**	**1,805**	**4,678**	**11,425**	**4,076**
Percentage	*0.05%*	*11.36%*	*5.60%*	*0.82%*	*2.11%*	*5.16%*	*1.84%*

Cayenne	Senegambia	Bight of Benin	Bight of Biafra	West Central Africa	
Period / Port	Saint-Louis	Ouidah	Bonny	Loango	Portuguese port
1676 - 1700	322	241			
1701 - 1725		436	125		
1726 - 1750		887			286
1751 - 1775		233			816
1776 - 1800	916	932	328		1,204
1801 - 1825	708		540	580	
Total	**1,946**	**2,729**	**993**	**580**	**2,306**
Percentage	*9.41%*	*13.20%*	*4.80%*	*2.81%*	*11.15%*
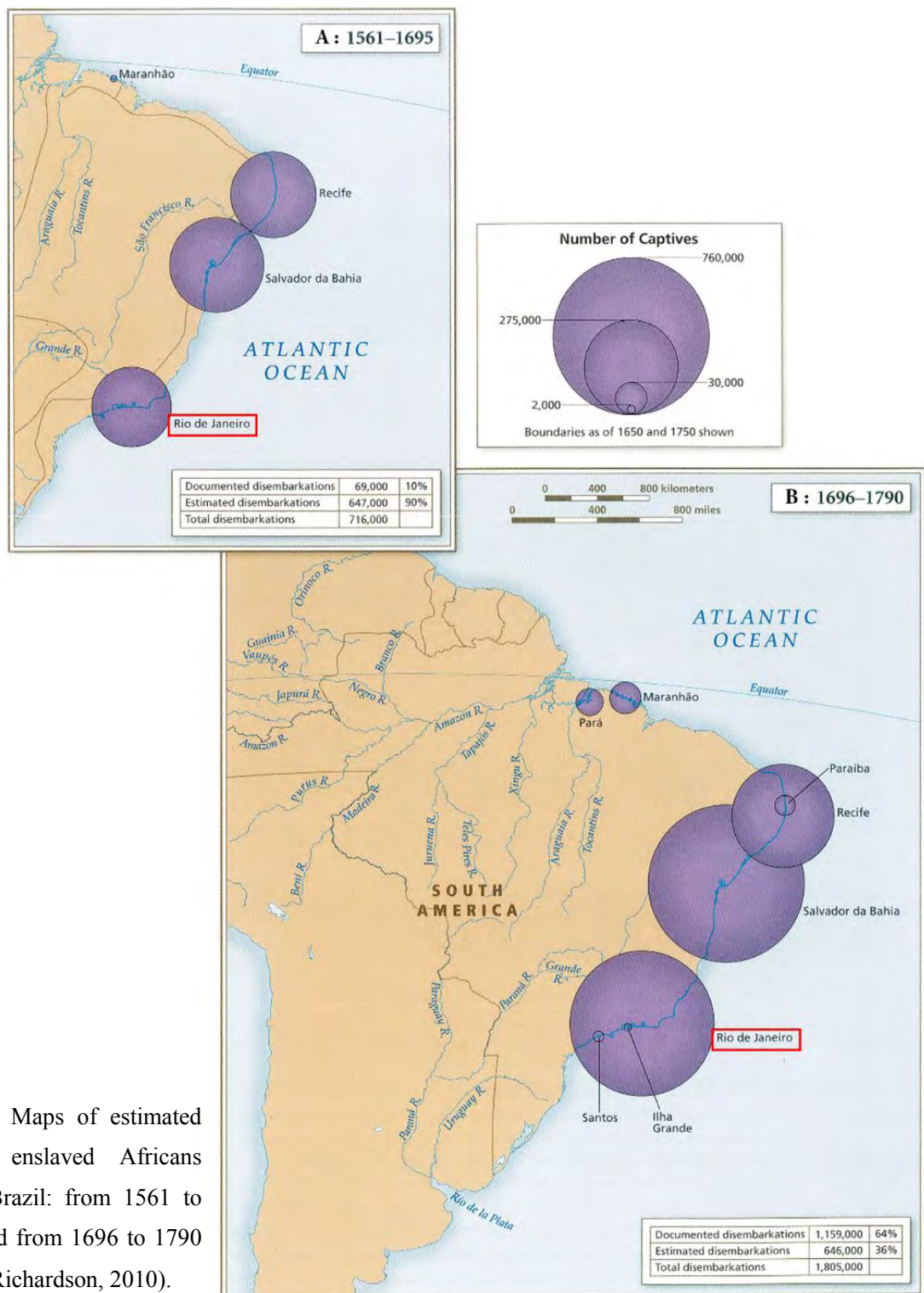
Figure 13: Maps of estimated number of enslaved Africans arriving in Brazil: from 1561 to 1695 (**A**), and from 1696 to 1790 (**B**) (Eltis & Richardson, 2010).

Furthermore, Rio de Janeiro, with over one million disembarked African captives (Figure 15), was the Atlantic Ocean's largest single outfitting port for slave ships, as well as the single most important disembarkation point in the Americas. So slave trading played a vital role in the growth of the city's wealth and population, allowing its merchants and city leaders to invest in the public spaces and buildings (Morgan, 2013). Most enslaved Africans entering the port were sold on to work in sugar, coffee, and, especially in the 18th century, gold production in Minas

Gerais and other provinces. Some of them remained in Rio de Janeiro itself, providing labour for urban crafts, household, and other services (Morgan, 2013). Throughout the Americas, enslaved Africans were used not only to produce sugar, coffee, gold, tobacco, cotton, and other goods for market, also to work as servants of European colonizer to display their owners' status, which means they were both, producers and symbols of their owners' wealth (Araujo, 2015).
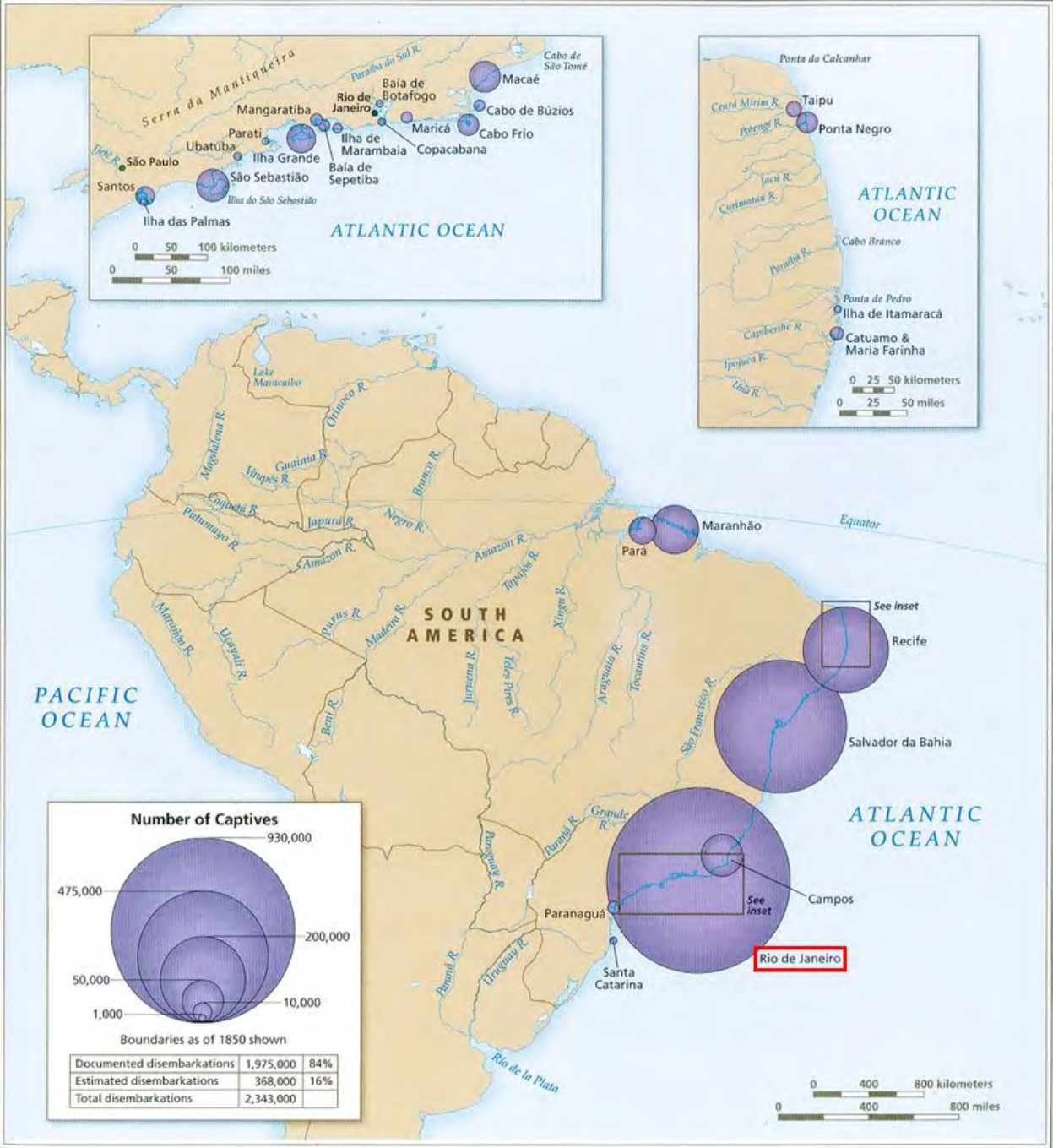
Figure 14: Maps of estimated number of African enslaved arriving in Brazil from 1791 to 1856. In the 19th century, more African captives arrived in Brazil than in any other colony or country to supply the labour needs of sugar and coffee plantations, especially in Rio de Janeiro (Eltis & Richardson, 2010).
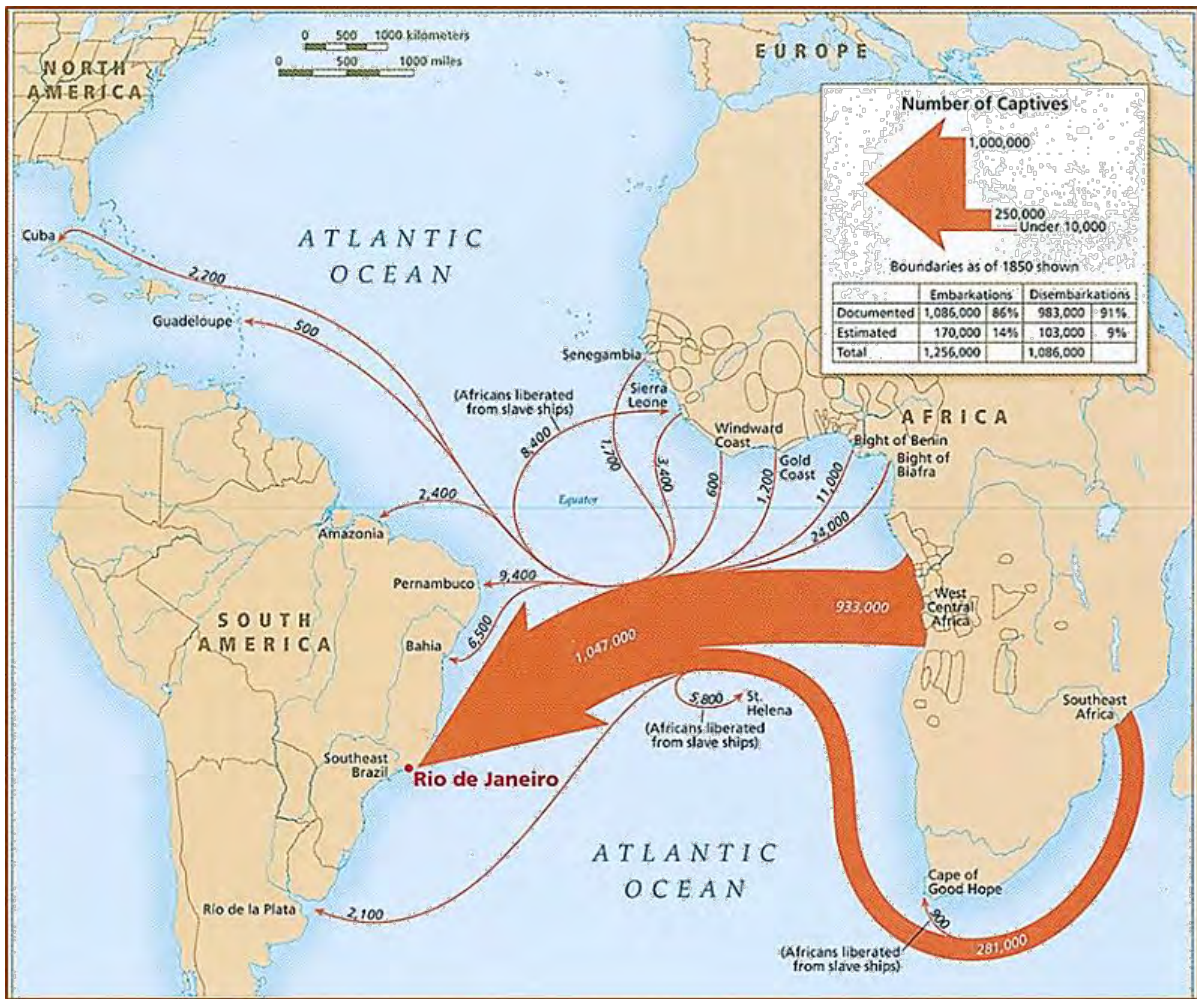
Figure 15: Map of African coastal origins of slaves carried to Rio de Janeiro (in Southeast Brazil) during the third phase of the transatlantic slave trade from 1808 to 1856. In the 1830s, British suppression and anti-slave trade are represented in part on the map by the return loops of Africans liberated from slave vessels and taken to Sierra Leone and St. Helena. The boundaries of historical African political units and ethnolinguistic groups are shown along the African coastal regions (Eltis & Richardson, 2010).
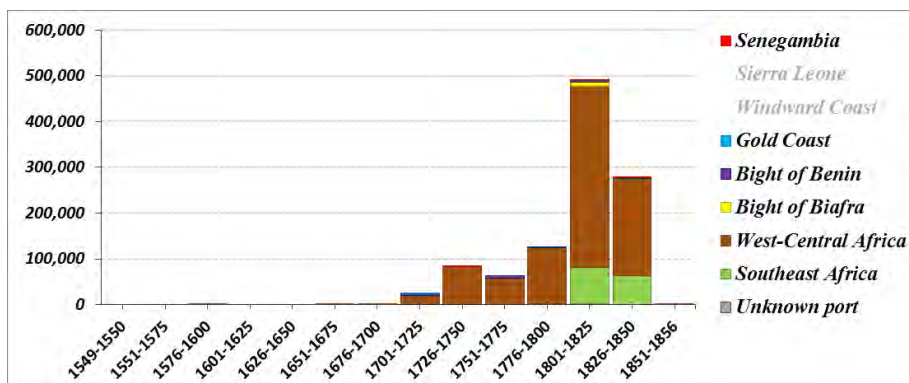
Figure 16: Estimated number of enslaved Africans disembarked on Rio de Janeiro from 1549 to 1856 leaving historical African coastal regions. There are no records of enslaved Africans arriving from Sierra Leone and Windward Coast to Rio de Janeiro (Voyages Database, 2009).

Increasing demand for coffee and sugar beginning in the 18th century sustained the Brazilian slave trade close to peak levels until the mid-18th century. In the late 1820s, more enslaved Africans were imported into Rio de Janeiro than into any other port in the history of the slave trade (Eltis & Richardson, 2010). After 1820, slave vessels addressed to Rio de Janeiro obtained enslaved Africans from more remote parts of the West Central African coast, instead of the traditional centres of Luanda and Benguela, also drew heavily from Mozambique in Southeast Africa (Table 4). Therefore, Southeast Africa region became an important source of captives, where around 13.5% African captives were taken and forced to move to Rio de Janeiro (Figure 16); so this region should be taken into consideration to trace the African origin of current African descendants from Rio de Janeiro.

Table 4: Estimated number and percentage of 1,100,570 enslaved Africans disembarked on the Brazilian port of Rio de Janeiro in Brazil from major African ports from 1676 to 1856 (Voyages Database, 2009).

Rio de Janeiro	Bight of Benin	Bight of Biafra	West Central Africa		Southeast Africa	
Period / Port	Ouidah	Bonny	Benguela	Luanda	Mozambique	Quilimane
1597 - 1600						
1651 - 1675						
1676 - 1700			105	2,726		
1701 - 1725	926		105	19,846		
1726 - 1750			3,000	79,413		
1751 - 1775	1,070		7,711	49,459	431	
1776 - 1800	230		39,154	84,581	1,278	186
1801 - 1825	139	734	106,640	165,996	45,675	26,699
1826 - 1850	335		41,843	63,037	25,856	23,884
1851 - 1856			563			
Total	2,700	734	199,121	465,058	73,240	50,769
Percentage	*0.25%*	*0.07%*	*18.09%*	*42.26%*	*6.65%*	*4.61%*

In the Spanish American mainland, after the conquest of the Azteca and Inca Empires in the 16th century, the Spanish consolidated an empire that comprised most of the Latin America. Enslaved Africans entered this empire mainly through Cartagena in Colombia (with around 196,000 captives), and through Veracruz in Mexico (with around 83,000 captives) (Figure 17). African slavery in Colombia began in the first decade of the 16th century, and by the 1520s, Africans captives were being imported steadily into Colombia to replace the rapidly declining Native American population. Cartagena was designated as an official port of the Spanish fleet system as early as 1537, and became by far the largest single slaving disembarkation port in the Spanish Americas (Almeida, 2008). Enslaved Africans arrived mainly from West Central Africa (45.2%) and Senegambia (38.9%) regions (Figure 18). In West Central Africa, Luanda was largely the first port were Africans were enslaved (around 25.4%), while Portuguese Guinea was the key port in Senegambia (18.3%) (Table 5).

They were forced to work in gold mines, sugar cane plantations, cattle ranches, and large *haciendas* (landed estates owned by colonists). Indeed, African labour was essential in different regions of Colombia, For instance, African workers were pioneered in the extracting of alluvial gold deposits and the growing of sugar cane in the states of Chocó, Antioquia, Cauca, Valle del Cauca, and Nariño in Western Colombia. However, with the arrival of numerous slaves to staff lowland mines and highland haciendas, the frequency of rebellions increased (Navarrete, 2012).

Figure 17: Map of African coastal origins of slaves carried to Cartagena (in Colombia) during the Spanish transatlantic slave trade from 1501 to 1641 (Eltis & Richardson, 2010).
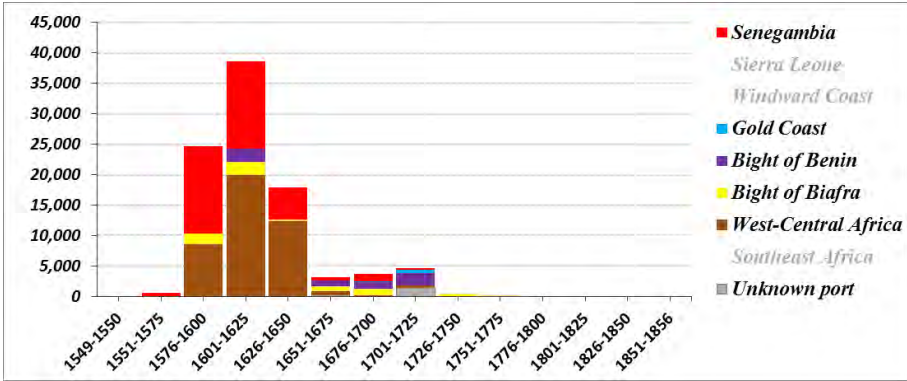
Figure 18: Estimated number of enslaved Africans disembarked on Cartagena (from 1549 to 1856), leaving African coastal regions. There are no records of enslaved Africans arriving from Sierra Leone, Windward Coast, and Southeast Africa to Cartagena (Voyages Database, 2009).

Table 5: Estimated number and percentage of 107,913 enslaved Africans disembarked on the Spanish port of Cartagena in Colombia from major African ports from 1564 to 1788 (Voyages Database, 2009).

Cartagena	Senegambia	Senegambia	Bight of Benin	Bight of Biafra	West Central Africa	
Period / Port	Portuguese Guinea	Cape Verde	Ouidah	Bonny	Luanda	Portuguese unknown
1564 - 1575		428				
1576 - 1600	8,607	5,554			6,752	1,632
1601 - 1625	8,460	3,488			9,588	10,268
1626 - 1650	2,669	743			11,037	1,400
1651 - 1675						650
1676 - 1700		313	944			251
1701 - 1725			1,955			364
1751 - 1775				344		
1776 - 1788				158		
Total	19,736	10,526	2,899	502	27,377	14,565
Percentage	18.29%	9.75%	2.69%	0.47%	25.37%	13.50%

B. Formation of the African-American communities

The sociologist Orlando Patterson defines slavery as *"the permanent, violent domination of natally alienated and generally dishonored persons"* (Patterson, 1990). This definition is important both for what it includes and does not include as part of the cultural baggage of the enslaved. Patterson excludes all references to slaves as a form of 'property' because he considers that *"to define slavery only as the treatment of human beings as property fails as a definition, since it does not really specify any distinct category of persons. Proprietary claims and powers are made with respect to many persons (such as wives or children in some cultures) who are clearly not slaves"*. Patterson places much more emphasis on 'natal alienation' and argues that *"slaves differed from other human beings in that they were not allowed freely to integrate the experience of their ancestors into their lives, to inform understanding of social reality with the inherited meanings of their forebears, or to anchor the living present in any conscious community of memory"*.

In fact, slavery was a degradation of human life and human values. Human beings were turned into working machines. The only way people could free themselves from that system was to be courageous and not to be afraid to sacrifice their lives, therefore wherever slavery flourished, so did resistance. For more than three centuries, in remote areas throughout the West Indies, Central America, South America, and North America, thousands of enslaved Africans managed to escape from the plantations of European colonizers, in search of freedom. These runaway slaves (generally called *"Maroons"*) formed independent African settlements of free people in South America and the Caribbean (such as Jamaica and Barbados), in a process known as

'marronage'. In fact, the English word; Maroons came from the Latin-American Spanish word; *"cimarrón"*, which means: wild, fugitive or runaway, and today this word has strong connotations with 'brave', 'fierceness' or 'unbroken spirit'. In plantations from different countries in Latin America, runaway enslaved Africans were also known as *Palenques* in the Spanish colonies, *Quilombos* or *Mocambos* in Brazil, and *Cumbes* in Venezuela. The initial Maroons in any New World colony hailed from a wide range of African societies in West and West Central Africa, and at the outset they shared major aspects of culture, including language (Agorsah, 1994).

These Maroon communities have subsequently emerged as free and independent societies that have forced colonial governments to sign treaties and pacts to guarantee: their freedom, their land, and their own political autonomy (Price, 2002). After centuries of survival and adaptation in the Amazon rainforest, they have developed a unique identity and history. Indeed, marronage represented a major form of slave resistance, whether accomplished by lone individuals, by small groups, or in great collective rebellions. Throughout the Americas, Maroon communities stood out as a heroic challenge to European authority, as the living proof of the existence of a slave consciousness that refused to be limited by the European' conception of the docile slave. It is no accident that throughout the Latin America, the historical Maroon has become today a touchstone of identity for the region's writers, artists, and intellectuals; and the ultimate symbol of resistance and the fight for freedom.

Today in South America, African descendants of some of the original Maroon communities still preserve a strong sense of their history, traditions, values, and identity; which are deeply rooted in their African past. For instance, the Noir Marron communities in French Guiana and Surinam, since their formation in the 18th century, have adapted to the Amazonian environment because of their cultural practices are inherited from African, Amerindian, and European cultural exchanges (Price, 1996).

B.1. Historical, cultural, and social backgrounds of the Noir Marron communities

The Noir Maroon from French Guiana and Surinam (formerly known as *"Bush Negroes"*, *"Bushinenge"*, or *"Buschinengué"*) have always been the largest Maroon population in South America. They are at once the most culturally, politically, and economically independent of all Maroon peoples in the Americas (Price & Price, 2001). They are the direct descendants of enslaved Africans who managed to flee from Dutch colonial oppression, and escaped from coastal plantations in the Dutch colony of Surinam (Figure 19A). They fled into the forested interior out of reach of both the planters and the colonial powers, where they regrouped into

small bands, and began forging a viable existence in this new and inhospitable environment (Price & Price, 2003).

Marronage on a grand scale, with individual fugitives banding together to create free communities, struck directly at the foundations of the plantation system, presenting military and economic threats that often taxed the colonists to their limits. Maroon communities, hidden near the fringes of the plantations or deep in the forest, periodically raided plantations looking for firearms, tools, and women; often permitting families formed during the slavery to be reunited in freedom. This daunting challenge was made even more difficult by the government's persistent and massive efforts to eliminate the threat they posed to the colony's thriving plantations. Therefore, the Dutch colonists reserved special punishments and a variety of deaths by torture written into law for recaptured slaves, such as hamstringing, amputation of limbs, castration, suspension from a meat hook through the ribs, slow roasting to death (Figure 19B, 19C, and 19D), whose punishments were designed *"to serve as an example to others"* (Stedman et al., 2010).

The Dutch government also established military expeditions against the nascent group of Maroons known as Saramaka. During the late 17th and the early 18th century, numerous small-scale military expeditions were mounted, sometimes at the personal expense of particular planters (Price, 1994). The organized pursuit of Maroons, and expeditions to destroy their settlements by the citizen's militia, ended up with a progressively costly warfare. Indeed, a typical expedition costed more than 100,000 Dutch guilders and had to traverse mountains and creeks (Figure 19E) before reaching the Maroons' hidden villages (Price, 1995).
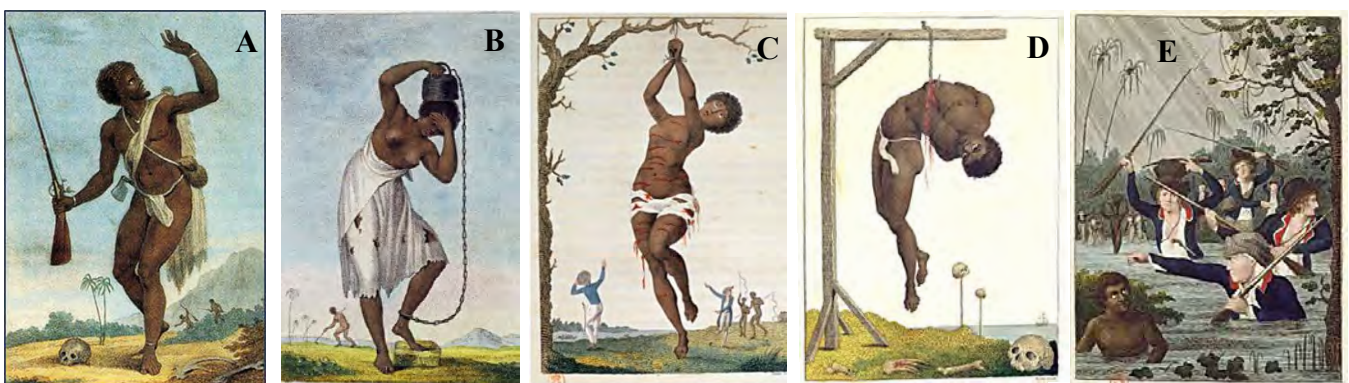
Figure 19: Representations of rebellions and punishments of African enslaved in the Dutch colony of Surinam in the 1770s: **A:** A rebel Marron man armed; **B:** Marron woman carrying a weight chained to her ankle; **C:** Flagellation of a Marron woman with deep lacerations; **D:** Marron man hung alive by the ribs to a gallows; and **E:** Through a swamp in pursuit of slaves (Stedman et al., 2010).

The increasingly costly warfare culminated in a decision by the colonists, during the late 1740s, to sue their former slaves for permanent peace. However, peace proved elusive, in 1754 they decided to mount another massive expedition against the Saramaka consisting of 500 men. In the 1760s, peace treaties were at last successfully concluded with the two largest Maroon groups, the Saramaka and the Ndjuka. Consequently, new slave revolts and the large-scale war of subsequent decades, for which an army of mercenaries was imported from Europe, led to the formation of the Aluku, as well as the smaller Paramaka, Matawai, and Kwinti groups (Price, 1995).

These runaway enslaved Africans are poorly documented, since most records were written by European settlers, explorers, and government officials; who wrote partial and subjective views of events. The majority of Noir Marron history is transmitted by oral traditions, through songs and poems; so there is variation in the some of the historical facts, notably in dates and the names of important individuals involved (Price & Price, 2003; Price, 2008). Currently, the Noir Marrons are not one unique group; they are members of six distinct communities disperse in separated territories. They also present remarkable different features in their history, languages, culture, religion, and social organization (Price & Price, 2003). For instance, each Noir Maroon people has its own understandings of the cosmos and magic world (Price, 2008).

The history of the Noir Marron began with the frequent marooning in Dutch Guiana during the 16th and the early 17th century. Slaves managed to escape during various occasions: during their dismemberment from the ships when a lack of control was exercised; during their transport to the plantations; and once on the farm. The dense equatorial forests in which the Dutch plantations were located, were ideal places to hide, especially as they held all the natural resources to ensure survival. Therefore, many slaves took refuge in the forest and small communities were slowly formed that were able to survive in this unknown environment. Individuals from the same plantation often regrouped (Price & Price, 2003).

The first Noir Marron group to form was the Saramaka. Its size continued to increase until the 1720s, with individuals coming mainly from the plantations controlled by Portuguese Jews, the first planters of Dutch Guiana. Posteriorly in 1710, the Ndjuka group was formed, located in the eastern region of Dutch Guiana. Currently, the Saramaka and Ndjuka groups are both the largest Noir Marron groups between Dutch and French Guiana. A third Noir Marron group known as Aluku was established from fugitive slaves in French Guiana in 1712, and was subjected to these partial collaborations between the Dutch and some Noir Marron groups. Located near the Cottica River, they constituted small communities led by Boni Okilifu, and they often attacked nearby plantations to obtain foods and weapons. The initial distribution of the fourth Noir Marron group, the Paramaka, was the Commewijne region in 1780. Finally, the

other Noir Marron communities, the Matawai and Kwinti, were established to the west of Saramaka territory. Today, the Matawai and Kwinti do not represent a major group in French Guiana (Price & Price, 2003).

Before the slavery was abolished in the French colonies (1848) and in the Dutch colonies (1863). Six Noir Marron communities had been established in the interior of Surinam and close to the political border with French Guiana at the end of the 18th century. The greatest cultural differences were between the Maroons of central Surinam (that means: Saramaka, Matawai, and Kwinti), and those of eastern Surinam and western French Guiana (that means: the Ndjuka, Aluku, and Paramaka). Although these new societies were formed under broadly similar historical and ecological conditions, they vary in everything from language, diet, and dress to patterns of marriage, place of residence, and migratory wage labour. Each community's legal territory is bounded by mountains, rivers, watersheds, and forests (Bilby, 2005).

In 1900, the Noir Marron population size in French Guiana was composed of approximately: 4,000 *Ndjuka*, 4,000 *Saramaka*, 600 *Matawai*, 400 *Aluku*, 400 *Paramaka*, and 200 *Kwinti* (Price & Price, 2003). By the end of the 1990s, an extensive anthropological fieldwork (INSEE, 2000) suggested that those figures required significant modifications, including more than doubling of the total population of Noir Marron. The new census indicated that between 36,500 and 46,500 Noir Marrons were living in French Guiana at that moment. Table 6 summarizes the consensus data reported by Price (2002) for Noir Marron populations residing in Surinam (60.6%) and French Guiana (31.6%). It include out-migration from all Maroon villages all over the interior, toward Paramaribo, the coast of French Guiana, and Netherlands, which is considered the main destination of Maroon diaspora in Europe. In general, the major encouragements to Maroon rapid demographic expansion would seem to be because of economic stability, access to health care facilities, and controlled disease like AIDS and malaria (Price, 2002). The Noir Marron population in the Netherlands jumped as a result of the Surinam civil war (1986-1992), and many of those living there today are illegal immigrants (Price, 2002).

Table 6: Demographic of the Noir Marron communities in Surinam, French Guiana (F.G.), and Netherlands by the end of the 1990s (Price, 2002).

Noir Marron	Surinam interior	Paramaribo	F.G. interior	F.G. coast	Netherlands	Total	Percentage
Ndjuka	24,000	8,000	3,000	11,000	4,500	50,500	*42.94%*
Saramaka	25,000	7,000	-	14,500	4,000	50,500	*42.94%*
Aluku	-	-	3,900	2,000	100	6,000	*5.10%*
Paramaka	2,300	500	500	2,300	400	6,000	*5.10%*
Matawai	1,000	2,900	-	-	100	4,000	*3.40%*
Kwinti	170	400	-	-	30	600	*0.51%*
Total	**52,470**	**18,800**	**7,400**	**29,800**	**9,130**	**117,600**	***100,00%***

Currently, Noir Marron communities present a remarkable African heritage in their culture, language, political organization, and religion; which cannot be easily associate with a unique African source. They are the result of mixed of diverse African groups, and this complex legacy has been maintained for several centuries until nowadays (Price & Price, 2003). Furthermore, they developed a new African identity in the process of creolisation according to their adaptation to a new environment and circumstances, and they integrated elements influenced by European and Native American culture.

All six groups are running their own political and judicial affairs under the authority of paramount chiefs and village captains, which functioned as *"states within a state"*. They are known for such interesting practices as polygyny, oracular divination, spirit possession, body scarification, and ancestor worship, as well as distinctive styles of music, dance, and plastic arts, and countless other aspects of daily life that reflected their uncompromised heritage of independence and their radical difference from the other populations of French Guiana and Surinam. Maroons' dealings with the outside world were largely limited to the men's wage-labour trips, which provided the cash needed to buy soap, salt, tools, cloth, kerosene, kitchenware, and other necessities for life back in the villages of the rainforest. Maroons felt tremendous pride in the accomplishments of their heroic ancestors and, on the whole, remained masters of their forest realm (Price & Price, 2003).

The period of the 1960s-1970s has undergone dramatic transformations and gradual modernization of the world of these people; such as out-board motors that facilitated mobility within and beyond the interior, the construction of airstrips in the interior, radios and tape recorders that allowed closer communication with the coast, gasoline-powered generators in some of the villages that brought electric lights and the occasional refrigerator, and an increase in the missionary schools that prepared boys and sometimes girls for contacts with Creoles and other non-Maroons. All of these changes were monitored by public consensus, and through community meetings (Price & Price, 2001).

In the 1970s, there were even more intense transformations. Surinam moved away from its ties to Europe, becoming an independent republic. In French Guiana, Paris targeted it for rapid development in connection with the establishment of the European Space Center in Kourou. These shifts eventually had profound consequences for Maroons communities in terms of territorial sovereignty, political independence, cultural integrity, and economic opportunities such as basic issues of health and personal dignity (Price & Price, 2001).

Since its independence in 1975, Surinam has been pursuing an increasingly militant and destructive policy against Maroons, stripping them of their rights to land and its potential riches and endangering their right to exist as distinctive people. In 1980, the army seized power in a

coup d'état, and the country began a downward spiral from which it has never recovered. For instance, at that time Surinam endured a plummeting economy, a massive brain drain, and a notable increase in poverty, drugs, and crime (Price, 1995). In 1986, civil war broke out between Maroons and the national Creole-run military, sending thousands of Maroons fleeing across the border into French Guiana. At least 10,000 Ndjukas, recognized as refugees, were confined to camps enclosed by barbed wire; and countless other Maroons (mainly Saramaka), as clandestine attempting to build a new life while remaining invisible to French authorities, charged with the expulsion of illegals. The fighting that raged from 1986 to 1992 pitted Maroons against the national army of Surinam, bringing back to life many of the horrors of their early ancestors' struggles for freedom. African medicine bundles that had lain buried for 200 years were unearthed and carried into battle. Maroon men and boys, often armed with shotguns, confronted the army's automatic weapons, tanks, and helicopter gunships dropping napalm. Whole villages, particularly in the Cottica Ndjuka region, were razed as soldiers killed hundreds of women and children with machetes and bullets (Price & Price, 2001).

In 1992, the civil war was concluded, the refugee camps in French Guiana were shut down, and their occupants were either "regularized" or sent back to Surinam. As for the remaining "illegals", who number in the thousands, the quality of life generally rises and falls with immigration policy decisions made in Paris. Meanwhile, those Maroons who are officially French citizens by virtue of having been born east of the Marowijne (Maroni) and Lawa Rivers, have been adapting to an aggressive program of "Francisation". This assimilationist program disseminates the language and culture of the French state, provides generous welfare benefits, redefines the nature of Maroon political leadership and land ownership, encourages consumerism both in the stores of French Guiana and through European mail order catalogues, and redefines Maroon visual and performative arts as part of the cultural patrimony of overseas France (Bilby, 2005).

In the Surinam post-civil war, Maroon life has been transformed, perhaps irreparably, with rampant poverty and malnutrition. The official restoration of peace in 1992 came at a price, as the Maroons were pushed into signing a treaty largely focused on rights to land, minerals, and other natural resources. The Surinam state has now embarked on a rigorous program aimed at the legal unification, uniformity, and ultimately the integration of Maroon minorities (Price & Price, 2001). However, the overall decline in Surinam's prosperity during the past 20 years has had strong trickle-down effects on Maroon communities. The basic rights of Surinamese Maroons (such as to be free from discrimination, to own and enjoy their lands and natural resources, to participate in decision making, and to practice their cultures) are routinely violated in policy and laws; for instance issuing of logging and mining concessions without any

consultation, environmental degradation, dispossession, and by ignoring of legal agreements such as Maroon peace treaties (Kambel & MacKay, 1999).

In neighbouring French Guiana, the economic and social situation contrasts sharply with that of Surinam, however the threats to the survival of Maroon cultures are equally distressing. French Guiana territory is distinguished strongly for its high diversity ethnic and linguistic than any other French department, particularly for both African American and Amerindian diversity. Nevertheless the presence of Marron communities in French territory represents an important governmental question in France (Belkacemi, 1999). After the slave trade period, French government disagree systematically with international convections or declarations to recognise specific rights to determinate ethnic groups in reason of their membership to minority groups (Gery et al., 2014). For instance, unlike Brazil and Colombia, France has never ratified the Indigenous and Tribal Peoples Convention N°169 (or ILO 169) concerning to:

"Peoples in independent countries who are regarded as indigenous on account of their descent from the populations which inhabited the country, or a geographical region to which the country belongs, at the time of conquest or colonisation or the establishment of present state boundaries and who, irrespective of their legal status, retain some or all of their own social, economic, cultural, and political institutions." Article 1b. Part I (ILO, 1989).

ILO 169 is based on the principle that indigenous and tribal peoples should *"enjoy as much control as possible over their own economic, social and cultural development"*. It recognizes that indigenous and tribal peoples *"have the right to decide their own priorities for the process of development as it affects their lives, beliefs, institutions and spiritual wellbeing and the lands they occupy or otherwise use, and to exercise control, to the extent possible, over their own economic, social and cultural development"* (ILO, 1989). It also contains over six articles on indigenous and tribal land rights, basing these rights on traditional occupation and use of land and resources rather than on grants from the state, and a number of provisions relating to consultation and participation in decision making.

The vast majority of American states have ratified international human rights treaties such as ILO 169 that obligate them to respect the rights of individuals and certain groups. While Maroon groups' rights has been addressed in countries such as Colombia (1999), Ecuador (1997), and Surinam (1983, 1985). The position of French government is even against the European Parliament Resolution (in 1994, Strasbourg) that recognized indigenous peoples' right (such as Noir Marron) to autonomous control over their territory, political status, and culture; and called on host governments (France) to secure indigenous land rights and develop specific measures to protect their rights (Anaya, 2004).

Besides, in Surinam (independent republic since 1975) the constitution of 1987 specifies that all non-titled land and resources belong to the State. The constitution also denies specifically the possibility that an indigenous or 'Maroon people' could have a juridical personality and therefore collective rights to property (or to anything else). For instance, during the 1990s the Saramaka suddenly found their territory invaded by Chinese and other multinational logging and mining companies, which were extracting resources with the explicit permission of the State. Hence, the Maroon communities in French Guiana and Surinam, the most resilient of all Maroon cultures in the Americas, endure ever into the 21st century.

B.1.a.   Formation of the Saramaka community

The ancestors of the Saramaka (or *Saamaka* in their own pronunciation) were slaves on the plantations of Surinam between 1690 and 1760, who fought against their oppressors and freed themselves from the inhumane system of slavery. The white plantation owners (mainly Dutch, French, or Portuguese Jews) undertook many expeditions to try to recapture their enslaved Africans, however these expeditions were mainly unsuccessful. In 1762, they had to close reluctantly an agreement in which they recognized the Saramaka as *"free negroes"*, and the Saramaka could live in the interior in whatever way they pleased and that they would receive annual gifts by which the government; in effect, the government paid 'protection money' to forestall raids on the plantations (Price & Price, 2003).

The Saramaka have a long history with French Guiana, since 1860 many Saramaka have resided temporarily or permanently in French Guiana, and for a long time, their numbers exceeded those of the other Noir Marron groups (Price & Price, 2003). At the early 18th century, the first Saramaka arrived in French Guiana during the period of the first gold rush and they established three centres in French Guiana. The first settlement was established in the small town of Mana alongside the Mana River in western French Guiana. They came to quickly monopolize river transport, and did so for about a hundred years (Figure 20). After several armed conflicts, they forced to the Dutch colons to sign their freedom and independence in 1762. At that moment, they had already created their own culture, language, and social structure (Price, 1994).

Figure 20: Saramaka Maroon village in Suriname River, photo took in 1955. © 1955 Ted Hill.

At the end of the 19ᵗʰ century, many Saramaka also settled on the Approuague River and later in Regina in Eastern French Guiana, then a centre for the exploitation of gold and rosewood. The Saramaka soon took over the river transport on that river and found employment in forestry activities. At its height, there were as many as 400 Saramaka men and the Saramaka maintained a considerable presence in the region until the 1950s. Saint-Georges-de-l'Oyapok was a third Saramaka centre in French Guiana. At round the 1900s, it became an important port of departure for gold searchers headed for the Carsewene region in Brazil. As in the other regions, the Saramaka quickly monopolized the river transport on the Oyapok, and later they also established their own village, Tampak, a few kilometres from Saint-Georges. Many Saramaka men and their French Guianese creole wives settled in that village, which at its height numbered over three hundred people and had the greatest Saramaka ancestor shrine (Price, 1994).

The current geographic distribution of Saramaka in French Guiana is concentred mainly in major cities of Cayenne, Kourou, Saint-Georges, and Saint-Laurent-du-Maroni, as well as in its periphery between Iracoubo and Mana, and in the delta regions of Aprouague and Saint-Georges-de-l'Oyapock. Moreover, there are a considerable proportion of Saramaka in Surinam, and numerous emigrants to Netherlands and United States (Renault-Lescure & Goury, 2009).

B.1.b.   Formation of the Ndjuka community

The current geographic distribution of Ndjuka (also known as *Ndyuka*, *Okanisi* or *Aukaners*) in French Guiana is concentred in Gran-Santi around the Maroni in the region of Saint-Laurent and Mana (Renault-Lescure & Goury, 2009). The first runaways (called *lowe nenge*) found refuge close to Tapanahony in Surinam, and they were in permanent war with the Dutch government (Price, 1996).

In 1760, the Ndjuka, who lived in the east of Surinam, signed an accord with the plantation owners to recognize them as a free group, two years earlier than the Saramaka's agreement with the colonial government. The *Okanisi* or *Aukaners* words came from the name 'Auka', which is the name of the plantation where Ndjuka and the Dutch signed this peace treaty. Although, it is still not clear the origin of the Ndjuka word (Price & Price, 2003).

B.1.c.   Formation of the Aluku community

The Aluku (also known as *Boni*, which is the name of the founding hero of the Aluku, Boni Okilifu) were the first Noir Marron community permanently established in French Guiana. In a historical perspective, the Aluku were probably the greatest victims of colonial repression (Hoogbergen, 1990). During the first years of their formation, they lived on the forest in disperse clans of runways from Surinamese plantations, mainly from Cottica and Commewijne regions (Moomou, 2004). The names of each clan were the same than the name of each plantation of owner that they belonged previously. In 1760, the three major clans (rebels of Cottica, clan Kawina, and clan Dju) established one alliance of freedom, and started a common society. In 1777, Aluku group were moved over the Marowijne into French Guiana after a year-long guerrilla war against the Dutch troops. In 1789, the Aluku again went to war against the Surinamese plantation owners until 1793, and carried out raids on the plantations. A considerable number were killed, including the leader's Aluku, Kormantin Kodjo, who is still considered a symbol of rebellion and untamed soul. In 1793, only about 100 Aluku were left, of the 400 living in 1789. In the 1830s, the remaining Aluku settled on the French bank of the Lawa River, which forms the border between French Guiana and Surinam. All different bands and clans accorded to create an unique community, and at that moment they started the Aluku community (Hoogbergen, 1990; Price, 1996).

The traditional Aluku territory is now divided between three major communes: Maripasoula, Grand-Santi-Papaïchton, and Apatou. Aluku society is divided into several named matrilineal clans, which means a woman's children automatically belonged to her clan by birth. These clans, called *lo*, formed the basic unit of Aluku social organization. Most Aluku *lo* were concentrated in single villages founded by clan members, however some *lo* were divided between two or three villages. Each village had its clan chiefs, called *kapiten*, responsible for the members of their *lo*, as well as for any other villagers happening to live there. The *kapiten*, in turn, answered to the *gaanman*, the paramount chief, whose authority extended to the entire tribe (Price & Price, 2003). Daily social life, however, is regulated primarily at the village level, with the *kapiten* playing an important part in most public transactions (Price, 1996).

In 1989, French Guianese authorities launched in Cayenne the program entitled *"Sur les traces de Boni"* (or "On the trail of Boni") in conjunction with the bicentennial of the French Revolution. The program officially recognized the liberation struggles of the Aluku ancestors, and aims to foster widespread appreciation of the Aluku cultural heritage (culture, art, dance, and music). Indeed, the Aluku had a rich repertoire of songs, dances, and drumming styles to accompany the major rites and crises of life (Price & Price, 2003). As other Noir Marron communities, the Aluku play the drum-based genre known as *aleke*, based on a three-drum ensemble with other percussion (Bilby, 2001). According with musicological analysis, there is a close resemblance between the interlocking rhythmic patterns played by drummers in Aluku ceremonies in French Guiana and Surinam, and those characteristic of certain drum ensemble genres still played in Akan-speaking groups from Gold Coast (Bilby, 1995, 2009). However, we lack historical documentation musical forms existed in the Guianas during the 17th, 18th or 19th centuries, that could allow us to confirm that association through musicological analysis based on written notation.

B.1.d. Formation of the Paramaka community

The Paramaka (or *Paamaka* in their own pronunciation) is a small community originated after other major Marron groups. Originally, Paramaka distribution was the Commewijne region in 1780, far away enough to avoid on the one hand European colonizers, and on the other hand the Ndjuka group. They moved slowly to the south, where they established new villages. Their presence in Maroni Islands such as Langatabiki dates back to 1863, year of the abolition of slave trade in Surinam. Through the 19th century, they created their final distribution between the Ndjuka community to the north and south, and the Saramaka community to the west. Currently, the Kwinti is direct descendent of Paramaka community (Price and Price, 2003).

B.2. Formation of the Afro-Brazilian communities

About 44% of all Africans forced into the slave trade ended their lives in Brazil (Figure 1). Currently, Brazil has the largest population of Africans outside of Africa, generally known as Quilombos (or *Mocambos*). *Palmares* is the most famous of all Quilombos. Indeed, *Palmares* is a federation of Maroon communities whose population was estimated to be 20,000 and even 30,000 people (Schwartz, 1995). Its several constituent settlements were located on the *Serra da Barriga*, a mountain chain in the periphery of what was then the captaincy of Pernambuco, in the northeast of Brazil, an area that now belongs to the state of Alagoas. *Palmares* itself was originally created in the late 16th century by rebellious slaves from a large sugar plantation near

Porto Calvo, on the coast of Pernambuco. Fugitives usually killed or maltreated masters, overseers, and members of their families; burned fields; and stole arms, ammunition, and foodstuffs before escaping to the woods, swamps, or mountains (Schwartz, 1992).

In the mid-17th century, the population of *Palmares* grew from both natural reproduction and the incorporation of such slaves who had been liberated from sugar plantations by Quilombo raiders. In the long run, *Palmares* became the prototype of the Quilombo in Brazilian historical narratives (Anderson, 1996; Schwartz, 1995). Besides, Quilombos were also founded in more remote regions of Brazil, located next to Minas Gerais, Goiás, and Mato Grosso. During the 18th century gold rush, the mining region of Minas Gerais was the setting for the formation of dozens of Quilombos of between 100 and 300 inhabitants each. They were working for independent small-time miners, and merchants to buy their gold or exchange it for foodstuffs, firearms, ammunition, and other products (Schwartz, 1992).

Conversely to Surinam, French Guiana, and Colombia, where Maroon communities forced treaties with colonial governments, became politically autonomous, and persist into nowadays; in Rio de Janeiro, slave flight did not always lead to the formation of Quilombos societies. Brazilian cities such as Rio de Janeiro, Recife, Salvador, Porto Alegre, and Vila Rica (present-day Ouro Preto) were surrounded by small and large bands of fugitive slaves, whose mud houses, campsites, and subsistence grounds were periodically raided and destroyed by the police, only to reappear later on. Fugitives often escaped individually or in small groups and disguised themselves as freed blacks or mestizos, especially in larger urban settlements. They sometimes settled in remote areas, however unlikely in most cases were located near large urban centres or near plantation, ranching, and mining zones, from which they could extract part of their subsistence. Runaway enslaved Africans frequently dedicated themselves to subsistence agriculture; although they also sold their excess production in nearby local markets, or sold their labour to local planters, farmers, and miners (Gomes, 1995).

Interestingly, in Brazil the formation of Quilombos did not always mean a complete withdrawal from captivity. Many rebel slaves organized themselves in Quilombos to negotiate from a position of force to obtain better terms of labour and living under slavery. For instance, Maroons from the Santana plantation in Bahia even produced a detailed *"peace treaty"*, as they called it, consisting of several demands relating to the work routine and to bargain for better terms of bondage (Anderson, 1996).

In the year Brazil celebrated the centennial of the abolition of slavery, the Brazilian Constitution of 1988 for the first time recognized the right of descendants of slave-era, Quilombos, to receive lands from the state: *"The definitive property rights of remnants of Quilombos that have been occupying the same lands are hereby recognized, and the state shall*

grant them title to such lands." (Art. 68). Further, the Brazilian Ministry of Culture established the *Fundação Cultural Palmares* (or FCP) dedicated to *"promoting the cultural, social and economic values resulting from black influence in the formation of Brazilian society"*. One of the FCP's specific actions is to carry out research, studies, and surveys about Afro-Brazilian cultural legacies to preserve the Afro-Brazilian patrimony. The FCP also is in charge of protecting the legal rights of Quilombos and pulling together the documentation necessary to support their historical justification (Landers et al., 2015).

In 1988, when these rights were finally recognised, Brazil was emerging from two decades of military rule. The new constitution was marked by a will to fully include the various minorities that until then had been excluded from citizenship. In fact, the recognition of Quilombo remnant communities formed part of this wish for a renewed vision of Brazilian society. As the country was celebrating the centennial of abolition, the aim was to rehabilitate the historical resistance of its black population. However, today only about thirty of remnant African communities have been recognized by the Ministry of Culture, of more than 700 that have been identified throughout Brazil, and just a small minority have received their land titles (Price, 1998).

B.3. Formation of the Afro-Colombian communities

The formation of an Afro-Colombian identity built upon a tradition of resistance can be attributed to a continuous history of slave rebellions and to the existence of independent Maroon communities called *Palenques*, formed by escaped enslaved Africans (Jordan, 2004). The term *Palenque* is translated literally in English as *'stockade'*. The strength of these free communities in Colombia was centred on their capacity: to increase in number; to maintain market relations with neighbouring communities; to exercise control and dominion over immediate agricultural, hunting, and fishing territories; to receive fugitive slaves from mines and haciendas; and to maintain autonomy against slave-owners and authorities wishing to reduce them to the status of illegal squatters and vagabonds (Navarrete, 2011).

The discoveries of gold, silver, and later platinum attracted miners to the Chocó region, who brought large numbers of enslaved Africans to extract the precious metals. Gold-seeking raiders killed hundreds of natives, burned native villages, attempted to establish fortified settlements, and were repeatedly driven away. Therefore, enslaved Africans replaced the dramatic decline among the native populations (Landers et al., 2015). Nowadays, the Chocó region has the second largest population of African descendants in Colombia after *Palenque de San Basilio* community, a small town on the Colombian Caribbean coast. The almost exclusive African community in San Basilio de Palenque, generally known as *Palenque(ro)* or *Palenque de San*

Basilio, is located about 60 Km inland from the city of Cartagena de Indias. This Maroon community, established between 1655 and 1674, successfully managed to resist multiple attacks by Spanish militia. It was the first official *Palenque* in the Americas declared as a free village in 1713, after the King of Spain abandoned efforts to attack the walled village and to recapture slaves (Ferrari, 2012). Today, this former Maroon community speaks Spanish and their own unique language called *Palenquero*, a type of Spanish-derived Creole language that combines Spanish and Bantu languages from West-Central Africa such as Kikongo and its dialects (Schwegler, 2006, 2011).

Many of the formerly enslaved in Quibdó purchased their freedom with gold mined on days off or stolen from their owners, and by the 18th century the Chocó department was home to a large free population of African descent. In the middle of the 18th century, gold mining in the Antioquia region reached greatest productivity. Enslaved Africans imports grew, and also did African creole populations. The acute demand for labourers allowed many enslaved Africans to rent themselves to someone other than their master on off days in exchange for cash. Ultimately, money accumulated was used to purchase their own freedom or that of family members. Indeed, hundreds of enslaved Africans in Colombia bought their freedom, and started migrating to places such as the Baudó Valley, where they formed largely black towns with cultural and social characteristics similar to the *Palenque* towns established by runaway enslaved Africans. (Landers et al., 2015; Navarrete, 2012).

Finally, slavery was abolished in Colombia in 1851. According to the last census (DANE, 2005), today in Colombia the percentage of people who self-identify as African descendant is around 11%, which includes people called: Afro-Colombian, African descendant, black, *mulato(a)*, *negro(a)*, *Palenquero* (or Afro-Colombian from San Basilio de Palenque), and *Raizal* (or Afro-Colombian from archipelago of San Andrés, Providencia, and Santa Catalina). It means that the current Afro-Colombian population is about 4,395,649 people of 41,468,384 total inhabitants in Colombia. However, the distribution of African descendants in Colombia is very uneven. For instance, 95.32% was estimated in Quibdó, 56.98% in San Andrés, 36.47% in Cartagena, and 6.48% in Medellín, while 1.5% was estimated in Bogotá D.C. (DANE, 2005).

C.   Linguistic influences in the formation of Marron Creole languages

C.1.   Africa's linguistic diversity

Africa populations not only have the highest levels of genetic diversity in the world, but also a considerable amount of linguistic diversity. Currently, about 2,138 living languages exist in Africa, representing about a third of the world's languages (Paul et al., 2015). They are classified into four major linguistic families (Figure 21): *i*) *Niger-Kordofanian* spoken primarily by agriculturalist populations located in large contiguous regions of sub-Saharan Africa from western Africa to eastern and southern Africa, and Bantu language is the major linguistic subfamily within this phylum; *ii*) *Nilo-Saharan* spoken predominantly by pastoralist populations in central and eastern Africa; *iii*) *Afroasiatic* spoken predominantly by pastoralist and agro-pastoralists populations in northern and eastern Africa; and *iv*) *Khoesan* a language family that contains click consonants spoken by hunter-gatherer San populations in southern Africa as well as the Hadza and Sandawe hunter-gatherers in Tanzania (Campbell et al., 2014; Gomez et al., 2014; Scheinfeldt et al., 2010). Furthermore, there are other non-African languages such as *Indo-European* and *Malayo-Polynesian* that are resulting for ancient interactions with other non-African populations, such as European, Ottoman, Semitic, Arabian, and Polynesian groups (Figure 21).

It is estimated that there are more than 600 Bantu languages spoken in sub-Saharan Africa, making this one of the largest and most widespread language group in the world (Paul et al., 2015). Indeed, modern Bantu-speakers are numbering over 220 million people (that means ~28% of Africans), who spread now over almost 9 million Km². The Bantu languages are generally divided into three major groups: north western Bantu (with subgroup A, B, and C), eastern Bantu (with subgroup E, F, G, J, N, P, and S), and western Bantu (with subgroup H, K, L, R, D, and M) (Currie et al., 2013).

The most outstanding migration event in Africa is considered the geographic expansion of Bantu-speakers across East and South Africa from their homeland in the Cross River Valley, near southeast Nigeria and western Cameroon border, where the Bantoid languages are most closely related to narrow Bantu (Hyman & Hombert, 1999). The expansion of Bantu-speakers took two main routes from its starting point: *i*) a western route, throughout the west coast of Africa, having arrived to Angola, South Africa and Botswana around 3.5 kya; and *ii*) an eastern route, towards the Great Lakes in Eastern Africa, reaching the region of Uganda about 2.5 kya, where they remained for a couple thousand years, expanding later into the south, reaching Mozambique by ~1.8 kya (Diamond & Bellwood, 2003; Salas et al., 2002).
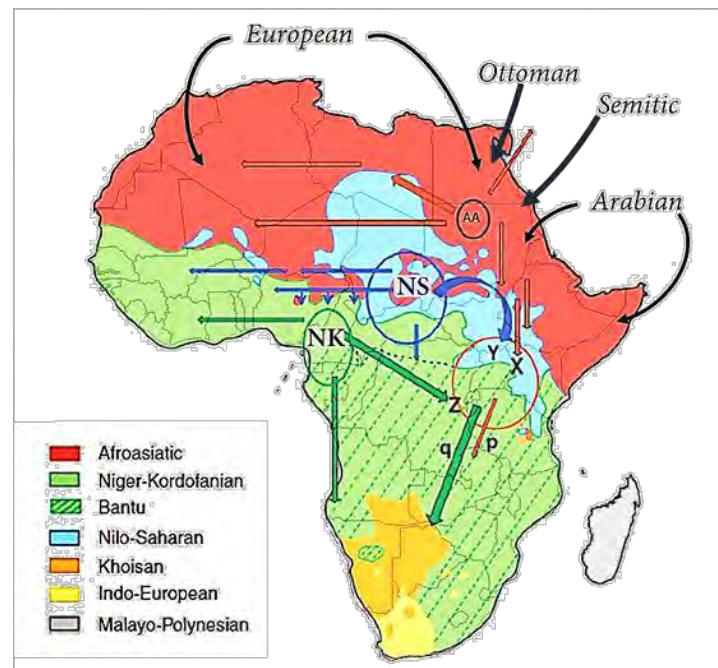
Figure 21: Geographic distributions of major linguistic phylum in Africa. Geographic range occupied by Bantu speakers is shown. Putative centres of origin and initial expansion of language families are also shown: **AA**, *Afroasiatic* (14 kya, thousands years ago); **NS**, *Nilo-Saharan*; and **NK**, *Niger-Kordofanian* (5 kya). Afroasiatic-speaking pastoralists were the first food-producing populations to migrate into East Africa circa 5 kya (X); followed by Nilo-Saharan-speaking pastoralists circa 3 kya (Y); and later Bantu-speaking agriculturalists after 2.5 kya (Z). Ultimately, after 2 kya initial expansion to southern Africa from East Africa of pastoralists (p) and later Bantu speaking agriculturalists (q). Image modified from Gomez et al. (2014).

The spread of West African people across sub-Saharan Africa changed the demographic, linguistic, and genetic scenario for generations until today. Geneticist, archaeologists, and linguistics, all point to the extensive distribution of this language family being the result of a population dispersal that began around 3,000–5,000 years ago (Campbell et al., 2014; Currie et al., 2013; de Filippo et al., 2012; Ehret, 2001; Grollemund et al., 2015). However, there is less agreement about the routes taken by Bantu groups as they spread out over the rest of the continent (Barbieri et al., 2013a; Li et al., 2014; Marks et al., 2015; Russell et al., 2014). These debates have crucial implications regarding to the origin and spread of important cultural innovations, such as metallurgy, cattle-keeping, and farming (Holden, 2002; Pakendorf et al., 2011).

To understand this extremely high African cultural and linguistic diversity is a key point to shed new light on the African origin of the captives who embarked from one particular African region during the slave-trade era. However, it is very complex to distinguish between the geography of major linguistic families and of historic migration events took in Africa during

that period. For instance, in French and Dutch Guiana, historic slave trade records indicate the numerical dominance of: Bantu speakers (such as Kikongo from West Central Africa) and Gbe speakers in the late 17th century; only Gbe speakers in the early 18th century; Akan speakers in the 1720s and 1730s; and 'Upper Guinea' (and to a lesser extent Bantu) speakers thereafter (Huttar, 2010). Indeed, Smith (2002) showed that during 1675-1714, the formative period of Creole languages, about 50% of the enslaved Africans in the Guiana regions were originated from the Gold Coast and the Bight of Benin. Therefore, Gbe varieties from West Africa (such as *Fongbe*), which come from the Gold Coast and Bight of Benin, were the base language in the formation of Creoles languages spoken today in French Guiana and Surinam rather than Bantu or other linguistic groups (Essegbey et al., 2013a; Huttar, 2010; Lefebvre, 2013).

C.2. English-based Creole's linguistic diversity

The English-based Creole languages known as Ndjuka, Saramaka, Paramaka, and Aluku (language name is the same as Noir Marron tribe that belonged) include lexical items from many distinct African and European sources (Huttar, 2012; Smith & Cardoso, 2004). The Saramaka established the first creole language in Surinam (Renault-Lescure & Goury, 2009), which is characterized for a high influence of Portuguese words (language of Portuguese Jewish masters) (see Table 7). The English language was introduced in Surinam between 1650 and 1667, during the British colonization (Huttar, 2010). Currently, Noir Marron groups present a lot of characteristics in their culture and language (Price, 2001). For instance, the linguistic similitudes between Saramaka and Ndjuka languages are very close, as for the linguistic similitudes between Aluku, Paramaka, and Ndjuka languages (Huttar & Huttar, 1994; Leglise et al., 2013). Interestingly, none Noir Marron group spoke *Taki-taki*, which is the mother tongue of diverse ethnic groups in Surinam such as Afro-Surinam Creoles known as *Sranan* and *Tongo* (Price & Price, 2003).

In the course of the past decade, an increasing number of historical linguists have used newly-developed classification techniques, borrowed from biology techniques, in order to provide a classificatory tool, which helps uncover the evolutionary histories of Creole language types. Several attempts at classifying creoles languages based on shared features have been undertaken (Baker, 1999; Baker & Huber, 2001; Hancock, 1987). However, none have successfully achieved to present a clear picture of relationships between these languages, mainly because of the extremely complex sociohistorical conditions under which creole languages were developed.

FRANÇAIS	SARAMACCAN		NDYUKA/ALUKU/PARAMAKA	
	dérivé de		*dérivé de*	
bouche	buka	port. *boca*	mofu	angl. *mouth*
main	mau	port. *mão*	ana	angl. *hand*
homme	womi	port. *homem*	man	angl. *man*
femme	muyee	port. *mulher*	uman	angl. *woman*
fleuve	lio	port. *rio*	liba	angl. *river*
dormir	duumi	port. *dormir*	siibi	angl. *sleep*
boire	bebe	port. *beber*	diingi	angl. *drink*
cuiller	kuyee	port. *colher*	supun	angl. *spoon*
autel des ancêtres	faaka pau	angl. *flag* + port. *pau*	faaka tiki	angl. *flag* + *stick*
eau	wata	angl. *water*	wata(a)	angl. *water*
maison	osu	angl. *house*	osu	angl. *house*
aller	go	angl. *go*	go	angl. *go*
venir	ko	angl. *come*	kon	angl. *come*
dents	tanda	holl. *tand*	tifi	angl. *teeth*
joli	hanse	angl. *handsome*	moi	holl. *mooi*
viande	gwamba	(origine précise inconnue) [afr.]	switi mofu	angl. *sweet mouth*
manger	nyan	(diverses langues africaines)	nyan	(diverses langues africaines)
tambour parlant	apinti	akan [afr.] *apentemma*	apinti	akan [afr.] *apentemma*
récipient couvert	apaki	twi [afr.] *apákyi*	pakiba	twi [afr.] *apákyi*
pierres à foyer	makuku	kikoongo [afr.] *makukwa*	makuku (vieilli) ou doti dii futu	angl. *dirt three feet*
calebasse	kuya	tupi [amérind.] *kuia*	kaabasi	angl. *calabash*
presse à manioc	matapi	kali'na [amérind.] *matapi*	matapi	kali'na [amérind.] *matapi*
tablier de jeune fille	koyo	trio [amérind.] *kwayu*	kwei	kali'na [amérind.] *kwai*

Table 7: Origin of several words from Creoles languages spoken for Noir Marron and their influence from European (English (angl.), Duth (holl.), and Portuguese (port.)), Amerindian (amérind.) and African (afr.) languages (Price & Price, 2003).

To unravel questions about evolutionary histories of Creole language types, Daval-Markussen & Bakker (2011) provided a classification of English-based creoles based on a selection of lexical and typological features encoded as binary pairs ("1" versus "0"). These data was used to provide a classification according to the estimated evolutionary scenario for the development of English-based creoles. This approach adopted the network-based method of split-decomposition for computing phylogenies (Bandelt & Dress, 1992) and was applied with the software SplitsTree (Huson & Bryant, 2006), which has been developed to estimate evolutionary splits or phylogenetic network based on the degree of similarity between biological species.

Based on that approach, Daval-Markussen & Bakker (2011) studied the relationships between 33 Atlantic English-based creoles: Saramaka (Sar), Boni or Aluku (Bon), Paramaka (Par), Ndjuka (Dju), Matawai (Mat), Kwinti (Kwi), Sranan (Sra), Guyana (Guy), Bahamas (Bah), Sea-Islands Creole English or Gullah (Sea), Afro-Seminole (Afr), Providencia (Pro), Belize (Bel), Cayman (Cay), Jamaica (Jam), Saint-Thomas (StT), Saint-Eustatius (StE), Saba (Sab), Saint-Kitts (StK), Antigua (Ant), Saint-Vincent (StV), Carriacou (Car), Grenada (Gre), Tobago (Tob), Barbados (Bar), Trinidad (Tri), Liberia (Lib), Cameroon (Cam), Nigeria (Nig), Krio (Kri), two Pacific varieties Norfolk (Nor), American Black English (Bla), and Hawaii (Haw). The geographical locations of these languages represent the most extensive database on English-based creoles (Figure 22).

Figure 22: Map of geographic distribution of Atlantic English-based creole languages in the Americas, except Lib, Cam, Nig, Kri, Nor, Bla, and Haw (Daval-Markussen & Bakker, 2011).

In that study, they reported different phylogenetic networks to reflect the most likely evolutionary developments and affinities among English-based creoles (Figure 23 and 24). Interestingly, there are three main consistent clusters reflecting affinities among English-based creoles in the split network reported (Figure 23). The most distant and best differentiated group

is made up of the creoles of French Guiana and Surinam (clustered as Sar/Mat, Dju/Bon, Kwi/Par, and Sra; the latter branching off further away from the rest of the cluster). The languages in this group have been shown to be highly related historically (Price, 1996; Price & Price, 2003). The second group can be decomposed into four less obvious subgroups: the restructured vernaculars of West Africa (Kri/Cam/Nig), the North American varieties (Sea/Afr), the Eastern Caribbean cluster (StK/Tob and Ant/Guy), and the Western Caribbean cluster (Pro/Bel/Jam/StV and Bah). The third group is made up of five clusters: the Pacific cluster (with vernaculars Haw/Nor), two Eastern Caribbean clusters (Bar/Tri and Gre/Car), and two other clusters seemingly unrelated (Sab/Cay/StE and StT/Bla/Lib). This unrooted graph estimated the most feasible evolution of the different languages relatively to one another, and therefore better reflected the interactions that may have occurred in the course of their respective histories.
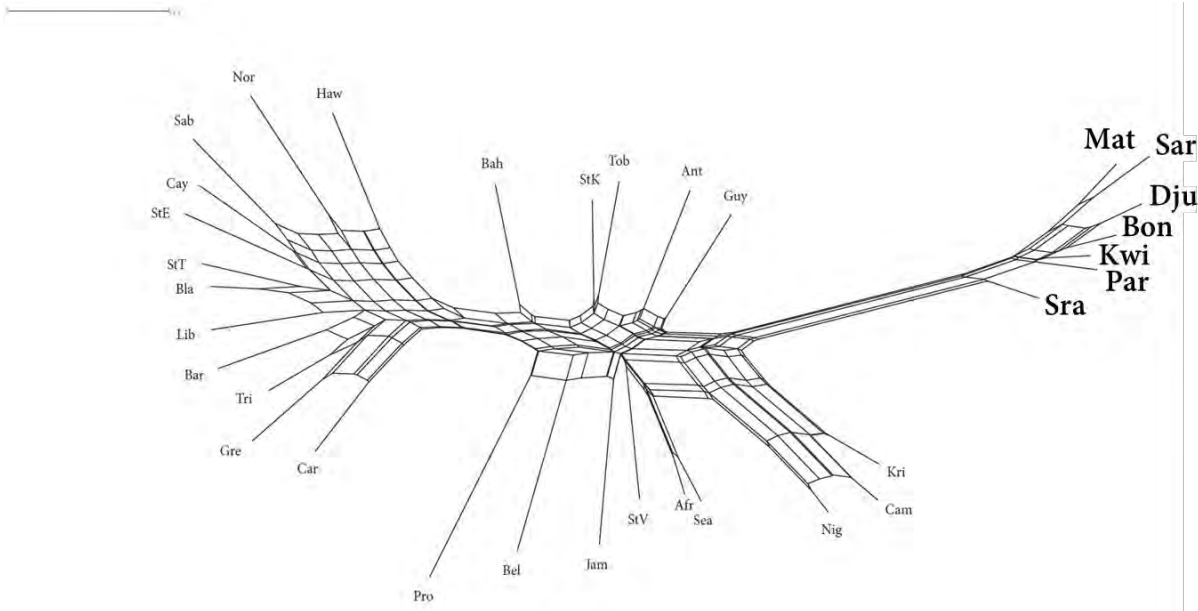
Figure 23: Phylogenetic network for form and structure of principal creoles languages in the Americas. Image modified from Daval-Markussen & Bakker (2011).

In that study, 122 linguistic features were used as dataset, which were classified and divided according to their formal properties (such as historical, phonetics/phonology, semantics, and specific forms) and structural properties (such as categories, morphology, nominal syntax, sentential syntax, tense, mood, and aspect). Those features were coded as binary oppositions ("1" versus "0") (Daval-Markussen & Bakker, 2011). Furthermore, in order to root the network, English language (Eng) was included in the dataset as outgroup (Figure 24). As expected the most European Creoles varieties appeared near the root of the tree with English, and languages belonged to the Noir Marron communities appeared noteworthy far away from the root in a differentiated branch (Daval-Markussen & Bakker, 2011).
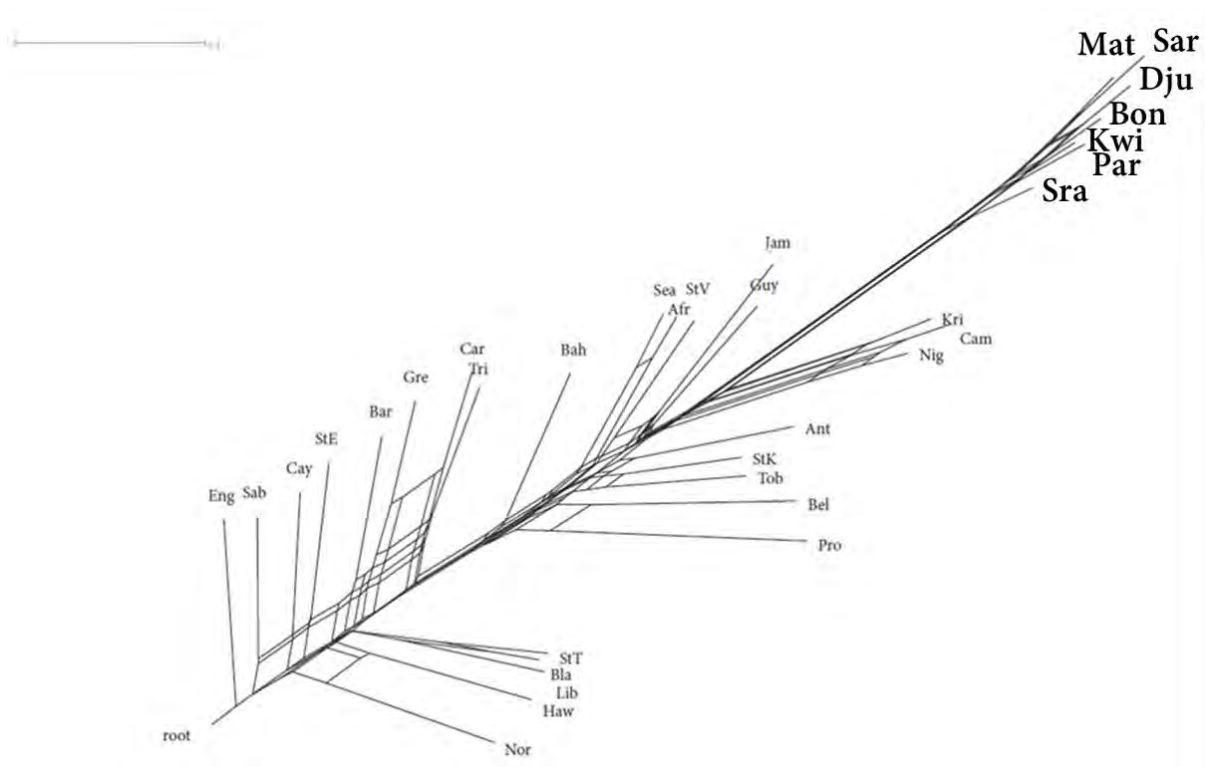
Figure 24: Rooted phylogenetic network for form and structure of principal creoles languages in the Americas. Image modified from Daval-Markussen & Bakker (2011).

Therefore, Marron Creole languages spoken today in French Guiana and Surinam are the most widely accepted branching of the Atlantic English-based creoles, and retained more African identity than any other linguistic group (Bakker et al., 2011; Daval-Markussen & Bakker, 2011). This split can also be explained based on the sociohistorical background of these languages. While the speakers of Sranan are the descendants of the slaves brought in by the British, principally from Barbados, who firmly established in Surinam within the first sixteen years of the settlement of the colony and remained on the plantations (Arends, 1995); the ancestors of the speakers of the Noir Marron groups were runaway enslaved Africans who had escaped from plantations in the 17th and the 18th century (Price, 1996).

The distribution of these languages in both phylogenetic networks implied that the pair Saramaka and Matawai (Sar/Mat) developed separately from the other subgroups, however using a similar original linguistic input, which is supported by the historical scenario (Smith, 1987). The clusters Ndjuka and Boni (Dju/Bon) and Kwinti and Paramaka (Kwi/Par) suggested a parallel development from a common source, on a par with the last member of the group Sranan (Sra). Linguistically, the presence of these subgroups can be explained as reflecting the influence of Portuguese as well as the African elements in these languages (Smith, 1987). The chronology of this group is partially reflected in the phylogenetic networks. In the beginning of the 18th century, Saramaka, and its offshoot Matawai, separated from other Marron Creoles.

Developed in the late 18th century, remnant runaway enslaved Africans were re-joined by other fugitive bands, some of whom founded the Ndjuka group, Aluku, Paramaka, and Kwinti (Price, 1996; Smith, 1987).

Although Sranan was probably not the first Creole vernacular spoken in Surinam, it appears further from the other members of the group (Figure 23) and nearer the root (Figure 24). It is most likely that Sranan developed on the plantations in continued contact with English (Arends, 2001). It resulted in a less abrupt creolization than its closest relatives, which were more rapidly cut-off from their lexifier (Bakker et al., 2011; Daval-Markussen & Bakker, 2011).

D. *Genetic systems to study ancestry and admixture in the New World*

Genetic ancestry profiles of human populations are a valuable tool to understand the dynamics of migration and colonisation events, as well as to determine admixture patterns of populations. Historically, some of the first studies of genetic admixture at the molecular level were those that analysed the frequencies of different blood group protein alleles in African Americans, comparing them to European Americans and Africans (Glass & Li, 1953; Glass, 1955). Nowadays, genetic data has proved to be a useful way of supplementing the inadequate historical documents regarding to the origin and ethnicity of enslaved Africans (Shriver & Kittles, 2004). Indeed, genetic data is increasingly being used to reconstruct ancestral origins and to identify familial ties, even when they extend back for hundreds of years (Novembre & Ramachandran, 2011; Underhill & Kivisild, 2007). Heritage is, however, too complex to be reduced to simple genetic sequences. With respect to the transatlantic African diaspora, genetic data are proving to be important complement to historical, linguistic, ethnographic, and archaeological data in family tree reconstructions (Royal et al., 2010; Stefflova et al., 2011). In conjunction with data from other disciplines, genetic data can give more robust information on: *i*) the African diaspora and African migrations; *ii*) evidence for gene flow between Africans groups and non-Africans groups; and *iii*) evidence for genetic drift and in some cases founder effects (Jobling et al., 2014).

Historical records indicate that enslaved Africans came from various regions throughout coastal African regions, and with the advent of modern anthropological genetics techniques we have now the opportunity to determine their specific regional ancestries with more accuracy. Molecular genetic studies have been used to trace African regional origins of many of their descendants, and to reconstruct the proportions of ancestry derived from different coastal African regions, and within the African continent (Bryc et al., 2010a; Ely et al., 2006; Shriver

& Kittles, 2004). For this approach, genetics and molecular anthropologists employ the more varied types of genes and genetic systems to reconstruct the African ancestry of African descendants, which allow stronger regional associations with specific African groups. For instance, previous genetic studies tracing the genetic identity of the Noir Marron communities in French Guiana have shown that these African-American descendant communities still preserve a remarkably high conservation of the African gene pool in all their genetic systems: 99% in their mtDNA, and similar high percentages in the Y chromosome, GM polymorphisms of immunoglobulins, and also in their human T-cell lymphotropic virus (HTLV) types (Brucato et al., 2009; Brucato et al., 2010).

Nevertheless, there is still a significant lack of knowledge on the ancestry of African-American populations in South America (see review in Fortes-Lima & Dugoujon (2015)), especially when the disparity of markers and sampling criteria that have been used in different publications is taken into consideration (Bortolini et al., 1995; Bortolini et al., 1999; Salzano & Bortolini, 2002). Additional studies, using larger random samples and new populations, can be very helpful in determining new aspects of the genetic history of African Americans, particularly for populations in South America.

There are two pieces of the human genome that are quite useful in enlightening the human history: the mitochondrial DNA and the Y chromosome. These are the only two parts of the genome that are uniparental transferred to the descendants, and also they are not jumbled up by the evolutionary mechanisms that generate diversity with each generation. Instead, these elements are passed down intact (Kundu & Ghosh, 2015). So, by studying the genomic variation in these genetic systems among different population, and compare them with biparental markers, we can trace historic footprints of modern human populations, particularly human migration routes result of the transatlantic slave trade.

D.1. Y chromosome genetic system

The Y chromosome is one of the smallest chromosomes in the human genome (about 60 Mb) and represents around 2 or 3% of a haploid genome. Cytogenetic observations based on chromosome-banding studies allowed different three Y regions to be identified: the PseudoAutosomal Region (or PAR, divided into two regions: PAR1 and PAR2), the euchromatic region (~30 Mb), and heterochromatic region (~30 Mb) (Quintana-Murci & Fellous, 2001).

The PAR1 is located at the terminal region (~2.6 Mb) of the short arm (Yp), and the PAR2 (~0.32 Mb) at the tip of the long arm (Yq). Both regions are located where the Y chromosome pairs exchange genetic material with the pseudoautosomal region of the X chromosome during male meiosis, particularly PAR1. The PAR derives from the ancient origin of the mammalian sex chromosomes as a pair of homologous autosomes some 300 million years ago. Consequently, genes located within the PAR are inherited in the same manner as autosomal genes in every male meiosis (Quintana-Murci & Fellous, 2001). While PAR1 and PAR2 represent the 5% of the entire chromosome, the majority of the length of the Y chromosome (95%) is made by the so-called Non-Recombining portion of the Y chromosome (or NRY). The NRY includes the euchromatic and heterochromatic regions of the Y chromosome. The heterochromatic region is considered genetically inert, although the euchromatic region has numerous highly repeated sequences (Figure 25), and also contains some genes (such as gene *SRY*) coding for 27 proteins that are responsible of important biological functions.

Figure 25: Idiogram of a G-banded Y chromosome shows the distribution of the variant positions across the MSY region. The region most densely populated by Y-SNPs (chrY: 2,880,000-20,900,000) is shown in red (Scozzari et al., 2014). Modified from UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly (http://genome-euro.ucsc.edu/index.html).

The process of divergence of sex chromosomes started when the Y chromosome acquired a male sex-determination function (such as spermatogenesis) early in mammalian evolution, followed by a repression of recombination (and the inevitable disappearance of the Y chromosome itself) (Bachtrog, 2013). Since then, much of the common ancestral gene content has been lost in the Y chromosome, and remaining genes have been amplified into multiples copies. In contrast to the lack of genes, the Y chromosome is enriched for many different types of repeats, including long and short interspersed nuclear elements (or LINE and SINE, respectively), and highly diverse Short Tandem Repeats polymorphisms (or Y-STRs, also known as microsatellite polymorphisms) (Figure 26). Interestingly, the NRY retains a record of the mutational events that have occurred along male lineages throughout evolution, which has not only demonstrated to be extremely informative in disentangling the history of human populations but it also has essential biological roles that make this chromosome an important component of the human genome (Jobling et al., 2014; Quintana-Murci & Fellous, 2001).

Indeed, the Male-Specific portion of the NRY (or MSY) is enriched in intra- and inter-chromosomal segmental duplications, and displays extensive structural rearrangements even in comparison with our closest living relatives, the great apes (Hughes et al., 2010). Additionally, the MSY is an invaluable tool to investigate many issues about population history (Francalacci et al., 2015; Jobling & Tyler-Smith, 2003) and forensic genetics (Ballantyne et al., 2012; Jobling, 2001). Both the lack of meiotic recombination and the uniparental inheritance imply that the MSY differentiation may only be generated by the sequential accumulation of new mutations along radiating male-borne lineages (Underhill & Kivisild, 2007). This process creates monophyletic and evolutionary stable entities known as haplogroups, defined by biallelic markers such as Single Nucleotide Polymorphisms (also known as SNPs). Y-SNPs markers (Figure 26) are sometimes referred to as Unique Event Polymorphisms (or UEPs; also known as Unique Mutation Events, or UMEs) because of they have a much lower rate of mutation than Y-STRs ($\sim 10^{-8}$ versus $\sim 10^{-3}$ mutations per generation) (Ballantyne et al., 2014; de Knijff, 2000). Therefore, Y-SNPs are characterized for a low mutation rate (Wilson Sayres et al., 2014; Xue et al., 2009), which can be arranged in an unambiguous maximum parsimony phylogenetic tree (Karafet et al., 2008; van Oven et al., 2014).
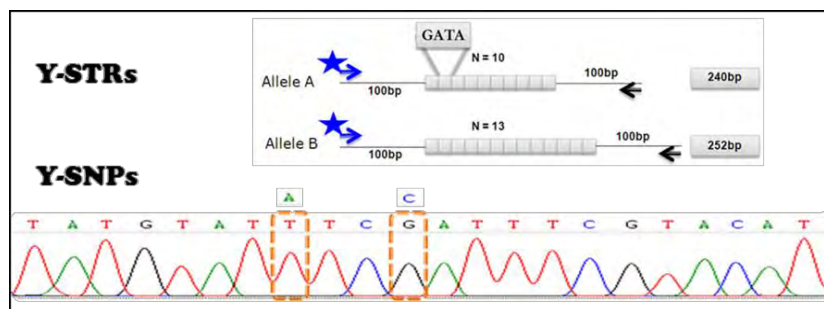
Figure 26: Molecular differences between Y Short Tandem Repeats (or Y-STRs) and Y Single Nucleotide Polymorphisms (or Y-SNPs). The number of repeats is variable among Y-STRs, while in Y-SNPs is variable the type of nucleotide.

The first biallelic marker found on the Y chromosome was an *Alu* insertion (DYS287), abbreviated YAP for Y chromosome *Alu Polymorphism*, which is associated with haplogroup E. YAP occurs in high frequencies within African populations and is rare in most non-African populations (Hammer, 1994). In 1992, Roewer & Epplen (1992) described the first polymorphic Y chromosome marker, Y-27H39, known now as the STR locus DYS19 (which means D; DNA, Y; Y chromosome, S; segment, and number of locus). In 1997, the European forensic community settled on a core set of Y-STR markers useful to identify males and male lineages, which was called "minimal haplotype" and includes DYS19, DYS389I/II, DYS390, DYS391,

DYS392, DYS393, and DYS385 a/b (Kayser et al., 1997). Haplotype was defined as a specific segment of DNA sequence that is inherited in blocks as a unit. Since the first discovery, hundreds of microsatellites have been identified from the reference sequence (GRCh37/hg19). The rapid growth in the discovery of new Y-STR markers is a direct result of the availability of DNA sequence information from the Human Genome Project and improved bioinformatics tools for searching DNA sequence databases. Besides, genome sequencing projects have led to the discovery of many thousands of Y-SNPs (Wei et al., 2013). Indeed, information on the paternally inherited Y-STRs and Y-SNPs has been extensively applied in population genetics and evolution studies to track male specific movements and admixture as well as mating behaviour, particularly between African populations (Ansari Pour et al., 2013; de Filippo et al., 2011).

A comprehensive description of genetic diversity of the MSY (Figure 27), through high-coverage next-generation sequencing, has been proposed for thousands of worldwide Y chromosomes phylogeny (Hallast et al., 2015; Karmin et al., 2015; Poznik et al., 2013; Trombetta et al., 2015), and more specifically in the African continent (Cruciani et al., 2011; Scozzari et al., 2014). The major haplogroup A, B, and mainly E, are found across Africa, associated with ancient African populations. In European populations, the subhaplogroup R1 is highly represented, while in Native American populations the subhaplogroup Q is highly represented more than any other population (Battaglia et al., 2013).
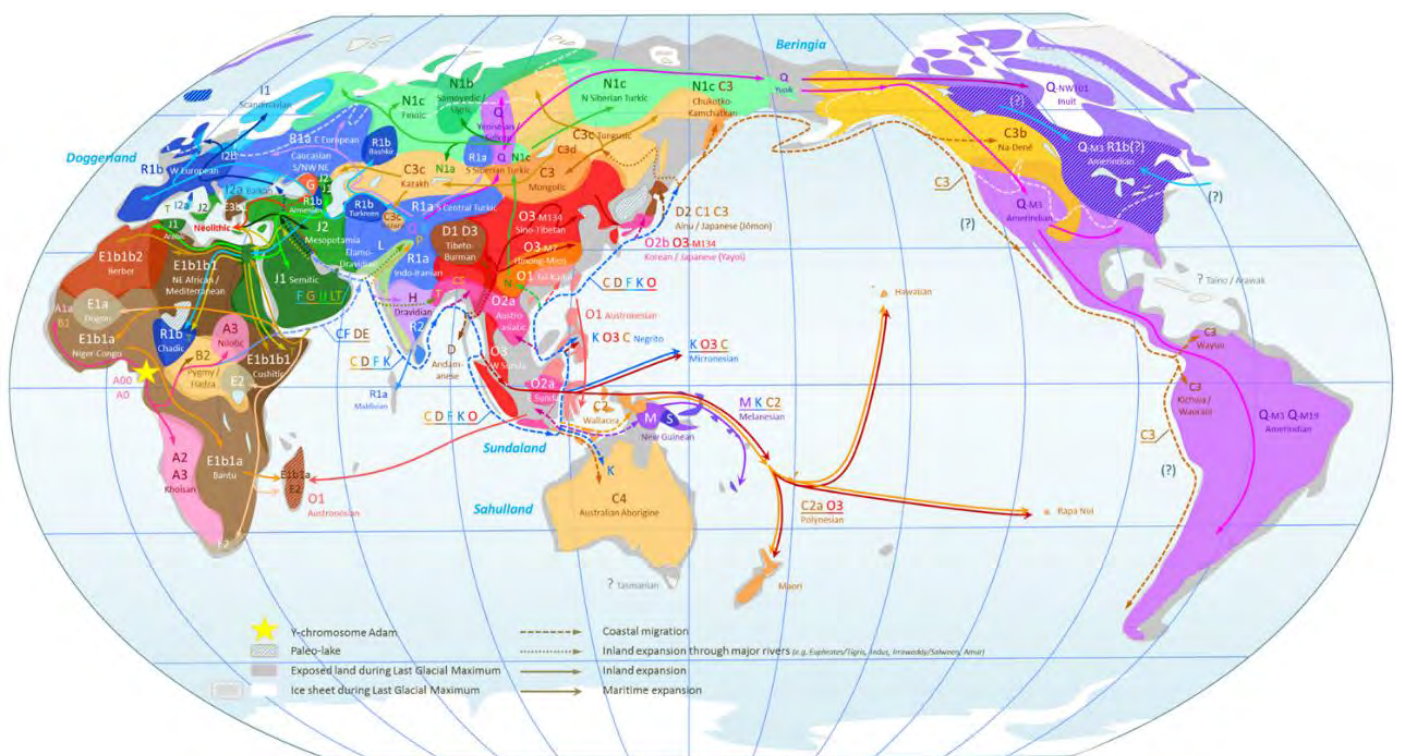
Figure 27: World map of Y chromosome haplogroups shows dominant haplogroups in pre-colonial populations with possible migrations routes (Dienekes' Anthropology blog, 2013).

Complex nomenclatures are used to describe Y chromosome haplogroups (Jobling & Tyler-Smith, 2003). This complexity is required to maintain some order of changes to the phylogeny additional variant and haplogroups are discovered. Currently, the nomenclature is based on cladistic relationships of the haplogroups. Major clades are identified by single capital letters (for example, haplogroup E), sublineages within these clades are given numerical suffixes (for example, subhaplogroup E1), and this can continued using alternating lowercase letters and numbers until all lineages have been named (for example, E1b1a1a1c1a). An alternative adopted for Y haplogroups is to use the name of the derived variant furthest from the root (such as M191 for the last example).

D.2. *Mitochondrial DNA genetic system*

Human mitochondrial DNA (or mtDNA) is present in hundreds to thousands of copies in each cell, not within the nucleus, but within the cell's energy-generating organelles, the mitochondria. MtDNA is a circular double-stranded molecule, 16,569 base pairs (bp) in length, which codes for 37 genes: 2 ribosomal RNAs (rRNAs), 13 subunits of the oxidative phosphorylation system, and 22 transfer RNAs (tRNAs) (see Figure 28). All these genes are tightly packed within the 16.5 Kb circular genome, encoding proteins involve in energy production and mitochondrial protein synthesis.

Figure 10.1, J.M. Butler (2005) *Forensic DNA Typing*, 2nd Edition © 2005 Elsevier Academic Press
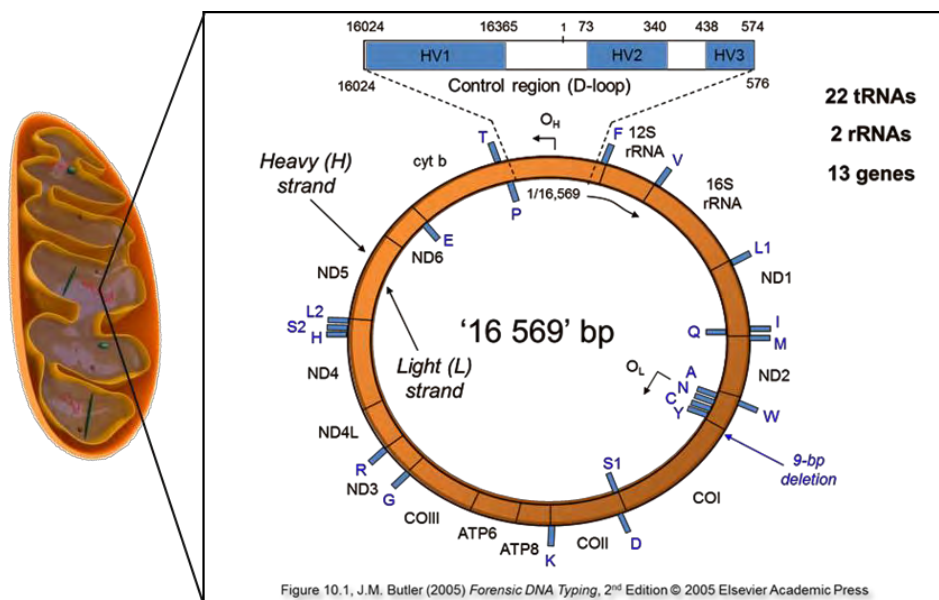
Figure 28: Molecular structure of whole mitochondrial DNA. In detail, position of hypervariable segments (HV1, HV2, and HV3) in the control region. Image modified from Butler (2005).

The mitochondrial genome also presents a substantial stretch of noncoding sequence known as 'control region' or 'D-loop', which contains a replication origin. The control region contains also three hypervariable segments (known as HVS-I, HVS-II, and HVS-III) that are commonly assayed for their high variability. Compared to nuclear genes, mtDNA sequences do not have introns and much noncoding sequence around them, and are densely packed (93%) with coding regions. A large proportion of positions in these genes are known to be highly conserved across different species, implying strong purifying selection in large human cohorts, likely because of being fatally deleterious or associated with disease (see MITOMAP (Ruiz-Pesini et al., 2007)).

Maternal inheritance (Hutchison et al., 1974), fast mutation rate (Brown et al., 1979), high copy number per cell (Piko & Matsumoto, 1976), and the lack of recombination (Hagstrom et al., 2014) were the features that brought mtDNA at the focus of evolutionary genetic studies in the 1980s and 1990s, when the human genome sequencing had not been completed yet and the idea of whole nuclear genome level population genetics was only a daydream for population geneticists (Pakendorf & Stoneking, 2005). Those features allowed the data from coding and noncoding regions of mtDNA to be combined into the shape of a phylogenetic tree (Cann et al., 1987). The branches of this ever-growing tree were assigned with alphabetic indicatives that became to be known as mtDNA haplogroups.

The nomenclature of mtDNA haplogroups was introduced in the mid-1990s with A-G labels assigned to variation observed in Asian and American lineages (Torroni et al., 1993), H-K to Europe (Torroni et al., 1994), while only a single letter, L, was assigned to describe the highest level of variation observed in Africa lineages. Curiously, the early study which defined African haplogroups L (L1 and L2) assumed Asian origins of human mtDNA variation and hence misplaced the root of the human phylogeny (Chen et al., 1995). Since then, diversity within mtDNA has been studied in approaching a million individuals by sequencing the most variable segments of the control region (i.e., HVS-I and HVS-II), often complemented by typing a number of informative SNPs from the coding region. Currently, the mtDNA nomenclature has a robust branch structure (see mtDNA tree, available online at http://www.phylotree.org/), which has been determined through the rigorous and detailed analyses of the whole mtDNA genomes (van Oven & Kayser, 2009; van Oven, 2010).

The first full mitochondrial genome sequence was determined in 1981 from human placenta in the laboratory of Fred Sanger in Cambridge, UK, which belonged to haplogroup H2a2a and became known as the Cambridge Reference Sequence (or CRS) (Anderson et al., 1981). Subsequently, the nucleotide numbering of mtDNA sequences was based on a revised and corrected version of the CRS, and known as the revised CRS (or rCRS) (Andrews et al., 1999). Behar et al. (2012) have proposed to adopt a reconstructed ancestral sequence instead, which

was called the Reconstructed Sapiens Reference Sequence (RSRS).

The topological details of the mtDNA phylogeny have been revealed step by step over the last two decades, thanks to the contributions of many groups in covering with data ever increasing numbers of populations across the world, and thanks to the advances in technology that eventually have led to the use of whole mtDNA sequencing as a routine approach in the field (Kivisild, 2015). In recent years, phylogeographic studies have produced detailed knowledge on the worldwide distribution of mtDNA variants, linking specific clades of the mtDNA phylogeny with certain geographic areas (van Oven et al., 2011). The first worldwide survey of mtDNA whole genome sequences showed, with a robust bootstrap support of the internal branches, that the root of the human mtDNA variation lies in Africa with TMRCA date of $171,500 \pm 50,000$ years, and that the age of the youngest clade with African and non-African sequences was $52,000 \pm 27,500$ years (Ingman et al., 2000). Other whole mtDNA studies based on global sampling have generally agreed with these structural findings, and revealed more details of: the regional patterns of diversity (Mishmar et al., 2003), time scale of the accumulation of diversity (Behar et al., 2012), and the female effective population size changes over time (Lippold et al., 2014).

As it has been repeatedly shown, the root of the mtDNA phylogeny and the most diverse branches are restricted to African populations (Figure 29). By using the maximum molecular resolution enabled by the analysis of whole mtDNA genomes, the first seven bifurcations in this tree define the distinction of strictly sub-Saharan African branches (from L0 to L6) from those that are shared by Africans and non-African populations (Kivisild, 2015).
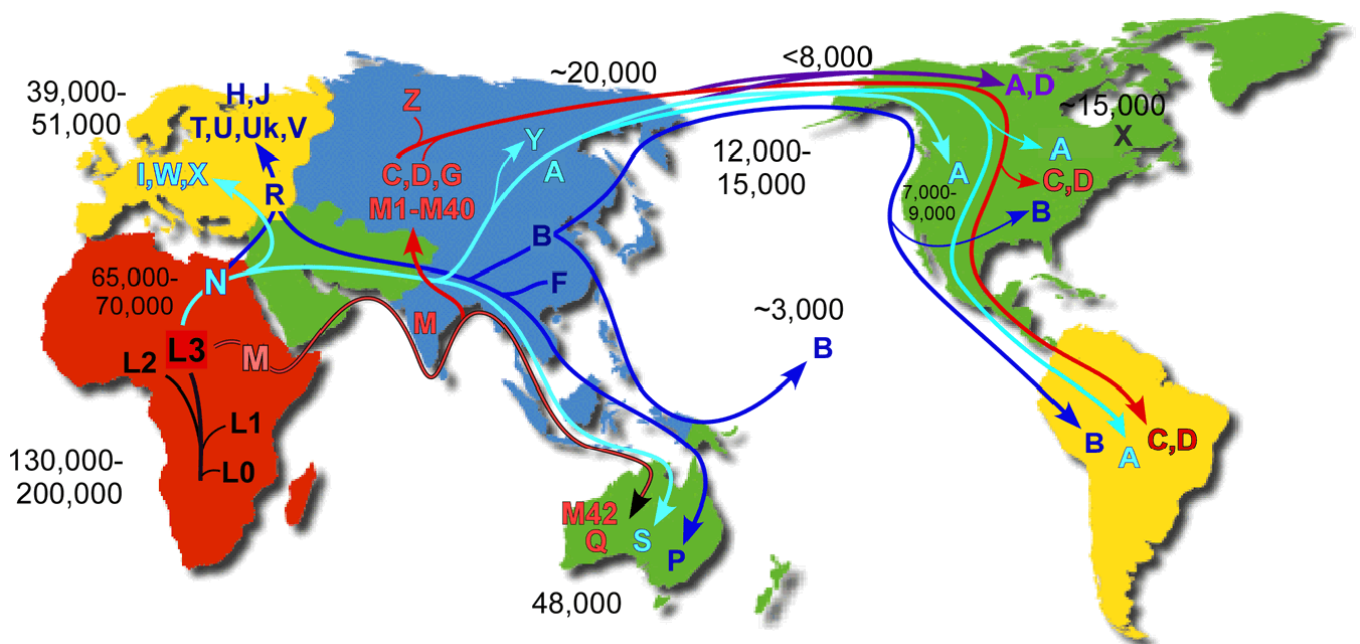
Figure 29: Map outlines the migratory history of the major mtDNA haplogroups. Image available online at http://www.mitomap.org/.

Analyses of whole mtDNA sequences of sub-Saharan Africans have revealed an old divergence (circa 90 to 150 ky) of the sublineages L0d and L0k that are specific to the Khoisan populations from South Africa. It has been estimated that during this time period at least six additional sublineages existed in Africa with living descendants (Barbieri et al., 2013b; Barbieri et al., 2014a). In contrast to the overall high basal clade diversity and geographic structure, some terminal branches from haplogroups L0a, L1c, L2a, and L3e show recent coalescent times and wide geographical distribution in Africa, likely because of the recent Bantu expansion (Barbieri et al., 2014b; Marks et al., 2015). Given the complexity of admixture of the Bantu-speaking populations, the use of whole mtDNA sequences in these studies have been instrumental in revealing the distinct autochthonous sources and ancient substructure at the background of the overall high genetic homogeneity of the Bantu speakers (de Filippo et al., 2012).

Outside Africa, lineages L0 to L6 are extremely rare and restricted to geographic areas that have received historic gene flow from Africa, such as Mediterranean Europe, West Asia, and the Americas. On the basis of analyses of high resolution whole mtDNA sequences, it has been estimated that approximately two thirds of the rare African lineages L, which are found at combined frequency of less than 1% in Europe, were brought in Europe from Africa during Roman times, Arab conquests, and the transatlantic slave trade; while just one third are more likely to have been introduced in Europe earlier during pre-historic times (Cerezo et al., 2012).

The fact that virtually every non-African mtDNA lineage derives from just one of the two subclades of the African haplogroup L3 has been interpreted as an evidence of a major bottleneck of mtDNA diversity at the onset of the out of Africa dispersal (Underhill & Kivisild, 2007). The magnitude of this bottleneck has been estimated from the whole mtDNA sequence data yielding the estimates of the effective population size which range between several hundred (Macaulay et al., 2005) and only few tens of females (Lippold et al., 2014). The separation of these two subclades, M and N (Figure 29), from their African sister-clades in L3 can be dated back to 62 to 95 kya (Fu et al., 2013). While the internal coalescent time estimates of the M and N founders have been estimated in the range of 40 to 70 ky (Soares et al., 2013). This estimation suggests that the dispersal of the M and N founders occurred probably after rather than before the eruption of Mount Toba 74 kya in Indonesia, one of the Earth's largest known volcanic events in human history (Smith et al., 2011).

Even though haplogroups M and N are widely spread in Asia, Australia, Oceania and Americas, the geographic distribution of each of their subclades has more specific regional configuration (Figure 29). In Eurasia, haplogroups U, HV, I, JT, N, W and X are today common in Europe, Southwest Asia and North Africa (Soares et al., 2010); haplogroups R5-R8, M2-M6

and M4'67 are restricted to South Asia (Thangaraj et al., 2008); while haplogroups A-G, Z and M7-M9 are widespread in East Asia (Stoneking & Delfin, 2010). In Native Americans, mtDNA variation primarily falls to haplogroups A to D and X (Tamm et al., 2007). They originated, with the exclusion of X, form a subset of the East Asian diversity (Achilli et al., 2013). Despite the clear and distinct geographic spread patterns in extant populations, it is not simple and straightforward to make inferences about the origin of these patterns, and to associate the haplogroup labels with specific prehistoric events or time periods (Kivisild, 2015).

Overall, mtDNA has offered the opportunity to explore genealogical relationships among individuals and to study the frequency differences of matrilineal clades among human populations at continental and regional scales. Consequently, mtDNA has been a widely used tool in human evolutionary and population genetic studies over the past three decades. Moreover in the era of whole nuclear genome sequencing, mitochondrial genomes are continuing to be informative as a unique tool for the assessment of female-specific aspects of the demographic history of human populations (Kivisild, 2015).

D.3. Genome-wide SNP data diversity

In the 1977, Sanger et al. (1977) and Maxam & Gilbert (1977) developed methods to sequence DNA by chain termination and fragmentation techniques, respectively. This transformed biology by providing the tools to decipher complete genes and, later, entire genomes. The technique developed by Sanger et al. (1977), commonly referred to as 'Sanger sequencing', required less handling of toxic chemicals and radioisotopes than Maxam and Gilbert's method, and as a result it became the prevailing DNA sequencing method for the next 30 years. A growing demand for increased throughput led to laboratory automation and process parallelization, which eventually resulted in the establishment of factory-like outfits with hundreds of sequencing instruments. Thanks to these advances, the Sanger sequencing ultimately enabled the completion of the first human genome sequence in 2004 (International Human Genome Sequencing Consortium, 2004).

Nevertheless, the Human Genome Project required vast amounts of time and resources; it was clear that faster, higher throughput, and cheaper technologies were required. For this reason, research institutes and companies developed and commercialized the Next-Generation Sequencing technologies (or NGS), as opposed to the previous methods. These new sequencing methods share three major improvements. First, instead of requiring bacterial cloning of DNA fragments they rely on the preparation of NGS libraries in a cell free system. Second, instead of hundreds, from thousands to many millions of sequencing reactions are produced in parallel.

Third, the sequencing output is directly detected without the need for electrophoresis; base interrogation is performed cyclically and in parallel (van Dijk et al., 2014). The enormous numbers of reads generated by NGS enabled the sequencing of entire genomes at an unprecedented speed.

The first NGS technology to be released in 2005 was the pyrosequencing method by 454 Life Sciences (now Roche 454) (Margulies et al., 2005). The 454 Genome Sequencer generated about 200,000 reads (~20 Mb) of 110 base-pairs (bp). During the past decade, tremendous progress has been made in terms of speed, read length, and throughput, along with a sharp reduction in per-base cost. Today, NGS platforms such as Roche 454, SOLiD, and Illumina, provide cheaper and larger genome-wide SNP data, which has new extraordinary applications in research areas such as clinical diagnostics, agro-genomics, and forensic science (van Dijk et al., 2014). For instance, Illumina human whole-genome genotyping microarrays provide large datasets with a genomic coverage from around 250,000 SNP markers to the whole genome (Figure 30). Interestingly, after the whole genome sequencing, Illumina HumanOmni5 microarray delivers the most comprehensive up-to-date coverage of the genome.
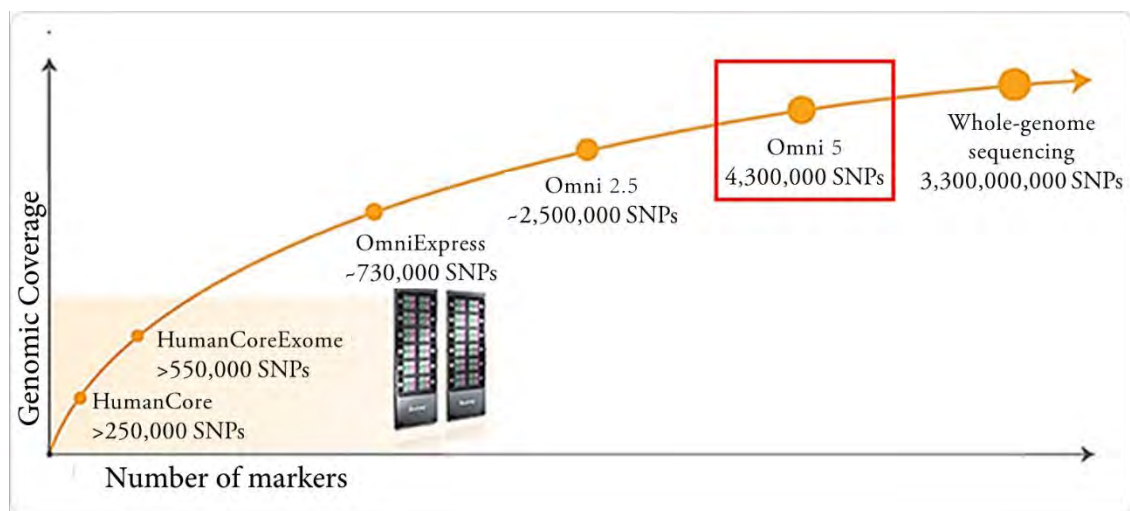
Figure 30: Illumina Omni SNP microarrays can perform from thousands to millions of markers. Illumina HumanOmni5 SNP microarray (red square) delivers the most comprehensive coverage of the genome, with the exception of whole genome sequencing. © 2015 Illumina, Inc.

Initially, with only a few hundred autosomal microsatellites, genome-wide polymorphism data have established differences in allele frequency among continental regions (Jakobsson et al., 2008; Rosenberg et al., 2002). More recently, genome-wide SNP data derived from SNP arrays found evidence of appreciable fine-scale structure within continents, and even within countries. For instance, genome-wide SNP data have revealed strong statistical evidence for genetic

substructure within Europe (Lao et al., 2008; Novembre et al., 2008), the Near East (Li et al., 2008), India (Reich et al., 2009), East Asia (Tian et al., 2008), the Americas (Risch et al., 2009), and Africa (Henn et al., 2011; Henn et al., 2012). Currently, these studies are based on Ancestry Informative Markers (or AIMs), which are genetic markers that show substantial differences in allele frequency across population groups.

Generally, patterns of genetic substructure suggest that geographic barriers to gene flow play a key role. By using nearly 200,000 common SNPs, Novembre et al. (2008) reconstructed a 'genetic map' of Europe by projecting the first two principal axes of genetic variation onto geographic coordinates. This phenomenon can clearly extend to variation within other continents or countries. For instance, it has been reported genetic evidence for fine-scale substructure within United Kingdom (Leslie et al., 2015), Finland (Jakkula et al., 2008), Mexico (Silva-Zolezzi et al., 2009), Puerto Rico (Tang et al., 2007), Ethiopia (Pagani et al., 2012), Madagascar (Pierron et al., 2014), and South Africa (de Wit et al., 2010; Schlebusch et al., 2012). However, not all regions appear to have clearly differentiated populations. For instance, West Africa is striking for having very little fine-scale structure, at least at the level of resolution captured by common SNP data (Bryc et al., 2010a; Zakharia et al., 2009), even though these data included populations of Bantu and Non-Bantu Niger-Kordofanian, Afro-Asiatic and Nilo-Saharan speakers spread out over broad geographic regions (Henn et al., 2010).

Although many initial large-scale genetic association studies have focused primarily on homogeneous populations, increasingly studies are addressing samples in which individuals have more complex backgrounds, including admixed ancestry of African American (Bryc et al., 2010a; Hinch et al., 2011; Perera et al., 2013; Wegmann et al., 2011). Such studies depend crucially on accurate and unbiased ancestry inference both at a genome-wide level as well as at each locus in the genome (Pasaniuc et al., 2013). Hence, inference of ancestry from genetic data is a critical aspect of genetic studies, with applications ranging from the inference of population history to the estimation of population structure (Novembre et al., 2008; Rosenberg et al., 2003).

Recent methodological and technical advances in genomic technologies and computing resources have made possible the emergence of genome-wide studies, whose main advantage for demographic inference is allowing to identify and quantify admixture event among populations with different ancestries (Novembre & Ramachandran, 2011). Hence, we can apply genome-wide studies to accurately infer overall ancestry, as well as ancestry at a fine-scale across an individual's genome. Ancestry estimation is a frequently encountered problem and has been used in a variety of applications such as tracing someone's geographic origin in forensic investigations (Kayser & de Knijff, 2011), correcting for population stratification in genome-wide association studies (Bush & Moore, 2012), and developing personalized

approaches to treatment (Huser et al., 2014). Importantly, in genetic association studies, ancestry inference can be used to account for the effects of population stratification which is a serious confounding factor and can lead to elevated rates of false positives (Price et al., 2010).

There are currently two different paradigms underlying ancestry inference: global ancestry estimation and local ancestry estimation. Global ancestry inference involves estimating the proportion of ancestry contributed by different populations averaged across the entire genome. Such methods have been applied to study population structure in humans (Lao et al., 2014; Pritchard et al., 2000; Rosenberg et al., 2002; Wollstein & Lao, 2015), and also other mammals such as chimpanzee populations (Becquet et al., 2007). In contrast, in local ancestry inference, we interpret each chromosome in an individual's genome as a mosaic of segments that originate from different ancestral populations and the goal is to find the ancestral population of origin at each position (Tang et al., 2006).

Local ancestry-based methods, such as LAMP (Sankararaman et al., 2008), HAPMIX (Price et al., 2009), RFMix (Maples et al., 2013), and PCAdmix (Brisbin et al., 2012), devolve ancestry at each locus in the genome and provide individual-level information about ancestry and admixture mapping. While these methods provide valuable insights into the recent history of populations, they have reduced power to detect older events. Local ancestry inference methods have been used mainly to study recently admixed populations such as African Americans (Bryc et al., 2010a; Kidd et al., 2012) and Hispanic populations (Bryc et al., 2010b; Johnson et al., 2011; Moreno-Estrada et al., 2013).

The most commonly used methods for studying global ancestry are the Principal Component Analysis (or PCA) (Patterson et al., 2006; Price et al., 2006), and model-based clustering methods such as STRUCTURE (Pritchard et al., 2000), FRAPPE (Tang et al., 2005), and ADMIXTURE (Alexander et al., 2009). They are also the most powerful tools for detecting population substructure. ADMIXTURE employs the same model as STRUCTURE but uses a maximum likelihood estimation procedure involving high-dimensional optimization algorithms. ADMIXTURE is over an order of magnitude faster than STRUCTURE and produces estimates of similar accuracy (Alexander et al., 2009).

The PCA was firstly introduced to the study of genetic data almost thirty years ago by Menozzi et al. (1978), since then the PCA has become as a standard tool in genetics, especially to study scenarios of genetic geographic variation and population structure (Cavalli-Sforza et al., 1994; Cavalli-Sforza & Feldman, 2003). EIGENSTRAT (Patterson et al., 2006; Price et al., 2006) is a well-known program that implements PCA, which seeks to construct projections in lower dimensional space that capture a large fraction of the variation in the marker genotypes. The projections inferred by such approach tend to be highly correlated with the geographic locations

from where individuals were sampled (Novembre et al., 2008; Wang et al., 2012). For dense polymorphism datasets such as those obtained from sequencing, haplotype based analysis has the potential to leverage this information and provide improved ability to detect population substructure. ChromoPainter and fineSTRUCTURE are recently developed programs that aim to make use of haplotype structure for high quality PCA and population structure inference respectively (Lawson et al., 2012).

D.4. Genetic landscape in the African populations

Africa has been considered the cradle of mankind for a long time. Both, genetic data and fossil evidence suggest that anatomically modern humans originated in this continent (McDougall et al., 2005), spreading later all over the globe (Groucutt et al., 2015). Furthermore, Africa is a region of great linguistic, cultural, phenotypic, and genetic diversity. It contains more than 2,000 distinct ethno-linguistic groups, practicing a wide range of subsistence patterns including agriculture, pastoralism, and hunting-gathering. The pattern of genetic variation in modern African populations is influenced by their demographic history (such as changes in population size, short- and long-range migration events, and admixture process) as well as locus-specific forces such as natural selection, recombination, and mutation. For instance, the Bantu expansion throughout sub-Saharan Africa and subsequent admixture with indigenous populations has had a major impact on patterns of variation in modern African populations (Campbell & Tishkoff, 2008).

Several studies have highlighted that Sub-Saharan African populations present the highest levels of genetic diversity (Rosenberg, 2011; Tishkoff & Verrelli, 2003). European populations show intermediate values, and East Asian populations show the lowest (Figure 31). This pattern has been confirmed with autosomal microsatellite variation (Rosenberg et al., 2002), and large-scale SNP genotyping (Li et al., 2008). Such observations emphasize the importance of studying diversity within Africa.

A study of ~800 microsatellites and ~400 insertion or deletion polymorphisms (or INDEL) genotyped in >2,500 African people indicated high levels of populations structure, which correlates with six ancestral population clusters (Figure 32). Most African populations have mixed ancestry from these different clusters, reflecting high levels of migration and admixture among ethnically diverse groups. In a Bayesian clustering analysis of those clusters, the most geographically widespread cluster extends from the Senegalese Mandenka group of West Africa through Central Africa to the Bantu-speaking group Xhosa of South Africa. This cluster corresponds closely to the distribution of the Niger-Kordofanian language family, and seems

compatible with the expansion of Bantu-speakers (Tishkoff et al., 2009). Hence, the West and West-Central populations shared common ancestry with Niger-Kordofanian speakers groups across African coastal regions. Therefore, it is a challenging task to infer the ancestral origin of the African Americans, who are descents of West and West-Central populations.

Genetic studies of Y chromosome (Figure 33A) and mitochondrial DNA (Figure 33B) diversity across Africa have identified particular lineages associated with specific African regions (Figure 33C). Africa is notable for being the cradle of the deepest-rooting Y haplogroups (A and B) and mtDNA haplogroups (L0 to L6). Both, Y haplogroups A and B are present at only 13% overall in a sample of 3,255 chromosomes (Jobling et al., 2014). There is one predominant African lineage, haplogroup E1b1a1, which accounts for over 59% of the African chromosomes, and about 80% of the chromosomes in Bantu speakers from Cameroon and West Africa. Haplogroup E1b1a1 is also found at lower frequencies among Pygmy populations, who have retained their hunter-gatherer lifestyles but experienced some admixture from agriculturalists. Besides, mtDNA diversity highlights haplogroups L0a, L2a, L3b, and L3e as signature lineages of West and West-Central African populations. They are found also in high frequencies in modern Bantu-speaking populations (Jobling et al., 2014).
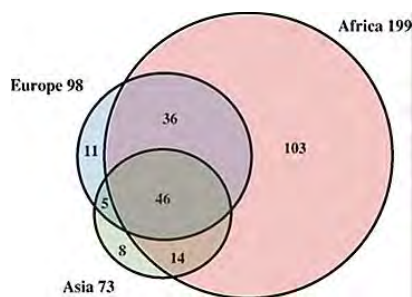
Figure 31: African and African-American populations present the highest levels of within-population genetic diversity in comparison with non-African populations (Tishkoff & Verrelli, 2003).

Figure 32: Geographic distributions of six major clusters based on genome-wide SNP diversity within African populations. Each colour represents one of six clusters produced by using a Bayesian clustering analysis assuming no admixture. The yellow cluster corresponds well with the distribution of Niger-Kordofanian languages (Tishkoff et al., 2009).
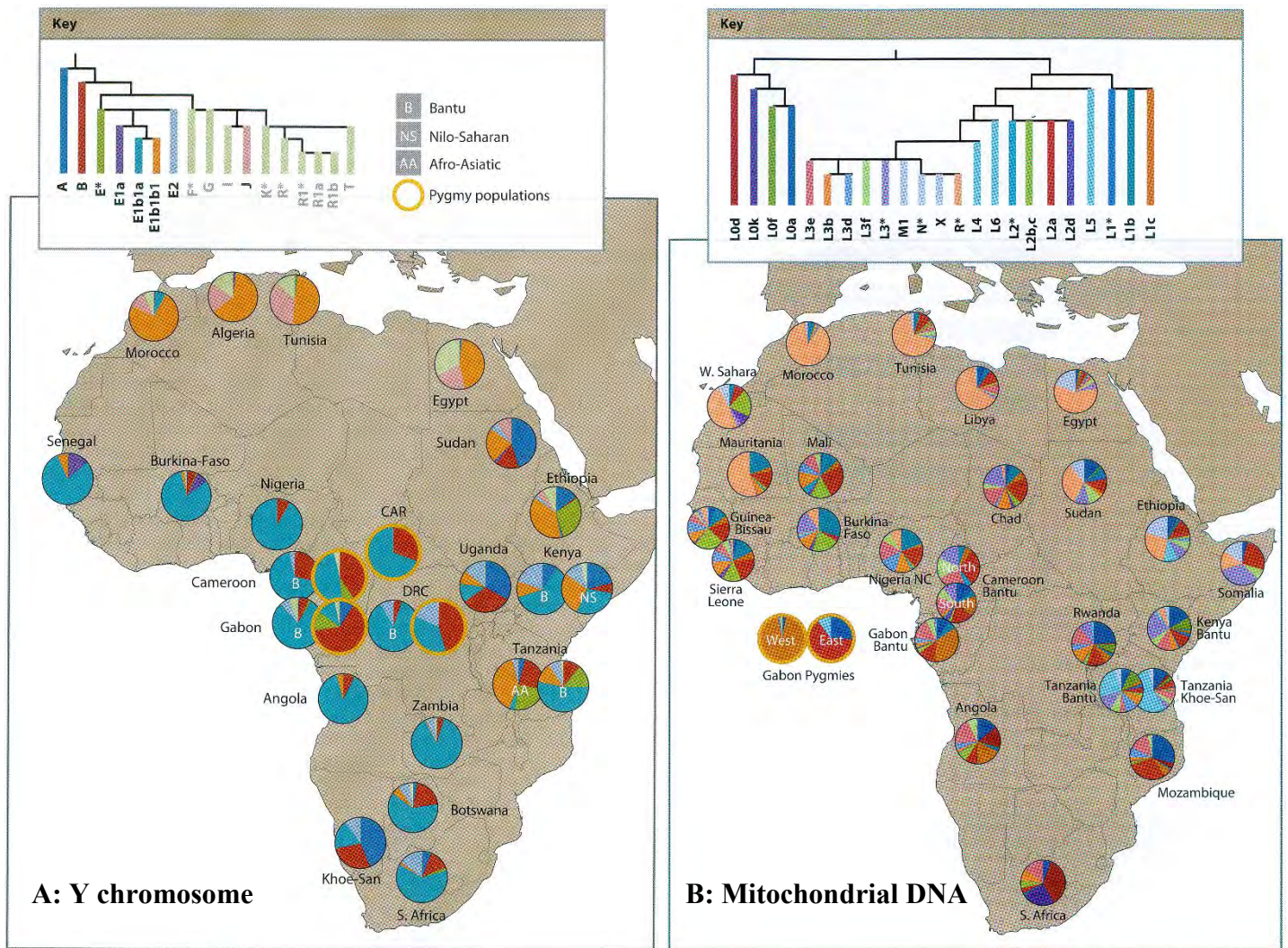
A: Y chromosome

B: Mitochondrial DNA

C
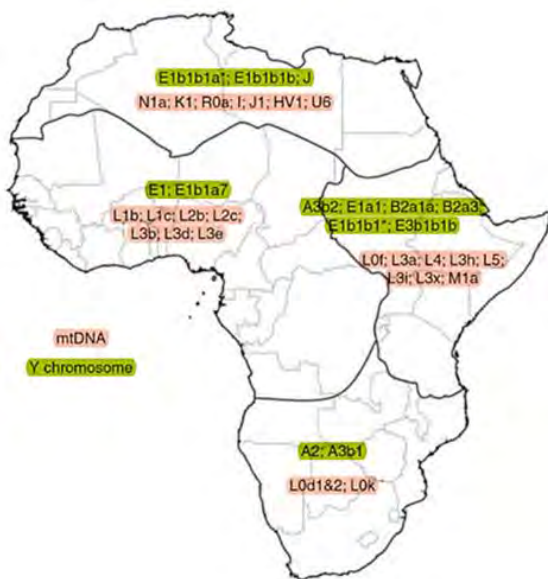
Figure 33: Geographic distributions of major Y-chromosomal (**A**) and mtDNA (**B**) haplogroups in Africa. Pie charts show the frequencies of different haplogroups represented on the respective phylogenetic tree. The trees show the phylogenetic relationships of the haplogroups, with colour coding corresponding to the pie charts. For some regions, both Bantu-speaking (B) and non-Bantu-speaking (NS and AA) populations are shown. Pygmy populations are indicated by yellow edging. Y chromosome data from Arredi et al. (2004) and de Filippo et al. (2011). MtDNA data from Rosa & Brehem (2011). Images modified from Jobling et al. (2014).

Below, (**C**) current consensus distribution of major mtDNA and Y-chromosomal haplogroups in African regions: North, West, East, and South (Gomez et al., 2014).

D.5. Genetic landscape in the African-American populations

The admixture process in Latin America started after Christopher Columbus first disembarked at Hispaniola (1492), as soon as the 39 men that he left on the island had sexual intercourse with the local Native American women (Morner, 1967). Therefore, a maximum of roughly 21 generations of admixing may be established, with some variance because of regional difference. Heyer et al. (1997) identified up to 19 generations of admixing based on pedigrees of descended from males who lived in the 17th century. Whereas Wang et al. (2008), based on 13 mixed Latin American populations, estimated that the average time since the first admixture allowed for 6 to 14 generations, however these low estimations excluded the Caribbean region.

Genetics, as an independent source of data, can complement existing historical reports. For instance, among Afro-Brazilians in São Paulo (Brazil) was suggested that the African ancestral contributions are from West (43%), West-Central (45%), and Southeast Africa (12%) (Goncalves et al., 2008). This data is in good agreement with historical documents (Klein & Vinson, 2007). Hence, merging such valuable genetic data with non-genetic sources (such as historical, archaeological, and linguistic data) can yield important details and specificity to our reconstructions of the diverse events associated with the transatlantic African diaspora and its aftermath (Jackson & Borgelin, 2010).

Historical records on the origin of African Americans in the United States have estimated that about 64% were from West Africa, 35% from West-Central Africa, and 1% from Southeast Africa (McMillin, 2012; Thomas, 1999). In agreement with these historical data, Salas et al. (2004) estimated the quantitative contribution of the different African regions to the formation of the New World mtDNA gene pool. According to their estimated admixture coefficients reported, 53.0% of shared mtDNA sequence types found in North America have a West African origin, and 34.5% a West-Central African contribution. These values are significantly different from those obtained for Central America (69.1% West and 21.2% West-Central Africa), and South America (32.1% West and 58.5% West-Central Africa). By using the same kind of approach, but with substantially more data (an African database of 4,860 mtDNA sequences, and a database of 1,148 mtDNA sequences for African Americans from the U.S. that also contained 1,053 mtDNA sequences for sub-Saharan ancestry), Salas et al. (2005a) estimated that >55% of the U.S. mtDNA lineages have a West African ancestry, with <41% coming from West Central or Southwest Africa.

A recent study about the autosomal diversity of genetic ancestry of 5,269 self-described African Americans in the U.S. indicates that the highest levels of African ancestry are found in the South, especially South Carolina and Georgia, and the lowest proportions of African ancestry in the Northeast, the Midwest, the Pacific Northwest and California (Bryc et al., 2015),

consistent with previous studies (Parra et al., 2001; Zakharia et al., 2009). In that study, genome-wide ancestry estimates of African Americans showed average proportions of 73.2% African, 24.0% European, and 0.8% Native American ancestry (Bryc et al., 2015), which agrees with previous studies (Bryc et al., 2010a; Lind et al., 2007; Tishkoff et al., 2009).

In Afro-Caribbean populations, several genetic studies have estimated they have around 65–95% West African, 4–27% European, and 0–6% Native American ancestries (Deason et al., 2012; Kidd et al., 2012; Murray et al., 2010; Simms et al., 2011; Simms et al., 2013). Although, pooled individuals from the Caribbean have a high proportion of African ancestry, fine-scale genetic structure has been observed within and between islands particularly in Dominica, Grenada, St. Kitts, St. Lucia, St. Thomas, St. Vincent, Jamaica, and Trinidad, because of regional differences in levels of African and European ancestries (Torres et al., 2013).

Similarly, a study of genetic admixture within Puerto Ricans showed that levels of African ancestry varied geographically with the highest proportion occurring in the eastern part of the island where enslaved Africans and their descendants historically engaged in sugar production (Via et al., 2011). In addition, genome-wide SNP data have suggested that patterns of genetic ancestry in Cuba, Puerto Rico, and Hispaniola (the Greater Antilles) are consistent with a model of two pulses of African migration events from different regions of western Africa, implying that Afro-Caribbean populations have mixed African ancestry. The first pulse involved coastal West African regions in early stages of the transatlantic slave trade. The second pulse involved present-day West-Central African populations, supporting historical records of later transatlantic deportation (Moreno-Estrada et al., 2013).

These results are also congruent with Y chromosome studies that found diverse haplotypes in Afro-Caribbean from the Bahamas, Haiti, and Jamaica that were inferred to originate from different ethnic groups within West and West-Central Africa (Simms et al., 2011; Simms et al., 2012). Furthermore, isotope data from skeletal remains of enslaved Africans in Barbados suggested that first generation captives had different dietary histories likely because of differences in their geographic origins in Africa (Schroeder et al., 2009).

Interestingly, various patterns of sex-biased gene flow were found in a number of ex-colonial populations throughout the Americas (Simms et al., 2013; Stefflova et al., 2009). The diverse socio-cultural histories of North, Central, and South America are reflected in sex-specific admixture. By using mtDNA, NRY, and AIMs, Stefflova et al. (2011) confirmed the presence of sex-specific admixture as well as the existence of remarkable differences across the Americas (Figure 34). That study pointed out that European males, rather than females, are predominantly responsible for the European genomic contribution to American populations of African descendants, and also both Native American females and European males provided a greater

68

contribution to South American populations (represented by Brazil) compared to the U.S. admixed populations. Consequently, the combination of the maternal (Figure 34a) and paternal (Figure 34b) admixture proportions were approximately reflected in the autosomal genome admixture proportions (Figure 34c).
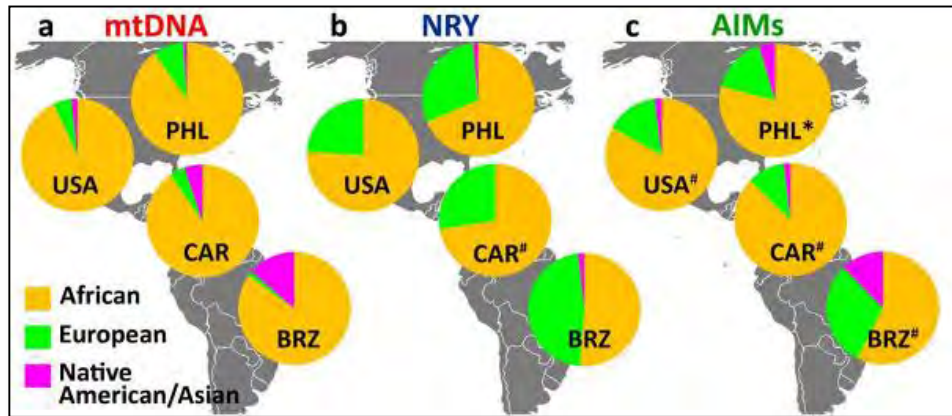
Figure 34: Pan-continental group ancestry of African descendants. MtDNA (**a**) and NRY (**b**) reflecting maternal and paternal admixture respectively, while AIMs (**c**) reflecting the autosomal genome. These genetic systems show the relative ratio of three continental populations that significantly contribute to the admixed populations of the United States (the entire U.S. and by a sample from Philadelphia, PHL), the Caribbean Islands (CAR), and Brazil (BRZ) (Stefflova et al., 2011).

Overall, during the transatlantic slave trade, African-American communities have been an endpoint of migration for hundreds of years, resulting in diverse genetic patterns. Indeed, genetic studies pointed out that populations with African ancestry have a complex history resulting in genetic heterogeneity within African-American populations and within African populations. As a result of the complexity of past migration events, additional studies across a broader geographic range of the Americas are needed to fully understand the extent of genetic variability, and the different demographic processes that have contributed to it in African descendants particularly from South America.

E. Objectives of present study

According to historical, linguistic, and genetic sources in African descendants, there are different ancestral contributions from large African regions from West, West-Central, South-West, and Southeast Africa (Figure 5). Their proportions substantially differ between North, Central, and South America (Table 2). Moreover, the relative genetic homogeneity between West and West-Central African populations makes a challenge significant discrimination of populations that allows determinate African origin (Bryc et al., 2010a; Ely et al., 2006; Hunemeier et al., 2007; Jackson & Borgelin, 2010; Salas et al., 2005b; Tishkoff et al., 2009).

Previous researches have shown that the Noir Marron communities have linguistic richness (Bakker et al., 2011; Huttar et al., 2007), as well as conserved considerable African genetic ancestry (Brucato et al., 2009). Both, linguistic and genetic studies suggest strong ancestral links of the Noir Marron communities with populations inhabiting the Bight of Benin, such as Fon and Yoruba (Brucato et al., 2010; Essegbey et al., 2013a; Migge & Winford, 2013). However, there is an important caveat about genetic variation within the Noir Marron communities, and also about original ancestry of runaway enslaved Africans that played an important role in the formation of each community. Furthermore, there is a significant lack of knowledge about genetic structure of historical African coastal regions that may identify the African ancestry of African Americans, as well as about the effect of admixture events in shaping the genetic diversity in the Americas.

In the present study, we aim;

i) First, to verify the distinctive African heritage of four Noir Marron communities from French Guiana and Surinam: Aluku, Paramaka, Ndjuka, and Saramaka.

ii) Second, to explore the African genetic identity of other descendants with noteworthy African heritage from the Chocó department in northwest Colombia, and from Rio de Janeiro in southeaster Brazil. Additionally, in order to achieve a high contextualization of the genetic diversity in their African ancestors, we also study the genetic diversities of West African populations from Benin, Ivory Coast, and Mali; and from other studied African populations across the continent.

iii) Third, to estimate likely sex-biases and to compare paternal and maternal genetic legacies among the African-American populations, a high resolution of the uniparental markers could be considered.

iv) Fourth, to evaluate likely admixture patterns and gene flow with non-African populations that may add ancestral source of variation an extended analysis of the bi-parental markers could be considered to compare their African autosomal legacy.

v) Finally, to reconnect their African regional and ethnic roots based on phylogenetic reconstructions and multidimensional analysis, African ancestral contributions could be estimated for different historical slave regions across Africa that brought African captives to South America and to establish probable slave routes.

In a nutshell, the inference of biogeographical ancestry of African American can provide useful information about the slave trade. This study could add new population data that could characterize the genetic ethnicity of African populations throughout the Atlantic world to further understand their demographic histories. We focus on enslaved African descendants in South America, and particularly in French Guiana and Surinam, as well as their likely West African ancestors. The new findings could shed new light on one of the darkest chapters of world human history.

II. MATERIAL AND METHODS

A. *Population samples and sampling procedures*

We performed a genetic analysis of African descendant populations from South America (Figure 35). We compared the results of this analysis with the results obtained for populations with West African ancestry (from Benin, Ivory Coast, and Mali). Biological samples were collected in French Guiana for this study, and other were collected previously in different countries in collaboration with different institutions.

The present study complies with the Helsinki Declaration of Ethical Principles established by 59th World Medical Association General Assembly in Seoul, October 2008 (World Medical Association, 2013). The informed consent was obtained from all participants prior to their participation in accordance with local ethical recommendations. This study was performed after authorization of the *Commission Nationale de l'Informatique et des Libertés* (CNIL), *le Comité Consultatif de Protection des Personnes dans la Recherche Biomédicale Outre-Mer III* (*Hôpital Purpan* in Toulouse, France), *l'Agence Française de Sécurité Sanitaire des Produits de Santé* (AFSSAPS).
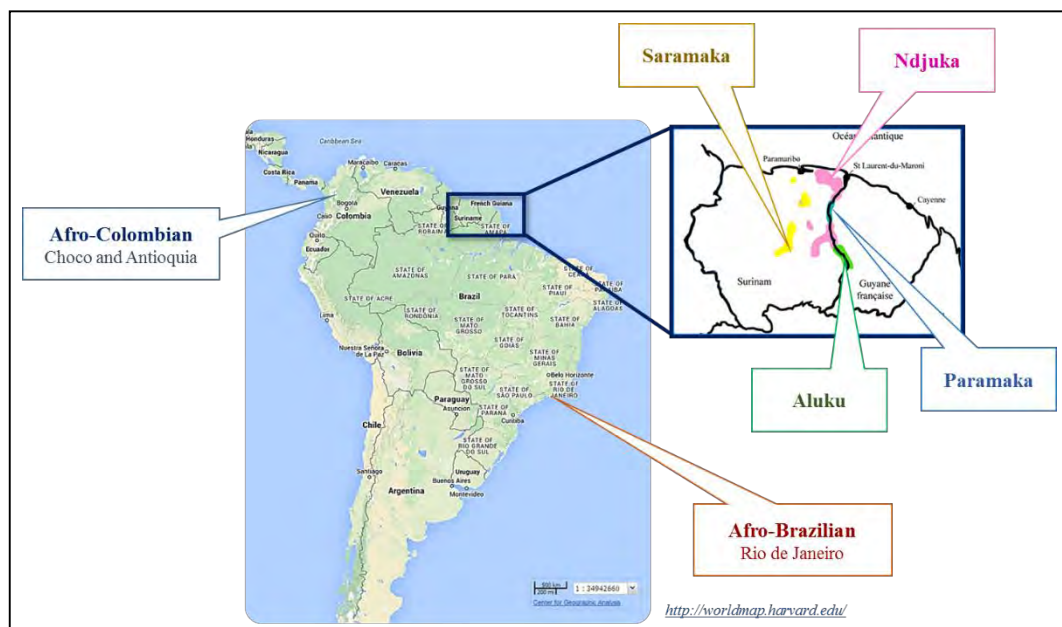
Figure 35: Geographic location of studied African American communities in South America. Locations of the Noir Marron communities of French Guiana and Surinam are according to Price (1996).

A.1. *African-American populations*

In French Guiana, samples from four Noir Marron communities (Aluku, Ndjuka, Saramaka and Paramaka) were collected during previous collaborative studies in Saint-Laurent du Maroni (5°29'35.5"N, 54°01'36.1"W), Maripasoula (3°38'38.2"N, 54°01'54.1"W) and Papaichton (3°48'33.0"N, 54°09'00.0"W) across the Maroni river region in the border between French Guiana and Surinam, by Antoine Gessain (*Department of Virology, Unit of Epidemiology and Physiopathology of Oncogenic Viruses, Institut Pasteur, Paris, France*) (Plancoulaine et al., 1998; Plancoulaine et al., 2000; Plancoulaine et al., 2006).

A new sampling of the various Noir Marron communities was carried out in the border between French Guiana and Surinam in order to increase the sample size previously analysed by Brucato et al. (2009) and Brucato et al. (2010). The sampling strategy was designed to be as representative as possible of the four communities, and to include individuals with different geographical origins (see Table 8 and Figure 36). Participants were asked about the origins of their parents and grandparents, and their genetic relationships with other donors contributing to this study. Blood samples were taken from participants representing at least the third generation who were born in the same geographical region.

Table 8: Geographic origin of Noir Marron people analysed in the present study. Each participant was carefully informed about the goals of this research project.

Geographic origin	Latitude	Longitude	Ndjuka	Aluku	Saramaka	Paramaka
Paramaribo	5°51'00.00"N	55°12'00.00"W			X	
Wia Wia	5°53'24.00"N	54°28'48.00"W	X			
Moengo	5°36'36.00"N	54°23'60.00"W	X			
St. Laurent du Maroni	5°30'00.00"N	54°01'48.00"W	X	X	X	X
Apatou	5°09'00.00"N	54°21'24.00"W	X	X		X
Brokopondo	5°01'31.70"N	54°59'34.20"W			X	
Nasson	4°51'36.00"N	54°28'12.00"W				X
Stoelmans Eiland	4°21'01.66"N	54°24'31.76"W	X			
Grand Santi	4°15'00.00"N	54°22'48.00"W	X			
Abouna Sounga	4°02'60.00"N	54°21'00.00"W	X			
Kajana	3°54'00.00"N	55°40'48.00"W			X	
Boniville	3°49'48.00"N	54°10'48.00"W		X		
Loka	3°49'48.00"N	54°12'00.00"W	X	X		
Papaichton	3°48'36.00"N	54°08'60.00"W	X	X		
Lawatabiki	3°40'48.00"N	54°05'24.00"W				X

Figure 36: Geographic origin of studied Noir Marron communities between French Guiana and Surinam. Image modified from Google Earth Pro (Google, Inc.).

In Brazil, Maria Cátira Bortolini and her collaborators (*Laboratório de Evolução Humana e Molecular, Departamento de Genética, Universidade Federal do Rio Grande do Sul, Brazil*) recruited participants from the Afro-Brazilian population in Rio de Janeiro city (22°54'36.0"S, 43°12'00.0"W), the capital of the state of Rio de Janeiro, southeaster Brazil. Ethical approval for the use of these samples in evolutionary and demographic studies was provided by the Research Ethics Committee of the *Universidade Federal do Rio Grande do Sul* (Resolution no. 98002/1998). Afro-Brazilian samples were obtained from individuals who self-classified as 'black' based on their physical appearance.

In Colombia, Gabriel Bedoya and his collaborators (*Laboratorio de Genética Molecular, Universidad de Antioquia, Medellin, Colombia*) recruited participants from the Afro-Colombian population based in Quibdó (5°41'44.0"N, 76°38'52.9"W) in the Chocó department and Medellin (6°15'16.2"N, 75°34'29.1"W) in the Antioquia department; both departments are located in northwest Colombia. Ethical approval for the use of these samples was provided by the Ethical committee of the Institute of Biology of the Universidad de Antioquia.

A.2. West African populations

Populations with West African ancestry were analysed from different locations in Benin: the Bariba population in North Benin (Parakou region; 9°21'00.0"N, 2°37'12.0"E), the Yoruba population in Central Benin (Ketou region; 7°21'49.1"N, 2°36'21.4"E), and the Fon population in South Benin (Gbeto, Whydah; 6°22'12.0"N, 2°04'48.0"E, and Cotonou; 6°22'00.2"N,

2°25'48.6"E). In Ivory Coast, we analysed the Ahizi population in the southeast (Nigui-Saff; 5°15'00.0"N, 4°36'36.0"W), and the Yacouba population from western region (Danané, Glanlé; 7°15'50.5"N, 8°09'25.0"W). In Mali, the Bwa population from Cercle de Tominian in the Southeast (Segou region; 13°15'00.0"N, 4°25'00.0"W) was collected by Gil Bellis and André Chaventré (*Institut National d'Études Démographiques*, INED, Paris, France). Geographic locations of these populations are shown in Figure 37.
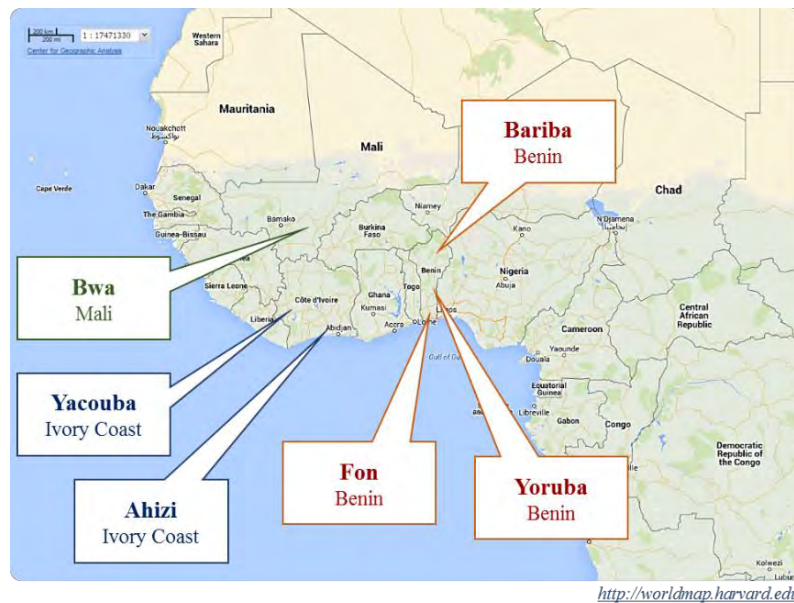
Figure 37: Geographic locations of studied West African populations.

A.3. *West European populations*

West European populations have played a significant role in South America during the transatlantic slave trade (Klein & Vinson, 2007), and their genetic identity might still be present in the African-American populations that were involved in Spanish and Portuguese colonisations. Hence, we extended the Y-chromosomal analysis to West European populations form the Iberian Peninsula (Figure 38A).

In Spain, we analysed samples from: Galician population in northwest Spain collected by Antonio Salas (*Unidade de Xenética, Instituto de Ciencias Forenses and Dept de Anatomía Patolóxica e Ciencias Forenses*, University of Santiago de Compostela, Galicia, Spain); Catalan population from Barcelona in northeast Spain collected by Pedro Moral (*Institut de Biodiversitat, Dept Biologia Animal-Antropologia, Facultat de Biologia*, University of Barcelona, Barcelona, Spain); and Andalusian populations of Huelva and Granada in south Spain collected by Rosario Calderón (D*epartamento de Zoología y Antropología Física*, Complutense University of Madrid, Madrid, Spain).

Portuguese samples from north and south Portugal were collected by Luisa Pereira (*Institute of Molecular Pathology and Immunology*, IPATIMUP, University of Porto, Porto, Portugal).

A.4. Native American populations

We also extended the Y-chromosomal analysis to three Amerindian groups in French Guiana (Figure 38B): Kalinya (or *Kaliña*), Palikur (or *Palikour*), and Oyampi (or *Wayampi*). These samples were collected by Georges Larrouy (*Laboratoire d'Anthropologie Moléculaire et Imagerie de Synthèse*, AMIS CNRS UMR-5288, University Paul Sabatier, Toulouse, France) and Stéphane Mazieres (*Anthropologie Bioculturelle, Droit, Éthique et Santé*, ADES UMR7268, University Aix-Marseille, Marseille, France). These populations spoke different languages: Cariban in the Kalinya population, Maipurean in the Palikur population, and Tupian in the Wayampi population (Paul et al., 2015). Each of them had different historical backgrounds.
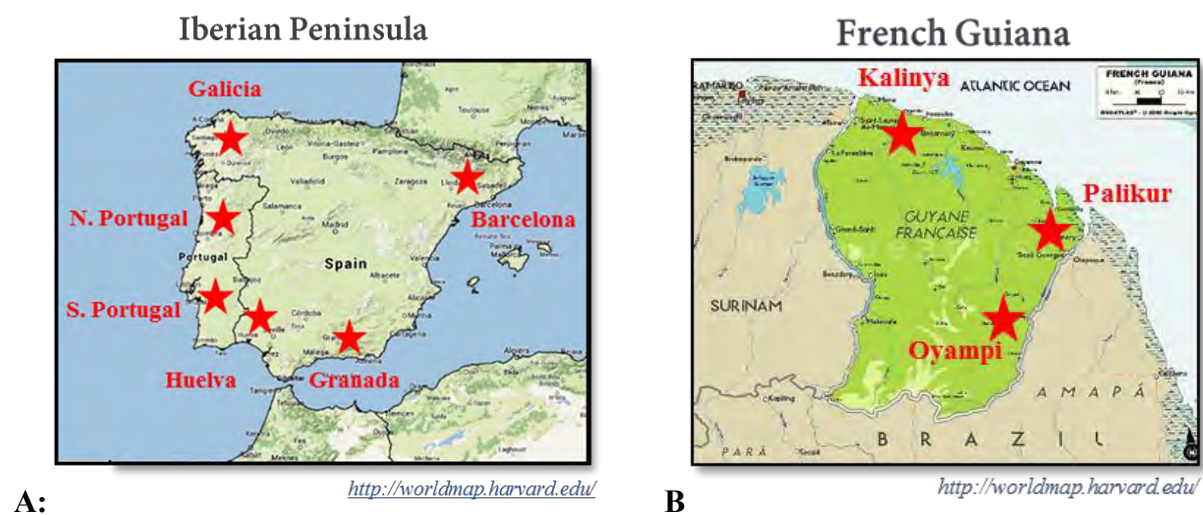
Figure 38: Geographic locations of studied West European populations from the Iberian Peninsula (**A**); and Native American populations from French Guiana (**B**).

A.5. DNA extraction

Peripheral blood samples were collected in tubes with the anticoagulant ethylenediaminetetraacetic acid (or EDTA). The genomic DNA of Noir Marron populations was extracted from 200 μl of blood samples by using the QIAmp Blood DNA Mini kit (made by Qiagen based in Courtaboeuf, France) according to the manufacturer instructions. This purification did not requires alcohol precipitation or phenol/chloroform methods, and is considerably faster.

Initially, DNA was diluted in QIAGEN Protease K buffer (free of DNase activity) to digest cell membranes and proteins; this dilution also helps to inactivate nucleases that might otherwise degrade the DNA during purification. Afterwards, four steps of purification procedures were carried out using QIAamp Mini spin columns in a standard microcentrifuge (Figure 39). The lysate buffering conditions were adjusted to allow optimal binding of the DNA to the silica-based membrane included in the columns. Wash conditions were performed using two different wash buffers, Buffer AW1 and Buffer AW2, to remove any residual contaminants without affecting DNA binding. Purified DNA was eluted from the QIAamp Mini spin column in standard 1.5 ml eppendorf tubes, and it was stored in a concentrated form with Tris-EDTA (TE) buffer at -25 C°.

Genomic DNA of remaining populations was extracted previously using the standard phenol/chloroform method (Gill et al., 1985), and was stored in a concentrated form with Tris-EDTA (TE) buffer at -25 C°.

Extracted DNA was quantified by optical density absorption measurement using a NanoDrop Spectrophotometer 2000C (NanoDrop products, Wilmington, DE), and normalised to a concentration of 5 ng/µl for Y-STR typing, 50 ng/µl for Y-SNP and autosomal DNA typing, and 10 ng/µl for mtDNA sequencing.
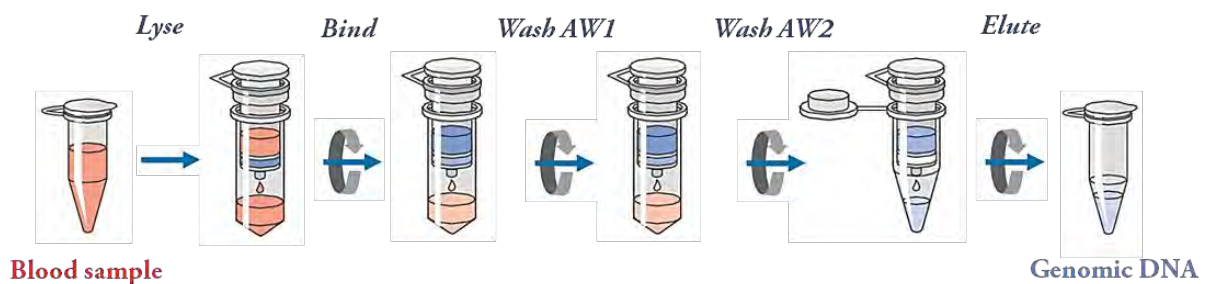
Figure 39: Genomic DNA was extracted using four microcentrifuge spins following the QIAamp Spin procedure.

A.6. *Datasets elaborated*

We analysed the African-American and West African samples by using three different genetic systems: Y chromosome (379 samples in total for Y-STR, and 346 samples in total for Y-SNP), whole mitochondrial DNA (273 samples in total), and autosomal DNA (229 samples in total) (see Table 9).

Table 9: Total population size analysed in the present study with Y chromosome (17 Y-STRs and 96 Y-SNPs), whole mitochondrial DNA genome, and autosomal DNA (4.5 million SNP), for African-American descendants and West African populations.

Region	Population	Y-STRs	Y-SNPs	mtDNA	Autosomal
French Guiana	*Aluku*	*8*	*16*	*15*	*23*
	Ndjuka	*21*	*11*	*17*	*23*
	Paramaka	*5*	*1*	*8*	*19*
	Saramaka	*16*	*2*	*3*	*6*
Total Noir Marron communities		50	30	43	71
Brazil	Afro-Brazilian	24	28	39	16
Colombia	Afro-Colombian	19	20	30	20
Total African Americans		**93**	**78**	**112**	**107**
Benin	Fon	78	63	36	19
	Yoruba	54	55	32	24
	Bariba	51	57	32	24
Ivory Coast	Ahizi	49	47	22	20
	Yacouba	41	32	16	17
Mali	Bwa	13	14	23	18
Total West Africans		**286**	**268**	**161**	**122**
TOTAL		**379**	**346**	**273**	**229**

For Y-chromosomal analysis, we used a wide number of genetic markers for each individual (96 Y-SNPs and 17 Y-STRs). Paternal genetic markers having such a high resolution and specificity for each individual is not currently available in other publications. Therefore, to study possible admixtures and gene flow from European and Amerindian populations to African Americans, we analysed individuals with West European and Native American ancestry with the same set of genetic markers (see Table 10).

In French Guiana, we compared the Noir Marron communities with Amerindian groups analysed in the present study. These Native American populations are descendants of original Native American groups who were present in French Guiana territories during the slave trade (Grenand & Grenand, 1985). Their geographical proximity made it very likely that there might be some evidence of gene flow between the Noir Marron communities and Native Americans during the slave trade, and even afterwards. If there is no gene flow, then this may further support the high genetic isolation hypothesis proposal in previous studies (Brucato et al., 2009; Brucato et al., 2010).

Additionally, we included Native American descendants from South America to analyse signals of gene flow in the past with their African American neighbours in Colombia and Brazil.

Yfiler profiles (17 Y-STRs) were obtained for the Emberá-Chamí population from Antioquia in Colombia (as reported by Roewer et al. (2013)), and for Karitiana population in Brazil (as reported by Xu et al. (2015)) (see Table 10). The Karitiana population is often used as a reference population for Native American populations from South America (Li et al., 2008).

Table 10: Summary of Y chromosome datasets for the African-American populations and assembled reference panels for West African, West European, and Native American ancestry.

Population	ID Pop	Sample size	Y-STR	Y-SNP
Noir Marron (F.G.) [1]	**GUF_NM**	**54**	**50**	**30**
Aluku	*NM_A*	*19*	*8*	*16*
Ndjuka	*NM_N*	*18*	*21*	*11*
Paramaka	*NM_P*	*5*	*5*	*1*
Saramaka	*NM_S*	*12*	*16*	*2*
Afro-Brazilian	**Af-BRA**	**35**	**24**	**28**
Afro-Colombian	**Af-COL**	**20**	**19**	**20**
Total African American	*AAM*	*109*	*93*	*78*
Benin	**BEN**	**198**	**183**	**175**
Bariba [2]	*BEN_B*	*59*	*51*	*57*
Fon [1]	*BEN_F*	*79*	*78*	*63*
Yoruba [2]	*BEN_Y*	*60*	*54*	*55*
Ivory Coast	**CIV**	**90**	**90**	**79**
Ahizi [1]	*CIV_A*	*49*	*49*	*47*
Yacouba [1]	*CIV_Y*	*41*	*41*	*32*
Mali - *Bwa*	**MLI_B**	**14**	**13**	**14**
Total West African	*AFR*	*302*	*286*	*268*
Kalinya (F.G.) [3]	*GUF_K*	*23*	*23*	*3*
Oyampi (F.G.) [3]	*GUF_O*	*29*	*25*	*3*
Palikour (F.G.) [3]	*GUF_P*	*43*	*41*	*7*
Emberá-Chamí (COL) [4]	*COL-EC*	*24*	*24*	*-*
Karitiana (BRA) [5]	*BRA-Ka*	*17*	*17*	*-*
Total Native American	*NAM*	*136*	*130*	*13*
Spain	**ESP**	**101**	**101**	**80**
Galicia	*ESP_Ga*	*30*	*21*	*21*
Barcelona	*ESP_Ba*	*21*	*30*	*26*
Granada [6]	*ESP_Gr*	*25*	*25*	*20*
Huelva [6]	*ESP_Hu*	*25*	*25*	*13*
Portugal	**PRT**	**30**	**30**	**30**
North Portugal	*PRT_N*	*7*	*7*	*7*
South Portugal	*PRT_S*	*23*	*23*	*23*
Total West European	*EUR*	*131*	*131*	*110*
TOTAL Sample Size		**678**	**640**	**469**

Note: Table competed with Yfiler profiles from [1] (Brucato et al., 2010), [2] (Fortes-Lima et al., 2015), [3] (Mazieres et al., 2011), [4] (Roewer et al., 2013), [5] (Xu et al., 2015), and [6] (Ambrosio et al., 2012).

B.   Y chromosome analysis

B.1.   Genotyping techniques for Y-STRs

We analysed Y-chromosomal microsatellites for 469 samples from: African American, West African, West European, and Native American populations (see Table 10). We carried out the Polymerase Chain Reaction (or PCR) amplification in the *Laboratory of Molecular Anthropology and Image Synthesis* (AMIS CNRS-UMR 5288), University Paul Sabatier, Toulouse, France. Haplotypes were genotyped with the widely used Y chromosome markers included in the AmpF*l*STR® Yfiler® PCR Amplification Kit (Life Technologies) (subsequently referred to as Yfiler kit), for genotyping 17 Y-STRs loci: DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS385 a/b, DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYS635, and GATA H4 (see Figure 40).

Figure 40: Relative positions of Y chromosome microsatellites present in the Yfiler kit. The 17 Y-STRs are presented in the NRY (Butler et al., 2012).
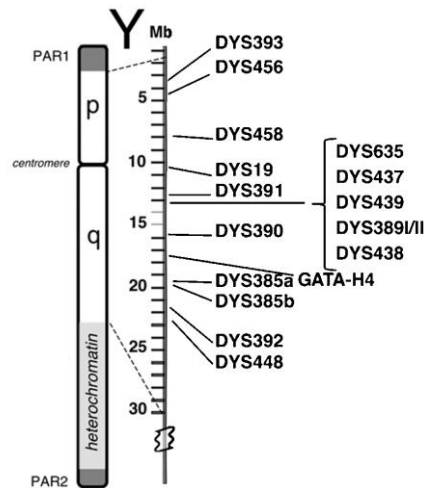
The PCR amplification of Yfiler kit was performed in a reaction volume of 6.25µl containing: 0.2µl of AmpliTaq Gold® DNA polymerase (Applied Biosystems), 2.3µl of Yfiler PCR reaction mix, 1.25µl of Yfiler primer set, and a maximum volume of 1µl of target DNA. The standard thermal cycling conditions in a Gene Amp® PCR System 2700 thermocycler (Applied Biosystems) consisted of enzyme activation at 95°C for 11 min, followed by 30 cycles of denaturation at 94°C for 1 min, annealing at 61°C for 1 min, and extension at 72°C for 1 min. A final extension was performed at 60°C for 80 min, with a 4°C temperature hold, if the PCR product was to remain in the thermal cycler.

The Yfiler kit employs a five-dye set system consisting of 6-FAM™, VIC®, NED™, PET®, and LIZ®, which makes it easier to analyse Y-STRs results in four colour's groups (i.e., blue, green, yellow, and red). Y-STRs in this kit present different tetranucleotide repeats and mutation rates that occur at high frequencies (see Table 11). They are commonly applied to forensics and evolutionary studies (Diegoli, 2015; Roewer, 2009).

Genotyping was carried out following the manufacturer's recommendations by using Applied Biosystems 3130xl Genetic Analyser (Life Technologies), which is provided by the genotyping services of the Genomic and Transcriptome platform GeT-Genotoul, Toulouse, France. Data analysis was performed in our laboratory by using GeneMapper® IDSoftware v.4.0 (Applied Biosystems, Inc.). We identified high peaks for each loci in each Y-STR for each individual. Allele peaks were interpreted when they were greater than 50 relative fluorescence units (RFUs). Figure 41 shows results obtained using GeneMapper® for the allelic ladder and for one individual. The Y-STR alleles were designated according to the ISFG recommendations (Gusmao et al., 2006). Only complete Y-STR profiles were used for subsequent analyses. DYS389II alleles were encoded by the difference between the total number of repeats at DYS389II and the number of repeats at DYS389I (henceforth was labelled as DYS389II-I).

Table 11: Summary of locus configuration of 17 Y-STRs typing using Yfiler kit. Bayesian median mutation rates were reported by Ballantyne et al. (2010).

Label	N° Alleles	Mutation rate	Repeat Structure	GeneBank Accession N°	Dye
DYS389I	9 - 17	$5.51\,E^{-03}$	$(TCTG)_3(\textbf{TCTA})_{\textbf{6-14}}$	AC004617	6-FAM
DYS389II	24 - 34	$3.83\,E^{-03}$	$\textbf{(TCTG)}_{\textbf{4-5}}\textbf{(TCTA)}_{\textbf{10-14}}N_{28}(TCTG)_3\textbf{(TCTA)}_{\textbf{6-14}}$	AC004617	6-FAM
DYS390	17 - 28	$1.52\,E^{-03}$	$(TCTG)_8(\textbf{TCTA})_{\textbf{9-4}}(TCTG)_1(TCTG)_4$	AC011289	6-FAM
DYS456	13 - 18	$4.94\,E^{-03}$	$\textbf{(AGAT)}_{\textbf{11-23}}$	AC010106	6-FAM
DYS19	10 - 19	$4.37\,E^{-03}$	$(TAGA)_3(TAGG)_1(\textbf{TAGA})_{\textbf{6-16}}$	AC017019	VIC
DYS385 a	7 - 28	$2.08\,E^{-03}$	$(AAGG)_4N_{14}(AAAG)_3N_{12}(AAAG)_3N_{29}(\textbf{AAGG})_{\textbf{6-7}}\textbf{(GAAA)}_{\textbf{7-23}}$	AC022486	VIC
DYS385 b	7 -28	$4.14\,E^{-03}$	$(AAGG)_4N_{14}(AAAG)_3N_{12}(AAAG)_3N_{29}(\textbf{AAGG})_{\textbf{6-7}}\textbf{(GAAA)}_{\textbf{7-23}}$	AC022486	VIC
DYS458	13 - 20	$8.36\,E^{-03}$	$\textbf{(GAAA)}_{\textbf{11-24}}$	AC010902	VIC
DYS391	7 - 13	$3.23\,E^{-03}$	$(TCTG)_3(\textbf{TCTA})_{\textbf{6-15}}$	AC011302	NED
DYS392	6 - 17	$9.70\,E^{-04}$	$\textbf{(TAT)}_{\textbf{4-20}}$	AC011745	NED
DYS393	9 - 17	$2.11\,E^{-03}$	$\textbf{(AGAT)}_{\textbf{7-18}}$	AC006152	NED
DYS439	9 - 14	$3.84\,E^{-03}$	$(GATA)_3N_{32}(\textbf{GATA})_{\textbf{5-19}}$	AC002992	NED
DYS635	20 - 26	$3.85\,E^{-03}$	$(TCTA)_4(TGTA)_2(TCTA)_2(TGTA)_2(TCTA)_2\textbf{(TATG)}_{\textbf{0-2}}\textbf{(TCTA)}_{\textbf{4-17}}$	AC004772	NED
DYS437	13 - 17	$1.53\,E^{-03}$	$\textbf{(TCTA)}_{\textbf{4-12}}(TCTG)_2(TCTA)_4$	AC002992	PET
DYS438	8 - 13	$9.56\,E^{-04}$	$\textbf{(TTTTC)}_{\textbf{7-16}}$	AC002531	PET
DYS448	20 - 26	$3.94\,E^{-04}$	$\textbf{(AGAGAT)}_{\textbf{11-13}}N_{42}\textbf{(AGAGAT)}_{\textbf{8-9}}$	AC025227	PET
Y-GATA H4	8 - 13	$3.22\,E^{-03}$	$(TAGA)_3N_{12}(TAGG)_3(\textbf{TAGA})_{\textbf{8-15}}N_{22}(TAGA)_4$	AC011751	PET

Figure 41: Electropherograms show the entire allelic ladder for each loci of each Y-STR analysed (**up**), and profile of one individual (**down**) amplified with the Yfiler kit (see Table 11). Both were obtained using GeneMappers ID software v.4.0. The four rows correspond to 6-FAM, VIC, NED, and PET dye-labeled peaks. The haplotype is showing with the allele number displayed underneath each peak.
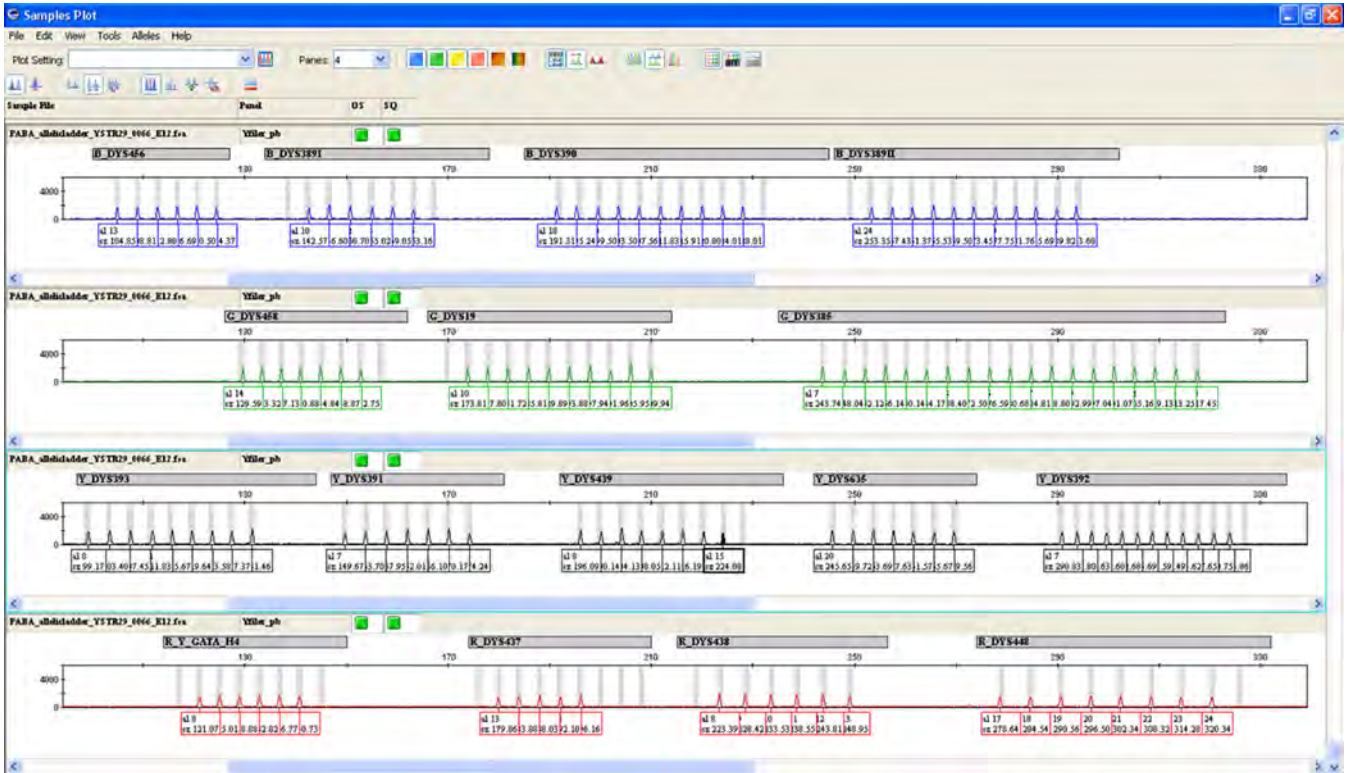
B.2. Genotyping techniques for Y-SNPs

We genotyped 96 Y-SNPs (see Table S1) for 446 samples from: African American, West African, West European, and Native American populations (see Table 10). For each Y-SNP, three primers were designed (per strand) by using Primer3 (Untergasser et al., 2012): allele specific 1 (AS1), allele specific 2 (AS2), and common reverse. Designs were checked for specificity by In-Silico PCR (http://genome.ucsc.edu/cgi-bin/hgPcr). The two tailed AS-PCR primers and the common reverse were ordered for synthesis to Integrated DNA Technologies, Inc. (Leuven, Belgium).

We carried out the AS-PCR (Myakishev et al., 2001) by using the BioMark™ HD system (Fluidigm Corporation, USA) in a microfluidic multiplex array chip. This high-throughput genotyping system was previously applied for the Malagasy and Indonesian populations by Kusuma et al. (2015). The system is developed for SNP genotyping assays and able to perform 9,216 Real-Time PCRs on a single chip (96 primers × 96 samples). The results were provided by the platform IntegraGen (Genopole Campus, Evry, France).

The 96-plex was done for all individuals to avoid any analytical bias. A 14-cycle pre-amplification reaction was performed for each sample in 5µl by pooling 96 common primer pairs, 1.25µl genomic DNA, and 2.5µl 2x QIAGEN Multiplex PCR Master Mix (QIAGEN). For each individual assay, 5µl 10X Assay Mix containing 100X Primers mix (final concentration 10X, composed of AS1, AS2 and common reverse), and 1X Assay Loading Reagent were loaded into one of the Assay Inlets on the chip.

The following solution (5µl) was loaded in sample inlets: 1µl Preamplified sample previously diluted to 1:5 in low TE Buffer, 2.5µl 2X Fast probe QPCR master mix (Biotium), 0.25µl 20X Gene Expression Sample Loading Reagent (Fluidigm), 0.07µl 50X Rox Reference Dye, and 0.11µl 50X Allele Specific Universal primer.

The Biomark's specific cycling program was used to amplify fragments:

Hot start at 95°C for 5 min.

PCR 6 Touch Down cycles;	95°C 15sec → 72°C 45sec (-2°C/cycle) →	72°C 15sec
PCR 30 cycles ;	95°C 15sec → 60°C 45sec	→ 72°C 15sec

Finally, genotypes were obtained by clustering by using the Fluidigm SNP Genotyping Analysis v.4.1.3 based on the k-means clustering analysis method (Wang et al., 2009). Y-SNP haplogroups were designated according to the current indications of ISOGG Y-DNA Haplogroup Tree 2015 (Y Chromosome Consortium, 2002) (updated on January 2015, http://www.isogg.org/tree/). Table S1 includes the list of studied 96 Y-SNP markers, the mutation types of each SNP, its haplogroup affiliations, and further information about the

sequences amplified.

We elaborated a comprehensive phylogenetic tree for Y haplogroups that can be identified with the 96 Y-SNPs genotyped (see Figure 42); special attention was given to Unique Event Polymorphism (or UEP) from African ancestry (Ansari Pour et al., 2013; Tishkoff et al., 2007). Y haplogroups and internal nodes following the latest classification proposed for ISOGG Y-DNA Haplogroup Tree 2015 (updated on April 2015) (Y Chromosome Consortium, 2002), and were compared with the minimal reference phylogeny for the human Y chromosome (update version on November 2014) (van Oven et al., 2014).

We analysed 446 samples by using the 96-plex genotyping described above. Additionally, the Y haplogroups affiliation was identified for 23 samples by using genome-wide SNP data described in section II.D.1. We used PLINK v.1.9 (Chang et al., 2015; Purcell et al., 2007) to extract the same set of Y-SNPs markers, which are localised in the same position of Y chromosome and present the same reference SNP ID number (or rs) (see Table S1).

B.3. *Statistical analysis of Y chromosome*

With the entire Y-SNP database obtained for 469 samples, we estimated Y haplogroup frequencies for each population and elaborated a Y chromosome tree. Molecular diversity indices were estimated for each population and each ancestry group by using Arlequin v.3.5.2.1 software (Excoffier & Lischer, 2010). Haplotype diversity (HD) was calculated for each population sample by using the following equation:

$HD = (1 - \sum q_i^2) (n/n - 1)$, where n is the sample size and q_i is the haplotype frequency.

Gene diversity (h) is equivalent to HD but referred to a single locus; and h was also computed as an average over loci. The discriminatory capacity (DC) was determined by dividing the number of different haplotypes by the total number of samples in a given population (Kayser et al., 1997). We estimated intra- and inter-population molecular diversity index one for seventeen Yfiler markers set, another one for the set of tetra- and trimeric Y-STRs set constituting the 'minimal haplotype' or minHt (i.e., DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, and DYS393; without DYS385 a/b) (Kayser et al., 1997).

Figure 42: Phylogeny Y chromosome tree obtained with 96 Y-SNPs analysed, and based on ISOGG Y-DNA Haplogroup Tree 2015. Genotyped Y-SNPs are indicated in blue. SNPs with phyloequivalent alternatives are shown in parentheses. Recurrent SNPs are indicated with a # suffix. E1b1 and K2b branches are shown in separated boxes.
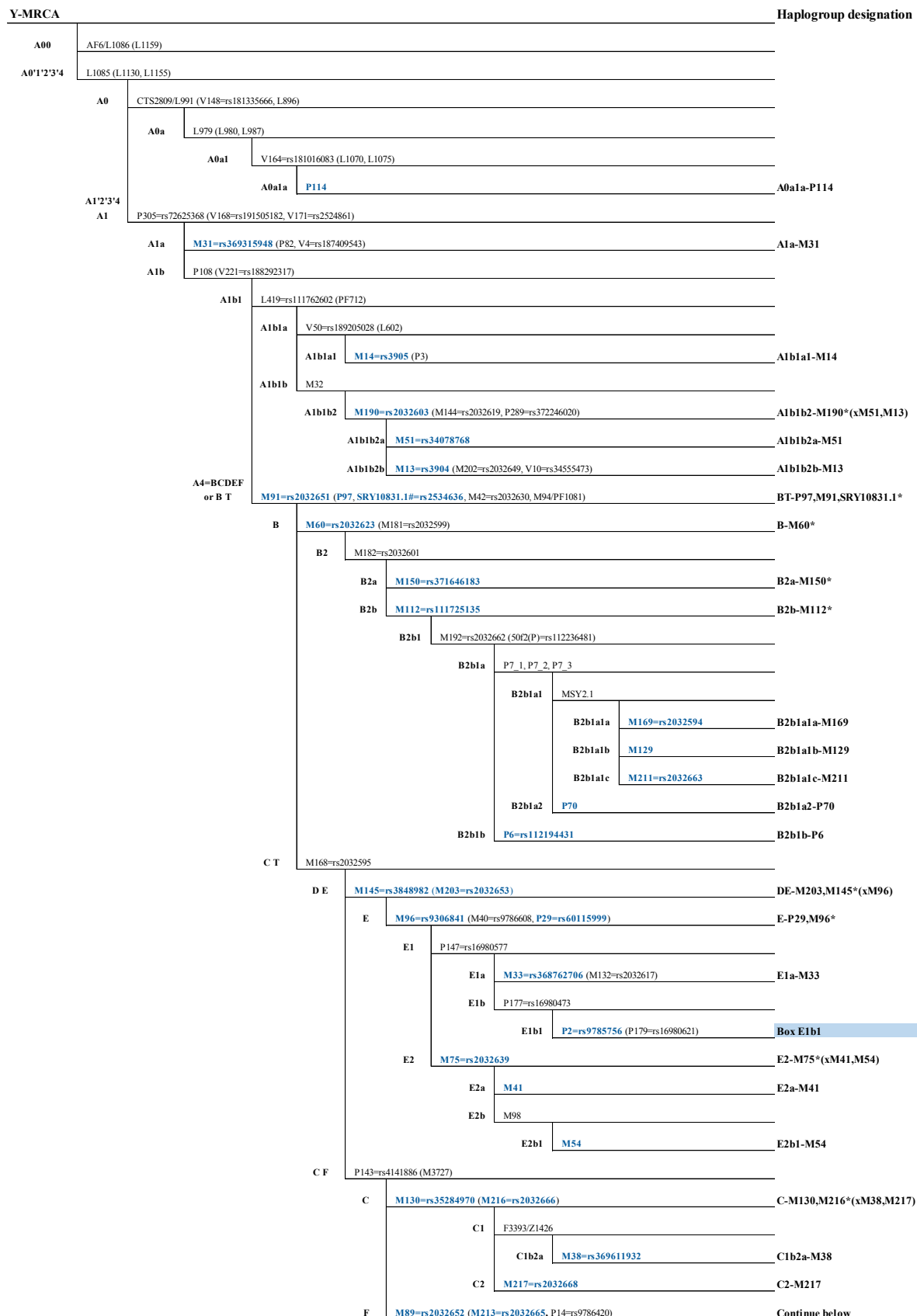
F   M89=rs2032652 (M213=rs2032665, P14=rs9786420)   **F-M89,M213***

G H I J K   F1329/M3658/PF2622

     G   M201=rs2032636 (P257=rs2740980)   **G-M201*(xP287)**

       G2   P287=rs4116820   **G2-P287*(xP15)**

         G2a   P15=rs370167410   **G2a-P15**

     H I J K   M578=rs73614810

       H   M3035=rs74378870 (L901)

         H1   M69=rs2032673 (M370, L902)   **H-M69*(xM52)**

           H1a   M52=rs376769460   **H1a-M52*(xM82)**

             H1a1   M82=rs2032675   **H1a1-M82**

       I J K   M522=rs9786714 (M523=rs9786139)

         I J   M429=rs17306671 (P126=rs17250163)

           I   M170=rs2032597 (M258=rs9341301, U179=rs2319818)

             I1   P203.2#=rs13447354 (M253=rs9341296, L80=rs35960273)   **I1-P203.2**

           J   M304=rs13447352 (P209=rs17315835)   **J-P209,M304*(xM267,M172)**

             J1   M267=rs9341313   **J1-M267***

               J1a   Z2215=rs376267849

                 J1a2   L136

                   J1a2b   P58=rs34043621   **J1a2b-P58**

             J2   M172=rs2032604 (L228=rs371968167)   **J2-M172*(xM410,M12)**

               J2a   M410=rs371079691   **J2a-M410***

                 J2a1   L26=rs34459399

                   J2a1b   M67=rs2032628   **J2a1b-M67**

               J2b   M12=rs3903   **J2b-M12*(xM241)**

                 J2b2   M241=rs8179022   **J2b2-M241***

         KLT   M9=rs3900 (P128/PF5504=rs17250121)   **KLT-M9,P128/PF5504***

           L T (or K1)   L298/P326=rs372687543 (L811)

             L (or K1a)   M20/PF5570=rs3911   **L-M20/PF5570***

               L1   M295   **L1-M295*(xM76,M357)**

                 L1a   M2481

                   L1a1   M76=rs377001539 (M27=rs376303746)   **L1a1-M76**

                   L1a2   M357=rs377156966   **L1a2-M357**

             T (or K1b)   M184=rs20320 (M272/PF5667=rs9341308)   **T-M184,M272/PF5667***

               T1   M193=rs2032676 (L206)

                 T1a   M70=rs2032672   **T1a-M70*(xL131)**

                   T1a2   L131=rs2215828   **T1a2-L131**

          K2   M526=rs2033003   **K -M526***

            NO (or K2a)   M214=rs2032674

              O (or K2a2)   P186=rs16981290 (M175=rs2032678)   **O-P186***

               O1   MSY2.2

                 O1a   M119=rs72613040   **O1a-M119*(xP203,M50)**

                   O1a1   P203.1#=rs13447354   **O1a1-P203.1**

                   O1a2   M50=rs2032632 (M103=rs2032609)   **O1a2-M50**

               O2   M268=rs13447443 (P31=rs200861659)

                 O2a   PK4=rs367562925

                   O2a1   M95=rs2032650   **O2a1-M95*(xM88)**

                     O2a1a   M88=rs2032645 (M111)   **O2a1a-M88**

               O3   M122=rs78149062   **O3-M122***

                 O3a   M324=rs13447361 (P197=rs17276358, P200=rs17316592)

                   O3a2   P201=rs2267801   **O3a2-P201***

                   O3a2b   M7=rs3898   **O3a2b-M7**

                   O3a2c   P164=rs17316007

                     O3a2c1   M134=rs200634940   **O3a2c1-M134**

          K2b   M1221/P331/PF5911   **Box K2b**

E1b1   P2=rs9785756 (P179=rs16980621)   **E1b1-P2*(xV38,M215)**

   E1b1a   V38=rs768983 (L222.1#)   **E1b1a-V38*(xM2)**

      E1b1a1   M2=rs9785941 (P1=rs73626946, P189#=rs9786819, V95)   **E1b1a1-M2***

         E1b1a1a1   M180=rs2032598

            E1b1a1a1c   L485=rs9786118

               E1b1a1a1c1   L514=rs9786371

                  E1b1a1a1c1a   M191=rs2032590   **E1b1a1a1c1a-M191*(xU174)**

                     E1b1a1a1c1a1   U174=rs16980586   **E1b1a1a1c1a1-U174*(xP115,P116)**

                        E1b1a1a1c1a1b   P115   **E1b1a1a1c1a1b-P115**

                        E1b1a1a1c1a1c   CTS8030/Z1704

                           E1b1a1a1c1a1c1a   P116   **E1b1a1a1c1a1c1a-P116**

               E1b1a1a1d   U175=rs16980588   **E1b1a1a1d-U175*(xU209)**

                  E1b1a1a1d1   U209=rs16980502   **E1b1a1a1d1-U209*(xU290,M154)**

                     E1b1a1a1d1a   U290=rs16980406   **E1b1a1a1d1a-U290*(xU181)**

                        E1b1a1a1d1a1   U181=rs16980589   **E1b1a1a1d1a1-U181**

                     E1b1a1a1d1c   M154   **E1b1a1a1d1c-M154**

   E1b1b   M215=rs2032654   **E1b1b-M215*(xM35.1)**

      E1b1b1   M35.1=rs375228668   **E1b1b1-M35.1***

         E1b1b1a   V68=rs373648041

            E1b1b1a1   M78=rs368977028   **E1b1b1a1-M78**

         E1b1b1b   Z827=rs367988033

            E1b1b1b1   M310=rs13447357

               E1b1b1b1a   M81=rs2032640   **E1b1b1b1a-M81**

            E1b1b1b2   Z830=rs369637510

               E1b1b1b2a   M123=rs371143248   **E1b1b1b2a-M123*(xM34)**

                  E1b1b1b2a1   M34=rs373666971   **E1b1b1b2a1-M34**

K2b   M1221/P331/PF5911

   K2b1   P397

      M (or K2b1d)   P256 (Page93=rs34486382)   **M-P256*(xM186)**

         M1   M186=rs2032681 (M4=rs3895, M5=rs3896, M106=rs2032611)   **M1-M186**

   P (or K2b2)   P295/PF5866/S8

      P1   M45=rs2032631 (M74=rs2032635)   **P1-M45*(xM207)**

         Q   M242=rs8179021   **Q-M242**

         R   M207=rs2032658   **R-M207*(xM173)**

            R1   M173=rs2032624 (M306=rs1558843)   **R1-M173*(xM434)**

               R1a   M434 (M420=rs17250535)   **R1a-M434***

                  R1a1   SRY10831.2# (M459)   **R1a1-SRY10831.2**

                     R1a1a   M17=rs3908 (M198=rs2020857)   **R1a1a-M17***

                        R1a1a1   M417=rs17316771

                           R1a1a1b   Z645/S224=rs111731595 (Z647/S441=rs112284571)

                             R1a1a1b1   Z283/S339=rs112309702

                               R1a1a1b1a1   M458=rs375323198   **R1a1a1b1a1-M458**

               R1b   M343=rs9786184

                  R1b1   M415=rs9786194   **R1b1-M415***

                     R1b1a   P297=rs9785702

                        R1b1a2   M269=rs9786153   **R1b1a2-M269**

            R2   M479=rs372157627

               R2a   M124=rs372706460   **R2a-M124**

Patterns of haplotype variations within the African-American populations with different continental ancestry were investigated applying the Median-joining (or MJ) Network method (Bandelt et al., 1999). Phylogenetic relationships of Y-STR haplotypes were elaborated using Network program v.4.6.1.3 (Fluxus Technology Ltd); the relationships were based on nine Y-STRs (DYS19, DYS389I, DYS389II-I, DYS390, DYS391, DYS392, DYS393, and DYS385a/b). We assigned weights to each individual Y-STR locus as inversely proportional to the variance observed in our dataset (Berniell-Lee et al., 2009). Individuals that had STR missing values were excluded from the analysis.

Pairwise F_{ST} and R_{ST} genetic distances between populations

We estimated genetic differentiation between pairwise populations by using F_{ST} distances based on Y-SNPs frequencies. Although for microsatellite data is necessary statistics based on models that take into account features of microsatellite evolution. We therefore calculated R_{ST} distances between pairwise populations based on twelve Y-STRs. These distances are analogous to F_{ST} although based on mutational dynamics of microsatellites related to the stepwise model (Slatkin, 1995), where the size of new mutant alleles depends on its progenitor (Di Rienzo et al., 1994). This measure allows more accurate calculations of population genetic parameters based on microsatellite data (Goodman, 1997; Rousset, 1996). Finally, we calculated a matrix of coancestry coefficients defined by Reynolds et al. (1983) as t/N= -ln (1-F_{ST}), based on twelve Y-STRs. Both genetic distances, F_{ST} and R_{ST} were calculated using Arlequin with 10,000 permutations (Excoffier et al., 1992).

B.3.a. *Admixture estimation of paternal ancestry*

We quantified genetic contributions of continental parental sources (African, European, and Native American) among the African Americans by using three admixture-model methods. First, we estimated genetic contributions based on Y haplogroup frequencies (mY_{SNP}) using ADMIX program v.2.0 (Dupanloup & Bertorelle, 2001), which takes mutation and sampling error into account. This analysis employed the coalescent-based approach proposed by Bertorelle & Excoffier (1998). Time since admixture was set to 450 years, without taking into account molecular distances between Y haplogroups. The bootstrap procedure was set to 10,000 repetitions. Mutation rate for Y-SNPs was set to 1.0×10^{-9}/bp/year estimated from NGS analysis of MSY sequences in a deep-rooting pedigree (Xue et al., 2009). Input files for admixture software have been created using AdFiT v.1.7 software (Gourjon et al., 2014).

Second, in the Lineage sharing (or LS) analysis (Trejaut et al., 2014) we analysed shared STR lineages between populations based on a set of seven Y-STRs (i.e., DYS19, DYS389I,

DYS389II-I, DYS390, DYS391, DYS392, and DYS393). We used Arlequin software to calculate the number of shared haplotypes between each parental source and each African-American population, and corrected for differences in sample size by dividing by the number of different haplotypes present in each parental population. These values were normalised to obtain the relative contribution of each parental population (Maca-Meyer et al., 2004).

Finally, to compare estimations calculated with different methods, we evaluated admixture proportions based on twelve Y-STRs by using the STRUCTURE software v.2.3.4 (Pritchard et al., 2000). To estimate the ancestral membership proportions, a supervised analysis was performed by using prior information on the geographic origin of the reference samples from Africa, Europe, and South America. A tri-hybrid contribution was assumed for African, European, and Native American ancestries. The STRUCTURE runs comprised three replicates of 10,000 burning steps followed by 1,000 Markov Chain Monte Carlo (MCMC) iterations (Falush et al., 2003). This approach has been previously applied for autosomal microsatellites to study admixtures among the African Americans (Falush et al., 2003; Tishkoff et al., 2009), and also for comparison of Y haplotypes among Y haplogroups (Wang et al., 2015).

B.3.b.   *Geographic patterns of Y chromosome in Africa*

To characterize subdivisions in populations with different continental ancestry, we carried out the Discriminant Analysis of Principal Components (or DAPC) (Jombart et al., 2010) based on seventeen Y-STRs. We analysed African-American populations and populations with different continental ancestry (Table 10). This method has been previously applied for worldwide populations to identify clusters between group structures and to unravel complex population structures (Jombart et al., 2010). We also represented relationships among African American and African coastal regions involved in the transatlantic slave trade by the PCA analysis based on Y haplogroup frequencies. We used the statistical software R (R Core Team, 2014) and the packages: *ade4* (Dray & Dufour, 2007) for the PCA analysis; *adegenet* (Jombart, 2008) for the DAPC analysis; and *ggplot2* (Wickham, 2009) for plotting both analyses.

Finally, to determine the geographical origin of African descendant populations, we plotted contour maps for geographic patterns of Y haplogroups across African and African-American populations in South America. We analysed frequencies of three Y haplogroups highly represented in West and West-Central African populations (E1b1a1-M2*, E1b1a1a1c1a-M191, and E1b1a1a1d-U175). We applied the Kriging method (Chiles & Delfiner, 2008), and we used SURFER programme v.12 (Golden Software, Inc.).

C. Mitochondrial DNA analysis

C.1. Sequencing techniques of mtDNA

We analysed the whole mitochondrial genome for the following 273 samples: *i*) six African-American populations: 43 Noir Marron individuals from Aluku, Ndjuka, Paramaka, and Saramaka communities, 30 Afro-Colombians, and 39 Afro-Brazilians; and *ii*) six West African populations: 22 Ahizi, 16 Yacouba, 23 Bwa, 36 Fon, 32 Bariba, 32 Yoruba (Table 9).

Libraries of mtDNA were pooled and sequenced according to the manufacturer's instructions for Single Read Multiplex sequencing (Maricic et al., 2010) by using a multiplex method developed for the Illumina Genome Analyzer IIx platform (Meyer & Kircher, 2010), with post processing step by using the Illumina software followed by the Improved Base Identification System (or IBIS, (Kircher et al., 2009)). IBIS considerably reduced the error rate and increased the output of millions of short sequencing reads generated in the Illumina Genome Analyzer. The run was processed with RTA v.1.5. (Illumina Inc.). The high-throughput sequencing was carried out at the *Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology*, Leipzig, Germany, in collaboration with Prof. Mark Stoneking and Dr. Roland Schröder. We applied the criteria proposed by Li et al. (2010) to detect heteroplasmy, and to avoid false positives arising because of sequencing errors in complete human mtDNA genomes from the reads generated by the Illumina GAII technology. The Binary Alignment/Map files (or BAM files), containing the mapping information and reads, were processed by using MitoBam Annotator (Zhidkov et al., 2011) and SAMTOOLS v.1.0 (Li et al., 2009). Both software provide various tools for manipulating alignments, including sorting, merging, indexing, generating alignments, and creating a consensus sequence. Mitochondrial genomes were assembled separately for each library, corresponding to an individual tag. We estimated frequencies of heteroplasmic sites and structural variance of the mitochondrial genome by using MitoSeek v.1.3 (Guo et al., 2013). Furthermore, we used a custom Perl script to estimate the number of unpaired and paired reads, the length of the consensus mtDNA sequence, and the minimum, maximum, and average coverage per position.

The reads for each samples were mapped to the rCRS (GenBank accession number NC012920.1; (Andrews et al., 1999)) by using the iterative mapping assembler MIA (Green et al., 2008). By convention, the nucleotide positions of each mtDNA genome were numbered from 1 to 16,569 according to the rCRS. Multiple sequence alignments of mtDNA genomes with rCRS were carried out using Clustal Omega tool (Sievers et al., 2011), which is available on the EMBL-EBI bioinformatics website (http://www.ebi.ac.uk/Tools/msa/clustalo/) (Li et al., 2015). We classified the mtDNA polymorphisms present in each mtDNA genome by using the

mtDNA GeneSyn computer tool (Pereira et al., 2009). This computer tool identifies the positions that are variable relative to the rCRS.

Finally, sequences were assigned to the mtDNA haplogroup affiliation for the whole genome and the HVS-I and HVS-II segments according to Phylotree.org Build 16 ((van Oven & Kayser, 2009); available online at http://www.phylotree.org/). We used the Haplogrep tool ((Kloss-Brandstatter et al., 2011); available online at https://haplogrep.uibk.ac.at/) and a custom Perl script. Sequences were assigned to the closest matching haplogroup (with more than 85% haplogroup predicted score) for which all mutations that defined the haplogroup were observed in that sequence. We also considered private mutations present in some individuals, and expected but missing mutations for that haplogroup.

C.2. *Statistical analysis of mtDNA*

We used mtDNA GeneSyn software v.1.0 (Pereira et al., 2009) to convert sequences into haplotypes, and also to calculate the following: *i*) the number of non-synonymous and synonymous mutations in the protein-coding genes; and *ii*) the number of mutations in the r-RNA genes, t-RNA genes, and noncoding regions.

For mtDNA subhaplogroups and major clades, we estimated mtDNA frequencies for each population and we elaborated a comprehensive mtDNA tree with detected mtDNA haplogroups. Mitochondrial sequence diversity indices were estimated using Arlequin v.3.5.2.1 software (Excoffier & Lischer, 2010). We carried out the Tajima's D test or Neutrality test. This test, described by Tajima (1993), compares two estimators of the population parameter θ: one is based on the number of segregating sites in the sample (or θ S), and the other is based on the mean number of pairwise differences between haplotypes (or θ Pi). This test is based on the infinite-site model without recombination; which is appropriate for mtDNA sequences. Significant *P*-values can be because of factors such as population expansion, bottlenecks, or heterogeneity of mutation rates, rather than selective effects (Tajima, 1996).

We estimated the frequencies of major mtDNA haplogroups L0-L3 based on segments HVS-I+HVS-II for our dataset of African-American and West African populations, as well as other African populations across the continent and Near Eastern populations reported in previous studies (see Table S7). Moreover, we estimated mtDNA frequencies for the African-American populations from North America (AWS in USA) and the Caribbean (ACB in Barbados) by using mtDNA genomes reported by (1000 Genomes Project Consortium, 2015).

We computed genetic distances between pairs of populations (F_{ST} with 10,000 permutations) based on mtDNA haplogroup frequencies using Arlequin. The segments of the mtDNA control region considered corresponded to the same range previously used by mtDNA studies: the entire HVS-I (positions 16012–16400), and part of HVS-II (positions 73–263) (Bandelt et al., 2006; Brucato et al., 2010). We represented relationships among the African-American and African populations residing in historic regions involved in the transatlantic slave trade by the PCA analysis based on mtDNA haplogroup frequencies. We used the statistical software R (R Core Team, 2014) and the packages: *ade4* (Dray & Dufour, 2007) for the PCA analysis; and for plotting, we used *ggplot2* (Wickham, 2009).

C.2.a. *Phylogenetic reconstruction of African mtDNA lineages*

In order to study the phylogeographic origins of the maternal lineages of the African-American populations belonging to the major African mtDNA haplogroups L, we compiled an exhaustive worldwide dataset containing 2,558 complete mtDNA genomes from the haplogroups L0 to L3. We assembled this dataset by using published mtDNA genomes in reference sequence databases, such as GenBank® (Benson et al., 2013), the 1000 Genomes Project (1000 Genomes Project Consortium, 2015), and unpublished mtDNA genomes from other studies in process. Additionally, we included 263 new whole mtDNA genomes belonging to the macrohaplogroup L0-L3 obtained for the African-American and West African populations.

We built four comprehensive phylogeographic trees for each haplogroup L (e.g. L0, L1, L2, and L3) by using complete mtDNA genomes. The branching structure was established and based on the reduced median algorithm (Bandelt et al., 1995) by using the Network program v.4.6.1.3 (Fluxus Technology Ltd). Branching order was manually constructed using Excel (Microsoft, Inc.). This approach allowed us to create new subbranches in terminal parts of the tree that are not present in the Phylotree.org Build 16 (van Oven & Kayser, 2009).

We estimated the splitting time of each nodes and clades in the phylogenetic tree by using three calibrations methods: *i*) *rho* statistic for complete mtDNA genome rate, *ii*) *rho* statistic for synonymous mutations rate, and *iii*) maximum likelihood (or ML) for complete mtDNA genome rate. The *rho* (or ρ) statistics estimates the average of mutational steps from a given ancestral node to the tips of the phylogeny purely based on a given mutation rate; it does not including any evolutionary model (Forster et al., 1996). We applied the *rho* statistic by using a mutation rate estimate for the complete mtDNA sequence of one substitution in every 3,624 years, further corrected for purifying selection by using the mtDNA clock provided by Soares et al. (2009). We also applied the *rho* statistic by using a synonymous mutation rate of one

substitution in every 7,884 years (Soares et al., 2009). Finally, we estimated branch ages by ML using PAML v.3.13 (Yang, 2007), assuming the HKY85 mutation model with gamma-distributed rates (approximated by a discrete distribution with 32 categories) and the same whole-mtDNA genome clock (Soares et al., 2009). For each calibration, we estimated standard errors with the Saillard method (Saillard et al., 2000). These calibration methods have recently been applied successfully to mtDNA trees of other mammalian species (Soares et al., 2013) and to evaluate the timing of the spread of the first humans in Africa (Fortes-Lima et al., 2014; Rito et al., 2013).

C.2.b. *Geographic patterns of mtDNA in Africa*

To determine the geographical origins of the African-American populations, we generated spatial frequency distribution maps of the major African haplogroups and subhaplogroups L (i.e., L0, L0a, L0a1b, L0b, L0d, L0f, L0k, L1, L1b, L1b2, L1c, L2, L2a, L2a1a3, L2b, L2b1, L2c, L2c1, L2d, L2e, L3, L3a, L3b, L3c, L3d, L3e, L3e1, L3e2, L3e2b, L3f, L3h, L3i, and L3x). The dataset elaborated is based on HVS-I and HVS-II data, and it included 17,687 individuals distributed across 51 countries across Africa and the Near East; it also included six African-American populations from French Guiana, Brazil, and Colombia (see Figure 43 and Table S7). These frequencies provided the most comprehensive up-to-date picture of the genetic landscape in the African continent and African Diaspora. Each contour map was generated based on the Kriging method for interpolations in such maps (Chiles & Delfiner, 2008), which were graphically plotted using the SURFER software v.12 (Golden Software, Inc.).

To obtain a better understanding of the geographic patterns of major clades and subclades across Africa, we combined the results obtained from the spatial distribution frequency analysis with the phylogenetic analysis. By using this approach, we shed new light about the African origin of the African-American populations in South America.
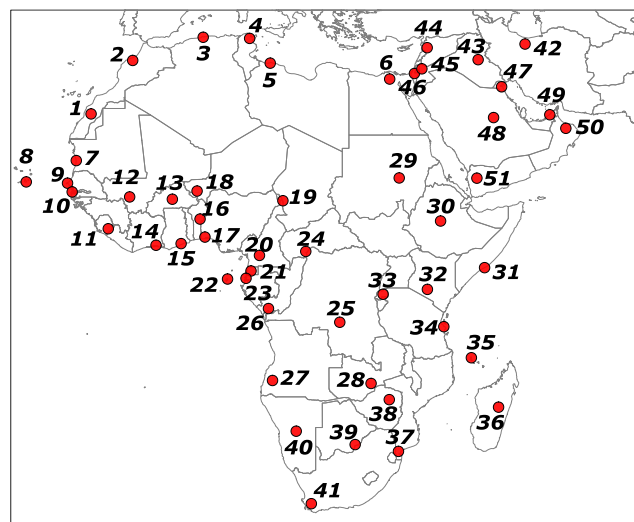
Figure 43: Geographic location of African and Near Est countries analysed using mtDNA position HVS-I and HVS-II (see Table S7).

We quantified genetic contributions of continental African, European, and Native American maternal sources (mY_{mtDNA}) among the African Americans based on mtDNA haplogroup frequencies using the ADMIX program v.2.0 (Dupanloup & Bertorelle, 2001). We analysed the African-American and West African populations indicated in the Table 9, in addition with West European and Native American frequencies were taken from Marques et al. (2015) and Mazieres et al. (2008), respectively. Likewise, we estimated ancestral contribution of historical African coastal regions (Figure 5). Each African region was determined by using an up-to-date database of African populations analysed with HVSI+HVSII (Table S8). Time since admixture was set to 450 years, without taking into account molecular distances between mtDNA haplogroups. The bootstrap procedure was set to 10,000 repetitions. Mutation rate for mtDNA was set to 2.14×10^{-8} /bp/year estimated for modern and ancient mtDNA genomes (Rieux et al., 2014). Input files for admixture software have been created using AdFiT v.1.7 software (Gourjon et al., 2014).

D.   *Genome-wide SNP data analysis*

The majority of previous studies addressing ancestry estimates among the African-American populations were based on uniparental markers (Benn Torres et al., 2007; Madrilejo et al., 2015; Simms et al., 2012). To evaluate admixture proportions among the African-American populations, we used autosomal markers because autosomes undergo recombination in every generation; they record the histories of both, male and female lineages. It is expected that the ancestry proportions captured correspond approximately to the average values indicated by mtDNA and Y chromosome markers.

D.1. *Genotyping techniques of genome-wide SNP data*

To estimate the continental ancestral contribution from African and non-African groups to African American communities, we analysed genome-wide SNP data for the following 229 samples: *i*) six African-American populations; 71 Noir Marron people from people from Aluku, Ndjuka, Paramaka, and Saramaka, 20 Afro-Colombian, and 16 Afro-Brazilian; and *ii*) six West African populations; 20 Ahizi, 17 Yacouba, 18 Bwa, 19 Fon, 24 Bariba, and 24 Yoruba (Table 9). We genotyped 4,301,332 SNP markers per sample (including autosomal DNA, sexual chromosomes, and mitochondrial DNA). This high-throughput SNP genotyping assay delivers robust high-quality genotyping data (Steemers et al., 2006) that allow to create a fine population-scale structure (Lachance & Tishkoff, 2013).

We used *Illumina HumanOmni5 Quad BeadChips^{TM}*, which is consider as the most powerful

whole-genome array (with the exception of whole genome sequencing) (Figure 30). The genotyping was performed using the *Illumina Infinium Assay*^{TM} (Gunderson et al., 2005) in the SNP&SEQ Technology platform at Uppsala University, Sweden. The results and genotyping quality were analysed by using *GenomeStudio*^{TM} software v.2011.1 (Illumina, Inc). DNA strand positions were built according to the Genome Reference Consortium Human genome build 37 (or GRCh37). Three individual were genotyped twice to quantify the reproducibility of this genotyping.

D.2. *Statistical analysis of genome-wide SNP data*

D.2.a. *Quality control procedure and assembled datasets*

In order to select autosomal SNPs that are highly informative of continental ancestry, we performed genotyping quality control procedure (or QC) of genome-wide SNP data. The QC was carried out using PLINK v.1.9 (Chang et al., 2015; Purcell et al., 2007), according to the protocol published by Anderson et al. (2010). Initially, we removed SNPs with high genotype error (GENO > 1%), individuals with high missing genotype rates (MIND > 1%), and monomorphic alleles (MAF < 5%). We performed the Hardy-Weinberg test statistics for each SNP (Wigginton et al., 2005). Finally, we pruned SNPs that were in linkage disequilibrium (or LD) to generate a pruned subset of SNPs that were in approximate linkage equilibrium with one another. For the pruning process, we followed a series of PLINK command steps. First, we considered a window size of 50 SNPs; then, we calculated the LD between each pair of SNPs in the window. Subsequently, we removed one SNP from a pair of SNPs if the LD was greater than the threshold $r^2 = 0.2$. Finally, we shifted the window 5 SNPs forward and repeated the procedure. This highly restrictive threshold removed SNP markers in LD, and also SNPs in splicing sites, SNPs in 5′ and 3′ untranslated regions of mRNAs (or UTR), and singleton SNPs (Ke et al., 2008).

We applied the QC procedure using the following PLINK commands:

plink1.9 --bfile --geno 0.01 --mind 0.01 --maf 0.05 --hwe 0.01 --indep-pairwise 50 5 0.2 --out

To gain a better understanding on admixture patterns from hundreds of years ago, we analysed genome-wide SNP data variations in African-American populations by comparing them with reference populations included in different reference population datasets. We elaborated three genome-wide SNP datasets by using different reference populations.

First, to detect presence of population structure and admixed individuals in African-American descendants from South America, we compared the new genome-wide SNP data of African-

American and West African populations with populations included in the Human Genome Diversity Cell Line (or HGDP) panel (also known as the HGDP-CEPH panel) (Cann et al., 2002; Cavalli-Sforza et al., 1991).

The HGDP panel is the general reference dataset used in previous studies related to ancestry in human populations (Montinaro et al., 2015; Pugach & Stoneking, 2015). The HGDP panel includes a dataset of 650,000 SNPs determined with the Illumina Bead technology; this dataset has been genotyped for different populations from Africa, Europe, the Middle East, South and Central Asia, East Asia, Oceania, and the Americas. Therefore, this panel reveals a well-known geographical structure for human populations at the continental level, and allows a detailed characterization of worldwide genetic variations (Jakobsson et al., 2008; Li et al., 2008).

We analysed 107 African Americans from South America (Table 9) and 149 individuals from 17 reference populations for African (N= 60), European (N= 52), and Native American (N= 37) continental ancestry, which are included in the HGDP panel (see Table 12 and Figure 44); and 122 new West African samples (Table 9) were also analysed. We merged the African-American populations with reference populations by using PLINK v.1.9 (Chang et al., 2015; Purcell et al., 2007). After the QC procedure and genotyping pruning, we ended up with about 90,600 SNPs to carry out subsequently analysis.

Second, we estimated continental ancestry by using a comprehensive up-to-date reference population for genome-wide analysis. Currently, whole genome sequences are becoming available to evaluate worldwide human genetic diversity. For instance, the 1000 Genomes Project Phase 3 version 5a (1000 Genomes Project Consortium et al., 2012; 1000 Genomes Project Consortium, 2015) sequenced whole genomes of large population samples originating from Africa, America, Europe, and Asia. This international project provides free genome-wide SNP data of populations with different continental ancestry and genetic backgrounds. Hence, to increase resolution and the number of SNP that are highly informative of ancestry, we used as a reference for continental ancestry only those populations that were included in the last released of 1000 Genomes Project Phase 3 (1000 Genomes Project Consortium, 2015).
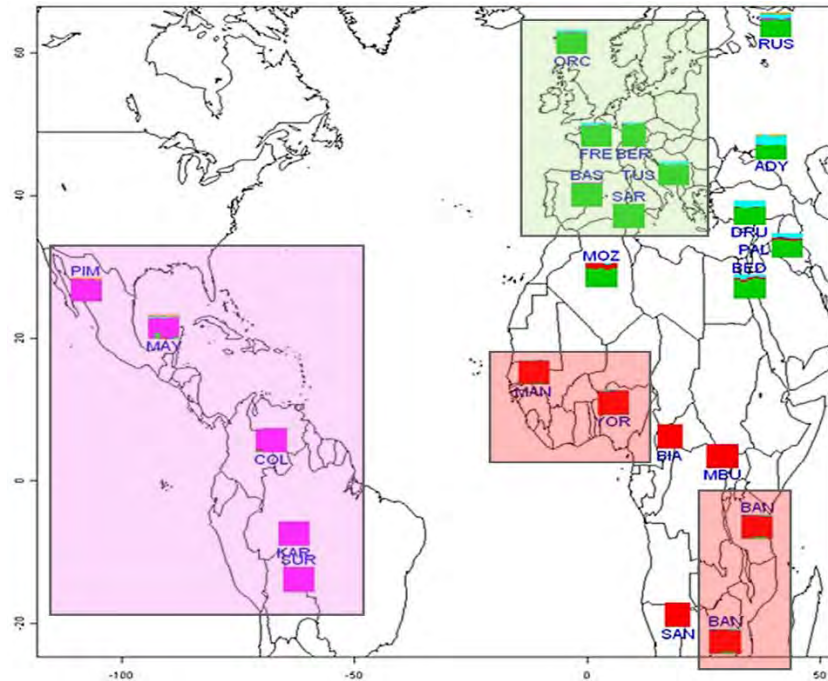
Figure 44: Geographic location of populations from the HGDP Panel used as source population of African (MAN, YOR, and BAN), European (FRE, BAQ, ORC, TUS, and SAR), and Native American (PIM, MAY, COL, KAR, and SUR) ancestries. Pygmy populations (BIA and MBU) from Central Africa and San population (SAN) from South Africa were included in the TreeMix analysis. Populations are labelled according to Table 12. Image modified from Lopez Herraez et al. (2009).

Table 12: List of populations from the HGDP panel and geographic origin used in the ADMIXTURE and PCA analysis. The HGDP panel is freely available online at http://www.cephb.fr/en/hgdp_panel.php.

Geographic Origin	Population samples	Id Pop	Coordinates	N
Sub-Saharan African				
Senegal	Mandenka	MAN	12N, 12W	21
Nigeria	Yoruba	YOR	6-10N, 2-8E	21
Kenya (Bantu speakers)	Bantu N.E.	BAN	3S, 37E	10
S. Africa (Bantu speakers)	Pedi (l), Sotho (1), Tswana (2), Zulu (1), Herero (2), Ovambo (1)	BAN	29S, 30E	8
West European				
France	French (various regions)	FRE	46N, 2E	28
France	Basque	BAQ	43N, 0E	24
Orkney Islands	Orcadian	ORC	59N, 3W	16
Italy	Tuscan	TUS	43N, 11E	8
Italy	Sardinian	SAR	40N, 9E	28
Native American				
Mexico	Pima	PIM	29N, 108W	8
Mexico	Maya	MAY	19N, 91W	21
Colombia	Piapoco and Curripaco	COL	3N, 68W	5
Brazil	Karitiana	KAR	10S, 63W	2
Brazil	Surui	SUR	11S, 62W	1

We downloaded whole genome sequences from this large dataset (last released available online in February 2015, at ftp://ftp-trace.ncbi.nih.gov/1000genomes/ftp/release/20130502/). Genomes were available in separate files for each chromosome, in variant call format (or VCF format) (Danecek et al., 2011). We selected only populations with African (including African American), Native American, European, and East Asian ancestries (Figure 45). After filtering, the VCF files for each chromosome were converted to PED/MAP files and merged using PLINK v.1.9 (Chang et al., 2015; Purcell et al., 2007). For each whole genome sequence, we extracted the same set of ~4.3 million SNPs present in the *Illumina HumanOmni5 Quad BeadChips^TM* by using PLINK.

We merged genome-wide SNP data for African-American and West African populations (Table 9) with the reference populations selected in the 1000 Genomes Project (see Table 13). Additionally, we included one population of Bantu-speaking people residing in Soweto, South Africa, which has been recently genotyped with the same whole-genome array (May et al., 2013). After merging, we had approximately four million SNPs for 2,042 individuals in 31 worldwide populations.

We performed the restrictive QC procedure described above by using PLINK v.1.9 (see all steps in Figure 47). At the end of the QC procedure, we had approximately 240,000 SNPs for 2,038 individuals (see Table 13). This dataset was called 'high-density' SNP dataset (in short 240K dataset). The 240K dataset was employed to analyse genetic ancestry and gene flow in African Americans from various parts of the Americas—North America (ASW in USA), the Caribbean (ACB in Barbados), and South America (in French Guiana, Colombia, and Brazil) by using different admixture-model approaches.

Finally, to estimate fine-scale population structure within African populations, we built a dataset of African Americans and the biggest representation of African populations from different geographical locations across the continent, as well as with different historical, linguistic, and cultural backgrounds (see Figure 46). We analysed representative ethnolinguistic groups in North Africa (Henn et al., 2012), West Africa (Present study; (Bryc et al., 2010a)), West-Central Africa (Bryc et al., 2010a; Patin et al., 2014), East Africa (Pagani et al., 2012), and South Africa (May et al., 2013; Schlebusch et al., 2012). Additionally, we included African populations in the 1000 Genomes Project (1000 Genomes Project Consortium, 2015), and also populations with the highest European (IBS, TSI, GBR, and CEU), Native American (MXL and PEL), and East Asian (CHB and CHS) ancestries. Furthermore, we included one Qatari population from the Middle East (Hunter-Zinck et al., 2010).
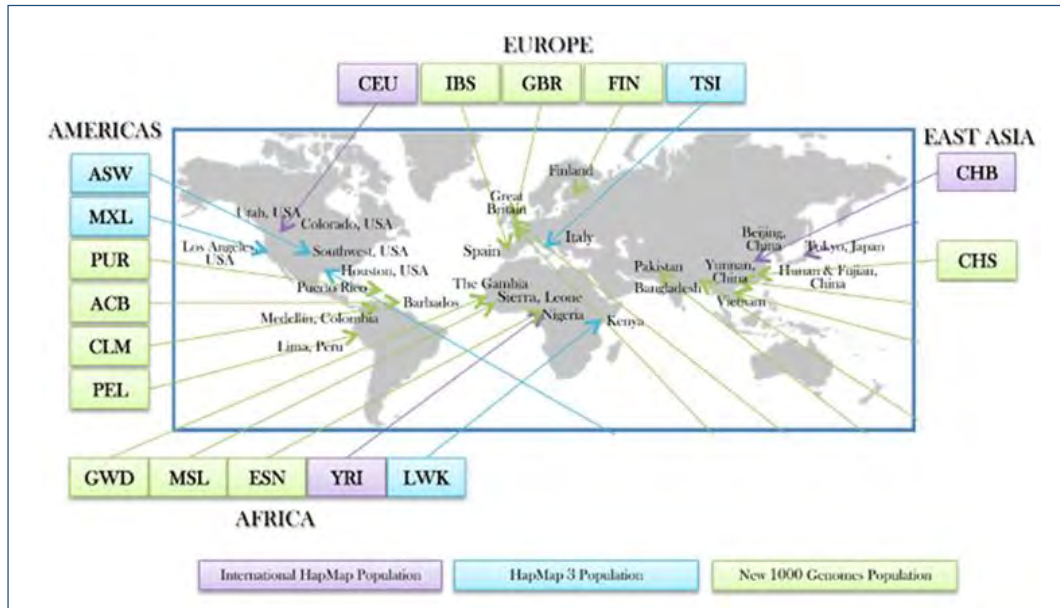
Figure 45: Geographic location of populations included in the 1000 Genomes Project Consortium Phased 3 (1000 Genomes Project Consortium, 2015) that were analysed in the present study. This new panel also included reference populations from International HapMap Project (International HapMap Consortium, 2005) and HapMap 3 (International HapMap 3 Consortium et al., 2010). Populations are labelled according to Table 13.
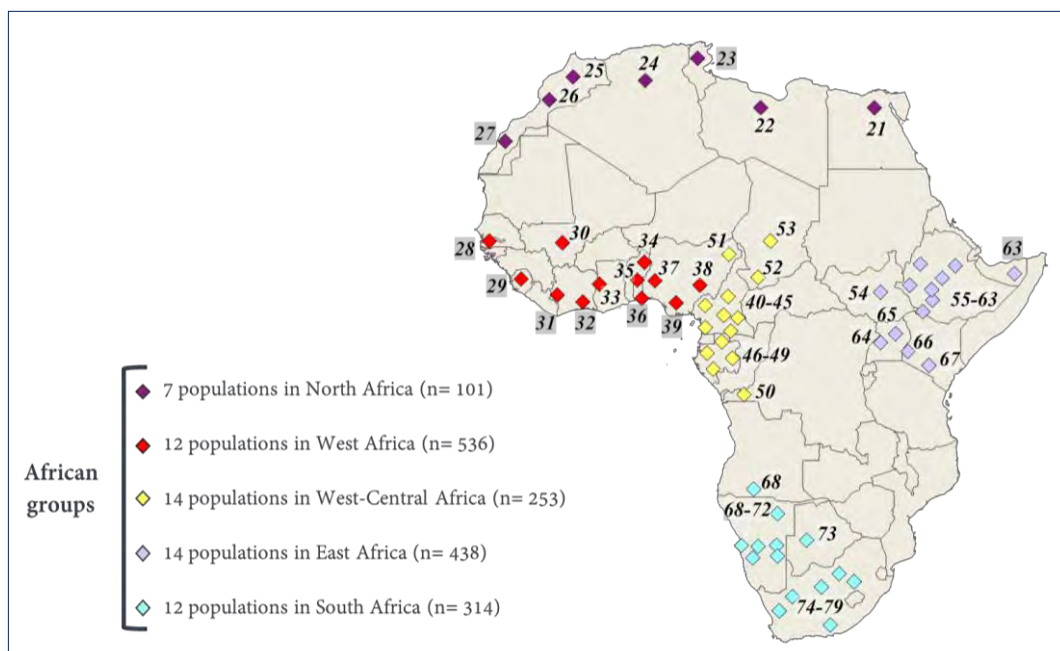
Figure 46: Geographic location of 59 African populations included in ADMIXTURE analysis using thee low-density SNP dataset. Populations are labelled according to Table 13.
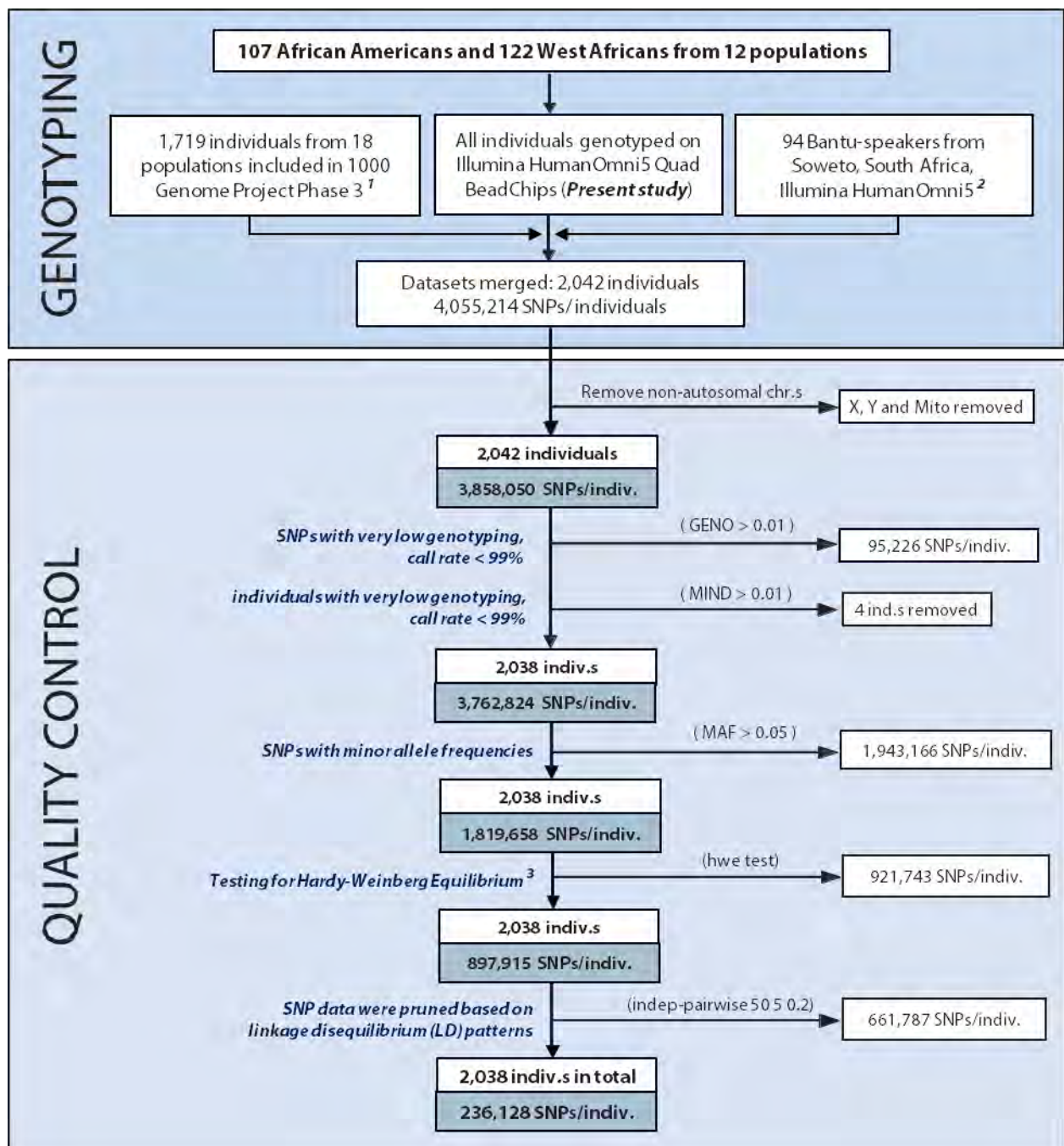
Figure 47: Flowchart of data genotyping and quality control procedure performed. It shows inclusions and exclusions of SNP/individual followed to obtain a high-density SNP dataset. References: [1] (1000 Genomes Project Consortium, 2015), [2] (May et al., 2013), and [3] (Wigginton et al., 2005).

After merging all these populations in one unique dataset, we carried out the same restrictive QC procedure described above. We ended up with around 50,000 SNPs dataset for 2,812 individuals in 76 populations (see Table 13). This dataset was called 'low-density' SNP dataset (in short 50K dataset). We used the whole 50K datasets for all African regions to compare results obtained with the 240K dataset. Furthermore, we selected only African populations from historical African coastal regions involved in the transatlantic slave trade to analyse the African origin of the African-American populations from South America.

Table 13: Populations analysed based on genome-wide SNP data using a high-density SNP (~240K SNPs) and a low-density SNP (~50K SNPs) dataset. The geographical locations are indicated in Figure 45 and Figure 46.

Code	Continental ancestry	Country	Population	ID Pop	Sample Size	High-density (240K)	Low-density (50K)	Number of SNPs	Platform	Ref *
AA	North America	USA, Southwest	African American (ASW)	1	66	61	61	Whole genome	1000 GP Phase 3	1
	The Caribbean	Barbados	African American (ACB)	2	96	96	96	Whole genome	1000 GP Phase 3	1
	South America	French Guiana	Noir Marron, Aluku	3	23	23	23	4,300,000	Illumina HumanOmni5	2
	South America	French Guiana	Noir Marron, Ndjuka	4	23	22	23	4,300,000	Illumina HumanOmni5	2
	South America	French Guiana	Noir Marron, Paramaka	5	6	6	6	4,300,000	Illumina HumanOmni5	2
	South America	French Guiana	Noir Marron, Saramaka	6	19	18	18	4,300,000	Illumina HumanOmni5	2
	South America	Colombia	Afro-Colombian	7	20	20	20	4,300,000	Illumina HumanOmni5	2
	South America	Brazil	Afro-Brazilian	8	16	16	16	4,300,000	Illumina HumanOmni5	2
	Total African American populations (AFA)				**269**	**262**	**263**			
NAM	Central America	USA, California	Mexican ancestry (MXL)	9	67	64	64	Whole genome	1000 GP Phase 3	1
	Central America	Puerto Rico	Puerto Rican (PUR)	10	105	104	-	Whole genome	1000 GP Phase 3	1
	South America	Colombia	Colombian in Medellin (CLM)	11	95	94	-	Whole genome	1000 GP Phase 3	1
	South America	Peru	Peruvian in Lima (PEL)	12	86	85	85	Whole genome	1000 GP Phase 3	1
	Total Native American populations (NAM)				**353**	**347**	**149**			
EUR	South Europe	Spain	Iberian (IBS)	13	107	107	107	Whole genome	1000 GP Phase 3	1
	South Europe	Italy	Toscani in Italy (TSI)	14	108	107	107	Whole genome	1000 GP Phase 3	1
	North Europe	England and Scotland	British (GBR)	15	92	91	91	Whole genome	1000 GP Phase 3	1
	North Europe	Finland	Finnish (FIN)	16	100	99	-	Whole genome	1000 GP Phase 3	1
	North-West Europe	USA	Utah residents (CEU)	17	103	99	99	Whole genome	1000 GP Phase 3	1
	Total European populations (EUR)				**510**	**503**	**404**			
MDE	Middle East	Qatar	Qatari	18	168	-	140	500,000	Affymetrix 500K	3
EAS	Central Asia	China	Han Chinese in Bejing (CHB)	19	106	103	103	Whole genome	1000 GP Phase 3	1
	Central Asia	China	Southern Han Chinese (CHS)	20	112	105	105	Whole genome	1000 GP Phase 3	1
	Total Asian populations (MDE and EAS)				**386**	**208**	**348**			
NAF	North Africa	Egypt	Egyptian	21	19	-	19	730,000	Affymetrix 6.0	4
	North Africa	Libya	Libyan	22	17	-	16	730,000	Affymetrix 6.0	4
	North Africa	Tunisia	Tunisian Berber	23	18	-	16	730,000	Affymetrix 6.0	4
	North Africa	Algeria	Algerian	24	19	-	15	730,000	Affymetrix 6.0	4
	North Africa	Morocco	North Moroccan	25	18	-	16	730,000	Affymetrix 6.0	4
	North Africa	Morocco	South Moroccan	26	16	-	14	730,000	Affymetrix 6.0	4
	North Africa	West Sahara	Saharawi	27	18	-	14	730,000	Affymetrix 6.0	4
	Total North African populations (NAF)				**125**	**0**	**110**			
WAF	West Africa	Gambia	Gambian in W. Division (GWD)	28	113	113	113	Whole genome	1000 GP Phase 3	1
	West Africa	Sierra Leone	Mende (MSL)	29	85	85	85	Whole genome	1000 GP Phase 3	1
	West Africa	Mali	Bwa	30	18	17	17	4,300,000	Illumina HumanOmni5	2
	West Africa	Ivory Coast	Yacouba	31	17	16	16	4,300,000	Illumina HumanOmni5	2
	West Africa	Ivory Coast	Ahizi	32	20	20	20	4,300,000	Illumina HumanOmni5	2
	West Africa	Ghana	Brong	33	8	-	4	500,000	Affymetrix 500K	5
	West Africa	Benin	Bariba	34	24	24	24	4,300,000	Illumina HumanOmni5	2
	West Africa	Benin	Yoruba_B	35	24	24	24	4,300,000	Illumina HumanOmni5	2
	West Africa	Benin	Fon	36	19	19	19	4,300,000	Illumina HumanOmni5	2
	West Africa	Nigeria	Yoruba_N (YRI)	37	116	108	108	Whole genome	1000 GP Phase 3	1
	West Africa	Nigeria	Esan (ESN)	38	99	99	99	Whole genome	1000 GP Phase 3	1
	West Africa	Nigeria	Igbo	39	17	-	7	500,000	Affymetrix 500K	5
	Total West African populations (WAF)				**560**	**525**	**536**			
WCA	West-Central Africa	Cameroon	Hausa	40	13	-	7	500,000	Affymetrix 500K	5
	West-Central Africa	Cameroon	Mada	41	12	-	9	500,000	Affymetrix 500K	5
	West-Central Africa	Cameroon	Fang	42	18	-	5	500,000	Affymetrix 500K	5
	West-Central Africa	Cameroon	Bamoun	43	20	-	11	500,000	Affymetrix 500K	5
	West-Central Africa	Cameroon	Nzime	44	53	-	53	930,000	Illumina HumanOmni1	6
	West-Central Africa	Cameroon	Baka_Hunter-gatherer	45	58	-	57	930,000	Illumina HumanOmni1	6
	West-Central Africa	Gabon	Baka_Hunter-gatherer	46	16	-	16	930,000	Illumina HumanOmni1	6
	West-Central Africa	Gabon	Bongo East_Hunter-gatherer	47	22	-	22	930,000	Illumina HumanOmni1	6
	West-Central Africa	Gabon	Bongo South_Hunter-gatherer	48	24	-	24	930,000	Illumina HumanOmni1	6
	West-Central Africa	Gabon	Nzebi	49	20	-	20	930,000	Illumina HumanOmni1	6
	West-Central Africa	D.R.C.	Kongo	50	9	-	4	500,000	Affymetrix 500K	5
	Total West-Central African populations (WCA)				**265**	**0**	**228**			

Code	Continental ancestry	Country	Population	ID Pop	Sample Size	High-density (240K)	Low-density (50K)	Number of SNPs	Platform	Ref *
	Central Africa	Nigeria	Fulani_Mbororo	51	13	-	6	500,000	Affymetrix 500K	5
CAF	Central Africa	Chad	Kaba	52	16	-	8	500,000	Affymetrix 500K	5
	Central Africa	Chad	Bulala	53	15	-	11	500,000	Affymetrix 500K	5
	Total Central African populations (CAF)				**44**	**0**	**25**			
	East Africa	Sudan	Sudanese_South	54	24	-	17	930,000	Illumina HumanOmni1	7
	East Africa	Ethiopia	Afar	55	12	-	7	930,000	Illumina HumanOmni1	7
	East Africa	Ethiopia	Amhara	56	26	-	19	930,000	Illumina HumanOmni1	7
	East Africa	Ethiopia	Anuak	57	23	-	20	930,000	Illumina HumanOmni1	7
	East Africa	Ethiopia	Ari_BlackSmith	58	17	-	14	930,000	Illumina HumanOmni1	7
	East Africa	Ethiopia	Ari_Cultivator	59	24	-	19	930,000	Illumina HumanOmni1	7
EAF	East Africa	Ethiopia	Gumuz	60	19	-	7	930,000	Illumina HumanOmni1	7
	East Africa	Ethiopia	Oromo	61	21	-	17	930,000	Illumina HumanOmni1	7
	East Africa	Ethiopia	Tygray	62	21	-	17	930,000	Illumina HumanOmni1	7
	East Africa	Somalia	Somali	63	23	-	13	930,000	Illumina HumanOmni1	7
	East Africa	Uganda	Batwa_Hunter-gatherer	64	32	-	32	930,000	Illumina HumanOmni1	6
	East Africa	Uganda	Bakiga	65	35	-	35	930,000	Illumina HumanOmni1	6
	East Africa	Kenya	Luhya in Webuye (LWK)	66	101	99	99	Whole genome	1000 GP Phase 3	1
	East Africa	Kenya	Maasai in Kinyawa (MKK)	67	133	-	133	Whole genome	1000 GP Phase 3	1
	Total East African populations (EAF)				**511**	**99**	**449**			
	South Africa	Namibia and S. Africa	!Xun_San	68	19	-	19	2,400,000	Illumina Omni 2.5M	8
	South Africa	Namibia	Herero_Bantu	69	12	-	12	2,400,000	Illumina Omni 2.5M	8
	South Africa	Namibia	Ju/'hoansi_San	70	18	-	18	2,400,000	Illumina Omni 2.5M	8
	South Africa	Namibia	Nama_Khoe	71	20	-	20	2,400,000	Illumina Omni 2.5M	8
	South Africa	Namibia and S. Africa	Khwe_San	72	17	-	17	2,400,000	Illumina Omni 2.5M	8
SAF	South Africa	Botswana	/Gui and //Gana_San	73	15	-	15	2,400,000	Illumina Omni 2.5M	8
	South Africa	South Africa	Karretjie_San	74	20	-	20	2,400,000	Illumina Omni 2.5M	8
	South Africa	South Africa	Khomani_San	75	39	-	39	2,400,000	Illumina Omni 2.5M	8
	South Africa	South Africa	Coloured_Colesberg	76	20	-	20	2,400,000	Illumina Omni 2.5M	8
	South Africa	South Africa	Coloured_Wellington	77	20	-	20	2,400,000	Illumina Omni 2.5M	8
	South Africa	South Africa	Bantu speakers groups	78	20	-	20	2,400,000	Illumina Omni 2.5M	8
	South Africa	South Africa	Bantu speakers in Soweto	79	94	94	94	4,300,000	Illumina HumanOmni5	9
	Total South African populations (SAF)				**295**	**94**	**314**			
	TOTAL			79	**3,318**	**2,038**	**2,826**			

* References: [1] (1000 Genomes Project Consortium, 2015), [2] *Present study*, [3] (Hunter-Zinck et al., 2010), [4] (Henn et al., 2012), [5] (Bryc et al., 2010a), [6] (Patin et al., 2014), [7] (Pagani et al., 2012), [8] (Schlebusch et al., 2012), and [9] (May et al., 2013).

Note: Population gathered according to historical African coastal regions; 28 for Senegambia; 29 for Sierra Leone; 30 and 31 for Windward Coast; 33 for Gold Coast; 34-38 for Bight of Benin; 39-44 and 49 Bight of Biafra; 50, 69, and 79 for West Central Africa; and 66 for Southeast Africa.

D.2.b. *Inbreeding coefficient and runs of homozygosity*

The inbreeding coefficient is the probability of homozygosity by descent having common ancestors, that is to say the probability that a zygote obtains copies of the same ancestral gene from both its parents (Frankham et al., 2010). For instance, if the parents of an individual are related, it is possible for the individual to receive at one locus two identical-by-descent (or IBD) alleles that are copies of a single allele carried by a common ancestor of the parents. The inbreeding coefficient measures the probability of this event and increases with increasing relatedness between the parents (Li & Horvitz, 1953; Wright, 1933).

Similarly, if individuals inherit the same ancestral mutations identically by descent, then they probably also share adjacent DNA segments on which the mutation first arose. For a recessive phenotype in affected inbred individuals, the homozygous risk locus probably resides in an unusually long homozygous region. Deleterious recessive variants can therefore be identified in affected inbred individuals by detecting long homozygous regions or runs of homozygosity (also known as ROH) (Broman & Weber, 1999; Clark, 1999).

With the availability of genome-wide SNP data, it is possible to compute a genome-based inbreeding coefficient and to identify inbred individuals in a population from the observed patterns of homozygosity (Gazal et al., 2014). A commonly used measure to quantify IBD in an individual is the inbreeding coefficient (or F), a quantity that reflects not only consanguinity, but also other aspects of kinship in the population to which the individual belongs (Pemberton & Rosenberg, 2014).

We examined the relationship between the genomic estimates of the inbreeding coefficient and the population patterns of genetic variations in eight African-American populations in the Americas. We analysed the dataset obtained after merging African-American populations from North America (in USA), the Caribbean (in Barbados), and South America (in French Guiana, Colombia, and Brazil). We analysed 262 African-American people in total.

We used the PLINK software v.1.9 (Chang et al., 2015; Purcell et al., 2007) to estimate both F and ROH of African-American populations. F was based on the observed versus expected number of homozygous genotypes.

We applied the following PLINK command to estimate F:

plink1.9 --bfile --het --out

Similarly, we applied the following PLINK command to estimate ROH:

plink1.9 --bfile --homozyg-window-snp 50 --homozyg-snp 25 --homozyg-kb 500 --homozyg-gap 100 --homozyg-window-missing 5 --homozyg-window-threshold 0.05 --homozyg-window-het 1 --homozyg-density 50 --out

D.2.c.   *Estimation of admixture events*

To test admixture models and demographic scenarios in African Americans from the Americas, we built three maximum likelihood trees based on the approach proposed by Pickrell & Pritchard (2012). This method can infer patterns of population splits and admixtures events of a set of populations. Hence, these trees largely recapitulate the relationships among population groups and identify aspects of ancestry and migratory events. Each tree was

elaborated for different SNP datasets (50K, 90K, and 240K) using the TreeMix software v.1.12 (Pickrell & Pritchard, 2012). For increasing the confidence in each tree topology, we also applied the bootstrap method by generating a single bootstrap replicated by resampling blocks of 500 SNPs.

First, we elaborated one maximum likelihood tree based on allele frequencies of 90K SNP dataset by using African-American populations from South America, West African populations, and other populations with different continental ancestries included in the HGDP panel. We selected the San population from South Africa as outgroup. These indigenous hunter-gatherer people, together with Khoe people, are the oldest known lineage of modern humans according to mtDNA (Chan et al., 2015; Morris et al., 2014), Y chromosome (Marks et al., 2015), and genome-wide studies (Kim et al., 2014; Pickrell et al., 2012; Schuster et al., 2010). We also included in this analysis Pygmy populations (Biaka and Mbuti) from Central Africa.

Second, we elaborated another maximum likelihood tree based on the high-density SNP dataset by using African-American populations from South America, North America, and the Caribbean, and other populations with different continental ancestries included in the 1000 Genomes Project (1000 Genomes Project Consortium, 2015). The 240K dataset enabled us to analyse the increased number of West African populations (with GWD, MSL, ESN, and YRI) with a higher resolution of SNPs. In this tree, we used Bantu speakers from South Africa (May et al., 2013) as outgroup. The Bantu-speakers population is the most geographically distant from West Africa.

Finally, we elaborated another maximum likelihood tree based on low-density SNP dataset by using African-American populations and African populations across the entire continent; this was done to determine the genetic links of African Americans within African populations. In this tree, we used the Batwa Pygmy population from east-central Africa as outgroup because this rainforest hunter-gatherer population presented significantly higher |iHS| and F_{ST} values than any other rainforest hunter-gatherer populations in West or East Africa (Perry et al., 2014).

Furthermore, we calculated allele frequency correlations among populations by using the $f3$-statistics, which allowed us to make inferences about the demographic history of populations (Moorjani et al., 2011; Reich et al., 2009). We carried out the three-population test by using the *threepop* application of the TreeMix software v.1.12 (Pickrell & Pritchard, 2012). The three-population test is a formal test of admixture, which provides clear evidence of admixture, even if the gene flow events occurred hundreds of generations ago (Durand et al., 2011). This test is of the form $f3$ (X;A,B), where a negative value of the $f3$-statistic implies that the target population (population X) was the result of an admixture event between the two ancestral source populations (A and B populations) (Patterson et al., 2012). We used African-American

populations as a target population and two groups of source populations: one for African and European populations and another one for African and Native American populations.

D.2.d. ADMIXTURE analysis

To address admixture events that have occurred many generations ago, several clustering methods have been proposed (Libiger & Schork, 2012; Padhukasahasram, 2014). Clustering methods, such as ADMIXTURE (Alexander et al., 2009), STRUCTURE (Pritchard et al., 2000), and FRAPPE (Tang et al., 2005), are model-based approaches for global ancestry inference in unrelated individuals. These clustering methods can take multi-locus genotypes of individuals from several populations, and apportion them into resolved clusters that are differentiated from one another. A number of model-based clustering methods have been devised that determine the cluster number, the frequency of any given allele in each cluster, and the proportion to which each individual's genome owes ancestry to each cluster. With this approach each individual genome is considered to be a palimpsest of multi-layered history of successive periods of drift and admixture events in multiple ancestral populations (Jobling et al., 2014).

We estimated individual African, European and Native American ancestry proportions in African-American populations by using the ADMIXTURE v.1.23 software (Alexander et al., 2009). We carried out the unsupervised method and performed runs at K-groups values of 2, 3, 4, and 5, where K is the number of unknown ancestral groups. This model-based method takes into account basic demographic assumptions, such as the presence of the Hardy-Weinberg equilibrium in the allelic frequencies of the K 'ancestral' population(s) (Alexander et al., 2009). The main advantage of ADMIXTURE is its model-based approach. The Expectation-Maximization (or EM) algorithm (Zhou et al., 2011) is incorporated in ADMIXTURE, which runs faster and with greater accuracy than STRUCTURE (Pritchard et al., 2000) and EIGENSTRAT (Patterson et al., 2006; Price et al., 2006) programs.

We analysed the African-American and West African populations (Table 9) with reference populations included in the HGDP panel (90K dataset). We also analysed the African-American and West African populations with reference populations included in the 1000 Genomes Project with both high- and low-density SNP datasets (240k and 50K datasets).

The ADMIXTURE model assumes the Hardy-Weinberg equilibrium and linkage equilibrium among loci. Therefore, this model-based approach does not explicitly consider the LD between markers. It is especially relevant in recently admixed populations, such as African Americans, which have a high degree of 'admixture LD' (Alexander et al., 2009). The easiest way to avoid

the effects of LD between relatively close SNPs is by pruning SNPs in LD by using PLINK v.1.9 (Chang et al., 2015; Purcell et al., 2007). Hence, in the QC procedure of each dataset, we removed each SNP that had a correlation threshold value (or r^2) greater than 0.2 with any other SNP within a 50-SNP sliding window, and advanced by five SNPs each time.

The other fundamental question in clustering methods is to decide upon the most likely number of clusters of ancestral populations. We carried out the cross-validation (or CV) test (Alexander & Lange, 2011) to answer that question. We estimated the CV error running the ADMIXTURE software with the --cv flag for K values between 2 and 10 for 50K and 90K datasets, and for K values between 2 and 20 for 240K dataset.

For the 50K dataset, we ran ten replicates using a random seed and kept the Q (ancestral cluster proportions) and P (inferred ancestral cluster allele frequencies) matrices from the run with the best log likelihood. We used the Q matrix from each K to estimate the most likely cluster proportions because of the high number of probable ancestral populations in the 50K dataset. We used the CLUMPAK (or Cluster Markov Packager Across K) software (Kopelman et al., 2015) to selected the preferred value of K according to the methods of Pritchard et al. (2000).

We elaborated ADMIXTURE plots for each dataset based on membership frequencies of each individual by using the R software (R Core Team, 2014). Additionally, we elaborated an ADMIXTURE plot by averaging admixture's membership frequencies across individuals for each population. Likewise, we calculated the average membership frequencies across African populations that are representative of historical African coastal regions.

We elaborated maps of admixture coefficients of the African-American and African populations across the continent by using the 50K SNP dataset. We analysed patterns of distributions of Q admixture coefficients for West-Central Africa and Bantu ancestry. We plotted these maps by using R functions based on the Kriging method proposed by Francois (2013). For mapping, we used Surfer v.12 (Golden Software, Inc.). This approach has been recently applied for African populations by Hodgson et al. (2014).

D.2.e. *Principal component analysis*

The PCA is a statistical method commonly used in population genetics to identify structure in the distribution of genetic variations across geographical locations and ethnic backgrounds (McVean, 2009). The PCA seeks to construct projections in the lower dimensional space that captures a large fraction of the variation in the marker genotypes. We carried out a Eigenstrat PCA analysis to study genetic variations of a worldwide dataset of 1,172 individuals from African-American and West African populations (Table 9); the reference populations were from

the HGDP panel with all geographical regions represented. We included 940 individuals from 53 different populations from Africa, Europe, the Middle East, South and Central Asia, East Asia, Oceania and the Americas (see Table S10).

To compare the results obtained in the ADMIXTURE analysis, we performed another Eigenstrat PCA based on African-American, African, European and Native American populations from the HGDP panel used in the 90K SNP dataset. We also performed an Eigenstrat PCA based on African-American populations from French Guiana and African populations from historical coastal regions.

The PCA analyses were carried out using SMARTPCA in the EIGENSOFT package v.6.0.1 (Patterson et al., 2006) and by plotting using the R packages *ggplot2* (Wickham, 2009). The EIGENSOFT software implements PCA analysis for the purposes of detecting the population stratification in genome-wide association studies (or GWAS) (Price et al., 2006), and the population structure in genetic studies (Patterson et al., 2006). This analysis is based on eigenanalysis method proposed by Price et al. (2006) and Patterson et al. (2006) to project genetic data from individuals into a low-dimensional space formed from the eigenvectors of the genetic sample. Because of they do not assume a specific model for population evolution, they can be used in a variety of evolutionary scenarios, particularly two-population admixture scenarios (Shringarpure & Xing, 2014).

III. RESULTS

We characterised three genetic systems in the Noir Marron communities and West African populations by using a high genetic resolution which were compared with other African-American descendants in South America, the Caribbean and North America. We also compared the new results obtained with other African populations as well as populations with European and Native American continental ancestry, in order to infer the African origin and to shed new light on admixture patterns in African-American descendants.

A. Y chromosome results

A.1. Y-SNP diversity detected

Through the Y-SNP genotyping, we obtained a large representation of Y lineages with 36 haplogroups from clades A to R (Figure 48); most haplogroups were for clade E that is associated with sub-Saharan African populations (de Filippo et al., 2011). We built a complete Y chromosome tree with all terminal clades and their internal nodes obtained (see Figure 49). We estimated relative frequencies for each Y haplogroup in each population, and we also compared African American frequencies with frequencies obtained for reference populations with West African, West European, and Native American ancestry.

African Y haplogroups from A to E1b1a were not detected in European or Native American populations (with the exception of one E1b1a1-M2 in Oyampi tribe). In sharp contrast to African-American populations, whose African haplogroups represented high values in Noir Marron (90.0%), and middle values in Afro-Colombians (50.0%) and Afro-Brazilians (53.6%). This set of African Y haplogroups represents between 94.7% and 100% in West African populations (see Figure 48). Hence, to belong to these haplogroups is a valuable indicator of West African ancestry, which is highly presented in the Noir Marron communities and is in good agreement with Brucato et al. (2010). The remaining African Y haplogroups belonged to the clades E1b1b1 and E2, which have widespread distribution in sub-Saharan Africa, North Africa, and even the Near East (de Filippo et al., 2011).

Y chromosome diversity in West African populations indicated a large representation of African lineages in Benin and Ivory Coast, and less diversity in Mali because of its lower sample size (Figure 48). The degree of variation is substantially higher in the Bariba population (HD = 0.833 +/-0.023; Figure 49) with seven African Y haplogroups observed (Fortes-Lima et al., 2015).
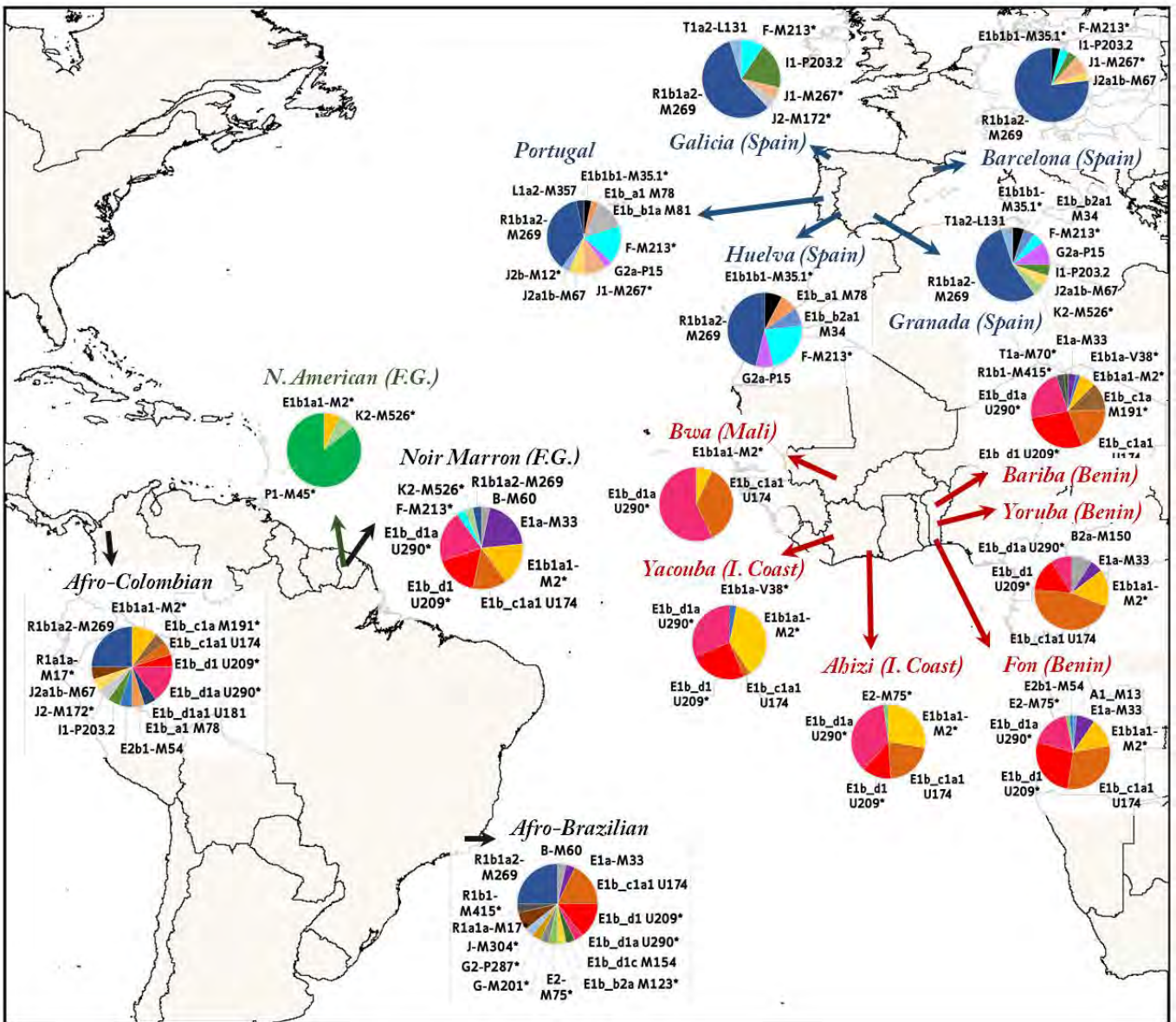
Figure 48: Map of frequency distributions of Y-SNP haplogroups for African American, Native American, European and West African populations analysed. Y haplogroup nomenclature is according to ISOGG Y-DNA Haplogroup Tree 2015 (Y Chromosome Consortium, 2002). Results from Benin and Ivory Coast ethnic groups were reported by Fortes-Lima et al. (2015). Note: 'E1b_' means 'E1b1a1a1'.

The most frequent haplogroups belonged to the clade E1b1a1-M2* (see Figure 48), which is mainly associated with West African populations (de Filippo et al., 2011). The highest frequency (45.45%) was detected for the haplogroup E1b1a1a1c1a1-U174* in the Beninese Yoruba population (Fortes-Lima et al., 2015), in sharp contrast to the frequency reported for the same haplogroup (17.02%) in the Nigerian Yoruba population (de Filippo et al., 2011). The haplogroup E1b1a1a1d1a-U290* was detected at high frequencies in Bwa (57.14%), Ahizi (36.17%), and Yacouba (31.25%). This was in good agreement with values the previously reported in other West Africa populations, such as Asante (24.47%) and Ewe (27.27%) from

Ghana and Ekoi from Nigeria (27.66%), and the values reported in West-Central Africa populations, such as Bembe from Congo (28.44%) and Aghem from Cameroon (31.03%) (Ansari Pour et al., 2013).

The haplogroup E2-M75* , common in Western and Eastern Bantu speakers (de Filippo et al., 2011), was detected in Ahizi and Fon populations at low frequencies (2.13% and 1.59%, respectively). It might suggest some recent gene flow inherited from individuals of Bantu-speaking groups. Similarly, the presence of B2a-M150 in the Yoruba population (9.09%) adds further support to a potential genetic input from Bantu-speakers in West African populations (Batini et al., 2011a).

The presence of two Eurasian haplogroups in the Bariba ethnic group, such as R1b-M415* (3.51%) and T1a-M70*(xL131) (1.75%), suggested contacts with pastoralist populations from the Sahel region with populations in North Benin (Cerny et al., 2011). These haplogroups might also indicate migration routes from the Near East to sub-Saharan Africa (Mendez et al., 2011), or a back-migration from Asia to Africa (Cruciani et al., 2010).

In addition, West European populations from Andalusia and Portugal evidence gene flow with African populations because of the presence of the haplogroup E1b1b1-M35.1 (10.0% Granada, 23.1% Huelva, and 20.0% Portugal), most likely from North African populations. This haplogroup and its subhaplogroups arrived in the Iberian Peninsula through the Strait of Gibraltar associated with the Islamic expansion during the VIII century. This was previously stated by Ambrosio et al. (2010) and (2012) in both, Huelva and Granada. In Portugal, Beleza et al. (2006) and Santos et al. (2014) reached the same conclusions. Interestingly, its subhaplogroup E1b1b1a1-M78 was found in Afro-Brazilian (5.0%) and West European populations (Figure 49), and this might suggest that Iberian populations brought this North African subhaplogroup to South America during the slave trade.

In Native American tribes from French Guiana (Kalinya, Oyampi, and Palikour), the haplogroup P1-M45* was detected in high frequency (85.7%); it most likely belonged to the haplogroup Q, which is virtually the only branch of the Y-phylogeny observed in modern-day Amerindians of Central and South America (Battaglia et al., 2013). However, none of the Y-SNPs included in the 96-plex genotyping (Table S1) can be associated with the haplogroup Q. The haplogroup P1-M45* was not detected in the African-American populations, not even in the geographically close Noir Marron communities, and this might indicated a high genetic isolation between the Noir Marron communities and Native American tribes in French Guiana. Nevertheless, only in the Oyampi tribe very low gene flow was detected with African Americans (one individual belonged to the African haplogroup E1b1a1-M2*).

Figure 49: Simplified Y phylogenetic tree and table of Y haplogroup frequencies and haplogroup diversity (HD ± Standard Deviation) estimated for each population in South America (SAM), West Africa (WAF), and West Europe (WEU). Y haplogroup nomenclature is according to ISOGG Y-DNA Haplogroup Tree 2015 (Y Chromosome Consortium, 2002). Results from Benin and Ivory Coast ethnic groups were reported by Fortes-Lima et al. (2015).

Region	Country	Population	N	HD±S.D.	A1b1b2b	B	B2a	E1a	E1b1a	E1b1a1*	E1b1a1c1a*	E1b1a1c1a1*	E1b1a1d1*	E1b1a1d1a*	E1b1a1d1a1	E1b1a1d1c	E1b1b1*	E1b1b1a1*	E1b1b1a	E1b1b2a*	E1b1b2a1	E2*	E2b1
SAM	French Guiana	Native American	13	0.2747 +/-0.1484																			
SAM	French Guiana	Noir Marron	30	0.8713 +/-0.0260		0.0333		0.2000		0.1667		0.1333	0.1667	0.2000									
SAM	Colombia	Afro-Colombian	20	0.9263 +/-0.0431						0.1000	0.0500	0.0500	0.0500	0.1500	0.0500			0.0500					0.0500
SAM	Brazil	Afro-Brazilian	28	0.8995 +/-0.0365		0.0357		0.0357				0.1786	0.1429	0.0357		0.0357				0.0357		0.0357	
WAF	Benin	Bariba	57	0.8333 +/-0.0230				0.0351	0.0175	0.0702	0.1228	0.1930	0.2807	0.2281									
WAF	Benin	Yoruba	55	0.7394 +/-0.0461			0.0909	0.0545		0.1636		0.4545	0.1450	0.0909									
WAF	Benin	Fon	63	0.7952 +/-0.0238	0.0159			0.0794		0.1270		0.3016	0.2698	0.1746								0.0159	0.0159
WAF	Ivory Coast	Ahizi	47	0.7465 +/-0.0286						0.2826		0.2174	0.1087	0.3696								0.0217	
WAF	Ivory Coast	Yacouba	32	0.7798 +/-0.0362					0.0333	0.2333		0.0333	0.4000	0.3000									
WAF	Mali	Bwa	14	0.5824 +/-0.0919						0.0714		0.3571		0.5714									
WEU	Spain	Galicia	21	0.6524 +/-0.1020																			
WEU	Spain	Barcelona	26	0.4123 +/-0.1194													0.0385						
WEU	Spain	Granada	20	0.7053 +/-0.1114													0.0500				0.0500		
WEU	Spain	Huelva	13	0.7692 +/-0.1031													0.0769				0.0769		
WEU	Portugal	Portugal	30	0.8276 +/-0.0515													0.0333	0.0333	0.1333				

Phylogenetic tree internal-node labels (root to tips):

CF — F-M213, M89 — GHIJK — G-M201 — G2-P287 — G2a-P15; HIJK — IJ — I-P203 — I1-P203; LJ — J-M304, P209 — J1-M267 — J2-M172 — J2a-M410 — J2a1b-M67 — J2b-M12; K-M9, P128 — L,T (or K1) — L-M20 (or K1a) — L1-M295 — L1a — L1a2-M357; T-M184, M272 (or K1b) — T1 — T1a-M70 — T1a2-L131; K2-M526 — P1-M45 (or K2b2) — P1-M45* (xM207) — R-M207 — R1-M173 — R1a-M434 — R1a1-SRY10831.2 — R1a1a-M17 (xM458); R1b1-M415 — R1b1a2-M269

Haplogroup terminal labels: F-M213*/M89* · G-M201* · G2-P287* · G2a · I1-P203.2 · J-M304* (xM267,M172) · J1-M267 · J2-M172* (xM410,M12) · J2a1b · J2b-M12* (xM241) · L1a2-M357 · T1a-M70* (xL131) · T1a2-L131 · K2-M526* · P1*-M45* (xM207) · P1* (Q-M242) · R1a1a*-M17* (xM458) · R1b1-M415* · R1b1a2-M269

Region	Country	Population	N	HD ± S.D.	F*	G*	G2*	G2a	I1	J*	J1	J2*	J2a1b	J2b*	L1a2	T1a*	T1a2	K2*	P1* (Q-M242)	R1a1a*	R1b1*	R1b1a2
SAM	French Guiana	Native American	13	0.2747 +/-0.1484														0.0714	0.8571			
SAM	French Guiana	Noir Marron	30	0.8713 +/-0.0260	0.0333													0.0333		0.0500		0.0333
SAM	Colombia	Afro-Colombian	20	0.9263 +/-0.0431					0.0500			0.0500	0.0500									0.2500
SAM	Brazil	Afro-Brazilian	28	0.8995 +/-0.0365		0.0357	0.0357			0.0357										0.0714	0.0357	0.2500
WAF	Benin	Bariba	57	0.8333 +/-0.0230												0.0175					0.0351	
WAF	Benin	Yoruba	55	0.7394 +/-0.0461																		
WAF	Benin	Fon	63	0.7952 +/-0.0238																		
WAF	Ivory Coast	Ahizi	47	0.7465 +/-0.0286																		
WAF	Ivory Coast	Yacouba	32	0.7198 +/-0.0362																		
WAF	Mali	Bwa	14	0.5824 +/-0.0919																		
WEU	Spain	Galicia	21	0.6524 +/-0.1020	0.0952				0.1905		0.0476	0.0476	0.0385				0.0476					0.5714
WEU	Spain	Barcelona	26	0.4123 +/-0.194	0.0385				0.0385		0.0769		0.0500									0.7692
WEU	Spain	Granada	20	0.7053 +/-0.1114	0.0500			0.1000	0.0500								0.0500	0.0500				0.5500
WEU	Spain	Huelva	13	0.7692 +/-0.1031	0.2308			0.0769														0.4615
WEU	Portugal	Portugal	30	0.8276 +/-0.0515	0.1667			0.0333			0.1000		0.0667	0.0333	0.0333							0.3667

A.1.a. Y-SNP diversity in the Noir Marron communities

Among African Americans, nine Y haplogroups were detected among the Noir Marron communities. Those Y chromosome markers revealed a high West and Central African component (50.0% of E1b1a1a1*)—even higher than in Afro-Brazilian (35.7%) and Afro-Colombian populations (25.0%). In Noir Marron, the highest value (20.0%) was detected for subhaplogroup E1b1a1a1d1a-U290*, which was also highly frequent among the West African populations analysed (Figure 49). We estimated pairwise F_{ST} distances between the Noir Marron and other populations analysed (Table 14). Pairwise F_{ST} distances indicated considerable genetic similarity between the Noir Marron and the Beninese populations; Fon (0.012) and Bariba (0.019).

Table 14: Comparison of pairwise F_{ST} genetic distances for Y haplogroup frequencies estimated for 10,000 permutations. Populations are labelled according to Table 10.

ID Pop	N		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
GUF_NM	30	1	0.000														
Af-COL	20	2	0.034	0.000													
Af-BRA	28	3	0.047	0.000	0.000												
NAM	13	4	0.369	0.368	0.366	0.000											
CIV_A	47	5	0.029	0.078	0.119	0.416	0.000										
CIV_Y	32	6	0.041	0.092	0.147	0.443	0.015	0.000									
MLI_B	14	7	0.107	0.143	0.175	0.569	0.039	0.160	0.000								
BEN_Y	55	8	0.065	0.123	0.086	0.419	0.079	0.151	0.133	0.000							
BEN_B	57	9	0.019	0.062	0.060	0.375	0.044	0.064	0.100	0.067	0.000						
BEN_F	63	10	0.012	0.082	0.059	0.385	0.039	0.076	0.105	0.020	0.007	0.000					
ESP_Ga	21	11	0.215	0.067	0.091	0.514	0.294	0.311	0.379	0.297	0.250	0.265	0.000				
ESP_Ba	26	12	0.334	0.178	0.185	0.642	0.399	0.426	0.518	0.398	0.352	0.364	0.027	0.000			
ESP_Gr	20	13	0.190	0.049	0.063	0.486	0.271	0.287	0.351	0.275	0.228	0.243	0.015	0.022	0.000		
ESP_Hu	13	14	0.155	0.034	0.050	0.484	0.244	0.259	0.326	0.249	0.199	0.215	0.008	0.086	0.024	0.000	
PRT	30	15	0.135	0.031	0.048	0.400	0.215	0.227	0.279	0.220	0.175	0.190	0.036	0.111	0.018	0.019	0.000

A low European contribution (3.3%) was detected associated with the haplogroup R1b1a2-M269. Moreover, low frequencies of haplogroups F-M213* and K2-M526* were detected (3.3% each), which have both a widespread distribution in Europe and Asia. This would be in good agreement with the low European contribution (2.4%) reported by Brucato et al. (2010).

We did not detect a Native American contribution to the Noir Marron. However, we found low frequencies of the haplogroup K2-M526 in Noir Marron (3.3%) and Native American (7.1%) populations, which might suggest a low gene flow of Southeast Asian populations. The haplogroup K2-M526 is likely originated in Southeast Asia, and is monophyleticly related to

haplogroups R and Q (Karafet et al., 2015). Therefore, it might indicate slight gene flow from Southeast Asian populations to African Americans and Native Americans because of recent migrations of Chinese populations in French Guiana (Poston & Mei, 1990).

A.1.b. Y-SNP diversity in the Afro-Brazilian population

In the study of Y chromosome markers among Afro-Brazilian, 14 different Y haplogroups were detected (Figure 49). We found 53.6% of African Y haplogroups, which is even lower than the previous estimation (69.0%) in Afro-Brazilian, also from Rio de Janeiro (Hunemeier et al., 2007). Only 42.9% of them belonged to the major African haplogroup E1b1*(P2), with notably 39.3% of the subhaplogroup E1b1a1*(M2), which is more frequent in West Africa (de Filippo et al., 2011). Moreover, two African subhaplogroups E1b1a1a1c1a1*(U174) and E1b1a1a1d1*(U209), which are widely present across West and Central African (de Filippo et al., 2011), have high frequencies (17.9% and 14.3%, respectively). We detected 46.4% of the Eurasian haplogroups G, J, and R, most of them belonged to haplogroup R (35.7%). The highest European percentage was observed for R1b1a2-M269 (25.0%), which is widely present across Western European (Balaresque et al., 2010). There was also a noteworthy presence of haplogroups G2 (7.1%) among Afro-Brazilians; this haplogroup is largely spread in the Middle East and the Caucasus (Rootsi et al., 2012). We did not identify any Native American contribution.

We detected the recurrent mutation SRY10831 (rs2534636) in two Afro-Brazilians that belonged to R1a1a-M17*. They presented positive signals for R1a1 associated with SNP SRY10831.2 (T). Interestingly, this mutation is also present in one individual from Benin, whose haplotype belongs to the haplogroup A1b1b2b-M13. As Hammer et al. (1998) demonstrated in the Y chromosome phylogeny, both an ancient T to C transition associated with the haplogroup BT (SRY10831.1), and a more recent C to T reversion associated with R1a1 (SRY10831.2) occurred at the nucleotide position SRY10831. Therefore, the individual from Benin shows the ancestral SNP (T), while the two Afro-Brazilians present the recurrent mutation SRY10831, which derived SNP is also T.

A.1.c. Y-SNP diversity in the Afro-Colombian population

In Afro-Colombian, thirteen Y haplogroups were detected, of which notably 45.0% belonged to the major African clade E1b1a1, which is extraordinarily frequent in West Africa (de Filippo et al., 2011), while 5.0% belonged to African subhaplogroup E2b1-M54. Conversely, 45.0%

belonged to European and Eurasian haplogroups I, J and mainly R, and Amerindian contribution was not detected (Figure 49). We also found low frequency (5.0%) for European and Near Eastern subhaplogroup E1b1b1a1-M78.

The African haplogroup E1b1a1a1d1a*(U290) has a high frequency (15.0%), which is also widely present across West-Central African (de Filippo et al., 2011). The highest percentages were observed for European subhaplogroup R1b1a2-M269 (25.0%), which is widely present across Western Europeans (Balaresque et al., 2010). Similar to the Afro-Brazilians, pairwise F_{ST} distances exhibited similar values between Afro-Colombian and West African populations, than between Afro-Colombian and West European populations (see Table 14).

A.2.   Y-STR diversity detected

All populations presents a high-level of genetic diversity for 17 Y-STRs (HD = 1.00 to 0.90), with the exception of Native American populations of Oyampi (0.69) and Palikour (0.88). In Native American in general, diversity values were lower for both minimum and Yfiler haplotypes (particularly for the discrimination capacity; 0.315 and 0.454, respectively) (see Table 15). Among the Noir Marron communities, haplotype diversity was higher in Ndjuka (HD= 0.995 +/0.017, see Table 16) than in other Marron communities.

Several Y-STRs present patterns of variation in repeats number associated with specific continental population (Figure 50). For instance, repeat 13 of DYS19 was mainly detected in Native American populations, while repeats 15, 16 and 17 of DYS19 were detected in high frequency in the African-American and African populations. Conversely, we detected other repeats present in high frequencies in European populations and admixed populations, such as repeat 11 of DYS391, repeat 15 of DYS437, and repeat 12 of DYS438. In the case of DYS385 a/b, we detected a noticeable contrast between the frequency for 11,14 belonged mainly to European, Afro-Brazilian, and Afro-Colombian, with the frequency for 16,17 belonged mainly to African, Noir Marron, and Afro-Colombian (Figure 50).

NETWORK analysis revealed phylogenetic relationships between microsatellite haplotypes belonging to West African and African-American populations (Figure 51). Determinate microsatellite haplotypes cluster in close correspondence to Y haplogroups identified (Figure 49). For instance, haplotypes associated with E1b1a1a1d1a-U290 shows a 'star-like' structure indicative of expansion from one source, 19 haplotypes (3%) lie in its central node, and this node plus its single-step mutational neighbours together comprise 56 haplotypes (8.8%). As expected, in this cluster were detected West African and African-American populations. Other haplotypes associated with Y haplogroups E1b1a1-M2*, E1b1a1a1c1a1-U174, E1b1a1a1d1-

U209 present strong phylogenetic affiliations for each Y haplogroup. While some haplotypes spread far away from those clusters. While remaining haplogroups are indicative of phylogenetic branches.

Similarly, haplotypes associated with R1b1a2-M269 shows a 'star-like' structure (Balaresque et al., 2010), which 14 haplotypes (2.2%) lie in its central node, and this node plus its single-step mutational neighbours together comprise 40 haplotypes (6.3%). In this cluster were detected a high representation of West European individuals, and also substantial number of Afro-Brazilian and Afro-Colombian individuals. Both African-American populations present clear evidence of admixture, because of their dual distribution within African and European clusters.

Native American populations presents a different picture with multiple mutational steps between neighbour haplotypes, however haplotype substructure was not found. Native Americas haplotypes from French Guiana belonging to haplogroup P-M45* are close to haplotypes associate with subhaplogroup Q1. The haplogroup P-M45* is ancestor of that subhaplogroup; therefore, it is feasible to argument that those Native American groups are also relate to subhaplogroup Q1. Remarkable, Native Americas haplotypes are not closely relate to African American haplotypes, not even populations from the same country. Therefore, there is no evidence of sex-specific between those groups.

To identify clusters between-group structure and to confirm population demographic patterns, we analysed the variation of among Y-STRs polymorphisms using DAPC (Jombart et al., 2010) and STRUCTURE analysis (Pritchard et al., 2000). DAPC plot shows three major clusters among populations with different continental ancestry (see Figure S1). The a-score was 0.749, which means that the probability of re-assignment of populations to true clusters is three times higher than to randomly permuted clusters (Jombart et al., 2010). The first discriminant function splits populations with Native American ancestry, while the second discriminant function splits populations with African and European ancestries. Both, Afro-Brazilian and Afro-Colombian populations present inertia ellipses around African and European populations. Therefore, Y-STRs adds further support to admixture patterns occurred in descendants from Brazil and Colombia. Conversely, Noir Marron populations are within populations with African ancestry, with the exception the Paramaka that inertia ellipse is also associate with European populations. It may indicate gene flow between the Paramaka and European, however the sample size of this population is very low (N= 5) that it cannot be support.

Table 15: Intra and inter-population molecular diversity index for minHt.

minHt (7 Y-STR)	ID Pop	N	K	DC	θ *Hom*	HD	*h*	MNPD
Populations								
Aluku	*NM_A*	8	5	*0.625*	7.08	0.893 +/- 0.086	0.342 +/- 0.236	2.393 +/- 1.450
Ndjuka	*NM_N*	21	16	*0.762*	39.19	0.976 +/- 0.020	0.521 +/- 0.307	3.648 +/- 1.924
Paramaka	*NM_P*	5	3	*0.600*	3.17	0.800 +/- 0.164	0.571 +/- 0.401	4.000 +/- 2.399
Saramaka	*NM_S*	16	8	*0.500*	4.64	0.850 +/- 0.075	0.342 +/- 0.219	2.392 +/- 1.372
Total Noir Marron	GUF_NM	50	26	*0.520*	21.82	0.959 +/- 0.013	0.477 +/- 0.276	3.340 +/- 1.743
Afro-Colombian	Af-COL	19	18	*0.947*	168.05	0.994 +/- 0.019	0.582 +/- 0.339	4.076 +/- 2.126
Afro-Brazilian	Af-BRA	24	21	*0.875*	89.09	0.989 +/- 0.015	0.609 +/- 0.349	4.264 +/- 2.190
Ahizi	CIV_A	49	31	*0.633*	36.40	0.975 +/- 0.009	0.479 +/- 0.277	3.350 +/- 1.748
Yacouba	CIV_Y	41	24	*0.585*	12.68	0.934 +/- 0.025	0.371 +/- 0.225	2.600 +/- 1.421
Fon	BEN_F	78	51	*0.654*	52.75	0.982 +/- 0.006	0.497 +/- 0.284	3.476 +/- 1.792
Yoruba	BEN_Y	54	30	*0.556*	32.99	0.972 +/- 0.009	0.463 +/- 0.269	3.240 +/- 1.697
Bariba	BEN_B	51	34	*0.667*	37.04	0.975 +/- 0.011	0.488 +/- 0.281	3.414 +/- 1.775
Bwa	MLI_B	13	4	*0.308*	1.43	0.654 +/- 0.106	0.286 +/- 0.193	2.000 +/- 1.204
Galicia	ESP_Ga	21	14	*0.667*	18.37	0.952 +/- 0.030	0.445 +/- 0.268	3.114 +/- 1.683
Barcelona	ESP_Ba	30	21	*0.700*	36.74	0.975 +/- 0.015	0.470 +/- 0.277	3.290 +/- 1.740
Granada	ESP_Gr	25	20	*0.800*	34.71	0.973 +/- 0.022	0.514 +/- 0.301	3.600 +/- 1.890
Huelva	ESP_Hu	25	20	*0.800*	47.16	0.980 +/- 0.018	0.563 +/- 0.325	3.940 +/- 2.042
Portugal North	PRT_N	7	6	*0.857*	18.37	0.952 +/- 0.096	0.619 +/- 0.399	4.333 +/- 2.436
Portugal South	PRT_S	23	20	*0.870*	60.38	0.984 +/- 0.019	0.589 +/- 0.340	4.126 +/- 2.132
Kalinya (F.G.)	GUF_K	23	12	*0.522*	11.58	0.929 +/- 0.032	0.519 +/- 0.305	3.636 +/- 1.912
Oyampi (F.G.)	GUF_O	25	7	*0.280*	0.71	0.487 +/- 0.121	0.269 +/- 0.177	1.880 +/- 1.111
Palikour (F.G.)	GUF_P	41	10	*0.244*	4.68	0.851 +/- 0.028	0.417 +/- 0.248	2.922 +/- 1.564
Emberá-Chamí	COL-EC	24	9	*0.375*	6.42	0.884 +/- 0.037	0.549 +/- 0.319	3.844 +/- 2.002
Karitiana	BRA-KA	17	6	*0.353*	3.05	0.794 +/- 0.063	0.299 +/- 0.197	2.096 +/- 1.230
Groups								
African American	AAM	93	57	*0.613*	54.95	0.983 +/- 0.005	0.550 +/- 0.309	3.850 +/- 1.952
West African	AFR	286	125	*0.437*	46.62	0.980 +/- 0.003	0.471 +/- 0.269	3.296 +/- 1.700
West European	EUR	131	78	*0.595*	36.26	0.974 +/- 0.007	0.530 +/- 0.298	3.708 +/- 1.886
Native American	NAM	130	41	*0.315*	16.00	0.946 +/- 0.010	0.538 +/- 0.302	3.768 +/- 1.912

Note: N = Sample size, **K** = Number of different haplotypes, **DC** = Discrimination capacity, *θ Hom* = Theta (*Hom*), **HD** = Haplotype diversity, *h* = Gene diversity over loci, **MNPD** = Mean number of pairwise differences.

Table 16: Intra and inter-population molecular diversity index for Yfiler.

Yfiler (17 Y-STR)	ID Pop	N	K	DC	*θ Hom*	HD	*h*	MNPD
Populations								
Aluku	*NM_A*	8	6	*0.750*	11.53	0.929 +/- 0.084	0.456 +/- 0.271	7.750 +/- 4.042
Ndjuka	*NM_N*	21	20	*0.952*	207.04	0.995 +/- 0.017	0.586 +/- 0.312	9.962 +/- 4.746
Paramaka	*NM_P*	5	4	*0.800*	7.71	0.900 +/- 0.161	0.612 +/- 0.394	10.400 +/- 5.729
Saramaka	*NM_S*	16	11	*0.688*	17.38	0.950 +/- 0.036	0.473 +/- 0.260	8.042 +/- 3.943
Total Noir Marron	GUF_NM	50	39	*0.780*	91.32	0.989 +/- 0.006	0.562 +/- 0.291	9.548 +/- 4.453
Afro-Colombian	Af-COL	19	19	*1.000*	-	1.000 +/- 0.000	0.648 +/- 0.345	11.023 +/- 5.242
Afro-Brazilian	Af-BRA	24	24	*1.000*	-	1.000 +/- 0.000	0.671 +/- 0.351	11.402 +/- 5.359
Ahizi	CIV_A	49	34	*0.694*	48.28	0.980 +/- 0.009	0.475 +/- 0.249	8.074 +/- 3.813
Yacouba	CIV_Y	41	34	*0.829*	79.10	0.988 +/- 0.010	0.436 +/- 0.231	7.413 +/- 3.538
Fon	BEN_F	78	74	*0.949*	747.76	0.999 +/- 0.002	0.549 +/- 0.282	9.329 +/- 4.332
Yoruba	BEN_Y	54	45	*0.833*	156.05	0.994 +/- 0.005	0.524 +/- 0.272	8.904 +/- 4.168
Bariba	BEN_B	51	44	*0.863*	138.72	0.993 +/- 0.006	0.498 +/- 0.260	8.461 +/- 3.979
Bwa	MLI_B	13	6	*0.462*	2.61	0.769 +/- 0.103	0.307 +/- 0.178	5.218 +/- 2.699
Galicia	ESP_Ga	21	19	*0.905*	102.08	0.991 +/- 0.018	0.566 +/- 0.302	9.619 +/- 4.593
Barcelona	ESP_Ba	30	27	*0.900*	142.05	0.993 +/- 0.011	0.534 +/- 0.281	9.083 +/- 4.300
Granada	ESP_Gr	25	25	*1.000*	-	1.000 +/- 0.000	0.612 +/- 0.322	10.403 +/- 4.909
Huelva	ESP_Hu	25	23	*0.920*	147.05	0.993 +/- 0.013	0.627 +/- 0.329	10.653 +/- 5.020
Portugal North	PRT_N	7	7	*1.000*	-	1.000 +/- 0.000	0.650 +/- 0.386	11.048 +/- 5.725
Portugal South	PRT_S	23	23	*1.000*	-	1.000 +/- 0.000	0.662 +/- 0.348	11.261 +/- 5.305
Kalinya (F.G.)	GUF_K	23	14	*0.609*	20.34	0.957 +/- 0.022	0.508 +/- 0.271	8.632 +/- 4.139
Oyampi (F.G.)	GUF_O	25	10	*0.400*	1.70	0.690 +/- 0.102	0.291 +/- 0.163	4.953 +/- 2.494
Palikour (F.G.)	GUF_P	41	12	*0.293*	6.26	0.882 +/- 0.026	0.438 +/- 0.232	7.449 +/- 3.553
Emberá-Chamí	COL-EC	24	13	*0.542*	10.70	0.924 +/- 0.032	0.542 +/- 0.288	9.214 +/- 4.390
Karitiana	BRA-KA	17	10	*0.588*	9.96	0.919 +/- 0.044	0.229 +/- 0.135	3.890 +/- 2.053
Groups								
African American	AAM	93	82	*0.882*	326.10	0.997 +/- 0.002	0.621 +/- 0.316	10.550 +/- 4.850
West African	AFR	286	235	*0.822*	519.52	0.998 +/- 0.001	0.512 +/- 0.262	8.703 +/- 4.030
West European	EUR	131	124	*0.947*	1,213.44	0.999 +/- 0.001	0.605 +/- 0.308	10.280 +/- 4.722
Native American	NAM	130	59	*0.454*	33.20	0.972 +/- 0.006	0.546 +/- 0.280	9.280 +/- 4.292

Note: N = Sample size, **K** = Number of different haplotypes, **DC** = Discrimination capacity, *θ Hom* = Theta (*Hom*), **HD** = Haplotype diversity, *h* = Gene diversity over loci, **MNPD** = Mean number of pairwise differences.

Figure 50: Frequencies of 17 Y-STRs markers using in Yfiler. Population from African-American (Noir Marron, Afro-Colombian, and Afro-Brazilian), African (Yacouba, Ahizi, and Fon), European (Granada, Huelva, Barcelona, Galicia, and Portugal), and Native American (Kaliña, Palikur, and Oyampi) populations.
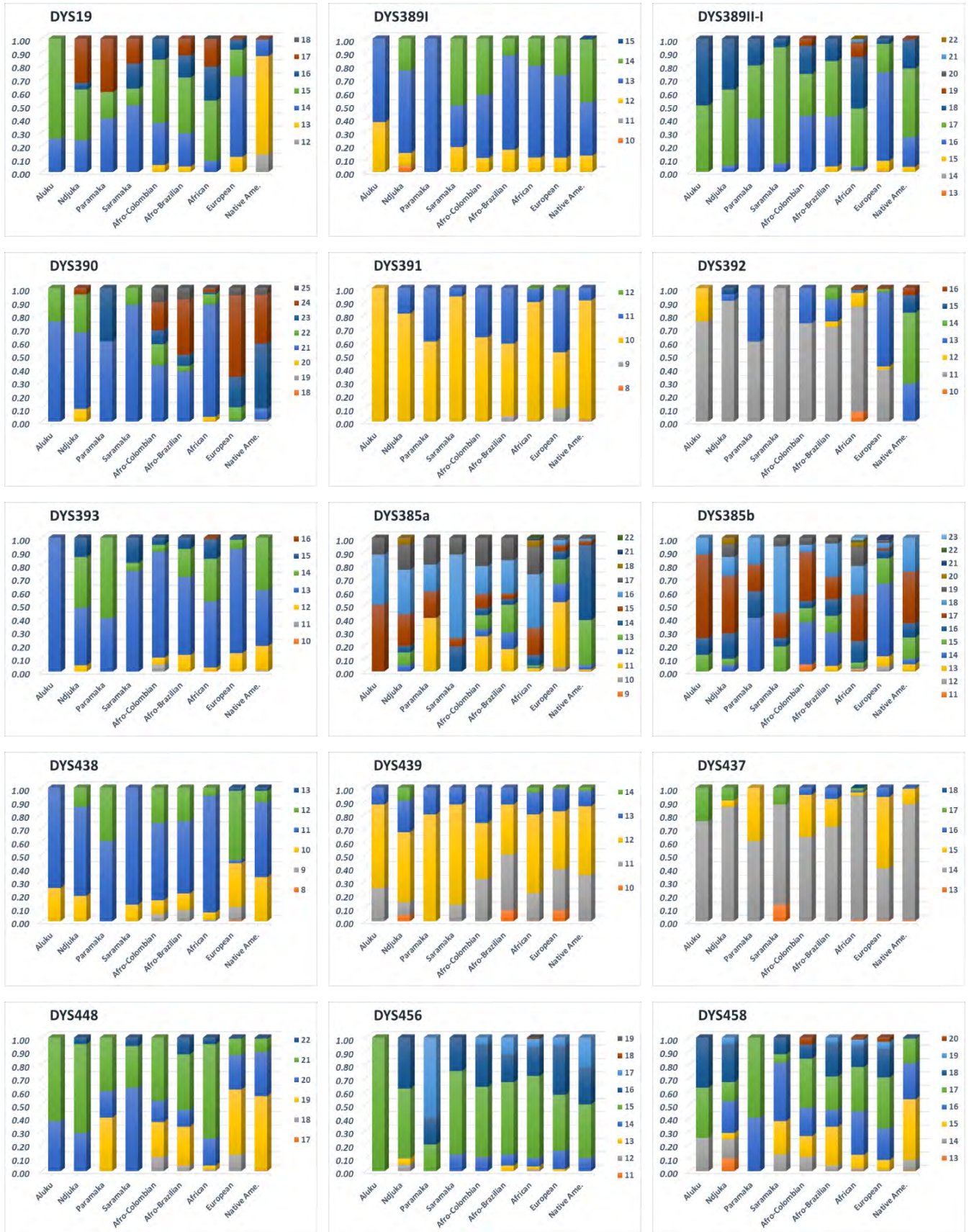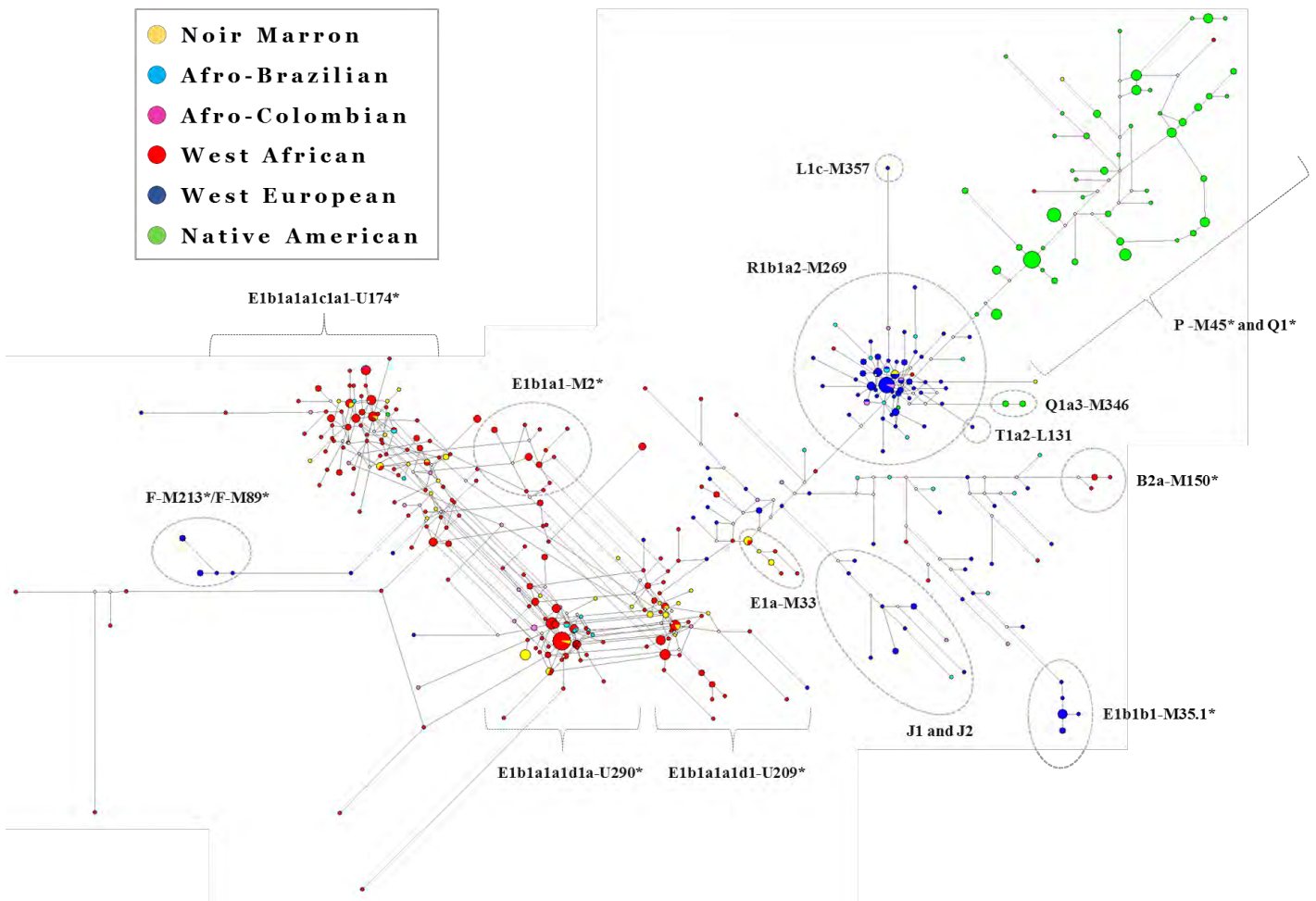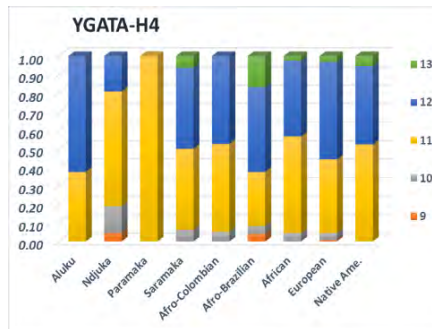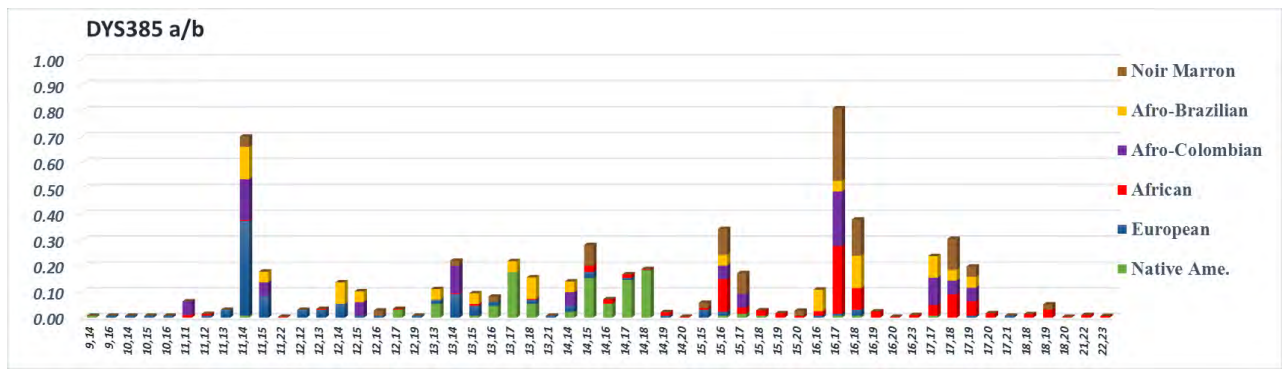
Figure 51: Median-joining Network of 640 Y microsatellite haplotypes between African-American, West African, West European, and Native American populations. Molecular relationships based on 9 Y-STR haplotypes. Circles represent haplotypes with area proportional to frequency, and coloured according to population label. Lines between circles represent microsatellite mutational steps.

STRUCTURE analysis distinguished among populations with different continental ancestry (Figure S2). This Bayesian model-based clustering approach use microsatellite data to infer population structure and assign individuals to unknown populations (Pritchard et al., 2000). Individuals from the Noir Marron communities were highly assigned to West African populations (78%) (Table S3). Both, Afro-Brazilians and Afro-Colombians present African (~51%) and European (~43%) proportions that are consistent with previous results based on Y haplogroup frequencies (Figure 49). Interestingly, the Amerindian population of Emberá-Chamí from Colombia shows a high Native American component (Roewer et al., 2013) and substantial proportion of European component (22.8%) that may reflect European paternal gene flow. However, some individuals were erroneously identify with different continental ancestry and may indeed suggest a bias. Despite this method can detect strong signals of population structure, individuals might not be accurately clustered.

A.3. *Spatial distribution of the African Y haplogroups*

We elaborated contour maps based on a comprehensive dataset of 23 African populations (Table S6) in order to gain a better understanding of African origin of major Y haplogroup (Figure 52). The haplogroup E1b1a1-M2* presented a West African distribution associated with Senegambia region based on the high-levels detected in Senegal and Guinea-Bissau populations (de Filippo et al., 2011). The haplogroup E1b1a1a1c1a1-M191 (including its subhaplogroup E1b1a1a1c1a1-U174) depicted mainly a Central African distribution associated with Bight of Biafra, with dominant focus in Cameroon, Gabon, C.A.R., and D.R.C. Both haplogroups present low frequencies among African Americans that are close to isoclines in other West African regions involved in the slave trade.

Instead, the haplogroup E1b1a1a1d1-U175 presented a wide distribution from West to South Africa, with high-levels in Angola (62.7%) (Coelho et al., 2009; de Filippo et al., 2011) and Mali (58.3%). It is particularly observed along coastal African regions involved in the transatlantic slave trade, and therefore presents in high frequencies among Noir Marron, Afro-Colombian, and Afro-Brazilian populations. Specially, frequencies observed in Windward Coast (51.3%), Gold Coast (37.0%), and Bight of Benin (39.4%) regions match closely with the Noir Marron communities (41.7%).

Hg E1b1a1-M2*
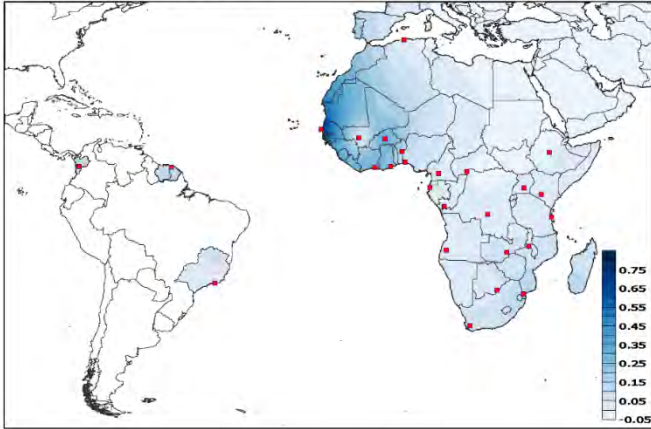
Hg E1b1a1a1c1a-M191
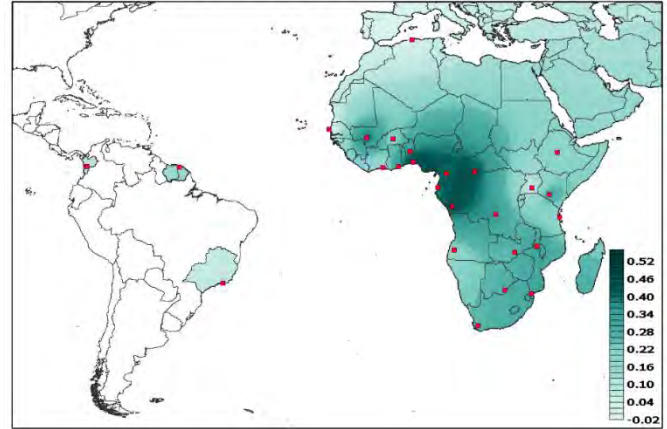
Hg E1b1a1a1d-U175
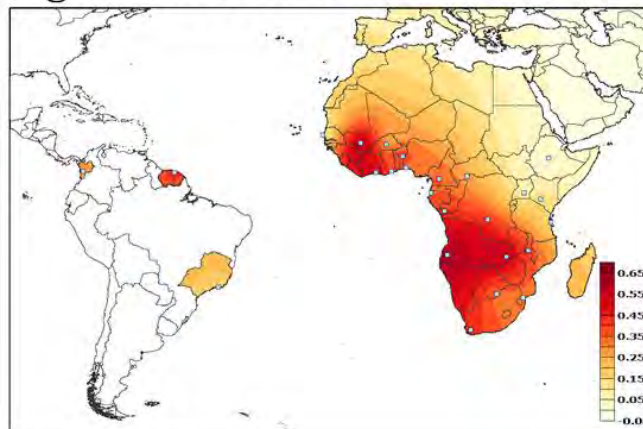
Figure 52: Patterns of geographic distributions of haplogroup E1b1a1-M2*, E1b1a1a1c1a-M191, and E1b1a1a1d-U175, in the African-American and African populations across the continent.

A.4. *Origin of African Y chromosome diversity among the African Americans*

We study the African origin of the African-American populations within the African continent from the Y chromosome perspective using both Y-STRs and Y-SNPs markers. We calculated the matrix of pairwise R_{ST} genetic distances (Slatkin, 1995) and coancestry coefficients (Reynolds et al., 1983) based on 17 Y-STRs. Both parameters indicated closed similarities among West African populations and African-American groups, especially with populations from Benin and Ivory Coast (Table 18). Conversely, African Americans presented the biggest dissimilarities with European and Native American groups, with the exception of European with the Afro-Brazilian and Afro-Colombian populations. Lastly, the biggest dissimilarities were detected among groups with different continental ancestry, which means African, European and Native American.

We determined the proportions of continental ancestry in African Americans based on both Y chromosome markers. Both admixture analysis based on seven Y-STR haplotypes and Y haplogroup frequencies indicated the highest African ancestry (>86%) in the Noir Marron communities. In contrast, both Afro-Brazilian and Afro-Colombian population showed high European ancestry, which is extraordinary elevate (67.9%) in Afro-Colombian according to Y-STRs haplotypes. Native American contribution was very low or null among African Americans.

Admixture estimates of African Americans calculated based on Y-STR haplotypes frequencies indicated high (89.1%) West African ancestry in the Noir Marron, while Afro-Brazilians (41.4%) and Afro-Colombians (67.9%) showed high West European ancestry. Likewise, we obtained similar values in each population for mY_{SNP} estimations (see Table 17).

Table 17: Gene contribution estimates between populations based on variation in Y chromosome markers; Y-STR haplotypes frequencies obtained by using Lineage sharing (L.S.) analysis, and Y haplogroup frequencies (Figure 49). Admixture proportions ($mY_{Y-SNP} \pm$ S.D.) were estimated by using ADMIX v.2.0. Populations analysed are indicated in Table 10.

Y-STRs (L.S.)	N	*West African* 286	*West European* 131	*Native American* 130
Noir Marron	50	0.8911	0.0604	0.0486
Afro-Brazilian	24	0.5494	0.4143	0.0363
Afro-Colombian	19	0.3211	0.6789	0.0000

Y-SNPs (mY_{Y-SNP})	N	*West African* 268	*West European* 110	*Native American* 13
Noir Marron	30	0.8603 ±*0.0127*	0.1288 ±*0.0123*	0.0109 ±*0.0013*
Afro-Brazilian	28	0.5269 ±*0.0206*	0.4658 ±*0.0211*	0.0073 ±*0.0013*
Afro-Colombian	20	0.5049 ±*0.0214*	0.4880 ±*0.0220*	0.0070 ±*0.0012*

Table 18: Matrix of pairwise R_{ST} genetic distances (in the lower diagonal) and coancestry coefficients (in the upper diagonal) between African American and refence populations. Both parameters were estimated for pairwise population and based on 17 Y-STRs.

Groups	ID Pop	N	African American						West African						West European						Native American				
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
NM_A	1	8		0.0056	0.3345	0.0597	0.1787	0.1790	0.1532	0.1463	0.0438	0.0374	0.1167	0.4030	0.6993	0.6658	1.0989	0.8437	0.7336	0.4663	0.6662	1.1794	0.7698	1.0786	1.4331
NM_N	2	21	0.0056		0.1289	0.0000	0.1518	0.1518	0.0820	0.0931	0.0000	0.0000	0.0220	0.0538	0.6235	0.6018	0.9066	0.7045	0.5926	0.4588	0.4773	0.7788	0.6470	0.8513	0.8618
NM_P	3	5	0.2843	0.1210		0.2590	0.0000	0.0000	0.3989	0.3358	0.1941	0.2006	0.3350	0.5012	0.2220	0.2188	0.4737	0.3128	0.1825	0.1113	0.3984	0.6579	0.4965	0.6894	0.9177
NM_S	4	16	0.0579	-0.0123	0.2281		0.1777	0.1810	0.1147	0.1735	0.0048	0.0252	0.0314	0.1412	0.7246	0.6912	1.0747	0.8619	0.7666	0.5360	0.5624	0.9699	0.6935	1.0367	1.2175
Af-COL	5	19	0.1637	0.1408	-0.0611	0.1628		0.0000	0.4030	0.2393	0.2139	0.2332	0.3209	0.3535	0.1520	0.1442	0.3514	0.2027	0.1202	0.0911	0.2746	0.4059	0.3239	0.4424	0.4318
Af-BRA	6	24	0.1639	0.1409	-0.0372	0.1655	-0.0264		0.3750	0.2455	0.2058	0.2170	0.2977	0.3356	0.1710	0.1517	0.3630	0.2172	0.1233	0.0941	0.2562	0.4052	0.3361	0.4189	0.4210
CIV_A	7	49	0.1421	0.0787	0.3290	0.1084	0.3317	0.3127		0.1746	0.0426	0.0586	0.0312	0.0396	0.9240	0.8852	1.1663	1.0067	0.9361	0.7424	0.6446	0.9520	0.8456	1.0203	1.0185
CIV_Y	8	41	0.1361	0.0889	0.2853	0.1593	0.2129	0.2177	0.1602		0.0560	0.0671	0.1049	0.2924	0.8161	0.7894	1.1353	0.9489	0.9442	0.6027	0.7651	1.1317	0.8599	1.1392	1.2155
BEN_F	9	78	0.0429	-0.0077	0.1764	0.0048	0.1925	0.1860	0.0417	0.0544		0.0008	0.0000	0.0255	0.6920	0.6680	0.9105	0.7510	0.6932	0.5497	0.4875	0.7210	0.6467	0.8245	0.7934
BEN_Y	10	54	0.0367	-0.0108	0.1818	0.0249	0.2080	0.1950	0.0569	0.0649	0.0008		0.0098	0.0469	0.7031	0.6839	0.9401	0.7654	0.6865	0.5498	0.5574	0.7957	0.7099	0.8861	0.8479
BEN_B	11	51	0.1101	0.0217	0.2847	0.0309	0.2745	0.2575	0.0307	0.0996	-0.0005	0.0097		0.0419	0.8779	0.8467	1.1492	0.9746	0.9197	0.6976	0.6425	0.9576	0.8077	1.0644	1.0853
MLI_B	12	13	0.3317	0.0524	0.3942	0.1316	0.2978	0.2851	0.0388	0.2535	0.0252	0.0458	0.0410		0.9964	0.9463	1.3808	1.1977	1.1507	0.7546	0.8652	1.4411	1.0301	1.3894	1.7900
ESP_Gr	13	25	0.5031	0.4640	0.1991	0.5155	0.1410	0.1571	0.6031	0.5579	0.4994	0.5049	0.5844	0.6308		0.0000	0.0235	0.0000	0.0000	0.0194	0.5205	0.5143	0.4417	0.4253	0.4749
ESP_Hu	14	25	0.4862	0.4522	0.1965	0.4990	0.1343	0.1407	0.5874	0.5459	0.4873	0.4953	0.5712	0.6118	-0.0159		0.0560	0.0033	0.0000	0.0000	0.4760	0.4537	0.4161	0.3837	0.4359
ESP_Ba	15	30	0.6668	0.5961	0.3773	0.6586	0.2963	0.3044	0.6885	0.6787	0.5977	0.6094	0.6831	0.7486	0.0232	0.0545		0.0218	0.0291	0.1091	0.7489	0.7319	0.6494	0.5626	0.6951
ESP_Ga	16	21	0.5699	0.5057	0.2686	0.5776	0.1835	0.1952	0.6346	0.6128	0.5281	0.5348	0.6227	0.6981	-0.0273	0.0033	0.0215		0.0000	0.0274	0.6286	0.6616	0.5445	0.5193	0.6453
PRT_N	17	7	0.5198	0.4471	0.1669	0.5354	0.1132	0.1160	0.6079	0.6110	0.5000	0.4967	0.6014	0.6836	-0.0558	-0.0409	-0.0287	-0.0392		0.0000	0.6419	0.7525	0.5758	0.5741	0.7556
PRT_S	18	23	0.3727	0.3680	0.1053	0.4149	0.0870	0.0898	0.5240	0.4527	0.4229	0.4229	0.5022	0.5298	0.0192	-0.0086	0.1033	0.0270	-0.0401		0.4199	0.4362	0.4110	0.3707	0.3804
GUF_K	19	23	0.4863	0.3796	0.3286	0.4301	0.2401	0.2260	0.4751	0.5347	0.3859	0.4273	0.4740	0.5790	0.4058	0.3788	0.5271	0.4667	0.4737	0.3429		0.2902	0.1109	0.2697	0.4750
GUF_O	20	25	0.6925	0.5410	0.4821	0.6209	0.3337	0.3332	0.6141	0.6775	0.5137	0.5487	0.6162	0.7633	0.4021	0.3647	0.5190	0.4840	0.5288	0.3535	0.2519		0.1900	0.4504	0.7758
GUF_P	21	41	0.5369	0.4764	0.3914	0.5002	0.2767	0.2855	0.5707	0.5768	0.4762	0.5083	0.5541	0.6430	0.3571	0.3404	0.4776	0.4199	0.4377	0.3370	0.1049	0.1730		0.2509	0.3458
COL-EC	22	24	0.6599	0.5731	0.4981	0.6454	0.3575	0.3422	0.6395	0.6799	0.5615	0.5877	0.6551	0.7508	0.3464	0.3187	0.4303	0.4051	0.4368	0.3097	0.2364	0.3627	0.2219		0.1875
BRA-KA	23	17	0.7614	0.5776	0.6003	0.7040	0.3507	0.3436	0.6389	0.7035	0.5477	0.5717	0.6622	0.8330	0.3781	0.3533	0.5010	0.4755	0.5303	0.3164	0.3781	0.5397	0.2924	0.1710	

The PCA analysis of the African-American and African populations residing in African coastal regions showed close genetic variation between Bight of Biafra and West Central Africa regions (Figure 53). The first principal component (or PC1) mainly split Senegambia and Sierra Leone regions from Bight of Biafra and West Central Africa regions. The second principal component (or PC2) split populations from remaining regions.

We could not establish a unique genetic link with any specific African region. In this bidimensional space, the African-American populations are close to several West African regions, suggesting multiple African sources in genetic pool of African descendants. Populations from Senegambia region might not contribute in this parental variation. Only Afro-Brazilian populations was slightly close to Southeast African region, in agreement with historical data (Eltis & Richardson, 2010). Interestingly, African Americans from Barbados (Table S5) are close to the Aluku, while African Americans from USA (Table S5) are close to Afro-Colombians.

Figure 53: PCA analysis of the African-American and African populations from different historical coastal regions based on Y haplogroup frequencies. Populations are labelled according to Table S6. PC1 (27.91%) and PC2 (14.31%) constitute 42.22% of the total variance. European and Native American proportions were removed in admixed African-American populations and recalculated their frequencies.

B. Mitochondrial DNA results

The increasing availability of complete mtDNA genome sequences from humans has greatly refined the human mtDNA phylogenetic tree, and provided new insights into the phylogeography of particular mtDNA haplogroups (Torroni et al., 2006). Recently, new methods have been developed for high-throughput and low-cost sequencing of complete mtDNA genomes by using a parallel tagged sequencing approach and new sequencing platforms (Maricic et al., 2010). Here, we have applied this approach to obtain 273 complete mtDNA genome sequences for African-American and West African populations. A total of 27,415,099 unpaired and 276,891,367 paired reads were obtained per each individual. Both unpaired and paired reads were mapped to the rCRS. The average coverage per position for the total mtDNA sequences in the final dataset was 1,931 (range 266–3,901), with an average minimum coverage of 508 (range 47–1,784) and an average maximum coverage of 2,765 (range 437–5,515) (Figure 54). All positions were covered more than the minimum requirement of 20X.
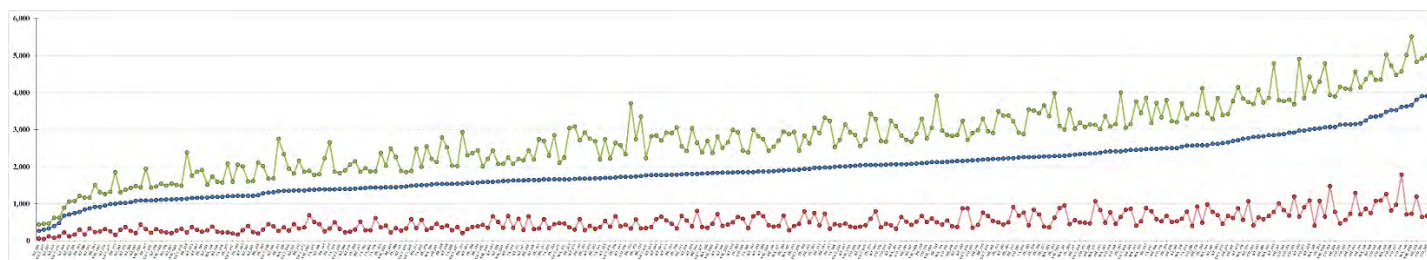
Figure 54: Average coverage (blue dots), minimum (red dots), and maximum coverage (green dots) per position for the African-American and West African mtDNA genome sequences.

B.1. Heteroplasmy calling detected

Heteroplasmy refers to the existence of different types of mtDNA in the same individual. Typically, there are approximately 100 mitochondria in each mammalian cell, and each mitochondrion harbours 2–10 copies of mtDNA (Robin & Wong, 1988). Therefore, somatic mutation on mtDNA are relatively frequent in a cell containing a mixture of normal and mutant mtDNA copies (Irwin et al., 2009). It has been found that heteroplasmies are randomly occurring on the mitochondrial genome and relatively common in healthy individuals (Ramos et al., 2013). Moreover, the frequency of heteroplasmic variants varies considerably between different tissues in the same individual (He et al., 2010). In a study based on earlier sequencing technology, heteroplasmies (or heteroplasmic sites) as low as 5% were detectable (Li et al., 2010). Later studies have shown that with a read depth of tens of thousands, mtDNA heteroplasmies as low as 1% could be detected (Guo et al., 2012). Therefore, to detect

heteroplasmies less than 1%, it is important to significantly increase the depth of coverage.

We detected a high number of heteroplasmic sites in the mtDNA genomes analysed. However, these heteroplasmic sites were not in essential sites to obtain the haplogroup affiliation. Figure 55 summarises heteroplasmic sites detected across the mtDNA genome over all samples.
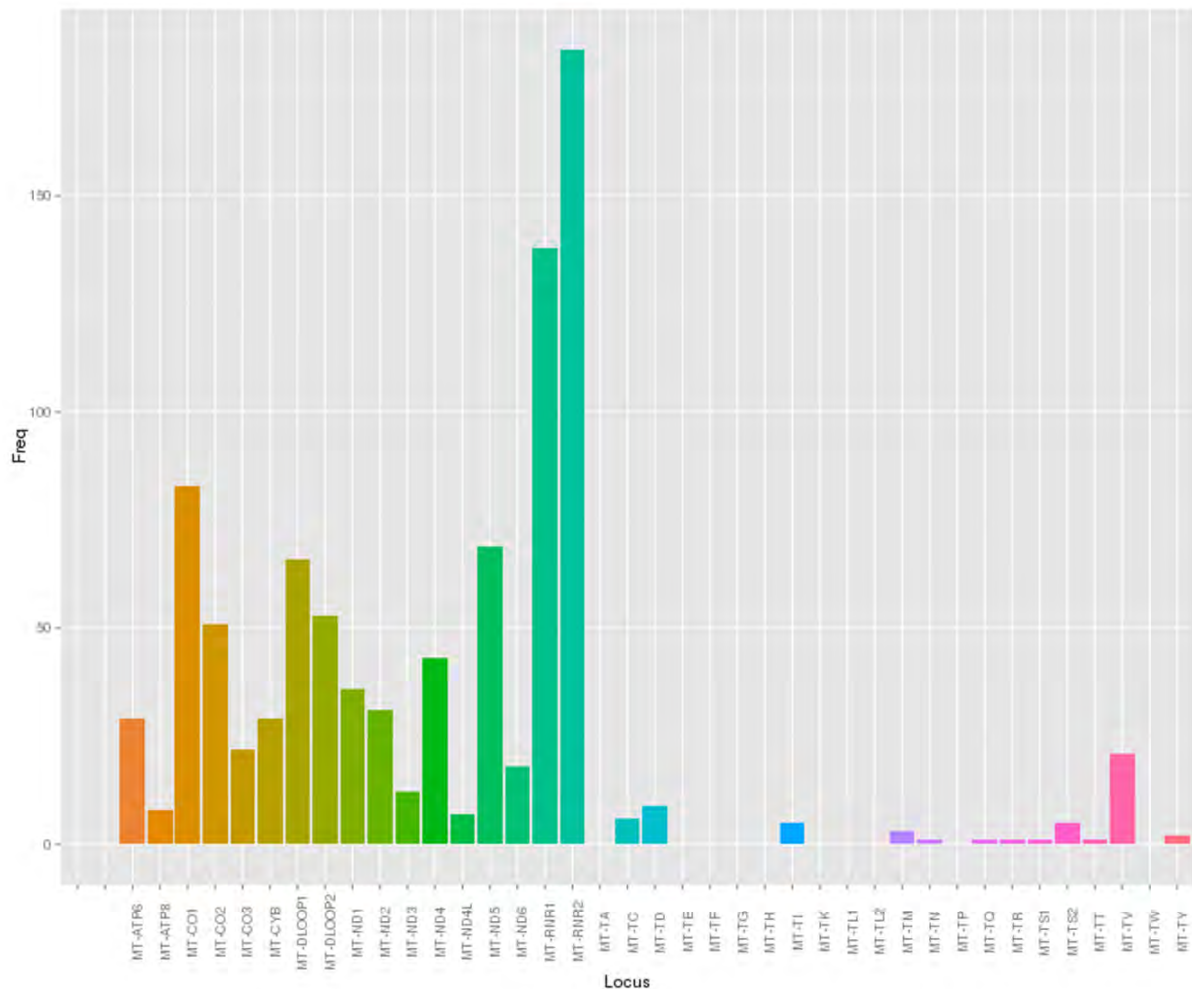
Figure 55: Heteroplasmic sites grouped according their loci on the mitochondrial genome over all samples.

The level of heteroplasmy detectable in mtDNA is heavily dependent on the depth of coverage (Ye et al., 2014). These parameters are good indicators of the depth of coverage and lowered sequencing error rates that we obtained in this high- throughput sequencing. We calculated the average heteroplasmy count over all samples (103 ±50, range 9–331), and for each individual. The frequency of heteroplasmic sites for all samples was low (0.620 ±0.301%, range 0.054–1.998%) because of the high coverage obtained with the high-throughput sequencing performed (Maricic et al., 2010). Besides, 27 of the 273 samples presented frequencies of heteroplasmic sites over 1.0%, and four of the 273 samples presented frequencies of heteroplasmic sites over 1.5%.

B.2. *Mitochondrial DNA diversity detected*

Sequence diversity indices indicated a high haplotype diversity in African Americans (range 0.9770–0.9933) and West African populations (range 0.9249–0.9952). As expected, the estimations were higher for the control region than the whole mtDNA genome (see Table 19). The Tajima's *D* statistics presented significant *P*-values for Afro-Brazilians and populations from Benin (see Table 19). The Tajima's D values were significant in Afro-Brazilians and Beninese populations for the mtDNA genome analysis, however we did not obtained significant negative Tajima's D values in any population for the control region analysis. In all analysed populations the nucleotide diversity (or π) and mean number of pairwise differences (or MNPD) were considerable higher for the mtDNA genome analysis than for the control region analysis because of its high number of loci under study (see Table 19).

We calculated a number of recurrent sites of synonymous and nonsynonymous polymorphisms in coding regions of mtDNA genome for each populations (Table 20). As expected, we obtained more synonymous sites (6,419) than nonsynonymous sites for over samples (2,528). Noir Marron populations present a substantial number of synonymous sites (1,055) and nonsynonymous sites (433). The ratio of nonsynonymous polymorphisms per nonsynonymous site to synonymous polymorphisms per synonymous site (*p*N/*p*S) was higher in Noir Marron populations (0.410) than in Afro-Colombian and Afro-Brazilian populations (0.381 in both populations). It is noteworthy to consider these polymorphisms in the phylogenetic analysis of mtDNA genomes, because of the mtDNA phylogeny shows a higher proportion of synonymous mutations in ancient than in young branches; this trend is known as 'purifying selection' (Elson et al., 2004). For instance, the youngest branches present a higher proportion of nonsynonymous mutations in protein-coding genes and substitutions in RNA genes; the purifying selection acts gradually over time on weakly deleterious characters. However, slightly deleterious mutations can persist for some time in the population (Kivisild et al., 2006; Soares et al., 2009).

Table 19: Sequence diversity indices estimated for each population based on whole genome DNA and D-loop region (1-576; 16024-16569).

	Population	Sample size	k	N° loci	HD	π	MNPD	Theta Hom	Theta k	Theta S	Theta Pi	Tajima's D test	P-value
MtDNA genome	**African American**												
	Noir Marron	43	31	16584	0.9823 +/-0.0093	0.0042 +/-0.0020	68.8549 +/-30.2646	53.58	48.36	98.00	68.85	-1.351	0.072
	Afro-Brazilian	39	35	16581	0.9933 +/-0.0080	0.0041 +/-0.0020	68.2861 +/-30.0794	145.25	116.61	116.61	68.29	-1.682	*0.018**
	Afro-Colombian	30	22	16573	0.9770 +/-0.0145	0.0038 +/-0.0019	63.5448 +/-28.2026	40.68	35.67	92.89	63.54	-1.463	0.050
	West African												
	Ahizi	22	18	16575	0.9697 +/-0.0278	0.0035 +/-0.0018	58.6494 +/-26.3386	30.24	43.88	64.47	58.65	-0.549	0.317
	Yacouba	16	14	16576	0.9833 +/-0.0278	0.0041 +/-0.0021	68.7500 +/-31.3044	57.13	49.89	91.62	68.75	-1.231	0.106
	Fon	36	33	16576	0.9952 +/-0.0078	0.0037 +/-0.0018	61.2032 +/-27.0431	207.04	186.67	106.11	61.2	-1.734	*0.009**
	Yoruba	32	27	16571	0.9879 +/-0.0115	0.0033 +/-0.0016	54.8750 +/-24.3542	79.76	78.78	87.90	54.88	-1.545	*0.025**
	Bariba	32	28	16571	0.9919 +/-0.0099	0.0037 +/-0.0018	60.7298 +/-26.9182	121.06	103.46	93.86	60.73	-1.436	*0.049**
	Bwa	23	12	16571	0.9249 +/-0.0324	0.0031 +/-0.0016	51.3399 +/-23.0555	10.87	9.41	60.96	51.34	-0.739	0.233
MtDNA D-loop	**African American**												
	Total Noir Marron	43	37	1126	0.9922 +/-0.0071	0.0152 +/-0.0076	17.0703 +/-7.7416	126.06	122.86	22.19	17.07	-0.898	0.191
	Afro-Brazilian	39	39	1128	0.9960 +/-0.0076	0.0152 +/-0.0077	17.1187 +/-7.7786	244.03	345.06	24.60	17.12	-1.179	0.107
	Afro-Colombian	30	30	1126	1.0000 +/-0.0086	0.0148 +/-0.0075	16.6707 +/-7.6342	-	-	22.72	16.67	-1.078	0.132
	West African												
	Ahizi	22	18	1126	0.9697 +/-0.0278	0.0130 +/-0.0068	14.6585 +/-6.8234	30.24	43.88	15.91	14.66	-0.386	0.392
	Yacouba	16	15	1126	0.9917 +/-0.0254	0.0154 +/-0.0081	17.3607 +/-8.1484	117.07	109.78	20.79	17.36	-0.784	0.229
	Fon	36	35	1125	0.9984 +/-0.0070	0.0140 +/-0.0071	15.7050 +/-7.1753	627.01	606.45	21.94	15.71	-1.106	0.124
	Yoruba	32	32	1125	1.0000 +/-0.0078	0.0118 +/-0.0060	13.2384 +/-6.1156	-	-	19.62	13.24	-1.267	0.089
	Bariba	32	32	1125	1.0000 +/-0.0078	0.0132 +/-0.0067	14.8151 +/-6.8069	-	-	20.36	14.82	-1.077	0.130
	Bwa	23	15	1125	0.9565 +/-0.0238	0.0121 +/-0.0063	13.5785 +/-6.3327	20.34	17.61	15.71	13.58	-0.608	0.299

Note: k = number of haplotypes; Number of loci with less than 5% missing data; **HD**= Haplotype diversity (+/-C.I.); π = Nucleotide diversity; **MNPD** = Mean number of pairwise differences; Theta **Hom**= θ value based on expected homocigosity; Theta k= θ value based on number of alleles; Theta **S**= θ value based on the number of segregating sites; Theta **Pi**; θ value based on the average number of pairwise differences; and **Tajima's D test** or Neutrality test (Tajima, 1996); *: significant *P*-values under a threshold of 0.05.

Table 20: Number of recurrent Synonymous (or *Syn*) and Non-synonymous sites (or *N-Syn*) polymorphisms in the African-American and West African populations.

Syn sites	N	ATP6	ATP8	COX1	COX2	COX3	CYTB	ND1	ND2	ND3	ND4L	ND4	ND5	ND6	Total
African Americans															
Noir Marron	43	32	17	156	63	83	94	102	74	26	23	168	177	40	**1,055**
Afro-Col.	30	11	12	93	34	59	55	54	54	13	13	121	106	26	**651**
Afro-Braz.	39	37	21	151	45	71	76	86	71	13	29	166	158	45	**969**
West Africans															
Fon	36	15	16	117	42	71	73	65	66	18	13	134	127	29	**786**
Yoruba	32	12	10	109	36	58	68	56	57	19	12	116	119	28	**700**
Bariba	32	16	15	104	41	62	70	60	69	12	15	122	108	36	**730**
Ahizi	22	16	15	90	25	36	50	48	36	10	10	81	90	27	**534**
Yacouba	16	11	8	57	15	28	30	32	33	5	11	61	65	19	**375**
Bwa	23	14	12	90	35	54	51	54	50	17	10	98	109	25	**619**
Total	**273**	**164**	**126**	**967**	**336**	**522**	**567**	**557**	**510**	**133**	**136**	**1,067**	**1,059**	**275**	**6,419**

N-Syn sites	N	ATP6	ATP8	COX1	COX2	COX3	CYTB	ND1	ND2	ND3	ND4L	ND4	ND5	ND6	Total
African Americans															
Noir Marron	43	91	6	27	9	4	125	10	14	51	0	6	78	12	**433**
Afro-Col.	30	67	3	10	3	4	76	13	16	26	0	6	19	5	**248**
Afro-Braz.	39	94	7	28	5	5	97	13	15	41	0	5	47	12	**369**
West Africans															
Fon	36	84	5	14	4	1	98	13	16	39	0	3	46	6	**329**
Yoruba	32	78	3	14	4	1	89	3	14	36	0	1	36	7	**286**
Bariba	32	80	3	15	1	2	85	10	15	34	0	3	47	6	**301**
Ahizi	22	51	5	19	4	5	50	12	4	28	0	1	24	7	**210**
Yacouba	16	37	2	14	3	1	43	8	8	21	0	1	22	5	**165**
Bwa	23	53	4	11	1	3	55	6	11	23	0	0	18	2	**187**
Total	**273**	**635**	**38**	**152**	**34**	**26**	**718**	**88**	**113**	**299**	**0**	**26**	**337**	**62**	**2,528**

Note: Ratio pN/pS: Noir Marron= 0.410, Afro-Colombian= 0.381, Afro-Brazilian= 0.381, Fon= 0.419, Yoruba= 0.409, Bariba= 0.412, Ahizi= 0.393, Yacouba= 0.440, and Bwa= 0.302.

All mtDNA haplogroups were phylogenetically identified following the latest classification (van Oven & Kayser, 2009) (see Figure S3). We estimated relative frequency distributions for each haplogroup in each population and each region (Figure 56). In Noir Marron, all mtDNA haplogroups belonged to the major African haplogroup L, while it was detected in 94.8% of Afro-Brazilians and 83.3% of Afro-Colombians (Table 21). Furthermore, in Noir Marron the highest percentages were observed for L2c (25.6%) and L2a1 (18.60%), in contrast with the Afro-Brazilian and Afro-Colombian populations that present the highest percentages for L2a1 (i.e., 26.67% and 20.51%, respectively).

In West African populations, L2a is highly present (Figure 56), with the highest percentages in the Bwa population from Mali (56.5%) and the Ahizi population from Ivory Coast (36.4%). The Fon population from Benin presented a wide representation of African diversity with fourteen subhaplogroups, including one subhaplogroup belonging to L4b1a, which is not present in other West African populations in the present study. However, the subhaplogroup L4b1a is also present in other West African population, such as in Burkina Faso (Barbieri et al., 2012). In Ivory Coast, the Yacouba population presented more subhaplogroups than the Ahizi population (12 versus 7). Interestingly, both West African populations, the Fon and Yacouba presented the subhaplogroup U6a3, which is mainly associated with North African (Maca-Meyer et al., 2003) and Mediterranean populations (Hernandez et al., 2014). This subhaplogroup might reflects the Maghreb expansion to West Africa, which occurred around 15 or 20 kya (Secher et al., 2014).

Table 21: Frequencies of major mtDNA haplogroups per population estimated with the whole mtDNA.

Population	African American			West African					
	F. Guiana	Colombia	Brazil	Benin			Ivory Coast		Mali
	N. Marron	Afro-Col.	Afro-Bra.	Fon	Bariba	Yoruba	Ahizi	Yacouba	Bwa
273 mtDNA	**43**	**30**	**39**	**36**	**32**	**32**	**22**	**16**	**23**
L0a	0.047	0.100	0.051	0.056	0.063	0.031		0.063	0.130
L0d			0.026						
L0f			0.026						
L1b	0.070	0.067	0.103	0.056	0.125	0.031	0.091	0.125	0.043
L1c	0.163		0.128	0.056	0.031	0.063	0.182	0.125	
L2a1	0.186	0.267	0.205	0.194	0.188	0.281	0.364	0.188	0.565
L2b	0.047	0.067	0.026	0.083	0.063	0.063	0.045		0.087
L2c	0.256	0.033		0.056	0.063	0.031		0.063	
L2d		0.033							
L3b	0.047			0.111	0.031	0.125		0.063	
L3d			0.077	0.084	0.188	0.125		0.063	0.174
L3e1	0.093	0.033	0.077	0.028				0.063	
L3e2	0.070	0.167	0.128	0.111	0.094	0.094	0.227	0.063	
L3e3			0.026	0.028	0.031	0.094		0.063	
L3e5					0.063		0.045		
L3f1b	0.023	0.033	0.077	0.083	0.063	0.063	0.045	0.063	
L3h1b		0.033							
L4b1a				0.028					
M2a1			0.026						
K1a1			0.026						
U6a3				0.028				0.063	
A2		0.033							
B2d		0.100							
C1d		0.033							
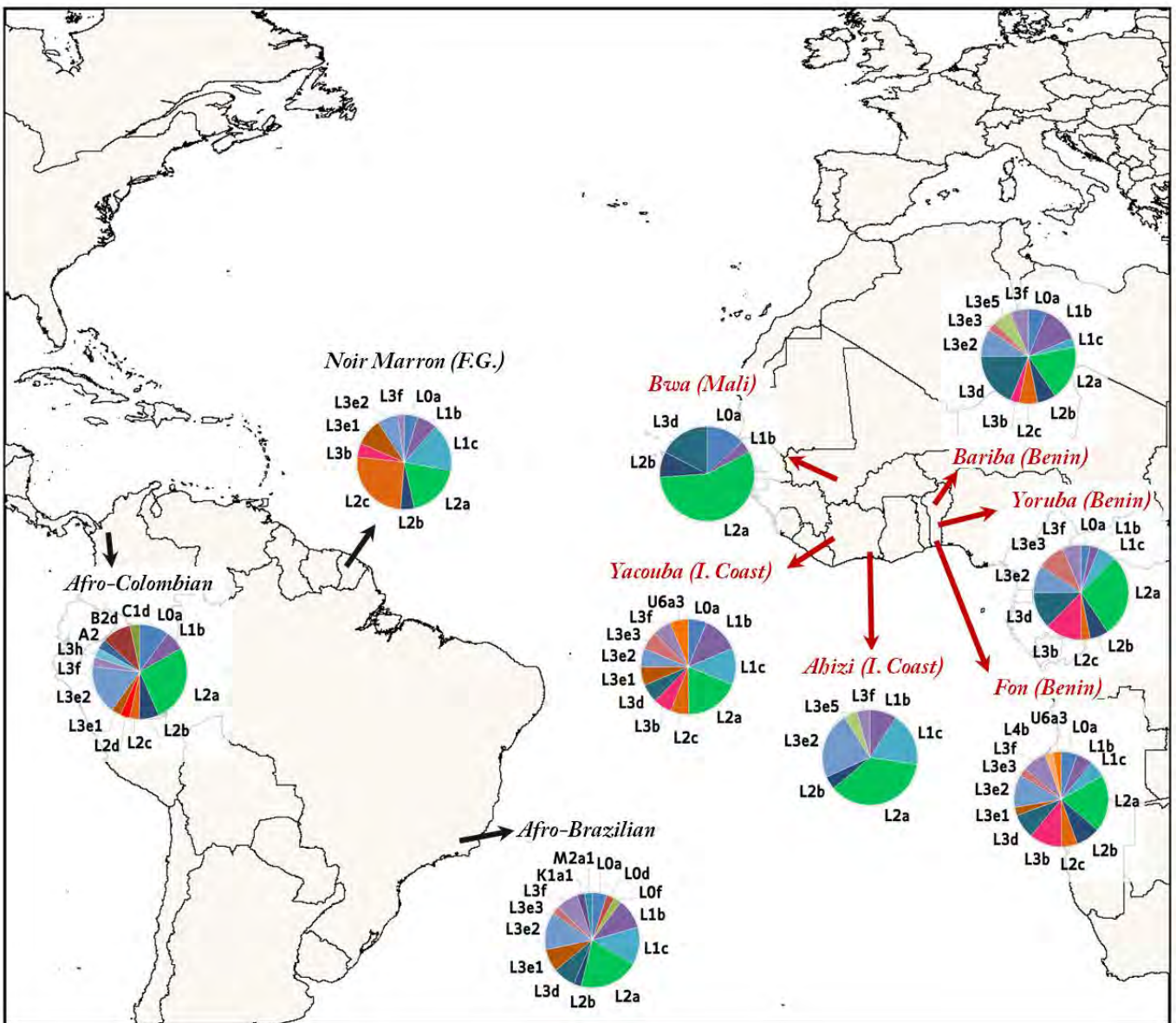
Figure 56: Map of frequency distributions of mtDNA haplogroups for the African-American and West African populations from the present study.

We estimated admixture proportions in African Americans based on mtDNA haplogroup frequencies (Table 22). In the Noir Marron communities, mY_{mtDNA} was remarkably high (97.8%) for West African ancestry, and almost null for West European and Native American ancestry. Afro-Brazilians and Afro-Colombians also presented high (>83%) West African ancestries. However, evidence of West European ancestry (7.8%) was detected in Afro-Brazilians, and evidence of Native American ancestry (16.3%) was detected in Afro-Colombians.

Table 22: Gene contribution estimates between populations based on variation in mtDNA haplogroup frequencies (Figure 56). Admixture proportions ($mY_{mtDNA} \pm$ S.D.) were estimated by using ADMIX v.2.0.

mtDNA genome	N	*West African* 161	*West European* 292	*Native American* 160
Noir Marron	43	0.9780 ±*0.0182*	0.0110 ±*0.0091*	0.0110 ±*0.0091*
Afro-Brazilian	30	0.9042 ±*0.0167*	0.0780 ±*0.0103*	0.0178 ±*0.0077*
Afro-Colombian	39	0.8373 ±*0.0169*	0.0002 ±*0.0062*	0.1625 ±*0.0134*

B.2.a.   *MtDNA diversity in the Noir Marron communities*

We estimated mtDNA haplogroup frequencies in all Noir Marron communities together based on new whole mtDNA genomes (N= 43) and for each Noir Marron community based on HVSI+HVSII segments (N= 183) by using both the new mtDNA data and the mtDNA data reported previously by Brucato et al. (2010) (Table 23). According to whole mtDNA genome data, the highest frequencies are for the three subhaplogroups: L2c (25.6%), L2a1 (18.6%), and L1c (16.3%). This pattern is noteworthy different from other African-American populations from South America; for instance, L2c is absent in Afro-Brazilian population (Figure 56).

According to HVSI+HVSII segments, the highest frequencies were detected for L2a1 in Aluku, Ndjuka, and Saramaka communities (28.0%, 30.9%, and 31.0%, respectively). The clade L2a is geographically widespread and highly frequent throughout Africa (Figure 58) and accounts for more than 70% of all L2 branches (Rosa & Brehem, 2011; Salas et al., 2002). The subclade L2a1 is the most complex subclade within L2a. This subclade harbours sublineages from all African regions, as well as sublineages from other continents, including non-African branches, such as L2a1l2a, connected to the Ashkenazi Jewish Diaspora (Costa et al., 2013) and the exclusively European L2a1k (Malyarchuk et al., 2008). Hence, we cannot established a unique African origin for L2a1 for those communities, however phylogenetic reconstruction of complete mtDNA genomes could give a strong indication of its African origin and rule out the European and Jewish origins.

In Paramaka, we detected the highest frequency for the subhaplogroup L2c (31.6%) that might suggest an African origin from Senegambia, because this subhaplogroup is rare in other African regions. The subhaplogroup L2c is also highly present in the Mandenka population from Senegal (37.1%) (Graven et al., 1995) and several populations from Guinea Bissau (range 15.0–23.0%) (Rosa et al., 2004).

Table 23: mtDNA haplogroup frequencies detected in Noir Marron. All these subhaplogroups present sub-Saharan African distribution.

mtDNA clade	mtDNA genome		HVSI + HVSII *			
	All Noir Marron		*Aluku*	*Ndjuka*	*Paramaka*	*Saramaka*
	N	43	25	97	19	42
L0a	2	*0.047*	*0.080*	*0.010*	*0.000*	*0.119*
L1b	3	*0.070*	*0.080*	*0.103*	*0.211*	*0.167*
L1c	7	*0.163*	*0.080*	*0.144*	*0.000*	*0.071*
L2a1	8	*0.186*	*0.280*	*0.309*	*0.211*	*0.310*
L2b	2	*0.047*	*0.000*	*0.072*	*0.000*	*0.048*
L2c	11	*0.256*	*0.160*	*0.113*	*0.316*	*0.048*
L3b	2	*0.047*	*0.000*	*0.041*	*0.000*	*0.024*
L3d	0	*0.000*	*0.000*	*0.010*	*0.000*	*0.095*
L3e*	0	*0.000*	*0.040*	*0.010*	*0.000*	*0.048*
L3e1	4	*0.093*	*0.160*	*0.062*	*0.053*	*0.024*
L3e2	3	*0.070*	*0.080*	*0.072*	*0.211*	*0.048*
L3f1	1	*0.023*	*0.040*	*0.052*	*0.000*	*0.000*
Total	**43**	***1.000***	***1.000***	***1.000***	***1.000***	***1.000***

Note: HVSI+HVSII frequencies were obtained after merged mtDNA data from present study with mtDNA data reported by Brucato et al. (2010).

B.2.b.   MtDNA diversity in the Afro-Brazilian population

In Afro-Brazilians were detected mtDNA haplogroups belonged mainly to the major African haplogroup L (94.9%). Besides low frequencies of Eurasian subhaplogroups (5.1% in total) were detected with the haplogroups K1a1a (found mainly in Europe and Central Asia) and M2a (found mainly in Europe and South Asia) (Table 21). As mentioned for Y chromosome haplogroups, the likely Amerindian contribution was not detected. The highest percentages were observed for clades L2a (20.5%) and L1c (12.8%), which are widely present in sub-Saharan African populations. Their subhaplogroups L2a1 (20.5%) and L1c1 (7.7%) were found mainly in West African populations (Rosa & Brehem, 2011; Salas et al., 2002). This West African genetic link was also supported by the frequencies of the subhaplogroups: L1b (10.3%), L3e2 (12.8%), L3d (7.7%), and L3f1 (7.7%) (see Table 21). The high proportion of African mtDNA lineages in the population of Rio de Janeiro is in accordance with studies related to the state of Rio de Janeiro (Bernardo et al., 2014) and other Brazilian states (Alves-Silva et al., 2000; Bortolini et al., 1997a; Bortolini et al., 1997b; Bortolini et al., 1999).

B.2.c. MtDNA diversity in the Afro-Colombian population

In Afro-Colombians, mtDNA haplogroups belonged mainly to the major African macrohaplogroup L (83.3%), mainly for clade L2 (40.0%) (Table 21). These values are in sharp contrast to lower frequencies detected previously in Afro-Colombian populations in the Chocó department of northern Colombia (53.0%) (Rojas et al., 2010), and in the Cauca department of South Colombia (72.6%) (Salas et al., 2008). Interestingly, we detected an important presence (16.6%) of Native American haplogroups: A2, B2d, and C1d, which are also present in substantial high frequencies in the Emberá-Chamí population from Antioquia (Xavier et al., 2015) suggesting gene flow. We did not detect any evidence of maternal European admixture, in contrast with Afro-Brazilian population.

B.3. Spatial distribution of the African mtDNA haplogroups

The patterns in the matrilineal variation within African ethnicgroups are remarkably complex (Rosa & Brehem, 2011; Salas et al., 2002). The increasing availability of complete mtDNA genome sequences has greatly refined the human mtDNA phylogenetic tree and provided new insights into the phylogeography of particular haplogroups. To shed new light about the African origins of mtDNA haplogroups detected in the African-American populations, we redefined the patterns of geographic distributions of mtDNA subhaplogroups L across the African continent. We elaborated a comprehensive mtDNA phylogenetic trees for the major clades L0, L1, L2, and L3 to obtain a better understanding of geographical association of major clades and their subclades in Africa and among African descendants. The mtDNA phylogenetic trees were based on 2,821 complete mtDNA genomes from haplogroup L0 to L3, which represent a large worldwide mtDNA diversity (Table 24). The mtDNA phylogeny drawn from the complete mtDNA genomes is remarkably more robust than using only variable segments of the control region; and new branches were identified in the tree.

In each tree, we included the new mtDNA genomes obtained for the African-American and West African populations. We estimated the ages of divergence for each node by using three different methods (*rho* statistic for complete mtDNA genome rate, *rho* statistic for synonymous mutations rate, and maximum likelihood). The phylogenetic trees also showed the mutation(s) that identified each clade or subclades (Figure S3). These calibration methods have recently also been applied successfully to mtDNA trees of other mammalian species (Soares et al., 2013) and to evaluate the timing of spread of the first humans in Africa (Fortes-Lima et al., 2014; Rito et al., 2013). They are in very good agreement with the recent recalibrations that use ancient mtDNA samples (Fu et al., 2013; Rieux et al., 2014).

These trees are extraordinarily large and are not included in this manuscript. They are available on request to the principal author (email address: *cesar@eurotast.eu*) and after the acceptance of the standard confidentiality agreement. Figure 57 shows a simplification of elaborated phylogenetic trees for each major clade L0, L1, L2, and L3; with ages calculated for each node in the tree (Table S9). Phylogenetically, the most ancient mtDNA branch, that is L0, seems to have a southern Africa distribution, and (probable) origin. L0 is divided between L0d (with southern African distribution) and L0a'b'f'k, where L0k is more frequent in southern African and L0a'b'f has mainly an eastern African distribution (Rito et al., 2013). The other major branch of the human mtDNA tree has a much more complex genealogy and distribution, with subclades distributed throughout central, eastern, and western Africa; and more recently to North Africa and the rest of the world. This L1'6 clade is also much more frequent overall than L0 throughout Africa, even in most of southern Africa where L0 is found at its highest frequencies. L1'6 then splits into L1, mainly found in West-Central Africa, and L2'3'4'5'6 (or L2'6) (Figure 57).

The latter then splits once again into L2'6 and further divides into L2 and L3 (or L3'4'6). Clade L2 is most likely originated from Central or West Africa and L3'6 from eastern Africa (Silva et al., 2015). Haplogroup L3 also includes major subclades that were most likely of eastern (L3a, L3c, L3h, L3i, and L3x), central (L3b), western (L3e2), and southern (L3e1) African origin. These geographical splits in the tree represented the most ancient dispersals taking place in the Middle Stone Age (or MSA) (Soares et al., 2012).

Table 24: Worldwide distribution of individuals belonging to the macrohaplogroup L (from L0 to L3). Columns show population size and frequencies, and histograms show frequencies for each row. Table also includes our African Americans from Latin America and West African populations. West African frequencies noticeably match with the African-American populations from the U.S. and Latin America.

World distribution	N	L0	L1	L2	L3	L0	L1	L2	L3	L0	L1	L2	L3
North Africa	106	8	16	31	51	0.011	0.034	0.038	0.066				
West Africa	829	29	123	390	287	0.038	0.263	0.474	0.370				
West-Central Africa	198	41	101	26	30	0.054	0.216	0.032	0.039				
Central Africa	89	13	2	6	68	0.017	0.004	0.007	0.088				
East Africa	145	44	8	28	65	0.058	0.017	0.034	0.084				
South Eastern Africa	218	65	59	38	56	0.086	0.126	0.046	0.072				
Southern Africa	533	491	10	20	12	0.649	0.021	0.024	0.015				
United States	258	22	51	127	58	0.029	0.109	0.155	0.075				
Latin America	148	11	24	68	45	0.015	0.051	0.083	0.058				
Europe	99	1	42	33	23	0.001	0.090	0.040	0.030				
Near East	80	12	4	22	42	0.016	0.009	0.027	0.054				
Others	118	19	27	33	39	0.025	0.058	0.040	0.050				
TOTAL	2,821	756	467	822	776	1.000	1.000	1.000	1.000				

Phylogenetic tree analyses suggested that the geographical origin of all mtDNA genomes cannot be easily traced back to a single geographical region in Africa, although some geographical inferences can be made. Among individuals belonging to the clade L0, one Afro-Brazilian (RJ_029) presented the haplogroup L0d1b2a1, which had only been found in Khoisan populations (Khoe, Tuu, and Kx'a) from Namibia and Botswana (Barbieri et al., 2013b); one Afro-Brazilian (RJ_080) belonged to the subhaplogroup L0a1b1a1, which matched with a specific subbranch of Khoisans from southern Africa (Behar et al., 2008). So, it might indicate connections with Khoisan groups.

Interestingly, in Noir Marron, one Aluku (GUY_026) belonged to the subhaplogroup L0a2a2a1, which matched with a specific subbranch of other African Americans from the U.S. (ASW), and populations from Zambia (Barbieri et al., 2013a). In addition, one Aluku (GUY_027) belonged to the subhaplogroup L0a2a2a, which matched with several groups from southeaster Africa (such as Mozambique). Three Afro-Colombians (CO_093, CO_086, and CO_011) belonged to the subhaplogroup L0a1a2, which is widely presented in West African populations, such as Mali, Burkina Faso (Barbieri et al., 2012), Benin, and Nigeria (ESN). It might reflect connections with Bight of Benin and Bight of Biafra.

In individuals belonging to clade L1, we detected more variations. Several subhaplogroups in African Americans (such as L1b1a4a, L1b1a7, and L1b2a) matched with terminal branches associated with West African populations from Benin, Ivory Coast, Burkina Faso (Barbieri et al., 2012), and Nigeria (YRI).

Specifically, one Aluku (GUY_032) belonging to L1c5 matched with one mtDNA detected in the Fon population from Benin (BE_124). In addition, two Alukus (GUY_069 and GUY_014) belonging to the subbranch L1c1a, which has been only found in Pygmy groups from Gabon and Cameroon. Complete mtDNA sequences for L1c1a depicted this clade to be autochthonous to Central Africa, which its most recent branches shared exclusively between Pygmies and some farmers (Batini et al., 2011b; Quintana-Murci et al., 2008). We found the same association with one Saramaka who belonged to L1c1b. It might reveal connections with Bight of Biafra. Interestingly, one Ndjuka (GUY_012) belonging to L1c2a3 matched with mtDNA genomes found in the Dama population from South Africa (Behar et al., 2008). One Ndjuka (NM_02601) belonging to L1c3b2 matched with populations from Burkina Faso (Barbieri et al., 2012).

Likewise, one Afro-Brazilian (RJ_083) belonging to L1c2a1 matched with southeaster African populations from Mozambique (Behar et al., 2008) and Kenya (Ingman et al., 2000). It might reflect connections of Afro-Brazilians with Southeast Africa.

For mtDNA genomes associated with the clade L2, we detected a high genetic variations and geographic distributions, particularly for the subbranches: L2b, L2c, and L2a1. The latter is

highly present in the African-American populations analysed in the present study as well as in populations from North America (ASW) and the Caribbean (such as Haiti and Barbados). These subbranches are highly frequent in West and West-Central African populations, particularly in populations with different linguistic and cultural backgrounds (Silva et al., 2015). For instance, one mtDNA genome belonging to the subhaplogroup L2a1f was found in twelve individuals of different populations: one Ndjuka (NM_07401), one Afro-Brazilian (RJ_135), five African American from the U.S. (Just et al., 2008), one Gambian (GWD), one individual from Burkina Faso (Barbieri et al., 2012), one Yoruba from Nigeria (YRI), one individual from Botswana (Barbieri et al., 2014a), and one individual from South Africa (Mishmar et al., 2003).

For mtDNA genomes associated with the clade L3, we also detected a high genetic variation and a wide geographical distribution in sub-Saharan populations. The subbranch L3e2 was the most present in African Americans from the Americas in general, as well as in West African populations. Interestingly, the subhaplogroup L3e1e, which is associated with south eastern African populations (Barbieri et al., 2013a), was found in one Afro-Brazilian (RJ_181), two Alukus (GUY_035 and GUY_063), two African Americans from Puerto Rico (PUR), and one from the U.S. (ASW).

We also summarised the current data of about African mtDNA haplogroups based on their frequencies estimated using only the HVSI+HVSII segments. We calculated frequencies for major clades and its subclades for each West African population analysed; we grouped them according to their country affiliations. Furthermore, we calculated mtDNA haplogroup frequencies for each African country by using the published mtDNA data of African populations across the continent. We also included Near Eastern countries. We analysed 8,343 people from 186 populations in 40 African countries, and 3,114 people in 10 Near Eastern countries (see Table S7).

Figure 58 shows the different distributions of major mtDNA haplogroups in the African continent and the neighbouring Near Eastern countries. Substantially, frequency distributions of major African mtDNA haplogroups L based on HVSI+HVSII segments agree with geographic patterns obtained with whole mtDNA genomes explained above (Figure 57). These mtDNA data were used to analyse geographical patterns of major clades across Africa in order to established genetic links with the African-American populations from South America. First, we elaborated contour maps across African populations and from Near Eastern populations (Figure 57) to represent the genetic landscape of major African mtDNA lineages. Second, by using the same approach, we elaborated the same contour maps including the African-American populations, to track back their African origin.
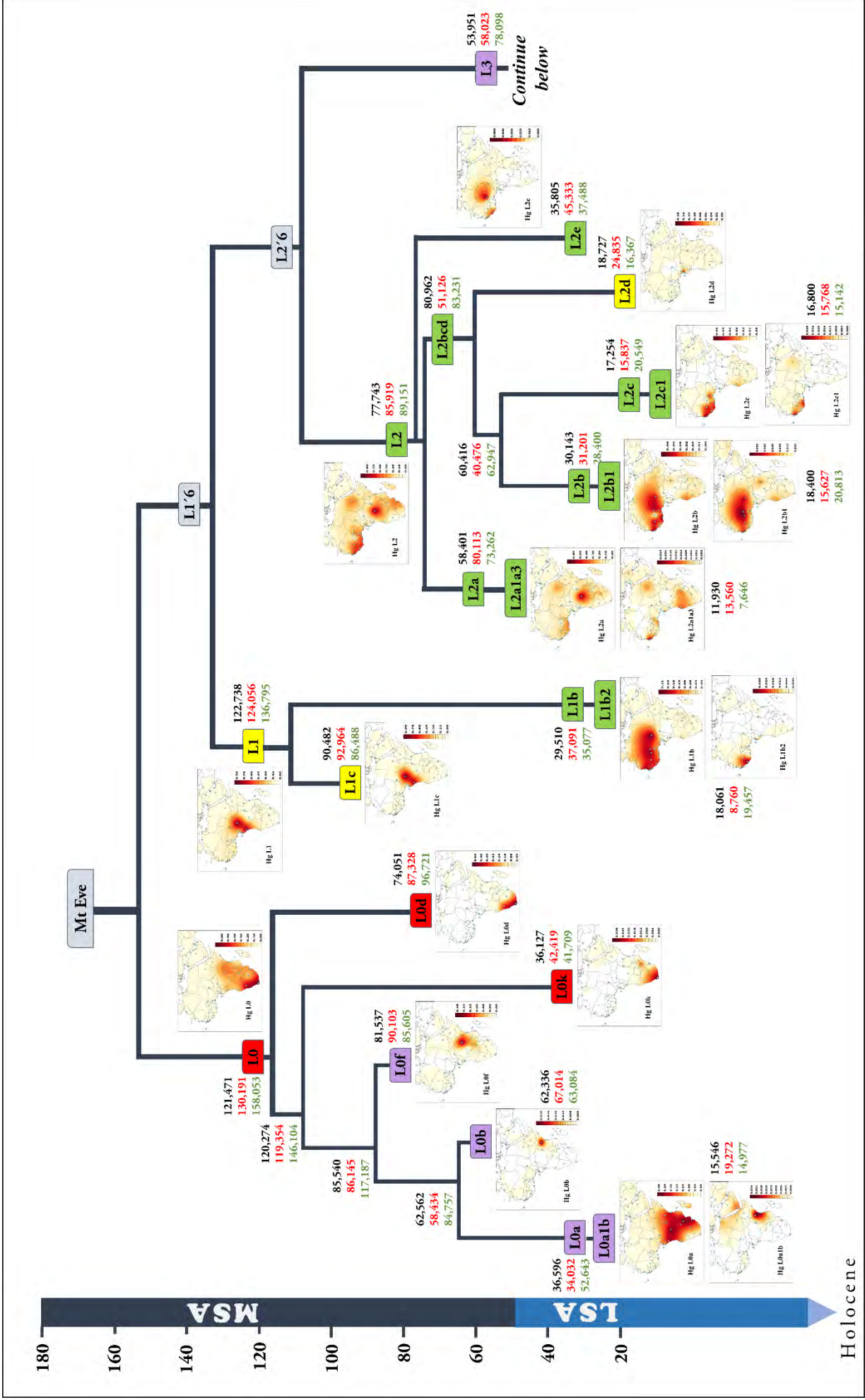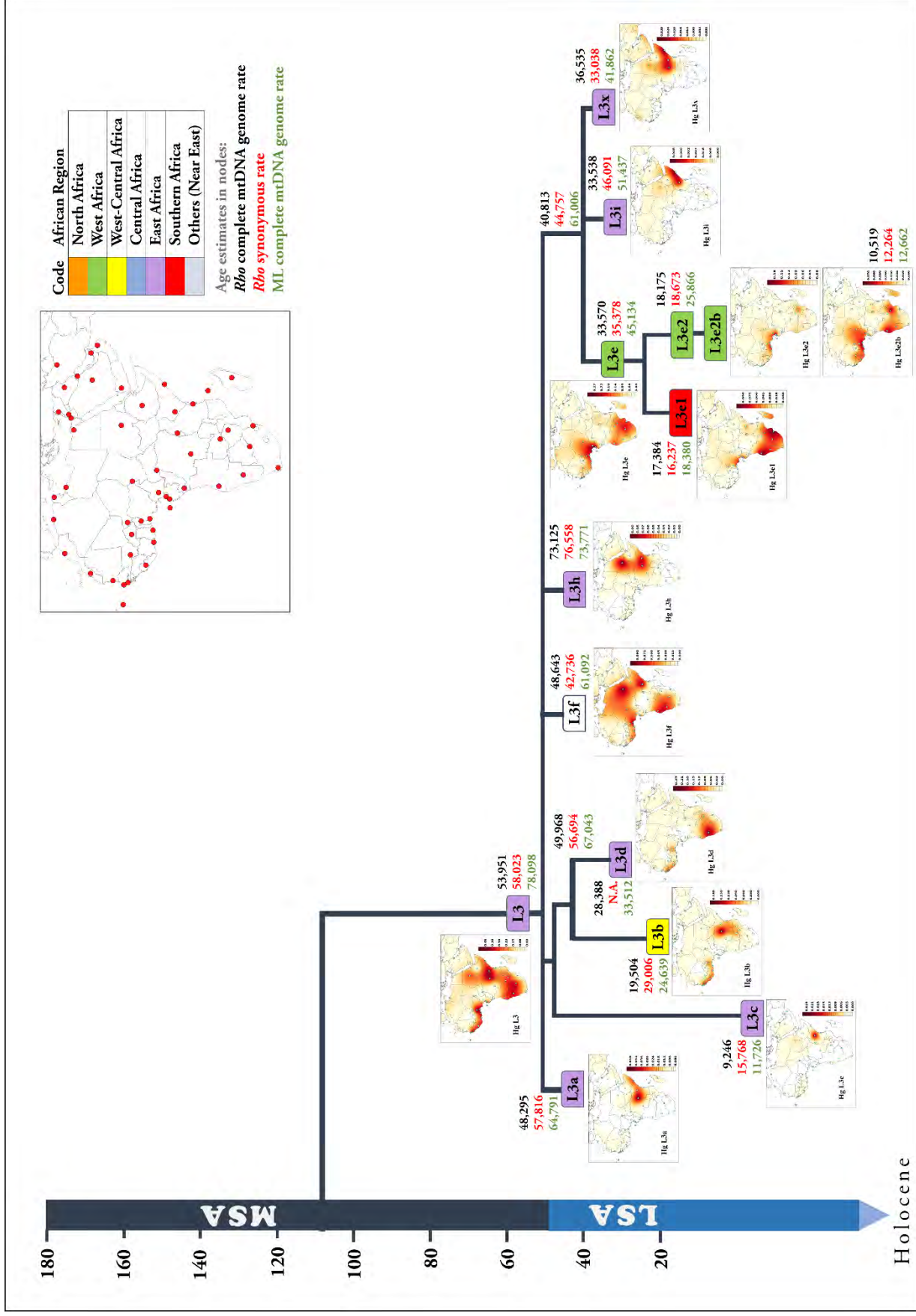
Figure 57: Geographic distributions and phylogenetic tree of major African mtDNA haplogroups L across the African continent. Ages estimates in each node (Table S9) were estimated using *rho* statistic of complete mtDNA genome rate (black), *rho* statistic of synonymous mutations rate (red), and maximum likelihood (ML) of complete mtDNA genome rate (green). Between African periods: Middle Stone Age (or MSA) and Later Stone Age (or LSA).
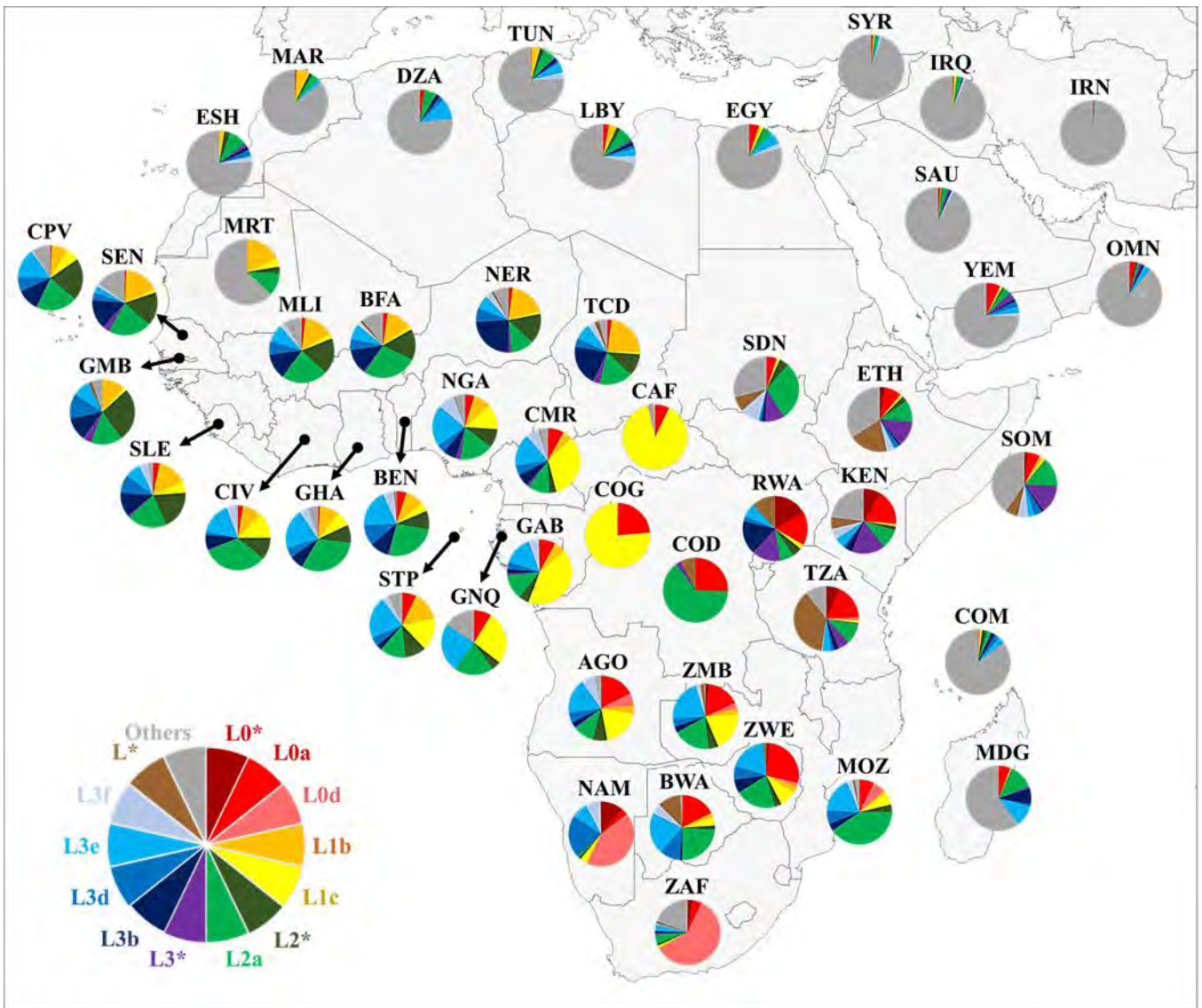
Figure 58: Frequency distributions of the major African mtDNA haplogroups L in sub-Saharan African and Near Eastern countries. Countries names according to ISO code (see Table S7).

Distribution of haplogroup L0 in Africa and South America

We extensively analysed the phylogeography of major African mtDNA macrohaplogroup L0 and its subhaplogroup L0a (Figure 59). We observed patterns of geographical variations among South African populations associated with L0 and eastern Africa mainly associated with L0a. The African L0a lineage depicted a wide distribution associated with eastern, central, and southern African regions, despite the eastern origins of this subhaplogroup (Rito et al., 2013). The distribution of L0a, similar to L2, was earlier linked to Bantu movements (Atkinson et al., 2009). However, recent evidence supports an earlier expansion of L0a to Central Africa during the Pleistocene/Holocene transition (Rito et al., 2013). We did not observed genetic links of lineage L0 with the African-American populations that we analysed, although some individuals

belonging to specific subhaplogroups were associated with specific African regions involved in the transatlantic slave trade.

Distribution of haplogroup L1 in Africa and South America

We analysed the phylogeography of major African mtDNA macrohaplogroup L1 and its subhaplogroups L1b, L1b2, and L1c (Figure 60). We observed that L1 is remarkably located in the West-Central African populations. In strong agreement with previous studies, the geographic distributions of the clade L1b and its subclade L1b2 are also common in West African populations, particularly along the coastal areas (Rosa & Brehem, 2011; Watson et al., 1997). They peak in the Senegal and Sierra Leone (Jackson et al., 2005) and among the Fulani people in Burkina-Faso, Chad, and South Cameroon (Cerny et al., 2011). Clade L1c is frequent in Central and West populations, representing over 70% of the maternal legacy of many Pygmy groups (Destro-Bisol et al., 2004; Quintana-Murci et al., 2008).

We analysed the spatial distributions of the clade L1b across the African-American populations, particularly in the Noir Marron communities. L1b2 is mainly frequent in the Upper Guinea coastal region and in Afro-Brazilian and Afro-Colombian communities. This might suggest a westward African connection during the slave trade.

Distribution of haplogroup L2 in Africa and South America

We widely analysed the phylogeography of major African mtDNA macrohaplogroup L2 and its subhaplogroups L2a, L2b, and L2c (Figure 61). Clade L2 originated in Western/Central Africa, but is nowadays spread across the entire continent. L2 movements were previously postulated to be related to the Bantu expansion (Pereira et al., 2001; Salas et al., 2002), but L2 expansions eastwards probably occurred much earlier. Recently, Silva et al. (2015) have proposed three moments of expansion associated with L2 from a Central African source: *i*) a migration at 70–50 kya into eastern or southern Africa; *ii*) postglacial movements at 15–10 kya into Eastern Africa; and *iii*) the southward Bantu expansion in the last five kya. We observed the consequences of those splits in the geographic pattern of this clade and its subclades. Additionally, L2 was also reported in North Africa probably because of trans-Saharan slave trade occurring in the last few centuries (Harich et al., 2010).

The subhaplogroup L2a spread into Central Africa. It is the most frequent and widespread mtDNA clade in sub-Saharan Africa (Figure 61), reaching over 40% among Tuareg from Niger/Nigeria and Mali (Salas et al., 2002; Veeramah et al., 2010; Watson et al., 1997), Fali from North Cameroon (Coia et al., 2005), Western Pygmies from Gabon (Quintana-Murci et al., 2008), and Bantu from Mozambique (Pereira et al., 2001). L2a is highly represented in Afro-

Brazilian (20.5%) and Afro-Colombian (26.7%) populations; however, it is difficult to identify the likely geographic origins of African-American descendants.
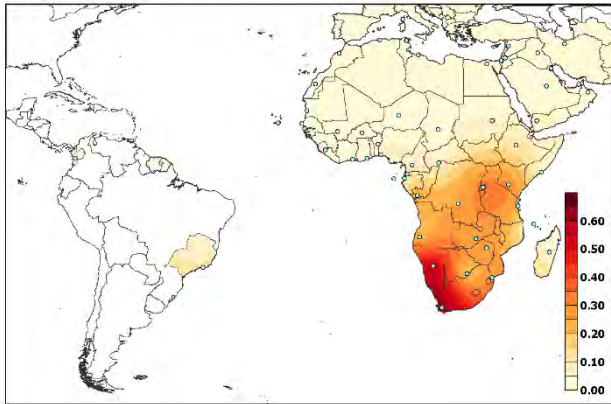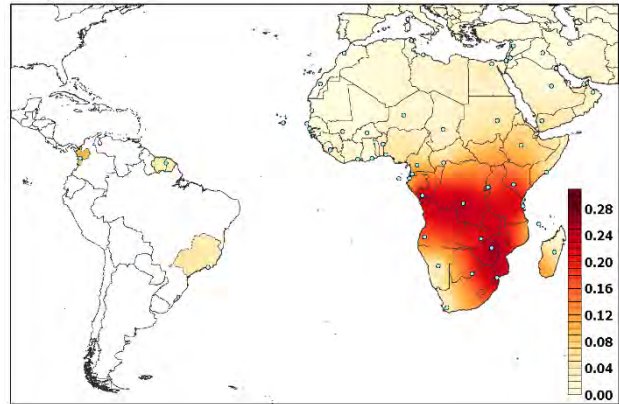
Figure 59: Spacial patterns of the African haplogroup L0 and its subhaplogroup L0a, within African, Near Eastern, and African-American populations.
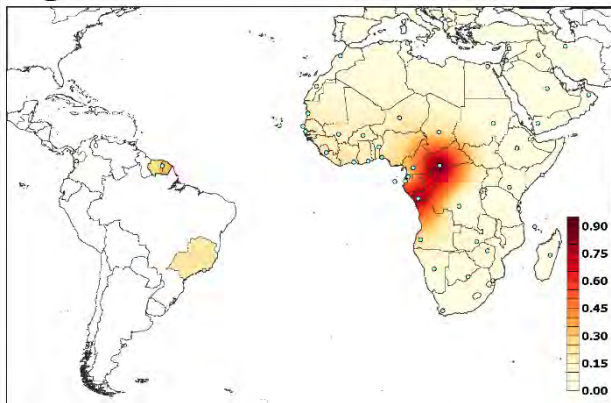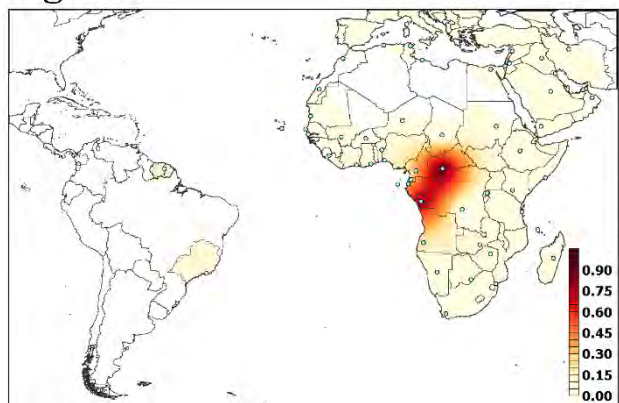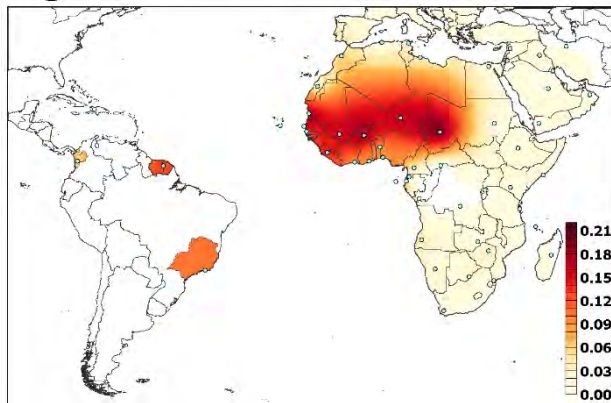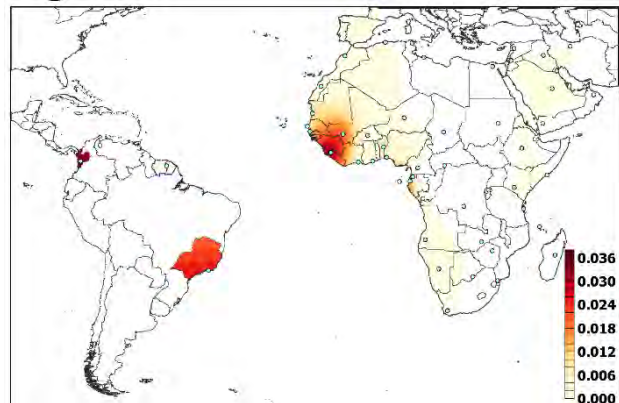
Figure 60: Spacial patterns of the African haplogroup L1 and its subhaplogroups: L1b, L1b2, and L1c, within African, Near Eastern, and African-American populations.

The subhaplogroups L2b and L2c spread into West Africa. However, both were present in low frequencies in the African-American and West African populations; with the exception of Noir Marron whose presented a high frequency (25.6%) of L2c, associated with the Paramaka community (31.6%). Therefore, the Paramaka community seems to be linked with the Upper Guinea region.
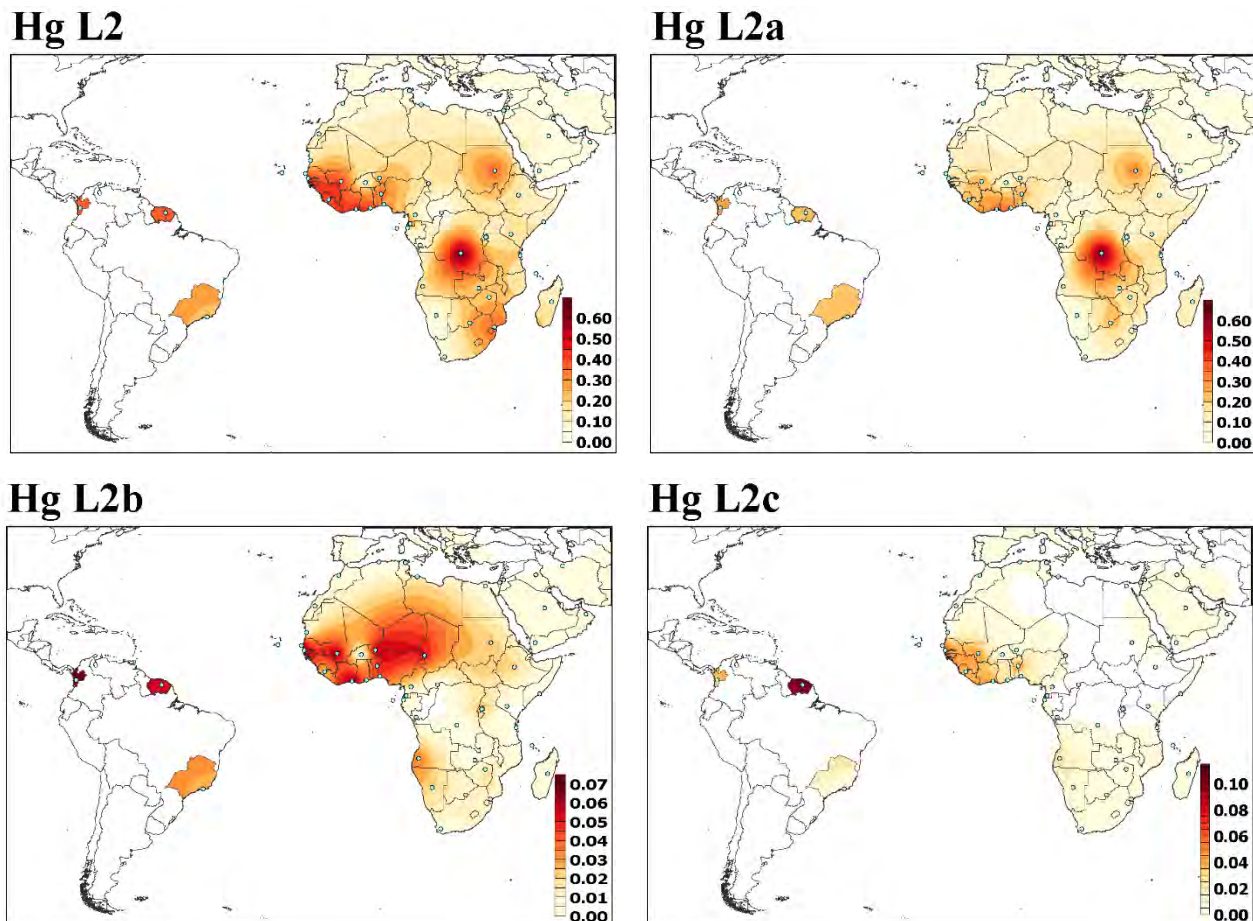
Figure 61: Spacial patterns of the African haplogroup L2 and its subhaplogroups: L2a, L2b, and L2c, within African, Near Eastern, and African-American populations.

Distribution of haplogroup L3 in Africa and South America

We broadly analysed the phylogeography of major African mtDNA macrohaplogroup L3 and its subhaplogroups L3e, L3e1, L3e2b, and L3f (Figure 62). Clade L3 is also highly frequent and widespread across Africa. The subclade L3f is mainly present in eastern populations, while L3e and L3e types are mainly concentrated in Central and South Africa. As was previously highlighted by Soares et al. (2012), eastern Africa was the source of most of the ancient L3 variation, although some subclades (L3b, L3d, and L3e) most probably emerged in Central Africa. In good agreement, we observed that a large set of African-American populations belonged to the haplogroup L3e and L3e types, which were primarily found in West and Central

Africa (Rosa et al., 2004; Salas et al., 2002; Soares et al., 2009; Watson et al., 1997). Therefore, the lineage L3e from different West and Central African sources contributed significantly to the genetic landscape of the African-American populations from South America.
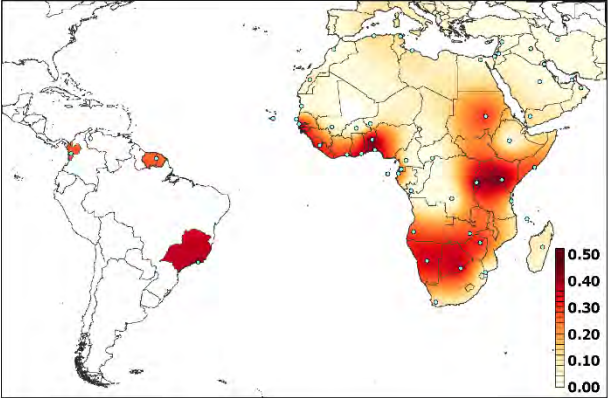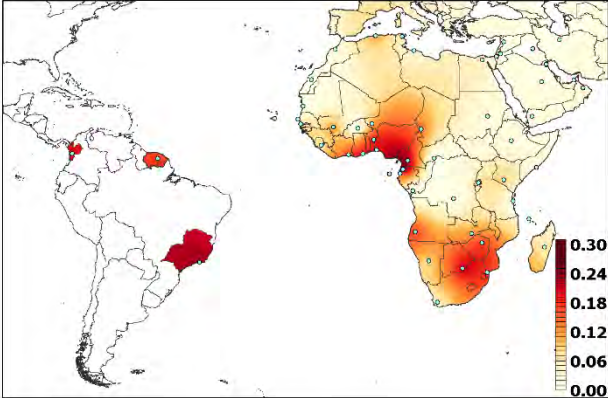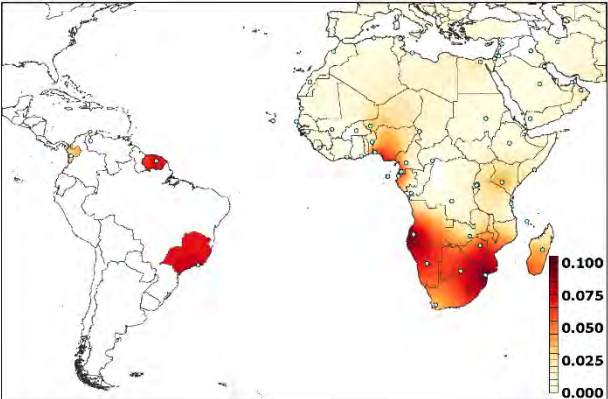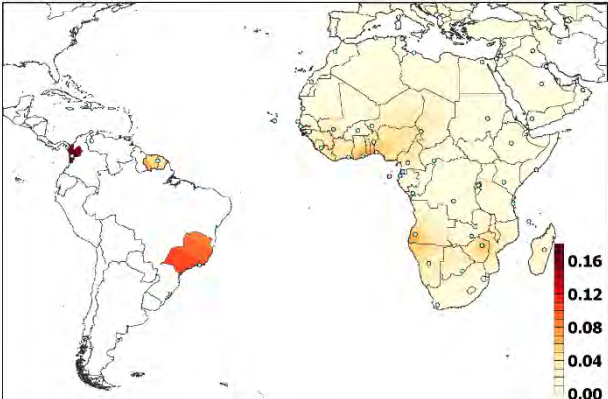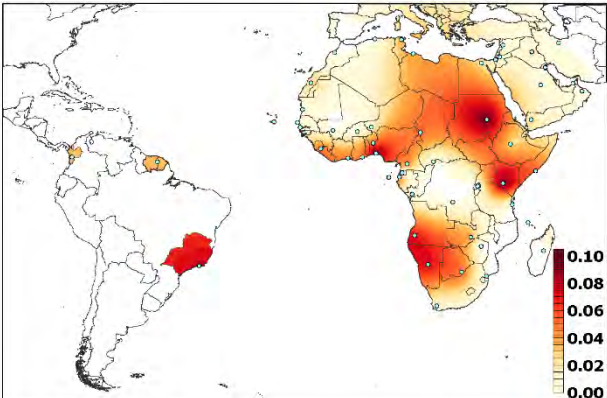
Figure 62: Spacial patterns of the African haplogroup L3 and its subhaplogroups: L3e, L3e1, L3e2b, and L3f, within African, Near Eastern, and African-American populations.

B.4. *Origin of African mtDNA diversity among the African Americans*

In the PCA analysis based on mtDNA haplogroup frequencies of the African-American populations and African populations, we obtained a genetic differentiation of populations associated with different African coastal regions involved in the transatlantic slave trade (Figure 63). The PC1 splits populations located in Senegambia, Sierra Leone, Windward Coast, Gold Coast, and Bight of Benin from populations located in West Central Africa and Southeast Africa. The PC2 mainly splits populations in Bight of Biafra from populations in other regions.

In African Americans, Ndjuka, Saramaka, and Afro-Colombian were close to populations from Bight of Benin and Windward Coast, while Paramaka was close to the Mende population in Sierra Leone (Jackson et al., 2005). Interestingly, Aluku and Afro-Brazilian populations were close to populations in Bight of Biafra and West Central Africa. Only Afro-Brazilians fall in the left upper part of the PCA plot slightly close with populations from Southeast Africa, similar association was detected in the PCA plot of Y haplogroup frequencies (Figure 53).

We observed different genetic links in other African-American populations from the New World. African Americans from Barbados were close to populations in Bight of Benin and Sierra Leone, while African Americans from the U.S. fall in the right upper part of the PCA plot far-off from populations from Senegambia, Bight of Benin, and Sierra Leone.
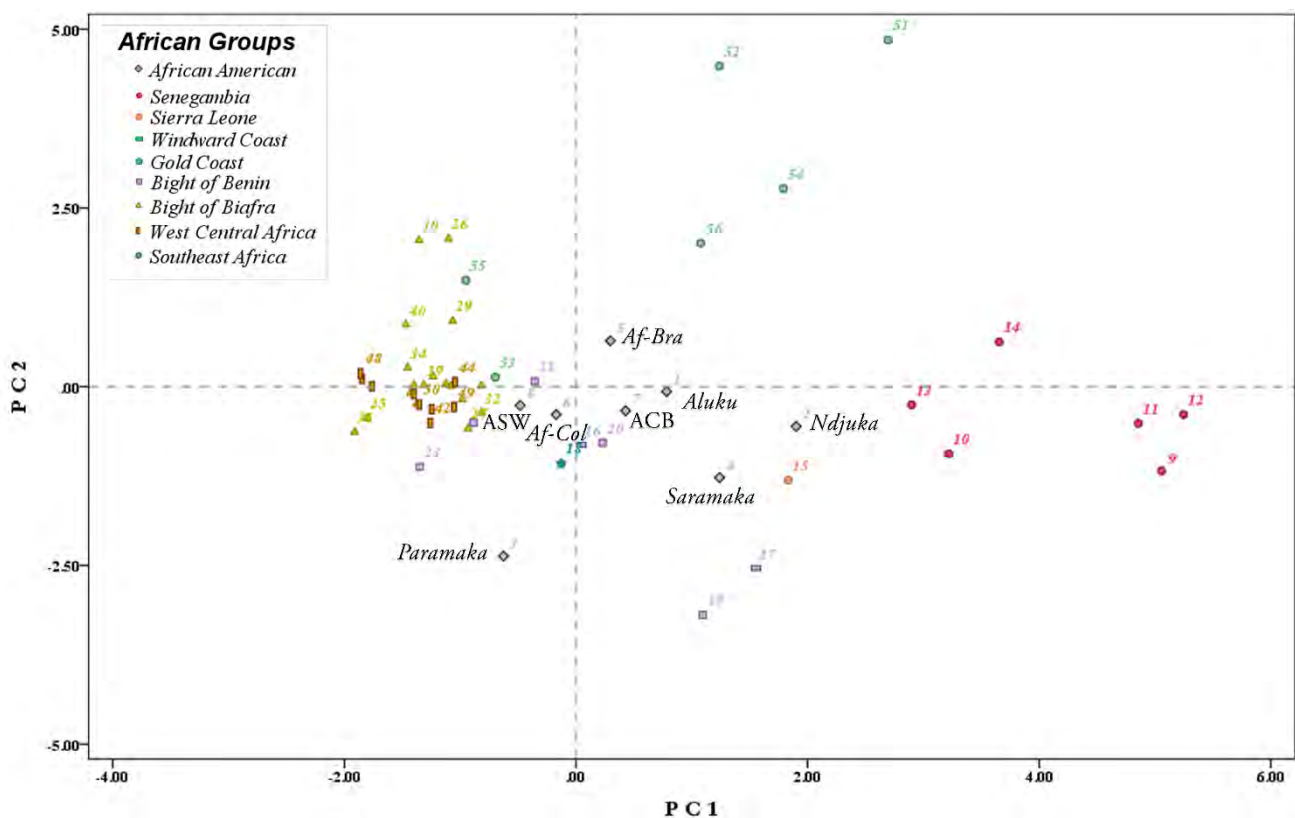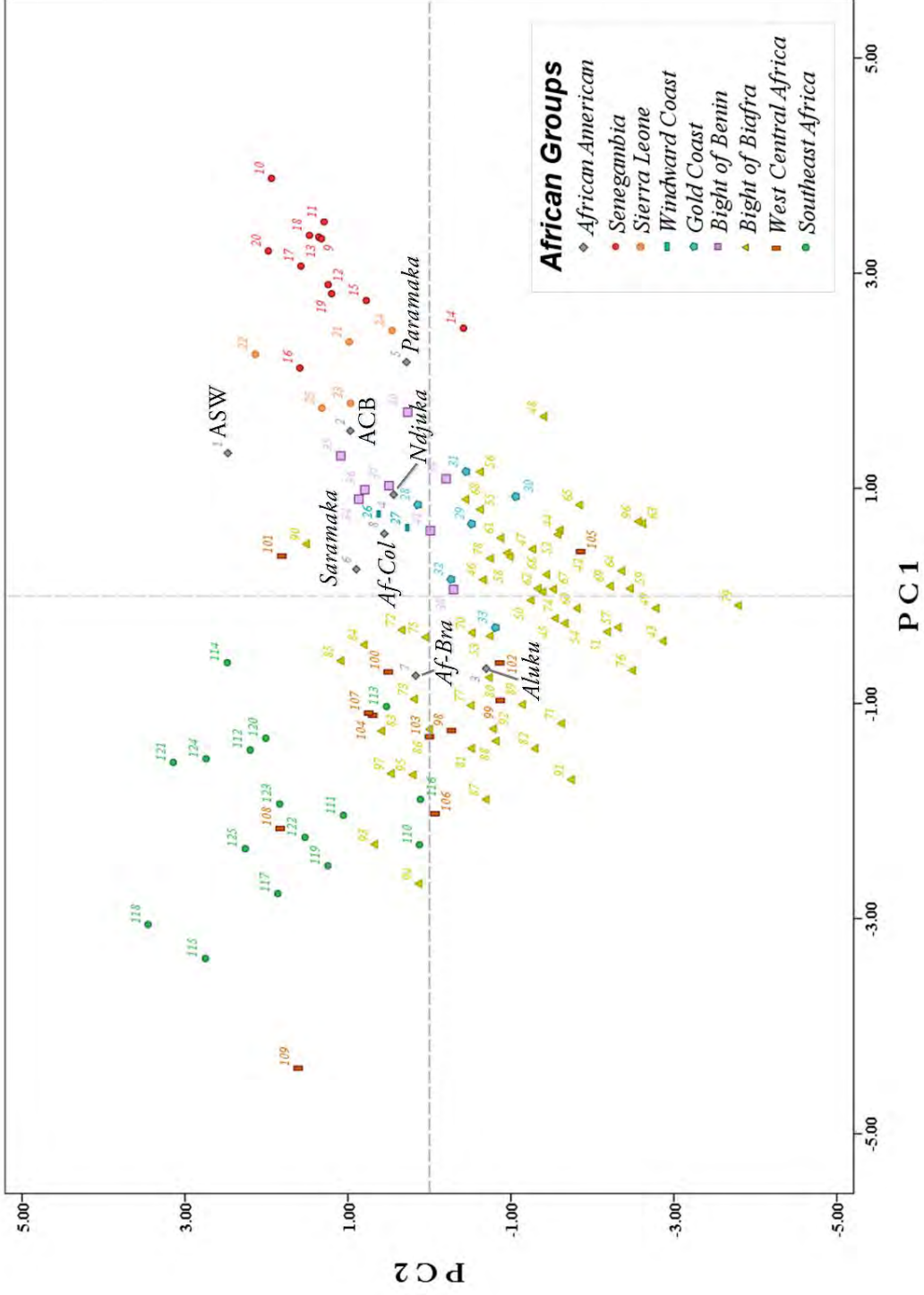
Figure 63: PCA analysis of the African Americans and African populations from different historical coastal regions, and based on mtDNA haplogroup frequencies of major African subhaplogroups. Populations are labelled according to Table S8. PC1 (14.28%) and PC2 (11.64%) constitute 25.92% of the total variance.

C. Genome-wide SNP data results

C.1. Genotyping Quality of genome-wide SNP data

The Illumina BeadArray technology used in this study provides a high quality of genome-wide amplification (Gunderson et al., 2005). We quantified the genotyping quality obtained by measuring the following parameters:

Total SNP call rates

We calculated fraction of SNP markers with sample call rate classified as: *i*) High (or ≥ 98%), was estimated 98.945% (4,255,964 / 4,301,332); *ii*) Low (or ≥ 90%), was estimated 99.981% (4,300,506 / 4,301,332); and iii) Very low (or < 90%:), was estimated 0.019% (826 / 4,301,332).

Individual SNP call rates

We estimated average call rate per individual with call rate: 99.695%. Only one sample presented SNP markers with very low SNP call rate (YA_036; 84.75%), three samples presented SNP markers with low SNP call rate (NM_6002; 91.24%, MA_420; 91.53%, GUY_090; 97.26%), and 225 samples presented SNP markers with high SNP call rate (from 98.88% to 99.94%).

Population SNP call rates

We estimated average SNP call rate of genotyped per population: 99.695% (see Table 25). We obtained a high SNP call rate for each population, which is a rewarding indicative of the remarkable quality of this important genotyping. Only the Yacouba population presented a SNP call rate in the edge of high threshold (98.997%), because of just one individual (YA_036) presented a very low SNP call rate (84.75%).

Reproducibility

We compared results for three individual genotyped twice. We calculated the followed parameters: *i*) Fraction of successful duplicate genotyping, which was estimated 99.991% (760 conflicts in 8,584,922 duplicate tests); *ii*) Average fraction of duplicate genotyping per marker, which was estimated 0.91%; *iii*) N° duplicate tested SNP markers without errors, which was estimated 4,298,028; and *iv*) N° duplicate tested SNP markers with errors was 717.

Identity test

Genotype data for each sample was pairwise compared, and one pairwise sample (NM_6002 vs YA_036) was identified similarity in genotype data (threshold ~80%). Both samples are from different populations, NM_6002 is a Noir Marron from French Guiana and YA_036 is from Yacouba population in Ivory Coast. However, both samples had a low SNP call rate (91.48% and 85.73%, respectively). It might indicate that the similarity between both genotype data

could be unreliable, and it is likely to have influenced for the low SNP call rate.

Table 25: Table of sample size analysed for the African-American and African populations by using Illumina platform. Table also shows the number of females (F) and males (M), and the average SNP call rate obtained per population (%).

Region	Population	N	F	M	SNP call rate
French Guiana	*Aluku*	23	18	5	*99.807*
	Ndjuka	23	16	7	*99.703*
	Saramaka	19	15	4	*99.408*
	Paramaka	6	6	0	*99.808*
Total Noir Marron populations		71	55	16	99.667
Colombia	Afro-Colombian	20	10	10	*99.876*
Brazil	Afro-Brazilian	16	4	12	*99.866*
Total African American		**107**	**69**	**38**	***99.803***
Benin	Fon	19	9	10	*99.813*
	Yoruba	24	12	12	*99.850*
	Bariba	24	12	12	*99.823*
Ivory Coast	Ahizi	20	5	15	*99.886*
	Yacouba	17	2	15	*98.997*
Mali	Bwa	18	8	10	*99.428*
Total West African		**122**	**48**	**74**	***99.633***
TOTAL		**229**	**117**	**112**	***99.695***

C.2. *Inbreeding coefficient and runs of homozygosity test*

A commonly used measure to quantify IBD in an individual is the inbreeding coefficient. This coefficient reflects not only inbreeding or consanguinity, but also other aspects of kinship in the population to which the individual belongs (Pemberton & Rosenberg, 2014). We examined the relationship between genomic estimates of the inbreeding coefficient and population patterns of genetic variations in African-American populations.

We observed the highest inbreeding coefficients in all Noir Marron communities (higher than 0.05%), with a much larger variation in the Saramaka population (Figure 64A). In other African Americans, the inbreeding coefficients were considerable lower (less than 0.03%). The ROH reflected also inbreeding patterns. The ROH analysis showed the highest values in the Noir Marron communities, with the exception of the Paramaka population whose variation is similar to other African-American populations (Figure 64B).

These results highlighted an elevated long-term genetic isolation in the Noir Marron communities, in comparison with other African-American descendants (Figure 64). It might suggest that the Noir Marron communities still preserve a high African genetic identity, which has been maintained after the abolition of slave trade and formation of these communities. In sharp contrast, other African-American populations have less inbreeding coefficients and ROH because of they have elevate gene flow with African and non-African groups.
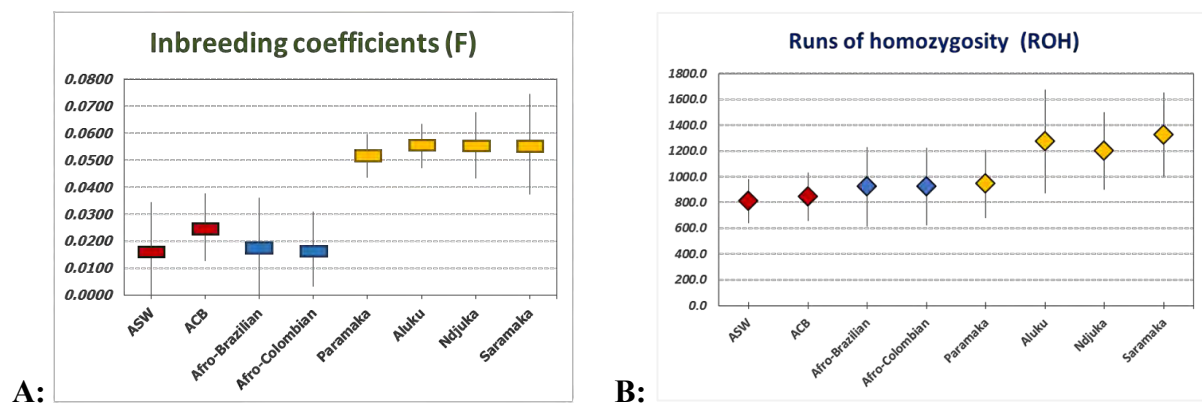
Figure 64: Inbreeding coefficients (**A**) and run of Homozygosity (**B**) estimated in the African-American populations in the Americas. Vertical lines indicate the range for each population.

C.3. *Admixture models by using the TreeMix analysis*

We analysed patterns of population splits and admixtures in the African-American populations by using TreeMix analyses (Pickrell & Pritchard, 2012). First, we elaborated a maximum likelihood tree based on 90K SNPs by using the African-American and West African populations as well as reference populations included in the HGDP panel (Cann et al., 2002). We observed one arrow from the Native American branch that evidenced admixture between these populations with Afro-Colombian (Figure 65A). Conversely, we observed another arrow from the European branch that evidenced admixture events took place with Afro-Brazilian. In this admixture model, the details of the demographic histories of the populations are absorbed into the branch lengths of the tree and the arrow's colour is indicative of intensive gene flow between different populations. Thus arrow's colour highlights stronger gene flow among Afro-Brazilians and European populations than between Afro-Colombians and Native Americans.

Second, we elaborated a maximum likelihood tree based on high-density SNPs by using the African-American and West African populations analysed, as well as reference populations included in the 1000 Genomes Project (1000 Genomes Project Consortium, 2015) and the Bantu-speakers from South Africa (May et al., 2013). Here, we extended this analysis to African

Americans from North America (ASW in the U.S.) and the Caribbean (ACB in Barbados). We detected three admixture events with different intensities between European populations and Afro-Brazilians, as well as with ASW and ACB (Figure 65B). Notably, there is only one arrow that indicated admixture between Afro-Colombians and Native Americans (PEL). In both TreeMix analyses, we did not detect any evidence of admixture events between the Noir Marron communities and European or Native American populations.

Finally, we elaborated a maximum likelihood tree based on low-density SNP dataset by using the African-American and West African populations as well as an extended representation of African populations across the continent, in order to establish the phylogenetic position of African Americans within the African tree. We identified the Noir Marron communities clustering together with West African populations (Figure 66). However, we could not identify a unique African source for each community in this large West African branch. We analysed the residual fit between each pair of populations, and the plot obtained indicated that West African populations presented the closest genetic affinities with Noir Marron populations (Figure S4). Other African-American populations were localised out of the West African branch. As expected, populations from Central Africa were not included within the West African branch. Furthermore, agriculturalist populations from Gabon and Cameroon were remarkably separated from their neighbours rainforest hunter-gatherers, as was previously reported by Patin et al. (2014).

To confirm those gene flow and admixture events observed between African Americans and continental groups, we ran three-population tests based on high-density SNP dataset. This approach examines patterns of allele frequency correlations across populations, which can provide a robust evidence of admixture events (Patterson et al., 2012). The $f3$-statistics and Z-scores did not show admixture evidence between Noir Marrons and populations included in the 1000 Genomes Project with the highest European (GRB, IBS, and TSI) and Native American (PEL) continental ancestries (Figure 67).

In sharp contrast, other African-American populations in the Americas present negative $f3$-statistics values indicative of admixture patterns with reference populations (see Table S11 and Table S12). The highest peaks signalling admixtures with European populations were found in Afro-Brazilian and ASW populations (mean $f3$-statistic= -0.0028 and -0.0026, respectively). Interestingly, the Afro-Colombian population had the highest peaks signalling an admixture with Native Americans (-0.0031). The lowest values for the Native American source was detected in ACB (-0.0009), which was basically close to zero. This emphasises the limited gene flow of the Native American population in African Americans from Barbados (Benn-Torres et al., 2008).
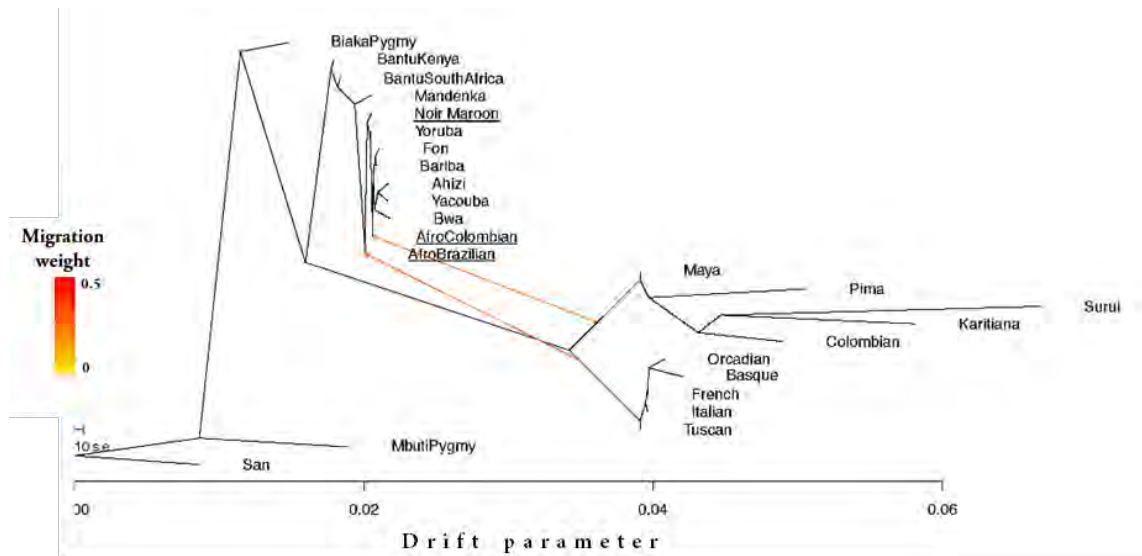
A:
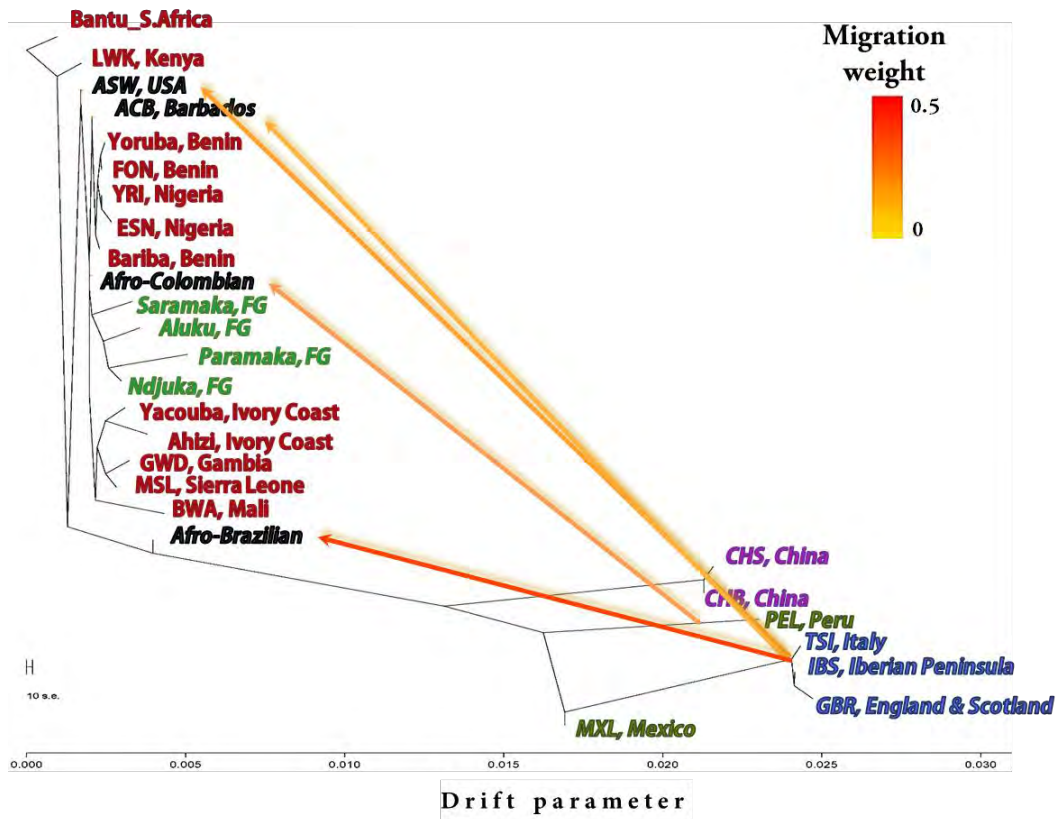
B:

Figure 65: Maximum likelihood trees for the African-American and West African populations using two reference panel to represent populations with different continental ancestry. **A:** Populations include in HGDP panel (Cann et al., 2002). **B:** Populations included in 1000 Genomes Project (1000 Genomes Project Consortium, 2015) and Bantu speakers from South Africa (May et al., 2013).
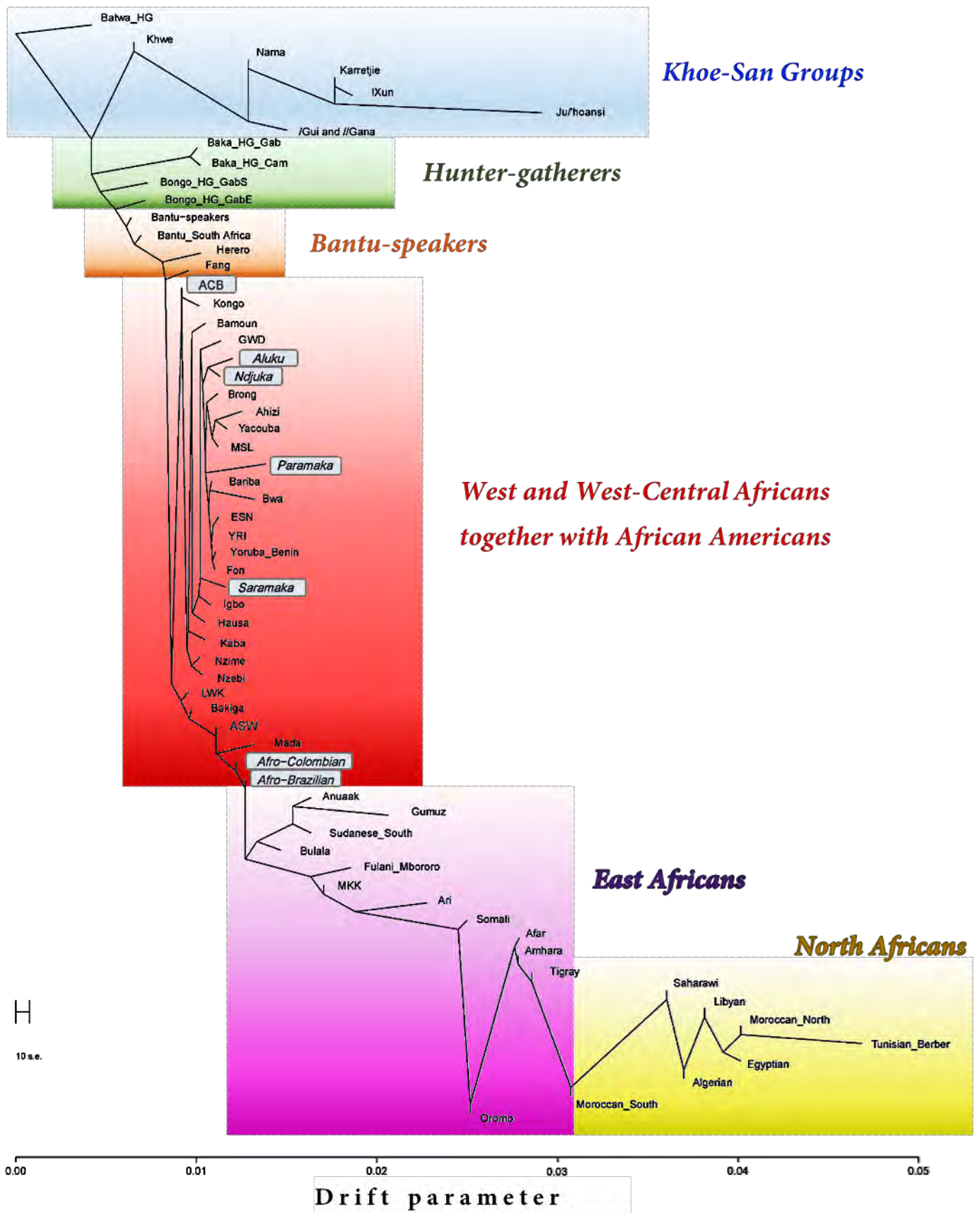
Figure 66: *TreeMix* analysis of the African-American and West African populations. The Noir Marron communities and other African-American populations are detected among West and West-Central African populations, which are far-off other African groups with different backgrounds (see Table13).
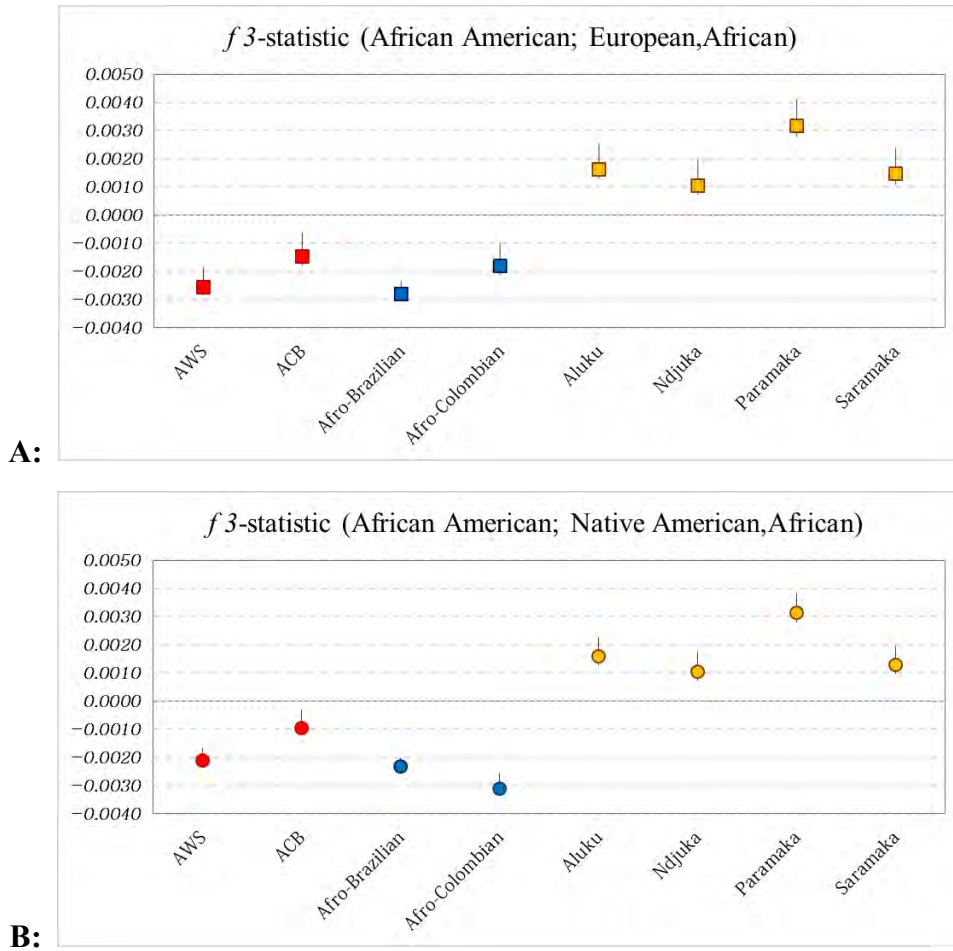
Figure 67: Average *f3*-statistic estimated in the African-American populations in the Americas. **A**: European and African populations used as a source (see Table S11). **B**: Native America and African populations were used as a source (see Table S12).

C.4. *Admixture models by using the ADMIXTURE analysis*

To obtain a finer characterization of the ancestry landscape of admixed African-American populations, we used the model-based approach to estimate global ancestry based on different datasets (50K, 90K, and 240K SNPs).

First, we estimated continental ancestral contributions to the current African-American populations by using the new genome-wide SNP data obtained for African Americans and West African populations, and reference populations included in the HGDP panel (Cann et al., 2002) for African, European, and Native American ancestries. Pygmy populations (Biaka and Mbuti) from Central Africa and the San population from South Africa, which were included in the HGDP panel (Figure 44), were not included in the ADMIXTURE analysis. They are hunter-

gatherer groups that frequently appeared as an ancient and external group in phylogenetic analysis (see Figure 65A and previous studies (Li et al., 2008; Pickrell & Pritchard, 2012)). Thus these small and isolated groups might not be a good representative of African populations involved in the slave trade, and they could altered the estimation of African ancestry.

We used the unsupervised algorithm ADMIXTURE, which assumes a specified number of hypothetical populations and provides a maximum likelihood estimate of allele frequencies for each population and admixture proportion for each individual. We performed ADMIXTURE analysis for the 90K SNPs dataset from K-groups= 2 to 10. According to the cross-validation (CV) test (Alexander & Lange, 2011), three is the number of underlying ancestral groups that best fit with the degree of differentiation between the populations under study (Figure 68).
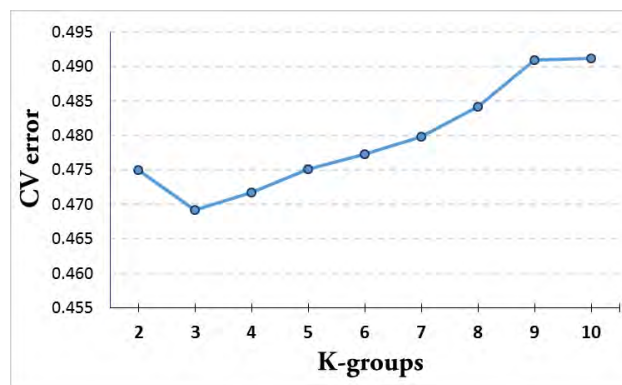
Figure 68: Cross-validation error (CV) versus K-groups from 2 to 10 identified accurately; K-groups= 3 as the number of underlying ancestral populations that best fit with the degree of differentiation between the populations under study because of presented the lowest CV error ($CV_{K=3}$: 0.4692).

As expected, ADMIXTURE plot at K= 2 showed close relatedness of African-American and African groups that clearly split up with non-African groups. It is also pointed out the presence of admixture of Afro-Brazilian and Afro-Colombian with non-African groups (Figure 69).

We compared the continental ancestry contributions for unsupervised ADMIXTURE analysis at K-groups= 3 (Figure 69). In the Noir Marron, the ADMIXTURE analysis indicated the main ancestral origin primarily came from African ancestry (98.88%), and there was a very low admixture proportions for European (0.82%) and Native American (0.30%) ancestries. In Afro-Brazilians, the ADMIXTURE analysis indicated a major component for African ancestry (72.4%), remarkable European proportions (21.6%), and low Native American ancestry (6.0%), suggesting that there is a major European admixture in the Afro-Brazilian population. Conversely, in Afro-Colombians showed a major component of African ancestry (77.2%), and similar proportions of European (10.3%) and Native American (12.5%) ancestries (Table S13).

K = 2, CV: 0.47495.
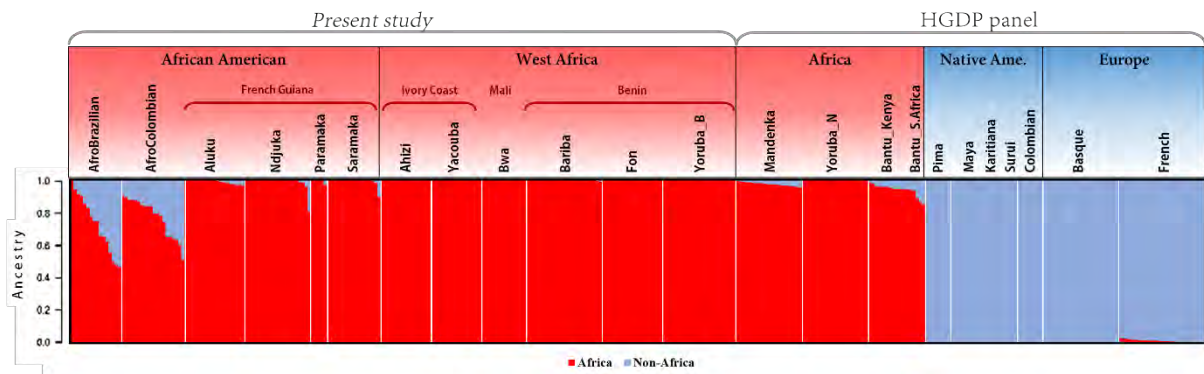
K = 3, CV: 0.46916.
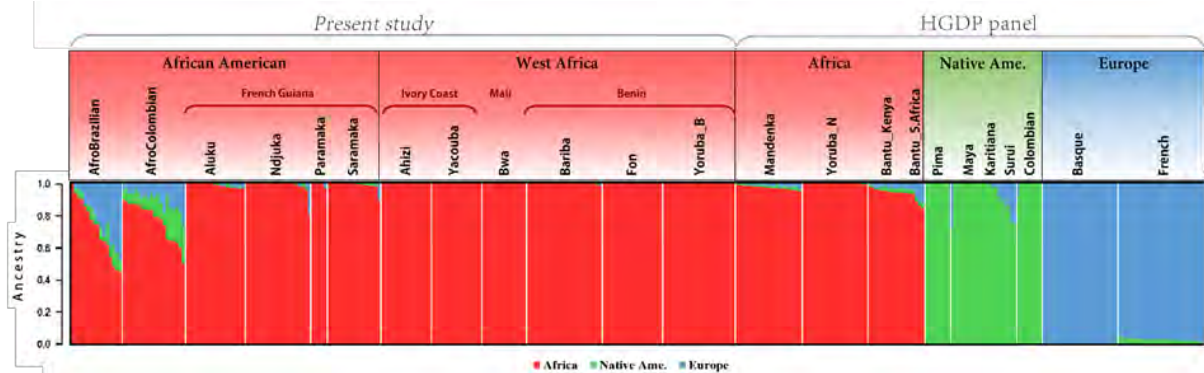
K = 4, CV: 0.47179.
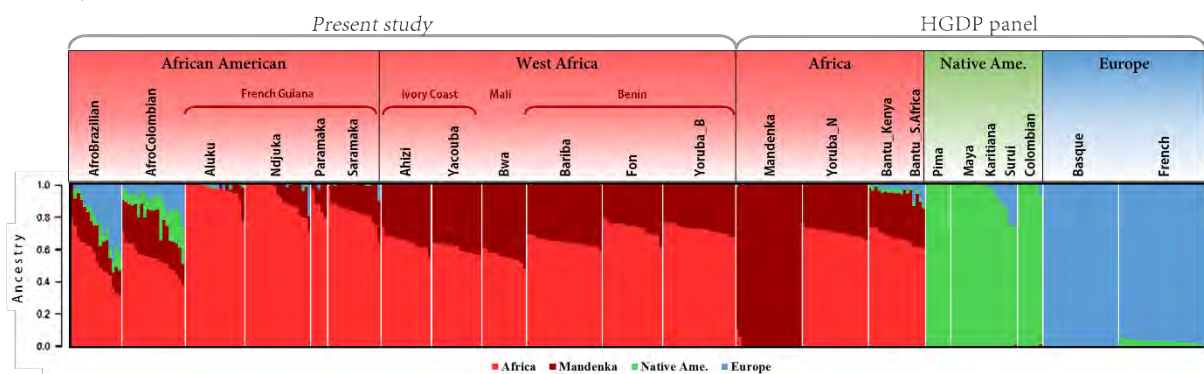
K = 5, CV: 0.47513.
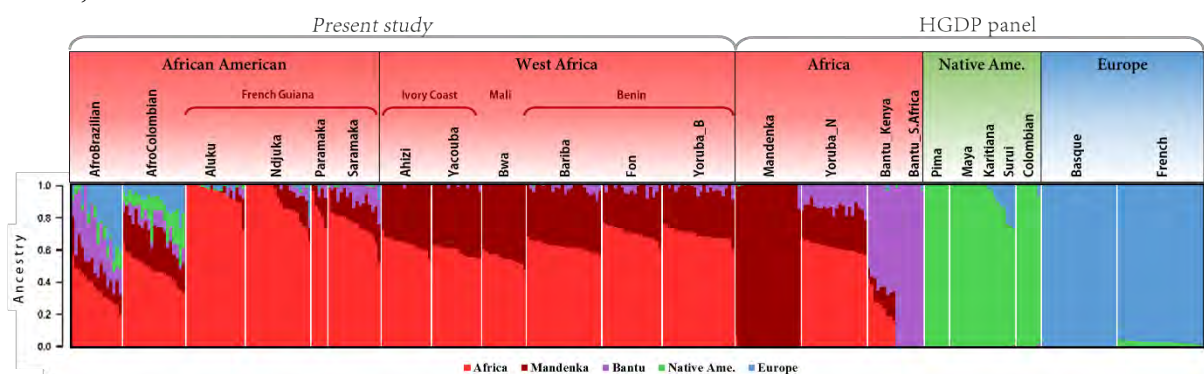
Figure 69: ADMIXTURE plots performed from K-groups= 2 to 5 for African American, West African, and continental reference populations included in the HGDP panel. Ancestry proportions for African (red), Mandenka (dark red), Bantu (purple), Native American (green), and European (blue) ancestries were indicated in each plot.

In the unsupervised ADMIXTURE analysis at K-groups= 4, we detected an ancestral group associated with the Mandenka population in the Senegal. At K-groups= 5, we detected an ancestral group associated with the Bantu-speakers populations from Kenya and South Africa. Both African ancestral groups are also related to African Americans (Figure 69). Interestingly, Mandenka contribution is slightly high in Afro-Colombians (21.9%), and Bantu contribution is slightly high in Afro-Brazilians (24.0%), suggesting different African sources for these populations. However, in the Noir Marron communities, we detected high proportions of West African ancestry (see Table S13).

Second, we ran ADMIXTURE analysis from K-groups= 3 to 7 using the new reference populations included in the last released of the 1000 Genomes Project Phase 3 (1000 Genomes Project Consortium, 2015). This unsupervised ADMIXTURE analysis based on a high-density SNP dataset (~240K SNPs) provides accurate continental ancestry estimations, and we can make comparison with other African-American populations. The CV test revealed that the CV error decreased with more K-groups, probably reflecting the high pairwise F_{ST} between populations (Figure 70A). ADMIXTURE plots confirmed previous findings and emphasized the highest African proportion detected in the Noir Marron (Figure 71). At K-groups= 3, the ADMIXTURE plot depicted a high African continental ancestry in the African-American populations. In this plot, Native American and East Asian populations appeared together because of these groups presented a much more recent common origin. This has been proved in several studies using linguistic data (Greenberg et al., 1986), craniofacial shape variations (Gonzalez-Jose et al., 2008), mtDNA data (Achilli et al., 2013), X chromosome (Bourgeois et al., 2009), and Y chromosome data (Bisso-Machado et al., 2011). In fact, recent genome-wide studies indicated that Native American groups descended from at least two streams of Asian gene flow (Raghavan et al., 2015; Reich et al., 2012; Skoglund et al., 2015).

At K-groups= 4, the ADMIXTURE plot indicated different continental ancestry between the African-American populations in the Americas, so thereafter we only focus on continental admixture proportions obtained in this analysis. In African-American groups, the Noir Marron communities presented the highest African ancestry (ranging from 98.2% in Aluku and Ndjuka to 99.1% in Paramaka), and subsequently the lowest European (range from 0.6 to 1.3%), Native American (range from 0.1 to 1.0%), and East Asian (range from 0.06 to 0.2%) ancestries (see Table 26). These results add further support to the high preservation of African heritage of these communities, which have been isolated from non-African populations for centuries.

In the Afro-Brazilian and Afro-Colombian populations, we detected high African ancestry (range from 72.5 to 76.6%), remarkable European (range from 10.4 to 22.3%) and Native American (range from 4.9 to 12.2%) ancestry, and very low East Asian (range from 0.2 to 0.8%) ancestry (see Table 26). We noticed that among the African-American populations analysed in the Americas: *i*) Noir Marrons account for the highest African proportions (98.3%), *ii*) Afro-Brazilians account for the highest European proportions (22.3%), *iii*) Afro-Colombians account for the highest Native American proportions (12.2%), and *vi*) ASW in the U.S. account for the highest East Asian proportions (1.1%). The ASW might indicate a recent gene flow between ASW and East Asian populations.

Table 26: Mean individual ancestry proportions and standard deviation of African American, African, European, Native American, and East Asian ancestries estimated using ADMIXTURE analysis at K-groups= 4 (Figure 71).

Group	Population	N	AFR	EUR	NAM	EAS
African American	*Aluku, French Guiana*	23	98.182 ± 1.574	1.172 ± 1.195	0.455 ± 0.582	0.191 ± 0.386
	Ndjuka, French Guiana	22	98.291 ± 4.368	1.314 ± 4.194	0.322 ± 0.454	0.072 ± 0.124
	Saramaka, French Guiana	18	98.318 ± 2.169	0.627 ± 2.058	0.996 ± 0.670	0.059 ± 0.156
	Paramaka, French Guiana	6	99.142 ± 1.614	0.667 ± 1.279	0.124 ± 0.196	0.066 ± 0.159
	Total Noir Marron	69	98.340 ± 2.850	1.030 ± 2.680	0.530 ± 0.610	0.110 ± 0.250
	Afro-Brazilian, Brazil	16	72.534 ±17.946	22.333 ±15.546	4.915 ± 3.590	0.217 ± 0.298
	Afro-Colombian, Colombia	20	76.603 ±11.366	10.382 ± 6.265	12.208 ± 6.811	0.807 ± 1.693
	ASW, USA	61	75.893 ±15.708	19.295 ± 9.107	3.664 ±10.723	1.148 ± 1.524
	ACB, Barbados	96	88.005 ± 7.798	11.076 ± 7.310	0.336 ± 0.423	0.583 ± 1.659
African	GWD, Gambia	113	97.664 ± 1.288	1.904 ± 1.292	0.172 ± 0.255	0.260 ± 0.348
	MSL, Sierra Leone	85	99.730 ± 0.349	0.079 ± 0.223	0.083 ± 0.168	0.109 ± 0.207
	Bwa, Mali	17	99.788 ± 0.321	0.129 ± 0.270	0.060 ± 0.167	0.023 ± 0.049
	Ahizi, Ivory Coast	20	99.974 ± 0.083	0.001 ± 0.000	0.019 ± 0.080	0.006 ± 0.024
	Yacouba, Ivory Coast	16	99.997 ± 0.000	0.001 ± 0.000	0.001 ± 0.000	0.001 ± 0.000
	Fon, Benin	19	99.947 ± 0.140	0.031 ± 0.130	0.021 ± 0.065	0.001 ± 0.000
	Bariba, Benin	24	99.636 ± 0.520	0.203 ± 0.485	0.097 ± 0.181	0.064 ± 0.195
	Yoruba, Benin	24	99.938 ± 0.159	0.001 ± 0.000	0.005 ± 0.014	0.056 ± 0.160
	YRI, Nigeria	108	99.873 ± 0.260	0.031 ± 0.189	0.056 ± 0.134	0.040 ± 0.133
	ESN, Nigeria	99	99.907 ± 0.202	0.008 ± 0.069	0.053 ± 0.129	0.032 ± 0.117
	LWK, Kenya	99	94.517 ± 1.013	4.134 ± 0.979	0.091 ± 0.168	1.259 ± 0.587
	Bantu, South Africa	94	97.386 ± 2.923	1.792 ± 2.283	0.137 ± 0.245	0.684 ± 0.848
European	TSI, Italy	107	0.220 ± 0.296	99.091 ± 0.485	0.035 ± 0.108	0.654 ± 0.462
	IBS, Iberian Peninsula	107	1.372 ± 1.192	98.333 ± 1.255	0.091 ± 0.254	0.204 ± 0.314
	GBR, England and Scotland	91	0.001 ± 0.000	99.573 ± 0.400	0.408 ± 0.396	0.018 ± 0.071
	FIN, Finland	99	0.002 ± 0.007	91.960 ± 1.356	2.688 ± 0.618	5.349 ± 1.100
	CEU, Utah, USA	99	0.017 ± 0.113	99.434 ± 0.566	0.416 ± 0.425	0.132 ± 0.325
American	PUR, Puerto Rico	104	14.430 ± 9.813	70.809 ±10.255	14.355 ± 4.004	0.406 ± 0.484
	MXL, Mexico	64	3.960 ± 2.423	43.538 ±19.982	49.672 ±20.297	2.830 ± 1.636
	CLM, Colombia	94	8.241 ± 8.072	63.246 ±13.920	27.926 ±10.273	0.587 ± 0.606
	PEL, Peru	85	2.338 ± 5.989	15.430 ±13.009	81.242 ±16.126	0.990 ± 4.439
East Asian	CHB, China	103	0.001 ± 0.000	0.210 ± 0.574	0.610 ± 0.727	99.180 ± 1.091
	CHS, China	105	0.001 ± 0.000	0.001 ± 0.004	0.022 ± 0.153	99.976 ± 0.153

The ADMIXTURE plot indicated difference continental ancestry between Native American and East Asian populations. Remarkably, Native American populations in the 1000 Genomes Project (1000 Genomes Project Consortium, 2015) present different degrees of continental ancestry; this is indicative of diverse patterns of gene flow (see Table S13). For instance, Puerto Rican and Colombian populations included in this reference panel have a high European ancestry (74.18% and 69.47%, respectively) with a noteworthy admixture with Native American populations in the past, because of the homogeneous Amerindian proportions in these populations (11.61% and 22.63%, respectively). The highest African proportion in Native Americans was detected in Puerto Ricans (14.43%) and the lowest in Peruvians (2.34%). We obtained similar ancestry values in the Puerto Rican population (N= 104) as Via et al. (2011) obtained with the analysis of 93 AIMs in a large Puerto Rican sample size (N= 642). They reported 63.7% European, 15.2% Native American, and 21.2% African ancestry. Peruvian and Mexican populations presented the highest Native American ancestry (81.24% and 49.67%, respectively). Therefore, they were included as Native American reference populations in the ADMIXTURE analysis by using a low-density SNP dataset (~50K SNPs).

Among European populations, we detected a low East Asian (5.35%) and Native American (2.69%) ancestries in Finns (FIN population). Both ancestral components are not present in other analysed European populations, and they are relatively homogeneous in that population (N= 99). It might indicate an ancient gene flow between Siberian populations with East Asian populations, and ancestral populations from Northeast Europe (Allentoft et al., 2015).

In the ADMIXTURE analysis at K-groups= 5, we detected two within-Africa ancestry clusters associated with West African and Bantu populations; this seemed to gradually decrease across West African populations. At K-groups= 6, we detected two within-European ancestry clusters associated to South and North European populations, which might indicate a geography pattern in European populations (Novembre et al., 2008).

Finally, we extended the ADMIXTURE analysis to a large representation of African populations by using a low-density SNP dataset (~50K SNPs) in order to associate admixture patterns with geographic ancestry and to identify the African origin of African-American populations. We used the CLUMPAK (or Cluster Markov Packager Across K) software (Kopelman et al., 2015) to select the preferred value of K according to the methods of Pritchard et al. (2000). We identified K-groups= 14 as the best number of clusters (or ancestral groups) to describe the membership coefficients of each individual (Figure 70B).
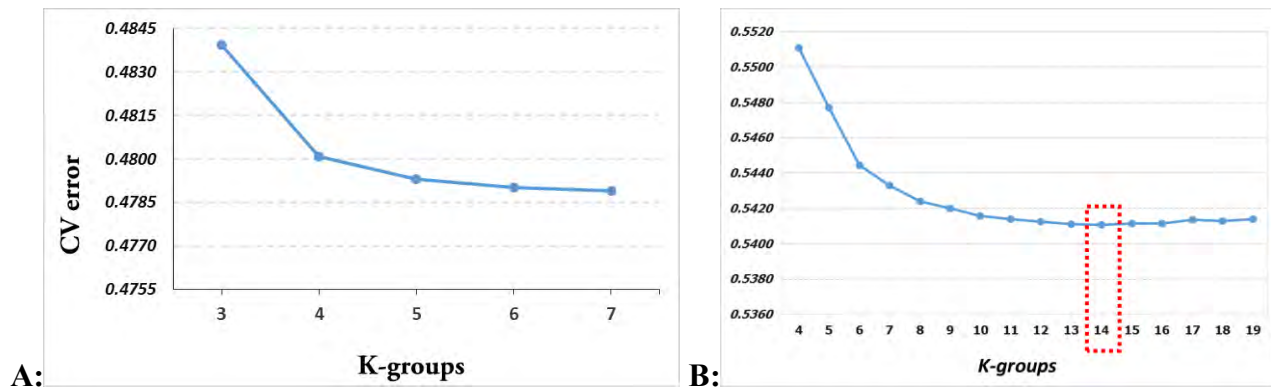
Figure 70: Cross-validation test for the ADMIXTURE analysis for **A:** high-density SNP dataset, and **B:** low-density SNP dataset. In the latter, K-groups= 14 is the best number of clusters to describe the ancestral source of genetic diversity detected.

We elaborated one ADMIXTURE plot for K-group= 14 using the membership frequencies of each individual based on low-density dataset (Figure 72A), and we gathered them according to their geographic locations. We identified the geographic origins of each populations that were associated with fourteen ancestral groups: North Africa, West Africa, West-Central Africa, Pygmies (except Batwa), Batwa from Uganda, Khoe-San from South Africa, Bantu, Masaai from Kenya, East Africa, Ari from Ethiopia, Middle Eastern from Qatar, European, East Asian from China, and Native American.

We calculated pairwise F_{ST} between estimated ancestral groups based on Wright's F-statistics method (Wright, 1965) (Table 27). In African groups, we observed substantial variation in pairwise F_{ST} among ancestral groups from different African regions, suggesting high genetic intra-continental heterogeneity. Nonetheless, notable genetic similarity (0.019) was detected between West African and West-Central African ancestry groups.

We calculated the average membership frequencies of each population based on individual frequencies (Figure 72B). In the African-American populations, we identified the same admixture patterns described above, based on the high-density SNP dataset. Moreover, we detected a high Bantu component in Afro-Brazilians, which can be associated with West-Central or South African populations. In Noir Marron communities, we detected the highest West-Central component, which matches with populations from Benin.

In West African populations, we identified a noteworthy genetic gradient for those populations that strongly matched with their geographic distribution of populations from Gambia to Nigeria (Figure 72B). This genetic gradient was associated with a progressive decrease of the West African component. Likewise, this gradient was related to a progressive increase of the West-Central African component.

K = 3, CV: 0.4839.
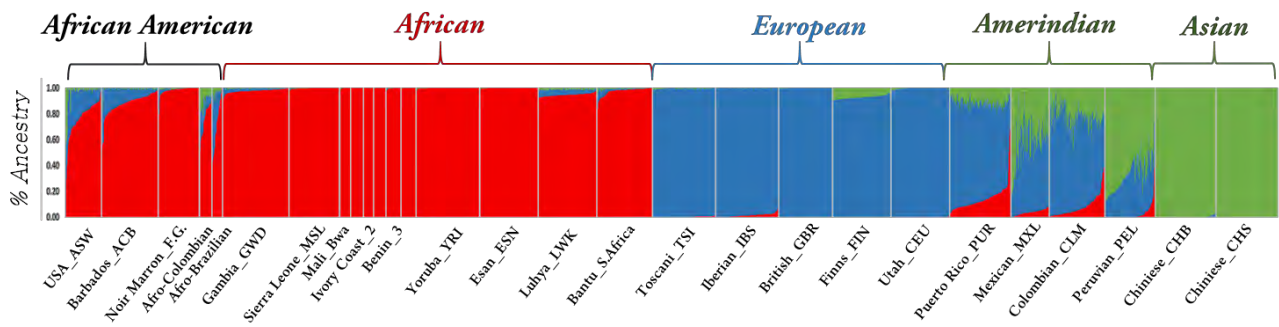
K = 4, CV: 0.4801.
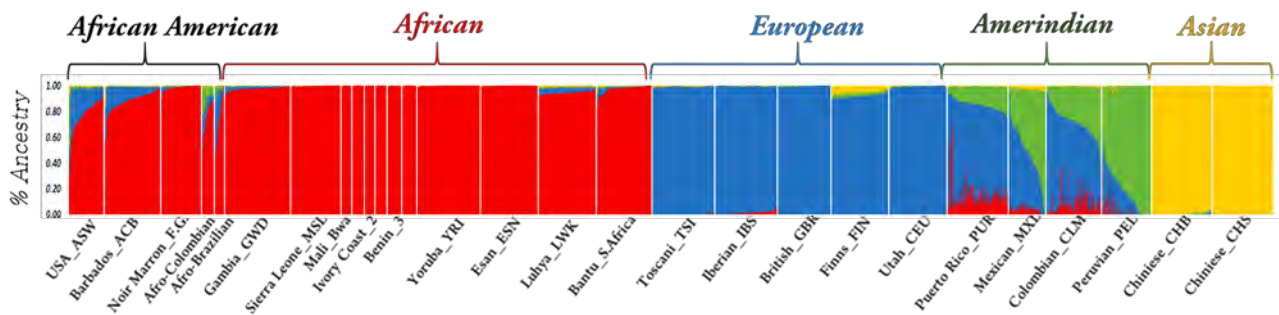
K = 5, CV: 0.4793.
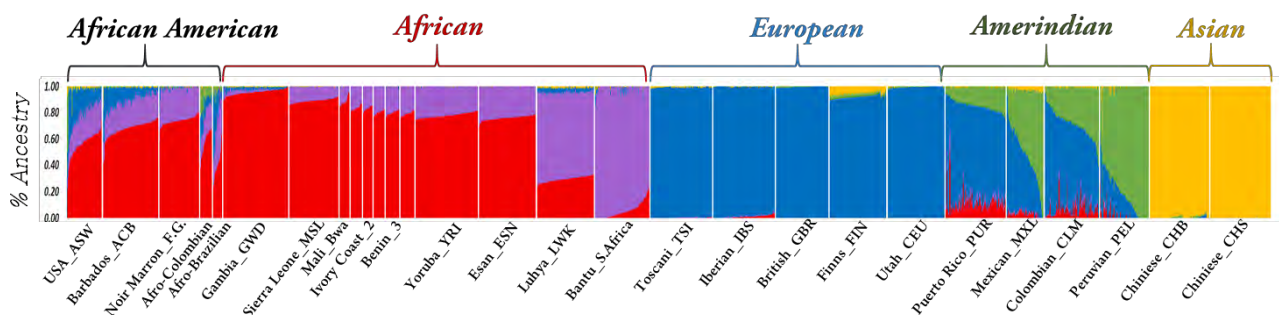
K = 6, CV: 0.4790.
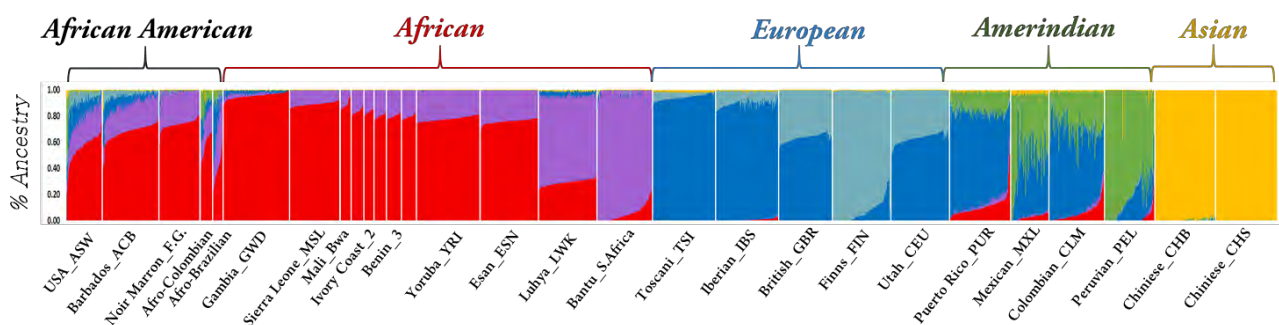
Figure 71: ADMIXTURE plot performed from K-groups= 3 to 6 for African American and reference continental populations. Ancestral component of African (red), Bantu (purple), European (blue), Amerindian (green), and Asian (yellow) were indicated in each plot.

In the other African regions, other patterns of variation were detected, as reflected by the greatest level of regional substructure in Africa (Tishkoff et al., 2009). For instance, in West-Central African populations we found substantial proportions of different ancestry proportions between agriculturalist and rainforest hunter-gatherer populations. This is in good agreement with previous outcomes (Patin et al., 2014; Perry et al., 2014). We also confirmed the substantial difference in ancestry proportions between the Ari population and other Ethiopian populations (Pagani et al., 2012; van Dorp et al., 2015). Among North African (Henn et al., 2012) populations were detected the highest amount of non-African composition, which accounts for both the uppermost European ancestry, suggesting a key role for the Mediterranean region in supporting gene flow into Africa (Botigue et al., 2013; Moorjani et al., 2011). Among Ethiopian populations (Pagani et al., 2012) were detected the highest amount of Middle East ancestry that pointed out historical sources of admixture and subsequent population movements (Hodgson et al., 2014; Quintana-Murci et al., 1999). While the Fulani population (Bryc et al., 2010a), nomadic pastoralists that speak a Niger-Kordofanian language, presented extraordinary diversity of ancestral groups from North, West, West-Central, and East Africa, because of the wide migration behaviour of this population across Africa (Cerny et al., 2006; Scheinfeldt et al., 2010).

To better understand the genetic structure of African-American populations in South America and to determine their African ancestry, we calculated the average membership frequencies of African populations based on historical African coastal regions. Each African regions represented the genetic diversity of African populations that descended from African captives during the slave trade. Similarly, patterns of African substructure are still present in African-American descendants. We used ancestral proportions estimated in the ADMIXTURE analysis for each historical region involved in the slave trade to discern differences in demographic history among African populations. We removed non-African components (European, Native American, Middle Eastern, and East Asian) in Afro-Brazilian and Afro-Colombian, and we recalculated the ancestral proportions for only African groups. We detected strong genetic links between all Noir Marron communities and the Bight of Benin region. Conversely, Afro-Colombians indicated genetic links with Gold Coast, and Afro-Brazilians with West Central Africa region (Figure 73).
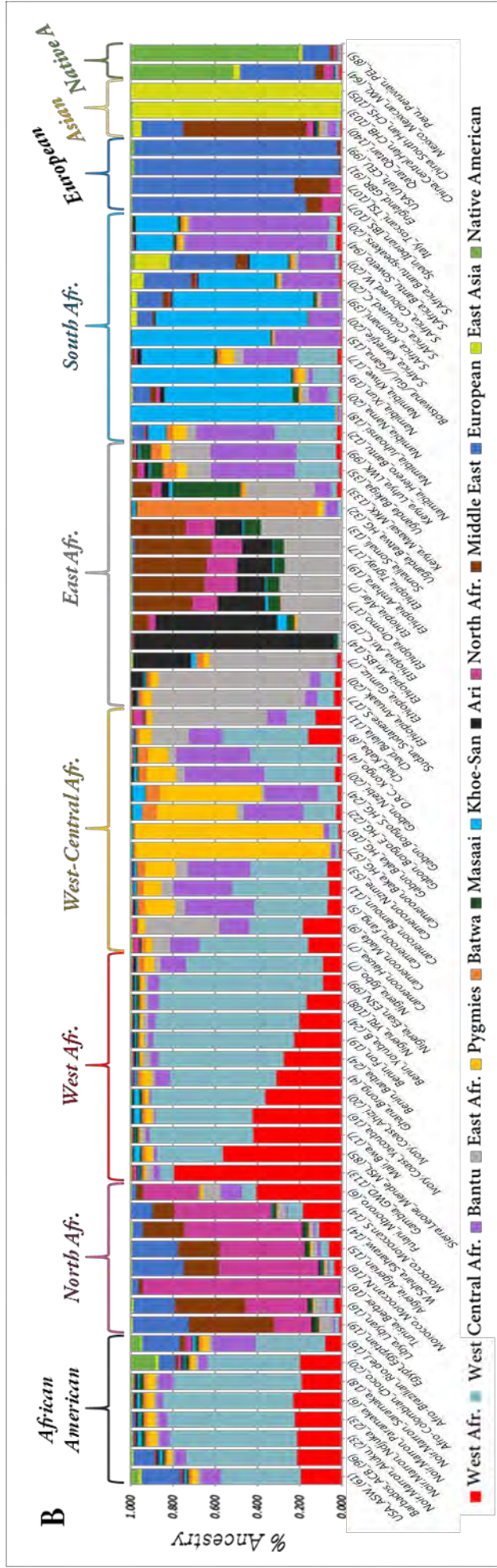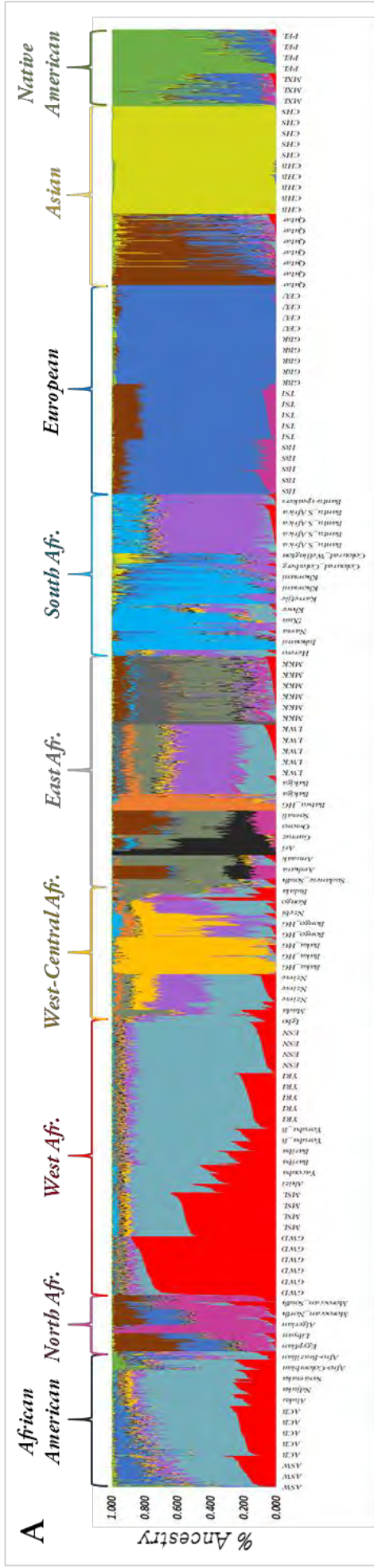
Figure 72: ADMIXTURE plots performed at K-groups= 14 for African continent and African-American populations. **A:** Based on frequencies for individual by individual. **B:** Based on frequencies for population by population. These plots represent the genome-wide structure across African regions.

Table 27: Pairwise F_{ST} genetic distances between ancestry groups estimated in the ADMIXTURE analysis at K= 14. The *F*-statistic values quantify the genetic differentiation between populations included in each group.

Ancestral group		1	2	3	4	5	6	7	8	9	10	11	12	13	14
West Africa	1	0.000													
W.C. Africa	2	0.019	0.000												
Bantu	3	0.025	0.021	0.000											
Pygmies	4	0.052	0.048	0.052	0.000										
Batwa, HG	5	0.074	0.071	0.074	0.069	0.000									
East Africa	6	0.053	0.051	0.057	0.076	0.091	0.000								
Masaai	7	0.075	0.072	0.076	0.091	0.106	0.078	0.000							
Ari	8	0.082	0.081	0.083	0.094	0.106	0.080	0.084	0.000						
Khoe-San	9	0.103	0.101	0.103	0.081	0.103	0.121	0.131	0.125	0.000					
North Africa	10	0.120	0.119	0.123	0.142	0.152	0.122	0.110	0.114	0.177	0.000				
Middle East	11	0.137	0.137	0.140	0.157	0.166	0.138	0.120	0.124	0.190	0.056	0.000			
European	12	0.136	0.135	0.138	0.155	0.164	0.136	0.121	0.124	0.188	0.046	0.040	0.000		
East Asian	13	0.158	0.157	0.160	0.176	0.186	0.159	0.154	0.153	0.209	0.122	0.119	0.105	0.000	
Native American	14	0.191	0.190	0.193	0.208	0.219	0.192	0.187	0.186	0.241	0.151	0.147	0.128	0.103	0.000
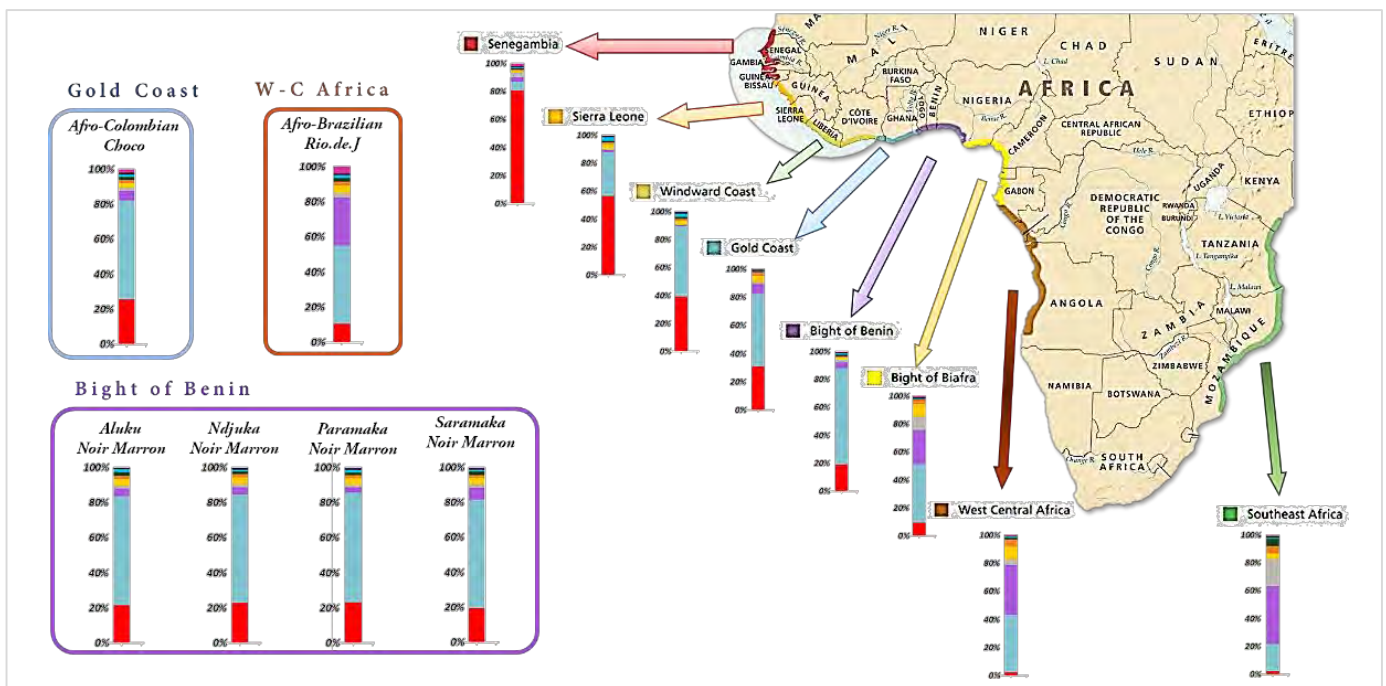
Figure 73: Average membership frequencies of the African-American and African populations that are representative of historical African coastal regions. We indicated genetic links detected in African Americans with different history African region associated to the slave trade.

C.5. *Admixture models by using the PCA analysis*

We investigated patterns of the population structure by using multidimensional analysis based on the PCA implemented with EIGENSOFT. First, the Eigenstrat PCA analysis was applied to detect genetic variations in a large database of worldwide geographic regions (Figure 74). As expected, Eigenvector 1 distinguishes between African and Non-African regions, while Eigenvector 2 differentiates between the European and Asian regions. This is good agreement with previous studies (Gravel et al., 2013; Kidd et al., 2012). The Eigenstrat PCA plot showed: *i*) the Noir Marron were very close into African groups, *ii*) Afro-Brazilians were in African groups and marginally extended to European groups, and *iii*) Afro-Colombians were in African groups and marginally extended to Asian or Native American groups. We obtained analogous results in the simplified Eigenstrat PCA base on African-American populations and only one reference population for African, European, and Native American continental ancestries included in the HGDP panel (Cann et al., 2002) (Figure 75).
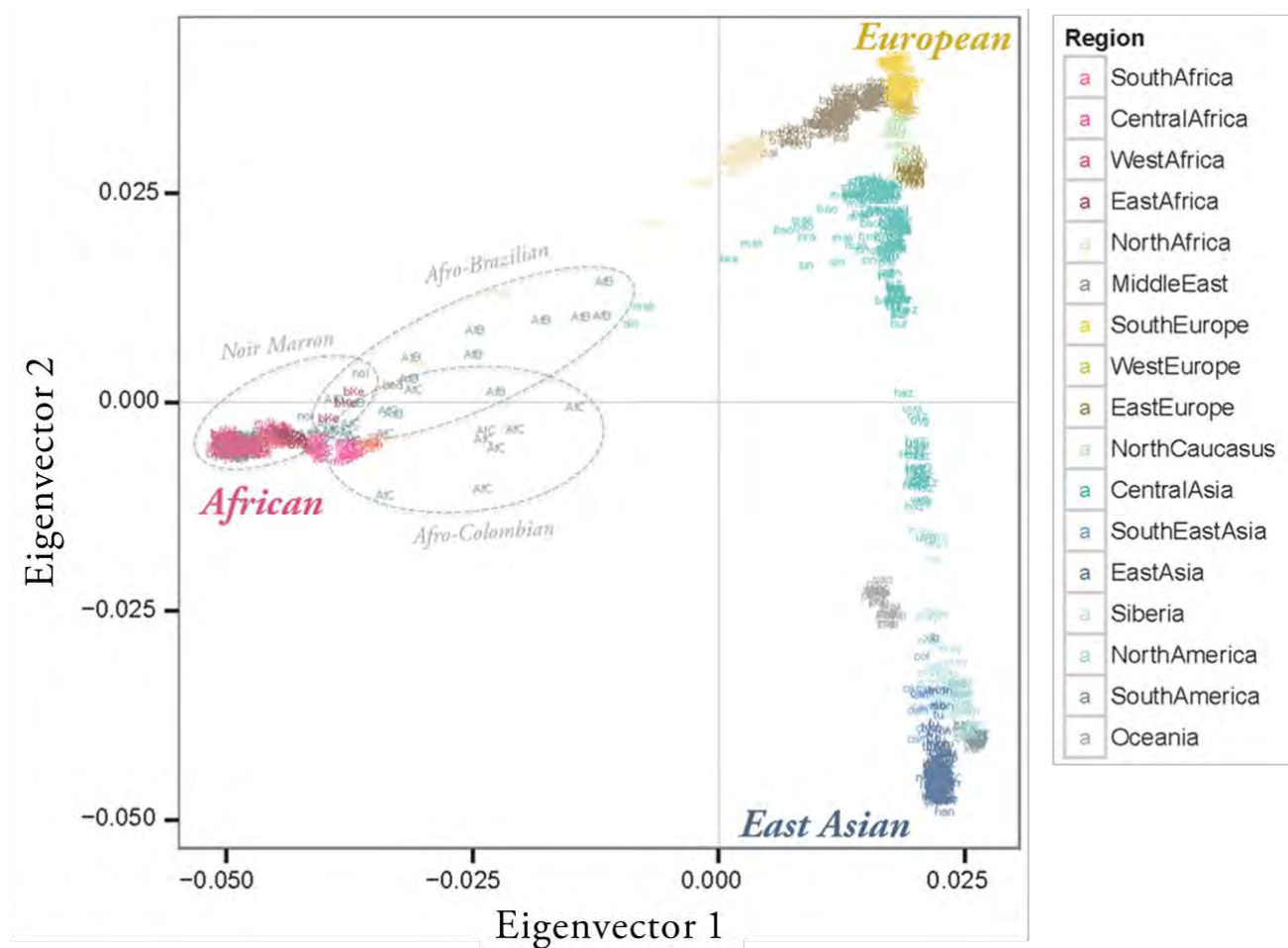
Figure 74: Eigenstrat PCA based on worldwide geographic regions (Table S10) including the African-American populations of the Noir Marron (*noi*), Afro-Brazilians (*AfB*), and Afro-Colombians (*AfC*). Eigenvector 1 (5.23%) and Eigenvector 2 (2.19%) constitute 7.42% of the total variance.

To investigate whether we could reliably identify ancestry among the Noir Marron communities and to find likely genetic links with different African regions in the Atlantic world. We carried out an Eigenstrat PCA analysis for the Noir Marron and historical African coastal regions involved in the slave trade. We confirmed that all Noir Marron communities were together in one cluster within the Bight of Benin cluster in the upper left (Figure 76). Interestingly, the Noir Marron cluster is genetically close to the Beninese populations (Yoruba, Bariba, and Fon) than Nigerian populations (ESN, YRI, and Igbo) within the Bight of Benin cluster. Other populations from different historical regions are notably split in the bidimensional space. Furthermore, we observed a genetic structure in all African regions that could be associated with the same geographic pattern across the African continent (for this we turned the Eigenstrat PCA plot by 90°).
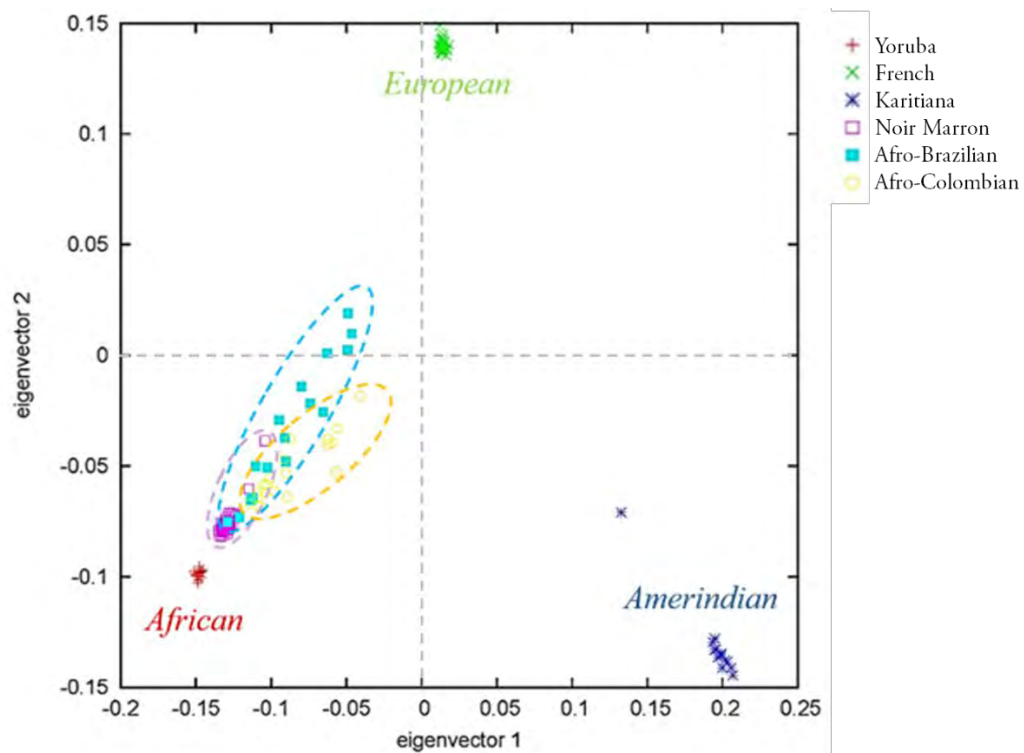
Figure 75: Eigenstrat PCA based on African-American populations and only one reference population for African (Yoruba from Nigeria), European (French from France), and Native American (Karitiana from Brazil) included in the HGDP panel (see Table 12).
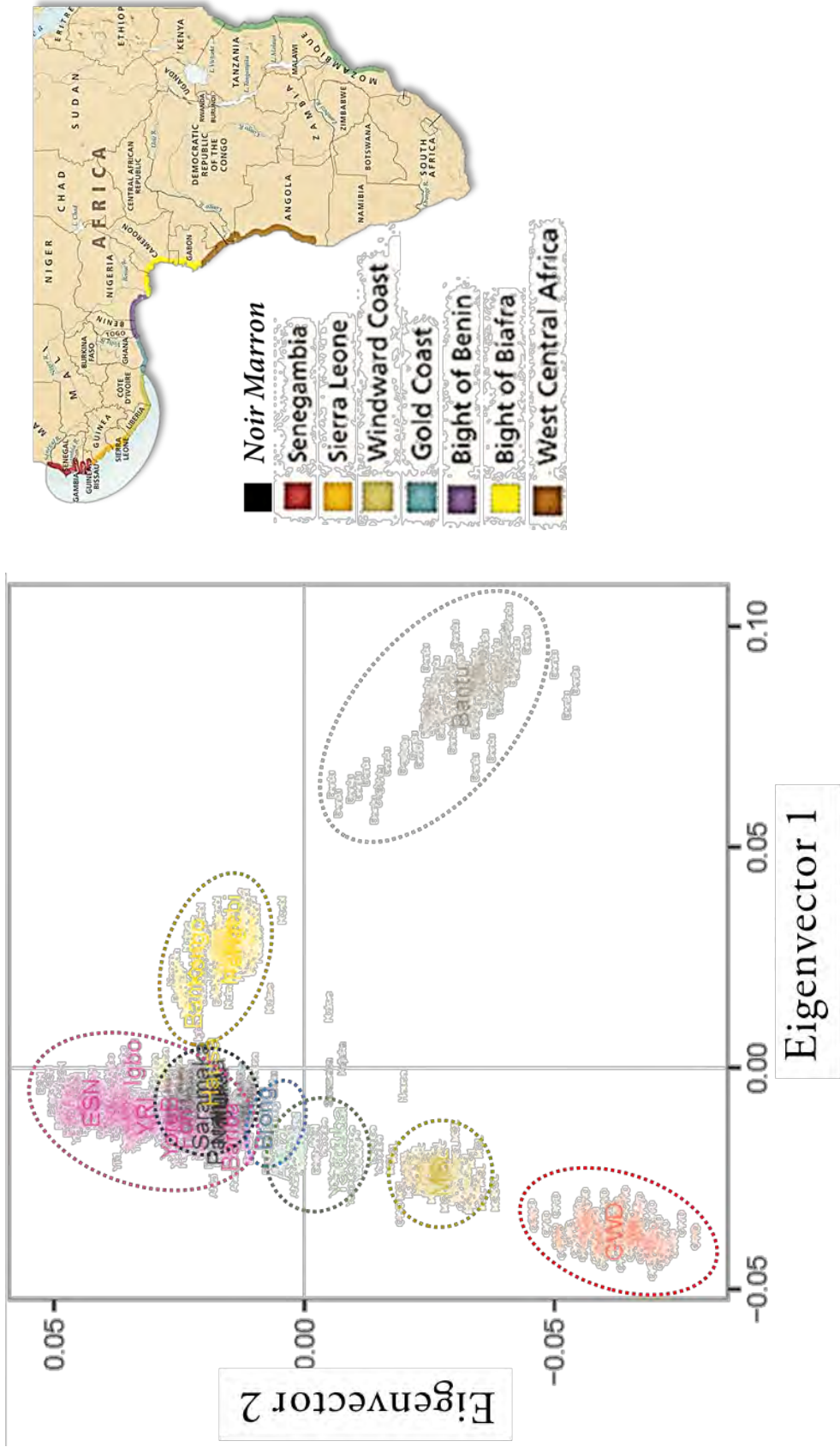
Figure 76: Eigenstrat PCA plot obtained for Noir Marron populations related with historical African coastal regions (Southeast Africa region was not included). All Noir Marron communities are together in one cluster within the Bight of Benin cluster. Eigenvector 1 (7.74%) and Eigenvector 2 (4.86%) constitute 12.6% of the total variance.

IV. DISCUSSION

The present study identifies a strong characterization of the genetic diversities in the four Noir Marron communities, which was compared with other African-American descendants in South America, the Caribbean, and North America. To establish their African origin, we also compared their genetic diversities with West African populations from Benin, Mali, and Ivory Coast as well as with other sub-Saharan African populations involved in the slave trade.

A. *Genetic history of the African-American populations in South America*

Recent findings further support the theory of West African origins of the Noir Marron communities in French Guiana associated with the Bight of Benin; and estimated admixture proportions are roughly in agreement with previous studies on the genetic identities of the Noir Marron by using different genetic systems (Brucato et al., 2009; Brucato et al., 2010). These communities still preserve remarkably high traces of African ancestry and an extremely low admixture with the non-African populations (Figure 77). This study provides new maternal and paternal genetic data to achieve a more complex picture of the genetic variations within the Noir Marron communities. We also analysed genome-wide SNP data in them and other African-American populations in Colombia and Brazil to attempt to shed new light on the slave trade history and its implications on the current genetic landscape in South America.
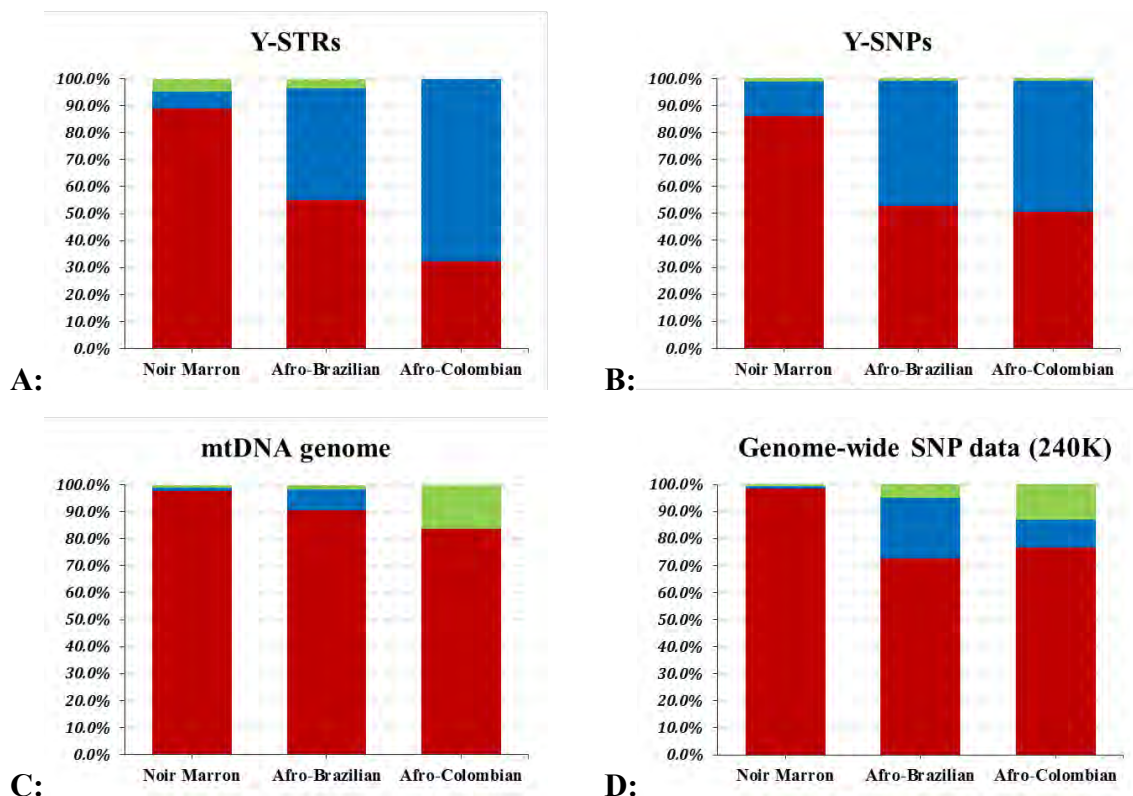
Figure 77: Admixture proportions estimated in African Americans based on different genetic system analysed: Y-STR haplotype frequencies (**A**), Y haplogroup frequencies (**B**), mtDNA haplogroup frequencies (**C**), and admixture coefficients of allele frequencies from 240K SNPs (**D**). Estimates are suitable for African (red), European (blue), and Native American (green) continental ancestries. For all genetic systems analysed, the Noir Marron communities presented elevated African proportions than Afro-Brazilians and Afro-Colombians (see further notes in section IV.A.1). Both Y chromosome markers, Y-STRs and Y-SNPs, presented notable European ancestry in Afro-Brazilians and Afro-Colombians (see further notes in section IV.A.2).

A.1. *Genetic history of the Noir Marron communities*

The Noir Marron communities presented the highest African ancestry in all genetic systems analysed: 89.11% for Y-STRs, 86.03% for Y-SNPs, 97.80% for mtDNA, and 98.34% for genome-wide SNP data (Figure 77). These values are considerably higher than in other African-American populations in the Americas when analysed them with the same high genetic resolution. Therefore, they still present a remarkably strong African identity preserved in the post-abolition period of the slave trade. Within the uniparental markers, the mtDNA genomes showed a higher African component than the Y chromosome markers, which was likely because of the Noir Marron communities presenting a matrilineal system of marriage behaviours (Price & Price, 2003).

We carried out population cross-comparisons to analyses both, the maternal and paternal genetic diversities of the Noir Marron communities. The geographic distributions of the African Y haplogroups depicted close genetic affiliation of the Noir Marron communities to the West African populations (Figure 52). Similarly, the study detected similar genetic links in the geographic distributions of the African mtDNA lineages: L2 (Figure 61) and L3 (Figure 62); and also African Y chromosome lineages such as E1b1a1a1d-U175 (Figure 52). These findings are in good agreement with the genetic distribution analysis of paternal (Figure 78A) and maternal (Figure 78B) lineages as highlighted previously by Brucato et al. (2010).

We found high inbreeding coefficient (Figure 64A) and ROH (Figure 64B) values for the Noir Marron communities, which were remarkably different from the other African-American populations in the Americas (with the exception of the ROH in the Paramaka). It is understood that the population history and cultural factors can affect the levels of inbreeding and homozygosity in individual genomes (Pemberton et al., 2012). In some populations, cultural practices that promote consanguineous marriages or endogamy can result in elevated inbreeding levels, and consequently, high levels of homozygosity even when the overall population size is large (Woods et al., 2006). However, according to historical and anthropological studies, the

Noir Marron communities do not promote consanguineous marriage (Price, 1996, 2001; Price & Price, 2003). Therefore, in these Noir Marron communities, even in the absence of overt endogamy, inbreeding and homozygosity can be high. This is attributed to historical bottlenecks or geographic isolations, which have led to high levels of relatedness among members of a population. It was similarly pointed out among the aborigines of Taiwan (Li et al., 2006), the isolated population of the Orkney Isles in northern Scotland (McQuillan et al., 2008), among religious Lebanese communities within which numerous consanguineous marriages occur (Jalkh et al., 2015), and the small endogamous Tunisian community of Douiret (Ben Halim et al., 2015).

The high degree of variance in ROH among African Americans (Figure 64B) is an additional indicator of substructure among those populations and suggests a large variance in historical population sizes. Comparing the ROH patterns, the Noir Marron might be characterized by small effective population sizes under a model with a likely bottleneck, in agreement with the high ROH found in Native American groups within Mexico (Moreno-Estrada et al., 2014).

These findings highlight interesting questions regarding the formation of these Creole communities and the survival of African cultures in the Amazonian rainforest. For instance, how could they avoid gene flow with non-African groups for around four centuries? Why was it so important for the Noir Marron communities conserve their African identities after the abolition of slave trade?
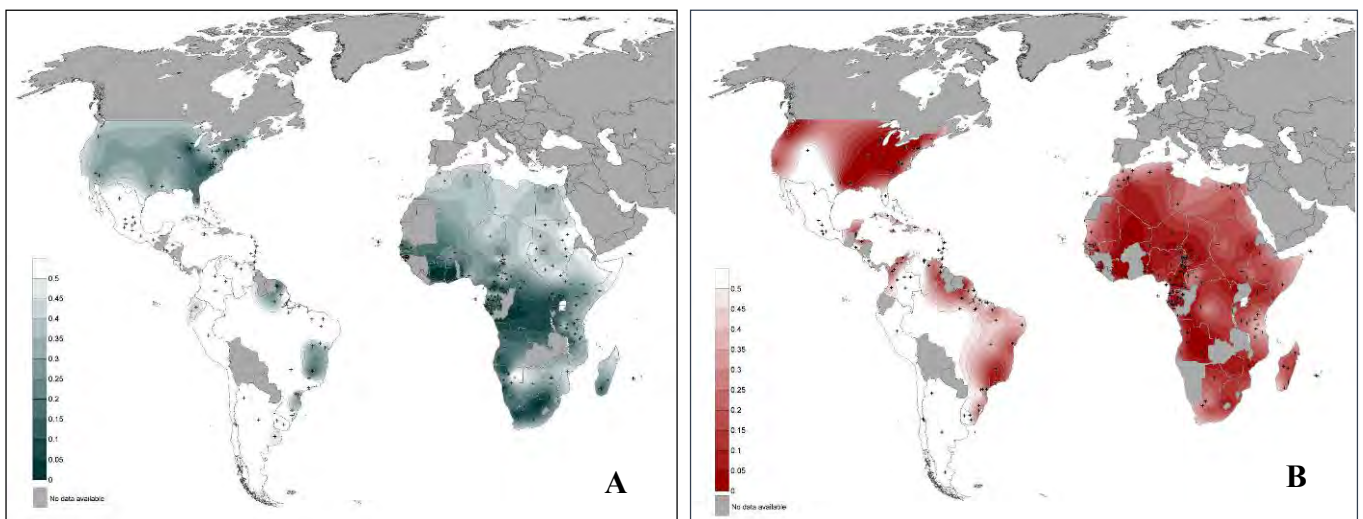
Figure 78: Maps of F_{ST} genetic distances show the genetic distribution of paternal (**A**) and maternal (**B**) lineages in Noir Marron, other African Americans, and African populations. Maps reported by Brucato et al. (2010).

The three-population test estimated in African-American populations was in good agreement with inbreeding coefficients and ROH. The Noir Marron communities are apparently shut out to gene flow with European and Native American populations (Figure 67). In the same way, these communities present high values of inbreeding coefficients and ROH probably because of their genetic isolation. We therefore confirmed the high African diversity in Noir Marron that reveal non-altered inheritance from their enslaved ancestors as it was pointed out by Brucato et al. (2010), despite four centuries in the Americas, neighboured by European settlers and Amerindian ethnic groups with whom the Noir Marron made cultural exchanges. Large linguistic influences are detectable today in the Noir Marron communities: their languages are composed of English, Portuguese, and Amerindian words (Table 7), and the structure of their villages (Figure 20) are inherited from their knowledge of Amerindian culture (Price & Price, 2003). In most African-American groups, cultural exchanges are generally followed by gene flow, however the Noir Marron have conserved a large African genetic inheritance. In same way, they are a unique African ethnic group living today in the Americas, which has preserved their African identity and genetic background after centuries of forced migration. Resistance is a phenomenon that cannot be separated from slavery or oppression, especially in these communities. These Noir Maroon communities conceptualized their freedom in the worst conditions, and they continue to maintain rituals inherited from West African groups particularly from Ghana, Senegal, and Nigeria (Price, 2008).

A.2. *Models of sex-biased gene flow in African Americans*

We detected different patterns of sex-biased gene flow in the Afro-Brazilian and Afro-Colombian populations. Both populations presented high European parental heritage (50.0% and 46.4%, respectively). Conversely, low European maternal heritage was seen among the Afro-Brazilians (5.2%) while there was considerable Native American maternal heritage among the Afro-Colombians (16.6%). The distribution of genetic variations within a population may reflect underlying social practices regarding the choice of mates during and after the slave trade era, which could have contributed to the heterogeneity of the degree of admixtures. These patterns of sex-biased gene flow are indicative of different preferential marriage behaviours in the American colonies that are associated with their different colonial pasts (Portuguese and Spanish, respectively) as highlighted previously by historical data (Eltis & Richardson, 2008; Klein & Vinson, 2007).

The maximum likelihood trees confirmed both gene flow from Europeans to Afro-Brazilians, and also from the Native Americans to the Afro-Colombians (Figures 65A and 65B). Further,

estimations of *f3*-statistics provided additional support to the dissimilar admixture patterns in African-American populations, and the null admixture patterns among the Marron communities (Figures 67A and 67B). This was in sharp contrast to the other Marron communities in the Caribbean that presented considerable gender-specific admixture patterns (Benn Torres et al., 2007; Simms et al., 2012).

Among the Afro-Brazilians, the ADMIXTURE analysis indicated a major component representative of African ancestry (72.5%), remarkable European component (22.3%), and low traces of Native American component (4.6%), which pointed out the importance of the European admixture in the Afro-Brazilian population. This fact is in close agreement with the genetic frequencies of the African (49%), European (39%), and Native American (11%) ancestries estimated recently among the Brazilians with the self-perception of "black" ethnicity by Ruiz-Linares et al. (2014). This was done by using 30 AIMs. Likewise, these results were consistent with the ancestral membership proportions obtained for the self-declared Afro-descendants residing in Rio de Janeiro by using 46 AIM-Indels. The use of 46 AIM-Indels here, revealed a similar African (51.6%), European (39.0%), and Native American (9.4%) ancestries (Manta et al., 2013). Conversely, in admixture proportions estimated in Brazilian populations from different regions by using AIMs, other studies detected considerable high European proportions (62.0–77.1%) (Lins et al., 2010; Rodrigues de Moura et al., 2015).

Interestingly, the Afro-Brazilian population showed the lowest values (72.5%) for the African ancestry proportion among all African-Americans analysed (Table 26). Despite the fact that enslaved Africans came to Brazil during the last period of the slave trade, and at least 798,309 enslaved Africans disembarked at Rio de Janeiro between 1801 and 1856 (Table 2). This meant that 72.54% of the total numbers of African captives disembarked in that region. Besides, Brazil was the last country in the Western world to abolish slavery in 1888. By that time, around four million slaves had been forced to move from Africa to Brazil and 40% of the total number of slaves had been brought to the Americas (Eltis & Richardson, 2010).

The predominant initial migration of European men resulted in the union between European men and African women (Klein, 2002). The entrance of the Europeans and Africans into Brazil is historically well-documented and genetically well-supported by previous studies, which have used uniparental markers (Abe-Sandes et al., 2004; Alves-Silva et al., 2000; Bortolini et al., 1999; Carvalho-Silva et al., 2001; Domingues et al., 2007; Hunemeier et al., 2007; Manta et al., 2013; Santos et al., 2010). Both, the mtDNA and the Y chromosome showed strong evidence of asymmetric mating patterns involving predominantly, the European men and African women (Pena et al., 2009; Salzano, 2004). This is supported by historical data regarding the formation of the Brazilian population (Klein, 2002). Recently, an asymmetric mating pattern was also

observed between the African men and the Native American women in the Afro-Brazilian communities of Maranhão in the Brazilian Amazonia (Pereira et al., 2012).

A high European ancestry (>55%) was also reported in autosomal studies for admixed Brazilian populations particularly in Rio de Janeiro (Manta et al., 2013) and Brazil in general (Callegari-Jacques et al., 2003; Lins et al., 2010; Parra et al., 2003; Pena et al., 2011). Recently, Kehdy et al. (2015) reported a higher mean proportion of European ancestry in South (Pelotas: 76.0%) and Southeast Brazil (Bambuí: 77.0%), in comparison to Northeast Brazil (Salvador: 43.0%) based on above 330K SNPs. Regarding to their African ancestry, the Pelotas, Bambuí, and Salvador communities presented mean proportions of 16.0, 16.0, and 50.0%, respectively, while the mean proportion of continental Native American ancestry were similar and low for all populations: 8.0, 7.0, and 7.0%, respectively. In general terms, European and African admixture proportions varied greatly among Brazilian populations. It was noteworthy at the level of the entire country. At the regional level, however, the European ancestry was found to be the major contributor to the genetic background of the Brazilians, even among those who were African descendants (Rodrigues de Moura et al., 2015).

Among the Afro-Colombians, the ADMIXTURE analysis indicated a major component for African ancestry (76.6%), and notable proportions for European (10.4%) and Native American (12.2%) ancestries. These values are in close agreement with the genetic frequencies of African (69%), European (19%), and Native American (12%) ancestries estimated recently for the Colombians with self-perception of "black" ethnicity as studied by Ruiz-Linares et al. (2014) through the use of 30 AIMs.

Colombian populations has been the focus of a large number of genetic studies (Carvajal-Carmona et al., 2003; Rojas et al., 2010; Salas et al., 2005b). However, the first study of the Afro-Colombians in Chocó was focused on the polymorphisms of forensic interest based on the autosomal STR genetic variations (Bravo et al., 2001). One of the first attempts to unravel the issues of Afro-Colombian ancestry was performed by Rodas et al. (2003) through an analyses of the mtDNA variations. Low frequencies of lineages L in five different Afro-Colombian populations were reported with results ranging from 21.4% in Quibdó to 52.5% in Providencia.

Focusing on a phylogeographic approach of the mtDNA, Salas et al. (2008) investigated the genetic ancestry of admixed Colombian groups including: *"mestizos"* (Colombian term to designate individuals of European and Native American co-ancestry), *"mulatos"* (Colombian term to designate individuals of African and European co-ancestry), and Afro-Colombians. Salas et al. (2008) reported that 72.6% of the lineages belonged to the macrohaplogroup L, which is still lower than what has been found in the present study (83.4%; Table 21).

In good concordance with the ADMIXTURE and TreeMix analyses, the analysed mtDNA genomes in Afro-Colombian indicated a noteworthy Native-American maternal heredity, which is associated with the mtDNA haplogroups: A, B, and C (16.6% in total; Table 21). Those Native American clades were present in Colombia during the pre-Colombian period before the European and African populations arrived to this region. It has been confirmed in the Colombian Amerindian populations using both, modern (Keyeux et al., 2002) and ancient mtDNA (Casas-Vargas et al., 2011). Clade A and B were more frequent in northern Colombia, while the haplogroup C and D were more frequent in southern and South-western Colombia (Noguera-Santamaria et al., 2015). Hence, the Afro-Colombian population presented evidence of maternal Native American admixtures mainly from the northern Colombian gene flow (13.3%) than from southern Colombia (3.3%). In contrast with the Afro-Brazilian population, Afro-Colombian did not show any evidence of maternal European admixture, which is in agreement with mY_{mtDNA} values (Table 22).

With regard to the Y chromosome markers, the Afro-Colombian population presented 67.9% of European ancestry associated with Y-STRs, and 48.8% of the European admixture proportions as associated with the Y haplogroup frequencies (Table 17). In the same way, within the *Palenque* population from San Basilio de Palenque (Bolivar department), Noguera et al. (2013) pointed out a high (38.5%) European male component in this population. This *Palenque* population, founded in the second half of the 17th century, is considered the best example of an African free community among the African descendants in Colombia (Schwegler, 2011), however this population presented higher European male component than the Noir Marron communities. Likewise, high frequencies of European Y haplogroups were detected in other admixed populations living in Bolivar (range 57.1–81.8%) (Noguera et al., 2013). This was in close agreement with the sum of European Y haplogroups (36.29%) reported in the Afro-Colombians from the Cauca region located in South-West Colombia (Acosta et al., 2009). Among the Afro-Colombians from Chocó in North-West Colombia, a remarkable presence of European Y-lineages (52.4%) (Rojas et al., 2010) was also detected. These converging results were concordant with the European paternal lineages (50.0%) detected in the analysed Afro-Colombian population from Chocó (Figure 49), and in disagreement with a genetic study of eight *Alu* insertion polymorphisms in other Afro-Colombian population residing in Antioquia that determined null contribution of the European component (Gomez-Perez et al., 2010).

Overall, the present study adds further support to the asymmetric mating during the slave trade, and highlights differences between the African descendants with different colonial pasts. Only in Colombia, preferential marriages were detected between African men (belonging to the mtDNA macrohaplogroup L) and Native American women (belonging to mtDNA

subhaplogroups: A2, B2d, and C1d). A noteworthy tendency of preferential marriages was detected between the European men (belonging to the Y subhaplogroup R1b1a2-M269) and the African women (belonging to the mtDNA macrohaplogroup L) in both, Brazil and Colombia. Before the 19th century, the vast majority of European immigrants arriving in South America were masculine (Bryant et al., 2012; Klein & Vinson, 2007). Nevertheless, the high proportion of European male-biased asymmetric gene flow detected was atypical for colonial settings (Wilkins & Marlowe, 2006). This suggests that the admixture occurred during a period when African people were no longer enslaved. This is thoroughly consistent with historical information from Brazil and Colombia. Similar patterns of sex-biased gene flow were described in populations from other regions in South America, such as, Argentina (Corach et al., 2010), Chile (Cifuentes et al., 2004), Uruguay (Bravi et al., 1997), Brazil (Carvalho-Silva et al., 2001; Goncalves et al., 2008), and Costa Rica (Carvajal-Carmona et al., 2003). Indeed, it may prove a general characteristic of the Iberian colonization of the Latin American region, however, several studies on African-American populations founded by fugitive slaves have reported variable or null proportions of European ancestry (Bortolini et al., 1995; Bortolini et al., 1999; Cotrim et al., 2004; Da Silva et al., 1999; Sans et al., 2002). This is suggestive of partial isolation or variability in the marriage behaviours of those communities.

B.   Genetic African heritage of the slave trade in South America

The transatlantic slave trade created a new distribution of African genetic diversity in the Americas, and generated new opportunities for gene flow through local interactions with non-Africans. Interestingly, interactions among Africans from various regions were also contributory factors to this phenomenon. However, the Middle Passage was, in some way, a genetic bottleneck for enslaved Africans. It restricted the African genetic diversity, and enslaved Africans arriving in the New World represented a limited fraction of the African continent. Survivors of the Middle Passage were able to regenerate much of this original diversity primarily through the gene flow of the African populations in the New World. This was achieved through African tribal and regional restrictions on marriage that were not in force and secondly, because of the gene flow with non-African populations.

Indeed, African-American descendant groups in the Americas still had strong genetic affinities with West, West Central, and Southeast Africans, despite four centuries of separation. These genetic links were underscored by new evidence from maternally inherited mtDNA, paternally inherited Y chromosome, and the autosomal DNA associated with African ancestry. We carried out population cross-comparisons for both, the maternal and the paternal genetic diversities of

African-American and African populations to trace back their African origins. First, we analysed patterns of geographic distributions of major African lineages across the African continent and the African-American populations. We used contour maps for both, the mtDNA (see section III.A.3) and the Y chromosome haplogroup frequencies (see section III.B.3). We also elaborated a comprehensive phylogenetic tree for the mtDNA genomic variation among the African and the African-American populations (see simplified tree in Figure 57). Second, we estimated paternal (mY $_{Y-SNP}$) and maternal (mY $_{mtDNA}$) ancestral contributions of historical African coastal regions among the African Americans (Table 28). In agreement with historic data (Eltis & Richardson, 2013; Morgan, 2013), those living in African regions of the Gold Coast, Bight of Benin, Bight of Biafra, and West Central Africa proved highly related to the African descendants from South America for both genetic systems.

Table 28: Ancestral contribution of historical African regions, European, and Native American ancestries in African Americans. Ancestry proportions are based on Y haplogroup frequencies (mY $_{Y-SNP}$) and mtDNA haplogroup frequencies (mY $_{mtDNA}$).

mY $_{Y-SNP}$	N	Noir Marron (N= 60)	Afro-Brazilian (N= 24)	Afro-Colombian (N= 19)
Senegambia	252	0.0007 ±0.0004	0.0554 ±0.0075	0.0500 ±0.0066
Sierra Leone	42	0.0855 ±0.0092	0.1731 ±0.0066	0.0057 ±0.0074
Windward Coast	90	0.0011 ±0.0003	0.0534 ±0.0076	0.2078 ±0.0057
Gold Coast	205	0.1815 ±0.0079	0.0749 ±0.0021	0.0294 ±0.0069
Bight of Benin	296	0.2879 ±0.0050	0.0098 ±0.0008	0.1757 ±0.0097
Bight of Biafra	805	0.1564 ±0.0063	0.1747 ±0.0098	0.2447 ±0.0114
West Central Africa	280	0.2630 ±0.0057	0.2488 ±0.0051	0.1873 ±0.0094
Southeast Africa	318	0.0029 ±0.0027	0.0507 ±0.0061	0.0388 ±0.0040
Europe	110	0.0205 ±0.0045	0.1588 ±0.0017	0.0605 ±0.0012
Native American	13	0.0005 ±0.0002	0.0004 ±0.0001	0.0001 ±0.0007
mY $_{mtDNA}$	N	Noir Marron (N= 183)	Afro-Brazilian (N= 39)	Afro-Colombian (N= 30)
Senegambia	656	0.0241 ±0.0036	0.0176 ±0.0071	0.1253 ±0.0275
Sierra Leone	362	0.0270 ±0.0054	0.0075 ±0.0007	0.0480 ±0.0066
Windward Coast	192	0.0673 ±0.0086	0.0573 ±0.0074	0.0405 ±0.0076
Gold Coast	429	0.2233 ±0.0066	0.1062 ±0.0057	0.1250 ±0.0021
Bight of Benin	485	0.2551 ±0.0074	0.2776 ±0.0069	0.0320 ±0.0078
Bight of Biafra	2,714	0.2570 ±0.0092	0.3128 ±0.0097	0.2346 ±0.0098
West Central Africa	555	0.0990 ±0.0065	0.0105 ±0.0014	0.1756 ±0.0051
Southeast Africa	404	0.0460 ±0.0031	0.1200 ±0.0094	0.0871 ±0.0061
Europe	100	0.0007 ±0.0006	0.0735 ±0.0040	0.0003 ±0.0002
Native American	100	0.0005 ±0.0003	0.0170 ±0.0013	0.1316 ±0.0241

In agreement with the PCA analysis of haplogroup frequencies (Figures 53 and 63), it may not be feasible to determine a unique source of African captives by using uniparental markers. According to *mY Y-SNP*, West Central Africa slaving region is associated with the Noir Marron and Afro-Colombian populations in sharp contrast to previous analyses (Figure 53). It may suggest close genetic backgrounds of sub-Saharan African populations because of the Bantu expansion (de Filippo et al., 2011), which increased the homogeneity of the haplogroup composition among the Africans and the African Americans by means of two African paternal markers such as E1b1a1a1c1a1-U174 and E1b1a1a1d-U175.

B.1. *Paternal genetic legacy of the African-American descendants*

Among the Noir Marron communities, we detected close genetic connections with populations from Benin residing in the historical region of the Bight of Benin. Despite the high European male lineage among the Afro-Brazilians we found close genetic connections with populations from Angola in West Central Africa region (Table 28). The Afro-Colombian presented close genetic connections with populations from West African regions. Likewise, the *Palenque* population presented a close genetic proximity with the population from Cape Verde as well as similar admixed proportions of European and African contributions (Noguera et al., 2013). The high European male frequencies detected within the African descendants in Brazil and Colombia (46.4% and 50.0%, respectively) were in sharp contrast with frequencies detected among African Americans residing in different geographic locations of the U.S. (~26.4%) (Hammer et al., 2006). In both, African-American populations from South America were also detected the highest frequencies for the subhaplogroup R1b1a2-M269 (25.0%). In Colombia, this European subhaplogroup could have come from founders, who predominantly arrived from the Iberian Peninsula (Bedoya et al., 2006; Carvajal-Carmona et al., 2000; Carvajal-Carmona et al., 2003). In Brazil though, this was most likely related to the Portuguese population (Beleza et al., 2006). Nevertheless, it is estimated that over 110 million European men belong to this subhaplogroup (Balaresque et al., 2010), which is carried by two thirds of Western European men (Sole-Morata et al., 2014). According to the MJ Network analysis, most Spanish and Portuguese haplotypes belonged to R1b1a2-M269, and shared the same or a very similar haplotype (Figure 51). Therefore, we cannot establish genetic differences between the Spanish and Portuguese haplotypes to infer the European source of parental ancestry among the African descendants based on the Y chromosome markers.

The presence of the Y subhaplogroup R1a1a-M17 detected among the Afro-Brazilians and the Afro-Colombians is also noteworthy (7.1% and 5.0%, respectively). The spatial frequency

distributions of this subhaplogroup have been primarily found in Europe and confined to Central and South Asia (Underhill et al., 2010; Underhill et al., 2015). It might indicate European and Asian gene flow to African descendants after the abolition of slave trade.

Taking into account the African Y haplogroups, the low presence (3.3%) of haplogroup B-M60 was detected among the Noir Marron, which was in good agreement with the B-M60 frequency (4.0%) reported recently among the Accompong Town Maroon community from western Jamaica (Madrilejo et al., 2015). This haplogroup belonged to one of the oldest Y-lineages in sub-Saharan Africa (Batini et al., 2011a). This finding highlighted the fact that African-Americans could preserve the signatures of very early human Y chromosome lineages. Interestingly, Mendez et al. (2013) found the most ancient human lineage among the African Americans, called the haplogroup A00. This lineage A00 carried the ancestral state of all SNPs that were highly divergent from other known African lineages, and defined the basal portion of the Y chromosome phylogenetic tree. Therefore, today some African Americans present strong signatures of the most recent common ancestor (MRCA) for the Y tree.

B.2. *Maternal genetic legacy of the African-American descendants*

The most common specific mtDNA subhaplogroup among the analysed African Americans was L2a1 (range 18.6–26.7%). We detected numerous mtDNA haplogroups (see Table 21) and low frequencies of non-Africans mtDNA haplogroups (range 5.2–16.6%), suggestive of present-day African-American maternal lineages representing a large mixtures of the African mtDNA variants, in close correspondence with the values reported for the African Americans in the U.S. (Allard et al., 2005; Just et al., 2015), and in the Caribbean (Benn Torres et al., 2007; Deason et al., 2012). In the PCA analysis of mtDNA haplogroup frequencies, the African Americans from Colombia, Barbados, the U.S., and French Guiana (with the exception of the Aluku community) were found to be closely related to the populations from the West African regions (Figure 63). Conversely, the Aluku and Afro-Brazilians were close to the Bight of Biafra and West Central African regions.

Indeed, the current Brazilian population is the most important representative of African mtDNA lineages outside Africa. It has been estimated that at least 90 million individuals in Brazil, independent of their physical appearances, have a sub-Saharan African origins according to their mtDNA (Pena et al., 2009). Within Brazil today, there is quite a bit of regional variation in the proportions of African descendants among the local population. For instance, in Rio Grande do Sul, around 16% of the individuals self-identified as "white" present African maternal origins (Marrero et al., 2005).

The phylogenetic analysis of mtDNA haplotypes indicated complex geographic associations among African Americans. In Afro-Brazilians, the detection of the mtDNA lineages from Angola and Mozambique were consistent with the historical records, since Brazil was one of the main destinations for African captives from Portuguese colonies (Araujo, 2015; Landers et al., 2015), and most of the Brazilian enslaved, who arrived in Rio de Janeiro were mainly from West Central Africa region (see Table 4). Among the Afro-Brazilians living in Rio de Janeiro and Porto Alegre, a previous study did indicate that their mtDNA variants were originated from West-Central and Southeast African regions (69% and 82%, respectively) (Hunemeier et al., 2007). This was in close agreement with genetic links detected in the phylogenetic reconstructions of their mtDNA genomes.

Surprisingly, in Tocaña, the isolated Afro-Bolivian community in the province of *Nor* (North) Yungas in La Paz department in Bolivia, was found to have maternal genetic links to the South-eastern African populations and with the Near Eastern countries, such as Saudi Arabia, Yemen, and Oman (Heinz et al., 2015). Those results were explained by the Arab slave trade having initiated contacts between these regions. In the analysis of geographic patterns of major mtDNA clades L in the African and Near Eastern countries (Figure 59 to Figure 62) and phylogenetic reconstruction of African lineages L0–L3 by using mtDNA genomes, this study did not detect any evidence to support the fact that the Arab slave trade marked the genetic structures of the African-American populations of South America.

B.3. *Autosomal genetic legacy of the African-American descendants*

Model-based methods for estimating ancestry have traditionally focused on the relationships within and between populations while quantifying the admixture proportions in the admixed populations (Liu et al., 2013; Novembre & Ramachandran, 2011). The three main factors that determine the accuracy and precision of ancestry estimations are: *i*) number and resolution of markers used, *ii*) quality of the reference databases (both in geographical spread and number of samples included), and *iii*) levels of genetic differentiations between populations in the African regions being considered (Shriver & Kittles, 2004). We applied model-based methods by using different datasets for different continental reference populations (50K, 90K, and 240K SNPs datasets) to estimate ancestry proportions among the African Americans. Despite sharing an African ancestry, admixed African-American populations exhibited large variations in the admixture proportions associated with ancestral continental populations (Parra et al., 1998; Via et al., 2011). In all, the ADMIXTURE analyses evidenced that the highest African ancestry was detected within the Noir Marron communities. Different admixture patterns were determined

among the remaining African Americans (see Table 26).

As of late, several relevant research efforts have focused on identifying a small set of AIMs that can be used to infer biogeographical ancestry and admixture proportions (Galanter et al., 2012; Halder et al., 2008; Londin et al., 2010; Pardo-Seco et al., 2014). However, the majority of the AIM panels are designed to determine admixture proportions between only three source populations (African, European, and Native American). We estimated admixture proportions using three source populations based on the 90K SNPs dataset (Figure 69). In order to obtain more accurate estimations, we also used four source populations based on 240K SNPs (Figure 71). Interestingly in both ADMIXTURE analyses, we detected population substructures within the African populations. We characterized the African structure by using a large representation of the African populations based on the 50K SNPs dataset (Figure 72). We also estimated genetic ancestry at different scales of complexity: at the individual level (Figure 72A), at the population level (Figure 72B), and at the regional level (Figure 73).

We also represented West-Central Africa and Bantu ancestry proportions as obtained in the ADMIXTURE analysis. We used the same approach to analyse the Q proportion applied to the African populations by Hodgson et al. (2014). We observed a high West-Central African ancestry among the Noir Marron and Afro-Colombians (Figure 79). Conversely, the Afro-Brazilians presented a high Bantu ancestry from D.R.C. to South Africa. This Bantu ancestry is likely associated with the populations from Angola or Mozambique, however there is not available genome-wide SNP data from Angola or Mozambique to confirm that genetic link.

In order to discern differences in demographic history among the African-American populations in the Americas, we also analysed populations from the U.S. and Barbados. The continental ancestry proportions among the African Americans from Barbados (Table 26) were found to be in good agreement with the high African ancestry previously detected by Benn-Torres et al. (2008) who used 28 AIMs. They reported 89.6% African, 10.2% European, and 0.2% Native American ancestries for this population. This contrasted with those of African descent from Jamaica, which had the highest mean rank for Native American ancestry (3.2 to 5.9%) (Benn-Torres et al., 2008). These differences in the Native American component were attributed to admixture patterns between indigenous communities and the Maroon population residing in Jamaica. In contrast, there was almost null admixture detected among the Maroon population in Barbados, because of they lived in regions of the island that were difficult to access (Benn-Torres et al., 2008). We found a similar lack of admixture in the Noir Marron communities living in the isolated regions of the Amazonian rainforest of French Guiana.

We detected a very low East Asian component in African Americans from the U.S. (1.15%), which might indicate a very recent gene flow between these populations in North America.

Interestingly, this pattern was not observed in the Noir Marron communities, despite substantial recent migrations of the East Asian populations in French Guiana that originated in South China (Brucato et al., 2012; Tapp, 2005).

West-Central ancestry
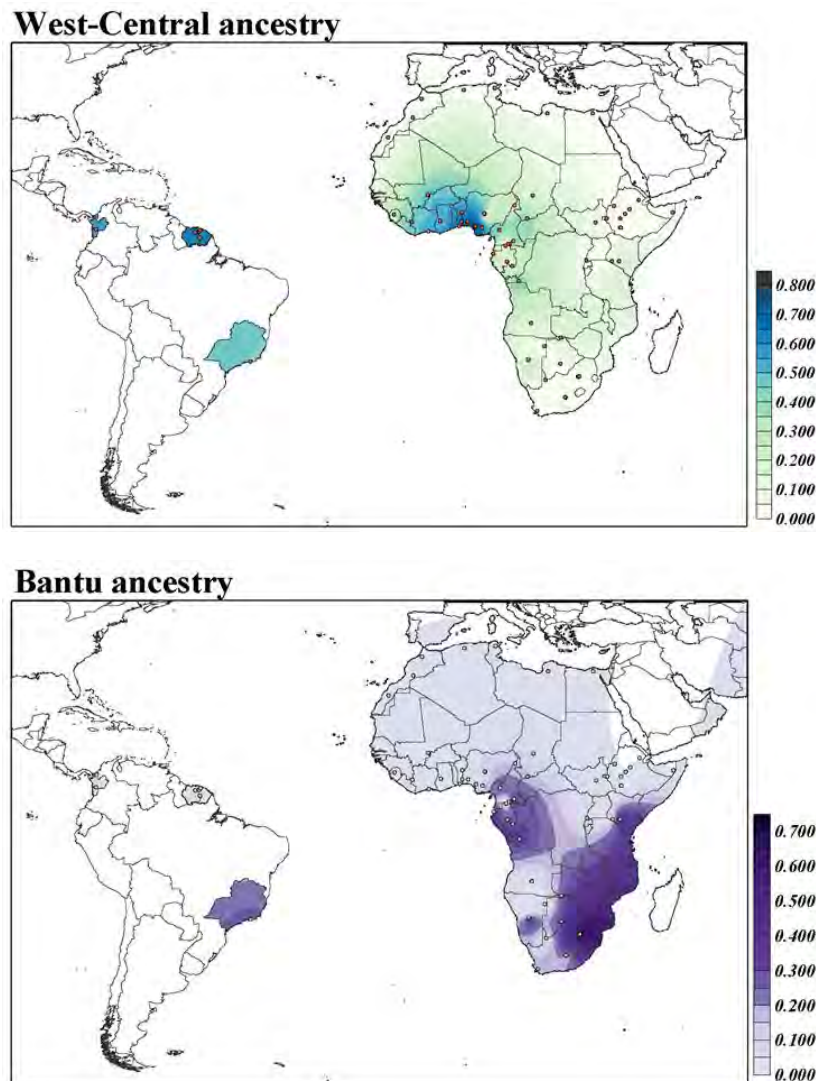
Bantu ancestry

Figure 79: Population structure for West-Central and Bantu ancestry of African American and Africa populations based on ancestry components obtained in ADMIXTURE analysis (Figure 72B).

C.   *Genetic landscape of West African ancestry in African Americans*

Despite the patterns of sex-biases and admixtures identified in African Americans, studies of nuclear DNA diversity revealed that important African genetic identities have persisted among the African Americans derived from the indigenous groups in West and West-Central Africa. For instance, a high number of markers of nuclear genes with African ancestry are commonly found today among the African Americans, such as the Duffy null allele (*Fy0*) (Howes et al., 2011), the sickle cell allele (and its various molecular forms) (Fong et al., 2013; Lemos Cardoso & Farias Guerreiro, 2006), and certain specific HLA polymorphisms (Hanchard et al., 2006).

Other genetic traits show similar patterns, including those related to body sizes and energy stores (Luke et al., 2001), hypertension susceptibilities (Rotimi et al., 1996), and the HLA-QA1 allele (Zimmerman et al., 1995). These retained nuclear DNA polymorphisms suggest that, in conjunction with the mtDNA and Y chromosome evidence, the African-descendant groups are still strongly linked, genetically, to a western African heritage, even after four centuries of separation (around 15 generations).

However, it is very complex to identify the African origins within the West and West-Central African populations (Shriver & Kittles, 2004). The genetic landscapes in West and West-Central Africa is characterised by strong inter-individual differentiation but relatively low inter-population differentiations (Tishkoff et al., 2009). Pairwise F_{ST} distances between African groups indicate the lowest values (0.019) for the West and West-Central African populations (Table 27). Several studies add a better understanding of genetic backgrounds of these populations by using the mtDNA (Montano et al., 2013; Salas et al., 2002) or Y chromosome markers (de Filippo et al., 2011; Scozzari et al., 2014). However, there is still scarce knowledge about genetic differences among the West and West-Central African populations based on bi-parental markers that could shed new light on genetic identities of the African Americans.

Recent studies evaluated nuclear ancestry in American admixed populations (Montinaro et al., 2015) or geographic population origin (Elhaik et al., 2014) based on the same admixture-based methods. Nevertheless, those studies present large number of European reference populations and considerable low number of African reference populations, especially with West and West-Central African ancestries (Figure 80).
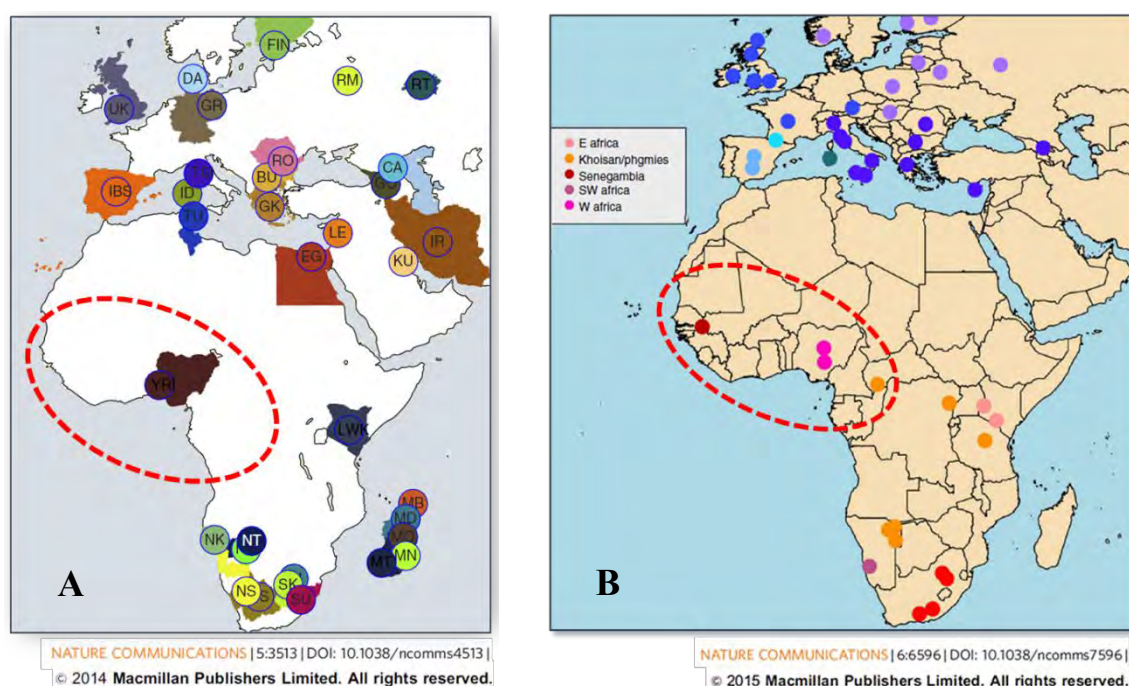
To obtain a relevant African genetic picture that could shed new light about African ancestry, we analysed a large representation of African populations by using an up-to-date dataset of African populations in addition to the new West African data (Figure 46). In the ADMIXTURE analysis, we identified a West African structure that matched a geographic distribution as a genetic gradient (Figure 72). Interestingly, all the Noir Marron communities presented identical ancestry proportions than the African populations residing in the Bight of Benin region (Figure 73). This pointed out a strong contribution from the Bight of Benin despite enslaved Africans were imported in great number to French Guiana and Surinam from different slave regions (Figures 11 and 12). The Eigenstrat PCA analysis corroborated both the African genetic identities of the Noir Marron communities and also the West African structure within West African populations (Figure 76). All the Noir Marron communities clustered within the Bight of Benin especially with populations from Benin, and these were more distant from the Nigerian populations residing in the Bight of Benin region. Conversely, we detected maternal contributions from the Bight of Biafra and West Central Africa regions in the Aluku community (Figure 63).

Otherwise, the ADMIXTURE analysis for the Afro-Colombians showed identical African ancestry from the population localized in the Gold Coast region, and the Afro-Brazilians with populations from West Central Africa region. Interestingly in Colombia, historic data witnessed a considerably low number of enslaved Africans from the Gold Coast region (Figure 18), in contrast with the high numbers of African captives who arrived from Senegambia (28.0%) and West Central Africa (38.9%) regions (Table 5). Besides, currently, only one population presented genome-wide SNP data available from the Gold Coast region, the Brong population from modern Ghana (Bryc et al., 2010a), and its sample size is considerably low (N= 4) to establish that genetic link. Therefore, the study does not clarify the fact that the African origins of the Afro-Colombians may be strongly associated with the Gold Coast region or that their African ancestry proportions could be a combination of separate African regions (Table 28). Both, the African-American populations from Colombia and Brazil are deeply admixed compared to the Noir Marron communities (Figure 77), and they do not present a clear association with one African region in the multidimensional space (PCA plot not included). Their sub-continental ancestry could be identified by using the PCA based on local ancestry methods, such as ancestry-specific PCA (or ASPCA) (Moreno-Estrada et al., 2013).

Currently, it is increasingly common for the descendants of the enslaved Africans to want to contextualize their ancestry, by understanding more about who their ancestors were, where they came from, what conditions initiated their movements within the African continent and their forced migrations to the Americas and what genetic events they may have experienced. Each of these inquiries is ultimately designed to deduce the health status, ancestral backgrounds, and potential disease susceptibilities of their contemporary descendants (Jackson & Borgelin, 2010).

The popularity of ancestry and genealogical research has grown rapidly over the past ten years. In the U.S., genealogical research has become the fastest growing hobby in many communities. The burgeoning number of genetics companies that now offer fee-for-service tests for genetic ancestry, such as Ancestry.com, Africanancestry.com, 23andMe.com or Oxfordancestors.com (see more details in http://www.isogg.org/wiki/Autosomal_DNA_testing_comparison_chart), indicates the desire of many African Americans to identify their ancestral origins (Shriver & Kittles, 2004). These companies have enjoyed profits by connecting regional groups of African Americans mainly from the U.S. with specific areas of Africa. However, they do not provide accurate estimations about one's African roots for particular customers. Caution is warranted when testing companies tracing African ancestry using a small set of uniparental markers (Bandelt et al., 2008). Ancestry tests can give some clues about African genetic inheritance, however, they cannot be sure about West African ancestry, because of the important caveat that the modern reference populations might not be the same as the historical populations who lived in the same locations at the time of the Atlantic slave trade. For instance, Ancestry.com makes determinations of African affiliations based on six ethnicity regions of West African ancestry including: *i*) Senegal, *ii*) Mali, *iii*) Ivory Coast/Ghana, *iv*) Benin/Togo, *v*) Nigeria, and *vi*) Cameroon/Congo (Figure 81) (Granka, 2013).

However, this division of the West Africa groups into modern countries disagrees with the historical information of African geographic distribution during the period when the slave trade flourished (Figures 5 and 7). The social implications of genetic ancestry tests highlight the need for a better picture of West African regions that correspond with the available historical information (Figure 82).
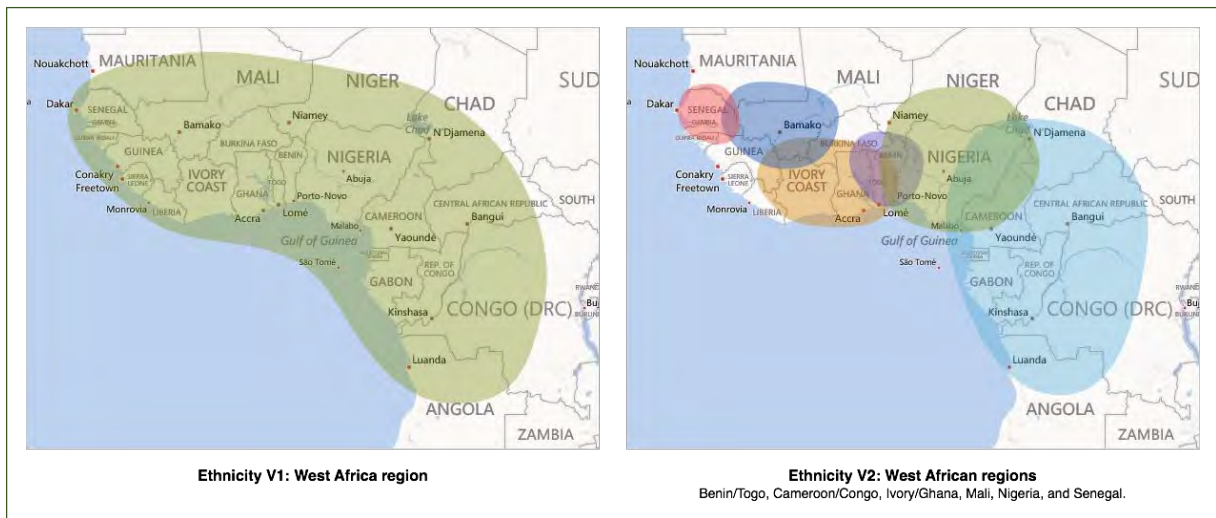
Figure 81: Finer-resolution genetic ethnicity estimate for individuals with West African ancestry (Ethnicity V1). This company analyse the likely African origin from six West African regions (Ethnicity V2) (Granka, 2013). © 2013 Ancestry.com™.
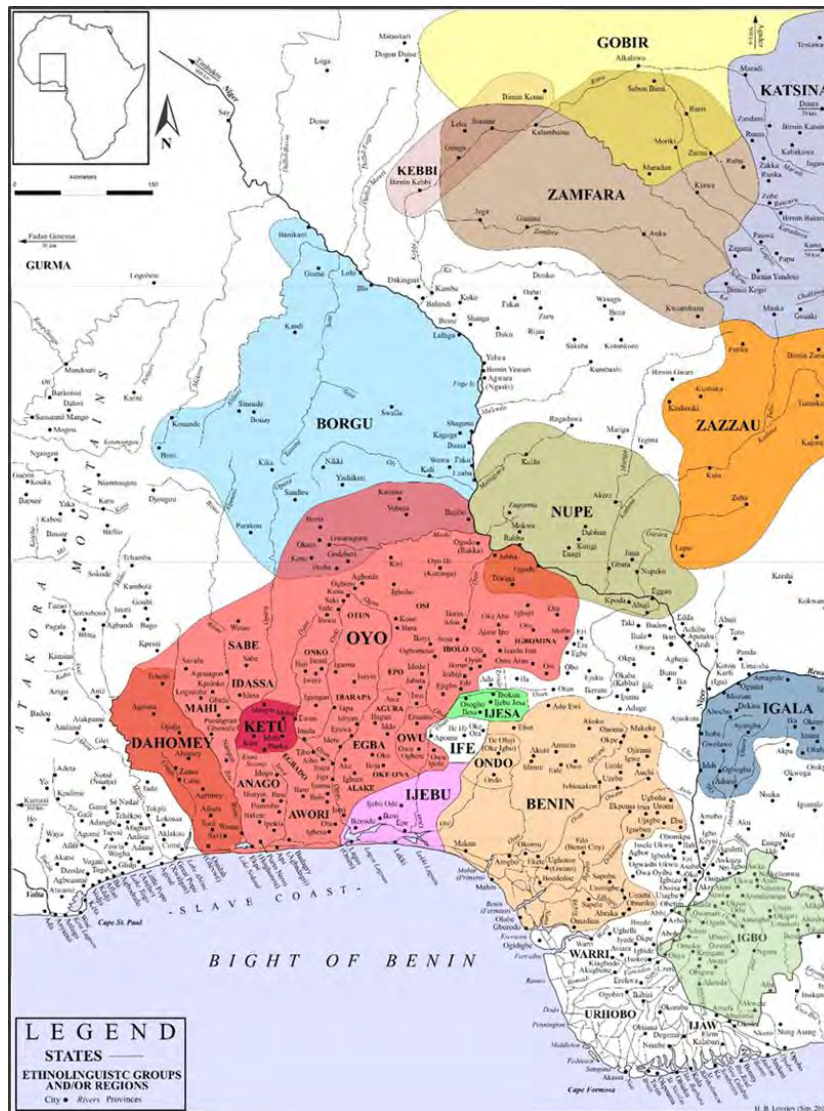
Figure 82: Map of historic West African kingdoms located in the Bight of Benin region during the 17ᵗʰ and the 18ᵗʰ century (Lovejoy, 2013).

Therefore, the most reliable method for inferring ancestry is to combine DNA evidence with other forms of historical and linguistic data. In the analysis of a large representation of African populations with different cultural and linguistic backgrounds, we compared the Noir Marron communities with African populations that were representative of historical African kingdoms during the period of the slave trade. For instance, in the Bight of Benin region, the most likely direct descendants of the African captives were include in the genetic analyses. Therefore, the West African populations residing in Benin and Nigeria were considered for the ADMIXTURE and Eigenstrat PCA analyses because of they are located in the same geographic region than their ancestors during the period of the slave trade (Figure 82), such as the Fon, the descendants for the Dahomey Kingdom; the Bariba, the descendants for the Borgu Kingdom; the Beninese Yoruba, the descendants for Ketu Kingdom; the Nigerian Yoruba, the descendants for the Oyo Kingdom; the Esan, the descendants for the Benin Kingdom; and the Igbo, the descendants for the Igbo Kingdom (Law, 1991, 2004; Lovejoy, 2013; Lovejoy, 2011). In some way, we explore the genetic links between the West African descendants and the African descendants living in French Guiana and Surinam today.

Indeed, the Eigenstrat PCA analysis suggested a common ancestry for the Noir Marron communities and the populations having descended from the Dahomey, Borgu, and Ketu Kingdoms localized in modern Benin. In good agreement with linguistic studies, it was suggested close structural and functional similarities between the *Saamaka* (Creole spoke for Saramaka people) and *Fongbe*, the language spoken by the Fon population in modern Benin (Essegbey et al., 2013a; Essegbey et al., 2013b; Migge & Winford, 2013). These findings lend further support to the ethnic origins of enslaved African descendants deported to French Guiana and Surinam around four centuries ago.

V. CONCLUSIONS

i. This study provides a strong characterization of the genetic diversities within African-American populations residing in South America, particularly in four Noir Marron communities: Aluku, Paramaka, Ndjuka, and Saramaka.

ii. The Noir Marron communities presented the highest African ancestry in all genetic systems analysed: 89.11% for Y-STRs, 86.03% for Y-SNPs, 97.80% for mtDNA, and 98.34% for genome-wide SNP data. These values are considerably higher than in other African-American populations in the Americas.

iii. Uniparental markers revealed a significant degree of sex-biases. According to Y chromosome, between Afro-Brazilian and Afro-Colombian women (macrohaplogroup L) and European men (haplogroup R1b1a2-M269). According to mtDNA, between Afro-Colombian men (haplogroup E1b1a1) and Native American women (haplogroups: A2, B2d, and C1d). This fact highlighted asymmetric mating patterns involving predominantly European men and African women within populations with different colonial pasts (that means Spanish and Portuguese).

iv. Genome-wide SNP data of African ancestry had outlined a new understanding of historical African coastal regions where Africans slaves were kidnaped several centuries ago. ADMIXTURE and PCA analyses allowed the reconstruction of genetic links in Noir Marron communities with historic populations residing today in Bight of Benin region.

v. The new findings linked the African-American genetic diversities with historical African coastal regions, and suggested different slave trade routes in the Noir Marron associated with the Bight of Benin, in the Afro-Brazilians with the West Central Africa, and in the Afro-Colombians with the Gold Coast. These genetic links are in agreement with their colonial historic records, with the exception of the Afro-Colombian association.

vi. Overall, the present study contributed to better understanding of unique African legacy among the Noir Marron communities, and highlighted that within an appropriate historical framework, genetic ancestry analyses add further understanding of ethnicity in African populations throughout the Atlantic world.

VI. BIBLIOGRAPHY

1000 Genomes Project Consortium, Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T. & McVean, G.A. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature,* 491(7422); 56-65.

1000 Genomes Project Consortium (2015). A global reference for human genetic variation. *Nature,* 526(7571); 68-74.

Abe-Sandes, K., Silva, W.A., Jr. & Zago, M.A. (2004). Heterogeneity of the Y chromosome in Afro-Brazilian populations. *Hum Biol,* 76(1); 77-86.

Achilli, A., Perego, U.A., Lancioni, H., Olivieri, A., Gandini, F., Hooshiar Kashani, B., Battaglia, V., Grugni, V., Angerhofer, N., Rogers, M.P., Herrera, R.J., Woodward, S.R., Labuda, D., Smith, D.G., Cybulski, J.S., Semino, O., Malhi, R.S. & Torroni, A. (2013). Reconciling migration models to the Americas with the variation of North American native mitogenomes. *Proc Natl Acad Sci U S A,* 110(35); 14308-14313.

Acosta, M.A., Blanco-Verea, A., Lareu, M.V., Brion, M. & Carracedo, A. (2009). The genetic male component of two South-Western Colombian populations. *Forensic Sci Int Genet,* 3(2); e59-61.

Agorsah, E.K. (1994). *Maroon Heritage: Archaeological, Ethnographic, and Historical Perspectives.* Kingston, Jamaica: Canoe Press, University of the West Indies.

Alexander, D.H., Novembre, J. & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res,* 19(9); 1655-1664.

Alexander, D.H. & Lange, K. (2011). Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics,* 12; 246.

Allard, M.W., Polanskey, D., Miller, K., Wilson, M.R., Monson, K.L. & Budowle, B. (2005). Characterization of human control region sequences of the African American SWGDAM forensic mtDNA data set. *Forensic Sci Int,* 148(2-3); 169-179.

Allentoft, M.E., Sikora, M., Sjogren, K.-G., Rasmussen, S., Rasmussen, M., Stenderup, J., Damgaard, P.B., Schroeder, H., Ahlstrom, T., Vinner, L., Malaspinas, A.-S., Margaryan, A., Higham, T., Chivall, D., Lynnerup, N., Harvig, L., Baron, J., Casa, P.D., Dabrowski, P., Duffy, P.R., Ebel, A.V., Epimakhov, A., Frei, K., Furmanek, M., Gralak, T., Gromov, A., Gronkiewicz, S., Grupe, G., Hajdu, T., Jarysz, R., et al. (2015). Population genomics of Bronze Age Eurasia. *Nature,* 522(7555); 167-172.

Almeida, A. (2008). The Foundation of the System: A Reassessment of the Slave Trade to the Spanish Americas in the Sixteenth and Seventeenth Centuries. In: *Extending the Frontiers: Essays on the New Transatlantic Slave Trade Database.* Eltis, D. & Richardson, D. (eds.). New Haven, USA: Yale University Press.

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

Ambrosio, B., Dugoujon, J.M., Hernandez, C., De La Fuente, D., Gonzalez-Martin, A., Fortes-Lima, C.A., Novelletto, A., Rodriguez, J.N. & Calderon, R. (2010). The Andalusian population from Huelva reveals a high diversification of Y-DNA paternal lineages from haplogroup E: Identifying human male movements within the Mediterranean space. *Ann Hum Biol,* 37(1); 86-107.

Ambrosio, B., Novelletto, A., Hernandez, C., Dugoujon, J.M., Fortes-Lima, C., Rodriguez, J.N. & Calderon, R. (2012). Y-STR genetic diversity in autochthonous Andalusians from Huelva and Granada provinces (Spain). *Forensic Sci Int Genet,* 6(2); e66-71.

Anaya, S.J. (2004). *Indigenous Peoples in International Law.* New York, USA: Oxford University Press, Inc.

Anderson, C.A., Pettersson, F.H., Clarke, G.M., Cardon, L.R., Morris, A.P. & Zondervan, K.T. (2010). Data quality control in genetic case-control association studies. *Nat Protoc,* 5(9); 1564-1573.

Anderson, R.N. (1996). The Quilombo of Palmares: A New Overview of a Maroon State in Seventeenth-Century Brazil. *Journal of Latin American Studies* 28; 545-566.

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., Schreier, P.H., Smith, A.J., Staden, R. & Young, I.G. (1981). Sequence and organization of the human mitochondrial genome. *Nature,* 290(5806); 457-465.

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

Ansari Pour, N., Plaster, C.A. & Bradman, N. (2013). Evidence from Y-chromosome analysis for a late exclusively eastern expansion of the Bantu-speaking people. *Eur J Hum Genet,* 21(4); 423-429.

Araujo, A.L. (2014). *Shadows of the Slave Past: Memory, Heritage, and Slavery.* London, UK: Routledge. Routledge Studies in Cultural History.

Araujo, A.L. (2015). *African Heritage and Memories of Slavery in Brazil and the South Atlantic World.* New York: Cambria Press.

Arends, J. (1995). Demographic factors in the formation of Sranan. In: *The early stages of creolization.* Arends, J. (ed. Amsterdam: John Benjamins Publishing Company.

Arends, J. (2001). Social stratification and network relations in the formation of Sranan. In: *Creolization and Contact.* Smith, N. & Veenstra, T. (eds.). Amsterdam: John Benjamins.

Arredi, B., Poloni, E.S., Paracchini, S., Zerjal, T., Fathallah, D.M., Makrelouf, M., Pascali, V.L., Novelletto, A. & Tyler-Smith, C. (2004). A predominantly neolithic origin for Y-chromosomal DNA variation in North Africa. *Am J Hum Genet,* 75(2); 338-345.

Atkinson, Q.D., Gray, R.D. & Drummond, A.J. (2009). Bayesian coalescent inference of major human mitochondrial DNA haplogroup expansions in Africa. *Proc Biol Sci,* 276(1655); 367-373.

Bachtrog, D. (2013). Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nat Rev Genet,* 14(2); 113-124.

Baker, P. (1999). Investigating the origin and diffusion of shared features among the Atlantic English Creoles. In: *St. Kitts and the Atlantic Creoles. The text of Samuel Augustus Mathews in Perspective.* Baker, P. & Bruyn, A. (eds.). London: University of Westminster Press.

Baker, P. & Huber, M. (2001). Atlantic, Pacific, and world-wide features in English-lexicon contact languages. *English World-Wide,* 22; 157–208.

Bakker, P., Daval-Markussen, A., Parkvall, M. & Plag, I. (2011). Creoles are typologically distinct from non-creoles. *Journal of Pidgin and Creole Languages,* 26(1); 5-42.

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

Bales, K., Trodd, Z. & Williamson, A.K. (2009). *Modern Slavery: The Secret World of 27 Million People.* London: Oneworld Publications.

Ballantyne, K.N., Goedbloed, M., Fang, R., Schaap, O., Lao, O., Wollstein, A., Choi, Y., van Duijn, K., Vermeulen, M., Brauer, S., Decorte, R., Poetsch, M., von Wurmb-Schwark, N., de Knijff, P., Labuda, D., Vezina, H., Knoblauch, H., Lessig, R., Roewer, L., Ploski, R., Dobosz, T., Henke, L., Henke, J., Furtado, M.R. & Kayser, M. (2010). Mutability of Y-chromosomal microsatellites: rates, characteristics, molecular bases, and forensic implications. *Am J Hum Genet,* 87(3); 341-353.

Ballantyne, K.N., Keerl, V., Wollstein, A., Choi, Y., Zuniga, S.B., Ralf, A., Vermeulen, M., de Knijff, P. & Kayser, M. (2012). A new future of forensic Y-chromosome analysis: rapidly mutating Y-STRs for differentiating male relatives and paternal lineages. *Forensic Sci Int Genet,* 6(2); 208-218.

Ballantyne, K.N., Ralf, A., Aboukhalid, R., Achakzai, N.M., Anjos, M.J., Ayub, Q., Balazic, J., Ballantyne, J., Ballard, D.J., Berger, B., Bobillo, C., Bouabdellah, M., Burri, H., Capal, T., Caratti, S., Cardenas, J., Cartault, F., Carvalho, E.F., Carvalho, M., Cheng, B., Coble, M.D., Comas, D., Corach, D., D'Amato, M.E., Davison, S., de Knijff, P., De Ungria, M.C., Decorte, R., Dobosz, T., Dupuy, B.M., et al. (2014). Toward male individualization with rapidly mutating y-chromosomal short tandem repeats. *Hum Mutat,* 35(8); 1021-1032.

Bandelt, H.J. & Dress, A.W. (1992). Split decomposition: a new and useful approach to phylogenetic analysis of distance data. *Mol Phylogenet Evol,* 1(3); 242-252.

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

Bandelt, H.J., Forster, P. & Rohl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol,* 16(1); 37-48.

Bandelt, H.J., Richards, M. & Macaulay, V. (2006). *Human Mitochondrial DNA and the Evolution of Homo sapiens.* Heidelberg, Germany: Springer-Verlag Berlin Heidelberg. Nucleic Acids and Molecular Biology.

Bandelt, H.J., Yao, Y.G., Richards, M.B. & Salas, A. (2008). The brave new era of human genetic testing. *Bioessays,* 30(11-12); 1246-1251.

Barbieri, C., Whitten, M., Beyer, K., Schreiber, H., Li, M. & Pakendorf, B. (2012). Contrasting maternal and paternal histories in the linguistic context of Burkina Faso. *Mol Biol Evol,* 29(4); 1213-1223.

Barbieri, C., Butthof, A., Bostoen, K. & Pakendorf, B. (2013a). Genetic perspectives on the origin of clicks in Bantu languages from southwestern Zambia. *Eur J Hum Genet,* 21(4); 430-436.

Barbieri, C., Vicente, M., Rocha, J., Mpoloka, S.W., Stoneking, M. & Pakendorf, B. (2013b). Ancient substructure in early mtDNA lineages of southern Africa. *Am J Hum Genet,* 92(2); 285-292.

Barbieri, C., Guldemann, T., Naumann, C., Gerlach, L., Berthold, F., Nakagawa, H., Mpoloka, S.W., Stoneking, M. & Pakendorf, B. (2014a). Unraveling the complex maternal history of Southern African Khoisan populations. *Am J Phys Anthropol,* 153(3); 435-448.

Barbieri, C., Vicente, M., Oliveira, S., Bostoen, K., Rocha, J., Stoneking, M. & Pakendorf, B. (2014b). Migration and interaction in a contact zone: mtDNA variation among Bantu-speakers in Southern Africa. *PLoS One,* 9(6); e99117.

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

Batini, C., Lopes, J., Behar, D.M., Calafell, F., Jorde, L.B., van der Veen, L., Quintana-Murci, L., Spedini, G., Destro-Bisol, G. & Comas, D. (2011b). Insights into the demographic history of African Pygmies from complete mitochondrial genomes. *Mol Biol Evol,* 28(2); 1099-1110.

Battaglia, V., Grugni, V., Perego, U.A., Angerhofer, N., Gomez-Palmieri, J.E., Woodward, S.R., Achilli, A., Myres, N., Torroni, A. & Semino, O. (2013). The first peopling of South America: new evidence from Y-chromosome haplogroup Q. *PLoS One,* 8(8); e71390.

Becquet, C., Patterson, N., Stone, A.C., Przeworski, M. & Reich, D. (2007). Genetic structure of chimpanzee populations. *PLoS Genet,* 3(4); e66.

Bedoya, G., Montoya, P., Garcia, J., Soto, I., Bourgeois, S., Carvajal, L., Labuda, D., Alvarez, V., Ospina, J., Hedrick, P.W. & Ruiz-Linares, A. (2006). Admixture dynamics in Hispanics: a shift in the nuclear genetic ancestry of a South American population isolate. *Proc Natl Acad Sci U S A,* 103(19); 7234-7239.

Behar, D.M., Villems, R., Soodyall, H., Blue-Smith, J., Pereira, L., Metspalu, E., Scozzari, R., Makkan, H., Tzur, S., Comas, D., Bertranpetit, J., Quintana-Murci, L., Tyler-Smith, C., Wells, R.S., Rosset, S. & Genographic, C. (2008). The dawn of human matrilineal diversity. *Am J Hum Genet,* 82(5); 1130-1140.

Behar, D.M., van Oven, M., Rosset, S., Metspalu, M., Loogvali, E.L., Silva, N.M., Kivisild, T., Torroni, A. & Villems, R. (2012). A "Copernican" reassessment of the human mitochondrial DNA tree from its root. *Am J Hum Genet,* 90(4); 675-684.

Beleza, S., Gusmao, L., Lopes, A., Alves, C., Gomes, I., Giouzeli, M., Calafell, F., Carracedo, A. & Amorim, A. (2006). Micro-phylogeographic and demographic history of Portuguese male lineages. *Ann Hum Genet,* 70(Pt 2); 181-194.

Belkacemi, N. (1999). Les autochtones français : populations ou peuples ? *Droit et cultures,* 37(1); 25-52.

Ben Halim, N., Nagara, M., Regnault, B., Hsouna, S., Lasram, K., Kefi, R., Azaiez, H., Khemira, L., Saidane, R., Ammar, S.B., Besbes, G., Weil, D., Petit, C., Abdelhak, S. & Romdhane, L. (2015). Estimation of Recent and Ancient Inbreeding in a Small Endogamous Tunisian Community Through Genomic Runs of Homozygosity. *Ann Hum Genet*.

Benn-Torres, J., Bonilla, C., Robbins, C.M., Waterman, L., Moses, T.Y., Hernandez, W., Santos, E.R., Bennett, F., Aiken, W., Tullock, T., Coard, K., Hennis, A., Wu, S., Nemesure, B., Leske, M.C., Freeman, V., Carpten, J. & Kittles, R.A. (2008). Admixture and population stratification in African Caribbean populations. *Ann Hum Genet,* 72(Pt 1); 90-98.

Benn Torres, J., Kittles, R.A. & Stone, A.C. (2007). Mitochondrial and Y chromosome diversity in the English-speaking Caribbean. *Ann Hum Genet,* 71(Pt 6); 782-790.

Benson, D.A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J. & Sayers, E.W. (2013). GenBank. *Nucleic Acids Res,* 41(Database issue); D36-42.

Bernardo, S., Hermida, R., Desiderio, M., Silva, D.A. & de Carvalho, E.F. (2014). MtDNA ancestry of Rio de Janeiro population, Brazil. *Mol Biol Rep,* 41(4); 1945-1950.

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

Bertorelle, G. & Excoffier, L. (1998). Inferring admixture proportions from molecular data. *Mol Biol Evol,* 15(10); 1298-1311.

Bilby, K. (1995). Introducing the Popular Music of Suriname. In: *Caribbean Currents: Caribbean Music from Rumba to Reggae*. Manuel, P., Bilby, K. & Largey, M. (eds.). Philadelphia, USA: Temple University Press.

Bilby, K. (2001). Aleke: New Music and New Identities in the Guianas. *Latin American Music Review,* 22(1); 31-47.

Bilby, K. (2005). *True-Born Maroons.* Gainesville, USA: University Press of Florida.

Bilby, K. (2009). African American Memory at the Crossroads: Grounding the Miraculous with Tooy. *Small Axe,* 13(2); 185-199.

Bisso-Machado, R., Jota, M.S., Ramallo, V., Paixao-Cortes, V.R., Lacerda, D.R., Salzano, F.M., Bonatto, S.L., Santos, F.R. & Bortolini, M.C. (2011). Distribution of Y-chromosome Q lineages in Native Americans. *Am J Hum Biol,* 23(4); 563-566.

Bortolini, M.C., Weimer Tde, A., Salzano, F.M., Callegari-Jacques, S.M., Schneider, H., Layrisse, Z. & Bonatto, S.L. (1995). Evolutionary relationships between black South American and African populations. *Hum Biol,* 67(4); 547-559.

Bortolini, M.C., Salzano, F.M., Zago, M.A., Da Silva Junior, W.A. & Weimer Tde, A. (1997a). Genetic variability in two Brazilian ethnic groups: a comparison of mitochondrial and protein data. *Am J Phys Anthropol,* 103(2); 147-156.

Bortolini, M.C., Zago, M.A., Salzano, F.M., Silva-Junior, W.A., Bonatto, S.L., da Silva, M.C. & Weimer, T.A. (1997b). Evolutionary and anthropological implications of mitochondrial DNA variation in African Brazilian populations. *Hum Biol,* 69(2); 141-159.

Bortolini, M.C., Da Silva, W.A.J.W.A., De Guerra, D.C., Remonatto, G., Mirandola, R., Hutz, M.H., Weimer, T.A., Silva, M.C., Zago, M.A. & Salzano, F.M. (1999). African-derived South American populations: A history of symmetrical and asymmetrical matings according to sex revealed by bi- and uni-parental genetic markers. *Am J Hum Biol,* 11(4); 551-563.

Botigue, L.R., Henn, B.M., Gravel, S., Maples, B.K., Gignoux, C.R., Corona, E., Atzmon, G., Burns, E., Ostrer, H., Flores, C., Bertranpetit, J., Comas, D. & Bustamante, C.D. (2013). Gene flow from North Africa contributes to differential human genetic diversity in southern Europe. *Proc Natl Acad Sci U S A,* 110(29); 11791-11796.

Bourgeois, S., Yotova, V., Wang, S., Bourtoumieu, S., Moreau, C., Michalski, R., Moisan, J.P., Hill, K., Hurtado, A.M., Ruiz-Linares, A. & Labuda, D. (2009). X-chromosome lineages and the settlement of the Americas. *Am J Phys Anthropol,* 140(3); 417-428.

Bravi, C.M., Sans, M., Bailliet, G., Martinez-Marignac, V.L., Portas, M., Barreto, I., Bonilla, C. & Bianchi, N.O. (1997). Characterization of mitochondrial DNA and Y-chromosome haplotypes in a Uruguayan population of African ancestry. *Hum Biol,* 69(5); 641-652.

Bravo, M.L., Moreno, M.A., Builes, J.J., Salas, A., Lareu, M.V. & Carracedo, A. (2001). Autosomal STR genetic variation in negroid Choco and Bogota populations. *Int J Legal Med,* 115(2); 102-104.

Brion Davis, D. (1999). *The Problem of Slavery in the Age of Revolution, 1770-1823.* Oxford, USA: Oxford University Press.

Brisbin, A., Bryc, K., Byrnes, J., Zakharia, F., Omberg, L., Degenhardt, J., Reynolds, A., Ostrer, H., Mezey, J.G. & Bustamante, C.D. (2012). PCAdmix: principal components-based assignment of ancestry along each chromosome in individuals with admixed ancestry from two or more populations. *Hum Biol,* 84(4); 343-364.

Broman, K.W. & Weber, J.L. (1999). Long homozygous chromosomal segments in reference families from the centre d'Etude du polymorphisme humain. *Am J Hum Genet,* 65(6); 1493-1500.

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

Brucato, N., Tortevoye, P., Plancoulaine, S., Guitard, E., Sanchez-Mazas, A., Larrouy, G., Gessain, A. & 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(10); 917-926.

Brucato, N., Cassar, O., Tonasso, L., Tortevoye, P., Migot-Nabias, F., Plancoulaine, S., Guitard, E., Larrouy, G., Gessain, A. & 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.

Brucato, N., Mazieres, S., Guitard, E., Giscard, P.H., Bois, E., Larrouy, G. & Dugoujon, J.M. (2012). The Hmong Diaspora: preserved South-East Asian genetic ancestry in French Guianese Asians. *C R Biol,* 335(10-11); 698-707.

Bryant, S.K., O'Toole, R.S. & Vinson, B. (2012). *Africans to Spanish America: Expanding the Diaspora.* Springfield: University of Illinois Press.

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. & Bustamante, C.D. (2010a). Genome-wide patterns of population structure and admixture in West Africans and African Americans. *Proc Natl Acad Sci U S A,* 107(2); 786-791.

Bryc, K., Velez, C., Karafet, T., Moreno-Estrada, A., Reynolds, A., Auton, A., Hammer, M., Bustamante, C.D. & Ostrer, H. (2010b). Colloquium paper: genome-wide patterns of population structure and admixture among Hispanic/Latino populations. *Proc Natl Acad Sci U S A,* 107 Suppl 2; 8954-8961.

Bryc, K., Durand, E.Y., Macpherson, J.M., Reich, D. & Mountain, J.L. (2015). The genetic ancestry of African Americans, Latinos, and European Americans across the United States. *Am J Hum Genet,* 96(1); 37-53.

Bush, W.S. & Moore, J.H. (2012). Chapter 11: Genome-wide association studies. *PLoS Comput Biol,* 8(12); e1002822.

Butler, J.M. (2005). *Forensic DNA Typing: Biology, Technology, and Genetics of STR Markers.* Burlington, USA: Elsevier Academic Press. Forensic DNA Typing Series.

Butler, J.M., Hill, C.R. & Coble, M.D. (2012). Variability of new STR loci and kits in US population groups. Promega Corporation Website; http://france.promega.com/resources/profiles-in-dna/2012/variability-of-new-str-loci-and-kits-in-us-population-groups/.

Callegari-Jacques, S.M., Grattapaglia, D., Salzano, F.M., Salamoni, S.P., Crossetti, S.G., Ferreira, M.E. & Hutz, M.H. (2003). Historical genetics: spatiotemporal analysis of the formation of the Brazilian population. *Am J Hum Biol,* 15(6); 824-834.

Campbell, C.J. (2015). Making Abolition Brazilian: British Law and Brazilian Abolitionists in Nineteenth-Century Minas Gerais and Pernambuco. *Slavery & Abolition,* 36(3); 521-543.

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

Campbell, M.C., Hirbo, J.B., Townsend, J.P. & Tishkoff, S.A. (2014). The peopling of the African continent and the diaspora into the new world. *Curr Opin Genet Dev,* 29; 120-132.

Cann, H.M., de Toma, C., Cazes, L., Legrand, M.F., Morel, V., Piouffre, L., Bodmer, J., Bodmer, W.F., Bonne-Tamir, B., Cambon-Thomsen, A., Chen, Z., Chu, J., Carcassi, C., Contu, L., Du, R., Excoffier, L., Ferrara, G.B., Friedlaender, J.S., Groot, H., Gurwitz, D., Jenkins, T., Herrera, R.J., Huang, X., Kidd, J., Kidd, K.K., Langaney, A., Lin, A.A., Mehdi, S.Q., Parham, P., Piazza, A., et al. (2002). A human genome diversity cell line panel. *Science,* 296(5566); 261-262.

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

Carvajal-Carmona, L.G., Soto, I.D., Pineda, N., Ortiz-Barrientos, D., Duque, C., Ospina-Duque, J., McCarthy, M., Montoya, P., Alvarez, V.M., Bedoya, G. & Ruiz-Linares, A. (2000). Strong Amerind/white sex bias and a possible Sephardic contribution among the founders of a population in northwest Colombia. *Am J Hum Genet,* 67(5); 1287-1295.

Carvajal-Carmona, L.G., Ophoff, R., Service, S., Hartiala, J., Molina, J., Leon, P., Ospina, J., Bedoya, G., Freimer, N. & Ruiz-Linares, A. (2003). Genetic demography of Antioquia (Colombia) and the Central Valley of Costa Rica. *Hum Genet,* 112(5-6); 534-541.

Carvalho-Silva, D.R., Santos, F.R., Rocha, J. & Pena, S.D. (2001). The phylogeography of Brazilian Y-chromosome lineages. *Am J Hum Genet,* 68(1); 281-286.

Casas-Vargas, A., Gomez, A., Briceno, I., Diaz-Matallana, M., Bernal, J.E. & Rodriguez, J.V. (2011). High genetic diversity on a sample of pre-Columbian bone remains from Guane territories in northwestern Colombia. *Am J Phys Anthropol,* 146(4); 637-649.

Cavalli-Sforza, L.L., Wilson, A.C., Cantor, C.R., Cook-Deegan, R.M. & King, M.C. (1991). Call for a worldwide survey of human genetic diversity: a vanishing opportunity for the Human Genome Project. *Genomics,* 11(2); 490-491.

Cavalli-Sforza, L.L., Menozzi, P. & Piazza, A. (1994). *The History and Geography of Human Genes.* Princeton, USA: PrincetonUniversityPress.

Cavalli-Sforza, L.L. & Feldman, M.W. (2003). The application of molecular genetic approaches to the study of human evolution. *Nat Genet,* 33 Suppl; 266-275.

Cerezo, M., Achilli, A., Olivieri, A., Perego, U.A., Gomez-Carballa, A., Brisighelli, F., Lancioni, H., Woodward, S.R., Lopez-Soto, M., Carracedo, A., Capelli, C., Torroni, A. & Salas, A. (2012). Reconstructing ancient mitochondrial DNA links between Africa and Europe. *Genome Res,* 22(5); 821-826.

Cerny, V., Hajek, M., Bromova, M., Cmejla, R., Diallo, I. & Brdicka, R. (2006). MtDNA of Fulani nomads and their genetic relationships to neighboring sedentary populations. *Hum Biol,* 78(1); 9-27.

Cerny, V., Pereira, L., Musilova, E., Kujanova, M., Vasikova, A., Blasi, P., Garofalo, L., Soares, P., Diallo, I., Brdicka, R. & Novelletto, A. (2011). Genetic structure of pastoral and farmer populations in the African Sahel. *Mol Biol Evol,* 28(9); 2491-2500.

Chan, E.K., Hardie, R.A., Petersen, D.C., Beeson, K., Bornman, R.M., Smith, A.B. & Hayes, V.M. (2015). Revised timeline and distribution of the earliest diverged human maternal lineages in southern Africa. *PLoS One,* 10(3); e0121223.

Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M. & Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience,* 4; 7.

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

Childs, G. (2007). The history of smallpox and its spread around the world. In: *Encyclopedia of Plague and Pestilence: From Ancient Times to the Present.* New York, USA: Infobase Publishing.

Chiles, J.-P. & Delfiner, P. (2008). Kriging. In: *Geostatistics.* Hoboken, NJ, USA: John Wiley & Sons, Inc.

Cifuentes, L., Morales, R., Sepulveda, D., Jorquera, H. & Acuna, M. (2004). DYS19 and DYS199 loci in a Chilean population of mixed ancestry. *Am J Phys Anthropol,* 125(1); 85-89.

Clark, A.G. (1999). The size distribution of homozygous segments in the human genome. *Am J Hum Genet,* 65(6); 1489-1492.

Coelho, M., Sequeira, F., Luiselli, D., Beleza, S. & 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.

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

Corach, D., Lao, O., Bobillo, C., van Der Gaag, K., Zuniga, S., Vermeulen, M., van Duijn, K., Goedbloed, M., Vallone, P.M., Parson, W., de Knijff, P. & Kayser, M. (2010). Inferring continental ancestry of argentineans from Autosomal, Y-chromosomal and mitochondrial DNA. *Ann Hum Genet,* 74(1); 65-76.

Costa, M.D., Pereira, J.B., Pala, M., Fernandes, V., Olivieri, A., Achilli, A., Perego, U.A., Rychkov, S., Naumova, O., Hatina, J., Woodward, S.R., Eng, K.K., Macaulay, V., Carr, M., Soares, P., Pereira, L. & Richards, M.B. (2013). A substantial prehistoric European ancestry amongst Ashkenazi maternal lineages. *Nat Commun,* 4; 2543.

Cotrim, N.H., Auricchio, M.T., Vicente, J.P., Otto, P.A. & Mingroni-Netto, R.C. (2004). Polymorphic Alu insertions in six Brazilian African-derived populations. *Am J Hum Biol,* 16(3); 264-277.

Cruciani, F., Trombetta, B., Sellitto, D., Massaia, A., Destro-Bisol, G., Watson, E., Beraud Colomb, E., Dugoujon, J.M., Moral, P. & Scozzari, R. (2010). Human Y chromosome haplogroup R-V88: a paternal genetic record of early mid Holocene trans-Saharan connections and the spread of Chadic languages. *Eur J Hum Genet,* 18(7); 800-807.

Cruciani, F., Trombetta, B., Massaia, A., Destro-Bisol, G., Sellitto, D. & 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(6); 814-818.

Currie, T.E., Meade, A., Guillon, M. & Mace, R. (2013). Cultural phylogeography of the Bantu Languages of sub-Saharan Africa. *Proc Biol Sci,* 280(1762); 20130695.

Da Silva, W.A., Jr., Bortolini, M.C., Meyer, D., Salzano, F.M., Elion, J., Krishnamoorthy, R., Schneider, M.P., De Guerra, D.C., Layrisse, Z., Castellano, H.M., Weimer, T.D. & Zago, M.A. (1999). Genetic diversity of two African and sixteen South American populations determined on the basis of six hypervariable loci. *Am J Phys Anthropol,* 109(4); 425-437.

DANE (2005). Censo General 2005. Departamento Administrativo Nacional de Estadística (DANE). Website; http://www.dane.gov.co/index.php/es/poblacion-y-registros-vitales/censos/censo-2005.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., McVean, G., Durbin, R. & 1000 Genomes Project Analysis Group (2011). The variant call format and VCFtools. *Bioinformatics,* 27(15); 2156-2158.

Daval-Markussen, A. & Bakker, P. (2011). A phylogenetic networks approach to the classification of English-based Atlantic creoles. *English World-Wide,* 32(2); 115-146.

Davis, D.B. (2006). *Inhuman Bondage: The Rise and Fall of Slavery in the New World.* New York: Oxford University Press.

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

de Filippo, C., Bostoen, K., Stoneking, M. & Pakendorf, B. (2012). Bringing together linguistic and genetic evidence to test the Bantu expansion. *Proc Biol Sci,* 279(1741); 3256-3263.

de Knijff, P. (2000). Messages through bottlenecks: on the combined use of slow and fast evolving polymorphic markers on the human Y chromosome. *Am J Hum Genet,* 67(5); 1055-1061.

de Wit, E., Delport, W., Rugamika, C.E., Meintjes, A., Moller, M., van Helden, P.D., Seoighe, C. & Hoal, E.G. (2010). Genome-wide analysis of the structure of the South African Coloured Population in the Western Cape. *Hum Genet,* 128(2); 145-153.

Deason, M.L., Salas, A., Newman, S.P., Macaulay, V.A., St, A.M.E.Y. & Pitsiladis, Y.P. (2012). Interdisciplinary approach to the demography of Jamaica. *BMC Evol Biol,* 12; 24.

Destro-Bisol, G., Donati, F., Coia, V., Boschi, I., Verginelli, F., Caglia, A., Tofanelli, S., Spedini, G. & Capelli, C. (2004). Variation of female and male lineages in sub-Saharan populations: the importance of sociocultural factors. *Mol Biol Evol,* 21(9); 1673-1682.

Di Rienzo, A., Peterson, A.C., Garza, J.C., Valdes, A.M., Slatkin, M. & Freimer, N.B. (1994). Mutational processes of simple-sequence repeat loci in human populations. *Proc Natl Acad Sci U S A,* 91(8); 3166-3170.

Diamond, J. & Bellwood, P. (2003). Farmers and their languages: the first expansions. *Science,* 300(5619); 597-603.

Diegoli, T.M. (2015). Forensic typing of short tandem repeat markers on the X and Y chromosomes. *Forensic Sci Int Genet*.

Dienekes' Anthropology blog (2013). World map of Y-chromosome haplogroups. Website; http://dienekes.blogspot.fr/2013/09/world-map-of-y-chromosome-haplogroups.html.

Domingues, P.M., Gusmao, L., da Silva, D.A., Amorim, A., Pereira, R.W. & de Carvalho, E.F. (2007). Sub-Saharan Africa descendents in Rio de Janeiro (Brazil): population and mutational data for 12 Y-STR loci. *Int J Legal Med,* 121(3); 238-241.

Dray, S. & Dufour, A.B. (2007). The ade4 package: Implementing the duality diagram for ecologists. *Journal of Statistical Software,* 22(4); 1-20.

Drescher, S. (2009). *Abolition: A History of Slavery and Antislavery.* Cambridge, UK: Cambridge University Press.

Dupanloup, I. & Bertorelle, G. (2001). Inferring admixture proportions from molecular data: extension to any number of parental populations. *Mol Biol Evol,* 18(4); 672-675.

Durand, E.Y., Patterson, N., Reich, D. & Slatkin, M. (2011). Testing for ancient admixture between closely related populations. *Mol Biol Evol,* 28(8); 2239-2252.

Ehret, C. (2001). Bantu Expansions: Re-Envisioning a Central Problem of Early African History. *The International Journal of African Historical Studies,* 34(1); 5-41.

Elhaik, E., Tatarinova, T., Chebotarev, D., Piras, I.S., Maria Calo, C., De Montis, A., Atzori, M., Marini, M., Tofanelli, S., Francalacci, P., Pagani, L., Tyler-Smith, C., Xue, Y., Cucca, F., Schurr, T.G., Gaieski, J.B., Melendez, C., Vilar, M.G., Owings, A.C., Gomez, R., Fujita, R., Santos, F.R., Comas, D., Balanovsky, O., Balanovska, E., Zalloua, P., Soodyall, H., Pitchappan, R., Ganeshprasad, A., Hammer, M., et al. (2014). Geographic population structure analysis of worldwide human populations infers their biogeographical origins. *Nat Commun,* 5; 3513.

Elson, J.L., Turnbull, D.M. & Howell, N. (2004). Comparative genomics and the evolution of human mitochondrial DNA: assessing the effects of selection. *Am J Hum Genet,* 74(2); 229-238.

Eltis, D., Behrendt, S.D., Richardson, D. & Klein, H.S. (1999). The trans-Atlantic slave trade a database on CD-ROM. Cambridge, UK: Cambridge University Press.

Eltis, D. & Richardson, D. (2008). Extending the frontiers essays on the new transatlantic slave trade database. New Haven, Conn. ; London: Yale University Press,.

Eltis, D. & Richardson, D. (2010). *Atlas of the transatlantic slave trade.* New Haven, Conn: Yale University Press. The Lewis Walpole series in eighteenth-century culture and history.

Eltis, D. & Richardson, D. (2013). *Routes to Slavery: Direction, Ethnicity and Mortality in the Transatlantic Slave Trade.* New York, USA: Routledge.

Ely, B., Wilson, J.L., Jackson, F. & Jackson, B.A. (2006). African-American mitochondrial DNAs often match mtDNAs found in multiple African ethnic groups. *BMC Biol,* 4; 34.

Essegbey, J., Migge, B. & Winford, D. (2013a). Cross-linguistic influence in language creation: Assessing the role of the Gbe languages in the formation of the Creoles of Suriname. *Lingua,* 129; 1-8.

Essegbey, J., van den Berg, M. & van de Vate, M. (2013b). Possibility and necessity modals in Gbe and Surinamese creoles. *Lingua,* 129; 67-95.

Excoffier, L., Smouse, P.E. & Quattro, J.M. (1992). Analysis of Molecular Variance Inferred from Metric Distances among DNA Haplotypes: Application to Human Mitochondrial DNA Restriction Data. *Genetics,* 131(2); 479-491.

Excoffier, L. & Lischer, H.E. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour,* 10(3); 564-567.

Falush, D., Stephens, M. & Pritchard, J.K. (2003). Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics,* 164(4); 1567-1587.

Fernandes, V., Triska, P., Pereira, J.B., Alshamali, F., Rito, T., Machado, A., Fajkosova, Z., Cavadas, B., Cerny, V., Soares, P., Richards, M.B. & Pereira, L. (2015). Genetic stratigraphy of key demographic events in Arabia. *PLoS One,* 10(3); e0118625.

Ferrari, L. (2012). San Basilio de Palenque (Colombia): un performance de libertad. In: *Palenque (Colombia): oralidad, identidad y resistencia. Un enfoque interdisciplinario.* Maglia, G. & Schwegler, A. (eds.). Bogota: Instituto Caro y Cuervo & Universidad Javeriana.

Fong, C., Lizarralde-Iragorri, M.A., Rojas-Gallardo, D. & Barreto, G. (2013). Frequency and origin of haplotypes associated with the beta-globin gene cluster in individuals with trait and sickle cell anemia in the Atlantic and Pacific coastal regions of Colombia. *Genet Mol Biol,* 36(4); 494-497.

Forster, P., Harding, R., Torroni, A. & Bandelt, H.J. (1996). Origin and evolution of Native American mtDNA variation: a reappraisal. *Am J Hum Genet,* 59(4); 935-945.

Fortes-Lima, C., Pereira, L., Stoneking, M. & Dugoujon, J.M. (2014). Hunting the oldest human mitochondrial DNA lineages in Africa. Conference: The African human fossil record. Toulouse. Conference Poster. Website; http://www.tahfr.cnrs.fr/wp-content/uploads/2014/09/TAHFR_symposium_program.pdf; 8-9.

Fortes-Lima, C., Brucato, N., Croze, M., Bellis, G., Schiavinato, S., Massougbodji, A., Migot-Nabiase, F. & Dugoujon, J.M. (2015). Genetic population study of Y-chromosome markers in Benin and Ivory Coast ethnic groups. *Forensic Sci Int Genet,* 19; 232-237.

Fortes-Lima, C. & Dugoujon, J.M. (2015). Revisiting genetic ancestry in African American communities from French Guiana, Colombia and Brazil. In: *Trends in Biological Anthropology, 2.* Gerdau-Radonic, K. & Mcsweeney, K. (eds.). Oxford, UK: Oxbow Books for the British Association for Biological Anthropology and Osteoarchaeology. Proceeding paper accepted.

Francalacci, P., Sanna, D., Useli, A., Berutti, R., Barbato, M., Whalen, M.B., Angius, A., Sidore, C., Alonso, S., Tofanelli, S. & Cucca, F. (2015). Detection of phylogenetically informative polymorphisms in the entire euchromatic portion of human Y chromosome from a Sardinian sample. *BMC Res Notes,* 8; 174.

Francois, O. (2013). How to display admixture coefficients (Q matrix) spatially? Website; http://membres-timc.imag.fr/Olivier.Francois/admix_display.html.

Frankham, R., Ballou, J.D. & Briscoe, D.A. (2010). *Introduction to Conservation Genetics.* UK: Cambridge University Press.

Fu, Q., Mittnik, A., Johnson, P.L., Bos, K., Lari, M., Bollongino, R., Sun, C., Giemsch, L., Schmitz, R., Burger, J., Ronchitelli, A.M., Martini, F., Cremonesi, R.G., Svoboda, J., Bauer, P., Caramelli, D., Castellano, S., Reich, D., Paabo, S. & Krause, J. (2013). A revised timescale for human evolution based on ancient mitochondrial genomes. *Curr Biol,* 23(7); 553-559.

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., Jimenez-Sanchez, G., Zolezzi, I.S., Torres, M., Ponte, C.R., Ruiz, Y., Salas, A., Nguyen, E., Eng, C., Borjas, L., Zabala, W., Barreto, G., Gonzalez, F.R., Ibarra, A., Taboada, P., Porras, L., Moreno, F., Bigham, A., Gutierrez, G., Brutsaert, T., Leon-Velarde, F., et al. (2012). Development of a panel of genome-wide ancestry informative markers to study admixture throughout the Americas. *PLoS Genet,* 8(3); e1002554.

Gazal, S., Sahbatou, M., Perdry, H., Letort, S., Genin, E. & Leutenegger, A.L. (2014). Inbreeding coefficient estimation with dense SNP data: comparison of strategies and application to HapMap III. *Hum Hered,* 77(1-4); 49-62.

Gery, Y., Gruner, C. & Mathieu, A. (2014). *Les abandonnés de la République. Vie et mort des Amérindiens de Guyane française.* Paris, France: Albin Michel. Essais Doc.

Gill, P., Jeffreys, A.J. & Werrett, D.J. (1985). Forensic application of DNA 'fingerprints'. *Nature,* 318(6046); 577-579.

Glass, B. & Li, C.C. (1953). The dynamics of racial intermixture; an analysis based on the American Negro. *Am J Hum Genet,* 5(1); 1-20.

Glass, B. (1955). On the unlikelihood of significant admixture of genes from the North American Indians in the present composition of the Negroes of the United States. *Am J Hum Genet,* 7(4); 368-385.

Gomes, F. (1995). *História de quilombolas: mocambos e comunidades de senzalas no Rio de Janeiro, século XIX.* Rio de Janeiro, Brazil: Arquivo Nacional.

Gomez-Perez, L., Alfonso-Sanchez, M.A., Perez-Miranda, A.M., Garcia-Obregon, S., Builes, J.J., Bravo, M.L., De Pancorbo, M.M. & Pena, J.A. (2010). Genetic admixture estimates by Alu elements in Afro-Colombian and Mestizo populations from Antioquia, Colombia. *Ann Hum Biol,* 37(4); 488-500.

Gomez, F., Hirbo, J. & Tishkoff, S.A. (2014). Genetic variation and adaptation in Africa: implications for human evolution and disease. *Cold Spring Harb Perspect Biol,* 6(7); a008524.

Goncalves, V.F., Carvalho, C.M., Bortolini, M.C., Bydlowski, S.P. & Pena, S.D. (2008). The phylogeography of African Brazilians. *Hum Hered,* 65(1); 23-32.

Gonzalez-Jose, R., Bortolini, M.C., Santos, F.R. & Bonatto, S.L. (2008). The peopling of America: craniofacial shape variation on a continental scale and its interpretation from an interdisciplinary view. *Am J Phys Anthropol,* 137(2); 175-187.

Goodman, S.J. (1997). RST Calc: a collection of computer programs for calculating estimates of genetic differentiation from microsatellite data and determining their significance. *Molecular Ecology,* 6(9); 881-885.

Gourjon, G., Saliba-Serre, B. & Degioanni, A. (2014). Admixture Indicative Interval (AII): a new approach to assess trends in genetic admixture. *Genetica,* 142(5); 473-482.

Granka, J. (2013). AncestryDNA Makes Scientific Breakthrough in West African Ethnicity. Website; http://blogs.ancestry.com/.

Gravel, S., Zakharia, F., Moreno-Estrada, A., Byrnes, J.K., Muzzio, M., Rodriguez-Flores, J.L., Kenny, E.E., Gignoux, C.R., Maples, B.K., Guiblet, W., Dutil, J., Via, M., Sandoval, K., Bedoya, G., Genomes, P., Oleksyk, T.K., Ruiz-Linares, A., Burchard, E.G., Martinez-Cruzado, J.C. & Bustamante, C.D. (2013). Reconstructing Native American migrations from whole-genome and whole-exome data. *PLoS Genet,* 9(12); e1004023.

Graven, L., Passarino, G., Semino, O., Boursot, P., Santachiara-Benerecetti, S., Langaney, A. & Excoffier, L. (1995). Evolutionary correlation between control region sequence and restriction polymorphisms in the mitochondrial genome of a large Senegalese Mandenka sample. *Mol Biol Evol,* 12(2); 334-345.

Green, R.E., Malaspinas, A.S., Krause, J., Briggs, A.W., Johnson, P.L., Uhler, C., Meyer, M., Good, J.M., Maricic, T., Stenzel, U., Prufer, K., Siebauer, M., Burbano, H.A., Ronan, M., Rothberg, J.M., Egholm, M., Rudan, P., Brajkovic, D., Kucan, Z., Gusic, I., Wikstrom, M., Laakkonen,

L., Kelso, J., Slatkin, M. & Paabo, S. (2008). A complete Neandertal mitochondrial genome sequence determined by high-throughput sequencing. *Cell,* 134(3); 416-426.

Greenberg, J.H., Ii, C.G.T., Zegura, S.L., Campbell, L., Fox, J.A., Laughlin, W.S., Szathmary, E.J.E., Weiss, K.M. & Woolford, E. (1986). The Settlement of the Americas: A Comparison of the Linguistic, Dental, and Genetic Evidence [and Comments and Reply]. *Current Anthropology,* 27(5); 477-497.

Grenand, P. & Grenand, F. (1985). La question amérindienne en Guyane française : Les Wayapi. *Ethnies,* 1(1/2); 25-26.

Grollemund, R., Branford, S., Bostoen, K., Meade, A., Venditti, C. & Pagel, M. (2015). Bantu expansion shows that habitat alters the route and pace of human dispersals. *Proceedings of the National Academy of Sciences*.

Groucutt, H.S., Petraglia, M.D., Bailey, G., Scerri, E.M., Parton, A., Clark-Balzan, L., Jennings, R.P., Lewis, L., Blinkhorn, J., Drake, N.A., Breeze, P.S., Inglis, R.H., Deves, M.H., Meredith-Williams, M., Boivin, N., Thomas, M.G. & Scally, A. (2015). Rethinking the dispersal of Homo sapiens out of Africa. *Evol Anthropol,* 24(4); 149-164.

Gunderson, K.L., Steemers, F.J., Lee, G., Mendoza, L.G. & Chee, M.S. (2005). A genome-wide scalable SNP genotyping assay using microarray technology. *Nat Genet,* 37(5); 549-554.

Guo, Y., Cai, Q., Samuels, D.C., Ye, F., Long, J., Li, C.I., Winther, J.F., Tawn, E.J., Stovall, M., Lahteenmaki, P., Malila, N., Levy, S., Shaffer, C., Shyr, Y., Shu, X.O. & Boice, J.D., Jr. (2012). The use of next generation sequencing technology to study the effect of radiation therapy on mitochondrial DNA mutation. *Mutat Res,* 744(2); 154-160.

Guo, Y., Li, J., Li, C.I., Shyr, Y. & Samuels, D.C. (2013). MitoSeek: extracting mitochondria information and performing high-throughput mitochondria sequencing analysis. *Bioinformatics,* 29(9); 1210-1211.

Gusmao, L., Butler, J.M., Carracedo, A., Gill, P., Kayser, M., Mayr, W.R., Morling, N., Prinz, M., Roewer, L., Tyler-Smith, C., Schneider, P.M. & Genetics, D.N.A.C.o.t.I.S.o.F. (2006). DNA Commission of the International Society of Forensic Genetics (ISFG): an update of the recommendations on the use of Y-STRs in forensic analysis. *Forensic Sci Int,* 157(2-3); 187-197.

Hagstrom, E., Freyer, C., Battersby, B.J., Stewart, J.B. & Larsson, N.G. (2014). No recombination of mtDNA after heteroplasmy for 50 generations in the mouse maternal germline. *Nucleic Acids Res,* 42(2); 1111-1116.

Halder, I., Shriver, M., Thomas, M., Fernandez, J.R. & Frudakis, T. (2008). A panel of ancestry informative markers for estimating individual biogeographical ancestry and admixture from four continents: utility and applications. *Hum Mutat,* 29(5); 648-658.

Hallast, P., Batini, C., Zadik, D., Maisano Delser, P., Wetton, J.H., Arroyo-Pardo, E., Cavalleri, G.L., de Knijff, P., Destro Bisol, G., Dupuy, B.M., Eriksen, H.A., Jorde, L.B., King, T.E., Larmuseau, M.H., Lopez de Munain, A., Lopez-Parra, A.M., Loutradis, A., Milasin, J., Novelletto, A., Pamjav, H., Sajantila, A., Schempp, W., Sears, M., Tolun, A., Tyler-Smith, C., Van Geystelen, A., Watkins, S., Winney, B. & Jobling, M.A. (2015). The Y-chromosome tree bursts into leaf: 13,000 high-confidence SNPs covering the majority of known clades. *Mol Biol Evol,* 32(3); 661-673.

Hammer, M.F. (1994). A recent insertion of an alu element on the Y chromosome is a useful marker for human population studies. *Mol Biol Evol,* 11(5); 749-761.

Hammer, M.F., Karafet, T., Rasanayagam, A., Wood, E.T., Altheide, T.K., Jenkins, T., Griffiths, R.C., Templeton, A.R. & Zegura, S.L. (1998). Out of Africa and back again: nested cladistic analysis of human Y chromosome variation. *Mol Biol Evol,* 15(4); 427-441.

Hammer, M.F., Chamberlain, V.F., Kearney, V.F., Stover, D., Zhang, G., Karafet, T., Walsh, B. & 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(1); 45-55.

Hanchard, N.A., Hambleton, I., Harding, R.M. & McKenzie, C.A. (2006). Predicted declines in sickle allele frequency in Jamaica using empirical data. *Am J Hematol,* 81(11); 817-823.

Hancock, I.F. (1987). A preliminary classification of the anglophone Atlantic creoles with syntactic data from thrity-three representative dialects. In: *Pidgin and Creole Languages. Essays in Memory of John E. Reinecke*. Gilbert, G.G. (ed. Honolulu: University of Hawaii Press.

Harich, N., Costa, M.D., Fernandes, V., Kandil, M., Pereira, J.B., Silva, N.M. & 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.

He, Y., Wu, J., Dressman, D.C., Iacobuzio-Donahue, C., Markowitz, S.D., Velculescu, V.E., Diaz, L.A., Jr., Kinzler, K.W., Vogelstein, B. & Papadopoulos, N. (2010). Heteroplasmic mitochondrial DNA mutations in normal and tumour cells. *Nature,* 464(7288); 610-614.

Heinz, T., Cardenas, J.M., Alvarez-Iglesias, V., Pardo-Seco, J., Gomez-Carballa, A., Santos, C., Taboada-Echalar, P., Martinon-Torres, F. & Salas, A. (2015). The Genomic Legacy of the Transatlantic Slave Trade in the Yungas Valley of Bolivia. *PLoS One,* 10(8); e0134129.

Henn, B.M., Gravel, S., Moreno-Estrada, A., Acevedo-Acevedo, S. & Bustamante, C.D. (2010). Fine-scale population structure and the era of next-generation sequencing. *Hum Mol Genet,* 19(R2); R221-226.

Henn, B.M., Gignoux, C.R., Jobin, M., Granka, J.M., Macpherson, J.M., Kidd, J.M., Rodriguez-Botigue, L., Ramachandran, S., Hon, L., Brisbin, A., Lin, A.A., Underhill, P.A., Comas, D., Kidd, K.K., Norman, P.J., Parham, P., Bustamante, C.D., Mountain, J.L. & Feldman, M.W. (2011). Hunter-gatherer genomic diversity suggests a southern African origin for modern humans. *Proc Natl Acad Sci U S A,* 108(13); 5154-5162.

Henn, B.M., Botigue, L.R., Gravel, S., Wang, W., Brisbin, A., Byrnes, J.K., Fadhlaoui-Zid, K., Zalloua, P.A., Moreno-Estrada, A., Bertranpetit, J., Bustamante, C.D. & Comas, D. (2012). Genomic ancestry of North Africans supports back-to-Africa migrations. *PLoS Genet,* 8(1); e1002397.

Hernandez, C.L., Reales, G., Dugoujon, J.M., Novelletto, A., Rodriguez, J.N., Cuesta, P. & Calderon, R. (2014). Human maternal heritage in Andalusia (Spain): its composition reveals high internal complexity and distinctive influences of mtDNA haplogroups U6 and L in the western and eastern side of region. *BMC Genet,* 15; 11.

Heyer, E., Tremblay, M. & Desjardins, B. (1997). Seventeenth-century European origins of hereditary diseases in the Saguenay population (Quebec, Canada). *Hum Biol,* 69(2); 209-225.

Hinch, A.G., Tandon, A., Patterson, N., Song, Y., Rohland, N., Palmer, C.D., Chen, G.K., Wang, K., Buxbaum, S.G., Akylbekova, E.L., Aldrich, M.C., Ambrosone, C.B., Amos, C., Bandera, E.V., Berndt, S.I., Bernstein, L., Blot, W.J., Bock, C.H., Boerwinkle, E., Cai, Q., Caporaso, N., Casey, G., Cupples, L.A., Deming, S.L., Diver, W.R., Divers, J., Fornage, M., Gillanders, E.M., Glessner, J., Harris, C.C., et al. (2011). The landscape of recombination in African Americans. *Nature,* 476(7359); 170-175.

Hodgson, J.A., Mulligan, C.J., Al-Meeri, A. & Raaum, R.L. (2014). Early back-to-Africa migration into the Horn of Africa. *PLoS Genet,* 10(6); e1004393.

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

Hoogbergen, W.S.M. (1990). *The Boni Maroon Wars in Suriname.* Leiden, The Netherlands: Brill.

Howes, R.E., Patil, A.P., Piel, F.B., Nyangiri, O.A., Kabaria, C.W., Gething, P.W., Zimmerman, P.A., Barnadas, C., Beall, C.M., Gebremedhin, A., Menard, D., Williams, T.N., Weatherall, D.J. & Hay, S.I. (2011). The global distribution of the Duffy blood group. *Nat Commun,* 2; 266.

Hughes, J.F., Skaletsky, H., Pyntikova, T., Graves, T.A., van Daalen, S.K., Minx, P.J., Fulton, R.S., McGrath, S.D., Locke, D.P., Friedman, C., Trask, B.J., Mardis, E.R., Warren, W.C., Repping, S., Rozen, S., Wilson, R.K. & Page, D.C. (2010). Chimpanzee and human Y chromosomes are remarkably divergent in structure and gene content. *Nature,* 463(7280); 536-539.

Hunemeier, T., Carvalho, C., Marrero, A.R., Salzano, F.M., Junho Pena, S.D. & 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(2); 854-867.

Hunter-Zinck, H., Musharoff, S., Salit, J., Al-Ali, K.A., Chouchane, L., Gohar, A., Matthews, R., Butler, M.W., Fuller, J., Hackett, N.R., Crystal, R.G. & Clark, A.G. (2010). Population genetic structure of the people of Qatar. *Am J Hum Genet,* 87(1); 17-25.

Huser, V., Sincan, M. & Cimino, J.J. (2014). Developing genomic knowledge bases and databases to support clinical management: current perspectives. *Pharmgenomics Pers Med,* 7; 275-283.

Huson, D.H. & Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol,* 23(2); 254-267.

Hutchison, C.A., 3rd, Newbold, J.E., Potter, S.S. & Edgell, M.H. (1974). Maternal inheritance of mammalian mitochondrial DNA. *Nature,* 251(5475); 536-538.

Huttar, G.L. & Huttar, M.L. (1994). *Ndyuka.* London: Routledge.

Huttar, G.L., Essegbey, J. & Ameka, F.K. (2007). Gbe and other West African sources of Suriname creole semantic structures: Implications for creole genesis. *Journal of Pidgin and Creole Languages,* 22(1); 57-72.

Huttar, G.L. (2010). Successive African and European Contributions to Suriname Creole Lexicons. In: *A mosaic of languages and cultures: studies celebrating the career of Karl J. Franklin.* Mcelhanon, K.A. & Reesink, G. (eds.). Dallas, USA: SIL International.

Huttar, G.L. (2012). The African lexical contribution to Ndyuka, Saramaccan, and other creoles: Implications for how creoles develop. In: *Black through White: African words and calques which survived slavery in Creoles and transplanted European languages.* Bartens, A. & Baker, P. (eds.). London, UK: Battlebridge Publications.

Hyman, L.M. & Hombert, J.M. (1999). *Bantu Historical Linguistics: Theoretical and Empirical Perspectives.* Stanford, USA: CSLI Publications Stanford University. Center for the Study of Language and Information - Lecture Notes.

ILO (1989). C169 Indigenous and Tribal Peoples Convention. Geneva, Switzerland: General Conference of the International Labour Organisation.

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

INSEE (2000). *Évolutions démographiques 1982-1990-1999: Données déflnitives. 973 Guyane.* Paris: Institut National de la Statistique et des Études Économiques (INSEE).

International HapMap 3 Consortium, Altshuler, D.M., Gibbs, R.A., Peltonen, L., Altshuler, D.M., Gibbs, R.A., Peltonen, L., Dermitzakis, E., Schaffner, S.F., Yu, F., Peltonen, L., Dermitzakis, E., Bonnen, P.E., Altshuler, D.M., Gibbs, R.A., de Bakker, P.I., Deloukas, P., Gabriel, S.B., Gwilliam, R., Hunt, S., Inouye, M., Jia, X., Palotie, A., Parkin, M., Whittaker, P., Yu, F., Chang, K., Hawes, A., Lewis, L.R., Ren, Y., et al. (2010). Integrating common and rare genetic variation in diverse human populations. *Nature,* 467(7311); 52-58.

International HapMap Consortium (2005). A haplotype map of the human genome. *Nature,* 437(7063); 1299-1320.

International Human Genome Sequencing Consortium (2004). Finishing the euchromatic sequence of the human genome. *Nature,* 431(7011); 931-945.

Irwin, J.A., Saunier, J.L., Niederstatter, H., Strouss, K.M., Sturk, K.A., Diegoli, T.M., Brandstatter, A., Parson, W. & Parsons, T.J. (2009). Investigation of heteroplasmy in the human mitochondrial DNA control region: a synthesis of observations from more than 5000 global population samples. *J Mol Evol,* 68(5); 516-527.

Jackson, B.A., Wilson, J.L., Kirbah, S., Sidney, S.S., Rosenberger, J., Bassie, L., Alie, J.A., McLean, D.C., Garvey, W.T. & Ely, B. (2005). Mitochondrial DNA genetic diversity among four ethnic groups in Sierra Leone. *Am J Phys Anthropol,* 128(1); 156-163.

Jackson, F.L. & Borgelin, L.F. (2010). How Genetic s Can Provide Detail to the Transatlantic African Diaspora. In: *The African diaspora and the disciplines*. Olaniyan, T. & Sweet, J.H. (eds.). Bloomington: Indiana University Press.

Jakkula, E., Rehnstrom, K., Varilo, T., Pietilainen, O.P., Paunio, T., Pedersen, N.L., deFaire, U., Jarvelin, M.R., Saharinen, J., Freimer, N., Ripatti, S., Purcell, S., Collins, A., Daly, M.J., Palotie, A. & Peltonen, L. (2008). The genome-wide patterns of variation expose significant substructure in a founder population. *Am J Hum Genet,* 83(6); 787-794.

Jakobsson, M., Scholz, S.W., Scheet, P., Gibbs, J.R., VanLiere, J.M., Fung, H.C., Szpiech, Z.A., Degnan, J.H., Wang, K., Guerreiro, R., Bras, J.M., Schymick, J.C., Hernandez, D.G., Traynor, B.J., Simon-Sanchez, J., Matarin, M., Britton, A., van de Leemput, J., Rafferty, I., Bucan, M., Cann, H.M., Hardy, J.A., Rosenberg, N.A. & Singleton, A.B. (2008). Genotype, haplotype and copy-number variation in worldwide human populations. *Nature,* 451(7181); 998-1003.

Jalkh, N., Sahbatou, M., Chouery, E., Megarbane, A., Leutenegger, A.L. & Serre, J.L. (2015). Genome-wide inbreeding estimation within Lebanese communities using SNP arrays. *Eur J Hum Genet,* 23(10); 1364-1369.

Jobling, M.A. (2001). Y-chromosomal SNP haplotype diversity in forensic analysis. *Forensic Sci Int,* 118(2-3); 158-162.

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

Jobling, M.A., Hollox, E., Hurles, M., Kivisild, T. & Tyler-Smith, C. (2014). *Human Evolutionary Genetics.* New York: Garland Science, Taylor & Francis Group.

Johnson, N.A., Coram, M.A., Shriver, M.D., Romieu, I., Barsh, G.S., London, S.J. & Tang, H. (2011). Ancestral components of admixed genomes in a Mexican cohort. *PLoS Genet,* 7(12); e1002410.

Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics,* 24(11); 1403-1405.

Jombart, T., Devillard, S. & Balloux, F. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet,* 11; 94.

Jordan, J. (2004). Afro-Colombian. *Souls: A Critical Journal of Black Politics, Culture, and Society,* 6(2); 19-30.

Just, R.S., Diegoli, T.M., Saunier, J.L., Irwin, J.A. & Parsons, T.J. (2008). Complete mitochondrial genome sequences for 265 African American and U.S. "Hispanic" individuals. *Forensic Sci Int Genet,* 2(3); e45-48.

Just, R.S., Scheible, M.K., Fast, S.A., Sturk-Andreaggi, K., Rock, A.W., Bush, J.M., Higginbotham, J.L., Peck, M.A., Ring, J.D., Huber, G.E., Xavier, C., Strobl, C., Lyons, E.A., Diegoli, T.M., Bodner, M., Fendt, L., Kralj, P., Nagl, S., Niederwieser, D., Zimmermann, B., Parson, W. & Irwin, J.A. (2015). Full mtGenome reference data: development and characterization of 588 forensic-quality haplotypes representing three U.S. populations. *Forensic Sci Int Genet,* 14; 141-155.

Kambel, E.R. & MacKay, F. (1999). The Rights of Indigenous Peoples and Maroons in Suriname. Copenhagen, Dermark: International Work Group for Indigenous Affairs.

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

Karafet, T.M., Mendez, F.L., Sudoyo, H., Lansing, J.S. & Hammer, M.F. (2015). Improved phylogenetic resolution and rapid diversification of Y-chromosome haplogroup K-M526 in Southeast Asia. *Eur J Hum Genet,* 23(3); 369-373.

Karmin, M., Saag, L., Vicente, M., Sayres, M.A., Jarve, M., Talas, U.G., Rootsi, S., Ilumae, A.M., Magi, R., Mitt, M., Pagani, L., Puurand, T., Faltyskova, Z., Clemente, F., Cardona, A., Metspalu, E., Sahakyan, H., Yunusbayev, B., Hudjashov, G., DeGiorgio, M., Loogvali, E.L., Eichstaedt, C., Eelmets, M., Chaubey, G., Tambets, K., Litvinov, S., Mormina, M., Xue, Y., Ayub, Q., Zoraqi, G., et al. (2015). A recent bottleneck of Y chromosome diversity coincides with a global change in culture. *Genome Res,* doi: 10.1101/gr.186684.114.

Kayser, M., Caglia, A., Corach, D., Fretwell, N., Gehrig, C., Graziosi, G., Heidorn, F., Herrmann, S., Herzog, B., Hidding, M., Honda, K., Jobling, M., Krawczak, M., Leim, K., Meuser, S., Meyer, E., Oesterreich, W., Pandya, A., Parson, W., Penacino, G., Perez-Lezaun, A., Piccinini, A., Prinz, M., Schmitt, C., Roewer, L. & et al. (1997). Evaluation of Y-chromosomal STRs: a multicenter study. *Int J Legal Med,* 110(3); 125-133, 141-129.

Kayser, M. & de Knijff, P. (2011). Improving human forensics through advances in genetics, genomics and molecular biology. *Nat Rev Genet,* 12(3); 179-192.

Ke, X., Taylor, M.S. & Cardon, L.R. (2008). Singleton SNPs in the human genome and implications for genome-wide association studies. *Eur J Hum Genet,* 16(4); 506-515.

Kehdy, F.S., Gouveia, M.H., Machado, M., Magalhaes, W.C., Horimoto, A.R., Horta, B.L., Moreira, R.G., Leal, T.P., Scliar, M.O., Soares-Souza, G.B., Rodrigues-Soares, F., Araujo, G.S., Zamudio, R., Sant Anna, H.P., Santos, H.C., Duarte, N.E., Fiaccone, R.L., Figueiredo, C.A., Silva, T.M., Costa, G.N., Beleza, S., Berg, D.E., Cabrera, L., Debortoli, G., Duarte, D., Ghirotto, S., Gilman, R.H., Goncalves, V.F., Marrero, A.R., Muniz, Y.C., et al. (2015). Origin and dynamics of admixture in Brazilians and its effect on the pattern of deleterious mutations. *Proc Natl Acad Sci U S A,* 112(28); 8696-8701.

Keyeux, G., Rodas, C., Gelvez, N. & Carter, D. (2002). Possible migration routes into South America deduced from mitochondrial DNA studies in Colombian Amerindian populations. *Hum Biol,* 74(2); 211-233.

Kidd, J.M., Gravel, S., Byrnes, J., Moreno-Estrada, A., Musharoff, S., Bryc, K., Degenhardt, J.D., Brisbin, A., Sheth, V., Chen, R., McLaughlin, S.F., Peckham, H.E., Omberg, L., Bormann Chung, C.A., Stanley, S., Pearlstein, K., Levandowsky, E., Acevedo-Acevedo, S., Auton, A., Keinan, A., Acuna-Alonzo, V., Barquera-Lozano, R., Canizales-Quinteros, S., Eng, C., Burchard, E.G., Russell, A., Reynolds, A., Clark, A.G., Reese, M.G., Lincoln, S.E., et al. (2012). Population genetic inference from personal genome data: impact of ancestry and admixture on human genomic variation. *Am J Hum Genet,* 91(4); 660-671.

Kim, H.L., Ratan, A., Perry, G.H., Montenegro, A., Miller, W. & Schuster, S.C. (2014). Khoisan hunter-gatherers have been the largest population throughout most of modern-human demographic history. *Nat Commun,* 5; 5692.

Kircher, M., Stenzel, U. & Kelso, J. (2009). Improved base calling for the Illumina Genome Analyzer using machine learning strategies. *Genome Biol,* 10(8); R83.

Kivisild, T., Shen, P., Wall, D.P., Do, B., Sung, R., Davis, K., Passarino, G., Underhill, P.A., Scharfe, C., Torroni, A., Scozzari, R., Modiano, D., Coppa, A., de Knijff, P., Feldman, M., Cavalli-Sforza, L.L. & Oefner, P.J. (2006). The role of selection in the evolution of human mitochondrial genomes. *Genetics,* 172(1); 373-387.

Kivisild, T. (2015). Maternal ancestry and population history from whole mitochondrial genomes. *Investig Genet,* 6; 3.

Klein, H.S. (2002). As origens africanas dos escravos brasileiros. In: *Homo brasilis: aspectos genéticos, lingüísticos, históricos e socioantropológicos da formação do povo brasileiro*. Pena, S.D.J. (ed. Ribeirão Preto: FUNPEC Editora.

Klein, H.S. & Vinson, B. (2007). *African Slavery in Latin America and the Caribbean.* New York: Oxford University Press.

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

Kopelman, N.M., Mayzel, J., Jakobsson, M., Rosenberg, N.A. & Mayrose, I. (2015). Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Mol Ecol Resour*.

Kundu, S. & Ghosh, S.K. (2015). Trend of different molecular markers in the last decades for studying human migrations. *Gene,* 556(2); 81-90.

Kusuma, P., Cox, M.P., Pierron, D., Razafindrazaka, H., Brucato, N., Tonasso, L., Suryadi, H.L., Letellier, T., Sudoyo, H. & Ricaut, F.X. (2015). Mitochondrial DNA and the Y chromosome suggest the settlement of Madagascar by Indonesian sea nomad populations. *BMC Genomics,* 16; 191.

Lachance, J. & Tishkoff, S.A. (2013). SNP ascertainment bias in population genetic analyses: why it is important, and how to correct it. *Bioessays,* 35(9); 780-786.

Landers, J., Gomez, P., Polo-Acuna, J. & Campbell, C.J. (2015). Researching the history of slavery in Colombia and Brazil through ecclesiastical and notarial archives. In: *From Dust to Digital: Ten Years of the Endangered Archives Programme*. Kominko, M. (ed. Cambridge, UK: Open Book.

Lane, P. & MacDonald, K.C. (2011). Introduction: Slavery, Social Revolutions and Enduring Memories. In: *Slavery in Africa: Archaeology and Memory*. Lane, P. & Macdonald, K.C. (eds.). London, UK: British Academy.

Lao, O., Lu, T.T., Nothnagel, M., Junge, O., Freitag-Wolf, S., Caliebe, A., Balascakova, M., Bertranpetit, J., Bindoff, L.A., Comas, D., Holmlund, G., Kouvatsi, A., Macek, M., Mollet, I., Parson, W., Palo, J., Ploski, R., Sajantila, A., Tagliabracci, A., Gether, U., Werge, T., Rivadeneira, F., Hofman, A., Uitterlinden, A.G., Gieger, C., Wichmann, H.E., Ruther, A., Schreiber, S., Becker, C., Nurnberg, P., et al. (2008). Correlation between genetic and geographic structure in Europe. *Curr Biol,* 18(16); 1241-1248.

Lao, O., Liu, F., Wollstein, A. & Kayser, M. (2014). GAGA: a new algorithm for genomic inference of geographic ancestry reveals fine level population substructure in Europeans. *PLoS Comput Biol,* 10(2); e1003480.

Law, R. (1991). *The Slave Coast of West Africa, 1550–1750: the impact of the Atlantic slave trade on an African society.* Oxford, UK: Clarendon Press. Oxford Studies in African Affairs.

Law, R. (2004). *Ouidah: The Social History of a West African Slaving Port 1727-1892.* Suffolk, UK: James Currey. Western African Studies.

Lawson, D.J., Hellenthal, G., Myers, S. & Falush, D. (2012). Inference of population structure using dense haplotype data. *PLoS Genet,* 8(1); e1002453.

Lefebvre, C. (2013). A comparison of the nominal structures of Saramaccan, Fongbe and English with reference to Haitian Creole: Implications for a relabelling-based account of creole genesis. *Lingua,* 129; 32-66.

Leglise, I., Renault-Lescure, O., Launey, M. & Migge, B. (2013). Langues de Guyane et langues parlées en Guyane. In: *Histoire sociale des langues de France.* Kremnitz, G. (ed. Rennes: Presses Universitaires de Rennes.

Lemos Cardoso, G. & Farias Guerreiro, J. (2006). African gene flow to north Brazil as revealed by HBB*S gene haplotype analysis. *Am J Hum Biol,* 18(1); 93-98.

Leslie, S., Winney, B., Hellenthal, G., Davison, D., Boumertit, A., Day, T., Hutnik, K., Royrvik, E.C., Cunliffe, B., Wellcome Trust Case Control, C., International Multiple Sclerosis Genetics, C., Lawson, D.J., Falush, D., Freeman, C., Pirinen, M., Myers, S., Robinson, M., Donnelly, P. & Bodmer, W. (2015). The fine-scale genetic structure of the British population. *Nature,* 519(7543); 309-314.

Li, C.C. & Horvitz, D.G. (1953). Some methods of estimating the inbreeding coefficient. *Am J Hum Genet,* 5(2); 107-117.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. & Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics,* 25(16); 2078-2079.

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. & Myers, R.M. (2008). Worldwide human relationships inferred from genome-wide patterns of variation. *Science,* 319(5866); 1100-1104.

Li, L.H., Ho, S.F., Chen, C.H., Wei, C.Y., Wong, W.C., Li, L.Y., Hung, S.I., Chung, W.H., Pan, W.H., Lee, M.T., Tsai, F.J., Chang, C.F., Wu, J.Y. & Chen, Y.T. (2006). Long contiguous stretches of homozygosity in the human genome. *Hum Mutat,* 27(11); 1115-1121.

Li, M., Schonberg, A., Schaefer, M., Schroeder, R., Nasidze, I. & Stoneking, M. (2010). Detecting heteroplasmy from high-throughput sequencing of complete human mitochondrial DNA genomes. *Am J Hum Genet,* 87(2); 237-249.

Li, S., Schlebusch, C. & Jakobsson, M. (2014). Genetic variation reveals large-scale population expansion and migration during the expansion of Bantu-speaking peoples. *Proc Biol Sci,* 281(1793).

Li, W., Cowley, A., Uludag, M., Gur, T., McWilliam, H., Squizzato, S., Park, Y.M., Buso, N. & Lopez, R. (2015). The EMBL-EBI bioinformatics web and programmatic tools framework. *Nucleic Acids Res*.

Libiger, O. & Schork, N.J. (2012). A Method for Inferring an Individual's Genetic Ancestry and Degree of Admixture Associated with Six Major Continental Populations. *Front Genet,* 3; 322.

Lind, J.M., Hutcheson-Dilks, H.B., Williams, S.M., Moore, J.H., Essex, M., Ruiz-Pesini, E., Wallace, D.C., Tishkoff, S.A., O'Brien, S.J. & Smith, M.W. (2007). Elevated male European and female African contributions to the genomes of African American individuals. *Hum Genet,* 120(5); 713-722.

Lins, T.C., Vieira, R.G., Abreu, B.S., Grattapaglia, D. & Pereira, R.W. (2010). Genetic composition of Brazilian population samples based on a set of twenty-eight ancestry informative SNPs. *Am J Hum Biol,* 22(2); 187-192.

Lippold, S., Xu, H., Ko, A., Li, M., Renaud, G., Butthof, A., Schroder, R. & Stoneking, M. (2014). Human paternal and maternal demographic histories: insights from high-resolution Y chromosome and mtDNA sequences. *Investig Genet,* 5; 13.

Liu, Y., Nyunoya, T., Leng, S., Belinsky, S.A., Tesfaigzi, Y. & Bruse, S. (2013). Softwares and methods for estimating genetic ancestry in human populations. *Hum Genomics,* 7; 1.

Londin, E.R., Keller, M.A., Maista, C., Smith, G., Mamounas, L.A., Zhang, R., Madore, S.J., Gwinn, K. & Corriveau, R.A. (2010). CoAIMs: a cost-effective panel of ancestry informative markers for determining continental origins. *PLoS One,* 5(10); e13443.

Lopez Herraez, D., Bauchet, M., Tang, K., Theunert, C., Pugach, I., Li, J., Nandineni, M.R., Gross, A., Scholz, M. & Stoneking, M. (2009). Genetic variation and recent positive selection in worldwide human populations: evidence from nearly 1 million SNPs. *PLoS One,* 4(11); e7888.

Lovejoy, H.B. (2013). Redrawing historical maps of the Bight of Benin Hinterland, c. 1780. *Canadian Journal of African Studies,* 47(3); 443-463.

Lovejoy, P.E. (2011). *Transformations in Slavery. A History of Slavery in Africa.* New York, USA: Cambridge University Press.

Luke, A., Guo, X., Adeyemo, A.A., Wilks, R., Forrester, T., Lowe, W., Jr., Comuzzie, A.G., Martin, L.J., Zhu, X., Rotimi, C.N. & Cooper, R.S. (2001). Heritability of obesity-related traits among Nigerians, Jamaicans and US black people. *Int J Obes Relat Metab Disord,* 25(7); 1034-1041.

Maca-Meyer, N., Gonzalez, A.M., Pestano, J., Flores, C., Larruga, J.M. & Cabrera, V.M. (2003). Mitochondrial DNA transit between West Asia and North Africa inferred from U6 phylogeography. *BMC Genet,* 4; 15.

Maca-Meyer, N., Arnay, M., Rando, J.C., Flores, C., Gonzalez, A.M., Cabrera, V.M. & Larruga, J.M. (2004). Ancient mtDNA analysis and the origin of the Guanches. *Eur J Hum Genet,* 12(2); 155-162.

Macaulay, V., Hill, C., Achilli, A., Rengo, C., Clarke, D., Meehan, W., Blackburn, J., Semino, O., Scozzari, R., Cruciani, F., Taha, A., Shaari, N.K., Raja, J.M., Ismail, P., Zainuddin, Z., Goodwin, W., Bulbeck, D., Bandelt, H.J., Oppenheimer, S., Torroni, A. & Richards, M. (2005). Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Science,* 308(5724); 1034-1036.

Madrilejo, N., Lombard, H. & Torres, J.B. (2015). Origins of marronage: Mitochondrial lineages of Jamaica's Accompong Town Maroons. *Am J Hum Biol,* 27(3); 432-437.

Malyarchuk, B.A., Derenko, M., Perkova, M., Grzybowski, T., Vanecek, T. & Lazur, J. (2008). Reconstructing the phylogeny of African mitochondrial DNA lineages in Slavs. *Eur J Hum Genet,* 16(9); 1091-1096.

Manta, F.S., Pereira, R., Caiafa, A., Silva, D.A., Gusmao, L. & Carvalho, E.F. (2013). Analysis of genetic ancestry in the admixed Brazilian population from Rio de Janeiro using 46 autosomal ancestry-informative indel markers. *Ann Hum Biol,* 40(1); 94-98.

Maples, B.K., Gravel, S., Kenny, E.E. & Bustamante, C.D. (2013). RFMix: a discriminative modeling approach for rapid and robust local-ancestry inference. *Am J Hum Genet,* 93(2); 278-288.

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.J., Chen, Z., Dewell, S.B., Du, L., Fierro, J.M., Gomes, X.V., Godwin, B.C., He, W., Helgesen, S., Ho, C.H., Irzyk, G.P., Jando, S.C., Alenquer, M.L., Jarvie, T.P., Jirage, K.B., Kim, J.B., Knight, J.R., Lanza, J.R., Leamon, J.H., Lefkowitz, S.M., Lei, M., Li, J., et al. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature,* 437(7057); 376-380.

Maricic, T., Whitten, M. & Paabo, S. (2010). Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS One,* 5(11); e14004.

Marks, S.J., Montinaro, F., Levy, H., Brisighelli, F., Ferri, G., Bertoncini, S., Batini, C., Busby, G.B., Arthur, C., Mitchell, P., Stewart, B.A., Oosthuizen, O., Oosthuizen, E., D'Amato, M.E., Davison, S., Pascali, V. & Capelli, C. (2015). Static and moving frontiers: the genetic landscape of Southern African Bantu-speaking populations. *Mol Biol Evol,* 32(1); 29-43.

Marques, S.L., Goios, A., Rocha, A.M., Prata, M.J., Amorim, A., Gusmao, L., Alves, C. & Alvarez, L. (2015). Portuguese mitochondrial DNA genetic diversity-An update and a phylogenetic revision. *Forensic Sci Int Genet,* 15; 27-32.

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 Junior, W.A. & 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,* 17(4); 496-506.

Maxam, A.M. & Gilbert, W. (1977). A new method for sequencing DNA. *Proc Natl Acad Sci U S A,* 74(2); 560-564.

May, A., Hazelhurst, S., Li, Y., Norris, S.A., Govind, N., Tikly, M., Hon, C., Johnson, K.J., Hartmann, N., Staedtler, F. & Ramsay, M. (2013). Genetic diversity in black South Africans from Soweto. *BMC Genomics,* 14; 644.

Mazieres, S., Guitard, E., Crubezy, E., Dugoujon, J.M., Bortolini, M.C., Bonatto, S.L., Hutz, M.H., Bois, E., Tiouka, F., Larrouy, G. & Salzano, F.M. (2008). Uniparental (mtDNA, Y-chromosome) polymorphisms in French Guiana and two related populations--implications for the region's colonization. *Ann Hum Genet,* 72(Pt 1); 145-156.

Mazieres, S., Callegari-Jacques, S.M., Crossetti, S.G., Dugoujon, J.M., Larrouy, G., Bois, E., Crubezy, E., Hutz, M.H. & Salzano, F.M. (2011). French Guiana Amerindian demographic history as revealed by autosomal and Y-chromosome STRs. *Ann Hum Biol,* 38(1); 76-83.

McDougall, I., Brown, F.H. & Fleagle, J.G. (2005). Stratigraphic placement and age of modern humans from Kibish, Ethiopia. *Nature,* 433(7027); 733-736.

McMillin, J.A. (2012). *The Final Victims: Foreign Slave Trade to North America, 1783-1810.* Columbia, USA: University of South Carolina Press. Carolina Lowcountry and the Atlantic World.

McQuillan, R., Leutenegger, A.L., Abdel-Rahman, R., Franklin, C.S., Pericic, M., Barac-Lauc, L., Smolej-Narancic, N., Janicijevic, B., Polasek, O., Tenesa, A., Macleod, A.K., Farrington, S.M., Rudan, P., Hayward, C., Vitart, V., Rudan, I., Wild, S.H., Dunlop, M.G., Wright, A.F., Campbell, H. & Wilson, J.F. (2008). Runs of homozygosity in European populations. *Am J Hum Genet,* 83(3); 359-372.

McVean, G. (2009). A genealogical interpretation of principal components analysis. *PLoS Genet,* 5(10); e1000686.

Mendez, F.L., Karafet, T.M., Krahn, T., Ostrer, H., Soodyall, H. & Hammer, M.F. (2011). Increased resolution of Y chromosome haplogroup T defines relationships among populations of the Near East, Europe, and Africa. *Hum Biol,* 83(1); 39-53.

Mendez, F.L., Krahn, T., Schrack, B., Krahn, A.M., Veeramah, K.R., Woerner, A.E., Fomine, F.L., Bradman, N., Thomas, M.G., Karafet, T.M. & Hammer, M.F. (2013). An African American paternal lineage adds an extremely ancient root to the human Y chromosome phylogenetic tree. *Am J Hum Genet,* 92(3); 454-459.

Menozzi, P., Piazza, A. & Cavalli-Sforza, L. (1978). Synthetic maps of human gene frequencies in Europeans. *Science,* 201(4358); 786-792.

Meyer, M. & Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb Protoc,* 2010(6); pdb prot5448.

Migge, B. & Winford, D. (2013). Fact-type complements in Gbe and the Surinamese Creoles. *Lingua,* 129; 9-31.

Mishmar, D., Ruiz-Pesini, E., Golik, P., Macaulay, V., Clark, A.G., Hosseini, S., Brandon, M., Easley, K., Chen, E., Brown, M.D., Sukernik, R.I., Olckers, A. & Wallace, D.C. (2003). Natural selection shaped regional mtDNA variation in humans. *Proc Natl Acad Sci U S A,* 100(1); 171-176.

Montano, V., Marcari, V., Pavanello, M., Anyaele, O., Comas, D., Destro-Bisol, G. & Batini, C. (2013). The influence of habitats on female mobility in Central and Western Africa inferred from human mitochondrial variation. *BMC Evol Biol,* 13; 24.

Montinaro, F., Busby, G.B., Pascali, V.L., Myers, S., Hellenthal, G. & Capelli, C. (2015). Unravelling the hidden ancestry of American admixed populations. *Nat Commun,* 6; 6596.

Moomou, J. (2004). *Le monde des Marrons du Maroni en Guyane (1772-1860).* French Guiana, France: Ibis Rouge Editions. Histoire.

Moorjani, P., Patterson, N., Hirschhorn, J.N., Keinan, A., Hao, L., Atzmon, G., Burns, E., Ostrer, H., Price, A.L. & Reich, D. (2011). The history of African gene flow into Southern Europeans, Levantines, and Jews. *PLoS Genet,* 7(4); e1001373.

Moreno-Estrada, A., Gravel, S., Zakharia, F., McCauley, J.L., Byrnes, J.K., Gignoux, C.R., Ortiz-Tello, P.A., Martinez, R.J., Hedges, D.J., Morris, R.W., Eng, C., Sandoval, K., Acevedo-Acevedo, S., Norman, P.J., Layrisse, Z., Parham, P., Martinez-Cruzado, J.C., Burchard, E.G., Cuccaro, M.L., Martin, E.R. & Bustamante, C.D. (2013). Reconstructing the population genetic history of the Caribbean. *PLoS Genet,* 9(11); e1003925.

Moreno-Estrada, A., Gignoux, C.R., Fernandez-Lopez, J.C., Zakharia, F., Sikora, M., Contreras, A.V., Acuna-Alonzo, V., Sandoval, K., Eng, C., Romero-Hidalgo, S., Ortiz-Tello, P., Robles, V.,

Kenny, E.E., Nuno-Arana, I., Barquera-Lozano, R., Macin-Perez, G., Granados-Arriola, J., Huntsman, S., Galanter, J.M., Via, M., Ford, J.G., Chapela, R., Rodriguez-Cintron, W., Rodriguez-Santana, J.R., Romieu, I., Sienra-Monge, J.J., del Rio Navarro, B., London, S.J., Ruiz-Linares, A., Garcia-Herrera, R., et al. (2014). Human genetics. The genetics of Mexico recapitulates Native American substructure and affects biomedical traits. *Science,* 344(6189); 1280-1285.

Morgan, P.D. (2013). The Cultural Implications of the Atlantic Slave Trade: African Regional Origins American Destinations and New World Developments. In: *Routes to Slavery: Direction, Ethnicity and Mortality in the Transatlantic Slave Trade*. Eltis, D. & Richardson, D. (eds.). New York, USA: Routledge.

Morner, M. (1967). *Race Mixture in the History of Latin America.* Boston: Little, Brown and Co.

Morris, A.G., Heinze, A., Chan, E.K., Smith, A.B. & Hayes, V.M. (2014). First ancient mitochondrial human genome from a prepastoralist southern African. *Genome Biol Evol,* 6(10); 2647-2653.

Murray, T., Beaty, T.H., Mathias, R.A., Rafaels, N., Grant, A.V., Faruque, M.U., Watson, H.R., Ruczinski, I., Dunston, G.M. & Barnes, K.C. (2010). African and non-African admixture components in African Americans and an African Caribbean population. *Genet Epidemiol,* 34(6); 561-568.

Myakishev, M.V., Khripin, Y., Hu, S. & Hamer, D.H. (2001). High-throughput SNP genotyping by allele-specific PCR with universal energy-transfer-labeled primers. *Genome Res,* 11(1); 163-169.

Navarrete, M.C. (2011). Los cimarrones de la provincia de Cartagena de Indias en el siglo XVII: Relaciones, diferencias y politicas de las autoridades. *Revista Interdisciplinaria de Trabajos sobre las Américas,* 5; Available online at www.revue-rita.com.

Navarrete, M.C. (2012). Palenques: cimarrones y castas en el Caribe colombiano – Sus relaciones sociales (siglo XVII). In: *Palenque (Colombia): oralidad, identidad y resistencia. Un enfoque interdisciplinario*. Maglia, G. & Schwegler, A. (eds.). Bogota: Instituto Caro y Cuervo & Universidad Javeriana.

Noguera-Santamaria, M.C., Anderson, C.E., Uricoechea, D., C., D., Briceno-Balcazar, I. & Bernal-Villegas, J. (2015). Mitochondrial DNA analysis suggests a Chibchan migration into Colombia. *Univ Sci,* 20(2); 261-278.

Noguera, M.C., Schwegler, A., Gomes, V., Briceno, I., Alvarez, L., Uricoechea, D., Amorim, A., Benavides, E., Silvera, C., Charris, M., Bernal, J.E. & Gusmao, L. (2013). Colombia's racial crucible: Y chromosome evidence from six admixed communities in the Department of Bolivar. *Ann Hum Biol,* 41(5); 453-459.

Novembre, J., Johnson, T., Bryc, K., Kutalik, Z., Boyko, A.R., Auton, A., Indap, A., King, K.S., Bergmann, S., Nelson, M.R., Stephens, M. & Bustamante, C.D. (2008). Genes mirror geography within Europe. *Nature,* 456(7218); 98-101.

Novembre, J. & Ramachandran, S. (2011). Perspectives on human population structure at the cusp of the sequencing era. *Annu Rev Genomics Hum Genet,* 12; 245-274.

Padhukasahasram, B. (2014). Inferring ancestry from population genomic data and its applications. *Front Genet,* 5; 204.

Pagani, L., Kivisild, T., Tarekegn, A., Ekong, R., Plaster, C., Gallego Romero, I., Ayub, Q., Mehdi, S.Q., Thomas, M.G., Luiselli, D., Bekele, E., Bradman, N., Balding, D.J. & Tyler-Smith, C. (2012). Ethiopian genetic diversity reveals linguistic stratification and complex influences on the Ethiopian gene pool. *Am J Hum Genet,* 91(1); 83-96.

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

Pakendorf, B., de Filippo, C. & Bostoen, K. (2011). Molecular Perspectives on the Bantu Expansion: A Synthesis. *Language Dynamics and Change,* 1(1); 50-88.

Pardo-Seco, J., Martinon-Torres, F. & Salas, A. (2014). Evaluating the accuracy of AIM panels at quantifying genome ancestry. *BMC Genomics,* 15; 543.

Parra, E.J., Marcini, A., Akey, J., Martinson, J., Batzer, M.A., Cooper, R., Forrester, T., Allison, D.B., Deka, R., Ferrell, R.E. & Shriver, M.D. (1998). Estimating African American admixture proportions by use of population-specific alleles. *Am J Hum Genet,* 63(6); 1839-1851.

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., McKeigue, P.M., Kamboh, M.I., Ferrell, R.E., Pollitzer, W.S. & Shriver, M.D. (2001). Ancestral proportions and admixture dynamics in geographically defined African Americans living in South Carolina. *Am J Phys Anthropol,* 114(1); 18-29.

Parra, F.C., Amado, R.C., Lambertucci, J.R., Rocha, J., Antunes, C.M. & Pena, S.D. (2003). Color and genomic ancestry in Brazilians. *Proc Natl Acad Sci U S A,* 100(1); 177-182.

Pasaniuc, B., Sankararaman, S., Torgerson, D.G., Gignoux, C., Zaitlen, N., Eng, C., Rodriguez-Cintron, W., Chapela, R., Ford, J.G., Avila, P.C., Rodriguez-Santana, J., Chen, G.K., Le Marchand, L., Henderson, B., Reich, D., Haiman, C.A., Gonzalez Burchard, E. & Halperin, E. (2013). Analysis of Latino populations from GALA and MEC studies reveals genomic loci with biased local ancestry estimation. *Bioinformatics,* 29(11); 1407-1415.

Patin, E., Siddle, K.J., Laval, G., Quach, H., Harmant, C., Becker, N., Froment, A., Regnault, B., Lemee, L., Gravel, S., Hombert, J.M., Van der Veen, L., Dominy, N.J., Perry, G.H., Barreiro, L.B., Verdu, P., Heyer, E. & Quintana-Murci, L. (2014). The impact of agricultural emergence on the genetic history of African rainforest hunter-gatherers and agriculturalists. *Nat Commun,* 5; 3163.

Patterson, N., Price, A.L. & Reich, D. (2006). Population structure and eigenanalysis. *PLoS Genet,* 2(12); e190.

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T. & Reich, D. (2012). Ancient admixture in human history. *Genetics,* 192(3); 1065-1093.

Patterson, O. (1990). *Slavery & Social Death - A Comparative Study.* Cambridge, USA: Harvard University Press.

Paul, L.M., Gary, F., Simons, G.F. & Fennig, C.D. (2015). Ethnologue: Languages of the World, Seventeenth edition. Texas: SIL International. Website: http://www.ethnologue.com.

Pemberton, T.J., Absher, D., Feldman, M.W., Myers, R.M., Rosenberg, N.A. & Li, J.Z. (2012). Genomic patterns of homozygosity in worldwide human populations. *Am J Hum Genet,* 91(2); 275-292.

Pemberton, T.J. & Rosenberg, N.A. (2014). Population-genetic influences on genomic estimates of the inbreeding coefficient: a global perspective. *Hum Hered,* 77(1-4); 37-48.

Pena, S.D., Bastos-Rodrigues, L., Pimenta, J.R. & Bydlowski, S.P. (2009). DNA tests probe the genomic ancestry of Brazilians. *Braz J Med Biol Res,* 42(10); 870-876.

Pena, S.D., Di Pietro, G., Fuchshuber-Moraes, M., Genro, J.P., Hutz, M.H., Kehdy Fde, S., Kohlrausch, F., Magno, L.A., Montenegro, R.C., Moraes, M.O., de Moraes, M.E., de Moraes, M.R., Ojopi, E.B., Perini, J.A., Racciopi, C., Ribeiro-Dos-Santos, A.K., Rios-Santos, F., Romano-Silva,

M.A., Sortica, V.A. & Suarez-Kurtz, G. (2011). The genomic ancestry of individuals from different geographical regions of Brazil is more uniform than expected. *PLoS One,* 6(2); e17063.

Pereira, L., Macaulay, V., Torroni, A., Scozzari, R., Prata, M.J. & Amorim, A. (2001). Prehistoric and historic traces in the mtDNA of Mozambique: insights into the Bantu expansions and the slave trade. *Ann Hum Genet,* 65(Pt 5); 439-458.

Pereira, L., Freitas, F., Fernandes, V., Pereira, J.B., Costa, M.D., Costa, S., Maximo, V., Macaulay, V., Rocha, R. & Samuels, D.C. (2009). The diversity present in 5140 human mitochondrial genomes. *Am J Hum Genet,* 84(5); 628-640.

Pereira, R., Phillips, C., Pinto, N., Santos, C., dos Santos, S.E., Amorim, A., Carracedo, A. & Gusmao, L. (2012). Straightforward inference of ancestry and admixture proportions through ancestry-informative insertion deletion multiplexing. *PLoS One,* 7(1); e29684.

Perera, M.A., Cavallari, L.H., Limdi, N.A., Gamazon, E.R., Konkashbaev, A., Daneshjou, R., Pluzhnikov, A., Crawford, D.C., Wang, J., Liu, N., Tatonetti, N., Bourgeois, S., Takahashi, H., Bradford, Y., Burkley, B.M., Desnick, R.J., Halperin, J.L., Khalifa, S.I., Langaee, T.Y., Lubitz, S.A., Nutescu, E.A., Oetjens, M., Shahin, M.H., Patel, S.R., Sagreiya, H., Tector, M., Weck, K.E., Rieder, M.J., Scott, S.A., Wu, A.H., et al. (2013). Genetic variants associated with warfarin dose in African-American individuals: a genome-wide association study. *Lancet,* 382(9894); 790-796.

Perry, G.H., Foll, M., Grenier, J.C., Patin, E., Nedelec, Y., Pacis, A., Barakatt, M., Gravel, S., Zhou, X., Nsobya, S.L., Excoffier, L., Quintana-Murci, L., Dominy, N.J. & Barreiro, L.B. (2014). Adaptive, convergent origins of the pygmy phenotype in African rainforest hunter-gatherers. *Proc Natl Acad Sci U S A,* 111(35); E3596-3603.

Pickrell, J.K., Patterson, N., Barbieri, C., Berthold, F., Gerlach, L., Guldemann, T., Kure, B., Mpoloka, S.W., Nakagawa, H., Naumann, C., Lipson, M., Loh, P.R., Lachance, J., Mountain, J., Bustamante, C.D., Berger, B., Tishkoff, S.A., Henn, B.M., Stoneking, M., Reich, D. & Pakendorf, B. (2012). The genetic prehistory of southern Africa. *Nat Commun,* 3; 1143.

Pickrell, J.K. & Pritchard, J.K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet,* 8(11); e1002967.

Pierron, D., Razafindrazaka, H., Pagani, L., Ricaut, F.X., Antao, T., Capredon, M., Sambo, C., Radimilahy, C., Rakotoarisoa, J.A., Blench, R.M., Letellier, T. & Kivisild, T. (2014). Genome-wide evidence of Austronesian-Bantu admixture and cultural reversion in a hunter-gatherer group of Madagascar. *Proc Natl Acad Sci U S A,* 111(3); 936-941.

Piko, L. & Matsumoto, L. (1976). Number of mitochondria and some properties of mitochondrial DNA in the mouse egg. *Dev Biol,* 49(1); 1-10.

Plancoulaine, S., Buigues, R.P., Murphy, E.L., van Beveren, M., Pouliquen, J.F., Joubert, M., Remy, F., Tuppin, P., Tortevoye, P., de The, G., Moreau, J.P. & Gessain, A. (1998). Demographic and familial characteristics of HTLV-1 infection among an isolated, highly endemic population of African origin in French Guiana. *Int J Cancer,* 76(3); 331-336.

Plancoulaine, S., Gessain, A., Joubert, M., Tortevoye, P., Jeanne, I., Talarmin, A., de The, G. & Abel, L. (2000). Detection of a major gene predisposing to human T lymphotropic virus type I infection in children among an endemic population of African origin. *J Infect Dis,* 182(2); 405-412.

Plancoulaine, S., Gessain, A., Tortevoye, P., Boland-Auge, A., Vasilescu, A., Matsuda, F. & Abel, L. (2006). A major susceptibility locus for HTLV-1 infection in childhood maps to chromosome 6q27. *Hum Mol Genet,* 15(22); 3306-3312.

Poston, D.L., Jr. & Mei, Y. (1990). The Distribution of the Overseas Chinese in the Contemporary World. *International Migration Review,* 24(3); 480-508.

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. & Bustamante, C.D. (2013). Sequencing Y chromosomes resolves discrepancy in time to common ancestor of males versus females. *Science,* 341(6145); 562-565.

Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A. & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet,* 38(8); 904-909.

Price, A.L., Tandon, A., Patterson, N., Barnes, K.C., Rafaels, N., Ruczinski, I., Beaty, T.H., Mathias, R., Reich, D. & Myers, S. (2009). Sensitive detection of chromosomal segments of distinct ancestry in admixed populations. *PLoS Genet,* 5(6); e1000519.

Price, A.L., Zaitlen, N.A., Reich, D. & Patterson, N. (2010). New approaches to population stratification in genome-wide association studies. *Nat Rev Genet,* 11(7); 459-463.

Price, R. (1994). *Les Premiers Temps: la conception de l'histoire des Marrons Saramaka.* Paris, France: Seuil. Recherches anthropologiques.

Price, R. (1995). Executing Ethnicity: The Killings in Suriname. *Cultural Anthropology,* 10(4); 437-471.

Price, R. (1996). *Maroon Societies: rebel slave communities in the Americas.* Baltimore, Maryland: John Hopkins University Press.

Price, R. (1998). Scrapping Maroon History: Brazil's Promise, Suriname's Shame. *New West Indian Guide,* 72; 233-255.

Price, R. (2001). The miracle of creolization : a retrospective. *New West Indian Guide,* 75(2); 35-64.

Price, R. (2002). Maroons in Suriname and Guyane: how many and where. *New West Indian Guide,* 76; 81-88.

Price, R. & Price, S. (2003). *Cultures en Guyane: Les Marrons.* France: Châteauneuf-le-Rouge: Vents d'ailleurs.

Price, R. (2008). *Travels with Tooy: History, Memory, and the African American Imagination.* Chicago, USA: University of Chicago Press.

Price, S. & Price, R. (2001). Maroons under assault in Suriname and French Guiana. *Cultural Survival,* 25.4 (Winter 2001); Maroons in the Americas.

Pritchard, J.K., Stephens, M. & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics,* 155(2); 945-959.

Pugach, I. & Stoneking, M. (2015). Genome-wide insights into the genetic history of human populations. *Investig Genet,* 6; 6.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J. & Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet,* 81(3); 559-575.

Quintana-Murci, L., Semino, O., Bandelt, H.J., Passarino, G., McElreavey, K. & Santachiara-Benerecetti, A.S. (1999). Genetic evidence of an early exit of Homo sapiens sapiens from Africa through eastern Africa. *Nat Genet,* 23(4); 437-441.

Quintana-Murci, L. & Fellous, M. (2001). The Human Y Chromosome: The Biological Role of a "Functional Wasteland". *J Biomed Biotechnol,* 1(1); 18-24.

Quintana-Murci, L., Quach, H., Harmant, C., Luca, F., Massonnet, B., Patin, E., Sica, L., Mouguiama-Daouda, P., Comas, D., Tzur, S., Balanovsky, O., Kidd, K.K., Kidd, J.R., van der Veen, L., Hombert, J.M., Gessain, A., Verdu, P., Froment, A., Bahuchet, S., Heyer, E., Dausset, J., Salas, A. & Behar, D.M. (2008). Maternal traces of deep common ancestry and asymmetric gene flow between Pygmy hunter-gatherers and Bantu-speaking farmers. *Proc Natl Acad Sci U S A,* 105(5); 1596-1601.

R Core Team (2014). R: A language and environment for statistical computing. Vienna, Austria: http://www.R-project.org/.

Raghavan, M., Steinrucken, M., Harris, K., Schiffels, S., Rasmussen, S., DeGiorgio, M., Albrechtsen, A., Valdiosera, C., Avila-Arcos, M.C., Malaspinas, A.S., Eriksson, A., Moltke, I., Metspalu, M., Homburger, J.R., Wall, J., Cornejo, O.E., Moreno-Mayar, J.V., Korneliussen, T.S., Pierre, T., Rasmussen, M., Campos, P.F., Damgaard Pde, B., Allentoft, M.E., Lindo, J., Metspalu, E., Rodriguez-Varela, R., Mansilla, J., Henrickson, C., Seguin-Orlando, A., Malmstrom, H., et al. (2015). Genomic evidence for the Pleistocene and recent population history of Native Americans. *Science,* 349(6250); aab3884.

Ramos, A., Santos, C., Mateiu, L., Gonzalez Mdel, M., Alvarez, L., Azevedo, L., Amorim, A. & Aluja, M.P. (2013). Frequency and pattern of heteroplasmy in the complete human mitochondrial genome. *PLoS One,* 8(10); e74636.

Reich, D., Thangaraj, K., Patterson, N., Price, A.L. & Singh, L. (2009). Reconstructing Indian population history. *Nature,* 461(7263); 489-494.

Reich, D., Patterson, N., Campbell, D., Tandon, A., Mazieres, S., Ray, N., Parra, M.V., Rojas, W., Duque, C., Mesa, N., Garcia, L.F., Triana, O., Blair, S., Maestre, A., Dib, J.C., Bravi, C.M., Bailliet, G., Corach, D., Hunemeier, T., Bortolini, M.C., Salzano, F.M., Petzl-Erler, M.L., Acuna-Alonzo, V., Aguilar-Salinas, C., Canizales-Quinteros, S., Tusie-Luna, T., Riba, L., Rodriguez-Cruz, M., Lopez-Alarcon, M., Coral-Vazquez, R., et al. (2012). Reconstructing Native American population history. *Nature,* 488(7411); 370-374.

Renault-Lescure, O. & Goury, L. (2009). *Langues de Guyane.* Marseille: Vents d'ailleurs. Cultures en Guyane.

Reynolds, J., Weir, B.S. & Cockerham, C.C. (1983). Estimation of the coancestry coefficient: basis for a short-term genetic distance. *Genetics,* 105(3); 767-779.

Rieux, A., Eriksson, A., Li, M., Sobkowiak, B., Weinert, L.A., Warmuth, V., Ruiz-Linares, A., Manica, A. & Balloux, F. (2014). Improved calibration of the human mitochondrial clock using ancient genomes. *Mol Biol Evol,* 31(10); 2780-2792.

Risch, N., Choudhry, S., Via, M., Basu, A., Sebro, R., Eng, C., Beckman, K., Thyne, S., Chapela, R., Rodriguez-Santana, J.R., Rodriguez-Cintron, W., Avila, P.C., Ziv, E. & Gonzalez Burchard, E. (2009). Ancestry-related assortative mating in Latino populations. *Genome Biol,* 10(11); R132.

Rito, T., Richards, M.B., Fernandes, V., Alshamali, F., Cerny, V., Pereira, L. & Soares, P. (2013). The first modern human dispersals across Africa. *PLoS One,* 8(11); e80031.

Robin, E.D. & Wong, R. (1988). Mitochondrial DNA molecules and virtual number of mitochondria per cell in mammalian cells. *J Cell Physiol,* 136(3); 507-513.

Rodas, C., Gelvez, N. & Keyeux, G. (2003). Mitochondrial DNA studies show asymmetrical Amerindian admixture in Afro-Colombian and Mestizo populations. *Hum Biol,* 75(1); 13-30.

Rodrigues de Moura, R., Coelho, A.V., de Queiroz Balbino, V., Crovella, S. & Brandao, L.A. (2015). Meta-analysis of Brazilian genetic admixture and comparison with other Latin America countries. *Am J Hum Biol,* 27(5); 674-680.

Roewer, L. & 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(2); 163-171.

Roewer, L. (2009). Y chromosome STR typing in crime casework. *Forensic Sci Med Pathol,* 5(2); 77-84.

Roewer, L., Nothnagel, M., Gusmao, L., Gomes, V., Gonzalez, M., Corach, D., Sala, A., Alechine, E., Palha, T., Santos, N., Ribeiro-Dos-Santos, A., Geppert, M., Willuweit, S., Nagy, M., Zweynert, S., Baeta, M., Nunez, C., Martinez-Jarreta, B., Gonzalez-Andrade, F., Fagundes de Carvalho, E., da Silva, D.A., Builes, J.J., Turbon, D., Lopez Parra, A.M., Arroyo-Pardo, E., Toscanini, U., Borjas, L., Barletta, C., Ewart, E., Santos, S., et al. (2013). Continent-wide decoupling of Y-chromosomal genetic variation from language and geography in native South Americans. *PLoS Genet,* 9(4); e1003460.

Rojas, W., Parra, M.V., Campo, O., Caro, M.A., Lopera, J.G., Arias, W., Duque, C., Naranjo, A., Garcia, J., Vergara, C., Lopera, J., Hernandez, E., Valencia, A., Caicedo, Y., Cuartas, M., Gutierrez, J., Lopez, S., Ruiz-Linares, A. & Bedoya, G. (2010). Genetic make up and structure of Colombian populations by means of uniparental and biparental DNA markers. *Am J Phys Anthropol,* 143(1); 13-20.

Rootsi, S., Myres, N.M., Lin, A.A., Jarve, M., King, R.J., Kutuev, I., Cabrera, V.M., Khusnutdinova, E.K., Varendi, K., Sahakyan, H., Behar, D.M., Khusainova, R., Balanovsky, O., Balanovska, E., Rudan, P., Yepiskoposyan, L., Bahmanimehr, A., Farjadian, S., Kushniarevich, A., Herrera, R.J., Grugni, V., Battaglia, V., Nici, C., Crobu, F., Karachanak, S., Hooshiar Kashani, B., Houshmand, M., Sanati, M.H., Toncheva, D., Lisa, A., et al. (2012). Distinguishing the co-ancestries of haplogroup G Y-chromosomes in the populations of Europe and the Caucasus. *Eur J Hum Genet,* 20(12); 1275-1282.

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

Rosa, A. & Brehem, A. (2011). African human mtDNA phylogeography at-a-glance. *J Anthropol Sci,* 89; 25-58.

Rosenberg, N.A., Pritchard, J.K., Weber, J.L., Cann, H.M., Kidd, K.K., Zhivotovsky, L.A. & Feldman, M.W. (2002). Genetic structure of human populations. *Science,* 298(5602); 2381-2385.

Rosenberg, N.A., Li, L.M., Ward, R. & Pritchard, J.K. (2003). Informativeness of genetic markers for inference of ancestry. *Am J Hum Genet,* 73(6); 1402-1422.

Rosenberg, N.A. (2011). A population-genetic perspective on the similarities and differences among worldwide human populations. *Hum Biol,* 83(6); 659-684.

Rotimi, C., Puras, A., Cooper, R., McFarlane-Anderson, N., Forrester, T., Ogunbiyi, O., Morrison, L. & Ward, R. (1996). Polymorphisms of renin-angiotensin genes among Nigerians, Jamaicans, and African Americans. *Hypertension,* 27(3 Pt 2); 558-563.

Rousset, F. (1996). Equilibrium values of measures of population subdivision for stepwise mutation processes. *Genetics,* 142(4); 1357-1362.

Royal, C.D., Novembre, J., Fullerton, S.M., Goldstein, D.B., Long, J.C., Bamshad, M.J. & Clark, A.G. (2010). Inferring genetic ancestry: opportunities, challenges, and implications. *Am J Hum Genet,* 86(5); 661-673.

Ruiz-Linares, A., Adhikari, K., Acuna-Alonzo, V., Quinto-Sanchez, M., Jaramillo, C., Arias, W., Fuentes, M., Pizarro, M., Everardo, P., de Avila, F., Gomez-Valdes, J., Leon-Mimila, P., Hunemeier, T., Ramallo, V., Silva de Cerqueira, C.C., Burley, M.W., Konca, E., de Oliveira, M.Z., Veronez, M.R., Rubio-Codina, M., Attanasio, O., Gibbon, S., Ray, N., Gallo, C., Poletti, G., Rosique, J., Schuler-Faccini, L., Salzano, F.M., Bortolini, M.C., Canizales-Quinteros, S., et al. (2014). Admixture in Latin America: geographic structure, phenotypic diversity and self-perception of ancestry based on 7,342 individuals. *PLoS Genet,* 10(9); e1004572.

Ruiz-Pesini, E., Lott, M.T., Procaccio, V., Poole, J.C., Brandon, M.C., Mishmar, D., Yi, C., Kreuziger, J., Baldi, P. & Wallace, D.C. (2007). An enhanced MITOMAP with a global mtDNA mutational phylogeny. *Nucleic Acids Res,* 35(Database issue); D823-828.

Russell, T., Silva, F. & Steele, J. (2014). Modelling the spread of farming in the Bantu-speaking regions of Africa: an archaeology-based phylogeography. *PLoS One,* 9(1); e87854.

Saillard, J., Forster, P., Lynnerup, N., Bandelt, H.J. & Norby, S. (2000). mtDNA variation among Greenland Eskimos: the edge of the Beringian expansion. *Am J Hum Genet,* 67(3); 718-726.

Salas, A., Richards, M., De la Fe, T., Lareu, M.V., Sobrino, B., Sanchez-Diz, P., Macaulay, V. & Carracedo, A. (2002). The making of the African mtDNA landscape. *Am J Hum Genet,* 71(5); 1082-1111.

Salas, A., Richards, M., Lareu, M.V., Scozzari, R., Coppa, A., Torroni, A., Macaulay, V. & Carracedo, A. (2004). The African diaspora: mitochondrial DNA and the Atlantic slave trade. *Am J Hum Genet,* 74(3); 454-465.

Salas, A., Carracedo, A., Richards, M. & Macaulay, V. (2005a). Charting the ancestry of African Americans. *Am J Hum Genet,* 77(4); 676-680.

Salas, A., Richards, M., Lareu, M.V., Sobrino, B., Silva, S., Matamoros, M., Macaulay, V. & Carracedo, A. (2005b). Shipwrecks and founder effects: divergent demographic histories reflected in Caribbean mtDNA. *Am J Phys Anthropol,* 128(4); 855-860.

Salas, A., Acosta, A., Alvarez-Iglesias, V., Cerezo, M., Phillips, C., Lareu, M.V. & Carracedo, A. (2008). The mtDNA ancestry of admixed Colombian populations. *Am J Hum Biol,* 20(5); 584-591.

Salzano, F.M. & Bortolini, M.C. (2002). *The evolution and genetics of Latin American populations.* Cambridge ; New York: Cambridge University Press.

Salzano, F.M. (2004). Interethnic variability and admixture in Latin America--social implications. *Rev Biol Trop,* 52(3); 405-415.

Sanger, F., Nicklen, S. & Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A,* 74(12); 5463-5467.

Sankararaman, S., Sridhar, S., Kimmel, G. & Halperin, E. (2008). Estimating local ancestry in admixed populations. *Am J Hum Genet,* 82(2); 290-303.

Sans, M., Weimer, T.A., Franco, M.H., Salzano, F.M., Bentancor, N., Alvarez, I., Bianchi, N.O. & Chakraborty, R. (2002). Unequal contributions of male and female gene pools from parental populations in the African descendants of the city of Melo, Uruguay. *Am J Phys Anthropol,* 118(1); 33-44.

Santos, C., Fregel, R., Cabrera, V.M., Alvarez, L., Larruga, J.M., Ramos, A., Lopez, M.A., Pilar Aluja, M. & Gonzalez, A.M. (2014). Mitochondrial DNA and Y-chromosome structure at the Mediterranean and Atlantic facades of the Iberian Peninsula. *Am J Hum Biol,* 26(2); 130-141.

Santos, N.P., Ribeiro-Rodrigues, E.M., Ribeiro-Dos-Santos, A.K., Pereira, R., Gusmao, L., Amorim, A., Guerreiro, J.F., Zago, M.A., Matte, C., Hutz, M.H. & Santos, S.E. (2010). Assessing individual interethnic admixture and population substructure using a 48-insertion-deletion (INSEL) ancestry-informative marker (AIM) panel. *Hum Mutat,* 31(2); 184-190.

Scheinfeldt, L.B., Soi, S. & Tishkoff, S.A. (2010). Colloquium paper: working toward a synthesis of archaeological, linguistic, and genetic data for inferring African population history. *Proc Natl Acad Sci U S A,* 107 Suppl 2; 8931-8938.

Schlebusch, C.M., Skoglund, P., Sjodin, P., Gattepaille, L.M., Hernandez, D., Jay, F., Li, S., De Jongh, M., Singleton, A., Blum, M.G., Soodyall, H. & Jakobsson, M. (2012). Genomic variation in seven Khoe-San groups reveals adaptation and complex African history. *Science,* 338(6105); 374-379.

Schroeder, H., O'Connell, T.C., Evans, J.A., Shuler, K.A. & Hedges, R.E. (2009). Trans-Atlantic slavery: isotopic evidence for forced migration to Barbados. *Am J Phys Anthropol,* 139(4); 547-557.

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., Alkan, C., Kidd, J.M., Sun, Y., Drautz, D.I., Bouffard, P., Muzny, D.M., Reid, J.G., Nazareth, L.V., Wang, Q., Burhans, R., Riemer, C., Wittekindt, N.E., Moorjani, P., Tindall, E.A., Danko, C.G., Teo, W.S., Buboltz, A.M., Zhang, Z., Ma, Q., Oosthuysen, A., et al. (2010). Complete Khoisan and Bantu genomes from southern Africa. *Nature,* 463(7283); 943-947.

Schwartz, S.B. (1992). Rethinking Palmares: Slave resistance in Colonial Brazil. In: *Slaves, Peasants and Rebels*. Schwartz, S.B. (ed. Urbana, USA: University of Illinois Press.

Schwartz, S.B. (1995). *Slaves, Peasants and Rebels: Reconsidering Brazilian Slavery.* Illinois, USA: University of Illinois Press. Blacks in the New World.

Schwegler, A. (2006). Bantu elements in Palenque (Colombia). In: *African Re-genesis: Confronting Social Issues in the Diaspora*. Haviser, J.B. & Macdonald, K.C. (eds.). Abingdon, UK: UCL Press.

Schwegler, A. (2011). Palenque-ro. The search for its African substrate. In: *Creoles, their Substrates, and Language Typology*. Lefebvre, C. (ed. Amsterdam: John Benjamins Publishing Company.

Scozzari, R., Massaia, A., Trombetta, B., Bellusci, G., Myres, N.M., Novelletto, A. & Cruciani, F. (2014). An unbiased resource of novel SNP markers provides a new chronology for the human Y chromosome and reveals a deep phylogenetic structure in Africa. *Genome Res,* 24(3); 535-544.

Secher, B., Fregel, R., Larruga, J.M., Cabrera, V.M., Endicott, P., Pestano, J.J. & Gonzalez, A.M. (2014). The history of the North African mitochondrial DNA haplogroup U6 gene flow into the African, Eurasian and American continents. *BMC Evol Biol,* 14; 109.

Shringarpure, S. & Xing, E.P. (2014). Effects of sample selection bias on the accuracy of population structure and ancestry inference. *G3 (Bethesda),* 4(5); 901-911.

Shriver, M.D. & Kittles, R.A. (2004). Genetic ancestry and the search for personalized genetic histories. *Nat Rev Genet,* 5(8); 611-618.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., Thompson, J.D. & Higgins, D.G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol,* 7; 539.

Silva-Zolezzi, I., Hidalgo-Miranda, A., Estrada-Gil, J., Fernandez-Lopez, J.C., Uribe-Figueroa, L., Contreras, A., Balam-Ortiz, E., del Bosque-Plata, L., Velazquez-Fernandez, D., Lara, C., Goya, R., Hernandez-Lemus, E., Davila, C., Barrientos, E., March, S. & Jimenez-Sanchez, G. (2009). Analysis of genomic diversity in Mexican Mestizo populations to develop genomic medicine in Mexico. *Proc Natl Acad Sci U S A,* 106(21); 8611-8616.

Silva, M., Alshamali, F., Silva, P., Carrilho, C., Mandlate, F., Jesus Trovoada, M., Cerny, V., Pereira, L. & Soares, P. (2015). 60,000 years of interactions between Central and Eastern Africa documented by major African mitochondrial haplogroup L2. *Sci Rep,* 5; 12526.

Simms, T.M., Martinez, E., Herrera, K.J., Wright, M.R., Perez, O.A., Hernandez, M., Ramirez, E.C., McCartney, Q. & Herrera, R.J. (2011). Paternal lineages signal distinct genetic contributions from British Loyalists and continental Africans among different Bahamian islands. *American Journal of Physical Anthropology,* 146(4); 594-608.

Simms, T.M., Wright, M.R., Hernandez, M., Perez, O.A., Ramirez, E.C., Martinez, E. & Herrera, R.J. (2012). Y-chromosomal diversity in Haiti and Jamaica: contrasting levels of sex-biased gene flow. *Am J Phys Anthropol,* 148(4); 618-631.

Simms, T.M., Wright, M.R., Martinez, E., Regueiro, M., McCartney, Q. & Herrera, R.J. (2013). Y-STR diversity and sex-biased gene flow among Caribbean populations. *Gene,* 516(1); 82-92.

Skoglund, P., Mallick, S., Bortolini, M.C., Chennagiri, N., Hunemeier, T., Petzl-Erler, M.L., Salzano, F.M., Patterson, N. & Reich, D. (2015). Genetic evidence for two founding populations of the Americas. *Nature*.

Slatkin, M. (1995). A measure of population subdivision based on microsatellite allele frequencies. *Genetics,* 139(1); 457-462.

Smith, N. (1987). The genesis of the creole languages of Surinam. Amsterdam, Netherlands: University of Amsterdam.

Smith, N. (2002). The history of the Surinamese Creoles II: origin and differentiation. In: *Atlas of the Languages of Surinam*. Carlin, E. & Arends, J. (eds.). Leiden, Netherlands: KITLV Press.

Smith, N. & Cardoso, H. (2004). A new look at the Portuguese element in Saramaccan. *Journal of Portuguese Linguistics,* 3(2); 115-147.

Smith, V.C., Pearce, N.J.G., Matthews, N.E., Westgate, J.A., Petraglia, M.D., Haslam, M., Lane, C.S., Korisettar, R. & Pal, J.N. (2011). Geochemical fingerprinting of the widespread Toba tephra using biotite compositions. *Quaternary International,* 246(1–2); 97-104.

Soares, P., Ermini, L., Thomson, N., Mormina, M., Rito, T., Rohl, A., Salas, A., Oppenheimer, S., Macaulay, V. & Richards, M.B. (2009). Correcting for purifying selection: an improved human mitochondrial molecular clock. *Am J Hum Genet,* 84(6); 740-759.

Soares, P., Achilli, A., Semino, O., Davies, W., Macaulay, V., Bandelt, H.J., Torroni, A. & Richards, M.B. (2010). The archaeogenetics of Europe. *Curr Biol,* 20(4); R174-183.

Soares, P., Alshamali, F., Pereira, J.B., Fernandes, V., Silva, N.M., Afonso, C., Costa, M.D., Musilova, E., Macaulay, V., Richards, M.B., Cerny, V. & Pereira, L. (2012). The Expansion of mtDNA Haplogroup L3 within and out of Africa. *Mol Biol Evol,* 29(3); 915-927.

Soares, P., Abrantes, D., Rito, T., Thomson, N., Radivojac, P., Li, B., Macaulay, V., Samuels, D.C. & Pereira, L. (2013). Evaluating purifying selection in the mitochondrial DNA of various mammalian species. *PLoS One,* 8(3); e58993.

Sole-Morata, N., Bertranpetit, J., Comas, D. & Calafell, F. (2014). Recent radiation of R-M269 and high Y-STR haplotype resemblance confirmed. *Ann Hum Genet,* 78(4); 253-254.

Stedman, J.G., Price, R. & Price, S. (2010). *Narrative of a Five Years Expedition against the Revolted Negroes of Surinam.* Bloomington, USA: iUniverse.

Steemers, F.J., Chang, W., Lee, G., Barker, D.L., Shen, R. & Gunderson, K.L. (2006). Whole-genome genotyping with the single-base extension assay. *Nat Methods,* 3(1); 31-33.

Stefflova, K., Dulik, M.C., Pai, A.A., Walker, A.H., Zeigler-Johnson, C.M., Gueye, S.M., Schurr, T.G. & Rebbeck, T.R. (2009). Evaluation of group genetic ancestry of populations from Philadelphia and Dakar in the context of sex-biased admixture in the Americas. *PLoS One,* 4(11); e7842.

Stefflova, K., Dulik, M.C., Barnholtz-Sloan, J.S., Pai, A.A., Walker, A.H. & Rebbeck, T.R. (2011). Dissecting the within-Africa ancestry of populations of African descent in the Americas. *PLoS One,* 6(1); e14495.

Stoneking, M. & Delfin, F. (2010). The human genetic history of East Asia: weaving a complex tapestry. *Curr Biol,* 20(4); R188-193.

Tajima, F. (1993). Measurement of DNA polymorphism. In: *Mechanisms of Molecular Evolution. Introduction to Molecular Paleopopulation Biology*. Takahata, N. & Clark, A.G. (eds.). Tokyo, Japan: Japan Scientific Societies Press, Sinauer Associates, Inc.

Tajima, F. (1996). The amount of DNA polymorphism maintained in a finite population when the neutral mutation rate varies among sites. *Genetics,* 143(3); 1457-1465.

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., Fedorova, S.A., Golubenko, M.V., Stepanov, V.A., Gubina, M.A., Zhadanov, S.I., Ossipova, L.P., Damba, L., Voevoda, M.I., Dipierri, J.E., Villems, R. & Malhi, R.S. (2007). Beringian standstill and spread of Native American founders. *PLoS One,* 2(9); e829.

Tang, H., Peng, J., Wang, P. & Risch, N.J. (2005). Estimation of individual admixture: analytical and study design considerations. *Genet Epidemiol,* 28(4); 289-301.

Tang, H., Coram, M., Wang, P., Zhu, X. & Risch, N. (2006). Reconstructing genetic ancestry blocks in admixed individuals. *Am J Hum Genet,* 79(1); 1-12.

Tang, H., Choudhry, S., Mei, R., Morgan, M., Rodriguez-Cintron, W., Burchard, E.G. & Risch, N.J. (2007). Recent genetic selection in the ancestral admixture of Puerto Ricans. *Am J Hum Genet,* 81(3); 626-633.

Tapp, N. (2005). Hmong Diaspora. In: *Encyclopedia of Diasporas: immigrant and refugee cultures around the world.* New York: Springer.

Thangaraj, K., Chaubey, G., Kivisild, T., Selvi Rani, D., Singh, V.K., Ismail, T., Carvalho-Silva, D., Metspalu, M., Bhaskar, L.V., Reddy, A.G., Chandra, S., Pande, V., Prathap Naidu, B., Adarsh, N., Verma, A., Jyothi, I.A., Mallick, C.B., Shrivastava, N., Devasena, R., Kumari, B., Singh, A.K., Dwivedi, S.K., Singh, S., Rao, G., Gupta, P., Sonvane, V., Kumari, K., Basha, A., Bhargavi, K.R., Lalremruata, A., et al. (2008). Maternal footprints of Southeast Asians in North India. *Hum Hered,* 66(1); 1-9.

Thiaw, I. (2010). L'archéologie de l'île de Gorée au Sénégal : Chaque maison a une histoire. In: *La construction transatlantique d'identités noires : entre Afrique et Amériques*. Sansone, L., E., S. & Barry, B. (eds.). Paris, France: Éditions Karthala.

Thiaw, I. (2011). Slaves without Shackles: An Archaeology of Everyday Life on Gorée Island, Senegal. In: *Slavery in Africa: Archaeology and Memory*. Lane, P. & Macdonald, K.C. (eds.). London, UK: British Academy.

Thomas, H. (1999). *The slave trade: the history of the Atlantic slave trade: 1440–1870.* London, UK: Simon & Schuster.

Tian, C., Kosoy, R., Lee, A., Ransom, M., Belmont, J.W., Gregersen, P.K. & Seldin, M.F. (2008). Analysis of East Asia genetic substructure using genome-wide SNP arrays. *PLoS One,* 3(12); e3862.

Tishkoff, S.A. & Verrelli, B.C. (2003). Patterns of human genetic diversity: implications for human evolutionary history and disease. *Annu Rev Genomics Hum Genet,* 4; 293-340.

Tishkoff, S.A., Gonder, M.K., Henn, B.M., Mortensen, H., Knight, A., Gignoux, C., Fernandopulle, N., Lema, G., Nyambo, T.B., Ramakrishnan, U., Reed, F.A. & Mountain, J.L. (2007). History of click-speaking populations of Africa inferred from mtDNA and Y chromosome genetic variation. *Mol Biol Evol,* 24(10); 2180-2195.

Tishkoff, S.A., Reed, F.A., Friedlaender, F.R., Ehret, C., Ranciaro, A., Froment, A., Hirbo, J.B., Awomoyi, A.A., Bodo, J.M., Doumbo, O., Ibrahim, M., Juma, A.T., Kotze, M.J., Lema, G., Moore, J.H., Mortensen, H., Nyambo, T.B., Omar, S.A., Powell, K., Pretorius, G.S., Smith, M.W., Thera, M.A., Wambebe, C., Weber, J.L. & Williams, S.M. (2009). The genetic structure and history of Africans and African Americans. *Science,* 324(5930); 1035-1044.

Torres, J.B., Stone, A.C. & Kittles, R. (2013). An anthropological genetic perspective on Creolization in the Anglophone Caribbean. *Am J Phys Anthropol,* 151(1); 135-143.

Torroni, A., Schurr, T.G., Cabell, M.F., Brown, M.D., Neel, J.V., Larsen, M., Smith, D.G., Vullo, C.M. & Wallace, D.C. (1993). Asian affinities and continental radiation of the four founding Native American mtDNAs. *Am J Hum Genet,* 53(3); 563-590.

Torroni, A., Lott, M.T., Cabell, M.F., Chen, Y.S., Lavergne, L. & Wallace, D.C. (1994). mtDNA and the origin of Caucasians: identification of ancient Caucasian-specific haplogroups, one of which is prone to a recurrent somatic duplication in the D-loop region. *Am J Hum Genet,* 55(4); 760-776.

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

Toyin, F. & Childs, M.D. (2004). *The Yoruba Diaspora in the Atlantic World.* Bloomington, USA: Indiana University Press.

Trejaut, J.A., Poloni, E.S., Yen, J.C., Lai, Y.H., Loo, J.H., Lee, C.L., He, C.L. & Lin, M. (2014). Taiwan Y-chromosomal DNA variation and its relationship with Island Southeast Asia. *BMC Genet,* 15; 77.

Trombetta, B., D'Atanasio, E., Massaia, A., Myres, N.M., Scozzari, R., Cruciani, F. & Novelletto, A. (2015). Regional Differences in the Accumulation of SNPs on the Male-Specific Portion of the Human Y Chromosome Replicate Autosomal Patterns: Implications for Genetic Dating. *PLoS One,* 10(7); e0134646.

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

Underhill, P.A., Myres, N.M., Rootsi, S., Metspalu, M., Zhivotovsky, L.A., King, R.J., Lin, A.A., Chow, C.E., Semino, O., Battaglia, V., Kutuev, I., Jarve, M., Chaubey, G., Ayub, Q., Mohyuddin, A., Mehdi, S.Q., Sengupta, S., Rogaev, E.I., Khusnutdinova, E.K., Pshenichnov, A., Balanovsky,

O., Balanovska, E., Jeran, N., Augustin, D.H., Baldovic, M., Herrera, R.J., Thangaraj, K., Singh, V., Singh, L., Majumder, P., et al. (2010). Separating the post-Glacial coancestry of European and Asian Y chromosomes within haplogroup R1a. *Eur J Hum Genet,* 18(4); 479-484.

Underhill, P.A., Poznik, G.D., Rootsi, S., Jarve, M., Lin, A.A., Wang, J., Passarelli, B., Kanbar, J., Myres, N.M., King, R.J., Di Cristofaro, J., Sahakyan, H., Behar, D.M., Kushniarevich, A., Sarac, J., Saric, T., Rudan, P., Pathak, A.K., Chaubey, G., Grugni, V., Semino, O., Yepiskoposyan, L., Bahmanimehr, A., Farjadian, S., Balanovsky, O., Khusnutdinova, E.K., Herrera, R.J., Chiaroni, J., Bustamante, C.D., Quake, S.R., et al. (2015). The phylogenetic and geographic structure of Y-chromosome haplogroup R1a. *Eur J Hum Genet,* 23(1); 124-131.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M. & Rozen, S.G. (2012). Primer3--new capabilities and interfaces. *Nucleic Acids Res,* 40(15); e115.

van Dijk, E.L., Auger, H., Jaszczyszyn, Y. & Thermes, C. (2014). Ten years of next-generation sequencing technology. *Trends Genet,* 30(9); 418-426.

van Dorp, L., Balding, D., Myers, S., Pagani, L., Tyler-Smith, C., Bekele, E., Tarekegn, A., Thomas, M.G., Bradman, N. & Hellenthal, G. (2015). Evidence for a Common Origin of Blacksmiths and Cultivators in the Ethiopian Ari within the Last 4500 Years: Lessons for Clustering-Based Inference. *PLoS Genet,* 11(8); e1005397.

van Oven, M. & Kayser, M. (2009). Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat,* 30(2); E386-394.

van Oven, M. (2010). Revision of the mtDNA tree and corresponding haplogroup nomenclature. *Proc Natl Acad Sci U S A,* 107(11); E38-39; author reply e40-31.

van Oven, M., Ralf, A. & Kayser, M. (2011). An efficient multiplex genotyping approach for detecting the major worldwide human Y-chromosome haplogroups. *Int J Legal Med,* 125(6); 879-885.

van Oven, M., Van Geystelen, A., Kayser, M., Decorte, R. & Larmuseau, M.H. (2014). Seeing the wood for the trees: a minimal reference phylogeny for the human Y chromosome. *Hum Mutat,* 35(2); 187-191.

Veeramah, K.R., Connell, B.A., Ansari Pour, N., Powell, A., Plaster, C.A., Zeitlyn, D., Mendell, N.R., Weale, M.E., Bradman, N. & Thomas, M.G. (2010). Little genetic differentiation as assessed by uniparental markers in the presence of substantial language variation in peoples of the Cross River region of Nigeria. *BMC Evol Biol,* 10; 92.

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., Ziv, E., Risch, N., Burchard, E.G. & Martinez-Cruzado, J.C. (2011). History shaped the geographic distribution of genomic admixture on the island of Puerto Rico. *PLoS One,* 6(1); e16513.

Voyages Database (2009). Voyages: The Trans-Atlantic Slave Trade Database. Copyright 2009 Emory University: Website: http://slavevoyages.org/tast/database/search.faces.

Walvin, J. (2011). *The slave trade.* London, UK: Thames & Hudson Ltd. History Files.

Walvin, J. (2013). *Crossings: Africa, the Americas and the Atlantic Slave Trade.* London, UK: Reaktion Books.

Wang, C., Zollner, S. & Rosenberg, N.A. (2012). A quantitative comparison of the similarity between genes and geography in worldwide human populations. *PLoS Genet,* 8(8); e1002886.

Wang, C.C., Wang, L.X., Shrestha, R., Wen, S., Zhang, M., Tong, X., Jin, L. & Li, H. (2015). Convergence of Y Chromosome STR Haplotypes from Different SNP Haplogroups Compromises Accuracy of Haplogroup Prediction. *J Genet Genomics,* 42(7); 403-407.

Wang, J., Lin, M., Crenshaw, A., Hutchinson, A., Hicks, B., Yeager, M., Berndt, S., Huang, W.Y., Hayes, R.B., Chanock, S.J., Jones, R.C. & Ramakrishnan, R. (2009). High-throughput single nucleotide polymorphism genotyping using nanofluidic Dynamic Arrays. *BMC Genomics,* 10; 561.

Wang, S., Ray, N., Rojas, W., Parra, M.V., Bedoya, G., Gallo, C., Poletti, G., Mazzotti, G., Hill, K., Hurtado, A.M., Camrena, B., Nicolini, H., Klitz, W., Barrantes, R., Molina, J.A., Freimer, N.B., Bortolini, M.C., Salzano, F.M., Petzl-Erler, M.L., Tsuneto, L.T., Dipierri, J.E., Alfaro, E.L., Bailliet, G., Bianchi, N.O., Llop, E., Rothhammer, F., Excoffier, L. & Ruiz-Linares, A. (2008). Geographic patterns of genome admixture in Latin American Mestizos. *PLoS Genet,* 4(3); e1000037.

Watson, E., Forster, P., Richards, M. & Bandelt, H.J. (1997). Mitochondrial footprints of human expansions in Africa. *Am J Hum Genet,* 61(3); 691-704.

Wegmann, D., Kessner, D.E., Veeramah, K.R., Mathias, R.A., Nicolae, D.L., Yanek, L.R., Sun, Y.V., Torgerson, D.G., Rafaels, N., Mosley, T., Becker, L.C., Ruczinski, I., Beaty, T.H., Kardia, S.L., Meyers, D.A., Barnes, K.C., Becker, D.M., Freimer, N.B. & Novembre, J. (2011). Recombination rates in admixed individuals identified by ancestry-based inference. *Nat Genet,* 43(9); 847-853.

Wei, W., Ayub, Q., Chen, Y., McCarthy, S., Hou, Y., Carbone, I., Xue, Y. & Tyler-Smith, C. (2013). A calibrated human Y-chromosomal phylogeny based on resequencing. *Genome Res,* 23(2); 388-395.

Wickham, H. (2009). *ggplot2: elegant graphics for data analysis.* New York: Springer.

Wigginton, J.E., Cutler, D.J. & Abecasis, G.R. (2005). A note on exact tests of Hardy-Weinberg equilibrium. *Am J Hum Genet,* 76(5); 887-893.

Wilkins, J.F. & Marlowe, F.W. (2006). Sex-biased migration in humans: what should we expect from genetic data? *Bioessays,* 28(3); 290-300.

Wilson Sayres, M.A., Lohmueller, K.E. & Nielsen, R. (2014). Natural selection reduced diversity on human y chromosomes. *PLoS Genet,* 10(1); e1004064.

Wollstein, A. & Lao, O. (2015). Detecting individual ancestry in the human genome. *Investig Genet,* 6; 7.

Woods, C.G., Cox, J., Springell, K., Hampshire, D.J., Mohamed, M.D., McKibbin, M., Stern, R., Raymond, F.L., Sandford, R., Malik Sharif, S., Karbani, G., Ahmed, M., Bond, J., Clayton, D. & Inglehearn, C.F. (2006). Quantification of homozygosity in consanguineous individuals with autosomal recessive disease. *Am J Hum Genet,* 78(5); 889-896.

World Medical Association (2013). World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA,* 310(20); 2191-2194.

Wright, S. (1933). Inbreeding and Homozygosis. *Proc Natl Acad Sci U S A,* 19(4); 411-420.

Wright, S. (1965). The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution,* 19(3); 395-420.

Xavier, C., Builes, J.J., Gomes, V., Ospino, J.M., Aquino, J., Parson, W., Amorim, A., Gusmao, L. & Goios, A. (2015). Admixture and genetic diversity distribution patterns of non-recombining lineages of Native American ancestry in Colombian populations. *PLoS One,* 10(3); e0120155.

Xu, H., Wang, C.C., Shrestha, R., Wang, L.X., Zhang, M., He, Y., Kidd, J.R., Kidd, K.K., Jin, L. & Li, H. (2015). Inferring population structure and demographic history using Y-STR data from worldwide populations. *Mol Genet Genomics,* 290(1); 141-150.

Xue, Y., Wang, Q., Long, Q., Ng, B.L., Swerdlow, H., Burton, J., Skuce, C., Taylor, R., Abdellah, Z., Zhao, Y., Asan, MacArthur, D.G., Quail, M.A., Carter, N.P., Yang, H. & Tyler-Smith, C. (2009). Human Y chromosome base-substitution mutation rate measured by direct sequencing in a deep-rooting pedigree. *Curr Biol,* 19(17); 1453-1457.

Y Chromosome Consortium (2002). A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res,* 12(2); 339-348.

Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol,* 24(8); 1586-1591.

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

Zakharia, F., Basu, A., Absher, D., Assimes, T.L., Go, A.S., Hlatky, M.A., Iribarren, C., Knowles, J.W., Li, J., Narasimhan, B., Sidney, S., Southwick, A., Myers, R.M., Quertermous, T., Risch, N. & Tang, H. (2009). Characterizing the admixed African ancestry of African Americans. *Genome Biol,* 10(12); R141.

Zhidkov, I., Nagar, T., Mishmar, D. & Rubin, E. (2011). MitoBamAnnotator: A web-based tool for detecting and annotating heteroplasmy in human mitochondrial DNA sequences. *Mitochondrion,* 11(6); 924-928.

Zhou, H., Alexander, D. & Lange, K. (2011). A quasi-Newton acceleration for high-dimensional optimization algorithms. *Statistics and Computing,* 21(2); 261-273.

Zimmerman, P.A., Phadke, P.M., Lee, A., Elson, L.H., Aruajo, E., Guderian, R. & Nutman, T.B. (1995). Migration of a novel DQA1* allele (DQA1*0502) from African origin to North and South America. *Hum Immunol,* 42(3); 233-240.

VII. ANNEXE

A. *Supplementary material of Y chromosome*

Table S1: Characteristics of 96 Y-SNP markers genotyped, and allelic frequencies obtained for each primer with different allele specific for all over the sequences.

Table S2: Frequencies of Y-STRs markers obtained in African American (Noir Marron, Afro-Colombian and Afro-Brazilian), West African, West European, and Native American populations.

Figure S1: DAPC of 17 Y-STRs of African American; Noir Marron (NM), Afro-Colombian (Af-COL) and Afro-Brazilian (Af-BRA) populations and continental populations with West African (CIVA, CIVY, MLI_B, BENB, BENY, and BENF), European (ESPBa, ESPGa, ESPGr, ESPHu, and PRT), and Native American (GUF_K, GUF_O, and GUF_P) ancestries.

Figure S2: STRUCTURE of 17 Y-STRs haplotypes showed the percentage of African, European and Native American ancestry in African Americans.

Table S3: Ancestry proportions of continental ancestry estimated for each population using STRUCTURE analysis based on 17 Y-STR.

Table S4: Pairwise F_{ST} genetic distances based on Y haplogroup frequencies.

Table S5: Table of Y-SNP frequencies of 18 populations included in 1000 Genome Project estimated by using the same Y-SNP set.

Table S6: Table of populations analysed based on major Y haplogroup frequencies in African Americans and historic African regions.

****

Table S1: Characteristics of 96 Y-SNP markers genotyped, and allelic frequencies obtained for each primer with different allele specific for all over the sequences.

	SNP	Haplogroup	Haplogroup and Y-SNP	RefSNP ID	Position relative to GRCh37	Mutation	Allele 1	Allele 2	Freq A1	Freq A2	Genomic sequence
1	P114	A0a1a	A0a1a-P114	-	15020655	G→A	A	G			GCTGGAGTGCAGTGGC [G/A] CGATCTCGGCTCACC
2	M31	A1a	A1a-M31	rs369315948	21739754	G→C	G	C	100.0%	0.0%	TTTTACACTGGGCAGT [G/C] TGGTAGCACAGAACAG
3	M14	A1b1a1	A1b1a1-M14	rs3905	21222268	T→C	C	T	0.0%	100.0%	TTATAAATGAAAAAATCAGTAACATG [T/C] TATAAGCAGTTAAAAAAGTTAATG
4	M190	A1b1b2	A1b1b2-M190*(xM51,M13)	rs2032603	14968527	A→G	A	G	99.8%	0.2%	TTTTAAGCTGATACTGAAAATCATTCT [A/G] AATTCTAAATAGTTTATTTTTTC
5	M51	A1b1b2a	A1b1b2a-M51	rs34078768	21868863	C→T	C	T	100.0%	0.0%	CCTAAACTCTGATCCCTGTTGGAAGC [C/T] TGTGTCTACTCTGCTTCAGGAGATA
6	M13	A1b1b2b	A1b1b2b-M13	rs3904	21722098	G→C	C	G	0.2%	99.8%	TAGGTAGGTTAAGGCAAGACGGTTA [G/C] ATCAGTTCTCGTTACTGTTGTATAAT
7	M91	BT	BT-M91*	rs2032651	21907538	del→A	-	A	100.0%	0.0%	CCCTACATTGCTATTCTGTTTTTTTT [-/A] AATCAGTTCACTACTGTAGTATCTT
8	P97	BT	BT-P97*	-	14886273	G→T	G	T	100.0%	0.0%	GGGCGTGGTGGTAGGT [G/T] CCTTTAGTCCCAGCT
9	SRY10831.1	BT	BT-SRY10831.1	rs2534636	2657176	T→C	C	T	98.7%	1.3%	CACATAGGTGAACCTTGAAAATGTTA [C/T] ACTGTGTGAAAAAGTCAGATACAAG
10	M60	B	B-M60*	rs2032623	21878072..21878073	del→T	T	-	1.4%	98.6%	TACATTTCAAAATGCATGACTTAAAG [-/T] ATCAGGCACACAGTGGTTACTCAAT
11	M150	B2a	B2a-M150	rs371646183	21869519	C→T	C	T	98.9%	1.1%	ACACACAGATAGAAGT [C/T] GGAGAAGAGGGCATG
12	M112	B2b	B2b-M112*	rs111725135	21767959	C→T	C	T	100.0%	0.0%	GCAGTACTGATAAAAA [C/T] ACTGCTTTGTTTTA
13	M169	B2b1a1a	B2b1a1a-M169	rs2032594	14813717	T→C	C	T	0.0%	100.0%	GATACTTCCACTGGAGGGAAACAGTT [T/C] AAAGGATATATGTGATACTATTAAT
14	M129	B2b1a1b	B2b1a1b-M129	-	21717717	C→T	C	T	100.0%	0.0%	GAAGATAAAAATTAGC [C/T] GGGCTTGGTGGTGTG
15	M211	B2b1a1c	B2b1a1c-M211	rs2032663	15544614	C→T	C	T	100.0%	0.0%	GTAGGCAAAAGGATGTTAACAACAAG [C/T] ACACCACTGAGACCAGTGAAACCTG
16	P70	B2b1a2	B2b1a2-P70	-	21612268	C→T	C	T	100.0%	0.0%	AATATCAAGACTGGAC [C/T] TCATCAACAGATGGT
17	P6	B2b1b	B2b1b-P6	rs112190431	6768265	G→C	G	C	100.0%	0.0%	TTCCACAGTTAAGTCT [G/C] TTTACCCTACCTCCA
18	M145	DE	DE-M145*(xM96)	rs3848982	21717208	A→G	G	A	67.9%	32.1%	CTTTCCTAGACACCAGAAAGAAAGGC [A/G] AGAGCCAGCCTTAGCCTAATCAAGA
19	M203	DE	DE-M203*(xM96)	rs2032653	15591537	G→C	C	G	69.7%	30.3%	GCGCGCGTAGTGAGCACTGCGCA [G/C] GATTTCATGGAAACAACAAATTTCC
20	M96	E	E-M96*	rs9306841	21778998	C→G	C	G	32.1%	67.9%	TTAATATTATACCTGAGTGTTTTATC [C/G] TATTATGAAGGACCTGTTTTCCAAG
21	P29	E	E-P29*	rs60115999	14497207	A→C	A	C	32.1%	67.9%	TCTCAATCCTCATCTAAAAATACCTTT [A/C] CAGCAAACATCTAGTTGTTGACCAAA
22	M33	E1a	E1a-M33	rs368762706	21740450	A→C	A	C	96.9%	3.1%	CATAAGTTACTAGTTA [A/C] ATCTCAGACATATTA
23	P2	E1b1	E1b1-P2*(xV38,M215)	rs9785756	21610831	G→A	A	G	63.5%	36.5%	TAAGAGAGGTGCCCCTAGGAGGAGAA [G/A] GCATTAGTTTTATTAATGCTCCTTT
24	V38	E1b1a	E1b1a-V38*(xM2)	rs768983	6818291	C→T	T	C	60.3%	39.7%	GAACGTTGTTTGGAATTGGTGGAGC [C/T] TACTCACAGTTCGTTTATAGTATGG
25	M2	E1b1a1	E1b1a1-M2*	rs9785941	14096577	A→G	A	G	39.2%	60.8%	TTATATTTCATTGTTAACAAAAGTCC [A/G] TGAGATCTGTGGAGGATAAAAGGGGG
26	M191	E1b1a1a1c1a	E1b1a1a1c1a-M191*(xU174)	rs2032590	15019613	T→G	G	T	18.8%	81.2%	ACATTTTTTTCTTTAACACTTGACTA [T/G] ATGAAAAATATGAGATATTTTGGTGT
27	U174	E1b1a1a1c1a1	E1b1a1a1c1a1-U174*(xP115,P116)	rs16980586	16251357	G→A	A	G	17.2%	82.8%	TTGTGTTGCATACCAGATTAACCCAT [G/A] CTGTGAAGAACAGAGACAGTGAAGC
28	P115	E1b1a1a1c1a1b	E1b1a1a1c1a1b-P115	-	14869250	G→A	G	A	100.0%	0.0%	TTTGGAACCTGGCTCA [G/A] AGTGATGATGTGCCT
29	P116	E1b1a1a1c1a1c	E1b1a1a1c1a1c-P116	-	14869690	A→G	A	G	100.0%	0.0%	ACAAAGTTGCTGGGATT [A/G] CAGGCGTCAGCCACC
30	U175	E1b1a1d	E1b1a1d-U175*(xU209)	rs16980588	16253694	G→A	A	G	30.9%	69.1%	ATGGGGCCACAGGTGCTAATGAAACC [G/A] TCTTGACTTCTCCTGGTCATTAGAT
31	U209	E1b1a1d1	E1b1a1d1-U209*(xU290,M154)	rs16980502	17294958	C→T	C	T	69.1%	30.9%	GGAATGCAAAAGATGTAATGAATTAA [C/T] TGATTTTAACTTGCAGTCTTCCATT
32	U290	E1b1a1d1a	E1b1a1d1a-U290*(xU181)	rs16980406	21646058	T→A	A	T	16.0%	84.0%	AGGATAGGTGTGGAATTGGTGGCGT [T/A] CCGGGACTGCCATCCTCCGT
33	U181	E1b1a1d1a1	E1b1a1d1a1-U181	rs16980589	16374426	C→T	C	T	99.8%	0.2%	ACTGAGTGTCTTTGTTTGGCAAGGA [C/T] AGGAACTGTATTCTGATACAAGGGT
34	M154	E1b1a1d1c	E1b1a1d1c-M154	-	21892657	T→C	T	C	99.8%	0.2%	AATTAGTGCGACACTA [T/C] TGTACTGAATATTAT
35	M215	E1b1b	E1b1b-M215*(xM35.1)	rs2032654	15467824	A→G	A	G	96.9%	3.1%	CATACTTGCTGCATTAAGACAAACTG [A/G] CTTTCTAACTGTTCCAGCTGATGCT
36	M35.1	E1b1b1	E1b1b1-M35.1*	rs375228668	21741703	G→C	G	C	96.8%	3.2%	TTCCTTTGGGACACTG [G/C] GACACAGGCAGAGAC
37	M78	E1b1b1a1	E1b1b1a1-M78	rs368977028	21893303	C→T	C	T	99.3%	0.7%	CAAAGATACTTCTTTC [C/T] GCCCTTCCAAATATT
38	M81	E1b1b1b1a	E1b1b1b1a-M81	rs2032640	21892572	C→T	C	T	99.1%	0.9%	AGAGGTAAATTTTGTCCTTTTTGAA [C/T] GTCATAGAGTATACTCACACAAACC
39	M123	E1b1b1b2a	E1b1b1b2a-M123*(xM34)	rs371143248	21764586	C→T	C	T	99.3%	0.7%	AGGTATTCAGGCGATG [C/T] TGATATGCTAGTTCA
40	M34	E1b1b1b2a1	E1b1b1b2a1-M34	rs373666971	21740716	G→T	G	T	99.6%	0.4%	GGCTTCCACCCAGGAG [G/T] CACATAGTGGTGTGTC
41	M75	E2	E2-M75*(xM41,M54)	rs2032639	21890177	G→A	A	G	1.1%	98.9%	GAAAAGACAAATTATCAAACCACATCC [G/A] TATATATACAGAGAAATACCTTTAT
42	M41	E2a	E2a-M41	-	2663889	C→A	A	C	100.0%	0.0%	ATTTTTAGTAGAGACC [C/A] AGTTTCACCATGTTG
43	M54	E2b1	E2b1-M54	rs2032620	21869264	G→A	A	G	0.4%	99.6%	AGGGCCTCACCCTCACGGCCAGCC [G/A] GGCAGCCTCTGCAGGGCAACCAGCA
44	M130	C	C-M130*(xM38,M217)	rs35284970	2734854	C→T	C	T	100.0%	0.0%	TTGCAGGGCAATAAACCTTGGATTTC [C/T] CTGCCCAGGAAAGGGCATTCGAC
45	M216	C	C-M216*(xM38,M217)	rs2032666	15437564	C→T	C	T	99.5%	0.5%	AAAATTCCTTTATTAAAGAAAATGTAA [C/T] ATTCAACAGGTATACATAACTAGCA
46	M38	C1b2a	C1b2a-M38	rs369611932	21742158	T→G	T	G	100.0%	0.0%	TTATGGAAAACCAACT [T/G] TTTTATTCAGTAAGC
47	M217	C2	C2-M217	rs2032668	15437333	A→C	A	C	100.0%	0.0%	AGGAGAATGAAAAAGTTGGGTGACAC [A/C] AACTCTTCAGAGGAAAAATACATA
48	M213	F	F-M213	rs2032665	15526751	T→C	C	T	28.6%	71.4%	CTATTACATATTATTAATAAGAAGTCA [T/C] GTAACGAGATGTTTTAAGTTCTGAA
49	M89	F	F-M89	rs2032652	21917313	C→T	C	T	70.7%	29.3%	CTTCCTAAGGTTATGTACAAAAATCT [C/T] ATCTCTCACTTTGCCTGAGTTGCAG

	SNP	Haplogroup	Haplogroup and Y-SNP	RefSNP ID	Position relative to GRCh37	Mutation	Allele 1	Allele 2	Freq A1	Freq A2	Genomic sequence
50	M201	G	G-M201	rs2032636	15027529	G→T	G	T	98.9%	1.1%	ATCTAATAATCCAGTATCAACTGAGG [G/T] TTTTCGTAATAGGTACTTAGTGTTG
51	P287	G2	G2-P287*(xP15)	rs4116820	22072097	G→T	G	T	99.1%	0.9%	GGAACCTCCTGACTGACATGGCCTTG [G/T] GACACTGAGCATTCAGTGCCAGTGG
52	P15	G2a	G2a-P15	rs370167410	23244026	C→T	C	T	99.1%	0.9%	GAGGTTCTGAATCTTA [C/T] GCCTGAAGGCAGATG
53	M69	H1	H-M69*(xM52)	rs2032673	21894058	T→C	C	T	0.0%	100.0%	TCACGGAGGCTGTTTACACTCCTGAAA [T/C] AAAATATATTCAGCAAGACAAAAGG
54	M52	H1a	H1a-M52*(xM82)	rs376769460	21753199	A→C	A	C	100.0%	0.0%	AAGAATATTGCCTGCA [A/C] GGATGTTTGATAGGT
55	M82	H1a1	H1a1-M82	rs2032675	21894174	ins→del-2bp	-	A	0.0%	100.0%	CTCATGCAATAAATTCTGCATTTTCAT [AT/-] GTTTCCAGGTAGGTTGTTTCTTCTCA
56	P203.2	I1	I1-P203.2	rs13447354	22750951	G→A	A	G	1.5%	98.5%	GGTTGGCTATTGAGTTAGCATAATCA [G/A] TCTGCCTACCATACTTAAAGTACCT
57	M304	J	J-M304*(xM267,M172)	rs13447352	22749853	A→C	A	C	96.9%	3.1%	GTGTTCAATTTGAAAGTAACTTGTGA [A/C] ACAACTGCGTGATATTTTGGTATAAG
58	P209	J	J-P209*(xM267,M172)	rs17315835	19179335	T→C	C	T	2.9%	97.1%	TTGATGTACTTCATCTCTGAGTCATT [T/C] TAATTGGATTCTGGTTTCTATCCGT
59	M267	J1	J1-M267*	rs9341313	22741818	T→G	G	T	1.3%	98.7%	ATTCGATGGAAGCATTTTTGTAAATA [T/G] ACGTTCAGTATTTTGTGGAAGAA
60	P58	J1a2b	J1a2b-P58	rs34043621	14486667	T→C	C	T	0.0%	100.0%	CCAGCAATGACATTTGTGTGCTTTGC [T/C] CAGAATGGCAAATATTGACCAGCTT
61	M172	J2	J2-M172*(xM410,M12)	rs2032604	14969634	T→G	G	T	1.6%	98.4%	TATCCCCAAACCCATTTTGATGCTT [T/G] ACTTAAAAGGTCTTCAATTATTATT
62	M410	J2a	J2a-M410*	rs371079691	2751678	A→G	A	G	99.8%	0.2%	ACATCTTGAGTTTTCTG [A/G] AGAAACATGTGCCAT
63	M67	J2a1b	J2a1b-M67	rs2032628	21878809	A→T	A	T	98.9%	1.1%	TCAAAATATGTGTAATTCAAAAAACA [A/T] ATATAGAGGGGTCCACGAACAAGTG
64	M12	J2b	J2b-M12*(xM241)	rs3903	7583480	G→T	G	T	99.8%	0.2%	CTAGCTAATTTTTTATTATTTGTAGA [G/T] ATGGGGTCACTATGTTGCTCAGGC
65	M241	J2b2	J2b2-M241*	rs8179022	15018459	G→A	G	G	0.0%	100.0%	CTAGAATTAAGTAGTAAATTCAGATG [G/A] CAAGATTTTTAAGTACAGTAGTATC
66	M9	K	KLT-M9*	rs3900	21730257	C→G	C	G	80.1%	19.9%	AAAGAAACGGCCTAAGATGGTTGAAT [C/G] CTCTTTATTTTTCTTTAATTTAGAC
67	P128	K	KLT-P128/PF5504*	rs17250121	20837553	C→T	C	T	80.1%	19.9%	CTGAATTATAGTTAAGTGTTTATAGA [C/T] TAGGTCTTGTTTCTCCCTCTTTGAAT
68	M526	K(xLT)	K -M526*	rs2033003	23550924	A→C	A	C	80.7%	19.3%	TGCTCTGTCATCAGGCTGAATCATAC [A/C] TCACTGTAGTGCCAGACTTTTGGGC
69	P256	M	M-P256*(xM186)	-	8685230	G→A	G	A	100.0%	0.0%	TACACTAGATAGAAAG [G/A] GGTTGAGTAGAAGGC
70	M186	M1	M1-M186	rs2032681	14923461	G→del	-	G	100.0%		ATAGGAAAACCACTTGAACACAGTTTGGG [-/G] AAGTTGTATAGAAGATCTCATTTCC
71	M45	P	P -M45*(xM207)	rs2032631	21867787	G→A	A	G	20.0%	80.0%	AAATTGGCAGTGAAAAATTATAGATA [G/A] GCAAAAGCTCCTTCTGAGGTCCAG
72	M207	R	R-M207*(xM173)	rs2032658	15581983	A→G	A	G	82.8%	17.2%	CAAATGTAAGTCAAGCAAGAAAATTTA [A/G] AAAGAGAATAACAATACCTTTTGAA
73	M173	R1	R1-M173*(xM434)	rs2032624	15026424	A→C	A	C	82.8%	17.2%	TCTTACAATTCAAGGGCATTTAGAAC [A/C] CTTTGTCATCTGTTAATTCAGAA
74	M434	R1a	R1a-M434*	-	14937992	G→A	G	A	100.0%	0.0%	ATCAGCCTCCCAGGTA [G/A] CTGGGATTTATAGGCG
	SRY10831.2	R1a1	R1a1-SRY10831.2	rs2534636	2657176	C→T	C	T	98.7%	1.3%	CACATAGGTGAACCTTGAAAATGTTA [C/T] ACTGTGTGAAAAAGTCAGATACAAG
75	M17	R1a1a	R1a1a-M17*	rs3908	21733168	4G→3G	-	G	0.4%	99.6%	AGTTTGTGGTTGCTGGTTGTTACGGG [G/-] TTTTTTAAGTGAATTTTGGGGTTT
76	M458	R1a1a1b1a1	R1a1a1b1a1-M458	rs375323198	24366464	A→G	A	G	100.0%	0.0%	ATAAATTCCATTTGAT [A/G] AGAGAGCCAGGAGGA
77	M415	R1b1	R1b1-M415*	rs9786194	9170545	C→A	A	C	17.0%	83.0%	CCCATGTGTTTACTGGCACTCTTTAA [C/A] CCATAATTTGATTGAACGGGAAACT
78	M269	R1b1a2	R1b1a2-M269	rs9786153	22739367	T→C	C	T	16.1%	83.9%	AGGGGAATGATCAGGTTTGGTTAAT [T/C] CTGGTAAATTGAAAACAATTTTTT
79	M124	R2a	R2a-M124	rs372706460	21764501	C→T	C	T	100.0%	0.0%	TCAGTATTATTAAAAC [C/T] GACTTCCCTGTTCCCC
80	P186	O	O-P186*	rs16981290	7568568	C→A	A	C			GGGCTCCTGAAATTTTCTGGGGAAC [C/A] AAAATCAGGGAAAATTCTATTATTT
81	M119	O1a	O1a-M119*(xP203,M50)	rs72613040	21762685	T→G	T	G	100.0%	0.0%	AATTCAGCATACAGGC [T/G] AAAATAGCAATTTTA
	P203.1	O1a1	O1a1-P203.1	rs13447354	22750951	G→A	A	G	1.5%	98.5%	GGTTGGCTATTGAGTTAGCATAATCA [G/A] TCTGCCTACCATACTTAAAGTACCT
82	M50	O1a2	O1a2-M50	rs2032632	21868672	T→C	C	T	0.0%	100.0%	AGCCAACAGCCTACCCAAACCACACC [C/T] GTCTTACCGAGCCCTTTCCAACCTG
83	M95	O2a1	O2a1-M95*(xM88)	rs2032650	21938444	C→T	C	T	100.0%	0.0%	GATAAGGAAAGACTACCATATTAGTG [C/T] TGGATGGCTAGCCTTTCCAACCTG
84	M88	O2a1a	O2a1a-M88	rs78149062	21764674	A→G	A	G	99.8%	0.2%	GCTTTTCTTATTCCTGCTTCTTCTGC [A/G] TATCCAAGATAGGTTTTCCTCTCTA
85	M122	O3	O3-M122*	rs20320	14898163	G→A	A	G	99.5%	0.5%	ATTTTCCCCTGAGAGC [A/G] TGAATTAGTATCTCA
86	P201	O3a2	O3a2-P201*	rs2267801	2828196	T→C	C	T	0.0%	100.0%	ATCTTGGTTAAGTCATTTGATCCTCAG [T/C] GGGCTTTAACTGGCTCTCACCCTTC
87	M7	O3a2b	O3a2b-M7	rs3898	4078217	C→G	C	G	100.0%	0.0%	TCCTCTCTGTATGTCAGATCTAACAA [C/G] AAGAAGAACAGTAACTCAACTACAA
88	M134	O3a2c1	O3a2c1-M134	rs200634940	21716218	C→del	C	-	100.0%		TTTGATCCCCACCAAT [C/-] ATACTTTCGTGGCC
89	M20	L	L-M20/PF5570*	rs3911	21733454	A→G	A	G	99.8%	0.2%	GGTTCAACCAACGTGGATTGAAAAT [A/G] TTTGAAAAATTAAAAATAGATGGTT
90	M295	L1	L1-M295*(xM76,M357)	rs9341318	22745051	T→C	C	T	0.2%	99.8%	GTTAGTAATACCAACCACCAGGTAGTT [T/C] ATAAATTACACTGATTAAAAGTCTC
91	M76	L1a	L1a-M76	rs377001539	21757316	T→G	T	G	100.0%	0.0%	AGCACCTTACACAGTT [T/G] AATTATTGCATAGAT
92	M357	L1c	L1c-M357	rs377156966	2888252	C→A	C	A	99.8%	0.2%	TTTTTTTTACCACAGG [C/A] CATCATTTTGTAGAA
93	M184	T	T-M184*	rs20320	14898163	G→A	A	G	0.4%	99.6%	ATAGAACAGCTGTAGAAAAATTACGA [G/A] CTGTTTGTTTGGGACCATGCAAAACT
94	M272	T	T-M272/PF5667*	rs9341308	22738775	A→G	A	G	99.3%	0.7%	TGGGCATTACTCTTTGCTCTCCCGAA [A/G] ACCCACTAACGGGTTGTGTGTATAA
95	M70	T1a	T1a-M70*(xL131)	rs2032672	21893881	A→C	A	C	99.3%	0.7%	TACTTTGGACTCATGTCTCCATGAGA [A/C] CTAAGACTACCACCAACAGAATCCCT
96	L131	T1a2	T1a2-L131	rs2215828	19372808	C→T	C	T	99.6%	0.4%	GGTCTTTGCCCIGGGTATTTCTTCCT [C/T] CTTTCTGTCTTTTTTGCATTCACAT

Table S2: Frequencies of Y-STRs markers obtained in African American (Noir Marron, Afro-Colombian and Afro-Brazilian), West African, West European, and Native American populations.

Aluku (n= 8)

Alleles	DYS19	DYS389I	DYS389II-I	DYS390	DYS391	DYS392	DYS393	DYS385a	DYS385b	DYS438	DYS439	DYS437	DYS448	DYS456	DYS458	DYS635	YGATAH4
9																	
10					1.000					0.250							
11						0.750				0.750	0.250						0.375
12		0.375				0.250					0.625						0.625
13		0.625					1.000				0.125						
14	0.250											0.750			0.250		
15	0.750							0.500	0.125				-	1.000	-		
16								0.375	0.125				-		-		
17			0.500					0.125	0.625				0.250		0.375		
18			0.500						0.125						0.375		
19																0.250	
20													0.375			0.375	
21				0.750									0.625			0.375	
22				0.250													
23																	
24																	
25																	
26																	

Ndjuka (n= 21)

Alleles	DYS19	DYS389I	DYS389II-I	DYS390	DYS391	DYS392	DYS393	DYS385a	DYS385b	DYS438	DYS439	DYS437	DYS448	DYS456	DYS458	DYS635	YGATAH4
9																	0.048
10		0.048			0.810					0.190	0.048						0.143
11		-			0.190	0.905				0.667	0.095						0.619
12		0.095				-	0.048	0.048		0.143	0.524			0.048			0.190
13		0.619				0.048	0.429	0.095			0.238			0.048	0.095		
14	0.238	0.238				-	0.381	0.048	0.048		0.095	0.857	-		0.143		
15	0.381					0.048	0.143	0.238	0.048			0.048		0.524	0.048		
16	0.048		0.048					0.333	0.190			-		0.381	0.238		
17	0.333		0.571					0.190	0.429			0.095			0.143		
18			0.381					0.048	0.143						0.286		
19									0.095						0.048	0.238	
20				0.095					0.048				0.286			-	
21				0.571									0.667			0.429	
22				0.286									0.048			0.286	
23				-												-	
24				0.048												0.048	
25																	
26																	

Paramaka (n= 5)

Alleles	DYS19	DYS389I	DYS389II-I	DYS390	DYS391	DYS392	DYS393	DYS385a	DYS385b	DYS438	DYS439	DYS437	DYS448	DYS456	DYS458	DYS635	YGATAH4
9																	
10					0.600												
11					0.400	0.600		0.400		0.600							1.000
12						-		-		0.400	0.800						
13		1.000				0.400	0.400	-			0.200						
14	0.400						0.600		0.400			0.600					
15	0.200							0.200	-			0.400		0.200			
16	-		0.400					0.200	0.200					0.200	0.400		
17	0.400		0.400					0.200	0.200					0.600	0.600		
18			0.200					0.200									
19													0.400				
20													0.200				
21				0.600									0.400			0.200	
22				-												0.200	
23				0.400												0.400	
24																0.200	
25																	
26																	

Saramaka (n= 16)

Alleles	DYS19	DYS389I	DYS389II-I	DYS390	DYS391	DYS392	DYS393	DYS385a	DYS385b	DYS438	DYS439	DYS437	DYS448	DYS456	DYS458	DYS635	YGATAH4
9																	
10					0.938					0.125							0.063
11					0.063	1.000				0.875	0.125						0.438
12		0.188									0.750						0.438
13		0.313					0.750				0.125						0.063
14	0.500	0.500					0.063	0.188				0.750		0.125	0.125		
15	0.125						0.188	0.063	0.188			-		0.625	0.250		
16	0.188		0.063					0.625	0.063			-		0.250	0.438		
17	0.188		0.875					0.125	0.188			0.125			0.063		
18			0.063						0.500						0.125		
19									0.063							0.125	
20													0.625			0.063	
21				0.875									0.313			0.500	
22				0.125									0.063			0.188	
23																0.125	
24																	
25																	
26																	

Afro-Colombian (n= 19)

Alleles	DYS19	DYS389I	DYS389II-I	DYS390	DYS391	DYS392	DYS393	DYS385a	DYS385b	DYS438	DYS439	DYS437	DYS448	DYS456	DYS458	DYS635	YGATAH4
9										0.053							
10					0.632					0.105							0.053
11					0.368	0.737	0.053	0.263	0.053	0.579	0.316						0.474
12		0.105				-	0.053	0.053	-	0.263	0.421						0.474
13	0.053	0.474				0.263	0.789	0.105	-		0.263						
14	0.316	0.421					0.053	0.053	0.316			0.632		0.105	0.105		
15	0.474						0.053	0.105	0.105			0.316		0.526	0.158		
16	0.158		0.421					0.211	0.053			0.053		0.316	0.211		
17			0.316					0.211	0.368					0.053	0.368		
18			0.211						0.053				0.105		0.105		
19			0.053						0.053				0.263		-		
20													0.158		0.053		
21				0.421									0.474			0.421	
22				0.158												0.158	
23				0.105												0.316	
24				0.211												0.053	
25				0.105												-	
26																0.053	

Afro-Brazilian (n= 24)

Alleles	DYS19	DYS389I	DYS389II-I	DYS390	DYS391	DYS392	DYS393	DYS385a	DYS385b	DYS438	DYS439	DYS437	DYS448	DYS456	DYS458	DYS635	YGATAH4
9					0.042					0.083							0.042
10					0.542					0.125	0.083						0.042
11					0.417	0.708		0.167		0.542	0.417						0.292
12		0.167				0.042	0.125	0.125		0.250	0.375						0.458
13	0.042	0.708				0.167	0.583	0.208	0.042		0.125			0.042			0.167
14	0.250	0.125				0.083	0.208	0.042	0.250			0.708		0.083	0.042		
15	0.417		0.042				0.083	0.042	0.125			0.208		0.542	0.292		
16	0.167		0.375					0.250	0.125			0.083		0.208	0.125		
17	0.125		0.417					0.167	0.167					0.125	0.250		
18			0.167						0.250				0.042		0.250		
19									0.042				0.292		0.042		
20													0.125			0.042	
21				0.375									0.417			0.417	
22				0.042									0.125			0.042	
23				0.083												0.333	
24				0.417												0.167	
25				0.083													
26																	

West African (n= 286)

Alleles	DYS19	DYS389I	DYS389II-I	DYS390	DYS391	DYS392	DYS393	DYS385a	DYS385b	DYS438	DYS439	DYS437	DYS448	DYS456	DYS458	DYS635	YGATAH4
8																	
9					0.007					0.010							
10					0.888	0.073	0.004			0.056	0.004						0.045
11					0.098	0.787	-	0.024	0.010	0.874	0.206			0.004			0.521
12		0.108				0.105	0.028	0.007	0.007	0.059	0.591			0.010			0.409
13		0.689				0.014	0.493	0.014	0.004		0.164	0.010		0.021	0.007		0.024
14	0.084	0.203	0.004			0.004	0.318	0.077	0.007		0.035	0.930		0.063	0.014		
15	0.451		0.007			0.007	0.147	0.203	0.038			0.024		0.615	0.101		
16	0.255		0.031			0.010	0.010	0.402	0.161			0.014		0.217	0.325		
17	0.206		0.434					0.206	0.346			0.017		0.059	0.332	0.014	
18	0.004		0.388	0.004				0.049	0.217			0.004	0.010	0.007	0.157	-	
19			0.105	-				-	0.143				0.031	0.004	0.049	0.024	
20			0.007	0.031				-	0.035				0.203		0.014	0.070	
21			0.021	0.843				0.010	-				0.706			0.612	
22			0.004	0.077				0.007	0.014				0.049			0.133	
23				0.014				0.010	0.017							0.108	
24				0.028												0.038	
25				0.004													
26																	

West European (n= 131)

Alleles	DYS19	DYS389I	DYS389II-I	DYS390	DYS391	DYS392	DYS393	DYS385a	DYS385b	DYS438	DYS439	DYS437	DYS448	DYS456	DYS458	DYS635	YGATAH4
8										0.008							
9					0.099			0.008		0.099							0.008
10					0.420			0.023		0.328	0.084						0.038
11					0.466	0.389		0.489		0.023	0.305						0.397
12		0.107			0.015	0.023	0.137	0.137	0.038	0.519	0.435						0.527
13	0.115	0.618	0.008			0.557	0.779	0.183	0.076	0.023	0.168	0.008		0.015			0.031
14	0.603	0.275	-			0.023	0.069	0.061	0.542		0.008	0.389		0.137	0.008		
15	0.198		0.076			-	0.015	0.046	0.191			0.534		0.420	0.076		
16	0.076		0.664			0.008		0.038	0.061			0.069		0.366	0.237		
17	0.008		0.214					0.015	0.015					0.061	0.382		
18			0.038						0.038				0.122		0.214		
19									0.023				0.489		0.053	0.008	
20									-				0.260		0.031	0.038	
21				0.008					0.015				0.122			0.237	
22				0.099									0.008			0.084	
23				0.229												0.496	
24				0.611												0.115	
25				0.053												0.015	
26																0.008	

Native American (n= 130)

Alleles	DYS19	DYS389I	DYS389II-I	DYS390	DYS391	DYS392	DYS393	DYS385a	DYS385b	DYS438	DYS439	DYS437	DYS448	DYS456	DYS458	DYS635	YGATAH4
8					0.008												
9					0.008			0.008									
10					0.892		-			0.331							
11					0.092	0.008	0.008	0.008		0.562	0.346						0.523
12	0.131	0.123				-	0.185	0.031		0.085	0.515						0.423
13	0.738	0.400				0.277	0.415	0.338	0.054	0.023	0.115	0.008			0.008		0.054
14	0.123	0.469				0.531	0.392	0.562	0.038		0.023	0.869		0.100	0.077		
15	-	0.008	0.038			0.131		0.031	0.162			0.115		0.400	0.454		
16	-		0.223			0.054		0.015	0.108			0.008		0.277	0.269		
17	0.008		0.515					0.008	0.385				0.008	0.223	0.185		
18			0.215						0.254				-				
19			0.008	0.015									0.554				
20				-									0.331				
21				0.085									0.100			0.008	
22				-									0.008			0.854	
23				0.485												0.131	
24				0.369												0.008	
25				0.046													
26																	

Figure S1: DAPC of 17 Y-STRs of African American; Noir Marron (NM), Afro-Colombians (Af-COL), and Afro-Brazilians (Af-BRA) and continental populations with West African (CIVA, CIVY, MLI_B, BENB, BENY, and BENF), European (ESPBa, ESPGa, ESPGr, ESPHu, and PRT), and Native American (GUF_K, GUF_O, and GUF_P) ancestries. Inertia ellipses are shown by different colours, while dots represent individual strains. Inertia ellipses are proportional to the internal variance of the clusters. In the left bottom corner the DA eigenvalues for discriminant functions, and in the right bottom corner the PCA eigenvalues for principal components are reported, 10 DA and 45 PCA axes were retained.
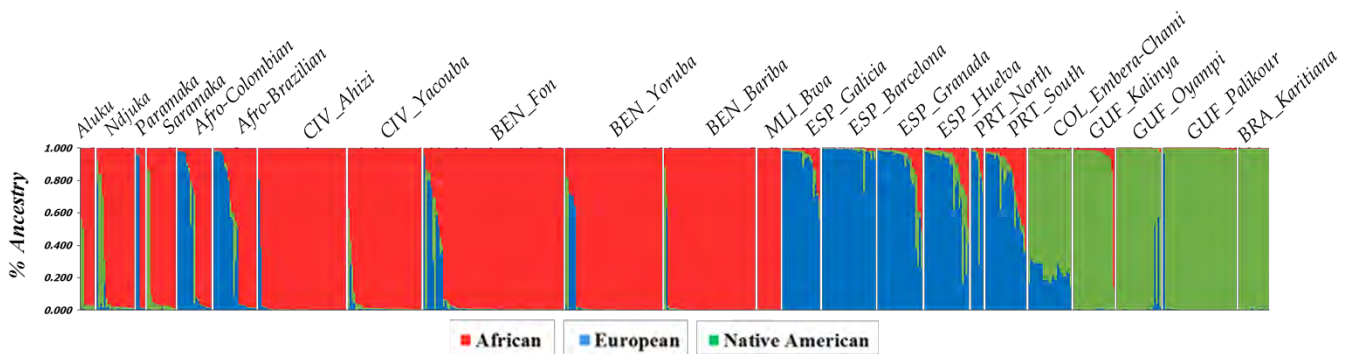
Figure S2: STRUCTURE of 17 Y-STRs haplotypes showed the percentage of African, European and Native American ancestry in African Americans. Mean value of r= 0.934 and mean value of alpha= 0.428.

Table S3: Ancestry proportions of continental ancestry estimated for each population using STRUCTURE analysis based on 17 Y-STR.

Group	Population	ID Pop	N	African	European	Native Ame.
African American	Noir Marron	GUF_NM	50	*0.780*	*0.114*	*0.106*
	Aluku	NM_A	8	*0.843*	*0.006*	*0.151*
	Ndjuka	NM_N	21	*0.809*	*0.055*	*0.136*
	Paramaka	NM_P	5	*0.606*	*0.390*	*0.004*
	Saramaka	NM_S	16	*0.862*	*0.005*	*0.133*
	Afro-Colombian	Af-COL	19	*0.511*	*0.433*	*0.056*
	Afro-Brazilian	Af-BRA	24	*0.514*	*0.442*	*0.044*
West African	Ahizi	CIV_A	49	*0.970*	*0.028*	*0.002*
	Yacouba	CIV_Y	41	*0.954*	*0.014*	*0.033*
	Fon	BEN_F	78	*0.892*	*0.075*	*0.033*
	Yoruba	BEN_Y	54	*0.913*	*0.055*	*0.032*
	Bariba	BEN_B	51	*0.965*	*0.014*	*0.021*
	Bwa	MLI_B	13	*0.994*	*0.002*	*0.004*
West European	Granada	ESP_Gr	25	*0.084*	*0.847*	*0.069*
	Huelva	ESP_Hu	25	*0.087*	*0.800*	*0.113*
	Barcelona	ESP_Ba	30	*0.005*	*0.977*	*0.018*
	Galicia	ESP_Ga	21	*0.058*	*0.906*	*0.036*
	Portugal North	PRT_N	7	*0.067*	*0.813*	*0.121*
	Portugal South	PRT_S	23	*0.170*	*0.758*	*0.072*
Native American	Kalinya	GUF_K	23	*0.062*	*0.005*	*0.933*
	Oyampi	GUF_O	25	*0.003*	*0.050*	*0.947*
	Palikour	GUF_P	41	*0.002*	*0.026*	*0.972*
	Emberá-Chamí	COL-EC	24	*0.006*	*0.228*	*0.766*
	Karitiana	BRA-KA	17	*0.004*	*0.004*	*0.992*

Table S4: Pairwise F_{ST} genetic distances based on Y haplogroup frequencies (bottom diagonal) and *P*-values (upper diagonal) of the African-American populations and African populations grouped according historical coastal regions. In italic and bold significant *P*-values.

African groups	N		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Aluku	17	1	*	***0.010***	***0.001***	0.180	***0.020***	0.183	0.407	0.088	***0.034***	***0.000***	***0.001***	0.275	0.246	***0.021***	***0.005***	***0.044***	***0.001***
Ndjuka	22	2	0.122	*	***0.001***	0.783	0.281	0.070	***0.027***	***0.009***	***0.047***	***0.001***	0.107	***0.016***	***0.001***	***0.000***	***0.000***	***0.000***	***0.000***
Paramaka	5	3	0.343	0.161	*	0.064	0.121	***0.003***	***0.011***	0.064	0.088	***0.000***	***0.002***	***0.003***	***0.002***	0.069	***0.048***	***0.009***	***0.013***
Saramaka	16	4	0.042	0.029	0.135	*	0.849	0.270	0.280	0.139	0.146	***0.001***	***0.044***	0.183	***0.015***	***0.021***	***0.004***	***0.002***	***0.014***
Noir_Marron	60	5	0.072	0.008	0.072	0.025	*	***0.030***	0.071	0.109	0.186	***0.000***	***0.000***	***0.002***	***0.000***	***0.001***	***0.000***	***0.000***	***0.000***
AfroBrazilian	24	6	0.028	0.044	0.256	0.013	0.046	*	0.705	0.086	0.110	***0.000***	***0.007***	0.228	***0.045***	***0.003***	***0.000***	***0.003***	***0.005***
AfroColombian	19	7	0.000	0.074	0.212	0.013	0.035	0.020	*	0.200	0.117	***0.000***	***0.000***	0.407	0.105	***0.041***	***0.018***	0.082	***0.031***
ACB	47	8	0.044	0.077	0.130	0.030	0.016	0.033	0.015	*	0.744	***0.000***	***0.000***	***0.013***	***0.001***	0.509	0.085	***0.012***	***0.000***
ASW	26	9	0.090	0.057	0.117	0.034	0.014	0.037	0.038	0.016	*	***0.000***	***0.001***	***0.027***	***0.003***	0.297	0.138	***0.011***	***0.000***
Senegambia	252	10	0.360	0.118	0.498	0.190	0.247	0.238	0.340	0.335	0.313	*	***0.012***	***0.000***	***0.000***	***0.000***	***0.000***	***0.000***	***0.000***
Sierra Leone	42	11	0.207	0.030	0.392	0.069	0.117	0.103	0.185	0.178	0.152	0.041	*	***0.000***	***0.000***	***0.000***	***0.000***	***0.000***	***0.000***
Windward Coast	90	12	0.010	0.071	0.272	0.021	0.051	0.011	0.002	0.042	0.052	0.281	0.127	*	0.108	***0.000***	***0.000***	***0.000***	***0.000***
Gold Coast	205	13	0.012	0.165	0.400	0.104	0.130	0.048	0.035	0.089	0.111	0.339	0.191	0.009	*	***0.000***	***0.000***	***0.000***	***0.000***
Bight of Benin	296	14	0.072	0.128	0.133	0.075	0.045	0.084	0.046	0.003	0.003	0.360	0.224	0.068	0.105	*	***0.002***	***0.000***	***0.000***
Bight of Biafra	805	15	0.096	0.136	0.128	0.097	0.063	0.094	0.056	0.012	0.015	0.349	0.237	0.092	0.127	0.010	*	***0.000***	***0.000***
West Central Africa	280	16	0.059	0.207	0.262	0.142	0.115	0.099	0.034	0.039	0.064	0.431	0.296	0.074	0.085	0.029	0.031	*	***0.000***
Southeast Africa	318	17	0.086	0.081	0.139	0.049	0.053	0.044	0.032	0.050	0.071	0.276	0.190	0.100	0.157	0.087	0.080	0.112	*

Table S5: Table of Y-SNP frequencies of 18 populations included in the 1000 Genomes Project Consortium (2015) estimated by using the same Y-SNP set.

Population (N) Y-Haplogroup	ACB 47	ASW 26	GWD 55	MSL 42	YRI 52	ESN 53	LWK 44	IBS 54	TSI 53	GBR 46	FIN 38	CEU 49	PEL 41	CLM 43	PUR 54	MXL 32	CHB 46	CHS 52
A1b-V221	0.0227		0.0741															
B2a-M150							0.0455								0.0196			
B2b-M112							0.0227											
E1a-M33	0.0455		0.2037	0.0500	0.0769										0.0196			
E1b1a1-M2*	0.1136	0.1667	0.5185	0.6250	0.0192		0.0227								0.0392	0.0313		
E1b1a1c1a-M191*(xU174)	0.0227				0.0192	0.0377	0.0227											
E1b1a1c1a1-U174	0.2500	0.2917		0.1000	0.4615	0.4528	0.6136							0.0238	0.0196			
E1b1a1c1a1d-U175*(xU209)		0.0417	0.1111	0.0250		0.4717	0.0682											
E1b1a1a1d1-U209*(xU290)	0.1591	0.0417	0.0370		0.1538	0.0377	0.0909								0.0196			
E1b1a1a1d1a-U290*(xU181)	0.1136	0.1250		0.1750	0.2500									0.0476				
E1b1b1-M35*	0.0227		0.0370				0.0455	0.0192										
E1b1b1a1-M78	0.0227	0.0417					0.0227		0.0755				0.0244	0.0476	0.0980	0.0313		
E1b1b1b1a-M81								0.0192					0.0244		0.0588	0.0938		
E1b1b1b2a1-M34														0.0714	0.0392			
E2-M75*(xM54)		0.0417			0.0192													
E2b-M54			0.0185				0.0455											
C2-M217								0.0385	0.0377						0.0392	0.0625	0.0652	0.0192
G2a-P15		0.0417		0.0250					0.0377			0.0667				0.0313	0.0217	
I1-P203.2	0.0682							0.0192	0.0377	0.1333	0.2432	0.1778						
J1a2b-P58								0.0192	0.0189		0.0270	0.0222	0.0732	0.0714	0.0196			
J1-M267*(xP58)													0.0244	0.0714				
J2a-M410*(xM67)								0.0577	0.0566				0.0244		0.0392			
J2a1b-M67								0.0192	0.1509						0.0196			
J2b-M12*(xM241)								0.0192	0.0189									
J2b2-M241								0.0192	0.0566	0.0222		0.0222						
K-M526*											0.6216						0.1304	
P-M45*(xM207)													0.5854	0.0714		0.3125		
R1-M173*(xM434)													0.0244		0.0196			
R1a1-SRY10831.2								0.0192	0.0377			0.0222	0.0244	0.0714				
R1a1a1b1-Z283										0.0889	0.0811	0.0222						
R1b1a2-M269	0.1591	0.2083						0.7500	0.4528	0.7556	0.0270	0.6667	0.1951	0.5714	0.5098	0.4063	0.0217	
O-P186*																	0.1522	0.0962
O1a-M119																	0.0870	0.0385
O2a1-M95*(xM88)																	0.0217	0.0192
O2a1a-M88																		0.0385
O3-M122*																	0.1522	0.3462
O3a2-P201*(xM134)																	0.0217	0.2692
O3a2c1-M134																	0.3261	0.1731
T1a-M70*(xL131)									0.0377					0.0238	0.0196	0.0313		
T1a2-L131								0.0192	0.0189						0.0196			

Table S6: Table of populations analysed based on major Y haplogroup frequencies in African Americans and historic African regions.

Group	Population	N	A	B	E*-M96	E1a-M33	E1b1a1*-M2	E1b1a1a1c1a*-M191	E1b1a1a1c1a1-U174	E1b1a1a1d-U175	E1b1b1-M35	E2-M75	R1b1a2-M269	Others	Refs*
African American															
1	*Aluku*	17	0.0000	0.0000	0.0000	0.1765	0.1765	0.0000	0.0588	0.4706	0.0000	0.0000	0.0588	0.0588	[1, 2]
2	*Ndjuka*	22	0.0455	0.0909	0.0000	0.0455	0.4091	0.1364	0.1364	0.1364	0.0000	0.0000	0.0000	0.0000	[1, 2]
3	*Paramaka*	5	0.0000	0.0000	0.0000	0.0000	0.0000	0.2000	0.4000	0.0000	0.0000	0.0000	0.0000	0.4000	[1, 2]
4	*Saramaka*	16	0.0000	0.0000	0.0000	0.1250	0.3125	0.1875	0.1250	0.2500	0.0000	0.0000	0.0000	0.0000	[1, 2]
	Noir Marron	60	0.0114	0.0227	0.0000	0.0867	0.2245	0.1310	0.1800	0.2142	0.0000	0.0000	0.0147	0.1147	[1, 2]
5	AfroBrazilian	24	0.0000	0.0000	0.0000	0.0000	0.1250	0.0000	0.0417	0.1667	0.0833	0.0417	0.2917	0.2500	[1]
6	AfroColombian	19	0.0000	0.0000	0.0000	0.0000	0.1053	0.0526	0.0526	0.2632	0.0526	0.0526	0.2632	0.1579	[1]
7	ACB	47	0.0227	0.0000	0.0000	0.0455	0.1136	0.0227	0.2500	0.2727	0.0455	0.0000	0.1591	0.0682	[3]
8	ASW	26	0.0000	0.0000	0.0000	0.0000	0.1667	0.0000	0.2917	0.2083	0.0417	0.0417	0.2083	0.0417	[3]
Senegambia		252	0.0395	0.0000	0.0141	0.1161	0.7129	0.0028	0.0000	0.0275	0.0710	0.0064	0.0000	0.0097	
9	Balanta	28	0.0357	0.0000	0.0357	0.1071	0.7857	0.0000	0.0000	0.0000	0.0357	0.0000	0.0000	0.0000	[4]
10	Bijagos	24	0.0000	0.0000	0.0000	0.0417	0.7917	0.0000	0.0000	0.0000	0.1250	0.0000	0.0000	0.0417	[4]
11	Felupe-Djola	34	0.0294	0.0000	0.0294	0.1765	0.7059	0.0000	0.0000	0.0000	0.0588	0.0000	0.0000	0.0000	[4]
12	Fulbe	51	0.0980	0.0000	0.0196	0.1176	0.7255	0.0000	0.0000	0.0000	0.0196	0.0196	0.0000	0.0000	[4]
13	Mandenka	60	0.0000	0.0000	0.0000	0.0500	0.7500	0.0167	0.0000	0.0167	0.1500	0.0000	0.0000	0.0167	[4, 5]
14	GWD	55	0.0741	0.0000	0.0000	0.2037	0.5185	0.0000	0.0000	0.1481	0.0370	0.0185	0.0000	0.0000	[3]
Sierra Leone		42	0.0000	0.0000	0.0000	0.0500	0.6250	0.0000	0.1000	0.2000	0.0000	0.0000	0.0000	0.0250	
15	MSL	42	0.0000	0.0000	0.0000	0.0500	0.6250	0.0000	0.1000	0.2000	0.0000	0.0000	0.0000	0.0250	[3]
Windward Coast		90	0.0102	0.0000	0.0122	0.0000	0.2912	0.0732	0.1244	0.4786	0.0000	0.0102	0.0000	0.0000	
16	Ahizi	49	0.0204	0.0000	0.0000	0.0000	0.2653	0.0000	0.2245	0.4694	0.0000	0.0204	0.0000	0.0000	[1, 2]
17	Yacouba	41	0.0000	0.0000	0.0244	0.0000	0.3171	0.1463	0.0244	0.4878	0.0000	0.0000	0.0000	0.0000	[1, 2]
Gold Coast		205	0.0000	0.0000	0.0000	0.0000	0.2976	0.0000	0.1073	0.5951	0.0000	0.0000	0.0000	0.0000	
18	Bimoba clans	205	0.0000	0.0000	0.0000	0.0000	0.2976	0.0000	0.1073	0.5951	0.0000	0.0000	0.0000	0.0000	[6]
Bight of Benin		296	0.0025	0.0182	0.0035	0.0460	0.0948	0.0410	0.3706	0.3989	0.0000	0.0114	0.0000	0.0131	
19	Bariba	57	0.0000	0.0000	0.0175	0.0351	0.0702	0.1228	0.1930	0.5088	0.0000	0.0000	0.0000	0.0526	[1]
20	Fon	79	0.0127	0.0000	0.0000	0.0633	0.2025	0.0253	0.2911	0.3544	0.0000	0.0380	0.0000	0.0127	[1, 2]
21	Yoruba.B	55	0.0000	0.0909	0.0000	0.0545	0.1818	0.0000	0.4545	0.2182	0.0000	0.0000	0.0000	0.0000	[1]
22	YRI	52	0.0000	0.0000	0.0000	0.0769	0.0192	0.0192	0.4615	0.4038	0.0000	0.0192	0.0000	0.0000	[3]
23	ESN	53	0.0000	0.0000	0.0000	0.0000	0.0000	0.0377	0.4528	0.5094	0.0000	0.0000	0.0000	0.0000	[3]

Group	Population	N	A	B	E*-M96	E1a-M33	E1b1a1*-M2	E1b1a1a1c1a*-M191	E1b1a1c1a1-U174	E1b1a1a1d-U175	E1b1b1-M35	E2-M75	R1b1a2-M269	Others	Refs.*
Bight of Biafra		**805**	**0.0047**	**0.0874**	**0.0000**	**0.0199**	**0.0442**	**0.0000**	**0.3676**	**0.3558**	**0.0012**	**0.0663**	**0.0519**	**0.0012**	
24	Akele	50	0.0000	0.0200	0.0000	0.0000	0.0800	0.0000	0.3600	0.4000	0.0000	0.1200	0.0200	0.0000	[7]
25	Ateke	48	0.0000	0.0000	0.0000	0.0000	0.0208	0.0000	0.3542	0.4375	0.0000	0.0833	0.1042	0.0000	[7]
26	Bakola	22	0.0000	0.3636	0.0000	0.0455	0.0000	0.0000	0.4091	0.1364	0.0000	0.0000	0.0455	0.0000	[7]
27	Benga	48	0.0000	0.0417	0.0000	0.0625	0.0000	0.0000	0.5208	0.2917	0.0000	0.0208	0.0417	0.0208	[7]
28	Duma	46	0.0217	0.0652	0.0000	0.0000	0.0000	0.0000	0.3913	0.4130	0.0000	0.0870	0.0217	0.0000	[7]
29	Eshira	42	0.0000	0.1667	0.0000	0.0000	0.1190	0.0000	0.1667	0.4286	0.0000	0.1190	0.0000	0.0000	[7]
30	Eviya	24	0.0000	0.2083	0.0000	0.0417	0.0417	0.0000	0.2917	0.2083	0.0000	0.2083	0.0000	0.0000	[7]
31	Fang	60	0.0000	0.0167	0.0000	0.0000	0.0333	0.0000	0.5833	0.1667	0.0000	0.0000	0.2000	0.0000	[7]
32	Galoa	47	0.0000	0.0213	0.0000	0.0213	0.1064	0.0000	0.2553	0.5319	0.0213	0.0426	0.0000	0.0000	[7]
33	Kota	53	0.0000	0.0566	0.0000	0.0189	0.0377	0.0000	0.4906	0.3208	0.0000	0.0189	0.0566	0.0000	[7]
34	Makina	43	0.0000	0.1395	0.0000	0.0000	0.0000	0.0000	0.4651	0.3721	0.0000	0.0233	0.0000	0.0000	[7]
35	Ndumu	36	0.0278	0.0278	0.0000	0.0556	0.0278	0.0000	0.3611	0.3333	0.0000	0.0556	0.1111	0.0000	[7]
36	Nzebi	57	0.0175	0.0702	0.0000	0.0000	0.0526	0.0000	0.3158	0.4737	0.0000	0.0351	0.0351	0.0000	[7]
37	Obamba	47	0.0000	0.0426	0.0000	0.0000	0.1277	0.0000	0.2979	0.3191	0.0000	0.1064	0.1064	0.0000	[7]
38	Orungu	21	0.0000	0.0000	0.0000	0.0952	0.0000	0.0000	0.1905	0.5714	0.0000	0.0952	0.0476	0.0000	[7]
39	Punu	58	0.0000	0.0862	0.0000	0.0172	0.0862	0.0000	0.3276	0.3276	0.0000	0.0345	0.1207	0.0000	[7]
40	Shake	43	0.0000	0.1628	0.0000	0.0000	0.0465	0.0000	0.4186	0.2558	0.0000	0.0930	0.0233	0.0000	[7]
41	Tsogo	60	0.0167	0.0833	0.0000	0.0000	0.0167	0.0000	0.4167	0.4167	0.0000	0.0500	0.0000	0.0000	[7]
West Central Africa		**280**	**0.0000**	**0.0527**	**0.0000**	**0.0000**	**0.0311**	**0.0041**	**0.2804**	**0.5442**	**0.0000**	**0.0874**	**0.0000**	**0.0000**	
42	Ang.Other.Bantu	27	0.0000	0.0370	0.0000	0.0000	0.0370	0.0370	0.2222	0.5926	0.0000	0.0741	0.0000	0.0000	[8]
43	Ganguela	11	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1818	0.6364	0.0000	0.1818	0.0000	0.0000	[8]
44	Kuvale	25	0.0000	0.1200	0.0000	0.0000	0.0800	0.0000	0.1600	0.6000	0.0000	0.0400	0.0000	0.0000	[8]
45	Mbala	11	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.3636	0.4545	0.0000	0.1818	0.0000	0.0000	[5]
46	Mbuun	9	0.0000	0.1111	0.0000	0.0000	0.0000	0.0000	0.4444	0.4444	0.0000	0.0000	0.0000	0.0000	[5]
47	Nyaneka.Nkhumbi	74	0.0000	0.0811	0.0000	0.0000	0.0270	0.0000	0.1757	0.6892	0.0000	0.0270	0.0000	0.0000	[8]
48	Pende	10	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.3000	0.5000	0.0000	0.2000	0.0000	0.0000	[5]
49	Umbundu	93	0.0000	0.0753	0.0000	0.0000	0.0860	0.0000	0.2258	0.5806	0.0000	0.0323	0.0000	0.0000	[8]
50	Yansi	20	0.0000	0.0500	0.0000	0.0000	0.0500	0.0000	0.4500	0.4000	0.0000	0.0500	0.0000	0.0000	[5]
Southeast Africa		**318**	**0.0114**	**0.1109**	**0.0072**	**0.0994**	**0.0481**	**0.0713**	**0.1268**	**0.2177**	**0.2051**	**0.0659**	**0.0000**	**0.0363**	
51	Burunge	23	0.0435	0.2609	0.0000	0.2609	0.0000	0.0000	0.0000	0.0435	0.3478	0.0435	0.0000	0.0000	[9]
52	Chagga	24	0.0000	0.0000	0.0417	0.0000	0.0417	0.3333	0.0000	0.1667	0.1667	0.0417	0.0000	0.2083	[10]
53	Datoga	31	0.0000	0.1290	0.0000	0.0645	0.0323	0.0000	0.0323	0.0645	0.5484	0.1290	0.0000	0.0000	[9]
54	Maputo	78	0.0128	0.1538	0.0000	0.0000	0.0256	0.0000	0.2436	0.4487	0.0128	0.0769	0.0000	0.0256	[11]
55	Masai	12	0.0000	0.0000	0.0000	0.0000	0.0833	0.2500	0.0000	0.5000	0.0833	0.0833	0.0000	0.0000	[10]
56	Mbugwe	14	0.0000	0.0714	0.0000	0.3571	0.0000	0.0000	0.2857	0.0714	0.2143	0.0000	0.0000	0.0000	[9]
57	Sandawe	86	0.0465	0.1163	0.0233	0.0116	0.0000	0.0581	0.1628	0.1977	0.2558	0.0349	0.0000	0.0930	[9, 10]
58	Sukuma	30	0.0000	0.1667	0.0000	0.0000	0.1000	0.0000	0.2667	0.2667	0.0667	0.1333	0.0000	0.0000	[9]
59	Turu	20	0.0000	0.1000	0.0000	0.2000	0.1500	0.0000	0.1500	0.2000	0.1500	0.0500	0.0000	0.0000	[9]

* References of the populations included in the Table S6.

[1] *Present study*.

[2] Brucato N, Cassar O, Tonasso L, Tortevoye P, Migot-Nabias F, Plancoulaine S, et al. (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.

[3] The 1000 Genomes Project Consortium (2015). A global reference for human genetic variation. *Nature*, 526(7571); 68-74.

[4] Rosa A, Ornelas C, Jobling MA, Brehm A, Villems R (2007). Y-chromosomal diversity in the population of Guinea-Bissau: a multiethnic perspective. *BMC Evol Biol*, 7; 124.

[5] de Filippo C, Barbieri C, Whitten M, Mpoloka SW, Gunnarsdottir ED, Bostoen K, et al. (2011). Y-chromosomal variation in sub-Saharan Africa: insights into the history of Niger-Congo groups. *Mol Biol Evol*, 28(3); 1255-69.

[6] Sanchez-Faddeev H, Pijpe J, van der Hulle T, Meij HJ, van der Gaag KJ, Slagboom PE, et al. (2013). The influence of clan structure on the genetic variation in a single Ghanaian village. *Eur J Hum Genet*, 21(10); 1134-9.

[7] Berniell-Lee G, Calafell F, Bosch E, Heyer E, Sica L, Mouguiama-Daouda P, et al. (2009). Genetic and demographic implications of the Bantu expansion: insights from human paternal lineages. *Mol Biol Evol*, 26(7); 1581-9.

[8] Coelho M, Sequeira F, Luiselli D, Beleza S, 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.

[9] Tishkoff SA, Gonder MK, Henn BM, Mortensen H, Knight A, Gignoux C, et al. (2007). History of click-speaking populations of Africa inferred from mtDNA and Y chromosome genetic variation. *Mol Biol Evol*, 24(10); 2180-95.

[10] Xu H, Wang CC, Shrestha R, Wang LX, Zhang M, He Y, et al. (2015). Inferring population structure and demographic history using Y-STR data from worldwide populations. *Mol Genet Genomics*, 290(1); 141-50.

[11] Rowold D, Garcia-Bertrand R, Calderon S, Rivera L, Benedico DP, Alfonso Sanchez MA, et al. (2014). At the southeast fringe of the Bantu expansion: genetic diversity and phylogenetic relationships to other sub-Saharan tribes. *Meta Gene*, 2; 670-85.

B. *Supplementary material of mitochondrial DNA*

Figure S3: Phylogeny of mtDNA trees L0 to L4 for mtDNA haplogroups detected in the African-American and West African populations based on the PhyloTree mtDNA tree Build 16 (19 February 2014) (van Oven & Kayser, 2009).

Table S7: Estimated frequencies of major African mtDNA subhaplogroups L for our dataset of the African-American and African populations from West Africa and other African populations across the African continent and the Near East from previous studies.

Table S8: African-American and African populations analysed in the PCA analysis based on mtDNA haplogroup frequencies, and grouped according historical African coastal regions.

Table S9: Age estimated and confidence intervals for the mtDNA clades analysed in the phylogenetic tree and calculated by using three calibration methods.

****

Figure S3: Phylogeny of mtDNA trees L0 to L4 for mtDNA haplogroups detected in the African-American and West African populations based on the PhyloTree mtDNA tree Build 16 (19 February 2014) (van Oven & Kayser, 2009). Nucleotide position numbers are consistent with both the rCRS and the RSRS. Mutations are given in forward evolutionary time direction in the format [ancestral base][position number][derived base]. In case of a transversion the derived allele is shown in lowercase instead of uppercase. Insertions are indicated by a dot following the respective position number, deletions by the letter "d" following the position number(s) involved. Mutations that are reversions to an ancestral state (back mutations) are indicated with an exclamation mark (!), two exclamation marks for a double back mutation (!!), etc. Coding-region mutations (np 577-16023) are shown in black; control-region mutations (np 16024-576) in blue. Mutations between brackets () are recurrent/unstable within the respective clade, or are yet uncertain based on current data. Mutation motifs in *italic* are preliminary and are likely to be further refined as additional sequences become available. The mutations 309.1C(C), 315.1C, AC indels at 515-522, 16182C, 16183C, 16193.1C(C) and 16519 were not considered for phylogenetic reconstruction and are therefore excluded from these trees.

mt-MRCA

L0 G263A C1048T C3516a T5442C T6185C C9042T A9347G G10589A G12007A A12720G

L0a'b'f'k A189G T4586C C9818T T16172C

L0a'b'f G73A G185A C195T A263G! A2245G C5603T A11641G C15136T G15431A

L0a'b A93G(A95c) T236C C8428T A8566G G9755A C16148T

L0a C146T G5231A G5460A G11176A T14308C C16188g T16278C C16320T

L0a1'4 C16168T

L0a1 T5096C

L0a1a (C64T) T3866C — EU092665 — L0a1a+200

A200G — NA19311

L0a1a1 T2759C — EU092714 — L0a1a1

L0a1a2 C152T G2245c — JX303911 — L0a1a2 — 3

L0a1a3 A9181G — JQ044943 — L0a1a3

A16293G

L0a1b C5911T A14007G C16278T!

L0a1b1 C152T A8191G — AF381988

L0a1b1a G12127A — EU935434

L0a1b1a1 T961C — EU092688 — L0a1b1a1 — 1

L0a2 C64T A185G! G5147A A5711G G6257A 8281-8289d A8460G A11172G A16129G

L0a2a C11143T A14755G

L0a2a2 T204C G207A A9545G G9554A — FJ157838

L0a2a2a C13116T — EU092787 — L0a2a2a — 1 — 1

L0a2a2a1 G8701A G14905A — JX303835 — L0a2a2a1 — 1

L0f (G207A) C4964T T7148C T9581C C9620T A13470G C14109T C14620T T15852C C16169T C16327T T16368C

L0f1 C146T C151T A7364G T16187C C16354T — EU092870 — L0f1 — 1

L0d G1438A T4232C T6815C C8113a G8152A G8251A T12121C G15466A G15930A T15941C T16243C

L0d1'2 C498d A3756G G9755A T16278C

L0d1 G719A G2706A G3438A A6266G G13759A

L0d1b T3618C

L0d1b2 T7283C C14315T C14659T C16294T

L0d1b2a A6692G G8790A A14280G G14560A T15449G G16153A (G16474t)

L0d1b2a1 T9887C — KC346152 — L0d1b2a1 — 1

L1'2'3'4'5'6 C146T C182T T4312C T10664C C10915T A11914G G13276A G16230A

L1 G3666A A7055G T7389C T13789C T14178C G14560A

L1b G185t A357G G709A T710C G1438A T1738C T2352C A2768G T3308C G3693A C6548T T6827C A6989G C7867T A8248G T12519C A14769G T15115C T16126C A16129G C16264T C16270T (A16293G)

L1b1 A5036G G5046A T5655C C13880a A14203G — DQ112737

L1b1a T5393C — JN214480 — L1b1a — 1

L1b1a1'4 C16114a

L1b1a4 G8790A — JQ045107 — L1b1a4 — 1 — 1

L1b1a2 T7954C — EU092672

A189G — JN214471

L1b1a3 G13980A — DQ304908 — L1b1a3 — 1

L1b1a3a C7915T — DQ304914 — L1b1a3a

L1b1a3a1 T11254C — DQ304906 — L1b1a3a1

L1b1a9 G6446A — EU092737 — L1b1a9

L1b1a15 G15077A C15103T — NA19462 — L1b1a15

L1b1a7 T6378C — EU092716 — L1b1a7 — 2 — 1

L1b2'3 C195T

L1b2 A189G C12891T A13893G G14323A C16239T — NA19909 — L1b2

L1b2a T146C! A12171G C16111T T16264C! — HM771162 — L1b2a — 1 — 1

L1c C151T C186a A189c G316A A2395d A5951G T6071C G8027A A9072G G10586A A12810G A13485G T14000a C14911T C16294T C16360T

L1c1'2'4'5'6 A297G

L1c1'2'4'6 C198T T10321C

L1c1 A3796t A3843G A14148G A16293G

L1c1a'b'd T11899C

L1c1a T198C! T4454a T8087C T14088C — JQ705275 — L1c1a+@198 — 1

L1c1a2 A93G A95c T151C! T236C A2755G T2863C C3513T A3927G A4506G A7202G T9647C A12768G G16274A — L1c1a2 — 2

L1c1b'd T16086C

L1c1b T1291C T4688C T5553C C8619T T9861C T10084C T12681C A14393G A16241G — NA19917 — L1c1b — 1

L1c1d G5460A G11914A! G15301A C15626T A16038G — NA19385 — L1c1d

L1c1c T198C! A249d G6267A G8387A A8389G T9233C C11335T T12879C T16172C T16187C T16294C! — EU092717 — L1c1c

L1c2'4 5899.1C C12049T A13149G

L1c2 G6150A T6253C A7076G G7337A A8784G T8877C A10792G C10793T A11654G A16265c G16286g C16527T

L1c2a T1420C 2156.1A C15016T T15784C

L1c2a1 G16145A G16213A

L1c2a1a T7744C G8251A C13212T T13281C C14812T C16071T C16234T — AF346992 — L1c2a1a — 1

L1c2a3 T471C G6723A A11329G G14861A — EU092848 — L1c2a3 — 1

L1c5 A291t G709A A5390G G7762A T8143C 8281-8289d T9899C G11150A A12425G G12630A G13359A G13368A T15449C G15553A T15941C (C16114a) C16261T T16294C! — JX303797 — L1c5 — 1

L1c3 C195T T6221a G6917A G7055A! C11302T A15226G T15905C C15978T

L1c3a G6260A G7498A G7789A G9966A C12019T G12501A T16187C A16215G — EU092703 — L1c3a

L1c3a1 A3105G

L1c3a1a T2387C T12616C — EU935458 — L1c3a1a — 1

L1c3a1b T8668C C16355T — JX303871 — L1c3a1b — 2

L1c3b'c C2283T A16293G

L1c3b G8251A C13981T C14794T A16163G

L1c3b1 T629C C3210T A3434G T4755C C8417T A12400G C12542T T16017C — HG01378

L1c3b1a A11317G G13708A T16209C — HG01080 — L1c3b1a — 1

L1c3b1b A1686G T2083C — JQ702600 — L1c3b1b

L1c3b2 T2283C! T3096C T6297C A6353G G7805A A7844G A14128G G14831A A15244G A15924G T16086C — NA19256 — L1c3b2 — 1

L1c3c A93G T195C! A248G C458T 745.1T T3027C C3525T C3600T G4853A G11852A T11984C A12507G A14669G — EU273493 — L1c3c

L2'3'4'5'6 C152T A2758G C2885T G7146A T8468C

	N. Marron	Colombia	Brazil
Total	12	5	13

238

Haplogroup / mutations	Example	mtDNA genome	N. Marron 43	Colombia 30	Brazil 39
L2'3'4'5'6 C152T A2758G C2885T G7146A T8468C					
L2'3'4'6 C195T A247C A825t T8655C A10688G C10810T G13105A T13506C G15301A A16129G T16187C C16189T					
L2 T146C! C150T T152C! T2416C G8206A A9221G T10115C G13590A C16311T G16390A					
L2a'b'c'd T195C! T11944C					
L2a T150C! T7175C					
L2a1'2'3'4 C2789T C7274T A7771G G11914A! A13803G A14566G C16294T					
L2a1 T182C! A12693G T15784C A16309G					
L2a1a G3918A A5285G A15244G T15629C	DQ304928	L2a1a			
L2a1a1 T6152C C15391T T16368C	NA19266	L2a1a1			
L2a1a2 C16286T	JN214449	L2a1a2			
L2a1a2a T10454C	DQ304975				
L2a1a2a1 C15211T	NA20287				
L2a1a2a1a A15421G	NA19117	L2a1a2a1a			
L2a1a2c T5492C C8841T A13615G	HQ425645	L2a1a2c			
L2a1a3 G143A (T16093C)	EU092711				
L2a1a3a T9165C	EU092890	L2a1a3a			
L2a1a3c C9449T C16256T	HM771168	L2a1a3c		1	1
T16189C! (C16192T)	EU092806				
L2a1f A5581G	DQ304959	L2a1f	2	2	1
L2a1f1 A731G	DQ304961	L2a1f1	1	1	
L2a1f2 C11692T	DQ112698	L2a1f2			1
L2a1f3 C198T	KC533500	L2a1f3			
G143A					
L2a1c G3010A A6663G	EU092733	L2a1c			
T16086C	JQ044973				
L2a1c1 C198T G930A T3308C T8604C	FJ460560	L2a1c1			
L2a1c1a C6311T					
L2a1c1a1 T3338C	JQ705046	L2a1c1a1			
L2a1c2 C10903T A15924G G16213A	JQ045108				
L2a1c2a G513A C16193T C16239T	HM771169	L2a1c2a		1	
L2a1c4 A12172G	JN214432				
L2a1c4a T12354C	DQ304942	L2a1c4a			2
L2a1c4a1 G5252A	DQ304951	L2a1c4a1			
G16129A!	JN858955				
L2a1c5 T13260C	NA18908	L2a1c5			
G16309A!					
L2a1h A3505G T4772C C12976T	EU092676	L2a1h	2		1
L2a1e C3495a G8790A G12630A	DQ304945	L2a1e		1	
L2a1e1 A143G! G8541A A14599G	DQ304930	L2a1e1			3
T16189C! (C16192T)	FJ460527				
G16309A!	EU092793				
L2a1i T15229C T16362C	JQ044958	L2a1i			
L2a1i1 C6164T C10920T	JX303853	L2a1i1			1
L2a1l C534T		L2a1l			
L2a1l1 T12408C	JQ044956	L2a1l1			
L2a1l1a T5580C	EU092807	L2a1l1a	1		1
L2a1m A13884G	JQ045040	L2a1m			
L2a1n A4317d G5147A	DQ304941	L2a1n			1
L2b'c'd C2332T					
L2b'c C198T G1442A T7624a G12236A G15110A G15217A					
L2b T204C C1706T A2358G A4158G T4370C A4767G C5027T C5331a T5814C C6713T C8080T G8387A A12948G A14059G C16114a G16129A! G16213A					
L2b1 C418T G6026A T10828C C13924T T16362C	EU092747				
L2b1a C146T!! C16355T	JQ044854	L2b1a			
L2b1a2 A7569G	EU092722	L2b1a2		1	
L2b1a3 G8856A	DQ304978	L2b1a3			1
L2b1b A385G A6629G	FJ228403	L2b1b			
L2b2 T6614C A6806G T8503C	EU092692	L2b2			
L2b3 15944.1T		L2b3			
L2b3a C4185T G5744A A8925G G14544A A15236G G15326A	JQ702626	L2b3a	2	1	
L2b3b T2626C	EU092734	L2b3b			
L2c A93G C325T T680C G709A T3200a G13928c G13958c C15849T	JQ044941	L2c	2		
L2c1 A16318G	JQ044858	L2c1	1		
L2c2 T1040C C16264T	NA19229	L2c2	4	1	
L2c2a G15043A	DQ304988	L2c2a			
L2c2b A183G					
L2c2b1 T8772C T15313C					
L2c2b1b T8567C A9063G T10790C T16311C!	KC533455	L2c2b1b	1		
L2c3 G513A	JQ705626	L2c3	3		
L2c4 C13440T	JQ044914	L2c4			
L2d C152T!! T182C! C456T C870T T2159C C3254a A3434G G3693A C6231T G9554A A9941G C10955T T11353C C14845T G15777c T16189C! T16223C A16300G C16354T A1699G	JQ045050	L2d+16129			
L3'4'6 G4104A A7521G					
L3'4 T182C! T3594C T7256C T13650C T16278C					
L4 T195C! G5460A T16362C					
L4a C198T G3357A G10373A T11253C A11344G T11485C T12414C T13174C T14302C C16260T					
L4a1 C325T T6167C C7376T G7775A A11653G T14000a A16207t	FJ460531				
L4a1a (C150T) G7762A T8473C A8631G	DQ341064	L4b1a			
L3 A769G A1018G C16311T					
Total			33	17	22

239

	Example	mtDNA genome	N. Marron 43	Colombia 30	Brazil 39
L3 A769G A1018G C16311T					
L3b'f T15944d					
L3b C3450T G5773A T6221C C9449T A10086G A13105G! C13914a A15311G A15824G T16124C C16278T! T16362C					
L3b1 G10373A					
L3b1a A11002G	EU092726	**L3b1a**	1		
L3b1a1 A11800G	JQ705783				
L3b1a1a G15883A	DQ304995	**L3b1a1a**			
L3b1a5 T5063C	JQ044986	**L3b1a5**	1		
L3b1a5a G5417A	NA18865	**L3b1a5a**			
T152C!	JN655790				
L3b1a6 A3385G C5255T G8155A	EU092958	**L3b1a6**			
C16124T!	EU092768				
L3b1a7 A3441G C5211T A5581G T15299C	JX666328				
L3b1a7a G9477A	DQ341073	**L3b1a7a**			
L3b1a8 C2332T G16145A	NA19222	**L3b1a8**			
L3f T3396C T4218C T15514C T16209C					
L3f1 C5601T T9950C					
L3f1b A189G (A200G) T1822C C7819a A8527G C8932T G11440A A14769G T16311C!		**L3f1b+16292**			
		L3f1b+16292+150		1	
C16292T	DQ341077				
L3f1b1 C16295T	JQ044831	**L3f1b1**	1		
L3f1b1a C8410T C10070T G16129A!	DQ305036	**L3f1b1a**			
L3f1b1a1 A272G	JX303774	**L3f1b1a1**			1
C150T	NA19172				
L3f1b3 T711C A6806G A7158G	EU092883	**L3f1b3**			
L3f1b4 A3505G					
L3f1b4a A13167G T16292C!	EU092865	**L3f1b4a**			1
L3f1b4a1 A8799G	EU092912	**L3f1b4a1**			1
L3f1b4c C16218T	EU092704	**L3f1b4c**			
L3e'd T152C! A13105G!					
L3d G5147A A7424G T8618C T13886C C14284T T16124C	DQ112884				
L3d1'2'3'4'5'6 T921C	EU092898	**L3d1'2'3'4'5'6**			
L3d1 T6680C					
L3d1a G4048A C7648T G11887A	EU092796				
L3d1a1'2 G16319A					
L3d1a1 G1503A					
L3d1a1a C150T A4203G G5471A T10640C T10915C!	JN655800	**L3d1a1a**			1
L3d1a1b T146C! T195C! T5162C G9452A C10837T G12406A A13884G G15106A	EU092876	**L3d1a1b**			
L3d1a2 A15061G	HM771228	**L3d1a2**			
L3d1b G5046A	JQ045112				
L3d1b3 T146C! T14634C G15110A	JQ045051	**L3d1b3**			
L3d1b3a C16256T	JX303821	**L3d1b3a**			1
L3d1c A3203G T9111C A11239G C12870T A13542G A16166G	EU092762				
L3d1c1 A5372C A9254G	EU092830	**L3d1c1**			
L3d1d A7765G A9151G (C16256T) (T16368C)	AY195782	**L3d1d**			
L3d2 C14272T T14584C C16111T					
L3d2b T199C	JQ705019	**L3d2b**			
L3d3 G1719A T4688C A15061G					
L3d3a C3498g G8251A A10899G A11404G (T16189C) (C16278T) (T16304C) (T16311C)	DQ112716				
L3d3a1 A15208G	JX303864	**L3d3a1**			1
L3d5 A15799G T16362C	JN214479				
L3d5a A7158G T12311C G15930A	NA19121	**L3d5a**			
L3e'i'k'x C150T A10819G					
L3e T2352C T14212C					
L3e1 A189G A200G T6221C C6587T A14152G T15670C T15942C C16327T	EU092827	**L3e1**			1
L3e1a C16185T		**L3e1a**		1	
L3e1e T10370C	JQ045096	**L3e1e**	2		1
L3e1f 8281-8289d	JQ704728	**L3e1f**	1		1
L3e1f1 T16189C!	HM771135				
L3e1f1a A4395G C16260T	HM771172	**L3e1f1a**	1		
L3e1g G1442A A4844G A5498G T5618C T8093C T10490C T11318C T12215C T16172C A16399G	JQ044847	**L3e1g**			
L3e2 T195C! G14905A C16320T					
L3e2a T4823C A13105G! G14869A	EU092769	**L3e2a**			
L3e2a1 C198T					
L3e2a1a A1737G C3852T C7196T A15924G	DQ305026	**L3e2a1a**	1		
L3e2a1b T6413C	DQ305029	**L3e2a1b**	1		
L3e2a1b1 A16399G	DQ305031	**L3e2a1b1**			1
L3e2b T16172C T16189C!	DQ341071	**L3e2b**	1		2
		L3e2b+152		2	1
L3e2b1 A9377G	DQ305022				
L3e2b1a T2483C	JQ703138				
L3e2b1a1 T5580C	DQ305021	**L3e2b1a1**			
T152C!	DQ304998				
L3e2b3 A6932G	DQ305008	**L3e2b3**		2	1
L3e2b4 T10861C	DQ305000	**L3e2b4**		1	
L3e2b5 G6261A	NA19238	**L3e2b5**			
L3e3'4'5 G750A					
L3e3,4 G5262A					
L3e3 T195C! C2000T T6524C G9554A T10667C A10816G A13101c A16265t					
L3e3b G4655A A12248G C13197T A13651G G15812A	DQ305011	**L3e3b**			
L3e3b3 C16465T	JQ044850	**L3e3b3**			1
L3e5 T398C G8392A A16041G	KF358485	**L3e5**			
L3e5a G13317A	JN214452	**L3e5a**			
L3h T7861C G9575A					
L3h1 G1719A A4388G C5300T T9509C A11590G T16311C!					
L3h1b A189c T195C! A10044G G14410A C16256a					
L3h1b2 C151T T152C! T294C A606G T990C A8842G T9758C C12882T T13437C G16129A! T16362C	EU092736	**L3h1b2**		1	
Total			10	8	15

240

Table S7: Estimated frequencies of major African mtDNA subhaplogroups L for the dataset of: African Americans (N= 252), African populations across the continent (N= 14,321) and Near East populations (N=3,114). There are included 17,687 individuals in total.

n°	Country / Population	ISO Code	Region	N	Lat.	Long.	L0*	L0a	L0d	L1b	L1c	L2*	L2a	L3*	L3b	L3d	L3e	L3f	L*	Others	References
A	**Noir Marron (F.G.)**	**GUF**	**AAM**	**183**	**5.500**	**-54.030**	**0.000**	**0.044**	**0.000**	**0.126**	**0.186**	**0.175**	**0.213**	**0.000**	**0.027**	**0.027**	**0.169**	**0.033**	**0.000**	**0.000**	**[1, 2]**
	Aluku	GUF-A	AAM	25	3.830	-54.18	0.000	0.080	0.000	0.080	0.280	0.160	0.080	0.000	0.000	0.000	0.280	0.040	0.000	0.000	[1, 2]
	Ndjuka	GUF-N	AAM	97	4.250	-54.38	0.000	0.010	0.000	0.103	0.144	0.186	0.309	0.000	0.041	0.010	0.144	0.052	0.000	0.000	[1, 2]
	Paramaka	GUF-P	AAM	19	4.860	-54.47	0.000	0.000	0.000	0.211	0.000	0.316	0.211	0.000	0.000	0.000	0.263	0.000	0.000	0.000	[1, 2]
	Saramaka	GUF-S	AAM	42	5.850	-55.20	0.000	0.119	0.000	0.167	0.310	0.095	0.071	0.000	0.024	0.095	0.119	0.000	0.000	0.000	[1, 2]
B	**Afro-Brazilian**	**BRA**	**AAM**	**39**	**-22.910**	**-43.200**	**0.026**	**0.051**	**0.026**	**0.103**	**0.128**	**0.026**	**0.205**	**0.000**	**0.000**	**0.077**	**0.231**	**0.077**	**0.000**	**0.051**	**[1]**
C	**Afro-Colombian**	**COL**	**AAM**	**30**	**5.690**	**-76.660**	**0.000**	**0.100**	**0.000**	**0.067**	**0.000**	**0.133**	**0.267**	**0.033**	**0.000**	**0.000**	**0.200**	**0.033**	**0.000**	**0.167**	**[1]**
1	Western Sahara	ESH	NAF	110	25.178	-13.779	0.000	0.000	0.000	0.027	0.000	0.027	0.091	0.018	0.018	0.027	0.009	0.027	0.000	0.755	[3, 4]
2	Morocco	MAR	NAF	1,103	33.222	-7.522	0.004	0.003	0.000	0.064	0.002	0.015	0.036	0.000	0.000	0.000	0.021	0.000	0.000	0.855	[3-9]
3	Algeria	DZA	NAF	47	36.709	3.073	0.000	0.021	0.000	0.000	0.000	0.000	0.064	0.000	0.021	0.021	0.106	0.000	0.000	0.766	[3]
4	Tunisia	TUN	NAF	551	36.538	10.056	0.000	0.002	0.000	0.033	0.009	0.020	0.064	0.005	0.018	0.011	0.051	0.034	0.000	0.753	[3, 8, 10-12]
5	Libya	LBY	NAF	418	32.815	13.148	0.000	0.029	0.000	0.038	0.002	0.022	0.074	0.012	0.017	0.017	0.036	0.038	0.000	0.715	[13, 14]
6	Egypt	EGY	NAF	594	30.422	31.106	0.003	0.045	0.000	0.012	0.013	0.003	0.044	0.000	0.000	0.000	0.061	0.025	0.000	0.793	[6, 15-18]
7	Mauritania	MRT	WAF	94	18.158	-16.003	0.000	0.000	0.000	0.181	0.043	0.032	0.117	0.000	0.000	0.000	0.000	0.000	0.000	0.628	[4, 19]
8	Cape Verde	CPV	WAF	292	14.950	-23.522	0.000	0.007	0.000	0.079	0.068	0.209	0.205	0.007	0.106	0.072	0.147	0.007	0.000	0.092	[20]
9	Senegal	SEN	WAF	293	14.760	-17.318	0.000	0.007	0.000	0.171	0.020	0.167	0.218	0.034	0.147	0.038	0.034	0.017	0.000	0.147	[4, 21-23]
10	Gambia	GMB	WAF	113	13.442	-16.620	0.000	0.000	0.000	0.115	0.018	0.265	0.159	0.044	0.115	0.124	0.088	0.000	0.018	0.053	[24]
11	Sierra Leone	SLE	WAF	276	7.899	-11.197	0.000	0.033	0.000	0.147	0.059	0.191	0.202	0.011	0.118	0.092	0.077	0.048	0.000	0.022	[24, 25]
12	**Mali**	**MLI**	**WAF**	**328**	**12.690**	**-7.980**	**0.000**	**0.018**	**0.000**	**0.143**	**0.028**	**0.180**	**0.232**	**0.000**	**0.128**	**0.067**	**0.088**	**0.028**	**0.000**	**0.088**	**[1, 19, 26-28]**
	Mali-Bwa	MLI	WAF	23	12.690	-7.980	0.000	0.130	0.000	0.043	0.000	0.087	0.565	0.000	0.000	0.174	0.000	0.000	0.000	0.000	[1]
13	Burkina Faso	BFA	WAF	427	12.345	-1.564	0.000	0.023	0.000	0.126	0.019	0.155	0.272	0.009	0.122	0.056	0.077	0.019	0.012	0.110	[23, 27, 28]
14	**Ivory Coast**	**CIV**	**WAF**	**192**	**5.420**	**-4.016**	**0.000**	**0.026**	**0.000**	**0.094**	**0.130**	**0.109**	**0.328**	**0.005**	**0.073**	**0.016**	**0.161**	**0.047**	**0.000**	**0.010**	**[1, 2]**
	Yacouba	CIV-Y	WAF	63	7.270	-8.170	0.000	0.079	0.000	0.095	0.095	0.079	0.286	0.016	0.095	0.048	0.111	0.063	0.000	0.032	[1, 2]
	Ahizi	CIV-A	WAF	129	5.250	-4.610	0.000	0.000	0.000	0.093	0.147	0.124	0.349	0.000	0.062	0.000	0.186	0.039	0.000	0.000	[1, 2]
15	Ghana	GHA	WAF	238	5.666	-0.245	0.000	0.008	0.000	0.097	0.076	0.088	0.315	0.004	0.067	0.050	0.197	0.046	0.000	0.050	[29]
16	**Benin**	**BEN**	**WAF**	**214**	**9.360**	**2.601**	**0.000**	**0.049**	**0.000**	**0.092**	**0.046**	**0.092**	**0.261**	**0.007**	**0.092**	**0.105**	**0.190**	**0.056**	**0.007**	**0.007**	**[1, 2]**
	Beninese	BEN-Be	WAF	58	9.360	2.601	0.000	0.069	0.000	0.086	0.052	0.086	0.293	0.000	0.121	0.069	0.207	0.017	0.000	0.000	[1, 2]
	Bariba	BEN-Ba	WAF	32	9.350	2.620	0.000	0.063	0.000	0.125	0.031	0.125	0.188	0.000	0.031	0.188	0.188	0.063	0.000	0.000	[1]
	Yoruba	BEN-Y	WAF	32	7.360	2.600	0.000	0.031	0.000	0.031	0.063	0.094	0.281	0.000	0.125	0.125	0.188	0.063	0.000	0.000	[1]
	Fon	BEN-F	WAF	92	6.370	2.430	0.000	0.043	0.000	0.098	0.043	0.087	0.261	0.011	0.087	0.098	0.185	0.065	0.011	0.011	[1, 2]
17	Nigeria	NGA	WAF	1,554	6.649	3.336	0.001	0.046	0.000	0.086	0.128	0.093	0.168	0.024	0.066	0.042	0.205	0.085	0.003	0.053	[22, 24, 27, 29]
18	Niger	NER	WAF	233	13.574	2.155	0.000	0.021	0.000	0.185	0.013	0.150	0.116	0.017	0.240	0.069	0.069	0.034	0.009	0.077	[22, 27, 28]

n°	Country / Population	ISO Code	Region	N	Lat.	Long.	L0*	L0a	L0d	L1b	L1c	L2*	L2a	L3*	L3b	L3d	L3e	L3f	L*	Others	References
19	Chad	TCD	WCA	138	12.116	15.054	0.000	0.022	0.000	0.225	0.014	0.109	0.167	0.036	0.203	0.043	0.080	0.036	0.021	0.043	[22, 27]
20	Cameroon	CMR	WCA	859	3.893	11.532	0.000	0.081	0.000	0.045	0.332	0.035	0.108	0.013	0.061	0.049	0.162	0.066	0.003	0.044	[27, 29, 30]
21	Equatorial Guinea	GNQ	WCA	45	1.552	10.259	0.000	0.089	0.000	0.000	0.267	0.044	0.200	0.000	0.000	0.000	0.244	0.000	0.000	0.156	[31]
22	Sao Tomé	STP	WCA	153	0.339	6.736	0.000	0.072	0.000	0.144	0.163	0.105	0.131	0.000	0.026	0.046	0.209	0.033	0.000	0.072	[31, 32]
23	Gabon	GAB	WCA	946	0.446	9.510	0.000	0.080	0.000	0.054	0.424	0.050	0.125	0.011	0.020	0.037	0.145	0.050	0.005	0.000	[30]
24	C.A.R. (Pygmy, Biaka)	CAF	WCA	56	4.489	18.493	0.000	0.071	0.000	0.000	0.893	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.036	[30]
25	D.R.C. (Pygmy, Mbuti)	COD	SEA	39	-6.165	23.604	0.000	0.256	0.000	0.000	0.000	0.000	0.641	0.026	0.000	0.000	0.000	0.000	0.077	0.000	[30]
26	Congo (Pygmy, Biaka)	COG	SEA	17	-4.102	12.885	0.000	0.235	0.000	0.000	0.765	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	[22]
27	Angola	AGO	SEA	709	-14.923	13.493	0.004	0.171	0.062	0.041	0.193	0.071	0.103	0.007	0.025	0.051	0.172	0.068	0.000	0.032	[33-36]
28	Zambia	ZMB	SEA	541	-15.370	28.308	0.017	0.165	0.026	0.031	0.198	0.048	0.185	0.000	0.030	0.039	0.214	0.024	0.019	0.006	[37, 38]
29	Sudan	SDN	EAF	178	15.545	32.538	0.000	0.051	0.000	0.006	0.011	0.045	0.298	0.096	0.017	0.011	0.006	0.096	0.073	0.292	[15, 39]
30	Ethiopia	ETH	EAF	648	9.065	38.756	0.025	0.085	0.000	0.014	0.002	0.026	0.110	0.130	0.008	0.022	0.008	0.037	0.190	0.346	[39-42]
31	Somalia	SOM	EAF	183	2.050	45.370	0.016	0.060	0.005	0.016	0.011	0.000	0.148	0.142	0.011	0.038	0.033	0.055	0.062	0.402	[22, 39, 43]
32	Kenya	KEN	EAF	570	-1.217	36.775	0.091	0.170	0.005	0.005	0.005	0.018	0.098	0.168	0.037	0.011	0.051	0.053	0.058	0.230	[22, 24, 44, 45]
33	Rwanda	RWA	EAF	42	-1.951	30.110	0.167	0.167	0.000	0.000	0.024	0.048	0.071	0.143	0.167	0.024	0.071	0.000	0.119	0.000	[46]
34	Tanzania	TZA	EAF	378	-6.833	39.243	0.066	0.177	0.016	0.000	0.013	0.003	0.106	0.056	0.024	0.016	0.042	0.005	0.371	0.106	[43, 47]
35	Comoros Island	COM	ISL	88	-11.504	43.384	0.000	0.000	0.011	0.000	0.011	0.023	0.023	0.000	0.023	0.023	0.034	0.000	0.000	0.852	[48]
36	Madagascar	MDG	ISL	268	-18.938	47.472	0.000	0.060	0.000	0.000	0.007	0.004	0.119	0.011	0.078	0.011	0.104	0.000	0.000	0.604	[49, 50]
37	Mozambique	MOZ	SAF	338	-25.590	32.424	0.000	0.074	0.062	0.015	0.062	0.036	0.405	0.000	0.036	0.068	0.180	0.030	0.009	0.024	[51, 52]
38	Zimbabwe	ZWE	SAF	59	-17.805	31.044	0.017	0.271	0.017	0.034	0.085	0.034	0.203	0.000	0.068	0.068	0.186	0.000	0.017	0.000	[46]
39	Botswana	BWA	SAF	96	-24.567	25.915	0.000	0.177	0.000	0.021	0.042	0.021	0.240	0.000	0.010	0.104	0.208	0.052	0.115	0.010	[53]
40	Namibia	NAM	SAF	262	-22.552	17.056	0.111	0.034	0.424	0.011	0.031	0.008	0.015	0.004	0.000	0.206	0.080	0.073	0.004	0.000	[53]
41	South Africa	ZAF	SAF	631	-33.457	19.329	0.025	0.046	0.597	0.006	0.010	0.008	0.046	0.010	0.010	0.000	0.032	0.005	0.011	0.195	[54, 55]
42	Iran	IRN	NEA	738	35.687	51.452	0.000	0.004	0.000	0.000	0.001	0.000	0.001	0.001	0.000	0.000	0.001	0.000	0.000	0.991	[56, 57]
43	Iraq	IRQ	NEA	167	33.336	44.400	0.000	0.006	0.000	0.000	0.012	0.000	0.030	0.006	0.000	0.000	0.000	0.012	0.000	0.934	[58]
44	Syria	SYR	NEA	116	35.137	36.742	0.000	0.009	0.000	0.009	0.000	0.000	0.017	0.000	0.000	0.000	0.009	0.009	0.000	0.948	[58]
45	Jordan	JOR	NEA	142	31.962	35.943	0.000	0.014	0.000	0.007	0.000	0.007	0.028	0.014	0.028	0.007	0.000	0.007	0.000	0.887	[60]
46	Israel	ISR	NEA	216	31.257	34.826	0.000	0.005	0.000	0.000	0.000	0.000	0.032	0.019	0.005	0.014	0.023	0.005	0.000	0.898	[58, 59]
47	Kuwait	KWT	NEA	381	29.261	47.931	0.000	0.010	0.000	0.003	0.005	0.010	0.042	0.003	0.003	0.005	0.005	0.003	0.000	0.911	[61]
48	Saudi Arabia	SAU	NEA	553	24.628	46.722	0.000	0.011	0.000	0.002	0.004	0.004	0.031	0.011	0.004	0.005	0.000	0.007	0.000	0.922	[62, 63]
49	Dubai	SAU-D	NEA	249	25.050	55.175	0.000	0.052	0.000	0.004	0.012	0.004	0.048	0.000	0.000	0.000	0.032	0.008	0.000	0.839	[64]
50	Oman	OMN	NEA	196	23.012	57.584	0.000	0.031	0.000	0.000	0.000	0.010	0.010	0.015	0.010	0.000	0.036	0.000	0.000	0.888	[43]
51	Yemen	YEM	NEA	552	15.512	44.214	0.005	0.063	0.000	0.004	0.011	0.007	0.038	0.036	0.011	0.029	0.033	0.016	0.000	0.746	[40, 42, 43, 58, 65]

References of the populations included in the Table S7:   [1] *Present study.*

[2] Brucato N, Cassar O, Tonasso L, Tortevoye P, Migot-Nabias F, Plancoulaine S, et al. (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.

[3] Plaza S, Calafell F, Helal A, Bouzerna N, Lefranc G, Bertranpetit J, et al. (2003). Joining the pillars of Hercules: mtDNA sequences show multidirectional gene flow in the western Mediterranean. *Ann Hum Genet*, 67(Pt 4); 312-28.

[4] Rando JC, Pinto F, Gonzalez AM, Hernandez M, Larruga JM, Cabrera VM, et al. (1998). Mitochondrial DNA analysis of northwest African populations reveals genetic exchanges with European, near-eastern, and sub-Saharan populations. *Ann Hum Genet*, 62(Pt 6); 531-50.

[5] Rhouda T, Martinez-Redondo D, Gomez-Duran A, Elmtili N, Idaomar M, Diez-Sanchez C, et al. (2009). Moroccan mitochondrial genetic background suggests prehistoric human migrations across the Gibraltar Strait. *Mitochondrion*, 9(6); 402-7.

[6] Coudray C, Olivieri A, Achilli A, Pala M, Melhaoui M, Cherkaoui M, et al. (2009). The complex and diversified mitochondrial gene pool of Berber populations. *Ann Hum Genet*, 73(2); 196-214.

[7] Harich N, Costa MD, Fernandes V, Kandil M, Pereira JB, Silva NM, et al. (2010). The trans-Saharan slave trade - clues from interpolation analyses and high-resolution characterization of mitochondrial DNA lineages. *BMC Evol Biol*, 10; 138.

[8] Turchi C, Buscemi L, Giacchino E, Onofri V, Fendt L, Parson W, et al. (2009). Polymorphisms of mtDNA control region in Tunisian and Moroccan populations: an enrichment of forensic mtDNA databases with Northern Africa data. *Forensic Sci Int Genet*, 3(3); 166-72.

[9] Brakez Z, Bosch E, Izaabel H, Akhayat O, Comas D, Bertranpetit J, et al. (2001). Human mitochondrial DNA sequence variation in the Moroccan population of the Souss area. *Ann Hum Biol*, 28(3); 295-307.

[10] Cherni L, Fernandes V, Pereira JB, Costa MD, Goios A, Frigi S, et al. (2009). Post-last glacial maximum expansion from Iberia to North Africa revealed by fine characterization of mtDNA H haplogroup in Tunisia. *Am J Phys Anthropol*, 139(2); 253-60.

[11] Fadhlaoui-Zid K, Plaza S, Calafell F, Ben Amor M, Comas D, Bennamar El gaaied A (2004). Mitochondrial DNA heterogeneity in Tunisian Berbers. *Ann Hum Genet*, 68(Pt 3); 222-33.

[12] Loueslati BY, Cherni L, Khodjet-Elkhil H, Ennafaa H, Pereira L, Amorim A, et al. (2006). Islands inside an island: reproductive isolates on Jerba island. *Am J Hum Biol*, 18(1); 149-53.

[13] Ottoni C, Martinez-Labarga C, Loogvali EL, Pennarun E, Achilli A, De Angelis F, et al. (2009). First genetic insight into Libyan Tuaregs: a maternal perspective. *Ann Hum Genet*, 73(Pt 4); 438-48.

[14] Fadhlaoui-Zid K, Rodriguez-Botigue L, Naoui N, Benammar-Elgaaied A, Calafell F, Comas D (2011). Mitochondrial DNA structure in North Africa reveals a genetic discontinuity in the Nile Valley. *Am J Phys Anthropol*, 145(1); 107-17.

[15] Krings M, Salem AE, Bauer K, Geisert H, Malek AK, Chaix L, et al. (1999). mtDNA analysis of Nile River Valley populations: A genetic corridor or a barrier to migration? *Am J Hum Genet*, 64(4); 1166-76.

[16] Saunier JL, Irwin JA, Strouss KM, Ragab H, Sturk KA, Parsons TJ (2009). Mitochondrial control region sequences from an Egyptian population sample. *Forensic Sci Int Genet*, 3(3); e97-103.

[17] Kujanova M, Pereira L, Fernandes V, Pereira JB, Cerny V (2009). Near eastern neolithic genetic input in a small oasis of the Egyptian Western Desert. *Am J Phys Anthropol*, 140(2); 336-46.

[18] Stevanovitch A, Gilles A, Bouzaid E, Kefi R, Paris F, Gayraud RP, et al. (2004). Mitochondrial DNA sequence diversity in a sedentary population from Egypt. *Ann Hum Genet*, 68(Pt 1); 23-39.

[19] Gonzalez AM, Cabrera VM, Larruga JM, Tounkara A, Noumsi G, Thomas BN, et al. (2006). Mitochondrial DNA variation in Mauritania and Mali and their genetic relationship to other Western Africa populations. *Ann Hum Genet*, 70(Pt 5); 631-57.

[20] Brehm A, Pereira L, Bandelt HJ, Prata MJ, Amorim A (2002). Mitochondrial portrait of the Cabo Verde archipelago: the Senegambian outpost of Atlantic slave trade. *Ann Hum Genet*, 66(1); 49-60.

[21] Stefflova K, Dulik MC, Pai AA, Walker AH, Zeigler-Johnson CM, Gueye SM, et al. (2009). Evaluation of group genetic ancestry of populations from Philadelphia and Dakar in the context of sex-biased admixture in the Americas. *PLoS One*, 4(11); e7842.

[22] Watson E, Forster P, Richards M, Bandelt HJ (1997). Mitochondrial footprints of human expansions in Africa. *Am J Hum Genet*, 61(3); 691-704.

[23] Barbieri C, Whitten M, Beyer K, Schreiber H, Li M, Pakendorf B (2012). Contrasting maternal and paternal histories in the linguistic context of Burkina Faso. *Mol Biol Evol*, 29(4); 1213-23.

[24] The 1000 Genomes Project Consortium (2015). A global reference for human genetic variation. *Nature*, 526(7571); 68-74.

[25] Jackson BA, Wilson JL, Kirbah S, Sidney SS, Rosenberger J, Bassie L, et al. (2005). Mitochondrial DNA genetic diversity among four ethnic groups in Sierra Leone. *Am J Phys Anthropol*, 128(1); 156-63.

[26] Ely B, Wilson JL, Jackson F, Jackson BA (2006). African-American mitochondrial DNAs often match mtDNAs found in multiple African ethnic groups. *BMC Biol*, 4; 34.

[27] Cerny V, Pereira L, Musilova E, Kujanova M, Vasikova A, Blasi P, et al. (2011). Genetic structure of pastoral and farmer populations in the African Sahel. *Mol Biol Evol*, 28(9); 2491-500.

[28] Pereira L, Cerny V, Cerezo M, Silva NM, Hajek M, Vasikova A, et al. (2010). Linking the sub-Saharan and West Eurasian gene pools: maternal and paternal heritage of the Tuareg nomads from the African Sahel. *Eur J Hum Genet*, 18(8); 915-23.

[29] Veeramah KR, Connell BA, Ansari Pour N, Powell A, Plaster CA, Zeitlyn D, et al. (2010). Little genetic differentiation as assessed by uniparental markers in the presence of substantial language variation in peoples of the Cross River region of Nigeria. *BMC Evol Biol*, 10; 92.

[30] Quintana-Murci L, Quach H, Harmant C, Luca F, Massonnet B, Patin E, et al. (2008). Maternal traces of deep common ancestry and asymmetric gene flow between Pygmy hunter-gatherers and Bantu-speaking farmers. *Proc Natl Acad Sci U S A*, 105(5); 1596-601.

[31] Mateu E, Comas D, Calafell F, Perez-Lezaun A, Abade A, Bertranpetit J (1997). A tale of two islands: population history and mitochondrial DNA sequence variation of Bioko and Sao Tome, Gulf of Guinea. *Ann Hum Genet*, 61(Pt 6); 507-18.

[32] Trovoada MJ, Pereira L, Gusmao L, Abade A, Amorim A, Prata MJ (2004). Pattern of mtDNA variation in three populations from Sao Tome e Principe. *Ann Hum Genet*, 68(Pt 1); 40-54.

[33] Plaza S, Salas A, Calafell F, Corte-Real F, Bertranpetit J, Carracedo A, et al. (2004). Insights into the western Bantu dispersal: mtDNA lineage analysis in Angola. *Hum Genet*, 115(5); 439-47.

[34] Beleza S, Gusmao L, Amorim A, Carracedo A, Salas A (2005). The genetic legacy of western Bantu migrations. *Hum Genet*, 117(4); 366-75.

[35] Coelho M, Sequeira F, Luiselli D, Beleza S, 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.

[36] Barbieri C, Vicente M, Oliveira S, Bostoen K, Rocha J, Stoneking M, et al. (2014). Migration and interaction in a contact zone: mtDNA variation among Bantu-speakers in Southern Africa. *PLoS One*, 9(6); e99117.

[37] de Filippo C, Heyn P, Barham L, Stoneking M, Pakendorf B (2010). Genetic perspectives on forager-farmer interaction in the Luangwa valley of Zambia. *Am J Phys Anthropol*, 141(3); 382-94.

[38] Barbieri C, Butthof A, Bostoen K, Pakendorf B (2013). Genetic perspectives on the origin of clicks in Bantu languages from southwestern Zambia. *Eur J Hum Genet*, 21(4); 430-6.

[39] Soares P, Alshamali F, Pereira JB, Fernandes V, Silva NM, Afonso C, et al. (2012). The Expansion of mtDNA Haplogroup L3 within and out of Africa. Mol Biol Evol, 29(3); 915-27.

[40] Non AL, Al-Meeri A, Raaum RL, Sanchez LF, Mulligan CJ (2011). Mitochondrial DNA reveals distinct evolutionary histories for Jewish populations in Yemen and Ethiopia. *Am J Phys Anthropol*, 144(1); 1-10.

[41] Poloni ES, Naciri Y, Bucho R, Niba R, Kervaire B, Excoffier L, et al. (2009). Genetic evidence for complexity in ethnic differentiation and history in East Africa. *Ann Hum Genet*, 73(Pt 6); 582-600.

[42] Kivisild T, Reidla M, Metspalu E, Rosa A, Brehm A, Pennarun E, et al. (2004). Ethiopian mitochondrial DNA heritage: tracking gene flow across and around the gate of tears. *Am J Hum Genet*, 75(5); 752-70.

[43] Al-Abri A, Podgorna E, Rose JI, Pereira L, Mulligan CJ, Silva NM, et al. (2012). Pleistocene-Holocene boundary in Southern Arabia from the perspective of human mtDNA variation. *Am J Phys Anthropol*, 149(2); 291-8.

[44] Brandstatter A, Peterson CT, Irwin JA, Mpoke S, Koech DK, Parson W, et al. (2004). Mitochondrial DNA control region sequences from Nairobi (Kenya): inferring phylogenetic parameters for the establishment of a forensic database. *Int J Legal Med*, 118(5); 294-306.

[45] Boattini A, Castri L, Sarno S, Useli A, Cioffi M, Sazzini M, et al. (2013). mtDNA variation in East Africa unravels the history of Afro-Asiatic groups. *Am J Phys Anthropol*, 150(3); 375-85.

[46] Castri L, Tofanelli S, Garagnani P, Bini C, Fosella X, Pelotti S, et al. (2009). mtDNA variability in two Bantu-speaking populations (Shona and Hutu) from Eastern Africa: implications for peopling and migration patterns in sub-Saharan Africa. *Am J Phys Anthropol*, 140(2); 302-11.

[47] Knight A, Underhill PA, Mortensen HM, Zhivotovsky LA, Lin AA, Henn BM, et al. (2003). African Y chromosome and mtDNA divergence provides insight into the history of click languages. *Curr Biol*, 13(6); 464-73.

[48] Msaidie S, Ducourneau A, Boetsch G, Longepied G, Papa K, Allibert C, et al. (2011). Genetic diversity on the Comoros Islands shows early seafaring as major determinant of human biocultural evolution in the Western Indian Ocean. *Eur J Hum Genet*, 19(1); 89-94.

[49] Tofanelli S, Bertoncini S, Castri L, Luiselli D, Calafell F, Donati G, et al. (2009). On the origins and admixture of Malagasy: new evidence from high-resolution analyses of paternal and maternal lineages. *Mol Biol Evol*, 26(9); 2109-24.

[50] Capredon M, Brucato N, Tonasso L, Choesmel-Cadamuro V, Ricaut FX, Razafindrazaka H, et al. (2013). Tracing Arab-Islamic inheritance in Madagascar: study of the Y chromosome and mitochondrial DNA in the Antemoro. *PLoS One*, 8(11); e80932.

[51] Pereira L, Macaulay V, Torroni A, Scozzari R, Prata MJ, 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(Pt 5); 439-58.

[52] Salas A, Richards M, De la Fe T, Lareu MV, Sobrino B, Sanchez-Diz P, et al. (2002). The making of the African mtDNA landscape. *Am J Hum Genet*, 71(5); 1082-111.

[53] Barbieri C, Guldemann T, Naumann C, Gerlach L, Berthold F, Nakagawa H, et al. (2014). Unraveling the complex maternal history of Southern African Khoisan populations. *Am J Phys Anthropol*, 153(3); 435-48.

[54] Chen YS, Olckers A, Schurr TG, Kogelnik AM, Huoponen K, Wallace DC (2000). mtDNA variation in the South African Kung and Khwe-and their genetic relationships to other African populations. *Am J Hum Genet*, 66(4); 1362-83.

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

[56] Terreros MC, Rowold DJ, Mirabal S, Herrera RJ (2011). Mitochondrial DNA and Y-chromosomal stratification in Iran: relationship between Iran and the Arabian Peninsula. *J Hum Genet*, 56(3); 235-46.

[57] Metspalu M, Kivisild T, Metspalu E, Parik J, Hudjashov G, Kaldma K, et al. (2004). Most of the extant mtDNA boundaries in south and southwest Asia were likely shaped during the initial settlement of Eurasia by anatomically modern humans. *BMC Genet*, 5; 26.

[58] Richards M, Macaulay V, Hickey E, Vega E, Sykes B, Guida V, et al. (2000). Tracing European founder lineages in the Near Eastern mtDNA pool. *Am J Hum Genet*, 67(5); 1251-76.

[59] Behar DM, Metspalu E, Kivisild T, Rosset S, Tzur S, Hadid Y, et al. (2008). Counting the founders: the matrilineal genetic ancestry of the Jewish Diaspora. *PLoS One*, 3(4); e2062.

[60] Gonzalez AM, Karadsheh N, Maca-Meyer N, Flores C, Cabrera VM, Larruga JM (2008). Mitochondrial DNA variation in Jordanians and their genetic relationship to other Middle East populations. *Ann Hum Biol*, 35(2); 212-31.

[61] Scheible M, Alenizi M, Sturk-Andreaggi K, Coble MD, Ismael S, Irwin JA (2011). Mitochondrial DNA control region variation in a Kuwaiti population sample. *Forensic Sci Int Genet*, 5(4); e112-3.

[62] Abu-Amero KK, Gonzalez AM, Larruga JM, Bosley TM, Cabrera VM (2007). Eurasian and African mitochondrial DNA influences in the Saudi Arabian population. *BMC Evol Biol*, 7; 32.

[63] Abu-Amero KK, Larruga JM, Cabrera VM, Gonzalez AM (2008). Mitochondrial DNA structure in the Arabian Peninsula. *BMC Evol Biol*, 8; 45.

[64] Alshamali F, Brandstatter A, Zimmermann B, Parson W (2008). Mitochondrial DNA control region variation in Dubai, United Arab Emirates. *Forensic Sci Int Genet*, 2(1); e9-10.

[65] Cerny V, Mulligan CJ, Ridl J, Zaloudkova M, Edens CM, Hajek M, et al. (2008). Regional differences in the distribution of the sub-Saharan, West Eurasian, and South Asian mtDNA lineages in Yemen. *Am J Phys Anthropol*, 136(2); 128-37.

Table S8: African-American and African populations analysed in the PCA analysis based on mtDNA haplogroup frequencies, and grouped according historical African coastal regions.

ID Pop	Pop Name	ISO code	N	Country	Linguistic affiliation	Ref*
	African American		414			
1	African American (ASW)	ASW	66	USA, Southwest	-	[1]
2	African American (ACB)	ACB	96	Barbados	-	[1]
3	*Aluku*	GUF-A	25	French Guiana	-	[2, 3]
4	*Ndjuka*	GUF-N	97	French Guiana	-	[2, 3]
5	*Paramaka*	GUF-P	19	French Guiana	-	[2, 3]
6	*Saramaka*	GUF-S	42	French Guiana	-	[2, 3]
7	Noir Marron	GUF	183	French Guiana	-	[2, 3]
8	Afro-Brazilian	BRA	39	Brazil	-	[2]
9	Afro-Colombian	COL	30	Colombia	-	[2]
	Senegambia		656			
10	Mandenka	MAN1	22	Senegal	Mande	[4]
11	Mandenka	MAD2	78	Senegal	Mande	[5]
12	Serer	SER	23	Senegal	Atlantic	[6]
13	Woloff	WOL	48	Senegal	Atlantic	[6]
14	Gambian (GWD)	GWD	113	Gambia	West Africa	[1]
15	Baiote-Djola-Banhu-Cassanga-Beafada	BDA	50	Guinea_Bissau	Atlantic	[7]
16	Bijago	BJG	22	Guinea_Bissau	Atlantic	[7]
17	Balanta	BLE	62	Guinea_Bissau	Atlantic	[7]
18	Papel-Manjaco-Mancanha	PBO	77	Guinea_Bissau	Atlantic	[7]
19	Fulani	FUL	77	Guinea_Bissau	Atlantic	[7]
20	Mandinga-Mansonca-Landoma-Jancanca-Sussu	MNK	58	Guinea_Bissau	Atlantic	[7]
21	Nalu	NAJ	26	Guinea_Bissau	Atlantic	[7]
	Sierra Leone		362			
22	Limba	LIM	68	Sierra_Leone	Atlantic	[8]
23	Loko	LOK	32	Sierra_Leone	Mande	[8]
24	Mende (MSL)	MSL	85	Sierra Leone	West Africa	[1]
25	Mende	MEN	55	Sierra_Leone	Mande	[8]
26	Temne	TEM	122	Sierra_Leone	Atlantic	[8]
	Windward Coast		192			
27	Yacouba	YAC	63	Ivory Coast	Mande	[2, 3]
28	Ahizi	AHI	129	Ivory Coast	Kru	[2, 3]
	Gold Coast		429			
29	Akan	AKA	191	Ghana	Kwa	[9]
30	Enchi, AEW	GHE	20	Ghana	Kwa	[10]
31	Enchi, FEWR	GHF	59	Ghana	Kwa	[10]
32	Ho	GHH	87	Ghana	Kwa	[10]
33	Kibi	GHK	51	Ghana	Kwa	[10]
34	Sefwi-Wiawso	GHS	21	Ghana	Kwa	[10]
	Bight of Benin		485			
35	Beninese	BEN-Be	58	Benin	Miscellaneous	[2, 3]
36	Bariba	BEN-Ba	32	Benin	Gur	[2]
37	Yoruba	BEN-Y	32	Benin	Defoid	[2]
38	Fon	FON	92	Benin	Kwa	[2, 3]
39	Yoruba	YOR1	22	Nigeria	Defoid	[4]
40	Yoruba	YOR2	34	Nigeria	Defoid	[11]
41	Yoruba_N (YRI)	YRI	116	Nigeria	Defoid	[1]
42	Esan (ESN)	ESN	99	Nigeria	Edoid	[1]

ID Pop	Pop Name	ISO code	N	Country	Linguistic affiliation	Ref*
	Bight of Biafra		2,714			
43	Afaha Obong	ANA	37	Nigeria	Obolo	[10]
44	Ediene Abak	ANE	26	Nigeria	Obolo	[10]
45	Ikot Obioma	ANI	44	Nigeria	Obolo	[10]
46	Efut, Eniong	EFE	49	Nigeria	Obolo	[10]
47	Efut, Ikot	EFI	48	Nigeria	Obolo	[10]
48	Uwanse	EFO	48	Nigeria	Obolo	[10]
49	Akampka	EKA	17	Nigeria	Ekoid	[10]
50	Calabar	EKC	28	Nigeria	Ekoid	[10]
51	Ikom	EKI	38	Nigeria	Ekoid	[10]
52	Akampka	EKN	50	Nigeria	Ekoid	[10]
53	Afaha Eket	IAE	50	Nigeria	Obolo	[10]
54	Awa	IBA	28	Nigeria	Obolo	[10]
55	Itam	IBI	48	Nigeria	Obolo	[10]
56	Oku	IBO	48	Nigeria	Obolo	[10]
57	Idoma	IDO	37	Nigeria	Idomoid	[12]
58	Edienne Ikono	IEI	49	Nigeria	Obolo, Efik	[10]
59	Igala	IGA	41	Nigeria	Defoid	[12]
60	Calabar	IGC	96	Nigeria	Igbo	[10]
61	Enugu	IGE	54	Nigeria	Igbo	[10]
62	Nenwe	IGN	50	Nigeria	Igbo	[10]
63	Ntan Ibiono	INI	50	Nigeria	Obolo	[10]
64	Nnung Ndem	INN	50	Nigeria	Obolo	[10]
65	Oku-Iboku	IOI	50	Nigeria	Obolo	[10]
66	Obong Itam	ITA	50	Nigeria	Obolo	[10]
67	Ukpom Ette	IUE	50	Nigeria	Obolo	[10]
68	Western Nsit	IWN	36	Nigeria	Obolo	[10]
69	Afaha Okpo	OAO	28	Nigeria	Obolo	[10]
70	Afaha Ukwong	OAU	70	Nigeria	Obolo	[10]
71	Tiv	TIV	51	Nigeria	Tivoid	[12]
72	Bakaka	BAK	50	Cameroon	Narrow Bantu	[13]
73	Bamileke	BAM	48	Cameroon	Narrow Grassfields	[14]
74	Foumban	CAF	107	Cameroon	Narrow Grassfields	[10]
75	Wum	CAW	115	Cameroon	Narrow Grassfields	[10]
76	Bankim	CBT	34	Cameroon	Tikar	[10]
77	Bassa	BAS	47	Cameroon	Narrow Bantu	[13]
78	Ewondo	EWO	53	Cameroon	Narrow Bantu	[14]
79	Fang	FAC	39	Cameroon	Narrow Bantu	[15]
80	Tupuri	TUP	26	Cameroon	Adamawa	[13]
81	Fang	FAG	66	Gabon	Narrow Bantu	[15]
82	Benga	BEN	50	Gabon	Narrow Bantu	[15]
83	Beti	BET	48	Gabon	Narrow Bantu	[12]
84	Duma	DUM	47	Gabon	Narrow Bantu	[15]
85	Ewondo	EWD	25	Gabon	Narrow Bantu	[15]
86	Galoa	GAL	51	Gabon	Narrow Bantu	[15]
87	Eshira	GIS	40	Gabon	Narrow Bantu	[15]
88	Akele	KEL	48	Gabon	Narrow Bantu	[15]
89	Kota	KOT	56	Gabon	Narrow Bantu	[15]
90	Makina	MAK1	45	Gabon	Narrow Bantu	[15]
91	Ngoumba	NGO	44	Gabon	Narrow Bantu	[16]
92	Ngumba	NGU1	88	Gabon	Narrow Bantu	[15]
93	Obamba	OBA	47	Gabon	Narrow Bantu	[15]
94	Orungu	ORU	20	Gabon	Narrow Bantu	[15]
95	Sanga	SAN	30	Gabon	Narrow Bantu	[16]
96	Shake	SHA1	51	Gabon	Narrow Bantu	[15]
97	Tali	TAL	20	Gabon	Adamawa	[13]
98	Tsogo	TSO	64	Gabon	Narrow Bantu	[15]

ID Pop	Pop Name	ISO code	N	Country	Linguistic affiliation	Ref*
	West-Central Africa		555			
99	Ateke	TEK	54	Gabon	Narrow Bantu	[15]
100	Nzebi	NZE	63	Gabon	Narrow Bantu	[15]
101	Ndumu	NDU	39	Gabon	Narrow Bantu	[15]
102	Eviya	EVI	38	Gabon	Narrow Bantu	[15]
103	Punu	PUN	52	Gabon	Narrow Bantu	[15]
104	BatekeN	BAN	53	Congo	Narrow Bantu	[12]
105	BatekeS	BAT	50	Congo	Narrow Bantu	[16]
106	Manyanga	MAN	14	Democratic Republic of the Congo	Narrow Bantu	[17]
107	Nyaneka	NYA1	59	Angola	Narrow Bantu	[18]
108	Ganguela	GAN	20	Angola	Narrow Bantu	[18]
109	Kuvale	KUV	53	Angola	Narrow Bantu	[18]
	Southeast Africa		404			
110	Yao	YAO	10	Mozambique_N and Tanzania	Narrow Bantu	[19, 20]
111	Tonga	TON	28	Mozambique_SE	Narrow Bantu	[19, 20]
112	Shangaan	SHA2	57	Mozambique_SW	Narrow Bantu	[19, 20]
113	Chopi	CHO	39	Mozambique_SE	Narrow Bantu	[19, 20]
114	Chwabo	CHW	23	Mozambique_NE	Narrow Bantu	[19, 20]
115	Lomwe	LOM	20	Mozambique_NE	Narrow Bantu	[19, 20]
116	Makonde	MAK2	19	Mozambique_N and Tanzania	Narrow Bantu	[19, 20]
117	Makhuwa	MAK3	22	Mozambique_NE	Narrow Bantu	[19, 20]
118	Ndau	NDA	23	Mozambique_Central	Narrow Bantu	[19, 20]
119	Nguni	NGU2	11	Mozambique_W and Malawi	Narrow Bantu	[19, 20]
120	Nyungwe	NYU	20	Mozambique_W and Zimbabwe	Narrow Bantu	[19, 20]
121	Nyanja	NYA2	20	Mozambique_N and Malawi	Narrow Bantu	[19, 20]
122	Ronga	RON	42	Mozambique_S and Swaziland	Narrow Bantu	[19, 20]
123	Shona	SHO	18	Mozambique_C and Zimbabwe	Narrow Bantu	[19, 20]
124	Sena	SEN	25	Mozambique_Central	Narrow Bantu	[19, 20]
125	Tswa	TSW	27	Mozambique_SE	Narrow Bantu	[19, 20]

References of the populations included in the Table S8:

[1] The 1000 Genomes Project Consortium (2015). A global reference for human genetic variation. *Nature*, 526(7571); 68-74.

[2] *Present study*.

[3] Brucato N, Cassar O, Tonasso L, Tortevoye P, Migot-Nabias F, Plancoulaine S, et al. (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.

[4] Cann HM, de Toma C, Cazes L, Legrand MF, Morel V, Piouffre L, et al. (2002). A human genome diversity cell line panel. *Science*, 296(5566); 261-2.

[5] Graven L, Passarino G, Semino O, Boursot P, Santachiara-Benerecetti S, Langaney A, et al. (1995). Evolutionary correlation between control region sequence and restriction polymorphisms in the mitochondrial genome of a large Senegalese Mandenka sample. *Mol Biol Evol*, 12(2); 334-45.

[6] Rando JC, Pinto F, Gonzalez AM, Hernandez M, Larruga JM, Cabrera VM, et al. (1998). Mitochondrial DNA analysis of northwest African populations reveals genetic exchanges with European, near-eastern, and sub-Saharan populations. *Ann Hum Genet*, 62(Pt 6); 531-50.

[7] Rosa A, Brehm A, Kivisild T, Metspalu E, Villems R (2004). MtDNA profile of West Africa Guineans: towards a better understanding of the Senegambia region. *Ann Hum Genet*, 68(4); 340-52.

[8] Jackson BA, Wilson JL, Kirbah S, Sidney SS, Rosenberger J, Bassie L, et al. (2005). Mitochondrial DNA genetic diversity among four ethnic groups in Sierra Leone. *Am J Phys Anthropol*, 128(1); 156-63.

[9] Fendt L, Rock A, Zimmermann B, Bodner M, Thye T, Tschentscher F, et al. (2012). MtDNA diversity of Ghana: a forensic and phylogeographic view. *Forensic Sci Int Genet*, 6(2); 244-9.

[10] Veeramah KR, Connell BA, Ansari Pour N, Powell A, Plaster CA, Zeitlyn D, et al. (2010). Little genetic differentiation as assessed by uniparental markers in the presence of substantial language variation in peoples of the Cross River region of Nigeria. *BMC Evol Biol*, 10; 92.

[11] Watson E, Bauer K, Aman R, Weiss G, von Haeseler A, Paabo S (1996). MtDNA sequence diversity in Africa. *Am J Hum Genet*, 59(2); 437-44.

[12] Montano V, Marcari V, Pavanello M, Anyaele O, Comas D, Destro-Bisol G, et al. (2013). The influence of habitats on female mobility in Central and Western Africa inferred from human mitochondrial variation. *BMC Evol Biol*, 13; 24.

[13] Coia V, Destro-Bisol G, Verginelli F, Battaggia C, Boschi I, Cruciani F, et al. (2005). Brief communication: mtDNA variation in North Cameroon: lack of Asian lineages and implications for back migration from Asia to sub-Saharan Africa. *Am J Phys Anthropol*, 128(3); 678-81.

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

[15] Quintana-Murci L, Quach H, Harmant C, Luca F, Massonnet B, Patin E, et al. (2008). Maternal traces of deep common ancestry and asymmetric gene flow between Pygmy hunter-gatherers and Bantu-speaking farmers. *Proc Natl Acad Sci U S A*, 105(5); 1596-601.

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

[17] Schlebusch CM, Lombard M, Soodyall H (2013). MtDNA control region variation affirms diversity and deep sub-structure in populations from southern Africa. *BMC Evol Biol*, 13; 56.

[18] Barbieri C, Vicente M, Oliveira S, Bostoen K, Rocha J, Stoneking M, et al. (2014). Migration and interaction in a contact zone: mtDNA variation among Bantu-speakers in Southern Africa. *PLoS One*, 9(6); e99117.

[19] Pereira L, Macaulay V, Torroni A, Scozzari R, Prata MJ, 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(Pt 5); 439-58.

[20] Salas A, Richards M, De la Fe T, Lareu MV, Sobrino B, Sanchez-Diz P, et al. (2002). The making of the African mtDNA landscape. *Am J Hum Genet*, 71(5); 1082-111.

Table S9: Age estimated and confidence intervals for the mtDNA clades analysed in the phylogenetic tree and calculated by using three calibration methods.

Haplogroup	N	*Rho* Total	*Rho* Synonymous	Maximum Likelihood
L0a1a+200	3	12,484 [7,557;17,545]	8,651 [6,472;10,829]	13,041 [8,944;17,228]
L0a1a1	1	12,261 [5,789;18,968]	14,191 [2,222;26,161]	13,041 [3,670;22,904]
L0a1a2	7	8,137 [6,019;10,282]	8,063 [5,506;10,620]	8,680 [6,201;11,195]
L0a1a3	1	14,067 [7,095;21,302]	17,739 [4,926;30,552]	10,081 [5,352;14,940]
L0a1b1a1	1	5,012 [2,857;7,197]	5,761 [1,435;10,088]	6,270 [3,137;9,465]
L0a2a2a	2	5,715 [3,666;7,790]	4,928 [1,652;8,203]	5,971 [3,649;8,328]
L0a2a2a1	1	6,816 [2,074;11,699]	1,577 [0;4,667]	4,427 [1,899;6,997]
L0d1b2a1	1	5,101 [0;10,561]	1,714 [0;3,729]	3,896 [246;7,635]
L0f1	1	72,328 [52,818;92,601]	92,637 [60,085;125,189]	81,456 [60,987;102,654]
L1b1a	1	11,725 [8,821;14,674]	13,806 [9,415;18,197]	13,744 [10,973;16,555]
L1b1a15	2	10,219 [5,505;15,062]	12,389 [5,068;19,711]	8,574 [3,953;13,322]
L1b1a3	2	5,337 [2,932;7,779]	8,780 [2,238;15,322]	4,841 [3,282;6,416]
L1b1a3a	1	4,687 [1,926;7,497]	7,884 [821;14,947]	3,799 [2,132;5,485]
L1b1a3a1	1	3,021 [479;5,606]	1,314 [0;3,889]	2,583 [577;4,615]
L1b1a4	2	7,742 [4,907;10,626]	8,760 [3,196;14,324]	8,385 [5,092;11,742]
L1b1a7	4	7,892 [3,777;12,109]	4,852 [734;8,969]	8,511 [2,937;14,272]
L1b1a9	3	9,246 [6,074;12,477]	10,512 [4,472;16,552]	9,330 [5,776;12,958]
L1b2	1	18,061 [9,968;26,482]	8,760 [4,554;12,966]	19,457 [9,900;29,461]
L1b2a	4	7,667 [3,755;11,673]	6,570 [1,752;11,388]	9,900 [5,001;14,938]
L1c1a+@198	1	57,244 [40,081;75,154]	52,433 [29,015;75,852]	55,051 [43,148;67,320]
L1c1a2	2	14,966 [7,683;22,531]	27,312 [9,730;44,895]	16,416 [8,851;24,276]
L1c1b	1	28,257 [17,464;39,533]	39,420 [19,141;59,699]	30,296 [18,561;42,580]
L1c1c	4	40,950 [25,735;56,938]	44,676 [19,973;69,379]	38,286 [26,856;50,166]
L1c1d	2	34,110 [22,535;46,183]	49,056 [26,802;71,310]	46,126 [31,906;60,959]
L1c2a1a	1	2,893 [1,279;4,524]	0 [0;0]	3,922 [1,780;6,093]
L1c2a3	1	22,618 [10,628;35,276]	27,594 [3,780;51,408]	22,339 [13,658;31,366]
L1c3a	2	20,447 [15,063;25,968]	22,338 [15,403;29,273]	20,324 [14,891;25,895]
L1c3a1a	1	15,234 [8,983;21,689]	17,082 [7,107;27,057]	14,912 [7,775;22,320]
L1c3a1b	2	15,367 [9,176;21,759]	18,021 [7,206;28,835]	14,912 [9,254;20,739]
L1c3b1a	1	6,358 [2,965;9,822]	6,758 [1,350;12,165]	8,197 [3,826;12,682]
L1c3b1b	1	10,154 [5,577;14,852]	13,140 [4,218;22,062]	11,940 [6,861;17,163]
L1c3b2	1	13,371 [7,089;19,867]	11,826 [2,363;21,289]	15,108 [7,228;23,318]
L1c3c	1	6,549 [793;12,514]	3,942 [0;11,668]	8,448 [281;17,026]
L1c5	2	6,549 [1,390;11,875]	3,942 [0;9,405]	5,828 [0;15,173]
L2a1a	7	11,163 [7,774;14,616]	12,647 [8,045;17,249]	7,646 [5,484;9,836]
L2a1a1	1	6,063 [4,073;8,078]	7,167 [3,726;10,608]	5,805 [3,668;7,972]
L2a1a2	2	8,406 [5,558;11,303]	11,638 [5,342;17,935]	6,539 [4,866;8,229]
L2a1a2a1a	2	6,459 [3,233;9,750]	5,256 [1,687;8,825]	4,530 [666;8,491]
L2a1a2c	1	6,103 [1,558;10,780]	5,256 [0;12,540]	3,866 [1,290;6,485]
L2a1a3a	2	7,507 [2,811;12,339]	5,631 [695;10,568]	4,954 [732;9,292]
L2a1a3c	3	9,586 [3,205;16,208]	10,841 [3,360;18,321]	5,136 [1,838;8,504]
L2a1c	1	19,885 [13,936;26,002]	20,361 [14,098;26,625]	17,080 [12,304;21,970]
L2a1c1	1	11,043 [6,808;15,380]	18,258 [7,716;28,799]	10,193 [6,671;13,787]
L2a1c1a1	1	9,518 [4,811;14,355]	12,614 [2,841;22,388]	7,153 [2,988;11,425]
L2a1c2a	4	8,307 [5,421;11,243]	8,490 [3,175;13,806]	8,759 [4,791;12,822]
L2a1c4a	2	4,905 [2,677;7,164]	8,812 [3,135;14,488]	5,744 [3,020;8,516]
L2a1c4a1	2	4,159 [1,264;7,108]	7,884 [973;14,795]	3,444 [876;6,057]

Haplogroup	N	Rho Total	Rho Synonymous	Maximum Likelihood
L2a1c5	1	3,895 [2,139;5,672]	2,816 [348;5,284]	4,046 [1,700;6,429]
L2a1e	2	15,404 [7,193;23,971]	24,010 [5,229;42,792]	11,135 [5,194;17,277]
L2a1e1	6	8,643 [5,275;12,078]	11,388 [4,850;17,926]	7,522 [3,113;12,051]
L2a1f	6	7,892 [5,434;10,386]	8,070 [5,678;10,461]	6,111 [4,324;7,917]
L2a1f1	2	8,714 [5,985;11,488]	13,026 [7,092;18,959]	5,805 [4,022;7,609]
L2a1f2	2	7,892 [2,899;13,037]	6,307 [126;12,488]	4,893 [0;10,331]
L2a1f3	1	1,719 [34;3,422]	0 [0;0]	588 [0;179,158]
L2a1h	3	15,608 [5,161;26,637]	25,229 [4,268;46,190]	13,606 [4,825;22,811]
L2a1i	2	8,680 [4,427;13,041]	9,527 [1,881;17,172]	10,507 [6,018;15,111]
L2a1i1	3	5,458 [1,574;9,439]	2,867 [57;5,676]	4,772 [0;9,838]
L2a1l	1	12,038 [7,699;16,482]	10,916 [3,167;18,665]	11,387 [8,083;14,752]
L2a1l1	2	11,067 [6,782;15,456]	9,198 [3,659;14,737]	10,319 [6,914;13,789]
L2a1l1a	5	6,549 [3,697;9,451]	5,068 [1,408;8,729]	7,646 [3,586;11,806]
L2a1m	2	7,219 [3,615;10,903]	12,483 [3,287;21,679]	8,884 [4,913;12,949]
L2a1m1a	1	6,281 [2,148;10,522]	7,884 [973;14,795]	6,050 [2,031;10,171]
L2a1n	2	7,555 [3,810;11,385]	2,957 [0;6,302]	6,968 [2,107;11,977]
L2b1a	1	12,780 [9,338;16,286]	15,433 [7,487;23,378]	17,210 [13,177;21,323]
L2b1a2	2	17,458 [11,794;23,282]	21,502 [11,974;31,030]	15,014 [10,657;19,469]
L2b1a3	2	4,015 [2,538;5,506]	2,867 [646;5,088]	4,530 [2,527;6,559]
L2b1b	1	11,395 [6,030;16,922]	11,263 [3,003;19,523]	14,372 [7,355;21,654]
L2b2	4	20,036 [13,684;26,581]	23,089 [12,218;33,960]	22,936 [15,358;30,774]
L2b3	1	-	-	-
L2b3a+207	3	8,383 [4,207;12,664]	5,017 [803;9,231]	8,387 [3,601;13,312]
L2b3b	1	11,709 [5,520;18,116]	6,307 [126;12,488]	16,367 [5,652;27,686]
L2c	4	17,254 [13,966;20,595]	15,837 [12,189;19,484]	20,549 [16,450;24,725]
L2c1	2	16,800 [10,300;23,514]	15,768 [4,509;27,027]	15,142 [9,801;20,633]
L2c2	5	13,623 [9,819;17,505]	17,440 [9,818;25,063]	14,500 [10,156;18,945]
L2c2a	1	11,434 [4,887;18,226]	6,307 [954;11,660]	9,693 [4,276;15,284]
L2c2b1b	1	6,995 [1,547;12,629]	5,256 [0;12,540]	3,986 [482;7,572]
L2c3	4	13,370 [9,189;17,646]	12,389 [5,767;19,012]	14,629 [8,936;20,493]
L2c4	1	10,953 [6,215;15,818]	18,725 [7,297;30,152]	19,824 [15,487;24,250]
L2d+16129	1	18,727 [10,830;26,931]	24,835 [7,960;41,709]	16,367 [9,152;23,851]
L3b1a	5	13,244 [10,897;15,621]	13,761 [9,838;17,685]	16,308 [12,735;19,947]
L3b1a1a	1	3,794 [730;6,921]	1,819 [0;3,878]	13,531 [8,205;19,011]
L3b1a5	1	11,709 [7,439;16,081]	14,191 [6,312;22,071]	14,404 [10,026;18,883]
L3b1a5a	1	6,103 [1,558;10,780]	2,628 [0;7,779]	9,228 [2,959;15,732]
L3b1a6	2	6,549 [2,357;10,851]	6,899 [1,104;12,693]	6,765 [3,245;10,363]
L3b1a7a	1	6,103 [1,558;10,780]	7,884 [0;16,806]	6,905 [2,181;11,769]
L3b1a8	1	9,699 [3,895;15,703]	10,512 [210;20,814]	11,295 [5,627;17,146]
L3d1-6	1	28,388 [22,594;34,317]	0 [0;0]	33,512 [26,771;40,420]
L3d1a1a	1	3,632 [1,348;5,952]	6,307 [954;11,660]	3,996 [1,197;6,846]
L3d1a1b	2	3,895 [770;7,085]	1,971 [0;5,834]	4,202 [83;8,433]
L3d1a2	1	15,233 [7,404;23,388]	13,140 [1,622;24,658]	19,199 [10,189;28,608]
L3d1b3	3	12,943 [5,783;20,388]	23,652 [8,199;39,105]	12,879 [0;33,546]
L3d1b3a	4	11,359 [4,957;17,995]	26,519 [8,365;44,673]	12,879 [6,612;19,363]
L3d1c1	4	6,549 [4,021;9,116]	5,913 [1,642;10,184]	8,945 [4,835;13,154]
L3d1d	1	11,526 [4,263;19,090]	13,140 [0;27,246]	18,528 [8,993;28,518]

Haplogroup	N	Rho Total	Rho Synonymous	Maximum Likelihood
L3d2	1	13,185 [9,000;17,466]	12,614 [6,099;19,130]	19,124 [12,224;26,256]
L3d3a1	1	4,015 [395;7,722]	1,433 [0;3,420]	4,477 [219;8,854]
L3d5a	2	11,002 [3,373;18,968]	13,515 [0;29,434]	10,579 [533;21,227]
L3e1	2	17,384 [13,317;21,534]	16,237 [10,308;22,167]	18,380 [13,485;23,392]
L3e1a	1	18,025 [9,786;26,603]	20,013 [8,008;32,018]	16,972 [11,266;22,842]
L3e1e	3	7,402 [3,198;11,714]	5,734 [0;12,016]	10,508 [4,460;16,768]
L3e1f	2	-	-	-
L3e1f1a	1	4,334 [528;8,236]	0 [0;0]	4,546 [0;11,438]
L3e1g	1	6,103 [1,558;10,780]	2,628 [0;7,779]	5,236 [562;10,052]
L3e2a	3	14,083 [8,498;19,835]	16,318 [6,722;25,914]	15,427 [10,465;20,517]
L3e2a1a	1	4,554 [1,167;8,017]	7,884 [158;15,610]	4,891 [1,016;8,863]
L3e2a1b	3	8,567 [4,086;13,169]	6,570 [3,043;10,097]	8,591 [2,746;14,643]
L3e2a1b1	5	6,740 [1,952;11,671]	4,505 [1,383;7,627]	6,138 [2,632;9,722]
L3e2b	4	10,519 [7,531;13,557]	12,264 [7,457;17,071]	12,662 [9,574;15,803]
L3e2b+152	5	5,942 [4,086;7,819]	9,795 [5,183;14,407]	7,255 [4,938;9,605]
L3e2b1a1	3	9,699 [4,377;15,189]	13,140 [0;26,328]	8,874 [5,308;12,515]
L3e2b3	3	4,838 [1,812;7,923]	4,505 [90;8,920]	5,305 [2,216;8,456]
L3e2b4	1	3,457 [0;7,713]	10,512 [0;23,129]	4,064 [80;8,154]
L3e2b5	1	10,610 [3,182;18,361]	7,884 [0;18,811]	6,068 [2,831;9,372]
L3e3b	6	7,339 [4,966;9,746]	7,278 [4,710;9,845]	8,098 [4,684;11,582]
L3e3b3	1	9,246 [3,112;15,603]	13,797 [2,208;25,386]	6,905 [3,114;10,785]
L3e5	1	11,352 [8,383;14,370]	10,441 [4,232;16,650]	12,662 [8,603;16,812]
L3e5a	2	9,617 [5,491;13,841]	5,734 [1,291;10,176]	9,938 [5,941;14,028]
L3f1b	2	13,408 [9,096;17,820]	13,744 [6,223;21,264]	14,112 [10,440;17,857]
L3f1b+150	1	13,989 [8,633;19,500]	14,016 [5,963;22,069]	13,024 [9,233;16,892]
L3f1b1	2	10,976 [3,491;18,785]	17,082 [0;34,708]	8,168 [3,944;12,501]
L3f1b1a	4	4,444 [2,613;6,298]	4,271 [1,949;6,592]	5,098 [2,945;7,280]
L3f1b1a1	1	857 [0;2,052]	1,314 [0;3,889]	1,073 [0;2,798]
L3f1b3	1	9,246 [4,001;14,654]	9,855 [0;20,076]	9,086 [4,973;13,299]
L3f1b4a	1	7,443 [3,283;11,709]	14,454 [3,835;25,073]	7,746 [3,667;11,927]
L3f1b4a1	1	4,334 [528;8,236]	5,256 [0;12,540]	5,029 [624;9,560]
L3f1b4c	1	2,585 [0;6,444]	0 [0;0]	4,271 [0;12,703]
L3h1b2	1	18,299 [11,780;25,027]	13,797 [4,737;22,857]	25,022 [15,041;35,440]
L4b1a *	1	17,114 [10,071;24,409]	24,928 [9,286;40,570]	16,914 [10,581;23,450]

* L4b1a haplogroup ages estimated in Fernandes et al. (2015) .

C. *Supplementary material of genome-wide SNP data*

Table S10: List of populations from different worldwide geographical regions included in PCA analysis.

Figure S4: Plot of residual fit between each pair of African-American and African populations analysed in the maximum likelihood tree (Figure 66).

Table S11: Three-population test calculated for Noir Marron communities. The $f3$-statistics and Z-score values show not admixture events with European and Native American populations from the 1000 Genomes Project (1000 Genomes Project Consortium, 2015).

Table S12: Three-population test calculated for African American from the U.S. (ASW), Barbados (ACB), Colombia, and Brazil. The negative $f3$-statistics and Z-score values show the robustness of the signal of admixture with European and Native American populations from the 1000 Genomes Project (1000 Genomes Project Consortium, 2015).

Table S13: Mean individual ancestry proportions and standard deviation of African American (AAM), African (AFR), Native American (NAM), and European (EUR) ancestries estimated by using ADMIXTURE analysis at K-groups= 3 (Figure 69).

****

Table S10: List of populations from different worldwide geographical regions included in PCA analysis.

Order	Id Population	Continent	Region	Dataset	Abbreviation	N
1	Noir Maroon	America	South America	*Present study*	*noi*	71
2	AfroBrazilian	America	South America	*Present study*	*AfB*	16
3	AfroColombian	America	South America	*Present study*	*AfC*	20
4	Fon	Africa	West Africa	*Present study*	*fon*	19
5	Bariba	Africa	West Africa	*Present study*	*bar*	25
6	Yoruba_Benin	Africa	West Africa	*Present study*	*yoB*	24
7	Ahizi	Africa	West Africa	*Present study*	*ahz*	20
8	Yacouba	Africa	West Africa	*Present study*	*yac*	17
9	Bwa	Africa	West Africa	*Present study*	*bwa*	20
10	San	Africa	South Africa	HGDP	*san*	5
11	Bantu_South Africa	Africa	South Africa	HGDP	*bSA*	8
12	Mbuti_Pygmy	Africa	Central Africa	HGDP	*mbP*	13
13	Biaka_Pygmy	Africa	Central Africa	HGDP	*biP*	21
14	Mandenka	Africa	West Africa	HGDP	*mdk*	22
15	Yoruba_Nigeria	Africa	West Africa	HGDP	*yoA*	21
16	Bantu_Kenya	Africa	East Africa	HGDP	*bKe*	11
17	Mozabite	Africa	North Africa	HGDP	*moz*	29
18	Bedouin	Asia	Middle East	HGDP	*bed*	46
19	Druze	Asia	Middle East	HGDP	*drz*	42
20	Palestinian	Asia	Middle East	HGDP	*pal*	46
21	North Italian	Europe	South Europe	HGDP	*itN*	12
22	Tuscan	Europe	South Europe	HGDP	*tus*	8
23	Sardinian	Europe	South Europe	HGDP	*sar*	28
24	Basque	Europe	South Europe	HGDP	*bas*	24
25	French	Europe	West Europe	HGDP	*fre*	28
26	Orcadian	Europe	West Europe	HGDP	*orc*	15
27	NW Russian	Europe	East Europe	HGDP	*ruN*	25
28	Adygei	Europe	North Caucasus	HGDP	*ady*	17
29	Uygur	Asia	Central Asia	HGDP	*uyg*	10
30	Hazara	Asia	Central Asia	HGDP	*haz*	22
31	Burusho	Asia	Central Asia	HGDP	*bur*	25
32	Balochi	Asia	Central Asia	HGDP	*bao*	24
33	Brahui	Asia	Central Asia	HGDP	*bra*	25
34	Kalash	Asia	Central Asia	HGDP	*kal*	23
35	Makrani	Asia	Central Asia	HGDP	*mak*	25
36	Pathan	Asia	Central Asia	HGDP	*pat*	22
37	Sindhi	Asia	Central Asia	HGDP	*sin*	24
38	Cambodian	Asia	South East Asia	HGDP	*cam*	10
39	Dai	Asia	East Asia	HGDP	*dai*	10
40	Daur	Asia	East Asia	HGDP	*dau*	9
41	Han	Asia	East Asia	HGDP	*han*	34
42	Han-NChina	Asia	East Asia	HGDP	*haC*	10
43	Hezhen	Asia	East Asia	HGDP	*hez*	8
44	Lahu	Asia	East Asia	HGDP	*lah*	8
45	Miao	Asia	East Asia	HGDP	*mia*	10
46	Naxi	Asia	East Asia	HGDP	*nax*	8
47	Oroqen	Asia	East Asia	HGDP	*oro*	9
48	She	Asia	East Asia	HGDP	*she*	10
49	Tu	Asia	East Asia	HGDP	*tu*	10
50	Tujia	Asia	East Asia	HGDP	*tuj*	10
51	Xibo	Asia	East Asia	HGDP	*xib*	9
52	Yi	Asia	East Asia	HGDP	*yi*	10
53	Japanese	Asia	East Asia	HGDP	*jap*	28
54	Mongolians	Asia	East Asia	HGDP	*mon*	10
55	Yakut	Asia	Siberia	HGDP	*yak*	25
56	Maya	America	North America	HGDP	*may*	21
57	Pima	America	North America	HGDP	*pim*	14
58	Colombian	America	South America	HGDP	*col*	7
59	Karitiana	America	South America	HGDP	*kar*	14
60	Surui	America	South America	HGDP	*sur*	8
61	Melanesian	Oceania	Oceania	HGDP	*mel*	10
62	Papuan	Oceania	Oceania	HGDP	*pap*	17
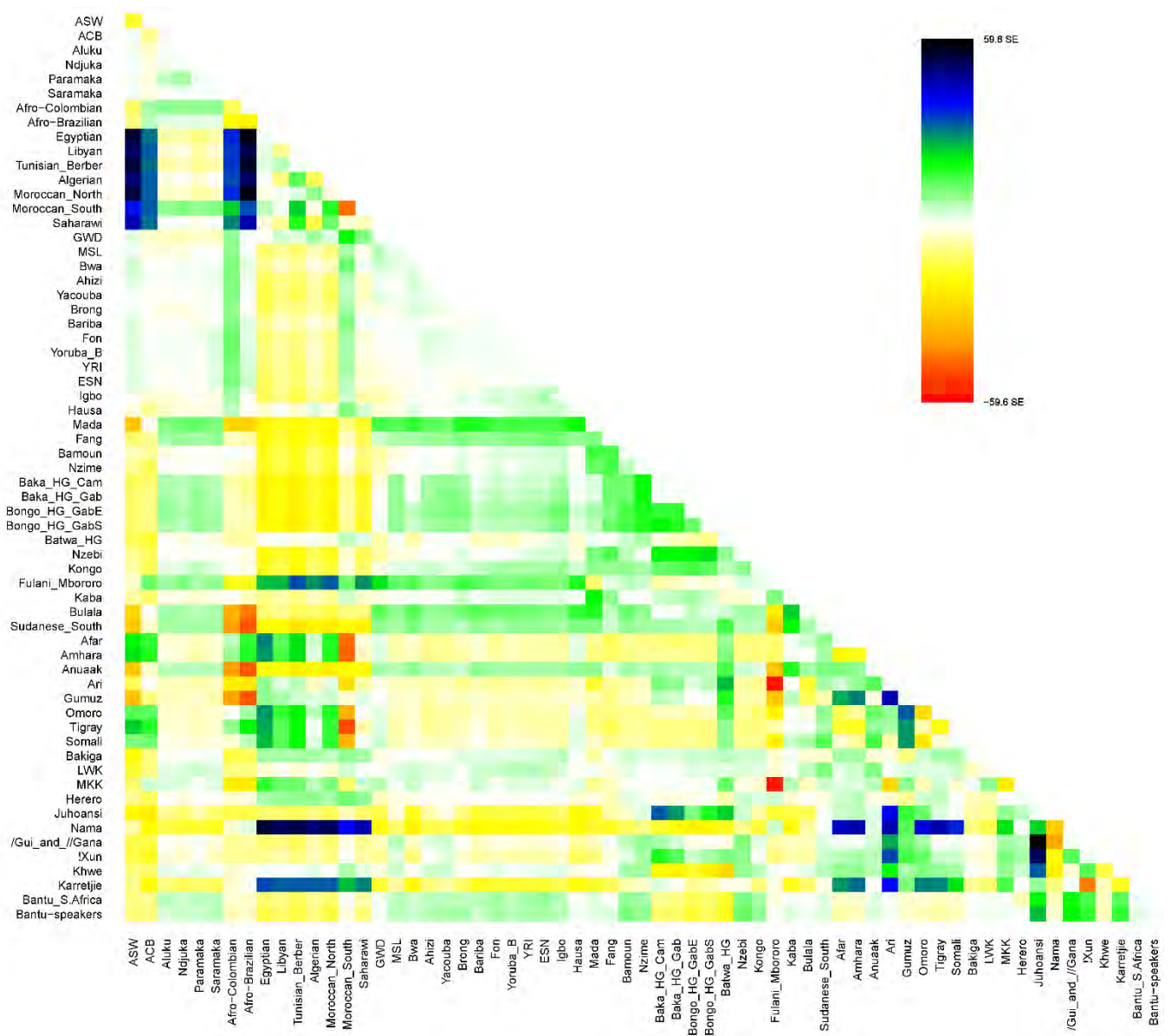
Figure S4: Plot of residual fit between each pair of African-American and African populations analysed in the maximum likelihood tree (Figure 66). Each cell represents the residual covariance between each pair of populations the average standard error across all pairs this scaled residual. Colours are described in the palette covariance on the right. Residuals above zero represent populations that are more closely related to each other in the data than in the best-fit tree, and are therefore candidates for admixture events.

Table S11: Three-population test calculated for Noir Marron communities. The *ƒ3*-statistics and Z-score values show not admixture events with European and Native American populations from the 1000 Genomes Project (1000 Genomes Project Consortium, 2015).

ƒ3-stat (AfrAme;European,African)		Aluku		Ndjuka		Paramaka		Saramaka	
European	African	ƒ3	Z-score	ƒ3	Z-score	ƒ3	Z-score	ƒ3	Z-score
GBR	GWD	0.0020	43.0363	0.0015	33.6832	0.0036	32.4423	0.0019	37.7926
GBR	MSL	0.0015	33.3470	0.0010	23.0279	0.0031	28.2439	0.0014	29.9837
GBR	Bwa	0.0016	30.0187	0.0010	18.7617	0.0031	26.5568	0.0014	25.4526
GBR	Ahizi	0.0013	25.5767	0.0007	16.1918	0.0029	25.2714	0.0012	22.6978
GBR	Yacouba	0.0014	26.3986	0.0008	16.4768	0.0028	24.0983	0.0012	23.4481
GBR	Fon	0.0013	26.3592	0.0008	16.5222	0.0029	25.7610	0.0011	22.7496
GBR	Bariba	0.0015	30.1333	0.0010	20.5551	0.0031	27.2259	0.0013	25.6474
GBR	Yoruba_B	0.0014	27.4087	0.0008	17.5551	0.0030	26.4179	0.0012	24.5194
GBR	YRI	0.0015	31.6947	0.0009	20.6992	0.0030	27.1190	0.0013	27.3598
GBR	ESN	0.0015	31.9303	0.0009	22.0846	0.0030	27.6661	0.0013	27.4032
GBR	LWK	0.0025	48.7209	0.0019	42.0746	0.0041	35.5139	0.0023	45.5454
GBR	Bantu_S.Africa	0.0020	39.0348	0.0014	32.2427	0.0036	31.3523	0.0018	36.0526
IBS	GWD	0.0021	43.7054	0.0015	34.4794	0.0036	32.9638	0.0020	38.5500
IBS	MSL	0.0015	33.6593	0.0010	23.0400	0.0031	28.3927	0.0014	29.9914
IBS	Bwa	0.0016	30.4270	0.0010	19.3158	0.0032	26.9374	0.0014	26.1069
IBS	Ahizi	0.0013	25.3633	0.0007	16.1588	0.0029	25.3661	0.0012	22.5701
IBS	Yacouba	0.0014	26.5870	0.0008	16.4431	0.0028	24.1493	0.0012	23.6610
IBS	Fon	0.0014	26.6341	0.0008	16.6745	0.0029	25.8072	0.0011	22.5164
IBS	Bariba	0.0016	30.4172	0.0010	20.8519	0.0031	27.5373	0.0013	25.6166
IBS	Yoruba_B	0.0014	27.6587	0.0008	18.0113	0.0030	26.6993	0.0012	24.9550
IBS	YRI	0.0015	31.9489	0.0009	21.1243	0.0030	27.3935	0.0013	27.8003
IBS	ESN	0.0015	32.0102	0.0009	22.3827	0.0031	27.9191	0.0013	27.6524
IBS	LWK	0.0025	49.2431	0.0020	42.8589	0.0041	35.7492	0.0023	46.4931
IBS	Bantu_S.Africa	0.0020	39.2906	0.0014	32.2828	0.0036	31.4466	0.0018	36.2694
TSI	GWD	0.0021	43.4532	0.0015	33.9018	0.0036	32.7917	0.0020	38.5456
TSI	MSL	0.0015	33.7630	0.0010	22.9242	0.0031	28.2751	0.0014	30.2806
TSI	Bwa	0.0016	30.6554	0.0010	18.8216	0.0031	26.7040	0.0014	25.6672
TSI	Ahizi	0.0013	25.7413	0.0007	16.2179	0.0029	25.2887	0.0012	23.0907
TSI	Yacouba	0.0014	27.0281	0.0008	16.3232	0.0028	24.0438	0.0012	23.7771
TSI	Fon	0.0014	27.1337	0.0008	16.7036	0.0029	25.8030	0.0011	22.9518
TSI	Bariba	0.0016	30.8209	0.0010	20.5192	0.0031	27.4753	0.0014	26.0445
TSI	Yoruba_B	0.0014	27.6198	0.0008	17.6107	0.0030	26.4726	0.0012	24.6823
TSI	YRI	0.0015	32.3367	0.0009	20.8518	0.0030	27.3217	0.0013	27.7749
TSI	ESN	0.0015	32.2479	0.0009	22.0010	0.0030	27.8042	0.0013	27.6794
TSI	LWK	0.0025	49.7086	0.0020	43.2052	0.0041	35.8416	0.0024	46.9537
TSI	Bantu_S.Africa	0.0020	39.6610	0.0014	32.4624	0.0036	31.3355	0.0018	36.7265
	Mean ƒ3-stat ± S.D.	0.0016	± 0.0003	0.0011	± 0.0004	0.0032	± 0.0004	0.0015	± 0.0004

ƒ3-stat (AfrAme;NativeAme,African)		Aluku		Ndjuka		Paramaka		Saramaka	
Native American	African	ƒ3	Z-score	ƒ3	Z-score	ƒ3	Z-score	ƒ3	Z-score
PEL	GWD	0.0019	37.2305	0.0014	30.7767	0.0035	32.0951	0.0017	31.5473
PEL	MSL	0.0015	31.1465	0.0010	22.8814	0.0030	28.8127	0.0013	25.8156
PEL	Bwa	0.0015	27.0681	0.0010	18.6648	0.0031	26.0908	0.0012	21.3349
PEL	Ahizi	0.0013	22.3232	0.0008	15.4973	0.0029	25.2181	0.0011	18.4047
PEL	Yacouba	0.0014	23.5881	0.0008	15.9526	0.0028	23.9612	0.0011	19.2866
PEL	Fon	0.0014	23.6822	0.0008	16.6391	0.0029	25.5607	0.0010	18.2697
PEL	Bariba	0.0015	27.4909	0.0010	19.9226	0.0030	27.3451	0.0012	21.5618
PEL	Yoruba_B	0.0014	24.4941	0.0008	17.5663	0.0029	26.7770	0.0011	20.0760
PEL	YRI	0.0015	28.7960	0.0009	21.6899	0.0030	27.7634	0.0011	23.7233
PEL	ESN	0.0015	29.4817	0.0010	22.9935	0.0031	28.0332	0.0012	23.0666
PEL	LWK	0.0023	41.5689	0.0017	38.4919	0.0038	34.0719	0.0020	36.8471
PEL	Bantu_S.Africa	0.0019	34.9240	0.0014	29.3484	0.0035	31.2280	0.0016	29.7763
	Mean ƒ3-stat ± S.D.	0.0016	± 0.0003	0.0010	± 0.0003	0.0031	± 0.0003	0.0013	± 0.0003

Table S12: Three-population test calculated for African American from the U.S. (ASW), Barbados (ACB), Colombia, and Brazil. The negative $f3$-statistics and Z-score values show the robustness of the signal of admixture with European and Native American populations from the 1000 Genomes Project (1000 Genomes Project Consortium, 2015).

$f3$-stat (AfrAme;European,African)		ASW		ACB		Afro-Brazilian		Afro-Colombian	
European	African	$f3$	Z-score	$f3$	Z-score	$f3$	Z-score	$f3$	Z-score
GBR	GWD	-0.0024	-101.2930	-0.0012	-52.0331	-0.0025	-57.3840	-0.0016	-39.7961
GBR	MSL	-0.0028	-123.0140	-0.0016	-80.3732	-0.0028	-68.8244	-0.0019	-52.8988
GBR	Bwa	-0.0027	-87.8769	-0.0016	-55.2751	-0.0028	-58.4726	-0.0019	-44.1308
GBR	Ahizi	-0.0028	-96.3049	-0.0017	-62.2022	-0.0029	-66.1339	-0.0021	-51.8970
GBR	Yacouba	-0.0028	-102.9420	-0.0017	-60.9063	-0.0029	-63.9994	-0.0021	-49.3526
GBR	Fon	-0.0028	-107.1220	-0.0018	-74.1389	-0.0030	-69.7506	-0.0021	-54.4996
GBR	Bariba	-0.0027	-101.5780	-0.0016	-65.4074	-0.0029	-69.8399	-0.0020	-49.2629
GBR	Yoruba_B	-0.0028	-109.6270	-0.0017	-74.4049	-0.0030	-72.1923	-0.0021	-53.0172
GBR	YRI	-0.0028	-133.3480	-0.0017	-96.2047	-0.0030	-77.1022	-0.0020	-56.0350
GBR	ESN	-0.0028	-131.3270	-0.0017	-94.2998	-0.0030	-79.0660	-0.0020	-54.9398
GBR	LWK	-0.0021	-88.7706	-0.0007	-31.7597	-0.0025	-59.9572	-0.0011	-27.9950
GBR	Bantu_S.Africa	-0.0024	-94.3819	-0.0011	-45.7313	-0.0029	-72.2002	-0.0015	-37.8224
IBS	GWD	-0.0023	-95.8825	-0.0011	-48.3291	-0.0024	-55.3191	-0.0014	-37.2928
IBS	MSL	-0.0026	-121.8930	-0.0016	-80.9161	-0.0027	-68.1282	-0.0019	-51.1155
IBS	Bwa	-0.0025	-85.2394	-0.0015	-54.5193	-0.0027	-57.3311	-0.0018	-41.8964
IBS	Ahizi	-0.0027	-96.0147	-0.0017	-61.3558	-0.0028	-65.4829	-0.0020	-50.9846
IBS	Yacouba	-0.0027	-101.0630	-0.0016	-60.8793	-0.0029	-62.6209	-0.0020	-47.6295
IBS	Fon	-0.0027	-103.9220	-0.0017	-71.0289	-0.0029	-69.0183	-0.0020	-52.1512
IBS	Bariba	-0.0025	-97.5835	-0.0015	-62.4013	-0.0028	-68.2725	-0.0019	-48.2241
IBS	Yoruba_B	-0.0027	-108.5380	-0.0017	-74.7802	-0.0029	-72.4312	-0.0020	-51.7145
IBS	YRI	-0.0027	-134.4440	-0.0017	-96.9901	-0.0029	-76.4063	-0.0019	-54.7289
IBS	ESN	-0.0027	-131.0720	-0.0017	-94.9818	-0.0029	-77.8095	-0.0019	-53.1647
IBS	LWK	-0.0019	-85.3005	-0.0006	-28.9490	-0.0024	-58.4878	-0.0010	-25.8531
IBS	Bantu_S.Africa	-0.0023	-93.7295	-0.0011	-44.3079	-0.0029	-71.5575	-0.0015	-36.5806
TSI	GWD	-0.0023	-95.4958	-0.0011	-48.1114	-0.0024	-55.3226	-0.0014	-36.9791
TSI	MSL	-0.0026	-123.3670	-0.0016	-80.8774	-0.0027	-67.3789	-0.0019	-50.7119
TSI	Bwa	-0.0025	-85.3769	-0.0015	-53.8069	-0.0028	-56.7490	-0.0018	-42.7177
TSI	Ahizi	-0.0026	-93.7285	-0.0017	-60.1206	-0.0028	-65.3311	-0.0020	-49.3873
TSI	Yacouba	-0.0027	-101.6540	-0.0016	-60.3181	-0.0029	-62.4999	-0.0020	-47.4195
TSI	Fon	-0.0026	-103.0530	-0.0017	-70.1490	-0.0029	-68.7815	-0.0020	-50.6465
TSI	Bariba	-0.0025	-97.9317	-0.0015	-62.0178	-0.0028	-66.8603	-0.0019	-47.4871
TSI	Yoruba_B	-0.0027	-107.8430	-0.0017	-73.1122	-0.0030	-71.6396	-0.0020	-50.6947
TSI	YRI	-0.0027	-135.2320	-0.0017	-95.5223	-0.0029	-75.3622	-0.0019	-54.0670
TSI	ESN	-0.0027	-132.0320	-0.0017	-93.2117	-0.0030	-77.2020	-0.0019	-52.4637
TSI	LWK	-0.0019	-80.4798	-0.0006	-26.9093	-0.0024	-56.8407	-0.0010	-24.3083
TSI	Bantu_S.Africa	-0.0023	-91.4596	-0.0011	-43.3425	-0.0029	-71.7031	-0.0014	-35.9604
	Mean $f3$-stat ± S.D.	**-0.0026**	**± 0.0002**	**-0.0015**	**± 0.0003**	**-0.0028**	**± 0.0002**	**-0.0018**	**± 0.0003**

$f3$-stat (AfrAme;NativeAme,African)		ASW		ACB		Afro-Brazilian		Afro-Colombian	
Native American	African	$f3$	Z-score	$f3$	Z-score	$f3$	Z-score	$f3$	Z-score
PEL	GWD	-0.0019	-82.3903	-0.0007	-30.9320	-0.0020	-39.5840	-0.0029	-59.2309
PEL	MSL	-0.0022	-91.4601	-0.0010	-48.2809	-0.0022	-44.6943	-0.0031	-65.9901
PEL	Bwa	-0.0021	-63.4344	-0.0010	-31.2226	-0.0023	-39.7153	-0.0032	-57.2767
PEL	Ahizi	-0.0022	-71.2173	-0.0011	-38.3628	-0.0024	-42.5206	-0.0033	-62.8414
PEL	Yacouba	-0.0022	-65.0234	-0.0011	-33.9567	-0.0024	-42.2862	-0.0033	-61.0965
PEL	Fon	-0.0022	-76.4036	-0.0011	-44.4963	-0.0024	-47.7313	-0.0033	-65.1103
PEL	Bariba	-0.0021	-75.1697	-0.0010	-39.1350	-0.0023	-45.8644	-0.0032	-63.7588
PEL	Yoruba_B	-0.0022	-77.8068	-0.0011	-43.3226	-0.0024	-46.6386	-0.0033	-62.9859
PEL	YRI	-0.0022	-100.0400	-0.0011	-64.9376	-0.0024	-49.3839	-0.0032	-66.9469
PEL	ESN	-0.0022	-97.7782	-0.0011	-60.2868	-0.0024	-50.1621	-0.0032	-66.4943
PEL	LWK	-0.0017	-68.2339	-0.0003	-14.1614	-0.0021	-43.1024	-0.0026	-50.6055
PEL	Bantu_S.Africa	-0.0019	-68.2601	-0.0006	-23.0962	-0.0025	-49.8939	-0.0028	-56.0340
	Mean $f3$-stat ± S.D.	**-0.0021**	**± 0.0002**	**-0.0009**	**± 0.0003**	**-0.0023**	**± 0.0001**	**-0.0031**	**± 0.0002**

Table S13: Mean individual ancestry proportions and standard deviation of African American (AAM), African (AFR), Native American (NAM), and European (EUR) ancestries estimated by using ADMIXTURE analysis at K-groups= 3 (Figure 69).

Population	Country	Dataset	Origin	N	AFRICAN	EUROPEAN	NATIVE A.
Aluku	F. Guiana	*Present study*	AAM	19	*0.989 ±0.013*	*0.007 ±0.011*	*0.004 ±0.006*
Ndjuka	F. Guiana	*Present study*	AAM	21	*0.987 ±0.042*	*0.011 ±0.041*	*0.002 ±0.005*
Paramaka	F. Guiana	*Present study*	AAM	5	*0.994 ±0.014*	*0.005 ±0.012*	*0.001 ±0.002*
Saramaka	F. Guiana	*Present study*	AAM	17	*0.990 ±0.024*	*0.006 ±0.023*	*0.004 ±0.005*
All Noir Marron	F. Guiana	*Present study*	AAM	62	*0.989 ±0.028*	*0.008 ±0.027*	*0.003 ±0.005*
Afro-Brazilian	Brazil	*Present study*	AAM	16	*0.724 ±0.177*	*0.216 ±0.151*	*0.060 ±0.040*
Afro-Colombian	Colombia	*Present study*	AAM	20	*0.772 ±0.115*	*0.103 ±0.060*	*0.125 ±0.070*
Ahizi	Ivory Coast	*Present study*	AFR	16	*1.000 ±0.000*	*0.000 ±0.000*	*0.000 ±0.000*
Yacouba	Ivory Coast	*Present study*	AFR	16	*1.000 ±0.000*	*0.000 ±0.000*	*0.000 ±0.000*
Bwa	Mali	*Present study*	AFR	14	*1.000 ±0.000*	*0.000 ±0.000*	*0.000 ±0.000*
Bariba	Benin	*Present study*	AFR	24	*0.999 ±0.003*	*0.000 ±0.002*	*0.000 ±0.002*
Fon	Benin	*Present study*	AFR	19	*1.000 ±0.000*	*0.000 ±0.000*	*0.000 ±0.000*
Yoruba_B	Benin	*Present study*	AFR	23	*1.000 ±0.000*	*0.000 ±0.000*	*0.000 ±0.000*
Mandenka	Senegal	HGDP	AFR	21	*0.974 ±0.011*	*0.014 ±0.010*	*0.012 ±0.007*
Yoruba_N	Nigeria	HGDP	AFR	21	*0.999 ±0.002*	*0.000 ±0.000*	*0.001 ±0.002*
Bantu_K	Kenya	HGDP	AFR	10	*0.959 ±0.013*	*0.024 ±0.011*	*0.018 ±0.007*
Bantu_S.A.	S. Africa	HGDP	AFR	8	*0.912 ±0.039*	*0.061 ±0.039*	*0.027 ±0.003*
Pima	Mexico	HGDP	NAM	8	*0.000 ±0.000*	*0.000 ±0.000*	*1.000 ±0.000*
Maya	Mexico	HGDP	NAM	21	*0.000 ±0.002*	*0.060 ±0.093*	*0.940 ±0.093*
Colombian	Colombia	HGDP	NAM	5	*0.003 ±0.006*	*0.015 ±0.034*	*0.982 ±0.039*
Karitiana	Brazil	HGDP	NAM	2	*0.000 ±0.000*	*0.000 ±0.000*	*1.000 ±0.000*
Surui	Brazil	HGDP	NAM	1	*0.000 ±0.000*	*0.000 ±0.000*	*1.000 ±0.000*
Basque	France	HGDP	EUR	24	*0.000 ±0.000*	*1.000 ±0.000*	*0.000 ±0.000*
French	France	HGDP	EUR	28	*0.000 ±0.000*	*0.965 ±0.010*	*0.035 ±0.010*

VIII. PUBLICATIONS

Fortes-Lima., C, and Dugoujon J.M. (201x). Revisiting genetic ancestry in African American communities from French Guiana, Colombia and Brazil. *Trends in Biological Anthropology series*. Proceeding paper accepted in Novembre 2014.

Fortes-Lima., C, Brucato N., Croze M., Bellis G., Schiavinato S., Massougbodji A., Migot-Nabias F., and Dugoujon J.M. (2015). Genetic population study of Y chromosome markers in Benin and Ivory Coast ethnic groups. *Forensic Sci Int Genet*. 9:232-237. doi: 10.1016/j.fsigen.2015.07.021. PMID: 26275614.

Fortes-Lima., C, Hernandez C., Dugoujon J.M., Novelletto A., Rodriguez J.N., Reales G., and Calderon R. (2014). Immunoglobulin genes in Andalusia (Spain). Genetic diversity in the Mediterranean space. *C R Biol*. 337(11):646-56. doi: 10.1016/j.crvi.2014.08.004. PMID: 25444709.

Afshan K., **Fortes-Lima C.A.**, Artigas P., Valero A.M., Qayyum M., and Mas-Coma S. (2014). Impact of climate change and man-made irrigation systems on the transmission risk, long-term trend and seasonality of human and animal fascioliasis in Pakistan. *Geospat Health*. 8(2):317-34. doi: http://dx.doi.org/10.4081/gh.2014.22. PMID: 24893010.

Reales G., Hernandez C., Dugoujon J.M., Novelletto A., **Fortes-Lima C.**, Rodriguez J.N., and Calderon R. (2014). New insights into the distribution of APOE polymorphism in the Iberian Peninsula. The case of Andalusia (Spain). *Ann Hum Biol*. 41(5):443-52. doi: 10.3109/03014460.2013.877966. PMID: 24502694.

Ambrosio B., Novelletto A., Hernandez C., Dugoujon J.M., **Fortes-Lima C.**, Rodriguez J.N., and Calderon R. (2012). Y-STR genetic diversity in autochthonous Andalusians from Huelva and Granada provinces (Spain). *Forensic Sci Int Genet*. 6(2):e66-71. doi: 10.1016/j.fsigen.2011.05.007. PMID: 21664894.

Ambrosio B., Hernandez C., Novelletto A., Dugoujon J.M., Rodriguez J.N., Cuesta P., **Fortes-Lima C.**, and Calderon R. (2010a). Searching the peopling of the Iberian Peninsula from the genetics of two Andalusian subpopulations: a study based on Y chromosome haplogroups J and E. *Collegium Antropologicum*. 4: 1215-1228. Link: http://hrcak.srce.hr/file/94109. PMID: 21874703.

Ambrosio B., Dugoujon J.M., Hernandez C., De La Fuente D., Gonzalez-Martin A., **Fortes-Lima C.A.**, Novelletto A., Rodriguez J.N., and Calderon R. (2010b). The Andalusian population from Huelva reveals a high diversification of Y-DNA paternal lineages from haplogroup E: Identifying human male movements within the Mediterranean space. *Ann Hum Biol*. 37:86-107. doi: 10.3109/03014460903229155. PMID: 19939195.

ABSTRACT

Background

The transatlantic slave trade, from the 15th to the 19th centuries, changed dramatically the demography of the Americas. Thousands of enslaved Africans managed to escape from the plantations of European colonizers, and formed independent African settlements of free people (or 'Marron'). Here, we study four *Noir Marron* communities from French Guiana and Surinam, as well as other populations with noteworthy African heritage in Brazil and Colombia, and West African populations in Benin, Ivory Coast, and Mali. To uncover different population histories, these populations were specifically characterized using different genetic markers based on 17 Y-STRs, 96 Y-SNPs, whole mtDNA genome, and genome-wide SNP data (4.5 million autosomal SNP).

Results

Paternally and maternally inherited DNA highlighted different patterns of sex-biased gene flow in both Afro-Brazilian and Afro-Colombian populations that suggest different preferential marriage behaviours. In sharp contrast, the Noir Marron communities presented the highest African ancestry in all genetic systems analysed (above 98%). These communities have apparently a null gene flow with non-African groups, and also present elevated inbreeding coefficients. In good agreement with linguistic studies, the Noir Marron communities showed a biogeographical ancestry associated with historical West African Kingdoms that existed in modern Benin during the slave trade. Afro-Colombians indicated genetic ancestry linked with the Gold Coast region. While Afro-Brazilian genetic ancestry was linked with the West Central African region, also supported by historical research.

Conclusions

This study provides specific genetic information in African Americans and thereby helps us to reconstruct broken links with their African past. The Noir Marron communities revealed a remarkably high African identity, which is still linked to Bight of Benin region. The Afro-Brazilian and Afro-Colombian populations present different demographic histories because of their different colonial pasts. Within an appropriate historical framework, genetic ancestry can add further understanding of ethnicity in African populations throughout the Atlantic world.

Keywords: *African American, Noir Marron, Afro-Colombian, Afro-Brazilian, West African, Y chromosome, mitochondrial DNA genome, genome-wide SNP data, admixture, global ancestry, population structure, transatlantic slave trade, African diaspora.*

RÉSUMÉ

Introduction

La traite transatlantique, du 15ième au 19ième siècle, a changé radicalement la démographie des Amériques. Des milliers d'esclaves africains ont réussi à échapper aux plantations des colonisateurs européens, et ont formé des colonies indépendantes de peuples libres (ou 'Marron'). Dans notre travail, nous étudions quatre communautés *Noir Marron* de la Guyane française et du Surinam, ainsi que d'autres populations ayant un héritage africain : Brésil et Colombie, ainsi que des populations d'Afrique de l'Ouest : Bénin, Côte-d'Ivoire et Mali. Afin de définir les différentes histoires démographiques, ces populations ont été caractérisées à l'aide de plusieurs marqueurs génétiques des lignées uniparentales: chromosome Y (17 Y-STR et 96 Y-SNP), ADN mitochondrial (génomes complet), et de données pan-génomiques (4,5 millions de SNP).

Résultats

Les ADN paternels et maternels ont mis en évidence différents modèles de biais sexuels dans les populations afro-brésiliennes et afro-colombiennes, ce qui suggère des comportements de mariages préférentiels. À l'opposé, les communautés Noir Marron présentent l'origine africaine la plus élevée pour tous les systèmes génétiques analysés (supérieure à 98%). Dans ces communautés, on note l'absence de flux génique avec les groupes non-africains, et également des coefficients de consanguinité très élevés. En accord avec les études linguistiques, les communautés Noir Marron montrent une origine géographique africaine associée aux royaumes historiques de l'Afrique de l'Ouest qui existaient au Bénin durant la traite des esclaves. En accord avec les études historiques, l'origine des afro-colombiens montre des liens génétiques avec la région de la Côte de l'Or, et celle des afro-brésiliens avec la région de l'Afrique centrale.

Conclusions

Cette étude fournit une importante information génétique sur les afro-américains et nous permet de reconstruire les liens brisés avec leur passé africain. Les communautés Noir Marron montrent une identité africaine très élevée, reliée au Golfe du Bénin. Les populations afro-brésiliennes et afro-colombiennes font apparaitre différentes histoires démographiques en raison de leur passé colonial différent. Confronté avec les études historiques, la génétique permet de mieux appréhender l'identité ethnique africaine sur les deux rives de l'Atlantique.

Mots-clés: *Afro-américain, Noir Marron, Afro-colombiens, Afro-brésiliens, Afrique de l'Ouest, chromosome Y, ADN mitochondrial, données pan-génomiques, métissage, ascendance globale, structure génétique des populations, traite transatlantique, diaspora africaine.*