THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par *Université Toulouse III (UT3 Paul Sabatier)*
Discipline ou spécialité : *Anthropologie génétique*

Présentée et soutenue par *GAYÀ-VIDAL Maria Magdalena*
Le *26 Septembre 2011*

Titre : *Genetic characteristics of the two main Native groups in Bolivia: Aymaras and Quechuas.*

JURY
LARROUY Georges, Professeur émérite
HERNANDEZ Miguel, Directeur de Recherche
FRANCALACCI Paolo, Directeur de Recherche
GARCIA MORO Clara, Directeur de Recherche
GIBERT Morgane, Chargé de Recherche
PEREIRA SANTOS Cristina, Chargé de Recherche
CALO Carla Directeur de Recherche

Ecole doctorale : *Biologie Santé Biotechnologie*
Unité de recherche : *Anthropologie Moléculaire et Imagerie de Synthèse, UMR 5288*
Directeur(s) de Thèse : *DUGOUJON Jean-Michel, MORAL Pedro*
Rapporteurs : *VARESI Laurent, CALO Carla*

Als meus pares,

na Rosa i en Raimon

“Homo sum; humani nil a me alienum puto”:
« Sóc un home i res del que és humà m’és aliè »

Terenci

"Une chose surtout donne de l'attrait à la pensée des hommes : c’est l'inquiétude."

« Sobretot una cosa fa suggestiu el pensament humà: la inquietud»

Anatole France

“No es faria ja cap altre descobriment si ens conforméssim amb el que ja sabem.”

Lucio Anneo Sèneca

Acknowledgements

Barcelona, 11 June, 2011

After all these years, I have had the opportunity to work in different laboratories, get to know new people, some of them are true friends. I appreciate all the support of those that have always been encouraging me.

First of all, I want to thank my supervisors, Dr. Pedro Moral and Dr. Jean-Michel Dugoujon. I am very grateful to Dr. Dugoujon for offering me the opportunity to work with an Aymara population for my Erasmus project several years ago, giving me such an amazing project and keeping me in contact with the lab of Toulouse. When I first arrived in Toulouse it was impossible for me to think that after all these years I would be so attached to this subject that has fascinated me and to the "*ville rose*" that has given me a lot and where I have lived unforgettable moments. *Merci beaucoup pour m'avoir dirigé la thèse avec compréhension et patience.*

I want to express a special acknowledge to Pedro with whom I have worked most of the time and who has always been available whenever I needed help. Thank you for always encouraging me, even in the most difficult moments. *Me has enseñado mucho tanto en el aspecto humano como profesional, muchas gracias.*

I would also like to thank all the people of both labs, at Moral's lab: Esther, Marc, Emili, Robert, Ares, Josep and Magda R, thanks for all your advice, help, exchange of ideas, jokes, conversations, coffee breaks, and everything that makes the work fun, you are the best company I could ever had. I want to especially thank PhD Athanasiadis (bazinga!) that has helped me a lot during all these years, always with a smile. *No encuentro "ni una sola palabra" para expresar mi agradecimiento, gracias por tu amistad.* I want also thank the master' students, Albert and Irene, and the visitors from other countries who have helped to break the routine of the lab, Wifak, Laura P, Daniela. Also, many thanks to the Moral people I met throughout these years: Toni, Neus V, Natalia, Meri. Thanks to rest of the people from the "Unitat d'Antropologia", Mireia, Neus M and Marta, the Lourdes' group (Bàrbara, Araceli, Mar, Sílvia, Marina, Ximena, etc.), the people from the other coin of the lab, Jordi, Bea, Ferràn, Laura, Mohammed, Andrés, and, of course, I want to thank all the bosses, Lourdes, Clara, Miquel, Alejandro, Daniel, and Txomin. *Moltes gràcies també al Dr. Pons pel seu bon*

humor. All of you are the main reason that I have really enjoyed to work in our lab all these years.

Although I have been more time in Barcelona than in Toulouse, I know that I have good friends there. At the lab of Toulouse, I would like to thank the PhD students I met when I arrived as an Erasmus student, Stéphane (I remember him encouraging me when I was trying my pulse not to trembling too much when charging a gel) and Clotilde for her patience and good explanations. The following years I met Harilanto, Nicolas, Nancy, Laure, Beatriz (another visitor, "la otra *españolita*"); in addition to their good company in and outside the lab, they have always helped me a lot to solve the "distance" issue. It's a pleasure to know that I have such good colleagues and friends over there. I want to thank also Mr. Larrouy, Denise L, Line H, André S, Morgane G, Eveline G, Cathie T, and especially to Mr. Crubézy who accepted me for doing my Erasmus' stage at the lab. *Merci beaucoup à tous pour me permettre partager une partie de ma thèse avec vous dans un laboratoire où je me suis sentie comme chez mois, pour votre convivialité, et bien sûr, pour tous les " apéros".*

I can not forget the three-moth stay in London. I am very grateful to Dr. Andrés Ruiz-Linares from the UCL for accepting me at his lab and making me feel like home. There, I meet Barbara, Tábita, Karla, Andrés, Desmond, Mari-Wyn, many thanks for sharing all those moments with me, without you that stay would have not been the same.

From outside the labs I have received a lot of support throughout these years from all my friends. In Barcelona, thanks to Clara, Maite, Marta, Ana, Diana, Laura G, Sergi, Laura S, Sílvia, for listening to me, and being comprehensive when I am stressed (only good friends can do that!). From Mallorca, I want to thank Marga S, Marga G and Vicky for their true friendship. Also I want to thank my class mates from the Bioinformatics master, for their help and support during these last months, Yorgos of course, José Luis, Pascal, David, etc.

Finally, and the most important of all, I want to thank my family, my parents (Rosa and Raimon), and my sister Mia for all their support and courage they have given to me. *Moltes gràcies a tots tres per sempre animar-me i estar al meu costat donant-me tant d'amor, fent que tot aquest camí fos més fàcil*; and special thanks to Jose, for his help and lots of patience. *Gràcies per compartir amb mi els moments més difícils, però també els més feliços, com aquest.*

TABLE OF CONTENTS

Foreword

The Andean highlands have been the cradle of complex societies and several amazing civilizations that still fascinate us. A legacy of that are two of the most spoken Native American languages (Aymara and Quechua). Another particularity of this region, as impressive as its population history, is an environment of altitude with hard living conditions, to which highlanders are biologically adapted.

In the past decades, as genetic studies emerged, a new world of possibilities appeared. Since then, a lot of studies have tried to elucidate the questions about human populations from a genetic point of view, complementing other more traditional approaches as archaeology, history, and linguistics. Several genetic studies on Andean highlanders have been focused on providing a better knowledge of the genetic basis of adaptation to altitude. However, other aspects of these populations, like their genetic relationships, are also interesting.

In this context, the present work is a genetic study of two populations of the two major Native linguistic groups (Aymaras and Quechuas) from the Bolivian Altiplano. The first section of this work, the Introduction, situates the present work into a general context. The introduction has seven parts: the first part is just an introduction to the Americas, the central Andes and Bolivia; the second part is a historical section, giving a brief revision to the history of the Americas based on archaeological and historical records; a third part treats the linguistics. These sections will introduce a general knowledge to the Americas, although focusing on South America and the Andean region in particular. The fourth part deals with cultural and environment aspects of Andean populations. The fifth part is dedicated to the human genetic variation, including a description of the most important concepts in human population genetics and the contribution of genetics to the history of the America, in particular South America and the Andean region. In the sixth section the models about the peopling of the Americas are revised, and the final part gives a detailed description of the two populations studied here.

The Introduction is followed by the goals of this work, in the Objetives section. The Results section contains four parts. The first one is the report of the supervisors about the quality of the published papers, and three remaining sections, each one corresponding to an article accompanied by a brief summary in Catalan. The first article is about *Alu* insertions in the two Bolivian samples, the second article, contains the results obtained for uniparental markers, for both the mtDNA and the Y-chromosome, and the last one consists of an article about the genetic variation on the *APOE/C1/C4/C2* gene cluster region. Finally, there is the Discussion section and the Conclusions. A summary in Catalan and French is also added.

Keywords

Aymaras

Quechuas

Native American

Andean populations

Bolivia

Mitochondrial DNA

Y-chromosome

Alu insertions

APOE/APOC1/APOC4/APOC2

INTRODUCTION

I.1. The place

I.1.1 The Americas

The Americas, also known as the New World, have always awakened a special interest in different disciplines, probably due to the fact that they were the last continent to be populated. However, this fact has not implied that unravelling the mysteries of the peopling of the Americas is easy. Far from that, controversy still exists in all fields that have tried to answer the questions related to this topic (archaeology, linguistics, morphological and genetic anthropology).

South America on its own bears such a high complexity at different levels, that a lot of studies in different fields have focused only on it. South America could be divided into several cultural geographical regions: Andes, Llanos, Amazonia, Chaco, Pampa, and Patagonia (Figure 1). The Andean region, presenting one of the most fascinating and incredible histories, as well as a particular environment of altitude requiring biological adaptations, is particularly interesting. Culturally, the Andes can be divided into three areas, North (Ecuador, Colombia, and Venezuela), Central (Peru, Bolivia and North Chile) and South (Chile and Argentina) (Figure 1).

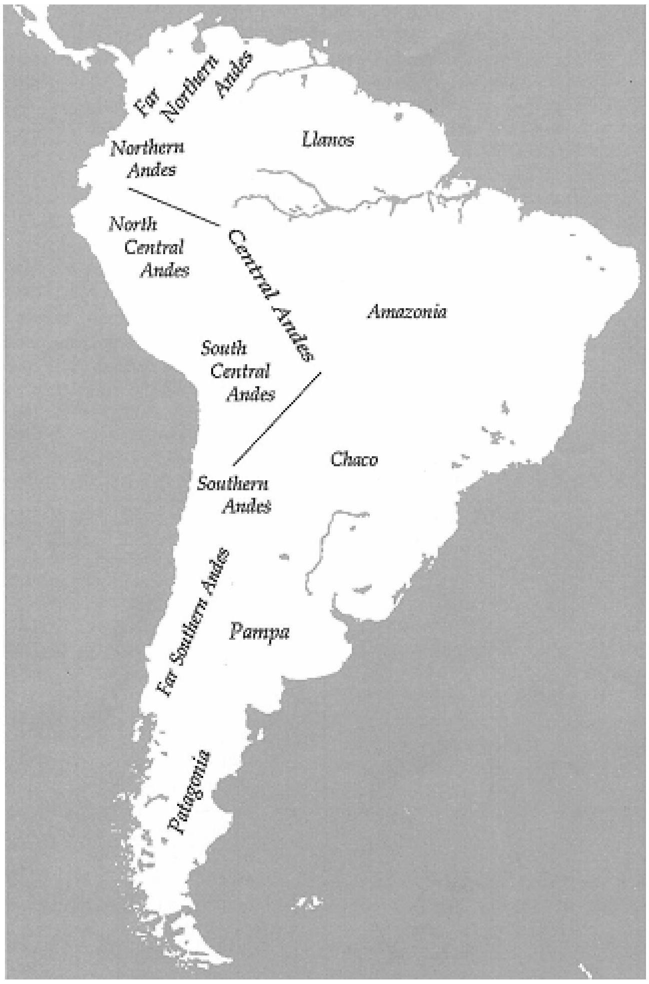


Figure 1. South America cultural areas (Stanish, 2001).

I.1.2 The central Andes

The central Andean region is geographically divisible into three areas: 1) the lowland area to the east of the Andes, heavily foliaged and connected to the westernmost part of the Amazonian jungle, 2) a dry, arid coastal plain crossed by several rivers flowing from east (Andes) to west (Pacific), and 3) the sierra or highlands, the rugged mountains of the Andean chain. It was in the last two areas where the first complex societies appeared and where the major civilizations of South America flourished (Stanish, 2001).

The Altiplano, a plateau enclosed between the two Andean chains, at an average altitude of ~3658 meters above sea level (m.a.s.l.), (3000-4500 m.a.s.l.), is where the Andes reach their maximum width. The most part is located in Bolivia, but also it occupies part of southern Peru and areas of North Chile (Figure 2). In the border between Peru and Bolivia, we find the Lake Titicaca at 3811 m.a.s.l., the larger lake in South America.

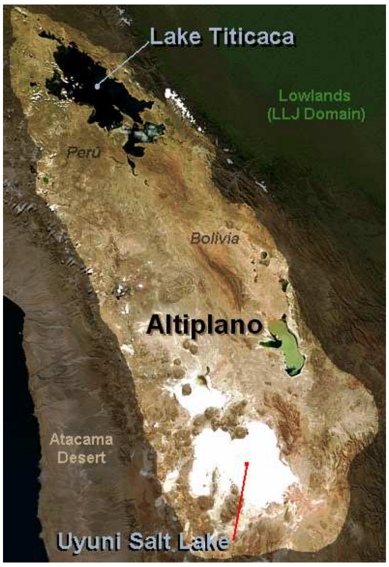


Figure 2. The Andean Altiplano.

The idea that the central Andean region presents a cultural homogeny is generally assumed, probably due to the fact that when the Europeans arrived, the Inca Empire covered this region entirely, promoting the cultural unity of the Empire, and the Quechua language was present in most areas (imposed by the Incas). Nevertheless, before the Inca Empire, distinct cultural, linguistic, and political areas characterized this region.

I.1.3 Bolivia

Bolivia is a country located in central South America bordered by Peru, Chile, Argentina, Paraguay and Brazil that has an area of 1,098,580 Km^2. Its landscape can be divided into three topographical and ecological parts, the Altiplano, the Yungas, and the Lowlands (Figure 3).

The Lowlands (Oriente) cover two-thirds of the country in the eastern (Chaco plains) and northern part of Bolivia.

The Yungas is a humid, rainy and warm area between the highlands and the lowlands in the eastern slope of the Cordillera Real of southern Peru and Bolivia. The valleys, mountains, and streams contribute to the presence of very diverse forests, becoming a rich environment.

The Andean Altiplano is located in the western part of the country. Almost half of the population lives on the plateau that contains the capital La Paz (3630 m.a.s.l.), as well as two other big cities, Oruro and Potosí (4090 m.a.s.l.). Distinct dry and rainy seasons give birth to the Puna grassland, the ecosystem found in the Altiplano as well as in the central Andean highlands. The Puna is found above the tree line at ~3500 m of altitude and below the snow line at 4500-5000 m.a.s.l., resulting into a cold region with low oxygen diffusion. It is drier than the *páramo* of the northern Andes. Native mammals of the Puna are alpacas, llamas, vicuñas, and guanacos.

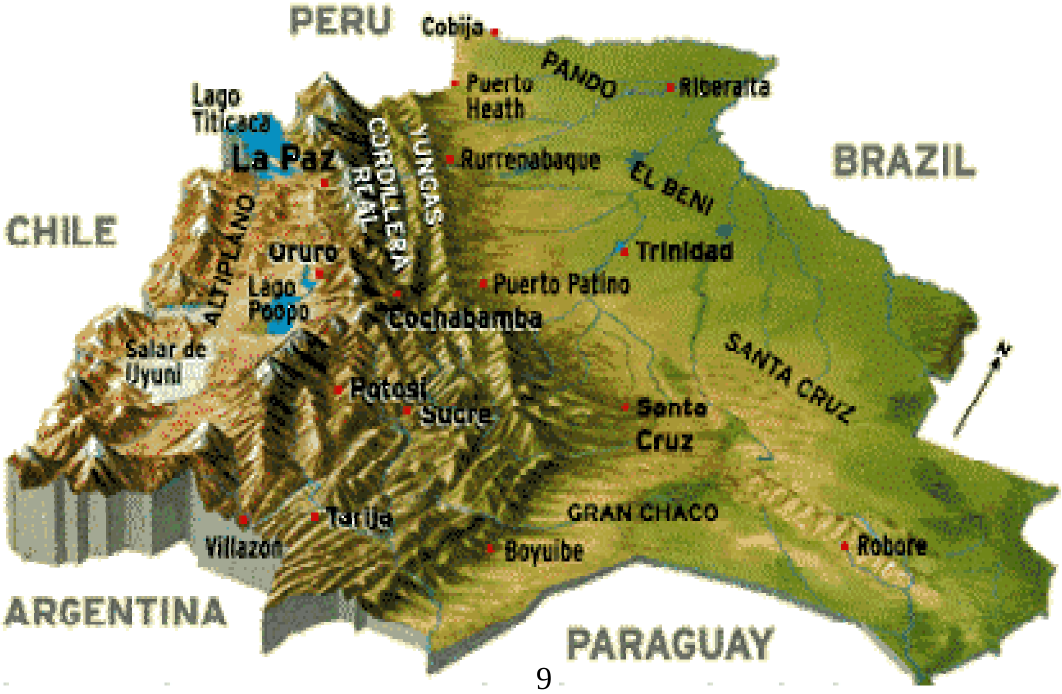


Figure 3. Bolivian landscape.

I.2. Brief history of Native Americans: archaeological and historical records.

The history of the population of the Americas can be split into two main periods: the Pre-Columbian period, extended from the first settlements of the Americas to the arrival of Columbus in 1492, and the historical period that can be divided into three stages: the conquest, the colonial and the post-colonial times.

I.2.1 The first Americans

In 1590, Friar José de Acosta already argued that Native inhabitants of the Americas must come from Asian populations (Acosta, 2002). Although nowadays it is widely accepted that Native Americans came from East Asia through the Bering Strait at some point during the last glacial maximum (LGM) when the sea level was much lower and Asia and America were connected, the time, routes, and the number of waves that entered the New World still remains uncertain (see the I.6 section).

The time of the entrance of the first humans in the Americas has been under discussion among archaeologists for decades. Until recently, it was thought that the first migrants entered the New World ~11,500 BP according to the Clovis-first-model. This model, based on the Clovis arrow points (stone tools that have been found in most archaeological sites in North America), proposed that hunter-gatherers colonized the continent leaving behind traces that have permitted to date their passage to Mexico, Peru, Chile until the Southernmost part of Argentina, the Patagonia, where remains were found dated back to 11,000 BP (Salzano and Callegary-Jacques, 1988) and that leaded to the extinction of the mammalian mega-fauna. These data indicated that the first settlers of the Americas colonized the whole continent from current Alaska to Tierra del Fuego in some thousand years.

However, new archaeological sites and revision of previous ones have proved a pre-Clovis settlement, opposing to the Clovis-first theory. One of the most important sites is Monte Verde in Chile dated to 14,700 BP (Dillehay, 1997). Other important pre-Clovis sites (Meadowcroft, Page-Ladson, and Paisley Cave) in North America also indicate a presence of humans in the Americas from ~15.2 to 14.1 ka. Although less certain, it is important to mention possible sites dating earlier than 15 ka (Cactus Hill) or even older, between 19 to 22 ka, (La Sena, Lovewell) (Goebel, Waters, and

O'Rourke, 2008). Finally, a recently discovered site in Texas dated back to 15.5 ka (Pringle, 2011).

Archaeological data have proved that the eastern part of Siberia was populated around 32 ka, indicating that modern humans had learnt how to subsist in this extreme environment. In the eastern part of Beringia, some remains have been proposed to have about 28 ka old, according to some archaeological sites. However, the earliest reliable remains date back to 14 ka in current Alaska. The coastal corridor seems to have deglaciated and permitted human habitation by 15 ka, and the interior corridor until 14 to 13.5 ka (Goebel et al., 2008).

After the discovery of the Monte Verde site, the fact that, before 15.7 ka, the ice covered completely Alaska and Canada and therefore there was a margin of only ~1000 years for the journey from Asia to Chile on foot, the Pacific-coast theory appeared. Fladmark (1979) proposed that crossing the two continents (12,000-mile trip) in only 1000 years could be possible travelling along the Pacific coast by boats. However, this theory is difficult to prove since sea level has risen, inundating the coastline. Another proposed model has been the Atlantic coast route, as the highest concentration of Clovis artefacts is found in the eastern part of North America, indicating a higher population density in the East than in the interior of the continent as expected according to the traditional model.

I.2.2 The central Andean region

According to archaeological records, the human settlement of the central Andean region could be traced back to around 10,000 BP and has continued until present.

Pre-historic times can be divided into several periods (Stanish, 2001): A pre-Ceramic and an Initial period followed by three Horizons -Early, Middle, and Late- which are characterized by pan-Andean cultures (Chavín, Tiwanaku-Wari, Inca). Between the Horizons, two intermediate periods (Early, Late) took place with flourishing regional cultures.

The **Pre-Ceramic** period (10,000-2000 BC) is characterized by the movement of hunter-gatherers as evidenced by archaeological remains in Peru around 8000 BC. In the late pre-ceramic period, several cultural sites with important monuments (pyramids, walled plazas, etc) have been found indicating the first sedentary people and the development of the first complex societies according to ranked societies reported in the Pacific coast of Peru (Guayaquil, Peru, 5000 BP; Caral, Peru, 4600 BP, etc). In the highlands, a ritual tradition, known as “Kotosh Religious Tradition”, was developed (Kotosh, La Galgada, Peru 2300 BC). The architecture of the coast and the highlands was different; however, exchange networks existed between the coast, the highlands and the eastern slopes.

The **Initial** period (1800 BC- 900 BC) is characterized by the development of new technologies (ceramic, metallurgic, agriculture and farming) as well as social institutions. The social complexity grew up to reach the formation of hierarchic societies. During this period, regional cultures (e.g. Kotosh, 2300-1200 BC; Cerro Sechin, 1000 BC; Paracas, 800 BC) took place. In the highlands, the important civilization of Chavín started in Chavín de Huantar in central highland Peru around 900 BC, becoming a centre of elite pottery, textile and stone art. It is important to mention the first construction in the south central highlands, the Chiripa site located in the south of the Titicaca basin (1300 BC).

The **Early Horizon** (900 BC- AD 200) corresponds to the first pan-Andean art style known as **Chavín** in the highlands and the coast, representing the first well-documented culture. A general collapse of polities occurred in the coast, while the cultures of the highlands in the north central Andes prospered and the site of Chavín increased in size and power. It has been suggested that by that time the population was 2000-3000 people, making Chavín an important political centre. Its influence reached the region of the current city of Ayacucho, Peru (Sondereguer and Punta, 1999). Other highland sites also grew in size and complexity; in the south central highlands the Pucara site dominated the northern Titicaca basin from 400 BC to AD 200. Pucara art shows links to the contemporary coastal Paracas (800-100 BC) in the Ica region, and Early Tiwanaku, with antecedents in Chavín. There is controversy about whether Chavín and Pucara should be considered states or not. Many agree that these sites were just ceremonial centres, while others consider them as complex chiefdoms or regional political spheres. In the southern Titicaca basin, Tiwanaku was occupied at this time, but its size and complexity is unknown.

The **Early Intermediate** (AD 200-600) is characterized by more regional cultures. In the north coast, the **Moche** culture appeared (AD 400). Moche site was unequivocally a true city, dominated by two main pyramids, *Huaca del Sol* and *Huaca de la Luna*. It may be the first Andean city and the first time there is evidence of royal tombs in the Andes. In the south coast the **Nazca** culture appeared (AD 100-800), whereas in the south central highlands, the **Pucara** ended as a political centre around AD 400. Finally, in the Titicaca basin of the Altiplano, the **Tiwanaku** site grew in importance and power, becoming a state.

The **Middle Horizon** (AD 600-1000) is characterized by the coexistence of the two first "states". As shown in Figure 4, the **Tiwanaku** civilization (100 BC-AD 1000) extended from Lake Titicaca to the south central Andes, while **Wari** (AD 700-1000) extended from Ayacucho in south central Peru to the northern highlands (Blom et al., 1998; Sondereguer and Punta, 1999).

In the south central highlands, around AD 600, Tiwanaku started an expansion process in the western part of Bolivia, Southern Peru, North Chile and North-Western Argentina. The site of Tiwanaku is a vast planned urban capital. In AD 800-900 Tiwanaku presented an impressive architectural core (pyramids, streets, temples, state buildings) surrounded by adobe houses of artisans, labourers, and farmers. It has been estimated that it covered an area of 4-6 km^2, with a population in the Tiwanaku valley ranging from 30,000 to 60,000 (Stanish, 2001). Areas of intensive agricultural production have been detected. Also, Tiwanaku colonies have been found in Moquegua, Cochabamba, Larecaja, and Arequipa. The Tiwanaku state seems to have controlled politically and militarily-strategic areas such as roads, rich agricultural areas and regions with high resources, indispensable to maintain such a population.

In the north central highlands, the Wari culture originated in Huanta (Ayacucho region) and expanded until reaching the Cuzco area in the South and Cajamarca in the North. The Wari urban complex has been calculated to cover an area up to 15 Km^2, the core site presenting a similar size than the contemporary Tiwanaku.

As a conclusion, around AD 500 the first states with a road network, warrior elite, and capitals, existed in the Andes. Around AD 1000, Tiwanaku and Wari had large populations, planned urban capitals, socioeconomic classes, expansionist policies, economical specialization, and colonial sites. The relationships between these two contemporaneous (coexisting around 500 years) and neighbour civilizations, whether they were in conflict, competition or just complementary are in great part hypothetical and controversial (Owen, 1994; Sondereguer and Punta, 1999; Isbell et al., 2008).

The high degree of complexity and the population density reached in the central Andean region was possible thanks to the "Vertical Archipelago Model" (Murra, 1972). This model explains how different environment regions (the coast, the highlands, and the yungas) provide abundant and varied resources.

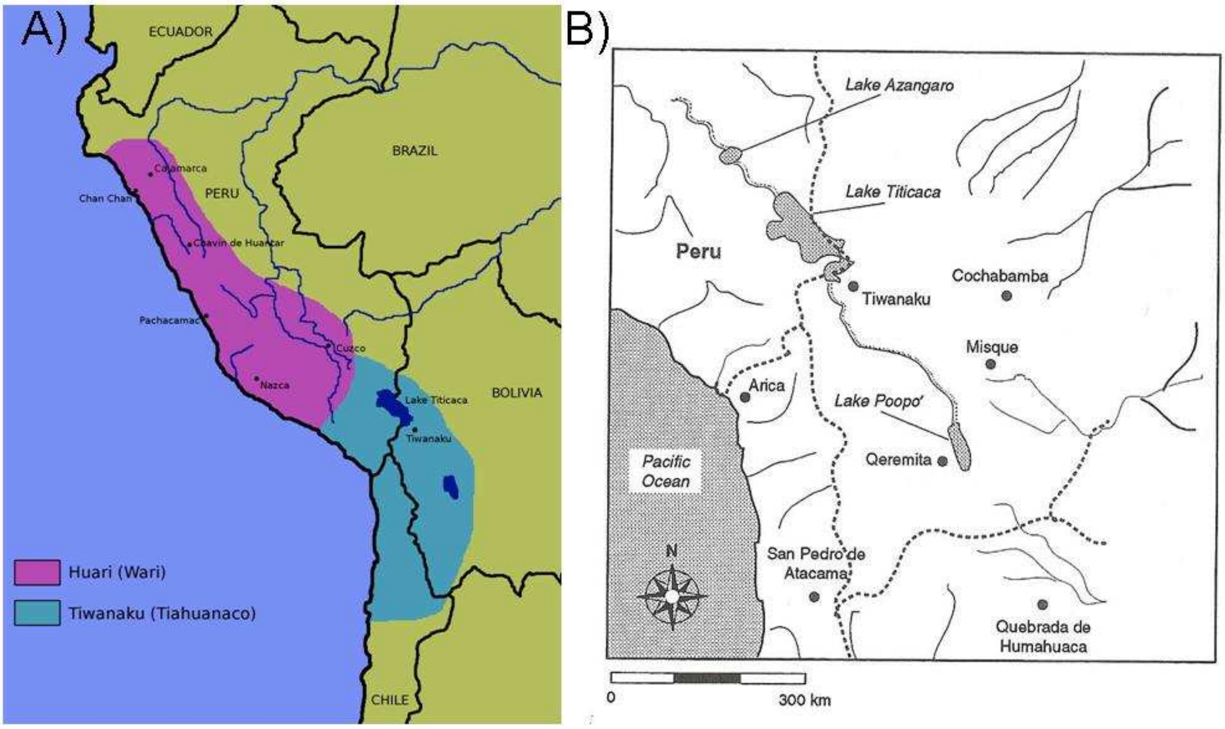


Figure 4. A) Extension of Wari and Tiwanaku civilizations (taken from Wikipedia-based on Heggarty, (2008)). B) Detail of the Tiwanaku influence area (Kolata, 1993).

The **Late Intermediate** or **late regional development** period (AD 1000-1476). At the end of Wari and Tiwanaku states, a period of regional cultures re-emerged (Chanca, Cajamarca, Chincha, Chimú, local states in the Altiplano, etc) (Figure 5A). The end of Tiwanaku, between AD 1000 and 1100, was due to the agricultural collapse because of a dramatic decrease in precipitations and the beginning of a drought period. The cities around the Titicaca Lake disappeared for nearly 400 years and human populations dispersed in smaller groups, starting a period of political instability. The twelve Aymara kingdoms or *señoríos* identified (Bouysse-Cassagne, 1986) established in a deep politic-economic-territorial reorganization are shown in Figure 5B. Two of these kingdoms (Qolla, Lupaqa) seem to have been organized almost at the state level with colonies in the coast and yungas. When the Incas irrupted by the mid fifteenth century, these two groups were in a battle for the political supremacy of the Lake District. Thus, if the Incas had not invaded these territories, probably a new Tiwanaku-style empire would have started (Kolata, 1993).

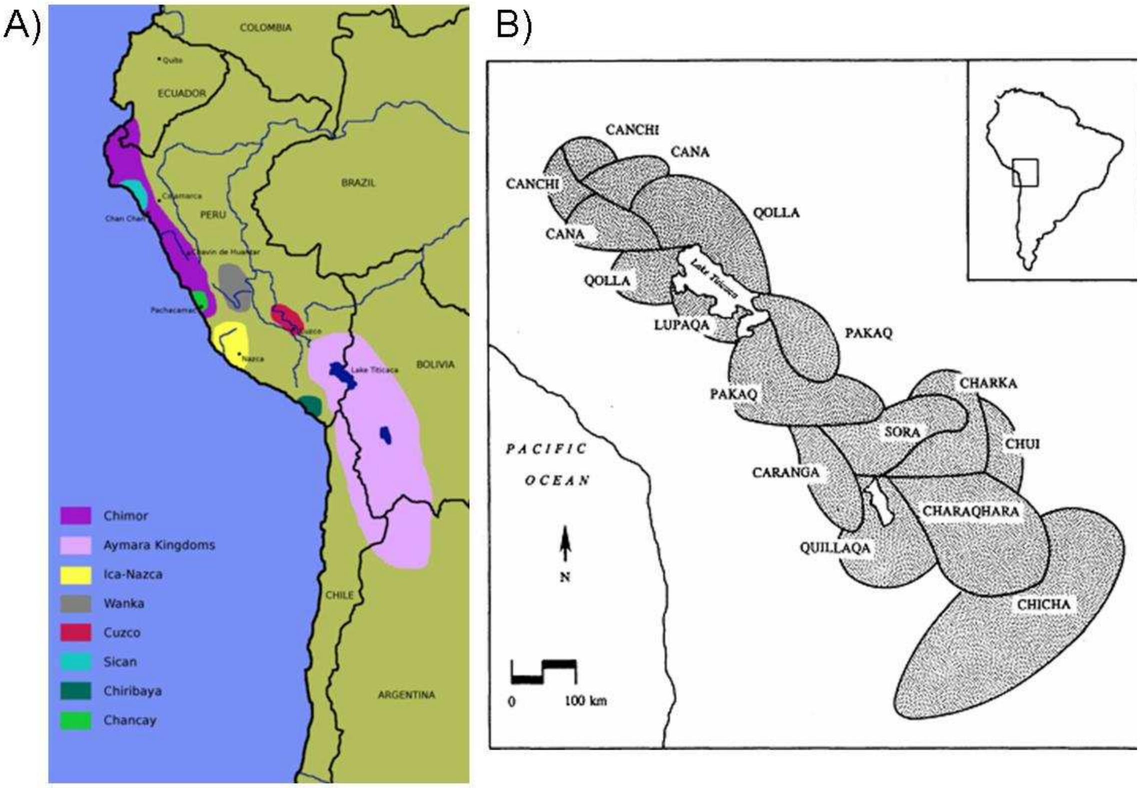


Figure 5. A) Map of the regional cultures of the late regional development period (image taken from Wikipedia). B) The location of the Aymara polities (image taken from Graffam, 1992).

The **Late Horizon** period (AD 1476-1535) corresponds to the **Inca Empire** (1438-1532AD), also known as *Tawatinsuyu*. A prevalent theory says that the Inca civilization derives from a family group probably from the Titicaca region that moved northwards to the Cuzco city in the 12th century (Markham, 1871). From Cuzco, the Inca Empire, in contrast to Tiwanaku, expanded its power towards the North and South with calculated military violence and coercive techniques such as language imposition

(Quechua) and community displacements (*mitma* system). These strategies inevitably awakened the hostility against the Inca Empire. At the arrival of the Europeans, the Inca territory expanded from the South of current Colombia in the North to Chile in the South (Figure 6A). This vast territory, the Tahuantin-suyu, was divided into four provinces or “suyus”: Chinchasuyu, Antisuyu, Contisuyu, and Collasuyu (Figure 6B).

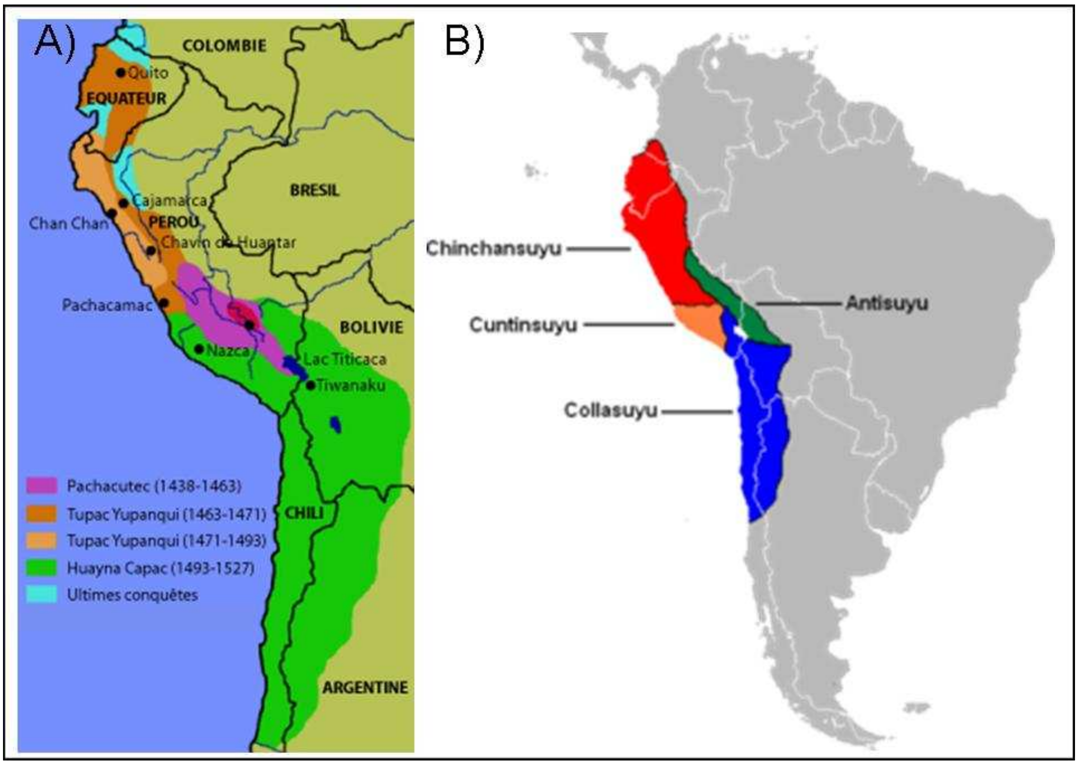


Figure 6. A) The Inca expansion. B) The four quarters of the Inca empire (taken from Wikipedia, based on: A: Rowe, 1963, and B: Kolata, 1993).

The Inca Empire was characterized by the construction of a complex road system that communicated the whole empire. Their architecture is also impressive; the constructions were built using stones sculpted that fit together exactly. The site of Machu Picchu is the most famous Inca construction.

The Incas carried out the *mitma* system (*mitmaq* meaning “outsider” or “newcomer”), deliberate resettlements of people, sometimes even entire villages, for different purposes (colonize new territories, work). They also used the *mita* system (*mit’a* meaning “turn” or “season”), a mandatory public service to carry out the projects of the government such as building the extensive road network, construction of

Emperor and noble houses, monuments, bridges, temples, working in mines and fields. Also military service was required some days out of a year.

The social organization was highly hierarchic. At the highest level of the social pyramid was the Emperor, the "son of the sun", with an absolute power and supposed to have a divine origin. The nobility occupied the highest administrative, military, and religious functions. The local chiefs could continue to be the authority if they were faithful to the sovereign. The labourers at the bottom of the pyramid included the farmers in general as well as the *mitimaes*.

The Conquest and Colonial period

In 1531, the Spaniard Francisco Pizarro arrived in Peru and in 1532 entered Cajamarca and captured the Inca Emperor Atahualpa. In a few years (1531-1536), the Spaniards succeeded in conquering the Inca Empire. The fact that in many regions the Europeans were seen as liberators from the Incas has been suggested as a reason for this rapid disintegration of the Inca Empire. With the arrival of the Spaniards, a lot of villages were funded (Chuquisaca, nowadays Sucre, La Paz, etc) including the capital, Lima.

In 1542, the Viceroyalty of Peru was created, including most Spanish-ruled South America (Figure 7A). Laws that protected the autochthons were created, although never applied. In 1568, several decades after the invasion, the true colonial period started with the arrival of Jesuits and of Viceroy Francisco de Toledo (Murra, 1984). In 1717, the Viceroyalty of New Granada was created, in 1742, the Viceroyalty of the *Capitaneria general de Chile*, and in 1776, the Viceroyalty of Rio de la Plata. The region of current Bolivia, known as Upper Peru, became part of the Viceroyalty of Rio de la Plata (Figure 7B).

During the colonial period, a big effort was made to convert the natives to Christianity; even native languages (Quechua, Guarani) were used, contributing to the expansion of these tongues and providing them with writing systems. Natives served as labour-force under the *mita* system taken from the Incas in mining industry or haciendas owned by Spanish colonials. In 1545, Potosí, a mining town, was founded and in few years became the largest city in the New World with a population estimated of 120,000 people, with constantly arriving people under the *mita* system from the entire Andes

(Cruz, 2006). From there, large amounts of silver were extracted from Cerro Rico, near Potosí, becoming an important revenue for the Spanish Empire.

In the end of the 18th century and beginning of the 19th, important revolts took place against the Spanish authorities in all the colonial territory.

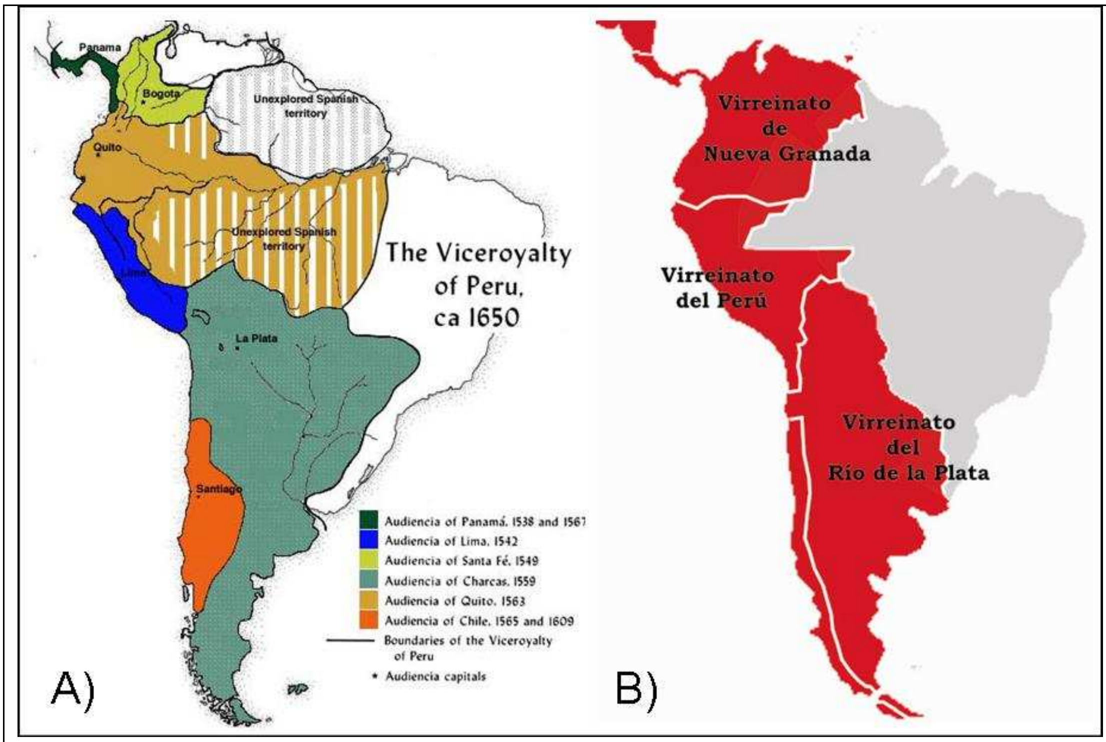

Panama
Bogota
Unexplored Spanish territory
Quito
Unexplored Spanish territory
Lima
La Plata
Santiago
The Viceroyalty of Peru, ca 1650
🟩 Audiencia of Panamá, 1538 and 1567
🟦 Audiencia of Lima, 1542
🟨 Audiencia of Santa Fé, 1549
🟩 Audiencia of Charcas, 1559
🟧 Audiencia of Quito, 1563
🟥 Audiencia of Chile, 1565 and 1609
— Boundaries of the Viceroyalty of Peru
• Audiencia capitals
A)
Virreinato de Nueva Granada
Virreinato del Perú
Virreinato del Río de la Plata
B)

Figure 7. A) Viceroyalty of Peru and its *Audiencias*. B) Viceroyalties in 1800.

The Republic Period

In the beginning of the 19th century, the independence from Spain was declared and after a long period of wars, the independence was established and several republic states were created (Paraguay, 1811; Uruguay, 1815). The insurrectionary army led by Simón Bolívar won the Independence of Nueva Granada (current Venezuela and Colombia), Ecuador and Bolivia (named in honour for Simón Bolívar) by 1825, and José de San Martín that of Argentina (1816), Chile (1818) and Peru (1821).

I.2.3 Bolivia

In Bolivia, the decades after the independence were characterized by political instability, where revolutions alternated with military dictatorships. Moreover, Bolivia had to face several conflicts with frontier countries (Chile, Paraguay, and Brazil) that led to the loss of half of its territory after one century of independence. In particular, the Pacific war (1879-1883) between Chile and Peru and Bolivia was especially cruel with Andean people as it took place in their territory. In this war, Bolivia lost the province of Atacama and their access to the sea. In the beginning of the twentieth century, Bolivia lost the Acre region against Brazil and ceded part of the Chaco region to Paraguay in the Chaco war (1933-1935).

In the mid-20th century democracy was established and the government carried out important programs promoting rural education and agrarian reforms. However, migration from the countryside to urban centres has been constant, depopulating the rural area. Currently, 67% of total population (10.1 million people) lives in urban centres. Additionally, as the highlands experienced an excess of population, highlanders started to migrate to the lowlands (Mojos plains in Beni department, Bolivia).

At present, Bolivians are mainly of a native origin (~55%), 30% being Quechua and 25% Aymara speakers, both groups located in the Andean region of the country. Other native groups are the Guaranis, Mojeños, Chimanes, etc, living in the lowlands. Around ~30% of Bolivians are mestizo (mixed Native and European ancestry) and around 15% from European ancestry (Sanchez-Albornoz, 1974; www.cia.gov on 5 May, 2011).

In recent years, movements have risen claiming the recognition of Native rights throughout the Americas and especially South America. A particular case is Bolivia, where in 2005 Evo Morales was the first indigenous descendant (from Aymara origin) elected president of the country.

I.2.4 Demographical Data

There is great disagreement on the size of the Native population of South America when Europeans arrived; the proposed population sizes range from 4.5 to 49 million. An approximation proposed in Crawford, (1998) is shown in Table 1.

Table 1. Population of the Americas at the arrival of the Europeans.

North America	2 million
Central America	25 million
Caribbean	7 million
South America	10 million
Total	44 million

At the arrival of the Europeans in South America, the high population density of the Pacific coastline and the mountains contrasted with other less populated areas (Sánchez-Albornoz, 1974). Considering the estimates proposed by Salzano, (1968) of about ~10 million South Americans, the inhabitants of the Inca Empire have been calculated to be about 3.5 million with the highest density in South America (10 people per square mile) (Crawford, 1998).

After the contact with Europeans, a dramatical depopulation took place due to war, epidemics caused by European diseases to which the Natives had no resistance, hard conditions by forced labour, etc. The depopulation ratios presented high variation depending on the region. Some groups became extinct while others could eventually recover.

The population of Native Americans decreased from over 44 million to 2 or 3 million in less than 100 years. This drastic reduction of the Native American gene pool from 1/3 to 1/25 of their previous sizes implied a great loss of genetic variation, that is, a genetic bottleneck. Therefore, the current Native populations, descendants of the survivors, may present different frequency distribution of some genetic traits (Crawford, 1998).

I.3 Linguistic data

As peoples migrate, separate, and isolate, their genetic patrimony diverges, but also their culture, including the language. From the study of the relationships among languages, the relationships of the peoples who spoke them can be inferred. Linguists have tried to classify world languages into families based on comparative linguistics. The field that studies the genealogical relationships between languages and family languages is the genetic linguistics. The linguist Johanna Nichols proposed that at least 20,000 years, possibly even 30,000 years would have been necessary to produce the large amount of diversity of languages among Native Americans. Although at present most linguists are pessimistic about the use of their data to reconstruct ancient population histories beyond about ~8 ka (Goebel et al., 2008), some works complementing linguistics and archaeology seem to be useful to reconstruct approximate time-scales and the geography of language expansion in the last thousands of years (Heggarty, 2007, 2008).

I.3.1 The Americas

One of the most controversial areas of genetic linguistics is the classification of Native American languages. Greenberg suggested three families, Eskimo-Aleut, Na-Dene and Amerind families (Figure 8). The Eskimo-Aleut would belong to the large, more ancient Eurasiatic family, together with the Indo-European, Uralic, Altaic, Korean-Japanese-Ainu, and Chukchi-Kamchatkan families. The Na-Dene family, first identified by Sapir in 1915, is a family included into a larger one called Dene-Caucasian, including also Basque, Caucasian, Burushaski, Sino Tibetan, and Yeniseian. The Amerind family groups together the Centre, South, and part of North American languages. This family includes eleven branches: Almosan-Keresiouan, Penutian, Hokan, and Central Amerind in North America, and Chibchan-Paezan, Andean, Macro-Tucanoan, Equatorial, Macro-Carib, Macro-Panoan, and Macro-Ge, most of them restricted to South America (Ruhlen, 1994).

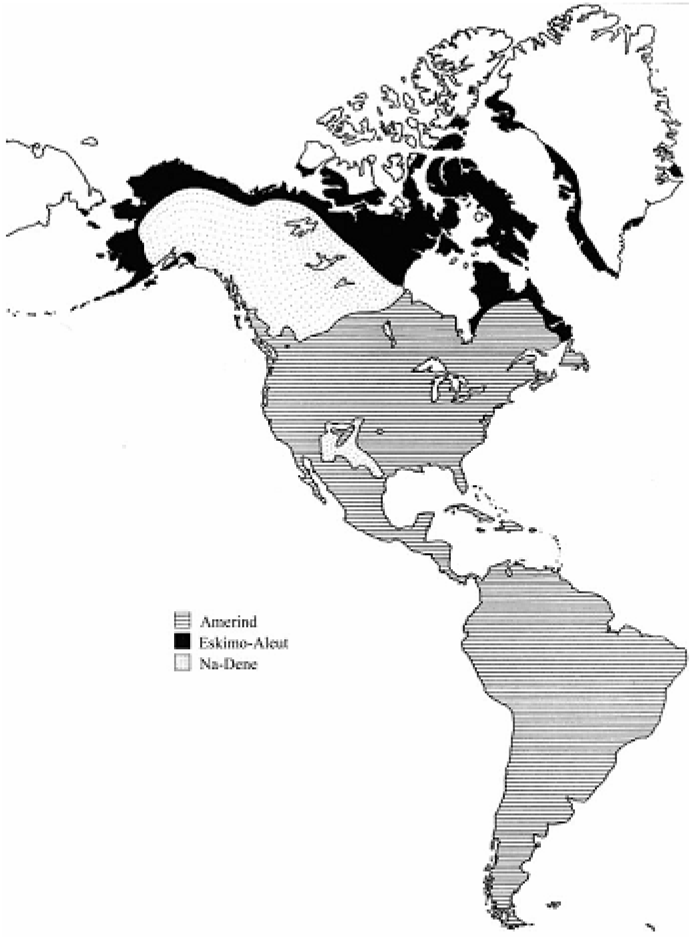


Figure 8. Distribution of the linguistic families in the Americas proposed by Greenberg et al., 1986 (image from Ruhlen, 1994).

Linguists have criticized this grouping into only three families. Most controversies are in the classification of the language families of South America. Campbell believes that this macrogrouping is unjustified, indicating that South America probably exhibits more genetic and typological diversity than North America and Mesoamerica put together. We will mention the three main classifications proposed by Greenberg (1987), Loukotka (1968), and Campbell (1997).

I.3.2 South America

Greenberg (1987) grouped all South American languages into one stock (Amerind) including four clusters (Chibchan-Paezan, Andean, Equatorial-Tucanoan, and Ge-Pano-Carib). His classification has mostly been used by human biologists to hypothesize about the peopling of America. Contrarily, Loukotka (1968) proposed 117 independent families grouped into three categories (Andean, Tropical Forest, and Paleo-American). Since he did not specify their relationships, it seems more a geographic or an anthropological clustering rather than a linguistic one (Adelaar and Muysken, 2004). On the other hand, Campbell (1997) proposed a conservative classification lacking an internal structure because, as most linguists, he thinks that the data and methods used by Loukotka and Greenberg are insufficient to establish these relationships.

However, the position of the Andean family in the Greeberg's classification is not clear since in his first classification (Greenberg, 1959), the South American languages were divided into three groups: Ge-Pano-Carib, Andean-Equatorian, and Chibchan-Paezan, the Andean and the Equatorian subfamilies grouped together because Andean languages showed a small distance with Arawak, suggesting the tropical forest as the cradle of the Andean family; and more recently, the Andean family has been grouped with the Chibchan-Paezan family resulting in the following three groups: Andean-Chibchan-Paezan, Equatorian-Tucanoan, Ge-Pano-Carib (Greenberg and Ruhlen, 2007).

I.3.3 The central Andes

The Andean family includes two of the most spoken Native American language families, Quechua and Aymara. Quechua is the most widely spoken language family in the Americas, with more than 10 million speakers in Ecuador, Peru, southern Bolivia, northern Chile, and Colombia; and Aymara has almost 2.5 million speakers, mainly in Bolivia, but also in parts of Peru and Chile (Lewis, 2009). Moreover, in the south central Andean highlands, other languages were widely spoken (Uru-Chipaya, Pukina) and nowadays are endangered or extinct. Linguists have proposed different theories for explaining their origin, antiquity, and relationships (see Browman, 1994; Itier, 2002; Goulder, 2003; Adelaar & Muysken, 2004; Heggarty, 2007, 2008).

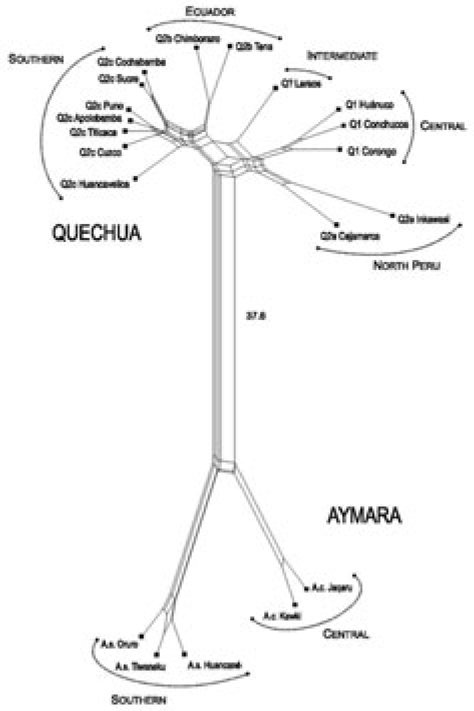


Figure 9. NeighborNet Showing semantic lexical divergence of Aymara and Quechua (Heggarty, 2008)

The relationship between Aymara and Quechua is still on debate. Some studies supported that the two languages were genetically related by an ancient proto-Quechumaran (Orr & Longacre, 1968), or that the two groups could be included into a family named Quechumaran (Mason, 1963). More recent linguistic reports suggested that they are separated enough to be considered separate families and if once there was a proto-Quechumara ancestor, it would be to an extremely ancient period that linguists can only speculate (Heggarty, 2008). Figure 9 shows the level of diversity within and between the two linguistic groups.

In any case, it is unquestionable that there has been a strong contact between Aymara and Quechua due to long periods of mutual influence from the beginning, resulting in high similarities (Heggarty, 2008).

Although the expansions of Quechua and Aymara have been attributed to the Inca period (Late Horizon) and Tiwanaku (Middle Horizon) respectively, there is strong consensus among Andean linguists that the ancestor language of the two families started diverging long before the Late Horizon and probably before the Middle Horizon too, maybe even by a millennium or so. Moreover, Quechua did not appear in Cuzco or Aymara in the Altiplano; rather, proto-Aymara and proto-Quechua languages would have originated in Central Peru (Heggarty, 2008).

I.3.3.1 Quechua

Although traditionally one talks about Quechua language and its dialects, Quechua is a linguistic family of unintelligible languages (Itier, 2002; Heggarty, 2007). The Quechua family is divided into two groups of languages (Figure 10): Quechua I (or Central Quechua) in central Peru and Quechua II (or Peripheral Quechua) including: i) II-A, in northern mountains of Peru, ii) II-B, in Ecuador (Kichwa), north Peru and Colombia, and iii) II-C, in South Peru, Bolivia and Argentina (Torero, 1983).

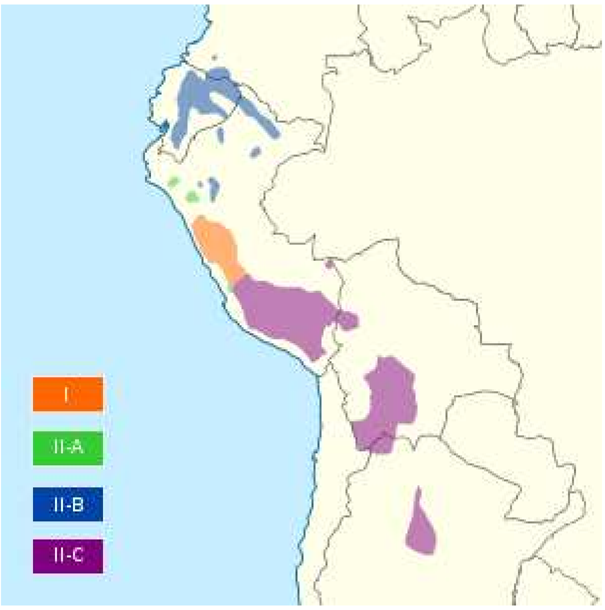


Figure 10. Distribution of Quechua dialects (from Wikipedia, based on Heggarty, 2007)

Quechua, always linked to the Incas, did not appear in Cuzco, and its expansion is very much older than the Inca period (Itier, 2002). In fact, it is likely that about a millennium before the Incas, Quechua had had a more ancient pan-Andean distribution. Heggarty (2007) proposes that Quechua would have arrived in Ecuador some centuries before the Incas as *lingua franca* for trade purposes, and the Ecuadorian populations having an origin language would have adopted Quechua.

The most probable geographical origin of Quechua seems to be Central Peru that corresponds to present Central Quechua. Finding a more precise location is difficult. Torero proposed the central coast, in the area of formation of proto-Chavín cultures, Cerrón-Palomino proposed a further inland location, even the early pre-ceramic site of Caral (3000-1600 BC) for a pre-proto-Quechua. According to Heggarty (2007), data on the degree of diversity per unit area points out to a region in the highlands near Lima, in the Yauyos province were the first language split it is supposed to occur.

The degree of divergence of Quechua family is comparable to that of Romance languages, therefore, a similar time-frame has been proposed. That is, the Early Horizon (900 BC-AD 200) that would coincide with the Chavín culture in the highlands and coast of north-central Peru, where a proto-Quechua form would have acted as lingua franca spreading to North and South occupying the distribution that nowadays presents Central Quechua and would have been an ancestor of the Quechua I spoken in Peru (Kolata, 1993; Goulder, 2002). However, possible dating could go back up to two millennia BP or even more (Heggarty, 2007). From there, Quechua spread to the North reaching Ecuador and to the South (Heggarty, 2008).

In the 15th century, the Incas adopted Quechua as the language of their empire probably since it already was a vehicular language; a language of communication between regions, also a language of privilege, and it was probably their second or perhaps their third language. In fact, some authors propose a regional form of Aymara as the language of the Inca nobility, and at an earlier stage Pukina (Heggarty, 2007).

In any case, the Incas spread the Quechua II-C dialect imposing it as the official language of the empire. Subsequently, during the Colonial period, the Spaniards continued to use it as *lingua franca* (it was spoken in the mines, on haciendas, and in commerce), thus, Quechua II-C expanded to the South, in Peru and Bolivia, where is currently spoken (Itier, 2002).

Despite nowadays Quechua is an official language, together with Spanish in Ecuador, and also with Aymara in Peru and Bolivia, it is relegated to a second place.

I.3.3.2 Aymara

Aymara is composed of two branches: a) Central Aymara or Tupino, spoken in central Peru, in the semi-desert mountains of Lima department, in the province of Yauyos, where there are two clusters of isolated villages that speaks Jaqaru and Kawki. Central Aymara is spoken by no more than a thousand people and are in process of extinction. Some authors consider them separate languages and others close dialects; b)

Southern Aymara or Collavino, spoken in the Altiplano with three dialects, Huancané, Tiwanaku, and Oruro (Cerrón-Palomino, 2000; Heggarty, 2008).

Nowadays, Aymara is spoken by ~1.6 million of Bolivian-speakers, 500,000 Peruvians and 30,000 speakers in Chile, although before the spreading of the Quechua language, the Aymara distribution covered a wider extension than today as it is shown in Figure 11 (Tschopik, 1963; Itier, 2002).

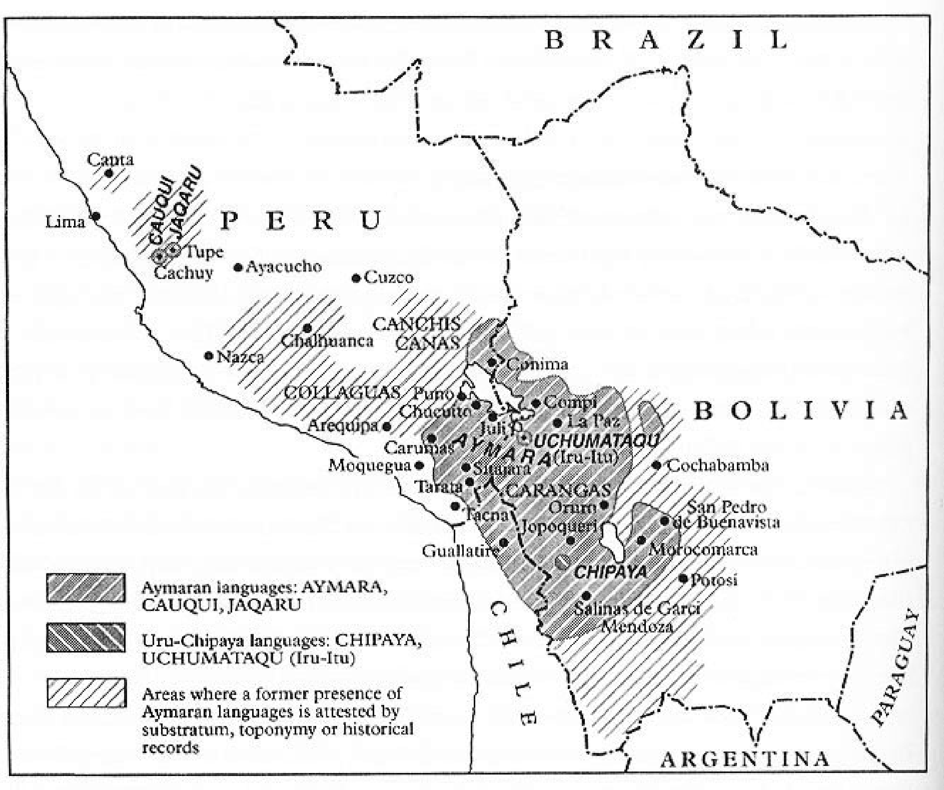


Figure 11. Distribution of Aymara and Uru-chipaya (Adelaar and Muysken, 2004).

The fact that the currently Aymara-speaking area closely overlaps with the influence area of Tiwanaku made some authors assume that Aymara was the language of Tiwanaku, and therefore the time-frame of the Aymara expansion would be the Early Intermediate and Middle Horizon, and the Aymara homeland, the Tiwanaku itself (Bird et al., 1984). However, the low diversity within Aymara in the Altiplano is inconsistent with an expansion as early as Middle Horizon and thus, with a Tiwanaku homeland. In this way, Central Aymara shows a higher within diversity per unit of area than Southern Aymara, indicating that the Central Aymara region is probably closer to the original Aymara homeland. There is agreement among important Andean linguists that Aymara did not originate in the Altiplano, but in the central-south coast of Peru. Torero (2002) proposed the southern coast (the area of Paracas culture). Moreover, the main Aymara expansion was likely not to be carried out by the Tiwanaku Empire. Both Torero (2002) and Cerrón-Palomino (2000) associate the other Middle Horizon civilization, Wari, with Aymara rather than with Southern Quechua.

According to Heggarty (2008), as the diversity analysis between the two Aymara branches is similar to that found in the Quechua family, the time-frame for the Aymara divergence would be similar to that of Quechua; that is, a span of more than one millennium, but probably less than three.

Torero (2002) associates the first Aymara stages with Paracas and Nazca. A first expansion towards the high sierras would have occurred around the fourth or fifth centuries due to the cultural influence exercised over the sierra, the "Nazcaisation" of the Ayacucho region that would have taken the language of the Nazcas because of its prestige.

In the sixth century, around Ayacucho, the Wari state extended to the south of the Cuzco area, and to the North towards the north-central highlands, and influenced the coast. The second and main wave of Aymara expansion, between 500/600 and 1000 AD, would have been carried out by the Wari expansion, promoting Aymara as state language. Aymara would be present throughout the central-south Andean mountains. Quechua remained in all the northern area. To the South there was a greater expansion of Uruquilla and Pukina according to Torero (Itier, 2002).

Then, which was the language of the Tiwanaku people? Kolata (1993) proposed a multilinguistic scenario in which, at the time of Tiwanaku, three languages were spoken in the area (Uru-Chipaya, Pukina, and proto-Aymara), the herders would be the Aymara-speakers but without indicating whether the proto-Aymara or Pukina was the original language of Tiwanaku. A multilingual scenario with several languages has also been proposed, each one restricted to an area of the society (Quechua: administration, Aymara: trade, Pukina: religion, and Uru: landless and lower classes) (see Browman, 1994). Other authors support that the people of Tiwanaku more probably spoke Pukina. According to Torero, Uruquilla was the first language of the Tiwanaku people. Although Aymara was not the language of Tiwanaku, during the late Tiwanaku period (contemporary with the Wari), Aymara was already being spoken in these areas because of the relationships between the two empires. Maybe Tiwanaku acquired Aymara in the process of development of the state and finally, the Wari empire contributed to the Aymara spread (Itier, 2002).

In the Altiplano, the Aymara would have spread relatively late, during the Late Intermediate from the southernmost Peru, where, by the end of Middle Horizon, there would be a strong Aymara presence due to the Wari expansion. Although a specific motor for an expansion at that time (Later Intermediate) is unclear (Heggarty, 2008), some authors have speculated about it. Cerron-Palomino (2002) suggested the expansion of the group of Aymaraes in the upper basin of the River Pachachaca (Apúrimac) displaced by Southern Quechua speakers. Torero (2002) also proposed stages and regions through which Aymara reached the Altiplano.

In any case, Aymara extended at the expense of other indigenous languages of the Altiplano, where two other families (Pukina and Uru-Chipaya) were widespread enough (Heggarty, 2008).

Although the Incas imposed the Quechua language, after the Spanish conquest Aymara was still spoken in small regions of southern Peru, a fact that suggests that the scenario was a continuum between the two present-day branches (Heggarty, 2008). However, both Incas, and later, Spaniards favoured the Quechua expansion.

I.3.3.3 Other minor languages

In the sixteenth century, the Andean Altiplano housed other linguistic groups like Uru, Pukina, Chipaya, Urukilla, Changos, and Camanchaca. The relationship between these languages is still controversial for linguists; whether they are separate languages or the same language is referred with different names. Uru-Chipaya and Pukina are grouped in the Arawak linguistic family. Some authors consider that if Uru and Pukina were considered the same linguistic group, they could be relicts of a previous “Pukina” civilization (Browman, 1994). According to Torero, the Pukina was the language of Pukara (city and ceremonial centre to the north of Lake Titicaca) that was important before the rise of Tiwanaku around 100 AD, and Uruquilla was the first language of the Tiwanaku people. Other authors propose Pukina as the best candidate to be the Tiwanaku language (Heggarty, 2008).

In any case, in the late-prehispanic period, Uru-Chipaya was spoken in the shores of Lake Titicaca, Lake Poopó and along the river Desaguadero, as well as on the Pacific coast and North Chile. These people fished and foraged and represented about 25% of the Colla region. At present, some one thousand people speak Uru-Chipaya in the shores of Lake Popoó. Until recently, there were also speakers in the Lake Titicaca but they have shifted to Aymara. Pukina is now an extinct language (Kolata, 1993; Browman, 1994; Itier, 2002), but during the Spanish colony, documents attest Pukina speakers in southernmost corner of Peru where some Pukina toponyms are found (Heggarty, 2008). It is important to note that some Andean languages have disappeared during the Spanish colony, Pukina in 1780 (another dating is 1910), even in the 20th century, Mochica in 1940 (Goulder, 2003).

Table 2. Association between languages and cultures (taken from Goulder, 2003).

Hegemony	**Language/ family**	**Main Dissemination**	**Comment**
Chavín-Pachacamac-Chincha-Inca- Spanish	Quechua	500 B.C.- 1940's A.D. (today)	9 lives! 9 golden periods!
Pukara to 1780	Pukina	600 B.C. -100 A.D. (c.1750)	Seminal to Tiwanaku and War. Also in Jesuit/Franciscan missions (See Churajon)
Nazca -Wari	Aymara	400-900 A.D (today)	In some ways now stronger than Quechua, not in numbers, but sense of unity, circuits of capital etc.
Moche	Mochica	0-600 A.D	Last speaker 1940's
Tiwanaku	Uruquilla	400-600 A.D (today)	Still spoken in Chipaya, Bolivia
Shipibo-Conibo	(Pano)	N.A.	Up and downstream from Pucallpa

I.4 Other cultural and environmental aspects of the Andean populations

I.4.1 Altitude environment

When talking about the Andean populations, we inevitably think about the altitude environment. Since the eighteenth century, naturalists and scientists have described how highlanders were more tolerant and better adapted to the hypoxic conditions extant above 3000 m than low-altitude natives who tend to suffer the mountain sickness. The lower rate of oxygen diffusion from air to blood implies several physiological responses in the following processes (the ventilation within lungs, oxygen diffusion, oxygen transport, and diffusion from blood to tissues; Beall, 2007).

A lot of studies have described the morphology and physiology of the Andean Natives. Several traits have been suggested as characteristic of these people, although not all of them are necessarily adaptive characteristics: enlarged chest, increased lung capacities, relatively hypoxia-tolerant VO_{2max}, blunted hypoxic ventilatory response, elevated haematocrit, increased pulmonary diffusion, preferential utilization of carbohydrates as fuel, etc (Rupert and Hochachka, 2001).

It is important to distinguish the process of acclimatization from genetic adaptation. Acclimatising to hypoxia consists on some physiological adjustments involving an increase respiratory rate as well as the heart rate with a faster distribution of the oxygen. In the long exposure to altitude, the production of red blood cells is increased, the oxygen affinity of blood is slightly decreased, and the number of capillaries is increased.

Since this area has been inhabited for more than 10,000 years, and is highly populated, it is reasonable to think that there has been enough time for natural selection to act, and thus, consider Andean people to be genetically adapted. Recent studies have tried to identify specific genes involved in the Altitude adaptation (Stobdan et al., 2008) and detect natural selection (Beall, 2007).

I.4.2 Economical organization

The rural Altiplano economy has been based on two important sectors: herding and agriculture, these two activities showing such an important interdependency that in the literature the most used term is “agropastoral economy”. Each community (*Ayllu*) is engaged in both pastoral and agricultural activities.

Even though at first glance, the windswept, arid plains of the Altiplano seem to be an inhospitable place to agriculture, crop agriculture in the lake basin has been important and highly productive for different kinds of potatoes, native grains like quinoa and cañiwa, and legumes. Since ancient times, terraces were constructed and sophisticated technologies were developed for intensifying agricultural production. Highlanders learned how to make the most of the hard climate conditions (the temperatures drastically vary from warm during the day to freezing at night) for food preservation. The dehydration (freeze-drying) of staple food, potatoes and other Andean tubers, permitted a long-term storage, necessary since their production is seasonal.

Pastoral activities predominate above 4000 m with llama and alpaca herding. This activity has played an important role in the economy of Aymara communities since ancient times. These native camelids provide not only meat, but also important products like wool for textiles, skin for leather, dung for fuel and fertilizer and they are also used for the transportation of their goods (Graffam, 1992).

Additionally, other minor resources have been taken from the lake and rivers (reeds, fish and fowl) and the dramatic ecological changes in close areas have also permitted to have access to other products as well as commercial activities like trade (although nowadays most products are sold, not traded) and wage labour (men have usually migrated seasonally for wage labour).

In fact, the *vertical* economic strategy refers to this distribution of activities depending on the altitude, herding above 4000 m, the potato, other tubers, and quinoa fields over 3000 m, and the maize, coca, and other warm lands crops in regions bellow 2000 m (Kolata, 1993).

I.4.3 Social organization

The social organization of Andean people is variable and has been adapted to special and temporary to political and economical forces. The Altiplano populations are usually grouped in communities, called “Ayllus”, the basic domestic unit. The Ayllu, an endogamous, patrilineal, corporate kin group, composed of one or several extended families (Graffam, 1992). Women tend to marry outside, while the recent married men stay with their parents until the new couple can be established in their own house. Traditionally, the new house is located in a territory offered by the father of the groom, but the increasing alternative economical possibilities have involved a decreasing of this dependent period until reaching the neolocal residence. Usually, the whole family collaborates economically. In weddings, baptisms, or other social events, other kin relationships are established by “compadrazo”. These *compadrazo* links can be horizontal or vertical, and people linked by *compadrazo* cannot get married.

I.5 Human genetic variation

Human diversity is shaped by demographical and biological factors. Darwin was the first, together with Wallace, to inquire this variation and to think which process or processes could be responsible for it. They proposed the mechanism of natural selection as the process of evolution. Until the 20th century, this diversity was defined in descriptive terms, focusing on visible traits, such as body and face morphology, hair features and pigmentation (physical anthropology). At the beginning of the 20th century, the ABO blood group system was discovered, permitting the definition of the first genetic polymorphism (Landsteiner, 1901). Subsequently, other blood-group systems were described and in the middle of the twentieth century, proteins were systematically studied showing differences among human groups.

These first molecular polymorphisms used in the study of human diversity are known as “classical” markers. Classical markers are products of DNA after genetic expression such as blood groups, enzymes and proteins, the human leukocyte antigen (HLA) system and immunoglobulin allotypes (Lewontin, 1972). These polymorphisms were detected by electhrophoretic or immunoprecipitation methods and in the 1960s and 1970s plenty of data on classical genetic markers were available for different human groups revealing human diversity (Mourant et al., 1976).

The work of Watson & Crick proposing the double-helix model of deoxyribonucleic acid (DNA) structure and hereditary mechanism in 1953 represented the birth of modern molecular biology. This new field of biology has experienced a flourishing development in the past 60 years strongly influencing many relative fields. Anthropology is one of the fields most deeply impacted by the theory and method of molecular biology. Thus, the terms *genetic* or *molecular anthropology* are used to designate the subfield that explores human genetic variation.

In the 1980s, new techniques appeared, like Restriction Fragment Length Polymorphism (RFLP) analysis and the Polymerase Chain Reaction (PCR). The detection of variation at a DNA level was available whether or not this variation was expressed, supposing the beginning of the DNA polymorphisms era. During the late 1980s and 1990s, most studies were focused on mtDNA polymorphisms of the Control Region, and Short Tandem Repeats (STRs).

In the early 1990s, the Human Genome Diversity Project (HGDP) was organized to explore human differences by sampling fragments of the genome from a number of populations across the globe. At the beginning of the 21st century, the Human Genome Project had as goal the sequencing of the complete human genome that concluded in 2003 (Collins et al., 2003).

The last decade has been characterized by extraordinary technological developments including automated sequencing techniques, allowing scientists to access human genetic diversity at an unprecedented rate. The DNA chip technology (also called DNA microarray technology) allows the analysis up to 2 million mutations in the genome or survey expression of tens of thousands of genes in one experiment. Finally, today's third generation sequence techniques will beyond any doubt represent a novel, very promising step for the determination of human genome diversity.

Homo sapiens is a relatively young species that presents less intraspecific variation than most of other species that have had more time to accumulate genetic variation. However, the variation among human groups is significant allowing interesting studies focused on demographical history reconstruction. The most common polymorphisms are single nucleotide polymorphisms (SNPs), but also, in recent years, a large amount of structural variation has been detected.

I.5.1 Human population genetics

Population genetics is a field that focuses in the study of populations, not individuals. Hence, it is important to establish a definition of "population". What is a population? There are different answers, depending on the context.

In population genetics a "population" is the ensemble of individuals showing a reproductively unity. All the inhabitants of a population have the same probability to interbreed among them and present less probability to breed with neighbour populations. Geographic and environmental factors imply that the distribution of individuals is not uniform, producing these population unities. In humans, culture is another factor conditioning this organisation of human beings in groups or populations. Therefore, in human population genetics, a population is a group of people of both sexes, of all ages, that share the same territory, interbreed, and share common rules of social behaviour and therefore, present a common genetic patrimony.

Human population genetics studies the genetic differences, specifically, allele frequency distributions between human groups and the mechanisms and processes that generate, reduce, and change them as the result of ancient and/or recent demographic events as well as differences in selective pressures.

I.5.1.1 Evolutionary processes

Population genetic variation is modulated by some factors called *evolutionary forces* driving the evolution of living organisms. The current genetic variation found in humans is the result of four evolutionary forces: mutation, selection, genetic drift and gene flow (or migration). These processes can introduce variation (mutation) while others remove it (selection, genetic drift). The study of these forces is complex and requires mathematical models (Powell, 2005).

Mutation

A mutation is a change in a DNA sequence. Mutations can be caused by mutagenic factors (radiation, viruses, and chemicals) or errors during meiosis or DNA replication. A *somatic* mutation appears in any cell of the body, except the germ cells (sperm and egg). These mutations are important in medicine as they can cause diseases (as for instance, about 95% of all cancers). On the other hand, a *germinal* mutation is the one that appears in the gonads and will be passed to the descendents. Population genetics focuses on germinal mutations.

The mutation process is the only one that generates new variation (*de novo* variation) as new alleles (mutations) appear. It is a random process, independent of the effects and consequences. These new variants will undergo the effect of the other evolutionary forces.

Genetic drift

Genetic drift is the random process by which the frequency of genetic variants (alleles) fluctuates from one generation to the next due to the fact that the number of genomes that pass to the next generation is only a sample from the genomes of the previous generation. The effects of chance are directly related to population size. The probability of losing alleles is higher in small-size populations, generating greater fluctuations and the reduction of the genetic variability over time.

The genetic drift is particularly important in two particular cases: 1) when catastrophes (natural catastrophes, epidemics, wars) cause a drastic reduction of the size of a population, called *bottleneck*, and 2) when a new territory is colonized by a small population, which does not include all the variation of the original population, called *founder effect*.

Selection

Natural selection refers to the differential reproduction of certain genotypes in successive generations. That is, more adapted individuals will present a higher capacity of survival, and therefore, will reproduce more than those not so well adapted. The term *fitness* refers to this capacity of reproduction, more precisely, to the number of offspring that reach sexual maturity, which is based on survival and fertility. The selection can only act on the phenotype, not on the genotype. Selective pressure can be positive or negative, favouring or acting against a certain phenotype, respectively.

Different subcategories of selection are distinguished. Sexual selection occurs when a phenotype is preferred by the opposite sex potential partners, implying higher chances of producing offspring. Ecological selection is the natural selection without the sexual selection.

The patterns of selection (the effect of selection on phenotypes) can be: i) disruptive, when the extreme phenotypes are favoured against intermediate ones, ii) stabilizing, favouring the intermediate characteristics, and iii) directional, favouring one extreme phenotype.

Migration

The movement of individuals from one population to another is known as migration or *gene flow*, although in this last case, migrants must contribute to the next generation. Migration affects the whole genome, which is the inclusion of genes into a population from one or more populations. Migration cannot change allelic frequencies at a species level but can change allele frequencies of populations. In the human species, migration is particularly important due to its complex demographic history.

I.5.2 Genetic markers

Human genetic variation among individuals and/or populations is studied using polymorphisms. A polymorphism refers to the presence of more than one variant in a locus in a population. However, since we tend to reserve the term polymorphism for when the less frequent allele has a frequency ≥1%, we use the more general term *genetic marker*. The minor allele frequency (MAF) in a given population is the frequency at which the rarest allele is present.

I.5.2.1 Brief description of the principal markers

There are different kinds of genetic markers regarding the amount of DNA involved, from a single nucleotide (sequence level) to whole chromosome(s) (cytogenetic level). The most common studied polymorphisms provided by the DNA from sequence to cytogenetic level are:

- **Single Nucleotide Polymorphisms (SNPs)** are polymorphisms due to a base substitution or an insertion/deletion of a single base. They are also called point mutations.
- **Small insertions/deletions (INDELS/DIPS).** Despite the fact that small INDELs (ranging from 1 to 10kb) are highly abundant in humans and cause a great amount of variation in human genes, they have received far less attention than SNPs and larger forms of structural variation (see review Mullaney et al., 2010).

  Different types of INDELS are:

  - **Variable Number of Tandem Repeats (VNTR).** These are genetic markers characterized by the repetition in tandem of 2-6bp (microsatellites) or 10-60bp (minisatellites). The microsatellites, also called Short Tandem Repeats (STRs), have been the most widely used.
  - **Mobile elements** represent a large amount of small INDEL variation in humans. Transposable elements (TEs) are pieces of DNA from 300bp to 10kb that are able to "jump" from one location to another in the chromosome. TEs are very diverse and are classified into two main groups according to the mode of transposition: a) class I: retrotransposons, encode a

reverse transcriptase (RT) and are restricted to eukaryote genomes, and b) class II, DNA transposons, encode a transposase (*Tnp*).

Approximately ~45% (Figure 12) of the human genome can currently be recognized as being derived from TEs, the majority of which are non-long terminal repeat (LTR) retrotransposons, such as LINE-1 (L1), Alu and SVA elements (Cordaux and Batzer, 2009).

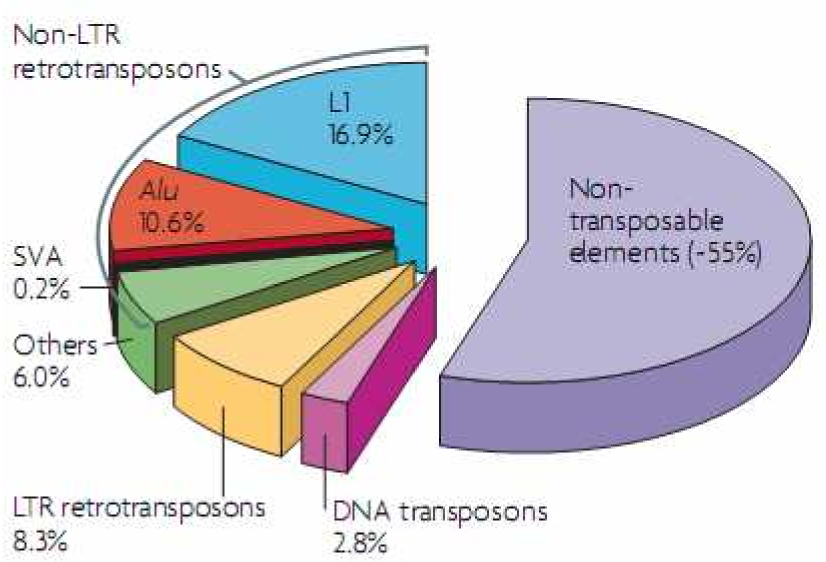


Figure 12. The TE content of the human genome (Cordaux and Batzer, 2009).

- **Structural Variation.** The large-scale genomic variation (>10 kb) includes insertions, deletions, duplications, inversions and rearrangements of DNA fragments of about several kilobases (kb) to megabases (Mb) (1kb-3Mb: submicroscopic, >3Mb: microscopic) that are found in general population and, apparently, have no phenotypic consequences for the carrier, although many of these variants have been associated with diseases. Some authors have suggested that Structural Variants (SVs) might have more impact in the phenotypic variation than SNPs (Korbel, 2007). The main category of SVs is known as Copy Number Variants (CNVs). CNVs are DNA segments of ≥1kb and an average size of 250kb, present at variable copy number in comparison with a reference genome. CNVs represent >15% of the human euchromatic genome and differences between any two genomes can reach up to 15Mb. CNVs are found in humans and other vertebrates. They can vary in gene dosage (Conrad et al., 2010).

Moreover, it is important to mention that different genomic compartments present different features, providing different, although complementary information. Uniparental markers are maternally (mtDNA) and paternally (Y chromosome) inherited and do not recombine. They are useful in detecting ancestry and tracing population movements, but, since they are just one locus, they are subject to stochastic errors. On the other hand, autosomal chromosomes provide multiple ancestry information and provide a more complete picture. Table 3 summarizes their characteristics.

Table 3. Genomic compartments from Cavalli-Sforza and Feldman, (2003).

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

I.5.2.2 Markers used in this work

In the present work different genetic markers have been determined, including: *Alu* insertions, SNPs, STRs, and the uniparental systems, mtDNA and Y chromosome.

Alu insertions

Alu elements are the most abundant short interspersed nuclear elements (SINEs), representing, with approximately 1.1 million of *Alu* copies, more than 10% of the human genome (Carroll et al., 2001) (Figure 12). The *Alu* name of these elements is due to the presence of a recognition site for the restriction enzyme *Alu*I in some members (Houck el al., 1979). Typically, an *Alu* insertion is a 300bp-long sequence ancestrally derived from the 7SL RNA gene inserted into the genome through an intermediate RNA single strand generated by RNA polymerase III transcription (Figure 13) (Batzer and Deininger, 2002) and distributed throughout the genome of primates. These elements appeared some 65MY ago, when primates expanded and diverged (Deininger and Daniels, 1986).

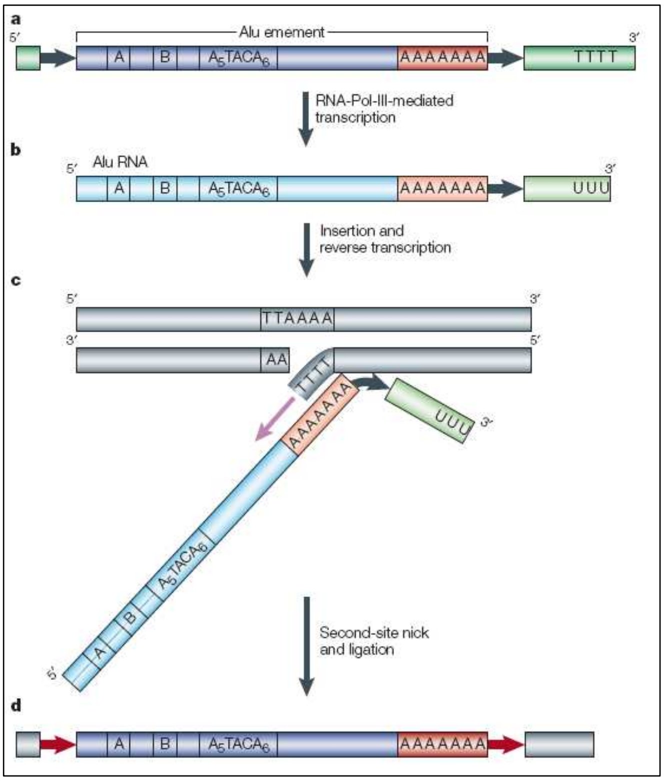


Figure 13. Mechanism of creation of an *Alu* insertion (Batzer and Deininger, 2002).

There is a small group of *Alu* elements called “Master” predisposed to the retroposition (Deininger et al., 1992). During the evolution of these *Alu* elements, mutations have appeared and have permitted the classification of these insertions into 12 subfamilies that have appeared at different times during the evolution of primates (Roy et al., 1999) and are classified into ancient (Jo and Jb), intermediate, and recent groups (Y, Yc1, Yc2, Ya5, Ya5a2, Ya8, Yb8 et Yb9) (Batzer and Deininger, 1991; Carroll et al., 2001; Roy-Engel et al., 2001). Some members of the youngest subfamilies are not yet fixed in humans and consequently are polymorphic for the presence or absence of the insertion (Roy et al., 1999), known as Polymorphic Alu Insertions (PAIs). Among these families, Ya5/8 and Yb8 include a large number of polymorphic insertions; their frequencies vary among populations.

Alu insertions present several advantages that make them very powerful tools as genetic markers for studying human evolutionary history and conduct human population genetic studies (Stoneking et al., 1997; Batzer and Deininger, 2002; Cordaux et al., 2007). PAIs are biallelic (insertion-lack of insertion) and considered as neutral markers. Moreover, two noteworthy features are: 1) the insertion is identical by descent, that is, when individuals share the same state at a locus it is almost certain to have inherited it from a common ancestor (it is practically unlikely that an independent retroposition occur at the same place considering that the rate of insertion and fixation of new *Alu* insertions has been estimated in ~100-200 per Myr (Willard et al., 1987; Britten, 1997) and a complete and precise excision of an element is extremely rare) (Batzer et al., 1994); 2) the ancestral state is known: in almost all cases it is the absence of the insertion. Thus, when comparing genomes, the absence of an element at a locus indicates that the individual carries an ancestral version and there is the possibility to root phylogenetic trees (Batzer et al., 1994)

Single Nucleotide Polymorphisms

SNPs are the simplest type of polymorphisms since only one base is involved. The base substitution can be classed into two groups:

- Transitions, when a pyrimidine base (T and C) is substituted by another pyrimidine or a purine (A and G) is exchanged for another purine.
- Transversions, when a purine is replaced by a pyrimidine or vice versa.

The insertion or deletion (indel) of a single base is also considered a SNP; however, the mechanisms that generate them and their analytical treatment are different from those of base replacement.

A single-base mutation is called polymorphism, thus, a SNP, when the less frequent allele reaches at least a frequency of 1% in the population. Usually, neutral mutations in general populations are called SNPs and the term *mutation* is mainly used when it implies variation in gene function. Also, the label *variant* is used when an allele is below the 1% frequency, although sometimes it is used as a general term for both polymorphism and mutation (Jobling, Hurles and Tyler-Smith, 2004).

Substitution mutations are usually generated by two main processes: misincorporation of nucleotides during replication that occurs at a frequency of about

10^{-9}-10^{-11} per nucleotide, or due to chemical or physical mutagenesis. Globally, mutation rates are between 10^{-8} -10^{-5} (see Jobling et al., 2004). This low rate implies that it is unlikely that at a given position a mutation have recurred or reverted during human evolution. Therefore, SNPs are considered identical by descent markers with some exceptions (Hipervariable Regions (HVI and HVII) of mtDNA).

The fact that SNPs are widely present in the human genome and regular distributed across the whole genome make them a very useful tool for the genetic mapping and association studies. SNPs can be located at coding or non-coding regions. Those located on non-coding regions are assumed to be neutral with no effect on phenotype, but still some of them can modulate the gene expression by altering promoters, enhancers, predisposing to specific disease or phenotype. SNPs located in coding regions can be highly deleterious or with no effect. Substitutions that do not change an amino acid are called *synonymous* or *silent-site* and those changing an amino acid, *non-synonymous,* or *missense* mutations. Substitutions *nonsense* result in a premature stop codon.

The HapMap is a catalogue of common genetic variants found in humans. It describes what these variants are, where they occur in our DNA, and their frequencies in different populations from different parts of the world (www.hapmap.org).

Short Tandem Repeats

Short tandem repeats (STRs, also known as microsatellites) consist of a unit of 1-6bp (called as mono-, di-, tri-, tetra-, penta-, and hexanucleotides) repeated in tandem between 2 and 50 times. They have a typical copy number of 10-30 and are very useful genetic markers. STRs are quite regularly present in the human genome, every 6 to 10kb (Beckman et Weber, 1992), which is a useful property in linkage analysis. The mechanism proposed for their generation is the slippage during the replication (di Rienzo et al., 1994). Although most microsatellites are neutral, some have clearly phenotypic effects.

Mononucleotides are not used in the genetic analysis, whereas dinucleotides (2bp repeats) are the most abundant in human genome and widely used. However, they present the so-called “stutter” problem (Figure 14) and since they mutate more rapidly than tri- and tetranucleotides, sometimes are extremely polymorphic (Jobling et al., 2004).

The mutation rate of STRs (10^{-3}-10^{-4} per locus per generation) is higher than that of SNPs, although with high variability depending on its structure and length (Brinkmann et al., 1998).

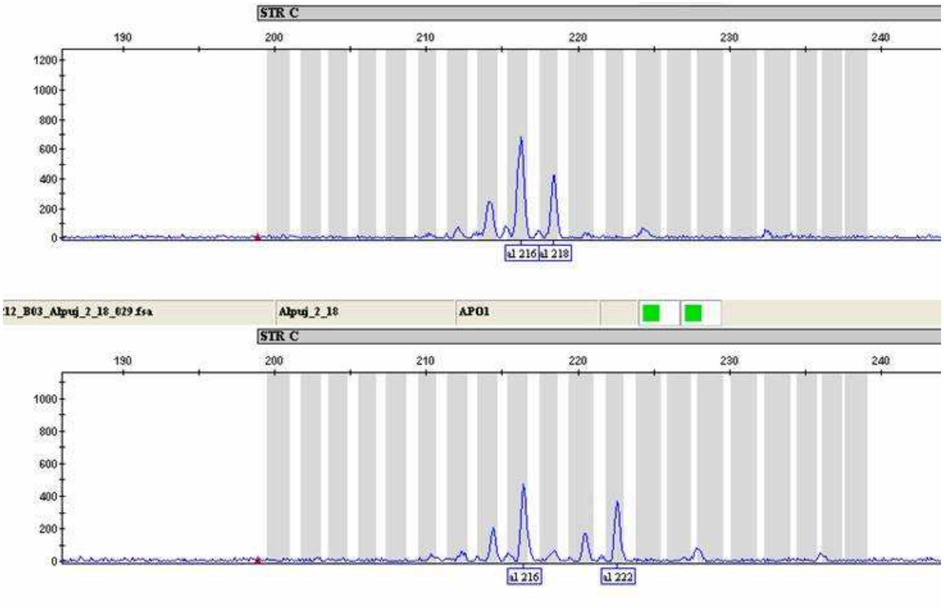


Figure 14. Electrophoregram (PCR product size -pb- and signal intensity) generated by GeneMapper software (Applied Biosystems) displaying the genotype of two individuals for the dinucleotide ss263192881 (ind_1: 216, 218pb alleles; ind_2: 216, 222pb alleles) (3rd manuscript). The stutter effect is visible.

Uniparental markers: mtDNA and Y chromosome

The two most commonly used systems in human population genetic studies in the last twenty or more years have been the mitochondrial DNA (mtDNA) and the non-recombining portion of the Y chromosome (NRY) that are genetic markers presenting a maternal and paternal pattern of inheritance, respectively. These two systems have the advantage of not recombining, thus, the analysis of their sequence variation provides direct haplotypes (combinations of mutations on the same DNA molecule). The definition of haplogroups (haplotypes defined by binary markers) and their sub-branches has allowed the reconstruction of phylogenetic trees, useful for detecting ancestry and tracing female (mtDNA) and male (Y chromosome) population movements separately. Moreover, estimates of the age of these lineages can be calculated through the measurement of the amount of variation accumulated in them and the knowledge of the mutation rates (Schurr and Sherry, 2004; Fagundes et al., 2008).

mtDNA

Mitochondria are organelles present in the cytoplasm of cells in a variable number (about 100-10,000 copies) and their main role is the production of energy. Although most DNA is packaged in chromosomes within the nucleus, mitochondria present a small amount of genetic material (mtDNA), a double-stranded circular DNA molecule that in humans is 16,569bp long, representing a small fraction of the total DNA present in cells (Anderson et al., 1981). MtDNA contains 37 genes, 13 protein-coding genes involved in oxidative phosphorylation, 22 tRNAs and 2 rRNAs, as well as the Control Region of the replication - or D-loop - of 1,121bp (Figure 15). In the control region there is a section between positions 16,001 and 16,540 that evolves faster and is called Hypervariable Region I (HVS-I or HVR-I), as well as the control region II (HVS-II or HVR-II) between positions 61 and 570.

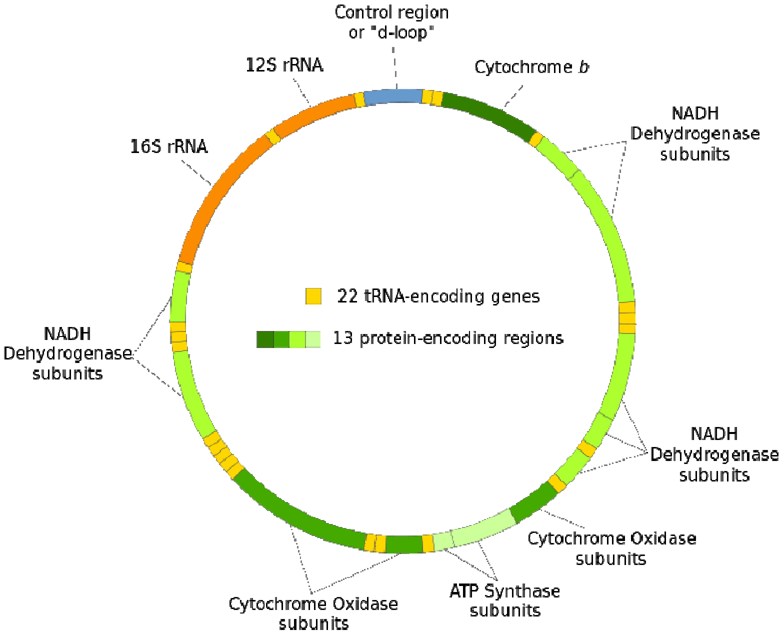


Figure 15. mtDNA diagram.

Mitochondrial DNA possesses several characteristics that make it unique:

a) It is maternally inherited, transmitted from the mother to her children. The cytoplasm provided by the mother's egg cell contains about 250,000 mitochondria in contrast to spermatozoids that contain few mitochondria located in the tail and are lost at the moment of fertilization. Therefore, all mitochondria of the zygote come from the mother's egg (Giles et al., 1980).

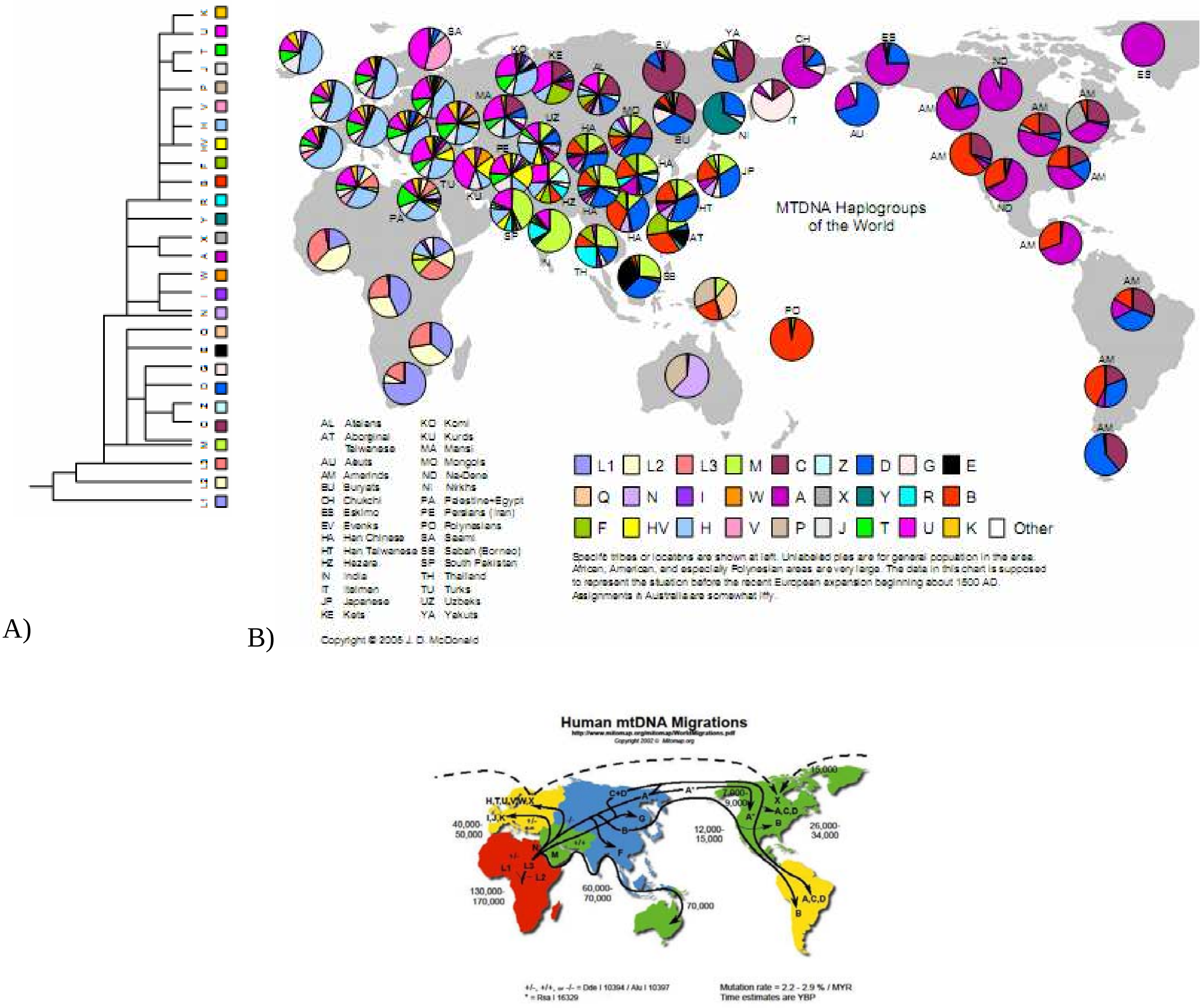
b) It does not undergo genetic recombination. That means that all modern mtDNA sequences descend from a single ancestral molecule at some point in the past and that they differ only by the accumulation of mutations (Merriwether et al., 1991).

c) It presents a higher mutation rate than nuclear DNA (about ten times higher), permitting discrimination even between closely related populations (Brown, George, and Wilson, 1979; Wallace et al., 1987)

d) There are a high number of copies per cell (100-10,000). Thus, fewer samples are required, which is particularly important in forensics and ancient DNA studies (Neckelman et al., 1987).

The exclusively maternal transmission, high mutation rate, lack of recombination and high number of copies make mtDNA an excellent tool for tracing back the history of females.

Mitochondrial haplogroups are defined by polymorphisms present in both the coding and control regions. The first mtDNA studies were performed using RFLPs, providing haplogroup frequencies in the different populations (Torroni et al., 1992, 1993). Then, most studies also sequenced the control region (mainly the HVR-I region), detecting sub-branches within haplogroups, and more recently, complete mtDNA sequences have permitted the construction of a more detailed phylogenetic tree (Oven & Kayser, 2009; Oven, 2010; Stoneking and Delfin, 2010). To date, a large data set on the mtDNA sequence variation in human populations has been accumulated (see www.mitomap.org/MITOMAP). Figure 16 shows the world distribution of mtDNA haplogroups (Fig.16B), a simplified phylogenetic tree (Fig.16A) and world migrations inferred from mtDNA information (Fig.16C).



A) B)

C)

Figure 16. A) Simplified tree of mtDNA haplogroups. B) World distribution of mtDNA haplogroups. C) Human mtDNA migrations (www.mitomap.org).

mtDNA in the Americas

Native American populations present five mtDNA haplogroups (A, B, C, D, and X); the first four are found in all the Americas, whereas haplogroup X, only in North America. The A-D haplogroups are also frequent in Asia, whereas haplogroup X has been suggested to be an ancient Euroasiatic haplogroup. These five haplogroups have also been found in ancient samples. Thus, they are considered to be the founding mtDNA lineages of Native Americans. Their frequency varies depending on the regions, and many groups lack one of those haplogroups that reflect the effects of genetic drift and founder events. Ancestral populations of the Na-Dene and Eskimo-Aleuts may have not possessed all four haplogroups since they present different mtDNA profiles than Amerindians, indicating different population histories (Schurr and Sherry, 2004).

Studies on mtDNA in the last twenty years have supported different models about the peopling of the Americas: some proposed a single and early entry the Americas (Merriwether et al., 1995; Bonatto and Salzano, 1997), whereas others supported two waves of migration, the first one (haplogroups A, C and D) from Central Asia about 26,000-34,000 years ago, and the second and more recent one (haplogroup B) by the Asiatic coast only 12-15,000 years ago (Quintana-Murci et al., 1999) or three waves of migration (Torroni et al., 1993), supporting the Greenberg's model.

Recent studies have provided a deeper knowledge on the genealogy of mtDNA haplogroups, providing a higher level of phylogenetic resolution that has allowed researchers the definition of sub-haplogroups. To date, from the first five haplogroups (A2, B2, C1, D1, X) 15 subhaplogroups have been described (Tamm et al., 2007; Achilli et al., 2008; Perego et al., 2010).

Y-chromosome

The Y-chromosome is one of the smallest human chromosomes of about ~60Mb. It is composed of the pseudoautosomal portion, two regions (PAR1 and PAR2) in the distal extremes, and the euchromatic and heterochromatic regions in between. Most of its length (>90%) corresponds to the non-recombining portion of the Y (NRPY, or NRY), except for the pseudoautosomal portions, which are homologous to the X chromosome. In mammals, it contains the SRY gene, which triggers embryonic male development, as well as other genes needed for normal sperm production.

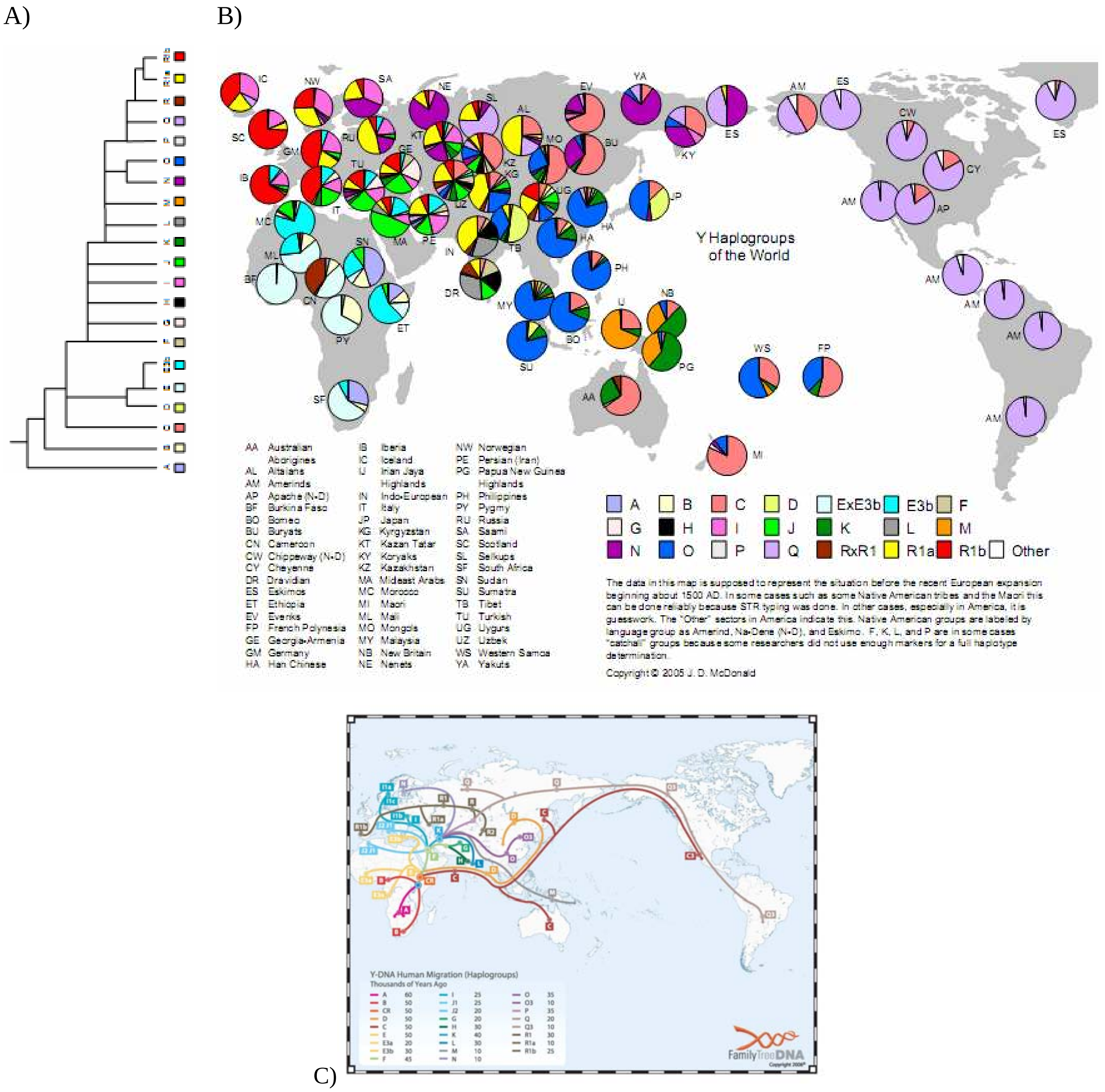
The paternal (from the father to his sons) transmission and the lack of recombination, make the NRY portion a useful tool for tracing the male history of populations. Both SNPs and STRs are used to define paternal lineages. The accumulation of mutations (4 to 8 times higher than in autosomes) derives direct haplotypes. The definition of haplogroups has allowed the reconstruction of male phylogenies. However, the first studies did not use common markers, producing some confusion, and thus, in 2002, the

Y Chromosome Consortium proposed a consensus nomenclature: haplogroups defined by a letter and the SNP that defines it. This consortium published a single parsimony tree showing the relationships among 153 haplogroups on the basis of 243 polymorphisms. Recently, Karafet et al. (2008) published a more complete tree containing 311 distinct haplogroups with the incorporation of 600 binary markers and presented age estimates for 11 of the major clades. Figure 17 shows the world distribution of Y haplogroups, and male migrations.

Y-Chromosome in the Americas

A high proportion of Y-chromosomes from European origin have complicated the task of identifying the native American-specific Y chromosomes. Analysis of both SNPs and STRs has identified two major Native Y-chromosome lineages, the haplogroups Q and C.

The identification of founding lineages within the haplogroups is necessary for the reconstruction of the population history of the Americas. Native American Y haplogroups derive from haplogroups present in Siberia. The two most common founding lineages of Native Americans are Q-M3 (also called Q1a3a) and C-M130 (also C-3b); Q-M3 is distributed in an increasing North-South cline on the New World, reaching frequencies of 100% in some populations and showing significant differences between North and South America according to the Q-M3 haplotype distributions. C-3b is present only in North America. This reduced variability of the Native Y-chromosomes reflects the genetic drift that Native populations have undergone (O’Rourke and Raff, 2010) and has mainly supported a single entry of Native American Y chromosomes into the Americas (Zegura et al., 2004).



The data in this map is supposed to represent the situation before the recent European expansion beginning about 1500 AD. In some cases such as some Native American tribes and the Maori this can be done reliably because STR typing was done. In other cases, especially in America, it is guesswork. The "Other" sectors in America indicate this. Native American groups are labeled by language group as Amerind, Na-Dene (N-D), and Eskimo. F, K, L, and P are in some cases "catchall" groups because some researchers did not use enough markers for a full haplotype determination.

Copyright © 2005 J. D. McDonald

Figure 17. A) Simplified tree of Y-chromosome haplogroups. B) World distribution of Y-chromosome haplogroups. C) Human Y-Chromosome migrations (http://www.kerchner.com/haplogroups-ydna.htm).

I.6 The peopling of the Americas from a multidisciplinary approach

The reconstruction of the biological history of Native American populations has been widely debated in the literature for the last three decades. Several attempts to reconstruct the peopling of the Americas have focused intensely on anthropological, linguistic, archaeological, and more recently, genetic information (Hazout et al., 1993; Dugoujon et al., 1995; Mourrieras et al., 1997).

Although it is widely accepted that the ancestors of Native Americans arrived from Asia via Beringia between 30,000-12,000 BP according to cultural, morphological and genetic similarities between American and Asian populations (Arnaiz-Villena et al., 2010), controversy still persists on the number, routes, and timing of the original migratory waves that moved in and southward into the continent (Wallace and Torroni, 1992; Cavalli-Sforza et al., 1994; Schurr and Sherry, 2004; Wang et al., 2007; Goebel, 2008; Rothhammer and Dillehay, 2009; O’Rouke and Raff, 2010).

I.6.1 Brief revision of the principal models

The first model proposed was the Clovis First/Single Origin model (Hrdlicka, 1937), although it was modified to incorporate new chronological and cultural data (Haynes, 2002). This model was based on archaeological data, but the above-mentioned discovery of archaeological sites dating before Clovis times (older than 11,500 years) has dismissed this hypothesis.

In the late 1980s, Greenberg et al. (1986) proposed a three-wave migration model based on linguistics, dental morphology, and classical genetic markers. This theory suggested that the three waves corresponded to the three linguistic families: Amerind, Na-Dene and Aleut-Eskimo, in this order. The earlier Amerindians entered and colonized the entire New World (12,000 BP), then, the Na-Dene speakers colonized the northwest Pacific coast (8,000 BP); finally, Eskimos occupied the Artic (6,000 BP). Studies on classical markers (Cavalli-Sforza et al., 1994; Parham and Ohta, 1996), and mtDNA (Schurr et al., 1990; Torroni et al., 1993) supported this theory. This hypothesis has been mainly criticized for claiming the existence of a linguistic unity of the Amerind family and has never been widely accepted (Rothhammer and Dillehay, 2009).

A two-migration model was also proposed (Neves & Hubbe, 2005) in the 1990s according to cranial morphology that suggested two successive migrations from different geographic origin and times, a first Paleoamerican migration, and afterwards an Amerindian migration. Paleoamericans tend to present more similarities with present-day Australians, Melanesians, and Sub-Saharan Africans, whereas Amerinds bear more resemblance to northern Asians. Paleoamericans were proposed to have entered the Americas from a non-differentiated Asian population by 15,000 BP and Amerindians by about 11,000 BP from the actual ancestral population to the modern Native Americans, both migrations entering by a terrestrial route. However, analyses controlling the effects of drift revealed that the first Americans are no longer differentiated from modern Amerindians (see Rothhammer and Dillehay, 2009).

I.6.2 The most recent hypothesis

Some recent studies have tackled these issues using an interdisciplinary approach (Goebel et al., 2008; Rothhammer and Dillehay, 2009; O'Rouke and Raff, 2010; Yanng et al., 2010). The most remarkable aspects are briefly mentioned here.

There is general consensus that Asian groups colonised northeast Siberia and parts of Beringia before the last glacial period. These populations probably remained isolated into refugial areas during the last glacial maximum (LGM), where they were genetically differentiated by drift; some alleles and haplotypes were lost, whereas novel ones appeared due to new mutations, often becoming highly frequent due to founder events (Tamm et al., 2007, Perego et al., 2009; Schroeder et al., 2009; Perego et al., 2010). The high presence of Native American private alleles and haplogroups in different genetic systems (mtDNA, Fagundes et al., 2008; Perego et al., 2009; Y-chromosome, Bortolini et al., 2003; Karafet et al., 2008; and autosomals, Schroeder et al., 2009) would support this hypothesis of a period of genetic differentiation in Beringia.

According to some authors, other groups from Beringia or eastern Siberia expanded into North America in the millennia after the initial migration into the Americas, contributing with novel haplogroups of an Asian origin into North America as supported by the presence of some the mtDNA haplogroups only in North America (Perego et al., 2010). This scenario has been also supported by nuclear DNA (Wang et al., 2009) and morphometric (González-José et al., 2008) data. Based on mtDNA haplogroups, the time of entry has been estimated around 15-18 kya for all subhaplogroups (A2, B2, C1b, C1c, D1, D4h3a, and X2) present in the Americas, with the exception for C1d, which has been estimated around 7.6-9.7 thousand years ago (kya) (Perego et al., 2009).

The time range of 18-15 kya includes the LGM (24.000-13.050 BP), a period of time when entry into the Americas through an interior corridor was not possible, therefore, other routes have been proposed. Shurr and Sherry (2004) proposed a Pacific coastal migration (containing the four A, B, C and D mtDNA haplogroups) around 20-15 kya, followed by a second wave (containing haplogroup X) into North America once the ice-free corridor was free. This model would be compatible with by the recent work of Perego et al. (2009).

Fagundes, et al. (2008) suggested a colonization of Eastern Siberia and Beringia prior to the LGM. During the LGM there would have been a drastically reduction of the population size, and between 19 and 15 kya the colonization of the Americas would have taken place via a coastal route. Archaeological data in Siberia supports this scenario (O’Rourke and Raff, 2010).

Mulligan, Kitchen and Miyamoto, (2008) suggested a three-stage model of the peopling of the Americas based on genetic, anthropological and paleoenvironmental information: i) divergence from Asian gene pool, ii) population isolation-occupation of Beringia (between 7,000 and 15,000 year pause), and iii) expansion into the Americas, around 16 kya. The entry from Asia to the Americas, interrupted by a period of isolation and stability, would have been by a population of 1,000-2,000 effective individuals.

A current model based on the Y chromosome postulates the Altai mountain region of Siberia as the starting point of a single, post-LGM migration between 17,2-10,1 kya according to the coalescent age of haplogroups Q and C (Zegura et al., 2004).

Figure 18 shows the three possible routes that have been proposed (O’Rourke and Raff, 2010) and a diagram combining molecular and archaeological data supporting a human dispersion from southern Siberia shortly after the LGM, arriving in the Americas as the Canadian ice sheets receded and the pacific coastal corridor opened, some 15 kya (Goebel et al., 2008).

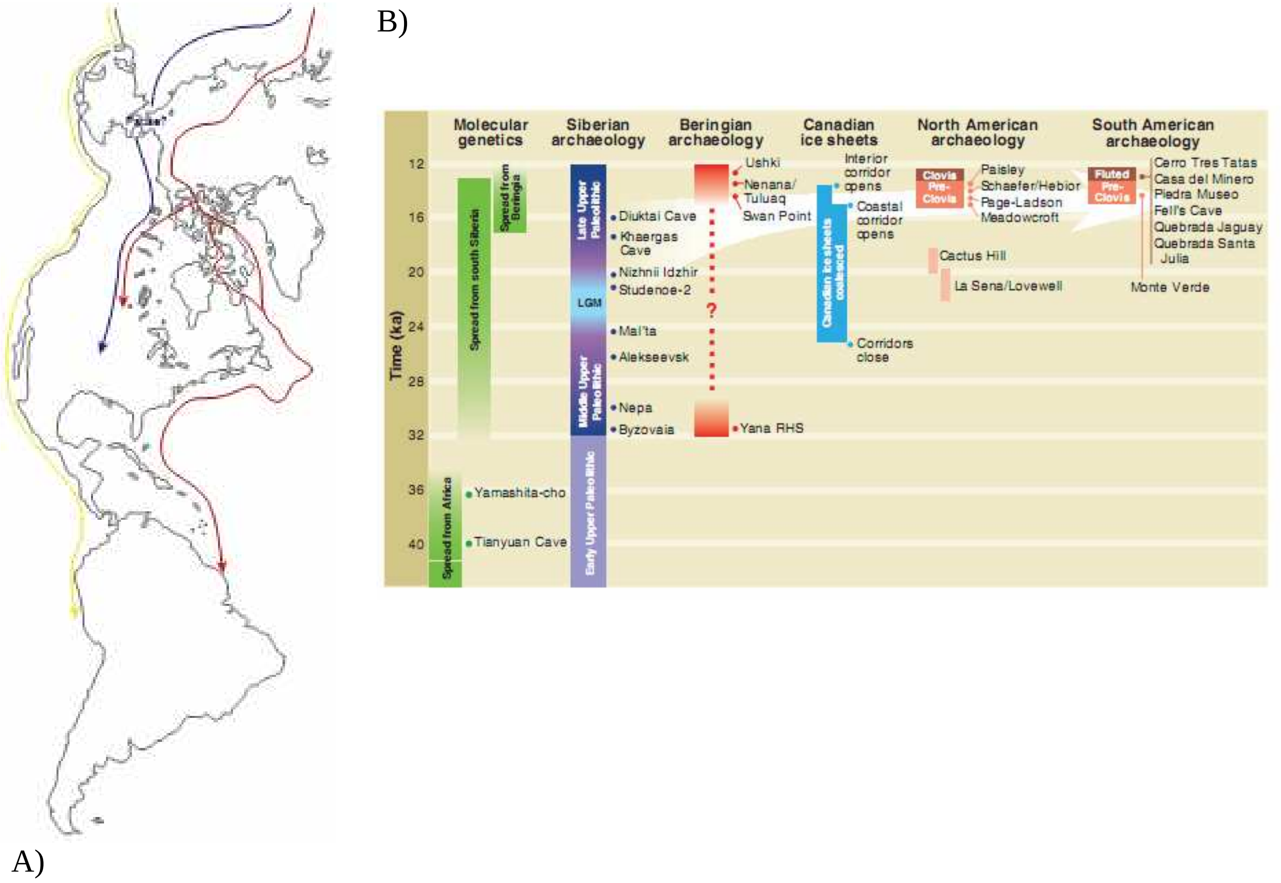


Figure 18. A) The three hypothesise routes for the entry into the Americas (O’Rourke and Raff, 2010). B) Combined molecular and archaeological records (Goebel et al., 2008).

I.6.3 The peopling of South America

The probable entry into South America has been proposed to have occured between 15,000 and 13,500 BP by one migration wave. The fact that eastern populations of South America exhibit lower levels of heterozygosity for different systems suggests an initial colonization of the western part of South America and a subsequent peopling of the eastern area by western subgroups (Rothhammer and Dillehay, 2009). These authors, based on the first model proposed by Bennett and Bird, (1964) propose a possible scenario of the human dispersal into South America. Hunter-gatherers, via the Isthmus of Panama could have entered the Andean Highlands by the Cauca and Magdalena rivers in Colombia. Some groups could also have migrated eastward by way of the Caribbean coast (Venezuela, the Guyanas, and north-east Brazil, and other into

Venezuela, and afterwards arriving into the Amazon basin along the large river systems. A Pacific coast route to Chile was also possible according to several data. From the Andean region of NorthWest Argentina, people could have spread throughout the Pampas and Patagonia (Figure 19).

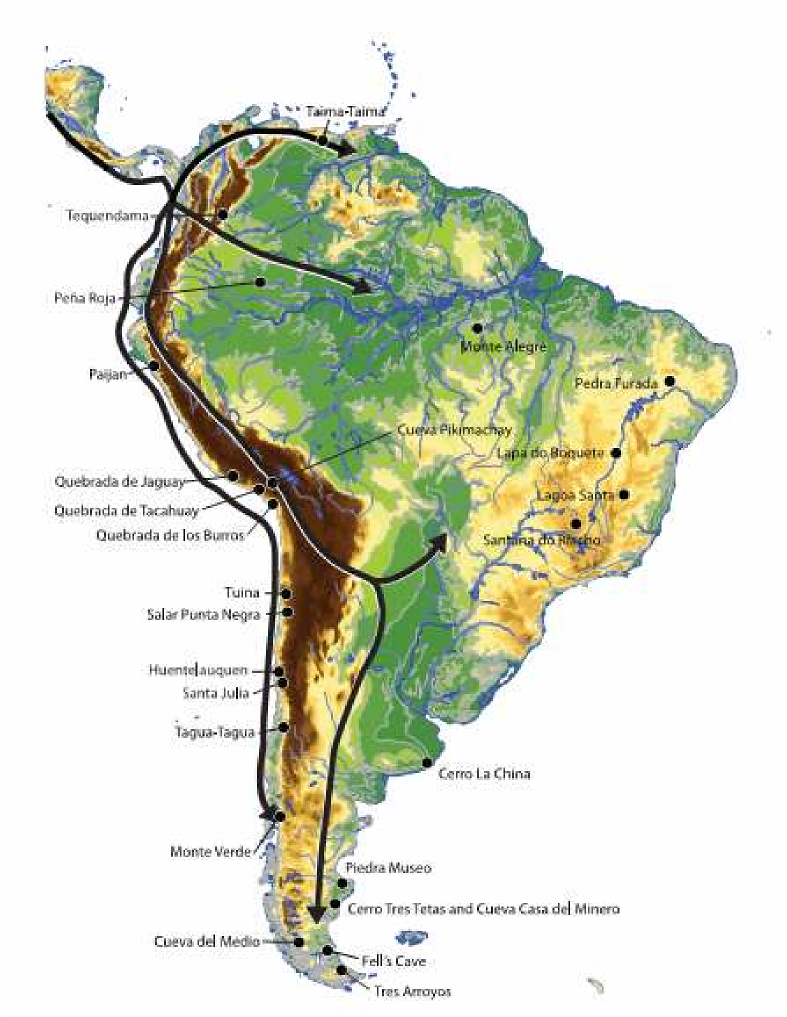


Figure 19. Hypothetical migration routes into South America and some major archaeological sites dating to the late Pleistocene period (Rothhammer and Dillehay, 2009)

According to this model, the Central Andean region would have been accessed from the Northern Andean region. An alternative route would be the access to the Andes from the Amazon, which was first proposed by Lathrap, (1970). The manioc cultivation could have started between 7000 and 9000 BP in the lowlands east of the

Andes. The development of agriculture increased the population density and people had to move. This model was supported on linguistics by the close distance between Andean and Arawak families (grouped under the Andean-equatorial linguistic family). Some of the first studies on genetic markers also suggested that the central Andean groups were related to the tropical forest tribes (Rothhammer and Silva, 1992).

More studies and an interdisciplinary approach is necessary to narrow down the timing and define the processes of the colonization of South America (see Rothhammer et al., 2001; Rothhammer and Dillehay, 2009 for more details).

I.7 Populations studied

This work focuses on two Native American populations from Bolivia belonging to the two main native linguistic families in the Andean region: Aymara and Quechua. Figure 20 shows their location. The collection of the two population samples was carried out by the IBBA (Instituto Boliviano de Biología de Altura) within the framework of a study of physiological adaptations to altitude lead by Prof. Mercedes Villena.

The IBBA has two lines of research: a) Adaptation to high altitude, and b) Human Biodiversity in Bolivia. This institution has done strong efforts in selecting the populations to be studied in order to know the actual effect of hypoxia in highlanders. The human biodiversity in the Bolivian program is applied to the rural area of the Bolivian Altiplano, and integrates biological (physiology, genetics) and social (anthropology, archaeology, history and linguistics) disciplines. Therefore, in addition to biological studies, as the present genetic diversity study, other studies have been carried out such as socio-demographical structures, genealogical, and fecundity and birth-rate patterns (Crognier et al., 2002a, 2002b, 2006).

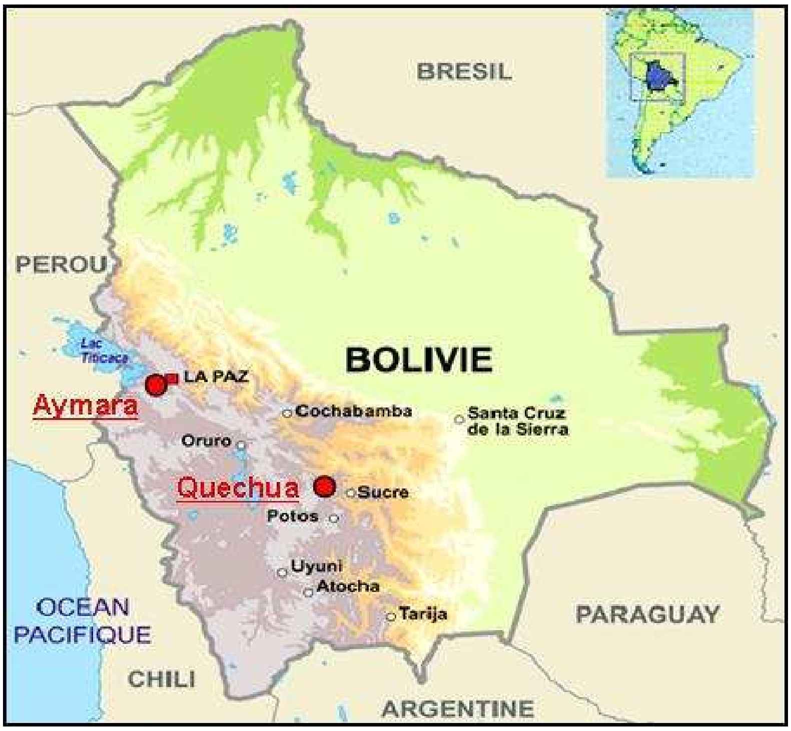


Figure 20. Location of the two studied populations.

I.7.1 The Aymara sample

The Aymara sample studied in this work corresponds to two agricultural communities (Tuni and Amachuma) located 3 km apart. The exchanges between them have been calculated to 25% (E.Crognier, personal communication). These two communities belong to a group comprising a dozen communities (or *Ayllus*) spread over a region of about 50 km^2. This group of peasant communities is located at ~4000 m, approximately 30-60 km southwest to the city of La Paz in the Bolivian Altiplano.

Within the framework of the study of physiological adaptations to altitude, a haematological query (blood extractions) was carried out and anthropometrical measures were also taken. The DNA extraction was done in the IBBA from La Paz from total blood (*buffy coat*) using the phenol-chloroform method.

For this population sample, very complete genealogical records were available thanks to a mission carried out. The reconstruction of genealogies was carried out by Emile Crognier, using the parochial register available that traces back to the XVIII century. From these genealogies, it was found that several families were related, confirming the endogamous characteristic of these populations, as shown in Figure 21. For the different genetic studies, non-related individuals were selected from genealogical records.

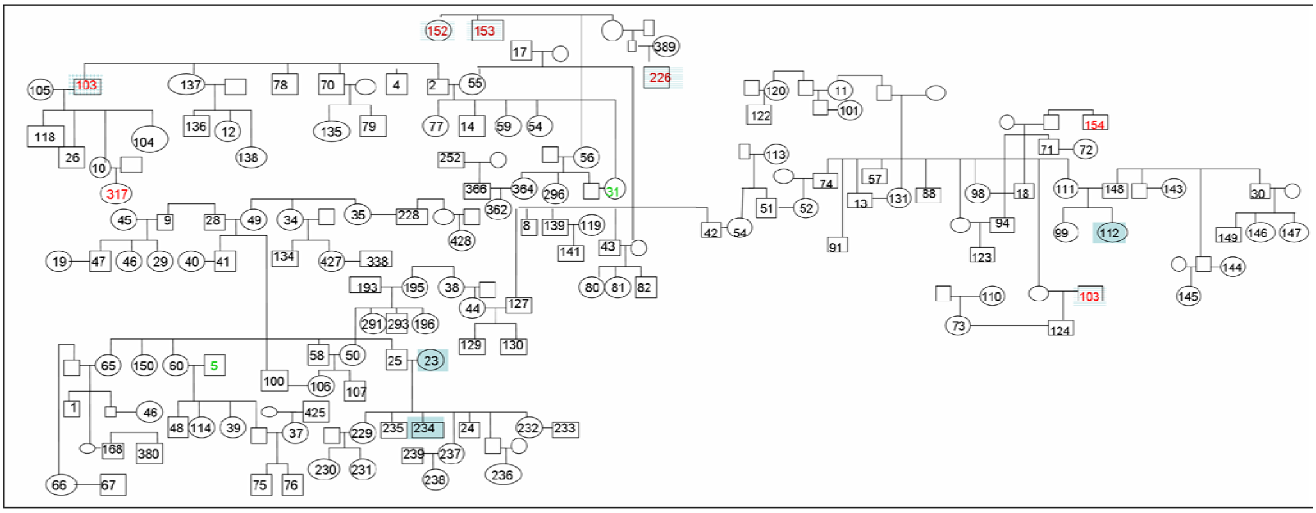


Figure 21. Complex genealogy in the Aymara population.

I.7.2 The Quechua sample

The Quechua sample corresponds to a dozen small agricultural communities (or *Ayllus*) surrounding the Tinguipaya city, in the northern of the Potosí department, at 3200 m.a.s.l.. Before the Inca Empire, this region belonged to the Charka-Qaraqara federation (Cruz, 2006). This sample was also collected within the framework of the study of physiological adaptation to altitude carried out by the IBBA of Potosí. This entity carried out the blood extractions as well as the DNA extractions with the phenol-chloroform method. From the available genealogies, most of them corresponding to nuclear families, non-related individuals were selected for the different studies.

II. OBJECTIVES

This work deals with the genetic characterisation of two Native American human populations based on different kinds of polymorphisms: *Alu* insertions from autosomal chromosomes and the X chromosome, SNPs and STRs from the autosomal region corresponding to the *APOE/C1/C4/C2* gene cluster, and the study of polymorphisms on the mtDNA and Y-chromosome.

The study of the genetic variation found for these markers was designed to address the following issues:

- The first main objective of this work was to characterise these populations for different kinds of genetic markers, since few data are available in Bolivian Andean samples. This genetic characterization will permit to address the following more concrete goals:

  a) Provide the first data on several markers on Native Americans.

  b) Estimate the intra-population variability on these populations.

  c) Estimate non-Native admixture of Aymaras and Quechuas.

  d) Contribute to the knowledge of the genetic variability of the populations from the Andean Altiplano.

- The second main objective was to assess the genetic relationships between these two populations. This global goal will be achieved by addressing the following specific questions:

  e) Are the genetic data in concordance with the linguistic data?

  f) Was the introduction of the Quechua language into Bolivia based on demographic or cultural processes?

  g) Have there been different population histories according to gender?

  h) Do the different markers studied here provide concordant results regarding the genetic relationship between the two Bolivian samples?

- The third main objective was to carry out a comparative genetic analysis in order to:

  i) Assess the position of the Bolivian samples with respect to other Andean samples. Is there genetic homogeneity as it has been suggested? Can we distinguish any genetic pattern according to ancient cultural data?

  j) Contribute with new data to the problem of the possible relationships/differences between the two main geographical areas, the Western (Andean) and the Eastern (Amazonian).

III. RESULTS

III.1 Supervisor's report on the quality of the published articles

The doctoral thesis "**Genetic characteristics of the two main native groups in Bolivia: Aymaras and Quechuas"** is based on the results obtained by Magdalena Gayà-Vidal and presented in three articles, two of them are already published in international peer-reviewed journals and the third one is ready for submission.

The importance of the obtained results is demonstrated by the quality of the two journals:

- *American Journal of Human Biology* is the official journal of the *Human Biology Association*. It is a journal indexed in the SCI and SSCI with an impact factor of 2.021 and classified in the first positions of the first quartile of the "Anthropology" field (position 6/75) and in the second quartile of "Biology" field (position 28/85).

- *American Journal of Physical Anthropology* is the official journal of the *American Association of Physical Anthropologists*. It is a journal indexed in the SCI and SSCI with an impact factor of 2.693 and classified in the first positions of the first quartile of the "Anthropology" field (position 4/75) and in the second quartile of "Evolutionary Biology" field (position 21/45).

Signed by Dr. Pedro Moral Castrillo and Dr. Jean-Michel Dugoujon
Barcelona, 26 Septembre 2011

Original Research Article

Autosomal and X Chromosome *Alu* Insertions in Bolivian Aymaras and Quechuas: Two Languages and One Genetic Pool

MAGDALENA GAYÀ-VIDAL,[1,2] JEAN-MICHEL DUGOUJON,[2] ESTHER ESTEBAN,[1] GEORGIOS ATHANASIADIS,[1] ARMANDO RODRÍGUEZ,[3] MERCEDES VILLENA,[3] RENÉ VASQUEZ,[4] AND PEDRO MORAL[1]*

[1]*Unitat d'Antropologia. Biologia Animal, Universitat de Barcelona, Barcelona, Spain*
[2]*Laboratoire d'Anthropologie Moléculaire et Imagerie de Synthèse, FRE 2960 CNRS et Université de Toulouse, Toulouse, France*
[3]*Instituto Boliviano de Biología de Altura (IBBA), La Paz, Bolivia*
[4]*Instituto Boliviano de Biología de Altura (IBBA), Potosí, Bolivia*

ABSTRACT Thirty-two polymorphic *Alu* insertions (18 autosomal and 14 from the X chromosome) were studied in 192 individuals from two Amerindian populations of the Bolivian Altiplano (Aymara and Quechua speakers: the two main Andean linguistic groups), to provide relevant information about their genetic relationships and demographic processes. The main objective was to determine from genetic data whether the expansion of the Quechua language into Bolivia could be associated with demographic (Inca migration of Quechua-speakers from Peru into Bolivia) or cultural (language imposition by the Inca Empire) processes. Allele frequencies were used to assess the genetic relationships between these two linguistic groups. Our results indicated that the two Bolivian samples showed a high genetic similarity for both sets of markers and were clearly differentiated from the two Peruvian Quechua samples available in the literature. Additionally, our data were compared with the available literature to determine the genetic and linguistic structure, and East–West differentiation in South America. The close genetic relationship between the two Bolivian samples and their differentiation from the Quechua-speakers from Peru suggests that the Quechua language expansion in Bolivia took place without any important demographic contribution. Moreover, no clear geographical or linguistic structure was found for the *Alu* variation among South Amerindians. Am. J. Hum. Biol. 22:154–162, 2010. © 2009 Wiley-Liss, Inc.

The Quechuas and the Aymaras are the two main Amerindian linguistic groups inhabiting the Andean Altiplano in Bolivia, an area where genetic studies have mainly focused on mtDNA (Corella et al., 2007; Sandoval et al., 2004). In a wider geographical context, most of the available genetic data on Andean populations derive from studies of uniparental markers, with small-sized samples, from populations geographically restricted to modern Peru (Fuselli et al., 2003; Lewis et al., 2007). The present study is focused on the genetic variability of these two main linguistic groups, through the analysis of significant-sized samples, to provide new autosomal data with a wide set of independent loci (32 *Alu* loci).

Archaeological and historical records suggest that modern Bolivian populations are the result of historic complex interactions among people of different languages and cultures. Most data point to the Central Andes (Bolivian Altiplano and Peru) as the heartland of the first complex societies of South America. It is commonly accepted that important civilizations/states such as Chavin (900–200 BC), Tiwanaku (100 BC–1200 AD), and Huari (700–1200 AD), existed before the establishment of the Inca Empire, which was conquered by the Spaniards around 1532 AD (Stanish, 2001). Specifically, in the South Central Andes, the Tiwanaku civilization, which originated in the Titicaca basin (in the Altiplano at 3,600 m a.s.l), extended its influence from Southern Peru to current Bolivia, Northern and Central Chile and North-Western Argentina, (Kolata, 1993). After the Tiwanaku collapse, the state fragmented into a number of Aymara polities or "*Señorios*" (Qolla, Lupaqa, Pakaq, Caranga, etc; see Bouysse-Cassagne, 1986) that persisted until their conquest by the Inca Empire (1300–1532 AD). From Cuzco, the Incas expanded its power towards the North and South using strategies such as language imposition (Quechua) and the *mitma* system (a deliberate movement of whole tribes from region to region around their vast Empire).

Linguistically in the Andes, two main Amerindian languages of the Andean subfamily (Greenberg, 1987), the Quechua (12 million speakers in Ecuador, Peru, Southern Bolivia, and Northern Chile), and the Aymara (1.5 million speakers mainly in Bolivia), are spoken along with other minor languages, such as Uru-Chipaya which is spoken around the shores of Lake Titicaca and Lake Poopó. It is important to note that this linguistic distribution seems to be relatively recent. Before the Inca period it is likely that an ancestral form of Quechua (technically referred to as proto-Quechua) was spoken in the Huari distribution area (around current Ayacucho), whereas a proto-Aymara, Pukina, and Uru were probably spoken in the influence area of the Tiwanaku civilization (Browman, 1994; Kolata, 1993; Stanish, 2001). Afterwards, the Incas spread the Quechua tongue and imposed it as the official language of the empire, which was subsequently promoted by the Spaniards as *lingua franca* (Rowe, 1963).

To gain new insights into the relationships between the two main Amerindian linguistic groups in Bolivia and the

Contract grant sponsor: Ministerio de Ciencia y Tecnología; Contract grant numbers: CGL2005-3391, CGL2008-03955; Contract grant sponsor: Generalitat de Catalunya; Contract grant number: 2005-SGR00252; Contract grant sponsor: Departament d'Educació i Universitats de la Generalitat de Catalunya i el Fons Social Europeu; Contract grant number: 2007_FI_B00108; Contract grant sponsors: "La Caixa", The French Government

*Correspondence to: Prof. Pedro Moral, Ph.D., Department of Biologia Animal-Antropologia. Fac. Biologia. Universitat de Barcelona, Av. Diagonal, 645. 08028 Barcelona, Spain. E-mail: pmoral@ub.edu

Received 25 March 2009; Revision received 5 May 2009; Accepted 30 May 2009

DOI 10.1002/ajhb.20967

Published online 10 July 2009 in Wiley InterScience (www.interscience.wiley.com).

demographic processes that may have affected these relationships, this study deals with the genetic variability of Aymara-speakers from the Titicaca Lake region and Quechua-speakers from the Northern Potosí department. The Bolivian Quechua population corresponds to the ancient Charaqara region, which was Aymara-speaking before the Inca expansion (Tschopik, 1963). Today, the Tomas Frias province, from where this sample originates, is 98% Quechua-speaker (Fabre, 2005). The aim is to assess the relative importance of the demographic and cultural processes of the Quechua expansion in Bolivia. Different studies have demonstrated that in many cases the use of a language by a population is a sign of genetic identity, but, in other cases it may simply be a cultural trait imposed by a political or economical power without any substantial effect on the genetic structure of the population itself (Belle and Barbujani, 2007; Cavalli-Sforza et al., 1992; Moral et al., 1994). The comparative analysis of our own data with other available data on some Peruvian Quechua-speaking populations, will allow us to determine the demographic implications of the movement of Quechua-speaking groups under the mitma system, under the assumption that Aymara-speakers were the main original inhabitants of the Bolivian Potosi department according to historical sources.

Our hypothesis is that if the Inca mitma system was an effective demographic process for the Andean southward expansion of the Quechua language, it might be recognizable through genetic similarity/difference patterns between current Bolivian Quechua-speakers and other northward Quechua populations from Peru. That is, if the mitma system was effective we can expect greater genetic differences between the two Bolivian groups than between Bolivian and Peruvian Quechua-speakers. The alternative scenario would predict an opposite pattern of genetic similarities. Additionally, our analysis will allow the genetic characterization of Bolivian populations in relation to other Native Americans, and provide new autosomal data to address general issues concerning the South American populations: the genetic East to West (or Amazon vs. Andes) differentiation suggested by some previous surveys (Lewis and Long, 2008; Tarazona-Santos et al., 2001); and the general correspondence between genetics and linguistics in this region (Hunley et al., 2007).

Population genetic analyses were performed using 32 Polymorphic *Alu* Insertions (also known as PAIs), the most abundant short interspersed nuclear elements (SINEs), representing more than 10% of the human genome (Carroll et al., 2001). Typically, an *Alu* insertion is a 300-bp-long sequence ancestrally derived from the 7SL RNA gene inserted into the genome through an intermediate RNA single strand generated by RNA polymerase III transcription. On the basis of the evolution changes into the original genes, these elements are grouped into subfamilies. Some members of the youngest subfamilies are not yet fixed in all human populations and consequently are polymorphic for the presence or absence of the insertion (Roy et al., 1999). These markers present two noteworthy features: (1) the insertion is identical by descent, and (2) the ancestral state is known. These characteristics make the *Alu* insertions a useful group of markers in the study of human population genetics (Cordaux et al., 2007; Resano et al., 2007).

As far as we know, previous data on *Alu* insertions variation in Native Americans range from a few loci (Antunez de Mayolo et al., 2002; Dornelles et al., 2004; Mateus-Pereira et al., 2005; Novick et al., 1998) to 12 loci (Battilana et al., 2006), that have been generally analyzed in population samples of quite small sizes. In the particular case of Andean populations, only two Peruvian Quechua groups have been previously tested (Battilana et al., 2006) but no *Alu* data are available on Aymaran populations. So, in relation to previous studies in the literature, this article represents: (i) the first *Alu* polymorphic survey carried out on Aymara populations, (ii) the first data on 14 X chromosome *Alu* polymorphic elements in Native Americans, and (iii) the first data on 8 out of the 18 autosomal PAI tested in this study in South Americans.

MATERIALS AND METHODS

Population samples

A total of 192 unrelated subjects originating from two linguistically different regions of Bolivia (96 from each population) were analyzed. These subjects were selected from a whole set of 686 individual samples according to available genealogical records. Blood samples were obtained with informed consent under the framework of the High-Altitude Adaptability Project of the IBBA (Instituto Boliviano de Biología de Altura) and with approval from the Ethical Committee of this institution. As an indicator of potential non-Native admixture, an analysis of the GM haplotypes showed around 1% of the specific European haplotype GM5*;3 (Dugoujon JM, personal communication). Also, the two samples analyzed here presented a frequency of 98% of the O group (ABO system).

The geographical location of the two samples studied is shown in Figure 1. The two population samples live in the Central Andes, in the Bolivian Altiplano. The Aymaran sample comes from two agricultural communities or "*Ayllus*" (an endogamous, patrilineal, corporate kin group), the Tuni and Amachuma, which are located 3-km apart between La Paz and the Titicaca Lake and show admixture of 25% (personal communication from pedigree data in Crognier et al., 2002). The Quechua-speaking subjects are inhabitants of rural areas from 13 ayllus near the Tinguipaya city, in the Potosi department.

Genotype determinations

DNA extracted from blood by classical phenol–chloroform method was used for PCR genotype determinations. Eighteen human-specific autosomal *Alu* polymorphic elements (ACE, APOA1, A25, B65, CD4, DM, D1, FXIIIB, PV92, TPA25, HS2.43, HS4.32, HS4.69, Sb19.12, Sb19.3, Yb8NBC120, Yb8NBC125, and Ya5NBC221) were genotyped using the primers and PCR conditions described with minor modifications (Gonzalez-Perez et al., 2003). From these 18 polymorphism, 8 PAIs (ACE, APOA1, FXIIIB, PV92, TPA25, D1, A25, and HS4.32) were selected to provide the best comparative data set regarding the published literature, whereas the remaining ones were included due to their discriminative power among local populations previously shown by other studies (Gonzalez-Perez et al., 2003, 2006; Resano et al., 2007). Additionally, 14 X chromosome *Alu* insertions (Ya5DP62, Ya5DP57, Yb8DP49, Ya5a2DP1, Yb8DP2, Ya5DP3, Ya5NBC37, Yd3JX437, Yb8NBC634, Ya5DP77, Ya5NBC491, Yb8NBC578, Ya5DP4, Ya5DP13) were also determined in each sample by using the primers and PCR

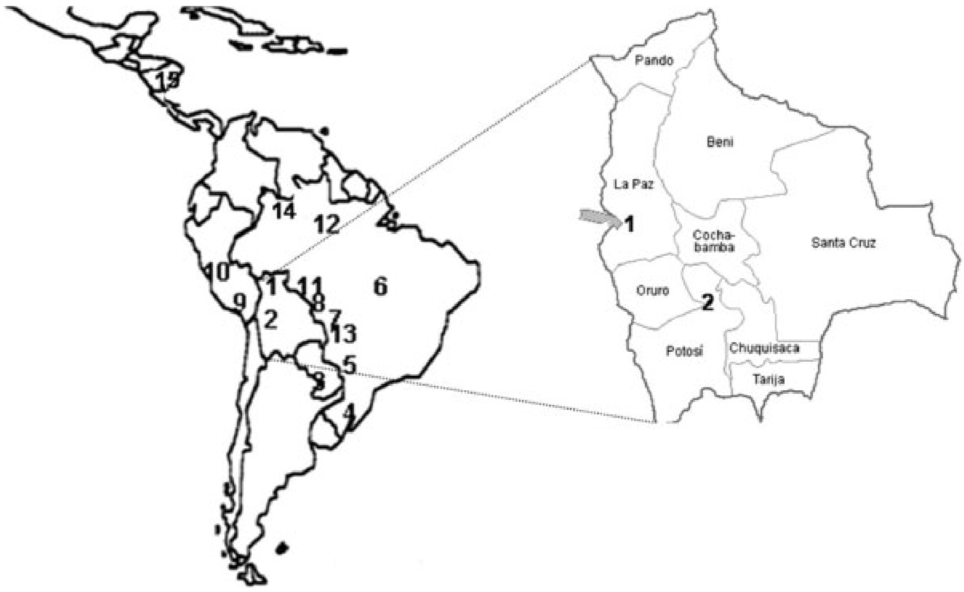


Fig. 1. Geographic location of the populations included into the analyses. 1: Aymara, 2: Quechua "Tin" (from Tinguipaya), 3: Aché, 4: Caingang, 5: Guaraní, 6: Xavante, 7: Cinta Larga, 8: Gavião, 9: Quechua "A" (from Arequipa), 10: Quechua "Tay" (from Tayacaja), 11: Surui, 12: Waiwai, 13: Zoró, 14: Yanomami, 15: Maya.

conditions according to Callinan et al. (2003), with minor modifications (Athanasiadis et al., 2007). Phenotypes were identified by electrophoresis of the PCR products, followed by ethidium bromide staining and observation under UV fluorescence. Positive and negative controls were used in all the PCR runs to assess the quality of the determinations.

Statistical analyses

Allele frequencies were computed by direct counting and Hardy-Weinberg equilibrium was tested by an exact test using the Genepop program (Raymond and Rousset, 1995) to assess data quality. Unbiased estimates of heterozygosity and its average across loci and populations were calculated according to the Nei's formula (Nei, 1978). For the X chromosome PAIs, H-W equilibrium and gene diversities were calculated from female genotype frequencies. The Bonferroni correction was applied in all analyses.

As a first approach to the genetic differentiation between the two Bolivian populations, an exact test based on the allele frequencies of all 32 individual loci was performed using Arlequin statistical package (Schneider et al., 2000).

For comparative purposes, *Alu* frequency data on 13 Native American populations were collected from the literature. These included 12 South Amerindian groups (Aché, Caingang, Guaraní, Cinta Larga, Gavião, Wai Wai, Xavante, Zoró, Quechua "A" from Arequipa, Quechua "Tay" from Tayacaja, Suruí, and Yanomami), and one Central-American population (Maya), whose geographical location is indicated in Figure 1. No North American samples were included in the comparisons due to their low number, high level of admixture, and their geographical irrelevance with our main hypothesis. For all these populations, data were available for 8 out of the 32 loci examined in the present study (Battilana et al., 2006). Using the joint variation in these loci, pairwise population relationships were determined by the analysis of the genetic distances using the Reynolds coefficient (Reynolds et al., 1983) for all the Amerindian populations available. These distance estimates were used (i) to quantify the genetic relationships between the two Bolivian linguistic groups in the framework of the relationships among other Native Americans, (ii) to compare the degree of genetic differentiation between the two Bolivian samples and other similar linguistic groups (Quechua) from Peru to obtain indirect evidence about the demographic impact of the Quechua expansion into Bolivia, and (iii) to approach the between-population variation in different geographical South American population groups (West vs. East). The distance relationships were depicted by a neighbor-joining tree (Saitou and Nei, 1987) and displayed in a Multi-Dimensional Scaling (MDS) graph. The reliability of the tree was tested by bootstrap resampling analysis (1,000 iterations).

The amount of genetic diversity in all Amerindian samples and in different sample groups according to geographical (West and East in South America) and linguistic criteria (Quechua-speakers) was assessed by the analysis of the molecular variance (AMOVA) of the allele frequencies using the Arlequin software (Schneider et al., 2000). Finally, the possible structuring of the genetic diversity in South America according to geography and linguistics was checked by hierarchical AMOVA analyses to test the potential general geography-genetics and linguistics-genetics correlations in South Native Americans.

RESULTS

Allele frequency distributions in Bolivia

Alu insertion frequencies for the 18 autosomal loci in the two Andean populations, the Aymaras and Quechuas from Bolivia, are shown in Table 1. From the 18 loci, 11 were polymorphic in both populations. Four loci were

TABLE 1. Autosomal PAI frequencies and gene diversities in two Bolivian samples and diversity range in other South American Native Populations (Battilana et al., 2006)

Autosomal PAIs	No. of chromosomes		Freq. insertion		Unbiased heterozygosity		Range *H* in Native S. Amer.
Populations	Aymara	Quechua	Aymara	Quechua	Aymara	Quechua	
ACE	170	188	0.853	0.809	0.252	0.311	0.000–0.476
HS4.32	186	186	0.473	0.419	0.501	0.490	0.185–0.470
FXIIIB	190	188	0.942	0.968	0.110	0.062	0.000–0.417
A25	192	182	0.094	0.104	0.171	0.188	0.000–0.365
D1	190	186	0.584	0.505	0.488	0.503	0.403–0.507
TPA 25	190	190	0.679	0.712	0.438	0.413	0.205–0.503
PV92	192	180	0.865	0.917	0.235	0.154	0.075–0.507
Yb8NBC120	192	184	0.630	0.582	0.469	0.489	
Sb19.3	192	184	0.635	0.636	0.466	0.466	
Yb8NBC125	188	184	0.011	0	0.021	0	
B65	192	182	0.219	0.374	0.344	0.471	
DM	192	184	0.031	0.044	0.061	0.084	
Ya5NBC221	192	178	1	1	0	0	
APOA1	192	172	1	1	0	0	0.000–0.131
Sb19.12	188	168	0	0	0	0	
CD4	192	180	1	1	0	0	
HS2.43	190	180	0	0.006	0	0.011	
HS4.69	192	164	0	0.018	0	0.036	
Average	–	–	–	–	0.198	0.204	

TABLE 2. X chromosome PAI frequencies, and gene diversities in the two studied populations and the range of Heterozygosities in other World Populations (Callinan et al., 2003, Athanasiadis et al., 2007)

X chromosome PAIs	No. of chromosomes		Freq. insertion		Unbiased heterozygosity		Range *H* in world populations
Populations	Aymara	Quechua	Aymara	Quechua	Aymara	Quechua	
Ya5NBC37	148	138	0.088	0.029	0.153	0.039	0.160–0.520
Ya5a2DP1	146	142	0.863	0.894	0.196	0.148	0.180–0.470
Yb8DP2	124	134	0.121	0.149	0.209	0.300	0.230–0.430
Yd3JX 437	152	139	0.520	0.554	0.503	0.481	0.080–0.500
Ya5DP62	127	100	0.921	0.860	0.172	0.132	0.080–0.430
Ya5DP77	137	109	0.628	0.624	0.478	0.390	0.000–0.500
Ya5DP57	42	140	1	0.993	0	0.019	0.060–0.410
Yb8DP49	147	144	1	0.986	0	0.037	0.080–0.380
Yb8NBC634	134	139	1	1	0	0	0.000–0.260
Ya5NBC491	131	133	1	1	0	0	0.000–0.500
Yb8NBC578	134	147	1	1	0	0	0.000–0.480
Ya5DP13	142	137	1	1	0	0	0.000–0.080
Ya5DP3	151	147	0	0.007	0	0.018	0.180–0.500
Ya5DP4	140	117	0	0	0	0	0.000–0.280
Average	–	–	–	–	0.122	0.113	0.075–0.377

monomorphic for either the *Alu* presence (Ya5NBC221, APOA1, and CD4) or absence (Sb19.12) of the insertion in the two samples. The absence of the *Alu* element was fixed for HS2.43 and HS4.69 in Aymaras and for Yb8NBC125 in Quechuas. All the polymorphic *Alu* frequency distributions (12 in Aymaras and 13 in Quechuas) fit the Hardy-Weinberg equilibrium. The autosomal *Alu* elements showing the highest gene diversities were HS4.32, D1, TPA25, Yb8NBC120, Sb19.3, and B65 (Table 1). The average heterozygosity for the 18 autosomal loci was $\sim$0.2 (Aymaras: 0.198 and Quechuas: 0.204).

The *Alu* insertion frequencies for the 14 X chromosome loci are displayed in Table 2. Six of them were polymorphic (Ya5NBC37, Ya5a2DP1, Yb8DP2, Yd3JX437, Ya5DP62, Ya5DP77) in both populations. Four were monomorphic for the insertion (Yb8NBC634, Yb8NBC578, Ya5DP13, and Ya5NBC491) and one for the absence (Ya5DP4) in the two samples. Finally, both the insertion for Ya5DP57, Yb8DP49 *Alu* and the absence for Ya5DP3 were fixed in Aymaras. Tests in female samples indicated that most of the observed distributions agree with the H-W equilibrium conditions. Only Ya5DP62 genotype distribution was significant ($P = 0.002$) after Bonferroni correction in the Aymaran population sample.

Two X chromosome *Alu* elements (Yd3JX437, Ya5DP77) showed a high heterozygosity compared with previous data (Athanasiadis et al., 2007; Callinan et al., 2003). The remaining markers examined in the two Andean samples exhibited diversity values that lie close to the lowest values worldwide (Table 2). The average gene diversities (Aymaras: 0.122, Quechuas: 0.113) for the X chromosome PAIs in these Amerindian populations were also low according to the heterozygosity range found in other populations.

The exact test of differentiation between the two populations showed no significant difference for any locus distribution.

Genetic comparisons with other Native Americans

The frequency distribution of eight PAIs (Table 3) was used to estimate the genetic relationships through Rey-

TABLE 3. Population size, allele frequency distribution for the 8 loci, average heterozygosity (considering the 8 loci) and linguistic affiliations of the 15 populations considered

Populations[a]	Ling[b]	*n*	Average *H* (8 loci)	ACE+	APO+	TPA 25+	FXIIIB+	PV92+	A25+	HS4.32+	D1+
1. Aymara (1)	AND	96	0.274	0.853	1	0.679	0.942	0.865	0.094	0.473	0.584
2. Quechua Tin (1)	AND	96	0.265	0.808	1	0.711	0.968	0.917	0.104	0.419	0.505
3. Aché (2)	ET	31–76	0.206	1	1	0.866	0.782	0.855	0.013	0.198	0.581
4. Caingang (2)	GPC	40	0.297	0.543	0.963	0.675	0.872	0.793	0.037	0.250	0.706
5. Guarani (2)	ET	34	0.268	0.829	0.941	0.710	0.935	0.783	0.097	0.130	0.394
6. Xavante (2)	GPC	33	0.305	0.683	1	0.417	1	0.813	0.234	0.242	0.532
7. Cinta larga (3)	ET	25	0.268	0.820	0.960	0.438	0.938	0.538	0.000	0.125	0.283
8. Gavião (3)	ET	28	0.177	0.926	1	0.793	1	0.897	0.000	0.154	0.589
9. Quechua A (3)	AND	21	0.297	0.826	1	0.643	1	0.696	0.136	0.357	0.361
10.Quechua Tay (3)	AND	22	0.312	0.630	0.978	0.714	0.891	0.739	0.043	0.300	0.675
11. Surui (3)	ET	23	0.220	0.870	1	0.409	1	0.938	0.083	0.214	0.286
12. Waiwai (3)	GPC	22	0.205	0.976	1	0.778	0.870	0.870	0.043	0.100	0.543
13. Zoro (3)	ET	28	0.212	0.962	0.983	0.692	1	0.833	0.018	0.217	0.583
14.Yanomami (3)	CP	21	0.216	0.750	1	0.685	1	0.962	0.000	0.241	0.333
15. Maya (3)	May	27	0.312	0.673	0.964	0.643	0.875	0.704	0.000	0.268	0.346

[a] References: (1): Present study, (2): Battilana et al., 2002, (3): Battilana et al., 2006.
[b] Linguistic filiations according to Greenberg (1987): AND: Andean, ET: Equatorial-Tucanoan, GPC: Gê-Pano-Carib, Chib: Chibcan, May: Mayan.

TABLE 4. Pairwise genetic distances between Native Americans

Distances	Aymara	QuechTi	Aché	Cainga	Guaraní	Xavante	Cint.L	Gavião	QuechA	QuechTay	Surui	Waiwai	Zoro	Yanom
Aymara														
QuechTin	0.007													
Aché	0.079	0.079												
Caingang	0.071	0.070	*0.120*											
Guaraní	*0.071*	*0.053*	0.065	0.082										
Xavante	0.072	0.067	0.164	0.066	0.067									
Cint.L	0.153	0.149	0.188	0.150	0.068	0.092								
Gavião	0.069	0.059	*0.039*	0.100	0.046	0.124	*0.173*							
QuechA	0.042	*0.035*	0.108	0.095	*0.032*	0.054	0.064	0.090						
QuechTay	0.046	0.049	*0.091*	0.007	0.064	0.065	0.131	0.078	0.066					
Surui	0.108	0.089	0.180	0.160	0.069	0.061	*0.087*	0.140	0.066	0.148				
Waiwai	*0.084*	*0.075*	0.017	*0.112*	0.035	0.126	0.150	*0.018*	0.090	0.090	0.130			
Zoro	*0.044*	0.044	*0.046*	*0.093*	0.038	0.090	0.127	0.013	0.058	*0.067*	0.102	*0.025*		
Yanomami	0.071	0.040	0.115	0.100	0.036	0.085	*0.123*	0.072	*0.052*	0.088	*0.056*	0.084	0.069	
Maya	0.069	0.057	0.108	*0.061*	0.030	0.065	0.050	0.099	*0.027*	*0.047*	0.079	0.093	0.079	*0.042*

In italics, genetic distance values not significantly different from zero.

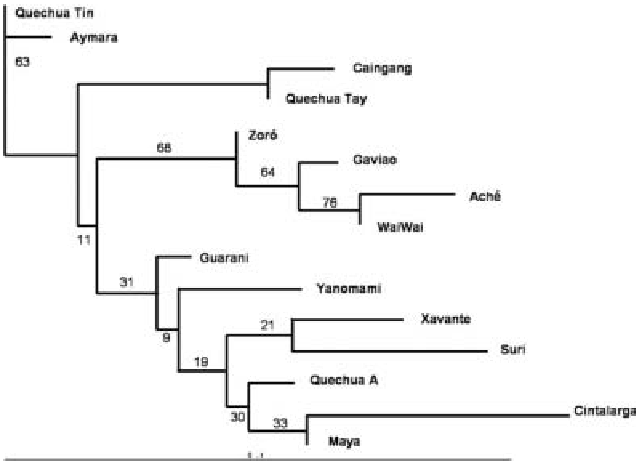


Fig. 2. Neighbor-joining tree obtained from Reynolds's distances. Bootstrap values based on 1,000 replications.

nolds's distances (Table 4). Distance errors (Table 4) indicated that around 76% of the distance values were significant. The highest distance was observed between Cinta Larga and the Aché South American groups (0.188). It is worth noting that the distance between the two Bolivian samples of this study was among the lowest values found (0.007). The average distance value between all pairs of South Americans was 0.082; the mean distance between groups of the Eastern region (10 samples) was 0.09, more than twice the value (0.04) of the Western (Andean) region (four samples). It is interesting to note that the distance between the two Bolivian samples examined is 11 times smaller than the average in South America, and 7 times smaller than the distance between any other pairs of Andean populations (range 0.035–0.066).

Population distance relationships were represented through a neighbor-joining tree (see Fig. 2). This tree highlights the similarity between the two Bolivian populations of the present study, grouping them into a tight cluster, clearly differentiated from the rest. The Zoró, Gavião, Aché, and WaiWai South Amerindian populations form another cluster. The Amazon population of Cinta Larga appeared as the most differentiated. Interestingly, the three Quechua samples appeared clearly separated in the tree in spite of sharing the same language and geographical proximity. In an attempt to avoid the dichotomy implied in the tree construction, a MDS analysis (see Fig. 3) was performed that illustrates the close position of the two Bolivian groups. The rest of South American populations appeared scattered in the plot, showing a distribution pattern similar to the tree topology.

Genetic structuring

A global analysis of the allele frequency variance in Central and South America indicated a significant variation of the PAI markers (Table 5). The global F_{st} for the 15 populations considered showed that ~5% ($P < 0.001$) of the variation could be ascribed to between-population differentiation. A value around 7% was found among South Americans.

Among the Andean populations, the analysis of the genetic variance showed a high similarity between the two Bolivian samples in this study ($F_{st} = -0.003$, $P = 0.91$), whereas the global F_{st} for the three Quechua populations (0.031) was statistically significant ($P = 0.014$).

A hierarchical F_{st} analysis was performed grouping the South American populations according to geographical criteria into two groups: the Western region (Aymara, Quechua Tinguipaya, Quechua Arequipa, and Quechua Tayacaja) and the Eastern region (Aché, Caingang, Guaraní, Cinta Larga, Gavião, WaiWai, Xavante, Zoró, Suruí). In this context, the total between-population diversity ($F_{st} = 0.049$, $P < 0.0001$) can be almost completely explained by the diversity within groups ($F_{sc} = 0.043$, $P < 0.0001$) indicating the absence of geographic structure of the PAI data in South America.

According to linguistic criteria (Greenberg, 1987) the South American population samples were grouped into three of the four linguistic subfamilies: Gê-Pano-Carib languages (Caingang, Xavante, Wai Wai), Equatorial-Tucanoan languages (Aché, Guaraní, Cinta Larga, Gavião, Surui, Zoró), and Andean languages (Aymara, Quechua Tinguipaya, Quechua Arequipa, and Quechua Tayacaja). In this analysis, the Yanomani population was not included because it belongs to a different linguistic subfamily (Chibchan-Paezan). As in the former result, the most important part of the diversity between populations ($F_{st} = 0.051$, $P < 0.0001$) can be attributed to the diversity within groups ($F_{sc} = 0.040$, $P < 0.0001$).

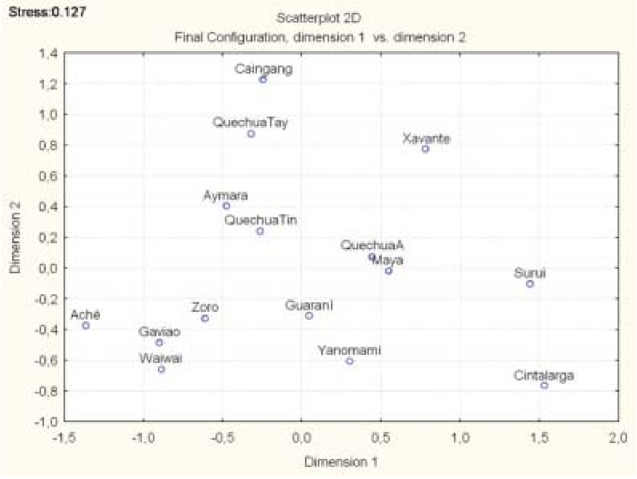


Fig. 3. Multidimensional scaling of Native Americans from Reynolds's distances. Raw stress value was 12.7%. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

DISCUSSION

The analysis of 32 *Alu* polymorphic insertions presented in this study allowed the determination of the genetic characterization of the two main linguistic groups from Bolivia, Aymara, and Quechua, and supplies new data on Native American genetic variability. In fact, so far as we know, the 14 X chromosome *Alu*s have been used for the first time in Amerindians, as well as 8 out of the 18 autosomal PAIs in South Amerindians. Moreover, an Aymara population has never previously been characterized for these markers.

Alu genetic features of the current Bolivian populations

Concerning the distinctiveness/characterization of the Native-American populations based on autosomal *Alu* frequency distributions, the two Bolivian populations show allele frequency patterns similar to other South Amerindian populations for the 10 markers for which data are available, (Antunez de Mayolo et al., 2002; Battilana et al., 2006; Dornelles et al., 2004; Mateus-Pereira et al., 2005; Novick et al., 1998; Tishkoff et al., 1996, 1998), except for the HS4.32 locus that displays the highest insertion frequencies in our study (Aymaras: 0.473, Quechuas: 0.419). For the eight autosomal loci examined for the first time in Native South Americans, it is interesting to note the extreme frequency values found in five loci near the fixation, for both absence (Yb8NBC125, HS2.43, HS4.69, Sb19.12) and presence (Ya5NBC221), as compared with other continents. The remaining three loci (Yb8NBC120, Sb19.3, and B65) present intermediate frequencies in relation to other populations. In general, our results for the eight PAIs previously studied in Amerindians are consistent with the pattern proposed by some authors indicating higher insertion frequencies in Native Americans and Asians than Africans (Mateus-Pereira et al., 2005; Stoneking et al., 1997); however, this trend is not clear for the remaining PAIs analyzed in this study.

Gene diversity variation for the eight *Alu* loci tested so far in South American populations appears to be remarkably high (Battilana et al., 2006; Novick et al., 1998; and present study). Some loci present a large heterozygosity

TABLE 5. Alu *frequency variance analyses in Native Americans*

			Hierarchical F_{st} analyses			
Nonhierarchical analyses	n	F_{st}	Population groups	Within groups	Among groups	Total F_{st}
Native Americans	15	0.045***	Geography: West (4)/East (10)	0.043***	0.006*	0.049***
South Americans	14	0.066***				
Western populations	4	0.015*	Linguistics: Gê-Pano-Carib (3)/ Equatorial-Tucanoan (6)/Andean (4)	0.040***	0.014*	0.051***
Quechua populations	3	0.031*				
Eastern populations	10	0.054***				

*$P < 0.05$; ***$P < 0.001$.

range, for example ACE (H = 0.0–0.5), PV92 (H = 0.0–0.5), and FXIIIB (H = 0.0–0.49). Other loci have a moderate range, for example Hs4.32 (H = 0.19–0.5), A25 (H = 0.0–0.36), and TPA25 (H = 0.21–0.5). For these loci the Bolivian populations are generally presented in high diversity values. In summary, 6 out of 11 of the polymorphic autosomal *Alu* markers in the two Bolivian samples exhibited important gene diversity higher than 0.4 (HS4.32, TPA25, Sb19.3, Yb8NBC120, D1, and B65).

The average gene diversity found in the two populations from the Bolivian Altiplano (H = 0.20) is similar to that described in other South Amerindian populations (H = 0.25; Battilana et al., 2006). Our data are consistent with a general worldwide trend that South Amerindians are the populations with the lowest heterozygosities, followed by Europeans (H = 0.29). This fact could most likely be explained by the genetic drift and bottleneck processes that occurred during the peopling of America, and especially of South America (Watkins et al., 2003).

Of interest is the frequency distribution of the PAIs on the X chromosome. For three loci (Ya5DP57, Yb8DP49, and Ya5a2DP1) Amerindians presented the highest insertion allele frequencies, whereas the Ya5NBC37 and Ya5DP3 Amerindian frequencies are the lowest so far reported in any human population, (Athanasiadis et al., 2007; Callinan et al., 2003). The highest allele frequency differences between the two Bolivian populations corresponded to the Ya5NBC37 locus.

The heterozygosities for most of the X chromosome *Alu* elements followed the general pattern previously described (Athanasiadis et al., 2007; Callinan et al., 2003) with an overall trend towards values lower than those for autosomal PAIs (average values of 0.12 vs. 0.20 in Bolivian populations) according to their chromosomal location. It is interesting to note that two PAIs (Ya5DP77 and Yd3JX437) showed the highest diversity values in the two Bolivian populations like in other African and Asian populations, in contrast with Europeans.

Linguistics vs. genetics in current Bolivian populations

One of the most evident results of this study is the high genetic similarity between the Aymara and Quechua linguistic groups from Bolivia. The two population samples showed very similar allele frequency distributions for the 32 loci analyzed. This close genetic similarity between the two Bolivian groups was also confirmed by the genetic distance and AMOVA analyses. In contrast, the comparison of the two Bolivian samples with other Andean groups underlines the genetic differentiation between Bolivian and Peruvian Quechua-speakers, showing genetic distances seven times higher than those between Aymaras and Quechuas from Bolivia. The high genetic similarity between the two Bolivian samples along with their clear differentiation from other Quechua-speaker peoples from Peru suggests a common genetic origin for the two main linguistic groups in the Bolivian Altiplano. This interpretation implies that the Quechua language expansion under the Inca power into the Bolivian Altiplano was due to cultural diffusion. However, an alternative explanation is also possible. A Quechua language expansion may have also been associated with an early movement of genetically different Quechua speaking people, and that the genetic signature of this movement was erased by subsequent gene flow from original local populations. This explanation is consistent with historical records describing frequent population movements in the Central Andes region during the Inca Empire and afterwards (Platt et al., 2006). However, it seems improbable that gene flow completely erased all genetic signatures in a relatively limited time period (around 500 years) unless an extremely high rate of gene flow was assumed. According to our demographic hypothesis, lower distance values would be expected on comparison of Bolivian vs. Peruvian Quechua-speakers (especially with Peruvian Quechua-speakers from Arequipa who share the same Quechua dialect; Cerron-Palomino, 2003), than to Aymara vs. Peruvian Quechua distances. Nevertheless, the result that the genetic distances of Bolivian Quechuas to Peruvian Quechua-speakers equal those of Aymaras supports that an erasure had to have been complete. On the other hand, the genetic distance between the two Peruvian Quechua-speaking groups is nine times higher than between the two Bolivian samples of this study, suggesting that gene flow in the Central Andes has not been high enough to erase all genetic differences between population groups. In general, the persistence of a certain degree of population divergence in the whole Andean region is shown by the variance of the *Alu* frequencies and is consistent with historically demonstrated (moderate) gene flow that has not completely eliminated the genetic particularities despite the important cultural integration undertaken by the Inca Empire and subsequently by the Spaniards.

Alu-based relationships among Native South Americans

Autosomal *Alu* variation is consistent with significant between-population diversity among South Americans. The N-J tree, MSD graph, and the AMOVA analysis fail to indicate strong clustering according to either geographical or linguistic criteria. However, the average genetic distances seem to indicate a different pattern of variation between the East and West regions of South America. The eastern populations show larger genetic distances and frequency variance than the western ones. Also the high *Alu* heterozygosities found in the Andean region seem to agree with higher within-population diversity as compared with the Eastern region. This could be consistent with different patterns of drift and gene flow, suggested elsewhere from mtDNA (Fuselli et al., 2003; Merriwether et al., 1995), Y chromosome (Tarazona-Santos et al., 2001), classical markers (Luiselli et al., 2000), and STR data (Wang et al., 2007). The Alu-based heterogeneity found in the Eastern South American populations is in agreement with other studies (Lewis and Long, 2008), indicating that they do not appear as a cohesive genetic group. This between-population higher diversity in the East is consistent with the suggested demographical scenario of lower effective population sizes in the East as compared with the West (Fuselli et al., 2003; Tarazona-Santos et al., 2001). Nevertheless, few and uneven population groups (10 from East vs. 4 from West); most of them exhibiting very low sample sizes do not allow a robust test of this hypothesis.

Although a detailed analysis of the correlation between linguistics and genetics in South Native Americans falls out of the scope of this study, it is worth noting that the simple approach of using genetic distance and frequency variance analyses indicates a clear absence of such a correlation. This result is consistent with some previous reports which revealed a positive correlation at a language level (Fagundes et al., 2002; Mateus-Pereira et al., 2005) and at a stock level (Mateus-Pereira et al., 2005), but none at a phyla level using the Loukotka language classification. This level corresponds to the linguistic subfamily level considered in the present work according to the Greenberg classification. The controversial results in the literature highlight the complexity of this subject, as discussed recently (Hunley et al., 2007). According to these authors, the observed absence of correlation can be expected considering deep linguistic branches of the Greenberg's classification. In this context, our results indicate that the autosomal *Alu* variation analyzed confirms the absence of genetic-linguistic congruence regarding these linguistic subfamilies in South Native Americans.

CONCLUSIONS

This genetic analysis confirmed the importance of using autosomal genetic markers, such as *Alu* insertions, to unravel the history of human populations. This work underlined the importance of new studies on additional populations to complete the genetic picture of the Andean and South American populations. Finally, this study has revealed the genetic similarity between Bolivian populations belonging to the two main linguistic groups of the region (Aymara and Quechua), reaffirming that languages may not be congruent with the genetic features of the populations. In this sense, the Quechua language, though the main language in the Andean region, is not a safe indicator of the genetic identity of this region.

ACKNOWLEDGMENTS

The authors are grateful to the Service de Coopération et d'Action Culturelle (in particular they thank Marie-Claude Senemaud) of French embassy in La Paz, and to Mr. Alain Fouquet (the French ambassador). This field work received support from Universidad Mayor de San Andres, La Paz. Genealogical information was collected by IBBA, with the help of Emile Crognier and André Sevin (CNRS).

LITERATURE CITED

Antunez-de-Mayolo G, Antunez-de-Mayolo A, Antunez-de-Mayolo P, Papiha SS, Hammer M, Yunis JJ, Yunis EJ, Damodaran C, Martinez de Pancorbo M, Caeiro JL, Puzyrev VP, Herrera RJ. 2002. Phylogenetics of worldwide human populations as determined by polymorphic Alu insertions. Electrophoresis 23:3346–3356.

Athanasiadis G, Esteban E, Via M, Dugoujon JM, Moschonas N, Chaabani H, Moral P. 2007. The X chromosome Alu insertions as a tool for human population genetics: data from European and African human groups. Eur J Hum Genet 15:578–583.

Battilana J, Bonatto SL, Freitas LB, Hutz MH, Weimer TA, Callegari-Jacques SM, Batzer MA, Hill K, Hurtado AM, Tsuneto LT, Petzl-Erler ML, Salzano FM. 2002. Alu insertions versus blood group plus protein genetic variability in four Amerindian populations. Ann Hum Biol 29:334–347.

Battilana J, Fagundes NJ, Heller AH, Goldani A, Freitas LB, Tarazona-Santos E, Munkhbat B, Munkhtuvshin N, Krylov M, Benevolenskaia L, Arnett FC, Batzer MA, Deininger PL, Salzano FM, Bonatto SL. 2006. Alu insertion polymorphisms in Native Americans and related Asian populations. Ann Hum Biol 33:142–160.

Belle EM, Barbujani G. 2007. Worldwide analysis of multiple microsatellites: language diversity has a detectable influence on DNA diversity. Am J Phys Anthropol 133:1137–1146.

Bouysse-Cassagne T. 1986. Urco and Uma: Aymara concepts of space. In: Murra JV, Wachtel N, Revel J, editors. Anthropological history of Andean polities. Cambridge: Cambridge University Press. p 201–227.

Browman DL. 1994. Titicaca Basin archaeolinguistics: Uru, Pukina and Aymara AD 750-1450. World Archaeol 26:235–251.

Callinan PA, Hedges DJ, Salem AH, Xing J, Walker JA, Garber RK, Watkins WS, Bamshad MJ, Jorde LB, Batzer MA. 2003. Comprehensive analysis of Alu-associated diversity on the human sex chromosomes. Gene 317:103–110.

Carroll ML, Roy-Engel AM, Nguyen SV, Salem AH, Vogel E, Vincent B, Myers J, Ahmad Z, Nguyen L, Sammarco M, Watkins WS, Henke J, Makalowski W, Jorde LB, Deininger PL, Batzer MA. 2001. Large-scale analysis of the Alu Ya5 and Yb8 subfamilies and their contribution to human genomic diversity. J Mol Biol 311:17–40.

Cavalli-Sforza LL, Minch E, Mountain JL. 1992. Coevolution of genes and languages revisited. Proc Natl Acad Sci USA 89:5620.

Cerron-Palomino R. 2003. Lingüística Quechua. Centro Bartolomé de las Casas: Cuzco, Peru.

Cordaux R, Srikanta D, Lee J, Stoneking M, Batzer MA. 2007. In search of polymorphic Alu insertions with restricted geographic distributions. Genomics 90:154–158.

Corella A, Bert F, Pérez-Pérez A, Gené M, Turbón D. 2007. Mitochondrial DNA diversity of the Amerindian populations living in the Andean Piedmont of Bolivia: Chimane, Moseten, Aymara and Quechua. Ann Hum Biol 34:34–55.

Crognier E, Villena M, Vargas E. 2002. Helping patterns and reproductive success in Aymara communities. Am J Hum Biol 14:372–379.

Dornelles CL, Battilana J, Fagundes NJ, Freitas LB, Bonatto SL, Salzano FM. 2004. Mitochondrial DNA and Alu insertions in a genetically peculiar population: the Ayoreo Indians of Bolivia and Paraguay. Am J Hum Biol 16:479–488.

Fabre A. 2005. Diccionario etnolingüístico y guía bibliográfica de los pueblos indígenas sudamericanos.

Fagundes NJR, Bonatto SL, Callegari-Jacques SM, Salzano FM. 2002. Genetic, geographic, and linguistic variation among South American Indians: possible sex influence. Am J Phys Anthropol 117:68–78.

Fuselli S, Tarazona-Santos E, Dupanloup I, Soto A, Luiselli D, Pettener D. 2003. Mitochondrial DNA diversity in South America and the genetic history of Andean highlanders. Mol Biol Evol 20:1682–1691.

Gonzalez-Perez E, Via M, Esteban E, López-Alomar A, Mazieres S, Harich N, Kandil M, Dugoujon JM, Moral P. 2003. Alu insertions in the Iberian Peninsula and North West Africa—genetic boundaries or melting pot? Coll Antropol 27:491–500.

Gonzalez-Perez E, Esteban E, Via M, García-Moro C, Hernández M, Moral P. 2006. Genetic change in the polynesian population of Easter Island: evidence from Alu insertion polymorphisms. Ann Hum Genet 70:829–840.

Greenberg JH. 1987. Language in the Americas. Stanford: Stanford University Press.

Hunley KL, Cabana GS, Merriwether DA, Long JC. 2007. A formal test of linguistic and genetic coevolution in Native Central and South America. Am J Phys Anthropol 132:622–631.

Kolata A. 1993. The Tiwanaku: portrait of an Andean civilization. The people of America. Cambridge: Blackwell Publishers.

Lewis CM, Long JC. 2008. Native South American genetic structure and prehistory inferred from hierarchical modelling of mtDNA. Mol Biol Evol 25:478–486.

Lewis CM, Lizarraga B, Tito RY, Lopez PW, Iannacone C, Medina A, Martinez R, Polo SI, De La Cruz AF, Caceres AM, Stone AC. 2007. Mitochondrial DNA and the peopling of South America. Hum Biol 79:159–178.

Luiselli D, Simoni L, Tarazona-Santos E, Pastor S, Pettener D. 2000. Genetic structure of Quechua-speakers of the Central Andes and geographic patterns of gene frequencies in South Amerindian populations. Am J Phys Anthropol 113:5–17.

Mateus-Pereira LH, Socorro A, Fernandez I, Masleh M, Vidal D, Bianchi NO, Bonatto SL, Salzano FM, Herrera RJ. 2005. Phylogenetic information in polymorphic L1 and Alu insertions from East Asians and Native American populations. Am J Phys Anthropol 128:171–184.

Merriwether DA, Rothhammer F, Ferrell RE. 1995. Distribution of the four founding lineage haplotypes in Native Americans suggests a single wave of migration for the New World. Am J Phys Anthropol 98:411–430.

Moral P, Marogna G, Salis M, Succa V, Vona G. 1994. Genetic data on Alghero population (Sardinia): contrast between biological and cultural evidence. Am J Phys Anthropol 93:441–453.

Nei M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89:583–590.

Novick GE, Novick CC, Yunis J, Yunis E, Antunez-de-Mayolo P, Scheer WD, Deininger PL, Stoneking M, York DS, Batzer MA, Herrera RJ. 1998. Polymprphic Alu insertions and the Asian origin of Native American populations. Hum Biol 70:23–39.

Platt T, Bouysse-Cassagne T, Harris O. 2006. Qaraqara-Charka, Mallku, Inka y Rey en la provincia de Charcas (siglos XV-XVII). Historia antropológica de una confederación aymara. Institut français d'études andines—IFEA; Plural editores; University of St. Andrews; University of London; Interamerican Foundation; Fundación Cultural del Banco Central de Bolivia, La Paz.

Raymond M, Rousset F. 1995. GENEPOP version 1.2: population genetics software for exact tests and ecumenicism. J Hered 86:248–249.

Resano M, Esteban E, González-Pérez E, Vía M, Athanasiadis G, Avena S, Goicoechea A, Bartomioli M, Fernández V, Cabrera A, Dejean C, Carnese F, Moral P. 2007. How many populations set foot through the Patagonian door? Genetic composition of the current population of Bahía Blanca (Argentina) based on data from 19 Alu polymorphisms. Am J Hum Biol 19:827–835.

Reynolds J, Weir BS, Cockerham CC. 1983. Estimation of the coancestry coefficient: basis for a short-term genetic distance. Genetics 105:767–779.

Rowe JH. 1963. Inca culture at the time of the Spanish conquest. In: Steward JH, editor. Handbook of South American Indians, Vol. 2. New York: Cooper Square. p 183–330.

Roy AM, Carroll ML, Kass DH, Nguyen SV, Salem AH, Batzer MA, Deininger PL. 1999. Recently integrated human Alu repeats: finding needles in the haystack. Genetica 107:149–161.

Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425.

Sandoval J, Delgado B, Rivas L, Bonilla B, Nugent D, Fujita R. 2004. Variants of mtDNA among islanders of the lake Titicaca: highest frequency of haplotype B1 and evidence of founder effect. Rev Peru Biol 11:161–168.

Schneider S, Roessli D, Excoffier L. 2000. Arlequin: a software for population genetics data analysis. Ver 2.000. Genetics and Biometry Laboratory, Department of Anthropology, University of Geneva.

Stanish C. 2001. The origin of state societies in South America. Annu Rev Anthropol 30:41–64.

Stoneking M, Fontius JJ, Clifford SL, Soodyall H, Arcot SS, Saha N, Jenkins T, Tahir MA, Deininger PL, Batzer MA. 1997. Alu insertion polymorphisms and human evolution: evidence for a larger population size in Africa. Genome Res 7:1061–1071.

Tarazona-Santos E, Carvalho-Silva DR, Pettener D, Luiselli D, De Stefano GF Labarga CM, Rickards O, Tyler-Smith C, Pena SD, Santos FR. 2001. Genetic differentiation in South Amerindians is related to environmental and cultural diversity: evidence from the Y chromosome. Am J Hum Genet 68:1485–1496.

Tishkoff SA, Dietzsch E, Speed W, Pakstis AJ, Kidd JR, Cheung K, Bonné-Tamir B, Santachiara-Benerecetti AS, Moral P, Krings M, Pääbo S, Watson E, Risch N, Jenkins T, Kidd KK. 1996. Global patterns of linkage disequilibrium at the CD4 locus and modern humans origins. Science 271:1380–1387.

Tishkoff SA, Goldman A, Calafell F, Speed WC, Deinard AS, Bonne-Tamir B, Kidd JR, Pakstis AJ, Jenkins T, Kidd KK. 1998. A global haplotype analysis of the myotonic dystrophy locus: implications for the evolution of modern humans and for the origin of myotonic dystrophy mutations. Am J Hum Genet 62:1389–1402.

Tschopik H. 1963. The Aymara. In: Steward JH, editor. Handbook of South American Indians, Vol. 2. New York: Cooper Square. p 501–573.

Wang S, Lewis CM Jr, Jakobsson M, Ramachandran S, Ray N, Bedoya G, Rojas W, Parra MV, Molina JA, Gallo C, Mazzotti G, Poletti G, Hill K, Hurtado AM, Labuda D, Klitz W, Barrantes R, Bortolini MC, Salzano FM, Petzl-Erler ML, Tsunedo LT, Llop E, Rothhammer F, Excoffier L, Feldman MW, Rosenberg NA, Ruiz-Linares A. 2007. Genetic variation and population structure in Native Americans. PLoS Genet 3:e185.

Watkins WS, Rogers AR, Ostler CT, Wooding S, Bamshad MJ, Brassington AM, Carroll ML, Nguyen SV, Walker JA, Prasad BV, Reddy PG, Das PK, Batzer MA, Jorde LB. 2003. Genetic variation among world populations: inferences from 100 Alu insertion polymorphisms. Genome Res 13:1607–1618.

mtDNA and Y-Chromosome Diversity in Aymaras and Quechuas From Bolivia: Different Stories and Special Genetic Traits of the Andean Altiplano Populations

Magdalena Gayà-Vidal,[1,2] Pedro Moral,[1] Nancy Saenz-Ruales,[2] Pascale Gerbault,[2] Laure Tonasso,[2] Mercedes Villena,[3] René Vasquez,[4] Claudio M. Bravi,[5] and Jean-Michel Dugoujon[2]*

[1]*Unitat d'Antropologia, Biologia Animal, Universitat de Barcelona, 08028, Spain*
[2]*Laboratoire d'Anthropologie Moléculaire et Imagerie de Synthèse (AMIS), FRE 2960 CNRS, Université de Toulouse III Paul Sabatier, Toulouse 31000, France*
[3]*Instituto Boliviano de Biología de Altura (IBBA), Universidad Mayor de San Andres, La Paz, Bolivia*
[4]*Instituto Boliviano de Biología de Altura (IBBA), Universidad Autonoma Tomas Frias, Potosí, Bolivia*
[5]*Laboratorio de Genética Molecular Poblacional, Instituto Multidisciplinario de Biología Celular (IMBICE), La Plata 1900, Argentina*

KEY WORDS uniparental markers; human populations; native Americans; Andean region

ABSTRACT Two Bolivian samples belonging to the two main Andean linguistic groups (Aymaras and Quechuas) were studied for mtDNA and Y-chromosome uniparental markers to evaluate sex-specific differences and give new insights into the demographic processes of the Andean region. mtDNA-coding polymorphisms, HVI-HVII control regions, 17 Y-STRs, and three SNPs were typed in two well-defined populations with adequate size samples. The two Bolivian samples showed more genetic differences for the mtDNA than for the Y-chromosome. For the mtDNA, 81% of Aymaras and 61% of Quechuas presented haplogroup B2. Native American Y-chromosomes were found in 97% of Aymaras (89% hg Q1a3a and 11% hg Q1a3*) and 78% of Quechuas (100% hg Q1a3a). Our data revealed high diversity values in the two populations, in agreement with other Andean studies. The comparisons with the available literature for both sets of markers indicated that the central Andean area is relatively homogeneous. For mtDNA, the Aymaras seemed to have been more isolated throughout time, maintaining their genetic characteristics, while the Quechuas have been more permeable to the incorporation of female foreigners and Peruvian influences. On the other hand, male mobility would have been widespread across the Andean region according to the homogeneity found in the area. Particular genetic characteristics presented by both samples support a past common origin of the Altiplano populations in the ancient Aymara territory, with independent, although related histories, with Peruvian (Quechuas) populations. Am J Phys Anthropol 145:215–230, 2011. © 2011 Wiley-Liss, Inc.

The present population of the Andean region in Bolivia is the result of complex processes over thousands of years. It was in the central Andes (Andean Altiplano and current Peru) where the first complex societies and civilizations in South America emerged (Chavin, 900–200 BC, Tiwanaku, 100 BC–1200 AD, Huari (700–1200 AD) as well as the first state; the Inca Empire that was conquered by the Spaniards around 1532 AD (Stanish, 2001). Specifically, in the south central Andes (southern Peru, Bolivian Altiplano, north Chile, and northwest Argentina), the Tiwanaku civilization, originating in the Titicaca basin, extended its influence over the south central Andes (Kolata, 1993). After the Tiwanaku collapse, the state fragmented into a number of Aymara polities or *Señorios* (Bouysse-Cassagne, 1986) that persisted until their conquest by the Inca Empire (1300–1532 AD) when they became grouped within the Kollasuyu Inca region. From Cuzco, the Incas expanded their power toward the north and south using strategies such as language imposition (Quechua) and the *mitma* system (a deliberate movement of whole tribes from region to region around their vast Empire).

Linguistically, two main groups are present in the Andean area, the Quechuas (10 million speakers in Ecuador, Peru, southern Bolivia, and northern Chile) and the Aymaras (around 2.5 million of speakers, mainly in Bolivia). Before the Inca period, it is likely that an ancestral form of Quechua (technically referred to as proto-Quechua) was spoken in the Huari distribution area (around current Ayacucho, Peru), whereas a proto-Aymara, together with Pukina and Uru, was probably

Additional supporting information may be found in the online version of this article.

Grant sponsors: The work of M. GV has been financed by a FI (2007FIB00108) and a BE grants from the "Departament d'Educació i Universitats de la Generalitat de Catalunya i el Fons Social Europeu". It benefited from funds of Centre National de la Recherche Scientifique (CNRS) and Université Paul Sabatier Toulouse III, France (J.-M.D.), and the CGL2008-03955 project of the Spanish Ministerio de Educacion y Ciencia (P.M.).

*Correspondence to: Jean-Michel Dugoujon, Laboratoire d'Anthropologie Moléculaire et Imagerie de Synthèse (AMIS), FRE 2960 CNRS, Université de Toulouse III Paul Sabatier, 37 Allées Jules Guesde, Faculté de Médecine, 31073 Toulouse cedex 3, Toulouse, France. Tel: 05 61 14 59 87. Fax: +33(0)5 61 14 59 79.
E-mail: dugoujon@cict.fr

Present address of Pascale Gerbault: Molecular and Cultural Evolution Laboratory, University College London (UCL), London, UK.

Received 25 August 2010; accepted 9 December 2010

DOI 10.1002/ajpa.21487
Published online 5 April 2011 in Wiley Online Library (wileyonlinelibrary.com).

spoken in the influence area of the Tiwanaku civilization (Kolata, 1993; Browman, 1994; Stanish, 2001). Afterward, the Incas spread the Quechua tongue and imposed it as the official language of the empire, which was subsequently promoted by the Spaniards as *lingua franca* (Rowe, 1963).

This study explores the genetic variability and the genetic relationships of two Bolivian populations belonging to the two main Andean linguistic groups (Aymaras and Quechuas) through the analysis of uniparental markers; mtDNA and Y-chromosome. A previous study (Gayà-Vidal et al., 2010) examined a total number of 32 polymorphic Alu insertions (PAIs) in these two samples. According to these autosomal and X-chromosome data, the two Bolivian populations showed a similar genetic structure and were significantly close to each other when they were compared to Peruvian Quechua-speakers from Tayacaja and Arequipa. This suggested that the arrival of the Quechua language into Bolivia was more likely the result of a cultural spread rather than a demographic expansion. Nevertheless, have there been different population histories according to gender?

The aim of this study is to evaluate sex-specific differences by analyzing maternal (mtDNA) and paternal (Y-chromosome) uniparental markers in these two populations to gain new insights into the relationships between these two linguistic groups in Bolivia and into the demographic processes that have shaped the current Bolivian populations. The existence of more extensive data for uniparental markers than for biparental PAIs will allow us to achieve more robust interpretations. The majority of the central Andean populations studied so far for the mtDNA control region (CR) are located in Peru (Fuselli et al., 2003; Lewis et al., 2005, 2007), but these studies only considered the HVI region. Also, several samples from northwest Argentina (Alvarez-Iglesias et al., 2007; Tamm et al., 2007) and Bolivia (Corella et al., 2007; Afonso Costa et al., 2010; Barbieri et al., 2011) have been studied. As for Bolivian samples, the samples from Corella et al. (2007) and Barbieri et al. (2011) were only studied for the HVI region. Additionally, those from Corella et al. (2007) corresponded to 10 Aymaras and 19 Quechuas that migrated from the highlands to the lowlands in the Beni department of Bolivia, so their original location is imprecise. On the other hand, Afonso Costa et al. (2010) studied a sample from La Paz with a remarkable sample size (106), but it was an urban sample, and thus, individuals may have different origins. As for previous Y-chromosome variation studies, Andean data come mainly from Peru and the Andean area of Argentina (Bianchi et al., 1998; Tarazona-Santos et al., 2001; Iannacone et al., 2005; Toscanini et al., 2008; Blanco Verea et al., 2010). Two Bolivian samples available (Lee et al., 2007) were described as Highlanders (from the Andean Altiplano) and Lowlanders (a mix of migrants from the Altiplano and natives from the Beni department), but SNPs were not analyzed to confirm the Native American haplogroups.

In this context, the present study (i) increases the number of Andean samples studied for both types of markers, providing haplogroup and haplotype data, (ii) covers the Bolivian Altiplano region, an area with a particular history in the Andean region, and (iii) provides data from two well-defined population samples with large sample sizes. Thus, it will allow us to obtain a more accurate understanding of the genetic relationships in the Andean region.

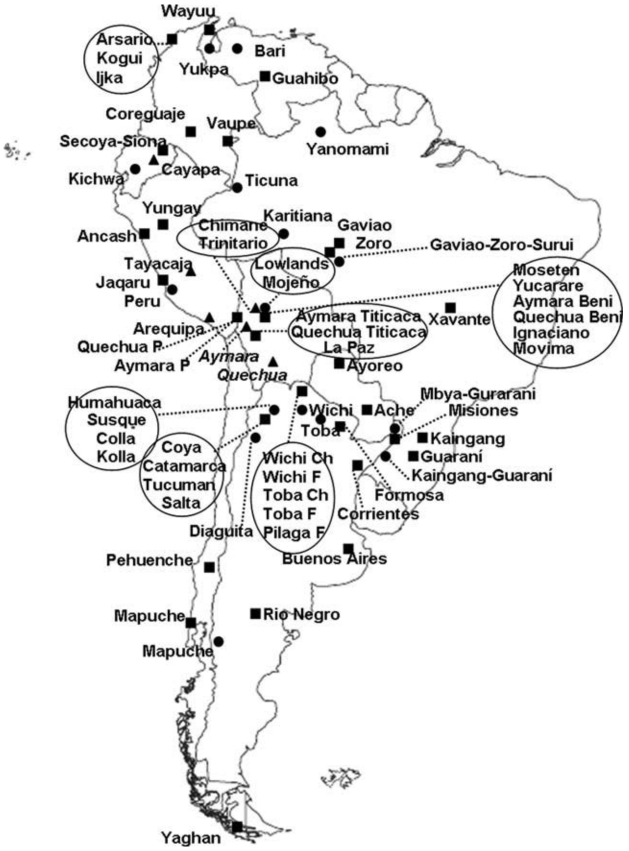


Fig. 1. Location of the populations included in the analyses. Circles, squares, and triangles indicate populations included in Y-chromosome, mtDNA, and both mtDNA and Y-chromosome analyses, respectively, *the two samples of this study.*

SUBJECTS AND METHODS

Population samples

Blood samples from two Native American Bolivian samples, Aymara-speakers from the Titicaca Lake area and Quechua-speakers from the northern Potosi department, a region that was Aymara-speaking before the Inca expansion (Tschopik, 1963; see Fig. 1), were collected with informed consent by the Instituto Boliviano de Biología de Altura (IBBA), with approval from the Ethical Committee of this institution. From the available genealogical records, a total of 189 (93 Quechuas and 96 Aymaras) and 114 (55 Quechuas and 59 Aymaras) unrelated individuals were analyzed for mtDNA and Y-chromosome, respectively. A more detailed description of these populations can be found in Gayà-Vidal et al. (2010).

mtDNA polymorphisms

A mtDNA segment including the HVS-I and most of the HVS-II mtDNA regions was amplified by polymerase chain reaction (PCR), using the primer pair F15973 and R296, and PCR conditions as described in Coudray et al. (2009). DNA purification was undertaken using QIAquick PCR purification Kit (QIAgen, Courtaboeuf, France). Both strands were sequenced with the Big Dye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) and run in an ABI PRISM 3730 sequencer (PE, Applied Biosystems). In cases of samples with C-stretch, typical of haplogroup B2, both strands were sequenced twice. The sequences were checked manually with Sequencing Analysis (ABI Prism v.3.7) and Chromas 2.13 software. The sequences were aligned and compared to the revised Cambridge Reference Sequence rCRS (Anderson et al., 1981; Andrews et al., 1999) using the Bioedit software (Tom Hall, Carlsbad, USA). To discard as many sequence artifacts as possible (Bandelt et al., 2001, 2002), each chromatogram was revised at least three times, sequences having unusual mutations were resequenced, and data entry and edition revised several times.

Coding region polymorphisms were typed to classify mtDNA into the four major Amerindian haplogroups (A–D) (Torroni et al., 1992). These four haplogroups were defined by restriction size or length polymorphisms (A: *Hae*III np663, B: 9bp-deletion, C: *Hinc*II np13259, and D: *Alu*I np5176). The primers and the PCR conditions were previously described in Mazieres et al. (2008). Genotype determinations were performed through 3% agarose gel electrophoresis and ethidium bromide staining. In addition, the SNP 6473T was typed to determine the B2 haplogroup, characteristic of Native Americans, and haplogroups A2, C1, and D1 were assigned according to the CR (Bandelt et al., 2003; Tamm et al., 2007).

Y-chromosome polymorphisms

A total of 20 markers were determined: 17 Y-chromosome short tandem repeat (Y-STR) polymorphic loci (DYS456, DYS389i/ii, DYS390, DYS458, DYS19, DYS385a/b, DYS393, DYS391, DYS439, DYS635, DYS392, GATAH4, DYS437, DYS438, and DYS448) and three SNPs (M242, M3, and M346).

The Y-STRs were analyzed according to AmpFISTR® Yfiler^{TM} PCR Amplification Kit (Applied Biosystems), using 1–2 ng of template DNA. The determinations were carried out in ABI 3730 with Genescan® and Genotyper® Analysis Softwares. Allele assignments were based on comparisons with the allelic ladders included in the kit using Genemapper software (Applied Biosystems). The quality of the determinations was assessed using the commercial allelic ladder and the DNA control supplied by Applied Biosystems.

To identify individuals carrying a Native American haplogroup, three biallelic markers were typed: (i) the polymorphic $C \rightarrow T$ transition (marker M242) that defines the haplogroup Q, present in Asia and America (Jobling and Tyler-Smith, 2003; Seielstad et al., 2003), using the methodology of Cinnioglu et al. (2004); (ii) the $C \rightarrow T$ transition (marker M3) in the DYS199 locus (Underhill et al., 1996), which defines the Q1a3a haplogroup, a lineage falling within the haplogroup Q, restricted to the Americas, and reaching a frequency of 100% in some populations; and (iii) for the Y-chromosomes possessing the M242 mutation, but not the M3, we sequenced the M346 $C \rightarrow T$ marker, downstream to M242 and upstream to M3 (Karafet et al., 2008). As in Bailliet et al. (2009), these chromosomes were considered to belong to the paragroup Q1a3*. To assign the most probable haplogroup to the non-Q samples and confirm the Q samples, we used Haplogroup Predictor (http://www.hprg.com/hapest5/) that assigns the most probable haplogroup from the Y-STR profiles.

Data analysis

For mtDNA analyses, we considered the fragment between the 16,017 and 249 positions, according to the rCRS (Anderson et al., 1981; Andrews et al., 1999). Haplogroups were assigned following criteria described in the literature (Torroni et al., 1992; Bandelt et al., 2003). Haplogroup and haplotype frequencies were calculated by direct counting. Various diversity indices were computed. To determine the genetic relationships between haplotypes found in the two samples, Median-Joining (MJ) networks (Bandelt et al., 1999) were constructed for each haplogroup. For haplogroup B, positions 16,182 and 16,183 were not considered, because they are dependent on the presence of C at site 16189 (Pfeiffer et al., 1999). Following the suggestions of Bandelt et al. (2000), higher weights were assigned to the least variable polymorphisms and lower weights to the more hypervariable sites in our data set.

As for the Y-chromosome, haplogroup and haplotype frequencies were calculated by direct counting. Taking into account only the individuals belonging to the Native American Q haplogroup (lineages Q1a3* and Q1a3a), various diversity indices were computed. MJ networks (Bandelt et al., 1999) were built with the MP postprocessing option (Polzin and Daneschmand, 2003) for the Q1a3a hg. STRs were given weights that were inversely proportional to their allele size variances.

Exact tests of population differentiation (Raymond and Rousset, 1995) were performed to detect whether significant differences in mtDNA haplogroup frequencies and in Y-STR allele frequencies existed between the two Bolivian samples. In addition, the two study populations were compared for both sets of markers using $F_{\rm st}$ indices, as a measure of population differentiation.

For comparative purposes, mtDNA data from 51 South American samples (Table 1) were collected from the literature on the basis of available sequences for the HVI region and a minimum sample size of nine individuals. Their geographical location is shown in Figure 1. For the analyses, only Native American haplogroups were considered. The comparisons were based on the HVI region between 16,051 and 16,362 positions. The San Martin de Pangoa sample (Fuselli et al., 2003) was not included, because it is composed of both Quechua and Nematsiguenga speakers. The Cayapa sample included was from Rickards et al. (1999), because Tamm et al. (2007) did not maintain the proportions of the haplogroups, because their focus was phylogeny. Additionally, the HVII CR (from position 73 to 249) was available for 22 of the 51 samples (Table 1), and the CR between positions 16,024 and 249 was available for 6 of the 51 samples.

Therefore, analyses were carried out considering the four sets of data separately: the HVI (53 samples), HVII (24 samples), HVI-HVII (24 samples) CRs, and the HVI, HVII and the intervening region, from now on designated as CR (eight samples). Haplotype (h) and nucleotide (π) diversity within groups were calculated using Nei's formulas (1987). Analyses of molecular variance (AMOVA) (Excoffier et al., 1992) and hierarchical AMOVA analyses under geographical criteria (also for haplogroup frequencies) were performed. Genetic distances between samples were estimated using the Tamura–Nei distance method (Tamura and Nei, 1993) with the α

TABLE 1. Populations included in the comparisons for the mtDNA and for the Y-STRs

mtDNA comparisons			Y-Chromosome comparisons		
Populations[a]	*N*[b]	References[c]	Populations	*N*[e]	References
Aymara	96	Present study	**Aymara****	57	Present study
Quechua	93	Present study	**Quechua****	45	Present study
Ignaciano	15/22	Bert et al., 2004, Bert et al., 2001	Colla**	10	Toscanini et al., 2008
Movima	12/22	Bert et al., 2004, Bert et al., 2001	Kaingang_Guarani**	27	Leite et al., 2008
Trinitario	12/35	Bert et al., 2004, Bert et al., 2001	Kichwa**	72	González-Andrade et al., 2007
Yucarare	15/20	Bert et al., 2004, Bert et al., 2001	Peru**	51	Iannacone et al., 2005
Quechua Beni	19/32	Corella et al., 2007/Bert et al., 2001	Toba**	44	Toscanini et al., 2008
Aymara Beni	10/33	Corella et al., 2007/Bert et al., 2001	Trinitario**	34	Tirado et al., 2009
Chimane	10/41	Corella et al., 2007/Bert et al., 2001	Chimane**	10	Tirado et al., 2009
Moseten	10/20	Corella et al., 2007/Bert et al., 2001	Mojeño**	10	Tirado et al., 2009
Ancash Quechua	33/33	Lewis et al., 2005	Kolla**	12	Blanco-Verea et al., 2010
Aymara Puno	14	Lewis et al., 2007	Diaguita**	9	Blanco-Verea et al., 2010
Quechua Puno	30	Lewis et al., 2007	Mapuche**	23	Blanco-Verea et al., 2010
Jaqaru Tupe	16	Lewis et al., 2007	Bari*	16	YHRD:YA003358[f]
Yungay Quechua	36	Lewis et al., 2007	Yanomami*	11	YHRD: YA002906[f]
Arequipa Quechua	22	Fuselli et al., 2003	Yukpa*	12	YHRD:YA003360[f]
Tayacaja Quechua	61	Fuselli et al., 2003	Cayapa	26	Tarazona-Santos et al., 2001
Toba Chaco	43/67[d]	Cabana et al., 2006	Tayacaja Quechua	44	Tarazona-Santos et al., 2001
Wichi Chaco	32/99[d]	Cabana et al., 2006	Arequipa Quechua	15	Tarazona-Santos et al., 2001
Pilaga Formosa	38	Cabana et al., 2006	Gaviao-Zoro-Surui	34	Tarazona-Santos et al., 2001
Toba Formosa	24/[d]	Cabana et al., 2006	Karitiana	8	Tarazona-Santos et al., 2001
Wichi Formosa	67/[d]	Cabana et al., 2006	Ticuna	32	Tarazona-Santos et al., 2001
Ayoreo	91	Dornelles et al., 2004	Mbyá-Guaraní	33	Altuna et al., 2006
Aché	63	Schmitt et al., 2004	Humahuaca	10	Bianchi et al., 1998
Gaviao	27	Ward et al., 1996	Wichi	12	Bianchi et al., 1998
Zoro	30	Ward et al., 1996	Susque	16	Bianchi et al., 1998
Xavante	25	Ward et al., 1996	Lowlands	97	Lee et al., 2007
Guarani	200	Marrero et al., 2007			
Kaingang	74	Marrero et al., 2007			
Quechua Titicaca	37	Barbieri et al., 2010			
Aymara Titicaca	20	Barbieri et al., 2010			
Coya	60	Alvarez-Iglesias et al., 2007			
Buenos Aires	89	Bobillo et al., 2010			
Corrientes	23	Bobillo et al., 2010			
Formosa	15	Bobillo et al., 2010			
Misiones	23	Bobillo et al., 2010			
RioNegro	30	Bobillo et al., 2010			
Guahibo	59	Vona et al., 2005			
Mapuche	34/11	Moraga et al., 2000			
Pehuenche	24/105	Moraga et al., 2000			
Yaghan	15/21	Moraga et al., 2000			
Cayapa	30	Rickards et al., 1999			
Arsario	47	Tamm et al., 2007			
Kogui	48	Tamm et al., 2007			
Ijka	29	Tamm et al., 2007			
Wayuu	42	Tamm et al., 2007			
Coreguaje	27	Tamm et al., 2007			
Vaupe	22	Tamm et al., 2007			
Secoya-Siona	12	Tamm et al., 2007			
Tucuman	9	Tamm et al., 2007			
Salta	18	Tamm et al., 2007			
Catamarca	25	Tamm et al., 2007			
La Paz	106	Afonso Costa et al., 2010			

[a] Underlined samples: HVI and HVII control regions available, Italic samples: Control Region (16,024–249) available.
[b] Individuals included for mtDNA sequences/haplogroup frequencies comparisons.
[c] References for mtDNA sequences/haplogroup frequency data.
[d] The two Wichi and the two Toba samples (Formosa and Chaco) were considered together for the haplogroup frequency comparisons.
[e] Considering only individuals belonging to a Native American haplogroup.
[f] Accession number from the YHRD database (http://www.yhrd.com).
* The minimal haplotype available.
** The 12 STRs available. *Note:* three different names (Coya, Kolla, and Colla) were used to differentiate the three Coya samples from the bibliography.

parameter set at 0.26 (Meyer et al., 1999). Distance matrices were visualized in a multidimensional scaling plot (MDS).

To compare the Y-STR data, 25 South American populations were selected from the literature (Table 1), taking into account only the haplotypes belonging to Native American Y-chromosomes. In cases of surveys where the haplogroups were not indicated, we inferred them using the Haplogroup Predictor web page (http://www.hprg.com/hapest5/). Because of the uneven number of Y-STRs

TABLE 2. mtDNA (from position 16,017 to 249) and Y-chromosome (17 STRs) haplogroup frequencies and diversity parameters

	Population	N	Haplogroup frequencies				Diversity parameters[a]							
mtDNA			A2	B2	C1	D1	H	K	π	h	P_w	Sub	Trans	In/dels
	Aymara	96	0.07	0.81	0.06	0.05	0.331	66	0.011	0.990	8.762	95	7	1
	Quechua	93	0.15	0.61	0.19	0.04	0.569	71	0.014	0.993	11.122	108	14	14 (8)[b]
Y-chromosome			Q1a3*	Q1a3a	Other		H	K	D	H	P_w			
	Aymara	59	0.10	0.86	0.03		0.192	42	0.465	0.988	7.91			
	Quechua	55	0.00	0.78	0.22		0.000	33	0.425	0.981	7.22			

N, number of individuals; H, gene diversity calculated from haplogroups; K, number of haplotypes; π, nucleotide diversity; h, haplotype diversity; P_w, mean number of pairwise differences; Sub, number of substitutions; Trans, number of transversions; In/dels, number of insertions and deletions; D, average gene diversity over loci.

[a] Y-chromosome diversity parameters calculated only considering Native American haplogroups.

[b] If we consider the 59-60d and the 206-211d as unique events, the number of in/dels would be reduced to eight.

analyzed in the other studies, three sets of data were analyzed including (i) the minimal haplotype (DYS19-DYS389i-DYS389ii-DYS390-DYS391-DYS392-DYS393-DYS385a/b), for which we collected 14 populations with a minimum of nine individuals; (ii) 12 STRs (the minimal haplotype plus DYS437, DYS438, and DYS439) with data for 11 of the 14 samples; and (iii) six STRs (DYS19-DYS389i-DYS389ii-DYS390-DYS391-DYS393) with data for 25 samples (the previous 14 plus 11 others) with at least eight individuals. The nomenclature of DYS389i and DYS389ii loci from Bianchi et al. (1998), Tarazona-Santos et al. (2001), and Lee et al. (2007) was homogenized with the rest of the studies. The six STRs were chosen to include in the comparisons the Tayacaja and Arequipa samples, the only Peruvian samples used in the autosomal study of Gayà-Vidal et al. (2010).

For each data set, diversity parameters, AMOVA (Excoffier et al., 1992) based on the sum of squared differences (Rst), hierarchical AMOVA analyses according to geographical criteria, and pairwise Rst genetic distances (depicted in a MDS) to evaluate genetic relationships among populations were calculated.

All the analyses were performed using the programs Arlequin 3.1 (Excoffier et al., 2005), Network v.4.5.1.6 (http://www.fluxus-engineering.com), Statistica (StatSoft 2001), and R.

RESULTS

Diversity in the two Bolivian populations

mtDNA. Haplogroup and mtDNA sequence variation data are shown in the additional file: Supporting Information Table S1. All the 189 individuals corresponded to one of the four major Native American mtDNA haplogroups. The four (A2, B2, C1, and D1) haplogroups were present in the two populations (Table 2). The haplogroup B2 was the most frequent in both samples, especially in the Aymaras (81%). The other three haplogroups showed frequencies of less than 10% in the Aymaras and 20% in Quechuas, and in both cases haplogroup D1 appeared at the lowest frequencies.

A total of 130 different haplotypes were found in the two Bolivian samples (66 in Aymaras and 71 in Quechuas), and only seven were common to both samples (Supporting Information Table S1). Haplotype 2 belonged to haplogroup A2, but it also presented the 9bp deletion that is characteristic of haplogroup B. Haplotype 51, shared by two Quechua individuals, presented several mutations between positions 59 and 71 in the HVII region that were considered as a 59–60 deletion plus two insertions (65 and 71 sites), which represent three events. Within population diversity (Table 2) was higher in Quechuas than in Aymaras, for all the parameters tested and for the number of transitions, transversions, and in/dels.

The MJ networks for the four haplogroups (see Fig. 2) indicated a clear starlike pattern for A2 and B2, with a central node corresponding to the haplotype presenting the characteristic mutations of the haplogroup. Haplogroups C1 and D1 lacked this central node. In general, the four MJ networks showed high-haplotype diversity, mainly for haplogroup B2. Most of the nodes were small, indicating single haplotypes or haplotypes shared by a few individuals, in most cases belonging to the same population.

Y-chromosome. Haplogroup and haplotype distributions are presented in additional file 2: Supporting Information Table S2. 96.6% and 78.2% of Aymara and Quechua individuals, respectively, carried a Native American haplogroup (lineages Q1a3* or Q1a3a) according to the SNPs tested. The remaining individuals belonged, according to the Haplogroup Predictor web page, mainly to haplogroup R1b, the most frequent in Western Europe (Jobling and Tyler-Smith, 2003). Considering only the Native American Y-chromosomes, 100% of Quechuas and 89% of Aymaras presented the haplogroup Q1a3a, the remaining 11% of Aymaras carried Q1a3*.

Taking into account the 17 STRs and all the 114 individuals tested, 87 different haplotypes were found, and 69 of them (79%) were unique. Only one haplotype (haplogroup Q1a3a) was shared between the two Andean populations. Considering only the Native American haplogroups (57 Aymaras and 43 Quechuas), 74 different haplotypes were found and 57 of these (77%) were unique. Haplotype diversity indices (Table 2) were carried out for these 100 individuals and indices were slightly higher in the Aymaras.

The MJ networks for the Q1a3* and Q1a3a haplogroups (see Fig. 3) revealed a high diversity of male lineages in the two Bolivian populations and showed that, in general, most of haplotypes at the end branches belonged to the Aymaras. The only haplotype shared by the two populations presented one of the biggest nodes.

Comparison of the two Bolivian populations

mtDNA. The exact tests of population differentiation for the mtDNA showed significant differences for both the

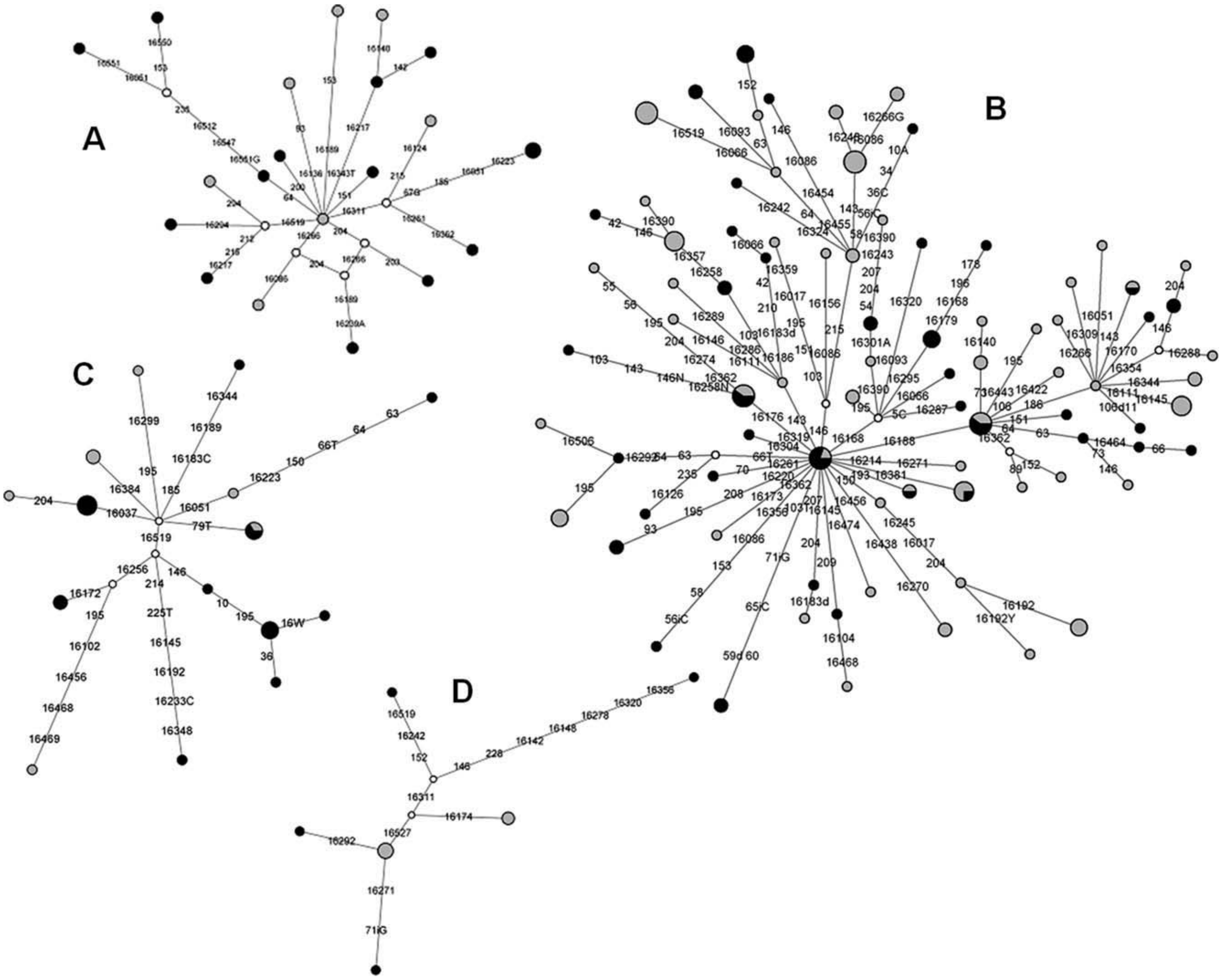


Fig. 2. Haplotype MJ Network for mtDNA haplogroups (**A–D**). Circle sizes are proportional to the number of individuals carrying the corresponding haplotype. Gray circles correspond to Aymaras and black circles to Quechuas. Small white circles correspond to hypothetical haplotypes. Mutational differences between haplotypes are identified as numbers. Central nodes in haplogroups A2 and B2 correspond to haplotypes with the mutations at nucleotide positions: A2 (16,111, 16,223, 16,290, 16,362, 64, 73, 146, 153, and 235); B2 (16,189, 16,217, 16,519, and 73). All the haplotypes in haplogroups C1 and D1 present the mutations: C1 (16,223, 16,298, 16,325, 16,327, 73, and 249d); D1 (16,223, 16,325, 16,362, 73).

mtDNA haplogroups (P = 0.012) and the mtDNA sequences ($P < 0.001$). The same results were obtained through F_{st} values (0.057, P = 0.003 for mtDNA haplogroups; 0.009, $P < 0.001$ for mtDNA sequences).

Y-chromosome. The exact tests of population differentiation did not show statistical differences in the allele frequency distributions for any of the 17 loci. Likewise, no significant differentiation of Y-STR haplotypes was found according to the F_{st} value (0.006, P = 0.239). On the contrary, slight, but significant differences were obtained for the Y-chromosome haplogroups (0.074, P = 0.035).

Comparison with other Native South Americans

mtDNA diversity. Estimates from haplotype and haplogroup data (Table 3) indicated moderate to high levels of diversity in the two Bolivian samples in agreement with other Andean populations. The lowest diversity values were observed in some Colombian samples as well as in the Aché, and the highest values were observed in the urban Argentinean samples.

The analyses of variance in South America indicated that 27% (HVI), 14.2% (HVII), 18.5% (HVI-HVII), and 9.7% (CR) of the variation could be ascribed to between-population differentiation (Table 4). Hierarchical F_{st} analyses (Table 4) from all sets analyzed, as well as from haplogroup frequencies, did not reveal a genetic structure (diversity among groups > within groups) for Andean versus non-Andean groups. In the same way, according to HVI region and haplogroup frequencies, no genetic structure was found for Andean versus nearby areas (Chaco and Bolivian Lowlands). For all sets analyzed, a significant genetic structure was found in the Andean region when Andean samples were divided into three groups (south Andes vs. central Andes vs. northwest Argentina). It is important to note that the Coya sample was included into the central Andean group and not into the northwestern Argentinean group, in which case, no significant or a weak genetic structure was

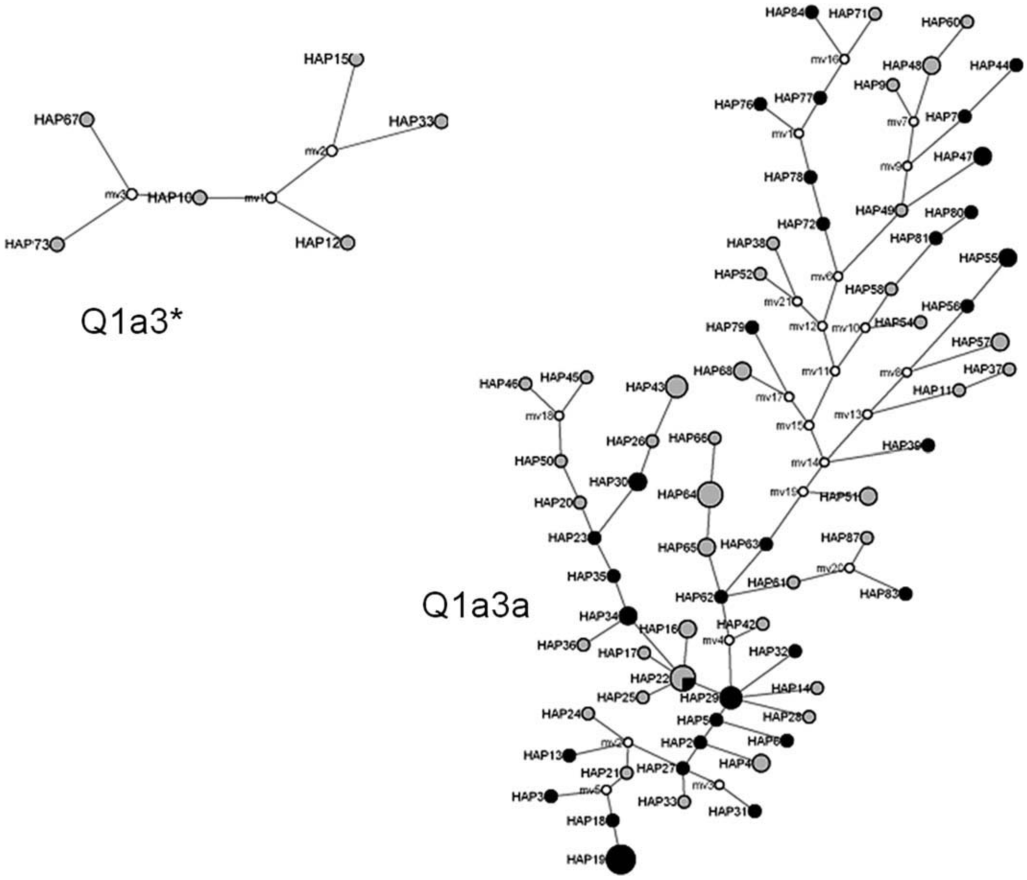


Fig. 3. Network for the Y-chromosome haplogroups Q1a3a and Q1a3*. Circle sizes are proportional to the number of individuals carrying the corresponding haplotype. Gray circles correspond to Aymaras and black circles to Quechuas. Small white circles correspond to hypothetical haplotypes.

found. Considering the central Andes for the HVI region, a slightly significant genetic structure was found for north central Andes versus south central Andes. The CR analysis revealed that most of the genetic differentiation between samples was attributed to between-group differentiation (south central Andes vs. Argentina). Similar results were obtained when the Cayapa sample was included into the central Andean group.

Genetic distances calculated with the Tamura and Nei formula showed that 84% (HVI), 74% (HVII), 82% (HVI-HVII), and 61% (CR) were statistically different from zero (data not shown). The average distance value between all pairs of samples was 0.21 (HVI), 0.16 (HVI-HVII, HVII), and 0.08 (CR). The mean distance between the non-Andean groups was 0.25 (HVI), 0.16 (HVI-HVII, HVII), and 0.01 (CR). The mean distance between the Andean groups was 0.13 (HVI), 0.12 (HVI-HVII), 0.08 (HVII), and 0.02 (CR), values that decreased to 0.09 (HVI), 0.02 (HVI-HVII), and 0.008 (HVII) when the southern Andean and northwest Argentinean samples were removed. It is worth mentioning that all distances between the Aymara sample and the other samples (HVI; HVI-HVII, CR) were statistically significant and showed the lowest distance with the Quechuas (0.036, HVI; 0.029, HVI-HVII; 0.024, CR), closely followed by La Paz (0.038, HVI; 0.03, HVI-HVII) and Aymara Titicaca (0.04, HVI). In contrast, the Quechua sample showed nonstatistically significant distances with Coya for the four sets, with La Paz (for HVI, HVI-HVII), and with Salta and Quechua Titicaca (for HVI).

MDS plots were built from genetic distances for the four sets of analyses (see Fig. 4). The HVI plot (Fig. 4A) shows most of the Andean samples in the upper left part of the plot (except Catamarca, Tucuman, and Yaghan). The Aymara sample is located at the left extreme, together with Aymara Titicaca and the two Bolivian samples from Corella et al. (2007). The Quechua sample is very close to La Paz, together with most of Andean samples. Wichi Formosa is located in the middle of these Andean samples, and Toba Chaco is very close to Quechua Titicaca. On the other hand, the Peruvian Quechuas from Tayacaja and Yungay, as well as Salta, Mapuche, and Pehuenche, are in the center of the plot, mixed with some non-Andean samples, mainly Chaco and Bolivian lowland samples. The Aché, Ayoreo, Guarani, and Ijka samples appear as the most distant samples. The HVII plot (Fig. 4B) shows the central Andean samples, except Catamarca and Salta, to be relatively grouped in the centre. In the HVI-HVII plot (Fig. 4C), the Quechua, Coya, and La Paz samples are grouped and separated from the Aymaras. The two HVII and HVI-HVII plots reveal the Ijka sample as the most separated population. The CR plot (Fig. 4D) reveals the five Argentinean samples to be practically aligned in the right part of the plot. On the contrary, the Andean samples are rather separated, the Aymara at the left extreme of the plot.

TABLE 3. mtDNA diversity data from contributing populations[a]

			HVI				Control region		
	HVI-HVII		(16,051–16,362)		HVII (73–249)		(16,024–249)		Haplogroups
Populations	h	π	h	π	H	π	H	π	H^b
Ijka	0.414	0.005	0.414	0.006	0.197	0.004			0.197
Secoya-Siona	0.818	0.012	0.818	0.016	0.485	0.005			0.53
Kogui	0.668	0.013	0.619	0.014	0.619	0.013			0.503
Arsario	0.790	0.013	0.730	0.014	0.674	0.011			0.472
Yaghan	0.886	0.014	0.886	0.017	0.819	0.008			0.524
Mapuche	0.890	0.014	0.838	0.018	0.793	0.007			0.636
Wayuu	0.820	0.015	0.810	0.017	0.745	0.011			0.675
Aymara	0.988	0.015	0.968	0.015	0.930	0.014	0.990	0.011	0.331
Catamarca	0.913	0.016	0.877	0.018	0.633	0.014			0.68
Pehuenche	0.928	0.016	0.902	0.019	0.859	0.012			0.617
Salta	0.980	0.016	0.967	0.020	0.556	0.010			0.752
Coreguaje	0.872	0.017	0.832	0.018	0.815	0.015			0.527
Cayapa	0.860	0.018	0.837	0.021	0.802	0.014			0.756
Vaupe	0.983	0.018	0.952	0.020	0.878	0.014			0.758
Guahibo	0.895	0.019	0.858	0.016	0.684	0.023			0.541
Quechua	0.988	0.019	0.952	0.020	0.918	0.016	0.993	0.013	0.568
LaPaz	0.993	0.019	0.952	0.020	0.958	0.018			0.579
Tucuman	1.000	0.019	0.972	0.019	0.944	0.019			0.639
Formosa	1.000	0.019	1.000	0.022	0.933	0.014	1.000	0.013	0.762
Misiones	0.984	0.020	0.964	0.022	0.885	0.015	0.984	0.013	0.755
RioNegro	0.991	0.020	0.984	0.022	0.947	0.016	0.991	0.014	0.687
Coya	0.996	0.020	0.980	0.021	0.943	0.017	0.997	0.013	0.556
Buenos Aires	0.993	0.022	0.988	0.023	0.938	0.020	0.994	0.015	0.752
Corrientes	0.984	0.023	0.968	0.022	0.933	0.024	0.988	0.016	0.735
Aché			0.204	0.003					0.175
AymaraBeni			0.667	0.006					0.119
Ayoreo			0.473	0.007					0.281
Guarani			0.764	0.008					0.283
Movima			0.894	0.009					0.571
Xavante			0.677	0.010					0.28
QuechuaBeni			0.673	0.011					0.417
AymaraTiticaca			0.947	0.012					0.195
Zoro			0.775	0.013					0.598
Gaviao			0.866	0.014					0.479
QuechuaYungay			0.954	0.016					0.644
AymaraPuno			0.967	0.016					0.484
Arequipa			0.978	0.016					0.524
JaqaruTupe			0.867	0.017					0.458
QuechuaPuno			0.975	0.017					0.591
Ancash			0.981	0.018					0.669
Kaingang			0.749	0.019					0.545
Chimane			0.800	0.019					0.571
TobaFormosa			0.906	0.019					
Tayacaja			0.968	0.019					0.734
WichiFormosa			0.881	0.020					
TobaChaco			0.888	0.020					0.671[c]
QuechuaTiticaca			0.954	0.020					0.632
Yucarare			0.952	0.021					0.742
Ignaciano			0.971	0.021					0.697
WichiChaco			0.738	0.022					0.689[c]
Moseten			0.844	0.022					0.563
Trinitario			0.985	0.022					0.697
PilagaFormosa			0.964	0.023					0.741

h, haplotype diversity; π, nucleotide diversity.

[a] Estimates are based on the mtDNA control region from positions 16,051–16,362 (HVI) and from 73 to 249 (HVII), therefore, they may be different from their original published source.

[b] Heterozygosities calculated from haplogroup frequencies.

[c] Includes also the Wichi or Toba samples from Formosa region.

Y-chromosome diversity. For the Y-chromosome haplotypes, diversity data parameters were calculated for the 27 populations (Table 5). Karitiana, Chimane, and Yukpa presented the lowest diversity values and the Andean samples the highest values. The Andean, Kichwa, Kolla, Peru, Arequipa, and Tayacaja presented the highest levels and Colla, Susque, and Humahuaca the lowest values. The two Bolivian samples of this work presented intermediate values, the Aymaras presenting slightly higher diversity values than the Quechuas.

The global F_{st} value was similar for the three sets of analysis (6-STRs: 0.24, $P < 0.001$; minimal haplotype and 12-STRs: 0.28, $P < 0.001$). Hierarchical analyses according to geographical criteria (Table 6) revealed a significant genetic structure when central Andes and nearby areas were grouped separately for all sets ana-

TABLE 4. Hierarchical AMOVA analyses with mtDNA sequences and haplogroup frequency data

Hierarchical F_{ST} Analyses				
		Values from HVI variation Values from haplogroup frequencies		
Population groups[a,b]	No. pops	Within groups	Among groups	Total F_{ST}
HVI: Global F_{st} = 0.270***				
Andes (21)/non-Andes (32)	53	0.23***	0.10***	0.31***
Andes (21)/non-Andes (30)	*51*	*0.24****	*0.09***	*0.31****
LowBol (6)/Andes (21)/Chaco (6)	33	0.11***	0.02NS	0.13***
LowBol (6)/Andes (21)/Chaco (4)	*31*	*0.14****	*0.01NS*	*0.14****
LowBol (6)/C Andes (18)/Chaco (6)	30	0.09***	0.03*	0.12***
LowBol (6)/C Andes (18)/Chaco (4)	*28*	*0.08****	*0.04**	*0.12****
S Andes (3)/C Andes (18)	21	0.09***	**0.11****	0.19***
S Andes (3)/C Andes (18)	*21*	*0.08****	***0.27*****	*0.33****
S Andes (3)/C Andes (15)/NWArg (3)	21	0.08***	**0.13*****	*0.19****
S Andes (3)/C Andes (15)/NWArg (3)	*21*	*0.05****	***0.26*****	*0.30****
S Andes (3)/C Andes (15)	18	0.07***	**0.13*****	*0.19****
S Andes (3)/C Andes (15)	*18*	*0.05****	***0.30*****	*0.34****
C Andes (15)/NWArg (3)	18	0.08***	**0.13*****	0.20***
C Andes (15)/NWArg (3)	*18*	*0.06****	***0.19*****	*0.23****
NC Andes (5)/SC Andes (10)	15	0.05***	**0.06****	0.11***
NC Andes (5)/SC Andes (10)	*15*	*0.04****	***0.04****	*0.08****
HVII: Global F_{st} = 0.142***				
Andes (10)/non-Andes (14)	24	0.10***	0.08**	0.17***
S Andes (3)/C Andes (4)/NW Arg (3)	10	0.01**	**0.07*****	0.08***
S Andes (3)/C Andes (7)	10	0.04***	0.02NS	0.06***
S Andes (3)/C Andes (4)	7	0.01***	**0.03***	0.04***
C Andes (4)/NW Arg (3)	7	0.01*	**0.10***	0.11***
HVI-HVII: Global F_{st} = 0.185***				
Andes (10)/Non-Andes (14)	24	0.13***	0.12***	0.23***
S Andes (3)/C Andes (7)	10	0.07***	0.11NS	0.17***
S Andes (3)/C Andes (4)/NW Arg (3)	10	0.03***	**0.15****	0.17***
S Andes (3)/C Andes (4)	7	0.02***	**0.15***	0.17***
C Andes (4)/NWArg (3)	7	0.02***	**0.16***	0.18***
Control Region: Global F_{st} = 0.097***				
SC Andes (3)/Argentina (5)	8	0.02**	**0.13***	0.15***

Andes: *South Andes:* Mapuche, Pehuenche, Yaghan; *Central Andes*: (i) *North Central Andes*: Yungay, Tayacaja, Arequipa, Jaqaru, and Ancash; (ii) *South Central Andes*: Aymara, Quechua, Aymara Beni, Quechua Beni, Aymara Puno, Quechua Puno, LaPaz, Coya, Aymara Titicaca, and Quechua Titicaca; *NW Argentina*: Catamarca, Tucuman, and Salta.

Non-Andes: Arsario, Ijka, Coreguaje, Kogui, Wayuu, Vaupe, Secoya-Siona, Guahibo, Zoró, Xavante, Ayoreo, Aché, Gaviao, Cayapa, Misiones, Corrientes, RioNegro, and Buenos Aires; *Bolivian Lowlands*: Ignaciano, Movina, Trinitario, Yucarare, Chimane, and Moseten; *Chaco*: Toba Chaco, Toba Formosa, Wichi Chaco, Wichi Formosa, Pilaga Formosa, and Formosa.

*** $P < 0.001$,

** $P < 0.01$,

* $P < 0.05$.

NS: nonsignificant.

Values in **bold:** significative cases where there was geographic structure.

[a,b] Number of populations included in each group.

lyzed. Focusing on the central Andes, the 6-STR analysis showed a genetic structure when samples were divided into two groups (north central Andes vs. middle and south central Andes). For the minimal haplotype and 12-STRs, we could not check this differentiation, because only one north sample was available (Kichwa). However, an absence of genetic structure for middle central Andes versus south central Andes was observed. Similar results were found when the Cayapa sample was included into the north central Andean group.

The Rst genetic distance matrices showed 71% (6-STRs and 12-STRs), and 81% (minimal haplotype) of significant distances. The average distance between all pairs of samples was 0.22 (6-STRs, minimal haplotype) and 0.17 (12-STRs). The mean distance between non-Andean samples was 0.27 (6-STRs), 0.32 (minimal haplotype), and 0.17 (12-STRs), seven, five, and almost three times higher than the values between Andean samples (0.04, 6-STRs; 0.07, minimal haplotype, and 12-STRs). Focusing on the Andean groups, all the analyses highlighted their proximity, with the exception of the Kichwas from Ecuador and the Mapuches. The 6-STR analysis showed that 69.6% of the distances were not statistically significant; the pairs made up of Kichwa were the most significant. Both, the minimal haplotype and the 12-STR analyses revealed nonsignificant distances between all pair of populations composed of Aymara, Colla, Kolla, Diaguita, and Peru. The Quechuas presented nonsignificant distances with Aymara, Kolla, and Diaguita. On the other hand, all distances between the Kichwa and the other Andean samples were significant, except with Colla and Mapuche.

The Rst distances depicted in the MDS plots (see Fig. 5) showed the Andean samples to be relatively grouped and most of the non-Andean populations scattered on the plots. The 6-STR plot (Fig. 5A) revealed the Chimane, Ticuna, Yanomami, Mbyá-Guarani, Bari, Yukpa, and Lowlands as the more distant populations. On the

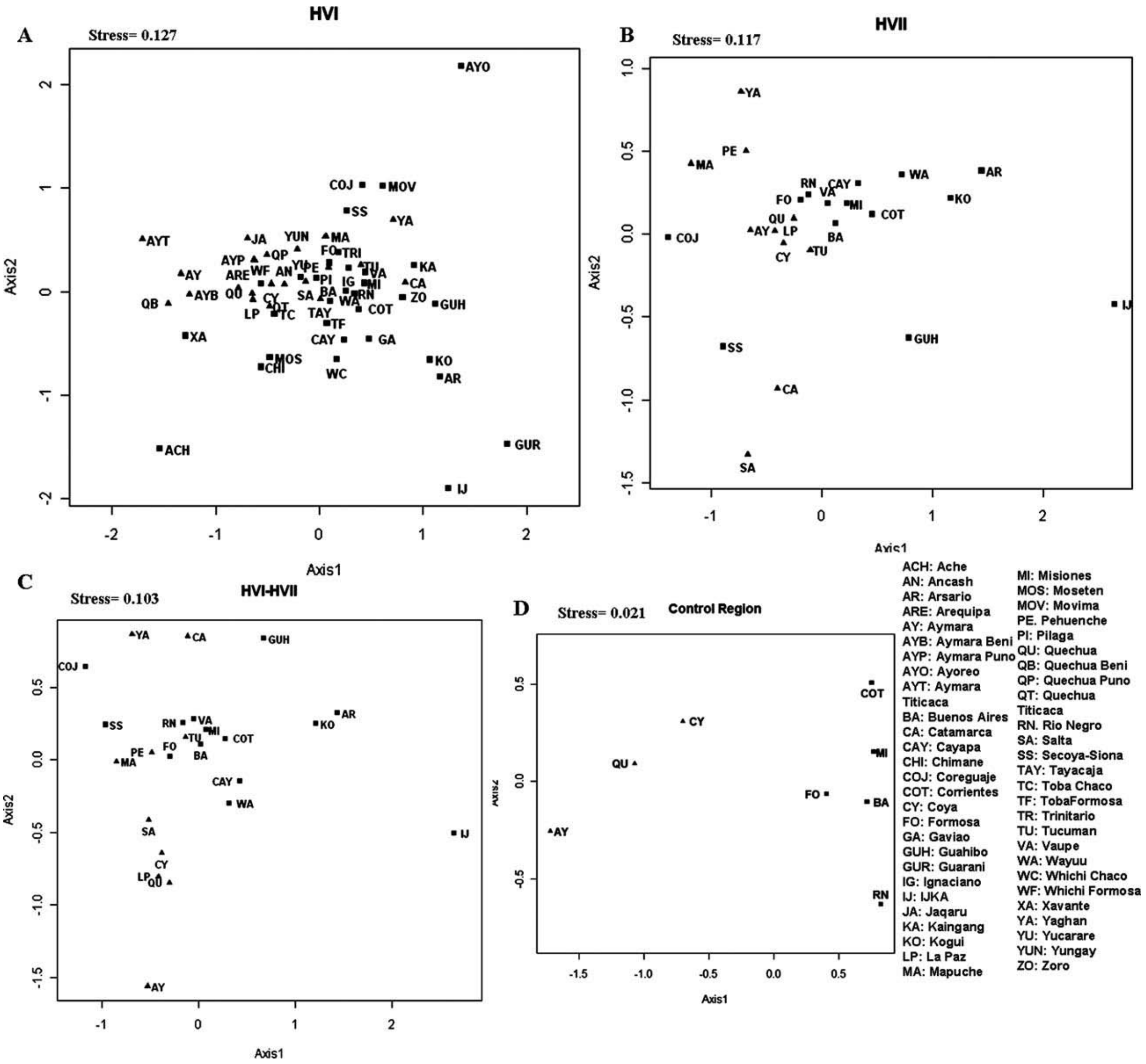


Fig. 4. MDS constructed from mtDNA Tamura and Nei genetic distances. **A:** mtDNA HVI region, (**B**) mtDNA HVII region, (**C**) HVI-HVII regions, (**D**) control region. Triangles and circles represent Andean and non-Andean samples, respectively.

contrary, the Toba, Wichi, Trinitario, Karitiana, and Gaviao-Zoro-Surui samples were the closest to the Andean group, especially to the Susque, Tayacaja, Kichwa, and Cayapa. The two Bolivian, the southernmost Peruvian, and the other northwest Argentinean samples were the most separated from the non-Andean samples. Both, the minimal haplotype and the 12-STR plots (Fig. 5B,C) showed the Kichwas (Ecuador) to be relatively separated from the other six central Andean samples, which formed a group. The Mojeño, Mapuche, Trinitario, Kichwa, and Kaingang-Guarani samples were relatively close to each other.

DISCUSSION

The analysis of both mtDNA and Y-chromosome in this study adds a new perspective to the autosomal data from Gayà-Vidal et al. (2010) for the genetic characterization of the two main linguistic groups in Bolivia, the Aymaras, and Quechuas and contributes new data on Native American genetic variability.

External contributions to the current gene pool of Bolivian populations

Previous studies on classical markers indicated low external admixture in the two Bolivian samples here examined; around 1% of the specific European haplotype GM5*;3 (Dugoujon JM, personal communication) and 98% of O group from the ABO system (hematological study by the IBBA). In the present study, differences were found by gender. The estimates of the non-Amerindian Y-chromosome and mtDNA haplogroups indicated a total absence of admixture for the mtDNA, but a certain proportion of Y-chromosomes admixture, especially in Quechuas which had 22% of non-Native American Y-chromosomes. This differentiation between the two types of markers is a general trend in all Native American

TABLE 5. Diversity data for Y-STRs haplotypes from populations included in the analyses, considering only Native American haplogroups

	Minimal haplotype		12-STRs		6-STRs	
Population	Haplotype diversity	Expected heterozygosity	Haplotype diversity	Expected heterozygosity	Haplotype diversity	Expected heterozygosity
Chimane	0.378	0.106	0.378	0.096	0.200	0.067
Yukpa	0.303	0.168			0.303	0.152
Bari	0.517	0.259			0.517	0.189
Yanomami	0.946	0.337			0.818	0.236
Toba	0.942	0.429	0.963	0.428	0.834	0.296
Kaingang-Guarani	0.906	0.460	0.914	0.406	0.886	0.451
Mojeño	0.889	0.472	0.889	0.438	0.889	0.421
Diaguita	0.972	0.500	0.972	0.481	0.944	0.398
Mapuche	0.988	0.502	0.988	0.467	0.964	0.430
Quechua	**0.960**	**0.516**	**0.976**	**0.462**	**0.780**	**0.446**
Trinitario	0.972	0.535	0.972	0.532	0.943	0.506
Colla	0.778	0.541	0.889	0.467	0.778	0.470
Aymara	**0.971**	**0.551**	**0.982**	**0.486**	**0.894**	**0.484**
Peru	0.989	0.570	0.995	0.507	0.962	0.506
Kolla	0.984	0.593	1.000	0.506	0.909	0.503
Kichwa	1.000	0.604	1.000	0.548	0.989	0.551
Karitiana					0.250	0.042
Wichi					0.909	0.316
Ticuna					0.698	0.323
Gaviao-Zoro-Surui					0.882	0.324
Mbya-Guarani					0.854	0.363
Humahuaca					0.978	0.385
Susque					0.967	0.443
Lowlands					0.954	0.461
Cayapa					0.963	0.501
Tayacaja					0.980	0.507
Arequipa					0.952	0.554

Values in bold: the two populations of this study.

populations; a consequence of the colonization by the Europeans. Nevertheless, it is worth noting that the Aymaras showed a remarkably low level of admixture with less than 2% of non-Native American Y-chromosomes compared to other Andean samples (Dipierri et al., 1998; Tarazona-Santos et al., 2001; González-Andrade et al., 2007) that highlights its isolation from non-Native Americans.

Genetic variation in Aymaras and Quechuas from Bolivia

mtDNA. The results revealed typical Andean characteristics in the two Bolivian samples as well as a certain degree of differentiation between them. Thus, haplogroup B2, the most frequent in the Andean region, was the most frequent in both the Aymaras (81%) and Quechuas (61%). It is interesting to note that around 60% of the diversity of the B2 haplotypes corresponded to groups defined by the variants 16,168, 16,188, 103–143, and 146–215. Particularly, the variant 16,188 was observed in 31% (Aymaras) and 21% (Quechuas) of B2 haplogroups. This variant seems to be characteristic of the Andean Altiplano, because it was also present in Aymara Titicaca (66%), Quechua Titicaca (38%), La Paz (44%), Coya (15%), Aymara Beni (11%), Quechua Beni (81%), and Arequipa (31%). Moreover, in a subgroup of these samples (Aymara, Quechua, Aymara Titicaca, Quechua Titicaca, La Paz, and Coya), the variant 16,188 was always combined with the variant 16183C. Within this subbranch, the 186 variant was present in half of the haplotypes and other minor clusters were defined by the 63–64 variants, the lack of variant 73, and so forth. The 186 variant and the lack of the 73 variant were also found in Coya and La Paz. The presence of the 16,188 variant in one individual of two populations from the Chaco region (Wichi Formosa and Pilaga) could indicate interactions between these two regions. Traces of contacts between different South American regions are also supported by the presence of two haplotypes typical of the Guaraní (combination 16,239A–16,266, Marrero et al., 2007) and northwest Argentinean lineages (combination 16,242–16,311, Tamm et al., 2007) in our Quechua sample.

On the other hand, it is interesting to discuss several particular characteristics found in the mtDNA of the two Bolivian samples. First, the Quechua haplotype 39 presented a 106–111d, also reported in one individual from La Paz (LPAZ070) sharing the same haplotype (considering the HVI and HVII regions separately). The 106–111d was proposed to be characteristic of Chibchan-speaking populations (Santos and Barrantes, 1994; Kolman et al., 1995). However, in those samples, the deletion occurred within haplogroup A2 and not B2 as in the two Andean cases, indicating a recurrent mutation rather than a trait restricted to a certain group. Note that this deletion is different from the 105 to 110d reported in the Coyas. Second, the haplotypes 18 and 19 (haplogroup A2) presented some mutations (lack of the haplogroup A diagnostic site 235, variants 16,512, 16,547, 16,551iG, and absence of the 64 variant) also found (except for 16551iG) in one Coya individual. Third, haplotype 51, with such a particular mutation combination, highlights the huge variability between the 55 and 71 positions in the HVII. These features highlight the importance of sequencing not only the HVI, nor the HVI and HVII separately, because interesting polymorphisms are located outside the classical segments.

TABLE 6. Hierarchical AMOVA analyses with Y chromosome 6 STRs, minimal haplotype, and 12 STRs data

		Hierarchical F_{ST} Analyses		
Population groups[a]	No. pops	Within groups	Among groups	Total F_{ST}
6 STRs Global F_{st} = 0.239***				
Andes (12)/non-Andes (14)	26	0.201***	0.015NS	0.213***
C Andes (11)/Chaco (2)/Bol_Low (3)	16	0.054***	**0.062****	0.113***
NC Andes (2)/MC Andes, SC Andes (9)	11	0.008NS	**0.056***	0.064***
MC Andes (4)/SC Andes (5)	9	0.000NS	0.008NS	0.008NS
Minimal haplotype Global F_{st} = 0.275***				
Andes (8)/non-Andes (8)	16	0.256***	0.048NS	0.292***
C Andes (7)/Bol_Low (3)	10	0.081***	**0.131***	0.202***
MC Andes (3)/SC Andes (3)	6	0.009NS	0.006NS	0.015NS
12 STRs Global F_{st} = 0.279***				
Andes (8)/non-Andes (5)	13	0.242***	0.095NS	0.314***
C Andes (7)/Bol_Low (3)	10	0.085***	**0.126***	0.200***
MC Andes (3)/SC Andes (3)	6	0.116NS	0.000NS	0.016NS

*** $P < 0.001$,
** $P < 0.01$,
* $P < 0.05$;
NS, nonsignificant. Values in **bold:** cases with geographic structure.
[a] In parentheses, number of populations included in each group.
Andes: Mapuche, *Central Andes:* (i) *North Central Andes:* Kichwa and Tayacaja; (ii) *Middle Central Andes*: Peru, Arequipa, Aymara, and Quechua; (iii) *South Central Andes:* Colla, Susque, Humahuaca, Kolla, and Diaguita.
Non-Andes: Bari, Toba, Kaingang_Guarani, Yanomani, Yukpa, Trinitario, Chimane, Mojeños, Gaviao-Zoro-Surui, Karitiana, Ticuna, Mbyá-Guaraní, Wichi, and Cayapa.

Considering the CR analysis, an important result was the high-mtDNA diversity observed in the two Bolivian samples, especially in the Quechuas, which was similar to the Coyas. This high-mtDNA CR diversity in the two Bolivian and other Andean samples confirm strongly the findings of Fuselli et al. (2003), suggesting a high-long-term effective population size in the Andean region. Higher values were found in some Argentinean samples, but these are most probably due to their mixed nature, because they correspond to a political subdivision.

Y-chromosome. All Native American Y-chromosomes in Quechuas and 89% of Aymaras belonged to haplogroup Q1a3a, which is the most frequent haplogroup in South America. The remaining Aymaras presented the para-group Q1a3* (11%), a value that is double that reported in other Bolivian samples (Bailliet et al., 2009). In any case, this study supports that the northwest border of South America harbors the highest frequencies of the Q1a3* lineage, as proposed in Bailliet et al. (2009). This high-Q1a3* frequency in Aymaras could be attributed to drift effects, but the high-diversity values (haplotype) observed in Aymaras, as well as in other Andean samples, is not consistent with this interpretation.

Concerning the Y-STR variation, an interesting result was the high frequency of the DYS393*14 allele in the two Bolivian samples; 56 and 58% of Aymaras and Quechuas, respectively. In the context of the central Andes, the average frequency of this allele was 19% (Ecuador), 40% (Peru), 42% (northwest Argentina), and 57% (Bolivia), which may indicate that its origin was in the Andean Altiplano with a subsequent expansion to the surrounding areas. These results support the study of Martínez-Marignac et al. (2001) that found that northwest Argentinean samples were characterized by a high frequency of this allele (38.9%), which was suggested to have a likely Altiplano origin, because most of surnames in the region were of Aymara origin. However, the South American distribution of this allele presents two discontinuous regions of high frequencies: on one hand, the central Andes with frequencies ranging from 12% in the Cayapas to 80% in Humahuaca, and on the other hand, the Venezuelan samples, Bari and Yukpa, with values around 90%. The total discontinuity between these two areas suggests two different events for the origin of this allele, and in order to verify this, we analyzed the composition of the haplotypes carrying this allele for the minimal haplotype. All the Andean samples shared haplotypes that were unique to their group (except a Toba individual that carried a haplotype also present in the Quechuas), and all the haplotypes in the Bari and Yukpa were unique. Moreover, when we removed the two most variable STRs (DYS385a/b), similarly the Andean samples shared haplotypes only with Andeans, except for one Aymara individual carrying a Bari haplotype, one Peru individual carrying a Yukpa haplotype, and the only Chimane individual carrying a haplotype also present in Kichwa. These results support the hypothesis of two independent origins.

Genetic relationships among Native South Americans

Our results revealed a similar value (around 25%) of between-population genetic diversity among South Americans for both sets of markers and failed to indicate a strong clustering of to the two main geographic areas in South America (west vs. east). However, different patterns of variation were observed in the Andean region compared to the east. Eastern samples presented larger within-group genetic distances and lower intrapopulation diversity parameters than for the Andean samples for both sets of markers. This is consistent with different patterns of drift and gene flow related to larger effective population sizes in the Andean area, as suggested by different kinds of data (mtDNA, Fuselli et al., 2003; Y-chromosome, Tarazona-Santos et al., 2001; classical markers, Luiselli et al., 2000; STRs, Wang et al., 2007, and PAIs, Gayà-Vidal et al., 2010). High diversities have also been reported in the Andean surrounding areas of Chaco and the Bolivian lowlands that may suggest a certain influence from the Andean region (also reflected in the MDS

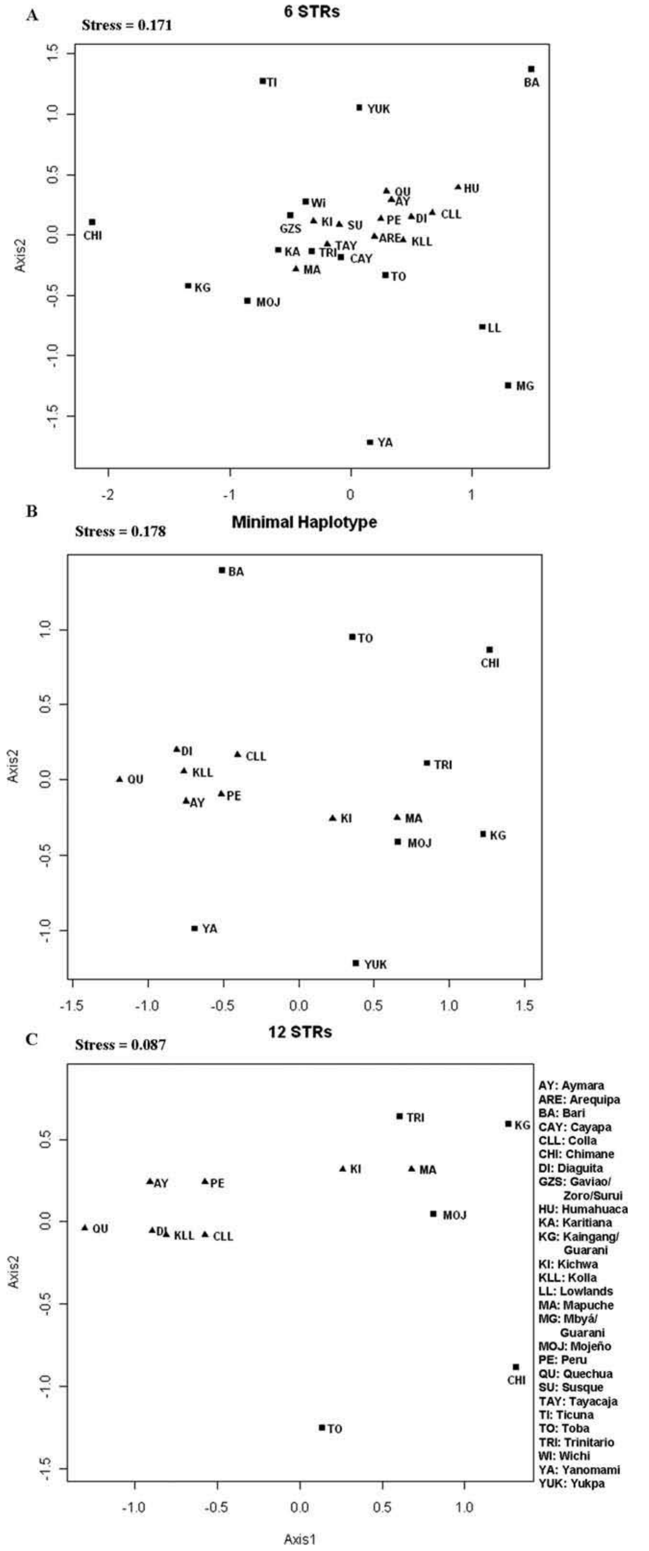


Fig. 5. MDS constructed from Y-chromosome Rst genetic distances. Considering: (**A**) 6 STRs, (**B**) the minimal haplotype, and (**C**) 12 STRs. Triangles and circles represent Andean and non-Andean samples, respectively.

plots), mainly for the mtDNA data, because no genetic structuring was found according to hierarchical AMOVA analyses. On the contrary, this Andean influence has not been strong enough to avoid significant genetic structuring for the Y-chromosome among these three areas.

Regarding the Andean range, the most important genetic differentiation was observed between south versus central Andean populations for both the mtDNA and Y-chromosome data, in agreement with geographical distance.

Concerning the central Andean region, both markers revealed a general homogeneity of this area according to the hierarchical AMOVA analyses and genetic distances if we exclude the Kichwa (Ecuador) for the Y-chromosome, and the three northwest Argentinean samples (Salta, Tucuman, and Catamarca) for the mtDNA. However, the Coyas, also in northwest Argentina, appeared very close to the Bolivian and Peruvian samples. This fact could be attributed to geographic distance (Tucuman and Catamarca) and political rather than ethnic subdivisions (Salta, Tucuman, and Catamarca), but it is most probably due to the Altiplano influence on the Coyas. In fact, the term Coya was used by the Incas to refer to the Aymara inhabitants. The Incas conquered the Aymara territories forming the southeastern provincial region of the Inca Empire, called "Collasuyu."

Focusing on our samples, for the HVI region, the mean genetic distance between the Peruvian samples (4) and the Aymaras, Quechuas, and Coyas was 0.17, 0.07, and 0.05, respectively. Moreover, the mean distance between the Peruvian samples and the Bolivian Aymaras (5) and the Bolivian Quechuas (4) was 0.12 and 0.08, respectively; the HVI plot showed the Aymaras slightly separated from the Peruvians and Coya. The HVI-HVII comparisons highlighted the separation of the Aymaras from the Quechuas, Coyas, and LaPaz, which clustered together. These results lead to the conclusion that (i) the Altiplano region, including the Aymaras, Quechuas, and Coyas present a certain degree of similarity, probably due to the ancient Aymara influence area; (ii) the Quechua and Coya samples would have received more Peruvian influences, probably during the Inca Empire that also imposed the Quechua language; and (iii) the Aymaras would have remained more isolated, thus maintaining certain mtDNA characteristics.

As for the Y-chromosome, the analyses of 6-STRs revealed a clear concordance between genetics and geography, with the highest genetic distance between the Kichwa and Humahuaca. All the analyses highlighted the differentiation of the Kichwas from the other central Andean samples, which formed a group. It is interesting to note that for the minimal haplotype most of the Andean samples (Aymara, Peru, Colla, and Diaguita) only shared haplotypes with other Andean samples. The minimal haplotype 13-14-31-23-10-16-14-15-18 was the most frequent among Andean samples, shared by Aymara, Quechua, Kolla, Colla, and Peru samples, and therefore, characteristic of the central Andes. The Quechua, Kolla, and Kichwa shared one haplotype with Toba from the Chaco area (when 12 STRs were considered, only the Quechua shared it), indicating gene flow into Andean populations from this area. These results indicate that male gene flow inside the Andean region, especially within the south central Andes, has been remarkably high.

Genetic relationships between the two Bolivian populations

In a previous study (Gayà-Vidal et al., 2010), the two Bolivian samples presented, according to 32 PAIs, a genetic similarity and a separation from the two Peruvian Quechua-speaker samples (Tayacaja and Arequipa) from the literature. This suggested a common origin of the two Bolivian populations and an expansion of the Quechua language mainly due to cultural rather than demographic processes. In the present study, the Tayacaja sample also appeared as one of the most differentiated Andean populations. However, a larger number of Andean samples was available for the comparisons, permitting more consistent conclusions, taking into account that the two systems are just two independent loci.

In this study, the comparison of the two Bolivian populations revealed more genetic differences for the mtDNA than for the Y-chromosome; that is, both markers reveal different histories. It is commonly accepted that the social organization of Andean populations was a patrilocal system. Under this assumption, more mtDNA similarities would be expected between the two Bolivian samples, unless a higher proportion of gene flow from external areas affected the Quechuas, as demonstrated by the differences in the frequency of haplogroup B2 and the presence of particular haplotypes from other non-Andean areas. Likewise, it is important to remember the presence in our populations, that is, in the Altiplano, of specific mtDNA features possibly related to high-long-term effective sizes since ancient times. The history of Y-chromosome is different. The distribution of the Y-chromosome variation indicates a clear genetic homogeneity inside the whole central Andean region. This homogeneity could be explained by the higher mobility of males than females across the entire region that might have been favored during the Inca Empire. This apparent controversy could be explained by the different nature of the markers analyzed to date.

CONCLUSION

We can hypothesize a demographic scenario to explain the information supplied by the three kinds of genetic data. According to the very low mutation rate of autosomal Alu markers, these data suggest a past common origin of the Altiplano populations, including the current Aymaras and Quechuas from Bolivia. The arrival of the Inca Empire stimulated the movement of people across the Andean region (probably by the *mitma* system). These movements were especially effective in changing the language (imposition of Quechua), but some regions presented important resistance, including the Titicaca Basin (Aymaras). The demographical consequences of these displacements would have been restricted to the beginning period, according to the very low-genetic distances between these two populations. But, the new Quechua-speaking areas would have been more permeable to the incorporation of foreigners. This is consistent with the closer genetic distances of the Quechuas to the Peruvians and Coyas and the presence of other South American lineages. Finally, in this context, the Y-chromosome homogeneity suggests an important male mobility in the Andean area. Nevertheless, data on additional central Andean samples and more markers are necessary to confirm this scenario.

ACKNOWLEDGMENTS

We are grateful to all the donors for providing DNA samples and to the people who contributed to the collection of samples. We acknowledge support from the CRGS platform of Genopole Toulouse Midi-Pyeénées, where mtDNA and NRY sequencing/genotyping work was performed. We are grateful to the Service de Coopération et d'Action Culturelle (in particular we thank Marie-Claude Senemaud) of French embassy in La Paz and to Mr. Alain Fouquet (the French ambassador). Genealogical information was collected by IBBA, with the help of Emile Crognier and André Sevin (CNRS). We particularly thank E. Guitard, C. Thèves, S. Mazières, C. Coudray, N. Brucato, and H. Razafindrazaka for their technical help and advice.

LITERATURE CITED

Afonso Costa H, Carvalho M, Lopes V, Balsa F, Bento AM, Serra A, Andrade L, Anjos MJ, Vide MC, Pantoja S, Vieira DN, Corte-Real F. 2010. Mitochondrial DNA sequence analysis of a native Bolivian population. J Forensic Leg Med 17:247–253.

Altuna ME, Modesti NM, Demarchi DA. 2006. Y-chromosomal evidence for a founder effect in Mbyá-Guaraní Amerindians from northeast Argentina. Hum Biol 78:635–639.

Alvarez-Iglesias V, Jaime JC, Carracedo A, Salas A. 2007. Coding region mitochondrial DNA SNPs: targeting East Asian and Native American haplogroups. Forens Sci Int Genet 1:44–55.

Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJ, Staden R, Young IG. 1981. Sequence and organization of the human mitochondrial genome. Nature 290:457–465.

Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. 1999. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat Genet 23:147.

Bailliet G, Ramallo V, Muzzio M, García A, Santos MR, Alfaro EL, Dipierri JE, Salceda S, Carnese FR, Bravi CM, Bianchi NO, Demarchi DA. 2009. Brief communication: restricted geographic distribution for Y-Q* paragroup in South America. Am J Phys Anthropol 140:578–582.

Bandelt HJ, Forster P, Röhl A. 1999. Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol 16:37–48.

Bandelt HJ, Herrnstadt C, Yao YG, Kong QP, Kisivild T, Rengo C, Scozzari R, Richards M, Villems R, Macaulay V, Howell N, Torroni A, Zhang YP. 2003. Identification of Native American founder mtDNAs through the analysis of complete mtDNA sequences: some caveats. Ann Hum Genet 67:512–524.

Bandelt HJ, Lahermo P, Richards M, Macaulay V. 2001. Detecting errors in mtDNA data by phylogenetic analysis. Int J Legal Med 115:64–69.

Bandelt HJ, Macaulay V, Richards M. 2000. Median networks: speedy construction and greedy reduction, one simulation, and two case studies from human mtDNA. Mol Phylogenet Evol 16:8–28.

Bandelt HJ, Quintana-Murci L, Salas A, Macaulay V. 2002. The fingerprint of phantom mutations in mitochondrial DNA data. Am J Hum Genet 71:1150–1160.

Barbieri C, Heggarty P, Castrì L, Luiselli D, Pettener D. 2011. Mitochondrial DNA variability in the Titicaca basin: matches and mismatches with linguistics and ethnohistory. Am J Hum Biol 23:89–99.

Bert F, Corella A, Gené M, Pérez-Pérez A, Turbón D. 2001. Major mitochondrial DNA haplotype heterogeneity in highland and lowland Amerindian populations from Bolivia. Hum Biol 73:1–16.

Bert F, Corella A, Gené M, Pérez-Pérez A, Turbón D. 2004. Mitochondrial DNA diversity in the Llanos de Moxos: Moxo, Movima and Yuracare Amerindian populations from Bolivia lowlands. Ann Hum Biol 31:9–28.

Bianchi NO, Catanesi CI, Bailliet G, Martinez-Marignac VL, Bravi CM, Vidal-Rioja LB, Herrera RJ, Lopez-Camelo JS.

1998. Characterization of ancestral and derived Y-chromosome haplotypes of New World Native populations. Am J Hum Genet 63:1862–1871.

Blanco Verea A, Jaime JC, Brión M, Carracedo A. 2010. Y-chromosome lineages in native South American population. Forens Sci Int Genet 4:187–193.

Bobillo MC, Zimmermann B, Sala A, Huber G, Röck A, Bandelt HJ, Corach D, Parson W. 2010. Amerindian mitochondrial DNA haplogroups predominate in the population of Argentina: towards a first nationwide forensic mitochondrial DNA sequence database. Int J Legal Med 124:263–268.

Bouysse-Cassagne T. 1986. Urco and Uma: Aymara concepts of space. In: Murra JV, Wachtel N, Revel J, editors. Anthropological history of Andean polities. Cambridge: Cambridge University Press. p 201–227.

Browman DL. 1994. Titicaca Basin archaeolinguistics: Uru, Pukina and Aymara AD 750–1450. World Archaeol 26:235–251.

Cabana GS, Merriwether DA, Hunley K, Demarchi DA. 2006. Is the genetic structure of Gran Chaco populations unique? Interregional perspectives on native South American mitochondrial DNA variation. Am J Phys Anthropol 131:108–119.

Cinnioglu C, King R, Kivisild T, Kalfõglu E, Atasoy S, Cavalleri GL, Lillie AS, Roseman CC, Lin AA, Prince K, Oefner PJ, Shen P, Semino O, Cavalli-Sforza LL, Underhill PA. 2004. Excavating Y-chromosome haplotype strata in Anatolia. Hum Genet 114:127–148.

Corella A, Bert F, Pérez-Pérez A, Gené M, Turbón D. 2007. Mitochondrial DNA diversity of the Amerindian populations living in the Andean Piedmont of Bolivia: Chimane, Moseten, Aymara and Quechua. Ann Hum Biol 34:34–55.

Coudray C, Olivieri A, Achilli A, Pala M, Melhaoui M, Cherkaoui M, El-Chennawi F, Kossmann M, Torroni A, Dugoujon JM. 2009. The complex and diversified mitochondrial gene pool of Berber populations. Ann Hum Genet 73:196–214.

Dipierri JE, Alfaro E, Martínez-Marignac VL, Bailliet G, Bravi CM, Cejas S, Bianchi NO. 1998. Paternal directional mating in two Amerindian subpopulations located at different altitudes in northwestern Argentina. Hum Biol 70:1001–1010.

Dornelles CL, Battilana J, Fagundes NJ, Freitas LB, Bonatto SL, Salzano FM. 2004. Mitochondrial DNA and Alu insertions in a genetically peculiar population: the Ayoreo Indians of Bolivia and Paraguay. Am J Hum Biol 16:479–488.

Excoffier L, Laval G, Schneider S. 2005. Arlequin ver. 3.0: an integrated software package for population genetics data analysis. Evol Bioinform Online 1:47–50.

Excoffier L, Smouse P, Quattro J. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131:479–491.

Fuselli S, Tarazona-Santos E, Dupanloup I, Soto A, Luiselli D, Pettener D. 2003. Mitochondrial DNA diversity in South America and the genetic history of Andean highlanders. Mol Biol Evol 20:1682–1691.

Gayà-Vidal M, Dugoujon JM, Esteban E, Athanasiadis G, Rodríguez A, Villena M, Vasquez R, Moral P. 2010. Autosomal and X chromosome Alu insertions in Bolivian Aymaras and Quechuas: two languages and one genetic pool. Am J Hum Biol 22:154–162.

González-Andrade F, Sánchez D, González-Solórzano J, Gascón S, Martínez-Jarreta B. 2007. Sex-specific genetic admixture of Mestizos, Amerindian Kichwas, and Afro-Ecuadorans from Ecuador. Hum Biol 79:51–77.

Iannacone GC, Tito RY, Lopez PW, Medina ME, Lizarraga B. 2005. Y-chromosomal haplotypes for the PowerPlex Y for twelve STRs in a Peruvian population sample. J Forens Sci 50:239–242.

Jobling MA, Tyler-Smith C. 2003. The human Y chromosome: an evolutionary marker comes of age. Nat Rev Genet 4:598–612.

Karafet TM, Mendez FL, Meilerman MB, Underhill PA, Zegura SL, Hammer MF. 2008. New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. Genome Res 18:830–838.

Kolata A. 1993. The Tiwanaku: portrait of an Andean civilization. The Peoples of America. Cambridge: Blackwell.

Kolman CJ, Bermingham E, Cooke R, Ward RH, Arias TD, Guionneau-Sinclair F. 1995. Reduced mtDNA diversity in the Ngöbé Amerinds of Panamá. Genetics 140:275–283.

Lee J, Ewis AA, Hurles ME, Kashiwazaki H, Shinka T, Nakahori Y. 2007. Y chromosomal STRs haplotypes in two populations from Bolivia. Legal Med (Tokyo) 9:43–47.

Leite FP, Callegari-Jacques SM, Carvalho BA, Kommers T, Matte CH, Raimann PE, Schwengber SP, Sortica VA, Tsuneto LT, Petzl-Erler ML, Salzano FM, Hutz MH. 2008. Y-STR analysis in Brazilian and South Amerindian populations. Am J Hum Biol 20:359–363.

Lewis CM Jr., Lizarraga B, Tito RY, Lopez PW, Iannacone C, Medina A, Martinez R, Polo SI, De La Cruz AF, Caceres AM, Stone AC. 2007. Mitochondrial DNA and the peopling of South America. Hum Biol 79:159–178.

Lewis CM Jr., Tito RY, Lizárraga B, Stone AC. 2005. Land, language, and loci: mtDNA in Native Americans and the genetic history of Peru. Am J Phys Anthropol 127:351–360.

Luiselli D, Simoni L, Tarazona-Santos E, Pastor S, Pettener D. 2000. Genetic structure of Quechua-speakers of the Central Andes and geographic patterns of gene frequencies in South Amerindian populations. Am J Phys Anthropol 113:5–17.

Marrero AR, Silva-Junior WA, Bravi CM, Hutz MH, Petzl-Erler ML, Ruiz-Linares A, Salzano FM, Bortolini MC. 2007. Demographic and evolutionary trajectories of the Guarani and Kaingang Natives of Brazil. Am J Phys Anthropol 132:301–310.

Martínez-Marignac VL, Bailliet G, Dipierri JE, Alfaro E, López-Camelo JS, Bianchi NO. 2001. Variabilidad y antigüedad de linajes holandricos en poblaciones jujeñas. Rev Arg Antropol Biol 3:65–77.

Mazières S, Guitard E, Crubézy E, Dugoujon JM, Bortolini MC, Bonatto SL, Hutz MH, Bois E, Tiouka F, Larrouy G, Salzano FM. 2008. Uniparental (mtDNA, Y-chromosome) polymorphisms in French Guiana and two related populations—implications for the region's colonization. Ann Hum Genet 72:145–156.

Meyer S, Weiss G, von Haeseler A. 1999. Pattern of nucleotide substitution and rate of heterogeneity in the hypervariable regions I and II of human mtDNA. Genetics 152:1103–1110.

Moraga ML, Rocco P, Miquel JF, Nervi F, Llop E, Chakraborty R, Rothhammer F, Carvallo P. 2000. Mitochondrial DNA polymorphisms in Chilean aboriginal populations: implications for the peopling of the southern cone of the continent. Am J Phys Anthropol 113:19–29.

Nei M. 1987. Molecular evolutionary genetics. New York: Columbia University Press.

Pfeiffer H, Brinkmann B, Huhne J, Rolf B, Morris AA, Steighner R, Holland MM, Forster P. 1999. Expanding the forensic German mitochondrial DNA control region database: genetic diversity as a function of sample size and microgeography. Int J Legal Med 112:291–298.

Polzin T, Daneschmand SV. 2003. On Steiner trees and minimum spanning trees in hypergraphs. Operations Res Lett 31: 12–20.

Raymond M, Rousset F. 1995. An exact test for population differentiation. Evolution 49:1280–1283.

Rickards O, Martínez-Labarga C, Lum JK, De Stefano GF, Cann RL. 1999. mtDNA history of the Cayapa Amerinds of Ecuador: detection of additional founding lineages for the Native American populations. Am J Hum Genet 65:519–530.

Rowe JH. 1963. Inca culture at the time of the Spanish conquest. In: Steward JH, editor. Handbook of South American Indians, Vol. 2. New York: Cooper Square. p 183–330.

Santos M, Barrantes R. 1994. D-loop mtDNA deletion as a unique marker of Chibchan Amerindians. Am J Hum Genet 55:413–414.

Schmitt R, Bonatto SL, Freitas LB, Muschner VC, Hill K, Hurtado AM, Salzano FM. 2004. Extremely limited mitochondrial DNA variability among the Aché Natives of Paraguay. Ann Hum Biol 31:87–94.

Seielstad M, Yuldasheva N, Singh N, Underhill P, Oefner P, Shen P, Wells RS. 2003. A novel Y-chromosome variant puts an upper limit on the timing of first entry into the Americas. Am J Hum Genet 73:700–705.

Stanish C. 2001. The origin of state societies in South America. Annu Rev Anthropol 30:41–64.

Tamm E, Kivisild T, Reidla M, Metspalu M, Smith DG, Mulligan CJ, Bravi CM, Rickards O, Martinez-Labarga C, Khusnutdinova EK, Fedorova SA, Golubenko MV, Stepanov VA, Gubina MA, Zhadanov SI, Ossipova LP, Damba L, Voevoda MI, Dipierri JE, Villems R, Malhi RS. 2007. Beringian standstill and spread of Native American founders. PLoS One 2:e829.

Tamura K, Nei M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol 10:512–526.

Tarazona-Santos E, Carvalho-Silva DR, Pettener D, Luiselli D, De Stefano GF Labarga CM, Rickards O, Tyler-Smith C, Pena SD, Santos FR. 2001. Genetic differentiation in South Amerindians is related to environmental and cultural diversity: evidence from the Y chromosome. Am J Hum Genet 68:1485–1496.

Tirado M, López-Parra AM, Baeza C, Bert F, Corella A, Pérez-Pérez A, Turbón D, Arroyo-Pardo E. 2009. Y-chromosome haplotypes defined by 17 STRs included in AmpFlSTR Yfiler PCR Amplification Kit in a multi ethnical population from El Beni Department (North Bolivia). Legal Med (Tokyo) 11:101–103.

Torroni A, Schurr TG, Yang CC, Szathmary EJ, Williams RC, Schanfield MS, Troup GA, Knowler WC, Lawrence DN, Weiss KM, Wallace DC. 1992. Native American mitochondrial DNA analysis indicates that the Amerind and the Nadene populations were founded by two independent migrations. Genetics 130:153–162.

Toscanini U, Gusmão L, Berardi G, Amorim A, Carracedo A, Salas A, Raimondi E. 2008. Y chromosome microsatellite genetic variation in two Native American populations from Argentina: population stratification and mutation data. Forens Sci Int Genet 2:274–280.

Tschopik H. 1963. The Aymara. In: Steward JH, editor. Handbook of South American Indians, Vol. 2. New York: Cooper Square. p 501–573.

Underhill PA, Jin L, Zemans R, Oefner PJ, Cavalli-Sforza LL. 1996. A pre-Columbian Y chromosome-specific transition and its implications for human evolutionary history. Proc Natl Acad Sci USA 93:196–200.

Vona G, Falchi A, Moral P, Calò CM, Varesi L. 2005. Mitochondrial sequence variation in the Guahibo Amerindian population from Venezuela. Am J Phys Anthropol 127:361–369.

Wang S, Lewis CMJr, Jakobsson M, Ramachandran S, Ray N, Bedoya G, Rojas W, Parra MV, Molina JA, Gallo C, Mazzotti G, Poletti G, Hill K, Hurtado AM, Labuda D, Klitz W, Barrantes R, Bortolini MC, Salzano FM, Petzl-Erler ML, Tsunedo LT, Llop E, Rothhammer F, Excoffier L, Feldman MW, Rosenberg NA, Ruiz-Linares A. 2007. Genetic variation and population structure in Native Americans. PLoS Genet 3: e185.

Ward RH, Salzano FM, Bonatto SL, Hutz MH, Coimbra CEA JR, Santos RV. 1996. Mitochondrial DNA polymorphism in three Brazilian Indian tribes. Am J Hum Biol 8:317–323.

III.4 Results III: Gayà-Vidal et al., (manuscript 3)

APOE/C1/C4/C2 Gene Cluster Diversity in Native Bolivian samples: Aymaras and Quechuas

Magdalena Gayà-Vidal[1,2], Georgios Athanasiadis[1], Robert Carreras-Torres[1], Marc Via[1], Esther Esteban[1], Mercedes Villena[3], Jean-Michel Dugoujon[2], Pedro Moral[1].

1. Biología Animal, Facultat de Biología, Universitat de Barcelona
2. Laboratoire AMIS, Université de Toulouse III, CNRS, Toulouse
3. Instituto Boliviano de Biología de Altura, La Paz, Bolívia

Abstract

Twenty five polymorphisms (10 STRs and 15 SNPs) from the *APOE/C1/C4/C2* gene cluster and the flanking region (24 markers in 108kb and one STR at 134kb from this region) were investigated in two Native American populations from the Bolivian Andean Altiplano (45 Aymaras and 45 Quechuas) to give new insights into the evolutionary history of this genomic region in Andean populations. In general, diversity in Bolivians was low, with nine out of 15 SNPs and seven out of 10 STRs being practically monomorphic. Part of this reduced diversity could be attributed to selection since the *APOE/C1/C4/C2* region presented a high degree of conservation compared to the flanking genes in both Bolivians and Europeans, which may be due to its physiological importance. Also, the lower genetic diversity in Bolivians compared to Europeans for some markers could be attributed to their different demographic histories. Regarding the *APOE* isoforms, besides the *APOE* ε3 (0.947%) and *APOE* ε4 (0.046%) the presence of *APOE* ε2 (0.007%) was also detected in Bolivians.

INTRODUCTION

Apolipoproteins play an important role in lipid metabolism. One of the most studied ones is apolipoprotein ApoE, involved in lipoprotein metabolism and lipid transport. *APOE* gene is located in chromosome 19q13.2, closely linked to the *APOC1*, *APOC4*, and *APOC2* genes forming the *APOE/C1/C4/C2* gene cluster which expands about 48kb. ApoE, C1, and C2 proteins are constituents of chylomicrons, very low density lipoproteins (VLDL), and high-density lipoproteins (HDL). ApoC4 plays an important role in the metabolism of circulating lipids as an activator of lipoprotein lipase (LPL) (Ken-Dror et al., 2010).

Three common isoforms with different physiological properties have been described for the ApoE: ε2, ε3, and ε4. These isoforms are determined by the variation in two SNPs (rs429358 and rs7412 in the coding residues 112 and 158) (Hanis et al., 1991). In comparison with the most frequent variant (*APOE* ε3), *APOE* ε4 has been associated with higher total and LDL cholesterol levels, and increased risks for cardiovascular and Alzheimer's diseases (Song et al., 2004; Wollmer, 2010). Other polymorphisms, such as -219G→T, located in the *APOE* promoter, have also been associated with cardiovascular, Alzheimer's and Parkinson diseases (Tycko et al., 2004; Artiga et al., 1998).

Although less known, variation in the other components of the cluster has also been described. However, its role in the risk for complex diseases is not clear (Kamboh et al., 2000). For instance, the most studied polymorphism in the *APOC1* gene is a CGTT insertion at position -317 in the promoter that has a negative effect on *APOC1* transcription levels (Xu et al., 1999). In recent years, detailed studies considering the variation in the whole cluster have been performed, mainly in populations from European origin, although most of them have been carried out in epidemiological surveys (Klos et al., 2008; Ken-Dror et al., 2010).

As far as we know, studies in Native Americans have been mainly carried out for the *APOE*, reporting isoform frequencies (Crews et al., 1993; Marin et al., 1997; Scacchi et al., 1997; Andrade et al., 2000; Jaramillo-Correa et al., 2001; and Demarchi et al., 2005). An additional study provided data for two polymorphisms in the *APOC1* and *APOC2* genes in five Brazilian samples (Andrade et al., 2002). Moreover, there is genomewide SNP data available for several Native American groups from the Human Genome Diversity Project (HGDP), although the sample sizes are very small (n<15) (Li et al., 2008; Jakobsson et al., 2008).

In light of all the above, this paper explores for the first time the variation of 25 polymorphisms in two Native Andean populations from Bolivia. Moreover, in order to provide an external group for comparison, a European sample has also been studied. Twenty-four of these markers lie within a 108kb-long region including the

APOE/C1/C4/C2 gene cluster as well as three flanking genes. The main goal of our work was to gain a deeper insight into the evolutionary history of this *APOE/C1/C4/C2* region in Andean populations by using polymorphisms (SNPs and STRs) with different mutation rates, providing the so far most complete data on the variation of this region in Native Americans. Due to the potential functional role of some variants in this region, a good knowledge of their variation in general populations is of great interest for both population and epidemiological purposes.

MATERIALS AND METHODS

Samples

Two Native American samples from Bolivia, belonging to the main native linguistic groups (Aymaras and Quechuas), were collected by the *Instituto Boliviano de Biología de Altura* (IBBA). The sampled Aymara individuals lived near Lake Titicaca and the Quechuas in the Potosí department, both regions located in the Andean Altiplano. The study included a total of 90 unrelated individuals of both sexes (45 for each population sample). A more detailed description of these samples was presented in Gayà-Vidal et al. (2010, 2011).

For broader context, a European sample from Catalonia, Spain (n= 42) was also included in the analyses. All subjects gave their informed consent and the study was approved by the ethical committees from the two institutions involved (IBBA and University of Barcelona).

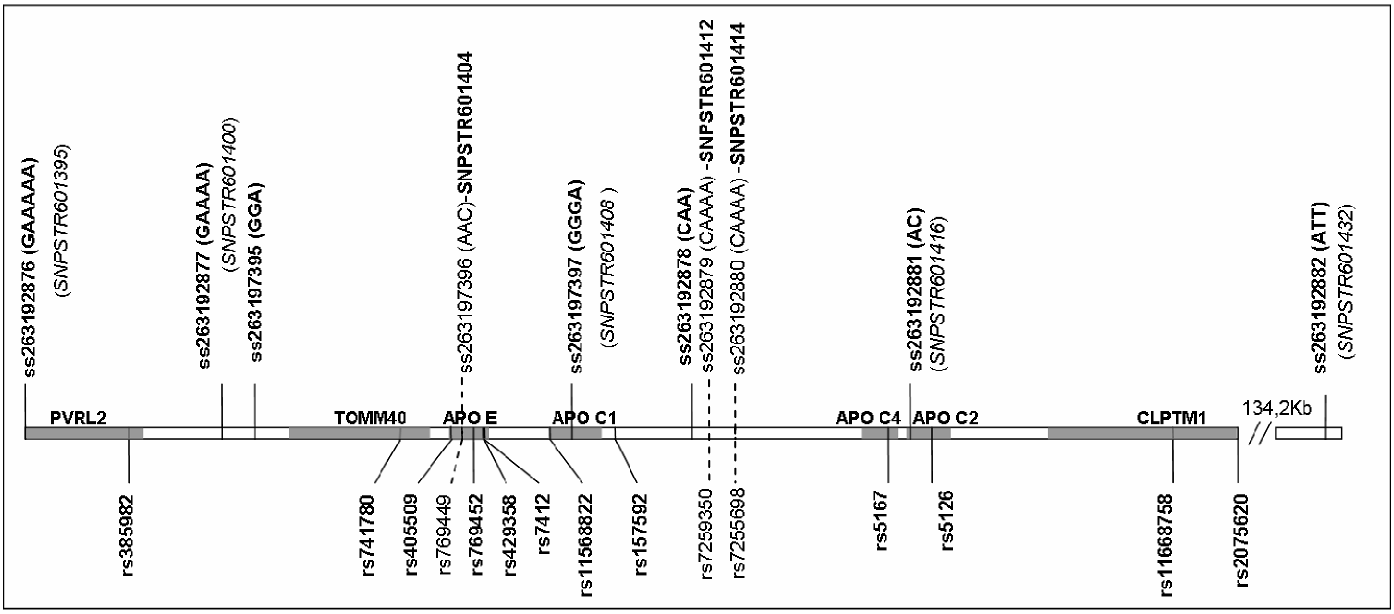
Polymorphisms

The 25 studied markers included 15 SNPs and 9 STRs distributed in the genomic region between positions 50063546 and 50172127 of chromosome 19, containing the *APOE/C1/C4/C2* gene cluster, and one STR located at 134.2kb 3'of this region. ID reference and their relative location on the genomic region is shown in Figure 1.

The SNPs were selected according to different criteria (functional importance, coverage of the flanking region of the *APOE/C1/C4/C2* gene cluster, and formation of SNPSTR markers; Mountain et al., 2002) as shown in Table 1.

Ten STRs were selected after scanning the *APOE/C1/C4/C2* region using the UCSC Genome Browser (http://genome.ucsc.edu/cgi-bin/hgGateway) and the SNPSTR database (Mountain et al., 2002). Nine STRs are located within the region and one located 134.2kb downstream of the region. The analysed STRs are described in Table 2.

Figure 1. Markers examined and their relative location.



In the upper part of the scheme there are the STRs and in the bottom part the SNPs.

Genotyping

With the exception of rs7412, all SNPs were typed with the iPLEX™ Gold assay on the Sequenom MassARRAY® Platform (Sequenom, San Diego, CA, USA). SNP rs7412 was genotyped by Real-Time PCR, using a TaqMan® SNP genotyping assay protocol (Applied Biosystems, Foster City, CA, USA) using a total volume of 5 μl per well. The RT-PCR, fluorescence measurements of the final products and data collection were carried out using an ABI PRISM 7900 HT Sequence Detection System (Applied Biosystems, Foster City, CA, USA).

STRs were genotyped by PCR amplification using standard conditions in a total volume of 10μl. One of the oligonucleotides was modified at the 5'-end with 6-FAM or HEX fluorescent dye. The diluted PCR product was analyzed in an Applied Biosystems 3130 Genetic Analyser using GeneScan™ 500ROX as size standard and genotypes were determined using the ABI Prism GeneMapper® v3.0 software (Applied Biosystems, Foster City, CA, USA). PCR protocols can be provided under request. Positive and negative controls were included in all runs, and the exact number of repeats was verified by direct sequencing.

Table 1. Allele frequencies and heterozygosities for the SNPs.

Gene Location	SNPSTR	SNPs	Aymara	H	Quechua	H	Europeans	H
PVRL2 intron		**rs385982**	(45)	0.533	(44)	0.364	(41)	0.342
		G	0.38		0.34		0.22	
		T	0.62		0.66		0.78	
TOMM40 intron		**rs741780**	(45)	0.511	(44)	0.455	(41)	0.585
		A	0.32		0.32		0.51	
		G	0.68		0.68		0.49	
APOE promoter		**rs405509**	(39)	0.436	(39)	0.461	(41)	0.537
		A	0.68		0.56		0.41	
		C	0.32		0.44		0.59	
APOE intron 2	601404	**rs769449**	(43)	0.070	(42)	0.119	(41)	0.098
		A	0.03		0.06		0.05	
		G	0.97		0.94		0.95	
APOE exon 3		**rs769452** (T/C)	(38)	-	(38)	-	41	-
		T	1		1		1	
APOE exon 4, determining the APOE isoforms		**rs429358**	(39)	0.103	(38)	0.105	(39)	0.103
		C	0.05		0.04		0.08	
		T	0.95		0.95		0.92	
		rs7412	(45)	0.022	(44)	-	(41)	0.195
		C	0.99		1		0.88	
		T	0.01		0		0.12	
APOC1 promoter		**rs11568822**	(45)	0.067	(43)	0.023	(41)	0.415
		-	0.94		0.99		0.77	
		CGTT	0.06		0.01		0.23	
Between *APOC1* and *APOC4*		**rs157592** (G/T)	(41)	-	(41)	-	(41)	-
		G	1		1		1	
	601412	**rs7259350** (C/T)	(45)	-	(44)	-	(41)	-
		C	1		1		1	
	601414	**rs7255698** (C/G)	(45)	-	(44)	-	(41)	-
		C	1		1		1	
APOC4 exon 3		**rs5167**	(45)	0.511	(44)	0.568	(41)	0.537
		G	0.32		0.35		0.37	
		T	0.68		0.65		0.63	
APOC2 exon 4		**rs5126** (A/C)	(39)	-	(40)	-	(41)	-
		A	1		1		1	
CLPTM1 introns		**rs11668758**	(45)	0.467	(44)	0.546	(41)	0.439
		C	0.68		0.66		0.68	
		T	0.32		0.34		0.32	
		rs207562	(39)	0.410	(39)	0.513	(41)	0.439
		A	0.69		0.67		0.68	
		G	0.31		0.33		0.32	
		Mean H*		0.313		0.315		0.369

Numbers inside brackets represent the number of individuals genotyped. H: Observed Heterozygosity. * Considering the 10 polymorphic loci.

Table 2. Allele frequencies and heterozygosities for the STRs.

Gene Location	SNPSTR	STRs	Aymara	H	Quechua	H	Europeans	H
PVRL2 intron	601395*	**ss263192876**	(42)	0.429	(38)	0.395	(39)	0.308
		(TTTTTC)$_6$	0.69		0.72		0.85	
		(TTTTTC)$_7$	0.31		0.28		0.15	
Between *PVRL2* and *TOMM40*	601400*	**ss263192877**	(43)	-	(31)	-	(36)	0.083
		(TTTTC)7	0		0		0.04	
		(TTTTC)$_8$	1		1		0.96	
		ss263197395	(28)	-	(24)	-	(39)	-
		(GGA)$_{10}$	1		1		1	
APOE intron2	601404	**ss263197396**	(24)	-	(29)	-	(38)	-
		(TTG)$_{13}$	1		1		1	
APOC1 intron 3	601408*	**ss263197397**	(19)	-	(20)	-	(10)	-
		(GGGA)$_9$	1		1		1	
Between *APOC1* and *APOC4*		**ss263192878**	(42)	0.024	(44)	-	(41)	0.024
		(AAC)$_7$	0		0		0.01	
		(AAC)$_9$	0.99		1		0.99	
		(AAC)$_{10}$	0.01		0		0	
	601412	**ss263192879**	(45)	0.267	(44)	0.205	(41)	0.268
		(CAAAA)$_3$	0		0		0.04	
		(CAAAA)$_4$	0.18		0.15		0.12	
		(CAAAA)$_5$	0.82		0.85		0.84	
	601414	**ss263192880**	(35)	-	(36)	-	(40)	-
		(TTTTG)$_5$	1		1		1	
APOC2 intron 1	601416*	**ss263192881**	(44)	0.750	(38)	0.763	(41)	0.829
		(TG)(AG)$_{17}$	0.44		0.40		0.16	
		(TG)(AG)$_{19}$	0.01		0		0	
		(TG)(AG)$_{21}$	0.01		0		0.09	
		(TG)(AG)$_{24}$	0.01		0		0	
		(TG)(AG)$_{25}$	0.18		0.18		0.04	
		(TG)(AG)$_{26}$	0.02		0		0	
		(TG)(AG)$_{27}$	0		0.01		0.05	
		(TG)(AG)$_{28}$	0.02		0.14		0.32	
		(TG)(AG)$_{29}$	0.27		0.22		0.13	
		(TG)(AG)$_{30}$	0.01		0.04		0.16	
		(TG)(AG)$_{31}$	0		0		0.05	
		(TG)(AG)$_{32}$	0.01		0		0	
		(TG)(AG)$_{33}$	0		0		0.01	
134.2 Kb 3’	601432*	**ss263192882**	(43)	0.093	(38)	0.053	(41)	0.439
		(ATT)$_8$	0.05		0.03		0.34	
		(ATT)$_9$	0.95		0.97		0.66	
		Average H*		0.313		0.283		0.325

Numbers inside brackets represent the number of individuals genotyped. H: Observed Heterozygosity. * Considering the polymorphic loci.

Statistical analysis

Allele frequencies and heterozygosities were calculated with Genetix v4.05.2 (Belkhir et al., 1998) and Hardy-Weinberg equilibrium (HWE) was assessed with Arlequin v3.1 (Excoffier et al., 2005). Pairwise linkage disequilibrium (LD) was measured as D'and r^2 with Haploview v.4.1 (Barrett et al., 2005) for biallelic markers. Also, the Black and Krafsur test (Black and Krafsur, 1985) and the LD test by 10000 permutations were performed with Genetix v4.05.2. The statistical significance of the non-random distribution of each pair of loci was tested by Fisher's exact test with Genepop v4.0 (Raymond & Rousset, 1995; Rousset, 2008). Haplotype phase was inferred with PHASE v2.1 (Stephens et al., 2001; Stephens & Donnelly, 2003). *APOE* isoforms were coded from rs429358 and rs7412 genotypes and allele and genotype frequencies were calculated with Genetix v4.05.2 (Belkhir et al., 1998).

Population differentiation between samples was tested using an exact G test (Genepop v4; Raymond & Rousset, 1995) for the allele frequencies of the 25 markers.

Comparisons were performed at different levels, depending on the available data from other Native Americans groups: i) three loci (rs405509, rs5167, rs11668758) in five populations fromHGDP-CEPH Project: Karitiana (14 individuals) and Surui (8) from Brazil; Piapoco and Curripaco from Colombia (7); and Maya (21) and Pima (14) from Mexico; ii) five loci (the previous three plus rs5126 and rs2075620) in 58 Mexicans from HapMap Project; iii) one locus (rs11568822) in 5 Brazilian samples (Andrade et al., 2002); and iv) *APOE* isoforms in 37 Native South American populations (see Figure 3).

RESULTS

Allele frequencies and heterozygosities

Allele frequencies for the 15 SNPs are shown in Table 1. All polymorphic SNPs (10/15) were in Hardy-Weinberg equilibrium. Three of the 5 monomorphic SNPs formed SNPSTR markers. In the two Bolivian samples, the polymorphic loci presented, in general, high diversity values (from 0.36 to 0.51) except for four SNPs (*APOE* and *APOC1*) that presented minor allele frequencies ≤ 6%. Average heterozygosity across the 10 polymorphic SNPs was slightly higher in the European sample than in Bolivians (Table 1), although non significant.

Regarding the STRs (Table 2), the variation was remarkably low: five were monomorphic, four were biallelic (one was triallelic in Europeans, and another one was triallelic when the three samples were grouped), and only one was multiallelic with 10 (Aymaras), 6 (Quechuas) and 9 (Europeans) different alleles. Heterozygosity values were relatively low in all samples; average heterozygosities were slightly lower in Bolivians (0.283, Quechua; 0.313, Aymaras) than in the Europeans (0.325), but without significant differences.

It is important to mention that the $(TG)_n(AG)_m$ was first selected as a $(TG)_n$ microsatellite from different databases. However, after sequencing five individuals for validation, variation in the number of repeats in the contiguous $(AG)_m$ was observed, the alleles of 17, 19, 24, 26, and 29 repeats presented different dinucleotide combinations according to the $(TG)_{9/12/17/18/21}(AG)_{7/8}$ pattern, indicating a compound STR.

No significant differences were found in allele frequencies between the two Bolivian populations for any marker after applying the Bonferroni correction, whereas between each Bolivian sample and the European population were significant for two markers (ss263192881, p<0.0001; ss263192882, p<0.0001), and differences existed between Quechuas and the European samples for two SNPs (rs7412, p<0.001; rs11568822, p<0.0001), and between Aymaras and the European sample for one locus (rs405509, p<0.001). These loci also showed differences between Bolivians pooled together and Europeans (rs7412, p<0.0001; rs11568822, p<0.0001; rs405509, p=0.0018).

As for the *APOE* isoforms, allele and genotype frequencies are shown in Table 3. The most common genotype was the homozygote for *APOE* ε3 (90%), the *APOE* ε3 allele presenting a very high frequency (~94%), followed by the *APOE* ε4 (~5%), and only 0.007 of *APOE* ε2 (only one allele in the Aymara sample). No significant differences were found when the two Bolivian samples were compared by an exact test.

Table 3. APOE isoform genotypes in the two Bolivian samples.

Population	Genotypes					Alleles		
	*E**2/ *E**3	*E**2/*E**4	*E**3/*E**3	*E**3/*E**4	*E**4/*E**4	*E**2	*E**3	*E**4
Aymaras (n=39)	0	1	35	3	0	0.013	0.936	0.051
Quechuas (n=38)	0	0	34	4	0	0	0.947	0.053
Total (n=77)	0	1	69	7	0	0.007	0.942	0.052

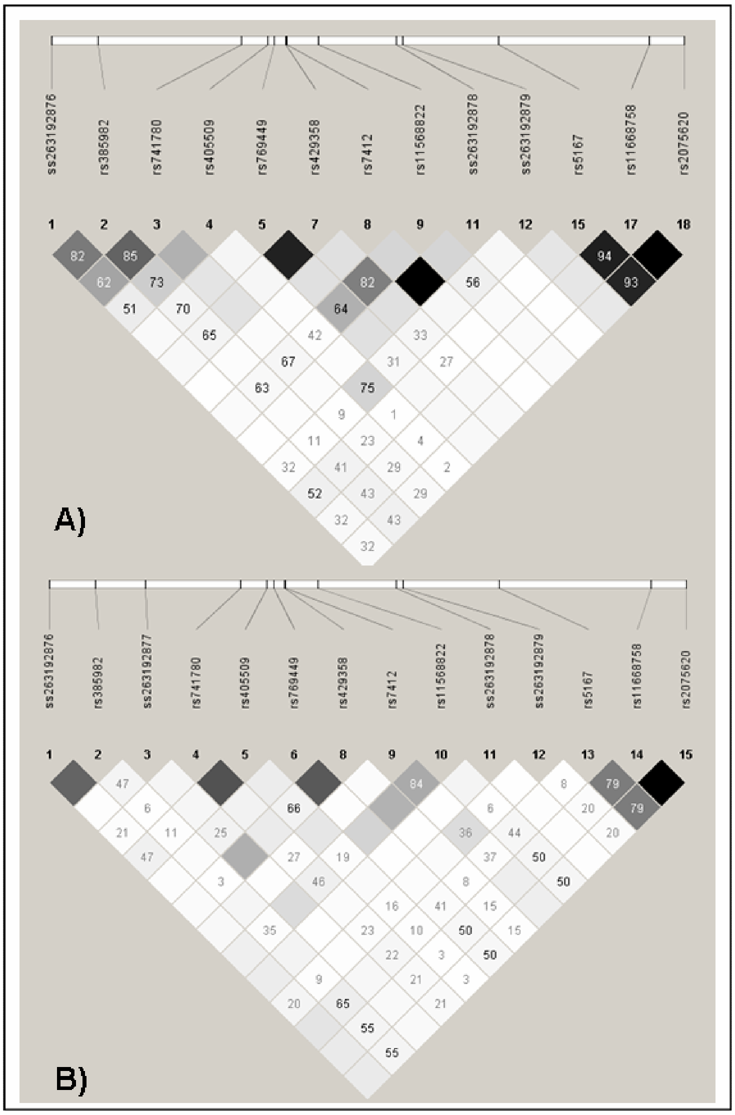
Linkage Disequilibrium and haplotypes

LD patterns (expressed as D' and r^2) between all pairs of biallelic markers (SNPs and STRs) for the pooled Bolivian and the European samples are shown in Figure 2. For Bolivians, the analysis revealed moderate to high LD (D’>0.62, $r^2 \geq 0.36$, p< 0.001) among STR ss263192876, rs385982 (*PVRL2*), and rs741780 (*TOMM40*) in the 5’-end. Likewise, complete linkage (D’=1, r^2=1, p<0.001) was observed in the 3’-end between rs11668758 and rs2075620 (*CLPTM1*), and very high LD between this gene and *APOC4* (rs5167; D’=0.94, r^2=0.86, p<0.001), and *APOC2* (ss263192881; p<0.0001). In contrast, although D’ suggested complete LD, a clear break was observed between *APOC1* (rs11568822) and *APOC4* (rs5167) according to the r^2 measure (0.018, p>0.05). LD results in the region including the *APOE* and *APOC1* genes are unclear, probably due to the low variability of some markers, such as rs7412 and rs11568822 showing only one copy of the minor allele in Aymaras and Quechuas, respectively. In fact, *APOC1* (rs11568822) presented moderate LD with rs429358 (r^2=0.50, p<0.0001), although with differences between Aymaras (r^2=0.79, p<0.0001) and Quechuas (r^2=0.24, p>0.05), and weaker LD with rs7412 (r^2=0.16; p>0.05). Within the *APOE* region, it is interesting to note the lack of LD between rs405509 in the promoter and the other *APOE* SNPs (r^2=0.036, rs764909; r^2=0.034, rs429358; r^2=0.011, rs7412; p>0.05) as well as between rs429358 and rs7412, determining the *APOE* isoforms, only 138pb apart (r^2 =0.12, p>0.05).

The LD analysis considering the *APOE* isoforms instead of the two determinant SNPs revealed strong LD with rs769449 (r^2=0.71, p<0.0001), moderate LD with rs11568822 (r^2 =0.61; p<0.0001), although it was almost complete in Aymaras, (r^2=0.99, p<0.0001), but low in Quechuas (r^2=0.24, p>0.05), and high LD with ss263192878 in Aymaras (r^2=0.97, p= 0.026).

In comparison, LD patterns in Europeans were, in general, similar to those of Bolivians, although weaker LD was observed in the 5’-end of the studied region, with high LD only between ss263192876 and rs385982 (*PVRL2*) (D’=1, r^2=0.61, p<0.0001), as well as in the 3’-end between *CLPTM1* and *APOC4* (rs5167) (D’=0.8, r^2=0.51, p<0.0001), and *APOC2* (ss263192881) (p=0.04). More homogenous LD values were observed between *APOC1* and *APOE* (r^2 = 0.33, p<0.001).

Figure 2. LD patterns from Haploview in Bolivians (A) and Europeans (B) for biallelic markers.



The colour scheme represents r^2 values (white: $r^2 = 0$, black: $r^2 =1$, shades of grey: $0 <r^2<1$). Numbers refers to D’ values (%), an empty cell being 100% (D’=1). To include the ss263192879 in Europeans, the allele $(CAAAA)_3$ was excluded for this analysis.

Table 4 shows the estimated haplotypes based on SNPs in the pooled Bolivian samples and in Europeans. Eighteen and 21 different haplotypes were found in Bolivians and Europeans, respectively, but the two samples only shared 7 haplotypes. In Bolivians, the four most common haplotypes accounted for 71% of the frequency whereas in Europeans for just 52%. An interesting result was that the two most common haplotypes in Bolivians were the third and fourth common haplotypes in Europeans, and the second most common haplotype in Europeans was not present in Bolivians.

Table 4. Frequency (± SE) of the estimated haplotypes in the two different world regions.

Haplotypes	Bolivians	Europeans
TGCGTC-TCA	0.230 ± 0.017	0.078 ± 0.027
GAAGTC-TCA	0.213 ± 0.015	0.122 ± 0.023
TGAGTC-GTG	0.163 ± 0.012	0
TGAGTC-TCA	0.101 ± 0.011	0
TGCGTC-GTG	0.084 ± 0.019	0.189 ± 0.022
GGAGTC-TCA	0.045 ± 0.005	0
GAAGTC-GTG	0.034 ± 0.014	0
GAAACC+TCA	0.022 ± 0.003	0
TAAGTC-GTG	0.022 ± 0.006	0.067± 0.020
GAAACC-TCA	0.017 ± 0.004	0
GGCGTC-GTG	0.017 ± 0.009	0
TGCGTC-GCA	0.011 ± 0.005	0.056± 0.012
TGCGTC-TTG	0.011 ± 0.004	0
GAAGCC+TCA	0.006 ± 0.002	0
GGCGTC-TCA	0.006 ± 0.006	0.067± 0.023
TAAACC-TCA	0.006 ± 0.004	0
TGAGTC-GCA	0.006 ± 0.003	0
TGCGTT+TCA	0.006 ± 0.003	0.056 ± 0.015
TAAGTC-TCA	0	0.133 ± 0.030
TACGTT+TCA	0	0.044 ± 0.013
TAAACC+GTG	0	0.033 ± 0.008
TAAGTC-TTG	0	0.022 ± 0.012
TACGCC+TCA	0	0.022 ± 0.008
TGCGTC+TCA	0	0.022 ± 0.010
GAAGTC-TTG	0	0.011 ± 0.007
GACGCC+TCA	0	0.011 ± 0.008
GGCGTC-GCA	0	0.011 ± 0.008
TAAACC+GCA	0	0.011 ± 0.006
TAAACC+TCA	0	0.011 ± 0.010
TAAGTC-GCA	0	0.011 ± 0.014
TACGTC+TCA	0	0.011 ± 0.006
TGCGTT-GCA	0	0.011 ± 0.005

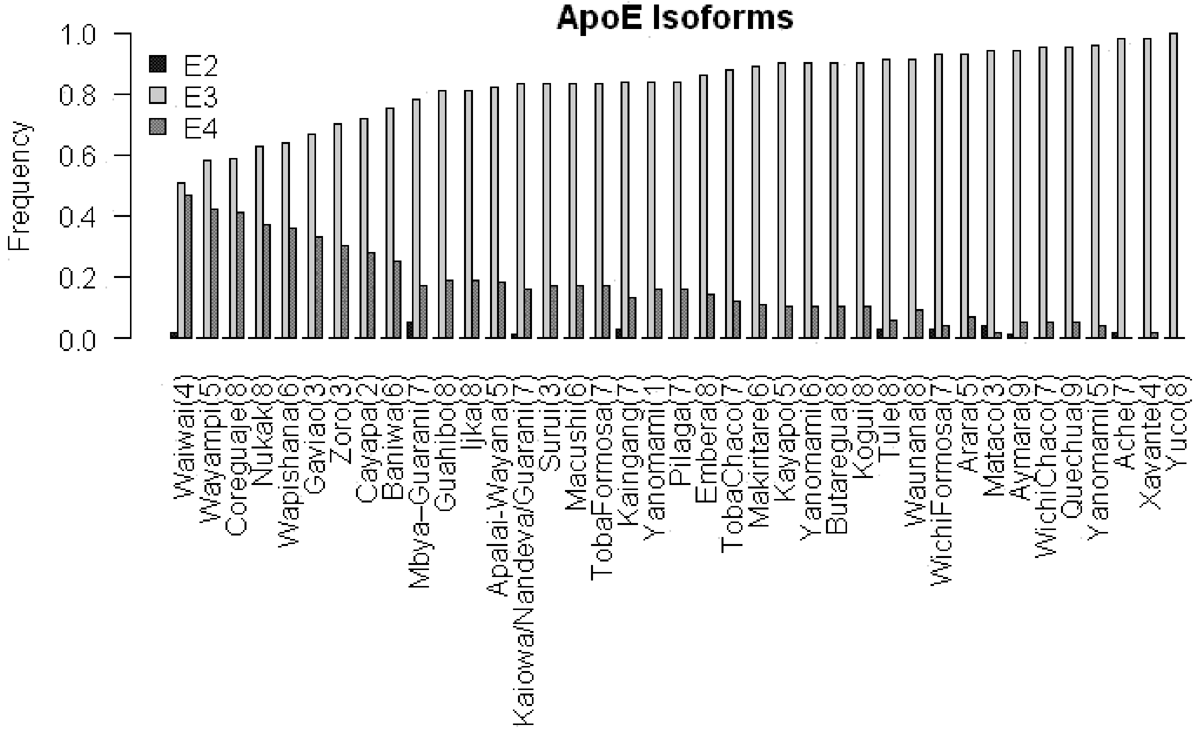
The SNPs shown are the polymorphic ones in this order: rs385982, rs741780, rs405509, rs769449, rs429358, rs7412, rs11568822, rs5167, rs11668758, rs2075620. Note: for the rs11568822, the "-" corresponds to the deletion and the "+" to the CGTT insertion.

Variation in South America

Allele and genotype frequencies for the *APOE* isoforms in the two Bolivian samples are shown in Table 3. The most common genotype was the homozygote for *APOE*ε3 (90%). No significant differences were found when the two Bolivian samples were compared by an exact test.

Figure 3 shows the *APOE* isoform frequencies in 39 South American samples. The *APOE* ε3 is the most frequent in all samples, ranging from 0.59 to 1. On the contrary, the absence of *APOE* ε2 is a common feature among South Americans, present only in 8 out of the 39 samples at very low frequencies (from 0.01 to 0.05). On the other hand, there is high variation in the frequency of *APOE* ε4 (from 0 to 0.47). Considering the 39 South American samples, between-population *APOE* isoform variation was significant (F_{ST} = 0.07, p<0.001), and the exact test of population differentiation revealed significant differences between the two Bolivian samples and 9 other groups (Cayapa, Gaviao, Zoro, WaiWai, Wayampi, Wapishana, Baniwa, Coreguaje and Nukak).

Figure 3. APOE isoform frequencies in Native South Americans.



Numbers refer to references: 1, Crews et al 1993; 2, Scacchi et al., 1997; 3, Andrade et al., 2000; 4, Andrade et al., 2002; 5, Marin et al., 1997; 6, JV Nel(Andrade 2000b); 7, Demarchi et al., 2005; 8, Jaramillo-Correa et al., 2001; 9, present work.

As for the available data from 3 loci in American samples from the HGDP, the exact test of population differentiation revealed that the Bolivian samples presented significant differences from Piapoco and Curripaco (Colombia), Kariatiana (Brazil) and Pima (Mexico) mainly due to differences in rs11668758 and rs5167. When five loci were considered, no significant differences were found between the two Bolivian samples and the Mexicans from HapMap.

DISCUSSION

This work studies the variability of the genomic region encompassing the *APOE/C1/C4/C2* gene cluster in two Andean populations from Bolivia. As so, it provides novel data for the genetic characterization of Native American populations. The 25 analyzed markers constitute the first data not only on Andean samples, but also, for most markers, on Native Americans in general, let alone the first data for nine out of the10 STRs examined in this genomic region.

In total, our results suggest low genetic variation in the Bolivian samples for both sets of markers as well as for the *APOE* isoforms. Nevertheless, a detailed analysis of both the differences in marker diversity, according to their location, and the allele frequency comparisons between Bolivian and European samples, indicates that some of this low diversity in Bolivia could be explained as a result of selection, while another part reflects the demographic history of Native American populations.

In our analysis there are indications that the low variability in the center of the genomic region studied can be attributed to selective factors. In this central region (from *TOMM40* to *APOC2*), 9 out of 12 SNPs present very low gene diversity both in Native Americans and Europeans, except for rs11568822 (Table 1), with no allele frequency differences found among populations, except for three loci. A similar observation is true for the STRs. Seven out of 10 markers with a more central position were monomorphic or showed very low gene diversities both in Bolivians and Europeans (Table 2). On the contrary, the SNPs with the highest diversities were those located at the extremes of the studied region both in Bolivians and Europeans, as well as for most of the HapMap populations (data not shown). This is reflected in the fact that the most common haplotypes differ mainly in changes in the extremes (Table 4), while the core formed by the *APOE* and *APOC1* polymorphisms remains practically invariable. As a result, the *APOE* ε3 isoform and the deletion in the *APOC1* promoter are the most common alleles worldwide. A similar situation is observed for the STRs, with ss263192876 at the 5'-end being one of the most diverse in Bolivians and Europeans, and ss263192882 at the 3'-end showing high diversity, although only in Europeans. At this point, it is worth mentioning the exception of the highly diverse compound STR $(TG)_n(AG)_m$ ss263192881. Although it presented significant differences in allele frequencies, the range of repeats detected in Bolivians (11 alleles from 17 to 32 repeats) was quite similar to that found in European populations (the Spanish sample used in this study: 9 alleles ranging from 17 to 33 repeats; a French sample from Zouali et al. (1999): 12 alleles from 17 to 34 repeats). Fornage et al. (1992) reported that diversity at this locus was attributable to length variation at both $(TG)_n$ and $(AG)_m$ motifs, although the variation in the $(AG)_m$ motif was restricted to only two alleles differing by one repeat unit. The sequencing of five individuals in the present study confirmed this observation, although other scenarios cannot be discarded due to the low number of sequences. These results could indicate that the *APOE/C1/C4/C2* region has been highly conserved, a fact most likely related to its biological importance.

On the other hand, the lower gene diversity observed in the Andean populations compared to Europeans for both kinds of markers could be rather related to the particular demographic history of Native South Americans, were drift and founder effects might have had an important impact. In particular, the STR located 134.2kb 3’ -end of the region had 5 to 8 times higher heterozygosity values in Europeans than in Quechuas and Aymaras, respectively. Also, concerning the SNP-based haplotypes, the Bolivians presented a lower number of haplotypes, some of them with extreme frequencies compared to Europeans. This was reflected in the LD patterns; even though Bolivians presented several similarities to those from HapMap, the Europeans here studied, and other studies (Klos et al., 2008; Ken-Dror et al., 2010), such as the presence of a LD block at the 3’of the region, high LD at the 5’-end and a break in LD between *APOC1* and *APOC4,* the higher degree of LD values in Bolivians compared to Europeans is consistent with a more recent origin and/or bottleneck. The strong LD between the *APOE* isoforms and *APOC1* observed in Bolivians was similar to the observations reported for other Native Americans (Andrade et al., 2002), although interethnic differences have been previously reported by Xu et al. (1999) and confirmed with the present work.

The actual comparisons of *APOE* isoforms among South Americans did not show a clear pattern of variation. The heterogeneity observed was mainly due to differences on the *APOE* ε4 frequencies. The two Bolivian samples of this work present a high incidence of *APOE* ε3 and, consequently, a very low frequency of *APOE* ε4, within the range observed in South Americans. The fact that one individual from the Aymara sample was a carrier of an *APOE* ε2 allele is important in the frame of the controversial question about the presence or absence of *APOE* ε2 in Native South Americans. Most Native South American samples studied so far (30 out of 39) did not have this allele and few samples had it at very low frequencies. Andrade et al. (2000) proposed two theories for explaining its presence: a) either it was present in the first founding populations at a very low frequency and it was lost in some groups or not identified due to restricted sample sizes, b) or its presence is due to admixture with non-Native Americans. The fact that the Aymara sample presented a very low degree of admixture according to different markers (Gayà-Vidal et al., 2011) supports the first theory, although more data are necessary for such an affirmation. Concerning the *APOE* ε4 distribution in Native Americans, two different explanations have been proposed: a) Corbo and Scacchi (1999) observed that some Native American populations had a very high *APOE* ε4 frequency as other Native populations from other world regions and suggested that this allele could have been favoured increasing the cholesterol absorption in populations with low-cholesterol diets; b) Andrade et al. (2000) suggested that in Native American populations, founder effects, isolation and genetic drift could have played a major role in determining *APOE* ε4 frequencies rather than natural selection. At this point it is important to notice that the *APOE* ε4 frequency of the two Bolivian samples was one of the lowest values (0.05) compared with other South American samples. In the case of the two Bolivian samples, their diet is mainly based on carbohydrates and not characterized by high cholesterol levels (their consumption of meat comes mainly from camelids, which is characterised by low cholesterol levels; Saadoun and Cabrera, 2008). Therefore, according to Corbo and Scacchi (1999), we should expect a higher frequency of *APOE* ε4; however, our results support the theory of Andrade et al, (2000), these differences being the result of founder effects and genetic drift that would have played a major role in the current distribution of this allele in Native Americans. Nevertheless, a lower mortality from coronary heart disease has been observed in populations living in areas of high altitude and some authors have proposed that this low incidence of cardiovascular diseases in high-altitude populations could be due to dislipidemia or hypoxia (Caen et al., 1974). Therefore, we can not discard selective factors related with the adaptation to the Altitude environment of these Altiplano populations.

Finally, it is worth mentioning that the absence of LD between the rs405509 (in the *APOE* promoter) and the *APOE* isoform alleles is in agreement with a recent studies (Ken-Dror et al., 2010), although it contrasts others studies that found partial LD (Fullerton et al.; 2000; Heijmans et al., 2002). These controversial results could be the result of using different LD measures (D' or r^2).

CONCLUSION

This study provides novel data on the *APOE/C1/C4/C2* region in two Andean populations. Our results showed a high degree of conservation of the *APOE/C1/C4/C2* gene cluster which may be due to its physiological importance. Nevertheless, certain differences detected between the Bolivian samples and Europeans would reflect their different demographic histories. The *APOE* isoform frequencies in Bolivians were within the range observed for South Americans, providing new insights into some controversial issues. The presence of *APOE* ε2 supports the idea that this allele was present in the founding populations rather than due to mixture and the low frequency of *APOE* ε4 suggests that its frequency in South Americans would be mainly due to demographic effects rather than the result of selection due to a low-cholesterol diet. Nevertheless, we cannot discard that the particular altitude environment could have plaid a role. More Andean samples should be studied to reach more robust conclusions. This paper demonstrates that the study of the variation in the *APOE/C1/C4/C2* region in general population is useful for both population and epidemiological purposes.

Acknowledgments

We are grateful to all the donors for providing DNA samples and to the people who contributed to the collection of samples. This research has been supported by the Ministerio de Ciencia y Tecnologia (CGL2005-3391, CGL2008-03955) and the Generalitat de Catalunya (2005-SGR00252). The work of MGV has been financed by an FI grant from the "Departament d'Educacio i Universitats de la Generalitat de Catalunya i el Fons Social Europeu".

REFERENCES

Andrade FM, Coimbra CE Jr, Santos RV, Goicoechea A, Carnese FR, Salzano FM, Hutz MH. 2000. High heterogeneity of apolipoprotein E gene frequencies in South American Indians. Ann Hum Biol 27:29-34.

Andrade FM, Ewald GM, Salzano FM, Hutz MH. 2002. Lipoprotein lipase and APOE/APOC-I/APOC-II gene cluster diversity in native Brazilian populations. Am J Hum Biol 14:511-518.

Artiga MJ, Bullido MJ, Frank A, Sastre I, Recuero M, García MA, Lendon CL, Han SW, Morris JC, Vázquez J, Goate A, Valdivieso F. 1998. Risk for Alzheimer's disease correlates with transcriptional activity of the APOE gene. Hum Mol Genet 7:1887-1892.

Barrett JC, Fry B, Maller J, Daly MJ. 2005. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21:263–265.

Belkhir K, Borsa P, Goudet J, Chikhi L, Bonhomme, F. 1998. GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5000. Université de Montpellier II, Montpellier, (France).

Black WC, Krafsur ES. 1985. A FORTRAN program for the calculation and analysis of two-locus linkage disequilibrium coefficients. Theor Appl Genet 70:491–496.

Corbo RM, Scacchi R. 1999. Apolipoprotein E (APOE) allele distribution in the world. Is APOE*4 a 'thrifty' allele? Ann Hum Genet 63:301-310.

Crews DE, Kamboh MI, Mancilha-Carvalho JJ, Kottke B. 1993. Population genetics of apolipoprotein A-4, E, and H polymorphisms in Yanomami Indians of northwestern Brazil: associations with lipids, lipoproteins, and carbohydrate metabolism. Hum Biol 65:211-224.

Demarchi DA, Salzano FM, Altuna ME, Fiegenbaum M, Hill K, Hurtado AM, Tsunetto LT, Petzl-Erler ML, Hutz MH. 2005. APOE polymorphism distribution among Native Americans and related populations. Ann Hum Biol 32:351-365.

Excoffier L, Laval G, Schneider S. 2005. Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evol Bioinform Online 1:47-50.

Fornage M, Chan L, Siest G, Boerwinkle E. 1992. Allele frequency distribution of the (TG)n(AG)m microsatellite in the apolipoprotein C-II gene. Genomics 12: 63-68.

Fullerton SM, Clark AG, Weiss KM, Nickerson DA, Taylor SL, Stengârd JH, Salomaa V, Vartiainen E, Perola M, Boerwinkle E, Sing CF. 2000. Apolipoprotein E variation at the sequence haplotype level: implications for the origin and maintenance of a major human polymorphism. Am J Hum Genet 67:881-900.

Gayà-Vidal M, Dugoujon JM, Esteban E, Athanasiadis G, Rodríguez A, Villena M, Vasquez R, Moral P. 2010. Autosomal and X chromosome Alu insertions in Bolivian Aymaras and Quechuas: two languages and one genetic pool. Am J Hum Biol 22:154-162.

Gayà-Vidal M, Moral P, Saenz-Ruales N, Gerbault P, Tonasso L, Villena M, Vasquez R, Bravi CM, Dugoujon JM. 2011. mtDNA and Y-chromosome diversity in Aymaras and Quechuas from Bolivia: Different stories and special genetic traits of the Andean Altiplano populations. Am J Phys Anthropol 145:215-230.

Hanis CL, Hewett-Emmett D, Douglas TC, Bertin TK and Schull WJ. 1991. Effects of the Apolipoprotein E polymorphism on levels of lipids, lipoproteins, and apolipoproteins among Mexican-Americans in Starr County, Texas. Arterioscler. Thromb 1:362–370.

Heijmans BT, Slagboom PE, Gussekloo J, Droog S, Lagaay AM, Kluft C, Knook DL, Westendorp RG. 2002. Association of APOE epsilon2/epsilon3/epsilon4 and promoter gene variants with dementia but not cardiovascular mortality in old age. Am J Med Genet 107:201-208.

Jakobsson M, Scholz SW, Scheet P, Gibbs JR, VanLiere JM, Fung HC, Szpiech ZA, Degnan JH, Wang K, Guerreiro R, Bras JM, Schymick JC, Hernandez DG, Traynor BJ, Simon-Sanchez J, Matarin M, Britton A, van de Leemput J, Rafferty I, Bucan M, Cann HM, Hardy JA, Rosenberg NA, Singleton AB. 2008. Genotype, haplotype and copy-number variation in worldwide human populations. Nature 451:998-1003.

Jaramillo-Correa JP, Keyeux G, Ruiz-Garcia M, Rodas C, Bernal J. 2001. Population genetic analysis of the genes APOE, APOB(3'VNTR) and ACE in some black and Amerindian communities from Colombia. Hum Hered 52:14-33.

Kamboh MI, Aston CE, Hamman RF. 2000. DNA sequence variation in human apolipoprotein C4 gene and its effect on plasma lipid profile. Atherosclerosis 152:193-201.

Ken-Dror G, Talmud PJ, Humphries SE, Drenos F. 2010. *APOE/C1/C4/C2* gene cluster genotypes, haplotypes and lipid levels in prospective coronary heart disease risk among UK healthy men. Mol Med 16:389-99.

Klos K, Shimmin L, Ballantyne C, Boerwinkle E, Clark A, Coresh J, Hanis C, Liu K, Sayre S, Hixson J. 2008. *APOE/C1/C4/C2* hepatic control region polymorphism influences plasma apoE and LDL cholesterol levels. Hum Mol Genet 17:2039-46.

Li JZ, Absher DM, Tang H, Southwick AM, Casto AM, Ramachandran S, Cann HM, Barsh GS, Feldman M, Cavalli-Sforza LL, Myers RM. 2008. Worldwide human relationships inferred from genome-wide patterns of variation. Science 319:1100-4.

Marin GB, Tavella MH, Guerreiro JF, Santos SEB and Zago MA. 1997. Absence of the E2 allele of apolipoprotein in Amerindians. Braz J Genet 20.

Mountain JL, Knight A, Jobin M, Gignoux C, Miller A, Lin AA, Underhill PA. 2002. SNPSTRs: empirically derived, rapidly typed, autosomal haplotypes for inference of population history and mutational processes. Genome Res 12:1766-1772.

Raymond M, Rousset F, 1995. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. J Heredity 86:248-249.

Rousset F. 2008. Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. Mol Ecol Resources 8: 103-106.

Saadoun A, Cabrera MC. 2008. A revier of the nutritional content and technological parameters of indigenous sources of meat in South America. Meat Science, doi:10.1016/j.meatsci.2008.03.027

Scacchi R, Corbo RM, Rickards O, Mantuano E, Guevara A, De Stefano GF. 1997. Apolipoprotein B and E genetic polymorphisms in the Cayapa Indians of Ecuador. Hum Biol 69:375-382.

Schull WJ and Rothhammer F. 1990. The Aymara. Strategies in Human Adaptation to a rigorous Environment. Kluwer Academic Publishers

Song Y, Stampfer MJ, Liu S. 2004. Meta-analysis: apolipoprotein E genotypes and risk for coronary heart disease. Ann Intern Med 141:137-147.

Stephens M, Smith NJ, Donnelly P. 2001. A new statistical method for haplotype reconstruction from population data. Am J Hum Genet, 68: 978-989.

Stephens M, Donnelly P. 2003. A comparison of bayesian methods for haplotype reconstruction from population genotype data. Am J Hum Genet 73:1162-1169.

Tycko B, Lee JH, Ciappa A, Saxena A, Li CM, Feng L, Arriaga A, Stern Y, Lantigua R, Shachter N, Mayeux R. 2004. *APOE* and *APOC1* promoter polymorphisms and the risk of Alzheimer disease in African American and Caribbean Hispanic individuals. Arch Neurol 61:1434-1439. Erratum in: Arch Neurol. 2005 62:332.

Wollmer MA. 2010. Cholesterol-related genes in Alzheimer's disease. Biochim Biophys Acta 1801:762-73. Epub 2010 May 24. Review.

Xu Y, Berglund L, Ramakrishnan R, Mayeux R, Ngai C, Holleran S, Tycko B, Leff T, Shachter NS. 1999. A common Hpa I RFLP of apolipoprotein C-I increases gene transcription and exhibits an ethnically distinct pattern of linkage disequilibrium with the alleles of apolipoprotein E J Lipid Res 40:50-58.

Zouali H, Faure-Delanef L and Lucotte G. 1999. Chromosome 19 locus apolipoprotein C-II association with multiple sclerosis. Multiple Sclerosis 5: 134-136.

IV. DISCUSSION

In this work different types of markers were studied in order to address different questions regarding the genetic features of the Bolivian Andean populations. This discussion is structured according to the aims proposed in the second section.

IV.1 Genetic characterization of the two Bolivian samples

The analysis of different markers has permitted to obtain a deep insight into the genetic characteristics of the two main linguistic groups in Bolivia, Aymaras and Quechuas, contributing also with new data on Native American genetic variability.

IV.1.1 Data provided for the first time in Native Americans

This work provides new genetic data on different kinds of markers for the Andean Altiplano populations. Most previous genetic studies on Andean populations were centred in Peruvian samples. Although the Altiplano region is a particular area regarding both the history and environment, few genetic population studies were carried out in this region, especially focusing on HLA (Arnaiz-Villena et al., 2005, 2010; Martinez-Laso et al., 2006) and mtDNA haplogroup frequencies (Merriwether et al., 1995; Bert et al., 2001; Sandoval et al., 2004). Also, few studies exist describing the HVI region (Corella et al., 2007, considering highland migrants; Barbieri et al., 2010), or the HVI-HVII regions in an urban sample (Afonso Costa et al., 2010). Moreover, Y-chromosome data exist for an urban sample (Bailliet et al., 2009) and in a highland Bolivian sample (Lee et al., 2007), but without any confirmation of Native Y-chromosomes with SNPs (the haplogroups inferred from the haplotypes revealed only seven out of 40 haplotypes that might belong to Native Y chromosome although with low probabilities, Gayà-Vidal et al., 2011). In particular, this work contributes with:

- The study of 32 *Alu* insertions on the two Bolivian samples, at the moment of the publication of Gayà-Vidal et al. (2010), represented:
    - The first data on X chromosome *Alu* insertions analysed in Native Americans.
    - The first data on eight out of the 18 autosomal PAIs determined in South Amerindians.
    - The first data on Bolivian samples and in an Aymara-speaking population.

- Regarding the study of the two most-studied systems in human population genetics, the mtDNA and Y-chromosome, at the time of publication of Gayà-Vidal et al. (2011), our study represented:
    - The first Bolivian samples to be studied for the HVII mtDNA region.
    - One of the few studies that considered well defined Native populations and adequate sample sizes, two important characteristics in these studies.
    - The first Bolivian Andean samples to be analysed for the Y chromosome.
- The study of the variation on the *APOE/C1/C4/C2* gene cluster region in the two Bolivian groups, at the time of the submission represented:
    - The first data on Andean samples for the *APOE* isoforms.
    - The first data on most SNP polymorphisms studied in Native Americans.
    - The first data in human populations for most STRs studied.

Many studies on the genetic variation of South America did not consider populations from the Altiplano (mtDNA control region: Fuselli et al., 2003; Cabana et al., 2006; Lewis et al., 2007, 2008; Y-chromosome: Bianchi et al., 1998; Tarazona-Santos et al., 2001), or even from the central Andean area (mtDNA: Vona et al., 2005; Y-Chromosome: Karafet et al., 1999; Lell et al., 2002; Bortolini et al., 2003)), leaving aside one of the most important regions in terms of population density and historical and environmental particularities. This study has shown that Altiplano populations slightly differ from other Andean regions for certain markers. Therefore, the two samples from the Bolivian Altiplano presented in this work will permit the inclusion of this region in future studies about the genetic diversity in South America.

IV.1.2 Within-population genetic variation

In general, the results of this thesis indicate a low degree of within-population gene diversity in the two Bolivian samples. This is shown by a rather high number of fixed markers (*Alu* insertions, Gayà-Vidal et al., 2010; SNPs and STRs, 3rd manuscript). Nevertheless, this general statement deserves some comments concerning whether these diversity values are considered in a worldwide or South American context.

The extreme allele frequency or fixation values found in many PAIs in the Bolivian samples (Gayà-Vidal et al., 2010), as well as the relatively low diversity found in SNPs and STRs of the *APOE/C1/C4/C2* region, is consistent with a general worldwide trend that Native Americans present a diversity reduction as compared to other world populations (Stoneking et al., 1997; Mateus-Pereira et al., 2005). This low within-diversity could be related to genetic drift and bottleneck processes that occurred during the peopling of America, and particularly that of South America (Watkins et al., 2003). The low number of mtDNA and especially of Y-chromosome haplogroups is also consistent with this particular demographic history of Native Americans.

In the South American context, different results were obtained depending on the studied polymorphisms. For the autosomal markers (*Alu* insertions and *APOE* isoforms), the gene diversities of the two Bolivian samples were within the range observed in South America. In the case of the *APOE* isoforms, the diversities were especially low, mainly due to the low frequency of the ε4 allele (3rd manuscript). The average gene diversity for the eight *Alu* loci also tested so far in other South American samples was moderate to high compared to other South American groups. For the mtDNA haplogroups, the two Bolivian samples also presented similar values to other South American groups, the Aymaras showing lower values than Quechuas because of the high frequency of haplogroup B2. However, when the Control Region was considered, the two samples showed high diversity values, especially the Quechuas, similar to that of the Coyas (Alvarez-Iglesias et al., 2007). Higher values were found in some Argentinean samples, but these are most probably due to their mixed nature, since they correspond to a political subdivision. Regarding the Y chromosome, the diversity values were also within the range observed in South Americans, again closer to the higher values, specially the Aymaras (Gayà-Vidal et al., 2011). At this point it is important to mention that, in general, Andean samples (including the two Bolivian populations) showed higher diversity values than populations from the eastern part of South America. More precisely, different patterns of intra- and inter-population diversity of the East and West areas of South America related to different demographic histories have been proposed (see section IV.3.3).

IV.1.3 Admixture of Native Bolivian populations

Previous data on classical markers, immunoglobulins and blood groups, in our Bolivian samples, showed the presence of around 1% of the specific European haplotype GM5*;3 (Dugoujon JM, personal communication) and a frequency of 98% of the O group (ABO system), which suggest a low external admixture of the two Bolivian samples studied here. The study of the mtDNA and Y-chromosome haplogroups also supports low admixture levels of these samples, but with differences depending on the gender and population. The two samples presented total absence of non-native admixture for the mtDNA. In contrast, certain degree of non-native admixture for the Y-chromosome was observed (presence of 3% and 22% of non-native Y-chromosomes in Aymaras and Quechuas, respectively; Gayà-Vidal et al., 2011). This differentiation between the two uniparental marker types is a general trend in Native American populations as a consequence of the colonization by the Europeans, carried out mainly by men.

On the other hand, the larger admixture levels found in Quechuas than in Aymaras according to the Y-chromosome haplogroups (22% versus 3% of non-Native American Y-chromosomes) is consistent with what was already reported by Salzano and Callegari-Jacques (1988). These authors estimated that the average Caucasian admixture in Quechuas was approximately 25%, whereas in Aymaras was approximately 8%. These values are larger than those obtained in the present work (22% and 3%, respectively), which could be related to the fact that the samples of this study are from rural areas. As Rupert and Hochachka (2001) noted, the values obtained by Salzano and Callegari-Jacques, (1988) may vary depending on the geographic proximity of the two groups (Native *vs* European origin) and the amount of time that they have been in contact. As Aymaras inhabit the highlands, Quechuas are found in both the highlands and lowlands, and Europeans mainly live in the lowlands, a higher European admixture in Quechuas than in Aymaras would be expected. In any case, a remarkably low level of admixture was observed in the Aymaras, with a mere 3% of non-Native American Y-chromosomes as compared with other Andean samples (Peruvian Quechuas from Tayacaja (30%) and Arequipa (20%), Susque-Humahuaca (5%), Tarazona-Santos et al. (2001); Kichwas (14%), González-Andrade et al. (2007);

Humahuaca (35%), Dipierri et al. (1998)), highlighting its isolation from non-Native Americans.

IV.1.4 Particular genetic features of the current Bolivian populations

The fact that some markers were studied in the two samples for the first time reduces the range of possible comparisons. Nevertheless, interesting particularities were found in the two Bolivian samples.

Regarding the characterization of the *Alu* insertions, despite the fact that the allele frequency distribution patterns of the two Bolivian samples were similar to other Native South Amerincan populations for the ten markers for which data were available (Tishkoff et al., 1996, 1998; Novick et al., 1998; Antunez de Mayolo et al., 2002; Dornelles et al., 2004; Mateus-Pereira et al., 2005; Battilana et al., 2006), the two Bolivian samples presented the highest insertion frequencies (Aymaras: 0.473, Quechuas: 0.419) for the HS4.32 locus, followed by the Quechuas of Arequipa (0.357). For the other markers, a general pattern of low diversity was observed. As for the X chromosome PAIs, Ya5DP77 and Yd3JX437 interestingly showed the highest diversity values in the two Bolivian populations like in other African and Asian populations, in contrast with Europeans. It would be interesting to have data on Native Americans to perform comparisons.

The most interesting genetic particularities of the Bolivian samples were found for the uniparental markers. Concerning the mtDNA, although our results revealed typical Andean characteristics in the two Bolivian samples (high frequency of haplogroup B2: 81% Aymaras and 61% Quechuas), the sequencing of the Control Region provided interesting characteristics. Some variants were at a high frequency within the haplogroup B2 (16168, 16188, 103-143, and 146-215 together representing 60% of B2 haplotypes). The most remarkable trait was the 16188 variant, also observed in other Andean populations, mainly from the Andean Altiplano, reaching frequencies up to 81% for haplogroup B2. However, the fact that in a subgroup of these samples, mainly Bolivians, (Aymara, Quechua, Aymara Titicaca, Quechua Titicaca, La Paz, and Coya) the 16188 variant was always combined with the variant 16183C confers a high degree of similarity among the Altiplano samples, especially among the Bolivian ones. In addition, other variants within this sub-branch were also found (186, 63-64, the lack of 73, etc), some of them (186 and the lack of the 73 variant) also observed in Coya and La Paz, confirming that some combinations of mutations seem to be characteristic of the Andean Altiplano.

In addition, it is worth mentioning other traits. First, one Quechua haplotype (hapl. 39) presented a 106-111deletion, also reported in one individual from La Paz (LPAZ070) that shared the same haplotype (at least when considering the HVI and HVII regions separately). This 106-111d in Bolivian samples occurred within haplogroup B2 in contrast to the deletion within haplogroup A2 found in some Chibchan-speaking populations (Santos and Barrantes, 1994; Kolman et al., 1995), indicating a recurrent mutation rather than a trait restricted to a certain group. Second, two haplotypes (18 and 19) of haplogroup A2 presented some interesting mutations (lack of the haplogroup A diagnostic site 235, variants 16512, 16547, 16551iG, and absence of the 64 variant) also found (except for 16551iG) in one Coya individual. Third, the haplotype 51 presented a particular mutation combination that has not been previously reported, probably because a lot of studies that sequenced the HVII region started from the position 73, that highlights the huge variability between the 55 and 71 positions in the HVII.

The high presence of rare mtDNA alleles in the two Bolivian samples and other Andean samples could be related to a high-long-term effective population size in the Andean region according to Fuselli et al. (2003).

For the Y-chromosome, the two Bolivian samples presented a pattern of Native American haplogroup frequencies in concordance with other South American groups, with a high frequency of haplogroup Q1a3a (100% of Quechuas and 89% of Aymaras), which is the most frequent haplogroup in South America. The most important particularity was the high frequency of the paragroup Q1a3* in Aymaras (11%), a value that doubled that reported in other Bolivian samples (Bailliet et al., 2009), supporting the fact that the northwest border of South America harbors the highest frequencies of the Q1a3* lineage, as proposed in Bailliet et al. (2009). This high Q1a3* frequency in Aymaras could be attributed to drift effects, but the high diversity values (haplotype) observed in Aymaras, as well as in other Andean samples, is not consistent with this interpretation.

Concerning the Y-STR variation, an interesting result was the high frequency of the DYS393*14 allele in the two Bolivian samples; 56% and 58% of Aymaras and

Quechuas, respectively. Even though this allele was present in central Andes (19% Ecuador; 40% Peru; 42% northwest Argentina), the Bolivian samples presented the highest frequencies (57%). This could indicate an Altiplano origin and a subsequent expansion to the surrounding areas as was already proposed by Martínez-Marignac et al. (2001) after finding a high frequency of this allele (38.9%) in northwest Argentinean samples, where most surnames were of Aymara origin. However, in the Venezuelan area, two samples (Bari and Yukpa) also presented this allele in extreme values (around 90%). The total discontinuity between these two areas suggested two different events for the origin of this allele. The Minimal Haplotype of the Y-chromosomes carrying this allele in both areas also supported the hypothesis of two independent origins (Gayà-Vidal et al., 2011).

Finally, regarding the *APOE/C1/C4/C2* region, the two Bolivian samples were characterized by a low diversity of the markers studied that can be attributed to both selective and historical reasons (3^{rd} manuscript). Regarding the APOE isoforms, the two samples presented a frequency distribution within the pattern observed in South America, a fact that a priori does not seem rather remarkable. However, it is important to take into account that these are the first data on Andean highland populations and that selective pressure could have acted upon this locus. It is well known that Andean highlands present low frequency of cardiovascular diseases (Caen et al., 1974; Mortimer et al., 1977) and that the ethnic origin has an important role in the lipid metabolism as well as the habitat (Bellido and Aguilar, 1992). Therefore, our results, showing a low frequency of APOε4 would support a normal absorption of cholesterol in the Andeans. In any case, to confirm this, studies should be carried out in more Andean samples. Another important trait was the presence of the APOε2 allele in the Aymaras, a fact that contributes to the debate about the origin of this allele in Native South Americans, supporting the hypothesis of the presence of this allele in the founding populations rather than a presence due to admixture.

IV.2 Genetic relationships between the two Bolivian samples

The different markers analysed, genetic distances, AMOVA analyses, exact tests of population differentiation and other analyses have permitted to shed light into the genetic relationships between the two Bolivian populations, especially within the Andean context.

IV.2.1 Linguistics *vs.* genetics in current Bolivian Andean populations

One of the most striking results of this work was the high genetic similarity between the Aymara and Quechua samples from Bolivia. This close genetic relationship was observed for all autosomal (*Alu* insertions and *APOE/C1/C4/C2* region) and X chromosome (*Alus*) markers since no significant differences in the allele frequency distribution were found between the two samples for any of those markers (Gayà-Vidal et al., 2010, 3rd manuscript).

In the case of the uniparental markers, more genetic differences were detected for the mtDNA than for the Y-chromosome. For the mtDNA, we expected more similarities between the two Bolivian samples because Andean populations present a patrilocal system. However, lower frequency of haplogroup B2 and the presence of particular haplotypes from other non-Andean areas were found in Quechuas. This differentiation could be due to a higher proportion of gene flow from external areas affecting the Quechuas. Concerning the Y-chromosome, despite the differences due to the paragroup Q1a3* present in Aymaras, there was a clear genetic homogeneity, not only for the Bolivian samples, but also for the whole central Andean region.

Therefore, the language (Aymara and Quechua) in Bolivia does not seem to imply an important genetic differentiation between these two groups; rather, the observed differentiation might be related to the degree of non-Native and non-Altiplano admixture. At this point, the Quechuas presented higher gene flow levels than the Aymaras, which would have been more isolated. This is quite expectable, taking into account that Quechua has been a language of intercommunication in that area, before and after the Spanish conquest. Therefore, disregarding the difference due to admixture, the high similarities observed for the autosomal markers and the genetic peculiarities of the uniparental markers in populations from the Altiplano suggest a common origin of the Altiplano samples or high levels of gene flow within the area.

IV.2.2 Introduction of Quechua into Bolivia

The *Alu*-based high genetic similarity between the two Bolivian samples and the differentiation from the Peruvian Quechua speakers supported that the Quechua language was expanded into Bolivia mainly by a cultural process, probably due to its imposition by the Incas and/or during the conquest promoted by the Spaniards as *lingua franca*, without an important demographic contribution by the Quechua-speakers from present Peru (Gayà-Vidal et al., 2010). In fact, there are well-documented cases of language replacement, such as that of the Macha of northern Potosí (precisely an area very close of the Quechuas' area studied in this work) who adopted Quechua in place of Aymara without changing their domestic economy (Kolata, 1993).

As stated in Gayà-Vidal et al. (2010), an alternative scenario was also possible: a hypothetically genetically different Quechua-speaker group could have entered into Bolivia and those differences would have been erased because of gene flow between the Aymara and the Quechua regions of Bolivia. This explanation would be consistent with historical records describing frequent population movements in the central Andean region during the Inca Empire and onwards (Platt, Bouysse-Cassagne and Harris, 2006). However, in such a case, lower genetic distances would be expected when comparing Bolivian *vs.* Peruvian Quechua-speakers (especially with Peruvian Quechua-speakers from Arequipa who share the same Quechua dialect (Cerron-Palomino, 2002)) than between Aymaras and Peruvian Quechuas, which was not the case according to the *Alu* insertions.

Nevertheless, the results on the mtDNA showed a closer affinity between the Quechuas of this work and the Peruvian samples. The Quechuas presented more than two times lower genetic distances with the Peruvian Quechuas than those with Aymaras for the HVI region. This could indicate the existence of certain demographic events from current Peru in the Quechua-speaker region of Bolivia.

These results seem to be contradictory; however, these conclusions come from different markers and the information they provide is complementary and needs to be integrated into a more complete and coherent scenario. In Gaya-Vidal et al. (2011), we proposed that the two Bolivian samples have a common origin and, since the Inca period, a higher movement of people would have taken place in the central Andes, with some areas adopting the Quechua language mainly by cultural processes, although certain demographic events cannot be discarded. In this way, populations that kept their language (Aymaras) would have been more isolated from the Peruvian influence.

IV.2.3 Sex-specific population histories

The study of the two uniparental systems revealed different aspects of the two Bolivian populations studied here. The distribution of the Y-chromosome variation indicated a clear genetic homogeneity inside the whole central Andean region. This homogeneity could be explained by the higher mobility of males than females across the entire region that might have been favored during the Inca Empire and Colonial times under the *mita* system. On the other hand, the mtDNA also revealed certain homogeneity, but in this case, it would be more restricted to the Altiplano area instead of the whole Andean region. However, this homogeneity is referred to the presence of certain particularities. It is important to remember that for the haplogroup frequencies, the two Bolivian samples presented significant differences. Since the Andean populations present a patrilocal system, we expected more similar haplogroup frequencies between the two Bolivian samples. The higher proportion of gene flow from external areas affecting the Quechua sample in the last centuries could explain this genetic differentiation for the mtDNA.

Moreover, our results on the uniparental markers indicated that the non-Native contribution to the Andean Bolivian gene pool would be almost exclusively by the male side.

IV.2.4 Concordance between different markers

Regarding the genetic relationships between the two Bolivian samples, there was high concordance among autosomal and X-chromosome data (*Alu* insertions and *APOE/C1/C4/C2* region). The main differences, as already mentioned, were found between the two uniparental systems, as well as among those and the autosomal loci. These differences could be explained by the different nature of the markers. First of all, it is important to have in mind that the mtDNA and Y-chromosome are just two loci and that they represent the history of just a part of the population. In contrast, autosomal data provide a more global picture of the population.

IV.3 Genetic relationships among South Americans

As for the amount of genetic variation, the two uniparental systems revealed a similar value around 25% of between-population genetic diversity among South Americans, a value much higher than that obtained from autosomal data (around 5% for PAIs and 7% for APOE isoforms).

IV.3.1 Genetic relationships between the two Bolivian samples and other Native Americans

The comparisons carried out in the three studies have been centred in South America. However, a Central American Mayan sample was included in Gaya-Vidal et al. (2010) and comparisons with Mayans and Pimas from Mexico were also carried out using a few SNPs in Gayà-Vidal et al. (3rd manuscript). The two works showed no differentiation between the Mayans and our Bolivian samples, in contrast to the differentiation from Pimas, a fact that agrees with previous findings that also reported short genetic distances between Mayan and Andean samples (Wang et al., 2007).

The relationships between the two Bolivian groups and other South American samples are highly related to the geographic distance or environmental differentiation. However, it is a fact that the Quechua sample has been in closer contact with people from other regions, according to certain mtDNA haplotypes characteristics of the Guarani and northwest Argentinean groups (Marrero et al., 2007; Tamm et al., 2007) and a shared Y-chromosome haplotype between Quechuas and Tobas. A more detailed explanation is provided in the following paragraphs.

IV.3.1 Genetic relationships among Andean populations

Regarding the Andean range, the most important genetic differentiation was observed between south *vs.* central Andean populations for both the mtDNA and Y-chromosome data (Gayà-Vidal et al., 2011). Unfortunately, no data on south Andean samples were available for autosomal markers (*Alu* inserions or APOE isoforms). This differentiation was rather expected due to the geographical distance separating these two regions and the genetic characteristics of the two areas, mainly for the mtDNA, (Andean region: high frequency of haplogroup B2, and Southern cone: high frequencies of haplogroups C and D).

IV.3.1.1 Genetic relationships among Central Andean populations

The central Andean region showed quite a genetic homogeneity, especially for the Y-chromosome, as demonstrated by the most frequent Y-chromosome haplotype in the region, and the DYS393*14 allele, two features that were practically present only in the central Andean area, as well as a high frequency of mtDNA haplogroup B2. Although during the last centuries different factors have contributed to the homogenization of this region, certain differentiation between some Andean groups was also detected, which might provide insights into the history of these groups.

This would be the case of the Kichwas from Ecuador, which appeared quite separated from the other central Andean samples that formed a group for the Y-chromosome. This separation between the Kichwas and the other Quechua-speakers from Peru or Bolivia could confirm that this population speaks a Quechua dialect that was adopted as *lingua franca* without important interactions with the south-central Andes.

At this point it is important to mention the Cayapas. This group inhabits the tropical forest of the Pacific coast of Ecuador and belongs to the Chibcha-Paezan linguistic branch. Their origin is controversial; the oral tradition suggests that they migrated from Ibarra (Andean highlands) to the forest to escape the Inca expansion, whereas some authors propose an Amazonian origin according to linguistics, and others suggest that in fact, they migrated from the Amazon to the Andean highlands and from there to the Pacific coast. This hypothesis is supported by their high adaptation to the tropical forest (see Rickards et al., 1999). In the present study, the inclusion of the Cayapas in the north-central Andean group, together with Kichwas, for the Y-chromosome analyses or with northern Peruvian samples for the mtDNA analysis, did not change the results, indicating certain genetic contribution from the Andean groups into the Cayapas. Nevertheless, significant differences for the APOE isoform frequencies were found between the Bolivian samples and the Cayapas that could indicate the action of selection pressure on the APOE isoforms since these samples inhabit different environments and are well adapted to it (Bolivians to high-altitude and Cayapas to the forest). Andrade et al. (2000) proposed drift and gene flow as the main factors of the APOE isoform frequencies observed in South America. In any case, more Andean samples are needed to confirm these hypotheses.

The case of the three northwest Argentinean samples (Salta, Tucuman and Catamarca) presenting strong differences from the Bolivian samples according to the mtDNA is also worth noting, especially the case of the Salta sample. In terms of geographical distance we could expect a closer position between these samples and the Bolivian ones, as well as among them, because all of them are located in the northwest of Argentina. However, an important aspect of these samples was that they represent political subdivisions rather than ethnic groups, a fact that implies that these samples may include people from different origins. In this way, it is important to distinguish which kind of data can contribute to the knowledge of the origin and history of Native American populations with consistent conclusions.

Thus, without considering the Kichwas and the northwest samples of Argentina, with the exception of the Coyas, the remaining central Andean populations were quite homogeneous, specially for the Y-chromosome (some samples only shared minimal haplotypes with other Andean groups and the minimal haplotype 13-14-31-23-10-16-14-15-18 was the most frequent among Andean samples, shared by the Aymara, Quechua, Kolla, Colla and Peru samples). As for the autosomal data, comparisons were carried out among only four samples (the two Bolivian and two from Peru) for eight *Alu* insertions. Contrary to the Y-chromosome, the analyses did not show a clear central Andean cluster. More precisely, the two Bolivian samples were very close and separated from the Peruvians (also separated). These results indicated that the *Alu* insertions might reveal genetic differences between central Andean groups that would be very interesting to shed light into their history. Nevertheless, more loci and more samples should be studied to obtain more robust conclusions.

Within the central Andean region, the Altiplano was particularly interesting. In addition to the separation from the Peruvian samples for the *Alu* insertions, it is a region that concentrates certain Andean genetic features as the highest frequencies of the mtDNA haplogroup B2, the Y-chromosome paragroup Q1a3*, the mtDNA 16188-16183C combination, and the DYS393*14 allele. The extremely high similarity between the Aymaras and Quechuas of this study, especially the Quechua sample, with the Coya sample in the north-west of Argentina was remarkably interesting, highlighting the high degree of the Altiplano influence in this region, probably since ancient times. In fact, the term Coya was used by the Incas to refer to the Aymara inhabitants. The Incas conquered the Aymara territories forming the southeastern provincial region of the Inca Empire (Collasuyu). This relationship with the Coyas contrast with that of the other northwest of Argentina, highlighting the importance of an ethnic definition of the samples for these kinds of studies.

In any case, more samples are necessary, mainly from Peru for the Y-chromosome and for the *Alu* insertions and APOE isoforms to obtain more robust results.

IV.3.2 Genetics *vs.* linguistics

The concordance between genetics and linguistics in South America has been studied for the *Alu* insertions (Gayà-Vidal et al., 2010). These markers failed to indicate strong clustering according to linguistic criteria in South America. Our results were in agreement with some previous reports which revealed a positive correlation at a language level (Fagundes et al., 2002; Mateus-Pereira et al., 2005) and at a stock level (Mateus-Pereira et al., 2005), but none at a phyla level, using the Loukotka language classification, which corresponds to the linguistic sub-family level according to the Greenberg classification, considered in Gayà-Vidal et al. (2010). The controversial results in the literature highlight the complexity of this subject, as discussed in Hunley et al. (2007). According to these authors, the observed absence of correlation could be expected, considering the deep linguistic branches of Greenberg's classification. In this context, our results for the autosomal *Alu* variation confirmed the absence of genetic-linguistic congruence regarding these linguistic sub-families in South Native Americans.

Concerning the Andean samples, the differentiation of the south Andes *vs.* the central Andes (both groups included into the Andean linguistic family) indicated the absence of such correspondence.

As for the central Andean region, the lack of genetic structure when the samples were grouped according to their language, Quechua or Aymara, indicated that the language is not a good indicator of the genetic background of the population. If some structure exists in this region it would rather depend on the geography. This result is probably related to the fact that Quechua-speakers come from different ethnic groups (Aymara speakers, and Peruvian groups belonging to different cultures that adopted the Quechua language). This would explain the high similarity of the Bolivian samples despite the fact that they speak different languages, and the quite differentiation of some Peruvian Quechua-speaking groups, such as the Tayacaja sample for both the *Alu* insertions and uniparental markers, despite sharing the same language. This is important because it means that even though high levels of gene flow have existed in the region, in part favoured by the intercommunication role of the Quechua language, they have not been enough to erase all genetic differences between population groups.

IV.3.3 Geography *vs*. genetics

The different markers studied in this work did not show a general correspondence of genetics compared to geography in South America. Nevertheless, it is important to mention that results can vary depending on the geographic level, the geographic area, and the markers considered.

None of the analyses carried out for the East-West differentiation showed a significant clustering of these two main areas in South America for any kind of marker (*Alus*, mtDNA and Y-chromosome, APOE isoforms). However, a different pattern of variation was observed in the Andean region compared to the East. The eastern populations showed larger within-group genetic distances and lower within-population diversity parameters as compared to the Andean region for both autosomal and uniparental markers (Gayà-Vidal, et al., 2010, 2011). This is consistent with different patterns of drift and gene flow previously suggested for different kinds of markers: mtDNA (Merriwether et al., 1995; Fuselli et al., 2003), Y chromosome (Tarazona-Santos et al., 2001), classical markers (Luiselli et al., 2000), and STR data (Wang et al., 2007) that would be related to larger effective population sizes in the Andean area (Fuselli et al., 2003). As in the study of Lewis and Long, (2008), our data revealed a clear heterogeneity among Eastern populations that did not form a cohesive genetic group, most of the samples exhibiting low sample sizes.

Regarding the Andean mountains, as already mentioned, the south Andes was differentiated from the central Andes according to mtDNA, in spite of sharing the Andean linguistic family. This difference is in agreement with the geographic distance. However, more data on other markers are necessary to confirm this separation.

The central Andean region presented significant differentiation from two close areas, Llanos and Bolivia Lowlands, according to the Y-chromosome, although some male gene flow was detected. In contrast, this structuring was not observed for the mtDNA data that revealed certain genetic relationships between these regions. This influence was reflected by the presence of the 16188 variant in two populations from the Chaco. These surrounding areas also presented high genetic diversities, similar to that of Andean samples that could reflect large population sizes (Gayà-Vidal et al., 2011).

Concerning the central Andean region, all results point to the persistence of certain degree of population differentiation in this region that would not depend on the linguistics as we have just said, but would be consistent with the history of the central Andean region, where different cultures/ethnics took place (in different geographic areas) in spite of the fact that several civilizations had a general influence across the region, mainly the acculturation by the Incas, and subsequently by the Spaniards. The differentiation of the Altiplano from the rest of the Andean region is especially important, indicating a particular history of this area.

V. CONCLUSIONS

The main conclusions obtained from this work are the following:

1. The variation of the analysed markers showed a clear genetic similarity between the two Bolivian samples studied in this work.
2. The Quechua expansion into Bolivia would mainly be the result of cultural rather than demographic processes.
3. The Quechua sample has been more permeable to the incorporation of non-Natives and other South Americans, mainly from Peru, in contrast to the Aymaras that have been more isolated throughout time.
4. The non-Native contribution to the genetic pool of current Bolivian Andeans is likely to have been produced mainly by males.
5. The two Bolivian samples presented, in general, low genetic diversity, in agreement with other data from Native Americans, probably due to the important genetic drift and founder effects that characterise the demographic history of the Americas, and particularly that of South America.
6. In a South American context, the two Bolivian samples presented moderate to high genetic diversity in concordance with other Andean samples.
7. The Andean samples present higher within-population and lower inter-population diversity compared to the eastern populations of South America, in consistence with larger population sizes in the Andean region.
8. The mouvements (gene flow) across the central Andean region must have been generalised, mainly for the Y-chromosome, resulting in a general genetic homogeneity of the Andean region, although not complete.
9. Differences between the central Andean samples were detectable despite this general genetic homogeneity of the central Andean region. These differences among central Andean samples support the fact that the Quechua language is spoken by different ethnic groups.
10. The Altiplano seems to have had an independent, although related, history from the rest of the central Andean area. The strongest influence of the Altiplano culture was over the northwest Argentina.

VI. REFERENCES

Achilli A, Perego UA, Bravi CM, Coble MD, Kong QP, Woodward SR, Salas A, Torroni A, Bandelt HJ. The phylogeny of the four pan-American MtDNA haplogroups:implications for evolutionary and disease studies. *PLoS One*, 2008; 3(3):e1764.

Acosta Jde. 2002. Natural and Moral History of the Indies, J.E. Mangan, ed. Durham: Duke University Press.

Adelaar, WFH, and Muysken, PC. 2004. The Languages of the Andes. Cambridge Language Surveys. Cambridge University Press.

Afonso-Costa H, Carvalho M, Lopes V, Balsa F, Bento AM, Serra A, Andrade L, Anjos MJ, Vide MC, Pantoja S, Vieira DN, Corte-Real F. Mitochondrial DNA sequence analysis of a native Bolivian population. *J Forensic Leg Med*, 2010; 17(5):247-53.

Alvarez-Iglesias V, Jaime JC, Carracedo A, Salas A. Coding region mitochondrial DNA SNPs: targeting East Asian and Native American haplogroups. *Forensic Sci Int Genet*, 2007; 1(1):44-55.

Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJ, Staden R, Young IG. Sequence and organization of the human mitochondrial genome. *Nature*, 1981; 290(5806):457-65.

Andrade FM, Coimbra CE Jr, Santos RV, Goicoechea A, Carnese FR, Salzano FM, Hutz MH. High heterogeneity of apolipoprotein E gene frequencies in South American Indians. *Ann Hum Biol*, 2000; 27(1):29-34.

Antunez-de-Mayolo G, Antunez-de-Mayolo A, Antunez-de-Mayolo P, Papita SS, Hammer M, Yunis JJ, Yunis EJ, Damodaran C, Martinez de Pancorbo M, Caeiro JL, Puzyrev VP, Herrera RJ. Phylogenetics of worldwide human populations as determined by polymorphic Alu insertions. *Electrophoresis*, 2002; 23: 3346-3356

Arnaiz-Villena A, Siles N, Moscoso J, Zamora J, Serrano-Vela JI, Gomez-Casado E, Castro MJ, Martinez-Laso J. Origin of Aymaras from Bolivia and their relationship with other Amerindians according to HLA genes. *Tissue Antigens,* 2005; 65(4):379-90.

Arnaiz-Villena A, Parga-Lozano C, Moreno E, Areces C, Rey D, Gomez-Prieto P. The Origin of Amerindians and the Peopling of the Americas According to HLA Genes: Admixture with Asian and Pacific People. *Curr Genomics*, 2010; 11(2):103-14.

Bailliet G, Rothhammer F, Carnese FR, Bravi CM, Bianchi NO. Founder mitochondrial haplotypes in Amerindian populations. *Am J Hum Genet,* 1994; 55(1):27-33.

Bailliet G, Ramallo V, Muzzio M, García A, Santos MR, Alfaro EL, Dipierri JE, Salceda S, Carnese FR, Bravi CM, Bianchi NO, Demarchi DA. Brief communication: Restricted geographic distribution for Y-Q* paragroup in South America. *Am J Phys Anthropol,* 2009; 140(3):578-82.

Barbieri C, Heggarty P, Castrì L, Luiselli D, Pettener D. Mitochondrial DNA variability in the Titicaca basin: Matches and mismatches with linguistics and ethnohistory. *Am J Hum Biol,* 2011; 23(1):89-99.

Battilana J, Fagundes NJR, Heller AH, Goldani A, Freitas LB, Tarazona-Santos E, Munkhbat B, Munkhtuvshin N, Krylov M, Benevolenskaia L, Arnett FC, Batzer MA, Deininger PL, Salzano FM, Bonatto SL. Alu insertion polymorphisms in Native Americans and related Asian populations. *Ann Hum Biol*, 2006; 33(2):142-60.

Batzer MA, and Deininger PL. A human-specific subfamily of Alu sequences. *Genomics*, 1991; 9:481-487.

Batzer MA, Stoneking M, Alegria-Hartman M, Bazan H, Kass DH, Shaikh TH, Novick GE, Ioannou PA, Scheer WD, Herrera RJ, et al. African origin of human-specific polymorphic Alu insertions. *Proc Natl Acad Sci U S A*, 1994; 91(25):12288-92.

Batzer MA, Deininger PL. Alu repeats and human genomic diversity. *Nat Rev Genet*, 2002; 3(5):370-9.

Beckman JS, Weber JL. Survey of human and rat microsatellites. *Genomics*, 1992; 12(4):627-31.

Beall CM. Two routes to functional adaptation: Tibetan and Andean high-altitude natives. *Proc Natl Acad Sci U S A*, 2007; 104 Suppl :8655-60.

Bellido D, Aguilar M. 1992. Estudio de lípidos en nativos habitantes de grandes alturas. II congreso SOLAT (Sociedad latinoamericana de Aterosclerosis).

Bennett WC and Bird JB. 1964. Andean Culture History. GardenCity, NY: The Natural History Press.

Bert F, Corella A, Gené M, Pérez-Pérez A, Turbón D. Major mitochondrial DNA haplotype heterogeneity in highland and lowland Amerindian populations from Bolivia. *Hum bio*, 2001; 73(1):1-16.

Bianchi NO, Catanesi CI, Bailliet G, Martinez-Marignac VL, Bravi CM, Vidal-Rioja LB, Herrera RJ, López-Camelo JS. Characterization of ancestral and derived Y-chromosome haplotypes of New World native populations. *Am J Hum Genet,* 1998; 63(6):1862-71.

Bird RM, Browman DL, Durbin ME. Quechua and maize: mirrors of Central Andean culture history. *J Steward Anthropol Soc,* 1884; 15(1 & 2), 187–240.

Blom DE, Hallgrimsson B, Keng L, Lozada MC, Buikstra JE. Tiwanaku “Colonization”: Bioarchaeological Implications for Migration in the Moquegua Valley, Peru. *World Arch,* 1998; 30:238–261.

Bonatto SL, Salzano FM. A single and early migration for the peopling of the Americas supported by mitochondrial DNA sequence data. *Proc Natl Acad Sci U S A*, 1997; 94(5):1866-71.

Bortolini MC, Salzano FM, Thomas MG, Stuart S, Nasanen SP, Bau CH, Hutz MH, Layrisse Z, Petzl-Erler ML, Tsuneto LT, Hill K, Hurtado AM, Castro-de-Guerra D, Torres MM, Groot H, Michalski R, Nymadawa P, Bedoya G, Bradman N, Labuda D, Ruiz-Linares A. Y-chromosome evidence for differing ancient demographic histories in the Americas. *Am J Hum Genet*, 2003; 73(3):524-39.

Bouysse-Cassagne T. 1986. Urco and Uma: Aymara concepts of space. In: Murra JV, Wachtel N, Revel J, editors. Anthropological history of Andean polities. Cambridge: Cambridge University Press. p 201–227.

Brinkmann B, Junge A, Meyer E, Wiegand P. Population genetic diversity in relation to microsatellite heterogeneity. *Hum Mutat,* 1998; 11(2):135-44.

Britten RJ. Mobile elements inserted in the distant past have taken on important functions. *Gen*e, 1997; 205(1-2):177-82.

Browman DL. Titicaca basin archaeolinguistics: Uru, Pukina and Aymara. *World Arch,* 1994; 26(2): 235-251

Brown WM, George M Jr, Wilson AC. Rapid evolution of animal mitochondrial DNA. *Proc Natl Acad Sci U S A*, 1979;76(4):1967-71.

Cabana GS, Merriwether DA, Hunley K, Demarchi DA. Is the genetic structure of Gran Chaco populations unique? Interregional perspectives on native South American mitochondrial DNA variation. *Am J Phys Anthropol*, 2006; 131(1):108-19.

Caen JP, Drouet L, Bellanger R, Michel H, Henon P. Thrombosis, platelet behaviour, fibrinolytic activity and diet on the Andes Plateau. *Haemostasis,* 1973/-1974; 2: 13-20.

Campbell L. 1997. American Indian Languages: The Historical Linguistics of Native America. Oxford: Oxford University Press.

Carrol ML, Roy-Engel AM, Nguyen SV, Salem AH, Vogel E, Bethaney V, Myers J, Ahmad Z, Nguyen L, Sammarco M, Watkins WS, Henke J, Makalowski W, Jorde LB, Deininger PL, Batzer M. Large-scale analysis of the Ya5 and Ya8 subfamilies and their contribution to human genomic diversity. *J Mol Biol*, 2001; 311:17-40.

Cavalli-Sforza LL, Menozzi R, Piazza A. 1994. The History and Geography of Human Genes. Princeton, NJ, Princeton University Press.

Cavalli-Sforza LL, Feldman MW. The application of molecular genetic approaches to the study of human evolution. *Nat Genet*, 2003; 33 Suppl(march):266-75.

Cerrón-Palomino R. 2000. Lingüística Aimara. Cuzco: C.E.R.A. "Bartolomé de Las Casa".

Collins FS, Green ED, Guttmacher AE, Guyer MS; US National Human Genome Research Institute. A vision for the future of genomics research. *Nature*, 2003; 422(6934):835-47.

Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y, Aerts J, Andrews TD, Barnes C, Campbell P, Fitzgerald T, Hu M, Ihm CH, Kristiansson K, Macarthur DG, Macdonald JR, Onyiah I, Pang AW, Robson S, Stirrups K, Valsesia A, Walter K, Wei J; Wellcome Trust Case Control Consortium, Tyler-Smith C, Carter NP, Lee C, Scherer SW, Hurles ME. Origins and functional impact of copy number variation in the human genome. *Nature*, 2010; 464(7289):704-12.

Cordaux R, Srikanta D, Lee J, Stoneking M, Batzer MA. In search of polymorphic Alu insertions with restricted geographic distributions. *Genomics,* 2007; 90(1):154-8.

Cordaux R, Batzer MA. The impact of retrotransposons on human genome evolution. *Nat Rev Genet,* 2009; 10(10):691-703.

Corella A, Bert F, Pérez-Pérez A, Gené M, Turbón D. Mitochondrial DNA diversity of the Amerindian populations living in the Andean Piedmont of Bolivia: Chimane, Moseten, Aymara and Quechua. *Ann Hum Biol,* 2007; 34(1):34-55.

Crawford MH. 1998. The Origins of Native Americans, Evidence from Anthropological Genetics. Cambridge University Press.

Crognier E, Villena M, Vargas E. Helping patterns and reproductive success in Aymara communities. *Am J Hum Biol*, 2002; 14(3):372-9.

Crognier E, Villena M, Vargas E. Reproduction in high altitude Aymara: physiological stress and fertility planning? *J Biosoc Sci*, 2002; 34(4):463-73.

Crognier E, Baali A, Hilali MK, Villena M, Vargas E. Preference for sons and sex ratio in two non-Western societies. *Am J Hum Biol*, 2006; 18(3):325-34.

Cruz P. Mundos permeables y espacios peligrosos. Consideraciones acerca de Punkus y Qaqas en el paisaje altoandino de Potosí, Bolivia. *Boletín del museo chileno de arte precolombino*, 2006; 11(2):33-50.

Deininger PL, Daniels, GL. The recent evolution of mammalian repetitive DNA elements. *Trends Genet*, 1986; 2:76-80.

Deininger, PL, Batzer, MA, Hutchison III, CA, Edgell, MH, Master genes in mammalian repetitive DNA amplification. *Trends Genet*, 1992; 8, 307–311.

Dillehay TD. 1997. Monte Verde. A Late Pleistocene Settlement in Chile. Vol. 2. The Archaeological Context and Interpretation, p. 1071. Washington, DC, London: Smithsonian Institution Press.

Di Rienzo A, Peterson AC, Garza JC, Valdes AM, Slatkin M, Freimer NB.Mutational processes of simple-sequence repeat loci in human populations. Proc *Natl Acad Sci U S A*, 1994; 91(8):3166-70.

Dipierri JE, Alfaro E, Martínez-Marignac VL, Bailliet G, Bravi CM, Cejas S, Bianchi NO. Paternal directional mating in two Amerindian subpopulations located at different altitudes in northwestern Argentina. *Hum Biol*, 1998; 70(6):1001-10.

Dornelles CL, Battilana J, Fagundes NJR, freitas LB, Bonatto SL, Salzano FM. Mitochondrial DNA and Alu insertions in a Genetically peculiar population: The Ayoreo Indians of Bolivia and Paraguay. *Am J Hum Biol*, 2004; 16:479-488.

Dugoujon JM, Mourrieras B, Senegas MT, Guitard E, Sevin A, Bois E, Hazout S. Human genetic diversity (immunoglobulin GM allotypes), linguistic data, and migrations of Amerindian tribes. *Hum Biol,* 1995; 67(2):231-49.

Dugoujon JM, Hazout S, Loirat F, Mourrieras B, Crouau-Roy B, Sanchez-Mazas A. GM haplotype diversity of 82 populations over the world suggests a centrifugal model of human migrations. *Am J Phys Anthropol,* 2004; 125(2):175-92.

Fagundes NJ, Bonatto SL, Callegari-Jacques SM, Salzano FM. Genetic, geographic, and linguistic variation among South American Indians: possible sex influence. *Am J Phys Anthropol,* 2002; 117(1):68-78.

Fagundes NJR, Kanitz R, Bonatto SL. A reevaluation of the Native American mtDNA genome diversity and its bearing on the models of early colonization of Beringia. *PloS one,* 2008; 3(9):e3157.

Fladmark, KR. Routes: Alternative Migration Corridors for Early Man in North America. *Am Antiquity*, 1979; 44, 55–69.

Fuselli S, Tarazona-Santos E, Dupanloup I, Soto A, Luiselli D, Pettener D. Mitochondrial DNA diversity in South America and the genetic history of Andean highlanders. *Mol Biol Evol*, 2003; 20(10):1682-1691.

Gayà-Vidal M, Dugoujon JM, Esteban E, Athanasiadis G, Rodríguez A, Villena M, Vasquez R, Moral P. Autosomal and X chromosome Alu insertions in Bolivian Aymaras and Quechuas: two languages and one genetic pool. *Am J Hum Biol,* 2010; 22(2):154-62.

Gayà-Vidal M, Moral P, Saenz-Ruales N, Gerbault P, Tonasso L, Villena M, Vasquez R, Bravi CM, Dugoujon JM. mtDNA and Y-chromosome diversity in

Aymaras and Quechuas from Bolivia: Different stories and special genetic traits of the Andean Altiplano populations. *Am J Phys Anthropol*, 2011; 145(2):215-30.

Giles RE, Blanc H, Cann HM, Wallace DC. Maternal inheritance of human mitochondrial DNA. *Proc Natl Acad Sci U S A*, 1980; 77(11):6715-9.

Goebel T, Waters MR, O'Rourke DH. The late Pleistocene dispersal of modern humans in the Americas. *Science,* 2008; 319(5869):1497-502.

González-Andrade F, Sánchez D, González-Solórzano J, Gascón S, Martínez-Jarreta B. Sex-specific genetic admixture of Mestizos, Amerindian Kichwas, and Afro-Ecuadorans from Ecuador. *Hum Biol,* 2007; 79:51–77.

González-José R, Bortolini MC, Santos FR, Bonatto SL. The peopling of America: craniofacial shape variation on a continental scale and its interpretation from an interdisciplinary view. *Am J Phys Anthropol,* 2008; 137(2):175-87.

Goulder P. Diasporas and Shared Knowledge: the case of regional, endangered, autochthonous languages. *Knowledge Creation Diffusion Utilization*. 2003:125-135.

Goulder P. The Languages of Peru: their past, present and future survival. *Review Literature And Arts Of The Americas*. 2003:166-196.

Graffam G. Beyond state collapse: rural history, raised fields, and pastoralism in South Andes. *Am Anthropol,* 1992; 94(4):882-904.

Greenberg M. 1959. Linguistic classification of South America. In Native peoples of South America. J.H. Steward and L.C. Faron, (eds.). New york: McGraw-Hill.

Greenberg JH, Turner CG II, Zegura SL. The settlement of the Americas: a comparison of the linguistic, dental and genetic evidence. *Curr Anthropol,* 1986; 27:477-497.

Greenberg JH. 1987. Language in the Americas. Stanford: Stanford University Press.

Greenberg JH, Ruhlen M. 2007. An Amerind Etymological Dictionary, Department of Anthropology, Stanford University Press.

Haynes G. 2002. The Early Settlement of North America: The Clovis Era. Cambridge University Press, Cambridge.

Hazout S, Guasp G, Loirat F, Maurieres P, Larrouy G, Dugoujon JM. A new approach for interpreting the genetic diversity in space: 'Mobile Site Method'. Application to Gm haplotype distribution of twenty-seven Amerindian tribes from North and Central America. *Ann Hum Genet,* 1993; 57(Pt 3):221-37.

Heggarty P. Linguistics for Archaeologists: Principles, Methods and the Case of the Incas. *Camb Archaeol J,* 2007; 17:3, 311–40.

Heggarty P. Linguistics for Archaeologists: a Case-study in the Andes. *Cambridge Archaeological Journal*. 2008; 18(01).

Hrdlicka, A. The origin and antiquity of the American Indian. *Annu Rep Smithson Inst Washington,* 1937; 1923, 481–494.

Hunley KL, Cabana GS, Merriwether DA, Long JC. A formal test of linguistic and genetic coevolution in Native Central and South America. *Am J Phys Anthropol,* 2007; 132:622–631.

Houck CM, Rinehart FP, Schmid CW. A ubiquitous family of repeated DNA sequences in the human genome. *J Mol Biol,* 1979; 132, 289–306.

Isbell WH, Schreiber KJ, Antiquity A, Jul N. Was Huari a State ? *Horizon,* 2008; 43(3):372-389.

Itier C. 2002. Languages of the Americas 2. Quechua, Aymara and Other Andean Languages: Historical, Linguistic and Socio-linguistic Aspects. Lecture given on Wednesday 16 January 2002 at 6.30 p.m.at Maison de l'Amérique Latine, 217 boulevard Saint Germain, 75007 Paris, organised by CECUPE (le Centre Culturel Péruvien / Peruvian Cultural Centre).

Jobling MA, Hurles ME, Tyler-Smith C. 2004. Human evolutionary genetics. Origins, peoples and disease. Garland Science, New York.

Karafet TM, Zegura SL, Posukh O, Osipova L, Bergen A, Long J, Goldman D, Klitz W, Harihara S, de Knijff P, Wiebe V, Griffiths RC, Templeton AR, Hammer MF. Ancestral Asian source(s) of new world Y-chromosome founder haplotypes. *Am J Hum Genet,* 1999; 64(3):817-31.

Karafet TM, Mendez FL, Meilerman MB, Underhill PA, Zegura SL, Hammer MF. New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Res,* 2008; 18(5):830-8.

Kolata A. 1993. The Tiwanaku: portrait of an Andean civilization. The people of America. Cambridge: Blackwell Publishers.

Kolman CJ, Bermingham E, Cooke R, Ward RH, Arias TD, Guionneau-Sinclair F. Reduced mtDNA diversity in the Ngöbé Amerinds of Panamá. *Genetics,* 1995; 140:275–283.

Landsteiner K. Ueber Agglutinationserscheinungen normalen menschlichen Blutes. *Wiener Klinische Wochenschrift,* 1901; 46,1132–1134.

Lathrap, D. 1970. The Upper Amazon. New York: Praeger Publishers.

Lee J, Ewis AA, Hurles ME, Kashiwazaki H, Shinka T, Nakahori Y. Y chromosomal STRs haplotypes in two populations from Bolivia. *Leg Med* (Tokyo), 2007; 9(1):43-7.

Lell JT, Sukernik RI, Starikovskaya YB, Su B, Jin L, Schurr TG, Underhill PA, Wallace DC. The dual origin and Siberian affinities of Native American Y chromosomes. *Am J Hum Genet,* 2002; 70(1):192-206.

Lewis CM Jr, Lizárraga B, Tito RY, López PW, Iannacone GC, Medina A, MartínezR, Polo SI, De La Cruz AF, Cáceres AM, Stone AC. Mitochondrial DNA and the peopling of South America. *Hum Biol,* 2007; 79(2):159-78.

Lewis CM, Long JC. Native South American genetic structure and prehistory inferred from hierarchical modeling of mtDNA. *Mol Biol Evol,* 2008; 25(3):478-86.

Lewis M. Paul (ed.). 2009. Ethnologue: Languages of the World, Sixteenth edition. Dallas, Tex.: SIL International. Online version: http://www.ethnologue.com/**.**

Lewontin RC. The apportionment of human diversity. *Evolutionary Biology,* 1972; 6 : 381-398.

Loukotka C. 1968. Clasification of South American Indian Languages. J. Wilbert, ed. Latin American Center-UCLA. Los Angeles.

Luiselli D, Simoni L, Tarazona-Santos E, Pastor S, Pettener D. Genetic structure of Quechua-speakers of the Central Andes and geographic patterns of gene frequencies in South Amerindian populations. *Am J Phys Anthropol,* 2000; 113(1):5-17.

Markham, CE. On the geographical positions of the tribes which formed the Empire of the Yncas, with an Appendix on the name “Aymara”. *J Roy Geog Soc*, 1871; vol. 41, pp. 281-338.

Marrero AR, Silva-Junior WA, Bravi CM, Hutz MH, Petzl-Erler ML, Ruiz-Linares A, Salzano FM, Bortolini MC. Demographic and evolutionary trajectories of the Guarani and Kaingang Natives of Brazil. *Am J Phys Anthropol*, 2007; 132:301–310.

Martinez-Laso J, Siles N, Moscoso J, Zamora J, Serrano-Vela JI, R-A-Cachafeiro JI, Castro MJ, Serrano-Rios M, Arnaiz-Villena A. Origin of Bolivian Quechua Amerindians: their relationship with other American Indians and Asians accordingto HLA genes. *Eur J Med Genet*. 2006; 49(2):169-85.

Martínez-Marignac VL, Bailliet G, Dipierri JE, Alfaro E, López-Camelo JS, Bianchi NO. Variabilidad y antigüedad de linajes holandricos en poblaciones jujeñas. *Rev Arg Antropol Biol,* 2001; 3:65–77.

Mason JA. 1963. The language of South American Indians. In Handbook of South American Indians. Vol.VI. J.H. Steward, (ed) pp.157–318. New York: Cooper Square Publ.Inc.

Mateus-Pereira LH, Socorro A, Fernandez I, Masleh M, Vidal D, Bianchi NO, Bonatto SL, Salzano FM, Herrera RJ. Phylogenetic information in polymorphic L1 and Alu insertions drom East Asians and Native American populations. *Am J Phys Anthropol,* 2005; 128:171-184.

Merriwether DA, Clark AG, Ballinger SW, Schurr TG, Soodyall H, Jenkins T, Sherry ST, Wallace DC. The structure of human mitochondrial DNA variation. *J MolEvol,* 1991; 33(6):543-55.

Merriwether DA, Rothhammer F, Ferrell RE. Distribution of the four founding lineage haplotypes in Native Americans suggests a single wave of migration for the New World. *Am J Phys Anthropol,* 1995; 98(4):411-30.

Mortimer EA, Monson RR, MacMahon B: Reduction in mortality from coronary heart disease in men residing at high altitude. *N Engl J Med,* 1977; 296:581-585.

Mourant AE, Lopel AC, Domaniewski-Sobezak K. 1976. The distribution of the human blood groups and other polymorphisms, Second edition, Oxford University Press.

Mourrieras B, Dugoujon JM, Buffat L, Hazout S. Assessment of genetic diversity in space by superimposition of a distorted geographic map with a spatial population clustering. Application to GM haplotypes of native Amerindian tribes. *Ann Hum Genet,* 1997; 61(Pt 1):37-47.

Mullaney JM, Mills RE, Pittard WS, Devine SE. Small insertions and deletions (INDELs) in human genomes. *Hum Mol Genet,* 2010; 19(2):131-136.

Mulligan CJ, Kitchen A, Miyamoto MM. Updated three-stage model for the peopling of the Americas. *PloS one,* 2008; 3(9):e3199.

Murra JV. 1972. El "control vertical" de un máximo de pisos ecológicos en la economía de las sociedades andinas. In I. Ortiz de Zúñiga (ed.) Visita de la provincia de león de Huanuco en 1562, Documentos para la historia y etnología de Huánaco y la Selva Central, 2, pp. 427-76, Huanuco, Perú: Universidad Nacional Hermilio Valdizán.

Murra JV. Andean Societies. *Annu Rev Anthropol,* 1984; 13: 119-141.

Neckelmann N, Li K, Wade RP, Shuster R, Wallace DC. cDNA sequence of a human skeletal muscle ADP/ATP translocator: lack of a leader peptide, divergence from a fibroblast translocator cDNA, and coevolution with mitochondrial DNA genes. *Proc Natl Acad Sci U S A,* 1987; 84(21):7580-4.

Neves WA, Hubbe M. Cranial morphology of early Americans from Lagoa Santa, Brazil: implications for the settlement of the New World. *Proc Natl Acad Sci U S A,* 2005; 102(51):18309-14.

Novick GE, Novick CC, Yunis J, Yunis E, Antunez-de-Mayolo P, Douglas Scheer W, Deininger PL, Stoneking M, York DS, Batzer MA, Herrera RJ. Polymprphic Alu insertions and the Asian origin of Native American populations. *Hum Biol,* 1998; 70 (1):23-39.

O'Rourke DH, Raff JA. The human genetic history of the Americas: the final frontier. *Current biology,* 2010; 20(4):R202-7.

Orr C, Longacre RE. Proto-Quechumaran. *Language,* 1968; 44(3):528.

Oven M van, Kayser M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat,* 2009; *30(2):E386-94.*

Oven M van. Revision of the mtDNA tree and corresponding haplogroup nomenclature. *Proc Natl Acad Sci U S A,* 2010; 107(11):E38-9; author reply e40-1.

Owen B. 1994. Were Wari and Tiwanaku in Conflict, Competition, or Complementary Coexistence? Survey Evidence from the Upper Osmore Drainage, Perú. ponencia presentada en el 59th Annual Meeting of the Society for American Archaeology, Anaheim.

Parham P, Ohta T. Population biology of antigen presentation by MHC class I molecules. *Science,* 1996; 272, 67-74.

Perego UA, Achilli A, Angerhofer N, Accetturo M, Pala M, Olivieri A, Kazan BH, Ritchie KH, Scozzari R, Kong QP, Myres NM, Salas A, Semino O, Bandelt HJ, Woodward SR, Torroni A. Distinctive Paleo-Indian migration routes from Beringia marked by two rare mtDNA haplogroups. *Curr Biol,* 2009; 19(1):1-8.

Perego UA, Angerhofer N, Pala M, Olivieri A, Lancioni H, Kashani BH, Carossa V, Ekins JE, Gómez-Carballa A, Huber G, Zimmermann B, Corach D, Babudri N, Panara F, Myres NM, Parson W, Semino O, Salas A, Woodward SR, Achilli A, Torroni A. The initial peopling of the Americas: a growing number of founding mitochondrial genomes from Beringia. *Genome Res,* 2010; 20(9):1174-9.

Platt T, Bouysse-Cassagne T, Harris O. 2006. Qaraqara-Charka, Mallku, Inka y Rey en la provincia de Charcas (siglos XV-XVII). Historia antropológica de una confederación aymara. Institut français d'études andines-IFEA; Plural editores; University of St. Andrews; University of London; Interamerican Foundation; Fundación Cultural del Banco Central de Bolivia, La Paz.

Pringle H. Archaeology. Texas site confirms pre-Clovis settlement of the Americas. *Science,* 2011; 331(6024):1512.

Quintana-Murci L, Veitia R, Santachiara-Benerecetti S, McElreavey K, Fellous M, Bourgeron T. L'ADN mitochondrial, le chromosome Y et l'histoire des populations humaines. *médecine/sciences,* 1999; 9(15): 974-82.

Rickards O, Martínez-Labarga C, Lum JK, De Stefano GF, Cann RL. mtDNA history of the Cayapa Amerinds of Ecuador: detection of additional founding lineages for the Native American populations. *Am J Hum Genet,* 1999; 65(2):519-30.

Rothhammer F, Silva C. Gene geography of South America: testing models of population displacement based on archeological evidence. *Am J Phys Anthropol,* 1992; 89(4):441-6.

Rothhammer F, Llop E, Carvallo P, Moraga M. Origin and evolutionary relationships of native Andean populations. *High Alt Med Biol,* 2001; 2(2):227-33.

Rothhammer F, Dillehay TD. The late Pleistocene colonization of South America: an interdisciplinary perspective. *Ann Hum Genet,* 2009; 73(Pt 5):540-9.

Rowe JH. 1963. Handbook of South American Indians. Vol.2: Inca culture at the time of the spanish conquest.Julian H. Steward, ed. New York: Cooper Square.

Roy AM, Carroll, ML, Kass, DH, Nguyen, SV, Salem, A-H, Batzer, MA, and Deininger, PL. Recently integrated human Alu repeats: Finding needles in the haystack. *Genetica,* 1999; 107:149–161.

Roy-Engel AM, Carroll ML, Vogel E, Garber RK, Nguyen SV, Salem AH, Batzer MA, Deininger PL. Alu insertion polymorphisms for the study of human genomic diversity. *Genetics,* 2001; 159(1):279-90.

Ruhlen Merritt. 1994. On the Origin of Languages. Studies in Linguistic Taxonomy. Stanford: Stanford University.

Rupert JL, Hochachka PW. The evidence for hereditary factors contributing to high altitude adaptation in Andean natives: a review. *High Alt Med Biol,* 2001; 2(2):235-56.

Salzano FM, Callegari-Jacques SM. 1988. South American Indians: A case study in evolution. Oxford: Clarendon Press.

Sanchez-Albornoz N. 1974. The Population of Latin America: A History – Unknown binding University of California Press.

Sandoval J, Delgado B, Rivas L, Bonilla B, Nugent D, Fujita R. Variants of mtDNA among islanders of the lake Titicaca: highest frequency of haplotype B1 and evidence of founder effect. *Rev Peru Biol,* 2004; 11(2):161-168.

Santos M, Barrantes R. D-loop mtDNA deletion as a unique marker of Chibchan Amerindians. *Am J Hum Genet*, 1994; 55:413–414.

Schroeder KB, Jakobsson M, Crawford MH, Schurr TG, Boca SM, Conrad DF, Tito RY, Osipova LP, Tarskaia LA, Zhadanov SI, Wall JD, Pritchard JK, Malhi RS, Smith DG, Rosenberg NA. Haplotypic background of a private allele at high frequency in the Americas. *Mol Biol Evol,* 2009; 26(5):995-1016.

Schurr TG, Ballinger SW, Gan YY, Hodge JA, Merriwether DA, Lawrence DN, Knowler WC, Weiss KM, Wallace DC. Amerindian mitochondrial DNAs have rare Asian mutations at high frequencies, suggesting they derived from four primary maternal lineages. *Am J Hum Genet,* 1990; 46(3):613-23.

Schurr TG, Sherry ST. Mitochondrial DNA and Y chromosome diversity and the peopling of the Americas: evolutionary and demographic evidence. *Am J Hum Biol,* 2004; 16(4):420-39.

Sondereguer C, Punta C. 1999. Amerindia. Introducción a la etnohistoria y las artes visuales precolombinas. Ed. Corregidor. Colección Amerindia Arte.

Stanish C. The origin of state societies in South America. *Annu Rev Anthropol,* 2001; 30:41–64.

Stobdan T, Karar J, Pasha MAQ. High altitude adaptation: genetic perspectives. *High Alt Med Biol*, 2008; 9(2):140-7.

Stoneking M, Fontius JJ, Clifford SL, Soodyall H, Arcot SS, Saha N, Jenkins T, Tahir MA, Deininger PL, Batzer MA. Alu insertion polymorphisms and human evolution: evidence for a larger population size in Africa. *Genome Res,* 1997; 7(11):1061-71.

Stoneking M, Delfin F. The human genetic history of East Asia: weaving a complex tapestry. *Curr Biol*, 2010; 20(4):R188-93.

Tamm E, Kivisild T, Reidla M, Metspalu M, Smith DG, Mulligan CJ, Bravi CM, Rickards O, Martinez-Labarga C, Khusnutdinova EK, Fedorova SA, Golubenko MV, Stepanov VA, Gubina MA, Zhadanov SI, Ossipova LP, Damba L, Voevoda MI, Dipierri E, Villems R, Malhi RS. Beringian standstill and spread of Native American founders. *PLoS One*, 2007; 2(9):e829.

Tarazona-Santos E, Carvalho-Silva DR, Pettener D, Luiselli D, De Stefano GF, Labarga CM, Rickards O, Tyler-Smith C, Pena SD, Santos FR. Genetic differentiation in South Amerindians is related to environmental and cultural diversity: evidence from the Y chromosome. *Am J Hum Genet*, 2001; 68(6):1485-96.

Tishkoff SA, Dietzsch E, Speed W, Pakstis AJ, Kidd JR, Cheung K, Bonné-Tamir B, Santachiara-Benerecetti AS, Moral P, Krings M, Pääbo S, Watson E, Risch N, Jenkins T, Kidd KK. Global patterns of linkage disequilibrium at the CD4 locus and modern humans origins. *Science*, 1996; 271:1380–1387.

Tishkoff SA, Goldman A, Calafell F, Speed WC, Deinard AS, Bonne-Tamir B, Kidd JR, Pakstis AJ, Jenkins T, Kidd KK. A global haplotype analysis of the myotonic dystrophy locus: implications for the evolution of modern humans and for the origin of myotonic dystrophy mutations. *Am J Hum Genet*, 1998; 62:1389–1402.

Torero A. 1983. La familia lingüística quechua. In: POTTIER, Bernard (ed.). América Latina en sus lenguas indígenas. Caracas: UNESCO y Monte Ávila, pp. 61-92.

Torero A. 2002. Idiomas de los Andes: Lingüística e Historia. Lima: Editorial Horizonte e IFEA.

Torroni A, Schurr TG, Yang CC, Szathmary EJ, Williams RC, Schanfield MS, Troup GA, Knowler WC, Lawrence DN, Weiss KM, et al. Native American mitochondrial DNA analysis indicates that the Amerind and the Nadene populations were founded by two independent migrations. *Genetics*, 1992; 130(1):153-62.

Torroni A, Schurr TG, Cabell MF, Brown MD, Neel JV, Larsen M, Smith DG, VulloCM, Wallace DC. Asian affinities and continental radiation of the four founding Native American mtDNAs. *Am J Hum Genet*, 1993; 53(3):563-90.

Tschopik H. 1963. Handbook of South American Indians. Vol.2: 501-573 The Aymara. Julian H. Steward, ed. New York: Cooper Square

Vona G, Falchi A, Moral P, Calò CM, Varesi L. Mitochondrial sequence variation in the Guahibo Amerindian population from Venezuela. *Am J Phys Anthropol,* 2005; 127(3):361-9.

Wallace DC. Maternal genes: mitochondrial diseases. *Birth Defects Orig Artic Ser*, 1987; 23(3):137-90.

Wallace DC, Torroni A. American Indian prehistory as written in the mitochondrial DNA: a review. *Hum Biol*, 1992; 64(3):403-16.

Wang S, Lewis CM, Jakobsson M, Ramachandran S, Ray N, Bedoya G, Rojas W, Parra MV, Molina JA, Gallo C, Mazzotti G, Poletti G, Hill K, Hurtado AM, Labuda D, Klitz W, Barrantes R, Bortolini MC, Salzano FM, Petzl-Erler ML, Tsuneto LT, Llop E, Rothhammer F, Excoffier L, Feldman MW, Rosenberg NA, Ruiz-Linares A. Genetic variation and population structure in native Americans. *PLoS Genet*, 2007; 3(11):e185.

Watkins WS, Rogers AR, Ostler CT, Wooding S, Bamshad MJ, Brassington AME, Carroll ML, Nguyen SV, Walker JA, Ravi Prasad BV, Reddy PG, Das PK, Batzer MA, Jorde LB. Genetic variation among world populations: Inferences from 100 Alu insertion polymorphisms. *Genome Res*, 2003; 13: 1607-1618.

Willard, C, Nguyen, HT, and Schmid, CW. Existence of at least three distinct Alu subfamilies. *J Mol Evol*, 1987; 26: 180–186.

Yang NN, Mazières S, Bravi C, Ray N, Wang S, Burley MW, Bedoya G, Rojas W, Parra MV, Molina JA, Gallo C, Poletti G, Hill K, Hurtado AM, Petzl-Erler ML, Tsuneto LT, Klitz W, Barrantes R, Llop E, Rothhammer F, Labuda D, Salzano FM, Bortolini MC, Excoffier L, Dugoujon JM, Ruiz-Linares A. Contrasting patterns of nuclear and mtDNA diversity in Native American populations. *Ann Hum Genet,* 2010; 74(6):525-38.

Zegura SL, Karafet TM, Zhivotovsky LA, Hammer MF. High-resolution SNPs and microsatellite haplotypes point to a single, recent entry of Native American Y chromosomes into the Americas. *Mol Biol Evol,* 2004; 21(1):164-75.

VII.I Appendix 1. Supplementary material of Gayà-Vidal et al., 2011

Table S1. mtDNA sequences from 16017 to 249 positions for the 189 Bolivians.

```
                 11111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111
                 66666666666666666666666666666666666666666666666666666666666666666666666666666666666666666666666666
H    Pops Hap    00000011111111111111111111111112222222222222222222222222222333333333333333333334444444444555555555                        11111111111111111111222222222222222
                 13568900122344444567777778888891122334444556677778888999999000112222444555568892345556667011245556 1133455555556666677778900000114455557889999000000111122344
      A  Q       7716632414660256868023469236892470339235868160148678902458914919045734846792140283456489462977011650646245667890345601399336789013601238563456034789024558529
              CRSTAAATTTCCTTTTCGACGCATCCCCAACCTCCTACACCTCCCACCCTGCCCTACCCCTACTATGCTTCACCCTTTTTGGTGTCGGATTGTTTCCTT-GATAGGTGTA-TTTTTC-GG-AGTAGGGAGCAGTCCTAAGCACTTAGTGTTATAAGGACA
 1    1     A2   ....C...T.........................T.........T........T.........A...........C....................-..........-.....T-..-G...........C...G...................G..
 2    1     A2   ........T...C...................C.T..................T.........A....T......C....................-..........-.....T-..-G...........C...G...................G..
 3    1     A2   ........T..C.......................T..................T.........A...........C....................-..........-.....T-..-G..G........C...G...................G..
 4    1     A2   ........T..........................T..................T.........A...........C....................-..........-.....T-..-G...........C...G...................G..
 5    1     A2   ........T..........................T..................T.........A...........C...............C....-..........-.....T-..-G...........C...G....T..............G..
 6    1     A2   ........T..................C.......T..................T.........A...........C....................-..........-.....T-..-G...........C.......................G..
 7       1  A2   ........T..........................T..................T.T.......A...........C...............C....-..........-.....T-..-G...........C...G...................G..
 8       1  A2   ........T..................C.......T.A.......T........T.........A...........C....................-..........-.....T-..-G...........C...G.........C.........G..
 9       1  A2   ........T..........................T..................T.........A...........C....................-..........-.....T-..-G...........C...G.......G...........G..
 10   1     A2   ........TC.........................T..................T........CA...........C....................-..........-.....T-..-G...........C...G................G..G..
 11      1  A2   ........T..........................T........T.........T........CA................................-..........-.....T-..-G...........C...G...................G..
 12      1  A2   ........T..........................T..................T.........A...........C....................-..........-.....T-..-G...........C...G........AC.........G..
 13      1  A2   ........T.......................C.T..................T.........A...........C...............C....-..........-.....T-..-G...........C...G..............C.G..G..
 14      1  A2   ........T..........................T..................T.........A...........C....................-..........-......-..-G...........C...G...................G..
 15      1  A2   ........T.......................C.T..................T.........A....T......C....................-..........-.....T-..-G...........C...G...................G..
 16      1  A2   ........T..........................T..................T.........A...........C....................-..........-.....T-..-G...........C.T.G...................G..
 17      1  A2   ........T.......................C.T..................T.........A....T......C....................-..........-.....T-..-G...........C...G...................GT.
 18      1  A2   ........T..........................T..................T.........A...........C..............C..TC.G..........-......-..-G...........C.....................................
 19      1  A2   ..G.....T..........................T..................T.........A...........C..............C..T.CG..........-......-..-G...........C...G.................................
 20      2  A2   ..G.....T.............................................T........CA...........C....................-..........-G....T-..-G...........C...G.A.................G..
 21   1     B2   ............C.............C.TC..C..........................................................C....-..........-......-..-.......................................
 22   4     B2   ........T.....A...........C.TC..C..........................................................C....-..........-......-..-G..................T...................
 23   1     B2   ..........................C.TC..C...........T..............................................C....-..........-......-..-G..................T...................
 24   1     B2   ..........................C.TC..C..............................................C...........C....-..........-......-..-G...A...............................
 25   1     B2   ..........................C.TC..C..........................................C...............C....-..........-......-..-G.C...................................
 26   1     B2   ..........................C.TC..C..........................................C...............C....-..........-......-..-G..............C.......................
 27   1     B2   ..........................C.TC..C..........................................................C....-..........-....CT-..-...........C.........................
 28   2  3  B2   ..........................C.TC..C..........................................................C....-..........-......-..-G......................................
 29   2     B2   ..........................C.TC..C..........................................................C....-..........-......-..-.......................................
 30   2     B2   ..........................C.TC..C....................................T.....................C....-..........-......-..-G..................T...................
 31   1     B2   ..........................C.TC..C..........................................................C....-..........-......-..-G..................T...................
 32   1     B2   ..........................C.TC..C................................................C.........C....-..........-......-..-G.....................C................
 33   1     B2   ..........................C.TC..C..................C...................T...................C....-..........-......-..-G..................T...................
 34   1     B2   ..........................C.TC..C......................................T...................C....-A.........-......-..-G...........C......T......C............
 35      2  B2   ..........................C.TC..C......................................T...................C....-..........-......-..-G...........C......T...................
 36   1  1  B2   ..........................C.TC..C..........................................................C....-..........-......-..-G..........A.......T...................
 37   1     B2   ..........................C.TC..C............................G.............................C....-..........-......-..-G..................T...................
 38   1     B2   ..G.......................C.TC..C..........................................................C....-..........-......-..-G..................T...................
 39      1  B2   ..........................C.TC..C..........................................................C....-..........-......-..-G...------.........T...................
 40      1  B2   ...................G......C.TC..C..........................................................C....-..........-......-..-G..................T...................
 41      1  B2   ..........................C.TC..C..........................................................C....-..........-......-..-G.............T........................
 42      1  B2   ..........................C.TC..C....................................................G.....C....-..........-....CT-A.-G......................................
 43      1  B2   ..........................C.TC..C..........................................................C....-..........-....CT-..-G......................................
 44      1  B2   ..........................C.TC..C....................................................G.....C....-..........-....CT-..-G......................................
 45      1  B2   ....C....................CC..C..C.......................................C..C...............C....-..........C.C....-..-G...............G......................
 46      1  B2   ....C....................CC..C..C.................................................TA.......C....-..........-......-..-G................................G.....
```

```
47     1  B2   .........................CC..C..C....T...........................C.........................C....-..........-......-..-G...........C....................G.....
48     2  B2   ..................T......CC..C..C..........................A..................A............C....-..........-......-..-G......................................
49     1  B2   ...G..............T......CC..C..C..........................................................C....-..........-......-..-G......................................
50     1  B2   ..................T......CC..C..C.................T........................................C....-.C........-......-..-G......................................
51     2  B2   .........................CC..C..C..........................................................C....-..........-..--..C..GG...T..................................
52     1  B2   .........................CC..C..C..........T...............................................C....-..........-......-.A-G......................................
53     2  B2   .........................CC..C..CG.........................................................C....-..........-......-..-G..G..................C.....C..........
54     1  B2   .........................CC..C..C..........................................................C....-..........-......-..-G...A......A...........................
55  5     B2   ...G.....................CC..C..C...............................................................-..........-.....T-..-G...........C....................G.....
56  2     B2   .........................CC..C..C............T..................................A...A......C....-..........-......-..-G......................................
57  1     B2   .........................CC..C..C..........................................................C....-..........-.....T-..-G...........C....................G.....
58  1     B2   .........................CC..C..C.......................................................A..C....-..........-......-..-G......................................
59  1     B2   .........................CC..C..C..........................................................C....-..........-......-..-G...........C....................G.....
60  1  1  B2   .........................CC..C..C..........................................................C....-..........-......-..-G......................................
61  1     B2   .........................CC..C..C..........................................................C....-..........-......-..-G............T.........................
62  2     B2   .........................CC..C..C..........................................................C....-..........-......-..-G..........AC....................G.....
63  2  1  B2   .......................T.CC..C..C..............................A...........................C....-..........-......-..-G......................................
64  1     B2   ....C............A.......CC..C..C..........................................................C....-..........-......-..-G...........C..........................
65  1     B2   .......T......A..........CC..C..C.....................................................C....C....-..........-......-..-G............................C.........
66  3     B2   C.........................C..CT.C......T...................................................C....-..........-......-..-G............T............C............
67  1     B2   C.........................C..CY.C......T...................................................C....-..........-......-..-G............T............C............
68  1     B2   C.........................C..C..C......T...................................................C....-..........-......-..-G............T............C............
69  1     B2   ..................T.......C..C..C.....C....................A...............................C....-.......A..-......-..-G.........................CA...........
70  1     B2   ..................T.......C..C..C.............................................A............C....-..........-......-..-G......................................
71  2     B2   ..................T.......C..C..C..........................................................C....-..........-......-..-G.....................C................
72     3  B2   ..................T.......C..C..C.......................T..................................C....-..........-......-..-G......................................
73     1  B2   .....C............T.......C..C..C...............................T..........................C....-..........-......-..-G......................................
74  2  1  B2   ..........................C..C..C...........................................C..............C....-..........-......-..-G......................................
75  1  1  B2   ..........................C..C..C..........................................................C....-..........-......-..-G...................G..................
76  1     B2   ..........................C..C..C.....................T..................................C.C....-..........-....CT-T.-G......................................
77  3     B2   ..........................C..C..C.....................T....................................C....-..........-....CT-T.-G.....................C................
78  1     B2   .....................T....C..C..C..........................................................C....-..........-......-..-G......................................
79  3     B2   ..........................C..C..C..........................................................C....-..........-......-..-G..........AC....................G.....
80  2     B2   ....C.....................C..C..C...........G..............................................C....-..........-......-..-G..........AC....................G.....
81  2     B2   ..........................C..C..C.......T..................................................C....-..........-......-..-G..........AC....................G.....
82  1     B2   ..........................C..C..C..........................................................C....-..........-......-..-G..........A...........................
83  1     B2   ........T......G..........C..C..C..........................................................C....-..........-......-..-G..........A...........................
84  4     B2   ..........................C..C..C.........G..............................C.................C....-..........-......-..-G...A......A...........................
85  1     B2   ..........................C..C..C.........G..............................C....A............C....-A.........-......-..-G...A......A...........................
86  1     B2   C.........................C..C..C.........................................C................C....-..........-......-..-G...A.......C.T.......C................
87  1     B2   .......................T..C..C..C..............A...............A...........C...............C....-........CG-......-..-G.....................C...C............
88     2  B2   .......................T..C..C..C..............................A...........................C....-..........-......-..-G......................................
89     1  B2   .......................T..C..C..C.........N....................A...........................C....-..........-......-..-G...A......AN..........................
90     1  B2   ..........................C..C..C..........................................................C....-..........-......-..-G.........................CA...........
91     1  B2   ..........................C..C..C...........................C..............................C....-..........-......-..-G......................................
92     1  B2   ..........................C..C..C.....................T....................................C....-..........-....CT-T.-G......................................
93     1  B2   ..........C...............C..C..C..........................................................C....-..........-......-T.-G...................................G..
94     1  B2   ..........................-..C..C..........................................................C....-..........-......-..-G.........................CA...........
95     1  B2   ...G......................-T.C..C..........................................................C....-......C...-......-..-G..........A..................G........
96     1  B2   ..........................-T.C..C..........................................................C....-......C...-......-..-G..........A..................G........
97     1  B2   ..........................C..C..C..........................................................C....-..........-....CT-..-G...........C....................G.....
98     1  B2   ........................T.C..C..C.......................T..................................C....-..........-......-..-G................G.....C...............
99     1  B2   ..........................C..C..C..........................................................C....-..........-......-..-G...A......A...........................
100    3  B2   ..........................C..C..C..........................................................C....-..........-....CT-..-G...........C..C.................G.....
101    1  B2   ..............A...........C..C..C..........................................................C....-..........-......-..-G............................C.........
102    2  B2   .....C....................C..C..C..........................................................C....-..........-.....T-..-G...........C....................G.....
103    3  B2   ..........................C..C..C..........................................................C....-..........-......-..-G......................................
104    1  B2   ..........................C..C..C.........G..............................C.................C....-......C...-......-..-G...A......AC..........................
105    1  B2   ..........................C..C..C..........................................................C....-..A.AC....C.C....-..-G...........C....................G.....
106 1     B2   ..........................C..C..C................T..G......................................C....-..........-......-..-G..........A...........................
```

107	1		B2	..........................C..C..C...C....-..........-......-..-G...........C....................G.....
108	1		B2	..........................C..C.TC.............C...C....-..........-......-..-G......................................
109	1		B2	..........................C..C..C...C...............C....-..........-......-..-G......................................
110		1	C1	..........................C..C....T......................C........CT.T...........................-..........-......-..-G.................A...................-
111		1	C1	..G..C........CT.............................-..........-....CT-T.-G............T........................-
112	1		C1	..G...............................T......................C........CT.............................-..........-......-..-G.....................................-
113		4	C1	.G................................T......................C........CT.............................-..........-......-..-G.....................................-
114		1	C1	.G................................T......................C........CT.............................-..........-......-..-G.........................C...........-
115	1		C1	..................................T......................CG.......CT.............................-..........-......-..-G.....................................-
116		1	C1	..................................T......................C........CT........................C....-..CW......-......-..-G...........C.........C...............-
117		3	C1	..................................T......................C........CT........................C....-..C.......-......-..-G...........C.........C...............-
118		1	C1	..................................T......................C........CT........................C....-..C..A....-......-..-G...........C.........C...............-
119		1	C1	..................................T......................C........CT........................C....-..........-......-..-G...........C.........................-
120		2	C1	....................C.............T......T...............C........CT........................C....-..........-......-..-G.....................................-
121	1		C1	......C...........................T......T...............C........CT.................A.CC...C....-..........-......-..-G.....................C...............-
122		1	C1	..............A...............T...TC.....................C........CT..T.....................C....-..........-......-..-G...............................G.T...-
123	2		C1	..................................T......................C........CT.........A...................-..........-......-..-G.....................................-
124	1	2	C1	..................................T......................C........CT.............................-..........-......-..-GT....................................-
125		1	D1	.............T..T.................T.............T.............C.T.C.....C..C................C....-..........-......-..-G...........C......................A...
126		1	D1	..................................T...................T...........C........C................CT...-..........-......-..-G......................................
127		1	D1	..................................T...........C...................C........C................CT...-..........-......-..GG..G...................................
128		1	D1	..................................T..T........................C...C........C.....................-..........-......-..-G..............C.......................
129	2		D1	......................T...........T...............................C........C................C....-..........-......-..-G......................................
130	3		D1	..................................T...............................C........C................CT...-..........-......-..-G......................................

H, Haplotype; A, Aymara population; Q, Quechua population; Hap, Haplogroup; CRS, Cambridge Reference Sequence (Anderson et al., 1981).

Table S2. Y_chromosome STR haplotypes in the Aymaras and Quechuas from Bolivia.

	n			DYS	DYS				DYS					Y_GAT					
Haplotype	**A**	**Q**	**Haplogroup[a]**	**389I**	**389II**	**DYS390**	**DYS456**	**DYS19**	**385a-b**	**DYS458**	**DYS437**	**DYS438**	**DYS448**	**H4**	**DYS391**	**DYS392**	**DYS393**	**DYS439**	**DYS635**
1	1		(R1b)	15	32	24	16	14	11-14	17	14	12	18	11	10	13	13	12	23
2		1	Q1a3a	14	32	23	15	13	15-19	17	14	11	20	12	10	16	14	13	22
3		1	Q1a3a	14	32	23	15	13	15-19	16	14	11	20	14	10	17	14	14	22
4	2		Q1a3a	14	32	23	15	13	15-19	16	14	11	20	12	10	16	14	14	22
5		1	Q1a3a	14	32	23	15	13	15-18	16	14	11	20	12	10	16	14	13	22
6		1	Q1a3a	14	32	21	15	13	14-18	17	14	11	21	12	10	14	13	13	22
7		1	Q1a3a	14	31	25	15	13	14-20	18	14	11	19	12	10	14	13	11	22
8		1	(R1b)	14	31	24	17	14	11-14	17	14	12	18	11	11	13	13	11	23
9	1		Q1a3a	14	31	24	15	13	16-17	18	14	12	20	12	11	14	13	11	22
10	1		Q1a3*	14	31	24	15	13	14-14	18	14	11	20	12	10	14	13	13	23
11	1		Q1a3a	14	31	24	15	13	14-14	17	14	11	20	12	10	14	13	13	23
12	1		Q1a3*	14	31	24	15	13	14-14	17	14	11	20	11	10	14	13	13	23
13		1	Q1a3a	14	31	23	16	13	15-20	16	14	11	19	12	10	16	14	14	22
14	1		Q1a3a	14	31	23	16	13	15-17	17	14	11	20	12	10	16	14	13	22
15	1		Q1a3*	14	31	23	15	13	16-19	17	14	11	20	12	10	16	14	13	22
16	2		Q1a3a	14	31	23	15	13	16-19	16	14	11	20	12	10	16	14	13	22
17	1		Q1a3a	14	31	23	15	13	15-20	16	14	11	20	12	10	16	14	13	22
18		1	Q1a3a	14	31	23	15	13	15-19	17	14	11	20	13	10	18	14	14	22
19		5	Q1a3a	14	31	23	15	13	15-19	17	14	11	20	12	10	18	14	14	22
20	1		Q1a3a	14	31	23	15	13	15-19	16	14	11	21	11	10	16	14	12	22
21	1		Q1a3a	14	31	23	15	13	15-19	16	14	11	20	13	10	16	14	14	22
22	3	1	Q1a3a	14	31	23	15	13	15-19	16	14	11	20	12	10	16	14	13	22
23		1	Q1a3a	14	31	23	15	13	15-19	16	14	11	20	12	10	16	14	12	22
24	1		Q1a3a	14	31	23	15	13	15-19	15	14	11	20	12	10	16	14	14	22
25	1		Q1a3a	14	31	23	15	13	15-19	15	14	11	20	12	10	16	14	13	22
26	1		Q1a3a	14	31	23	15	13	15-18	17	14	11	20	12	10	16	14	12	22
27		1	Q1a3a	14	31	23	15	13	15-18	16	14	11	20	12	10	16	14	14	22
28	1		Q1a3a	14	31	23	15	13	15-18	16	14	11	20	12	10	16	14	13	23
29		3	Q1a3a	14	31	23	15	13	15-18	16	14	11	20	12	10	16	14	13	22
30		2	Q1a3a	14	31	23	15	13	15-18	16	14	11	20	12	10	16	14	12	22
31		1	Q1a3a	14	31	23	15	13	15-18	16	14	11	20	12	10	14	14	14	23

32		1	Q1a3a	14	31	23	15	13	15-18	16	14	11	20	11	10	15	14	13	22
33	2		Q1a3a/Q1a3*	14	31	23	15	13	15-17	16	14	11	20	12	10	16	14	14	22
34		2	Q1a3a	14	31	23	15	13	14-19	16	14	11	20	12	10	16	14	13	22
35		1	Q1a3a	14	31	23	15	13	14-19	16	14	11	20	12	10	16	14	12	22
36	1		Q1a3a	14	31	23	15	13	14-18	16	14	11	20	12	10	17	14	13	22
37	1		Q1a3a	14	31	23	15	13	14-14	19	14	11	21	12	10	15	13	12	23
38	1		Q1a3a	14	30	25	17	13	15-19	17	14	11	19	11	11	15	13	12	26
39		1	Q1a3a	14	30	25	15	13	14-19	17	14	9	20	12	10	14	13	13	22
40		1	(R1b)	14	30	24	15	15	11-14	18	15	12	19	12	10	13	12	12	23
41		1	(R1b)	14	30	23	17	14	11-11	18	15	12	19	12	10	13	14	13	23
42	1		Q1a3a	14	30	23	15	14	15-18	16	14	11	20	12	10	16	14	13	22
43	3		Q1a3a	14	30	21	13	13	14-16	17	14	11	20	11	11	14	14	12	22
44		1	Q1a3a	14	29	25	15	13	14-20	18	14	11	19	12	10	14	13	11	22
45	1		Q1a3a	13	32	24	15	13	17-17	13	14	11	21	11	10	14	14	11	22
46	1		Q1a3a	13	32	23	16	13	15-18	16	14	11	20	11	10	14	13	11	22
47		2	Q1a3a	13	31	24	16	13	16-18	15	14	11	19	11	11	14	13	12	22
48	2		Q1a3a	13	31	24	15	14	14-17	18	14	11	20	12	11	14	13	11	22
49	1		Q1a3a	13	31	24	15	13	14-17	17	14	11	19	12	11	14	13	11	22
50	1		Q1a3a	13	31	23	15	13	15-19	16	14	11	21	11	10	14	13	11	22
51	2		Q1a3a	13	30	25	15	14	16-17	13	14	11	20	12	10	14	13	13	23
52	1		Q1a3a	13	30	25	15	12	14-20	17	14	11	19	11	10	14	13	11	23
53	1		(E1b1b)	13	30	24	16	13	17-17	15	14	10	20	12	10	11	13	13	23
54	1		Q1a3a	13	30	24	15	14	15-20	16	14	10	19	12	10	14	13	12	22
55		2	Q1a3a	13	30	24	15	13	14-15	17	14	12	20	12	11	14	13	13	22
56		1	Q1a3a	13	30	24	15	13	14-14	18	14	12	20	12	10	14	13	13	22
57	2		Q1a3a	13	30	24	15	13	14-14	18	14	11	20	12	10	14	13	14	22
58	1		Q1a3a	13	30	24	14	13	15-17	17	14	11	19	12	10	14	13	12	22
59		1	(T)	13	30	23	15	15	16-17	16	14	9	19	11	11	15	13	11	21
60	1		Q1a3a	13	30	23	15	14	12-14	19	14	11	20	11	10	14	13	11	22
61	1		Q1a3a	13	30	23	15	13	15-19	16	14	11	20	12	10	16	14	13	22
62		1	Q1a3a	13	30	23	15	13	15-18	16	14	11	20	12	10	16	14	13	22
63		1	Q1a3a	13	30	23	15	13	15-17	17	14	11	20	12	10	16	14	13	22
64	4		Q1a3a	13	30	23	15	13	13-19	16	14	11	20	12	10	16	14	14	22
65	2		Q1a3a	13	30	23	15	13	13-18	16	14	11	20	12	10	16	14	13	22
66	1		Q1a3a	13	30	23	11	13	13-19	16	14	11	20	12	10	16	13	14	22

67	1		Q1a3*	13	29	25	16	13	15-19	16	14	11	19	11	9	14	13	12	22
68	2		Q1a3a	13	29	25	15	14	14-17	17	14	10	21	12	10	14	13	12	22
69		1	(R1b)	13	29	26	15	14	11-11	16	15	12	19	12	10	13	13	12	23
70		1	(R1b)	13	29	24	16	14	11-14	16	15	11	19	11	10	13	13	12	23
71	1		Q1a3a	13	29	24	16	13	14-16	16	14	11	21	12	12	14	13	12	22
72		1	Q1a3a	13	29	24	15	13	14-17	17	14	11	19	12	11	14	13	11	22
73	1		Q1a3*	13	29	24	15	13	11-16	18	14	11	20	12	10	14	13	13	23
74		1	(J1)	13	29	23	15	14	14-17	17	14	10	20	11	10	11	12	13	20
75		2	(R1b)	13	29	23	15	14	11-11	17	14	12	18	11	11	13	13	13	23
76		1	Q1a3a	13	28	24	15	13	15-18	17	14	11	20	12	11	14	13	11	22
77		1	Q1a3a	13	28	24	15	13	15-17	17	14	11	20	12	11	14	13	12	22
78		1	Q1a3a	13	28	24	15	13	14-17	17	14	11	20	12	11	14	13	11	22
79		1	Q1a3a	12	30	24	15	13	16-18	17	14	11	21	11	10	14	13	12	22
80		1	Q1a3a	12	30	24	14	15	14-16	17	14	11	19	12	10	14	13	12	24
81		1	Q1a3a	12	30	24	14	14	14-16	17	14	11	19	12	10	14	13	12	24
82		1	(E1b1a)	12	30	21	15	17	16-17	18	14	11	21	12	10	11	13	12	22
83		1	Q1a3a	12	29	23	15	13	16-19	16	14	11	20	11	10	16	14	13	22
84		1	Q1a3a	12	28	24	15	13	14-18	16	14	11	21	12	11	14	13	13	22
85		1	(J)	12	28	23	15	14	13-17	17	14	9	21	11	10	11	12	10	22
86		1	(I1)	12	28	23	15	14	14-14	16	16	10	20	12	10	11	13	11	22
87	1		Q1a3a	12	28	21	16	13	14-19	18	14	11	20	12	10	14	14	12	22

[a] The haplogroups have been assigned by SNP genotyping, and those in parentheses have been inferred from the web page Haplogroup Predictor (http://www.hprg.com/hapest5/).

Résumé

Deux populations appartenant aux groupes linguistiques principaux de la Bolivie, Aymaras et Quechuas, ont été étudiées par différent marqueurs génétiques pour fournir information sur leurs relations génétiques et processus démographiques qui pourraient avoir souffert pendant leur histoire. Ce travail comprend trois parties: l'étude i) de marqueurs génétiques autosomiques (insertions Alu), ii) uniparentaux, l'ADN mitochondrial (ADNmt) et le chromosome Y, et iii) d'une région du chromosome 19 avec le *gene cluste*r des apolipoproteins APOE/C1/C4/C2.

Dans le premier travail, trente-deux insertions Alu polymorphiques (PAIs), 18 autosomiques et 14 du chromosome X, ont été étudiées. L'objectif principal de l'étude était d'aborder les relations génétiques entre ces deux populations et d'éclaircir d'après ces données génétiques si l'expansion de la langue Quechua dans la Bolivie pouvait être attribuée à des processus démographiques (migrations Incas de parlants Quechua de Pérou vers la Bolivie) ou culturel (imposition de la langue Quechua par les Incas). La relation génétique très proche observée entre les deux populations boliviennes ainsi que leur différentiation des Quechuas du Pérou suggère que l'expansion de la langue Quechua dans la Bolivie eu lieu sans une contribution démographique importante.

La deuxième partie concernant a été réalisé pour évaluer les possibles différences dépendant du genre et fournir plus de données pour éclaircir les processus démographiques de la région andine. Dans ce cas, les deux populations Boliviennes ont montré plus de différences génétiques pour l'ADNmt que pour le chromosome Y. Concernant l'ADNmt, les Aymaras semblent avoir été plus isolés au cours de leur histoire, fait qui aurait entrainé la conservation de certaines caractéristiques génétiques, tandis que les Quechuas aurait été plus perméables à l'incorporation des femmes étrangères et à l'influence péruvienne. Néanmoins, la mobilité des homes aurait été généralisée dans toute la région andine d'après l'homogénéité trouvée dans cette zone.

L'étude d'une région autosomique d'environ 108kb incluant le groupe de gènes APOE/C1/C4/C2 et les régions adjacentes, dans laquelle, vingt-cinq polymorphismes (10 STRs et 15 SNPs) ont été analysés pour éclaircir l'histoire évolutive de cette région génomique dans les populations andines. Une partie de cette diversité réduite pourrait être attribuée à l'effet de la sélection qui pourrait être due à son importance physiologique, mais aussi du à leur histoires démographiques.

Abstract

Two populations belonging to the two main Native linguistic groups of Bolivia, Aymaras and Quechuas, have been analysed for different genetic markers in order to provide relevant information about their genetic relationships and demographic processes. This work comprises three parts: the study of i) autosomal markers (*Alu* insertions), ii) uniparental markers, both mtDNA and Y-chromosome, and iii) a region including the APOE/C1/C4/C2 gene cluster that code for apolipoproteins that can have epidemiological implications.

In the first part, thirty-two polymorphic Alu insertions (18 autosomal and 14 from the X chromosome) were studied. The main objective was to determine from genetic data whether the expansion of the Quechua language into Bolivia could be associated with demographic (Inca migration of Quechua-speakers from Peru into Bolivia) or cultural (language imposition by the Inca Empire) processes. Our results indicated that the two Bolivian samples showed a high genetic similarity for both sets of markers and were clearly differentiated from the two Peruvian Quechua samples available in the literature. Additionally, our data were compared with the available literature to determine the genetic and linguistic structure, and East–West differentiation in South America. The close genetic relationship between the two Bolivian samples and their differentiation from the Quechua-speakers from Peru suggested that the Quechua language expansion in Bolivia took place without any important demographic contribution.

The second part, mtDNA and Y-chromosome uniparental markers were studied to evaluate sex-specific differences and give new insights into the demographic processes of the Andean region. In that case, the two Bolivian samples showed more genetic differences for the mtDNA than for the Y-chromosome. For the mtDNA, 81% of Aymaras and 61% of Quechuas presented haplogroup B2. Native American Y-chromosomes were found in 97% of Aymaras (89% hg Q1a3a and 11% hgQ1a3*) and 78% of Quechuas (100% hg Q1a3a). Our data revealed high diversity values in the two populations, in agreement with other Andean studies. The comparisons with the available literature for both sets of markers indicated that the central Andean area is relatively homogeneous. For mtDNA, the Aymaras seemed to have been more isolated throughout time, maintaining their genetic characteristics, while the Quechuas have

been more permeable to the incorporation of female foreigners and Peruvian influences. On the other hand, male mobility would have been widespread across the Andean region according to the homogeneity found in the area. Particular genetic characteristics presented by both samples support a past common origin of the Altiplano populations in the ancient Aymara territory, with independent, although related histories, with Peruvian (Quechuas) populations.

The study of the autosomal region of 108kb, including the APOE/C1/C4/C2 gene cluster and the flanking region in which twenty five polymorphisms (10 STRs and 15 SNPs) were analysed to give new insights into the evolutionary history of this genomic region in Andean populations. In general, diversity in Bolivians was low, with nine out of 15 SNPs and seven out of 10 STRs being practically monomorphic. Part of this reduced diversity could be attributed to selection since the APOE/C1/C4/C2 region presented a high degree of conservation compared to the flanking genes in both Bolivians and Europeans, which may be due to its physiological importance. Also, the lower genetic diversity in Bolivians compared to Europeans for some markers could be attributed to their different demographic histories.