Ancient DNA from pre-Columbian South America

Guido Marcelo Valverde Garnica

Australian Centre for Ancient DNA
Department of Genetics and Evolution
School of Biological Sciences
The University of Adelaide

A thesis submitted for the degree of Doctor of Philosophy at
The University of Adelaide

October 2015

TABLE OF CONTENTS

Thesis abstract

Ancient DNA (aDNA) research in the Americas represents a genetic strategy to investigate demographic and historical events of populations in the continent with the added bonus of having a direct and temporal perspective. This thesis aims to explore human mitochondrial DNA (mtDNA) diversity from a large number of pre-Columbian samples in a diachronic transect through time in order to refine our understanding of the genetic structure and diversity of ancient civilizations in the Central Andes of South America prior to the arrival of Europeans.

I used a combination of traditional PCR-based methods and the latest technological advances for DNA analysis (i.e. Next-Generation Sequencing – NGS) to generate high-resolution mtDNA data to explore the past genetic diversity of South American populations.

This work shows the perspective of aDNA research to identify temporal transitions in the genetic composition in the Central Andes of South America in real-time, since I aimed to incorporate samples from all cultural archaeological periods, improving the spatial and geographic coverage.

By comparing the results with genetic data from modern-day native populations, this thesis will also address the potential impact of the European colonization on indigenous populations to understand the evolutionary history of Native Americans. To that end, the acquisition of high-resolution genomic data from ancient specimens showcase the true potential of aDNA research to uncover (re-discover) lost genetic diversity or lost mtDNA lineages from pre-Columbian populations, which cannot be inferred from modern-day populations.

I aim to provide an accurate description of patterns of genetic diversity through time, reconciling and contrasting the genetic data with valuable archaeological information, and to test for demographic and population continuity or discontinuity in pre-contact South America. Finally, this thesis adds important perspective to the existing knowledge about mtDNA diversity and population prehistory in the Central Andes.

Research Aims of the project

a) To explore novel ancient mitochondrial DNA data from Native Americans from the Central Andes of South America across several archaeological periods in order to contrast this information with available HVR-I data from present-day and ancient populations to better understand the overall pre-Columbian mtDNA genetic diversity.

b) To generate complete mitochondrial DNA genome data at highest level of resolution from selected ancient samples from South America to characterize and explore the potential of mitochondrial genomes to unveil mtDNA genetic diversity in South America before the European arrival.

c) To reconcile genetic and archaeological information in the light of temporal sampling to reconstruct the population history of pre-Columbian ancient groups in restricted geographic locations in South America.

d) To combine the advantages of traditional and novel methods for aDNA analysis in order to develop and establish a new technique for genotyping ancient specimens exclusively from pre-contact Americas.

Thesis structure

This PhD thesis is written as a combination of chapters in publication format and a published paper. The content of the chapters and relevant supplemental materials are as follows:

Chapter 1: General Introduction

a. General description of the state-of-the art and current knowledge about population history in pre-Columbian Americas.
b. General introduction to ancient DNA field, applications and limitations.
c. General introduction to Next-Generation Sequencing technologies and applications in aDNA research.

Chapter 2: Ancient DNA from pre-Columbian populations in the Central Andes of South America: a diachronic study of mtDNA haplogroup diversity based on Hypervariable Region I

Exploration of the mtDNA genetic diversity with a large sample size in a broad geographical and chronological range, contrasting ancient and modern diversity in populations from South America based on analysis of the hypervariable region-I.

Chapter 3: High-resolution mitochondrial genome sequencing reveals a substantial loss of Pre-Columbian diversity

Application of Next-Generation Sequencing techniques to ancient samples from South America in order to retrieve mtDNA complete genomes by applying a temporal sampling through archaeological periods in the Central Andes of South America.

Exploration of mtDNA diversity in pre-Columbian Americas under a great level of genetic resolution.

Chapter 4: Ancient DNA analysis from the 'Huaca Pucllana' archaeological site in Central coastal Peru: Chronological study of mitochondrial DNA variation in the context of the Wari Empire expansion during the Middle Horizon

Exploration of changes in genetic diversity and micro-evolutionary processes in a transect through time in a restricted geographic area, i.e. coastal Peruvian populations, driven by the impact of pre-Columbian colonization led by the Wari Empire.

Chapter 5: AmericaPlex26 – A SNaPshot multiplex system for genotyping the main human mitochondrial founder lineages of the Americas (co-author)

Development of a laboratory analytical technique to perform an exhaustive screening of archaeological samples in the Americas for genotyping purposes and sample selection for Next-Generation Sequencing.

Published paper: Coutinho A, **Valverde G**, Fehren-Schmitz L, Cooper A, Barreto Romero MI, Espinoza IF, Llamas B, and Haak W. 2014. AmericaPlex26: a SNaPshot multiplex system for genotyping the main human mitochondrial founder lineages of the Americas. PLoS ONE 9(3):e93292.

Chapter 6: General Discussion

Summary and conclusion of the overall study.

Limitations of research, future perspectives and assessment for potential aDNA follow-up studies in South America.

Thesis declaration

This work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution to Guido Valverde and, to the best of my knowledge and belief, contains no material previously published or written by any other person, except where due reference has been made in the text.

I give consent to this copy of my thesis when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

The author acknowledges that copyright of published works contained within this thesis (as listed below) resides with the copyright holders(s) of those works.
I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library catalogue, the Australian Digital Theses Program (ADTP) and also through web research engines, unless permission has been granted by the University to restrict access for a period of time.

The published works of my thesis include:
Coutinho A, Valverde G, Fehren-Schmitz L, Cooper A, Barreto Romero MI, Espinoza IF, Llamas B, and Haak W. 2014. AmericaPlex26: a SNaPshot multiplex system for genotyping the main human mitochondrial founder lineages of the Americas. PLoS ONE 9(3):e93292.

Guido Valverde
October, 2015

Acknowledgements

The present work has been made possible with the financial support of the Australian Research Council funding (ARC), Discovery Project "DP1095782".
I would like to thank the School of Biological Sciences at the University of Adelaide for supporting my Candidature and the Australian Centre for Ancient DNA – ACAD, which provided the logistics to perform the project.

I would like to show my gratitude and admiration to my thesis supervisors Wolfgang Haak and Bastien Llamas for their understanding and patience guidance and encouragements during this time.
I am grateful to Alan Cooper for giving me the opportunity to work in this fascinating research field. I feel privileged to have been granted access to ancient DNA laboratories at ACAD and perform the laboratory analysis under the most impressive working conditions for ancient DNA research.
I would like to thank Lars Fehren-Schmitz for valuable advice and collaborative work.
I would like to thank Michael Herrera for valuable assistance and collaboration.
I would like to recognize all people and lab members at the Australian Centre for Ancient DNA for their intellectual, technical support and friendship.

I would like to thank the archaeologists at the Huaca Pucllana archaeological site in Peru; the Department of Humanities and Archaeology, Pontificia Universidad Católica del Perú; Centro de Investigaciones Arqueológicas del Museo de Ancón, Perú; Unidad de Arqueología y Museos (UDAM), Ministerio de Culturas de Bolivia; Instituto de Alta Investigación (IAI), Laboratorio de Arqueología y Paleoambiente, Universidad de Tarapacá, Arica, Chile, for providing the valuable samples for analysis and for relevant advise and archaeological information.

I would like to address a very special thanks to Ana and Thomas Ryan for making me feel like home and support me during my stay in Adelaide as my Australian family.
I owe my deepest gratitude to my family and my parents Guido and Mery and my brother Alvaro for their constant support and encouragements despite the distance and during the last stage of completion of the thesis.

CHAPTER 1

General Introduction

Ancient DNA from pre-Columbian South America

A) Population history of pre-Columbian Americas

The study of pre-Columbian Americas represents a fascinating and extensive research field. Various disciplines such as archaeology, linguistics, morphology and genetics have provided substantial insights into the history of the continent from different perspectives to refine our understanding of the complexity of societies, cultural development and structure of Native American populations before the European contact (Greenberg et al. 1986; Crawford 1998; Mulligan et al. 2004; Mann 2005; Livi-Bacci 2006; Tamm et al. 2007; Goebel et al. 2008; Kitchen et al. 2008; Dillehay 2009; Meltzer 2009; O'Rourke and Raff 2010; O'Fallon and Fehren-Schmitz 2011; Raff et al. 2011; Chatters et al. 2014; Rasmussen et al. 2014; Rasmussen et al. 2015a).

The American continents were the last ones to be colonized by anatomically modern humans. Two major demographic events seem to have played a central role in the settlement of the Americas and the population history of indigenous Americans: first, the initial migration of Native American ancestors into the continent through Beringia and second, the arrival of Europeans in the New World during the 15th century AD.

The recent coupling of archaeology and molecular anthropology in a "collaborative agreement" (Renfrew 2010), has provided the most significant contributions to our understanding about human migrations and dispersals across the world. Particularly in the Americas, the field of genetics has provided the key assumptions to explain long-standing questions regarding ancestry and relationships of the first human groups that settled and spread throughout the continent (Schurr et al. 1990; Torroni et al. 1992; Horai et al. 1993; Torroni et al. 1993; Torroni et al. 1994b; Merriwether et al. 1995; Bonatto and Salzano 1997; Silva et al. 2002; Rasmussen et al. 2010; Reich et al. 2012; Raghavan et al. 2014a; Raghavan et al. 2014b; Rasmussen et al. 2014; Raghavan et al. 2015; Rasmussen et al. 2015a). Nevertheless, explaining the human diaspora across the world still requires a multidisciplinary approach and support from other disciplines to generate a coherent synthesis of the many diverse processes involved.

Here, I present a summary of the current knowledge about the population history of pre-Columbian Americas from archaeological, linguistic, morphological and genetic evidence.

Archaeological evidence

The precise timing, colonization routes and number of migration followed by the first human groups that settled the Americas is one of the most debated topics in archaeology (Goebel et al. 2008). During the Last Glacial Maximum (LGM) ~21,000 years before present (YBP) (Clark et al. 2009), Asia and America were connected by a land bridge known as Beringia that corresponds what is now the Bering Strait, since the level of the sea was lower than it is today (Figure 1). However, the North American territories were covered by two glaciers; the Laurentide and the Cordilleran ice sheets, which would have formed an impenetrable barrier for human groups on their way towards the American continents (Meltzer 2009).

Human colonization of the Americas and the paths that the first settlers may have followed were influenced by environmental and climatic effects during the last glaciations period in the Last Pleistocene (Meltzer 1993; Mandryk et al. 2001). The archaeological evidence led to suggest hypotheses for the peopling of the Americas differing in terms of the timing (human incursions around the LGM period) and migration routes (i.e. Pacific coastal route vs. ice-free corridor (Mandryk et al. 2001)).

The two proposed entry routes into the Americas therefore are dependent on environmental conditions, chronology and isolation events of human groups before they spread from Beringia towards the Americas. The Pacific coastal route (Fladmark 1979) might have become available after deglaciation affected the Cordilleran ice sheet around 15,000 YBP (Taylor et al. 2014). Alternatively, the inland ice-free corridor (Dixon 1993) might have become passable for human explorers when both ice sheets began to separate on the eastern side of the Rocky Mountains ~13,500 YBP (Mandryk et al. 2001).

The Pacific coastal route of human colonization of the Americas implies access to marine resources, such as sea mammals, shellfish, etc. (Erlandson et al. 2007), mammal hunting and the ability to built transportations means (e.g. watercraft to traverse the edge of the Beringian landmass around the ice sheets until reaching unglaciated areas to the south – reviewed in (Schurr 2008)). In contrast, the passage of human groups through the ice-free corridor also required the accessibility and availability of food sources (e.g. American megafauna).

Archaeological evidence in the Americas, particularly in North America, covers widely dispersed habitation sites, which indicates early dispersal within the continent and different types of adaptation of the first populations to new environments in the continent, which laid the foundation for subsequent cultural developments (Goebel et al. 2008).

The earliest archaeological discoveries in the early 1930's became known as the Clovis cultural complex (Hrdlicka 1937). The Clovis culture, named after the eponymous archaeological site in New Mexico (USA), dates back to 13,500 – 12,000 YBP (Waters and Stafford 2007) and features a diverse set of characteristic stone tools constructed from a variety of stone types including stone projectiles known as the "Clovis points". These fluted projectiles and an accompanying array of lithic artefacts were associated with the hunting of the American megafauna – mammoth and mastodon (Howard 1933) as one of their main subsistence resources.

The discovery of several contemporaneous sites corresponding to the Clovis culture found throughout the United States led to propose the "Clovis First" Model, which refers to a late peopling of the Americas (i.e. after the LGM). Although the Clovis cultural complex was a widely distributed archaeological horizon in North America, it was soon replaced by other groups. The Folson tradition (12,900 – 12,000 YBP) for instance, was characterized by a different design of lithic artefacts and the sharp points mainly associated with the hunting of small mammals such as bison. Nevertheless, according to some authors the Clovis people might have been responsible for hunting at massive scale, which led to a sudden demise of mammal populations. This extinction model was termed the "Blitzkrieg" (rapid overkill) hypothesis of American colonization (Martin 1984).

The subsequent discoveries of other sites in North America, which are seemingly older than the Clovis culture, challenge the Clovis First Model. Reports of older archaeological sites such as Paisley Cave (14,600 YBP), La Sena/Lovewell (22,000 – 19,000 YBP), Meadowcroft (22,000 – 18,000 YBP) and Cactus Hill (20,000 – 18,000 YBP) – reviewed in (Goebel et al. 2008), suggest an even earlier human occupation of the continent. This implies that people of the Clovis complex were not the earliest migrants in North America, and this earlier settlement might otherwise represent an expansion of a successful cultural adaptation that developed among early settlers around the North American ice sheets of the LGM (O'Rourke and Raff 2010).

However, archaeological evidence from these sites is scarce, and rest on bones harbouring marks that could have potentially been made by humans. Since no other characteristic artefacts were found, this casts doubts on the actual early presence of humans in North America.

Monte Verde in South America (Puerto Montt, Southern Chile) represents another interesting pre-Clovis settlement, dated at 14,600 YBP (Dillehay 1997; Meltzer 1997; Dillehay et al. 2008). The remains found at Monte Verde are very rich and well preserved and include stone tools and organic materials such as bone, ivory and plants.

These data suggest a coastal type of livelihood relying on marine resources from the Pacific. Despite questions about how and when people reached the southern cone of South America without leaving much other evidence in the Americas, the presence of human remains more than 14,600 years-old at Monte Verde offers important clues to the peopling of the Americas (Dillehay et al. 2008). Indeed, this pre-Clovis settlement lends support to the hypothesis of a Pacific coastal route scenario, which opened ~15,000 YBP, as opposed to the inland ice-free corridor, which was not passable until 13,500 YBP.

Monte Verde has implications on the timing of the spread as it needs to explain the coverage of almost 10,000 miles from Alaska to Chile in a relatively short time period. Moreover, the latitudinal crossing from humans adapted to the Arctic in North America passing through tropical regions of the continent is one of the question that remains opened – see (Jobling et al. 2013).

A hypothesis put forward to explain this migration scenario is the so-called “Kelp highway”, which proposes that early maritime peoples were able to colonize the south part of the Americas based on a maritime ecology evidence along the Pacific coast (Erlandson et al. 2007). Overall, the sum of pre-Clovis sites nonetheless suggests that the first human incursions in the Americas dated back to shortly after the LGM (Goebel et al. 2008).

Figure 1. Map showing the main physical characteristics of the American continent during the Last Glacial Maximum (LGM), with a selection of the prehistoric sites mentioned in the text. After (Marangoni et al. 2013).

Linguistic evidence

The first multidisciplinary model to address the settlement of the Americas incorporating linguistic, dental and genetic evidence dates back to the mid 1980's. Upon comparison of languages spoken by Native American groups, Greenberg and collaborators proposed the "Tree-wave migration model" describing three language groups that reached the Americas at different times (Greenberg et al. 1986) (Figure 2).

The first wave gave rise to a large "Amerind" language family comprising population of South and Central America and parts of North America with a proposed entry time into the Americas around 11,000 YBP. A second migration wave comprising Na-Dene speakers (restricted to North America) occurred later, around 9,000 YBP. Finally, populations speaking Eskimo-Aleutian languages (restricted to the Arctic) migrated to the Americas around 4,000 YBP (Greenberg et al. 1986).

This model implies that the peopling of America involved a minimum of three independent and separated migration episodes during distinct pre-historical periods. However, it has been strongly challenged by physical anthropologists (i.e. 'see below') that proposed two waves of migration (Neves and Pucciarelli 1991), as well as by more recent models based on genetic data that support a single wave of migration (Merriwether et al. 1995; Bonatto and Salzano 1997; Silva et al. 2002). Moreover, the majority of linguists have criticized this model as broadly used in genetics (Bolnick et al. 2004).

Figure 2. Three-wave migration model. Native Americans descended from three groups of migrants belonging to distinct linguistic families: Amerindians, Na-denes and Eskimo-Aleutians (Marangoni et al. 2013).

Evidence from cranial morphology

Physical anthropology, and here specifically craniometrical variation, also provided important insights into the complexity of early Native American populations and proposed models describing the colonization of the continent (Powell and Neves 1999; Gonzalez-Jose et al. 2001; Jantz and Owsley 2001; Neves and Hubbe 2005). Significant contributions were made by Neves and colleagues in the late 1990's based on the study of human skeletal remains from the Lagoa Santa, Sumidouro Cave, Lapa Vermehla IV archaeological sites in Brazil (Neves and Pucciarelli 1991).

The comparative analysis of craniofacial morphology led to the proposal of two distinct morphological types, termed "Paleoamerican" and "Amerindian". This model became known as the "two main biological components / stock hypothesis" (Neves and Pucciarelli 1991; Neves and Hubbe 2005). The cranial morphology of Paleoamericans suggests a close affinity with Australo-Melanesian groups, whereas Amerindians show similarity with Northern Asian populations. These observations were interpreted as two distinct ancestral populations which colonized the Americas in separate waves, with the arrival of Paleoamericans around 15,000 YPB and Amerindians around 4,000 YPB (Pucciarelli 2004). As a result, Paleoamerican and Native American cranial types represent and should be considered as two separate populations.

An alternative craniometrical study based on the analysis of late Pleistocene/early Holocene modern skulls challenged the classical interpretation of Paleoamerican and Amerindian craniofacial patterns, suggesting that the cranial differences should be viewed as extremes of a continuous morphological variation (Gonzalez-Jose et al. 2008). The authors of this study also suggest that these differences should be understood and explained by means of micro-evolutionary changes, genetic drift, gene flow and maybe directional selection.

Another study that compared craniometrical data from ancient groups in Argentina ranging from 8,000 to 400 years BP has found that the oldest individuals displayed the Paleoamerican features, but harboured the same mtDNA variation as the oldest individuals with the Amerindian morphology. The authors also concluded that morphological differences might be result also of a local phenomena, random genetic drift and non-random factors such as selection and plasticity (Perez et al. 2009).

The latest study on cranial morphometry in Native Americans argues that there is no clear distinction between Paleoamerican and Amerindian groups (Chatters et al. 2014). Excavations of submerged cave sites at the "Hoyo Negro" in the Yucatan Peninsula in Mexico, revealed the complete human skeleton of a teenage individual dated as 12,000–13,000 year-old. Morphological analysis of the skull helped to explain physical differences between the first Paleoamericans and Native Americans that should be interpreted as a result of *'in situ'* evolution within the continent after the migration out of Asia (Chatters et al. 2014).

The authors conclude that despite differences in face and skull features, Paleoamericans and modern Native Americans were related, and probably deriving from the same gene pool. However, upon re-analysis by an independent group, the ancient DNA evidence of this publication had been criticized. The "Hoyo Negro" sample does not show the characteristic damage patterns of ancient DNA, which therefore questions the antiquity of the samples (Prüfer and Meyer 2015). The original authors nonetheless claim that irrespective of the genetic observation the proposed hypothesis is strong enough based on the cranio-morphological evidence alone (Kemp et al. 2015). Moreover a recent study of Raghavan et al. 2015 did not find genetic support for the two waves of migrations of Paleoamericans vs. Native Americans (Raghavan et al. 2015).

Most recently, the "Kennewick Man", one of the most iconic human remains in the Americas, has been analysed morphologically and genetically (Rasmussen et al. 2015a). The population affinities of this 9,000-year-old male skeleton discovered in Washington State (USA) have long been subject to a heated scientific debate. Some anthropologists have claimed that the cranio-morphological and facial features do not resemble those of Native Americans and that he might have a connection with populations from Asia (i.e. Ainu, a group from Japan) or groups from the Polynesia (Owsley and Jantz 2014).

However, morphological evidence of this individual has been not conclusive and requires further investigations. The recent morphological re-analysis of the skull argues that it is not possible to link the Kennewick Man to specific contemporary groups in the Americas and that the phenotypic similarities with Asian and Polynesians groups are more likely a result of adaptations and/or '*in vivo'* modifications, genetic drift or natural selection in the continent (Rasmussen et al. 2015a).

Genetic evidence

Modern genetics is a major contributor for the study of human origins, dispersals and diversity across the world. During the last 30 years, technical and methodological developments to analyse DNA have helped to clarify some of the disputed theories about human prehistory, of which the peopling of the Americas is most relevant here.

The genetic portrait of the colonization process of the Americas was based on the two uniparentally-inherited genetic marker systems: mitochondrial DNA and the non-recombining portion NRY of the Y-chromosome, from early studies in the mid 1990's (Torroni et al. 1993; Forster et al. 1996). Native American genetic diversity represents a subset of the one found in Northeast Asia. Moreover, comparing to their Northeast Siberian neighbours, Native American populations exhibit a low level of genetic variability, expressed in only four mtDNA haplogroups (A, B, C, D) (Schurr et al. 1990; Torroni et al. 1992; Torroni et al. 1993) and two founder Y-chromosome lineages (C, Q) (Karafet et al. 1999).

During the initial colonization of the Americas, which involves the crossing of the Bering land bridge, the populations entering the Americas must have gone through a genetic bottleneck that resulted in a loss of diversity, the so-called "founder effect". The latter implies that all lineages of modern Native Americans are derived from these few founder lineages.

a. Mitochondrial DNA

The mitochondrial DNA (mtDNA) has become a widely used genetic tool to study the human evolutionary history (Cann et al. 1987; Ingman et al. 2000). In general, mtDNA has a number of properties and features that make it an efficient and convenient genetic marker system to study population genetics, phylogeography and allow the reconstruction of population history:

1) MtDNA follows an exclusively maternal inheritance and is passed on directly from mother to offspring.
2) MtDNA does not undergo recombination, which means mtDNA molecules are transmitted with relatively few changes through generations. The haploid nature of the molecule implies that at least initially all mtDNAs from an individual are identical.

However, in vivo mutability over the course of the lifetime of a cell/tissue type/organism can result in slightly different types of mtDNAs (hetereoplasmy) in the same organism.

3) MtDNA is characterized by a relatively rapid evolutionary rate, which allows the tracking of the accumulation of mutations over time (Brown 1979; Giles 1980; Olivio 1983). The resulting mtDNA phylogeny has a hierarchical structure. This fact and the differential geographic distributions and frequencies of specific profiles of mutations across various populations (Schurr 2002), allow us to reconstruct peopling histories and migration patterns.

4) MtDNA represents less than 1% of the total DNA, however the large amount and number of copies ($10^3 – 10^4$) per cell makes it a molecule of choice for genetic studies in human populations.

The mtDNA is a circular double-stranded molecule with an extra-nuclear origin. Two strands compose the human mtDNA, the H strand (heavy strand), which is rich in guanines (G), and the L strand (light strand), which is rich in cytosines (C).
The complete human mtDNA genome is on average 16,569 base pairs (bp) long and encodes for 13 polypeptides, integral members of the mitochondrial respiratory chain, 22 distinct transfer RNAs, and 2 ribosomal RNAs (Pakendorf and Stoneking 2005).

The mtDNA genome can be divided into two main regions:

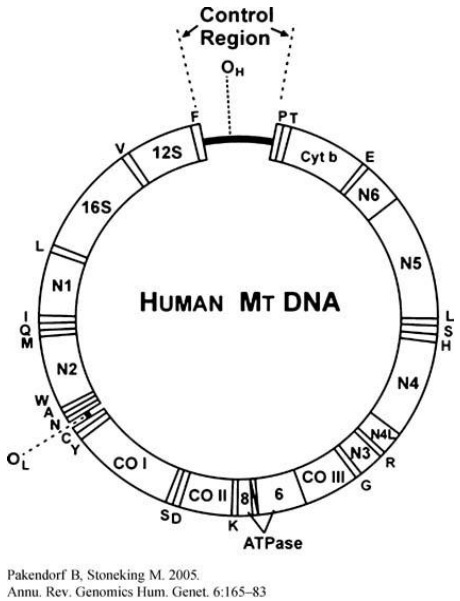
I) The Coding Region, represents the largest fraction and harbours functional genes underlying biochemical processes that take place inside the mitochondria.

II) The non-coding displacement loop (D-loop) or Control Region (CR) comprising about 1100 bp, which is located around the origin of replication and has mainly regulatory functions (Pakendorf and Stoneking 2005).

Sub-regions within the D-loop are called Hypervariable Regions (e.g. HVR I-II) and are characterized by a higher mutation rate compared to the rest of the mtDNA with high variability across populations, which is used to study population relationships (Stoneking et al. 1991; Schurr 2002). The base count starts at the replication origin on the heavy strand (O_H) at the position 1 in the middle of the CR (Figure 3).

The Control Region expands from position 16024 to 16569 (HVR-I) and continues from position 1 to 579 (HVR-II). The human mtDNA had been sequenced in its entirety for the first time by (Anderson et al. 1981), and since been termed "Cambridge Reference Sequence" (CRS). Due to some inconsistencies and misinterpretations of the sequence, a revision of such sequence was performed by (Andrews et al. 1999) modifying 18 nucleotides, now termed as "revised Cambridge Reference Sequence" (rCRS) and in use as mtDNA reference since. The fact that the rCRS represents a derived European mitochondrial sequence often leads to errors and confusion around the description of haplogroup defining SNPs (Single Nucleotide Polymorphisms).

A recent revision of mtDNA nomenclature based on in-depth analysis of modern human and available Neanderthal mtDNA genomes therefore suggested the use of a basal human reference sequence, termed "Reconstructed Sapiens Reference Sequence" (RSRS). Although not fully accepted by the community of mtDNA researchers, the RSRS provides a clearer perspective of human evolution and ancestry (Behar et al. 2012).



Pakendorf B, Stoneking M. 2005.
Annu. Rev. Genomics Hum. Genet. 6:165–83

Figure 3. Structure of human mitochondrial DNA (Pakendorf and Stoneking 2005)

b. Human mitochondrial DNA haplogroups

Researchers have estimated that the mtDNA has undergone a mutation every 10,000 years since the appearance of the first human, the so-called "Mitochondrial Eve" who lived ~200,000 years ago in Africa (Cann et al. 1987).

These selective mutations through time have been used to group mtDNA changes/variation into clusters termed haplogroups (hg). Haplogroups share a common recent ancestor and the set of mutations for a given lineage are transmitted to all descendants from the same maternal line. Human mitochondrial haplogroups are defined by the sum of a specific set of polymorphic sites or diagnostic substitutions (O'Rourke and Raff 2010). Haplotypes are defined as subsets of haplogroups carrying additional mutations or polymorphisms.

Historically, two main methods were used to characterize the variability of human mtDNA. The first mtDNA studies used restriction enzymes applying a method called RFLP-typing (Restriction Fragment Length Polymorphisms) to analyse DNA polymorphisms (Cann and Wilson 1983; Torroni et al. 1992). This was followed by the direct sequencing of the HVR-I and HVR-II (Vigilant et al. 1991) and later by the sequencing of complete mtDNA genomes (Kivisild et al. 2006).

Mitochondrial DNA substitutions have been accumulating throughout time along maternal lineages as human groups expanded towards different geographic regions around the globe. Therefore, mtDNA haplogroups have a particular geographic distribution and are continental-specific markers and could in theory be traced to specific points in time and space in human history (Schurr and Sherry 2004).

The nomenclature of mtDNA haplogroups follows a hierarchical structure and was introduced in the mid-1990's with the assignation of letters A-G to describe the variation in Asian and Native American lineages (Torroni et al. 1993), letters H-K for European lineages (Torroni et al. 1994a) and letter L to describe the variation in African populations (Figure 4). The standard reference for mtDNA haplogroup nomenclature and phylogeny is represented by "Phylotree" (van Oven and Kayser 2009).

The African mtDNA lineages comprises the Macro-haplogroup L, subdivided in L0 (L0a, L0d, L0f and L0k) with an East African origin being the oldest branch in the African mtDNA phylogeny. L0a and L0d are haplogroups traditionally found in Khoisan speakers (Tishkoff et al. 2007; Barbieri et al. 2014). L0a, L1 and L2 haplogroups are found in other sub-Saharan populations, including Bantu speakers and Pygmies (Tishkoff et al. 2007). Haplogroup L1 (L1a, L1b, L1c) is carried also by populations that are descendants of Khoisan speakers (Chen et al. 1995).

Haplogroup L2 (L2a, L2b, L2c, L2d) is distributed also in East Africa related with the Bantu expansion, represents the main variability in sub-Saharan populations (Salas et al. 2002). Haplogroups L3, M and N are also nested within Macro-haplogroup L. Haplogroup L3* (also known as N/M/L3) is restricted to the African continent (Salas et al. 2002) specific of sub-Saharan Africans, whereas M and N haplogroups originated in Eastern Africa but later dispersed into Europe and Asia around the time when anatomically modern humans colonized these regions (Quintana-Murci et al. 1999).

Mitochondrial DNA haplogroups H, I, J, K N1b, T, U, V and W are present in relatively similar frequency in many European populations and encompass virtually all European mtDNA lineages. All of them are derived from the Macro-haplogroup N (Torroni et al. 1996). Haplogroup U6, which is present in the Portuguese population is of North African origin (Pereira et al. 2000; Maca-Meyer et al. 2003; Pereira et al. 2010). Many Asian haplogroups are also derived from the Macro-haplogroup N (A, B, F, N9-21, O, S, P and Y), but most are derived from Macro-haplogroup M (M1-M40, C/Z, D, E, G) with different frequencies throughout Asian regions (Kivisild et al. 1999).

Australo-Melanesian haplogroups comprise region-specific M27, M28 and M29 lineages as well as Asian lineages in particular haplogroups B, F and E, while haplogroup N derived lineages O and S are exclusive to Australia, which shares haplogroup P with e.g. Papua New Guinea (Merriwether et al. 2005). Finally, Native American mtDNA haplogroup variation falls within A, B, C, D and X (see below).

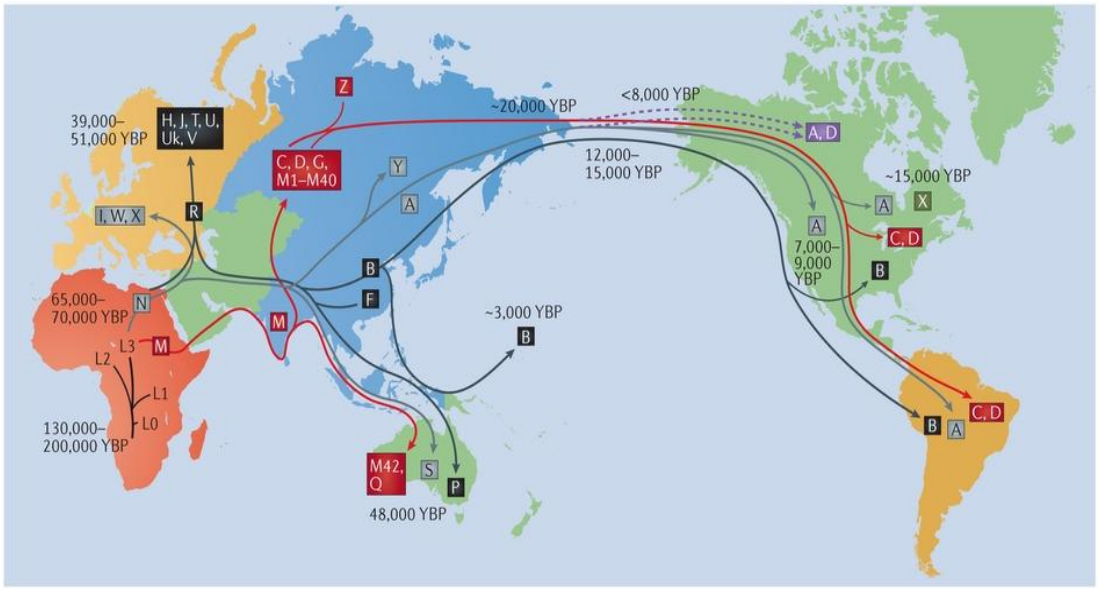


Figure 4. Map of mitochondrial DNA haplogroups with expansion times. After (Stewart and Chinnery 2015).

c. Mitochondrial DNA diversity in the Americas

The mtDNA variation in present-day Native Americans can be attributed primarily to five mtDNA haplogroups named A, B, C, D, and X (Torroni et al. 1992; Torroni et al. 1993; Torroni et al. 1994b; Schurr and Sherry 2004). A refinement of the resolution identified the so-called Pan-American haplogroups: A2, B2, C1b, C1c, D1, and D4h3a (Tamm et al. 2007; Achilli et al. 2008). In addition, the haplogroup composition across the Americas is compatible with a single wave of migration from Beringia (Merriwether et al. 1995; Achilli et al. 2008; Fagundes et al. 2008; Kitchen et al. 2008).

Recent studies based on high-resolution genomic data of complete mtDNA genomes have expanded the knowledge about mtDNA variation in the Americas by identifying several new sub-lineages. In consequence, the overall mtDNA diversity in Native Americans is better described as 15 American founder lineages (A2, A2a, A2b, B2, C1b, C1c, C1d, C1d1, C4, D1, D2a, D3, D4h3a, X2g, X2a) (Achilli et al. 2008; Perego et al. 2010; Bodner et al. 2012). The frequencies of mtDNA founding haplogroups A-D and X in the Americas show wide pattern of distribution from North to Central and South America (Malhi et al. 2002; Schurr 2002). Moreover, another study suggests that geographic and linguistic factors moderately influenced the mtDNA distributions (6% and 7%, respectively) and mtDNA haplogroups A and D correlated positively and negatively with latitude (Bisso-Machado et al. 2012).

Haplogroup A is found throughout the Americas, but occurs at highest frequencies in North American populations (Tamm et al. 2007; Perego et al. 2010), and decreases considerably in southern latitudes (Lalueza et al. 1997). Haplogroup A comprises sub-haplogroup A2a mainly found in Inuit, Na-Dene and Siberian populations (Achilli et al. 2013), while sub-haplogroup A2b can be found in Eskimo speaking populations across the Arctic (Gilbert et al. 2008a).

Frequencies of haplogroup B are mainly restricted to populations from South America, with a high concentration in the Andean region (Rodriguez-Delfin et al. 2001; Lewis et al. 2007; Afonso Costa et al. 2010; Barbieri et al. 2011; Gaya-Vidal et al. 2011; Sandoval JR et al. 2013; Taboada-Echalar et al. 2013), but with relatively low frequencies in North America (Schurr et al. 1990; Torroni et al. 1992; Torroni et al. 1993; Malhi et al. 2001).

Haplogroup C is present in Central and South American groups (Perego et al. 2010). Sub-haplogroups within the Americas includes C1b, C1c, C1d and C4c (Perego et al. 2010; Kumar et al. 2011). Haplogroup D is characteristic of populations in the southern parts of South America (Bodner et al. 2012; de Saint Pierre et al. 2012a). Sub-haplogroups within the Americas comprise D1, D2a, D3 and D4h3a (Perego et al. 2009). On the other hand, haplogroup X, with sub-haplogroups X2, X2a is found nearly exclusively in North America (Torroni et al. 1992; Torroni et al. 1993; Brown et al. 1998; Malhi et al. 2001). It has been suggested that this distribution likely reflects both the original pattern of settlement of the Americas and the subsequent genetic differentiation of Native American populations within the continental areas (Schurr and Sherry 2004).

d. Complete mtDNA genome data in the Americas

There has been increasing emphasis to generate genomic data from Native American populations. A better resolution of the genetic information is crucial to understand the details of population differentiation and dispersal within the continent.

To date, complete mtDNA genomes from ~600 modern-day Native Americans have been sequenced (Tamm et al. 2007; Achilli et al. 2008; Fagundes et al. 2008; Perego et al. 2010; Bodner et al. 2012; Cardoso et al. 2012; de Saint Pierre et al. 2012b; Achilli et al. 2013), with most studies aiming to test a suite of migration models and to estimate entry dates into the Americas. However, such mtDNA data focused only on specific sub-lineages of interest, and therefore cannot be considered as a representative source for the mtDNA genetic diversity of the continent. Nevertheless these mtDNA sequences have been used to refine and improve the human mitochondrial tree (van Oven and Kayser 2009). Mitochondrial DNA genome information from ancient population is still scarce and restricted to North America; amongst these are the Saqqaq, a Paleo-eskimo individual (Rasmussen et al. 2010), four Mid-Holocene individuals from north coast of British Columbia, Canada (Cui et al. 2013), the Anzick-1 genome from an individual associated with the Clovis culture (Rasmussen et al. 2014) and the Kennewick Man (Rasmussen et al. 2015a). A study in South America has reported five Early to Middle Holocene mtDNA genomes from Lauricocha individuals in highlands Peru (Fehren-Schmitz et al. 2015) and another study by Raghavan and colleagues features high coverage genomes of two Native American present-day Inuit, two Siberians, one Aleutian Islanders and two Athabascans (Raghavan et al. 2014a).

e. Y-chromosome

In humans, biological sex is determined by presence or absence of the two distinct sex chromosomes X and Y. Females have two copies of the X-chromosome (XX) and males one X-chromosome and one Y-chromosome (XY). The Y-chromosome is paternally inherited from father to son. It also contains a non-recombinant region (NRY) which, similarly to the mtDNA for the maternal history, makes it a highly effective means to study the paternal genetic history of humans (Jobling MA 1995).

Native Americans exhibit also a very limited amount of Y-chromosomal variation in contrast with Asian populations (Karafet et al. 1999; Zegura et al. 2004), which is consistent with the assumption that the peopling of the Americas involved a population bottleneck. Like mtDNA, the reduced genetic diversity of Y-chromosome markers in Native Americans is best explained as a consequence of a founder effect and according to Karafet and colleagues the current distribution throughout the Americas seems to reflect a genetic drift (Karafet et al. 1999).

Studies of paternally inherited NRY genetic variation in Native Americans consistently confirm the bottleneck hypothesis, as only a small subset of the Asian Y-chromosome haplogroups have been identified in the continent. The founder haplogroups C and Q were most probably carried into the Americas during a single migration process and occur in the majority of indigenous populations (Karafet et al. 1999; Schurr and Sherry 2004; Zegura et al. 2004).

High-resolution screening of haplogroup Q established a common ancestor between Southern Altaian and Native American groups and revealed important information about the potential the origins of population that settled the American continents (Dulik et al. 2012). A study focussing on Y-chromosome Single Nucleotide Polymorphisms (SNPs) analyses identified three major haplogroups: C, Q including R from a big sample size (n= 588) (Zegura et al. 2004). Haplogroup C was found in Amerind, Na-Dene and Aleut-Eskimo speakers and haplogroup Q found only in Na-Dene speakers, with both representing nearly 96% of Native American Y-chromosomes.

Moreover, demographic scenarios in Native Americans were also evaluated by studying Y microsatellite markers (Y-STRs), which have a much faster evolutionary rate than SNPs. Y-STRs have been a helpful tool in understanding of population and chromosome evolutionary histories (Bisso-Machado et al. 2012). In this study focused on the analysis of Y-chromosome haplogroups from 68 populations and 1,814 individuals plus Y-STR markers from 29 populations and 590 subjects, results showed that the

diversity of the markers makes it difficult to establish a general picture of Y-chromosome variability in the populations subjected to study. However, haplogroup Q1a3a* was almost always the most prevalent, whereas Q1a3* occurred equally in all regions. This suggested its prevalence among the early colonizers. The STR allele frequencies were used to derive a possible ancient Native American Q-clade chromosome haplotype, which showed significant geographic variation. Geography apparently plays a more important role than language in explaining the data for the Y-chromosome Q clade-STRs (Bisso-Machado et al. 2012).

Finally, the presence of other Y-haplotypes (*i.e.* R haplotypes) observed in Native American populations most likely represents the result of recent historic admixture with European and African populations (O'Rourke and Raff 2010).

f. Autosomal DNA

Autosomal DNA refers to the 22 pairs of chromosomes found in the nucleus excluding sex chromosomes X and Y, whereby one copy of each chromosome is biparentally inherited from the mother and one from the father and therefore reflects both male and female ancestry. Similar to the results from mtDNA and the Y-chromosome studies, geneticists have observed a reduced autosomal genetic diversity in extant Native Americans, showing a decline in genetic variability directly correlated with population distance from the Bering Strait (Wang et al. 2007; Halverson and Bolnick 2008).

A particular short tandem repeat (STR) marker studied in Native Americans, termed ‘D9S1120’ (Schroeder et al. 2007) reveals a high frequency in the continent and has been characterized as ubiquitous and universal for Native American groups, suggesting that D91120 resulted from a common descent. Due to the novelty of this allele it has been suggested that the Americas were colonized by a single founding population (Schroeder et al. 2009). The analysis of autosomal data showed also a clear north-to-south reduction in diversity when compared to Siberian populations, supporting the conclusions of the previous studies based on either mtDNA or Y-chromosome (Wang et al. 2007). A recent study addressed questions revolving around the timing and mode of the entry into the Americas. Researchers used genome-wide data from 52 Native American (n=493) and 17 Siberian groups (n=245) genotyping at 364,470 SNPs by using Illumina microarrays. They proposed that North and South America were populated by three migrations waves rather than just a single migration (Reich et al. 2012).

g. Ancient DNA analyses in the Americas

Ancient DNA (aDNA) data is now widely used to reconstruct origins, migration routes, settlement histories and relationships of past populations around the globe. During the last two decades, aDNA research has helped to clarify some questions about the peopling of the Americas and the evolution of the populations within the continent (Tamm et al. 2007; Gilbert et al. 2008a; Kitchen et al. 2008; Rasmussen et al. 2010; Rasmussen et al. 2014; Rasmussen et al. 2015a). Most important, aDNA studies in the Americas offer the possibility to shed light on demographic and historical events prior to the European contact. When compared with modern-day diversity, aDNA data can help to describe the evolution of the genetic pool of pre-Columbian populations over time.

For instance, aDNA was used to type blood groups of ancient Native Americans that showed a low genetic diversity associated with ABO blood group markers comparable to extant Native Americans (Halverson and Bolnick 2008). A study on aDNA of human coprolites from one of the earliest human settlements in the Americas, the Paisley 5 Mile Point Caves in Oregon, USA dated to 14,000 YBP, reported Native American founder mtDNA haplotypes A2 and B2 (Gilbert et al. 2008b).

A recent review from Raff and colleagues provides a comprehensive synthesis of genetic research of aDNA in the Americas, summarizing the genetic diversity of prehistoric populations throughout the continent (Raff et al. 2011). A comparison of aDNA mitochondrial frequencies with contemporary genetic variation confirms once more the existence of the founder A, B, C, D haplogroups in pre-contact times suggesting also that the overall geographic structure of mtDNA in Native Americas seems to have been established very early (O'Rourke and Raff 2010).

It is worth noting that, no other haplogroup has been identified from ancient Native Americans samples with the exception of the haplogroup X in North America (Malhi and Smith 2002), which is present in native populations in restricted areas of North America and some populations from South America (Ribeiro dos Santos et al. 1996). However, it is theoretically possible that early migrants into the Americas might have carried a higher genetic diversity or eventually other lineages.

A study by Malhi et al. 2007 identified the presence of mtDNA haplogroup M in two mid-Holocene individuals from North America. This study was met with scepticism as M is a common haplotype present in East Asia but has never been reported in ancient or modern population in the Americas (Malhi et al. 2007).

At the moment it is not clear how robust these results are, but it suggests that other lineages could have existed in pre-contact Americas, which entered the continent in low frequency and were lost by drift, or after the population demise following the European contact.

Research on aDNA was also used to investigate the formation of the genetic diversity of Native Americans through time. Although at global scale the broad genetic structure of prehistoric populations in the Americas seemed to been established rapidly (Reich et al. 2012), the observation of widespread genetic homogeneity among groups suggests a degree of population continuity over time. However, at smaller scale there is evidence for genetic discontinuity and population replacements in particular regions in the continent, e.g. Greenland and Xaltocan, Mexico (Gilbert et al. 2008a; Mata-Míguez et al. 2012).

By and large, genetic evidence in the Americas suggests population continuity between pre- and post-contact human groups, based on haplogroup composition (mtDNA and Y haplogroups) and shared ancestry (autosomal DNA data) through time. However, the resolution of pre-contact genetic diversity across the Americas is poorer than that of modern-day populations. Therefore more detailed analyses on a larger number of ancient populations are needed to fully characterize the pre-Columbian genetic pool of Native Americans and their relationship with present-day native groups throughout the continent.

The development of new analytical methods combined with methodological improvements offers the opportunity to generate more genomic data from ancient and modern samples. Massive amounts of genetic information can be nowadays obtained using the Next-Generation Sequencing (NGS) technology, which has generated a better resolution of the genetic information.

Recent genomic aDNA studies have provided important insights into the colonization of the Americas. For instance, the genetic analysis of a complete nuclear genome from a 24,000 year-old Upper Paleolithic individual in Siberia (Malt'a boy) has found a genetic link between ancient Siberians and the initial settlers of the Americas (Raghavan et al. 2014). Moreover, aDNA retrieved from the only archaeological human remains associated with the "Clovis" culture (Anzick-1) could clarify the genetic roots of Native American populations in the Americas (Rasmussen et al. 2014). Interestingly, the genetic signature from this individual shows a cline of gene flow with the Mal'ta boy.

The Mal'ta boy's genome offers the first global picture of genetic composition in Native Americans, suggesting that 18-38% of its genome variation is present in the current genetic pool of indigenous Americans across the continent. In addition, the report of a complete genome sequence from a 40,000-year-old early modern human from Tianyuan cave in China, has revealed a basal mtDNA 'haplogroup B' shared by Native American and present-day Asian populations (Fu et al. 2013).

Furthermore, aDNA analysis in the pre-Columbian Americas not only provides the opportunity to study past events in the continent; it has also been used to test the demographic impact of the European contact on Native American populations. Historical accounts report that the European colonization lead to a dramatic demise of Native Americas, and list a range of factors such as warfare, exploitation, disruption of the social structure, epidemics and diseases (Mann 2005).

By applying Bayesian analysis to a combined set of ancient and modern genetic data (O'Fallon and Fehren-Schmitz 2011) attempted to investigate the consequences of the European contact on Native American people. The study showed a dramatic episode of population reduction, shown clearly by a decline of the female effective population size, which coincides with the European arrival 500 years ago.

A recent study by (Llamas et al,. in press) offers the first genetic view of the role that European colonization played in reducing the overall Native American genetic pool to the low levels observed today. Gaining access to aDNA information and the power of a better resolution dataset was fundamental to refine this scenario, suggesting a local mass mortality and extinction of genetic diversity and lineages after the European contact. This implies that the European arrival in the early 16^{th} century has had a strong impact over Native Americans leading to a second population bottleneck event, which also resulted in a loss of diversity of the Amerindian genetic pool.

By applying the NGS approach, the field of ancient DNA will continue to generate new insights to elucidate essential questions about human evolutionary history in pre-Columbian Americas, plus a more extensive sampling either from ancient or modern-day population and in-depth sequencing efforts are needed to formally test this scenario.

h. Hypotheses on the peopling of the Americas based on modern genetic data

Genetic information has been proven essential in testing colonization scenarios into the Americas. The exact timing when the first people migrated into the Americas is critical for our understanding of the human evolutionary history of the continent. A common feature of all genetic studies, whether they focus on mtDNA, Y-chromosome or autosomal markers, is the fact that Native Americans show a reduced genetic diversity when compared to worldwide populations. This low genetic diversity links to a small group of individuals (founders) during the initial peopling of the Americas (Wang et al. 2007; Halverson and Bolnick 2008).

For instance, combining information gathered from mtDNA, Y-chromosome and autosomes, researchers have estimated that the effective size of the founding population for the Americas was fewer than 80 individuals, approximately 1% of the effective size of the estimated ancestral Asian population (Hey 2005). However, other estimates based on mtDNA data range from an effective female population size of 1,000– 2,000 (Mulligan et al. 2008) to 1,000–5,400 females (Kitchen et al. 2008).

Based on the fact that all four mtDNA haplogroups were present throughout the Americas, a single migration event was proposed (Merriwether et al. 1995; Forster et al. 1996), suggesting that the patterns of genetic variation across the Americas can be attributed to '*in-situ'* differentiation of Native American populations after the first colonization event rather than as a consequence of subsequent expansion waves (Schurr 2008). Based on the analysis of a continuous segment of 8.8 kb of the mtDNA coding region from 30 Native American individuals, Silva and colleagues argued also for a single migration wave into the Americas by observing a low diversity within the four Native American mtDNA haplogroups A, B, C and D (Silva et al. 2002). The authors calculated a time of differentiation at around 21,000 YBP and argued for a shared ancestry and common origin of all Native Americans, which suggested a colonization of the continent before the LGM.

On the other hand, multiple streams of migrations from Siberia into the Americas have been also proposed (Greenberg et al. 1986; Bortolini et al. 2003; Perego et al. 2009). For instance, early coalescence age estimates argued for the occurrence of two or more migrations based on older dates obtained for haplogroups A, C, D and X (between 35,000 – 20,000 YBP), that might have entered the continent in a first wave (Torroni et al. 1992; Torroni et al. 1993; Brown et al. 1998), and a younger date for haplogroup B (between 17,000 – 13,000 YBP), which suggests that this lineage might have arrived

later in the Americas in a separated migration. However, more recent studies using improved methodologies rather argue for a synchronicity of coalescence age estimates for all Native American lineages (Kumar et al. 2011)(Llamas et al. in press).

An alternative peopling scenario proposed by Schurr and Sherry suggested a Pacific coastal migration of people around 20,000 – 15,000 YBP containing only A-D mtDNA lineages, followed by a second migration via the ice-free corridor containing the haplogroup X (Schurr and Sherry 2004). This dual hypothesis (Figure 5) was further supported by Perego and colleagues based on the analysis of two rare mtDNA lineages, D4h3 distributed in populations along the Pacific coastal regions in North and South America, and X2a present only in populations from North America, which might have entered the Americas once the ice-free corridor opened up (Perego et al. 2009).

Furthermore, another model has suggested that Asian migrants experienced a long-term occupation in Beringia before moving on into the Americas (i.e. Beringian Incubation Model (Tamm et al. 2007)) (Figure 5).

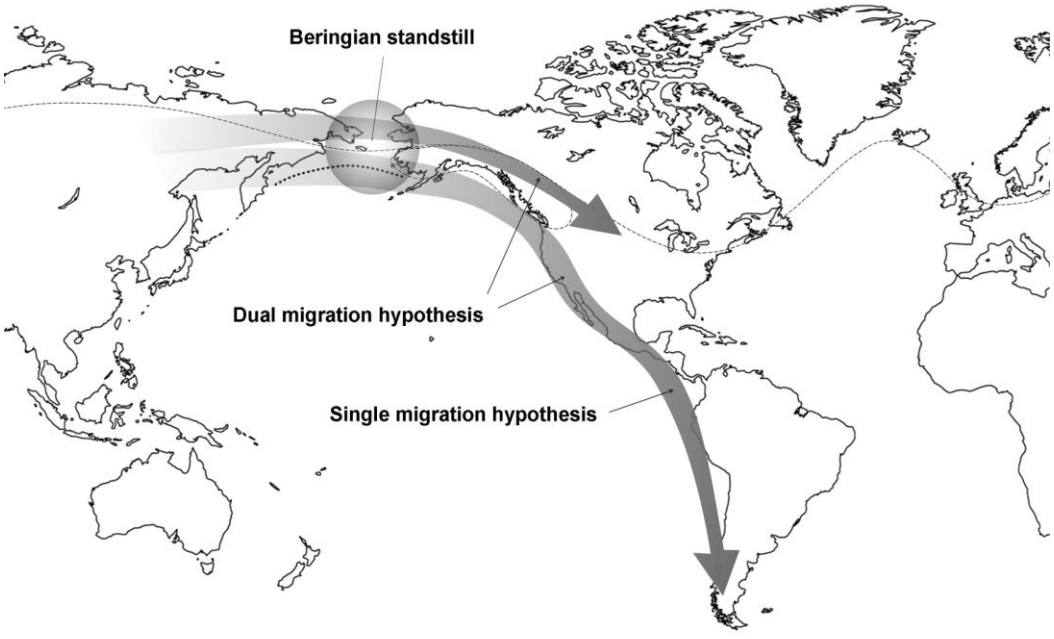


Figure 5. The single, dual and Beringian standstill hypothesis for migration into the Americas (Marangoni et al. 2013)

The generation of high-resolution mtDNA genome data from Native American populations with a broad geographic coverage, alongside the development of modelling algorithms (e.g. Bayesian phylogenetic analysis or Bayesian coalescent simulations), has

allowed researchers to explore more plausible peopling scenarios and the timing of population entry in the Americas in more detail (Tamm et al. 2007; Perego et al. 2010).

Tamm and colleagues proposed a refined model of the single demographic event for the peopling of the Americas, coined "Beringian Incubation Model" (BIM) (Tamm et al. 2007). The mtDNA genome data (n=623) available after sequencing Asian and Native American samples revealed a previously hidden diversity within the American founding haplogroups, as well as evidence for a potential recent gene flow and back migrations between the continents. According to the authors, the most parsimonious hypothesis was a long-term incubation of Native American founders around Beringia (comprising the main haplogroups A, B, C and D), where the population had sufficient time to differentiate from their Asian sister clades. Tamm and colleagues argue that the Beringian environment provided "glacial refugium" until the climatic and ecological conditions permitted entering the Americas, upon which humans rapidly spread into the continent, first via the coastal route and then via the ice-free corridor (Tamm et al. 2007).

Fagundes and colleagues also proposed a single pre-Clovis population migration around the LGM by analysing complete mtDNA genomes (n=86) and applying novel Bayesian methods that allowed the reconstruction and the timing of demographic events (Fagundes et al. 2008). The authors observed a strong population expansion right after the LGM (but pre-Clovis) between 18,000 and 15,000 YBP. Since these dates were not compatible with the dates of the opening of an inland corridor, which only became ice-free around 13,500 YBP (Mandryk et al. 2001), they argued for a Pacific coastal route, which was ice-free by 15,000 YBP (Taylor et al. 2014). This study was supported independently by Achilli and colleagues that obtained a coalescence time ranging from 18,000 to 21,000 years for haplogroups A2, B2, C1 and D1 after analysed (n=171) complete mtDNA genomes using a phylogenetic reconstruction approach (Achilli et al. 2008).

Ho and Endicott (2008) re-analysed the dataset of Fagundes and colleagues using different calibration techniques to estimate substitution rates in sequence data from Native American populations. They argued that the global substitution rate (1.26×10^{-8} subs/site/year) used by Fagundes et al. 2008, which was inferred by calibration with the human-chimpanzee split, is not adequate for the study of a human sub-population such as Native Americans. Instead, they applied a higher mutation rate (2.038×10^{-8}

subs/site/year), based on substitution rates derived from internal calibrations for mtDNA evolution (Endicott and Ho 2008). The resulting younger dates for the population growth commencing 13,000 – 12,000 YBP challenged the pre-Clovis population expansion proposed by Fagundes and colleagues (Ho and Endicott 2008).

An alternative study by Kitchen and colleagues explored mitochondrial (77 mtDNA coding region sequences, plus 812 mtDNA HVRI+II sequences) and nuclear data using the Bayesian skyline approach implemented in the software BEAST. This analytical tool infers effective population size backwards through time from a set of sequence data and has been demonstrated to be an effective means for reconstructing population dynamics (Drummond et al. 2005). Based on the shape of resulting demographic curve, the authors proposed a "Three-step model" with detailed timeframe with a three stage colonization event (Kitchen et al. 2008). The initial stage (~43,000 – 36,000 YBP) comprised an expansion and genetic divergence between populations from East Central Asia and the ancestors of Native Americans. The second stage was interpreted as occupation of Beringia, characterized by a period of isolated settlement and a population growth for the subsequent 20,000 years, which was consistent with the "Beringian standstill" hypothesis (Tamm et al. 2007). The third stage was the colonization of the continent by expansion throughout the Americas around 16,000 years ago.

Reich and colleagues give support to the evidence of three major migration waves into the Americas, sustaining the controversial interpretation of language relationship proposed by (Greenberg et al. 1986) . The authors argue that most Native Americans descent from an ancestral group they called "First Americans" that crossed the Beringian land bridge from Siberia. However, Eskimo-Aleut speakers from the Arctic region inherited almost half of their ancestry from a second stream of Asian genes/pool and the Chipewyan Na-Dene speakers from Canada, inherited roughly one-tenth of their ancestry from a third stream (Reich et al. 2012).

In terms of dissemination of people in the continent , the archaeological evidence suggests that humans peopled South America most likely between 15,000 and 13,500 YBP (Goebel et al. 2008). However, there is much debate regarding the routes and processes of early human migrations in South America. The settlement of the continent, according to some authors involved nomadic hunters, fishermen and gatherers who crossed the Isthmus of Panama and entered the northern Andean region spreading

towards the central and southern Andean region of South America (Bennett and Bird 1964). Other studies suggest two major migratory movements into South America. First, a migration into the Amazonas followed by a westward movement towards the highland regions in the Andes, suggesting that present-day Andean populations can trace their ancestry to Amazonian groups (Rothhammer et al. 2001). Another scenario for the peopling of South America, suggests a migration of populations heading first to the southern part of the continent along the Pacific coast, and then a subsequent migration towards the Andean highlands to finally move to the Pampas and Patagonia regions in the southern part of South America (Rothhammer and Dillehay 2009). However, based on differentiation in Amazonian and Andean populations it was observed that the overall genetic variance between these main two groups is low, which has led to postulate a single major migratory event for the settlement of South America (Moraga et al. 2000; Lewis et al. 2007).

A genetic study of South American populations identified two novel mtDNA subclades D1g and D1J, which are geographically restricted to populations from southern Argentina and Chile (Bodner et al. 2012). The estimated age of these mtDNA hg D1 subclades (18,000 - 14,000 YBP) coincides with the earliest dates of the archaeological site of Monte Verde in southern Chile. The authors proposed three models for the spread of these lineages to South America's southern cone whereby humans could have spread both eastward and westward after a period of incubation in north-western South America or expanded along the Pacific coast before crossing the Andes, with or without a subsequent gene flow between east and west (Figure 6).

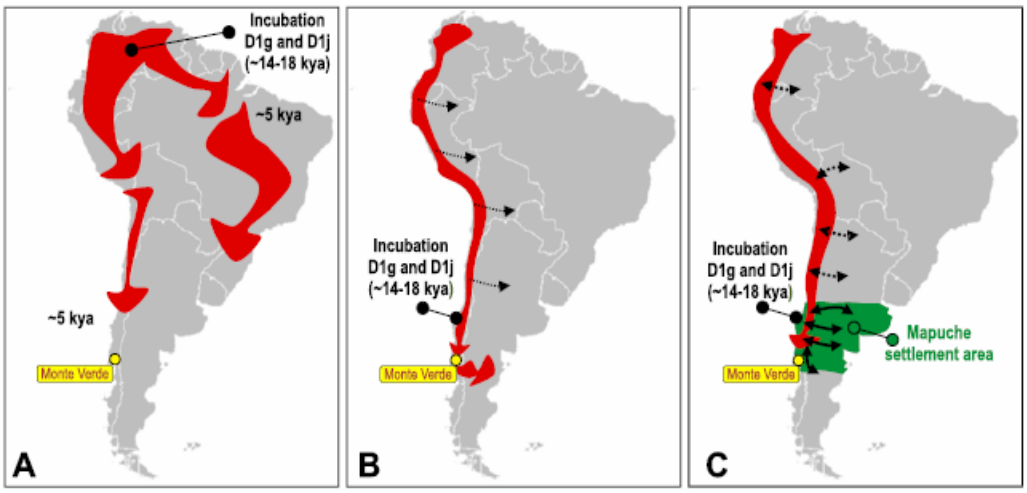


Figure 6. Maps showing the direction and timing of three alternative dispersal models in South America (Bodner et al. 2012).

The latest genetic analyses performed by two research teams focused on detailed exploration of genomic information from modern-day and ancient populations, have added another layer on the complexity of the peopling of the Americas. Both publications highlight the novel detection of traces of DNA signatures from people from Australia and Melanesia into the DNA diversity of some Native American groups (Raghavan et al. 2015; Skoglund et al. 2015).

The study by Skoglund and colleagues proposed a working model stating that a group of "First Americans" migrated into the continent around 15,000 YBP, while a second group tentatively called "Population Y" could have migrated before, after or even at the same time as the First Americans. Incidentally, this Population Y carries ancestry more closely related to indigenous Australians, New Guineans, and Andaman Islanders than to any present-day Eurasians or Native Americans. However, this diversity is only reflected in Amazonian populations from present-day Brazil in South America, the Karitiana and Suruí tribes whereas no other population from North America, Mesoamerica or western South America carries such genetic signature (Skoglund et al. 2015).

The study performed by Raghavan et al., 2015 postulated a single migratory wave into the Americas, which focused on the development of the Eurasian migration across the Bering Strait. The authors suggest that the migration across the land bridge of Beringia took place at the height of the last glacial period (Raghavan et al. 2015) . Native Americans diverged genetically from Eurasians around 23,000 YBP, however the Eurasian migrants were locked in the Bering Strait by the North American ice sheets for about 8,000 years. After the retreat of the glaciers about 14,000 years ago the first colonizers continued moving along the coast to the southernmost part of South America (Raghavan et al. 2015).

While this study did not performed an in-depth analysis of the Amazonian groups, the team did find a weak link between Australasians ancestry in some South American native populations, albeit not as strong as observed by Skoglund and colleagues. Both studies therefore suggest that the ancestry of the first Americans is a lot more complicated than scientists had envisioned, which leads to think a more diverse set of founding populations in the Americas than previously accepted (Skoglund et al. 2015).

Conclusion from genetic data

Genetic evidence combined with archaeological, linguistic and palaeoecological has allowed to researchers to propose plausible peopling scenarios describing the origin of Native Americans groups and migratory routes. The most likely hypothesis based on genetic and archaeological evidence suggests that the Americas were colonized 15,000 YBP after the deglaciation of the Cordilleran icesheet which opened up the Pacific coastal corridor (Goebel et al. 2008). Overall, genetic and linguistic data tend to support the hypothesis of (at least) three independent waves of dispersal into the Americas (Greenberg et al. 1986; Reich et al. 2012), although the linguistic theory is still controversial (Bolnick et al. 2004). Due to advances in sequencing technology it is now more feasible to address the long-standing question of the origins of Native Americans. In consequence, based on latest research it seems that large-scale genomic information (i.e. nuclear genomes) supports at least two big migrations of populations from Asia into the Americas (Skoglund et al. 2015). However, an agreement between mtDNA, Y-chromosome and autosomal evidence has not been reached so far with regard to hypothesis supporting colonization models (i.e. single vs. multiple streams of migration). Nevertheless, these models and their underlying assumptions will continue to be used as the framework for testing further hypothesis regarding peopling scenarios in the Americas (O'Rourke and Raff 2010).

Final remarks

Research on human population history in the Americas will continue. Ancient DNA represents a powerful tool to test long-standing questions about pre-Columbian Americas. The key advantage of aDNA techniques is to travel back in time, which enables us to assess genetic diversity in real-time, and to road-test existing hypotheses about the first settlements in the continent. Additional genetic studies of prehistoric human remains in the Americas are likely to reveal important insights into population history of Native Americans. However, is it important that future studies adopt a multi-disciplinary approach in order to incorporate knowledge from diverse disciplines and to attempt a better synthesis. With novel technological analytical advances, aDNA research can play a central role and the incorporation of large numbers of samples from ancient populations in a broader temporal and geographic context within the continent will refine our perspective and understanding of pre-Columbian civilizations.

B) Population history of Native Americans after the European arrival

For centuries, Native American populations lived and evolved in isolation from Old World influence. America before Columbus was characterized by highly-developed and sophisticated civilizations with technological know-how and complex cultural, socio-economic and political structures (Mann 2005). An estimated number of several million indigenous people lived in the Americas before in pre-Columbian times (Dobyns 1983; Mann 2005; Livi-Bacci 2006), and the catastrophic demographic decline of the indigenous population (Mann 2005) was estimated to have caused the death of almost 95% of the population during the first 130 years after the contact (Livi-Bacci 2006).

Same estimates for population counting have suggested that indigenous peoples in the Americas were reduced from over 44 million to 2 or 3 million in fewer than 100 years – reviewed in (Crawford 1998). Overall counts in the Americas following the so-called "approximation of populations densities" calculated the approximate size of pre-Columbian indigenous populations in South America ranging around 10 million peoples, estimating that almost 3,5 millions people were dispersed in the Central Andes (Steward and Faron 1959).

Even in South America lowlands it has been suggested that the European contact resulted in the extinction of 75% of societies and the loss of over 95% of the overall population in indigenous Amazonians from Brazil (Hamilton et al. 2014). While the precise scale and dimension of this catastrophe is still debated, the reduction of population size accounts for a significant population bottleneck, likely led to a considerable loss of genetic variability in indigenous populations. The main cause of such decline is likely due to violent deaths during the European colonization, but also the exposure to new infectious agents is thought to have played an important role in the Native American population collapse. In fact, diseases all over the world are known to have played a significant role in human history (Patterson and Runge 2002). In order to explain the particular vulnerability of Native Americans to European diseases two possible hypotheses have been suggested: the disease-free Eden or Virgin soil paradigm, and the Black Legend paradigm.

The Prehispanic "Disease-free-Eden" / Virgin soil Paradigm

For millennia, Europe/Eurasia and North Africa have been the centre of war and trade, facilitating a fertile ground for diverse human contacts and also allowed for the spread of transmissible and contagious diseases (Dobyns. 1993; Diamond 1997). In

addition, early domestication events involved close and frequent contact between animals and people, increasing chances for pathogens (virus or bacteria) from domestic animals to mutate and transfer to humans, fostered a source of new virulent diseases associated with the fauna (Diamond 1997). In contrast, the few domestic animals associated to the Native Americans populations (e.g. guinea pigs, llamas, alpacas, etc.) did not allow exposure to new diseases in comparison with the Europeans.

Long term exposure of a population to diseases leads to the development of immune resistance (Diamond 1997). Therefore Old World populations had acquired some resistance to their own diseases but Native Americans populations lacked the same immunity mechanisms. An alternative hypothesis is based on the observation that, prior to the initial peopling of the Americas, the founding population experienced a genetic bottleneck in Beringia. Therefore, assuming general reduction of genetic diversity would in turn explain a lack of immune response to new diseases in Native American populations. In other words, the lack of immune response is solely due to the overall reduced genetic diversity and not due to the absence of exposure to pathogens.

Old World diseases that were not present in the Americas before Columbus include smallpox (Figure 7), measles, chickenpox, influenza among others (Larsen 1994). However, the Americas were not free of endemic pathogens. There is documented evidence to suggest that endemic diseases such as tuberculosis or syphilis were present in the Americas before Columbus (Salo et al. 1994). Also, epidemic outbreaks were known to Native Americas people (e.g. Cocolitzli a type of haemorrhagic fever) mainly during periods of drought (Acuna-Soto et al. 2002).

The contrast between the Old World environment and the state of isolation of the Americas (disconnected from other continents since the initial peopling) is an ideal circumstance where a "biologically naïve" population exposed to new diseases succumbs at a high rate, resulting in a "virgin soil" epidemic (Crosby 1976). In this paradigm, the Old World infectious diseases brought by the Europeans to the Americas represent a major, if not the main factor behind the depopulation of the American continent during the European colonization. The concept that pre-Columbian populations lived in a "disease-free Eden" before the first contact represents a possible epidemiological model that explains in realistic terms the impact of Old World diseases on the 'naïve' immune system of the Native Americans (Livi-Bacci 2006). This would explain how the health of the American people was altered after 1492 (Mann 2005).

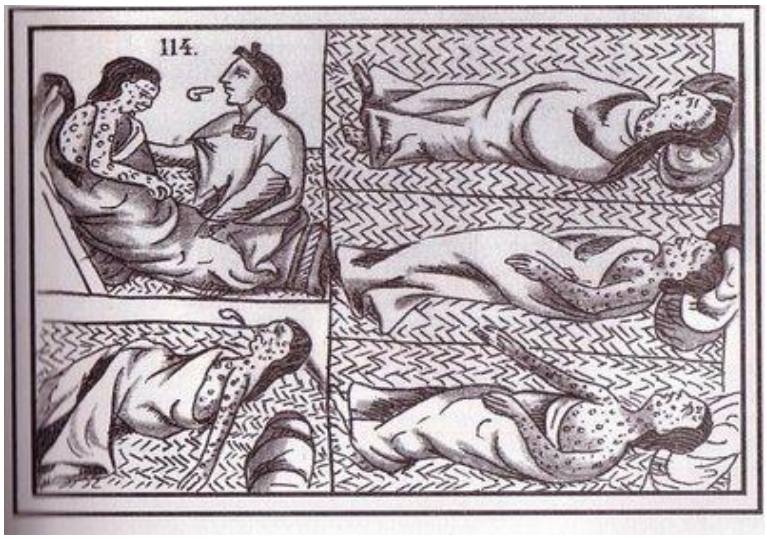


Figure 7. Native Americans suffering from smallpox (Sixteenth century Mexica drawings). After (Mann 2005).

Black Legend Paradigm

The arrival of the Spanish "Conquistadores" in the Americas in 1492 also meant the disruption and destruction of the Native American culture and way of living. The remarkable development of Native Americans in their own conceptions of the world, state, writing, languages, art, food, education, and religion were practically destroyed after the European contact. In contrast to the disease-driven demise, here the cruelty of the Spanish conquest and regimes' harsh conditions represent the main causes of the demographic catastrophe and is associated with environmental (destruction of infrastructure, deforestation, etc), political (wars, conflicts, warfare), social (dislocation and disruption of communities and cultures), economic (exploitation, confiscation) and demographic (abduction, separation) factors that strongly impacted the Native American societies (Figure 8) (Livi-Bacci 2006).

Moreover, the so-called "Indian reductions" or mission towns, established by the Jesuit missionaries in Central and South America were developed in order to relocate the indigenous populations. The main purpose of the reductions was to establish a direct control over Native populations through culturally transformation and religion (i.e. Christianization), establishing also in the taxation and exploitation of the indigenous communities. As consequence, the shift into the reductions had highly disruptive effects on the indigenous society, which experienced one more time the break of families and group relationships with the exposure to diseases and forced labour.

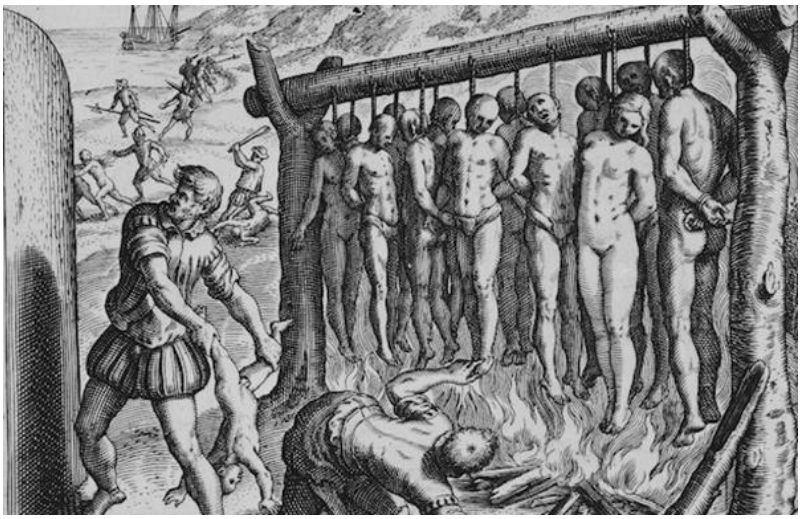


Figure 8. The Black Legend or "Leyenda Negra Española" is a term used to describe the cruelty and intolerance of the Spaniards during the Colonial period in the Americas. (Engraving by Theodor de Bry, from Bartolome De Las Casas's 'A Short Account of the Destruction of the Indies')

Another chapter of the colonization of the Americas by Europeans, represents the "the transatlantic African slave trade". Millions of people were forced and removed from Africa and brought to the "New World". The historical record of the African slave trade is based on scattered sources by the time (e.g. legislative notes, trade transactions, journals of slavers, etc.), making it difficult to account for the entire history (Morgan 1997). It is well known by historians that the slave trade began with the Spanish incursions into the Caribbean (Palmer 1995). As the records on such matter are scarce, another strategy to shed light on such events is to infer individual's origins and ancestry by applying the molecular technology (Lao et al. 2008).

Research has provided insights showing that genomic data can be used to trace the genetic ancestry of long-dead individuals, a finding that has important implications for archeology, especially in cases where historical information is missing (Schroeder et al. 2015). In this particularly scenario, poorly preserved DNA were analyzed from enslaved Africans in the Caribbean to determine where in Africa the individuals likely lived before they were captured; determining with high specificity the ethnic origins of such individuals. These analysis show the opportunity to understand the patterns of the trans Atlantic slave trade, offering new clues about the general practice of historical research (Schroeder et al. 2015).

Final remark

When two previously isolated populations make contact, the result has not only cultural, social and political consequences, but also can have a significant impact on health and disease. The moment when Christopher Columbus' fleets landed in the "New World" represents a turning point in the history of the Americas. The cultural and demographic encounter between the Old and New World had catastrophic consequences that caused the complete disruption of the complexity of pre-Columbian Americas.

C) Ancient DNA: Potentials and pitfalls of aDNA analyses

Ancient DNA (aDNA) refers to the retrieval/isolation and analysis of the genetic information from ancient/old material from historical or pre-historical sources or museum specimens. Some researchers have defined aDNA as any bulk or trace of DNA from a dead organism or parts thereof, as well as extracorporeal DNA from living organisms (Herrmann and Hummel 1994). Results arising from this field of research have helped to answer long-standing questions about evolutionary biology, human evolution, species extinctions, climate change, domestication processes and phylogenetic relationships of a variety of extinct and extant species.

Research on aDNA started in 1984 with the extraction and characterization of genetic material from a museum specimen known as Quagga (*Equus quagga*) kept at the Museum of Natural History in Mainz, Germany. This pioneering work reported the retrieval of the first aDNA sequence from an extinct species missing since 1883 (Higuchi et al. 1984). The reported 228 bp DNA fragment was compared with closely related species, determining that this extinct mammal was indeed phylogenetically related to modern zebras.

The following year, Svante Pääbo, a Swedish researcher at Uppsala University, reported a 3,400 bp DNA fragment from a 2,400 year old Egyptian mummy (Paabo 1985). The use of molecular cloning techniques revealed a repetitive sequence from human nuclear genome. However, the result of this work did not withstand scrutiny of today's criteria of authenticity and are now considered as modern DNA contamination rather than authentic DNA originating from the mummy (Del Pozzo et al. 1989). Nevertheless, the significance of both studies lies in the fact that DNA can potentially survive over long periods of time after death and be analysed.

These two seminal studies were performed in the "pre-PCR" (Polymerase Chain Reaction) era, and the amplification of aDNA sequences was based on bacterial cloning followed by Sanger sequencing of the cloned fragments (Sanger et al. 1977). However, the limitation of that approach was the lack or absence of reproducible results, which made it impossible to validate the authenticity of aDNA sequences (Paabo 1989). Moreover, contamination with "exogenous DNA", which refers to any DNA molecules that originate outside the study organism, was not considered at the time or was of little concern in the early days of aDNA work.

With the development of the Polymerase Chain Reaction (PCR) in the mid-eighties (Saiki et al. 1985; Mullis and Faloona 1987) began the "PCR era" for DNA research. The PCR made it possible to perform the amplification of DNA from very few or even a single starting molecule to obtain thousands or even millions of DNA copies at the end of the process. Since then, PCR became a methodological breakthrough in molecular biology and still represents the standard technique for DNA research.

The PCR methodology was soon also applied to the aDNA field, with the result that the field of aDNA research experienced a booming phase (Paabo and Wilson 1988; Paabo 1989). Despite the fragmentary nature of aDNA, the PCR allowed the amplification of millions of copies of aDNA, however, PCR is a stochastic process that depends solely of the quality and amount of target DNA subject to analysis.

The physico-chemical properties of aDNA and the process behind its '*post-mortem'* modification were investigated/studied by (Paabo 1989), inspired by Lindahl's pioneering work (1993). Pääbo analysed several sets of samples from different origins and ranging through time and observed that genetic material in ancient samples was modified by hydrolytic and oxidative processes (Paabo 1989; Lindahl 1993).

This study also reported that due to the great specificity of PCR, contemporary or exogenous DNA contamination was present in all ancient samples, an observation which formed the basis for the most inherent problems in aDNA research (Paabo et al. 1989).

Depending on the source of the aDNA sample (and its handling history), it is likely that endogenous and contaminant DNA will be co-extracted and amplified at the same time, leading to false positives in very extreme cases when authentic endogenous DNA forms the minority of molecules or is not preserved at all. This observation led to suggest a set of criteria for DNA authenticity in order to better assess the potential effects of contamination (Paabo et al. 1989).

Reanalysis of the Quagga mtDNA showed that two of the substitutions detected in the previously reported sequences from cloned fragments were not seen in the new sequences obtained with PCR analysis (Paabo and Wilson 1988).

In the early days of the aDNA research, working with human aDNA represented seemingly unsurmountable problems since contamination with modern molecules proved to be very difficult to evaluate and disprove. Human aDNA bears the intrinsic and ubiquitous peril of human contamination due to the similarity of DNA from worker (i.e. any persons handling the sample) and the sample itself. Therefore, any results obtained from human aDNA must withstand a set of authentication criteria, especially for work with human material (Gilbert et al. 2005). Continuously improved techniques and more stringent protocols help to discriminate between endogenous DNA and modern contamination (e.g. the software "Map damage" (Ginolhac et al. 2011)).

Not surprisingly, aDNA researchers preferably focussed on non-human organisms, in particular those that had gone extinct. Soon the field regularly reported aDNA sequences retrieved from extinct animal species, e.g. DNA from mammoth bones (Hagelberg et al. 1989; Hagelberg et al. 1994), the marsupial wolf (Thomas et al. 1989), the moa - a giant flyless birds from New Zealand (Cooper et al. 1992), and American ground sloths (Höss et al. 1996a). At the same different sources of biological material besides bones and teeth were explored for potential retrieval of aDNA. Examples include coprolites from a giant North America ground sloth (Poinar et al. 1998) or soft tissues such as brain tissue (Doran et al. 1986), and hair and nail (Bengtsson et al. 2011). The importance of hair shafts in particular as reliable source for aDNA was demonstrated by Gilbert with the subsequent sequencing of complete mitochondrial genomes (Gilbert et al. 2007).

Following the initial enthusiasm that the PCR-technique had unfolded in the field of aDNA research, some aDNA studies explored the upper temporal limits of DNA preservation. As a consequence aDNA sequences were reported from a variety of exceptionally/extremely old extinct species, e.g. a *cytochrome b* DNA sequences from a Cretaceous Dinosaur that had vanished 80 million years ago (Woodward et al. 1994). Other studies claimed to have sequenced DNA recovered from insects trapped in amber (DeSalle et al. 1992; Cano et al. 1993). However, all these rather sensational findings were later disproven showing that all of them were in fact products of contamination and/or sequencing artefacts (Hedges and Schweitzer 1995; Zischler et al. 1995; Austin et al. 1997).

Due to a series of 'high profile' studies, where could be shown were based on inconsistent or erroneous results, the emerging aDNA field was soon discredited and experienced a serious phase of scepticism and disregard by the research community. This was not surprising given the uncertainty and unreliability of the outcomes; an instance which could have led to the end of the young field.

Nevertheless, despite the lasting controversy and the 'bad days' of aDNA research, one positive effect was the establishment of strict "authenticity criteria" in an effort to guarantee that endogenous DNA / authentic DNA could indeed be retrieved from reliable ancient sources. Since then, several guidelines for authentication of aDNA have been suggested to allow peers, reviewers and general readers to assess the overall quality of aDNA results (Cooper and Poinar 2000; Hofreiter et al. 2001; Poinar 2003; Pääbo et al. 2004; Gilbert et al. 2005; Willerslev and Cooper 2005).

Substantial research efforts have focussed on understanding the characteristics of the ancient DNA molecules (Gilbert et al. 2003; Briggs et al. 2007; Brotherton et al. 2007). Moreover, factors related to preservation, survival and DNA damage has been explored in detail with the help of NGS methodologies (Ginolhac et al. 2011; Allentoft et al. 2012; Overballe-Petersen et al. 2012). Today even after three decades of aDNA research, contamination still remains the principal obstacle to overcome. With the advent of new technologies (e.g. Next-Generation Sequencing – NGS) techniques and laboratory protocols have improved to better prevent or minimize, monitor and detect aDNA contamination.

The DNA molecule

To better characterize the molecular processes resulting in '*post-mortem'* DNA degradation in ancient samples it is important to understand the chemical structure of DNA and its properties. The primary structure of DNA (Deoxyribonucleic acid) being a double-stranded helix molecule was first described by (Watson and Crick 1953) by means of X-ray diffraction. The DNA molecule can be explained as a polymer composed of a pentose sugar, a 2' deoxyribose, a phosphate group and four nitrous bases. The bases which form part of the DNA are two purines: Adenine (A) and Guanine (G) and two pyrimidines: Citosine (C) and Thymine (T). Purines and pyrimidines are classified as the two kinds of nitrogen-containing bases. Purines have a double-ring structure and pyrimidines only a single-string structure.

These bases are linked to phosphorylated sugars by means of glycosilic bonds. The phosphorylated sugars are 2'deoxiribose units linked to each other by means of phospodiester bonds. Hydrogen bonds between purines and pyrimidines adhere to a strict complementarity on opposite strands, always matching A with T only (and vice versa) by two hydrogen bonds and G with C only (and vice versa) via 3 hydrogen bonds. The ends of each of the strands of DNA are called 5'-P (phosphate) and 3'-OH (hydroxyl) in the deoxyribose. The two strains are aligned in parallel, but in opposite directions (both in 5 '→ 3' directions so that they reverse complement each other). In other words, since the interaction between the two strands is determined by the hydrogen bonds between the bases, this means that both DNA strands are located in an antiparallel sense (Figure 9).

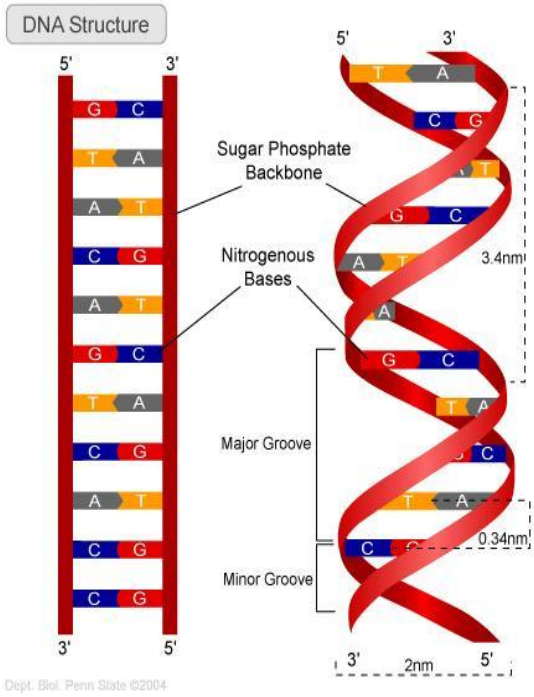


Figure 9. Structure of DNA. source: http://theinvestigation.yolasite.com/dna-structure.php

DNA degradation

After the death of an organism, biological molecules undergo severe '*post-mortem'* changes. Cellular instability starts immediately because metabolic pathways and especially the DNA repair mechanisms are no longer effective. Without these, DNA molecules become prone to degradation and undergo several physicochemical modifications. Endonucleases drive complex autolytic and enzymatic processes resulting in the breakdown of DNA molecules (Lindahl 1993). At the same time, external factors such as bacterial, fungal or insect colonization (Eglinton and Logan 1991), or

environmental factors such as heat, humidity, physical and chemical processes can also considerably affect the structure of DNA. Overall, the degradation process affects primarily the DNA backbone, and introduces a series of chemical modifications resulting in strand breaks, which leads to base modification and loss of very valuable information (i.e. nucleotide sequence), that will later affect the DNA retrieval and sequencing.

Ancient DNA characteristics

The study of DNA degradation has resulted in detailed descriptions of possible types of DNA damage, and the chemical properties of DNA have led to predictions as to how DNA molecules extracted from ancient tissues would look like under certain conditions (Lindahl et al. 1993). Direct evidence from aDNA studies have confirmed many of these observations and have led to the characterizations of various categories of DNA damage and bio-chemical alterations of (parts of) the DNA molecule, such as fragmentation of the DNA strands, blocking lesions, nucleotide modification and PCR inhibitors (Pääbo et al. 2004):

a) *DNA fragmentation*

Analyses of aDNA have shown that it is only possible to reliably amplify DNA fragments in the size range smaller than of 100-500 base pairs (bp) (Paabo 1989; Hofreiter et al. 2001). It has also been shown that there is an inverse relationship between amplification efficiency and length of the amplification products (Malmstrom et al. 2005), in clear linear relationship with the state of preservation of the sample. Therefore, amplification of aDNA ranging above 500 bp is likely to represent exogenous contamination (Paabo 1989).

b) *Blocking lesions*

Blocking lesions generally describe chemical alterations/modifications of nucleotides in the DNA strands that prevent amplification and sequencing of molecules by PCR.

-*Hydrolytic Damage:*

The DNA molecule is particularly prone to hydrolytic damage related to the presence of water. Hydrolytic processes are also responsible for breaks of phosphodiester and glycosidic bonds, which leads to the generation of baseless (i.e. abasic) sites, which are prone to strand-breaks and ultimately lead to DNA fragmentation (Lindahl 1993)(Figure 10).

The glycosidic bond is particularly susceptible to base protonation, this process is also known as depurination with a consequent formation of an apurinic/apyrimidinic site (AP site) (Lindahl and Nyberg 1972). In general apurinic sites are much more common than apyrimidic sites and can generate the break down the DNA molecule resulting in a DNA single-strand nick (Lindahl 1993).

-Oxidative Damage:

The oxidative damage occurs mainly though the action of free radicals (e.g. hydroxyl radicals -OH, peroxide radicals $-O_2$ and hydrogen peroxide H_2O_2) (Lindahl 1993). Bacterial or fungal process could be associated with the production of such radicals and oxidative processes mainly affect the double bound of purines and pyrimidines.

The phosphodiester bond is also prone to break down due to oxidative damage mainly because of the absence of 2'OH group in the ribose unit and following a break down generating a series of single-stranded nicks in the molecule (Lindahl 1993)(Figure 10).

c) Nucleotide modification

Ancient DNA sequences are subject to post-mortem damage in the form of nucleotide misincorporations. The main modifications described in aDNA are the deamination of nitrous bases. For example, deamination of Cytosine (C) to Uracil (U) an analogue base of Thymine (T) along with the deamination of Adenine (A) to hypoxanthine (HX), an analogue of Guanine (G) (Friedberg et al. 1995; Hofreiter et al. 2001). Such a DNA base modifications can lead to changes in the crucial information of the molecule, generating potentially misleading DNA sequences. Moreover because of the complementary of DNA strands, it has been noticed that the process of deamination of cytosine (C) leads to artificial C to T transitions (Lindahl 1993) but, if the complementary strand is sequenced, then it will be read as a G to A transition. Similarly, the deamination of A may be observed as change from A to G, or as T to C transition in the complement strand.

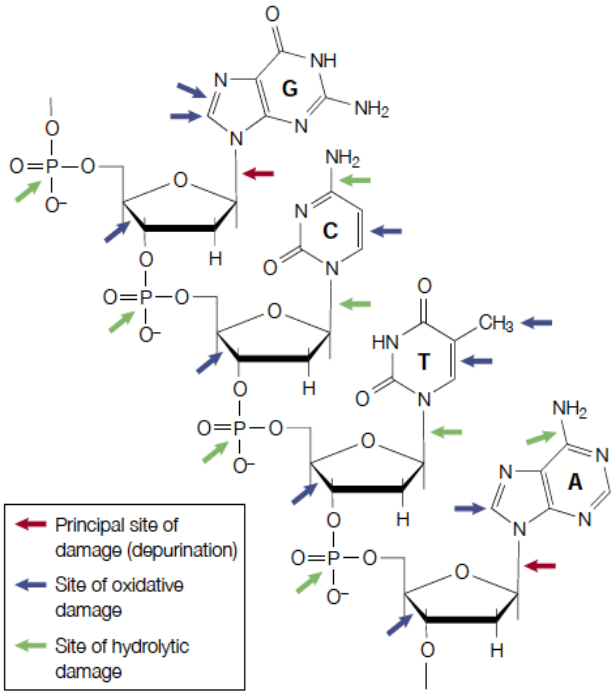


Figure 10. DNA damage in aDNA. Sites susceptible to hydrolytic attack are indicated by green arrows and those prone to oxidative damage by blue arrows. G, guanine; C, cytosine; T, thymine; A, adenine (After Hofreiter et al. 2001).

d) PCR inhibitors

Samples for aDNA analysis can come from a range of diverse sources (e.g. archaeological excavations, burial sites, ice or sediment cores, soils, museums, etc.) and are often contaminated with environmental substrates.

While minute amounts of DNA can be selectively amplified by PCR, this enzymatic step is also sensitive to inhibition, which can be caused by agents that are co-extracted alongside the DNA molecules.

These so-called "PCR inhibitors" are in general not associated with the ancient samples themselves, but form an inherent part of the environmental surrounding from which ancient material had been obtained. PCR inhibitors are mainly soil components or derived from products of soil degradation (e.g. fulvic, humic acids, tannins or complex iron molecules)(Paabo, 1989).

The exact mechanism of inhibition is not well understood, but it is assumed that the agents interact in a non-specific way leading to a loss or substantial decrease of the *Taq polymerase* activity during the PCR.

Other inhibitor are so-called Maillard products, which are derived from organic reactions such as sugar reduction/condensation (Willerslev and Cooper 2005) or which DNA is bound, or DNA derived from microorganisms, that is co-extracted with the target DNA, and could potentially contain competing primer binding sites.

e) "Jumping PCR"

A coding or blocking lesion in the DNA template can cause the *Taq polymerase* to discontinue strand elongation during PCR amplification. However, the partly synthesized strand can act as primer in the next PCR cycle, where the polymerase resumes on another DNA template, therefore producing a mixed strand of two potentially different templates. This process results in "chimeric" sequences, which could not only be misinterpreted as novel genetic variant, but could also generate drastically misleading phylogenetic information if not recognized as such (Pääbo et al. 2004; Willerslev and Cooper 2005).

Survival of DNA molecules

If we consider the process of DNA degradation as continue over time, theoretically all DNA molecules will be damaged, become more and more fragmented and will eventually too small to detect or will disintegrate entirely. However, contrary to this assumption, if degradation processes come to a halt, DNA molecules can survive/remain preserved under specific circumstances (i.e. particular environmental conditions). Such conditions include, low temperatures, fast inactivation of the nucleases activity, inhibition of micro-organisms or fast desiccation (Höss et al. 1996b). Other conditions include high salt concentration and neutral pH.

Temperature is probably the most important environmental factor in the preservation of the genetic material, since physicochemical reactions responsible for DNA damage take place with low rate at low temperatures (Höss et al. 1996b; Willerslev et al. 2003). Under such favourable conditions it is estimated that DNA molecules can survive for several millennia (Lindahl 1993). However, a study of modern papyri from Egypt with an age-span from 1,300-3,200 years BP showed that the DNA half-life in papyri is as short as 19-24 years (Marota et al. 2002).

It has been estimated that 100,000 years is the time before the DNA molecule will completely break down and will no longer be detectable/amplifiable by laboratory

methods (Lindahl 1993; Hofreiter et al. 2001). Consequently, Pääbo and colleagues estimated that all aDNA sequences older than 1 million years should be considered as artefacts (Pääbo et al. 2004).

A study focussed specifically on survival time of the aDNA used remains from 158 moa species and found that the total DNA from the remains of the moa is halved over 521 years when stored at a temperature of 13.1 degrees (Allentoft et al. 2012). When extrapolated over time, they concluded that DNA in bone could survive for 6.8 million years if kept constantly at a temperature of five degrees below zero.

Recent studies from exceptional sites have exceeded the 100,000 years mark. Ancient DNA from permafrost environments has yielded the oldest records for DNA survival for plant and invertebrate species with a date of approximately 800,000 years (Willerslev et al. 2003). Also, a study reporting aDNA from a 700,000-year-old Middle Pleistocene horse has dramatically extended the known limit of DNA survival for mammalian species (Orlando et al. 2011). New technological approach based on targeting ultrashort DNA fragments of DNA have made possible to report the oldest aDNA sequences outside the permafrost recovering the mtDNA genome of a <300,000-year-old from a cave bear in Spain (Dabney et al. 2013) and an early hominin remains ranging in age from the early Pleistocene to the Holocene from Sima de Huesos – 'pit of bones' (Meyer et al. 2014).

Contamination of DNA

While '*post-mortem'* DNA damage and degradation pose technical and analytical challenges, research on aDNA also encounters a series of experimental difficulties in the form of the ubiquitous peril of contamination by external modern DNA molecules. Exogenous or modern DNA contamination can be co-extracted and processed alongside the targeted endogenous DNA at all stages of the experiment, which could potentially lead to disputed outcomes and in the worst case scenario to false positive results if the analytical framework and laboratory protocols are not properly set up. Despite all precautions to minimize DNA contamination in aDNA analysis, it remains likely that contamination can occur at some step (Gilbert et al. 2005).

a) <u>Sample collection and handling</u>

   Correct handling of samples before and during sample collection is not only the first but also one the most important step in aDNA studies, on which most of the

ultimate success of aDNA retrieval depends. However, accession and collection of specimens cannot always be performed under ideal contamination-free conditions, or are difficult to monitor or maintain. This is of particular importance when archaeological samples have been excavated many years or decades ago, and have since been held at museum collections (Malmstrom et al. 2005). Therefore, the post-excavation history and archaeological context are fundamental to evaluate preservation and handling of samples for selection of reliable material (Gilbert et al. 2005).

In the best-case scenario, archaeologists and anthropologists trained in ancient DNA sampling are the first to recover '*in situ'* ancient specimens (i.e. freshly samples from and during archaeological excavations). Several guidelines to perform a careful collection in order to avoid any possible contamination have been proposed (Brown and Brown 1992; Yang and Watt 2005). The compulsory use of protective gear (e.g. face masks, face shields, full body suits, sterile latex gloves, etc.) during sample collection is recommended in order to minimize the risk of contamination. It has been shown that freshly excavated specimens are ideal in terms of preservation (Pruvost et al. 2007). Another important point in reducing contamination is to avoid contact of samples with water and do not wash them. If samples were collected from damp environment, it is important to either freeze or dry them to prevent microbial growth. It is generally recommended to store samples appropriately, either kept refrigerated or under cool conditions before analysis.

The use of preservatives of other kind of substances such as hardeners should be postponed until after the sampling as these might interfere with or cause inhibitions during the PCR. If feasible, it is recommended to genotype all people involved in peri- and post-excavation handling of the samples and in the sampling step itself. This will allow aDNA researchers to compare potential contaminants with genetic profiles of those persons who handled the samples and/or museum collections.

b) <u>Laboratory Procedures</u>

Ancient DNA analysis is only possible in a controlled environment. It is imperative to carry out aDNA research exclusively in purpose-built facilities. The aDNA laboratory must be separated from any modern and post-PCR laboratories.

Ancient DNA laboratories should be equipped with a positive air pressure system to reduce contamination from the surrounding environment, while spatial separation of all work areas and experimental steps is recommended as a standard procedure (Fulton 2012; Knapp et al. 2012). Before entering the laboratory, aDNA researchers must adhere to a set of protocols in order to keep contamination at a minimum (Paabo 1989; Cooper and Poinar 2000; Willerslev and Cooper 2005; Yang and Watt 2005). A mandatory measure to perform aDNA lab work is the use of protective gear (e.g. face mask, face shield, full body suit, sterile gloves and gum boots) (Fulton 2012). The use of chemical oxidants such as bleach and UV light irradiation is recommended as to decontaminate the sample surface (Yang and Watt 2005). Regular cleaning steps on every surface and work benches before and after any experiment must be performed. Also work areas must be cleaned with 3-5% bleach to destroy any potential contaminant followed by wiping with ethanol and nightly UV irradiation. Additional and thorough weekly cleaning is necessary to maintain a "DNA-free" work environment (Fulton 2012).

Regarding reagents and consumables, it is important to follow the same procedures and despite the fact that some are labelled as sterile or certified as DNA-free materials, all disposables and equipment must be decontaminated before use. Moreover, is important to keep a strict one-way system of consumables and disposals once they entered to the laboratory.

Criteria of authenticity

All features of aDNA presented above (fragmentation, PCR inhibition, nucleotide modification and contamination), can eventually pose problems in the subsequent analysis and interpretation of genetic data. Criteria of authenticity are based on a combination of precautions, interpretation of aDNA based on expected biochemical properties, and common sense. Ancient DNA researchers have suggested guidelines to standardize protocols and to validate the results (Cooper and Poinar 2000; Poinar 2003; Pääbo et al. 2004), although some of them are no longer applied.

Top 10 criteria of authenticity for DNA for ancient and degraded samples

Modified after (Poinar 2003)

1	**Physically isolated work area** Given the susceptibility of ancient material in each step of the analytical process, samples must be analysed in a purpose-built facilities equipped with positive air pressure inside, HEPA filters and work surfaces and specimens should be radiated with ultraviolet light and oxidant agents. All post-PCR steps must be performed in a separate laboratory.
2	**No template controls** The use of no template controls in extraction and amplification steps, as well as library preparations. Such controls contain all reagents, except the DNA template. Any positive result of negative controls is indicative of the introduction of contaminating DNA during the experiment.
3	**Appropriate molecular behaviour** Since DNA fragmentation only allows DNA amplification of fragments less than ~200-300 bp in size, larger products must be interpreted with caution. Recommended is the verification of the size of PCR molecules. We expect an inverse relationship between the number of surviving/amplified DNA fragments and the length of aDNA sequences.
4	**Quantitation** Measuring the amount of starting molecules of DNA by real time PCR is critical to assess whether or not an observed mutations in a sequence might result from one DNA copy with *post-mortem* oxidative modifications. Generally, when the number of starting DNA molecules exceeds > 10000 the probability of all sequence having a damaged base at the same site becomes low.
5	**Reproducibility** To achieve a consistency in the experiments, it is important to replicate results if samples are available. As a general rule, it is recommended to analyse at least two independent samples coming from the same source.
6	**Cloning** In the PCR era of ancient DNA research, cloning of the amplified aDNA was recommended to assess the state of initial degradation of the DNA molecules, detect

	contamination, and possible non-specific binding of the primers in the PCR reaction. Cloning allows detecting sequence heterogeneity present in the amplification products and can help identifying whether it is cause by contamination, molecular modifications and / or errors of the PCR polymerase.
7	**Independent replication** Perform the experiments in other research laboratories independently. The idea is to eliminate the possibility of systematic contamination in one lab. This is very important for studies on human remains, and when novel or unexpected results are observed.
8	**Biochemical preservation** This criterion uses independent observation of the survival of other biomolecules as indicator for the likelihood of survival of DNA molecules. Its purpose is to establish whether the general state of sample preservation is good enough to reasonably expect the preservation of authentic DNA molecules.
9	**Associated remains** The processing of associated faunal remains from the same site/location can help in assessing the overall chances of DNA preservation at a given site, and potentially monitor contamination (e.g. detection of human DNA in non-human species).
10	**Phylogenetic sense** Results should fall in appropriate places on a phylogenetic tree. The rationale is that very old samples are expected to fall onto basal branches (or even assume out-group status) and not on terminal twigs with the tree. When phylogeographic relationships are known, haplotype status is often an indicator for contamination. For example, given the known peopling history it is highly unlikely to observe typical European haplotypes in Pre-Columbian samples from the Americas.

The authentication of aDNA and assessment of contamination by external sources remains an issue in aDNA research despite the great major advances in aDNA analysis in the last years. Next-Generation Sequencing (NGS) techniques applied to aDNA requires a new set of guidelines to work with aDNA. In that regard, Knapp and colleagues, have proposed a new set of authenticity criteria to deal with the fundamental changes in sequencing strategies introduced by NGS (Knapp et al. 2015).

Ancient DNA: Applications and perspectives

The analysis of aDNA allowed archaeologists, anthropologists, palaeontologists, and geneticist to reconstruct the evolutionary history and systematics of a wide range of extinct and extant species, and interpret events in the past. In Table 1, I list a (subjective) selection of significant milestone studies resulting from aDNA research.

The ability to retrieve and characterize aDNA from ancient sources and materials has been evolving during the last decades resulting in continuously improving laboratory protocols and data analysis tools, otherwise it would not be possible to perform aDNA research without the developed, improvements and breakthrough aDNA methods – for a review see (Rizzi et al. 2012; Ermini et al. 2015).

a) Evolutionary and phylogenetic relationships of extinct and extant species

From the first study in aDNA research, the analysis of the Quagga museum specimen (Higuchi et al. 1984), which clarified its phylogenetic relationship to plain zebras, aDNA has helped to elucidate phylogenetic and taxonomic positioning of many species. Many such studies contributed to our understanding of species evolution and – to a degree – their extinctions. A famous example for the systematic classification of an extinct species through means of aDNA analyses is the Tasmanian tiger *(Thylacinus cynocephalus)*, which was found to be closer to Australian marsupials than to South American marsupials (Krajewski et al. 1997), suggesting an independent evolution of marsupials within the two continents. The moa, a giant flightless bird species from New Zealand was determined to be related more closely to Australian birds than to Kiwis from New Zealand (Cooper et al. 1992). Moreover, a recent analysis of aDNA extracted from two Madagascar elephant birds, has revealed its close genetic relationship with the kiwi, despite large differences in morphology and geography (Mitchell et al. 2014).

Coprolites from extinct ground sloths were genetically analysed to gain information about their diet and ecology of this species in the Americas (Poinar et al. 1998). The analyses of mammoth bones from the Pleistocene present another interesting case study, for which a phylogenetic relationship with the African elephant was first suggested (Yang et al. 1996). However, the analysis of complete mitochondrial genomes (Krause et al. 2006; Rohland et al. 2007) and

nuclear DNA finally solved this issue, determining that mammoths were indeed more closely related to Asian elephants (Miller et al. 2008).

Another important aspect of aDNA analysis is the ability to correlate species relationships with temporal and/or geographical variables in order to see how species evolved or expanded across space and time. Such an approach was used to obtain a view of population dynamics of Pleistocene bears in Alaska (Leonard et al. 2000). Later results clarified the ecology, distribution and relationship with extant bear species, and the genetic relationship with polar bear species (Miller et al. 2012).

Since its inception a huge variety of extant and extinct species have been characterized by aDNA, including hyenas (Rohland et al. 2005; Sheng et al. 2014), *Myotragus,* a cave goat from the Balearic Islands (Lalueza-Fox et al. 2000; Lalueza-Fox et al. 2005a; Ramirez et al. 2009), wild aurochs species (*Bos primigenius*) (Zeyland et al. 2013), chicken (Thomson et al. 2014) to name only a few – reviewed in (Hofreiter et al. 2001).

With the ability to generate massive high resolution data set through Next-Generation Sequencing approaches, many of these question have been revisited in recent times, leading to further clarification, especially where the ancient evidence was relying on mitochondrial DNA. Today, many complete mitochondrial and nuclear genomes are available for extinct species, which helped to track changes in evolutionary and population changes over time and resolved remaining phylogenetic questions. Accessing high-resolution data allows aDNA researchers to reconstruct past and present population dynamics of extinct species (Miller et al. 2009; Orlando et al. 2011; Dabney et al. 2013; Mitchell et al. 2014).

b) Molecular paleopathology

Ancient DNA studies have shown that is possible to retrieve DNA not only from the host (i.e. the organism carrying a disease), but also from the pathogens associated with a particular disease, in which case there must be traces of viral or bacterial infections.

The alleged survival of DNA from microorganisms (bacteria, fungus, viruses) has been reported and discussed in several publications, – for a review see (Harkins and Stone 2015). To study diseases that affect by human populations in the past,

to understand human-pathogen interactions through time and the role played by microorganisms in the history of human groups, it is important to assess its origin, prevalence and evolution (Harkins and Stone 2015).

In screening and studying pathogens in past populations, the field of molecular paleopathology has reported interesting findings. For instance, the tuberculosis – TB – causing agent *Mycobacterium tuberculosis* has been detected in pre-Columbian mummified remains (Salo et al. 1994; Arriaza et al. 1995; Braun et al. 1998), in East Central Europe (Faerman and Jankauskas 2000; Haas et al. 2000; Fletcher et al. 2003), and in Egyptian mummies (Zink et al. 2001; Zink et al. 2003; Crubezy et al. 2006).

With the advent of NGS technology, a recent study on tuberculosis has postulated a novel hypothesis regarding the origin and spread of the disease. By analysing human remains from South America, researchers found that the ancient TB strains from Peru did not resemble human-adapted forms, but were very similar to strands from *Mycobacterium pinnipedii*, which are adapted to seals and sea lions. These findings suggest that seals might have picked up the disease in Africa and carried it across the ocean to pre-Columbian Americas, meaning that tuberculosis was present before the European contact (Bos et al. 2014).

The presence of *Yersinia pestis* was reported by (Drancourt et al. 1998) by analysing teeth extracted from skeletons excavated from 16th and 18th century French graves. This particular pathogen, which is thought to have caused the Black Death (1348-1350 AD), one of the most devastating pandemics of plague in human history, was later studied by using NGS methods retrieving the genome from plague victims from medieval mass graves (Bos et al. 2011; Schuenemann et al. 2011). These studies suggest that ancient *Y. pestis* strains do not differ from modern-day bacterial phenotypes in terms of virulence; however as the causative agent of the Black Death, other factors such as environment, vector dynamics or host susceptibility might have played a role in the spread of such pandemics in medieval times.

A very recent aDNA study identified the oldest evidence of *Y. pestis* in human teeth from Asia and European individuals ranging from 2,800 – 5,000 years ago.

The results suggest that plague infections in the past might be endemic of Eurasian populations, pushing back the origins of the disease by some 3,000 years (Rasmussen et al. 2015b).

From an historical and epidemiological point of view the origins of pathogen or epidemics is often unclear or debated. For example, a European or American origin of the syphilis disease (caused by *Treponema palladium*) had been suggested, and thus been studied in ancient samples (Kolman et al. 1999; von Hunnius et al. 2007). Recent investigations showed that syphilis was introduced to the New World by the Spanish (de Melo et al. 2010) based on the finding of the oldest reported case of syphilis in the Old World, suggesting a European origin of the disease (Montiel et al. 2012).

Dental calculus – calcified plaque, has received particular attention in recent years as a special type of outlasting sample that traps and binds DNA during its formation. With the study of microbial DNA from dental calculus it has been possible to generate important insights into the diet and cultural changes, and the emergence of new pathogens related to the introduction of agriculture and the associated changes in microbial diversity it brought about (Adler et al. 2013; Warinner et al. 2014).

However, ancient samples with highly degraded DNA in which pathogenic taxa represent often a minute component, performing a sequence-based metagenomic characterization for these kinds of studies represents a costly and time consuming approach. Latest methods to analyse pathogens have incorporated the use of the "Microbial Detection Array technology", tested successfully with previous ancient pathogens/samples, showing that a less expensive and more efficient paleo-pathological screening is available, without the lengthy analysis associated with high-throughput sequencing. Based on this, it is possible now to obtain an informative "snapshot" of microbial diversity in complex samples (Devault et al. 2014). Further research will focus on the mechanisms and processes of host/pathogen interactions to better understand pathogen evolution and related functional adaptations through time.

c) Hominin evolution

The evolution of archaic forms of humans poses a key question in aDNA studies, as the information gained provides insight into the evolutionary puzzle concerning the origins and dispersals of anatomically modern humans (AMH) and their interaction with archaic humans. The most iconic of these, the so-called

Neanderthals, were a group of humans that existed in Western Asia and Europe until they became extinct around 30,000 years ago (Hublin 2009). In an evolutionary context, Neanderthals represent human's closest relatives.

The breakthrough for aDNA research came with the retrieval of the first Neanderthal sequences from the Feldhofer-Cave specimen from the Duessel valley in Germany (Krings et al. 1997), which opened up avenues to explore the genetic relationship of AMHs and extinct hominids. The successfully amplified and cloned sequence of the mtDNA – hypervariable Region I (HVR-I) showed that modern humans and Neanderthals shared some similar genetic characteristic. However, there were enough differences in the Neanderthal mtDNA sequence that suggested that Neanderthals fell outside the genetic variation from modern humans (Krings et al. 1997). This results was later confirmed with sequences from HVR-II from the same individual (Krings et al. 1999).

Since this first report many more Neanderthal samples were screened by several research teams, leading to sequencing of complete (Green et al. 2008) and partial mitochondrial genomes (Ovchinnikov et al. 2000; Lalueza-Fox et al. 2005b; Caramelli et al. 2006; Briggs et al. 2007; Burbano et al. 2010).

With the advent of NGS techniques it was also possible to retrieve genomic data, which allowed answering the all-important question whether or not there was gene-flow between AMHs and Neanderthals (Green et al. 2010; Prufer et al. 2014).

When compared to a set of worldwide modern populations, the study found signals of 1-4 % Neanderthal genetic contribution in all present-day non-African human populations (Green et al. 2010).

Upon screening of more samples identified as potential Neanderthal, researchers from the Max Planck Institute in Leipzig found the sequence of an unknown hominid. This archaic hominid was later named as Denisovan after the Denisova cave in the Altai Mountains of Southern Siberia where its remains were found. The archaeological site contained unique cultural layers that indicate presence of archaic humans from 280,000 years ago onwards. The finger bone from which the unknown sequence was derived was found in a layer dated between 50,000 and 30,000 years ago. With no fossil record other than and a small fragment of a finger phalanx and two molars, and no phenotypic information, the sequence analysis shows that these elusive individuals represents an archaic human group

closely related to Neanderthals and modern humans (Reich et al. 2010; Meyer et al. 2012). In consequence, Denisovans added another layer of complexity to the puzzle of human evolution (Reich et al. 2010). A comparison between Neanderthal, Denisovans and modern human genomes, revealed a basal genetic signature shared by the three forms. It has also been shown that a contribution of 5% of Denisovan ancestry can be seen in today's populations from Papua New Guinea and the Pacific, suggesting independent events of interbreeding between archaic forms of humans (Reich et al. 2011). More recently, aDNA of hominin remains from Sima de Huesos – 'pit of bones'– in Spain revealed that it is closely related to the lineage leading to mtDNA genomes of Denisovans (Meyer et al. 2014).

d) Human evolutionary history

The study of human origin and dispersal has received special attention in aDNA research in order to better understand the processes of migration that led to the colonization of almost the entire world. The generation of genomic data, as reported in many recent studies – reviewed in (Ermini et al. 2015), bears the most promising prospect to evaluate and reconstruct relationships from past and modern-day human populations, and to reconstruct critical events in human evolutionary history.

From the earliest works in human populations studying the colonization of Japan (Horai et al. 1996; Oota et al. 1999), the peopling of the Pacific Islands (Hagelberg and Clegg 1993), population genetics studies of prehistoric Italian populations (Vernesi et al. 2004), or studies performed to explain the genetic variation in Iberians and Sardinians populations (Sampietro et al. 2006; Caramelli et al. 2007), aDNA was used to investigate population affinities from ancient individuals all over the world. Studies have also targeted complete mtDNA genomes such as the Paleoeskimo from the Saqqaq culture in Greenland (Rasmussen et al. 2010), or an Aboriginal Australians (Rasmussen et al. 2011).

A series of studies addressed the origins of Native Americans populations by aDNA aiming to clarify questions about the peopling of the Americas and the subsequent history of migrations and relationships among Native American groups (Stone and Stoneking 1993; Stone and Stoneking 1998; Stone and Stoneking 1999; Gilbert et al. 2008b). This also involved sampling strategies to

address regional questions, such as the peopling of the Caribbean (Lalueza-Fox et al. 2001; Lalueza-Fox et al. 2003) or studies on ancient human mummies from Andean populations in South America (Shimada et al. 2004; Moraga et al. 2005; Shinoda et al. 2006; Carnese et al. 2010; Fehren-Schmitz et al. 2011; Baca et al. 2012; Fehren-Schmitz et al. 2015).

A specimen of special interest to Europeans is the Tyrolean ice mummy, a 5,300-year-old male individual from the Copper Age, known as "Ötzi", named after the site in Ötztal Alps in Italy, where he was discovered (Handt et al. 1994). The key factor for successful aDNA retrieval from this individual was preservation in permafrost conditions. Further molecular analysis of the complete mtDNA genome revealed the European mtDNA haplogroup subcluster K1ö (Rollo et al. 2006). Despite initial fruitless efforts in trying to retrieve nuclear DNA data, this was possible by applying NGS technologies as shown in a study reporting a low coverage version of the complete genome, which revealed closely affinities to populations from Corsica and Sardinia (Keller et al. 2012).

Given the temperate and cooler climates favourable for DNA preservation, aDNA studies from Europe continued to provide new insights relevant to long-standing archaeological and anthropological research questions, for example whether ancient Neolithic remains from central Europe would support a Palaeolithic or Neolithic ancestry/origin of modern-day European populations (Haak et al. 2005). Soon after, strong signals for genetic discontinuity were shown between hunter-gatherers and the first farmers during the Mesolithic-Neolithic transition in Europe (Bramanti et al. 2009; Malmström et al. 2009), while the origin of the first farmers could be traced back to the Near East (Haak et al. 2010).

This could soon be confirmed in Western and Central Europe with the generation of SNP data and mtDNA genomes from Mesolithic hunter-gatherers and early farmers (Sanchez-Quinto et al. 2012; Skoglund et al. 2012; Lazaridis et al. 2014; Olalde et al. 2014), highlighting the dramatic impact of farming during the Neolithic period and the resulting demographic changes during the transition from foraging to agricultural societies. Moreover, autosomal genome data has shed light on phenotypic traits in order to unveil specific and adaptive features (e.g. eye color, hair and skin pigmentation genes) or diet and immunity genes from ancient individuals (Olalde et al. 2014).

However, aDNA studies have been shown to be particularly powerful when applied over wider temporal windows, as so-called "transect through time" or "diachronic" studies. One such study reports ancient mitochondrial genome from 39 individuals belonging to haplogroup H, the dominant lineage in present-day Europe. This study describes the steady rise in frequency starting with the early Neolithic, identifies incoming lineages in the Late Neolithic, and more generally highlights the power of high resolution mitogenome sequencing (Brotherton et al. 2013).

Another study reports mtDNA profiles from ~360 prehistoric individuals in Central Europe and details subsequent changes in genetic composition following the initial Mesolithic-Neolithic period, which were characterized by a resurgence of hunter-gather lineages in the Middle Neolithic and another marked changeover during the transition from Late Neolithic to Early Bronze Age cultures (Brandt et al. 2013; Brandt et al. 2015).

Ancient DNA studies from other parts of the world include the report of mtDNA genome and the entire non-repetitive portion of chromosome 21 from a 40,000-year-old individual in Tianyuan cave in China. This study described the basal form of the mitochondrial B haplogroup, ancestral to present-day B haplogroup, which is common in East Asian / Native American populations today (Fu et al. 2013).

A remarkable finding was the genetic evidence from the Upper Palaeolithic Malt'a boy (Raghavan et al. 2014b), discovered in Siberia, which shows genomic signatures that link to Mesolithic populations in Western and Central Eurasia, but also to present-day Asians and Native Americans. This research provides an invaluable piece of the puzzle in the connection between Siberian populations and the initial colonizers of the Americas. A related study reporting the complete genome of the 'Anzick-1' individual from the Clovis culture in North America dated to (10,705±35 ^{14}C years BP) confirms the scenario about the peopling of the Americas from eastern Siberia (Rasmussen et al. 2014).

Another recent study from a Siberian individual named Ust'-Ishim retrieved the oldest genomic data thus far from an anatomically modern human (AMH) with an age of 45,000 years. The data reveals a non-differentiated population in northern Asia during the Palaeolithic, which offers insight into the timing and mode of out-of-Africa journey of AMHs and gene-flow with Neanderthals (Fu et al. 2014).

A landmark study, that reconciled the available ancient data with high resolution data from 2,400 modern-day Europeans, but also generated three high coverage genomes from ancient Europeans, (two hunter-gatherers and one early farmer) could show that nearly present-day Europeans can trace their ancestry to three different groups: indigenous hunter-gatherers who colonized Europe in Palaeolithic times, Middle Eastern farmers who expanded west with the rise of agriculture and animal husbandry, and a mysterious third population, related to the Siberian Mal'ta boy (termed 'Ancient North Eurasian'; ANE) that spanned across northern Europe and Siberia (Lazaridis et al. 2014).

This ancestry composition was further explored in follow up study that utilised a novel SNP capture method to generate genome wide SNP data from 69 prehistoric European ranging from 3,000 to 8,000 years old. The study could shed further light on the third component, which could be traced back to cattle herders from the Eurasian steppes, which expanded west reaching Central Europe around 4,500 years ago. Since the Eurasian steppes are one of the proposed homelands of the "proto-Indo-European" language family, the substantial genetic turnover observed in Europe 4500 years ago suggests that at least some of the Indo-European languages must have been brought to Europe with this large-scale expansion (Haak et al. 2015).

Latest research involving genetic and morphological analyses has shed light on the genetic affinity of the 8,500-year-old "Kennewick Man", America's most iconic prehistoric skeleton. The ancestry and origin of Kennewick Man has been subject of a long-held dispute, but ancient DNA evidence suggests that Kennewick Man is more closely related to Native Americans than any other population (Rasmussen et al. 2015a).

Lately, the first complete nuclear genome from a 4,500-year-old human skeleton from the "Mota" cave in southern Ethiopian highlands in Africa has been sequenced to a mean coverage of 12.5X (Llorente et al. 2015). Resulting analyses suggest that the DNA from this skeleton is different from today's Africans, and does not show signals for Eurasian backflow from populations closely related to Early Neolithic farmers that spread into Africa 3,000 years ago. This signal is otherwise widely spread across the continent. Therefore, these results highlight the importance of "unadmixed" reference baseline data to reconstruct

demographic events in population studies. Despite not very favourable climatic conditions for sample preservation for most parts Africa, this study raises hope for future with aDNA studies.

Next-Generation Sequencing (NGS)

The introduction and development of 2nd, 3rd and Next-Generation Sequencing (NGS) technologies has revolutionized the fields of molecular biology and genetic research (Margulies et al. 2005; Bentley et al. 2008). Despite the great accomplishments made possible with classical Sanger sequencing (Sanger et al. 1977), including sequencing of the first complete human genome (Lander et al. 2001; Venter et al. 2001), it had been quickly replaced by new sequencing techniques in the last decade. Next-Generation Sequencing techniques are capable of generating massive amounts of genomic information in a more efficient way, rendering the generation of genomic data by several orders of magnitude, therefore increasing the prospects for research in a variety of newest applications and benefits (Mardis 2008; Shendure and Ji 2008; Metzker 2010).

Applications of Next-Generation Sequencing technologies

The wide range of NGS applications includes data generation for *de novo* whole-genome sequencing, exome sequencing, methylation patterns of DNA, RNA analysis, epigenomics, proteomics, metabolomics, and metagenomics (Mardis 2008; Pareek et al. 2011).

a) *de novo* whole-genome sequencing

The most common application of NGS is the generation of complete genome sequence data (Wheeler et al. 2008). Next-generation sequencing platforms are capable of producing massive amount of genetic information in a single run, which has led to the development of large-scale sequencing projects such as the 1000 genomes project (The 1000 Genomes Project Consortium 2010), the HapMap project (Consortium 2005), the Human ENCODE project (Dunham et al. 2012), or the Human Microbiome Project (Gevers et al. 2012), aiming to increase the knowledge of medical genetics and how genetic variation affects health and disease in humans. Moreover, NGS has opened the possibility to explore the genomic information encoded in most of living organisms and consequently a large number of animals, plant and microbial species genomes have been

sequenced. The overall quality and degree of accuracy for a complete genome is measured in terms of "sequencing depth" or "fold-coverage" parameters (e.g. 20X), which is the average number of reads for a given nucleotide or the number of times each nucleotide is sequenced.

For ancient DNA, the quality of a genome depends on the integrity and preservation of samples. Genomes obtained from modern samples usually reach a deep fold-coverage above 40X, typically representing 99% of the overall genetic information. According to estimates it has been proposed that 8X (times) sequence coverage represents an acceptable genome quality (Millar et al. 2008). However, for poorly preserved ancient specimens, NGS has dealt with sequencing fold-coverage ranging from 0.7X from mammoth bones (Poinar et al. 2006), 1.3X from first Neanderthal genome (Green et al. 2006), spanning to 6.4X coverage of Australian aboriginal genome (Rasmussen et al. 2011) to an genomic coverage of 20X from an inhabitant of Greenland from the Saqqaq culture (Rasmussen et al. 2010). In maximising sequencing efforts, a significant increase in coverage could be achieved for the archaic hominins, including a 30X of genome of the Denisovan individual (Meyer et al. 2012) and the 52X genome from a Neanderthal from the Altai Mountains (Prufer et al. 2014).

b) Re-sequencing or targeted sequencing

Target sequencing approaches have been used to delimit and focus on a specific segment or 'informative' regions of the entire genome. Performing a 'DNA enrichment' of the targeted segment or particular loci (Mertes et al. 2011) can substantially increase the amount of genetic data for the desirable genomic region. Thus, targeted enrichment has been considered one of the most interesting methodological improvements of the NGS era (Harismendy et al. 2009; Mamanova et al. 2010). This was exemplified by a study of (Hodges et al. 2007), which used a highly-sensitive microarray method to capture desired region of the human genome (e.g. protein-coding exons). Targeted sequencing also allows the screening of common and rare genetic, and unknown variants within the target region in an organism (Ng et al. 2009; Liu and Leal 2010), and identification of new genes associated with both rare and common diseases (Ng et al. 2009). It facilitates the generation of exclusive panels of genes, either functional or phenotypic SNPs (e.g. high-altitude (Yi et al. 2010)), or genomic regions

associated to a disease, which can be updated on regular terms, and provide new insights into the complexity of regulatory human mechanisms (Biesecker 2010; Liu and Leal 2010).

Researchers investigating gene expression and functional genomics have taken advantage of this approach studying for instance evolutionary differences among species. Applying this methodology to ancient DNA it was possible to study functional variants in genes associated to speech ability in humans and Neanderthals targeting the denominated FOXP2 gene, which has shed light on speech and language acquisition in humans (Krause et al. 2007). (Briggs et al. 2009) used a technique called 'Primer extension capture' (PEC) to enrich for heavily degraded aDNA samples from a high background of microbial DNA. A more recent DNA enrichment approach used biotinylated baits in solution, sequence complete ancient mtDNA genomes (Maricic et al. 2010). This method was also successfully used to sequence five complete mtDNA genomes from Neanderthals (Briggs et al. 2009), while a novel microarray-based sequence capture method was developed to retrieve target regions from Neanderthal DNA even in the presence of substantial bacterial contamination (Burbano et al. 2010). Overall, enrichment techniques are more cost-effective, and highly sensitive to target minute amounts of endogenous DNA from ancient samples.

c) DNA Methylation.

DNA methylation is a key process in molecular development in higher organisms. Methylation patterns are studied to understand the regulation of processes such as cell differentiation, development, and disease occurrence (Bock et al. 2010). Various sequencing approaches are used to identify global patterns of methylation (methylome) with high reliability, in a short time and at minimal cost per base when compared to traditional sequencing (Harris et al. 2010).

d) Metagenomics

Metagenomics refers to the retrieval of genome sequences from a variety of organisms, especially microorganisms such as bacteria, fungi and viruses, mainly from environmental samples, with the extraction and characterization of DNA from the entire population comprising the samples. The genomic information is known as *metagenome or metagenomic DNA*, which represents the overall

genetic profile from the global set of species in such a sample. The main aim of metagenomic studies is to characterize the biodiversity derived from samples, with special focus on those species, which cannot be cultured with classical microbiological protocols. However, the resulting challenge lies in the accurate identification of the total number of organisms (and their diversity) present in a given sample, e.g. water, soil, sediments, ice cores, etc.

As a consequence, researchers working with metagenomics analyses find many practical applications to study complex environmental microbial communities and in a comprehensive manner offer the opportunity to observe how ecosystems develop and respond to e.g. environmental change (Edwards et al. 2006). The information provided by metagenomic analysis is constantly increasing and finds direct applications in industry, such water research or agencies engaged in environmental sustainability.

Ancient DNA and Next-Generation Sequencing

Ancient DNA samples were traditionally analysed using standard molecular biology methods, such as bacterial cloning, Polymerase Chain Reaction (PCR) (Mullis and Faloona 1987) and Sanger sequencing (Sanger et al. 1977). Due to the fragmentary nature of ancient DNA molecules, aDNA research was largely restricted to the analysis of short fragments of DNA, mainly from the mitochondria (mtDNA) due the higher abundance of copies per cell compared to nuclear DNA.

Amplification of DNA via PCR helped to overcome the low amounts of endogenous DNA in ancient samples, but at the same time introduced new problems, such as the potential (and often preferential) co-amplification of contaminating exogenous DNA. Traditional PCR based techniques have been unable to fully resolve the molecular characteristics of aDNA. Before applying NGS a novel single primer extension (SPEX)-based technique has provided new insight into the molecular nature of aDNA damage and fragmentation (Brotherton et al. 2007). However, recent developments in DNA sequencing methodologies applied to ancient samples were able to study characteristic aDNA damage profiles in more detail (Ginolhac et al. 2011). Nevertheless, PCR-based protocols are still widely used in laboratories with no or limited access to NGS platforms.

For instance, in order to recover as much information as possible from aDNA, traditional approaches rely on sets of overlapping primer pairs to analyse aDNA from a small and specific locus (e.g. mtDNA Hypervariable Regions HVR-I, HVR-II).

Targeting the minute amount of DNA molecules is sufficient to gain valuable information and allow a valid haplogroup call (Haak et al. 2010). Another popular approach to analyse degraded and aDNA using PCR techniques, is the so-called "Multiplex PCR" (Sanchez and Endicott 2006). This technique differs from standard PCR in that the amplification is performed simultaneously using multiple primers for multiple loci in the DNA molecule instead of just one. Multiplex PCR approaches are advantageous as less template DNA is needed compared to multiple simplex PCRs (Butler et al. 2003; Haak et al. 2010).

Multiplexing techniques are particular suitable for SNP-typing because the small amplicon sizes are only minimally affected by the degree of DNA fragmentation in ancient samples. Multiplex PCR typing assays proved to be an important tool in targeting Y-chromosome SNPs from modern-day (Sanchez et al. 2003; Berniell-Lee et al. 2007; Martínez-Cruz et al. 2011), forensic (Sanchez et al. 2006; Sanchez et al. 2008) and ancient samples (Bouakaze et al. 2007; Haak et al. 2010). Furthermore, (Seidenberg et al. 2012) adapted STR multiplexing techniques to analyse short tandem repeats (STRs) from ancient and degraded DNA (i.e. miniSTR heptaplex system) for individuals identification and for reconstruction of kinship.

Specific aDNA multiplex PCR systems have been successfully applied in National Geographic's Genographic Project to screen worldwide mtDNA genetic diversity in human populations using the GenoCoRe22 SNP assay (Haak et al. 2010; Martínez-Cruz et al. 2012), and with particular relevance to this thesis, to screen genetic diversity in ancient human populations from South America using the AmericaPlex26 assay (Coutinho et al. 2014)(Chapter 5).

New NGS techniques applied to the field of aDNA has now moved away from targeted PCR amplification to the generation of (amplified) complete DNA genomic libraries from ancient extracts. Despite the fact that NGS was not developed for aDNA research, this field has benefitted greatly from this technology (Millar et al. 2008; Knapp and Hofreiter 2010), because NGS platforms have currently only a limited read length, which requires that the DNA to be fragmented before it can be prepared for sequencing. The short fragment length of aDNA therefore proves in fact advantageous as the DNA does not need to be mechanically or enzymatically fragmented before the genomic

library construction (Knapp and Hofreiter 2010; Meyer and Kircher 2010; Briggs and Heyn 2012).

In many cases it is only possible to retrieve minute amounts of endogenous DNA from ancient samples, and the average amount of endogenous aDNA varies greatly but is usually small and represents no more than the 5%, while the remaining DNA in an aDNA extract comes from environmental sources and microbes, such as bacteria or fungi (Green et al. 2006; Millar et al. 2008). Provided that sufficiently complex aDNA libraries can be generated, massively parallel sequencing made it possible to generate complete mtDNA genomes and nuclear genomes from ancient specimens.

A recent methodological development in aDNA research for genomic library preparation focussed on DNA libraries from "single-stranded" DNA molecules. The technique allows a substantial improvement in the retrieval of aDNA molecules, mainly from fragmented and degraded aDNA molecules with single-strand breaks, which are likely present in ancient samples. The advantage of such technique lies in the additional capture of single-strand DNA molecules, which would not be possible with the standard double-strand library preparation approach (Gansauge and Meyer 2013). Moreover, another method was also developed to target the small amount of "endogenous molecules" using biotinylated RNA baits synthesis transcribed form DNA libraries. Carpenter and colleagues were able to capture DNA fragments increasing the reads mapped to human genomes. In consequence, this represents an effective DNA enrichment approach, making it less expensive and applicable to large numbers of samples (Carpenter et al. 2013).

The deep coverage of NGS data does in theory allow discrimination of contamination from endogenous aDNA, and novel bioinformatics tools allow the examination of aDNA characteristics. Patterns of nucleotide misincorporation using the NGS approach have been elucidated (Stiller et al. 2006) and an in-depth exploration of nucleotide misincorporations were studied to distinguish aDNA in a wide range of ancient remains (Sawyer et al. 2012).

New bioinformatics tools such as the "Map of aDNA damage" software (Ginolhac et al. 2011) allows to calculate and plot the rate of damage according to the main types (e.g. nucleotide misincorporation and fragmentation patterns) using NGS sequencing reads. This tool is also embedded in a novel analysis pipeline called "Paleomix" which was established to handle aDNA data from the raw reads to the mapping reads to reference genomes (Schubert et al. 2014). Next-Generation Sequencing

technologies applied to aDNA research have provided the field with an unprecedented volume of genetic information (i.e. complete genomes) from ancient and extinct specimens (Ho and Gilbert 2010; Knapp and Hofreiter 2010; Krause 2010; Shapiro and Hofreiter 2010; Orlando et al. 2011; Paijmans et al. 2012).

The first study to report of NGS technology applied to aDNA was the large-scale sequencing of 27,000 bp from a Pleistocene cave bear (Noonan et al. 2005). A year later, 13 billion (base pairs) bp from the nuclear genome of a woolly mammoth permafrost sample were published under the suggestive title "Metagenomics to paleogenomics" (Poinar et al. 2006).

The reported first 1 million bp from a Neanderthal marked a major breakthrough in the field of ancient human DNA (Green et al. 2006). The data analysis led Green and colleagues to suggest some degree of admixture (i.e. genetic contribution from one population to another) between humans and Neanderthals by the time both forms co-existed in Europe around 40,000 years ago according to the archaeological record.

The study also estimated the time of divergence of both groups to be 516,000 years BP based on SNP composition (Green et al. 2006). A parallel study using the same Neanderthal DNA extract reported 75,000 bp of the nuclear genome (Noonan et al. 2006), and estimated a divergence time for modern humans and Neanderthals of 370,000 years. Concerning the specific question of admixture, Noonan and colleagues pointed out that little or no interbreeding had occurred. Since both studies differed in the time of divergence and interbreeding, this lead to suggest that at least one of the publications was erroneous. A reanalysis of the data by (Wall and Kim 2007) reported some incongruence in the Green et al. data mainly identified as longer sequence reads, that resembles modern human contamination. Soon after, the same research group published the first complete mitochondrial genome from the same specimen (Green et al. 2008) and the first draft genome of the Neanderthal with 1.3X fold coverage, representing almost 60% of the genome. Learning from the previous attempt to sequence aDNA from a closely related human species and the omnipresent DNA contamination risk, this work also showed the necessity to validate the authenticity of Neanderthal genome in this new era of massive DNA sequencing (Green et al. 2009), and highlighted critical steps such as the preparation and amplification of the genomic libraries in a sterile environment (Green et al. 2009). The NGS based analysis of genomic data from the Denisovan specimen indicated that Neanderthal had a sister taxon, and suggested a genetic contribution of 4-6% from this archaic human to the gene pool of present-day populations from Papua

New Guinea and Oceania (Reich et al. 2010). Later, in an outstanding technological development coupled with NGS, researchers were able to obtain a high coverage (30X) complete genome sequence from the same Denisovan individual (Meyer et al. 2012).

Here, the single-stranded library preparation method was applied first and allowed to generate additional sequence data, which helped to complete a full genome sequence similar in quality to modern human genomes. The ultimate research effort in aDNA completing a long-standing endeavour was the publication of the high coverage Neanderthal genome (Prufer et al. 2014). A toe phalanx also found in the Denisova Cave in the Altai Mountains from Siberia represents the best preserved Neanderthal sample, with sufficient complexity to obtain a high-resolution DNA nuclear genome. The comparison with present-day humans and Denisovans has established a complex pattern of gene flow in the past between these archaic humans.

Final remarks

Since the establishment of the ancient DNA research field in the mid 80's, the retrieval of aDNA molecules has astonished the research community and impressive scientific achievements were accomplished. Ancient DNA offers a window into the past and allows reconstructing historic events, which would not be possible to infer by analysing only modern genetic data. With the advent of new methodological and technological advances in recent years (e.g. NGS, optimization of extraction, library preparation protocols, capture techniques, and not least massive advances in bioinformatics), the field has truly come of age and confirmed the true potential of aDNA research, undertaking large-scale genetic studies, and making great progress in answering long-standing research questions. The NGS era has changed the landscape of genomic projects, pushing the boundaries to what it is achievable in terms of data generation, ultimately leading to field of "Paleogenomics". While technology is moving forward, the field of ancient DNA is not from its inherent problems and it remains important for aDNA researchers to continuously improve strategies that guarantee the validation of the results. Working with ancient human samples is particularly challenging. New opportunities to study arise (e.g. gene expression or positive selection of genes, refinement of population genetics involving ancestral populations, climate and environmental change and evolution of diseases). In more general terms, biologists have made progress in understanding the processes of evolution for animals, plants and human species.

However, since the sequencing of genomes has already been accomplished for many of the extinct species, it is now important to fully understand such information in order to interpret ancient genomic data with care.

Table 1. Subjective milestones in aDNA research

Date	Reference	Publication
1984	(Higuchi et al. 1984)	First aDNA sequences from Quagga species
1985	(Paabo 1985)	Egyptian mummy DNA
1987	(Mullis and Faloona 1987)	Polymerase Chain Reaction - PCR
1992	(Cooper et al. 1992)	aDNA from Moa species
1994	(Woodward et al. 1994)	Dinosaur DNA
1997	(Krings et al. 1997)	First Neanderthal mtDNA sequence
2000	(Cooper and Poinar 2000)	Ancient DNA guidelines 'Do it right or not at all'
2006	(Green et al. 2006)	Neanderthal DNA (1 million bp)
2006	(Poinar et al. 2006)	Mammoth metagenomics (NGS aDNA)
2008	(Ermini et al. 2008)	Tyrolean Iceman mtDNA genome sequence
2008	(Miller et al. 2008)	Woolly mammoth genome sequence
2009	(Miller et al. 2009)	Tasmanian tiger (Thylacine) genome sequence
2008	(Green et al. 2008)	Neanderthal mtDNA complete sequence
2008	(Gilbert et al. 2008c)	'Paleoskimo' mtDNA complete genome
2009	(Briggs et al. 2009)	Targeted retrieval and analysis of Neanderthal mtDNA genomes
2010	(Green et al. 2010)	Neanderthal: Draft genome sequence
2010	(Reich et al. 2010)	Genetic history of an archaic hominin group from Denisova Cave in Siberia.
2010	(Krause et al. 2010)	mtDNA genome of an Early Modern Human from Kostenki, Russia
2010	(Rasmussen et al. 2010)	Ancient human genome sequence of an extinct Palaeo-Eskimo individual
2011	(Schuenemann et al. 2011)	DNA from from victims of the Black Death
2011	(Orlando et al. 2011)	aDNA sequences from a permafrost-preserved Pleistocene horse bone (3rd Generation NGS)
2012	(Keller et al. 2012)	New insights into the Tyrolean Iceman's origin

		and phenotype as inferred by whole-genome sequencing
2012	(Meyer et al. 2012)	High-coverage complete genome sequence from 'Denisovan' individual
2013	(Fu et al. 2013)	DNA from an early human in Tianyuan, China
2013	(Dabney et al. 2013)	Mitochondrial genome reconstructed from ultra-short DNA fragments
2013	(Gansauge and Meyer 2013)	Single-stranded DNA library preparation for ancient or damaged DNA
2014	(Raghavan et al. 2014b)	Upper Palaeolithic Siberian Genome (Malt'a boy)
2014	(Meyer et al. 2014)	Middle Pleistocene hominin genome sequence from 'Sima de los Huesos'
2014	(Rasmussen et al. 2014)	Late Pleistocene genome from 'Clovis' burial
2014	(Prufer et al. 2014)	High resolution Neanderthal genome sequence
2014	(Olalde et al. 2014)	Mesolithic genome: ancestral pigmentation genes
2014	(Seguin-Orlando et al. 2014)	Genomic structure in Europeans dating back at least 36,200 years
2014	(Fu et al. 2014)	Genome sequence of a 45,000-year-old modern human from western Siberia
2014	(Lazaridis et al. 2014)	Genomes from Mesolithic and Neolithic ancient Europeans
2014	(Bos et al. 2014)	Pre-Columbian tuberculosis in the New World
2014	(Devault et al. 2014)	Microbial detection Array for ancient pathogens
2015	(Haak et al. 2015)	Massive migration from the steppe was a source for Indo-European languages in Europe
2015	(Allentoft et al. 2015)	Population genomics of Bronze Age Eurasia
2015	(Rasmussen et al. 2015a)	Ancestry of 'Kennewick Man'
2015	(Raghavan et al. 2015)	Genomic evidence for the Pleistocene and recent population history of Native Americans
2015	(Llorente et al. 2015)	First ancient genome from African remains

REFERENCES

Acuna-Soto R, Stahle DW, Cleaveland MK, and Therrell MD. 2002. Megadrought and Megadeath in 16th Century Mexico. Emerging Infectious Diseases 8(4):360-362.

Achilli A, Perego U, and Bravi C. 2008. The Phylogeny of the Four Pan-American MtDNA Haplogroups: Implications for Evolutionary and Disease Studies. PLoS ONE 3(3):e1764.

Achilli A, Perego UA, Lancioni H, Olivieri A, Gandini F, Hooshiar Kashani B, Battaglia V, Grugni V, Angerhofer N, Rogers MP et al. . 2013. Reconciling migration models to the Americas with the variation of North American native mitogenomes. Proceedings of the National Academy of Sciences 110(35):14308-14313.

Adler CJ, Dobney K, Weyrich LS, Kaidonis J, Walker AW, Haak W, Bradshaw CJA, Townsend G, Soltysiak A, Alt KW et al. . 2013. Sequencing ancient calcified dental plaque shows changes in oral microbiota with dietary shifts of the Neolithic and Industrial revolutions. Nat Genet 45(4):450-455.

Afonso Costa H, Carvalho M, Lopes V, Balsa F, Bento AM, Serra A, Andrade L, Anjos MJ, Vide MC, Pantoja S et al. . 2010. Mitochondrial DNA sequence analysis of a native Bolivian population. J Forensic Leg Med 17(5):247-253.

Allentoft ME, Collins M, Harker D, Haile J, Oskam CL, Hale ML, Campos PF, Samaniego JA, Gilbert MTP, Willerslev E et al. . 2012. The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. Proceedings of the Royal Society B: Biological Sciences.

Allentoft ME, Sikora M, Sjögren K-G, Rasmussen S, Rasmussen M, Stenderup J, Damgaard PB, Schroeder H, Ahlström T, Vinner L et al. . 2015. Population genomics of Bronze Age Eurasia. Nature 522(7555):167-172.

Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F et al. . 1981. Sequence and organization of the human mitochondrial genome. Nature 290:457–465.

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

Arriaza BT, Salo W, Aufderheide AC, and Holcomb TA. 1995. Pre-Columbian tuberculosis in Northern Chile: Molecular and skeletal evidence. Am J Phys Anthropol 98(1):37-45.

Austin JJ, Ross AJ, Smith AB, Fortey RA, and Thomas RH. 1997. Problems of reproducibility--does geologically ancient DNA survive in amber-preserved insects? ProcRSocLond B BiolSci 264(1381):467-474.

Baca M, Doan K, Sobczyk M, Stankovic A, and Weglenski P. 2012. Ancient DNA reveals kinship burial patterns of a pre-Columbian Andean community. BMC Genet 13(1):30.

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

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

Behar DM, van Oven M, Rosset S, Metspalu M, Loogvali EL, Silva NM, Kivisild T, Torroni A, and Villems R. 2012. A "Copernican" reassessment of the human mitochondrial DNA tree from its root. Am J Hum Genet 90(4):675-684.

Bengtsson CF, Olsen ME, Brandt Lÿ, Bertelsen MF, Willerslev E, Tobin DJ, Wilson AS, and Gilbert MTP. 2011. DNA from Keratinous Tissue, Part I: Hair and Nail. Annals of Anatomy - Anatomischer Anzeiger In Press, Accepted Manuscript.

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

Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, Hall KP, Evers DJ, Barnes CL, Bignell HR et al. . 2008. Accurate whole human genome sequencing using reversible terminator chemistry. Nature 456(7218):53-59.

Berniell-Lee G, Sandoval K, Mendizabal I, Bosch E, and Comas D. 2007. SNPlexing the human Y-chromosome: a single-assay system for major haplogroup screening. Electrophoresis 28(18):3201-3206.

Biesecker LG. 2010. Exome sequencing makes medical genomics a reality. Nat Genet 42(1):13-14.

Bisso-Machado R, Bortolini MC, and Salzano FM. 2012. Uniparental genetic markers in South Amerindians. Genetics and Molecular Biology 35(2):365-387.

Bock C, Tomazou EM, Brinkman AB, Muller F, Simmer F, Gu H, Jager N, Gnirke A, Stunnenberg HG, and Meissner A. 2010. Quantitative comparison of genome-wide DNA methylation mapping technologies. Nat Biotechnol 28(10):1106-1114.

Bodner M, Perego UA, Huber G, Fendt L, Rock AW, Zimmermann B, Olivieri A, Gomez-Carballa A, Lancioni H, Angerhofer N et al. . 2012. Rapid coastal spread of First Americans: novel insights from South America's Southern Cone mitochondrial genomes. Genome Research 22(5):811-820.

Bolnick DA, Shook B, Campbell L, and Goddard I. 2004. Problematic Use of Greenberg's Linguistic Classification of the Americas in Studies of Native American Genetic Variation. Am J Hum Genet 75(3):519-523.

Bollongino R, and Vigne J-D. 2008. Temperature monitoring in archaeological animal bone samples in the Near East arid area, before, during and after excavation. Journal of Archaeological Science 35(4):873-881.

Bonatto S, and Salzano F. 1997. A single and early migration for the peopling of the Americas supported by mitochondrial DNA sequence data. Proc Natl Acad Sci USA 94::1866–1871.

Bortolini M-C, Salzano FM, Thomas MG, Stuart S, Nasanen SPK, Bau CHD, Hutz MH, Layrisse Z, Petzl-Erler ML, Tsuneto LT et al. . 2003. Y-Chromosome Evidence for Differing Ancient Demographic Histories in the Americas. Am J Hum Genet 73(3):524-539.

Bos KI, Harkins KM, Herbig A, Coscolla M, Weber N, Comas I, Forrest SA, Bryant JM, Harris SR, Schuenemann VJ et al. . 2014. Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. Nature 514(7523):494-497.

Bos KI, Schuenemann VJ, Golding GB, Burbano HA, Waglechner N, Coombes BK, McPhee JB, DeWitte SN, Meyer M, Schmedes S et al. . 2011. A draft genome of Yersinia pestis from victims of the Black Death. Nature 478(7370):506-510.

Bouakaze C, Keyser C, Amory S, Crubezy E, and Ludes B. 2007. First successful assay of Y-SNP typing by SNaPshot minisequencing on ancient DNA. Int J Legal Med 121(6):493-499.

Bramanti B, Thomas MG, Haak W, Unterlaender M, Jores P, Tambets K, Antanaitis-Jacobs I, Haidle MN, Jankauskas R, Kind CJ et al. . 2009. Genetic discontinuity between local hunter-gatherers and central Europe's first farmers. Science 326(5949):137-140.

Brandt G, Haak W, Adler CJ, Roth C, Szecsenyi-Nagy A, Karimnia S, Moller-Rieker S, Meller H, Ganslmeier R, Friederich S et al. . 2013. Ancient DNA reveals key stages in the formation of central European mitochondrial genetic diversity. Science 342(6155):257-261.

Brandt G, Szécsényi-Nagy A, Roth C, Alt KW, and Haak W. 2015. Human paleogenetics of Europe – The known knowns and the known unknowns. Journal of Human Evolution 79:73-92.

Braun M, Cook DC, and Pfeiffer S. 1998. DNA from Mycobacterium tuberculosis Complex Identified in North American, Pre-Columbian Human Skeletal Remains. Journal of Archaeological science 25:271-277.

Briggs AW, Good JM, Green RE, Krause J, Maricic T, Stenzel U, Lalueza-Fox C, Rudan P, Brajkovic D, Kucan Z et al. . 2009. Targeted retrieval and analysis of five Neandertal mtDNA genomes. Science 325(5938):318-321.

Briggs AW, and Heyn P. 2012. Preparation of next-generation sequencing libraries from damaged DNA. Methods Mol Biol 840:143-154.

Briggs AW, Stenzel U, Johnson PLF, Green RE, Kelso J, Prüfer K, Meyer M, Krause J, Ronan MT, Lachmann M et al. . 2007. Patterns of damage in genomic DNA sequences from a Neandertal. Proc Natl Acad Sci U S A 104(37):14616-14621.

Brotherton P, Endicott P, Sanchez JJ, Beaumont M, Barnett R, Austin J, and Cooper A. 2007. Novel high-resolution characterization of ancient DNA reveals C > U-type base modification events as the sole cause of post mortem miscoding lesions. Nucleic Acids Res 35(17):5717-5728.

Brotherton P, Haak W, Templeton J, Brandt G, Soubrier J, Jane Adler C, Richards SM, Sarkissian CD, Ganslmeier R, Friederich S et al. . 2013. Neolithic mitochondrial haplogroup H genomes and the genetic origins of Europeans. Nature Communications 4:1764.

Brown MD, Hosseini SH, Torroni A, Bandelt HJ, Allen JC, Schurr TG, Scozzari R, Cruciani F, and Wallace DC. 1998. mtDNA haplogroup X: An ancient link between Europe/Western Asia and North America? Am J Hum Genet 63(6):1852-1861.

Brown TA, and Brown KA. 1992. Ancient DNA and the archaeologist. Antiquity 66:10-23.

Brown WM, George, M. Jr & Wilson, A. C. . 1979. Rapid evolution of animal mitochondrial DNA. . Proc Natl Acad Sci USA 76, 1967±1971

Burbano HA, Hodges E, Green RE, Briggs AW, Krause J, Meyer M, Good JM, Maricic T, Johnson PLF, Xuan Z et al. . 2010. Targeted Investigation of the Neandertal Genome by Array-Based Sequence Capture. Science 328(5979):723-725.

Butler J, Shen Y, and McCord B. 2003. The development of reduced size STR amplicons as tools for analysis of degraded DNA. Journal of Forensic Sciences 48(5):1054 - 1064.

Cann RL, Stoneking M, and Wilson AC. 1987. Mitochondrial DNA and human evolution. Nature 325(6099):31-36.

Cann RL, and Wilson AC. 1983. Length mutations in human mitochondrial DNA. Genetics 104(4):699-711.

Cano RJ, Poinar HN, Pieniazek NJ, Acra A, and Poinar GO, Jr. 1993. Amplification and sequencing of DNA from a 120-135-million-year-old weevil. Nature 363(6429):536-538.

Caramelli D, Lalueza-Fox C, Condemi S, Longo L, Milani L, Manfredini A, de Saint Pierre M, Adoni F, Lari M, Giunti P et al. . 2006. A highly divergent mtDNA sequence in a Neandertal individual from Italy. Curr Biol 16(16):R630-632.

Caramelli D, Vernesi C, Sanna S, Sampietro L, Lari M, Castri L, Vona G, Floris R, Francalacci P, Tykot R et al. . 2007. Genetic variation in prehistoric Sardinia. Hum Genet 122(3-4):327-336.

Cardoso S, Alfonso-Sanchez MA, Valverde L, Sanchez D, Zarrabeitia MT, Odriozola A, Martinez-Jarreta B, and de Pancorbo MM. 2012. Genetic uniqueness of the Waorani tribe from the Ecuadorian Amazon. Heredity 108(6):609-615.

Carnese F, Mendisco F, and Keyser C. 2010. Paleogenetical study of pre-Columbian samples from Pampa Grande (Salta, Argentina). Am J Phys Anthropol 141:452 - 462.

Carpenter ML, Buenrostro JD, Valdiosera C, Schroeder H, Allentoft ME, Sikora M, Rasmussen M, Gravel S, Guillen S, Nekhrizov G et al. . 2013. Pulling out the 1%: whole-genome capture for the targeted enrichment of ancient DNA sequencing libraries. Am J Hum Genet 93(5):852-864.

Clark PU, Dyke AS, Shakun JD, Carlson AE, Clark J, Wohlfarth B, Mitrovica JX, Hostetler SW, and McCabe AM. 2009. The Last Glacial Maximum. Science 325(5941):710-714.

Consortium TIH. 2005. A haplotype map of the human genome. Nature 437(7063):1299-1320.

Cooper A, Mourer-Chauvire C, Chambers GK, von Haeseler A, Wilson AC, and Paabo S. 1992. Independent origins of New Zealand moas and kiwis 75. ProcNatlAcadSciUSA 89(18):8741-8744.

Cooper A, and Poinar HN. 2000. Ancient DNA: Do It Right or Not at All. Science(289):1139.

Coutinho A, Valverde G, Fehren-Schmitz L, Cooper A, Barreto Romero MI, Espinoza IF, Llamas B, and Haak W. 2014. AmericaPlex26: a SNaPshot multiplex system for genotyping the main human mitochondrial founder lineages of the Americas. PLoS ONE 9(3):e93292.

Crawford MH. 1998. The origins of Native Americans. Evidence from anthropological genetics. Cambridge University Press, Cambridge (UK).

Crosby A. 1976. Virgin soil epidemics as a factor in the aboriginal depopulation in America. William Mary Q 33:289-299.

Crubezy E, Legal L, Fabas G, Dabernat H, and Ludes B. 2006. Pathogeny of archaic mycobacteria at the emergence of urban life in Egypt (3400 bc). Infection, Genetics and Evolution 6(1):13-21.

Cui Y, Lindo J, Hughes CE, Johnson JW, Hernandez AG, and et al. 2013. Ancient DNA Analysis of Mid-Holocene Individuals from the Northwest Coast of North America Reveals Different Evolutionary Paths for Mitogenomes. PLoS ONE 8(7):e66948. doi:66910.61371/journal.pone.0066948.

Chatters JC, Kennett DJ, Asmerom Y, Kemp BM, Polyak V, Blank AN, Beddows PA, Reinhardt E, Arroyo-Cabrales J, Bolnick DA et al. . 2014. Late Pleistocene human skeleton and mtDNA link Paleoamericans and modern Native Americans. Science 344(6185):750-754.

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

Dabney J, Knapp M, Glocke I, Gansauge MT, Weihmann A, Nickel B, Valdiosera C, Garcia N, Paabo S, Arsuaga JL et al. . 2013. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. Proc Natl Acad Sci U S A 110(39):15758-15763.

de Melo FL, de Mello JCM, Fraga AM, Nunes K, and Eggers S. 2010. Syphilis at the Crossroad of Phylogenetics and Paleopathology. . PLoS Negl Trop Dis 4(1):e575. doi:510.1371/journal.pntd.0000575.

de Saint Pierre M, Bravi CM, Motti JM, Fuku N, Tanaka M, Llop E, Bonatto SL, and Moraga M. 2012a. An alternative model for the early peopling of southern South America revealed by analyses of three mitochondrial DNA haplogroups. PLoS ONE 7(9):e43486.

de Saint Pierre M, Gandini F, Perego UA, Bodner M, Gomez-Carballa A, Corach D, Angerhofer N, Woodward SR, Semino O, Salas A et al. . 2012b. Arrival of Paleo-Indians to the southern cone of South America: new clues from mitogenomes. PLoS ONE 7(12):e51311.

Del Pozzo G, Guardiola J, and 1989. Mummy DNA fragment identified. Nature 339:431–432.

DeSalle R, Gatesy J, Wheeler W, and Grimaldi D. 1992. DNA sequences from a fossil termite in Oligo-Miocene amber and their phylogenetic implications. Science 257(5078):1933-1936.

Devault AM, McLoughlin K, Jaing C, Gardner S, Porter TM, Enk JM, Thissen J, Allen J, Borucki M, DeWitte SN et al. . 2014. Ancient pathogen DNA in archaeological samples detected with a Microbial Detection Array. Sci Rep 4.

Diamond JM. 1997. Guns, Germs, and Steel: The Fates of Human Societies: W.W. Norton & Co.

Dillehay TD. 1997. The Archaeological Context and Interpretation, Volume 2: Monte Verde, a Late Pleistocene Settlement in Chile. (Washington, DC: Smithsonian Institute Press).

Dillehay TD. 2009. Probing deeper into first American studies. Proceedings of the National Academy of Sciences 106(4):971-978.

Dillehay TD, Ramirez C, Pino M, Collins MB, Rossen J, and Pino-Navarro JD. 2008. Monte Verde: seaweed, food, medicine, and the peopling of South America. Science 320(5877):784-786.

Dixon EJ. 1993. Quest for the Origins of the First Americans. Univ of New Mexico Press, Albuquerque

Dobyns HF. 1983. Their numbers become thinned: Native American population dynamics in eastern North America. Knoxville: University of Tenessee Press.

Dobyns. H. 1993. Disease transfer at contact. Annu Rev(Anthropol. 22:273–91).

Doran GH, Dickel DN, Ballinger WE, Jr., Agee OF, Laipis PJ, and Hauswirth WW. 1986. Anatomical, cellular and molecular analysis of 8,000-yr-old human brain tissue from the Windover archaeological site. Nature 323(6091):803-806.

Drancourt M, Aboudharam G, Signoli M, Dutour O, and Raoult D. 1998. Detection of 400-year-old Yersinia pestis DNA in human dental pulp: An approach to the diagnosis of ancient septicemia. Proc Natl Acad Sci U S A 95(21):12637-12640.

Drummond AJ, Rambaut A, Shapiro B, and Pybus OG. 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. Mol Biol Evol 22(5):1185-1192.

Dulik Matthew C, Zhadanov Sergey I, Osipova Ludmila P, Askapuli A, Gau L, Gokcumen O, Rubinstein S, and Schurr Theodore G. 2012. Mitochondrial DNA and Y Chromosome Variation Provides Evidence for a Recent Common Ancestry between Native Americans and Indigenous Altaians. The American Journal of Human Genetics 90(2):229-246.

Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, Doyle F, Epstein CB, Frietze S, Harrow J, Kaul R et al. . 2012. An integrated encyclopedia of DNA elements in the human genome. Nature 489(7414):57-74.

Edwards RA, Rodriguez-Brito B, Wegley L, Haynes M, Breitbart M, Peterson DM, Saar MO, Alexander S, Alexander EC, Jr., and Rohwer F. 2006. Using pyrosequencing to shed light on deep mine microbial ecology. BMC Genomics 7:57.

Eglinton G, and Logan G. 1991. Molecular preservation. Phil Trans R Soc Lond 333:315-328.

Endicott P, and Ho SY. 2008. A Bayesian evaluation of human mitochondrial substitution rates. Am J Hum Genet 82(4):895-902.

Erlandson JM, Graham MH, Bourque BJ, Corbett D, Estes JA, and Steneck RS. 2007. The Kelp Highway Hypothesis: Marine Ecology, the Coastal Migration Theory, and the Peopling of the Americas. The Journal of Island and Coastal Archaeology 2(2):161-174.

Ermini L, Der Sarkissian C, Willerslev E, and Orlando L. 2015. Major transitions in human evolution revisited: A tribute to ancient DNA. Journal of Human Evolution 79:4-20.

Ermini L, Olivieri C, Rizzi E, Corti G, Bonnal R, Soares P, Luciani S, Marota I, De Bellis G, Richards MB et al. . 2008. Complete mitochondrial genome sequence of the Tyrolean Iceman. Curr Biol 18(21):1687-1693.

Faerman M, and Jankauskas R. 2000. Paleopathological and molecular evidence of human bone tuberculosis in Iron Age Lithuania. AnthropolAnz 58(1):57-62.

Fagundes NJ, Kanitz R, Eckert R, Valls AC, Bogo MR, Salzano FM, Smith DG, Silva WA, Jr., Zago MA, Ribeiro-dos-Santos AK et al. . 2008. Mitochondrial population genomics supports a single pre-Clovis origin with a coastal route for the peopling of the Americas. Am J Hum Genet 82(3):583-592.

Fehren-Schmitz L, Llamas B, Lindauer S, Tomasto-Cagigao E, Kuzminsky S, Rohland N, Santos FR, Kaulicke P, Valverde G, Richards SM et al. . 2015. A Re-Appraisal of the Early Andean Human Remains from Lauricocha in Peru. PLoS ONE 10(6).

Fehren-Schmitz L, Warnberg O, and Reindel M. 2011. Diachronic investigations of mitochondrial and Y-chromosomal genetic markers in pre-Columbian Andean highlanders from South Peru. Ann Hum Genet 75:266 - 283.

Fladmark KR. 1979. Routes: Alternative Migration Corridors for Early Man in North America American. Antiquity 44:55-69.

Fletcher HA, Donoghue HD, Holton J, Pap I, and Spigelman M. 2003. Widespread occurrence of Mycobacterium tuberculosis DNA from 18th-19th century Hungarians. AmJPhysAnthropol 120(2):144-152.

Forster P, Harding R, Torroni A, and Bandelt HJ. 1996. Origin and evolution of Native American mtDNA variation: a reappraisal. AmJHumGenet 59(4):935-945.

Friedberg EC, Walker GC, and Siede W. 1995. Repair and Mutagenesis DNA. . ASM Press Washington DC:698.

Fu Q, Li H, Moorjani P, Jay F, Slepchenko SM, Bondarev AA, Johnson PL, Aximu-Petri A, Prufer K, de Filippo C et al. . 2014. Genome sequence of a 45,000-year-old modern human from western Siberia. Nature 514(7523):445-449.

Fu Q, Meyer M, Gao X, Stenzel U, Burbano HA, Kelso J, and Paabo S. 2013. DNA analysis of an early modern human from Tianyuan Cave, China. Proc Natl Acad Sci U S A 110(6):2223-2227.

Fulton TL. 2012. Setting up an ancient DNA laboratory. Methods Mol Biol 840:1-11.

Gansauge MT, and Meyer M. 2013. Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. Nature Protocols 8(4):737-748.

Gaya-Vidal M, Moral P, Saenz-Ruales N, Gerbault P, Tonasso L, Villena M, Vasquez R, Bravi CM, and 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(2):215-230.

Gevers D, Knight R, Petrosino JF, Huang K, McGuire AL, Birren BW, Nelson KE, White O, Methe BA, and Huttenhower C. 2012. The Human Microbiome Project: A Community Resource for the Healthy Human Microbiome. Plos Biology 10(8):e1001377.

Gilbert M, Bandelt H-J, Hofreiter M, and Barnes I. 2005. Assessing ancient DNA studies. Trends in Ecology and Evolution 20:541 - 544.

Gilbert MT, Kivisild T, Gronnow B, Andersen PK, Metspalu E, Reidla M, Tamm E, Axelsson E, Gotherstrom A, Campos PF et al. . 2008a. Paleo-Eskimo mtDNA genome reveals matrilineal discontinuity in Greenland. Science 320(5884):1787-1789.

Gilbert MT, Tomsho LP, Rendulic S, Packard M, Drautz DI, Sher A, Tikhonov A, Dalen L, Kuznetsova T, Kosintsev P et al. . 2007. Whole-genome shotgun sequencing of mitochondria from ancient hair shafts. Science 317(5846):1927-1930.

Gilbert MT, Willerslev E, Hansen AJ, Barnes I, Rudbeck L, Lynnerup N, and Cooper A. 2003. Distribution patterns of postmortem damage in human mitochondrial DNA. Am J Hum Genet 72(1):32-47.

Gilbert MTP, Jenkins DL, Gotherstrom A, Naveran N, Sanchez JJ, Hofreiter M, Thomsen PF, Binladen J, Higham TFG, Yohe RM et al. . 2008b. DNA from pre-Clovis human coprolites in Oregon, North America. Science 320(5877):786-789.

Gilbert MTP, Kivisild T, Grønnow B, Andersen PK, Metspalu E, Reidla M, Tamm E, Axelsson E, Götherström A, Campos PF et al. . 2008c. Paleo-Eskimo mtDNA genome reveals matrilineal discontinuity in Greenland. Science (New York, NY) 320:1787-1789.

Giles RE, Blanc, H., Cann, H. M. & Wallace, D. C. . 1980. Maternal inheritance of human mitochondrial DNA. Proc Natl Acad Sci USA 77 6715±6719.

Ginolhac A, Rasmussen M, Gilbert MT, Willerslev E, and Orlando L. 2011. mapDamage: testing for damage patterns in ancient DNA sequences. Bioinformatics 27(15):2153-2155.

Goebel T, Waters MR, and O'Rourke DH. 2008. The late Pleistocene dispersal of modern humans in the Americas. Science 319(5869):1497-1502.

Gonzalez-Jose R, Bortolini MC, Santos FR, and Bonatto SL. 2008. The peopling of America: craniofacial shape variation on a continental scale and its interpretation from an interdisciplinary view. Am J Phys Anthropol 137(2):175-187.

Gonzalez-Jose R, Dahinten SL, Luis MA, Hernandez M, and Pucciarelli HM. 2001. Craniometric variation and the settlement of the Americas: testing hypotheses by means of R-matrix and matrix correlation analyses. Am J Phys Anthropol 116(2):154-165.

Green RE, Briggs AW, Krause J, Prufer K, Burbano HA, Siebauer M, Lachmann M, and Paabo S. 2009. The Neandertal genome and ancient DNA authenticity. Embo J 28(17):2494-2502.

Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, Patterson N, Li H, Zhai W, Fritz MHY et al. . 2010. A Draft Sequence of the Neandertal Genome. Science 328(5979):710-722.

Green RE, Krause J, Ptak SE, Briggs AW, Ronan MT, Simons JF, Du L, Egholm M, Rothberg JM, Paunovic M et al. . 2006. Analysis of one million base pairs of Neanderthal DNA. Nature 444(7117):330-336.

Green RE, Malaspinas A-S, Krause J, Briggs AW, Johnson PLF, Uhler C, Meyer M, Good JM, Maricic T, Stenzel U et al. . 2008. A complete Neandertal mitochondrial genome sequence determined by high-throughput sequencing. Cell 134:416-426.

Greenberg J, Turner II C, Zegura S, and al. e. 1986. The Settlement of the Americas: A Comparison of the Linguistic, Dental, and Genetic Evidence. Current Anthropology , Vol. 27,(No. 5):pp. 477-497

Haak W, Balanovsky O, Sanchez JJ, Koshel S, Zaporozhchenko V, Adler CJ, Der Sarkissian CSI, Brandt G, Schwarz C, Nicklisch N et al. . 2010. Ancient DNA from European early neolithic farmers reveals their near eastern affinities. PLoS biology 8(11):e1000536.

Haak W, Forster P, Bramanti B, Matsumura S, Brandt G, Tanzer M, Villems R, Renfrew C, Gronenborn D, Alt KW et al. . 2005. Ancient DNA from the first European farmers in 7500-year-old Neolithic sites. Science 310(5750):1016-1018.

Haak W, Lazaridis I, Patterson N, Rohland N, Mallick S, Llamas B, Brandt G, Nordenfelt S, Harney E, Stewardson K et al. . 2015. Massive migration from the steppe was a source for Indo-European languages in Europe. Nature.

Haas CJ, Zink A, Molnar E, Szeimies U, Reischl U, Marcsik A, Ardagna Y, Dutour O, Palfi G, and Nerlich AG. 2000. Molecular evidence for different stages of tuberculosis in ancient bone samples from Hungary. AmJPhysAnthropol 113(3):293-304.

Hagelberg E, and Clegg JB. 1993. Genetic polymorphisms in prehistoric Pacific islanders determined by analysis of ancient bone DNA. Proc Biol Sci 252(1334):163-170.

Hagelberg E, Sykes B, and Hedges R. 1989. Ancient bone DNA amplified. Nature 342(6249):485.

Hagelberg E, Thomas MG, Cook CE, Jr., Sher AV, Baryshnikov GF, and Lister AM. 1994. DNA from ancient mammoth bones. Nature 370(6488):333-334.

Halverson M, and Bolnick D. 2008. An ancient DNA test of a founder effect in Native American ABO blood group frequencies. Am J Phys Anthropol 137(3):342-347.

Hamilton MJ, Walker RS, and Kesler DC. 2014. Crash and rebound of indigenous populations in lowland South America. Sci Rep 4:4541.

Handt O, Richards M, Trommsdorff M, Kilger C, Simanainen J, Georgiev O, Bauer K, Stone A, Hedges R, and Schaffner W. 1994. Molecular genetic analyses of the Tyrolean Ice Man. Science 264(5166):1775-1778.

Harismendy O, Ng PC, Strausberg RL, Wang X, Stockwell TB, Beeson KY, Schork NJ, Murray SS, Topol EJ, Levy S et al. . 2009. Evaluation of next generation sequencing platforms for population targeted sequencing studies. Genome Biol 10(3):R32.

Harkins KM, and Stone AC. 2015. Ancient pathogen genomics: insights into timing and adaptation. Journal of Human Evolution 79:137-149.

Harris RA, Wang T, Coarfa C, Nagarajan RP, Hong C, Downey SL, Johnson BE, Fouse SD, Delaney A, Zhao Y et al. . 2010. Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications. Nat Biotechnol 28(10):1097-1105.

Hedges SB, and Schweitzer MH. 1995. Detecting dinosaur DNA. Science 268(5214):1191-1192.

Herrmann B, and Hummel S. 1994. Ancient DNA. New York: Springer-Verlag.

Hey J. 2005. On the Number of New World Founders: A Population Genetic Portrait of the Peopling of the Americas. PLoS Biol 3(6): e193 doi:101371/journalpbio0030193 3(6): e193.(doi:10.1371/journal.pbio.0030193).
Higuchi R, Bowman B, Freiberger M, Ryder OA, and Wilson AC. 1984. DNA sequences from the quagga, an extinct member of the horse family. Nature 312(5991):282-284.
Ho S, and Endicott P. 2008. The crucial role of calibration in molecular date estimates for the peopling of the Americas. Am J Hum Genet 83: 142–146.(author reply 146–147.).
Ho SY, and Gilbert MT. 2010. Ancient mitogenomics. Mitochondrion 10(1):1-11.
Hodges E, Xuan Z, Balija V, Kramer M, Molla MN, Smith SW, Middle CM, Rodesch MJ, Albert TJ, Hannon GJ et al. . 2007. Genome-wide in situ exon capture for selective resequencing. Nat Genet 39(12):1522-1527.
Hofreiter M, Serre D, Poinar HN, Kuch M, and Paabo S. 2001. Ancient DNA. Nature Reviews Genetics 2(5):353-359.
Horai S, Kondo R, Nakagawahattori Y, Hayashi S, Sonoda S, and Tajima K. 1993. Peopling of the America, Founded by 4 Major Lineages of Mitochondrial-DNA. Molecular Biology and Evolution 10(1):23-47.
Horai S, Murayama K, Hayasaka K, Matsubayashi S, Hattori Y, Fucharoen G, Harihara S, Park KS, Omoto K, and Pan IH. 1996. mtDNA polymorphism in East Asian Populations, with special reference to the peopling of Japan. AmJHumGenet 59(3):579-590.
Höss M, Dilling a, Currant a, and Pääbo S. 1996a. Molecular phylogeny of the extinct ground sloth Mylodon darwinii. Proc Natl Acad Sci U S A 93:181-185.
Höss M, Jaruga P, Zastawny TH, Dizdaroglu M, and Pääbo S. 1996b. DNA damage and DNA sequence retrieval from ancient tissues. Nucleic Acids Res 24(7):1304-1307.
Howard EB. 1933. Association of artifacts with mammoth and bison in eastern New Mexico. Science 78:524.
Hrdlicka A. 1937. The origin and antiquity of the American Indian. Annu Rep Smithson Inst Washington 1923:481–494.
Hublin JJ. 2009. The origin of Neandertals. Proceedings of the National Academy of Sciences 106(38):16022-16027.
Ingman M, Kaessmann H, Paabo S, and Gyllensten U. 2000. Mitochondrial genome variation and the origin of modern humans. Nature 408(6813):708-713.
Jantz RL, and Owsley DW. 2001. Variation among early North American crania. Am J Phys Anthropol 114(2):146-155.
Jobling MA, Hurles M, Hollox E, Hurles ME, Kivisild T, and Tyler-Smith C. 2013. Human evolutionary genetics, 2nd Edition. Garland Science Taylor & Francis Group, New York and London.
Jobling MA T-SC. 1995. Fathers and sons: the Y chromosome and human evolution. . Trends Genet Nov;11(11):449–456.
Karafet TM, Zegura SL, Posukh O, Osipova L, Bergen A, Long J, Goldman D, Klitz W, Harihara S, de Knijff P et al. . 1999. Ancestral Asian Source(s) of New World Y-Chromosome Founder Haplotypes. AmJHumGenet 64:817-831.
Keller A, Graefen A, Ball M, Matzas M, Boisguerin V, Maixner F, Leidinger P, Backes C, Khairat R, Forster M et al. . 2012. New insights into the Tyrolean Iceman's origin and phenotype as inferred by whole-genome sequencing. Nat Commun 3:698.

Kemp BM, Lindo J, Bolnick DA, Malhi RS, and Chatters JC. 2015. Anthropology. Response to Comment on "Late Pleistocene human skeleton and mtDNA link Paleoamericans and modern Native Americans". Science 347(6224):1261188.

Kitchen A, Miyamoto M, and Mulligan C. 2008. A Three-Stage Colonization Model for the Peopling of the Americas. . PLoS ONE 3(2):(e1596. doi:10.1371/journal.pone.0001596).

Kivisild T, Bamshad MJ, Kaldma K, Metspalu M, Metspalu E, Reidla M, Laos S, Parik J, Watkins WS, Dixon ME et al. . 1999. Deep common ancestry of indian and western-Eurasian mitochondrial DNA lineages. Curr Biol 9(22):1331-1334.

Kivisild T, Metspalu M, Bandelt HJ, Richards M, and Villems R. 2006. The world mtDNA phylogeny", Human Mitochondrial DNA and the Evolution of Homo sapiens Bandelt, H J, Macaulay, V y Richards, M, eds Springer-Verlag, Berlin.

Knapp M, Clarke AC, Horsburgh KA, and Matisoo-Smith EA. 2012. Setting the stage - Building and working in an ancient DNA laboratory. Annals of Anatomy - Anatomischer Anzeiger In Press, Corrected Proof.

Knapp M, and Hofreiter M. 2010. Next Generation Sequencing of Ancient DNA: Requirements, Strategies and Perspectives. Genes 1(2):227-243.

Knapp M, Lalueza-Fox C, and Hofreiter M. 2015. Re-inventing ancient human DNA. Investig Genet 6(1).

Kolman CJ, Centurion-Lara A, Lukehart SA, Owsley DW, and Tuross N. 1999. Identification of Treponema pallidum subspecies pallidum in a 200-year- old skeletal specimen. JInfectDis 180(6):2060-2063.

Krajewski C, Buckley L, and Westerman M. 1997. DNA phylogeny of the marsupial wolf resolved: Proc Biol Sci. 1997 Jun 22;264(1383):911-7.

Krause J. 2010. From Genes to Genomes: What is New in Ancient DNA? Mitteilungen der Gesellschaft für Urgeschichte 19:11-33.

Krause J, Briggs AW, Kircher M, Maricic T, Zwyns N, Derevianko A, and Paabo S. 2010. A complete mtDNA genome of an early modern human from Kostenki, Russia. Curr Biol 20(3):231-236.

Krause J, Dear PH, Pollack JL, Slatkin M, Spriggs H, Barnes I, Lister AM, Ebersberger I, Paabo S, and Hofreiter M. 2006. Multiplex amplification of the mammoth mitochondrial genome and the evolution of Elephantidae. Nature 439(7077):724-727.

Krause J, Lalueza-Fox C, Orlando L, Enard W, Green RE, Burbano HA, Hublin JJ, Hanni C, Fortea J, de la Rasilla M et al. . 2007. The derived FOXP2 variant of modern humans was shared with Neandertals. Curr Biol 17(21):1908-1912.

Krings M, Geisert H, Schmitz RW, Krainitzki H, and Pääbo S. 1999. DNA sequence of the mitochondrial hypervariable region II from the neandertal type specimen. Proc Natl Acad Sci U S A 96:5581-5585.

Krings M, Stone A, Schmitz RW, Krainitzki H, Stoneking M, and Paabo S. 1997. Neandertal DNA Sequences and the Origin of Modern Humans. Cell 90:19-30.

Kumar S, Bellis C, Zlojutro M, Melton PE, Blangero J, and Curran JE. 2011. Large scale mitochondrial sequencing in Mexican Americans suggests a reappraisal of Native American origins. BMC Evol Biol 11(293):1471-2148.

Lalueza-Fox C, Bertranpetit J, Alcover JA, Shailer N, and Hagelberg E. 2000. Mitochondrial DNA from Myotragus balearicus, an extinct bovid from the Balearic Islands. J Exp Zool 288(1):56-62.

Lalueza-Fox C, Calderon FL, Calafell F, Morera B, and Bertranpetit J. 2001. MtDNA from extinct Tainos and the peopling of the Caribbean. Ann Hum Genet 65(Pt 2):137-151.

Lalueza-Fox C, Castresana J, Sampietro L, Marques-Bonet T, Alcover JA, and Bertranpetit J. 2005a. Molecular dating of caprines using ancient DNA sequences of Myotragus balearicus, an extinct endemic Balearic mammal. BMC evolutionary biology 5:70.

Lalueza-Fox C, Gilbert MT, Martinez-Fuentes AJ, Calafell F, and Bertranpetit J. 2003. Mitochondrial DNA from pre-Columbian Ciboneys from Cuba and the prehistoric colonization of the Caribbean. Am J Phys Anthropol 121(2):97-108.

Lalueza-Fox C, Sampietro ML, Caramelli D, Puder Y, Lari M, Calafell F, Martinez-Maza C, Bastir M, Fortea J, de la RM et al. . 2005b. Neandertal evolutionary genetics: mitochondrial DNA data from the iberian peninsula. MolBiolEvol 22(4):1077-1081.

Lalueza C, Perez-Perez A, Prats E, Cornudella L, and Turbon D. 1997. Lack of Founding Amerindian Mitochondrial DNA Lineages in Extinct Aborigines from Tierra del Fuego-Patagonia. Human Molecular Genetics 6(1):41-46.

Lander E, Linton M, Birren B, Nusbaum C, Zody C, Baldwin J, Devon K, Dewar K, and et al. 2001. Initial sequencing and analysis of the human genome. Nature 409(6822):860–921.

Lao O, Lu TT, Nothnagel M, Junge O, Freitag-Wolf S, Caliebe A, Balascakova M, Bertranpetit J, Bindoff LA, Comas D et al. . 2008. Correlation between genetic and geographic structure in Europe. Curr Biol 18(16):1241-1248.

Larsen. 1994. In the wake of Columbus: native population biology in the postcontact Americas. Yearbook Phys Anthropol 37:109-154.

Lazaridis I, Patterson N, Mittnik A, Renaud G, Mallick S, Kirsanow K, Sudmant PH, Schraiber JG, Castellano S, Lipson M et al. . 2014. Ancient human genomes suggest three ancestral populations for present-day Europeans. Nature 513(7518):409-413.

Leonard JA, Wayne RK, and Cooper A. 2000. Population genetics of ice age brown bears. ProcNatlAcadSciUSA 97(4):1651-1654.

Lewis C, Lizarraga B, and Tito R. 2007. Mitochondrial DNA and the peopling of South America. Hum Biol 79:159 - 178.

Lindahl T. 1993. Instability and Decay of the Primary Structure of DNA. Nature 362(6422):709-715.

Lindahl T, and Nyberg B. 1972. Rate of depurination of native deoxyribonucleic acid. Biochemistry 11(19):3610-3618.

Liu DJ, and Leal SM. 2010. A novel adaptive method for the analysis of next-generation sequencing data to detect complex trait associations with rare variants due to gene main effects and interactions. PLoS Genet 6(10):e1001156.

Livi-Bacci M. 2006. The Depopulation of Hispanic America after the Conquest. Population and Development Review 32(2):199-232.

Llorente MG, Jones ER, Eriksson A, Siska V, Arthur KW, Arthur JW, Curtis MC, Stock JT, Coltorti M, Pieruccini P et al. . 2015. Ancient Ethiopian genome reveals extensive Eurasian admixture throughout the African continent. Science.

Maca-Meyer N, Gonzalez AM, Pestano J, Flores C, Larruga JM, and Cabrera VM. 2003. Mitochondrial DNA transit between West Asia and North Africa inferred from U6 phylogeography. BMC Genet 4:15.

Malhi RS, Eshleman JA, Greenberg JA, Weiss DA, Schultz Shook BA, Kaestle FA, Lorenz JG, Kemp BM, Johnson JR, and Smith DG. 2002. The structure of diversity within New World mitochondrial DNA haplogroups: implications for the prehistory of North America. Am J Hum Genet 70(4):905-919.

Malhi RS, Kemp BM, Eshleman JA, Cybulski J, Smith DG, Cousins S, and Harry H. 2007. Mitochondrial haplogroup M discovered in prehistoric North Americans. Journal of Archaeological Science 34(4):642-648.

Malhi RS, Schultz BA, and Smith DG. 2001. Distribution of mitochondrial DNA lineages among Native American tribes of Northeastern North America. HumBiol 73(1):17-55.

Malhi RS, and Smith DG. 2002. Brief communication: Haplogroup X confirmed in prehistoric North America. Am J Phys Anthropol 119(1):84-86.

Malmström H, Gilbert MTP, Thomas MG, Brandström M, Storå J, Molnar P, Andersen PK, Bendixen C, Holmlund G, Götherström A et al. . 2009. Ancient DNA Reveals Lack of Continuity between Neolithic Hunter-Gatherers and Contemporary Scandinavians. Current Biology 19(20):1758-1762.

Malmstrom H, Stora J, Dalen L, Holmlund G, and Gotherstrom A. 2005. Extensive Human DNA Contamination in Extracts from Ancient Dog Bones and Teeth. Mol Biol Evol 22(10):2040-2047.

Mamanova L, Coffey AJ, Scott CE, Kozarewa I, Turner EH, Kumar A, Howard E, Shendure J, and Turner DJ. 2010. Target-enrichment strategies for next-generation sequencing. Nat Meth 7(2):111-118.

Mandryk CAS, Josenhans H, Fedje DW, and Mathewes RW. 2001. Late Quaternary paleoenvironments of Northwestern North America: implications for inland versus coastal migration routes. Quaternary Science Reviews 20(1–3):301-314.

Mann CC. 2005. 1491: New Revelations of the Americas Before Columbus. New York, NY: Knopf.

Marangoni A, Caramelli D, and Manzi G. 2013. Homo sapiens in the Americas. Overview of the earliest human expansion in the New World. Journal of anthropological sciences= Rivista di antropologia: JASS/Istituto italiano di antropologia 92:79-97.

Mardis ER. 2008. The impact of next-generation sequencing technology on genetics. Trends in Genetics 24(3):133-141.

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z et al. . 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature 437(7057):376-380.

Maricic T, Whitten M, and Paabo S. 2010. Multiplexed DNA Sequence Capture of Mitochondrial Genomes Using PCR Products. PLoS ONE 5(11):e14004.

Marota I, Basile C, Ubaldi M, and Rollo F. 2002. DNA decay rate in papyri and human remains from Egyptian archaeological sites. Am J Phys Anthropol 117(4):310-318.

Martin PS. 1984. Prehistoric overkill: The global model. In PS Martin & RG Klein (eds):Quaternary extinctions: A prehistoric revolution, University of Arizona Press, Tucson.

Martínez-Cruz B, Harmant C, Platt DE, Haak W, Manry J, Ramos-Luis E, Soria-Hernanz DF, Bauduer F, Salaberria J, Oyharçabal B et al. . 2012. Evidence of pre-Roman tribal genetic structure in Basques from uniparentally inherited markers. Molecular Biology and Evolution.

Martínez-Cruz B, Ziegle J, Sanz P, Sotelo G, Anglada R, Plaza S, and Comas D. 2011. Multiplex single-nucleotide polymorphism typing of the human Y chromosome using TaqMan probes. Investig Genet 2:13-13.

Mata-Míguez J, Overholtzer L, Rodríguez-Alegría E, Kemp BM, and Bolnick DA. 2012. The genetic impact of aztec imperialism: Ancient mitochondrial DNA evidence from Xaltocan, Mexico. Am J Phys Anthropol 149(4):504-516.

Meltzer DJ. 1993. Pleistocene peopling of the Americas. Evol Anthropol 1::157–169.
Meltzer DJ. 1997. Monte Verde and the Pleistocene Peopling of the Americas. Science 276(5313):754-755.
Meltzer DJ. 2009. First peoples in a New World. Colonizing Ice Age America. University of California Press London.
Merriwether DA, Hodgson JA, Friedlaender FR, Allaby R, Cerchio S, Koki G, and Friedlaender JS. 2005. Ancient mitochondrial M haplogroups identified in the Southwest Pacific. Proc Natl Acad Sci U S A 102(37):13034-13039.
Merriwether DA, Rothhammer F, and Ferrel RE. 1995. Distribution of the four founding lineage haplotypes in native Americans suggests a single wave of migration for the New World. J Phys Anthropol 98::411-430.
Mertes F, Elsharawy A, Sauer S, van Helvoort JM, van der Zaag PJ, Franke A, Nilsson M, Lehrach H, and Brookes AJ. 2011. Targeted enrichment of genomic DNA regions for next-generation sequencing. Brief Funct Genomics.
Metzker ML. 2010. Sequencing technologies - the next generation. Nature reviews Genetics 11(1):31-46.
Meyer M, Fu Q, Aximu-Petri A, Glocke I, Nickel B, Arsuaga JL, Martinez I, Gracia A, de Castro JM, Carbonell E et al. . 2014. A mitochondrial genome sequence of a hominin from Sima de los Huesos. Nature 505(7483):403-406.
Meyer M, and Kircher M. 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harb Protoc 2010(6):pdb prot5448.
Meyer M, Kircher M, Gansauge T, Racimo F, Prüfer K, de Filippo C, Fu Q, Siebauer M, Stenzel U, and al. e. 2012. A high-coverage genome sequence from an archaic denisovan individual. Science 1224344:Published online August 30, 2012.
Millar CD, Huynen L, Subramanian S, Mohandesan E, and Lambert DM. 2008. New developments in ancient genomics. Trends in ecology & evolution 23(7):386-393.
Miller W, Drautz DI, Janecka JE, Lesk AM, Ratan A, Tomsho LP, Packard M, Zhang Y, McClellan LR, Qi J et al. . 2009. The mitochondrial genome sequence of the Tasmanian tiger (Thylacinus cynocephalus). Genome Research 19(2):213-220.
Miller W, Drautz DI, Ratan A, Pusey B, Qi J, Lesk AM, Tomsho LP, Packard MD, Zhao F, Sher A et al. . 2008. Sequencing the nuclear genome of the extinct woolly mammoth. Nature 456(7220):387-390.
Miller W, Schuster SC, Welch AJ, Ratan A, Bedoya-Reina OC, Zhao F, Kim HL, Burhans RC, Drautz DI, Wittekindt NE et al. . 2012. Polar and brown bear genomes reveal ancient admixture and demographic footprints of past climate change. Proc Natl Acad Sci U S A 109(36):E2382-2390.
Mitchell KJ, Llamas B, Soubrier J, Rawlence NJ, Worthy TH, Wood J, Lee MS, and Cooper A. 2014. Ancient DNA reveals elephant birds and kiwi are sister taxa and clarifies ratite bird evolution. Science 344(6186):898-900.
Montiel R, Solorzano E, Diaz N, Alvarez-Sandoval BA, Gonzalez-Ruiz M, and et al. 2012. Neonate Human Remains: A Window of Opportunity to the Molecular Study of Ancient Syphilis. PLoS ONE 7(5):e36371. doi:36310.31371/journal.pone.0036371.
Moraga M, Rocco P, and Miquel J. 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.
Moraga M, Santoro C, Standen V, Carvallo P, and Rothhammer F. 2005. Microevolution in prehistoric Andean populations: chronologic mtDNA variation in the desert valleys of northern Chile. Am J Phys Anthropol 127:170 - 181.

Morgan PD. 1997. The cultural implications of the Atlantic slave trade: African regional origins, American destinations and new world developments. Slavery & Abolition 18(1):122-145.

Mulligan C, Kitchen A, and Miyamoto M. 2008. Updated Three-Stage Model for the Peopling of the Americas. PLoS ONE 3(9):e3199. doi:3110.1371/ journal.pone.0003199.

Mulligan CJ, Hunley K, Cole S, and Long JC. 2004. Population genetics, history, and health patterns in native americans. Annu Rev Genomics Hum Genet 5:295-315.

Mullis KB, and Faloona FA. 1987. Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. Methods Enzymol 155:335-350.

Neves WA, and Hubbe M. 2005. Cranial morphology of early Americans from Lagoa Santa, Brazil: implications for the settlement of the New World. Proc Natl Acad Sci U S A 102(51):18309-18314.

Neves WA, and Pucciarelli H. 1991. Morphological affinities of the first Americans: an exploratory analysis based on early South American human remains. J Hum Evol 21:261-273.

Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, Shaffer T, Wong M, Bhattacharjee A, Eichler EE et al. . 2009. Targeted capture and massively parallel sequencing of 12 human exomes. Nature 461(7261):272-276.

Noonan JP, Coop G, Kudaravalli S, Smith D, Krause J, Alessi J, Chen F, Platt D, Paabo S, Pritchard JK et al. . 2006. Sequencing and analysis of Neanderthal genomic DNA. Science 314(5802):1113-1118.

Noonan JP, Hofreiter M, Smith D, Priest JR, Rohland N, Rabeder G, Krause J, Detter JC, Paabo S, and Rubin EM. 2005. Genomic sequencing of Pleistocene cave bears. Science 309(5734):597-599.

O'Fallon BD, and Fehren-Schmitz L. 2011. Native Americans experienced a strong population bottleneck coincident with European contact. Proc Natl Acad Sci U S A.

O'Rourke DH, and Raff JA. 2010. The human genetic history of the Americas: the final frontier. Curr Biol 20(4):R202-207.

Olalde I, Allentoft ME, Sanchez-Quinto F, Santpere G, Chiang CW, DeGiorgio M, Prado-Martinez J, Rodriguez JA, Rasmussen S, Quilez J et al. . 2014. Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. Nature 507(7491):225-228.

Olivio PD, Van de Walle, M. J., Laipis, P. J. & Hauswirth, W. W. . 1983. Nucleotide sequence evidence for rapid genotypic shifts in the bovine mitochondrial DNA D-loop. Nature 306, 400±402

Oota H, Saitou N, Matsushita T, and Ueda S. 1999. Molecular genetic analysis of remains of a 2,000-year-old human population in China-and its relevance for the origin of the modern Japanese population. AmJHumGenet 64(1):250-258.

Orlando L, Ginolhac A, Raghavan M, Vilstrup J, Rasmussen M, Magnussen K, Steinmann KE, Kapranov P, Thompson JF, Zazula G et al. . 2011. True single-molecule DNA sequencing of a pleistocene horse bone. Genome research 21(10):1705-1719.

Ovchinnikov IV, Götherström A, Romanova GP, Kharitonov VM, and Liden K. 2000. Molecular analysis of Neanderthal DNA from the northern Caucasus. Nature 404:490–493. .

Overballe-Petersen S, Orlando L, and Willerslev E. 2012. Next-generation sequencing offers new insights into DNA degradation. Trends Biotechnol 30(7):364-368.

Owsley DW, and Jantz RL. 2014. Kennewick Man. Texas A&M Univ Press.

Paabo S. 1985. Molecular cloning of Ancient Egyptian mummy DNA. Nature 314(6012):644-645.
Paabo S. 1989. Ancient DNA: Extraction, characterization, molecular cloning, and enzymatic amplification. Proc Natl Acad Sci USA 86:1939-1943.
Paabo S, Higuchi RG, and Wilson AC. 1989. Ancient DNA and the polymerase chain reaction. The emerging field of molecular archaeology. J Biol Chem 264(17):9709-9712.
Pääbo S, Poinar H, Serre D, Jaenicke-Després V, Hebler J, Rohland N, Kuch M, Krause J, Vigilant L, and Hofreiter M. 2004. Genetic Analyses from Ancient DNA. Annual Review of Genetics 38(1):645-679.
Paabo S, and Wilson AC. 1988. Polymerase chain reaction reveals cloning artefacts. Nature 334(6181):387-388.
Paijmans JL, Gilbert MT, and Hofreiter M. 2012. Mitogenomic analyses from ancient DNA. Mol Phylogenet Evol.
Pakendorf B, and Stoneking M. 2005. Mitochondrial DNA and human evolution. Annu Rev Genomics Hum Genet 6:165-183.
Palmer CA. 1995. From Africa to the Americas: Ethnicity in the Early Black Communities of the Americas. Journal of World History 6(2):223-236.
Pareek CS, Smoczynski R, and Tretyn A. 2011. Sequencing technologies and genome sequencing. J Appl Genet 52(4):413-435.
Patterson KB, and Runge T. 2002. Smallpox and the Native American. The American Journal of the Medical Sciences 323(4):216-222.
Perego U, Achilli A, Angerhofer N, Accetturo M, Pala M, Olivieri A, Hooshiar Kashani B, Ritchie K, Scozzari R, Kong Q et al. . 2009. Distinctive Paleo-Indian migration routes from Beringia marked by two rare mtDNA haplogroups. Curr Biol 19:1-8.
Perego UA, Angerhofer N, Pala M, Olivieri A, Lancioni H, Kashani BH, Carossa V, Ekins JE, Gomez-Carballa A, Huber G et al. . 2010. The initial peopling of the Americas: a growing number of founding mitochondrial genomes from Beringia. Genome Research 20(9):1174-1179.
Pereira L, Prata MJ, and Amorim A. 2000. Diversity of mtDNA lineages in Portugal: not a genetic edge of European variation. Ann Hum Genet 64(6):491-506.
Pereira L, Silva NM, Franco-Duarte R, Fernandes V, Pereira JB, Costa MD, Martins H, Soares P, Behar DM, Richards MB et al. . 2010. Population expansion in the North African late Pleistocene signalled by mitochondrial DNA haplogroup U6. BMC Evol Biol 10(390):1471-2148.
Perez SI, Bernal V, Gonzalez PN, Sardi M, and Politis GG. 2009. Discrepancy between cranial and DNA data of early Americans: implications for American peopling. PLoS ONE 4(5):0005746.
Poinar H. 2003. The top 10 list: criteria of authenticity for DNA from ancient and forensic samples. Int Congr Ser 1239:575 - 579.
Poinar HN, Hofreiter M, Spaulding WG, Martin PS, Stankiewicz BA, Bland H, Evershed RP, Possnert G, and Paabo S. 1998. Molecular coproscopy: dung and diet of the extinct ground sloth Nothrotheriops shastensis. Science 281(5375):402-406.
Poinar HN, Schwarz C, Qi J, Shapiro B, Macphee RD, Buigues B, Tikhonov A, Huson DH, Tomsho LP, Auch A et al. . 2006. Metagenomics to paleogenomics: large-scale sequencing of mammoth DNA. Science 311(5759):392-394.
Powell JF, and Neves WA. 1999. Craniofacial morphology of the first Americans: Pattern and process in the peopling of the New World. Am J Phys Anthropol 29:153-188.

Prüfer K, and Meyer M. 2015. Comment on “Late Pleistocene human skeleton and mtDNA link Paleoamericans and modern Native Americans”. Science 347(6224):835.

Prufer K, Racimo F, Patterson N, Jay F, Sankararaman S, Sawyer S, Heinze A, Renaud G, Sudmant PH, de Filippo C et al. . 2014. The complete genome sequence of a Neanderthal from the Altai Mountains. Nature 505(7481):43-49.

Pruvost M, Schwarz R, Correia V, Champlot S, Braguier S, Morel N, Fernandez-Jalvo Y, Grange T, and Geigl E. 2007. Freshly excavated fossil bones are best for amplifcation of ancient DNA. Proccedings of the National Academy of Science USA 104(3):739 - 744.

Pucciarelli HM. 2004. Migraciones y variacion craniofacial humana en America. Complutum 15: 225–247.

Quintana-Murci L, Semino O, Bandelt HJ, Passarino G, McElreavey K, and Santachiara-Benerecetti AS. 1999. Genetic evidence of an early exit of Homo sapiens sapiens from Africa through eastern Africa. Nat Genet 23(4):437-441.

Raff JA, Bolnick DA, Tackney J, and O'Rourke DH. 2011. Ancient DNA perspectives on American colonization and population history. Am J Phys Anthropol.

Raghavan M, DeGiorgio M, Albrechtsen A, Moltke I, Skoglund P, Korneliussen TS, Gronnow B, Appelt M, Gullov HC, Friesen TM et al. . 2014a. The genetic prehistory of the New World Arctic. Science 345(6200):1255832.

Raghavan M, Skoglund P, Graf KE, Metspalu M, Albrechtsen A, Moltke I, Rasmussen S, Stafford TW, Jr., Orlando L, Metspalu E et al. . 2014b. Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. Nature 505(7481):87-91.

Raghavan M, Steinrücken M, Harris K, Schiffels S, Rasmussen S, DeGiorgio M, Albrechtsen A, Valdiosera C, Ávila-Arcos MC, Malaspinas A-S et al. . 2015. Genomic evidence for the Pleistocene and recent population history of Native Americans. Science.

Ramirez O, Gigli E, Bover P, Alcover JA, Bertranpetit J, Castresana J, and Lalueza-Fox C. 2009. Paleogenomics in a temperate environment: shotgun sequencing from an extinct mediterranean caprine. Plos One 4(5):e5670.

Rasmussen M, Anzick SL, Waters MR, Skoglund P, DeGiorgio M, Stafford TW, Jr., Rasmussen S, Moltke I, Albrechtsen A, Doyle SM et al. . 2014. The genome of a Late Pleistocene human from a Clovis burial site in western Montana. Nature 506(7487):225-229.

Rasmussen M, Guo X, Wang Y, Lohmueller KE, Rasmussen S, Albrechtsen A, Skotte L, Lindgreen S, Metspalu M, Jombart T et al. . 2011. An Aboriginal Australian Genome Reveals Separate Human Dispersals into Asia. Science.

Rasmussen M, Li Y, Lindgreen S, Pedersen JS, Albrechtsen A, Moltke I, Metspalu M, Metspalu E, Kivisild T, Gupta R et al. . 2010. Ancient human genome sequence of an extinct Palaeo-Eskimo. Nature 463(7282):757-762.

Rasmussen M, Sikora M, Albrechtsen A, Korneliussen TS, Moreno-Mayar JV, Poznik GD, Zollikofer CPE, Ponce de Leon MS, Allentoft ME, Moltke I et al. . 2015a. The ancestry and affiliations of Kennewick Man. Nature 523(7561):455-458.

Rasmussen S, Allentoft Morten E, Nielsen K, Orlando L, Sikora M, Sjögren K-G, Pedersen Anders G, Schubert M, Van Dam A, Kapel Christian Moliin O et al. . 2015b. Early Divergent Strains of Yersinia pestis in Eurasia 5,000 Years Ago. Cell 163(3):571-582.

Reich D, Green RE, Kircher M, Krause J, Patterson N, Durand EY, Viola B, Briggs AW, Stenzel U, Johnson PLF et al. . 2010. Genetic history of an archaic hominin group from Denisova Cave in Siberia. Nature 468(7327):1053-1060.

Reich D, Patterson N, Campbell D, Tandon A, Mazieres S, Ray N, Parra MV, Rojas W, Duque C, Mesa N et al. . 2012. Reconstructing Native American population history. Nature 488(7411):370-374.

Reich D, Patterson N, Kircher M, Delfin F, Nandineni MR, Pugach I, Ko AM-S, Ko Y-C, Jinam TA, Phipps ME et al. . 2011. Denisova Admixture and the First Modern Human Dispersals into Southeast Asia and Oceania. Am J Hum Genet.

Renfrew C. 2010. Archaeogenetics--towards a 'new synthesis'? Curr Biol 20(4):R162-165.

Ribeiro dos Santos A, Santos S, Machado A, Guapindaia V, and Zago M. 1996. Heterogeneity of mitochondrial DNA haplotypes in pre-columbian natives of the Amazon region. Am J Phys Anthropol 101:29-37.

Rizzi E, Lari M, Gigli E, De Bellis G, and Caramelli D. 2012. Ancient DNA studies: new perspectives on old samples. Genet Sel Evol 44:21.

Rodriguez-Delfin L, Rubin-de-Celis V, and Zago M. 2001. Genetic diversity in an Andean population from Peru and regional migration patterns of Amerindians in South America: data from Y chromosome and mitochondrial DNA. Hum Hered 51:97 - 106.

Rohland N, Malaspinas AS, Pollack JL, Slatkin M, Matheus P, and Hofreiter M. 2007. Proboscidean mitogenomics: chronology and mode of elephant evolution using mastodon as outgroup. PLoS Biology 5(8):e207.

Rohland N, Pollack JL, Nagel D, Beauval C, Airvaux J, Paabo S, and Hofreiter M. 2005. The Population History of Extant and Extinct Hyenas. MolBiolEvol.

Rollo F, Ermini L, Luciani S, Marota I, Olivieri C, and Luiselli D. 2006. Fine characterization of the Iceman's mtDNA haplogroup. Am J Phys Anthropol 130:557-564.

Rothhammer F, and Dillehay TD. 2009. The late Pleistocene colonization of South America: an interdisciplinary perspective. Ann Hum Genet 73(Pt 5):540-549.

Rothhammer F, Llop E, Carvallo P, and Moraga M. 2001. Origin and Evolutionary Relationships of Native Andean Populations. High Altitude Medicine and Biology 2((2)):227-233.

Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich HA, and Arnheim N. 1985. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science 230(4732):1350-1354.

Salas A, Richards M, De la Fe T, Lareu M-V, Sobrino B, Sánchez-Diz P, Macaulay V, and Carracedo Á. 2002. The Making of the African mtDNA Landscape. Am J Hum Genet 71(5):1082-1111.

Salo W, Aufderheide A, Buikstra J, and Holcomb T. 1994. Identification of Mycobacterium tuberculosis DNA in a pre-Columbian Peruvian mummy. Proc Natl Acad Sci U S A Mar 15;91(6):2091-4.

Sampietro ML, Caramelli D, Lao O, Calafell F, Comas D, Lari M, Agusti B, Bertranpetit J, and Lalueza-Fox C. 2006. The genetics of pre-Roman Iberian Peninsula: A mtDNA study of ancient Iberians. International Congress Series 1288:142-144.

Sanchez-Quinto F, Schroeder H, Ramirez O, Avila-Arcos MC, Pybus M, Olalde I, Velazquez AM, Marcos ME, Encinas JM, Bertranpetit J et al. . 2012. Genomic Affinities of Two 7,000-Year-Old Iberian Hunter-Gatherers. Current biology 22(16):1494-1499.

Sanchez JJ, Borsting C, Balogh K, Berger B, Bogus M, Butler JM, Carracedo A, Court DS, Dixon LA, Filipovic B et al. . 2008. Forensic typing of autosomal SNPs with a 29 SNP-multiplex-Results of a collaborative EDNAP exercise. Forensic Science International Genetics 2(3):176-183.

Sanchez JJ, Borsting C, Hallenberg C, Buchard A, Hernandez A, and Morling N. 2003. Multiplex PCR and minisequencing of SNPs--a model with 35 Y chromosome SNPs. Forensic Sci Int 137(1):74-84.

Sanchez JJ, and Endicott P. 2006. Developing multiplexed SNP assays with special reference to degraded DNA templates. Nature Protocols 1(3):1370-1378.

Sanchez JJ, Phillips C, Borsting C, Balogh K, Bogus M, Fondevila M, Harrison CD, Musgrave-Brown E, Salas A, Syndercombe-Court D et al. . 2006. A multiplex assay with 52 single nucleotide polymorphisms for human identification. Electrophoresis 27(9):1713-1724.

Sandoval JR, Lacerda DR, Jota MSA, Salazar-Granara A, Vieira PPR, and al. e. 2013. The Genetic History of Indigenous Populations of the Peruvian and Bolivian Altiplano: The legacy of the Uros. PLoS ONE 8(9):e73006.

Sanger F, Nicklen S, and Coulson AR. 1977. DNA sequencing with chain-terminating inhibitors. ProcNatlAcadSciUSA 74(12):5463-5467.

Sawyer S, Krause J, Guschanski K, Savolainen V, and Paabo S. 2012. Temporal Patterns of Nucleotide Misincorporations and DNA Fragmentation in Ancient DNA. Plos One 7(3):e34131. doi:34110.31371/journal.pone.0034131.

Schroeder H, Ávila-Arcos MC, Malaspinas A-S, Poznik GD, Sandoval-Velasco M, Carpenter ML, Moreno-Mayar JV, Sikora M, Johnson PLF, Allentoft ME et al. . 2015. Genome-wide ancestry of 17th-century enslaved Africans from the Caribbean. Proceedings of the National Academy of Sciences.

Schroeder KB, Jakobsson M, Crawford MH, Schurr TG, Boca SM, Conrad DF, Tito RY, Osipova LP, Tarskaia LA, Zhadanov SI et al. . 2009. Haplotypic Background of a Private Allele at High Frequency in the Americas. Molecular Biology and Evolution 26(5):995-1016.

Schroeder KB, Schurr TG, Long JC, Rosenberg NA, Crawford MH, Tarskaia LA, Osipova LP, Zhadanov SI, and Smith DG. 2007. A private allele ubiquitous in the Americas. Biol Lett Epub:218–223.

Schubert M, Ermini L, Der Sarkissian C, Jonsson H, Ginolhac A, Schaefer R, Martin MD, Fernandez R, Kircher M, McCue M et al. . 2014. Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. Nature Protocols 9(5):1056-1082.

Schuenemann VJ, Bos K, Dewitte S, Schmedes S, Jamieson J, Mittnik A, Forrest S, Coombes BK, Wood JW, Earn DJ et al. . 2011. Targeted enrichment of ancient pathogens yielding the pPCP1 plasmid of Yersinia pestis from victims of the Black Death. Proc Natl Acad Sci U S A.

Schurr T. 2002. Molecular Anthropological Perspective of the Peopling of the Americas. Athenea Review Vol.3, No.2

Schurr TG. 2008. The Peopling of the Americas as Revealed by Molecular Genetic Studies.

Schurr TG, Ballinger SW, Gan YY, Hodge JA, Merriwether DA, Lawrence DN, Knowler WC, Weiss KM, and Wallace DC. 1990. Amerindian mitochondrial DNAs have rare Asian mutations at high frequencies, suggesting they derived from four primary maternal lineages. Am J Hum Genet 46(3):613-623.

Schurr TG, and Sherry ST. 2004. Mitochondrial DNA and Y chromosome diversity and the peopling of the Americas: evolutionary and demographic evidence. Am J Hum Biol 16(4):420-439.

Seguin-Orlando A, Korneliussen TS, Sikora M, Malaspinas AS, Manica A, Moltke I, Albrechtsen A, Ko A, Margaryan A, Moiseyev V et al. . 2014. Paleogenomics. Genomic structure in Europeans dating back at least 36,200 years. Science 346(6213):1113-1118.

Seidenberg V, Schilz F, Pfister D, Georges L, Fehren-Schmitz L, and Hummel S. 2012. A new miniSTR heptaplex system for genetic fingerprinting of ancient DNA from archaeological human bone. Journal of Archaeological Science 39(10):3224-3229.

Shapiro B, and Hofreiter M. 2010. Analysis of ancient human genomes: using next generation sequencing, 20-fold coverage of the genome of a 4,000-year-old human from Greenland has been obtained. BioEssays : news and reviews in molecular, cellular and developmental biology 32:388-391.

Shendure J, and Ji H. 2008. Next-generation DNA sequencing. Nat Biotechnol 26(10):1135-1145.

Sheng GL, Soubrier J, Liu JY, Werdelin L, Llamas B, Thomson VA, Tuke J, Wu LJ, Hou XD, Chen QJ et al. . 2014. Pleistocene Chinese cave hyenas and the recent Eurasian history of the spotted hyena, Crocuta crocuta. Mol Ecol 23(3):522-533.

Shimada I, Shinoda K, Farnum J, Corruccini R, and Watanabe H. 2004. An integrated analysis of pre-Hispanic mortuary practices: a Middle Sican case study. Curr Anthropol 45:369–402.

Shinoda K, Adachi N, Guillen S, and Shimada I. 2006. Mitochondrial DNA analysis of ancient Peruvian highlanders. Am J Phys Anthropol 131(1):98-107.

Silva WA, Jr., Bonatto SL, Holanda AJ, Ribeiro-Dos-Santos AK, Paixao BM, Goldman GH, Abe-Sandes K, Rodriguez-Delfin L, Barbosa M, Paco-Larson ML et al. . 2002. Mitochondrial genome diversity of Native Americans supports a single early entry of founder populations into America. Am J Hum Genet 71(1):187-192.

Skoglund P, Malmström H, Raghavan M, Storå J, Hall P, Willerslev E, Gilbert MTP, Götherström A, and Jakobsson M. 2012. Origins and Genetic Legacy of Neolithic Farmers and Hunter-Gatherers in Europe. Science 336(6080):466-469.

Skoglund P, Mallick S, Bortolini MC, Chennagiri N, Hunemeier T, Petzl-Erler ML, Salzano FM, Patterson N, and Reich D. 2015. Genetic evidence for two founding populations of the Americas. Nature advance online publication.

Steward HJ, and Faron LC. 1959. Native peoples of South America. McGraw-Hill Co(New York).

Stewart JB, and Chinnery PF. 2015. The dynamics of mitochondrial DNA heteroplasmy: implications for human health and disease. Nat Rev Genet 16(9):530-542.

Stiller M, Green RE, Ronan M, Simons JF, Du L, He W, Egholm M, Rothberg JM, Keates SG, Ovodov ND et al. . 2006. Patterns of nucleotide misincorporations during enzymatic amplification and direct large-scale sequencing of ancient DNA. Proc Natl Acad Sci U S A 103(37):13578-13584.

Stone AC, and Stoneking M. 1993. Ancient DNA from a pre-Columbian Amerindian population. AmJPhysAnthropol 92(4):463-471.

Stone AC, and Stoneking M. 1998. mtDNA analysis of a prehistoric Oneota population: implications for the peopling of the New World 20. AmJHumGenet 62(5):1153-1170.

Stone AC, and Stoneking M. 1999. Analysis of ancient DNA from a prehistoric Amerindian cemetery. PhilosTransRSocLond B BiolSci 354(1379):153-159.

Stoneking M, Hedgecock D, Higuchi RG, Vigilant L, and Erlich HA. 1991. Population variation of human mtDNA control region sequences detected by enzymatic amplification and sequence-specific oligonucleotide probes. AmJHumGenet 48(2):370-382.

Taboada-Echalar P, Alvarez-Iglesias V, Heinz T, Vidal-Bralo L, Gomez-Carballa A, Catelli L, Pardo-Seco J, Pastoriza A, Carracedo A, Torres-Balanza A et al. . 2013. The genetic legacy of the pre-colonial period in contemporary Bolivians. PLoS ONE 8(3):e58980.

Tamm E, Kivisild T, Reidla M, Metspalu M, and Smith D. 2007. Beringian standstill and spread of Native American founders. PLoS ONE 2(9):e829.

Taylor MA, Hendy IL, and Pak DK. 2014. Deglacial ocean warming and marine margin retreat of the Cordilleran Ice Sheet in the North Pacific Ocean. Earth and Planetary Science Letters 403(0):89-98.

The 1000 Genomes Project Consortium. 2010. A map of human genome variation from population-scale sequencing. Nature 467(7319):1061-1073.

Thomas RH, Schaffner W, Wilson AC, and Paabo S. 1989. DNA phylogeny of the extinct marsupial wolf 83. Nature 340(6233):465-467.

Thomson VA, Lebrasseur O, Austin JJ, Hunt TL, Burney DA, Denham T, Rawlence NJ, Wood JR, Gongora J, Girdland Flink L et al. . 2014. Using ancient DNA to study the origins and dispersal of ancestral Polynesian chickens across the Pacific. Proc Natl Acad Sci U S A 111(13):4826-4831.

Tishkoff SA, Gonder MK, Henn BM, Mortensen H, Knight A, Gignoux C, Fernandopulle N, Lema G, Nyambo TB, Ramakrishnan U et al. . 2007. History of Click-Speaking Populations of Africa Inferred from mtDNA and Y Chromosome Genetic Variation. Molecular Biology and Evolution 24(10):2180-2195.

Torroni A, Huoponen K, Francalacci P, Petrozzi M, Morelli L, Scozzari R, Obinu D, Savontaus ML, and Wallace DC. 1996. Classification of European mtDNAs from an analysis of three European populations. Genetics 144(4):1835-1850.

Torroni A, Lott MT, Cabell MF, Chen YS, Lavergne L, and Wallace DC. 1994a. mtDNA and the origin of Caucasians: identification of ancient Caucasian-specific haplogroups, one of which is prone to a recurrent somatic duplication in the D-loop region. Am J Hum Genet 55(4):760-776.

Torroni A, Neel JV, Barrantes R, Schurr TG, and Wallace DC. 1994b. Mitochondrial-DNA Clock for the Amerinds and Its Implications for Timing Their Entry into North-America. Proc Natl Acad Sci U S A 91(3):1158-1162.

Torroni A, Schurr T, Cabell M, Brown M, Neel J, Larsen M, Smith D, Vullo C, and Wallace D. 1993. Asian affinities and continental radiation of the four founding Native American mtDNAs. Am J Hum Genet 53:563–590.

Torroni A, Schurr TG, Yang CC, Szathmary EJ, Williams RC, Schanfield MS, Troup GA, Knowler WC, Lawrence DN, Weiss KM et al. . 1992. Native American mitochondrial DNA analysis indicates that the Amerind and the Nadene populations were founded by two independent migrations. Genetics 130(1):153-162.

van Oven M, and Kayser M. 2009. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Hum Mutat 30(2):E386-394.

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA et al. . 2001. The sequence of the human genome. Science 291(5507):1304-1351.

Vernesi C, Caramelli D, Dupanloup I, Bertorelle G, Lari M, Cappellini E, Moggi-Cecchi J, Chiarelli B, Castrì L, Casoli A et al. . 2004. The Etruscans: A Population-Genetic Study. Am J Hum Genet 74(4):694-704.

Vigilant L, Stoneking M, Harpending H, Hawkes K, and Wilson AC. 1991. African populations and the evolution of human mitochondrial DNA. Science 253(5027):1503-1507.

von Hunnius TE, Yang D, Eng B, Waye JS, and Saunders SR. 2007. Digging deeper into the limits of ancient DNA research on syphilis. Journal of Archaeological Science 34(12):2091-2100.

Wall JD, and Kim SK. 2007. Inconsistencies in Neanderthal genomic DNA sequences. Plos Genetics 3(10):1862-1866.

Wang S, Lewis CM, Jakobsson M, Ramachandran S, Ray N, Bedoya G, Rojas W, Parra MV, Molina JA, Gallo C et al. . 2007. Genetic variation and population structure in native Americans. PLoS Genetics 3(11):e185.

Warinner C, Rodrigues JF, Vyas R, Trachsel C, Shved N, Grossmann J, Radini A, Hancock Y, Tito RY, Fiddyment S et al. . 2014. Pathogens and host immunity in the ancient human oral cavity. Nat Genet 46(4):336-344.

Waters MR, and Stafford TW, Jr. 2007. Redefining the age of Clovis: implications for the peopling of the Americas. Science 315(5815):1122-1126.

Watson JD, and Crick FH. 1953. Molecular structure of nucleic acids; a structure for desoxyribose nucleic acid. Nature 171:737-738.

Wheeler DA, Srinivasan M, Egholm M, Shen Y, Chen L, McGuire A, He W, Chen YJ, Makhijani V, Roth GT et al. . 2008. The complete genome of an individual by massively parallel DNA sequencing. Nature 452(7189):872-876.

Willerslev E, and Cooper A. 2005. Ancient DNA. ProcBiolSci 272(1558):3-16.

Willerslev E, Hansen AJ, Binladen J, Brand TB, Gilbert MT, Shapiro B, Bunce M, Wiuf C, Gilichinsky DA, and Cooper A. 2003. Diverse plant and animal genetic records from Holocene and Pleistocene sediments. Science 300(5620):791-795.

Woodward SR, Weyand NJ, and Bunnell M. 1994. DNA sequence from Cretaceous period bone fragments. Science 266(5188):1229-1232.

Yang DY, and Watt K. 2005. Contamination controls when preparing archaeological remains for ancient DNA analysis. Journal of Archaeological Science 32(3):331-336.

Yang H, Golenberg EM, and and Shoshani J. 1996. Phylogenetic Resolution within the Elephantidae Using Fossil DNA Sequence from the American Mastodon (Mammut americanum) As an Outgroup. Proc Natl Acad Sci USA 93:1190–1194.

Yi X, Liang Y, Huerta-Sanchez E, Jin X, Cuo ZX, Pool JE, Xu X, Jiang H, Vinckenbosch N, Korneliussen TS et al. . 2010. Sequencing of 50 human exomes reveals adaptation to high altitude. Science 329(5987):75-78.

Zegura SL, Karafet TM, Zhivotovsky LA, and Hammer MF. 2004. High-resolution SNPs and microsatellite haplotypes point to a single, recent entry of Native American Y chromosomes into the Americas. Molecular Biology and Evolution 21(1):164-175.

Zeyland J, Wolko Ł, Bocianowski J, Szalata M, Słomski R, Dzieduszycki AM, Ryba M, Przystałowska H, and Lipiński D. 2013. Complete mitochondrial genome of wild aurochs (Bos primigenius) reconstructed from ancient DNA. Polish Journal of Veterinary Sciences 16(2).

Zink A, Haas CJ, Reischl U, Szeimies U, and Nerlich AG. 2001. Molecular analysis of skeletal tuberculosis in an ancient Egyptian population. JMedMicrobiol 50(4):355-366.

Zink AR, Grabner W, Reischl U, Wolf H, and Nerlich AG. 2003. Molecular study on human tuberculosis in three geographically distinct and time delineated populations from ancient Egypt. EpidemiolInfect 130(2):239-249.

Zischler H, Hîss M, Handt O, and von Haeseler A. 1995. Detecting Dinosaur DNA. Science 268:1192-1193.

CHAPTER 2

Ancient DNA from pre-Columbian populations in the Central Andes of South America: a diachronic study of mtDNA haplogroup diversity based on Hypervariable Region I

Statement of Authorship

Chapter 2

Title of the Paper	**Ancient DNA from pre-Columbian populations in the Central Andes of South America: a diachronic study of mtDNA haplogroup diversity based on Hypervariable Region I**
Publication status	Published, Accepted for Publication, Submitted for Publication, **Publication style**
Publication details	

Author Contributions
By signing the Statement of Authorship, each author certifies that their stated contribution to the publication is accurate and that permission is granted for the publication to be included in the candidate's thesis.

Name of the principal Author	**Guido Valverde** (Candidate)	
Contribution to the paper	Performed DNA extractions, PCR amplifications, sequencing and downstream data processing on all samples, analysed the data, interpreted the results, created figures and tables and wrote the manuscript.	
Signature		Date: 8/06/15

Name of the co-author	Alexandra Coutinho	
Contribution to the paper	Performed DNA extractions, PCR amplifications, sequencing and downstream data processing of samples.	
Signature		Date: 02/06/2015

Name of the co-author	María Inés Barreto	
Contribution to the paper	Contributed archaeological material. Assisted with sample collection, provided archaeological and context information.	
Signature		Date: May 29 2015

Name of the co-author	Isabel Flores Espinoza	
Contribution to the paper	Provided samples for analysis. Contributed archaeological material. Assisted with sample collection, provided archaeological and context information.	
Signature		Date 02.06.2015

Name of the co-author	Lucia Watson	
Contribution to the paper	Contributed archaeological material. Assisted with sample collection, provided archaeological and context information.	
Signature		Date: 05-30-2015

Name of the co-author	Krzysztof Makowski	
Contribution to the paper	Provided samples for analysis. Contributed archaeological material. Assisted with sample collection, provided archaeological and context information.	
Signature		Date: 06.10.2015

Name of the co-author	Calogero Santoro	
Contribution to the paper	Provided samples for analysis. Contributed archaeological material. Assisted with sample collection, provided archaeological and context information.	
Signature		Date: June 1, 2015

Name of the co-author	Fabricio R. Santos	
Contribution to the paper	Contributed reagents, materials, analysis tools, assisted in the interpretation of results.	
Signature		Date: May 30th, 2015

Name of the co-author	Alan Cooper	
Contribution to the paper	Contributed reagents, materials, analysis tools assisted in the interpretation of results.	
Signature		Date: October, 3 2015

Name of the co-author	Lars Fehren-Schmitz	
Contribution to the paper	Provided samples for analysis. Contributed reagents, materials, analysis tools, provided archaeological and context information. Contributed with data generation and assisted in the interpretation of results.	
Signature		Date: 06/01/2015

Name of the co-author	Bastien Llamas	
Contribution to the paper	Co-developed the research concept and experimental design of the study, provided archaeological samples, gave advice on laboratory work, assisted with the interpretation of results and edited the manuscript	
Signature		Date: October, 23 2015

Name of the co-author	Wolfgang Haak	
Contribution to the paper	Co-developed the research concept and experimental design of the study, provided archaeological samples, gave advice on laboratory work, assisted with the interpretation of results and edited the manuscript	
Signature		Date: 20/10/2015

Ancient DNA from pre-Columbian populations in the Central Andes of South America: a diachronic study of mtDNA haplogroup diversity based on Hypervariable Region I

Guido Valverde[1], Alexandra Coutinho[1], Maria Inés Barreto[2], Isabel Flores Espinoza[2], Lucia Watson[3,4], Krzysztof Makowski[4], Calogero Santoro[5,6], Fabricio R. Santos[7], Lars Fehren-Schmitz[8,9], Alan Cooper[1], Bastien Llamas[1], Wolfgang Haak[1]

[1]Australian Centre for Ancient DNA, School of Biological Sciences, The University of Adelaide, Adelaide, South Australia 5005, Australia

[2] Proyecto Arqueológico Huaca Pucllana, Perú

[3] Centro de Investigaciones Arqueológicas del Museo de Sitio de Ancón, Perú

[4] Departamento de Letras y Ciencias Humanas de la Pontificia Universidad Católica del Perú, Lima, Perú

[5] Instituto de Alta Investigación (IAI), Laboratorio de Arqueología y Paleoambiente, Universidad de Tarapacá (UTA), Antofagasta, Arica, Chile.

[6] Centro de Investigaciones del Hombre en el Desierto (CIHDE)

[7] Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, MG, Brazil

[8] Department of Anthropology, University of California, Santa Cruz, CA 95064

[9] Historical Anthropology and Human Ecology, University of Goettingen, D-37073 Goettingen, Germany

ABSTRACT

Genetic data generated from past and present-day populations are now routinely used to reconstruct the origins, migrations and relationships of human populations around the world. We analysed mitochondrial DNA (mtDNA) from pre-Columbian coastal and highland human remains in the Central Andes of South America, incorporating samples from archaeological periods (Late Archaic: 8000–1900 BC to Late Horizon: 1476–1534 AD), in order to evaluate changes in the maternal genetic diversity through time. We typed mtDNA haplogroups and haplotypes from 302 samples (149 ancient Native American individuals), using three different approaches, including sequencing of the Hypervariable Region I, a coding-region single-nucleotide polymorphism (SNP) multiplex assay (GenoCoRe22), and a novel continent-specific SNP multiplex assay (AmericaPlex26) designed to further resolve the four major American mtDNA haplogroups A, B, C, and D.

Sequencing and SNP-typing results were successfully and reproducibly obtained from 72 individuals, and assigned to one of the four major Native American founder haplogroups (A2, B2, C1, D1). This study is the first to incorporate archaeological samples from all cultural periods in the Central Andes available to date, and provides a temporal description of demographic events in ancient South America. Mitochondrial DNA haplogroup frequencies and genetic distances (F_{ST}) reveal a population differentiation in the Late Archaic Period and two temporal transitions in the subsequent archaeological Early Horizon and Early Intermediate periods. In addition, the data suggests another temporal transition for the three successive Middle Horizon, Late Intermediate and Late Horizon periods, suggesting a process of genetic assimilation between highland and coastal populations. While we evidenced fluctuations in the haplogroup composition of ancient populations, we observe no major differences to modern-day populations, suggesting general population continuity in the Central Andean region of South America.

KEYWORDS: ancient DNA, mitochondrial DNA, Native American haplogroups, Central Andes.

1. INTRODUCTION

The human settlement of the Americas has been a subject of debate over the past decades, with a specific focus on questions regarding the timing, entry routes and migratory waves (Greenberg et al. 1986; Horai et al. 1993; Torroni et al. 1994; Merriwether et al. 1995; Bonatto and Salzano 1997; Tamm et al. 2007; Wang et al. 2007; Fagundes et al. 2008; Kitchen et al. 2008; O'Rourke and Raff 2010; Reich et al. 2012).

Two demographic events have had a major impact on the genetic diversity of human populations in the Americas. First, a massive population bottleneck during the initial peopling which shaped the genetic makeup of the First Americans crossing the Beringian land bridge from Siberia (Horai et al. 1993; Torroni et al. 1994; Goebel et al. 2008). A second significant episode in the history of the continent took place ~500 years ago with the arrival of Europeans. The European contact with Native Americans resulted in a dramatic episode of population collapse (Livi-Bacci 2006), impacting on the effective female population size (O'Fallon and Fehren-Schmitz 2011) and potentially on the genetic diversity of Native Americans.

It has been suggested that this bottleneck explains the restricted genetic diversity amongst Native Americans expressed in only five major mitochondrial DNA (mtDNA) haplogroups: A–D (Schurr et al. 1990; Horai et al. 1993; Torroni et al. 1993; Bandelt et al. 2003; Perego et al. 2010) and X (Brown et al. 1998), with only a particular set of lineages found exclusively in the Americas, namely the Pan-American haplogroups: A2, B2, C1b, C1c, D1, and D4h3a (Tamm et al. 2007; Achilli et al. 2008), which are absent in Asian populations. This reduced diversity could be due to stochastic events during the settlement process over thousands of years (which was coined “incubation” period in Beringia by (Szathmáry 1981; Tamm et al. 2007), such as the genetic drift expected from bottlenecks and/or small population sizes (Tamm et al. 2007; O'Rourke and Raff 2010). Moreover, it is hypothesized that a rapid population expansion into the Americas (Mulligan et al. 2008), resulted in the establishment of successful mtDNA lineages, and the formation of the initial genetic diversity in the Americas.

A study on complete mtDNA genomes has refined the founder hypothesis and describes the mtDNA diversity in Native Americans more precisely as so-called 15 American founder lineages (A2, A2a, A2b, B2, C1b, C1c, C1d, C1d1, C4, D1, D2a, D3,

D4h3a, X2g, X2a)(Perego et al. 2010), which form basal sister-clades within the aforementioned America-specific haplogroups.

Although all Native American mtDNA haplogroups are widely distributed across the continent with the exception of haplogroup X, which is restricted to North America (Brown et al. 1998; Dornelles et al. 2005), there is genetic variation across space that might link to a particular haplogroup frequency composition to geographic regions within the Americas.

Frequencies of haplogroup A are generally high in North American populations (Tamm et al. 2007; Perego et al. 2010) and decrease considerably in southern latitudes (Lalueza et al. 1997). Haplogroup B is mainly found in populations from South America, with a high concentration (>50%) in the Andean region (Rodriguez-Delfin et al. 2001; Lewis et al. 2007; Afonso Costa et al. 2010; Barbieri et al. 2011; Gaya-Vidal et al. 2011; Sandoval JR et al. 2013; Taboada-Echalar et al. 2013). Haplogroup C is present in populations in Mesoamerica (Perego et al. 2010) and populations in north-western South America, while haplogroup D is characteristic of populations in the southern parts of South America (Bodner et al. 2012; de Saint Pierre et al. 2012).

Research on ancient DNA (aDNA) in the Americas has developed considerably over the last years – for a review see (Raff et al. 2011), but focused mainly on small but informative portions of the mtDNA (e.g. Control Region), typical for PCR-based aDNA studies. Ancient DNA data from South American populations has accumulated to a large number of samples (n~500) from ~20 sites (Fehren-Schmitz et al. 2011a), but is limited in genetic resolution (Shimada et al. 2004; Moraga et al. 2005; Shinoda et al. 2006; Fehren-Schmitz et al. 2009; Carnese et al. 2010; Fehren-Schmitz et al. 2011a; Baca et al. 2012; Baca et al. 2014; Mendisco et al. 2014). More recently, genome wide data from a large number of modern-day populations has been published (e.g. Reich et al. 2012), but also the first genomic data from ancient Native Americans (Rasmussen et al. 2010; Raghavan et al. 2014; Rasmussen et al. 2014; Rasmussen et al. 2015), however restricted to North America.

Studies on ancient human remains from South America in particular are providing new insights into human prehistory, and a direct means to reconstruct the population history in this part of the world. Ancient DNA has been used to test different hypotheses on population relationships, to infer genetic continuity and discontinuity

associated with cultural transitions (Kemp et al. 2009), kinship relationships and social organization of Pre-Columbian populations (Baca et al. 2012; Baca et al. 2014), and to reconstruct micro-evolutionary processes and history of ancient populations (Moraga et al. 2005; Mendisco et al. 2014). Ancient DNA studies have the potential to explore cultural evolution and population dynamics to elucidate history of pre-Columbian societies (Fehren-Schmitz et al. 2010; Fehren-Schmitz et al. 2011b). Diachronic studies of ancient populations from Peru have shown significant differentiation between coastal and highland populations over time (Fehren-Schmitz et al. 2011b). Also, aDNA studies have shed light on the role of other factors such as climate influences and climate variability in the Andean and coastal regions from Peru that have driven population demography and cultural transitions over time (Fehren-Schmitz et al. 2014).

However, while some aDNA studies in South America have attempted limited temporal sampling (Moraga et al. 2005; Kemp et al. 2009; Fehren-Schmitz et al. 2011b), most are restricted to samples from a particular archaeological site or period. To date, there is no comprehensive diachronic/chronological study incorporating samples across the many cultural archaeological periods in South America.

Here, we present a detailed study of mitochondrial genetic diversity across time periods in the Central Andes of South America incorporating a temporal sampling strategy in order to understand demographic processes undergone by ancient populations through time, testing for population continuity, discontinuity and genetic maternal relationships between ancient and modern populations.

The ancient samples analysed for this study span from the Late Archaic period (8000–1900 BC) to the Late Horizon Period (1476–1534 AD). We have compiled the largest ancient DNA dataset available in order to contrast samples from coastal and Andean populations with modern indigenous South American populations. For reasons of compatibility and direct comparison, our study focuses on the HVR-I of the mtDNA from well-defined ancient and modern populations. This type of data is not yet available for complete mitochondrial genome studies, since most of these studies largely focus on phylogenetic refinements rather than population genetic analyses.

2. MATERIAL AND METHODS.

2.1. Sample description and DNA extraction

We collected a total of 302 samples from pre-Columbian archaeological sites/cultures across the Central Andes belonging to 149 individuals.

A minimum of two independent samples per individual were collected for the purpose of replication and authentication of aDNA results (Figure 1, Table 1). Samples were processed in dedicated aDNA facilities at the Australian Centre for Ancient DNA (ACAD), University of Adelaide, Australia. The laboratory employs standardized aDNA protocols and infrastructure to carry out aDNA work (positive air pressure, UV irradiation and regular cleaning with oxidating agents e.g. commercial bleach and Decon® to minimize contamination) (Cooper and Poinar 2000; Fulton 2012; Knapp et al. 2012).

The aDNA-specific workflow used in this study is summarized here:

1) Samples (bone or tooth) were decontaminated upon entry to the aDNA laboratory by exposure to UV light. The surface of the samples was gently wiped with 3% sodium hypochlorite solution (bleach) and then physically removed by abrasion using a Dremel® drill. A Mikro-dismembrator ball mill (Sartorius) was used to pulverize the sample and 0.2 g of bone powder were subsequently used in DNA extractions.

2) Samples were decalcified by incubation in 4 mL of 0.5 M EDTA (pH 8.0) overnight at 37°C on a rotor at ~30 rpm. Next, 70 µL Proteinase K (Invitrogen) was added and the lysis mix was incubated for 2 hours at 55°C. DNA was isolated using silicon dioxide solubilised in a Guanidinium buffer (Qiagen), as described previously (Brotherton et al. 2013; Der Sarkissian et al. 2013). DNA was resuspended in 200µL of TE buffer including 0.05% Tween-20 and stored at -20ºC. Since samples from different cultural periods and times vary in quality and preservation, we applied three PCR-based mtDNA typing techniques and compare the overall performance and effectiveness of each method to deal with degraded DNA (Table 2 - 3).

2.2. PCR amplification and sequencing of mtDNA Hypervariable Region I

Four sets of PCR primers pairs were used to amplify overlapping DNA fragments of the hypervariable region HVR-I. Primer names indicate forward (L-strand (L) and reverse primer orientation (H-strand (H). [L16055-H16142]: 126bp, [L16117-H16233]: 162bp, [L16209-H16348]: 179bp and [L16287-H16410]: 162bp (Figure 2), covering 354bp length (according to the Reconstructed Sapiens Reference Sequence, RSRS) (Behar et al. 2012). These short amplicons are suitable for highly fragmented DNA. PCRs were prepared in the aDNA laboratories at ACAD, while amplification and post-PCR procedures were performed in a standard modern DNA laboratory, isolated from the ancient DNA lab. DNA extract or extraction blank (3µL) was added to 1X PCR Gold Buffer (Applied Biosystems), 2.5 mM MgCl2 (Applied Biosystems), 0.5 mM deoxynucleoside triphosphate (dNTP) Mix (Invitrogen), 2U AmpliTaq Gold DNA Polymerase (Applied Biosystems), 0.2 µM primer pairs, and 1 mg/mL Rabbit Serum Albumin (RSA, Sigma).

PCR cycling was performed in a DNA thermocycler Tetrad 2 Peltier (BioRAD Laboratories) under the following conditions: Initial enzyme activation at 95°C for 6 min: 45 cycles of denaturation at 95°C for 30s, annealing at 56°C, for 30s, elongation at 72°C for 30s, followed by final elongation at 65°C for 10 min. PCR reactions (5 µL) were run on a 3.5% agarose gel at 100V for 30-40 min, post-stained in *GelRed*™ (Biotium), and visualised under UV light to confirm presence of bands of the expected lengths.

Post-PCR removal of non-incorporated dNTPs and primers were performed on successful amplicons by incubation of 5 µL PCR product with 0,8 U Exonuclease I (ExoI) and 1 U Shrimp Alkaline Phosphatase (SAP) for 45 min at 37°C, followed by heat inactivation for 15 min at 80°C. We used the BigDye Terminator® Technology (Applied Biosystems) to sequence purified products in forward and reverse direction on a 3130xl Genetic Analyzer.

2.3. Genotyping of mtDNA by Multiplex PCR (GenoCoRe22)

DNA extracts have also been analysed with a mtDNA coding region Single Nucleotide Polymorphism (SNPs) assay. The GenoCoRe22 is a multiplex reaction designed to target 22 informative SNPs that characterize basal branches of worldwide mtDNA haplogroups in one single reaction (Haak et al. 2010).

This multiplex PCR has been designed to target particularly short DNA fragments (average 60 – 80bp) and is suitable for degraded and ancient DNA. Genotyping of samples was performed using SBE reactions – Single-Base Extension (SNaPshot kit, Applied Biosystems) following the manufacturer's instructions and analysed as described in (Haak et al. 2010).

2.4. Genotyping of mtDNA by Multiplex PCR (AmericaPlex26)

DNA samples were analysed using a second multiplex coding region Single Nucleotide Polymorphism (SNPs) assay, coined AmericaPlex26, designed specifically to target Native American mtDNA haplogroups and their most common subclades (Coutinho et al. 2014)(Chapter 5). The AmericaPlex26 is a powerful alternative technique to the sequencing of the HVR-I because the small amplicon sizes cope with the degree of DNA fragmentation often observed in South American samples, and the degree of typing resolution is high. SNaPshot typing was performed using SBE reactions – Single-Base Extension (SNaPshot kit, Applied Biosystems) following the manufacturer's instructions and analysed as described in (Coutinho et al. 2014).

2.5. Sequence analysis

DNA sequences were edited, assembled, and aligned with Geneious Pro® Software v.6 (Biomatters Ltd) (Drummond et al. 2011). Ancient DNA sequences corresponding to 126bp, 162bp, 179bp, 162bp length were obtained with the overlapping primers respectively covering 354bp of the HVR-I. Sequences were aligned against the Reconstructed Sapiens Reference Sequence (RSRS) (Behar et al. 2012), from which contigs and a consensus sequence were generated. Haplotypes were determined using the online database Phylotree (http://www.phylotree.org) (mtDNA tree Build 16 [19 Feb 2014]) (van Oven and Kayser 2009) (Table 4).

2.6. Populations for comparative analyses

Ancient DNA sequences obtained in this study were compared to other ancient populations from South America and modern Native American populations obtained from the literature (Table 5). We trimmed or adjusted all DNA sequences to 354bp in length (np 16,055 to 16,410) to make the dataset compatible. We excluded substitutions at nucleotide positions 16,182 and 16,183, because they are dependent on the presence

of C at the position 16,189 (Horai et al. 1993), and also excluded insertions in poly-C stretch (due to uncertainty in the exact position of such mutations). The final dataset including our ancient samples comprises 92 populations (66 modern and 26 ancient).

2.7. Population genetics and statistical analyses

Population differentiation and molecular diversity indices, haplotype (*h*), nucleotide diversity (π) (Table 6-7), genetic distances (pairwise F_{ST}), and linearized Slatkin's F_{ST} values were calculated using Arlequin v3.5 (Excoffier and Lischer 2010) to measure genetic diversity and relationships between populations (Table 8).

Analysis of Molecular Variance (AMOVA) was performed in order to evaluate population structure (Table 9). Multidimensional Scaling (MDS) was performed in order to visualize genetic similarities and dissimilarities as measured by fixation indices (Slatkin's F_{ST}) (Slatkin 1995) in a bi-dimensional space using a customized script in R 2.14.1 version (www.r.project.org). Two datasets (modern & ancient combined; Figure 3 and 'ancient only'; Figure 4) were compiled to visualize genetic relationships between ancient samples and present-day populations as well as among only ancient populations.

From 92 populations, we excluded ten [modern] outliers from Amazonian groups (Waorani, Jamandi, Cinta Larga, Chibchan), populations from the Gran Chaco (Paraguay, Guarani), three populations from the highlands (Chipaya-Uru, Llapallapani-Uru, Puno-Quechua) and Central America (Ngoebe), and four additional [ancient] outliers, Patagonia, Pica-8, Lullaillaco and Lauricocha, due to small sample sizes and better MDS configuration. The MDS final plot comprises 78 populations.

Principal Component Analysis (PCA) was performed using a customized R script for samples grouped into archaeological periods in order to visualize genetic relationships among ancient populations in a reduced number of dimensions called 'components' to detect internal structure (Figure 5). Ward hierarchical clustering based on haplogroup frequencies was performed and plotted using a customized R script (Figure 6). The newly obtained HVR-I sequences have been deposited in GeneBank under accession numbers XXXXXX to XXXXXX.

2.8. Authentication of aDNA

Replication of results in aDNA research as part of the authentication criteria (Cooper and Poinar 2000) is fundamental to validate the analyses. All samples were collected following strict guidelines for aDNA research regarding contamination with exogenous DNA (Yang and Watt 2005). To minimize the risk of contamination by exogenous DNA, samples were processed in ACAD's purpose built laboratory aDNA facilities. Strict precautions to ensure the reliability of the results were applied including separate pre- and post-PCR laboratories and the use of protective gear (body suits, face mask, gumboots and disposable latex gloves). All working surfaces, workbenches and instruments were decontaminated by using DNA oxidants such as bleach (3%) Decon and clean-room 70% Isopropanol. For each individual, we extracted DNA from at least two independent samples to monitor the reproducibility of results (Table 3). Each batch of aDNA extraction consisted of twelve samples and two Extraction Blank Controls (EBCs) at a ratio of 1:6.

3. RESULTS and DISCUSSION

3.1. Performance of the three typing methods

From 302 samples analysed (149 individuals examined), we could generate HVR-I consensus sequences for 72 individuals (Table 4), which represents an overall success rate of 48.3% for extraction and sequencing of mtDNA. From the new 72 aDNA sequences reported in this study, 40 sequences were obtained and validated by using direct sequencing with four overlapping primers pairs for HVR-I. The remaining 32 sequences were trimmed and extracted from complete mitochondrial genome data obtained from a parallel study (Chapter 3).

Successful reported aDNA sequences take into account reproducibility from the two independent samples analysed from each individual as inclusion criteria. For HVR-I sequences, we observed partial and no reproducible results on the remaining 109 individuals. For example, partial profiles (i.e. only one sample with partial DNA sequence and no sequence data for the second sample), was observed in 46 individuals. Moreover, we observed that 63 individuals failed for aDNA extraction of both independent samples (Table 2-3). Sequencing of HVS-I is primarily intended to generate sequence haplotype data. However, without typing of sub-haplogroup

diagnostic SNPs, that are often found in the coding region, an unambiguous assignment at sub-haplogroup level is not always possible.

From 302 samples analysed with the GenoCoRe22 assay, 39 individuals were successfully genotyped with both independent samples yielding same haplotype (Table 2). Partial profiles were observed in 43 individuals and 67 individuals failed for genotyping of both independent samples (Table 2-3).

From 261 samples (130 individuals) analysed with the AmericaPlex26 assay, 32 individuals yielded consistent genotypes (Table 2). Partial genotype profiles were observed in 80 individuals and 18 individuals failed for genotyping of both independent samples. The AmericaPlex26 is a powerful alternative technique to the sequencing of the HVR-I because the small amplicon sizes cope with the degree of DNA fragmentation often observed in South American samples and cover the most common sub-haplogroups in modern-day Native Americans. Here, the Americaplex26 assay not only complements the HVSI sequencing, but can also serve as standalone tool for confident haplogroup calling.

Finally, we report a total of 60 consensus calls across all typing results from the three used methods (Table 2-3). Moreover, the three genotyping methods show consistent results and similar success rates and no significant differences when compare to each other $p < 0.01$, Chi-square test (Figure 9).

3.2. Contamination detection

The majority of samples collected for this study come from museum collections. Museum samples often lead to inefficient DNA recovery (Pruvost et al. 2007) and contamination due to handling of specimens by museum personnel (Malmstrom et al. 2005). Samples stored at Museum collections for long-time periods might not always be adequate or optimal for aDNA retrieval with the add bonus of displaying a very variable levels of contamination due to inadequate handling or post-excavation history (Yang and Watt 2005).

For instance, a sample from Huaca Pucllana 10776A (Wari culture) yielded a Native American B4 haplotype, however the second/independent sample 10777A from the same individual gave a European T2 haplotype. Another example, sample 10784A did not yield amplifiable aDNA, however the second sample 10783A reported a K haplotype. Both samples were excluded from subsequent analyses.

We found no aDNA haplotype sequences that match ACAD personnel, or people involved in the study. Samples were collected and processed by WH (mtDNA hg H1), BL (hg H3), and GV (hg B2), however this B2 haplotype harbours a set of unique mutations not seen in the overall B2 diversity of our Native American samples. The identification of non-Native American mtDNA in this case is likely due to manipulation from archaeologists of Eurasian origin. Nevertheless, we observed consistent Native American haplotypes across experiments. We also sequenced Extraction Blank Controls (EBCs) when they were positive, but could not find evidence for systematic PCR contamination, cross-contamination or "carryover" effects, which indicates a low level of contamination in the laboratory (Table 3).

3.3. Overall success rates of analysed samples

Results from well-preserved high altitude samples from Tiwanaku resulted in a high success rate (90%), likely due to cooler environmental conditions favouring the long-term DNA preservation (i.e. 7.7°C annual average temperature) which highlights also the importance of freshly excavated samples that are in general the best preserved material for aDNA studies (Pruvost et al. 2007).

Samples from the much older Lauricocha site confirm this observation, although the sample size is small and the preservation differs between samples from the same archaeological period. In contrast, samples from Peruvian coastal cultures (Huaca Pucllana), gave variable amplification success rates (Lima 60%, Wari 50%, Ychsma 88%, and Chancay 77%), while samples from some Peruvian lowland sites (Caral, Aspero and Puemape) failed to produce reliable haplotypes. This could be due to the older age of the samples (Late Archaic Period), as younger lowland samples, i.e. Pueblo Viejo-Pucara, Inca samples belonging to the Late Horizon Period showed a higher success rate (61%) (Table 2).

3.4. Haplogroup composition and frequency based genetic distances across cultural periods

This study incorporates archaeological samples from all cultural periods in the Central Andes of South America, aiming to improve the spatial and geographic coverage. All aDNA sequences could be assigned to haplogroups A, B, C, and D, which are well-described for Native American populations, with the exception of haplogroup

X, which was not observed in our samples. At the sub-haplogroup level, the ancient mtDNA sequence profiles could be assigned to A2 (10.4 %,), B2 (24.7%), B2b (16.8%), C1b (28.5%), C1c (6.5%), C1d (1.3%), and D1 (11.8%) (Table 4).

By tracking the haplogroup composition and genetic distances through time from the Late Archaic to Late Horizon period (Figure 7), we can explore dynamic processes and genetics shifts between archaeological cultures. Genetic composition of populations during the Late Archaic Period–LA (8000–1900 BC) Lauricocha and Pernil Alto show a high frequency of haplogroup A (43%) and also the lowest frequencies of haplogroup C (13%), which are distinct in comparison to the subsequent cultural periods. This might be due to low sample number per site (16 individuals in total). However, a similarly skewed composition in two archaic groups, Lauricocha from the highlands and Pernil Alto from the South coastal Nasca region, might represent the initial genetic structure in prehistoric populations in this region of South America, which had a higher percentage of haplogroup A.

The following two time periods Early Horizon–EH and Early Intermediate Period–EIP show a genetic proximity (F_{ST} = 0.03055) (Table 8). Interestingly, the most significant change in haplogroup composition is observed in individuals from the EH (900 BC–200 AD), where we observed a substantial increase in haplogroup D frequency (73%), especially in individuals from the Palpa region, suggesting a local genetic fluctuation, and a drop out of haplogroups A and B in individuals from Caverna6 and Palpa. Haplogroup C (15%) remains constant in comparison with the Late Archaic Period. The Early Intermediate Period–EIP (200–600 AD) represented by Pampa Grande, Monte Grande and Lima group, shows an increase in haplogroup C (26%) and B (23%) in contrast to the previous periods while the frequency of haplogroup D remains high (44%).

The Middle Horizon period–MH (600–1000 AD), represented by Conchopata, Wari, Laramate MH, Palpa MH and Tiwanaku populations, and the Late Intermediate Period–LIP (1000–1476 AD), represented by Laramate LIP, Huari, Palpa, Ychsma, Chancay, and Montegrande, show similar patterns of haplogroup composition. However, they mark another shift in haplogroup composition, with haplogroup B (46% and 49%) and haplogroup C (32% and 37%, respectively) being the dominant haplogroups at the time.

Samples from the Huaca Pucllana archaeological site in coastal Lima represent another key area in our study with individuals from three successive cultural periods until the arrival of Europeans; the Lima culture from the Early Intermediate Period (200–600 AD), followed by the Wari culture in the Middle Horizon (600–1000 AD), and the Ychsma culture in the Late Intermediate Period (1000–1476 AD). Huaca Pucllana represents an important cultural time transect and provides an opportunity for a regional and chronological study to explore further micro-evolutionary processes in the Central Andes (see Chapter 4).

Interestingly, samples from the Wari population of Huaca Pucllana resemble the genetic signature from Huari (LIP) and suggest genetic influx from high-altitude Huari population and the expansion to coastal sites in Central Peru during the Middle Horizon period (Menzel 1964; Lumbreras 1969).

The Middle Horizon characterises an important phase in political and cultural changes in the Central Andes and describes the rise of two competing empires, the Wari and Tiwanaku, which both expanded and led to political interactions and potential gene flow between groups.

The addition of new samples from the Tiwanaku culture, considered as one of the most important pre-Columbian civilizations in the Altiplano region, allowed us to contrast the genetic relationship with Wari populations during the Middle Horizon. It has been suggested that the dynamics of Tiwanaku and Wari expansions had played a crucial role in shaping and shuffling the genetic diversity which in this area, resulted in an increased homogenization of ancient populations from highlands and coastal regions in the Central Andes starting with the Middle Horizon (Isbell 2008).

To date, the genetic diversity of the Tiwanaku has been poorly investigated (Rothhammer et al. 2003), however a number of Wari samples had been reported by (Kemp et al. 2009), suggesting population continuity after the collapse of the Wari Empire. The Tiwanaku samples show low genetic distances (F_{ST} = 0.0398) with Wari populations, which suggest a direct interaction of both empires in the Andean region of Peru and Bolivia during the Middle Horizon. The suggested process of genetic homogenization during the Middle Horizon (Lewis 2009; Fehren-Schmitz et al. 2011b) is also supported by mtDNA haplogroup frequency analysis followed by Ward Clustering and PCA, which show a clear association between the two successive periods (Figures 5-6), but also embrace the Late Horizon period.

The Late Horizon–LH (1476–1534 AD) represented by Llullaillaco, Acchaymarca, Puca, Tompullo, and Pueblo Viejo is further characterized by higher proportion of haplogroup B (57%) in comparison with all preceding archaeological periods. This cultural period also marks the arrival of the Europeans in South America, thus it defines the ultimate Native American mtDNA diversity in pre-Columbian populations prior to the European contact. Note that the high frequency of haplogroup B in the Late Horizon is consistent with today's population diversity from this region in South America (Barbieri et al. 2011; Gaya-Vidal et al. 2011; Sandoval JR et al. 2013).

3.5. Genetic comparison of ancient and modern populations

To achieve a better representation of population affinities we grouped all 78 populations into a main geographic eco-regional pattern (Table 5). The resulting MDS plot shows a clustering in the bottom quadrants indicating and association between modern and ancient samples, although Coordinate 1 separates the large majority of modern samples from Central South America from the other regions.

Modern Central Andean populations are distributed mostly on the left, while ancient coastal groups and modern South populations fall on the right side of the MDS plot. The main cluster is observed at the bottom quadrants, which shows a closer relationship between ancient highland groups with coastal ancient populations. All ancient highland populations form a cluster on the bottom quadrants as for many of the coastal ones, suggesting also a closer relationship/connection between highlands and lowlands groups in ancient times. Also, Coordinate 2 separates all ancient highlands from coastal ancient populations.

Geographic pattern is visible on three populations from Tierra de Fuego that cluster together on the right hand side of the plot, in part visible on populations from modern southern South America that form a separate/regional cluster on the bottom right quadrant, and not visible on Amazonian populations and populations from the Gran Chaco, which do not form a cluster despite same geographic affinities (Figure 3).

3.6. Genetic comparison of ancient populations

Genetic relationships among the 22 ancient groups were explored in further detail by MDS (Figure 4). The position of groups is described in temporal order. Samples from the Late Archaic Period (Pernil Alto) and Early Horizon (Caverna6 and

Palpa) form outliers in the lower quadrants of the plot, based on substantial differences in haplogroup composition, which might be driven by sample sizes. Coastal samples from the subsequent Early Intermediate Period cluster in the right upper quadrant (Monte Grande and Palpa) and left upper quadrant, while the highland site (Pampa Grande) falls inside the lower left quadrant.

The most interesting observation is the similarity of Middle Horizon and Late Intermediate samples visualized in the MDS plot (Figure 4). We observe a strong genetic affinity of MH and LIP populations falling in the upper quadrants with no distinctions between coastal and highland groups. This is further supported by significantly negative values ($p<0.05$) of Tajimas's D, for these two periods, suggesting that MH and LIP populations have gone through demographic expansions in the past. The transition to the later successive Middle Horizon, Late Intermediate Period therefore represents a distinct temporal phase according to pairwise (F_{ST}) and heatmap (Figure 8). The proximity of samples from Huari (LIP) and Wari samples from Huaca Pucllana (MH) is consistent with Kemp et al. (Kemp et al. 2009), who suggested population continuity between the Middle Horizon and Late Intermediate period. Moreover, the Huari samples from Conchopata cluster also with samples from the Tiwanaku culture, confirming the proposed interaction between the Wari and Tiwanaku empires during the Middle Horizon period (Isbell 2008).

Samples from the Late Horizon (Pueblo Viejo, Tompullo, Acchaymarca, are positioned in the upper left quadrant, with the exception of Puca individuals which are separated by the second coordinate.

3.7. Analysis of Molecular Variance (AMOVA)

We used AMOVA framework to evaluate potential population structure under various groupings and to investigate patterns of association in the genetic dataset (Table 9). Populations were alternatively grouped according to chronological (time period, cultural period, ancient vs. modern populations) and geographic (distance, altitude) or geo-political (country) criteria. We expect groupings that reflect associations observed in our genetic data to show high "among groups" variation and low "within populations" variation.

First, the grouping based on geographic criteria (8 large geographic regions or eco-zones in South America: Amazon, Tierra de Fuego, South Andean, Central Andean,

Central Andean [highland], Central Andean [coast], Gran Chaco, North Western) shows the highest variance component with 8.63% ($p < 0.000$), which indicates that population structure in South America follows geography, which is visible in the MDS plot (Figure 3).

A second grouping based on modern samples was pooled according to geo-political criteria (present-day countries) and shows a lower but significant value of 6.09% variation ($p<0.0136$). This also supports a genetic structure of populations, which largely follows geography. The finding of a significant value for the current political grouping in South America can be explained by the fact that some eco-zones are unique to particular countries or in turn, that political territories circumscribe geographic regions.

For a third group we used only ancient sequences, which were grouped into six cultural archaeological periods: Archaic, Early Horizon, Early Intermediate Period, Middle Horizon, Late Intermediate Period and Late Horizon. The resulting variance component of 5.22% ($p<0.0225$) is lower than the above, but still points to a distinct signal per cultural entity, which we explored further with different analyses (see Haplogroup frequency, Ward Clustering and PCA).

A fourth AMOVA variant grouping tested Highland versus Lowland sites/populations, and reveals a low variance component of 3.70% ($p<0.0358$), suggesting a higher homogeneity in Andean population compared with coastal populations. We also explored a fifth grouping that contrasts Andean and Amazonian populations, following a previous study which had postulated a genetic relationship of Andean and Amazonian extant populations (Rothhammer et al. 2003). However, the resulting variance component is lowest (1.81%), which makes this scenario less plausible in comparison to other groupings tested by our AMOVA results.

Lastly, an all encompassing sixth grouping based on two time periods (ancient vs. modern) has a low variance component of 0.82% showing no differentiation through time and thus suggests a higher variance within each temporal group than between them. For this matter, the sufficient overlap between ancient and modern populations is enough to reject population discontinuity. Note that the geographic and temporal structure, even if significant, is never stronger than the within groups/between population structure.

4. CONCLUSION

The Central Andes of South America highlight an important geographic region to contrast archaeological and genetic information, in order to explore cultural and demographic events in pre-Columbian times. The sequencing of HVR-I still represents a valuable approach to study mtDNA genetic diversity, mainly because of the comprehensive amount of available modern and ancient data to perform comparative analysis. Although the level of resolution of HVR-I is low in comparison with genomic data (i.e. complete mtDNA genomes), we can gain important information about the female population history and test hypotheses concerning cultural transitions in this region of South America. Based on the analyses of mtDNA haplogroup frequencies (PCA and Ward clustering) and genetic distances (F_{ST}) we observe two temporal transitions in subsequent archaeological periods. The first transition describes the transition from the Late Archaic Period–LA (8000–1900 BC) to the Early Horizon–EH and Early Intermediate period–EIP. The second is marked by characteristic changes that define the Middle Horizon–MH, and the subsequent Late Intermediate–LIP and Late Horizon–LH periods, altogether suggesting a continuous process of genetic assimilation through time between highland and coastal populations.

With the advent of new technologies, such as Next-Generation Sequencing (NGS) applied to the aDNA field, the possibility to obtain larger amounts of high-resolution genomic information (e.g. complete mitochondrial and nuclear genomes) opens a new era, which will allow the detailed exploration of Native American genetic history.

Acknowledgments

This project was founded by Australian Research Council (ARC) Discovery Project "DP1095782". We thank team members of the "Proyecto Arqueológico Huaca Pucllana", Ministerio de Cultura, Bolivian Archaeology Office, Department of Humanities and Archaeology at the Pontificia Universidad Católica del Perú, Centro de Investigaciones Arqueológicas del Museo de Sitio de Ancón, Perú, Instituto de Alta Investigación, Laboratorio de Arqueología y Paleoambiente, at the Universidad de Tarapacá, Arica, Chile, for allowing access to the collections of pre-Columbian human remains. We thank Luis Pezo and Fabiana Martin for providing archaeological samples for analysis.

REFERENCES

Achilli A, Perego U, and Bravi C. 2008. The Phylogeny of the Four Pan-American MtDNA Haplogroups: Implications for Evolutionary and Disease Studies. PLoS ONE 3(3):e1764.

Afonso Costa H, Carvalho M, Lopes V, Balsa F, Bento AM, Serra A, Andrade L, Anjos MJ, Vide MC, Pantoja S et al. . 2010. Mitochondrial DNA sequence analysis of a native Bolivian population. J Forensic Leg Med 17(5):247-253.

Baca M, Doan K, Sobczyk M, Stankovic A, and Weglenski P. 2012. Ancient DNA reveals kinship burial patterns of a pre-Columbian Andean community. BMC Genet 13(1):30.

Baca M, Molak M, Sobczyk M, Weglenski P, and Stankovic A. 2014. Locals, resettlers, and pilgrims: A genetic portrait of three pre-Columbian Andean populations. Am J Phys Anthropol.

Bandelt HJ, Herrnstadt C, Yao YG, Kong QP, Kivisild T, Rengo C, Scozzari R, Richards M, Villems R, Macaulay V et al. . 2003. Identification of Native American founder mtDNAs through the analysis of complete mtDNA sequences: some caveats. AnnHumGenet 67(Pt 6):512-524.

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

Behar DM, van Oven M, Rosset S, Metspalu M, Loogvali EL, Silva NM, Kivisild T, Torroni A, and Villems R. 2012. A "Copernican" reassessment of the human mitochondrial DNA tree from its root. Am J Hum Genet 90(4):675-684.

Bodner M, Perego UA, Huber G, Fendt L, Rock AW, Zimmermann B, Olivieri A, Gomez-Carballa A, Lancioni H, Angerhofer N et al. . 2012. Rapid coastal spread of First Americans: novel insights from South America's Southern Cone mitochondrial genomes. Genome Research 22(5):811-820.

Bonatto S, and Salzano F. 1997. A single and early migration for the peopling of the Americas supported by mitochondrial DNA sequence data. Proc Natl Acad Sci USA 94::1866–1871.

Brotherton P, Haak W, Templeton J, Brandt G, Soubrier J, Jane Adler C, Richards SM, Sarkissian CD, Ganslmeier R, Friederich S et al. . 2013. Neolithic mitochondrial haplogroup H genomes and the genetic origins of Europeans. Nature Communications 4:1764.

Brown MD, Hosseini SH, Torroni A, Bandelt HJ, Allen JC, Schurr TG, Scozzari R, Cruciani F, and Wallace DC. 1998. mtDNA haplogroup X: An ancient link between Europe/Western Asia and North America? AmJ HumGenet 63(6):1852-1861.

Carnese F, Mendisco F, and Keyser C. 2010. Paleogenetical study of pre-Columbian samples from Pampa Grande (Salta, Argentina). Am J Phys Anthropol 141:452 - 462.

Cooper A, and Poinar HN. 2000. Ancient DNA: Do It Right or Not at All. Science(289):1139.

Coutinho A, Valverde G, Fehren-Schmitz L, Cooper A, Barreto Romero MI, Espinoza IF, Llamas B, and Haak W. 2014. AmericaPlex26: a SNaPshot multiplex system for genotyping the main human mitochondrial founder lineages of the Americas. PLoS ONE 9(3):e93292.

de Saint Pierre M, Bravi CM, Motti JM, Fuku N, Tanaka M, Llop E, Bonatto SL, and Moraga M. 2012. An alternative model for the early peopling of southern South America revealed by analyses of three mitochondrial DNA haplogroups. PLoS ONE 7(9):e43486.

Der Sarkissian C, Balanovsky O, Brandt G, Khartanovich V, Buzhilova A, Koshel S, Zaporozhchenko V, Gronenborn D, Moiseyev V, Kolpakov E et al. . 2013. Ancient DNA reveals prehistoric gene-flow from siberia in the complex human population history of North East Europe. PLoS Genetics 9(2):e1003296.

Dornelles CL, Bonatto SL, De Freitas LB, and Salzano FM. 2005. Is haplogroup X present in extant South American Indians? Am J Phys Anthropol 127(4):439-448.

Drummond A, Ashton B, Buxton S, Cheung M, Cooper A, Duran C, Field M, Heled J, Kearse M, Markowitz S et al. . 2011. Geneious v5.4. Available from http://www.geneious.com/.

Excoffier L, and Lischer HE. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour 10(3):564-567.

Fagundes NJ, Kanitz R, Eckert R, Valls AC, Bogo MR, Salzano FM, Smith DG, Silva WA, Jr., Zago MA, Ribeiro-dos-Santos AK et al. . 2008. Mitochondrial population genomics supports a single pre-Clovis origin with a coastal route for the peopling of the Americas. Am J Hum Genet 82(3):583-592.

Fehren-Schmitz L, Haak W, Machtle B, Masch F, Llamas B, Tomasto Cagigao E, Sossna V, Schittek K, Isla Cuadrado J, Eitel B et al. . 2014. Climate change underlies global demographic, genetic, and cultural transitions in pre-Columbian southern Peru. Proceedings of the National Academy of Sciences.

Fehren-Schmitz L, Hummel S, and Herrmann B. 2009. Who Were the Nasca? Population Dynamics in Pre-Columbian Southern Peru Revealed by Ancient DNA Analyses.159-172.

Fehren-Schmitz L, Llamas B, Tomasto E, and Haak W. 2011a. El ADN antiguo y la historia del poblamiento temprano de la parte oeste de Sudamérica: lo que hemos aprendido y hacia dónde vamos. Boletín de Arqueología PUCP, ISSN 1029-2004 Nº. 15:17-42.

Fehren-Schmitz L, Reindel M, Cagigao E, Hummel S, and Herrmann B. 2010. Pre-Columbian population dynamics in coastal southern Peru: A diachronic investigation of mtDNA patterns in the Palpa region by ancient DNA analysis. Am J Phys Anthropol 141:208 - 221.

Fehren-Schmitz L, Warnberg O, and Reindel M. 2011b. Diachronic investigations of mitochondrial and Y-chromosomal genetic markers in pre-Columbian Andean highlanders from South Peru. Ann Hum Genet 75:266 - 283.

Fulton TL. 2012. Setting up an ancient DNA laboratory. Methods Mol Biol 840:1-11.

Gaya-Vidal M, Moral P, Saenz-Ruales N, Gerbault P, Tonasso L, Villena M, Vasquez R, Bravi CM, and 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(2):215-230.

Goebel T, Waters MR, and O'Rourke DH. 2008. The late Pleistocene dispersal of modern humans in the Americas. Science 319(5869):1497-1502.

Greenberg J, Turner II C, Zegura S, and al. e. 1986. The Settlement of the Americas: A Comparison of the Linguistic, Dental, and Genetic Evidence. Current Anthropology , Vol. 27,(No. 5):pp. 477-497

Haak W, Balanovsky O, Sanchez JJ, Koshel S, Zaporozhchenko V, Adler CJ, Der Sarkissian CSI, Brandt G, Schwarz C, Nicklisch N et al. . 2010. Ancient DNA from European early neolithic farmers reveals their near eastern affinities. PLoS biology 8(11):e1000536.

Haak W, Forster P, Bramanti B, Matsumura S, Brandt G, Tanzer M, Villems R, Renfrew C, Gronenborn D, Alt KW et al. . 2005. Ancient DNA from the first European farmers in 7500-year-old Neolithic sites. Science 310(5750):1016-1018.

Handt O, Krings M, Ward RH, and Paabo S. 1996. The retrieval of ancient human DNA sequences. Am J Hum Genet 59(2):368-376.

Horai S, Kondo R, Nakagawahattori Y, Hayashi S, Sonoda S, and Tajima K. 1993. Peopling of the America, Founded by 4 Major Lineages of Mitochondrial-DNA. Molecular Biology and Evolution 10(1):23-47.

Isbell WH. 2008. Wari and Tiwanaku: International Identities in the Central Andean Middle Horizon, en: H. I. Silverman y W. H. Isbell (eds). Handbook of South American Archaeology Springer, New York.:731-759.

Kemp BM, Tung TA, and Summar ML. 2009. Genetic continuity after the collapse of the Wari empire: mitochondrial DNA profiles from Wari and post-Wari populations in the ancient Andes. Am J Phys Anthropol 140(1):80-91.

Kitchen A, Miyamoto M, and Mulligan C. 2008. A Three-Stage Colonization Model for the Peopling of the Americas. . PLoS ONE 3(2):(e1596. doi:10.1371/journal.pone.0001596).

Knapp M, Clarke AC, Horsburgh KA, and Matisoo-Smith EA. 2012. Setting the stage - building and working in an ancient DNA laboratory. Ann Anat 194(1):3-6.

Lalueza C, Perez-Perez A, Prats E, Cornudella L, and Turbon D. 1997. Lack of Founding Amerindian Mitochondrial DNA Lineages in Extinct Aborigines from Tierra del Fuego-Patagonia. Human Molecular Genetics 6(1):41-46.

Lewis C. 2009. Difficulties in rejecting a local ancestry with mtDNA Haplogroup data in the South Central Andes. Latin American antiquity.

Lewis C, Lizarraga B, and Tito R. 2007. Mitochondrial DNA and the peopling of South America. Hum Biol 79:159 - 178.

Livi-Bacci M. 2006. The Depopulation of Hispanic America after the Conquest. Population and Development Review 32(2):199-232.

Lumbreras LG. 1969. De los Pueblos, las Culturas y las Artes en el Antiguo Perú. Editorial Mancloa-Campodónico, Lima.

Malmstrom H, Stora J, Dalen L, Holmlund G, and Gotherstrom A. 2005. Extensive Human DNA Contamination in Extracts from Ancient Dog Bones and Teeth. Mol Biol Evol 22(10):2040-2047.

Mendisco F, Keyser C, Seldes V, Rivolta C, Mercolli P, Cruz P, Nielsen AE, Crubezy E, and Ludes B. 2014. Genetic diversity of a late prehispanic group of the quebrada de humahuaca, northwestern Argentina. Ann Hum Genet 78(5):367-380.

Menzel D. 1964. Style and time in the Middle Horizon. Ñawpa Pacha, Berkeley, California, Institute of Andean Studies 2:1-105.

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

Moraga M, Santoro C, Standen V, Carvallo P, and Rothhammer F. 2005. Microevolution in prehistoric Andean populations: chronologic mtDNA

variation in the desert valleys of northern Chile. Am J Phys Anthropol 127:170 - 181.

Mulligan C, Kitchen A, and Miyamoto M. 2008. Updated Three-Stage Model for the Peopling of the Americas. PLoS ONE 3(9):e3199. doi:3110.1371/ journal.pone.0003199.

O'Fallon BD, and Fehren-Schmitz L. 2011. Native Americans experienced a strong population bottleneck coincident with European contact. Proc Natl Acad Sci U S A.

O'Rourke DH, and Raff JA. 2010. The human genetic history of the Americas: the final frontier. Curr Biol 20(4):R202-207.

Perego UA, Angerhofer N, Pala M, Olivieri A, Lancioni H, Kashani BH, Carossa V, Ekins JE, Gomez-Carballa A, Huber G et al. . 2010. The initial peopling of the Americas: a growing number of founding mitochondrial genomes from Beringia. Genome Research 20(9):1174-1179.

Pruvost M, Schwarz R, Correia V, Champlot S, Braguier S, Morel N, Fernandez-Jalvo Y, Grange T, and Geigl E. 2007. Freshly excavated fossil bones are best for amplifcation of ancient DNA. Proccedings of the National Academy of Science USA 104(3):739 - 744.

Raff JA, Bolnick DA, Tackney J, and O'Rourke DH. 2011. Ancient DNA perspectives on American colonization and population history. Am J Phys Anthropol.

Raghavan M, DeGiorgio M, Albrechtsen A, Moltke I, Skoglund P, Korneliussen TS, Gronnow B, Appelt M, Gullov HC, Friesen TM et al. . 2014. The genetic prehistory of the New World Arctic. Science 345(6200):1255832.

Rasmussen M, Anzick SL, Waters MR, Skoglund P, DeGiorgio M, Stafford TW, Jr., Rasmussen S, Moltke I, Albrechtsen A, Doyle SM et al. . 2014. The genome of a Late Pleistocene human from a Clovis burial site in western Montana. Nature 506(7487):225-229.

Rasmussen M, Li Y, Lindgreen S, Pedersen JS, Albrechtsen A, Moltke I, Metspalu M, Metspalu E, Kivisild T, Gupta R et al. . 2010. Ancient human genome sequence of an extinct Palaeo-Eskimo. Nature 463(7282):757-762.

Rasmussen M, Sikora M, Albrechtsen A, Korneliussen TS, Moreno-Mayar JV, Poznik GD, Zollikofer CPE, Ponce de León MS, Allentoft ME, Moltke I et al. . 2015. The ancestry and affiliations of Kennewick Man. Nature.

Reich D, Patterson N, Campbell D, Tandon A, Mazieres S, Ray N, Parra MV, Rojas W, Duque C, Mesa N et al. . 2012. Reconstructing Native American population history. Nature 488(7411):370-374.

Rodriguez-Delfin L, Rubin-de-Celis V, and Zago M. 2001. Genetic diversity in an Andean population from Peru and regional migration patterns of Amerindians in South America: data from Y chromosome and mitochondrial DNA. Hum Hered 51:97 - 106.

Rothhammer F, Moraga M, Rivera M, Santoro CM, Standen VG, Garcıa F, Carvallo P, 35:269–, and 274. 2003. Analisis de ADNmt de restos esqueletales del sitio arqueologico de Tiwanaku y su relacion con el origen de sus constructores. Chungara (Arica) 35:269–274.

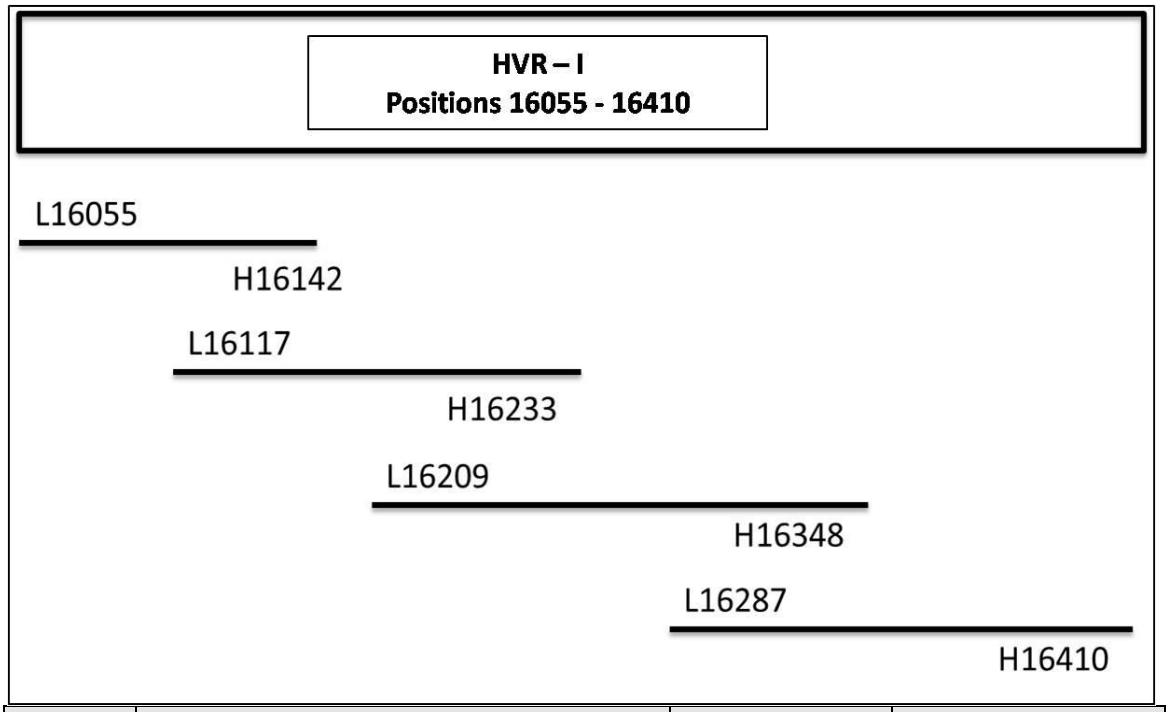
Sandoval JR, Lacerda DR, Jota MSA, Salazar-Granara A, Vieira PPR, and al. e. 2013. The Genetic History of Indigenous Populations of the Peruvian and Bolivian Altiplano: The legacy of the Uros. PLoS ONE 8(9):e73006.

Schurr TG, Ballinger SW, Gan YY, Hodge JA, Merriwether DA, Lawrence DN, Knowler WC, Weiss KM, and Wallace DC. 1990. Amerindian mitochondrial

DNAs have rare Asian mutations at high frequencies, suggesting they derived from four primary maternal lineages. Am J Hum Genet 46(3):613-623.
Shimada I, Shinoda K, Farnum J, Corruccini R, and Watanabe H. 2004. An integrated analysis of pre-Hispanic mortuary practices: a Middle Sican case study. Curr Anthropol 45:369–402.
Shinoda K, Adachi N, Guillen S, and Shimada I. 2006. Mitochondrial DNA analysis of ancient Peruvian highlanders. Am J Phys Anthropol 131(1):98-107.
Slatkin M. 1995. A measure of population subdivision based on microsatellite allele frequencies. Genetics 139:457-462.
Stone AC, and Mark Stoneking 1998. mtDNA Analysis of a Prehistoric Oneota Population: Implications for the Peopling of the New World. Am J Hum Genet 62:1153–1170. Szathmary, Emöke J. E.
Szathmáry EJE. 1981. Genetic markers in Siberian and northern North American populations. Yearb Phys Anthropol 24:37–73.
Taboada-Echalar P, Alvarez-Iglesias V, Heinz T, Vidal-Bralo L, Gomez-Carballa A, Catelli L, Pardo-Seco J, Pastoriza A, Carracedo A, Torres-Balanza A et al. . 2013. The genetic legacy of the pre-colonial period in contemporary Bolivians. PLoS ONE 8(3):e58980.
Tamm E, Kivisild T, Reidla M, Metspalu M, and Smith D. 2007. Beringian standstill and spread of Native American founders. PLoS ONE 2(9):e829.
Torroni A, Neel JV, Barrantes R, Schurr TG, and Wallace DC. 1994. Mitochondrial-DNA Clock for the Amerinds and Its Implications for Timing Their Entry into North-America. Proc Natl Acad Sci U S A 91(3):1158-1162.
Torroni A, Schurr T, Cabell M, Brown M, Neel J, Larsen M, Smith D, Vullo C, and Wallace D. 1993. Asian affinities and continental radiation of the four founding Native American mtDNAs. Am J Hum Genet 53:563–590.
van Oven M, and Kayser M. 2009. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Hum Mutat 30(2):E386-394.
Wang S, Lewis C, and Jakobsson M. 2007. Genetic variation and population structure in Native Americans. PLoS Genetics 3:e185.
Yang DY, and Watt K. 2005. Contamination controls when preparing archaeological remains for ancient DNA analysis. Journal of Archaeological Science 32(3):331-336.

List of Figures and Tables

Figure 1. Map of South America showing sites from which ancient human remains were sampled for this study. See Table 1 for detailed information about the sites and samples.
Lima [1], Wari [2], Ychsma [3], Miramar Necrópolis [4], Chancay [5], Aspero [6], Caral Supe [7], Pueblo Viejo-Pucara [8], Tablada de Lurín [9], Puemape [10], Palpa [11], Pernil Alto[12], Chullpa Botiriayocc [13], Lauricocha [14], Tiwanaku [15], Pica8-Tarapacá [16], Azapa-Tarapacá [17], Llullaillaco [18], Patagonia [19].



Primer	Primer Sequences 5' to 3'	Length in bp (incl./excl. primer)	Reference
L16055	GAAGCAGATTTGGGTACCAC	126bp	(Handt et al. 1996)
H16142	ATGTACTACAGGTGGTCAAG		(Stone 1998)
L16117	TACATTACTGCCAGCCACCAT	162bp	(Haak et al. 2005)
H16233	GCTTTGGAGTTGCAGTTGATGTGT		(Haak et al. 2005)
L16209	CCCCATGCTTACAAGCAAGT	179bp	(Handt et al. 1996)
H16348	ATGGGGACGAGAAGGGATTTG		(Haak et al. 2005)
L16287	ACTAGGATACCAACAAACCC	162bp	(Handt et al. 1996)
H16410	GCGGGATATTGATTTCACGG		(Handt et al. 1996)

Figure 2. Details of primers used for standard HVR-I amplification and sequencing. Set of four overlapping DNA primers to amplify 354 bp of the HVR-I (16,055 – 16,410).

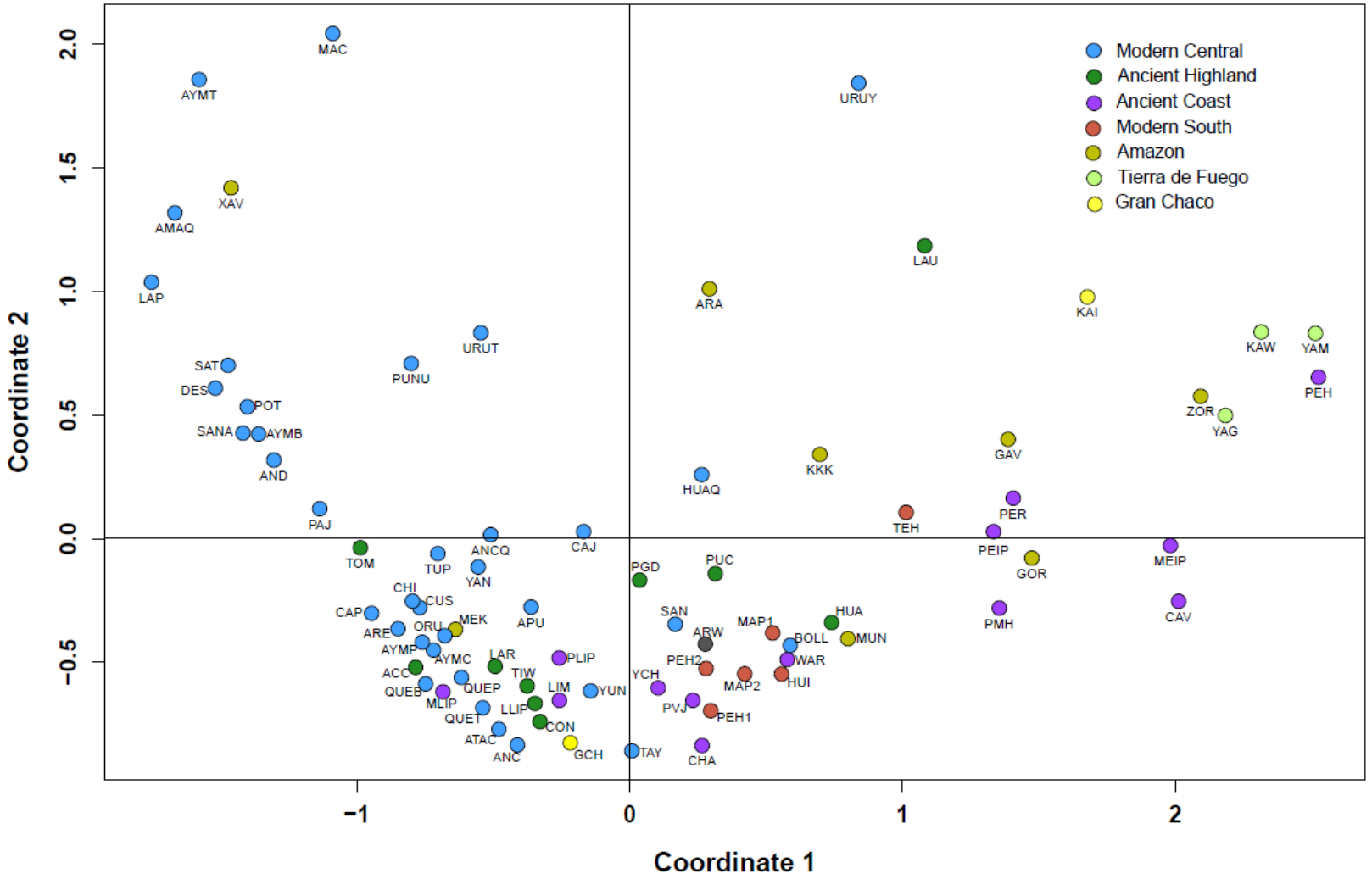

Modern Central
Ancient Highland
Ancient Coast
Modern South
Amazon
Tierra de Fuego
Gran Chaco
Coordinate 1
Coordinate 2
−1
0
1
2
−0.5
0.0
0.5
1.0
1.5
2.0
MAC
AYMT
XAV
AMAQ
LAP
SAT
DES
POT
SANA
AYMB
AND
PAJ
PUNU
URUT
TOM
TUP
ANCQ
CAJ
YAN
CHI
CUS
CAP
ARE
ORU
MEK
APU
AYMP
AYMC
LAR
PLIP
ACC
TIW
QUEB
QUEP
LIM
YUN
MLIP
LLIP
QUET
CON
ATAC
ANC
GCH
TAY
CHA
PVJ
PEH1
YCH
PEH2
MAP2
HUI
SAN
ARW
MAP1
HUA
BOLL
MUN
WAR
PGD
PUC
URUY
LAU
ARA
KAI
KAW
YAM
PEH
ZOR
YAG
GAV
KKK
HUAQ
TEH
PER
PEIP
GOR
PMH
MEIP
CAV

Figure 3. Genetic affinities of ancient and modern populations from South America.
Multidimensional scaling (MDS) plot based on Slatkin's *F*st.

Gran Chaco (GCH); Arequipa (ARE); San Martin (SAN); Tayacaja (TAY); Ancash (ANC); Kaingang (KAI); Mapuche (MAP1); Pehuenche (PEH1); Yaghan (YAG); Gaviao (GAV); Xavante (XAV); Zoro (ZOR); Yanomami (YAN); Puno_Quechua (QUEP); Yungay (YUN); Tupe (TUP); Puno_Aymara (AYMP); Arawaken (ARW); Titicaca_Quechua (QUET); Titicaca_Aymara (AYMT); Titicaca_Uros (URUT); Quechua_Bol (QUEB); Aymara_Bol (AYMB); Mapuche (MAP2); Yamana (YAM); Tehuelche (TEH); Atacameno (ATAC); Aymara_Chi (AYMC); Pehuenche (PEH2); Huilliche (HUI); Kawwesqar (KAW); Urus_Yanesha_ARA (URUY); Cajamarca_QUE (CAJ); Huancavelica_QUE (HUAQ); Apurimac_QUE (APU); Machiguenga_ARA (MAC); Puno_URU (PUNU); Amantani_QUE (AMAQ); SantaRosa_AYM (SAT); LaPaz_AYM (LAP); Andamarca_AYM (AND); Capachica_QUE (CAP); Chimu_AYM (CHI); Cusco_AYM (CUS); Desaguadero_AYM (DES); Pajchiri_AYM (PAJ); Oruro_AYM (ORU); Potosi_QUE (POT); Ancash_QUE (ANCQ); SantaAna_AYM (SANA); Gorotire (GOR); Kuben-Kran-Kegn (KKK); Munduruku (MUN); Arara (ARA); Mekranoti (MEK); Caverna6_EH (CAV); Palpa_MH (PMH); Palpa_EIP (PEIP); Laramate_LIP (LLIP); Palpa_EH (PEH); Conchopata_MH (CON); Huari_LIP (HUA); Laramate_MH (LAR); Pampa Grande_EIP-MH (PGD); MonteGrande_EIP (MEIP); Palpa_LIP (PLIP); MonteGrande_LIP (MLIP); Lauricocha-LA (LAU); Tompullo-Peru (TOM); Pucllana-Lima (LIM); Pucllana-Wari(WAR); Pucllana-Ychsma (YCH); Chancay (CHA); Tiwanaku-Bolivia (TIW); Pernil Alto-Arch (PER); Acchaymarca (ACC); Puca (PUC); Pueblo Viejo (PVJ); Bolivia-Lowlands (BOLL)

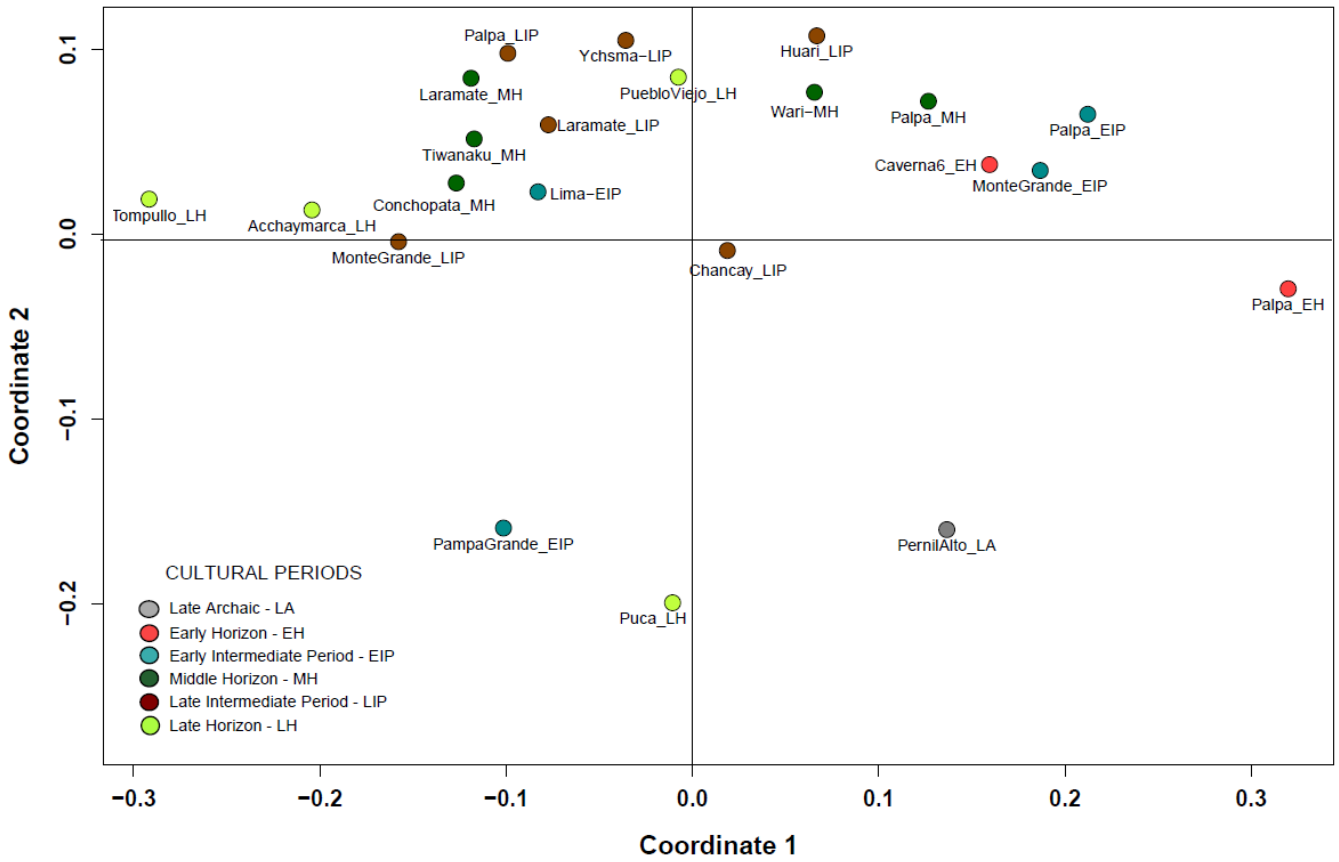


Figure 4. Genetic affinities of ancient populations from South America. Multidimensional scaling plot (MDS) based on Slatkin's Fst. LA—Late Archaic Period: Lauricocha, Pernil Alto; EH—Early Horizon: Caverna6, Palpa; EIP—Early Intermediate Period: Pampa Grande, Monte Grande, Lima; MH–Middle Horizon: Conchopata, Wari, Laramate, Palpa MH, Tiwanaku; LIP—Late Intermediate Period: Laramate, Huari, Palpa, Ychsma, Chancay, Montegrande; LH—Late Horizon: Llullaillaco, Acchaymarca, Puca, Tompullo, Pueblo Viejo.

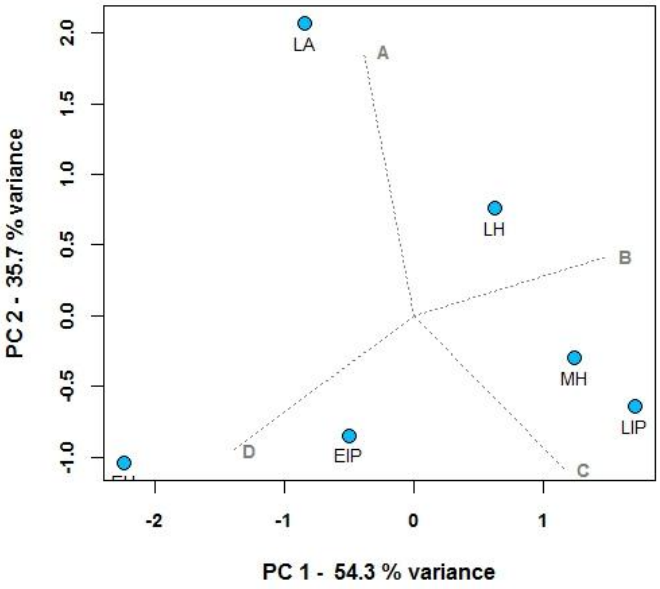


Figure 5. Principal component analysis (PCA) based on major mtDNA haplogroup frequencies from cultural periods. Late Archaic Period–LA, Early Horizon–EH, Early Intermediate Period–EIP, Middle Horizon–MH, Late Intermediate Period–LIP and Late Horizon–LH. In sum, PCA axes 1 and 2 describe 90% of the total variance.

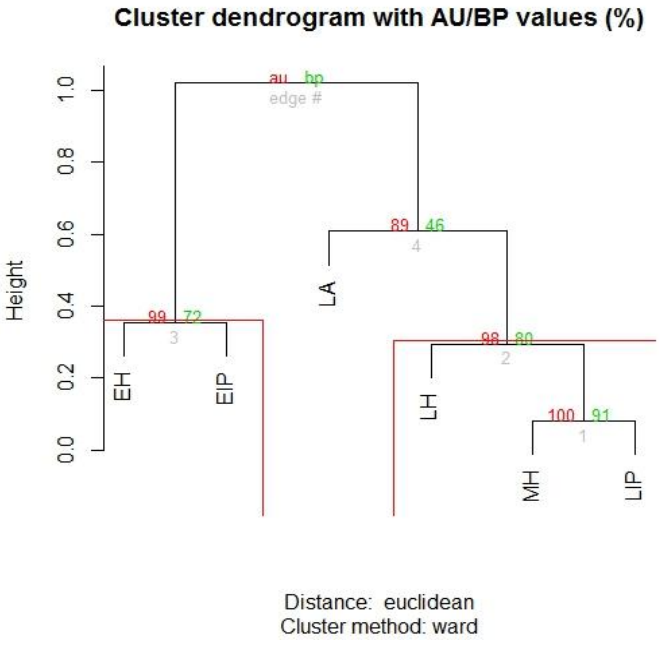


Figure 6. Ward clustering dendrogram of the six archaeological groups in South America. Late Archaic Period–LA, Early Horizon–EH, Early Intermediate Period–EIP, Middle Horizon–MH, Late Intermediate Period–LIP and Late Horizon–LH. Bootstrapping values (red/green) support three main clusters.

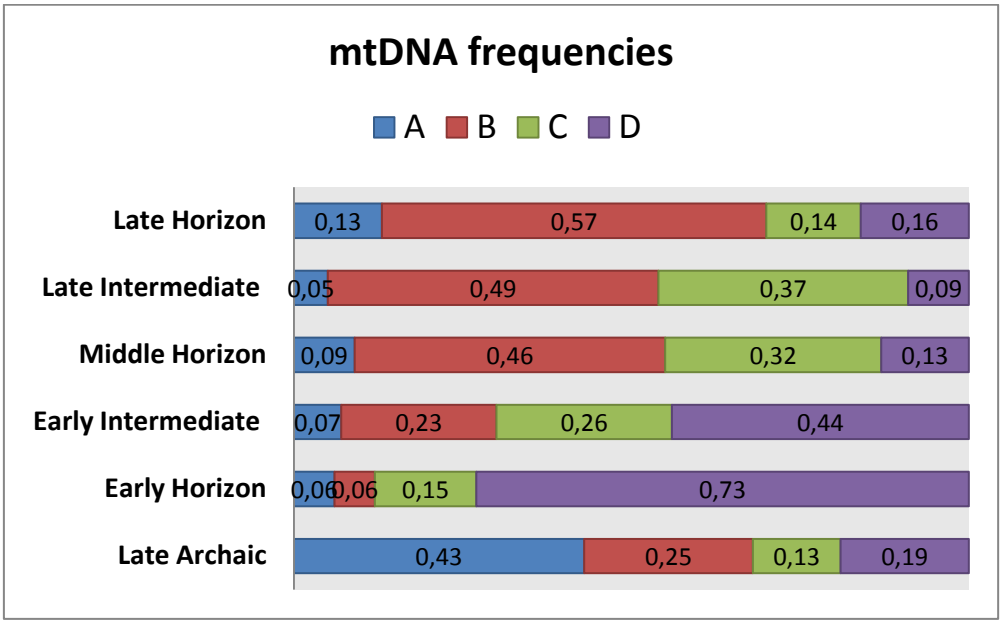


Figure 7. Variation of mtDNA frequencies from six distinct cultural periods in the Central Andes of South America through time.

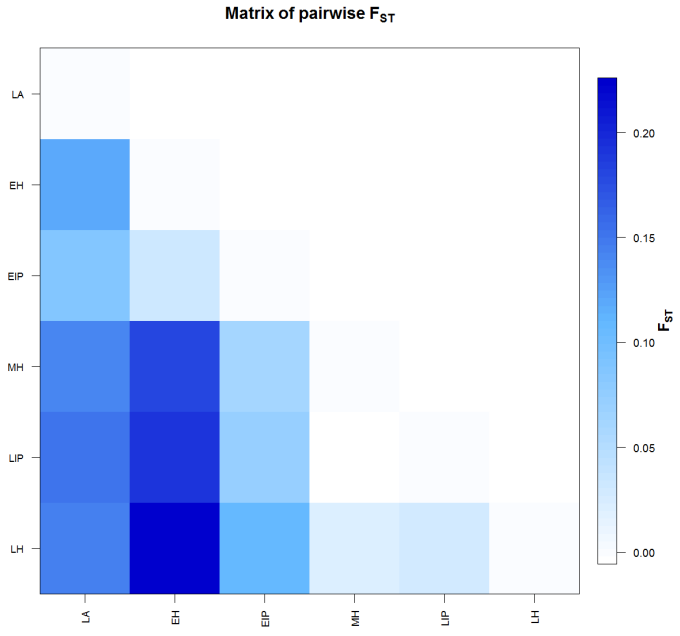


Figure 8. Genetic distances (pairwise (F_{ST}) between archaeological periods: Late Archaic Period–LA, Early Horizon–EH, Early Intermediate Period–EIP, Middle Horizon–MH, Late Intermediate Period–LIP and Late Horizon–LH. Increasing blue intensity indicates larger genetic distance. The heatmap was created using Arlequinv3.5.

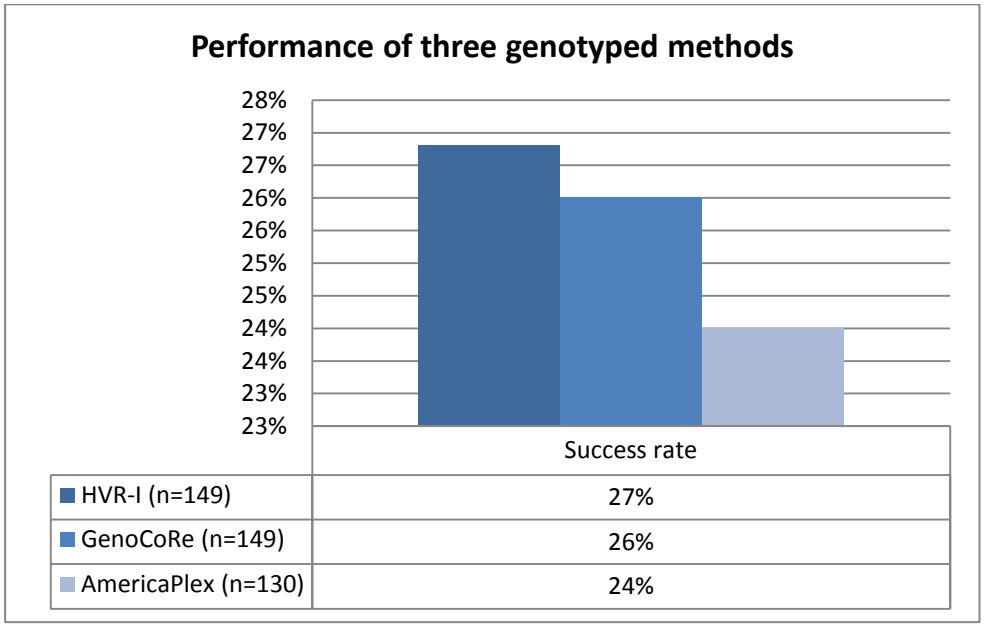


Figure 9. Comparison and haplotyping success of the three methods: HVR-I sequencing, GenoCoRe 22 and AmericaPlex26 multiplex assays.
The success rate is given in percentage of unambiguous haplotype calls for each of the three methods. Chi-square test show 0.1616 (p-value: 0.996907).
The result is not significant at $p < 0.01$

Table 1. List of samples and sites examined in this study

Site Name	Location	Coordinates	Age Range	Description/Notes	Culture (Samples/Individual)
Huaca Pucllana	Lima, Peru	12°06'37.01''S - 77°01'58.93"W	200 – 1475 AD	Huaca Pucllana is an archaeological site in coastal Lima Peru with three successive cultural periods described: First occupation by the Lima in the Early Intermediate Period (200 - 600 AD), a second settlement by the Wari Culture in the Middle Horizon (600 - 1000 AD) and finally occupation by the Ychsma in the Late Intermediate Period (1000 - 1476 AD)	Lima: (32/15) Wari: (43/20) Ychsma: (38/17)
Miramar Necropolis	Lima, Peru		500 – 1000 AD	Samples belong to the Wari culture from Peru. Samples were collected at the Museum Ancon in Lima Peru.	Wari: (10/5)
Pasamayo, Chancay	Lima, Peru		1000 – 1476 AD	Samples collected at the Pasamayo antic cemetery belong to the Chancay Culture (Late Intermediate period)	Chancay: (18/9)
Aspero	Peru	10°48'56.16"S 77°44'27.6"W	3000 – 1800 BC	Aspero and Caral civilization belong to the same late Archaic period, between 3000 and 1800 BC. Aspero was a fishing centre being only 500 meters from the sea.	Late Archaic: (32/16)
Caral Supe	Peru	10°53'36.95"S - 77°31'13.19"W	2500 – 1800 BC	Caral Supe is a ~5000-year-old archaeological site located on a dry desert terrace overlooking the valley of the Supe river. Caral dates back to the Late Archaic Period in the Central Andes and it considered the oldest centre of civilization in the Americas.	Caral Formative: (15/8)
Caral Supe	Peru	10°53'36.95"S - 77°31'13.19"W	2500 – 1800 BC		Caral Late Archaic: (10/5)
PuebloViejo Pucara	Peru	12°12'16.15"S 76°47'57.23"W	1476 – 1534 AD	Pueblo Viejo-Pucara, is a Inca site located on the left bank of the Lurin river that was occupied in Late Horizon period.	Inca: (35/17)
Tablada de Lurin	Peru	12°11'09.32"S 76°54'58.75"W	1476 – 1534 AD	Tablada de Lurin is a complex of prehispanic sites on the Central Coast of Peru, on the right bank of the river Lurin, south of Lima.	Inca: (14/7)
Puemape	Peru	7°31'09.80"S - 79°32'22.63"W		The Puemape site was a coastal fishing village on the North Coast of Peru. Comprises cemeteries and ceremonial structures of various occupations during the Formative periods.	Puemape (4/2)
Palpa	Peru	14°34'50.55"S 75°11'00.65"W	300 – 400 AD	Samples analysed for aDNA replication/authentication. The samples derive from the Early Intermediate Period settlement phase of the rural settlement Jauranga, associated with the Nasca Culture (Nasca 5, ~300-400 AD).	Palpa: (2/1)
Pernil Alto	Peru	14°48'36.35"S 75°05'24.78"W	5800 – 5000 cal BP	Samples analysed for aDNA replication/authentication Pernil Alto was a permanent settlement located in Southern Peru.	Pernil: (1)
Chullpa Botiriayocc	Peru	14°15'20.20"S 74°51'33.78"W	800 BP	Samples analysed for aDNA replication/authentication. Chullpa Botiriayocc is associated with a Late Intermediate settlement close to the modern town of Laramate in the South-Central Peruvian Highlands (Western Slope) above the Palpa Region.	Chullpa: (1)

Site Name	Location	Coordinates	Age Range	Description/Notes	Culture (Samples/Individual)
Lauricocha	Peru	10°19’10.81”S-76°43’43.74”W	6800 – 1500 BC	Lauricocha is a rock shelter in the Central Peruvian Highlands at approx. 4000m elevation. The site has been continuously used for accommodation and burials by humans from the Early to the Late Holocene. Samples derive from individuals found in three burial phases falling into the Early Archaic, ~6800-6300 BC; Late Archaic, ~3500 BC, and Initial Period, ~1800 BC (ages correlate with the burial events, not the ranges of the archaeological periods)	Lauricocha: (7/4)
Tiwanaku	La Paz, Bolivia	16°33'20.11"S - 68°40'24.04"W	724 – 1187 AD	Samples collected in association with the ‘Akapana Project 2009’ led by the ‘Unidad de Arqueologia y Museos - UDAM (Bolivian Archaeology Unit) in collaboration with the Ministerio de Culturas (Bolivian Ministry of Cultures) in the proximities of the Akapana Pyramid at the Tiwanaku archaeological site. Tiwanaku occupation during Epoch / Period V (724 – 1187AD).	Tiwanaku: (18/10)
Pica8–Tarapacá	Chile	20°29'20.45"S - 69°19'44.47"W	950 – 1450 AD	The Cultural Complex Pica-Tarapacá has developed during the Late Intermediate Period. With evidence primarily of ceramics and textiles and it has been defined largely by the remains recovered of the cemetery Pica-8.	Pica8: (8/4)
Azapa-Tarapacá	Chile	20°29'20.45"S - 69°19'44.47"W		Samples from Azapa are derived from the Cabuza Culture	Azapa: (2/1)
Llullaillaco	Argentina	24°43'13.13"S - 68°31'46.45"W	1476 – 1534 AD	Llullaillaco is the world's highest active volcano in the border between Chile and Argentina and belongs to the Puna de Atacama, a region with volcanic peaks on a high plateau within the Atacama Desert. Samples belong to Three mummies: The Lighting girl, The Doncella/Maiden and the Boy/El Niño.	Inca: (6/3)
Patagonia	Chile	51°34'23.9"S - 72°35’09.7"W	2158 – 2268 calBC	Human samples collected at Cueva Chica in the Patagonia region. The site is associated with Late Pleistocene fauna. Radiocarbon dates produced some of the oldest evidence for the presence of ancient mammal’s remains and other species.	Patagonia: (2/1)
Taino	Central America			Samples analysed for aDNA replication/authentication	Taino: (4/2)

Table 2. Genotypes of all samples performed with three different analytical methods, based on replication from two independent samples and extractions per individual.

Location	Culture	Analysed	Individuals	HVR-I Sequences	Success rate %	Genotyped HVR-I	Genotyped GenoCoRe22	Genotyped AmericaPlex
Huaca Pucllana	Lima	32	15	**9**	60%	3	2	7
Huaca Pucllana	Wari	43	20	**10**	50%	-	3	8
Huaca Pucllana	Ychsma	38	17	**15**	88%	11	6	8
Caral Formative		15	8	-	-	-	-	-
Caral (Archaic)		10	5	-	-	-	-	-
Aspero (Archaic)		32	16	-	-	-	-	-
PuebloViejo	Inca	35	17	**11**	64%	13	11	-
Miramar	Wari	10	5	-	-	-	-	1
Pasamayo	Chancay	18	9	**7**	77%	2	3	6
Llullaillaco	Inca	6	3	**3**	100%	3	3	-
Chullpa Botiriayocc		1	1	-	-	-	-	-
Pernil Alto		1	1	-	-	-	-	-
Patagonia		2	1	**1**	100%	1	1	-
Lauricocha	Lauricocha	7	4	**3**	75%	3	3	-
Pica8	Tarapacá	8	4	**3**	75%	-	-	-
Azapa	Tarapacá	2	1	-	-	-	-	-
Tiwanaku	Tiwanaku	18	10	**9**	90%	4	4	2
Palpa	Palpa	2	1	**1**	100%	-	1	-
Puemape	Puemape	4	2	-	-	-	-	-
Tablada Lurin	Inca	14	7	-	-	-	-	-
Taino	Taino	4	2	-	-	-	2	-
TOTAL		**302**	**149**	**72**		**40**	**39**	**32**

Table 3. Details of the haplotyping results for all ancient individuals.
Comparison of three typing methods and consensus haplogroup call

Ind	ACAD #	Ext	Culture	HVR-I	GenoCoRe	America Plex26	Consensus
1	**11115A**	95	Pica-Tarapacá	B?	B?	-	
1	**11116A**	100	Pica-Tarapacá	B4	-	-	
2	**11117A**	95	Pica-Tarapacá	?	B?	-	
2	**11118A**	100	Pica-Tarapacá	A2	-	-	
3	**11119A**	95	Pica-Tarapacá	?	?	-	
3	**11120A**	100	Pica-Tarapacá	-	-	-	
4	**11121A**	95	Pica-Tarapacá	B	?	-	
4	**11122A**	95	Pica-Tarapacá	?	?	-	
5	**11125A**	95	Azapa Late formative	?	?	-	
5	**11127A**	95	Azapa Late formative	D	-	-	
6	**10709A**	106	Ychsma, Pucllana	C1?	?	C1b	
6	**10710A**	85	Ychsma, Pucllana	C1	C?	M	
7	**10712A**	106	Ychsma, Pucllana	**B4**	?	**B2**	**B2**
7	**10713A**	85	Ychsma, Pucllana	**B4**	B	**B2**	**B2**
8	**10715A**	81	Ychsma, Pucllana	?	?	?	
8	**10715B**	82	Ychsma, Pucllana	B4?	B?	**B2**	**B2**
8	**10716A**	106	Ychsma, Pucllana	?	B	**B2**	**B2**
9	**10717A**	106	Ychsma, Pucllana	C1?	?	**M**	**M**
9	**10718A**	81	Ychsma, Pucllana	?	?	**M**	**M**
9	**10718B**	82	Ychsma, Pucllana	?	C?	?	
10	**10719A**	85	Ychsma, Pucllana	**B4**	B	B2	**B4**
10	**10720A**	106	Ychsma, Pucllana	**B4**	?	B2b	**B4**
11	**10722A**	106	Ychsma, Pucllana	**D1**	**D**	?	**D1**
11	**10723A**	85	Ychsma, Pucllana	**D1**	**D**	M?	**D1**
12	**10724A**	107	Ychsma, Pucllana	**C1**	**C**	**C1b**	**C1b**
12	**10725A**	85	Ychsma, Pucllana	**C1**	**C**	**C1b**	**C1b**
13	**10726A**	107	Ychsma, Pucllana	?	?	B2	
13	**10727A**	85	Ychsma, Pucllana	?	B	B2b	
14	**10728A**	81	Ychsma, Pucllana	**B4**	B?	**B2**	**B2**
14	**10728B**	82	Ychsma, Pucllana	**B4**	B?	?	
14	**10729A**	107	Ychsma, Pucllana	**B**	B?	**B2**	**B2**
15	**10730A**	84	Ychsma, Pucllana	**B4**	**B**	B2	**B4**
15	**10730Y**	107	Ychsma, Pucllana	**B4**	**B**	B2b	**B4**
16	**10731A**	84	Ychsma, Pucllana	**C1**	**C**	**C1b**	**C1b**
16	**10731Y**	107	Ychsma, Pucllana	**C1**	**C**	**C1b**	**C1b**
17	**10732A**	84	Ychsma, Pucllana	D?	**D**	?	**D**
17	**10732Y**	107	Ychsma, Pucllana	D1?	**D**	?	**D**
18	**10793A**	109	Ychsma, Pucllana	**B4**	?	**B2**	**B2**

Ind	ACAD	Ext	Culture	HVR-I	GenoCoRe	America Plex26	Consensus
18	**10794A**	83	Ychsma, Pucllana	**B4**	B	**B2**	**B2**
19	**10796A**	109	Ychsma, Pucllana	B4	?	B2b	
19	**10797A**	83	Ychsma, Pucllana	?	?	B2	
20	**10800A**	109	Ychsma, Pucllana	**B4**	B	B2b	**B4**
20	**10801A**	83	Ychsma, Pucllana	**B4**	?	B2	**B4**
21	**10804A**	81	Ychsma, Pucllana	**D1**	?	M	**D1**
21	**10804B**	82	Ychsma, Pucllana	**D1**	?	?	**D1**
21	**10805A**	111	Ychsma, Pucllana	**D1**	D	D1?	**D1**
22	**10809A**	111	Ychsma, Pucllana	**C1**	**C**	**C1b**	**C1b**
22	**10810A**	85	Ychsma, Pucllana	**C1**	**C**	**C1b**	**C1b**
23	**10733A**	81	Wari, Pucllana	?	C?	**C1b**	**C1b**
23	**10733B**	82	Wari, Pucllana	?	C?	**?**	
23	**10734A**	107	Wari, Pucllana	B4?	?	**C1b**	**C1b**
24	**10736A**	83	Wari, Pucllana	?	?	?	
24	**10737A**	107	Wari, Pucllana	A?	?	?	
25	**10738A**	107	Wari, Pucllana	?	?	?	
25	**10739A**	83	Wari, Pucllana	?	?	?	
26	**10741A**	81	Wari, Pucllana	?	B?	**B2**	**B2**
26	**10741B**	82	Wari, Pucllana	?	B?	**?**	
26	**10742A**	107	Wari, Pucllana	?	B	**B2**	**B2**
27	**10744A**	83	Wari, Pucllana	?	?	?	
27	**10745A**	107	Wari, Pucllana	?	B?	?	
28	**10747A**	83	Wari, Pucllana	?	**B**	?	**B**
28	**10748A**	107	Wari, Pucllana	?	**B**	B2	**B**
29	**10749A**	108	Wari, Pucllana	?	**B**	**B2**	**B2**
29	**10750A**	98	Wari, Pucllana	?	**B**	**B2**	**B2**
30	**10751A**	108	Wari, Pucllana	?	?	?	
30	**10752A**	83	Wari, Pucllana	?	D?	?	
31	**10753A**	108	Wari, Pucllana	?	**B**	**B2**	**B2**
31	**10754A**	98	Wari, Pucllana	B4?	**B**	**B2**	**B2**
32	**10756A**	83	Wari, Pucllana	B4?	?	?	
32	**10757A**	108	Wari, Pucllana	?	?	?	
33	**10758A**	85	Wari, Pucllana	?	?	?	
33	**10759A**	108	Wari, Pucllana	?	?	?	
34	**10761A**	85	Wari, Pucllana	?	?	?	
34	**10762A**	108	Wari, Pucllana	?	?	?	
35	**10763A**	108	Wari, Pucllana	C1?	?	C1b	
35	**10764A**	85	Wari, Pucllana	C4?	C?	?	
36	**10765A**	98	Wari, Pucllana	not B	C?	**C1b**	**C1b**
36	**10766A**	108	Wari, Pucllana	?	?	**C1b**	**C1b**
37	**10768A**	108	Wari, Pucllana	?	?	?	
37	**10769A**	81	Wari, Pucllana	not R	?	?	
37	**10769B**	82	Wari, Pucllana	?	?	?	

Ind	ACAD #	Ext	Culture	HVR-I	GenoCoRe	America Plex26	Consensus
38	**10770A**	108	Wari, Pucllana	C?	?	?	
38	**10771A**	98	Wari, Pucllana	not B	C?	C1b	
39	**10772A**	99	Wari, Pucllana	**C1**	?	**C1b**	**C1b**
39	**10773A**	108	Wari, Pucllana	**C1**	?	**C1b**	**C1b**
40	**10774A**	108	Wari, Pucllana	?	?	**A2**	**A2**
40	**10775A**	99	Wari, Pucllana	not B	A?	**A2**	**A2**
41	**10776A**	99	Wari, Pucllana	B4?	?	B2	
41	**10777A**	109	Wari, Pucllana	T2	T	?	
42	**10778A**	99	Wari, Pucllana	D?	?	**A2**	**A2**
42	**10779A**	109	Wari, Pucllana	?	?	**A2**	**A2**
43	**10780A**	81	Lima, Pucllana	?	?	?	
43	**10780B**	82	Lima, Pucllana	?	?	?	
43	**10781A**	109	Lima, Pucllana	?	?	?	
44	**10783A**	109	Lima, Pucllana	K?	?	-	
44	**10784A**	99	Lima, Pucllana	?	?	?	
45	**10785A**	99	Lima, Pucllana	?	?	?	
45	**10786A**	109	Lima, Pucllana	?	?	?	
46	**10787A**	83	Lima, Pucllana	?	?	**B2**	**B2**
46	**10788A**	109	Lima, Pucllana	B	?	**B2**	**B2**
47	**10789A**	109	Lima, Pucllana	B4	?	**B2b**	**B2b**
47	**10790A**	99	Lima, Pucllana	?	B?	**B2b**	**B2b**
48	**10791A**	109	Lima, Pucllana	**A2**	A	**A2**	**A2**
48	**10792A**	81	Lima, Pucllana	**A2**	?	**A2**	**A2**
48	**10792B**	82	Lima, Pucllana	A2	?	?	
49	**10798A**	109	Lima, Pucllana	?	?	?	
49	**10799A**	83	Lima, Pucllana	?	?	?	
50	**10802A**	83	Lima, Pucllana	?	?	**B2**	**B2**
50	**10803A**	111	Lima, Pucllana	?	?	**B2**	**B2**
51	**10806A**	111	Lima, Pucllana	?	?	M	
51	**10807A**	99	Lima, Pucllana	C1	?	?	
52	**10811A**	111	Lima, Pucllana	?	?	M	
52	**10812A**	85	Lima, Pucllana	?	?	?	
53	**10813A**	111	Lima, Pucllana	**B4**	**B**	**B2**	**B2**
53	**10814A**	85	Lima, Pucllana	**B4**	**B**	**B2**	**B2**
54	**10815A**	99	Lima, Pucllana	?	?	?	
54	**10816A**	111	Lima, Pucllana	?	?	?	
55	**10817A**	111	Lima, Pucllana	C1	C?	**C1b**	**C1b**
55	**10818A**	99	Lima, Pucllana	?	C	**C1b**	**C1b**
56	**10819A**	99	Lima, Pucllana	C?	C?	D4b1??	
56	**10820A**	111	Lima, Pucllana	?	C?	C1b	
57	**10821A**	111	Lima, Pucllana	**B4**	**B**	**B2**	**B2**
57	**10822A**	99	Lima, Pucllana	**B4**	**B**	**B2**	**B2**
58	**10824A**	100	Aspero late archaic	?	?	-	

Ind	ACAD #	Ext	Culture	HVR-I	GenoCoRe	America Plex26	Consensus
58	**10825A**	104	Aspero late archaic	?	?	-	
59	**10826A**	104	Aspero late archaic	?	?	-	
59	**10827A**	93	Aspero late archaic	?	?	-	
60	**10828A**	93	Aspero late archaic	?	?	-	
60	**10829A**	93	Aspero late archaic	D4?	?	-	
61	**10830A**	96	Aspero late archaic	A or D?	?	-	
61	**10831A**	104	Aspero late archaic	?	?	-	
62	**10832A**	104	Aspero late archaic	?	?	-	
62	**10833A**	93	Aspero late archaic	?	?	-	
63	**10835A**	84	Aspero late archaic	?	?	-	
63	**10836A**	104	Aspero late archaic	?	?	-	
64	**10839A**	96	Aspero late archaic	?	?	-	
64	**10840A**	104	Aspero late archaic	?	?	-	
65	**10842A**	104	Aspero late archaic	?	?	-	
65	**10843A**	96	Aspero late archaic	?	?	-	
66	**10844A**	84	Aspero late archaic	?	?	-	
66	**10845A**	104	Aspero late archaic	?	?	-	
67	**10846A**	96	Aspero late archaic	?	?	-	
68	**10847A**	96	Aspero late archaic	?	?	-	
68	**10848A**	104	Aspero late archaic	?	?	-	
69	**10849A**	105	Aspero late archaic	?	?	-	
69	**10850A**	84	Aspero late archaic	?	?	-	
70	**10851A**		Aspero late archaic	?	?	-	
70	**10852A**	105	Aspero late archaic	?	?	-	
71	**10853A**	96	Aspero late archaic	B4	?	-	
71	**10854A**	105	Aspero late archaic	?	?	-	
72	**10855A**	84	Aspero late archaic	A	?	-	
72	**10856A**	105	Aspero late archaic	?	?	-	
72	**10857A**	96	Aspero late archaic	not B	?	-	
73	**10858A**	105	Aspero late archaic	?	?	-	
73	**10859A**	96	Aspero late archaic	?	?	-	
74	**10860A**	84	Caral formative	?	?	-	
74	**10861A**	105	Caral formative	?	?	-	
75	**10862A**	105	Caral formative	?	?	-	
75	**10863A**	84	Caral formative	?	?	-	
76	**10864A**	93	Caral formative	?	?	-	
76	**10865A**	105	Caral formative	?	?	-	
77	**10866A**	84	Caral formative	?	?	-	
77	**10867A**	105	Caral formative	?	?	-	
78	**10868A**	105	Caral formative	?	?	-	
78	**10869A**	84	Caral formative	?	?	-	
79	**10870A**	93	Caral formative	?	?	-	
79	**10871A**	105	Caral formative	?	?	-	

Ind	ACAD #	Ext	Culture	HVR-I	GenoCoRe	America Plex26	Consensus
79	**10872A**	93	Caral formative	?	?	-	
80	**10873A**	105	Caral formative	B4	?	-	
80	**10874A**	93	Caral formative	?	?	-	
81	**10875A**	96	Caral formative	notB	?	-	
81	**10876A**	106	Caral formative	?	?	-	
82	**10877A**	84	Caral late archaic	?	?	-	
82	**10878A**	106	Caral late archaic	C1	?	-	
83	**10879A**	84	Caral late archaic	?	?	-	
83	**10880A**		Caral late archaic	?	?	-	
84	**10882A**	84	Caral late archaic	?	?	-	
84	**10883A**	106	Caral late archaic	?	?	-	
85	**10884A**	106	Caral late archaic	?	?	-	
85	**10885A**	93	Caral late archaic	?	?	-	
86	**10886A**	106	Caral late archaic	?	?	-	
86	**10887A**	84	Caral late archaic	?	?	-	
87	**13235A**	114	Tiwanaku, Bolivia	?	?	?	
87	**13236A**	114	Tiwanaku, Bolivia	B4	B	B2	
88	**13237A**	114	Tiwanaku, Bolivia	B	B	B2	
88	**13239A**	114	Tiwanaku, Bolivia	?	?	?	
89	**13240A**	114	Tiwanaku, Bolivia	B	B	B2	
90	**13241A**	114	Tiwanaku, Bolivia	C1	C	M	
90	**13243A**	114	Tiwanaku, Bolivia	?	?	?	
91	**13244A**	114	Tiwanaku, Bolivia	B	B	B2	
91	**13245A**	114	Tiwanaku, Bolivia	?	C?	C1b	
92	**13246A**	114	Tiwanaku, Bolivia	**A2**	**A**	**A2**	**A2**
92	**13247A**	114	Tiwanaku, Bolivia	**A2**	**A**	**A2**	**A2**
93	**13256A**	114	Tiwanaku, Bolivia	**B4**	**B**	**B2**	**B2**
93	**13264A**	115	Tiwanaku, Bolivia	**B4**	**B**	**B2**	**B2**
94	**13267A**	115	Tiwanaku, Bolivia	**B4**	**B**	?	**B4**
94	**13272A**	115	Tiwanaku, Bolivia	**B4**	**B**	B2	**B4**
95	**13275A**	115	Tiwanaku, Bolivia	B	B	B2	
96	**13278A**	115	Tiwanaku, Bolivia	**C1**	**C**	M	**C1**
96	**13279A**	115	Tiwanaku, Bolivia	**C1**	**C**	C1b	**C1**
97	**11152A**	87	Pueblo Viejo	D1	D?	-	
97	**11153A**	97	Pueblo Viejo	B4	D?	-	
98	**11154A**	87	Pueblo Viejo	**C1**	**C**	-	**C1**
98	**11155A**	97	Pueblo Viejo	**C1**	**C**	-	**C1**
99	**11156A**	97	Pueblo Viejo	**C1**	**C**	-	**C1**
99	**11157A**	87	Pueblo Viejo	**C1**	**C**	-	**C1**
100	**11158A**	97	Pueblo Viejo	**C1**	**C**	-	**C1**
100	**11159A**	87	Pueblo Viejo	**C1**	**C**	-	**C1**
101	**11160A**	87	Pueblo Viejo	B4	B	-	

Ind	ACAD #	Ext	Culture	HVR-I	GenoCoRe	America Plex26	Consensus
102	**11161A**	87	Pueblo Viejo	**C1b**	**C**	-	**C1b**
102	**11162A**	97	Pueblo Viejo	**C1b**	**C**	-	**C1b**
103	**11163A**	97	Pueblo Viejo	C1b	C	-	
103	**11164A**	87	Pueblo Viejo	B4	B	-	
104	**11165A**	97	Pueblo Viejo	**B4**	**B**	-	**B4**
104	**11166A**	87	Pueblo Viejo	**B4**	**B**	-	**B4**
105	**11167A**	97	Pueblo Viejo	**B4**	**B**		**B4**
105	**11168A**	87	Pueblo Viejo	**B4**	**B**	-	**B4**
106	**11169A**	97	Pueblo Viejo	B4	B	-	
106	**11170A**	95	Pueblo Viejo	D1	D?	-	
107	**11171A**	95	Pueblo Viejo	B4	B	-	
107	**11172A**	98	Pueblo Viejo	C1?	C	-	
108	**11173A**	98	Pueblo Viejo	**B4**	**B**	-	**B4**
108	**11174A**	95	Pueblo Viejo	**B4**	**B**	-	**B4**
109	**11175A**	97	Pueblo Viejo	**B4**	**B**	-	**B4**
109	**11176A**	98	Pueblo Viejo	**B4**	**B**	-	**B4**
110	**11177A**	97	Pueblo Viejo	**B4**	**B**	-	**B4**
110	**11178A**	98	Pueblo Viejo	**B4**	**B**	-	**B4**
111	**11179A**	97	Pueblo Viejo	**D1**	D?	-	**D1**
111	**11180A**	98	Pueblo Viejo	**D1**	D?	-	**D1**
112	**11181A**	87	Pueblo Viejo	**B4**	**B**	-	**B4**
112	**11182A**	98	Pueblo Viejo	**B4**	**B**	-	**B4**
113	**11183A**	87	Pueblo Viejo	**C1b**	**C**	-	**C1b**
113	**11184A**	98	Pueblo Viejo	**C1b**	**C**	-	**C1b**
114	**11185A**	87	Pueblo Viejo	**D1**	D?	-	**D1**
114	**11186A**	98	Pueblo Viejo	**D1**	D?	-	**D1**
115	**11187A**	103	Wari, Miramar	?	?	**B2**	**B2**
115	**11188A**	87	Wari, Miramar	?	?	**B2**	**B2**
116	**11189A**	87	Wari, Miramar	?	?	?	
116	**11190A**	103	Wari, Miramar	?	?	?	
117	**11191A**	103	Wari, Miramar	?	?	?	
117	**11192A**	87	Wari, Miramar	?	?	?	
118	**11193A**	87	Wari, Miramar	?	?	?	
118	**11194A**	103	Wari, Miramar	?	?	?	
119	**11195A**	87	Wari, Miramar	?	?	?	
119	**11196A**	103	Wari, Miramar	?	?	?	
120	**11197A**	88	Chancay, Pasamayo	?	**A**	**A2**	**A2**
120	**11198A**	103	Chancay, Pasamayo	?	**A**	**A2**	**A2**
121	**11199A**	103	Chancay, Pasamayo	**B4**	?	**B2**	**B2**
121	**11200A**	88	Chancay, Pasamayo	**B4**	B?	**B2**	**B2**
122	**11201A**	88	Chancay, Pasamayo	?	**B**	**B2**	**B2**
122	**11202A**	103	Chancay, Pasamayo	B4	**B**	**B2**	**B2**

Ind	ACAD #	Ext	Culture	HVR-I	GenoCoRe	America Plex26	Consensus
123	**11203A**	88	Chancay, Pasamayo	?	A	?	
123	**11204A**	103	Chancay, Pasamayo	?	?	?	
124	**11205A**	103	Chancay, Pasamayo	A?	?	**C1b**	**C1b**
124	**11206A**	88	Chancay, Pasamayo	?	C?	**C1b**	**C1b**
125	**11207A**	103	Chancay, Pasamayo	?	?	?	
125	**11208A**	88	Chancay, Pasamayo	B4	B	B2	
126	**11209A**	103	Chancay, Pasamayo	**C1**	**C**	**M**	**C1**
126	**11210A**	88	Chancay, Pasamayo	**C1**	**C**	**M**	**C1**
127	**11211A**	95	Chancay, Pasamayo	C	C	**M**	
127	**11212A**	104	Chancay, Pasamayo	C1?	?	**M**	
128	**11213A**	104	Chancay, Pasamayo	?	?	D4b1?	
128	**11214A**	95	Chancay, Pasamayo	?	?	D4b1?	
129	**6317A**	32	Llullaillaco, Lightning girl	**D**	**D**	-	**D**
129	**6318A**	32	Llullaillaco, Lightning girl	**D**	**D**	-	**D**
130	**6319A**	32	Llullaillaco, La Doncella	**D**	**D**	-	**D**
130	**6320A**	32	Llullaillaco, La Doncella	**D**	**D**	-	**D**
131	**6321A**	32	Llullaillaco, The Boy	**C**	**C**	-	**C**
131	**6322A**	32	Llullaillaco, The Boy	**C**	**C**	-	**C**
132	**10891A**	81	Patagonia, Chile	**B4**	**B**	-	**B4**
132	**10891B**	82	Patagonia, Chile	**B4**	**B**	-	**B4**
133	**6691A**	38	Taino	B?	**B**	-	**B**
133	**6692A**	55	Taino	?	**B**	-	**B**
134	**6693A**	38	Taino	?	**C**	-	**C**
134	**6694A**	55	Taino	?	**C**	-	**C**
135	**12686A**	110	Lauricocha (Lau1Esq9)	A2p	A	-	
136	**12687A**	110	Lauricocha (Lau3Esq6)	A2	A	-	
136	**12076A**	100	Lauricocha (Lau3Esq6)	A2?	A?	-	
137	**12077A**	100	Lauricocha (Lau4Esq2)	**A2**	A	-	
137	**12688A**	110	Lauricocha (Lau4Esq2)	**A2**	A	-	
138	**12689A**	110	Lauricocha (Lau5Esq1)	B4	B	-	
138	**12078A**	100	Lauricocha (Lau5Esq1)	B	?	-	
139	**12690A**	110	Chullpa Botigiriayocc	C1?	C?	-	
139	**12691A**	110	Pernil Alto, Peru	C1?	C?	-	
140	**12692A**	110	Jauranga (cyclops)	C1?	**C**	-	**C**
140	**12693A**	110	Jauranga (Palpa)	C1?	**C**	-	**C**

Ind	ACAD #	Ext	Culture	HVR-I	GenoCoRe	America Plex26	Consensus
141	**12303A**	113	Tablada de Lurin	?	?	-	
141	**12304A**	113	Tablada de Lurin	?	?	-	
142	**12306A**	113	Tablada de Lurin	?	?	-	
142	**12310A**	113	Tablada de Lurin	?	?	-	
143	**12317A**	113	Tablada de Lurin	?	?	-	
143	**12310A**	113	Tablada de Lurin	?	?	-	
144	**12322A**	113	Tablada de Lurin	?	?	-	
144	**12324A**	113	Tablada de Lurin	?	?	-	
145	**12327A**	113	Tablada de Lurin	?	?	-	
145	**12328A**	113	Tablada de Lurin	?	?	-	
146	**12329A**	113	Tablada de Lurin	?	?	-	
146	**12332A**	113	Tablada de Lurin	?	?	-	
147	**12303A**	113	Tablada de Lurin	?	?	-	
147	**12333A**	113	Tablada de Lurin	?	?	-	
148	**14117A**	119	Puemape	?	?	-	
148	**14118A**	119	Puemape	?	?	-	
149	**14119A**	119	Puemape	?	?	-	
149	**14120A**	119	Puemape	?	?	-	
	TOTAL	**302**		**40**	**39**	**32**	**60**
	EBC 10915	81		?	?	?	**-**
	EBC 10923	82		?	?	?	**-**
	EBC 10915	81		?	?	?	**-**
	EBC 10988	85		?	?	?	**-**
	EBC 12060	99		?	?	?	**-**

Note: Consensus haplogroups were called based on both replicates from independent extractions. (?)/(-): Insufficient or no sequence or genotyped information.
EBC: Extraction Blank Control.

Table 4. Haplogroup determination of 72 HVR-I sequences analysed in this study

#	ACAD	Populations	Location	Radiocarbon dates	Haplogroup
1	12686*	Lauricocha Lau1	Peru	8589 - 8482 calBP	**A2**
2	SO238*	Lauricocha Lau2	Peru	8589 - 8482 calBP	**A2**
3	12689*	Lauricocha Lau5	Peru	8700 - 8599 calBP	**B2**
4	13247	Tiwanaku	Bolivia		**A2**
5	13237	Tiwanaku	Bolivia		**B2**
6	13264*	Tiwanaku	Bolivia	1017 - 1155 calAD	**B2**
7	13272	Tiwanaku	Bolivia		**B2**
8	13240	Tiwanaku	Bolivia		**B2**
9	13275	Tiwanaku	Bolivia		**B2**
10	13241*	Tiwanaku	Bolivia	901 - 1024 calAD	**C1c**
11	13279	Tiwanaku	Bolivia		**C1b**
12	13245	Tiwanaku	Bolivia		**C**
13	11116	Pica8-Tarapacá	Chile		**B2**
14	11118	Pica8-Tarapacá	Chile		**B2**
15	11121	Pica8-Tarapacá	Chile		**B2**
16	11200	Chancay	Peru		**B2**
17	11211	Chancay	Peru		**C1b**
18	11205	Chancay	Peru		**C1b**
19	11209	Chancay	Peru		**C1c**
20	11213	Chancay	Peru		**D1**
21	11197	Chancay	Peru		**A2**
22	11208	Chancay	Peru		**B2b**
23	6322	Llullaillaco	Argentina		**C1b**
24	6320	Llullaillaco	Argentina		**D1**
25	6317	Llullaillaco	Argentina		**D**
26	10791*	Lima	Peru	584-660 calAD	**A2**
27	10814	Lima	Peru		**B2**
28	10802	Lima	Peru		**B2**
29	10789	Lima	Peru		**B2b**
30	10817*	Lima	Peru	534-642 calAD	**C1b**
31	10820	Lima	Peru		**C1b**
32	10806	Lima	Peru		**C1c**
33	10811	Lima	Peru		**D***
34	10821	Lima	Peru		**B2**
35	10774	Wari	Peru		**A2**
36	10754*	Wari	Peru	974-1220 calAD	**B2b**
37	10734*	Wari	Peru	776-968 calAD	**C1b**
38	10763	Wari	Peru		**C1b**
39	10771	Wari	Peru		**C1b**
40	10773	Wari	Peru		**C1b**
41	10778	Wari	Peru		**A2**

#	ACAD	Populations	Location	Radiocarbon dates	Haplogroup
42	10742	Wari	Peru		**B2b**
43	10750	Wari	Peru		**B2b**
44	10765	Wari	Peru		**C1b**
45	10713	Ychsma	Peru		**B2**
46	10729	Ychsma	Peru		**B2**
47	10720	Ychsma	Peru		**B2b**
48	10726	Ychsma	Peru		**B2b**
49	10709*	Ychsma	Peru	1244-1288 calAD	**C1b**
50	10717	Ychsma	Peru		**C1b**
51	10725	Ychsma	Peru		**C1b**
52	10731	Ychsma	Peru		**C1b**
53	10722*	Ychsma	Peru	1221-1278 calAD	**D1**
54	10732	Ychsma	Peru		**D***
55	10730	Ychsma	Peru		**B2b**
56	10805*	Ychsma	Peru	1223 - 1280 calAD	**D1**
57	10810*	Ychsma	Peru	1149 - 1249 calAD	**C1d**
58	10794	Ychsma	Peru		**B2**
59	10800	Ychsama	Peru		**B2b**
60	12692	Palpa- Jauranga	Peru		**C1c**
61	10891*	Patagonia	Chile	2158 – 2268 calBC	**B**
62	11165	Pueblo Viejo	Peru		**B2**
63	11168	Pueblo Viejo	Peru		**B2**
64	11173	Pueblo Viejo	Peru		**B2**
65	11182	Pueblo Viejo	Peru		**B2b**
66	11176	Pueblo Viejo	Peru		**B2b**
67	11154	Pueblo Viejo	Peru		**C1b**
68	11157	Pueblo Viejo	Peru		**C1b**
69	11161	Pueblo Viejo	Peru		**C1b**
70	11183	Pueblo Viejo	Peru		**C1c**
71	11180	Pueblo Viejo	Peru		**D1**
72	11185	Pueblo Viejo	Peru		**D1**

*: These sequences have direct radiocarbon dates; all other dates are based on the relevant archaeological period.

Cal AD: calibrated radiocarbon years Anno Domini.

Cal BP: calibrated radiocarbon years Before Present

Table 5. List of populations used to perform the comparative analysis

MODERN POPULATIONS	n	in MDS	Country	Region	Reference
Arequipa	**22**	ARE	Peru	Central Andean	Fuselli et al., 2003
San Martin	**21**	SAN	Peru	Central Andean	Fuselli et al., 2004
Tayacaja	**59**	TAY	Peru	Central Andean	Fuselli et al., 2005
Ancash	**33**	ANC	Peru	Central Andean	Lewis et al., 2005
Puno-Quechua	**30**	QUEP	Peru	Central Andean (Highland)	Lewis et al., 2007
Puno-Aymara	**14**	AYMP	Peru	Central Andean (Highland)	Lewis et al., 2007
Yungay	**36**	YUN	Peru	Central Andean	Lewis et al., 2007
Tupe	**16**	TUP	Peru	Central Andean	Lewis et al., 2007
Titicaca-Quechua	**37**	QUET	Peru	Central Andean (Highland)	Barbieri et al., 2011
Titicaca-Aymara	**20**	AYMT	Peru	Central Andean (Highland)	Barbieri et al., 2012
Titicaca-Uros	**7**	URT	Peru	Central Andean (Highland)	Barbieri et al., 2013
Cajamarca_QUE	**19**	CAJ	Peru	Central Andean	Sandoval et al., 2013
Huancavelica_QUE	**26**	HUAQ	Peru	Central Andean	Sandoval et al., 2013
Apurimac_QUE	**10**	APU	Peru	Central Andean	Sandoval et al., 2013
Machiguenga_ARA	**11**	MAC	Peru	Central Andean	Sandoval et al., 2013
Puno_URU	**25**	PUNU	Peru	Central Andean	Sandoval et al., 2013
Urus_Ar_Yanesha	**18**	URUY	Peru	Central Andean	Sandoval et al., 2013
Amantani_QUE	**26**	AMAQ	Peru	Central Andean	Sandoval et al., 2013
SantaRosa_AYM	**18**	SANA	Peru	Central Andean	Sandoval et al., 2013
Puno_QUE	**35**	PUNQ	Peru	Central Andean	Sandoval et al., 2013
Capachica_QUE	**15**	CAP	Peru	Central Andean	Sandoval et al., 2013
Chimu_AYM	**16**	CHI	Peru	Central Andean	Sandoval et al., 2013
Cusco_AYM	**36**	CUSA	Peru	Central Andean	Sandoval et al., 2013
Ancash_QUE	**10**	ANCQ	Peru	Central Andean	Sandoval et al., 2013
Gran Chaco	**204**	GCH	Paraguay	Gran Chaco	Cabana et al., 2006
Guarani	**200**	GUA	Paraguay	Gran Chaco	Marrero et al., 2007
Kaingang	**74**	KAI	Paraguay	Gran Chaco	Marrero et al., 2007
Paraguay	**63**	PAR	Paraguay	Gran Chaco	Schmitt et al., 2004
Mapuche	**34**	MAP1	Chile	South Andean	Moraga et al., 2000
Pehuenche	**24**	PEH2	Chile	South Andean	Moraga et al., 2000
Yaghan	**15**	YAG	Chile	Tierra de Fuego	Moraga et al., 2000
Mapuche	**18**	MAP2	Chile	South Andean	de Saint Pierre et al., 2012
Yamana	**21**	YAM	Chile	Tierra de Fuego	de Saint Pierre et al., 2012
Tehuelche	**23**	TEU	Chile	South Andean	de Saint Pierre et al., 2012
Atacameño	**28**	ATAC	Chile	Central Andean	de Saint Pierre et al., 2012

MODERN POPULATIONS	n	in MDS	Country	Region	Reference
Aymara	**39**	AYMC	Chile	Central Andean	de Saint Pierre et al., 2012
Pehuenche	**41**	PEH	Chile	South Andean	de Saint Pierre et al., 2012
Huilliche	**47**	HUI	Chile	South Andean	de Saint Pierre et al., 2012
Kawwesqar	**13**	KAW	Chile	Tierra de Fuego	Moraga et al., 2010
Bolivia Lowlands	**53**	BOLL	Bolivia	Central Andean	Bert et al., 2004
Quechua_Bol	**93**	QUEB	Bolivia	Central Andean (Highland)	Gaya-Vidal et al., 2011
Aymara_Bol	**97**	AYMB	Bolivia	Central Andean (Highland)	Gaya-Vidal et al., 2011
Llapallapani_URU	**5**	LLA	Bolivia	Central Andean	Sandoval et al., 2013
Chipaya_URU	**8**	CHI	Bolivia	Central Andean	Sandoval et al., 2013
LaPaz_AYM	**7**	LAP	Bolivia	Central Andean	Sandoval et al., 2013
Andamarca_AYM	**19**	AND	Bolivia	Central Andean	Sandoval et al., 2013
Desaguadero_AYM	**11**	DES	Bolivia	Central Andean	Sandoval et al., 2013
Pajchiri_AYM	**20**	PAJ	Bolivia	Central Andean	Sandoval et al., 2013
Oruro_AYM	**13**	ORU	Bolivia	Central Andean	Sandoval et al., 2013
Potosi_QUE	**29**	POT	Bolivia	Central Andean	Sandoval et al., 2013
SantaAna_AYM	**11**	SAT	Bolivia	Central Andean	Sandoval et al., 2013
CintaLarga	**8**	CIN	Brazil	Amazon	Ramallo et al., 2013
Gorotire	**11**	GOR	Brazil	Amazon	Ramallo et al., 2013
Jamamadi	**13**	JAM	Brazil	Amazon	Ramallo et al., 2013
Kuben-Kran-Kegn	**18**	KKK	Brazil	Amazon	Ramallo et al., 2013
Munduruku	**14**	MUN	Brazil	Amazon	Ramallo et al., 2013
Arara	**19**	ARA	Brazil	Amazon	Ramallo et al., 2013
Mekranoti	**24**	MEK	Brazil	Amazon	Ramallo et al., 2013
Gaviao	**27**	GAV	Brazil	Amazon	Ward et al., 1996
Xavante	**25**	XAV	Brazil	Amazon	Ward et al., 1996
Zoro	**30**	ZOR	Brazil	Amazon	Ward et al., 1996
Yanomami	**155**	YAN	Brazil	Amazon	Williams et al., 2002
Chibchan	**80**	CHI	Colombia	North Western	Melton et al., 2007
Arawaken	**29**	ARW	Colombia	North Western	Melton et al., 2007
Waorani	**25**	WAO	Brazil	North Western	Cardoso et al., 2012
Ngoebe	**15**	NGO	Panama	North Western	Kolman et al., 1995

ANCIENT POPULATIONS	n	in MDS	Country	Region	Reference
Caverna6_EH	**7**	CAV	Peru	Central Andean (Coastal)	Fehren-Schmitz et al, 2014
Palpa_MH	**11**	PMH	Peru	Central Andean (Coastal)	Fehren-Schmitz et al, 2014
Palpa_EIP	**56**	PEIP	Peru	Central Andean (Coastal)	Fehren-Schmitz et al, 2014

Laramate_LIP	**38**	LLIP	Peru	Central Andean (Highland)	Fehren-Schmitz et al, 2014
Palpa_EH	**26**	PEH	Peru	Central Andean (Coastal)	Fehren-Schmitz et al, 2014
MonteGrande_EIP	**11**	MEIP	Peru	Central Andean (Coastal)	Fehren-Schmitz et al, 2014
Laramate_MH	**39**	LAR	Peru	Central Andean (Highland)	Fehren-Schmitz et al, 2014
Palpa_LIP	**11**	PLIP	Peru	Central Andean (Coastal)	Fehren-Schmitz et al, 2014
MonteGrande_LIP	**11**	MLIP	Peru	Central Andean (Coastal)	Fehren-Schmitz et al, 2014
Conchapata_MH	**10**	CON	Peru	Central Andean (Highland)	Kemp et al., 2009
Huari_LIP	**17**	HUA	Peru	Central Andean (Highland)	Kemp et al., 2009
Pernil Alto	**13**	PER	Peru	Central Andean (Coastal)	Fehren-Schmitz et al, 2010
Tompullo2	**24**	TOM	Peru	Central Andean (Highland)	Fehren-Schmitz et al, 2014
Acchaymarca	**14**	ACC	Peru	Central Andean (Highland)	Fehren-Schmitz et al, 2014
Puca	**11**	PUC	Peru	Central Andean (Highland)	Fehren-Schmitz et al, 2014
Pampa Grande_EIP-MH"	**19**	PGD	Argentina	Central Andean (Highland)	Fehren-Schmitz et al, 2014
Huaca Pucllana - Lima	**9**	LIM	Peru	Central Andean (Coastal)	This Study
Huaca Pucllana - Wari	**7**	WAR	Peru	Central Andean (Coastal)	This Study
Huaca Pucllana - Ychsma	**15**	YCH	Peru	Central Andean (Coastal)	This Study
Chancay	**7**	CHA	Peru	Central Andean (Coastal)	This Study
Lauricocha_LA	**3**	LAU	Peru	Central Andean (Highland)	This Study
Pueblo Viejo	**11**	PVJ	Peru	Central Andean	This Study
Tiwanaku	**9**	TIW	Bolivia	Central Andean (Highland)	This Study
Patagonia	**1**	PAT	Argentina	Tierra de Fuego	This Study
Llullaillaco	**3**	LLU	Argentina	Central Andean (Highland)	This Study
Pica8 - Tarapacá	**3**	PIC	Chile	Central Andean	This Study

Table 6. Summary statistics of all ancient populations analysed in this study including those from published sources. Sites are listed in chronological order.

ANCIENT POPULATIONS	*n*	*H*	*h*	*π*	Tajima's D	p-value	Fu's FS	p-value	Reference
Lauricocha_LA	**3**	3	1.0000 (0.2722)	0.017045 (0.013918)	0.00000	0.75700	0.58779	0.39800	This Study
Pernil Alto_LA	**13**	13	1.0000 (0.0302)	0.013986 (0.008185)	-0.64183	0.26800	-9.50099	0.0000*	Fehren-Schmitz et al, 2010
Palpa_EH	**26**	26	1.0000 (0.0107)	0.011329 (0.006521)	-1.33453	0.07100	-25.88853	0.0000*	Fehren-Schmitz et al, 2014
Caverna6_EH	**7**	7	1.0000 (0.0764)	0.009470 (0.006321)	0.10944	0.54700	-4.09202	0.00700	Fehren-Schmitz et al, 2014
Palpa_EIP	**56**	56	1.0000 (0.034)	0.013609 (0.007486)	-1.23179	0.06700	-25.61216	0.0000*	Fehren-Schmitz et al, 2014
MonteGrande_EIP	**11**	11	1.0000 (0.0388)	0.011396 (0.006957)	-0.97069	0.17100	-8.14994	0.0000*	Fehren-Schmitz et al, 2014
Pampa Grande_EIP	**19**	9	0.9123 (0.0350)	0.015982 (0.008966)	1.16210	0.91400	0.14599	0.53700	Carnese et al., 2010
Lima_EIP	**9**	9	1.0000 (0.0302)	0.015115 (0.008770)	-0.55830	0.30800	-9.02299	0.0000*	This Study
Conchapata_MH	**10**	9	0.9778 (0.0540)	0.018598 (0.010946)	-0.23355	0.42000	-2.98332	0.04200	Kemp et al., 2009
Laramate_MH	**39**	39	1.0000 (0.0058)	0.014224 (0.007848)	-1.01739	0.16500	-25.52844	0.0000*	Fehren-Schmitz et al, 2014
Wari_MH	**11**	11	1.0000 (0.038)	0.017352 (0.010319)	-0.15220	0.46700	-7.06962	0.00100	This Study
Tiwanaku_MH	**9**	9	1.0000 (0.0524)	0.013683 (0.008406)	-.050318	0.33700	-5.17413	0.00500	This Study
Palpa_MH	**11**	11	1.0000 (0.038)	0.013157 (0.007887)	-0.68906	0.27000	-7.40723	0.0000*	Fehren-Schmitz et al, 2014
Laramate_LIP	**38**	38	1.0000 (0.0060)	0.016197 (0.008817)	-0.79081	0.22200	-25.34731	0.0000*	Fehren-Schmitz et al, 2014
Huari_LIP	**17**	13	0.9485 (0.0435)	0.015701 (0.008964)	-0.54383	0.31800	-5.02149	0.00600	Kemp et al., 2009
Palpa_LIP	**11**	11	1.0000 (0.0388)	0.013572 (0.008106)	-0.01494	0.53300	-7.25302	0.0000*	Fehren-Schmitz et al, 2014
MonteGrande_LIP	**11**	11	1.0000 (0.0388)	0.014090 (0.008378)	-0.66252	0.28500	-7.06962	0.0000*	Fehren-Schmitz et al, 2014
Ychsma_LIP	**15**	15	1.0000 (0.0388)	0.012913 (0.007756)	0.47532	0.72700	-7.48700	0.0000*	This Study

Chancay_LIP	**7**	7	1.0000 (0.0625)	0.015617 (0.009761	-0.41068	0.388000	-4.16080	0.00900	This Study
Pica8_LIP	**3**	3	1.0000 (0.2722)	0.001894 (0.002362)	0.00000	0.98800	-2.19722	0.0000*	This Study
Acchaymarca_LH	**14**	11	0.9560 (0.0447)	0.019209 (0.011203)	-0.81700	0.21400	-4.03828	0.02800	Baca et al., 2014
Puca_LH	**11**	8	0.9273 (0.0665)	0.022982 (0.013447)	0.23073	0.61900	-1.09443	0.25800	Baca et al., 2014
Pueblo Viejo_LH	**11**	11	1.0000 (0.0388)	0.011983 (0.007265)	1.56656	0.95600	-7.87008	0.0000*	This Study
Llullaillaco_LH	**3**	3	1.0000 (0.2722)	0.007576 (0.006799)	0.00000	0.84800	-0.34093	0.22600	This Study
Tompullo_LH	**24**	11	0.9203 (0.0304)	0.017348 (0.009909)	-0.53548	0.32800	-1.52982	0.27200	Baca et al., 2012

H: Number of haplotypes, h: Haplotype diversity, π: Nucleotide diversity, *: *statistically significant p-values (<0.05 for Tajima's D, p<0.02 for Fu's FS*

Table 7. Summary statistics of samples when grouped into archaeological periods

CULTURAL PERIOD	***n***	***H***	***h***	***π***	**Tajima's D**	**p-value**	**Fu's FS**	**p-value**	**mtDNA Hg frequencies**			
									A	**B**	**C**	**D**
Late Archaic	**16**	16	1.0000 (0.0221)	0.014441 (0.008283)	-0.63000	0.30300	-13.26769	0.00000*	0.43	0.25	0.13	0.19
Early Horizon	**33**	33	1.0000 (0.0075)	0.010955 (0.006280)	-1.21281	0.11100	-25.90656	0.00000*	0.06	0.06	0.15	0.73
Early Intermediate	**94**	84	0.9966 (0.0025)	0.015219 (0.008207)	-1.20122	0.09300	-25.38020	0.00000*	0.06	0.23	0.27	0.44
Middle Horizon	**79**	78	0.9997 (0.0020)	0.016183 (0.008759)	-1.55983	0.03300*	-25.44400	0.00000*	0.09	0.46	0.32	0.13
Late Intermediate	**103**	99	0.9987 (0.0018)	0.015631 (0.008469)	-1.56040	0.03200*	-25.44410	0.00000*	0.05	0.49	0.35	0.09
Late Horizon	**63**	44	0.9846 (0.0063)	0.020166 (0.011005)	-0.75500	0.26400	-25.53392	0.00000*	0.13	0.57	0.14	0.16

H: Number of haplotypes, h: Haplotype diversity, π: Nucleotide diversity, *: *statistically significant p-values (<0.05 for Tajima's D, p<0.02 for Fu's FS)*

Table 8. Genetic distances based on HVR-I sequences computed for ancient populations grouped into archaeological periods. The table provides pairwise F_{ST} (regular) and p-values (italicized).

	Late Archaic_LA	**Early Horizon_EH**	**Early Intermediate Period_EIP**	**Middle Horizon_MH**	**Late Intermediate Period_LIP**	**Late Horizon-LH**
Late Archaic-LA	0	*0.00000+-0.0000*	*0.00000+-0.0000*	*0.00000+-0.0000*	*0.00000+-0.0000*	*0.00000+-0.0000*
Early Horizon-EH	0.11798	0	*0.02703+-0.0139*	*0.00000+-0.0000*	*0.00000+-0.0000*	*0.00000+-0.0000*
Early Intermediate Period-EIP	0.08489	0.03055	0	*0.00000+-0.0000*	*0.00000+-0.0000*	*0.00000+-0.0000*
Middle Horizon-MH	0.13979	0.18023	0.06453	0	*0.72973+-0.0345*	*0.04505+-0.0203*
Late Intermediate Period-LIP	0.15416	0.19107	0.07283	-0.00560	0	*0.00901+-0.0091*
Late Horizon-LH	0.14599	0.22595	0.10870	0.01982	0.02979	0

Table 9. Population genetic structure estimated from the analysis of molecular variance (AMOVA) based on the HVR-I sequences of mtDNA (354bp).

			Variance components (%)			
Population grouping	**Number of groups**	**Number of populations**	**Among groups**	**Among populations/ within groups**	**Within populations**	**p**
No grouping	1	92		24.10	75.90	*0.000*
Time Period						
Ancient vs. Modern	2	92	**0.82%**	23.74%	75.44%	*0.2590*
Cultural periods (ancient only)*	6	27	**5.22%**	5.19%	89.25%	*0.0225*
Altitude (*highlands-lowlands*)	2	92	**3.70%**	21.85%	73.59%	*0.0358*
Region (Amazonian-Andean)	2	60	**1.81%**	12.60%	85.38%	*0.0269*
Geographic criteria**	8	92	**8.63%**	16.46%	73.89%	*0.0024*
Country (modern only)***	6	66	**6.09%**	20.80%	73.89%	*0.0136*

Significant p value ($p<0.05$)

*****Archaic, Early Horizon, Early Intermediate Period, Middle Horizon, Late Intermediate Period, Late Horizon

******Amazon, Tierra de Fuego, South Andean, Central Andean, Central Andean [highland], Central Andean [coast], Gran Chaco, North Western

******* Peru, Bolivia, Paraguay, Colombia, Chile, Brazil

CHAPTER 3

High-resolution mitochondrial genome sequencing from ancient pre-Columbian populations reveals new insights into South American genetic diversity

Statement of Authorship

Chapter 3

Title of the Paper	**High-resolution mitochondrial genome sequencing from ancient pre-Columbian populations reveals new insights into South American genetic diversity**
Publication status	Published, Accepted for Publication, Submitted for Publication, **Publication style**
Publication details	

Author Contributions

By signing the Statement of Authorship, each author certifies that their stated contribution to the publication is accurate and that permission is granted for the publication to be included in the candidate's thesis.

Name of the principal Author	**Guido Valverde** (Candidate)	
Contribution to the paper	Performed DNA extractions, PCR amplifications, sequencing and downstream data processing on all samples, interpreted the results, created figures and tables and wrote the manuscript.	
Signature		Date: 8/06/15

Name of the co-author	María Inés Barreto	
Contribution to the paper	Contributed archaeological material. Assisted with sample collection, provided archaeological and context information.	
Signature		Date: May 29 2015

Name of the co-author	Isabel Flores Espinoza	
Contribution to the paper	Provided samples for analysis. Assisted with sample collection, provided archaeological and context information.	
Signature		Date 02.06.2015

Name of the co-author	Lucia Watson	
Contribution to the paper	Contributed archaeological material. Assisted with sample collection, provided archaeological and context information.	
Signature		Date: 05-30-2015

Name of the co-author	Krzysztof Makowski	
Contribution to the paper	Provided samples for analysis. Contributed reagents, materials, analysis tools, assisted with sample collection, provided archaeological and context information.	
Signature		Date: 06.10.2015

Name of the co-author	Calogero Santoro	
Contribution to the paper	Provided samples for analysis. Contributed reagents, materials, analysis tools, assisted with sample collection, provided archaeological and context information.	
Signature		Date: June 1, 2015

Name of the co-author	Lars Fehren-Schmitz	
Contribution to the paper	Provided samples for analysis. Contributed reagents, materials, analysis tools, provided archaeological and context information. Contributed with data generation and assisted in the interpretation of results.	
Signature		Date: 06/01/2015

Name of the co-author	Alan Cooper	
Contribution to the paper	Contributed reagents, materials, analysis tools assisted in the interpretation of results.	
Signature		Date: October, 3 2015

Name of the co-author	Bastien Llamas	
Contribution to the paper	Co-developed the research concept and experimental design of the study, gave advice on laboratory work, assisted with the interpretation of results and edited the manuscript	
Signature		Date: October, 23 2015

Name of the co-author	Wolfgang Haak	
Contribution to the paper	Co-developed the research concept and experimental design of the study, gave advice on laboratory work, assisted with the interpretation of results and edited the manuscript	
Signature		Date: 

High-resolution mitochondrial genome sequencing from ancient Pre-Columbian populations reveals new insights into South American genetic diversity

Guido Valverde[1], Maria Inés Barreto[2], Isabel Flores Espinoza[2], Lucia Watson[3,4], Krzysztof Makowski[4], Calogero Santoro[5,6], Lars Fehren-Schmitz [7,8], Alan Cooper[1], Bastien Llamas[1], Wolfgang Haak[1]

[1]Australian Centre for Ancient DNA, School of Biological Sciences, The University of Adelaide, Adelaide, South Australia 5005, Australia

[2] Proyecto Arqueológico Huaca Pucllana, Perú

[3] Centro de Investigaciones Arqueológicas del Museo de Sitio de Ancón, Perú

[4] Departamento de Letras y Ciencias Humanas de la Pontificia Universidad Católica del Perú, Lima, Perú

[5] Instituto de Alta Investigación (IAI), Laboratorio de Arqueología y Paleoambiente, Universidad de Tarapacá (UTA), Antofagasta, Arica, Chile.

[6] Centro de Investigaciones del Hombre en el Desierto (CIHDE)

[7] Department of Anthropology, University of California, Santa Cruz, CA 95064

[8] Historical Anthropology and Human Ecology, University of Goettingen, D-37073 Goettingen, Germany

ABSTRACT

Ancient DNA (aDNA) from past cultures and civilizations can provide essential information to reconstruct human population history. The advent of new technological DNA sequencing strategies (e.g. Next-Generation Sequencing) has revolutionised the scope of genetic studies by generating high-resolution genomic data.

These techniques have also substantially improved DNA-based phylogenies, such as the human mitochondrial tree. We applied mitochondrial DNA (mtDNA) capture by hybridization and high-throughput sequencing with the aim of exploring the maternal genetic diversity in populations over time. We generated 92 complete mtDNA genomes from pre-Columbian human remains in South America spanning from the Archaic to Late Horizon cultures. Samples were screened for mtDNA preservation in advance, and yielded the four Native American founder haplogroups (A2, B2, C1, and D1).

Samples grouped into archaeological periods define a temporal transition during the Archaic period (8000–1900 BC), followed by a genetic assimilation / homogenization process over the course of the four subsequent periods, involving the Early Intermediate, Middle Horizon, Late Intermediate and Late horizon (EIP – LH: 900 BC–1534 AD).

Our analyses of mtDNA genetic diversity reveal ancestral lineages with a substantial amount of "private" mutations not reported in the modern genetic pool of Native American populations, suggesting a considerable loss of maternal diversity in the Americas after the European arrival. This study represents the largest effort in obtaining ancient mtDNA complete genomes from pre-Columbian populations to date, and offers an unprecedented level of resolution to assess mitochondrial diversity in ancient South Americans.

KEYWORDS: Ancient DNA, mtDNA genomes, South America

1. INTRODUCTION

By inferring human demographic events from genetic (Schurr et al. 1990; Merriwether et al. 1995; Schurr and Sherry 2004; Tamm et al. 2007; Kitchen et al. 2008; Mulligan et al. 2008; Perego et al. 2010; Reich et al. 2012; Raghavan et al. 2015; Skoglund et al. 2015), anthropological, morphological (Neves and Hubbe 2005; Gonzalez-Jose et al. 2008; Chatters et al. 2014; Rasmussen et al. 2015a), linguistic (Greenberg et al. 1986; Gruhn 1988; Sicoli and Holton 2014) and paleo-ecological data (Hoffecker et al. 2014), researchers have proposed a number of hypotheses for the peopling of the Americas. The consensus model claims that the first human settlers arrived in the Late Pleistocene from Northeast Siberia through the Bering Strait, crossing a land bridge that was passable during the last glacial maximum (LGM) ~16,500 years before present (YBP).

While an Asian origin is largely undisputed these days, the debate about the exact entry routes and number of migratory waves is still controversial (Forster et al. 1996; Schurr and Sherry 2004; Dillehay et al. 2008; Goebel et al. 2008; O'Rourke and Raff 2010; Reich et al. 2012). Recent studies based on high-resolution ancient DNA (aDNA) data have reinforced the consensus scenario of human migration into the Americas. Rasmussen et al., analysed the complete nuclear genome from a 12,000 YBP male individual from Montana USA termed Anzick-1. This individual is associated with the Clovis culture, which was the most widespread and oldest archaeological complex in North America (Waters and Stafford 2007; Goebel et al. 2008). The analysis of the complete genome of Anzick-1 suggests that all Native Americans are descendants of the first people that successfully settled the Americas as this genome is equally closely related to all indigenous populations in the Americas and not closer to any other group outside (Rasmussen et al. 2014).

Interestingly, the genetic signature from this individual also shows a gene flow from a 24,000 year-old Upper Palaeolithic individual from Siberia known as “Mal’ta boy” (Raghavan et al. 2014b). The Mal’ta boy’s genome offered the first global picture of genetic composition in Native Americans, suggesting that 18-38% of its genome variation is present in the current genetic pool of indigenous Americans across the continent (Raghavan et al. 2014b).

A related study reported the complete mtDNA genome sequence from an 40,000-year-old early modern human from the Tianyuan cave in China, which was shown to be basal mtDNA haplogroup B, which is shared by present-day Asian and Native American populations alike (Fu et al. 2013). Both studies could link extant Native Americans with these ancestral groups in Asia, reinforcing the East Asian and Siberian origins of all indigenous peoples in the Americas.

The two most recent studies based on nuclear genome data from modern and ancient samples in the Americas have provided new insights into the complex peopling scenario of the continent by detecting of traces of DNA diversity from some populations from Australia and Melanesia, linked with genetic diversity of some Amazonians groups (i.e. Suruí and Karitiana) from South America (Raghavan et al. 2015; Skoglund et al. 2015)

In recent years, research has increasingly focused also on large genomic SNP data to explore genetic diversity at a broader population scale. This approach has been effectively applied to genetic research on modern populations (Wang et al. 2007; Reich et al. 2012; Moreno-Estrada et al. 2014), while the ultimate resolution can be achieved by sequencing complete nuclear genomes. The latter has been applied in selected ancient specimens in order to tackle the most outstanding question regarding human evolution (Green et al. 2008; Rasmussen et al. 2010; Reich et al. 2010; Rasmussen et al. 2011; Keller et al. 2012; Meyer et al. 2012; Prufer et al. 2014; Raghavan et al. 2014a; Raghavan et al. 2014b; Rasmussen et al. 2014). Two studies on prehistoric Europeans were able to show that these new techniques can be applied to larger numbers of ancient samples (Allentoft et al. 2015; Haak et al. 2015), with great promise for population scale genetics on prehistoric individuals from the Americas.

Population genetic studies have described a restricted mtDNA diversity in Native American groups, and explained it as a consequence of stochastic and demographic events during the original peopling of the continent (Tamm et al. 2007; Kitchen et al. 2008; Mulligan et al. 2008). Native American mtDNA is characterized by four haplogroups identified as (A2, B2, C1, D1) spread widely all over the continent (Schurr et al. 1990; Horai et al. 1993; Torroni et al. 1993; Forster et al. 1996) and haplogroup X, which is less common and restricted to some populations in North America (Brown et al. 1998; Malhi et al. 2001; Dornelles et al. 2005).

The assignment to a particular haplogroup is defined by diagnostic substitutions (mostly single nucleotide polymorphisms – SNPs) on the mtDNA genome (Torroni et al. 1993; Kivisild et al. 2002; Bandelt et al. 2003). The confirmation of mtDNA haplogroups and refinement at sub-haplogroup level is possible by analysing diagnostic SNPs in the control region (CR) and the coding region of the mtDNA genome (Schurr et al. 1990; Torroni et al. 1993; Forster et al. 1996; Herrnstadt et al. 2002; Bandelt et al. 2003; Haak et al. 2010; van Oven et al. 2011). A refinement of the analyses has identified the so-called Pan-American mtDNA haplogroups (A2, B2, C1b, C1c, C1d, C1d1, D1 and D4h3a) (Tamm et al. 2007; Perego et al. 2009; Perego et al. 2010).

Recent studies based on complete mtDNA genomes have expanded our knowledge about mtDNA variation in the Americas by identifying several new sub-lineages. The overall mtDNA diversity in Native Americans is now described by a total of 15 so-called American founder lineages (A2, A2a, A2b, B2, C1b, C1c, C1d, C1d1, C4, D1, D2a, D3, D4h3a, X2g, X2a) (Tamm et al. 2007; Achilli et al. 2008; Perego et al. 2010; Bodner et al. 2012; Achilli et al. 2013).

Moreover, complete mtDNA genome data from Native American extant populations were sequenced in order to infer demographic events and reconstruct the population history of indigenous populations (Tamm et al. 2007; Achilli et al. 2008; Fagundes et al. 2008; Perego et al. 2009; Perego et al. 2010; Bodner et al. 2012; Cardoso et al. 2012; de Saint Pierre et al. 2012; Achilli et al. 2013).

Although the large majority of mtDNA genomes were generated using Sanger sequencing technology, the availability of new sequencing technologies (i.e. Next-Generation Sequencing – NGS (Margulies et al. 2005; Bentley et al. 2008), coupled with DNA hybridization-based enrichment techniques (Patel and Sive 2001), has increased the amount of mtDNA genomes on a global scale. Besides, hybridization techniques to retrieve mtDNA genomes have also been applied successfully to ancient samples (Briggs et al. 2009; Burbano et al. 2010; Maricic et al. 2010; Brotherton et al. 2013).

Mitochondrial genetic diversity in ancient Native American populations has been explored intensively – for a review see (Raff et al. 2011) – but only based on small segments of the mtDNA genome. Most studies on mtDNA in human populations focused on sequencing the Hypervariable Regions (HVR-I and II of the control region), which was cost-effective to study mtDNA genetic diversity.

Another advantage is the availability of large and comprehensive datasets of modern populations for comparative analysis.

However, the full potential of mtDNA studies is only reached when complete mtDNA genomes are sequenced, both from modern and ancient samples. Today it has been possible to obtain complete mtDNA sequence data from approximately ~600 modern-day Native Americans – according to "Phylotree" (van Oven and Kayser 2009). However, most data comes from studies, which focussed on phylogenetic refinements rather than population based investigations. Therefore, ancient mtDNA genomes from Native Americans is very limited with only a small number of samples from North America (Gilbert et al. 2008; Cui et al. 2013; Rasmussen et al. 2014; Rasmussen et al. 2015b), and some samples from South America (Fehren-Schmitz et al. 2015).

Here, we present the largest dataset of complete ancient human mtDNA genomes from South America sequenced so far, aiming to investigate the genetic relationship among ancient groups in a time transect spanning the Archaic to the Late Horizon periods. This in-depth exploration of the mtDNA diversity will shed light on the haplogroup composition and genetic diversity in pre-Columbian populations before the European arrival, and will explore potential changes through time in order to reconstruct past human history in the continent.

2. MATERIAL AND METHODS

2.1. Selection of samples for genomic library preparation

Samples were processed in the dedicated aDNA facilities of the Australian Centre for Ancient DNA (ACAD) at the University of Adelaide, Australia (85 samples), at the University of California at Santa Cruz's (UCSC) Human Paleogenomic lab (15 samples), and the Reich lab at Harvard Medical School (HMS)(Table 1). The laboratories employ standardized aDNA protocols and infrastructure (Cooper and Poinar 2000; Fulton 2012; Knapp et al. 2012).

We selected a total of 100 individual samples for whole mtDNA genome sequencing (SI Table 1), and the criteria for inclusion were based on reliable/consistent genetic profiles obtained from the two independent samples assessed previously for aDNA analysis (see Chapter 2). As for the 85 samples analysed at ACAD, preservation and suitability for genomic library preparation were tested in advance by sequencing the Hypervariable Region (HVR-I), and genotyping with the GenoCoRe22 (Haak et al. 2010) and the AmericaPlex26 (Coutinho et al. 2014) coding region SNPs multiplex assays.

2. 2. Radiocarbon dating

Radiocarbon dating was performed at the Oxford Radiocarbon Unit for AMS Oxford, England in order to confirm the cultural classification of eight samples from the Huaca Pucllana archaeological site, Lima, Peru, (ACAD 10709, 10722, 10734, 10754, 10791, 10805, 10810, 10817), two samples from the Tiwanaku archaeological site from Bolivia (ACAD 13241, 13264), and one sample from Patagonia (Chile). Radiocarbon dates were calibrated using the OxCal Program v.4.2 from Oxford Radiocarbon Accelerator Unit, using the `IntCal13' dataset (Reimer et al. 2013). The two samples from Arroyo Seco, Argentina (B9S14 and B2S19) were radiocarbon dated as part of a previous study (Politis et al. 2014). Samples from Lauricocha (ACAD 12686, 12687, 12688, 126899) were sent to the Curt-Engelhorn-Center for Archaeometry – MAMS for AMS radiocarbon dating as part of a re-examination of the Lauricocha archaeological site (Fehren-Schmitz et al. 2015) (SI Table 6).

2.3. DNA extraction library preparation and mitochondrial capture (ACAD)

Samples (bone or tooth) were decontaminated upon entry to the aDNA laboratory by exposure to UV light, wiping with 3% sodium hypochlorite solution (bleach) and physical removal of sample surface contaminants by abrasion using a Dremel® drill. A Mikrodismembrator ball mill (Sartorius) was used to pulverise the sample, and 0.2 grams of bone powder were used subsequently in DNA extraction. Each aDNA extraction was performed in parallel on twelve samples plus two Extraction Blank Controls (EBCs). Samples were decalcified by incubation in 4 ml of 0.5 M EDTA (pH 8.0) overnight at 37°C on a rotor at ~30 rpm. Next, 70μL Proteinase K (Invitrogen) was added to the lysis mix and incubated for 2 hours at 55°C. DNA was isolated using an in solution silica method as described previously (Brotherton et al. 2013; Der Sarkissian et al. 2013). DNA was resuspended in 200 μL TE buffer supplemented with 0.05% Tween-20 (Sigma), and stored at -20ºC.

2.4. Genomic Library preparation for degraded DNA samples (ACAD)

Genomic libraries were prepared using a modified Illumina® protocol. DNA extracts were converted into sequencing library following the multiplex protocol for ancient DNA (Meyer and Kircher 2010; Briggs and Heyn 2012).

All genomic libraries and enzymatic clean-up steps were performed in a contamination-free ancient DNA laboratory. Libraries were prepared including an Extraction Blank Control (EBC) every 8 samples analysed.

a) Standard DNA blunt-end repair: Genomic libraries were prepared as follows: 20µL of DNA extract, 4µL 10x PB Buffer (or NEB2 - New England Biolabs), 0.4µL 10mM dNTPs (100µM each), 0.8µL RSA (rabbit serum albumin) 10mg/ml (Sigma), 4µL 10mATP (1µM), 2µL T4-PNK (Polynucleotide Kinase – New England Biolabs, 10U/µL) and 1.5µL T4 DNA Polymerase (New England Biolabs, 3U/µL) and 40µL ddH_20. The reaction was incubated at 25°C for 30 min. After cycling, 10µL of EDTA pH8.0 was added to inactivate the enzymes. DNA purification was performed using the MinElute spin column purification (Enzymatic reaction clean-up kit, Qiagen) according to manufacturer's instructions. DNA was eluted with 22.5µL 0.1x EB buffer + 0.05% Tween-20.

b) Adaptor ligation reaction: Sample-specific truncated Illumina adaptors were ligated to libraries in order to allow differentiation among samples/individuals for the subsequent multiplex library sequencing (SI Table 3). The ligation reaction was prepared as follows: 20µL of blunt-end repaired DNA, 4µL 10x T4 Ligase Buffer (Fermentas), 4µL PEG-4000 (50% solution in water), 0.25µL T4 DNA Ligase (30U/µL Fermentas), 1µL of truncated barcoded P5 adapter (25mM), 1µL of truncated barcoded P7 adapter (25mM), ddH_2O q.s.p 40µL. The reaction was incubated at 22°C for 1h. DNA was purified using MinElute spin columns (Enzymatic reaction clean-up kit, Qiagen) according to manufacturer's instructions. DNA was eluted with 22.5µL 0.1x EB buffer + 0.05% Tween-20.

c) Bst-DNA polymerase fill-in reaction: The reaction was prepared as follows: 20µL of ligation reaction eluate, 4µL 10x Thermopol Buffer (New England Biosciences), 0.3µL dNTP (25mM each), 1.5µL Bst DNA Polymerase (New England Biosciences, 8 U/µL), ddH_2O q.s.p 40µL. The reaction was incubated at 37°C for 30min followed by an increase to 80°C for 10min to denaturate the Bst Polymerase. DNA was eluted with 30µL 0.1x EB buffer + 0.05% Tween-20.

d) Library PCR first amplification: PCR reactions were performed in quintuplicate to reduce amplification bias and to increase the amount of available DNA for subsequent capture by hybridisation. Each PCR reaction was prepared as follows: 3µL of library DNA, 2.5µL 10x Gold Buffer (Applied Biosystems),

2.5µL MgCl2 (25mM), 2.5U AmpliTaq Gold (Applied Biosystems), 0.25µL of dNTPs (Invitrogen, 25mM each), 1.25µL IS7_short_amp.P5 (10µM), 1.25µL IS8_short_amp.P7 (10µM) (SI Table 5), ddH_2O q.s.p 25µL. The DNA amplification and all subsequent steps were performed in a standard molecular biology laboratory at the University of Adelaide. The thermocycling profile consisted of 94 °C for 6min, followed by 12 cycles of 30sec at 94°C, 30sec at 60°C and 1min at 72°C, followed by a final 10min at 72°C.

All PCR products for each library were pooled together for a final volume of 125µL. A clean-up step using AMPure® XP PCR Purification (magnetic beads system AGENCOURT; Beckman Coulter) was performed adding 198µL AMPure, followed by 3 washes with 80% ethanol. Final DNA libraries were eluted in 30µL EB buffer + 0.05% Tween-20.

e) Library PCR re-amplification: Libraries were then subject to a second PCR amplification round. The second amplification and the DNA purification were performed as for the first amplification above. Finally, the amplification products were visualized following an electrophoresis on a 3,5% agarose gel. DNA quantification was performed with a Nanodrop 2000 (Thermo Scientific).

2.5. DNA Extraction and library preparation (UCSC Human Paleogenomic Lab)

DNA extraction was based on the protocol from Dabney et al. 2013 (Dabney et al. 2013) using 75 mg of bone or tooth powder. DNA was eluted twice in 16-30 µL 1X TE buffer (with 0.05% Tween-20). Aliquots were stored at -18°C until further use. Double-stranded Illumina libraries were built from 30 µl of each DNA extract following the protocol from Rohland et al. 2015 using truncated Illumina adapters with dual 7-mer internal barcodes (Rohland et al. 2015). Depending on the libraries, DNA was either not repaired, subject to partial UDG treatment (*Uracil-DNA-Glycosylase* + endoVIII) (Rohland et al. 2015), or subject to USER repair (New England Biolabs). DNA libraries were then sent to the Reich lab at HMS for mitochondrial capture by hybridization and sequencing.

2.6. Mitochondrial DNA enrichment using biotinylated RNA probes (ACAD)

Mitochondrial RNA baits for DNA capture by hybridization were prepared in-house. In a first step, the whole mitochondrial genome (WMG) of BL (haplogroup H3k1) was amplified in three overlapping fragments using the Expand Long Range

dNTPack kit (Roche), and following the manufacturer's protocol with an annealing temperature of 60ºC and the primers. From this step onward, all commercial kits were used according to the manufacturer's protocols.

For each amplicon, one of the primers was 5'-tailed with a T7 promoter sequence (5'- AATTGTAATACGACTCACTATAGGG-3') in order to perform an *in vitro* transcription using the T7 High Yield RNA Synthesis Kit (New England BioLabs). DNA was degraded with DNase after the *in vitro* transcription was completed. Resulting RNA was purified using the Ambion MEGAclear kit (Life Technology), and eluted in 2 x 50 µL of RNase-free molecular grade water. RNA was quantified using a Nanodrop spectrophotometer (FisherScientific), and RNA integrity was assessed using a TapeStation (Agilent). For each mitochondrial fragment, 40 µg of RNA was then fragmented using the NEBNext Magnesium RNA Fragmentation Module (New England BioLabs). Fragmented RNA was purified using the RNeasy Minelute Cleanup kit (Qiagen), and eluted in 15 µL of RNase-free molecular grade water. RNA was quantified using a Nanodrop spectrophotometer, and RNA fragmentation was assessed using a TapeStation. Finally, fragmented RNA was randomly biotinylated using the EZ-Link Psoralen-PEG3-Biotin kit (Thermo Scientific), performing as many assays as was possible given the concentration of starting template. Biotinylated RNA was purified using the RNeasy Minelute Cleanup kit (Qiagen), and eluted in 15 µL of RNase-free molecular grade water. Biotinylated RNA baits from the initial three mitochondrial fragments were pooled at a concentration of 50 ng/µl each (150 ng total) for capture by hybridisation.

2.7. Hybridization capture for mtDNA (ACAD)

The overall process is the following:

a) Tracer concentration: Ancient genomic libraries after the 2nd amplification were concentrated by evaporation in order to reach 70ng/µL required as initial concentration. 3µL of DNA libraries were pooled per mtDNA bait.

b) Block mastermix set-up: Mastermix was prepared for a final volume of 5µL adding 2.5µL Cot 1 DNA (1µg/1µL) (Block #1), 2.5µL salmon sperm DNA (1µg/1µL) (Block #2). 5µL of Block mastermix was added to each tube containing 3µL of pooled DNA libraries.

c) RNA-mix preparation: Reaction conditions for RNA-mix were prepared for at total volume of 7.5µL. Tubes were prepared containing 5µL of mtDNA baits, 1µL RNase Block (SUPERase-In), 0.5µL (Proprietary blocking agent) Block #3 (P5/P7 RNA)(50µM of each), 0.5µL of DNA Block IS5 (50µM of each) and 0.5µL DNA Block IS6 (50µM of each)

d) Hybridization solution mastermix preparation: Mastermix was prepared for a final volume of 18.4µL. Final reaction comprised: 10µL 20xSSC (saline sodium citrate) (Hyb#1), 0.4µL 0.5M EDTA (Hyb#2), 4µL 50X Denhardt's (Hyb#3) and 4µL 1% SDS (Hyb#4).

e) Touchdown PCR: The thermocycling step was performed using a modification of temperature profile (Touchdown) as follows: Step 1: 94ºC – 5 min, Step 2: 65ºC – 3 min, Step 3: 65ºC -2 min, Step 4: 65ºC – 9 h, Step 5: 63ºC – 9 h, Step 6; 61ºC – 9 h, Step 7: 59ºC – 9 h, Step 8: 57ºC – 9 h, Step 9: 60ºC – 3 h. Mastermixes were combined and incubated for 48 hours. DNA library and Block mastermix (8µL) was incubated at the start of Step 1. RNA-mix (6.5µL) was incubated at the start of Step 2. Hybridization buffer was incubated at the start of Step 3. 145µL of hybridization solution was added to each RNA-mix at the end of Step 3. Finally, hybridization buffer plus RNA-mix were added to each DNA library.

f) Streptavidin beads preparation: My One^{TM} C1 beads (Invitrogen) magnetic beads in solution were vortexed at the start of each wash for 5sec. After each step, beads were pellet using a magnetic stand and then the supernatant was discarded. For each hybridisation reaction 50µL of magnetic beads were transferred to a 1.5 mL tube (Eppendorf). Beads were subjected to a series of washes, using a saline sodium citrate (SSC) 2X. Two washes were performed using 500µL 2X SSC + 0.05% Tween-20 at room temperature. Beads were incubated in 500µL 2X SSC + 0.05% Tween-20 adding 100µg yeast tRNA for 30 min on rotor at room temperature. Another wash step using 500µL 2X SSC + 0.05% Tween-20 was performed at room temperature. Finally beads were resuspended in 200 µL 2X SSC + 0.05% Tween-20.

g) Post capture bead wash: Beads were vortexed at the start of each wash for 5sec. After each step, beads were pellet using a magnetic stand and then the supernatant was discarded. Hybridization solution was transferred to the beads and incubated for 30 min on rotor at room temperature. This creates a DNA-

capture library-bead complex. Beads were immobilised to the magnet stand and subjected to successively increased-stringency washes, to remove progressively non-hybridised DNA molecules, using saline solution SSC and temperature variation.

First washing using 500µL of saline sodium citrate 2X SSC + 0.05% Tween-20 was performed for 10 min at room temperature. Second washing using 500µL 0.75X SSC + 0.05% Tween-20 was performed for 10 min at 60ºC. Third washing using 500µL 0.75X SSC + 0.05% Tween-20 was performed for 10 min at 60ºC. Final washing using 500µL 0.2X SSC + 0.05% Tween-20 was performed for 10 min at 60ºC.

h) PCR set-up with full-length Illumina adaptors: Beads were pellet and dried down using a magnetic stand after the last wash. PCR reaction using the full-length Illumina adaptors was performed with a modified step **off-bead PCR** (Fisher et al. 2011) adding the PCR mastermix onto the dried streptavidin beads. Sets of short Illumina adaptors (IS_4 Universal short Adaptor) were used in concordance with long Illumina Adaptors (SI Table 4), which in permutation generate a unique combination for every single sample. The tagging process is important for the subsequent Illumina bioinformatic analysis. PCR reaction with full-length Illumina adaptors was performed in quintuplicate per original sample. Final reaction conditions comprised of 2.5µL of 10x Gold Buffer, 2.5µL MgCl2 (25mM), 0.25µL AmpliTaq Gold 2.5U (Applied Biosystems), 0.625µL dNTPs (10mM each) (Invitrogen), 1.25µL IS4_indPCR_P5, 1.25µL of GAII_Indexing_1-15 and 16.62µL H_2O. Mastermix was transferred into tubes after post mtDNA capture containing the enriched library plus the biotinylated beads for a final reaction volume of 25µL. The thermocycling profile consisted of 94 °C for 6min, followed by 18 cycles of 30sec at 94°C, 30sec at 60°C, 45sec at 72°C, followed by a final 10min at 72°C.

i) AMPure® PCR clean-up: All PCR products were cleaned using the AGENCOURT® AMPure® XP PCR Purification (magnetic beads system). Since each PCR reaction was performed using the biotinylated beads, a first step before pooling of samples, involves the removing the resulting beads using a magnetic rack. Once all beads resultant from the amplification were removed, only the supernatant was transferred into the tubes.

j) Measure of concentration: All prepared DNA libraries were quantified using the Nanodrop 2000 (Thermo Scientific) according to manufacturer's instructions. For the final quality control, samples were assessed for fragment size distribution and DNA concentration after mtDNA Hybridization Capture using Tapestation 2200 Instrument (High sensitive DNA concentration measure screen gel) (Agilent Technologies) following manufacturer's instructions. Finally, DNA pooled samples for Illumina sequencing were diluted in order to reach 2nM for final concentration.

2.8. Bioinformatics / DNA sequencing and sequence assembly

Pooled DNA libraries were sequenced on two lanes of an Illumina HiSeq2000 machine (Illumina Inc., San Diego CA) at the Australian Cancer Research Foundation (ACRF) Cancer Genomics Facility, Adelaide, South Australia.

Post-sequencing processing was performed as follows:

Illumina reads were automatically demultiplexed by index using Casava and a strict matching of the 7-mer indexes. An additional demultiplexing based on a strict matching of the two 5-mer internal barcodes and trimming of the barcodes were performed using the program Sabre 1.0 (https://github.com/najoshi/sabre). AdapterRemoval v1.5.2 (Lindgreen 2012) was used with default parameters to trim residual adapter sequences, merge overlapping reads, and discard reads shorter than 25 nt. Merged reads were aligned to the Reconstructed Sapiens Reference Sequence (RSRS) (Behar et al. 2012) sequence with bwa 0.7.5a-r405, using the parameters space recommended for ancient DNA (no seed, one gap opening, relaxed edit distance). Duplicate reads were removed using FilterUniqueSAMCons.py and final pileup statistics were calculated using SAMtools. Characteristic damage patterns were assessed using MapDamage v0.3.6 (Ginolhac et al. 2011) (Figure 6a).

2.9. Sequencing analysis and SNP calling

DNA sequence pileups were visually inspected and edited using Geneious Pro® Software V.6 (Biomatters Ltd) (Drummond et al. 2011). Single Nucleotide Polymorphisms (SNPs) were called using Annotate & predict (Find Variation) function in Geneious Pro®, and exported into table format.

Haplotype determination was performed manually in reference to the online phylotree database (www.phylotree.org) mtDNA tree Build 16 [19 Feb 2014] (van Oven and Kayser 2009).

Ancient DNA is usually characterized by an increase in 5' C-to-T substitutions, and 3' G-to-A substitutions in the case of the library construction method used herein. In order to avoid wrong SNP calls due to damage, we established a threshold of 3x minimum coverage for every SNP and a Minimum Variant Frequency of 0.75. Global "diagnostic" and local "private" SNPs were checked visually and confirmed by two researchers (GV and BL) in order to generate final consensus sequences.

2.10. Population genetics and statistical analyses

Population differentiation and molecular diversity indices, haplotype (*h*), nucleotide diversity (π), genetic distances (pairwise F_{ST}), and linearized Slatkin's F_{ST} values were calculated to measure genetic relationships among populations using the software Arlequin 3.5 (Excoffier and Lischer 2010) (Table 3).

A Median Joining (MJ) network (Bandelt et al. 1999) was built to evaluate genetic relationships among all individuals subjected to analysis. Clustering analyses was performed using the program Network (www.fluxus-engineering.com). Network analyses were performed separately for the four Native American mtDNA haplogroups. We excluded substitutions at nucleotide positions 16,182 and 16,183, because they are dependent on the presence of C at the position 16,189 (Horai et al. 1993). We did not consider positions 309.1C (C), 315.1C, AC indels at 523 and 524 and deletion 3107 for mtDNA genomes (van Oven and Kayser 2009).

Principal Component Analysis (PCA) was performed for samples grouped into archaeological periods to visualize genetic relationship among ancient populations in a reduced number of dimensions called 'components' to detect internal structure (Figures 4a,b). Ward hierarchical clustering based on haplogroup frequencies was performed and plotted using a customized R script (Figure 5).

In order to explore population demographic structure, distribution of the pairwise genetic distances and mismatch distribution (Rogers and Harpending 1992) were calculated. Samples grouped into archaeological periods were used to compare to the expected distributions of populations that has recently grown in size (population expansion) or has been stationary over a long time. Simulated and observed mismatch distributions models were tested using sum of squared difference (SSD) and

Harpending's raggedness index (Harpending 1994). Also neutrality test for population expansion, Fu's Fs (Fu 1997) and Tajima's D (Tajima 1989) were calculated in Arlequin 3.5 (Excoffier and Lischer 2010)(Table 5).

3. RESULTS

3.1. Mitochondrial lineages

From a total of 100 samples, we obtained 92 complete mitochondrial genomes and identified 84 haplotypes. All successful analysed samples could be assigned to one of the 'founder' mtDNA Native American haplogroups (A2: 16.3%, B2: 29.3%, B2b: 11.9%, C1b: 23.9%, C1c: 5.5%, C1d: 2.17%, D: 2.17%, D1: 8.7%)(Table 2). Haplogroup X was not observed in our dataset. We have obtained partial mtDNA from the remaining samples, however due to aDNA damage and low DNA quality sequence coverage, it was not possible to make reliable mtDNA haplogroup calls.

3.2. Data sequence quality

Samples come from a broad temporal range and varied in preservation. Genomic libraries yielded variable degrees of complexity, genome coverage and per-base redundancy (coverage at a given position), and sequence quality of ancient samples correlated with the observed genomic coverage (SI Table 1). In order to ensure comparable data quality, all sequences used in the study had less than 1% missing data. Missing data (i.e. no coverage at a given position) were called 'N'.

Highest average coverage depth per-position was observed in samples coming from Pueblo Viejo (678.5x), whereas a sample from Llullaillaco showed the lowest average coverage (3.5x). Incidentally, both samples belong to the latest pre-Columbian archaeological period (Late Horizon 1476–1534 AD). Due to aDNA damage and low number of mapped Illumina reads, samples 6317–Llullaillaco, 11187–Miramar, 11202–Chancay, 10891–Patagonia, 10758–Wari and a contaminated sample from Mexico, 13991–Candelaria that yielded a sample mix-up contamination pattern after NGS, were removed from the final analysis since no reliable mtDNA genome consensus sequence were obtained. From the most ancient samples (Archaic period) the highest coverage was (21.3x) observed in samples from Camarones, Chile (Chinchorro culture) and the lowest coverage of (10.5x) was observed in a sample from La Galgada, Peru.

3.3. Haplogroup composition and haplotype diversity

In order to test temporal changes in haplogroup composition, samples were pooled according to archaeological periods in South America (Figure 2). Samples from Arroyo-Seco2 belonging to the Early Middle Archaic (6480–6450 BP) were grouped with samples from Late Archaic period (La Galgada, Camarones and Lauricocha individuals) (8000 BC–1900 BC). Haplogroup A has the highest frequency (64%) during this period, while haplogroup C is absent in the Archaic period. Haplogroups B and D both show (18%). Despite being able to analyse samples from the EH – Early Horizon period (900 BC–200 AD) in previous study (Chapter 2), no samples for this period were suitable for complete mtDNA genome sequencing.

For the four subsequent archaeological periods, EIP–Early Intermediate (200–600 AD) that includes samples from Lima and Palpa populations, MH–Middle Horizon period (600–1000 AD) represented by Wari and Tiwanaku populations, LIP–Late Intermediate Period (1000–1476 AD), including samples from Ychsma, Chancay, Pica8, Chullpa Botigiriayocc and LH–Late Horizon (1476–1534 AD) including samples from Llullaillaco, San Sebastian and Pueblo Viejo, haplogroup composition remains relatively constant over time without major fluctuations (Figures 2a, 2b).

Haplotype diversity (h) for each period showed high values among archaeological periods: Archaic period (1.0000), EIP (0.9905), MH (0.9942), LIP (0.9933) and LH (0.9905). The same holds true for nucleotide diversity for each archaeological period (π): Archaic period (0.001655), EIP (0.002157), MH (0.002530), LIP (0.002372), and LH (0.002230) (Table 5).

3.4. Genetic distances and population affinities

Genetic distances were calculated for samples pooled according to archaeological periods. High F_{ST} values were observed between the Archaic period and all four subsequent periods, EIP (0.19071), MH (0.14892), LIP (0.15529) and LH (0.18296) (Figure 3, Table 4a, 4b).

Principal Component Analysis (PCA) based on haplogroup frequencies (for major haplogroups and sub-haplogroups) (Tables 4a, 4b; Figures 4a, 4b) for each of the five cultural periods also revealed two main clusters. This differentiation is also supported by hierarchical Ward Clustering, which separated samples from the Archaic period and grouped samples for the successive four periods (Figure 5).

Neutrality tests and values for Tajima's D suggest a population expansion for the four grouped periods EIP, MH, LIP and LH (-1.60071, $p<0.05$). Likewise Fu's FS values indicate also population expansion for the four groups (-24.14606, $p<0.02$). Moreover, mismatch distributions show a multimodal curve which supports the scenario that expansion (demographic and spatial) cannot be that recent anymore for populations grouped into archaeological periods, EIP, MH, LIP and LH. Samples from Mexico were removed from this analysis, as they do not share the geographical region with all the South American samples. (Figures 7a-d; Table 5).

3.5. Phylogenetic analysis of mtDNA genomes

In order to understand how our ancient samples sit within the modern mtDNA phylogeny, we placed our ancient samples using the "Phylotree" mtDNA phylogeny database as scaffold (van Oven and Kayser 2009) build 16 [19 Feb 2014]). We followed the hierarchical structure of phylotree to embed the new ancient mtDNA sequences.
We used the diagnostic SNPs defining specific haplogroups and sub-haplogroups, and SNPs not (yet) reported in phylotree were considered as "private" mutations (SI Table 2). All ancient sequences fall into the Native American mtDNA diversity and represent unique branches within haplogroups A2, B2, C1, and D1 (Figures SI 1a, 2a, 3a, 4a).

3.6. Haplotype sharing from complete mtDNA genomes

After network analyses we observed haplotype sharing in individuals from the same cultural layer, such as two individuals from Pica-8 (11116-Pica/11121-Pica) belonging to hg B2, two individuals from Lima (10817-Lima/10820-Lima) belonging to hg C1b, two individuals from Ychsma (10722-Ychsma /10805-Ychsma) belonging to hg D1, and another two individuals from Wari (10734-Wari/10765-Wari) belonging to hg C1b. We also observed shared haplotypes between individuals from different cultural layers. For instance, two individuals from Lima-EIP and Ychsma-LIP (10811-Lima/10732-Ychsma) belonging to hg D, and three samples from Wari and Ychsma (10763-Wari/10709-Ychsma/10725-Ychsma) belonging to hg C1b. Lastly, we observed haplotype sharing between two individuals from the same archaeological period (Late Horizon) but from different cultures (11180-Pueblo Viejo/6320- Llullaillaco (SI Figures 2b, 3b, 4b).

4. DISCUSSION

Next-Generation Sequencing technology has enabled the generation of genomic data, which offer novel insights about human evolutionary processes around the globe. This thesis chapter evaluates information from complete mtDNA genomes in contrast to limited mtDNA sequence information (e.g. D-loop sequences) (Chapter 2).

Complete mtDNA sequences can reveal aspects in genetic diversity which may not be evident by analysing small portions of DNA (Kivisild 2015). Despite the fact that mtDNA research approach reflects only a partial "snapshot", in fact the maternal part of the evolutionary history of human populations, the study of high-resolution mtDNA genomes of pre-Columbian populations represents a major advance in understanding historical events in the Americas, such as population bottlenecks, population collapse and population admixture after the European arrival.

The present study has generated novel ancient mtDNA complete sequences from ancient Native American populations, offering the first overview of the pre-Columbian genetic diversity under a high mtDNA genome resolution. Our complete mtDNA genomes confirm the shallow phylogenetic structure of the Native American branches since the mtDNA sequences from our samples fall into the diversity of the sub-haplogroups (A2, B2, B2b, C1b, C1c, C1d and D1). However, the description of a substantially large number of novel haplotypes defined by "private" mutations is the most important observation in all of the reported mtDNA complete genomes (SI Table 1) and attests a unexpectedly large haplotype diversity in pre-Columbian times.

Of note, all mtDNA genomes generated in this study represent unique mtDNA lineages that are not present in the modern-day diversity in the Americas. This means that these lineages have either become extinct or have not yet been observed in modern-day populations, a bias that can only be explained by severe under-sampling of modern-day individuals in our regions of interest. Given the availability of a number of investigations on the genetic make-up of Native Americans based on modern populations (Tamm et al. 2007; Perego et al. 2009; Perego et al. 2010; Achilli et al. 2013), the present study suggests that there is still a significant amount of "hidden" diversity that in some way impairs a full comprehension of the complexity of the evolutionary history in the Americas.

It is possible that these "lost" mitochondrial lineages might be explained by a substantial loss of mtDNA diversity in modern populations. Under this scenario, it is hypothesized that Native Americans had a much greater mtDNA diversity prior to the European contact (O'Rourke et al. 2000), and that this reduction of diversity was therefore a consequence of a genetic bottleneck after the arrival of European settlers 500 years ago (Hunley and Healy 2011) and the subsequent population collapse of Native Americans (Livi-Bacci 2006).

Moreover, O'Fallon and Fehren-Schmitz 2011 could show that the effective female population size of Native Americans was reduced by 50%, which could have led to a stochastic extinction of ancient lineages and fixation of mtDNA in Native American after the European landfall. Another potential explanation suggests that mtDNA genomes from today's Native Americans despite current sequencing efforts remain drastically under-sampled. The possibility that new mtDNA sequences belonging to modern-day populations would consistently match the ancient haplotypes reported in this study is still open. However, this requires extensive and systematic sampling and in-depth sequencing of modern Native American populations to fill these gaps in knowledge to exclude this scenario.

Although mtDNA variation and genetic structure in the Native American populations seems to have been established several thousand years ago (Raff et al. 2011), the overall global "genetic snapshot" depicted in this study suggests a complex evolutionary picture in the South American regions in pre-Columbian times.

Our mtDNA haplogroup frequency based results (PCA and Ward clustering) and genetic distances (F_{ST}) suggest only one temporal transition between Archaic populations and the subsequent archaeological periods Early Intermediate, Middle Horizon, Late Intermediate and Late horizon period, which form a separate, big cluster (Figures 4a-b and 5). In concordance with a parallel study (Chapter 2), mtDNA genomic data also suggest a process of population "assimilation" or "homogenization" in ancient South America in time periods post-dating the Late Archaic period.

Haplotype sharing occurred mostly between individuals from the same culture and same archaeological period, and is less often between different cultural layers. Moreover, maternal relationship was observed interestingly in two samples belonging to the Inca culture from the Late Horizon period (Llullaillaco and Pueblo Viejo). Although they do not share the same geographic location, this correlates the large expansion of Inca Empire before the European arrival.

Furthermore, we did not evidenced haplotype sharing from our ancient mtDNA sequences with any modern mtDNA sequences, which reinforces the scenario that pre-Columbian populations were genetically highly more diverse. Further research with additional mtDNA and nuclear genomes from ancient and modern samples will be needed to fully explore demographic models, population bottlenecks and genetic structure or admixture models in South American populations to better understand population history of pre-Columbian civilizations.

Acknowledgments

We thank the Australian Research Council (ARC) Discovery Project "DP1095782" for funding granted. We thank Cristina Valdiosera and Gustavo Politis for providing archaeological samples for analysis.

REFERENCES

Achilli A, Perego U, and Bravi C. 2008. The Phylogeny of the Four Pan-American MtDNA Haplogroups: Implications for Evolutionary and Disease Studies. PLoS ONE 3(3):e1764.

Achilli A, Perego UA, Lancioni H, Olivieri A, Gandini F, Hooshiar Kashani B, Battaglia V, Grugni V, Angerhofer N, Rogers MP et al. . 2013. Reconciling migration models to the Americas with the variation of North American native mitogenomes. Proceedings of the National Academy of Sciences 110(35):14308-14313.

Allentoft ME, Sikora M, Sjögren K-G, Rasmussen S, Rasmussen M, Stenderup J, Damgaard PB, Schroeder H, Ahlström T, Vinner L et al. . 2015. Population genomics of Bronze Age Eurasia. Nature 522(7555):167-172.

Bandelt HJ, Forster P, and Rohl A. 1999. Median-joining networks for inferring intraspecific phylogenies. MolBiolEvol 16(1):37-48.

Bandelt HJ, Herrnstadt C, Yao YG, Kong QP, Kivisild T, Rengo C, Scozzari R, Richards M, Villems R, Macaulay V et al. . 2003. Identification of Native American founder mtDNAs through the analysis of complete mtDNA sequences: some caveats. AnnHumGenet 67(Pt 6):512-524.

Behar DM, van Oven M, Rosset S, Metspalu M, Loogvali EL, Silva NM, Kivisild T, Torroni A, and Villems R. 2012. A "Copernican" reassessment of the human mitochondrial DNA tree from its root. Am J Hum Genet 90(4):675-684.

Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, Hall KP, Evers DJ, Barnes CL, Bignell HR et al. . 2008. Accurate whole human genome sequencing using reversible terminator chemistry. Nature 456(7218):53-59.

Bodner M, Perego UA, Huber G, Fendt L, Rock AW, Zimmermann B, Olivieri A, Gomez-Carballa A, Lancioni H, Angerhofer N et al. . 2012. Rapid coastal spread of First Americans: novel insights from South America's Southern Cone mitochondrial genomes. Genome Research 22(5):811-820.

Briggs AW, Good JM, Green RE, Krause J, Maricic T, Stenzel U, Lalueza-Fox C, Rudan P, Brajkovic D, Kucan Z et al. . 2009. Targeted retrieval and analysis of five Neandertal mtDNA genomes. Science 325(5938):318-321.

Briggs AW, and Heyn P. 2012. Preparation of next-generation sequencing libraries from damaged DNA. Methods Mol Biol 840:143-154.

Brotherton P, Haak W, Templeton J, Brandt G, Soubrier J, Jane Adler C, Richards SM, Sarkissian CD, Ganslmeier R, Friederich S et al. . 2013. Neolithic mitochondrial haplogroup H genomes and the genetic origins of Europeans. Nature Communications 4:1764.

Brown MD, Hosseini SH, Torroni A, Bandelt HJ, Allen JC, Schurr TG, Scozzari R, Cruciani F, and Wallace DC. 1998. mtDNA haplogroup X: An ancient link between Europe/Western Asia and North America? AmJ HumGenet 63(6):1852-1861.

Burbano HA, Hodges E, Green RE, Briggs AW, Krause J, Meyer M, Good JM, Maricic T, Johnson PLF, Xuan Z et al. . 2010. Targeted Investigation of the Neandertal Genome by Array-Based Sequence Capture. Science 328(5979):723-725.

Cardoso S, Alfonso-Sanchez MA, Valverde L, Sanchez D, Zarrabeitia MT, Odriozola A, Martinez-Jarreta B, and de Pancorbo MM. 2012. Genetic uniqueness of the Waorani tribe from the Ecuadorian Amazon. Heredity 108(6):609-615.

Cooper A, and Poinar HN. 2000. Ancient DNA: Do It Right or Not at All. Science(289):1139.

Coutinho A, Valverde G, Fehren-Schmitz L, Cooper A, Barreto Romero MI, Espinoza IF, Llamas B, and Haak W. 2014. AmericaPlex26: a SNaPshot multiplex system for genotyping the main human mitochondrial founder lineages of the Americas. PLoS ONE 9(3):e93292.

Cui Y, Lindo J, Hughes CE, Johnson JW, Hernandez AG, and et al. 2013. Ancient DNA Analysis of Mid-Holocene Individuals from the Northwest Coast of North America Reveals Different Evolutionary Paths for Mitogenomes. PLoS ONE 8(7):e66948. doi:66910.61371/journal.pone.0066948.

Chatters JC, Kennett DJ, Asmerom Y, Kemp BM, Polyak V, Blank AN, Beddows PA, Reinhardt E, Arroyo-Cabrales J, Bolnick DA et al. . 2014. Late Pleistocene human skeleton and mtDNA link Paleoamericans and modern Native Americans. Science 344(6185):750-754.

Dabney J, Knapp M, Glocke I, Gansauge MT, Weihmann A, Nickel B, Valdiosera C, Garcia N, Paabo S, Arsuaga JL et al. . 2013. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. Proc Natl Acad Sci U S A 110(39):15758-15763.

de Saint Pierre M, Gandini F, Perego UA, Bodner M, Gomez-Carballa A, Corach D, Angerhofer N, Woodward SR, Semino O, Salas A et al. . 2012. Arrival of Paleo-Indians to the southern cone of South America: new clues from mitogenomes. PLoS ONE 7(12):e51311.

Der Sarkissian C, Balanovsky O, Brandt G, Khartanovich V, Buzhilova A, Koshel S, Zaporozhchenko V, Gronenborn D, Moiseyev V, Kolpakov E et al. . 2013. Ancient DNA reveals prehistoric gene-flow from siberia in the complex human population history of North East Europe. PLoS Genetics 9(2):e1003296.

Dillehay TD, Ramirez C, Pino M, Collins MB, Rossen J, and Pino-Navarro JD. 2008. Monte Verde: seaweed, food, medicine, and the peopling of South America. Science 320(5877):784-786.

Dornelles CL, Bonatto SL, De Freitas LB, and Salzano FM. 2005. Is haplogroup X present in extant South American Indians? Am J Phys Anthropol 127(4):439-448.

Drummond A, Ashton B, Buxton S, Cheung M, Cooper A, Duran C, Field M, Heled J, Kearse M, Markowitz S et al. . 2011. Geneious v5.4. Available from http://www.geneious.com/.

Excoffier L, and Lischer HE. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour 10(3):564-567.

Fagundes NJ, Kanitz R, Eckert R, Valls AC, Bogo MR, Salzano FM, Smith DG, Silva WA, Jr., Zago MA, Ribeiro-dos-Santos AK et al. . 2008. Mitochondrial population genomics supports a single pre-Clovis origin with a coastal route for the peopling of the Americas. Am J Hum Genet 82(3):583-592.

Fehren-Schmitz L, Llamas B, Lindauer S, Tomasto-Cagigao E, Kuzminsky S, Rohland N, Santos FR, Kaulicke P, Valverde G, Richards SM et al. . 2015. A Re-Appraisal of the Early Andean Human Remains from Lauricocha in Peru. PLoS ONE 10(6).

Fisher S, Barry A, Abreu J, Minie B, Nolan J, Delorey TM, Young G, Fennell TJ, Allen A, Ambrogio L et al. . 2011. A scalable, fully automated process for construction of sequence-ready human exome targeted capture libraries. Genome Biol 12(1):R1.

Forster P, Harding R, Torroni A, and Bandelt HJ. 1996. Origin and evolution of Native American mtDNA variation: a reappraisal. AmJHumGenet 59(4):935-945.

Fu Q, Meyer M, Gao X, Stenzel U, Burbano HA, Kelso J, and Paabo S. 2013. DNA analysis of an early modern human from Tianyuan Cave, China. Proc Natl Acad Sci U S A 110(6):2223-2227.

Fu Y-X. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics 147:915-925.

Fulton TL. 2012. Setting up an ancient DNA laboratory. Methods Mol Biol 840:1-11.

Gilbert MT, Kivisild T, Gronnow B, Andersen PK, Metspalu E, Reidla M, Tamm E, Axelsson E, Gotherstrom A, Campos PF et al. . 2008. Paleo-Eskimo mtDNA genome reveals matrilineal discontinuity in Greenland. Science 320(5884):1787-1789.

Ginolhac A, Rasmussen M, Gilbert MT, Willerslev E, and Orlando L. 2011. mapDamage: testing for damage patterns in ancient DNA sequences. Bioinformatics 27(15):2153-2155.

Goebel T, Waters MR, and O'Rourke DH. 2008. The late Pleistocene dispersal of modern humans in the Americas. Science 319(5869):1497-1502.

Gonzalez-Jose R, Bortolini MC, Santos FR, and Bonatto SL. 2008. The peopling of America: craniofacial shape variation on a continental scale and its interpretation from an interdisciplinary view. Am J Phys Anthropol 137(2):175-187.

Green RE, Malaspinas AS, Krause J, Briggs AW, Johnson PL, Uhler C, Meyer M, Good JM, Maricic T, Stenzel U et al. . 2008. A complete Neandertal mitochondrial genome sequence determined by high-throughput sequencing. Cell 134(3):416-426.

Greenberg J, Turner II C, Zegura S, and al. e. 1986. The Settlement of the Americas: A Comparison of the Linguistic, Dental, and Genetic Evidence. Current Anthropology , Vol. 27,(No. 5):pp. 477-497

Gruhn R. 1988. Linguistic Evidence in Support of the Coastal Route of Earliest Entry Into the New World. Man, New Series Vol. 23, No. 1:pp. 77-100.

Haak W, Balanovsky O, Sanchez JJ, Koshel S, Zaporozhchenko V, Adler CJ, Der Sarkissian CSI, Brandt G, Schwarz C, Nicklisch N et al. . 2010. Ancient DNA from European early neolithic farmers reveals their near eastern affinities. PLoS biology 8(11):e1000536.

Haak W, Lazaridis I, Patterson N, Rohland N, Mallick S, Llamas B, Brandt G, Nordenfelt S, Harney E, Stewardson K et al. . 2015. Massive migration from the steppe was a source for Indo-European languages in Europe. Nature.

Harpending HC. 1994. Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. Human Biology 66:591-600.

Herrnstadt C, Elson JL, Fahy E, Preston G, Turnbull DM, Anderson C, Ghosh SS, Olefsky JM, Beal MF, Davis RE et al. . 2002. Reduced-median-network analysis of complete mitochondrial DNA coding-region sequences for the major African, Asian, and European haplogroups. Am J Hum Genet 70(5):1152-1171.

Hoffecker JF, Elias SA, and O'Rourke DH. 2014. Anthropology. Out of Beringia? Science 343(6174):979-980.

Horai S, Kondo R, Nakagawahattori Y, Hayashi S, Sonoda S, and Tajima K. 1993. Peopling of the America, Founded by 4 Major Lineages of Mitochondrial-DNA. Molecular Biology and Evolution 10(1):23-47.

Hunley K, and Healy M. 2011. The impact of founder effects, gene flow, and European admixture on native American genetic diversity. Am J Phys Anthropol.

Keller A, Graefen A, Ball M, Matzas M, Boisguerin V, Maixner F, Leidinger P, Backes C, Khairat R, Forster M et al. . 2012. New insights into the Tyrolean Iceman's origin and phenotype as inferred by whole-genome sequencing. Nature Communications 3:698.

Kitchen A, Miyamoto M, and Mulligan C. 2008. A Three-Stage Colonization Model for the Peopling of the Americas. . PLoS ONE 3(2):(e1596. doi:10.1371/journal.pone.0001596).

Kivisild T. 2015. Maternal ancestry and population history from whole mitochondrial genomes. Investig Genet 6(1).

Kivisild T, Tolk HV, Parik J, Wang Y, Papiha SS, Bandelt HJ, and Villems R. 2002. The emerging limbs and twigs of the East Asian mtDNA tree. MolBiolEvol 19(10):1737-1751.

Knapp M, Clarke AC, Horsburgh KA, and Matisoo-Smith EA. 2012. Setting the stage - building and working in an ancient DNA laboratory. Ann Anat 194(1):3-6.

Lindgreen S. 2012. AdapterRemoval: easy cleaning of next-generation sequencing reads. BMC research notes 5(337).

Malhi RS, Schultz BA, and Smith DG. 2001. Distribution of mitochondrial DNA lineages among Native American tribes of Northeastern North America. HumBiol 73(1):17-55.

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z et al. . 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature 437(7057):376-380.

Maricic T, Whitten M, and Paabo S. 2010. Multiplexed DNA Sequence Capture of Mitochondrial Genomes Using PCR Products. PLoS ONE 5(11):e14004.

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

Meyer M, and Kircher M. 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harb Protoc 2010(6):pdb prot5448.

Meyer M, Kircher M, Gansauge T, Racimo F, Prüfer K, de Filippo C, Fu Q, Siebauer M, Stenzel U, and al. e. 2012. A high-coverage genome sequence from an archaic denisovan individual. Science 1224344:Published online August 30, 2012.

Moreno-Estrada A, Gignoux CR, Fernández-López JC, Zakharia F, Sikora M, Contreras AV, Acuña-Alonzo V, Sandoval K, Eng C, Romero-Hidalgo S et al. . 2014. The genetics of Mexico recapitulates Native American substructure and affects biomedical traits. Science 344(6189):1280-1285.

Mulligan C, Kitchen A, and Miyamoto M. 2008. Updated Three-Stage Model for the Peopling of the Americas. PLoS ONE 3(9):e3199. doi:3110.1371/ journal.pone.0003199.

Neves WA, and Hubbe M. 2005. Cranial morphology of early Americans from Lagoa Santa, Brazil: implications for the settlement of the New World. Proc Natl Acad Sci U S A 102(51):18309-18314.

O'Rourke DH, Hayes MG, and Carlyle SW. 2000. Spatial and temporal stability of mtDNA haplogroup frequencies in native North America. Hum Biol 72(1):15-34.

O'Rourke DH, and Raff JA. 2010. The human genetic history of the Americas: the final frontier. Curr Biol 20(4):R202-207.

Patel M, and Sive H. 2001. PCR-Based Subtractive cDNA Cloning. Current Protocols in Molecular Biology: John Wiley & Sons, Inc.

Perego U, Achilli A, Angerhofer N, Accetturo M, Pala M, Olivieri A, Hooshiar Kashani B, Ritchie K, Scozzari R, Kong Q et al. . 2009. Distinctive Paleo-Indian migration routes from Beringia marked by two rare mtDNA haplogroups. Curr Biol 19:1-8.

Perego UA, Angerhofer N, Pala M, Olivieri A, Lancioni H, Kashani BH, Carossa V, Ekins JE, Gomez-Carballa A, Huber G et al. . 2010. The initial peopling of the Americas: a growing number of founding mitochondrial genomes from Beringia. Genome Research 20(9):1174-1179.

Politis GG, Gutiérrez MA, and Scabuzzo C. 2014. Estado actual de las investigaciones en el sitio arqueológico Arroyo Seco 2 (partido de Tre Arroyos, provincia de Buenos Aires, Argentina) (INCUAPA, Olavarría, 2014), Serie monográfica.

Prufer K, Racimo F, Patterson N, Jay F, Sankararaman S, Sawyer S, Heinze A, Renaud G, Sudmant PH, de Filippo C et al. . 2014. The complete genome sequence of a Neanderthal from the Altai Mountains. Nature 505(7481):43-49.

Raff JA, Bolnick DA, Tackney J, and O'Rourke DH. 2011. Ancient DNA perspectives on American colonization and population history. Am J Phys Anthropol.

Raghavan M, DeGiorgio M, Albrechtsen A, Moltke I, Skoglund P, Korneliussen TS, Gronnow B, Appelt M, Gullov HC, Friesen TM et al. . 2014a. The genetic prehistory of the New World Arctic. Science 345(6200):1255832.

Raghavan M, Skoglund P, Graf KE, Metspalu M, Albrechtsen A, Moltke I, Rasmussen S, Stafford TW, Jr., Orlando L, Metspalu E et al. . 2014b. Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. Nature 505(7481):87-91.

Raghavan M, Steinrücken M, Harris K, Schiffels S, Rasmussen S, DeGiorgio M, Albrechtsen A, Valdiosera C, Ávila-Arcos MC, Malaspinas A-S et al. . 2015. Genomic evidence for the Pleistocene and recent population history of Native Americans. Science.

Rasmussen M, Anzick SL, Waters MR, Skoglund P, DeGiorgio M, Stafford TW, Jr., Rasmussen S, Moltke I, Albrechtsen A, Doyle SM et al. . 2014. The genome of a Late Pleistocene human from a Clovis burial site in western Montana. Nature 506(7487):225-229.

Rasmussen M, Guo X, Wang Y, Lohmueller KE, Rasmussen S, Albrechtsen A, Skotte L, Lindgreen S, Metspalu M, Jombart T et al. . 2011. An Aboriginal Australian Genome Reveals Separate Human Dispersals into Asia. Science.

Rasmussen M, Li Y, Lindgreen S, Pedersen JS, Albrechtsen A, Moltke I, Metspalu M, Metspalu E, Kivisild T, Gupta R et al. . 2010. Ancient human genome sequence of an extinct Palaeo-Eskimo. Nature 463(7282):757-762.

Rasmussen M, Sikora M, Albrechtsen A, Korneliussen TS, Moreno-Mayar JV, Poznik GD, Zollikofer CPE, Ponce de Leon MS, Allentoft ME, Moltke I et al. . 2015a. The ancestry and affiliations of Kennewick Man. Nature 523(7561):455-458.

Rasmussen M, Sikora M, Albrechtsen A, Korneliussen TS, Moreno-Mayar JV, Poznik GD, Zollikofer CPE, Ponce de León MS, Allentoft ME, Moltke I et al. . 2015b. The ancestry and affiliations of Kennewick Man. Nature.

Reich D, Green RE, Kircher M, Krause J, Patterson N, Durand EY, Viola B, Briggs AW, Stenzel U, Johnson PLF et al. . 2010. Genetic history of an archaic hominin group from Denisova Cave in Siberia. Nature 468(7327):1053-1060.

Reich D, Patterson N, Campbell D, Tandon A, Mazieres S, Ray N, Parra MV, Rojas W, Duque C, Mesa N et al. . 2012. Reconstructing Native American population history. Nature 488(7411):370-374.

Reimer PJ, Bard E, Bayliss A, Beck JW, Blackwell PG, Bronk Ramsey C, Buck CE, Cheng H, Edwards RL, Friedrich M et al. . 2013. IntCal13 and Marine13 Radiocarbon Age Calibration Curves 0–50,000 Years cal BP.

Rogers AR, and Harpending H. 1992. Population growth makes waves in the distribution of pairwise genetic differences. MolBiolEvol 9(3):552-569.

Rohland N, Harney E, Mallick S, Nordenfelt S, and Reich D. 2015. Partial uracil–DNA–glycosylase treatment for screening of ancient DNA.

Schurr TG, Ballinger SW, Gan YY, Hodge JA, Merriwether DA, Lawrence DN, Knowler WC, Weiss KM, and Wallace DC. 1990. Amerindian mitochondrial DNAs have rare Asian mutations at high frequencies, suggesting they derived from four primary maternal lineages. Am J Hum Genet 46(3):613-623.

Schurr TG, and Sherry ST. 2004. Mitochondrial DNA and Y chromosome diversity and the peopling of the Americas: evolutionary and demographic evidence. Am J Hum Biol 16(4):420-439.

Sicoli MA, and Holton G. 2014. Linguistic Phylogenies Support Back-Migration from Beringia to Asia. PLoS ONE 9(3):e91722. doi:91710.91371/journal.pone.0091722.

Skoglund P, Mallick S, Bortolini MC, Chennagiri N, Hunemeier T, Petzl-Erler ML, Salzano FM, Patterson N, and Reich D. 2015. Genetic evidence for two founding populations of the Americas. Nature advance online publication.

Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123(3):585–595.

Tamm E, Kivisild T, Reidla M, Metspalu M, and Smith D. 2007. Beringian standstill and spread of Native American founders. PLoS ONE 2(9):e829.

Torroni A, Schurr T, Cabell M, Brown M, Neel J, Larsen M, Smith D, Vullo C, and Wallace D. 1993. Asian affinities and continental radiation of the four founding Native American mtDNAs. Am J Hum Genet 53:563–590.

van Oven M, and Kayser M. 2009. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Hum Mutat 30(2):E386-394.

van Oven M, Vermeulen M, and Kayser M. 2011. Multiplex genotyping system for efficient inference of matrilineal genetic ancestry with continental resolution. Investig Genet 2:6.

Wang S, Lewis C, and Jakobsson M. 2007. Genetic variation and population structure in Native Americans. PLoS Genetics 3:e185.

Waters MR, and Stafford TW, Jr. 2007. Redefining the age of Clovis: implications for the peopling of the Americas. Science 315(5815):1122-1126.

List of Figures and Tables

Figure 1. Map showing sites from which ancient human remains sampled for this study:

Lima Culture (1), Wari Culture (2), Ychsma Culture (3), Pasamayo-Chancay (4), Jauranga-Palpa (5), Pueblo Viejo-Inca (6), Lauricocha (7), Chullpa Botiriayocc (8), La Galgada (9), San Sebastián (10), Los Molinos (11), Huaca Prieta (12), Miramar-Ancon (13); Tiwanaku (14); Pica8-Tarapaca (15), Camarones (16); Llullaillaco (17), Patagonia (18), Arroyo Seco (19), Cueva Candelaria (20).

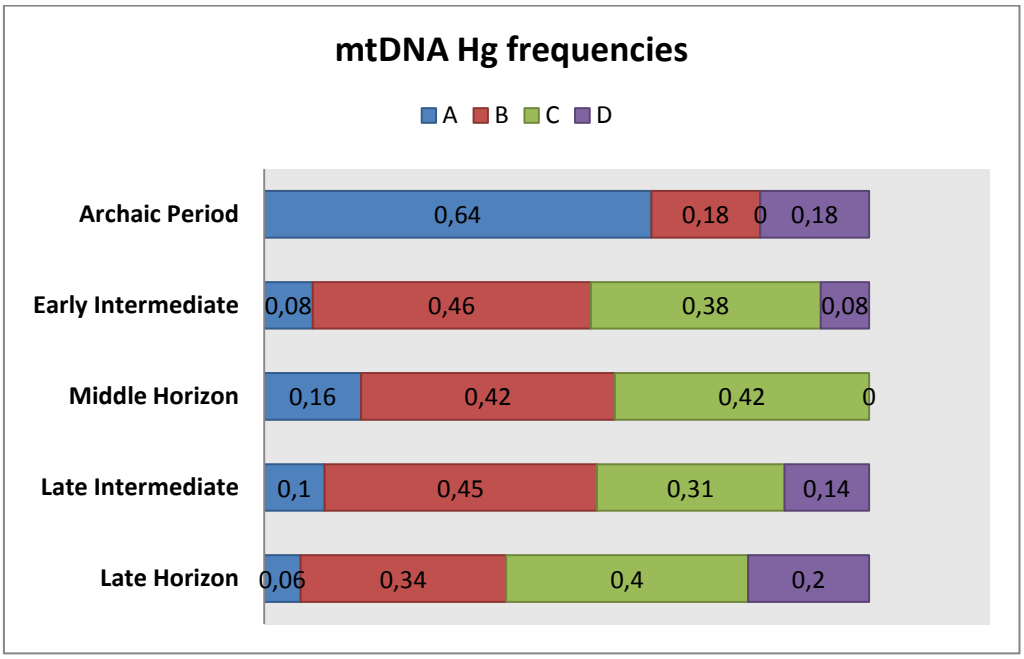


Figure 2a. Variation of mitochondrial DNA frequencies through time from five distinct cultural periods in South America excluding samples from Mexico (Major haplogroups).

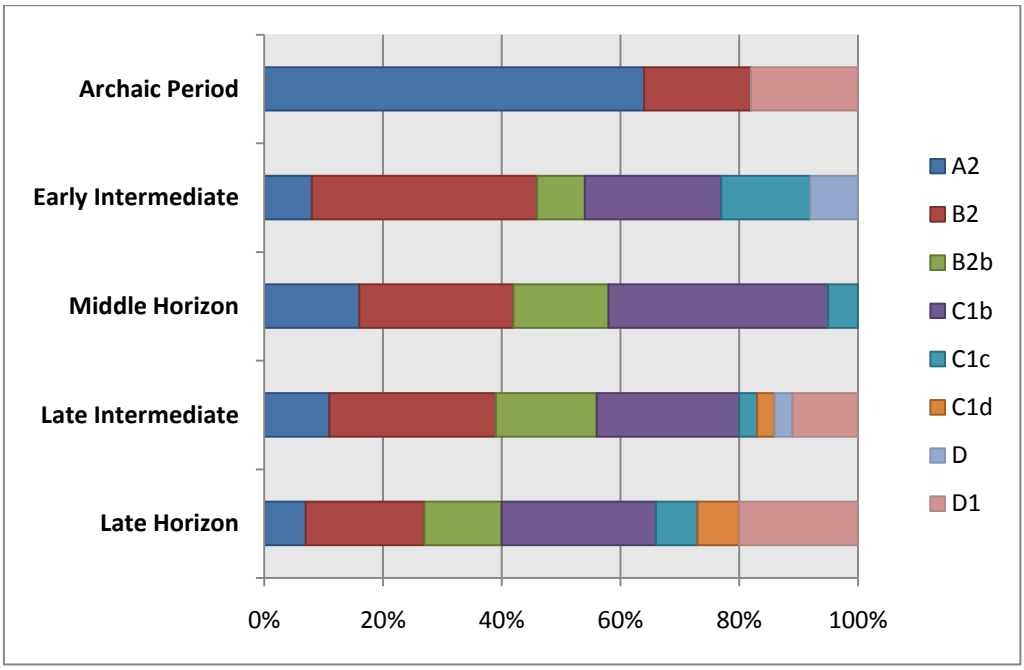


Figure 2b. Variation of mitochondrial DNA frequencies through time from five distinct cultural periods in South America, excluding samples from Mexico (Sub-haplogroups)

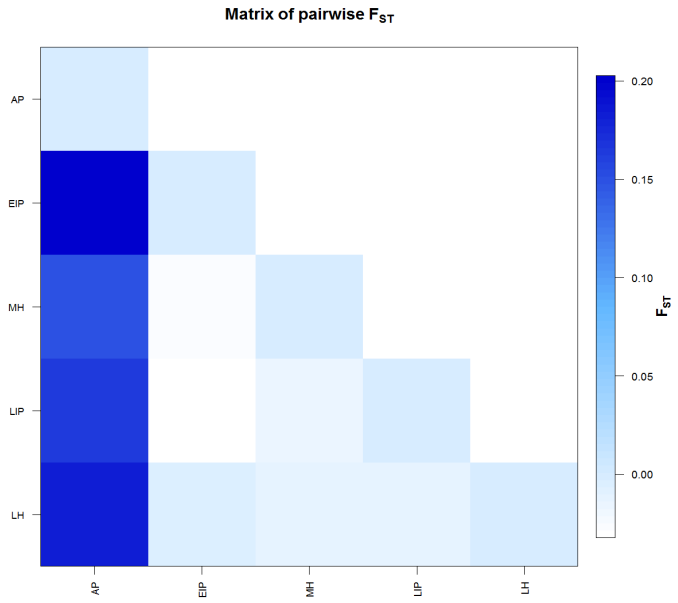


Figure 3. Genetic distances (pairwise (F_{ST}) between archaeological periods created with Arlequin v3.5. AP: Archaic Period; EIP: Early Intermediate Period; MH: Middle Horizon; LIP: Late Intermediate Period; LH: Late Horizon. Increasing blue intensity indicates larger genetic distance.

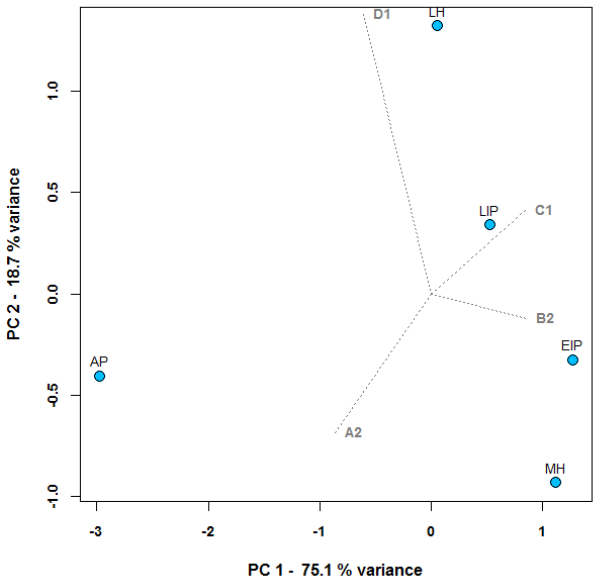


Figure 4a. Principal component analysis (PCA) based on major mtDNA (A2, B2, C1, D1) haplogroup frequencies according to five cultural periods. AP: Archaic Period; EIP: Early Intermediate Period; MH: Middle Horizon; LIP: Late Intermediate Period; LH: Late Horizon. In sum, PCA axes 1 and 2 describe 93.8 % of the total variance.

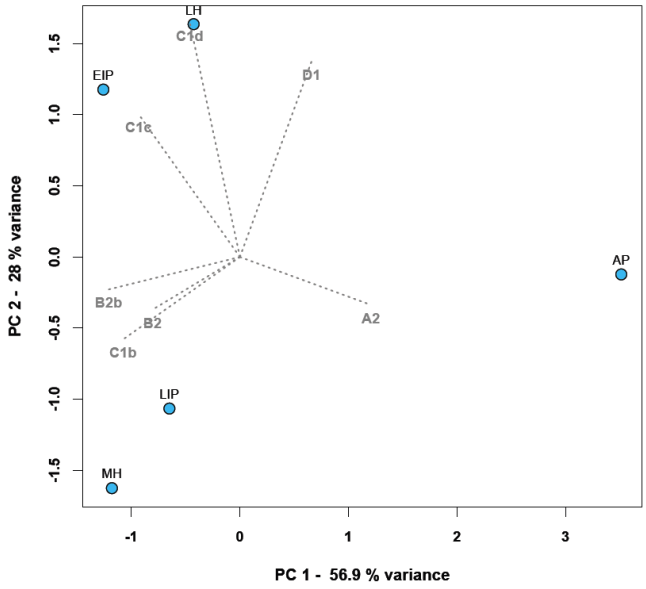


Figure 4b. Principal component analysis (PCA) based on mtDNA (A2, B2, B2b, C1b, C1c, C1d, D1) sub-haplogroup frequencies according to cultural periods. AP: Archaic Period; EIP: Early Intermediate Period; MH: Middle Horizon; LIP: Late Intermediate Period; LH: Late Horizon. In sum, PCA axes 1 and 2 describe 84.9 % of the total variance.

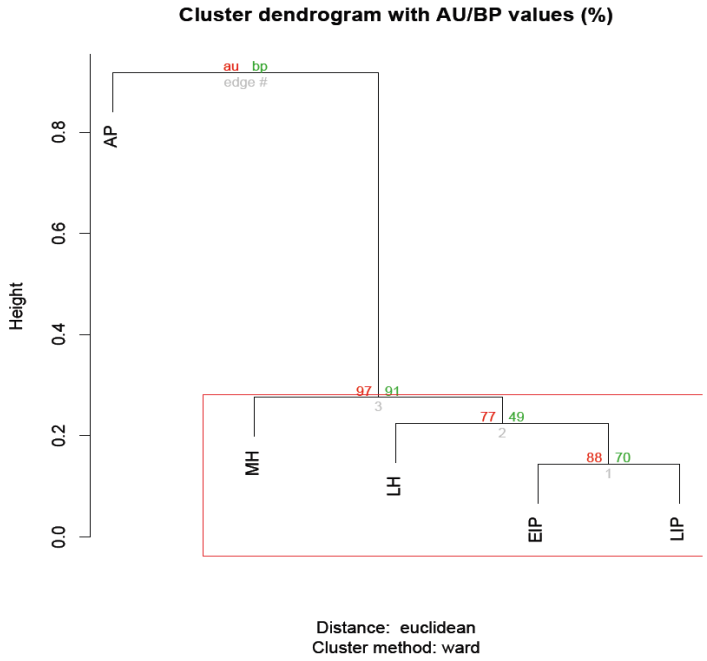


Figure 5. Ward clustering dendrogram of the five archaeological groups in South America. AP: Archaic Period; EIP: Early Intermediate Period; MH: Middle Horizon; LIP: Late Intermediate Period; LH: Late Horizon. Bootstrapping values (red/green) support two main clusters.

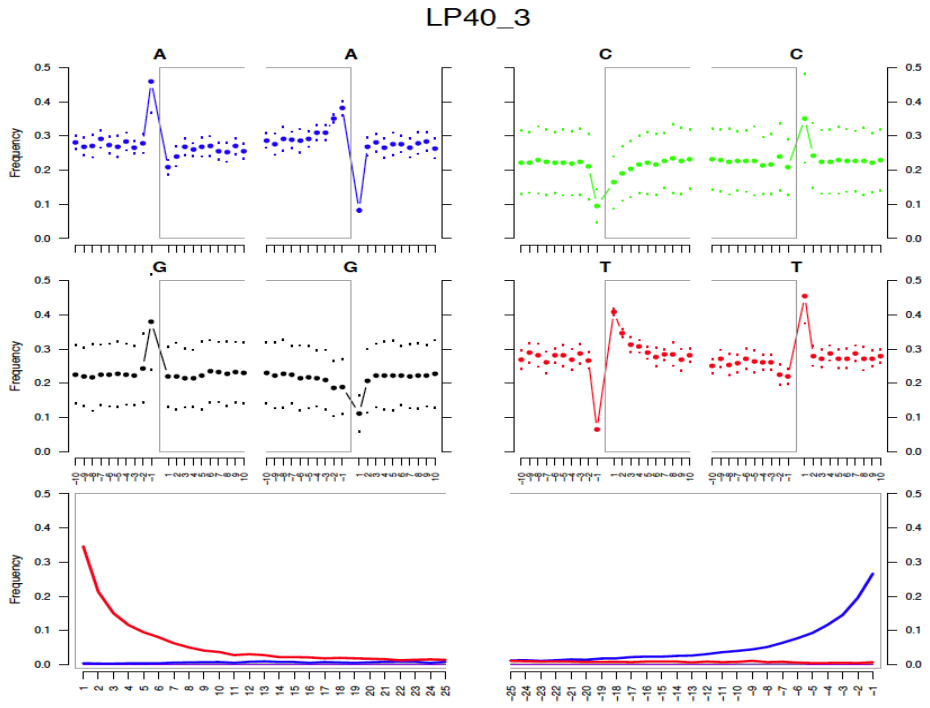


Figure 6a. Example of aDNA damage patterns (sample LP40.3/ACAD 10765-Wari).
Four upper plots: Frequencies of the A, C, G and T bases according to the nucleotide positions within the read (within the grey box) and outside the read (outside the grey box).
Two lower plots: Frequency distribution of specific substitutions from the 25-end (left) (C to T misincorporation) to the 25-end (right) (G to A misincorporation) of the read sequence.

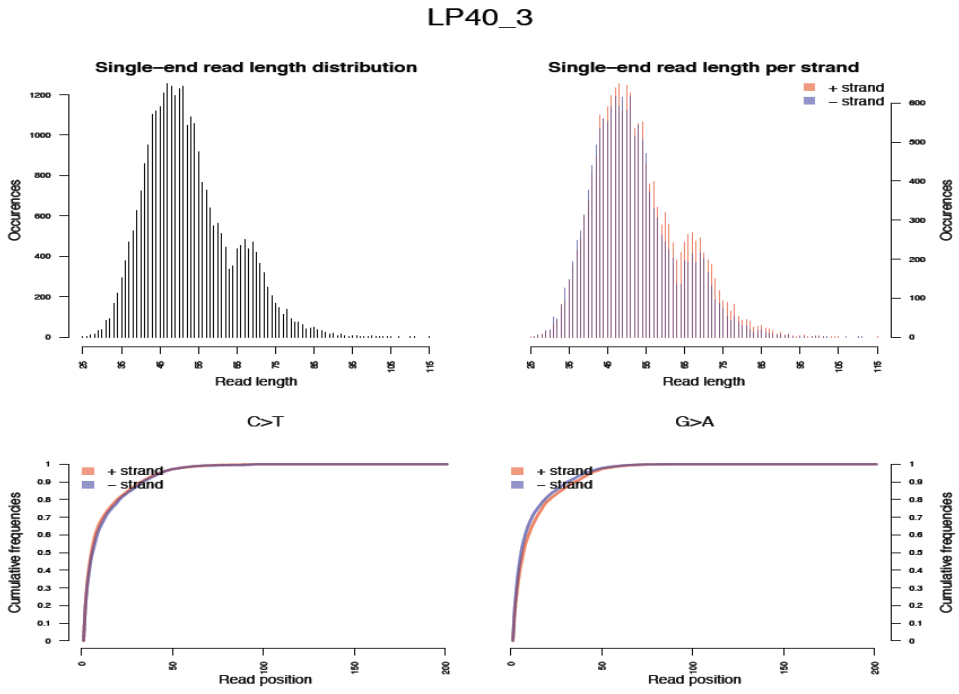


Figure 6b. Two upper plots: Example of read length distribution of sample ACAD 10765-Wari.
Two lower plots: Observed cumulative frequency of C to T and G to A misincorporations.

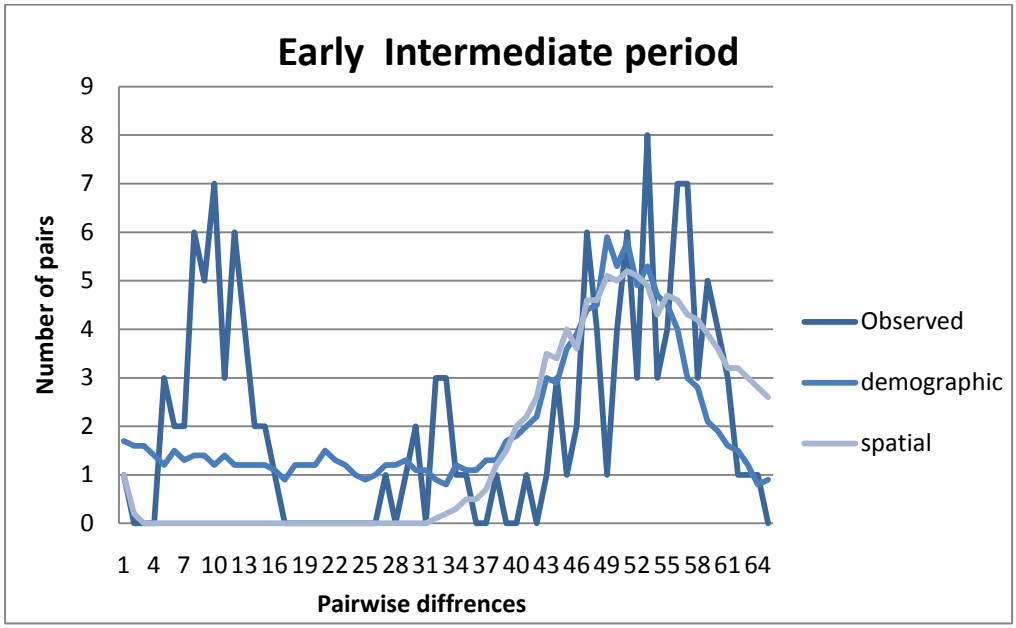


Figure 7a. Demographic and spatial expansion. Mismatch distribution patterns for the Early Intermediate period.

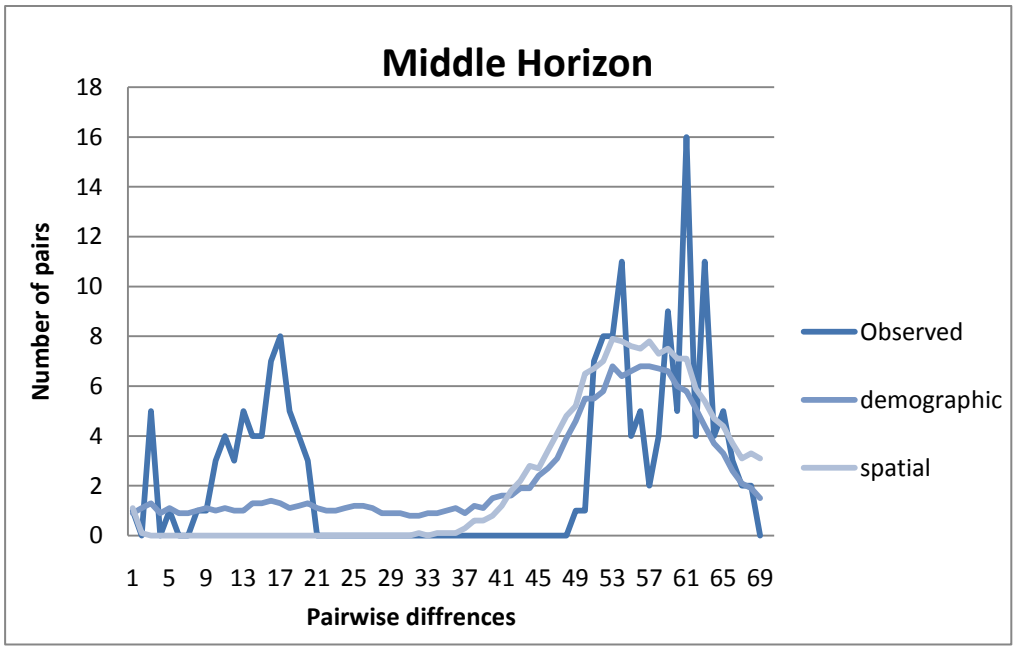


Figure 7b. Demographic and spatial expansion. Mismatch distribution patterns for the Middle Horizon period.

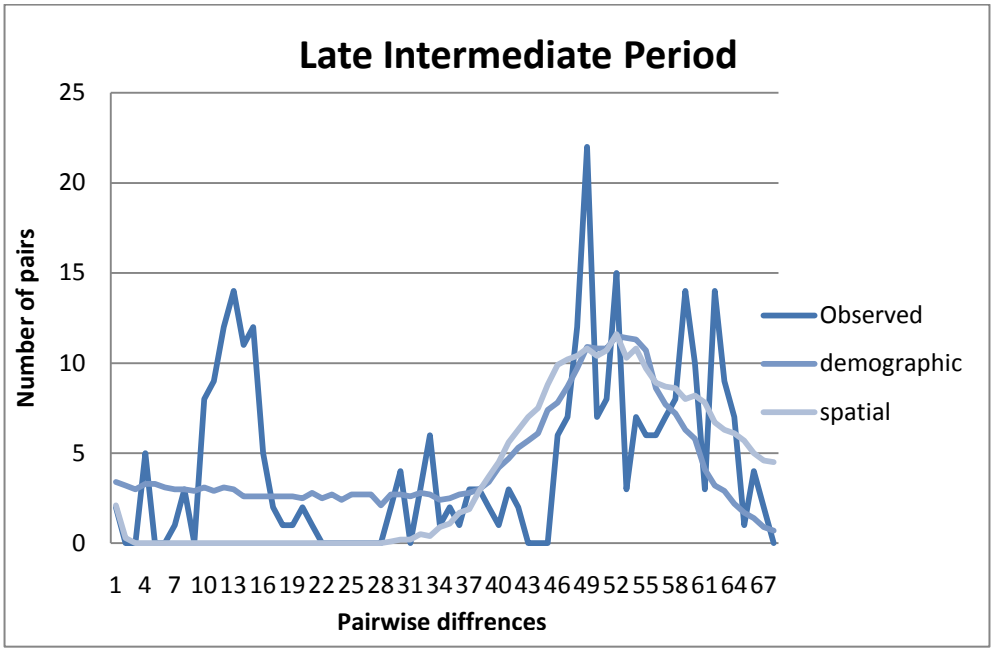


Figure 7c. Demographic and spatial expansion. Mismatch distribution patterns for the Late Intermediate period.

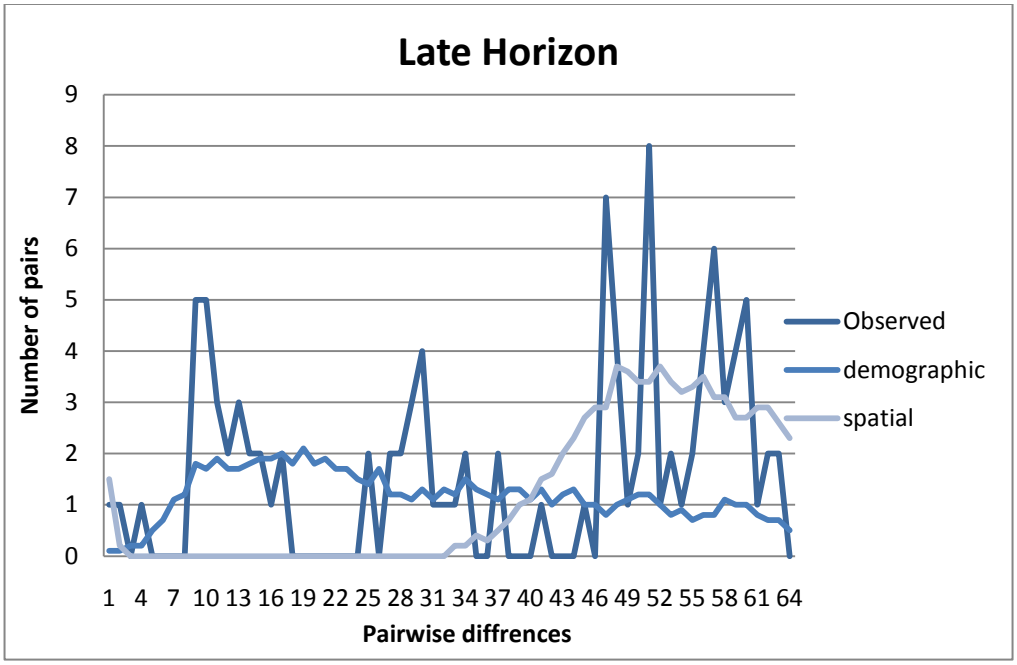


Figure 7d. Demographic and spatial expansion. Mismatch distribution patterns for the Late Horizon period.

Table 1. List of samples and sites examined in this study

Site Name	Location	Coordinates	Age Range	Description/Notes	Culture (Samples)
Huaca Pucllana	Lima, Peru	12°06'37.01''S - 77°01'58.93"W	200 – 1475 AD	Three successive cultural periods from Huaca Pucllana: a first occupation by the Lima in the Early Intermediate Period (200 - 600 AD), a second settlement by the Wari Culture in the Middle Horizon (600 - 1000 AD) and finally occupation by the Ychsma in the Late Intermediate Period (1000 - 1476 AD)	Lima (11) Wari (12) Ychsma (15)
Pasamayo, Chancay	Lima, Peru		1000 – 1476 AD	Samples collected at the Museum Ancon in Lima Peru.	Chancay (8)
Jauranga, Palpa	Peru	14°32'01.84"S - 75°11'07.07"W	1700 – 1600 BP	The samples derive from the Early Intermediate Period settlement phase of the rural settlement Jauranga, associated with the Nasca Culture (Nasca 5, ~300-400 AD).	Palpa: (3)
PuebloViejo, Pucara	Peru	12°12'16.15"S 76°47'57.23"W	1476 – 1534 AD	Pueblo Viejo-Pucara, is an Inca site located on the left bank of the Lurin river that was occupied in Late Horizon period.	Inca: (11)
Lauricocha	Peru	10°19'10.81"S-76°43'43.74"W	6800 – 1500 BC	Lauricocha is a rock shelter in the Central Peruvian Highlands at approx. 4000m elevation. The site has been continuously used for accommodation and burials by humans from the Early to the Late Holocene. Samples derive from individuals found in three burial phases falling into the Early Archaic, ~6800-6300 BC; Late Archaic, ~3500 BC, and Initial Period, ~1800 BC (ages correlate with the burial events, not the ranges of the archaeological periods)	Lauricocha: (4)
Chullpa Botiriayocc	Peru	14°15'20.20"S 74°51'33.78"W	800 BP	Chullpa Botiriayocc is associated with a Late Intermediate settlement close to the modern town of Laramate in the South-Central Peruvian Highlands (Western Slope) above the Palpa Region.	Chullpa: (2)
La Galgada	Peru	8°27'59.60"W 78°09'00.07"S	4000 BP	Late Archaic settlement site and ritual centre in the upper valleys of the Northern Peruvian Highlands. Individuals derive from several burial situations inside the ritual structure of La Galgada.	La Galgada (3)
San Sebastian	Peru		600 – 500 BP	Late Horizon – Inca rural settlement in the Rio Urubamba valley 10 miles from Machu Picchu, excavated by Bingham in 1918	San Sebastian (2)
Los Molinos	Peru	14°32'01.84"S – 75°11'07.07"W	2000 – 1800 BP	Early Intermediate urban settlement associated with the Nasca Culture (Regional Center in the Palpa Valley). Los Molinos dates to the Early Nasca Period (~0-200AD)	Los Molinos (1)
Huaca Prieta	Peru	7°55'11.06"S 79°18'30.75"W	1800 – 1000 BP	Huaca Prieta is a large and complex stone and earth platform mound located on the southern point of a Pleistocene terrace on the northern coastal desert in Peru, approximately 500 kilometres (310 miles) northwest of Lima.	Huaca Prieta (1)

Site Name	Location	Coordinates	Age Range	Description/Notes	Culture (Samples)
Miramar Necropolis	Lima, Peru		500 – 1000 AD	Samples collected at the Museum Ancon in Lima Peru. Samples belong to the Wari culture from Peru.	Wari (1)
Tiwanaku	La Paz, Bolivia	16°33'20.11"S - 68°40'24.04"W	724 – 1187 AD	Tiwanaku occupation during Epoch / Period V (724 – 1187AD). Samples were collected in association with the 'Akapana Project 2009' led by the 'Unidad de Arqueologia y Museos - UDAM (Bolivian Archaeology Unit) in collaboration with the Ministerio de Culturas (Bolivian Ministry of Cultures) in the proximities of the Akapana Pyramid at the Tiwanaku archaeological site.	Tiwanaku (10)
Tarapacá	Chile	20°29'20.45"S - 69°19'44.47"W	1000 – 500 BP	The Cultural Complex Pica-Tarapacá has developed during the Late Intermediate Period. With evidence primarily of ceramics and textiles and it has been defined largely by the remains recovered of the cemetery Pica-8.	Pica8: (6)
Camarones 15	Chile	20°29'20.45"S - 69°19'44.47"W	4000 – 3500 BP	Camarones 15 is a Late Archaic settlement and burial site of the Chinchorro culture	Chinchorro (1)
Llullaillaco	Argentina	24°43'13.13"S - 68°31'46.45"W	1476 – 1534AD	Llullaillaco is the world's highest active volcano in the border between Chile and Argentina and belongs to the Puna de Atacama, a region with volcanic peaks on a high plateau within the Atacama Desert. Samples belong to Three mummies: The Lighting girl, The Doncella/Maiden and the Boy/El Niño.	Inca: (3)
Patagonia	Chile	51°34'23.9"S - 72°35'09.7"W	2158 – 2268 calBC	Human sample collected at Cueva Chica in the Patagonia region. The site is associated with Late Pleistocene fauna. Radiocarbon dates produced some of the oldest evidence for the presence of remains of ancient mammals and other species.	Patagonia: (1)
Arroyo Seco 2	Argentina	33°09'59.21"S 60°29'56.15"W	6500+/- 60 BP	Arroyo Seco 2 is an open-air forager camp site in the Argentinean Pampas with a significant Paleo-Indian component. The camp site has been occupied over several thousand years from the Middle Archaic period on. Samples derive from the early phases of this site.	Arroyo Seco (2)
Cueva Candelaria	Mexico	5° 25′ 16″ N, 100° 57′ 40″ W	1000 –1600 AD	Cueva de la Candelaria (Candelaria Cave) is an archaeological site located in the state of Coahuila in northern Mexico. The Candelaria cave is an exceptional site due to conservation state of its organic remains and offers a particularly rich representation of funerary practices.	Candelaria (6)

Table 2. Haplogroup determination of 92 mtDNA genomes analysed in this study

#	ACAD #	Populations/Culture	Location	Radiocarbon dates	Haplogroup
1	12686	Lauricocha Lau1	Peru	**8589 - 8482 calBP**	**A2**
2	12687	Lauricocha Lau3	Peru	**3631 - 3513 calBP**	**A2**
3	12688	Lauricocha Lau4	Peru	**5936 - 5905 calBP**	**A2**
4	12689	Lauricocha Lau5	Peru	**8700 - 8599 calBP**	**B2**
5	13247	Tiwanaku	Bolivia		**A2**
6	13237	Tiwanaku	Bolivia		**B2**
7	13264*	Tiwanaku	Bolivia	**1017 - 1155 calAD**	**B2**
8	13272	Tiwanaku	Bolivia		**B2**
9	13240	Tiwanaku	Bolivia		**B2**
10	13275	Tiwanaku	Bolivia		**B2**
11	13241*	Tiwanaku	Bolivia	**901 - 1024 calAD**	**C1c**
12	13279	Tiwanaku	Bolivia		**C1b**
13	13245	Tiwanaku	Bolivia		**C**
14	11116	Pica8-Tarapaca	Chile		**B2**
15	11118	Pica8-Tarapaca	Chile		**B2**
16	11121	Pica8-Tarapaca	Chile		**B2**
17	11200	Chancay	Peru		**B2**
18	11211	Chancay	Peru		**C1b**
19	11205	Chancay	Peru		**C1b**
20	11209	Chancay	Peru		**C1c**
21	11213	Chancay	Peru		**D1**
22	11197	Chancay	Peru		**A2**
23	11208	Chancay	Peru		**B2b**
24	6322	Llullaillaco	Argentina		**C1b**
25	6320	Llullaillaco	Argentina		**D1**
26	10791*	Lima	Peru	**584-660 calAD**	**A2**
27	10814	Lima	Peru		**B2**
28	10802	Lima	Peru		**B2**
29	10789	Lima	Peru		**B2b**
30	10817*	Lima	Peru	**534-642 calAD**	**C1b**
31	10820	Lima	Peru		**C1b**
32	10806	Lima	Peru		**C1c**
33	10811	Lima	Peru		**D***
34	10821	Lima	Peru		**B2**
35	10774	Wari	Peru		**A2**
36	10754*	Wari	Peru	**974-1220 calAD**	**B2b**
37	10734*	Wari	Peru	**776-968 calAD**	**C1b**
38	10763	Wari	Peru		**C1b**
39	10771	Wari	Peru		**C1b**
40	10773	Wari	Peru		**C1b**
41	10778	Wari	Peru		**A2**

#	ACAD #	Populations/Culture	Location	Radiocarbon dates	Haplogroup
42	10742	Wari	Peru		**B2b**
43	10750	Wari	Peru		**B2b**
44	10765	Wari	Peru		**C1b**
45	10713	Ychsma	Peru		**B2**
46	10729	Ychsma	Peru		**B2**
47	10720	Ychsma	Peru		**B2b**
48	10726	Ychsma	Peru		**B2b**
49	10709*	Ychsma	Peru	**1244-1288 calAD**	**C1b**
50	10717	Ychsma	Peru		**C1b**
51	10725	Ychsma	Peru		**C1b**
52	10731	Ychsma	Peru		**C1b**
53	10722*	Ychsma	Peru	**1221-1278 calAD**	**D1**
54	10732	Ychsma	Peru		**D***
55	10730	Ychsma	Peru		**B2b**
56	10805*	Ychsma	Peru	**1223 - 1280 calAD**	**D1**
57	10810*	Ychsma	Peru	**1149-1249 calAD**	**C1d**
58	10794	Ychsma	Peru		**B2**
59	10800	Ychsma	Peru		**B2b**
60	12692	Palpa- Jauranga	Peru		**C1c**
61	12690	Botigiriayocc	Peru		**C1b**
62	11165	Pueblo Viejo, Inca	Peru		**B2**
63	11168	Pueblo Viejo, Inca	Peru		**B2**
64	11173	Pueblo Viejo, Inca	Peru		**B2**
65	11182	Pueblo Viejo, Inca	Peru		**B2b**
66	11176	Pueblo Viejo, Inca	Peru		**B2b**
67	11154	Pueblo Viejo, Inca	Peru		**C1b**
68	11157	Pueblo Viejo, Inca	Peru		**C1b**
69	11161	Pueblo Viejo, Inca	Peru		**C1b**
70	11183	Pueblo Viejo, Inca	Peru		**C1c**
71	11180	Pueblo Viejo, Inca	Peru		**D1**
72	11185	Pueblo Viejo, Inca	Peru		**D1**
73	13989	Candelaria	Mexico		**B2**
74	13990	Candelaria	Mexico		**B2**
75	13992	Candelaria	Mexico		**B2**
76	13994	Candelaria	Mexico		**B2**
77	13993	Candelaria	Mexico		**C1b**

#	UCSC Lab #	Populations	Location	Radiocarbon dates	Haplogroup
78	LGA42	La Galgada	Peru		**A2**
79	ASOB9S14*	ArroyoSeco	Argentina	**7928 -7591 calBP**	**D1**
80	ASOB2S9*	Arroyo Seco	Argentina	**7928 -7581 calBP**	**A2**
81	SO238*	Lauricocha Lau2	Peru	**8589 - 8482 calBP**	**A2**
82	CS1	Camarones	Chile		**A2**

#	UCSC Lab #	Populations	Location	Radiocarbon dates	Haplogroup
83	SSN1	San Sebastian	Peru		**A2**
84	PO3	Pica8	Chile		**A2**
85	PO3	Pica8	Chile		**A2**
86	BC12	Botigiriayocc	Peru		**B2**
87	LGA13	La Galgada	Peru		**B2**
88	JA14	Jauranga	Peru		**B2**
89	M21	Los Molinos	Peru		**B2**
90	HPA866	Huaca Prieta	Peru		**C1b**
91	SSN5	San Sebastian	Peru		**C1d1**
92	LGA2	La Galgada	Peru		**D1**

*: These sequences have direct radiocarbon dates; all other dates are based on the relevant archaeological period.

Cal AD: calibrated radiocarbon years Anno Domini.

Cal BP: calibrated radiocarbon years Before Present

Table 3. Summary statistics of all ancient populations analysed in this study. Sites are listed in chronological order

ANCIENT POPULATIONS	***n***	***H***	***h***	***π***	**Tajima's D**	**p-value**	**Fu's FS**	**p-value**	**Reference**
Arroyo Seco_EMA	**2**	2	1.0000 (0.5000)	0.002294 (0.002324)	0.00000	1.00000	3.63759	0.57600	This Study
Lauricocha_LA	**5**	5	1.0000 (0.1265)	0.001208 (0.000756)	-1.09667	0.095500	0.52104	0.35200	This Study
La Galgaga_LA	**3**	3	1.0000 (0.2722)	0.002818 (0.002127)	-0.00000	0.14700	2.73023	0.57600	This Study
Lima_EIP	**9**	9	1.0000 (0.0388)	0.002187 (0.001165)	-0.47719	0.29600	-1.10349	0.19000	This Study
Palpa_EIP	**4**	4	1.0000 (0.1768)	0.002405 (0.001595)	1.50258	0.89600	1.84734	0.48500	This Study
Wari_MH	**10**	9	0.9778 (0.0540)	0.002424 (0.001304)	0.15810	0.59600	1.09057	0.60200	This Study
Tiwanaku_MH	**9**	9	1.0000 (0.0524)	0.002648 (0.001442)	-0.39744	0.33800	-0.15258	0.27200	This Study
Ychsma_LIP	**15**	15	1.0000 (0.0302)	0.002160 (0.001131)	0.31165	0.67900	-1.81213	0.11100	This Study
Chancay_LIP	**7**	7	1.0000 (0.0764)	0.002651 (0.001505)	-0.47037	0.34400	0.54779	0.38200	This Study
Pica8_LIP	**6**	5	0.9333 (0.1217)	0.001944 (0.001147)	0.99147	0.86400	2.80180	0.82800	This Study
Botigiriayocc_LIP	**2**	2	1.0000 (0.5000)	0.003442 (0.003472)	0.00000	1.00000	4.04305	0.60100	This Study
Cueva_Candelaria_(LIP)	**5**	5	1.0000 (0.1265)	0.001703 (0.001056)	-1.09411	0.09100	0.91407	0.43000	This Study
Llullaillaco_LH	**2**	2	1.0000 (0.5000)	0.001751 (0.001781)	0.00000	1.00000	0.91407	0.43000	This Study
San Sebastian_LH	**2**	2	1.0000 (0.5000)	0.002958 (0.002988)	0.00000	1.00000	3.89182	0.61300	This Study
Pueblo Viejo_LH	**11**	11	1.0000 (0.0388)	0.002216 (0.001180)	0.24737	0.63800	-1.0866	0.17900	This Study

H: Number of haplotypes, h: Haplotype diversity, *π:* Nucleotide diversity*: statistically significant p-values (<0.05 for Tajima's D, p<0.02 for Fu's FS)*

Table 4a. Summary statistics for samples when grouped into archaeological periods (Major Haplogroups)

CULTURAL PERIOD	*n*	mtDNA Hg frequencies			
		A	**B**	**C**	**D**
Archaic Period	**11**	0.64	0.18	-	0.18
Early Intermediate	**13**	0.08	0.46	0.38	0.08
Middle Horizon	**19**	0.16	0.42	0.42	-
Late Intermediate	**29**	0.10	0.45	0.31	0.14
Late Horizon	**15**	0.06	0.34	0.40	0.20

Table 4b. Summary statistics for haplogroup composition of samples when grouped into five archaeological periods

CULTURAL PERIOD	*n*	mtDNA Hg frequencies							
		A2	**B2**	**B2b**	**C1b**	**C1c**	**C1d**	**D**	**D1**
Archaic Period	**11**	0.64	0.18	-	-	-	-	-	0.18
Early Intermediate	**13**	0.08	0.38	0.08	0.23	0.15	-	0.08	-
Middle Horizon	**19**	0.16	0.26	0.16	0.37	0.05	-	-	-
Late Intermediate	**29**	0.11	0.28	0.17	0.24	0.03	0.03	0.03	0.11
Late Horizon	**15**	0.07	0.20	0.13	0.26	0.07	0.07	-	0.20

Table 5. Population grouped in Archaeological Periods (Diversity data and neutrality indexes) for mtDNA genomes (samples from Mexico excluded)

			Genetic diversity		Neutrality test				Mismatch Distribution Demographic Expansion		Mismatch Distribution Spatial Expansion	
CULTURAL PERIOD	***n***	***H***	***h***	***π***	**Tajima's D**	**p-value**	**Fu's FS**	**p-value**	**SSD**	**HRI**	**SSC**	**HPI**
South America ALL	**87**	82	0.9987 (0.0020)	0.002318 (0.001127)	-1.66143	***0.02000****	-24.10417	***0.00000****	0.00495333 (0.47000000)	0.00156548 (0.91000000)	0.00922698 ***0.000000****	0.00156548 0.6000000
4 Periods (AP excluded)	**76**	71	0.9982 (0.0025)	0.002299 (0.001120)	-1.60071	***0.03000****	-24.14606	***0.00100****	0.00594206 (0.44000000)	0.00184882 (0.87000000)	0.01098722 ***0.0000000****	0.00184882 0.69000000
Archaic Period	**11**	11	1.0000 (0.0388)	0.001655 (0.000887)	-0.80436	0.23800	-0.87170	0.21800	0.05529960 (0.10000000)	0.10049587 (0.06000000)	0.05456860 ***0.0000000****	0.10049587 0.05000000
Early Intermediate	**13**	13	0.9905 (0.0281)	0.002157 (0.001116)	-0.50782	0.34500	-0.89171	0.28500	0.02232587 (0.29000000)	0.01750567 (0.72000000)	0.01935082 **0.06000000**	0.01750567 0.87000000
Middle Horizon	**19**	18	0.9942 (0.0193)	0.002530 (0.001284)	-0.70060	0.26300	-1.80845	0.15900	0.01624113 (0.21000000)	0.02116891 (0.16000000)	0.01895087 ***0.0000000****	0.02116891 0.25000000
Late Intermediate	**29**	29	0.9933 (0.0134)	0.002372 (0.001188)	-0.91209	0.17400	-3.55148	0.09400	0.01069635 (0.44000000)	0.01375801 (0.19000000)	0.01460366 ***0.01000000****	0.01375801 0.27000000
Late Horizon	**15**	14	0.9905 (0.0281)	0.002230 (0.001153)	-0.50628	0.34100	-0.81803	0.29600	0.02278091 (0.11000000)	0.02512472 (0.62000000)	0.02058531 ***0.05000000****	0.02512472 0.78000000

H: Number of haplotypes, h: Haplotype diversity, *π:* Nucleotide diversity, SSD: sum of squared difference, HRI: Hapending's Raggedness index. *- statistical significant p-values ($p<0.05$ for Tajima's D, $p<0.02$ for Fu's FS)

Table 6. Genetic distances based on mtDNA complete genome computed for ancient populations grouped into archaeological periods. The table provides pairwise F_{ST} (regular) and p-values (italicized).

	Archaic_Period	**Early Intermediate Period_EIP**	**Middle Horizon_MH**	**Late Intermediate Period_LIP**	**Late Horizon_LH**
Archaic_Period	0	*0.00000+-0.0000*	*0.00901+-0.0091*	*0.00000+-0.0000*	*0.00000+-0.0000*
Early Intermediate Period_EIP	0.19071	0	*0.80180+-0.0287*	*0.88288+-0.0266*	*0.88288+-0.0266*
Middle Horizon_MH	0.14892	-0.04209	0	*0.69369+-0.0334*	*0.47748+-0.0592*
Late Intermediate Period_LIP	0.15529	-0.03668	-0.01928	0	*0.54955+-0.0438*
Late Horizon_LH	0.18296	-0.03065	-0.01373	-0.01890	0

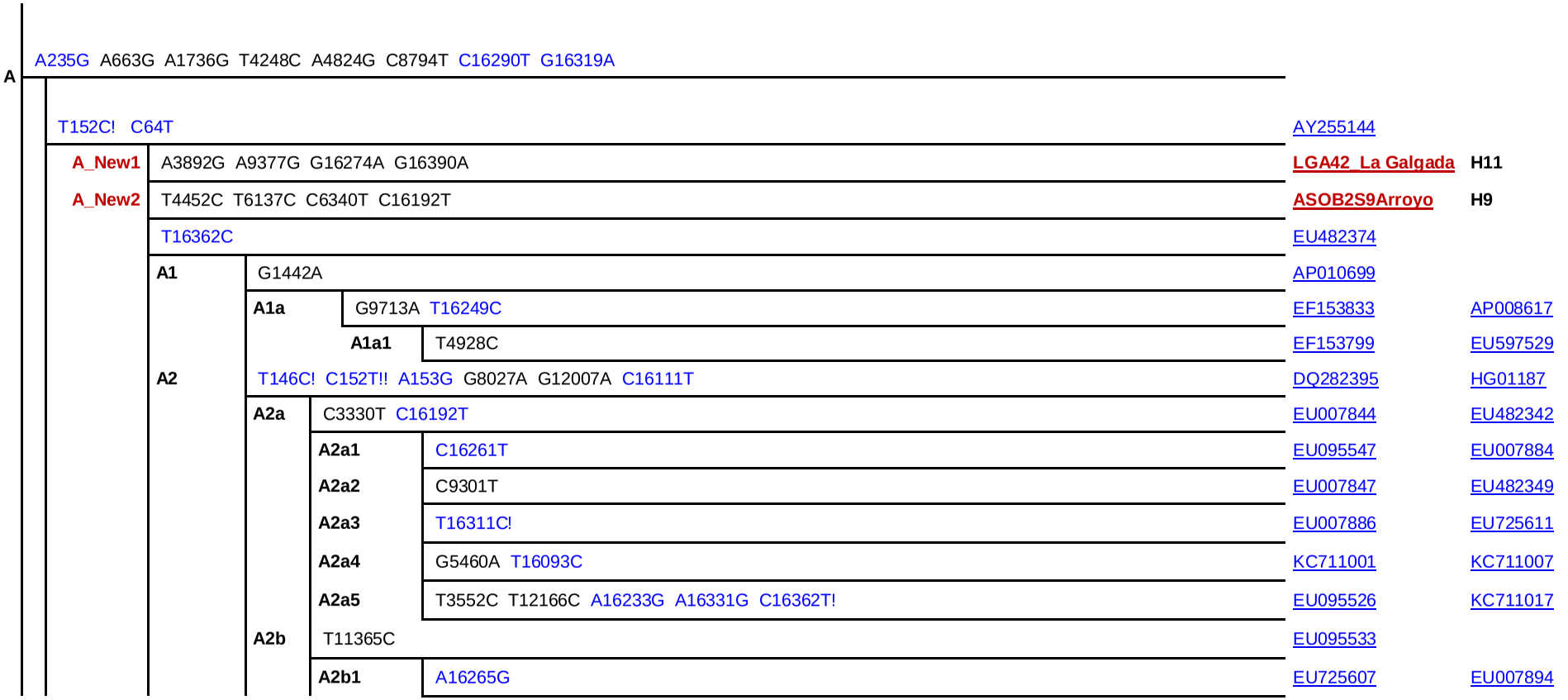
SUPLEMENTARY INFORMATION

Supplementary Figure 1a. Phylogenetic tree of newly sequenced ancient mitochondrial genomes **(Haplogroup A)**.

Nomenclature and topology are based on the PhyloTree mtDNA tree build 16 (19 Feb 2014) (van Oven and Kayser 2009).

Haplogroup A: New mtDNA haplotypes (15)

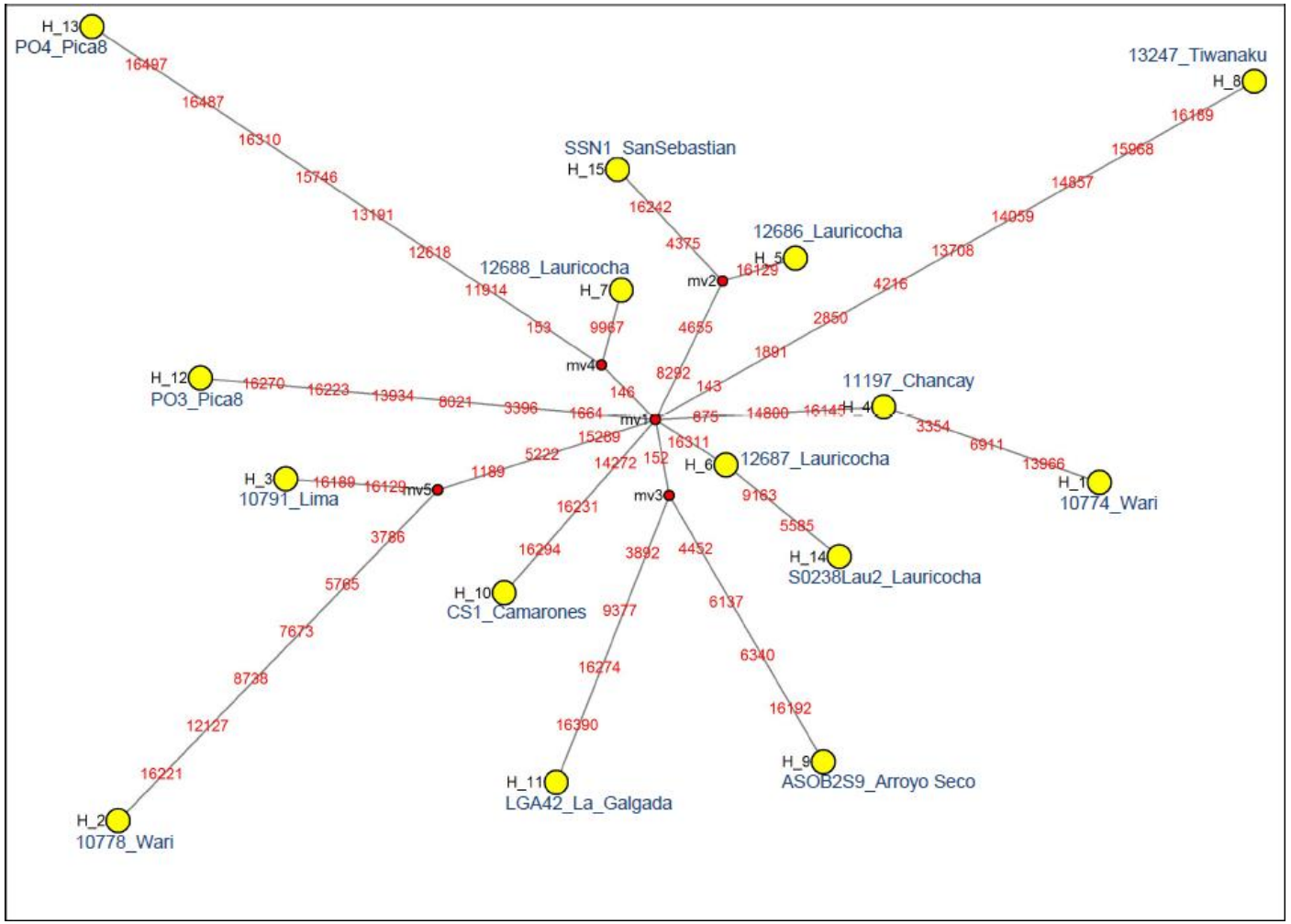
A2: T146C! C152T!! A153G G8027A G12007A C16111T



(C64T)		EU095538	DQ282394
A2_New1	G143A A1891G T2850C T4216C G13708A A14059G T14857C T15968C T16189C!	13247A Tiwanaku	H8
A2_New2	G1664A T3396C A8021G C13934T T16223C C16270T	PO3 Pica8	H12
A2_New3	C14272T T16231C C16294T	CS1_Camarones	H10
A2_New4	T16311C!	12687_Lauricocha	H6
A2_New5	G9163A G5585A	S0238 Lau2	H14
C146T!!			
A2_New6	T9967C	12688_Lauricocha	H7
A2_New7	G153A G11914A G12618A T13191C A15746G G16310A A16487G A16497G	P04_Pica8	H13
A2_New8	A675G C14800A G16145A	11197A_Pasamayo	H4
A2_New9	A3354G T6911C A13966G	10774A_Wari	H1
A2_New10	G4655A G8292A		
A2_New11	G16129A!	12686_Lauricocha	H5
A2_New12	C4375T C16242A	SSN1 San Sebastian	H15
T1189C A5222G T15289C			
A2_New13	356.1C G16129A! T16189C!	10791A_Lima	H3
	T3786C C5765T A7673G T8738C G12127A C16221T	10778A_Wari	H2
A2h	G16526A	EU095545	EU095202
A2h1	G1598A G1888A T12811C A16335G	HQ012079	JQ705534
A2i	G94A 960.XC 3307.1A T3308C C5165T T6620C A14280G T14470C C15386T T16325C	EU431080	GQ377757
A2j	T10595C A11548G	JQ705434	HQ012127
A2j1	A11314G	DQ282398	DQ282393
A2k	T3202C	DQ282417	
A2k1	A15924G	EU095552	

Supplementary Figure 1b. Mitochondrial DNA haplotype network analysis from complete genomes sequenced in this study

HAPLOGROUP A

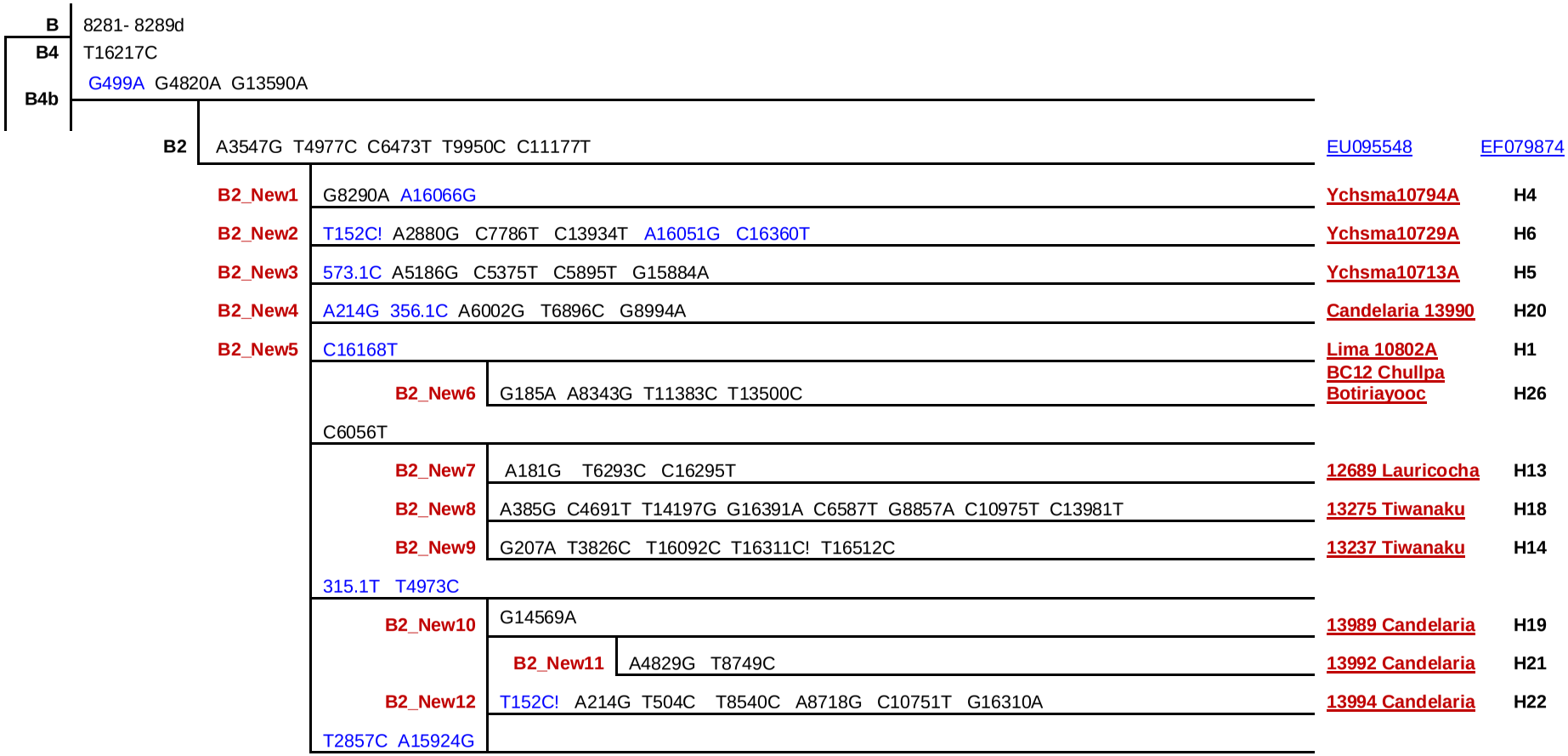


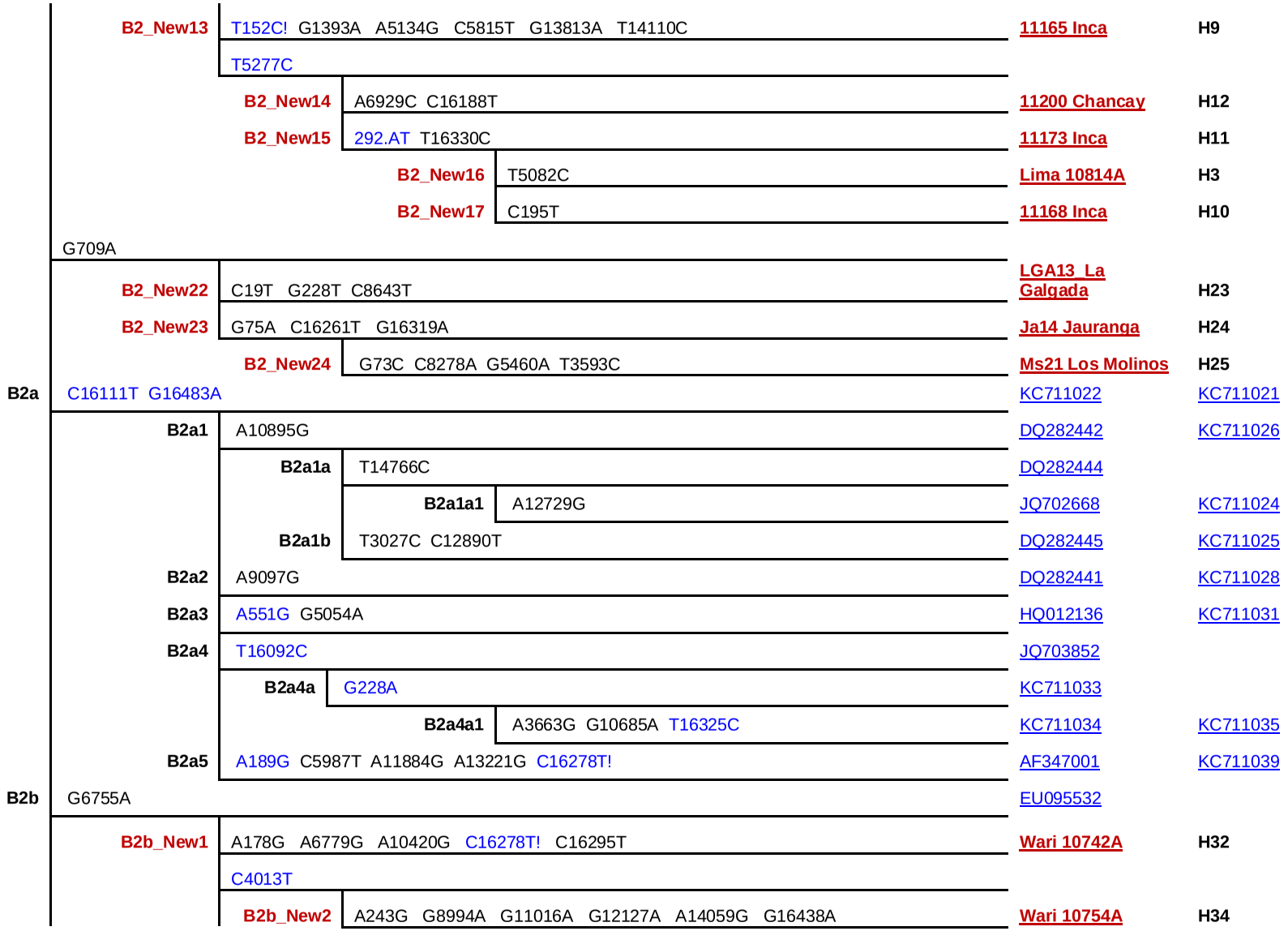
Supplementary Figure 2a. Phylogenetic tree of newly sequenced ancient mitochondrial genomes **(Haplogroup B)**.

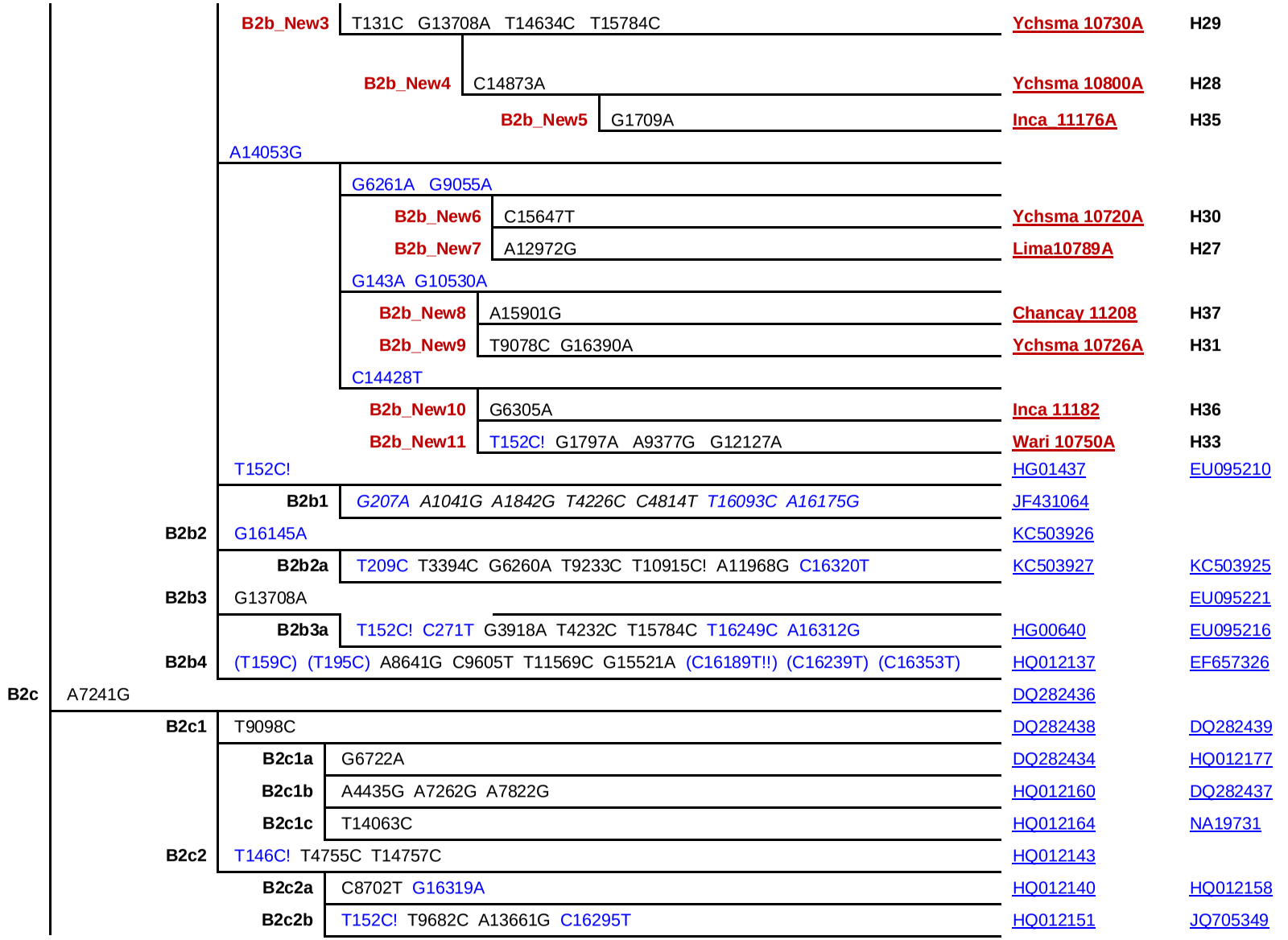
Nomenclature and topology are based on the PhyloTree mtDNA tree build 16 (19 Feb 2014) (van Oven and Kayser 2009).

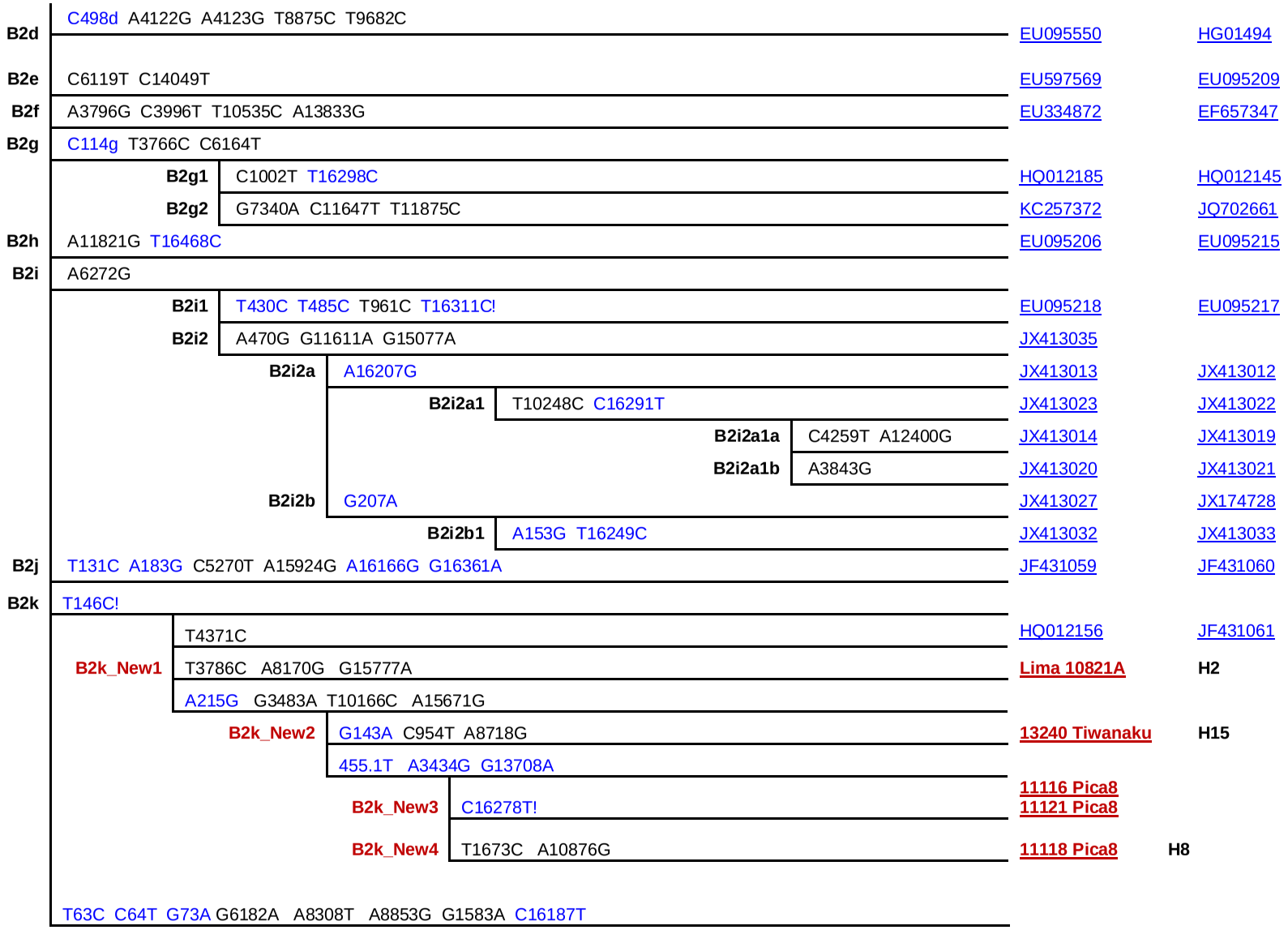
Haplogroup B: New mtDNA haplotypes (37)

B2: A3547G, T4977C, C6473T, T9950C, C11177T; **B2b:** G6755A




B2_New13
T152C! G1393A A5134G C5815T G13813A T14110C
11165 Inca
H9
T5277C
B2_New14
A6929C C16188T
11200 Chancay
H12
B2_New15
292.AT T16330C
11173 Inca
H11
B2_New16
T5082C
Lima 10814A
H3
B2_New17
C195T
11168 Inca
H10
G709A
B2_New22
C19T G228T C8643T
LGA13_La Galgada
H23
B2_New23
G75A C16261T G16319A
Ja14 Jauranga
H24
B2_New24
G73C C8278A G5460A T3593C
Ms21 Los Molinos
H25
B2a
C16111T G16483A
KC711022
KC711021
B2a1
A10895G
DQ282442
KC711026
B2a1a
T14766C
DQ282444
B2a1a1
A12729G
JQ702668
KC711024
B2a1b
T3027C C12890T
DQ282445
KC711025
B2a2
A9097G
DQ282441
KC711028
B2a3
A551G G5054A
HQ012136
KC711031
B2a4
T16092C
JQ703852
B2a4a
G228A
KC711033
B2a4a1
A3663G G10685A T16325C
KC711034
KC711035
B2a5
A189G C5987T A11884G A13221G C16278T!
AF347001
KC711039
B2b
G6755A
EU095532
B2b_New1
A178G A6779G A10420G C16278T! C16295T
Wari 10742A
H32
C4013T
B2b_New2
A243G G8994A G11016A G12127A A14059G G16438A
Wari 10754A
H34

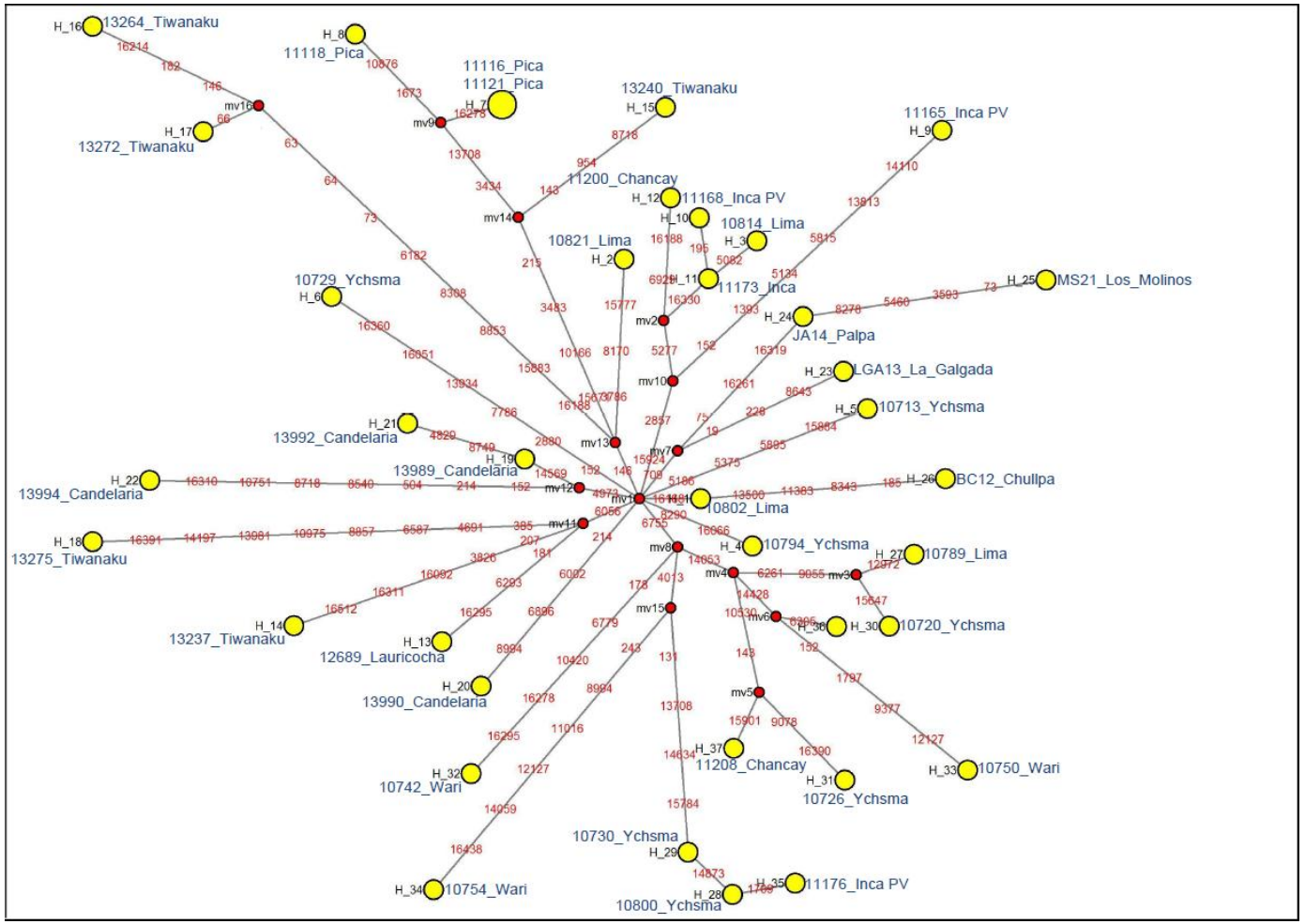

B2b_New3 T131C G13708A T14634C T15784C Ychsma 10730A H29
B2b_New4 C14873A Ychsma 10800A H28
B2b_New5 G1709A Inca_11176A H35
A14053G
G6261A G9055A
B2b_New6 C15647T Ychsma 10720A H30
B2b_New7 A12972G Lima10789A H27
G143A G10530A
B2b_New8 A15901G Chancay 11208 H37
B2b_New9 T9078C G16390A Ychsma 10726A H31
C14428T
B2b_New10 G6305A Inca 11182 H36
B2b_New11 T152C! G1797A A9377G G12127A Wari 10750A H33
T152C! HG01437 EU095210
B2b1 G207A A1041G A1842G T4226C C4814T T16093C A16175G JF431064
B2b2 G16145A KC503926
B2b2a T209C T3394C G6260A T9233C T10915C! A11968G C16320T KC503927 KC503925
B2b3 G13708A EU095221
B2b3a T152C! C271T G3918A T4232C T15784C T16249C A16312G HG00640 EU095216
B2b4 (T159C) (T195C) A8641G C9605T T11569C G15521A (C16189T!!) (C16239T) (C16353T) HQ012137 EF657326
B2c A7241G DQ282436
B2c1 T9098C DQ282438 DQ282439
B2c1a G6722A DQ282434 HQ012177
B2c1b A4435G A7262G A7822G HQ012160 DQ282437
B2c1c T14063C HQ012164 NA19731
B2c2 T146C! T4755C T14757C HQ012143
B2c2a C8702T G16319A HQ012140 HQ012158
B2c2b T152C! T9682C A13661G C16295T HQ012151 JQ705349


B2d C498d A4122G A4123G T8875C T9682C EU095550 HG01494
B2e C6119T C14049T EU597569 EU095209
B2f A3796G C3996T T10535C A13833G EU334872 EF657347
B2g C114g T3766C C6164T
B2g1 C1002T T16298C HQ012185 HQ012145
B2g2 G7340A C11647T T11875C KC257372 JQ702661
B2h A11821G T16468C EU095206 EU095215
B2i A6272G
B2i1 T430C T485C T961C T16311C! EU095218 EU095217
B2i2 A470G G11611A G15077A JX413035
B2i2a A16207G JX413013 JX413012
B2i2a1 T10248C C16291T JX413023 JX413022
B2i2a1a C4259T A12400G JX413014 JX413019
B2i2a1b A3843G JX413020 JX413021
B2i2b G207A JX413027 JX174728
B2i2b1 A153G T16249C JX413032 JX413033
B2j T131C A183G C5270T A15924G A16166G G16361A JF431059 JF431060
B2k T146C!
T4371C HQ012156 JF431061
B2k_New1 T3786C A8170G G15777A Lima 10821A H2
A215G G3483A T10166C A15671G
B2k_New2 G143A C954T A8718G 13240 Tiwanaku H15
455.1T A3434G G13708A
11116 Pica8
B2k_New3 C16278T! 11121 Pica8
B2k_New4 T1673C A10876G 11118 Pica8 H8
T63C C64T G73A G6182A A8308T A8853G G1583A C16187T

	B2k_New5	G66A	13272 Tiwanaku	H17
	B2k_New6	C182T!! C16214T C146T	13264 Tiwanaku	H16
B2l	T16422C		JQ702293	NA19785
B2m	A15766G A16164G C16519a		HQ012141	NA19761
B2n	A4191G G6383A		HQ012163	NA19795
B2o	T16092C		JQ703851	

Supplementary Figure 2b. Mitochondrial DNA haplotype network analysis from complete genomes sequenced in this study

HAPLOGROUP B

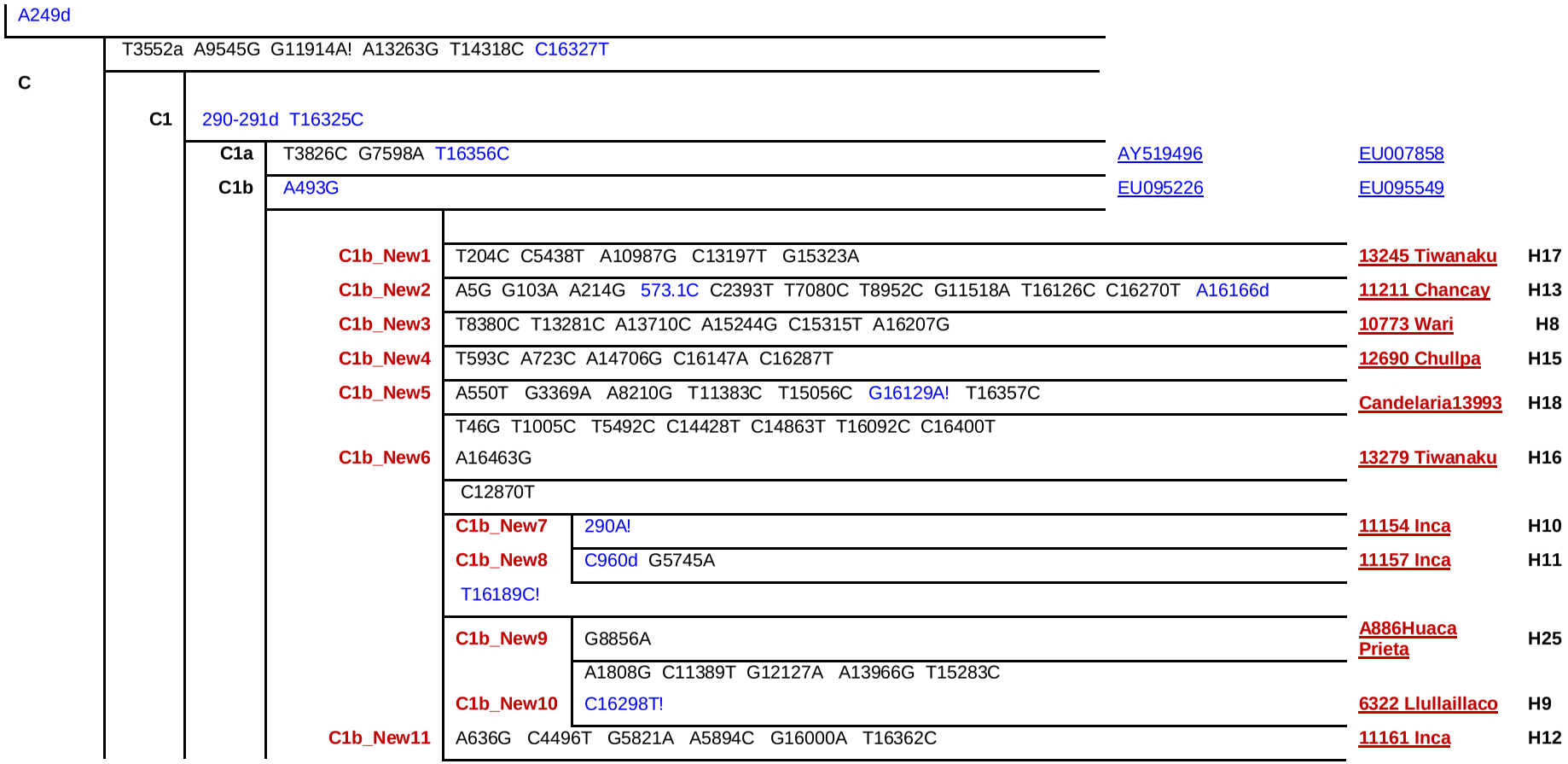


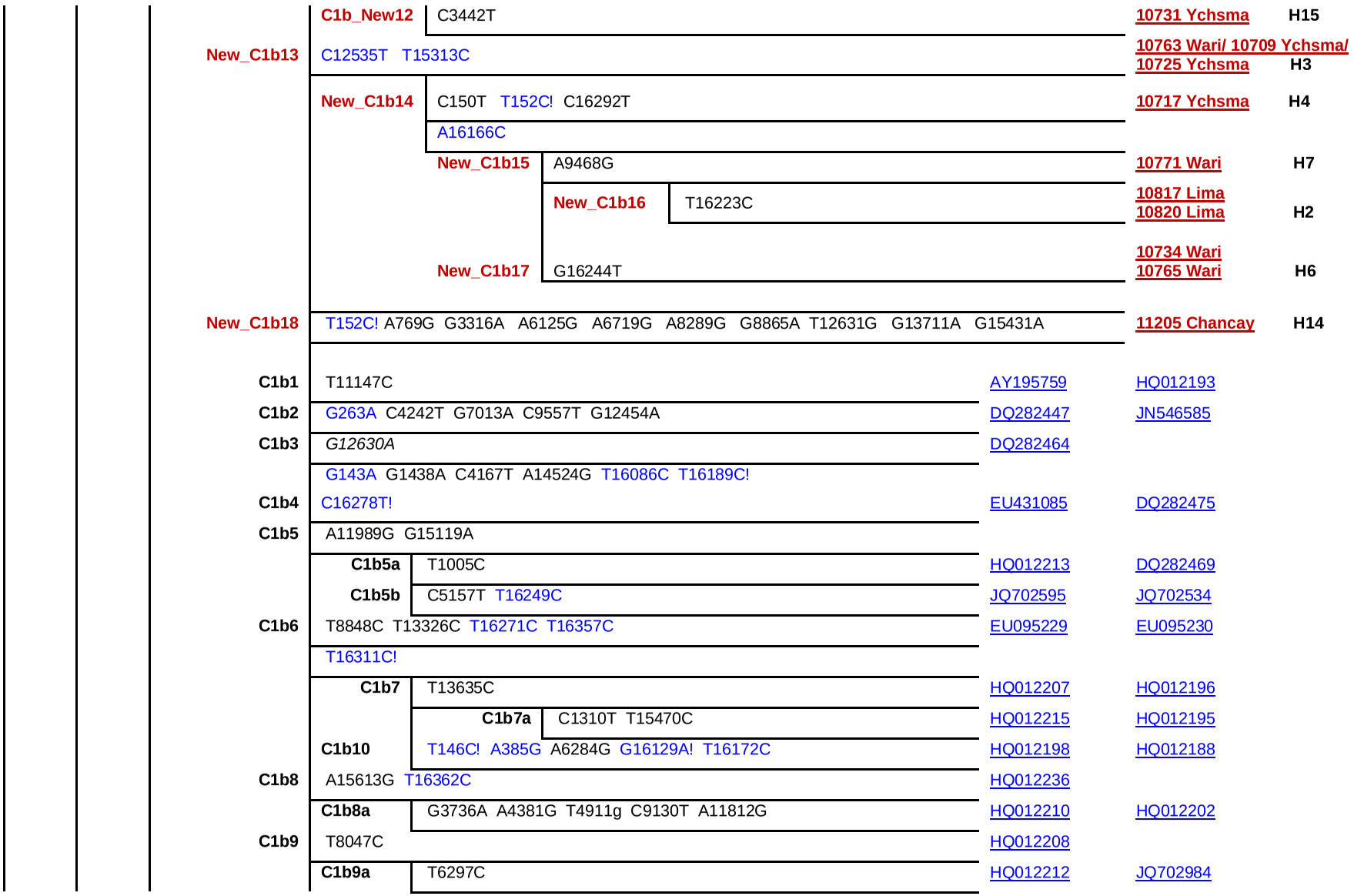
Supplementary Figure 3a. Phylogenetic tree of newly sequenced ancient mitochondrial genomes **(Haplogroup C)**.

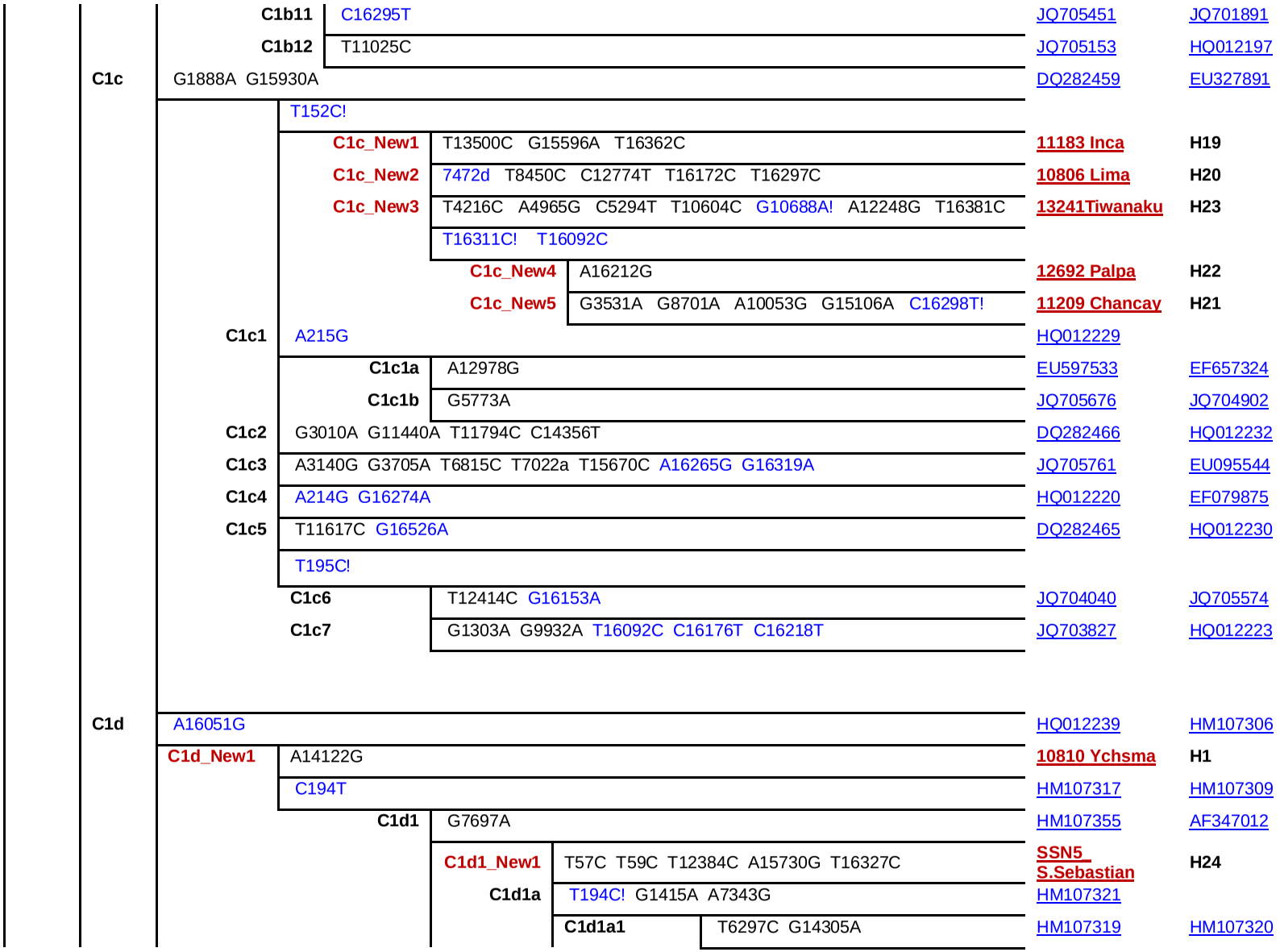
Nomenclature and topology are based on the PhyloTree mtDNA tree build 16 (19 Feb 2014) (van Oven and Kayser 2009).

Haplogroup C: New mtDNA haplotypes (25)

C1b: A493G, **C1c:** G1888A, G15930A, **C1d:** A16051G

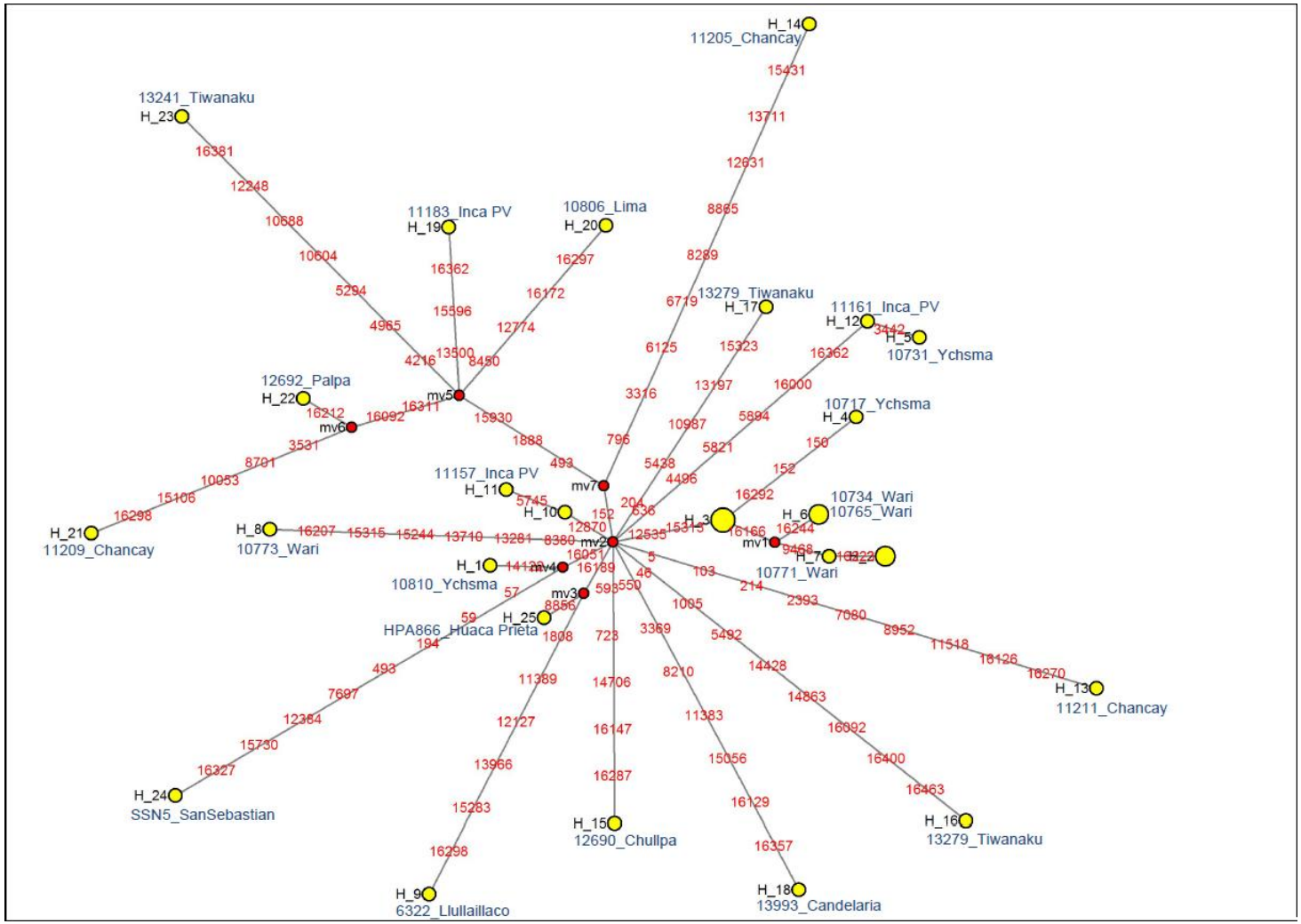



C1b_New12
C3442T
10731 Ychsma
H15
New_C1b13
C12535T T15313C
10763 Wari/ 10709 Ychsma/ 10725 Ychsma
H3
New_C1b14
C150T T152C! C16292T
A16166C
10717 Ychsma
H4
New_C1b15
A9468G
10771 Wari
H7
New_C1b16
T16223C
10817 Lima
10820 Lima
H2
New_C1b17
G16244T
10734 Wari
10765 Wari
H6
New_C1b18
T152C! A769G G3316A A6125G A6719G A8289G G8865A T12631G G13711A G15431A
11205 Chancay
H14
C1b1
T11147C
AY195759
HQ012193
C1b2
G263A C4242T G7013A C9557T G12454A
DQ282447
JN546585
C1b3
G12630A
G143A G1438A C4167T A14524G T16086C T16189C!
DQ282464
C1b4
C16278T!
EU431085
DQ282475
C1b5
A11989G G15119A
C1b5a
T1005C
HQ012213
DQ282469
C1b5b
C5157T T16249C
JQ702595
JQ702534
C1b6
T8848C T13326C T16271C T16357C
T16311C!
EU095229
EU095230
C1b7
T13635C
HQ012207
HQ012196
C1b7a
C1310T T15470C
HQ012215
HQ012195
C1b10
T146C! A385G A6284G G16129A! T16172C
HQ012198
HQ012188
C1b8
A15613G T16362C
HQ012236
C1b8a
G3736A A4381G T4911g C9130T A11812G
HQ012210
HQ012202
C1b9
T8047C
HQ012208
C1b9a
T6297C
HQ012212
JQ702984


C1b11 C16295T JQ705451 JQ701891
C1b12 T11025C JQ705153 HQ012197
C1c G1888A G15930A DQ282459 EU327891
T152C!
C1c_New1 T13500C G15596A T16362C 11183 Inca H19
C1c_New2 7472d T8450C C12774T T16172C T16297C 10806 Lima H20
C1c_New3 T4216C A4965G C5294T T10604C G10688A! A12248G T16381C 13241Tiwanaku H23
T16311C! T16092C
C1c_New4 A16212G 12692 Palpa H22
C1c_New5 G3531A G8701A A10053G G15106A C16298T! 11209 Chancay H21
C1c1 A215G HQ012229
C1c1a A12978G EU597533 EF657324
C1c1b G5773A JQ705676 JQ704902
C1c2 G3010A G11440A T11794C C14356T DQ282466 HQ012232
C1c3 A3140G G3705A T6815C T7022a T15670C A16265G G16319A JQ705761 EU095544
C1c4 A214G G16274A HQ012220 EF079875
C1c5 T11617C G16526A DQ282465 HQ012230
T195C!
C1c6 T12414C G16153A JQ704040 JQ705574
C1c7 G1303A G9932A T16092C C16176T C16218T JQ703827 HQ012223
C1d A16051G HQ012239 HM107306
C1d_New1 A14122G 10810 Ychsma H1
C194T HM107317 HM107309
C1d1 G7697A HM107355 AF347012
C1d1_New1 T57C T59C T12384C A15730G T16327C SSN5_S.Sebastian H24
C1d1a T194C! G1415A A7343G HM107321
C1d1a1 T6297C G14305A HM107319 HM107320

Supplementary Figure 3b. Mitochondrial DNA haplotype network analysis from complete genomes sequenced in this study

HAPLOGROUP C



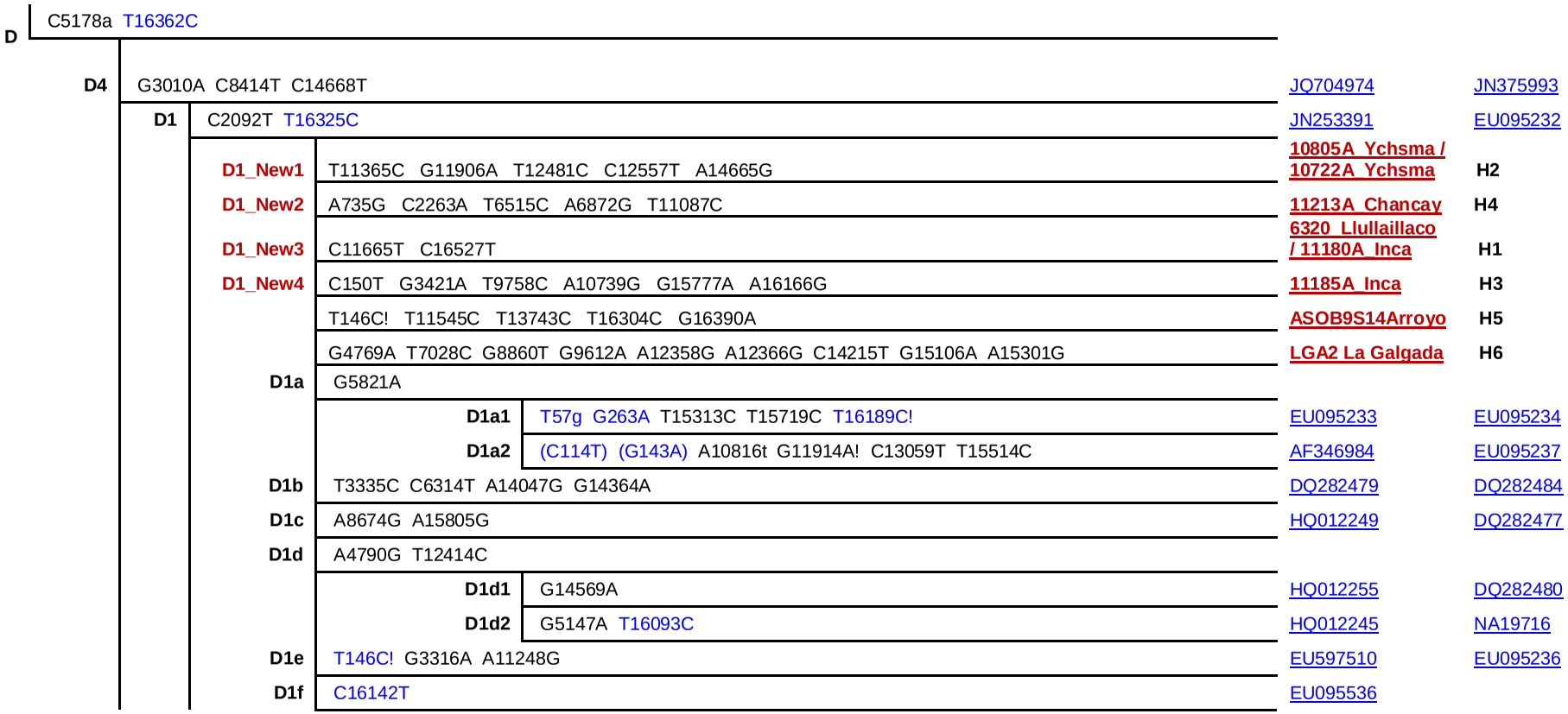
Supplementary Figure 4a. Phylogenetic tree of newly sequenced ancient mitochondrial genomes **(Haplogroup D)**.

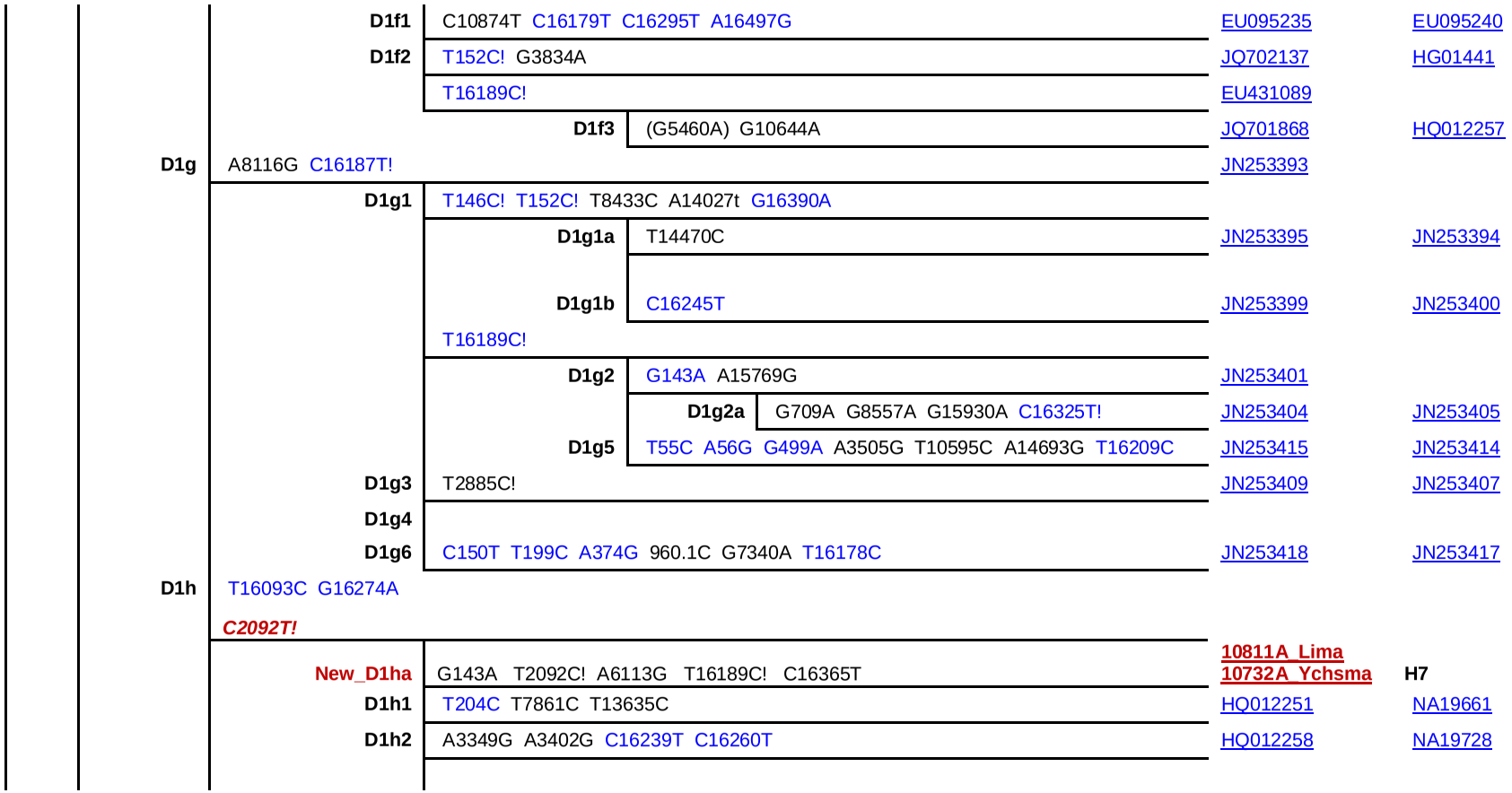
Nomenclature and topology are based on the PhyloTree mtDNA tree build 16 (19 Feb 2014) (van Oven and Kayser 2009).

Haplogroup D: New mtDNA haplotypes (7)

D1: C2092T, T16325C

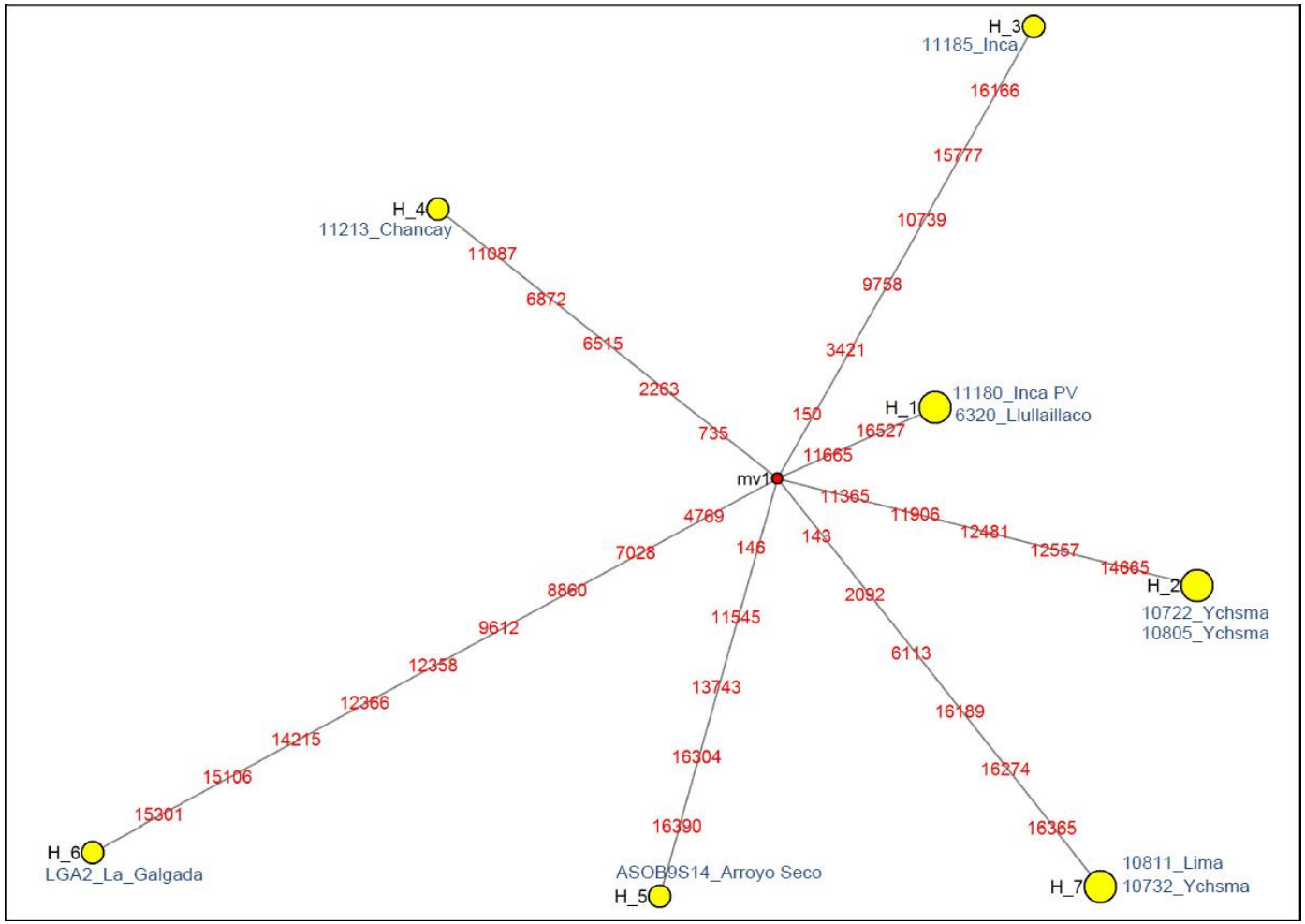
D*




D1f1 C10874T C16179T C16295T A16497G EU095235 EU095240
D1f2 T152C! G3834A JQ702137 HG01441
T16189C! EU431089
D1f3 (G5460A) G10644A JQ701868 HQ012257
D1g A8116G C16187T! JN253393
D1g1 T146C! T152C! T8433C A14027t G16390A
D1g1a T14470C JN253395 JN253394
D1g1b C16245T JN253399 JN253400
T16189C!
D1g2 G143A A15769G JN253401
D1g2a G709A G8557A G15930A C16325T! JN253404 JN253405
D1g5 T55C A56G G499A A3505G T10595C A14693G T16209C JN253415 JN253414
D1g3 T2885C! JN253409 JN253407
D1g4
D1g6 C150T T199C A374G 960.1C G7340A T16178C JN253418 JN253417
D1h T16093C G16274A
C2092T!
New_D1ha G143A T2092C! A6113G T16189C! C16365T 10811A_Lima 10732A_Ychsma H7
D1h1 T204C T7861C T13635C HQ012251 NA19661
D1h2 A3349G A3402G C16239T C16260T HQ012258 NA19728

Supplementary Figure 4b. Mitochondrial DNA haplotype network analysis from complete genomes sequenced in this study

HAPLOGROUP D



Supplementary Table 1. Haplogroup determination and genomic coverage from mtDNA complete genomes

ACAD #	Location	Population	Haplogroup	Coverage of the mitogenome	Average coverage depth per position	Std. Dev.	Min	Max	Private mutations
11116	Northern Chile	Pica8-Tarapacá	B2	100%	**165.5**	13.4	20	188	9
11118	Northern Chile	Pica8-Tarapacá	B2	99.99%	**43.6**	13	0	86	10
11121	Northern Chile	Pica8-Tarapacá	B2	100%	**180.6**	11.2	27	192	9
11200	Peru	Chancay	B2	100%	**69.4**	18.4	1	124	4
11211	Peru	Chancay	C1b	100%	**26.2**	8.8	0	52	8
11205	Peru	Chancay	C1b	100%	**111.6**	12	6	142	10
11209	Peru	Chancay	C1c	100%	**165.0**	14	7	188	8
11213	Peru	Chancay	D1	100%	**86.5**	19.6	3	143	6
11208	Peru	Chancay	B2b	99.80%	**27.6**	18.1	0	99	4
11202	Peru	Chancay	B2b	94.70%	**6.1**	4.7	0	27	4
11197	Peru	Chancay	A2	100%	**341.5**	194.3	9	991	3
6322	Argentina	Llullaillaco	C1b	100%	**62.9**	11.5	0	131	6
6317	Argentina	Llullaillaco	D1	94.50%	**3.5**	2.7	0	35	7
6320	Northern Argentina	Llullaillaco	D1	100%	**40.3**	11.5	0	131	3
10791	Huaca Pucllana	Lima	A2	100%	**122.7**	13.5	1	159	4
10814	Huaca Pucllana	Lima	B2	100%	**164.6**	14.3	9	186	5
10802	Huaca Pucllana	Lima	B2	100%	**63.4**	14.4	1	105	1
10789	Huaca Pucllana	Lima	B2b	100%	**100.2**	15.1	5	141	3
10817	Huaca Pucllana	Lima	C1b	100%	**132.2**	16.4	5	169	5
10820	Huaca Pucllana	Lima	C1b	100%	**84.6**	15.8	1	135	5
10806	Huaca Pucllana	Lima	C1c	100%	**73.4**	17.1	1	117	6
10811	Huaca Pucllana	Lima	D*	99.99%	**50.4**	14.9	0	97	6
10822	Huaca Pucllana	Lima	B2	99.90%	**82.8**	35.8	0	151	4
10821	Huaca Pucllana	Lima	B2	100%	**98.2**	19.9	1	149	4

ACAD #	Location	Population	Haplogroup	Coverage of the mitogenome	Average coverage depth per position	Std. Dev.	Min	Max	Private mutations
10774	Huaca Pucllana	Wari	A2	100%	**67.7**	18.1	1	115	6
10754	Huaca Pucllana	Wari	B2b	100%	**48.8**	17	1	99	7
10734	Huaca Pucllana	Wari	C1b	100%	**59.8**	17.3	1	104	4
10758	Huaca Pucllana	Wari	C1b	98.20%	**6.4**	4.1	0	22	8
10763	Huaca Pucllana	Wari	C1b	99.90%	**20.7**	9	0	55	2
10771	Huaca Pucllana	Wari	C1b	98.80%	**18.2**	8.2	0	50	4
10773	Huaca Pucllana	Wari	C1b	100%	**78.0**	15.2	0	119	6
10778	Huaca Pucllana	Wari	A2	97.60%	**20.5**	15.6	0	100	7
10765	Huaca Pucllana	Wari	C1b	99.80%	**105.2**	62.5	0	340	4
10750	Huaca Pucllana	Wari	B2b	98.70%	**22.3**	16.7	0	97	7
10742	Huaca Pucllana	Wari	B2b	100%	**84.3**	40.2	0	248	5
10713	Huaca Pucllana	Ychsma	B2	100%	**109.5**	21.9	2	163	4
10729	Huaca Pucllana	Ychsma	B2	100%	**101.6**	24.8	0	171	6
10720	Huaca Pucllana	Ychsma	B2b	100%	**171.6**	14.4	19	191	4
10726	Huaca Pucllana	Ychsma	B2b	99.96%	**38.5**	14.7	0	88	5
10709	Huaca Pucllana	Ychsma	C1b	100%	**146.6**	18.1	0	176	2
10717	Huaca Pucllana	Ychsma	C1b	100%	**175**	11.9	18	191	5
10725	Huaca Pucllana	Ychsma	C1b	100%	**177.9**	16.6	0	192	2
10731	Huaca Pucllana	Ychsma	C1b	100%	**182**	17.7	0	193	7
10722	Huaca Pucllana	Ychsma	D1	100%	**140.8**	21	0	181	5
10732	Huaca Pucllana	Ychsma	D*	100%	**146.6**	18.1	0	180	6
10730	Huaca Pucllana	Ychsma	B2b	99.80%	**46.0**	20.2	0	102	5
10805	Huaca Pucllana	Ychsma	D1	100%	**159.5**	13.1	7	186	6
10810	Huaca Pucllana	Ychsma	C1d	99.90%	**90.3**	39.4	0	166	2
10800	Huaca Pucllana	Ychsma	B2b	100%	**127.6**	17.2	4	166	6
10794	Huaca Pucllana	Ychsma	B2	99.60%	**16.8**	8.9	0	54	1

ACAD #	Location	Population	Haplogroup	Coverage of the mitogenome	Average coverage depth per position	Std. Dev.	Min	Max	Private mutations
13246	Bolivia	Tiwanaku	A2	100%	**51.7**	18.4	2	98	9
13247	Bolivia	Tiwanaku	A2	100%	**85.1**	17	0	130	9
13237	Bolivia	Tiwanaku	B2	100%	**138**	22.1	5	182	4
13264	Bolivia	Tiwanaku	B2	99.99%	**34.3**	9	0	62	10
13272	Bolivia	Tiwanaku	B2	100%	**189.7**	11.1	13	193	10
13240	Bolivia	Tiwanaku	B2	100%	**11.0**	6.1	0	36	8
13275	Bolivia	Tiwanaku	B2	99.10%	**5.5**	2.8	0	18	5
13241	Bolivia	Tiwanaku	C1c	100%	**180**	12	15	192	8
13245	Bolivia	Tiwanaku	C1*	99.3	**7.6**	3.6	0	21	4
13279	Bolivia	Tiwanaku	C1b	99.90%	**36.2**	15.6	0	94	8
13989	Mexico	Candelaria	B2	100%	**139.7**	24.5	9	180	3
13990	Mexico	Candelaria	B2	100%	**42.6**	20.7	1	110	6
13991	Mexico	Candelaria	**CONTAMINATION A2/B2**		**-**	-	-	-	-
13992	Mexico	Candelaria	B2	99.99%	**45.3**	14.8	0	89	5
13994	Mexico	Candelaria	B2	98.50%	**8.2**	5.1	0	29	2
13993	Mexico	Candelaria	C1b	99.96%	**25.3**	8.7	0	64	8
10891	Argentina	Patagonia	B2b	95.10%	**3.6**	2.1	0	13	3
12692	Peru	Palpa	C1c	99.99%	**26.5**	10.1	0	70	4
11187	Peru	Miramar-Huari	C1?	87.70%	**4.6**	4.3	0	32	1
11185	Peru	Pueblo Viejo	D1	99.90%	**44.7**	27.2	0	165	7
11183	Peru	Pueblo Viejo	C1c	99.80%	**44.9**	28.5	0	190	4
11182	Peru	Pueblo Viejo	B2b	99.50%	**45.4**	32	0	180	3
11180	Peru	Pueblo Viejo	D1	100%	**344.0**	159.2	5	814	3
11176	Peru	Pueblo Viejo	B2b	100%	**418.1**	187.8	8	903	6
11173	Peru	Pueblo Viejo	B2	100%	**188.6**	112.6	3	663	5
11168	Peru	Pueblo Viejo	B2	100%	**952.0**	409.2	11	2111	5

ACAD #	Location	Population	Haplogroup	Coverage of the mitogenome	Average coverage depth per position	Std. Dev.	Min	Max	Private mutations
11165	Peru	Pueblo Viejo	B2	100%	**678.5**	265.8	11	1587	8
11161	Peru	Pueblo Viejo	C1b	98.90%	**79.7**	49.6	0	280	6
11157	Peru	Pueblo Viejo	C1b	98.20%	**18.2**	13.1	0	81	1
11154	Peru	Pueblo Viejo	C1b	99.60%	**75.1**	47	0	267	2
12686	Peru	Lauricocha	A2	99.70%	**96.7**	41.2	0	179	3
12687	Peru	Lauricocha	A2	99.50%	**74.6**	36.6	0	148	1
12688	Peru	Lauricocha	A2	99.90%	**61.5**	30.2	0	123	2
12689	Peru	Lauricocha	B2	98.60%	**78.8**	9	32	134	4
12690	Peru	Botigiriayocc	C1b	99.30%	**76.6**	20.2	0	158	5
UCSC Lab #	**Location**	**Population**	**Haplogroup**	**Coverage of the mitogenome**	**Average coverage depth per position**	**Std. Dev.**	**Min**	**Max**	**Private mutations**
LGA42	Peru	La Galgada	A2						5
ASOB2S9	Argentina	Arroyo Seco	A2						5
SO238	Peru	Lau2	A2						3
CS1	Chile	Camarones	A2						3
SSN1	Peru	San Sebastian	A2						4
PO4	Chile	Pica8	A2						9
PO3	Chile	Pica8	A2						6
BC12	Peru	Botigiriayocc	B2						5
LGA13	Peru	La Galgada	B2						4
JA14	Peru	Jauranga	B2						4
M21	Peru	Los Molinos	B2						8
HPA866	Peru	Huaca Prieta	C1b						2
SSN5	Peru	San Sebastian	C1d1						5
ASOB9S14	Argentina	ArroyoSeco	D1						5
LGA2	Peru	La Galgada	D1						9

Supplementary Table 2. List of "private" variants (SNPs) from mtDNA complete genomes

Haplogroup A										
A2										
13247_Tiwanaku	C64T	G143A	A1891G	T2850C	T4216C	G13708A	A14059G	T14857C	T15968C	T16189C!
10791_Lima	C64T	356.1C	T1189C	A5222G	T15289C	G16129A!	T16189C!			
10774_Wari	C64T	A675G	A3354G	T6911C	A13966G	C14800A	G16145A			
12686_Lauricocha	C64T	G4655A	G8292A	G16129A!						
12687_Lauricocha	C64T	T16311C!								
12688_Lauricocha	C64T	C146T!!	T9967C							
10778_Wari	C64T	T1189C	T3786C	A5222G	C5765T	A7673G	T8738C	T15289C	C16221T	G12127A
11197_Chancay	C64T	A675G	C14800A	G16145A						
LGA42_La Galgada	C64T	T152C!	A3892G	A9377G	G16274A	G16390A				
ASOB2S9_Arroyo Seco	C64T	T152C!	T4452C	T6137C	C6340T	C16192T				
SO238_Lau2	C64T	G5585A	G9163A	T16311C!						
CS1_Camarones	C64T	C14272T	T16231C	C16294T						
SSN1_San Sebastian	C64T	C4375T	G4655A	G8292A	C16242A					
PO4_Pica8	C64T	C146T!!	G153A	G11914A	G12618A	T13191C	A15746G	G16310A	A16487G	A16497G
PO3_Pica8	C64T	G1664A	T3396C	A8021G	C13934T	T16223C	C16270T			
Haplogroup B										
B2										
13237_Tiwanaku	G207A	T3826C	C6056T	T16092C	T16512C	T16311C!				
13264_Tiwanaku	T63C	C64T	G73A	C182T!!	G6182A	A8308T	G15883A	A8853G	C16188T	C16214T
13272_Tiwanaku	T63C	C64T	G66A	G73A	T146C!	G6182A	A8308T	A8853G	G15883A	C16188T

13240_Tiwanaku	G143A	T146C!	A215G	C954T	G3483A	A8718G	T10166C	A15671G		
13275_Tiwanaku	A385G	C4691T	C6056T	T14197G	G16391A	C6587T	G8857A	C10975T	C13981T	
11116_Pica8-Tarapaca	T146C!	A215G	455.1T	A3434G	G3483A	T10166C	G13708A	A15671G	C16278T!	
11118_Pica8-Tarapaca	T146C!	A215G	455.1T	T1673C	A3434G	G3483A	T10166C	A10876G	G13708A	A15671G
11121_Pica8-Tarapaca	T146C!	A215G	455.1T	A3434G	G3483A	T10166C	G13708A	A15671G	C16278T!	
10814_Lima	T2857C	T5082C	T5277C	A15924G	T16330C					
10802_Lima	C16168T									
10794_Ychsma	G8290A	A16066G								
10713_Ychsma	573.1C	A5186G	C5375T	C5895T	G15884A					
10729_Ychsma	T152C!	A2880G	C7786T	C13934T	A16051G	C16360T				
11200_Chancay	T2857C	T5277C	A6929C	A15924G	C16188T					
13990_Candelaria	A214G	356.1C	A6002G	T6896C	G8994A					
13989_Candelaria	315.1T	T4973C	G14569A							
13992_Candelaria	315.1T	A4829G	T4973C	T8749C	G14569A					
13994_Candelaria	315.1T	T504C	T4973C	T8540C	A8718G	C10751T	G16310A	A214G	T152C!	
10821_Lima	T146C!	T3786C	A8170G	G15777A						
12689_Lauricocha	A181G	C6056T	T6293C	C16295T						
11165_Pueblo Viejo	T152C!	G1393A	T2857C	A5134G	C5815T	G13813A	T14110C	A15924G		
11168_ Pueblo Viejo	C195T	292.1AT	T2857C	T5277C	A15924G	T16330C				
11173_ Pueblo Viejo	292.1AT	T2857C	T5277C	A15924G	T16330C					
BC12_Botigiriayocc	G185A	A8343G	T11383C	T13500C	C16168T					
LGA13_La Galgada	C19T	G228T	G709A	C8643T						
JA14_Jauranga	G75A	G709A	C16261T	G16319A						
M21_Los Molinos	G73C	G75A	G709A	T3593C	G5460A	C8278A	C16261T	G16319A		

B2b										
10800_Ychsma	T131C	C4013T	G13708A	T14634C	C14873A	T15784C				
10789_Lima	G6261A	G9055A	A12972G	A14053G						
10720_Ychsma	G6261A	G9055A	A14053G	C15647T						
10726_Ychsma	G143A	T9078C	G10530A	A14053G	G16390A					
10730_Ychsma	T131C	C4013T	G13708A	T14634C	T15784C					
10754_Wari	A243G	C4013T	G8994A	G11016A	G12127A	A14059G	G16438A			
10742_Wari	A178G	A6779G	A10420G	C16278T!	C16295T					
10750_Wari	T152C!	G1797A	A9377G	G12127A	A14053G	C14428T				
11182_Inca	G6305A	A14053G	C14428T							
11208_Chancay	G143A	G10530A	A14053G	A15901G						
11176_Pueblo Viejo	T131C	G1709A	C4013T	G13708A	T14634C	C14873A	T15784C			
Haplogroup C										
C1b										
10817_Lima	A9468G	C12535T	T15313C	A16166C	T16223C					
10820_Lima	A9468G	C12535T	T15313C	A16166C	T16223C					
10709_Ychsma	C12535T	T15313C								
10717_Ychsma	C150T	T152C!	C12535T	T15313C	C16292T					
10725_Ychsma	C12535T	T15313C								
10731_Ychsma	A636G	C3442T	C4496T	G5821A	A5894C	G16000A	T16362C			
10734_Wari	C12535T	T15313C	A16166C	G16244T						
10763_Wari	C12535T	T15313C								
10771_Wari	A9468G	C12535T	T15313C	A16166C						

10773_Wari	T8380C	T13281C	A13710C	A15244G	C15315T	A16207G				
6322_Llullaillaco	A1808G	C11389T	G12127A	A13966G	T15283C	T16189C!	C16298T!			
11211_Chancay	A5G	A214G	G103A	573.1C	C2393T	T7080C	T8952C	G11518A	T16126C	A16166d C16270T
11205_Chancay	T152C!	A769G	G3316A	A6125G	A6719G	A8289G	G8865A	T12631G	G13711A	G15431A
13993_Candelaria	A550T	G3369A	A8210G	T11383C	T15056C	G16129A!	T16357C			
12690_Botigiriayocc	T593C	A723C	A14706G	C16147A	C16287T					
10765_Wari	C12535T	T15313C	A16166C	G16244T						
11154_ Pueblo Viejo	290A!	C12870T								
11157_ Pueblo Viejo	C960d	C12870T	G5745A							
11161_ Pueblo Viejo	A636G	C4496T	G5821A	A5894C	G16000A	T16362C				
13279_Tiwanaku	T46G	T1005C	T5492C	C14428T	C14863T	T16092C	C16400T	A16463G		
13245_Tiwanaku	T204C	C5438T	A10987G	C13197T	G15323A					
HPA866_Huaca Prieta	G8856A	T16189C!								
C1c										
13241_Tiwanaku	T152C!	T4216C	A4965G	C5294T	T10604C	G10688A!	A12248G	T16381C		
12692_Palpa	T152C!	T16092C	A16212G	T16311T!						
11209_Chancay	T152C!	G3531A	G8701A	A10053G	G15106A	T16092C	C16298T!	T16311C!		
10806_Lima	T152C!	7472d	T8450C	C12774T	T16172C	T16297C				
11183_Pueblo Viejo	T152C!	T13500C	G15596A	T16362C						
C1d/C1d1										
SSN5_San Sebastian	T57C	T59C	T12384C	A15730G	T16327C					
10810_Ychsma	A14122G									

Haplogroup D										
D1										
11213_Chancay	A735G	C2263A	T6515C	A6872G	T11087C					
10805_Ychsma	T11365C	G11906A	T12481C	C12557T	A14665G					
10722_Ychsma	T11365C	G11906A	T12481C	C12557T	A14665G					
6320_Llullaillaco	C11665T	C16527T								
11180_Pueblo Viejo	C11665T	C16527T								
11185_Pueblo Viejo	C150T	G3421A	T9758C	A10739G	G15777A	A16166G				
ASOB9S14_ArroyoSeco	T146C!	T11545C	T13743C	T16304C	G16390A					
LGA2_La Galgada	G4769A	T7028C	G8860T	G9612A	A12358G	A12366G	C14215T	G15106A	A15301G	
D										
10811A_Lima	G143A	T2092C!	A6113G	T16189C!	G16274A	C16365T	D1h sister branch			
10732A_Ychsma	G143A	T2092C!	A6113G	T16189C!	G16274A	C16365T	D7 (new branch)			

Supplementary Table 3. List of Adaptors (Barcodes) P5, P7 used for genomic library preparation

ACAD	Name	Sequence
N1	Sol_adap_P5_1_BC1	A*C*A*C*TCTTTCCCTACACGACGCTCTTCCGATCTA*A*C*A*T
N2	Sol_adap_P5_1_rev_BC1	A*T*G*T*TAGATCGGA*A*G*A*G
N3	Sol_adap_P5_1_BC2	A*C*A*C*TCTTTCCCTACACGACGCTCTTCCGATCTA*A*G*A*G
N4	Sol_adap_P5_1_rev_BC2	C*T*C*T*TAGATCGGA*A*G*A*G
N5	Sol_adap_P5_1_BC3	A*C*A*C*TCTTTCCCTACACGACGCTCTTCCGATCTA*A*G*C*C
N6	Sol_adap_P5_1_rev_BC3	G*G*C*T*TAGATCGGA*A*G*A*G
N7	Sol_adap_P5_1_BC4	A*C*A*C*TCTTTCCCTACACGACGCTCTTCCGATCTA*A*G*T*T
N8	Sol_adap_P5_1_rev_BC4	A*A*C*T*TAGATCGGA*A*G*A*G
N9	Sol_adap_P5_1_BC5	A*C*A*C*TCTTTCCCTACACGACGCTCTTCCGATCTA*A*T*G*C
N10	Sol_adap_P5_1_rev_BC5	G*C*A*T*TAGATCGGA*A*G*A*G
N11	Sol_adap_P5_1_BC6	A*C*A*C*TCTTTCCCTACACGACGCTCTTCCGATCTA*C*A*C*T
N12	Sol_adap_P5_1_rev_BC6	A*G*T*G*TAGATCGGA*A*G*A*G
N13	Sol_adap_P5_1_BC7	A*C*A*C*TCTTTCCCTACACGACGCTCTTCCGATCTA*C*C*G*G
N14	Sol_adap_P5_1_rev_BC7	C*C*G*G*TAGATCGGA*A*G*A*G
N15	Sol_adap_P5_1_BC8	A*C*A*C*TCTTTCCCTACACGACGCTCTTCCGATCTA*C*G*C*A
N16	Sol_adap_P5_1_rev_BC8	T*G*C*G*TAGATCGGA*A*G*A*G
N17	Sol_adap_P5_1_BC9	A*C*A*C*TCTTTCCCTACACGACGCTCTTCCGATCTA*C*G*T*G
N18	Sol_adap_P5_1_rev_BC9	C*A*C*G*TAGATCGGA*A*G*A*G
N19	Sol_adap_P5_1_BC10	A*C*A*C*TCTTTCCCTACACGACGCTCTTCCGATCTA*C*T*A*A
N20	Sol_adap_P5_1_rev_BC10	T*T*A*G*TAGATCGGA*A*G*A*G
N21	Sol_adap_P5_1_BC11	A*C*A*C*TCTTTCCCTACACGACGCTCTTCCGATCTA*C*T*C*G
N22	Sol_adap_P5_1_rev_BC11	C*G*A*G*TAGATCGGA*A*G*A*G
N23	Sol_adap_P5_1_BC12	A*C*A*C*TCTTTCCCTACACGACGCTCTTCCGATCTA*C*T*G*T
N24	Sol_adap_P5_1_rev_BC12	A*C*A*G*TAGATCGGA*A*G*A*G
N73	Sol_adap_P5_1_BC13	A*C*A*C*TCTTTCCCTACACGACGCTCTTCCGATCTa*g*g*a*c
N74	Sol_adap_P5_1_rev_BC13	g*t*c*c*tAGATCGGA*A*G*A*G
N75	Sol_adap_P5_1_BC14	A*C*A*C*TCTTTCCCTACACGACGCTCTTCCGATCTa*g*t*c*a
N76	Sol_adap_P5_1_rev_BC14	t*g*a*c*tAGATCGGA*A*G*A*G
N77	Sol_adap_P5_1_BC15	A*C*A*C*TCTTTCCCTACACGACGCTCTTCCGATCTa*g*t*g*g
N78	Sol_adap_P5_1_rev_BC15	c*c*a*c*tAGATCGGA*A*G*A*G
N79	Sol_adap_P5_1_BC16	A*C*A*C*TCTTTCCCTACACGACGCTCTTCCGATCTa*t*a*c*g
N80	Sol_adap_P5_1_rev_BC16	c*g*t*a*tAGATCGGA*A*G*A*G
N81	Sol_adap_P5_1_BC17	A*C*A*C*TCTTTCCCTACACGACGCTCTTCCGATCTa*t*a*g*c
N82	Sol_adap_P5_1_rev_BC17	g*c*t*a*tAGATCGGA*A*G*A*G
N83	Sol_adap_P5_1_BC18	A*C*A*C*TCTTTCCCTACACGACGCTCTTCCGATCTa*t*c*a*c

ACAD	Name	Sequence
N84	Sol_adap_P5_1_rev_BC18	g*t*g*a*tAGATCGGA*A*G*A*G
N85	Sol_adap_P5_1_BC19	A*C*A*C*TCTTTCCCTACACGACGCTCTTCCGATCTa*t*g*t*a
N86	Sol_adap_P5_1_rev_BC19	t*a*c*a*tAGATCGGA*A*G*A*G
N87	Sol_adap_P5_1_BC20	A*C*A*C*TCTTTCCCTACACGACGCTCTTCCGATCTa*t*t*c*c
N88	Sol_adap_P5_1_rev_BC20	g*g*a*a*tAGATCGGA*A*G*A*G
N89	Sol_adap_P7_2_BC1	G*T*G*A*CTGGAGTTCAGACGTGTGCTCTTCCGATCTa*t*t*g*a
N90	Sol_adap_P7_rev_2_BC1	t*c*a*a*tAGATCGGA*A*G*A*G
N91	Sol_adap_P7_2_BC2	G*T*G*A*CTGGAGTTCAGACGTGTGCTCTTCCGATCTc*a*a*t*t
N92	Sol_adap_P7_rev_2_BC2	a*a*t*t*gAGATCGGA*A*G*A*G
N93	Sol_adap_P7_2_BC3	G*T*G*A*CTGGAGTTCAGACGTGTGCTCTTCCGATCTc*a*t*c*t
N94	Sol_adap_P7_rev_2_BC3	a*g*a*t*gAGATCGGA*A*G*A*G
N95	Sol_adap_P7_2_BC4	G*T*G*A*CTGGAGTTCAGACGTGTGCTCTTCCGATCTc*a*t*t*c
N96	Sol_adap_P7_rev_2_BC4	g*a*a*t*gAGATCGGA*A*G*A*G
N97	Sol_adap_P7_2_BC5	G*T*G*A*CTGGAGTTCAGACGTGTGCTCTTCCGATCTc*c*g*a*g
N98	Sol_adap_P7_rev_2_BC5	c*t*c*g*gAGATCGGA*A*G*A*G
N99	Sol_adap_P5_1	A*C*A*C*TCTTTCCCTACACGACGCTCTTCCG*A*T*C*T
N100	Sol_adap_P5_1_rev	A*G*A*T*CGGA*A*G*A*G
N101	P5_short_RNAblock	ACACUCUUUCCCUACACGAC
N102	P7_short_RNAblock	GUGACUGGAGUUCAGACGUGU
N103	Sol_adap_P5_1_BC21	A*C*A*C*TCTTTCCCTACACGACGCTCTTCCGATCTc*g*g*t*t
N104	Sol_adap_P5_1_rev_BC21	a*a*c*c*gAGATCGGA*A*G*A*G
N105	Sol_adap_P5_1_BC22	A*C*A*C*TCTTTCCCTACACGACGCTCTTCCGATCTg*a*a*g*c
N106	Sol_adap_P5_1_rev_BC22	g*c*t*t*cAGATCGGA*A*G*A*G
N107	Sol_adap_P5_1_BC23	A*C*A*C*TCTTTCCCTACACGACGCTCTTCCGATCTg*t*a*c*c
N108	Sol_adap_P5_1_rev_BC23	g*g*t*a*cAGATCGGA*A*G*A*G
N109	Sol_adap_P5_1_BC24	A*C*A*C*TCTTTCCCTACACGACGCTCTTCCGATCTt*a*c*t*g
N110	Sol_adap_P5_1_rev_BC24	c*a*g*t*aAGATCGGA*A*G*A*G
N111	Sol_adap_P5_1_BC25	ACACTCTTTCCCTACACGACGCTCTTCCGATCTtgtgc
N112	Sol_adap_P5_1_rev_BC25	g*c*a*c*aAGATCGGA*A*G*A*G
N113	Sol_adap_P7_2_BC6	G*T*G*A*CTGGAGTTCAGACGTGTGCTCTTCCGATCTc*t*c*g*a
N114	Sol_adap_P7_rev_2_BC6	t*c*g*a*gAGATCGGA*A*G*A*G
N115	Sol_adap_P7_2_BC7	G*T*G*A*CTGGAGTTCAGACGTGTGCTCTTCCGATCTg*c*g*t*t
N116	Sol_adap_P7_rev_2_BC7	a*a*c*g*cAGATCGGA*A*G*A*G
N117	Sol_adap_P7_2_BC8	G*T*G*A*CTGGAGTTCAGACGTGTGCTCTTCCGATCTt*c*g*c*g
N118	Sol_adap_P7_rev_2_BC8	c*g*c*g*aAGATCGGA*A*G*A*G
N119	Sol_adap_P7_2_BC9	G*T*G*A*CTGGAGTTCAGACGTGTGCTCTTCCGATCTa*c*g*c*a
N120	Sol_adap_P7_rev_2_BC9	t*g*c*g*tAGATCGGA*A*G*A*G
N121	Sol_adap_P5_BC26	ACACTCTTTCCCTACACGACGCTCTTCCGATCTcgaat

ACAD	Name	Sequence
N122	Sol_adap_P5_rev_BC26	attcgAGATCGGAAGAG
N123	Sol_adap_P5_BC27	ACACTCTTTCCCTACACGACGCTCTTCCGATCTcgagg
N124	Sol_adap_P5_rev_BC27	cctcgAGATCGGAAGAG
N125	Sol_adap_P5_BC28	ACACTCTTTCCCTACACGACGCTCTTCCGATCTcgatc
N126	Sol_adap_P5_rev_BC28	gatcgAGATCGGAAGAG
N127	Sol_adap_P5_BC29	ACACTCTTTCCCTACACGACGCTCTTCCGATCTcgcac
N128	Sol_adap_P5_rev_BC29	gtgcgAGATCGGAAGAG
N129	Sol_adap_P5_BC30	ACACTCTTTCCCTACACGACGCTCTTCCGATCTcgcgt
N130	Sol_adap_P5_rev_BC30	acgcgAGATCGGAAGAG
N131	Sol_adap_P5_BC31	ACACTCTTTCCCTACACGACGCTCTTCCGATCTcgttg
N132	Sol_adap_P5_rev_BC31	caacgAGATCGGAAGAG
N133	Sol_adap_P5_BC32	ACACTCTTTCCCTACACGACGCTCTTCCGATCTctaag
N134	Sol_adap_P5_rev_BC32	cttagAGATCGGAAGAG
N135	Sol_adap_P5_BC33	ACACTCTTTCCCTACACGACGCTCTTCCGATCTctctc
N136	Sol_adap_P5_rev_BC33	gagagAGATCGGAAGAG
N137	Sol_adap_P5_BC34	ACACTCTTTCCCTACACGACGCTCTTCCGATCTctgcc
N138	Sol_adap_P5_rev_BC34	ggcagAGATCGGAAGAG
N139	Sol_adap_P5_BC35	ACACTCTTTCCCTACACGACGCTCTTCCGATCTctggt
N140	Sol_adap_P5_rev_BC35	accagAGATCGGAAGAG
N141	Sol_adap_P5_BC36	ACACTCTTTCCCTACACGACGCTCTTCCGATCTcttgg
N142	Sol_adap_P5_rev_BC36	ccaagAGATCGGAAGAG
N143	Sol_adap_P5_BC37	ACACTCTTTCCCTACACGACGCTCTTCCGATCTgactt
N144	Sol_adap_P5_rev_BC37	aagtcAGATCGGAAGAG
N145	Sol_adap_P5_BC38	ACACTCTTTCCCTACACGACGCTCTTCCGATCTgagca
N146	Sol_adap_P5_rev_BC38	tgctcAGATCGGAAGAG
N147	Sol_adap_P5_BC39	ACACTCTTTCCCTACACGACGCTCTTCCGATCTgatac
N148	Sol_adap_P5_rev_BC39	gtatcAGATCGGAAGAG
N149	Sol_adap_P5_BC40	ACACTCTTTCCCTACACGACGCTCTTCCGATCTgatga
N150	Sol_adap_P5_rev_BC40	tcatcAGATCGGAAGAG
N151	Sol_adap_P5_BC41	ACACTCTTTCCCTACACGACGCTCTTCCGATCTgccat
N152	Sol_adap_P5_rev_BC41	atggcAGATCGGAAGAG
N153	Sol_adap_P5_BC42	ACACTCTTTCCCTACACGACGCTCTTCCGATCTgccta
N154	Sol_adap_P5_rev_BC42	taggcAGATCGGAAGAG
N155	Sol_adap_P5_BC43	ACACTCTTTCCCTACACGACGCTCTTCCGATCTgctcc
N156	Sol_adap_P5_rev_BC43	ggagcAGATCGGAAGAG
N157	Sol_adap_P5_BC44	ACACTCTTTCCCTACACGACGCTCTTCCGATCTgctgg
N158	Sol_adap_P5_rev_BC44	ccagcAGATCGGAAGAG
N159	Sol_adap_P5_BC45	ACACTCTTTCCCTACACGACGCTCTTCCGATCTggaac

ACAD	Name	Sequence
N160	Sol_adap_P5_rev_BC45	gttccAGATCGGAAGAG
N161	Sol_adap_P5_BC46	ACACTCTTTCCCTACACGACGCTCTTCCGATCTggatg
N162	Sol_adap_P5_rev_BC46	catccAGATCGGAAGAG
N163	Sol_adap_P5_BC47	ACACTCTTTCCCTACACGACGCTCTTCCGATCTggcag
N164	Sol_adap_P5_rev_BC47	ctgccAGATCGGAAGAG
N165	Sol_adap_P5_BC48	ACACTCTTTCCCTACACGACGCTCTTCCGATCTggtgt
N166	Sol_adap_P5_rev_BC48	acaccAGATCGGAAGAG
N167	Sol_adap_P5_BC49	ACACTCTTTCCCTACACGACGCTCTTCCGATCTgtagt
N168	Sol_adap_P5_rev_BC49	actacAGATCGGAAGAG
N169	Sol_adap_P5_BC50	ACACTCTTTCCCTACACGACGCTCTTCCGATCTgtata
N170	Sol_adap_P5_rev_BC50	tatacAGATCGGAAGAG
N171	Sol_adap_P5_BC51	ACACTCTTTCCCTACACGACGCTCTTCCGATCTgtcaa
N172	Sol_adap_P5_rev_BC51	ttgacAGATCGGAAGAG
N173	Sol_adap_P5_BC52	ACACTCTTTCCCTACACGACGCTCTTCCGATCTgtcct
N174	Sol_adap_P5_rev_BC52	aggacAGATCGGAAGAG
N175	Sol_adap_P5_BC53	ACACTCTTTCCCTACACGACGCTCTTCCGATCTgtctg
N176	Sol_adap_P5_rev_BC53	cagacAGATCGGAAGAG
N177	Sol_adap_P5_BC54	ACACTCTTTCCCTACACGACGCTCTTCCGATCTgtgag
N178	Sol_adap_P5_rev_BC54	ctcacAGATCGGAAGAG
N179	Sol_adap_P5_BC55	ACACTCTTTCCCTACACGACGCTCTTCCGATCTgtgga
N180	Sol_adap_P5_rev_BC55	tccacAGATCGGAAGAG
N181	Sol_adap_P5_BC56	ACACTCTTTCCCTACACGACGCTCTTCCGATCTgttat
N182	Sol_adap_P5_rev_BC56	ataacAGATCGGAAGAG
N183	Sol_adap_P5_BC57	ACACTCTTTCCCTACACGACGCTCTTCCGATCTgttca
N184	Sol_adap_P5_rev_BC57	tgaacAGATCGGAAGAG
N185	Sol_adap_P5_BC58	ACACTCTTTCCCTACACGACGCTCTTCCGATCTtaaga
N186	Sol_adap_P5_rev_BC58	tcttaAGATCGGAAGAG
N187	Sol_adap_P5_BC59	ACACTCTTTCCCTACACGACGCTCTTCCGATCTtagct
N188	Sol_adap_P5_rev_BC59	agctaAGATCGGAAGAG
N189	Sol_adap_P5_BC60	ACACTCTTTCCCTACACGACGCTCTTCCGATCTtatag
N190	Sol_adap_P5_rev_BC60	ctataAGATCGGAAGAG
N191	Sol_adap_P5_BC61	ACACTCTTTCCCTACACGACGCTCTTCCGATCTtatcc
N192	Sol_adap_P5_rev_BC61	ggataAGATCGGAAGAG
N193	Sol_adap_P5_BC62	ACACTCTTTCCCTACACGACGCTCTTCCGATCTtcatt
N194	Sol_adap_P5_rev_BC62	aatgaAGATCGGAAGAG
N195	Sol_adap_P5_BC63	ACACTCTTTCCCTACACGACGCTCTTCCGATCTtccag
N196	Sol_adap_P5_rev_BC63	ctggaAGATCGGAAGAG
N197	Sol_adap_P5_BC64	ACACTCTTTCCCTACACGACGCTCTTCCGATCTtcgac

ACAD	Name	Sequence
N198	Sol_adap_P5_rev_BC64	gtcgaAGATCGGAAGAG
N199	Sol_adap_P5_BC65	ACACTCTTTCCCTACACGACGCTCTTCCGATCTtcgga
N200	Sol_adap_P5_rev_BC65	tccgaAGATCGGAAGAG
N201	Sol_adap_P5_BC66	ACACTCTTTCCCTACACGACGCTCTTCCGATCTtgaag
N202	Sol_adap_P5_rev_BC66	cttcaAGATCGGAAGAG
N203	Sol_adap_P5_BC67	ACACTCTTTCCCTACACGACGCTCTTCCGATCTtgacc
N204	Sol_adap_P5_rev_BC67	ggtcaAGATCGGAAGAG
N205	Sol_adap_P5_BC68	ACACTCTTTCCCTACACGACGCTCTTCCGATCTtgata
N206	Sol_adap_P5_rev_BC68	tatcaAGATCGGAAGAG
N207	Sol_adap_P5_BC69	ACACTCTTTCCCTACACGACGCTCTTCCGATCTtggat
N208	Sol_adap_P5_rev_BC69	atccaAGATCGGAAGAG
N209	Sol_adap_P5_BC70	ACACTCTTTCCCTACACGACGCTCTTCCGATCTtggtc
N210	Sol_adap_P5_rev_BC70	gaccaAGATCGGAAGAG
N211	Sol_adap_P5_BC71	ACACTCTTTCCCTACACGACGCTCTTCCGATCTtgtcg
N212	Sol_adap_P5_rev_BC71	cgacaAGATCGGAAGAG
N213	Sol_adap_P5_BC72	ACACTCTTTCCCTACACGACGCTCTTCCGATCTttact
N214	Sol_adap_P5_rev_BC72	agtaaAGATCGGAAGAG
N215	Sol_adap_P5_BC73	ACACTCTTTCCCTACACGACGCTCTTCCGATCTttagg
N216	Sol_adap_P5_rev_BC73	cctaaAGATCGGAAGAG

Supplementary Table 4. List of Indexed Illumina primers used for genomic library preparation

ACAD #	Name	Sequence
N25	GAII_Indexing_1	CAAGCAGAAGACGGCATACGAGAT**CCTGCGA**GTGACTGGAGTTCAGACGTGT
N26	GAII_Indexing_2	CAAGCAGAAGACGGCATACGAGAT**TGCAGAG**GTGACTGGAGTTCAGACGTGT
N27	GAII_Indexing_3	CAAGCAGAAGACGGCATACGAGAT**ACCTAGG**GTGACTGGAGTTCAGACGTGT
N28	GAII_Indexing_4	CAAGCAGAAGACGGCATACGAGAT**TTGATCC**GTGACTGGAGTTCAGACGTGT
N29	GAII_Indexing_5	CAAGCAGAAGACGGCATACGAGAT**ATCTTGC**GTGACTGGAGTTCAGACGTGT
N30	GAII_Indexing_6	CAAGCAGAAGACGGCATACGAGAT**TCTCCAT**GTGACTGGAGTTCAGACGTGT
N31	GAII_Indexing_7	CAAGCAGAAGACGGCATACGAGAT**CATCGAG**GTGACTGGAGTTCAGACGTGT
N32	GAII_Indexing_8	CAAGCAGAAGACGGCATACGAGAT**TTCGAGC**GTGACTGGAGTTCAGACGTGT
N33	GAII_Indexing_9	CAAGCAGAAGACGGCATACGAGAT**AGTTGGT**GTGACTGGAGTTCAGACGTGT
N34	GAII_Indexing_10	CAAGCAGAAGACGGCATACGAGAT**GTACCGG**GTGACTGGAGTTCAGACGTGT
N35	GAII_Indexing_11	CAAGCAGAAGACGGCATACGAGAT**CGGAGTT**GTGACTGGAGTTCAGACGTGT
N36	GAII_Indexing_12	CAAGCAGAAGACGGCATACGAGAT**ACTTCAA**GTGACTGGAGTTCAGACGTGT
N37	GAII_Indexing_13	CAAGCAGAAGACGGCATACGAGAT**TGATAGT**GTGACTGGAGTTCAGACGTGT
N38	GAII_Indexing_14	CAAGCAGAAGACGGCATACGAGAT**GATCCAA**GTGACTGGAGTTCAGACGTGT
N39	GAII_Indexing_15	CAAGCAGAAGACGGCATACGAGAT**CAGGTCG**GTGACTGGAGTTCAGACGTGT

Supplementary Table 5. List of Indexed Illumina short primers used for genomic library preparation

ACAD	Name	Sequence		
N56	IS1_adapter.P5	A*C*A*C*TCTTTCCCTACACGACGCTCTTCCG*A*T*C*T	HPLC	IDT
N57	IS2_adapter.P7	G*T*G*A*CTGGAGTTCAGACGTGTGCTCTTCCG*A*T*C*T	HPLC	IDT
N58	IS3_adapter.P5+P7	A*G*A*T*CGGAA*G*A*G*C	HPLC	IDT
N59	**IS4_indPCR.P5**	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT	HPLC	IDT
N60	IS5_reamp.P5	AATGATACGGCGACCACCGA	Desalt	IDT
N61	IS6_reamp.P7	CAAGCAGAAGACGGCATACGA	Desalt	IDT
N64	**IS7_short_amp.P5**	ACACTCTTTCCCTACACGAC	Desalt	IDT
N65	**IS8_short_amp.P7**	GTGACTGGAGTTCAGACGTGT	Desalt	IDT

Supplementary Table 6. Radiocarbon dates from selected ancient samples from South America

ACAD	site	culture	relative date	hg	ORAU#	Delta 13 C	uncal BP	IntCal13 cal BC/AD	2-sigma	SH13Cal13
10791	Huaca Pucllana	Lima	100-650 AD	A2	OxA-31118	-11.79	1420±29	584-660 calAD (95.4%)	584-660	603-760 cal AD (95.4%)
10817	Huaca Pucllana	Lima	100-650 AD	C1b	OxA-31120	-13.8	1493±30	534-642 calAD (90.6%)	435-642	549-652 cal AD (95.4%)
10734	Huaca Pucllana	Wari	500-1000 AD	C1b	OxA-31422	-12.65	1156±22	776-968 calAD (95.4%)	776-968	891-988 cal AD (95.4%)
10754	Huaca Pucllana	Wari	500-1000 AD	B2b	OxA-31423	-12.61	955±65	974-1220 calAD (95.4%)	974-1220	1016-1264 cal AD (95.4%)
10709	Huaca Pucllana	Ychsma	1100-1440 AD	C1b	OxA-31424	-9.6	745±23	1244-1288 calAD (92.9%)	1226-1288	1271-1315 calAD (72.9%)
10805	Huaca Pucllana	Ychsma	100-650 AD	D1	OxA-31462	-13,38	762±23	1223 - 1280 calAD (95.4%)	1223-1280	1231-1379 cal AD (95.3%)
10810	Huaca Pucllana	Ychsma	100-650 AD	C1d	OxA-31119	-12.18	866±28	1149-1249 calAD (80%)	1048-1249	1164-1272 cal AD (95.4%)
10722	Huaca Pucllana	Ychsma	1100-1440 AD	D1	OxA-31425	-12.49	773±24	1221-1278 calAD (95.4%)	1221-1278	1227-1301 calAD (93.9%)
13241	Tiwanaku	Tiwanaku	724-1187 AD	C1c	OxA-31463	-17,59	1056±23	901 - 1024 calAD (95.4%)	901-1024	954 - 1024 calAD (88.4%)
13264	Tiwanaku	Tiwanaku	724-1187 AD	B2	OxA-31443	-12,39	969±28	1017 - 1155 calAD (95.4%)	1017-1155	1032 – 1180 cal AD (95.4%)
10891	Patagonia Chile			B2b	OxA-31464	-19,41	3849±28	2158 - 2268 calBC (74.8%)	2458-2206	2260 - 2206 calBC (20.6%)

CHAPTER 4

Ancient DNA analysis from the ‘Huaca Pucllana’ archaeological site in Central coastal Peru: Chronological study of mitochondrial DNA variation in the context of the Wari Empire expansion during the Middle Horizon

Statement of Authorship

Chapter 4

Title of the Paper	**Ancient DNA analysis from the Huaca Pucllana archaeological site in Central coastal Peru: Chronological study of mitochondrial DNA variation in the context of the Wari Empire expansion during the Middle Horizon**
Publication status	Published, Accepted for Publication, Submitted for Publication, **Publication style**
Publication details	

Author Contributions
By signing the Statement of Authorship, each author certifies that their stated contribution to the publication is accurate and that permission is granted for the publication to be included in the candidate's thesis.

Name of the principal Author	**Guido Valverde** (Candidate)	
Contribution to the paper	Performed DNA extractions, PCR amplifications, sequencing and downstream data processing on all samples, analysed the data, interpreted the results, created figures and tables and wrote the manuscript.	
Signature		Date: 8/06/15

Name of the co-author	Maria Inés Barreto	
Contribution to the paper	Contributed archaeological material. Assisted with sample collection, provided archaeological and context information.	
Signature		Date: May 29 2015

Name of the co-author	Isabel Flores Espinoza	
Contribution to the paper	Contributed archaeological material. Assisted with sample collection, provided archaeological and context information.	
Signature		Date 02.06.2015

Name of the co-author	Alan Cooper	
Contribution to the paper	Contributed reagents, materials, analysis tools, and assisted with the interpretation of results.	
Signature		Date: October, 3 2015

Name of the co-author	Lars Fehren-Schmitz	
Contribution to the paper	Contributed reagents, materials, analysis tools, assisted in the interpretation of results.	
Signature		Date: 06/01/2015

Name of the co-author	Bastien Llamas	
Contribution to the paper	Co-developed the research concept and experimental design of the study, provided archaeological samples, gave advice on laboratory work, assisted with the interpretation of results and edited the manuscript	
Signature		Date: October, 23 2015

Name of the co-author	Wolfgang Haak	
Contribution to the paper	Co-developed the research concept and experimental design of the study, provided archaeological samples, gave advice on laboratory work, assisted with the interpretation of results and edited the manuscript	
Signature		Date: 20/10/2015

Ancient DNA analysis from the Huaca Pucllana archaeological site in Central coastal Peru: Chronological study of mitochondrial DNA variation in the context of the Wari Empire expansion during the Middle Horizon

Guido Valverde[1], María Inés Barreto[2], Isabel Flores Espinoza[2], Alan Cooper[1], Lars Fehren-Schmitz[3,4], Bastien Llamas[1], Wolfgang Haak[1]

[1] Australian Centre for Ancient DNA, School of Biological Sciences, University of Adelaide, Adelaide, South Australia 5005, Australia

[2] Proyecto Arqueológico Huaca Pucllana, Lima, Peru

[3] Department of Anthropology, University of California, Santa Cruz, CA 95064

[4] Historical Anthropology and Human Ecology, University of Goettingen, D-37073 Goettingen, Germany

ABSTRACT

The analysis of ancient human DNA from South American sites represents a unique opportunity to directly explore population history and test hypotheses about cultural transitions and demographic events in pre-Columbian times. The Middle Horizon (600–1000 AD) represents a particular transitional period harbouring a series of interregional interaction and socio-political changes associated with the development and expansion of ancient Andean empires such as the Wari and Tiwanaku, which have played important roles in demographic and cultural processes in Central South America.

We analysed three successive pre-Columbian cultures – Lima (Early Intermediate Period, 200–600 AD), Wari (Middle Horizon, 600–1000 AD) and Ychsma (Late Intermediate Period, 1000–1476 AD) - at the Huaca Pucllana archaeological site in Lima, Peru. We successfully sequenced 34 complete mitochondrial genomes to investigate population continuity during the Middle Horizon, and the potential genetic impact of the Wari Empire in the Central coast of Peru.

Mitochondrial haplogroup frequencies revealed an increase of haplogroup C1b during the Wari occupation in the Middle Horizon and a considerable decrease of haplogroup B2 in samples from the Wari culture. Thus, our results suggest demographic changes in the Central coast of Peru during the Middle Horizon. However, we caution that the impact of Wari imperialism and genetic influx on this particular archaeological site at Huaca Pucllana might not be representative for the entire Wari territory in the Peruvian Central coast.

KEYWORDS: ancient DNA, Huaca Pucllana, Middle Horizon, Wari Empire, Wari Imperialism, Central Peru

1. INTRODUCTION

Human dispersal into the Americas has been studied through diverse contexts and research disciplines. In particular, genetic research has been highly informative in explaining long-standing questions concerning initial and post-settlement population movements within the continents (Reich et al. 2012; Raghavan et al. 2014; Rasmussen et al. 2014; Fehren-Schmitz et al. 2015; Raghavan et al. 2015; Rasmussen et al. 2015; Skoglund et al. 2015). However, most of this evidence must be contrasted or described under a multidisciplinary scenario in order to clarify remaining uncertainties (O'Rourke and Raff 2010).

The Central Andes of South America represent a particular region of interest, defined by marked geographic contrasts between the Andean highlands and the Pacific coast, with each of the two regions characterized by its own combination of ecology and culture. Archaeological and genetic studies that contrasted population movements and settlements of the original groups within South America, have proposed a colonization event time in the continent around 15,000 – 13,5000 years BP (Dillehay 1997; Fuselli et al. 2003; Raghavan et al. 2015). This region is also closely associated with the history and interaction of iconic ancient civilizations developed before the European arrival. The rise and fall of pre-Columbian empires in South America such as the Wari, Tiwanaku and Inca (Williams 2002; Isbell 2008; Kemp et al. 2009), have left a complex structure of societies concurrent with ecological changes, populations dynamics, and cultural transitions (Williams 2002).

Archaeological records in the Central Andes have defined the Middle Horizon period (600–1000 AD) as a time of development in population dynamics, involving inter-regional interaction between highland and coastal populations, social stratification and socio-political changes (Isbell and Schreiber 1978; Schreiber 1992; Tung 2007; Isbell 2008; Jennings 2010; Marcone 2010). Cultural transitions are thought to be related to migration which correlates the social complexity of the Wari and Tiwanaku expansion and its influence over a large area in the Central Andes during the Middle Horizon (Isbell and Schreiber 1978; Shady and Ruiz 1979; Lumbreras 1980; Isbell 2008; Marcone 2010).

The Wari of the Middle Horizon

The Wari Empire is considered the first imperial state in South America, which had expanded from the Ayacucho heartland (i.e. Wari Capital city), the political and urban centre located in Central highland Peru, eventually covering a vast area of the Peruvian Andes (Isbell and Schreiber 1978; Glowacki 2012; Schreiber 2012) (Figure 1). The main features of the Wari hegemony are characterized by a well established military domination as a strategy of continuous conquests and colonization based on remarkable urban development of "control cities" and administrative centres (Isbell and McEwan 1991; Schreiber 1992).

Despite contact and interaction with the Tiwanaku Empire (centred in modern-day Bolivia), which also expanded in the Central Andes during the Middle Horizon (Goldstein 1989), the Wari and Tiwanaku should be considered as two contemporary but largely independent states that shared common roots and belief systems (Williams 2001; Williams and Isla 2002).

Moreover, Tiwanaku and Wari are also considered as two different political entities, and both had reflected their hegemony in the Central Andes due to expansionist characteristics (Menzel 1964; Lumbreras 1980). The Tiwanaku controlled the area around the Lake Titicaca basin and the Wari extended over most of the Central Andes toward the North coastal sites in Peru (Owen 1994), while the administrative centre at Cerro Baúl in the Moquegua Valley in Southern Peru is the only known frontier settlement occupied by both states (Goldstein 1989; Williams 2001) (Figure 1).

The Wari military strategies for expansion (Isbell 2008; Tung 2008) must have affected the population structure, and triggered genetic changes and migratory events in the area. The Wari state as a political entity seems to have been prevalent in many areas and held a dominant position in administrative centres such as Viracochapampa, Willkawaín, Castillo de Huarmey in the Northern region of Peru, Honcopampa, Jincamocco Azangaro, Cerro Baúl in Moquegua, and Pikillacta in Cuzco (Schreiber 1992; Williams and Isla 2002; Jennings 2010).

The Wari expansion during the Middle Horizon is also termed "Wari Imperialism", which describes the development of Wari culture and its wide-ranging influence in ancient Peru (Isbell and Schreiber 1978; Earle and Jennings 2012; Glowacki 2012). To understand the Wari phenomenon, different models have been proposed to explain the Wari expansionist process (Lumbreras 1969) alongside other kind of interregional interactions (Shady 1988).

The Wari imperial control can be understood as a complex socio-political authority that was coordinated from the political heartland in Ayacucho. The Wari imperialistic expansion has been described as a combination of religious indoctrination (Menzel 1964) and/or military campaigns (Lumbreras 1974) with strategies for achieving imperial dominance over other groups (Tung 2012). Relocation of groups that constituted the workforce, the construction of administrative centres according to defined architectural standards (Isbell and McEwan 1991), a Wari road network, and an extensive administrative system are considered as distinctive features of the Wari dominium in the Middle Horizon (Lumbreras 1974; Schreiber 1992; Williams and Isla 2002).

By investigating the impact of the Wari civilization over large territories in Peru, archaeological research has helped to understand the nature of the Wari expansion (Jennings 2010). In a general context, much of the debate regarding the development and characteristics of the Wari occupation/expansion during the Middle Horizon is dual. On the one hand, the Wari are a hegemonic imperialist state with presence in the Central coastal Peru (Menzel 1964; Lumbreras 1969; Marcone 2010; Segura and Shimada 2010). Alternatively, Central coastal cities were commercial states that interacted with the Wari people, without being politically dominated or assimilated (Shady 1982; Shady 1988). In the particular case of Wari interactions in Central coastal Peru, the main characteristics are the so-called Imperial architecture and funerary patterns not found in other Wari settlements elsewhere in Peru (McEwan 1990). Although archaeological data offers highly relevant information, so far no direct genetic research was attempted to explore the impact of Wari imperialism in Central coastal Peru, and the consequences of this internal form of colonialism in pre-Columbian times.

Genetic studies restricted to a particular geographic region have helped to understand micro-evolutionary processes involved in the genetic composition of ancient populations and their interaction with a particular environment (Moraga et al. 2005; Mendisco et al. 2014). In addition, ancient DNA (aDNA) research can provide an important temporal perspective when contrasting pre-historical genetic diversity with that of present-day populations from the same geographic area to evaluate population movements and test for population continuity (Moraga et al. 2005; Mendisco et al. 2014). Moreover, by incorporating other factors such as climatic variability in the

Andean and coastal regions from Peru, aDNA studies can help understanding demography and cultural transitions over time (Fehren-Schmitz et al. 2014).

Ancient DNA research has also incorporated temporal sampling, analysing individuals coming from different archaeological periods to study genetic changes over time transects in South America (Fehren-Schmitz et al. 2010; Fehren-Schmitz et al. 2011). Although, sample size and accession to specimens can be limited, the results are essential in identifying fluctuations and genetic transitions across time periods (Moraga et al. 2005; Kemp et al. 2009; Fehren-Schmitz et al. 2011).

For example, a study of the Palpa region in South coastal Peru, which incorporated diachronic sampling ranging from the Late Archaic Period (8000–1900 BC) to the Late Intermediate Period (1000–1476 AD), suggested population continuity at the coast until the Late Intermediate Period (Fehren-Schmitz et al. 2009; Fehren-Schmitz et al. 2010). Another diachronic study of ancient Peruvians has shown a differentiation between coastal and highland populations during the Middle Horizon (600–100 AD) (Fehren-Schmitz et al. 2011).

In this study, we explored the genetic and temporal population structure at the Huaca Pucllana archaeological site and tested for signs of population discontinuity especially related to the Wari expansion during the Middle Horizon, contrasting genetic profiles with archaeological contexts (e.g. material culture, funerary patterns and other characteristic of the people buried during the Wari occupation), which might differ from the preceding Lima occupation and the subsequent Ychsma culture.

The groups at Huaca Pucllana represent a particular diachronic cultural transect in a restricted geographic area (modern-day Lima) comparable to other important settlements associated with Wari expansion in the same region, such as Pachacamac, Cajamarquilla, Huaca San Marcos, Ancon, Huallamarca, and Catalina Huanca (Segura and Shimada 2010).

We used high-resolution genomic data from complete mtDNA genomes to characterise the genetic composition across a time transect consisting of three successive cultural periods: the Lima culture from the Early Intermediate Period - EIP (200–600 AD), the Wari culture from the Middle Horizon - MH (600–1000 AD), and the Ychsma culture from the Late Intermediate Period - LIP (1000–1476 AD). We further analysed the genetic signature of Huaca Pucllana cultures in the broader context of South American mtDNA past and present genetic diversity.

2. MATERIAL AND METHODS

2.1. Sample description and archaeological context

The archaeological site of Huaca Pucllana (12°06'37.01''S – 77°01'58.93''W) is located in the Miraflores District of Lima, Peru. A total of 115 samples from 52 individuals were collected at the 'Museo del Sitio Huaca Pucllana' (SI Table 1). We followed standard aDNA sampling guidelines to collect at least two independent samples per individual for replication and authentication purposes (Brown 1998; Yang and Watt 2005).

Samples come from three successive cultures, Lima (n=35), Wari (n=47) and Ychsma (n=33) (Lanning and Pollard 1967; Flores 1981). There is evidence at the Huaca Pucllana site that three layers developed during a slightly different time frame than the classical chronology established by (Rowe 1962). The Lima (500–700 AD) occupied the site during the late EIP and early MH. The Wari established as a state in Ayacucho ~600 AD but could have reached the Central coast and developed at Huaca Pucllana around 800–1000 AD. Finally, the time of the Ychsma (1000–1400 AD) is concordant with the classical chronology.

Despite rough environmental conditions in the coastal region of Peru, the development of the Lima culture suggests a "communal hub", which in this context refers to a highly important location/centre in the Central coast during the Early Horizon Period (900 BC–200 AD) (Rowe 1962; Lumbreras 1969), with a strong development of monumental architecture as seen in the construction of ceremonial temples, e.g. the Pachacamac Temple (Kaulicke 2000; Marcone 2001). The expansion of the Lima culture is described by its characteristic architecture, textiles, art style and ceramics found throughout the coastal valleys (Patterson 1966). Large structures and settlements could be found across the region dating to the Early Intermediate Period that demonstrate dominance of the Lima culture in the occupied territories (Marcone 2010).

The Lima period was also characterised by political and religious social elite who was responsible for the administration of different tasks in the administrative centres of the coastal valleys (Flores 2005). The religious elites are an important aspect of the Lima civilisation, which becomes evident from more than 20 settlements with many religious buildings such as temples and ceremonial centres (i.e. 'Huacas'), during the Early Intermediate Period and part of the Middle Horizon (Marcone 2010).

Many archaeological sites from the Lima civilization are still preserved in the modern city of Lima today, albeit threatened by the expansion of nearby urban developments (Marcone 2010; Segura and Shimada 2010). The main archaeological sites in coastal Lima are Huaca Pucllana and Pachacamac (of which the latter is the most important centre of the greater Lima region), that were maintained as ceremonial temples even after the fall of the civilization (Kaulicke 2000).

Huaca Pucllana is characterised by pyramidal platforms built of sun-dried bricks, or adobes (Patterson and Lanning 1964; Flores 1981). The main structure at the site is the Great Pyramid which is 400 m long, 60 m wide, and 20 m high, and includes seven platforms (Flores 2005; Flores et al. 2012; Navalte 2012)(Figure 2).

The construction of the main Lima ceremonial centre at Huaca Pucllana was done between the fourth and the seventh centuries AD. The Huaca Pucllana site was used for various activities related to the worship of divinities associated with the so-called sea-culture, which results from the proximity of the coast and it is considered as a ceremonial and administrative centre, with constant pilgrimage of people from surrounding regions to perform ceremonial rituals.

The Lima culture developed particular funerary patterns characterised either by burials, whereby single or multiple bodies were wrapped in cloth in extended position, or the practice of human sacrifices with no specific position of the body and with or without associated pottery and burial offerings (Barreto 2012). Human remains were often found in association with cultural elements, which can be attributed to a particular layer at the site. In addition, remains of ceremonial activities, such as offerings of pottery vessels, ritual banquets and sacrifices of young women were also identified (Barreto 2012).

Bioanthropological analysis of human remains at Huaca Pucllana shows a large number of violent deaths among the Lima individuals (Barreto 2012). Moreover, some of these findings suggest that individuals were sacrificed as part of rituals that correlates the last stages of the construction of the Great Pyramid at the site (Barreto et al. 2010). Huaca Pucllana appears to have been abandoned in the seventh century AD which was explained as a result of political and population changes driven by the expansion of the Wari, who arrived from the highland region of Ayacucho (Jennings 2010).

The Wari culture represents the second settlement stratum at Huaca Pucllana during the Middle Horizon (600–1000 AD), which partly destroyed and re-structured the former ceremonial Lima centre at the Great Pyramid, and used the place as an Imperial Wari elite cemetery (Flores 1981; Flores 2013). However, most of the Wari burials were also destroyed later on which altered the distribution of the elements of each burial, and intact mummies were only found in very few cases (Flores 2005). Most of the burials from this stratum represent individuals from the Wari elite characterized as "fardos" or funerary bundles, a method to preserve a corpse before burial as main distinctive feature found in the upper platforms of the Great Pyramid.

The presence of various elements such as clothing, household goods or ritual items and food as diagnostic elements of Wari mortuary practices also categorize these as elite Wari burials (Isbell 2004; Valdez et al. 2006). Some tombs and burials only contain a single individual, while others can contain more than four bodies. Some cases include human sacrifices such as young children next to the main body, which is described as typical Wari culture mortuary practice for important personalities as recently confirmed at the "Castillo de Huarmey" archaeological site (Pringle 2014).

Archaeologists consider the transition from the Middle Horizon (600–1000 AD) to the Late Intermediate period (1000–1476 AD) in the Central Andes of South America as another important cultural change, as it involves the collapse of the Wari Empire (Williams 2002; Finucane et al. 2007; Kemp et al. 2009). From ~1000 AD onward the Wari power gradually declined as their political centre increasingly lost control over cities and the territories that had extended in the Central Andes for four centuries. During this period of crisis, large urban centres such as Cajamarquilla and Pachacamac in the Central coast were abandoned (Segura and Shimada 2010), only to become occupied by the Ychsma culture soon after. It is assumed that the Huaca Pucllana site had been neglected for about 100 years from which there is no archaeological record.

However, causes for the collapse of the Wari Empire are not well understood. Some scholars argue that the Wari Empire succumbed to economic problems that prevented the political structure from meeting the growing needs of the empire (Willey 1991). Others believe that the Wari Empire collapsed as a consequence of a major drought and climate change (Williams 2002), likely alongside internal stress and social conflicts. With the decline and depopulation of the Wari centres starts the emancipation of the peoples from surrounding cities.

However, at Huaca Pucllana, it is assumed that some remaining local people reutilized the site around 1000–1100 AD. While there is characteristic evidence for an Ychsma occupation, this phase saw the addition of new constructions and the site was again used as cemetery. The subsequent Ychsma culture (1000–1476 AD) marks the third major period at Huaca Pucllana during which the site was again used for offerings and burials purposes.

The Ychsma culture is associated with a characteristic pottery style and similar Wari funerary patterns that also used the "fardos" as mortuary elements. The Ychsma culture represents and constitutes a distinctive cultural unit with political-economic autonomy in the Late Intermediate Period. While the presence of Ychsma was reported in other settlements /locations in the Central coast, e.g. Pachacamac (Eeckhout 2000), there are several elements that suggest a reconfiguration of Huaca Pucllana as religious space that was then continuously used during Ychsma times (Flores 1981).

2.2. Radiocarbon dates from Huaca Pucllana samples

Nine samples excavated at the site were sent for radiocarbon analysis at the University of Oxford Radiocarbon Accelerator Unit, Oxford, England. These dates were then used to confirm the assignment to cultural periods based on cultural contexts for all other Huaca Pucllana samples. Radiocarbon dates have been calibrated by using the Oxcal computer program (v4.2) of C. Bronk Ramsey, applying the `IntCal13' dataset – Northern Hemisphere (Reimer et al. 2013). Dates estimates were also calibrated with the Southern Hemisphere correction (SHCal13) (Hogg et al. 2013) (SI Table 4).

2.3. Sample preparation and DNA extraction

Samples were processed in a dedicated aDNA facility at the Australian Centre for Ancient DNA (ACAD) at the University of Adelaide, Australia. The laboratory employs standardized protocols and infrastructure for aDNA analysis (positive air pressure, UV irradiation and regular cleaning with oxidating agents, e.g. commercial bleach and Decon® to minimize contamination) (Cooper and Poinar 2000; Fulton 2012; Knapp et al. 2012).

Samples were decontaminated upon entry to the aDNA laboratory by exposure to UV light. The surface of the samples was gently wiped with 3% bleach and then physically removed by abrasion using a Dremel® drill.

A Mikro-dismembrator ball mill (Sartorius) was used to pulverize the sample and 0.2 g of bone powder were subsequently used in DNA extractions. For the extractions, samples were decalcified by incubation in 4 mL of 0.5 M EDTA (pH 8.0) overnight at 37°C and mixed constantly on a rotor mixer. Next, 70 µL Proteinase K (Invitrogen) was added and the lysis mix was incubated for 2 hours at 55°C under constant rotation. DNA was isolated using silicon dioxide solubilised in a Guanidinium buffer (Qiagen), as described previously (Brotherton et al. 2013; Der Sarkissian et al. 2013). DNA was resuspended in 200µL of TE buffer including 0.05% Tween-20 and stored at -20ºC until further use.

2.4. Library preparation and hybridization capture of mtDNA

DNA samples used for genomic library preparation were initially screened by sequencing the Hypervariable Region (HVR–I), and genotyping of two multiplex SNP assays, the GenoCoRe22 (Haak et al. 2010) and the AmericaPlex26 (Coutinho et al. 2014). Genomic libraries and enzymatic clean-up steps were performed also in a dedicated contamination-free room at ACAD for the Huaca Pucllana samples that showed replicable genotyping results. Library preparation and hybridization capture of mtDNA was conducted as described in (Fehren-Schmitz et al. 2015).

2.5. DNA sequencing and sequence assembly

DNA pooled libraries were sequenced on one Illumina flow cell lane on a Hi-Seq2000 Illumina machine (Illumina Inc., San Diego CA) at the Australian Cancer Research Foundation (ACRF) Cancer Genomics Facility, Adelaide, South Australia. Post-sequencing processing was performed using the ACAD pipeline:

Illumina reads were processed and filtered by means of demultiplexing of DNA sequences according to their index sequence, using the Illumina program Casava. Further demultiplexing and trimming of the internal barcodes was performed using the program Sabre 1.0 (https://github.com/najoshi/sabre). Residual adapter sequences were trimmed and read pairs were merged using Adapter Removal v1.5 (Lindgreen 2012) with default parameters. Merged reads were mapped to the reference sequence RSRS (Revised Sapiens Reference Sequences) (Behar et al. 2012) using the mapping program bwa 0.7.5a-r405 (Li and Durbin 2009), and duplicate reads were removed using FilterUniqueSAMCons.py (Kircher 2012). Ancient DNA damage patterns were assessed using MapDamage v0.3.6 (Ginolhac et al. 2011).

2.6. Sequence analysis

Read pileups were edited with Geneious Pro® Software V.6 (Biomatters Ltd) (Drummond et al. 2011) against the Reconstructed Sapiens Reference Sequence RSRS (16,569 bp) (Behar et al. 2012). Variant calling (SNPs) was based on a minimum coverage of 3x and an initial automated majority call of 75%, following by independent verification by eye by two researchers (GV and BL). Haplotype determination was performed using the online database phylotree (mtDNA tree Build 16 [19 Feb 2014]) (van Oven and Kayser 2009).

We excluded substitutions at nucleotide positions 16182 and 16183, because they are dependent on the presence of C at the position 16189 (Horai et al. 1993). For phylogenetic reconstruction and network analysis, we did not consider positions 309.1C(C), 315.1C AC indels at 523 and 524, deletion 3107 and position 16519 according to (van Oven and Kayser 2009).

2.7. Population genetic network and statistical analyses

Haplotype diversity (h), nuclear diversity (π), and pairwise genetic distances F_{ST} were calculated using the software Arlequin 3.5 (Excoffier and Lischer 2010). Median Joining (MJ) networks (Bandelt et al. 1999) were built using the program Network version 4.6 (www.fluxus-engineering.com) to evaluate genetic relationships among all individuals and to illustrate the extent of haplotype sharing within and among temporal periods and cultures.

We generated multidimensional scaling (MDS) plots to visualise genetic similarities and dissimilarities as measured by fixation indices (Slatkin's F_{ST}) (Slatkin 1995) in bi-dimensional space. Since genome data for modern-day populations in the Americas, is heavily biased towards phylogeographic questions that result in an excess of specific haplogroups, we had to restrict our analyses to HVR-I mtDNA sequences compiled from published sources to describe the genetic relationships between ancient samples from Huaca Pucllana and other ancient individuals (n=352) from 16 populations as well as present-day South Americans (n=1090) from 28 populations (SI Table 1).

3. RESULTS

3.1. Summary statistics and haplotype sharing

Sequence-based genetic distances F_{ST} among the tree cultural strata Lima, Wari and Ychsma show Lima–Wari (-0.05197), Lima–Ychsma (-0.06111) and Wari–Ychsma (-0.00770) (Table 1). Haplotype diversity (h) did not differ significantly between the three cultures Lima (0.9722), Wari (0.9778), Ychsma (0.9810).

The same happened for nucleotide diversity (π): Lima (0.002331), Wari (0.002492), Ychsma (0.002159) (Table 1). Genetic diversity indices and neutrality tests indicated that the Ychsma culture shows greater haplogroup diversity 0.9810 ($p>0.0308$) than the two previous occupation phases. Values for Tajima's D (neutrality test) only suggested a population expansion for the Lima culture, while Fu's FS values did not indicate population expansion for either of the three periods at Huaca Pucllana.

Phylogenetic and network analyses showed that haplotypes are shared between individuals from the same cultural layer, such as two individuals from Lima (10817-Lima/10820-Lima) belonging to mtDNA haplogroup C1b, two individuals from Ychsma (10722-Ychsma /10805-Ychsma) belonging to D1, and two individuals from Wari (10734-Wari/10765-Wari) belonging to C1b. We also observed shared haplotypes between individuals from different cultural layers, e.g. two individuals from Lima and Ychsma (10811-Lima/10732-Ychsma) belonging to D, and three samples from Wari and Ychsma (10763-Wari/10709-Ychsma/10725-Ychsma) belonging to C1b (Figure 4).

3.2. Mitochondrial DNA haplogroups and sequence coverage

From a total of 52 individuals collected from the Huaca Pucllana site, we obtained 34 complete mtDNA genome sequences: 9 for Lima, 10 for Wari, 15 for Ychsma. We have obtained sufficient sequencing coverage of mtDNA genomes (100%) for the large majority of samples and the average coverage depth per position shows (~97,4x) (range: 16,8x – 177,9x) (Table 2).

All 34 individuals could be assigned to one of the four "founder" mtDNA Native American haplogroups (A–D), defining 28 distinct and novel haplotypes (Table 3) not yet observed in modern-day populations. Detailed haplogroup frequencies for the populations analysed in this study were calculated per cultural period as reported in (Figure 3).

3.3. Huaca Pucllana mtDNA diversity in the South American context

We evaluated the relationship of cultural groups from Huaca Pucllana with modern and ancient mtDNA diversity in South America based on a genetic distance matrix and visualized as a Multidimentional Scaling Plot (MDS) (Figure 5).

Individuals from the EIP Lima culture show genetic proximity with highland Laramate populations from the MH and LIP and with Palpa coastal samples from the LIP. Interestingly, ancient Lima differed completely from the southern coastal Nasca populations (Palpa-EIP) from the same cultural horizon. Moreover, individuals from the MH Wari culture from Conchopata in the highlands also fall close to the coastal Lima individuals, whereas the Wari individuals from Huaca Pucllana form a cluster with the Huari-LIP from the highlands (Kemp et al. 2009) and two modern-day Mapuche populations (Mapuche and Mapuche1) from Chile, Individuals from the LIP Ychsma culture fall intermediate between the ancient Wari and Lima individuals and close to modern-day Yungay populations from Peru (Figure 5).

4. DISCUSSION

We sequenced 34 complete human mitochondrial genomes from the three successive cultural periods Lima, Wari and Ychsma found at Huaca Pucllana. This dataset represents a unique chronological transect in a restricted geographic area in South America, the Peruvian Central coast.

Low F_{ST} values between the three groups; Lima, Wari and Ychsma (Table 1) indicate no differentiation and that no dramatic changes in genetic composition over approximately ~1250 years have influenced the mtDNA diversity of the ancient populations from Huaca Pucllana. This therefore suggests population continuity and no major demographic shifts. Genetic diversity indices also do not differ substantially among the three cultural periods. Samples from the latest Ychsma period are the most diverse (Table 1), which is likely explained by the larger sample size, which may drive the statistic observation.

However, mitochondrial sub-haplogroup frequencies reveal interesting subtleties in the fluctuation and increase of haplogroup C1b and B2b in Wari individuals from the Middle Horizon period alongside a considerable drop of haplogroup B2 suggesting the arrival of new lineages in the Lima region during the occupation phase of the Wari Empire.

Interestingly, the haplogroup composition of the subsequent LIP Ychsma culture shows that haplogroup B2 lineages are present in frequencies similar to the EIP Lima culture, but absent in the Middle Horizon Wari.

The frequency of haplogroup B2b remains constant during the subsequent Wari and Ychsma periods, suggesting that following increase in the frequency of haplogroups C1b and B2b might be linked to the gene flow from the Andean region during the Middle Horizon. However, we note that haplogroup A2 lineages also increase with the Wari period, but do not seem to become established in the Lima region during Ychsma times, following the collapse of the Wari Empire (Figure 3).

We caution that haplogroup composition represents a very broad approach to investigate signals of population structure, in particular because frequencies are subject to genetic drift or stochastic processes, population movements and gene flow (Lewis 2009). Moreover, sample assignation confirmed by well-defined radiocarbon dates from the three cultural periods (Table 4), might not be representative for the whole population in Huaca Pucllana, therefore these results may be interpreted with care.

Network analysis showed that haplotype sharing occurred mostly between individuals from the same archaeological period, and less often between different cultural layers (only two instances, i.e. Wari-Ychsma, Lima-Ychsma). For example, a shared C1b lineage between Wari and the subsequent Ychsma is very plausible, while the shared Lima-Ychsma haplotypes suggests a local continuity of particular maternal lineages over several centuries, which implies that the Wari incursion was not a replacement but rather brought new lineages to the Lima region (Figure 4).

From the populations' affinities shown in the MDS plot, we propose a demographic scenario in which the Lima occupation in the EIP (i.e. before the arrival of Wari) in Peru's Central coast show affinities with ancient highland populations rather than other coastal populations from the same period (i.e. Palpa EIP and Montegrande EIP), suggesting that the population history of the Central coast is connected with the highlands and did not develop independently unlike Moche and Nasca populations (Reindel. M and Wagner. G 2009). However, this observation is independent from the archaeological records of the Central coast and should not be overinterpreted in terms of the cultural origins of the Lima Culture.

We also observed that the Wari influence makes the Ychsma culture fall intermediate between the Wari and the Lima substratum suggesting one more time the influence of highland groups (i.e. Wari migration/occupation in the Central coastal Peru during the Middle Horizon (Figure 5). More generally, the data support population continuity in the area and no major upheaval and a relatively homogenous mtDNA haplogroup composition across cultural periods in Central coastal Peru, even in the light of the expansion of the Wari Empire. Our results nonetheless hint at subtle demographic changes in coastal Peru during the Middle horizon, reinforcing the potential Wari expansion from the highlands towards the Central coast (i.e. Huaca Pucllana).

Moreover, it appears that the Wari did not adopt highly aggressive means of control such as direct colonization outside the political centre in Ayacucho, as suggested previously (Kaulicke 2000). Perhaps the independence of local and regional identities during the Middle Horizon co-existed within an atmosphere of population interaction and relationships implying a social complexity which is not yet understood in the Central Coast of Peru (Kaulicke 2000).

We caution that the situation at Huaca Pucllana in the greater Lima region might not be representative of the overall genetic influx of the Wari Empire in the Central coast, supporting the hypothesis that local settlements along the Central coast were not entirely dominated or affected by Wari imperialism (Shady 1982; Shady 1988). Consequently, definitive assertions about the Wari imperialism / colonialism in Central coastal Peru require further research.

Ancient DNA sequences from Huaca Pucllana contain a substantial number of novel haplotypes as defined by additional coding region SNPs within the mtDNA genome, suggesting that these sequences represent a unique pre-Columbian genetic diversity (Table 3). This can be explained under two premises: i) the mtDNA diversity of pre-Columbian populations was indeed substantially high, suggesting that our ancient mtDNA sequences constitute a sub-sample of the former diversity with the possibility of uncovering potential extinct lineages given the fact that they have not been described in modern Peruvian populations; or ii) there was a considerable reduction in the modern genetic diversity as a result of the dramatic losses amongst Native American populations during and after the European conquest (Livi-Bacci 2006), hence this substantial reduction of mtDNA lineages correlates with the impact of the Spanish colonization.

Within the Central Andes, ancient Peru counted the largest ensemble of indigenous peoples related to the vast expansion of the Inca Empire across South America (1432–1532 AD). While population densities must have differed in various regions of the Americas, the highlands of Central South America and Peru are among the densest (Crawford 1998).

As exemplified by the pre-Columbian expansions of Wari and Tiwanaku and later on the Inca Empire, the Central Andes have always been a dynamic environment with considerable population movements (Isbell 2008). Another large-scale internal migration movement took place in Peru in more recent times, with rural and indigenous population (e.g. Junín, Ayacucho, La Libertad, Ica, Lambayeque, Cajamarca, Piura, and to a lesser degree from other places) moving towards the urban centres like Lima since the 1940´s (Sandoval JR et al. 2013). We therefore suggest that the ancestry of people from large urban centres like today's Lima must have been constantly reshaped by newly incoming genetic diversity.

More sampling and high-resolution sequencing efforts are required to refine the analysis of ancient genetic diversity in Peru, but also from well-defined modern-day groups to be able to contrast ancient diversity with present-day populations. Such data would allow the detailed testing of demographic scenarios, replacements, and population bottlenecks on the basis on statistical inferences or simulations, which could elucidate the dynamic of pre-Columbian societies and the dramatic demise of Native American people that coincided with the time of the European arrival (O'Fallon and Fehren-Schmitz 2011).

5. CONCLUSION

This study represents the first investigation of ancient mtDNA genome data from the Peruvian Central coastal region. This transect through time study spanning ~1250 years covers three cultural periods allowing us to test the potential impact of the Wari expansion from the Ayacucho region in the highlands on the coastal Lima culture, and its legacy in the subsequent Ychsma period. The results indicate that haplogroup composition only slightly shifted across archaeological periods with no major changes, which rules out a complete population discontinuity or replacement driven by the Wari Empire.

We observed a relatively homogenous local genetic diversity from the Early Intermediate to the Late Intermediate period whereby the Wari influence only added to the diversity in the Central coast during the Middle Horizon, resulting in the highest diversity in the following Ychsma period. We suggest based on genetic data that the Wari imperialism did not produced significant demographic changes in Central coastal Peru at least at Huaca Pucllana level and the region around modern-day Lima.

Acknowledgments

This work was funded by the Australian Research Council (ARC). We thank team members of the "Proyecto Arqueológico Huaca Pucllana" for helping with accessing the collection of pre-Columbian human remains from the site. We thank Micaela Alvarez Calmet for providing archaeological information and helpful comments.

REFERENCES

Bandelt HJ, Forster P, and Rohl A. 1999. Median-joining networks for inferring intraspecific phylogenies. MolBiolEvol 16(1):37-48.

Barreto MI. 2012. Prácticas sacrificiales en el valle del Rímac durante el Período Intermedio temprano (150 - 650 d.C.). Tesis para obtener el Grado académico de Magister en Arqueología Andina Escuela de Post Grado de la Facultad de Ciencias Sociales Universidad Nacional Mayor de San Marcos 248 pp., Lima.

Barreto MI, Ccencho J, Chuchon H, and Silvera de la Torre H. 2010. Sacrificios humanos Lima asociados a la última etapa constructiva de Huaca Pucllana. Arqueología y Sociedad 22:55-77.

Behar DM, van Oven M, Rosset S, Metspalu M, Loogvali EL, Silva NM, Kivisild T, Torroni A, and Villems R. 2012. A "Copernican" reassessment of the human mitochondrial DNA tree from its root. Am J Hum Genet 90(4):675-684.

Brotherton P, Haak W, Templeton J, Brandt G, Soubrier J, Jane Adler C, Richards SM, Sarkissian CD, Ganslmeier R, Friederich S et al. . 2013. Neolithic mitochondrial haplogroup H genomes and the genetic origins of Europeans. Nature Communications 4:1764.

Brown K. 1998. Keeping it clean: the collection and storage of ancient DNA samples from the field. The Archaeologist 33:16 -17.

Cooper A, and Poinar HN. 2000. Ancient DNA: Do It Right or Not at All. Science(289):1139.

Coutinho A, Valverde G, Fehren-Schmitz L, Cooper A, Barreto Romero MI, Espinoza IF, Llamas B, and Haak W. 2014. AmericaPlex26: a SNaPshot multiplex system for genotyping the main human mitochondrial founder lineages of the Americas. PLoS ONE 9(3):e93292.

Crawford MH. 1998. The origins of Native Americans. Evidence from anthropological genetics. Cambridge University Press, Cambridge (UK).

Der Sarkissian C, Balanovsky O, Brandt G, Khartanovich V, Buzhilova A, Koshel S, Zaporozhchenko V, Gronenborn D, Moiseyev V, Kolpakov E et al. . 2013. Ancient DNA reveals prehistoric gene-flow from siberia in the complex human population history of North East Europe. PLoS Genetics 9(2):e1003296.

Dillehay TD. 1997. The Archaeological Context and Interpretation, Volume 2: Monte Verde, a Late Pleistocene Settlement in Chile. (Washington, DC: Smithsonian Institute Press).

Drummond A, Ashton B, Buxton S, Cheung M, Cooper A, Duran C, Field M, Heled J, Kearse M, Markowitz S et al. . 2011. Geneious v5.4. Available from http://www.geneious.com/.

Earle T, and Jennings J. 2012. Remodeling the political economy of the Wari Empire. Boletín de Arqueología PUCP(16):209-225.

Eeckhout P. 2000. The Palaces of the Lords of Ychsma. An Archaeological Reappraisal of the Function of Pyramids with Ramps at Pachacamac, Central Coast of Peru. Journal of American Archaeology 17-18-19:217-254.

Excoffier L, and Lischer HE. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour 10(3):564-567.

Fehren-Schmitz L, Hummel S, and Herrmann B. 2009. Who Were the Nasca? Population Dynamics in Pre-Columbian Southern Peru Revealed by Ancient DNA Analyses.159-172.

Fehren-Schmitz L, Llamas B, Lindauer S, Tomasto-Cagigao E, Kuzminsky S, Rohland N, Santos FR, Kaulicke P, Valverde G, Richards SM et al. . 2015. A Re-Appraisal of the Early Andean Human Remains from Lauricocha in Peru. PLoS ONE 10(6).

Fehren-Schmitz L, Reindel M, Cagigao E, Hummel S, and Herrmann B. 2010. Pre-Columbian population dynamics in coastal southern Peru: A diachronic investigation of mtDNA patterns in the Palpa region by ancient DNA analysis. Am J Phys Anthropol 141:208 - 221.

Fehren-Schmitz L, Warnberg O, and Reindel M. 2011. Diachronic investigations of mitochondrial and Y-chromosomal genetic markers in pre-Columbian Andean highlanders from South Peru. Ann Hum Genet 75:266 - 283.

Finucane BC, Valdez E, Calderon IP, Pomacanchari CV, Valdez L, and O'Connell T. 2007. The End of the Empire: New Radiocarbon dates from the Ayacucho Valley, Peru and their implications for the Collapse of the Wari State. Radiocarbon 49(2):579-592.

Flores I. 1981. "Investigaciones Arqueológicas en la Huaca Pucllana, Miraflores" Boletín de Lima, Editorial Los Pinos, Lima III (13):65-70.

Flores I. 2005. Pucllana, esplendor de la cultura Lima. Instituto Nacional de Cultura, Lima.

Flores I. 2013. Los Wari en Pucllana. La tumba de un sacerdote: Resultados del análisis bioantropológico. Ministerio de Cultura, Lima:87 - 142.

Flores I, Vargas P, Ccencho J, and Silvera H. 2012. Los patios con estructuras escalonadas de Huaca Pucllana: Caracterización y función de arquitectura ceremonial Lima. Arqueologia y Sociedad: Museo de Arqueología y Antropología, Universidad Nacional Mayor de San Marcos, Lima 25:57 - 88.

Fulton TL. 2012. Setting up an ancient DNA laboratory. Methods Mol Biol 840:1-11.

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

Ginolhac A, Rasmussen M, Gilbert MT, Willerslev E, and Orlando L. 2011. mapDamage: testing for damage patterns in ancient DNA sequences. Bioinformatics 27(15):2153-2155.

Glowacki M. 2012. Imperialismo en el Horizonte Medio: Una reevaluacion del paradigma clasico, Cuzco, Peru. Boletin de Arqueologia PUCP, ISSN 1029-2004 16:189-207.

Goldstein P. 1989. The Tiwanaku Occupation of Moquegua. In Ecology, Settlement and History in the Osmore Drainage, Peru, edited by D.S. Rice, C. Stanish, and P.R. Scarr. BAR. International Series: 545(i):219-255.

Haak W, Balanovsky O, Sanchez JJ, Koshel S, Zaporozhchenko V, Adler CJ, Der Sarkissian CSI, Brandt G, Schwarz C, Nicklisch N et al. . 2010. Ancient DNA from European early neolithic farmers reveals their near eastern affinities. PLoS biology 8(11):e1000536.

Hogg AG, Hua Q, Blackwell PG, Niu M, Buck CE, Guilderson TP, Heaton TJ, Palmer JG, Reimer PJ, Reimer RW et al. . 2013. SHCal13 Southern Hemisphere Calibration, 0–50,000 Years cal BP.

Horai S, Kondo R, Nakagawahattori Y, Hayashi S, Sonoda S, and Tajima K. 1993. Peopling of the America, Founded by 4 Major Lineages of Mitochondrial-DNA. Molecular Biology and Evolution 10(1):23-47.

Isbell WH. 2004. Mortuary Preferences : A Wari Culture Case Study from Middle Horizon Peru. Latin American Antiquity 15:3-32.

Isbell WH. 2008. Wari and Tiwanaku: International Identities in the Central Andean Middle Horizon, en: H. I. Silverman y W. H. Isbell (eds). Handbook of South American Archaeology Springer, New York.:731-759.
Isbell WH, and McEwan GF. 1991. Huari administrative structure: prehistoric monumental architecture and state government. Dumbarton Oaks Research Library and Collection, Washington, DC.
Isbell WH, and Schreiber KJ. 1978. Was Huari a State? American Antiquity 43:372-389.
Jennings J. 2010. Beyond Wari Walls : Regional Perspectives on Middle Horizon Peru. Albuquerque, NM, USA: University of New Mexico Press.
Kaulicke P. 2000. La Sombra de Pachacamac: Huari en la costa central. Boletin de Arqueologia PUEP 4:313-358.
Kemp BM, Tung TA, and Summar ML. 2009. Genetic continuity after the collapse of the Wari empire: mitochondrial DNA profiles from Wari and post-Wari populations in the ancient Andes. Am J Phys Anthropol 140(1):80-91.
Kircher M. 2012. Analysis of high-throughput ancient DNA sequencing data. Methods Mol Biol 840:197-228.
Knapp M, Clarke AC, Horsburgh KA, and Matisoo-Smith EA. 2012. Setting the stage - building and working in an ancient DNA laboratory. Ann Anat 194(1):3-6.
Lanning EP, and Pollard G. 1967. Peru before the Incas: Prentice-Hall Englewood Cliffs, NJ.
Lewis C. 2009. Difficulties in rejecting a local ancestry with mtDNA Haplogroup data in the South Central Andes. Latin American antiquity.
Li H, and Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25(14):1754-1760.
Lindgreen S. 2012. AdapterRemoval: easy cleaning of next-generation sequencing reads. BMC research notes 5(337).
Livi-Bacci M. 2006. The Depopulation of Hispanic America after the Conquest. Population and Development Review 32(2):199-232.
Lumbreras LG. 1969. De los Pueblos, las Culturas y las Artes en el Antiguo Perú. Editorial Mancloa-Campodónico, Lima.
Lumbreras LG. 1974. The Peoples and Cultures of Ancient Peru. Smithsonian Institution Press, Washington, DC.
Lumbreras LG. 1980. El imperio Wari. Historia del Perú Perú Antiguo, Lima, Editorial Mejía Baca 1:9-91.
Marcone G. 2001. El Complejo de los Adobitos y la Cultura Lima en el Santuario de Pachacamac. Boletín de Arqueología PUCP Nº 4:597-606; Huari y Tiwanaku: Modelos vs Evidencias Fondo Editorial Pontificia Universidad Católica del Perú, Lima.
Marcone G. 2010. What Role Did Wari Play in the Lima Political Economy?: The Peruvian Central Coast at the Beginning of the Middle Horizon. In J Jennings (ed), *Beyond Wari Walls Regional Perspectives on Middle Horizon Peru*, The University of New Mexico Press, Alburquerque Chapter 8:136-154.
McEwan GF. 1990. Some Formal Correspondences between the Imperial Architecture of the Wari and the Chimu Cultures of Ancient Peru. Latin American Antiquity, Washington, DC 1(2):97-116.
Mendisco F, Keyser C, Seldes V, Rivolta C, Mercolli P, Cruz P, Nielsen AE, Crubezy E, and Ludes B. 2014. Genetic diversity of a late prehispanic group of the quebrada de humahuaca, northwestern Argentina. Ann Hum Genet 78(5):367-380.
Menzel D. 1964. Style and time in the Middle Horizon. Ñawpa Pacha, Berkeley, California, Institute of Andean Studies 2:1-105.

Moraga M, Santoro C, Standen V, Carvallo P, and Rothhammer F. 2005. Microevolution in prehistoric Andean populations: chronologic mtDNA variation in the desert valleys of northern Chile. Am J Phys Anthropol 127:170 - 181.

Navalte PC. 2012. Secuencia constructiva de la Gran Pirámide de Huaca Pucllana en sus tiempos finales. Investigaciones sociales, UNMSM-IIHS, Lima, Peru 16(28):303-311.

O'Fallon BD, and Fehren-Schmitz L. 2011. Native Americans experienced a strong population bottleneck coincident with European contact. Proc Natl Acad Sci U S A.

O'Rourke DH, and Raff JA. 2010. The human genetic history of the Americas: the final frontier. Curr Biol 20(4):R202-207.

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

Patterson T. 1966. Pattern and Process in the Early Intermediate Period Pottery of the Central Coast of Perú. Publication in Anthropology, University of California Press, Berkeley, Los Angeles Vol 3.

Patterson T, and Lanning E. 1964. Changing Settlement Patterns on the Central Peruvian Coast Nawpa-Pacha, Instituto de Estudios Andinos Berkeley 2:113-123, .

Pringle H. 2014. Peru's Royal Wari Tomb. http://ngmnationalgeographiccom/2014/06/peru-tomb/pringle-text Retrieved August 15, 2015.

Raghavan M, Skoglund P, Graf KE, Metspalu M, Albrechtsen A, Moltke I, Rasmussen S, Stafford TW, Jr., Orlando L, Metspalu E et al. . 2014. Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. Nature 505(7481):87-91.

Raghavan M, Steinrücken M, Harris K, Schiffels S, Rasmussen S, DeGiorgio M, Albrechtsen A, Valdiosera C, Ávila-Arcos MC, Malaspinas A-S et al. . 2015. Genomic evidence for the Pleistocene and recent population history of Native Americans. Science.

Rasmussen M, Anzick SL, Waters MR, Skoglund P, DeGiorgio M, Stafford TW, Jr., Rasmussen S, Moltke I, Albrechtsen A, Doyle SM et al. . 2014. The genome of a Late Pleistocene human from a Clovis burial site in western Montana. Nature 506(7487):225-229.

Rasmussen M, Sikora M, Albrechtsen A, Korneliussen TS, Moreno-Mayar JV, Poznik GD, Zollikofer CPE, Ponce de Leon MS, Allentoft ME, Moltke I et al. . 2015. The ancestry and affiliations of Kennewick Man. Nature 523(7561):455-458.

Reich D, Patterson N, Campbell D, Tandon A, Mazieres S, Ray N, Parra MV, Rojas W, Duque C, Mesa N et al. . 2012. Reconstructing Native American population history. Nature 488(7411):370-374.

Reimer PJ, Bard E, Bayliss A, Beck JW, Blackwell PG, Bronk Ramsey C, Buck CE, Cheng H, Edwards RL, Friedrich M et al. . 2013. IntCal13 and Marine13 Radiocarbon Age Calibration Curves 0–50,000 Years cal BP.

Reindel. M, and Wagner. G. 2009. New technologies for archaeology: multidisciplinary investigations in Palpa and Nasca, Peru. Berlin-Heidelberg-New York: Springer.

Rowe J. 1962. Stages and Periods in Archaeological Interpretation. Southwestern Journal of Anthropology 8:40-54.

Sandoval JR, Lacerda DR, Jota MSA, Salazar-Granara A, Vieira PPR, and al. e. 2013. The Genetic History of Indigenous Populations of the Peruvian and Bolivian Altiplano: The legacy of the Uros. PLoS ONE 8(9):e73006.

Schreiber KJ. 1992. Wari imperialism in Middle Horizon Peru. Anthropological Papers of the Museum of Anthropology 87:University of Michigan, Ann Arbor.
Schreiber KJ. 2012. Una aproximacion a las investigaciones sobre Wari: Paradigmas y perspectivas sobre el Horizon Medio. Boletin de Arqueologia PUCP, ISSN 1029-2004 16:11-22.
Segura R, and Shimada I. 2010. The Wari Footprint on the Central Coast: A View from Cajamarquilla and Pachacamac. Beyond Wari walls : Regional perspectives on Middle Horizon PeruThe University of New Mexico Press Chapter 7:113-135.
Shady R. 1982. La Cultura Nievería y la interacción social en el mundo andino en la época Huari. Arqueológicas 19:15-108.
Shady R. 1988. Epoca Huari como interacción de las sociedades regionales. Revista Andina 6(1):67-99.
Shady R, and Ruiz A. 1979. Evidence for interregional relationships during the Middle Horizon on the North central coast of Peru. American Antiquity 44:676-684.
Skoglund P, Mallick S, Bortolini MC, Chennagiri N, Hunemeier T, Petzl-Erler ML, Salzano FM, Patterson N, and Reich D. 2015. Genetic evidence for two founding populations of the Americas. Nature advance online publication.
Tung T. 2012. Violence, Ritual, and the Wari Empire: A Social Bioarchaeology of Imperialism in the Ancient Andes. University Press of Florida.
Tung TA. 2007. Trauma and violence in the Wari empire of the Peruvian Andes: Warfare, raids, and ritual fights. Am J Phys Anthropol 133(3):NA-NA.
Tung TA. 2008. Dismembering bodies for display: a bioarchaeological study of trophy heads from the Wari site of Conchopata Peru. Am J Phys Anthropol 136:294-308.
Valdez LM, Bettcher KJ, Ochatoma JA, and Valdez JE. 2006. Mortuary preferences and selected references: a comment on Middle Horizon Wari burials. World Archaeology 38(4):672-689.
van Oven M, and Kayser M. 2009. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Hum Mutat 30(2):E386-394.
Willey GR. 1991. Horizontal integration and regional diversity: an alternating process in the rise of civilizations. American Antiquity 56:197-215.
Williams PR. 2001. Cerro Baul: A Wari Center on the Tiwanaku Frontier. Latin American Antiquity, Society for American Archaeology 12:67-83.
Williams PR. 2002. Rethinking disaster-induced collapse in the demise of the Andean highland states: Wari and Tiwanaku. World Archaeology 33(2):361-374.
Williams PR, and Isla J. 2002. Investigaciones arqueológicas en Cerro Baúl, un enclave Wari en el valle de Moquegua. Gaceta Arqueologica Andina 26:87-120.
Yang DY, and Watt K. 2005. Contamination controls when preparing archaeological remains for ancient DNA analysis. Journal of Archaeological Science 32(3):331-336.

List of Figures and Tables

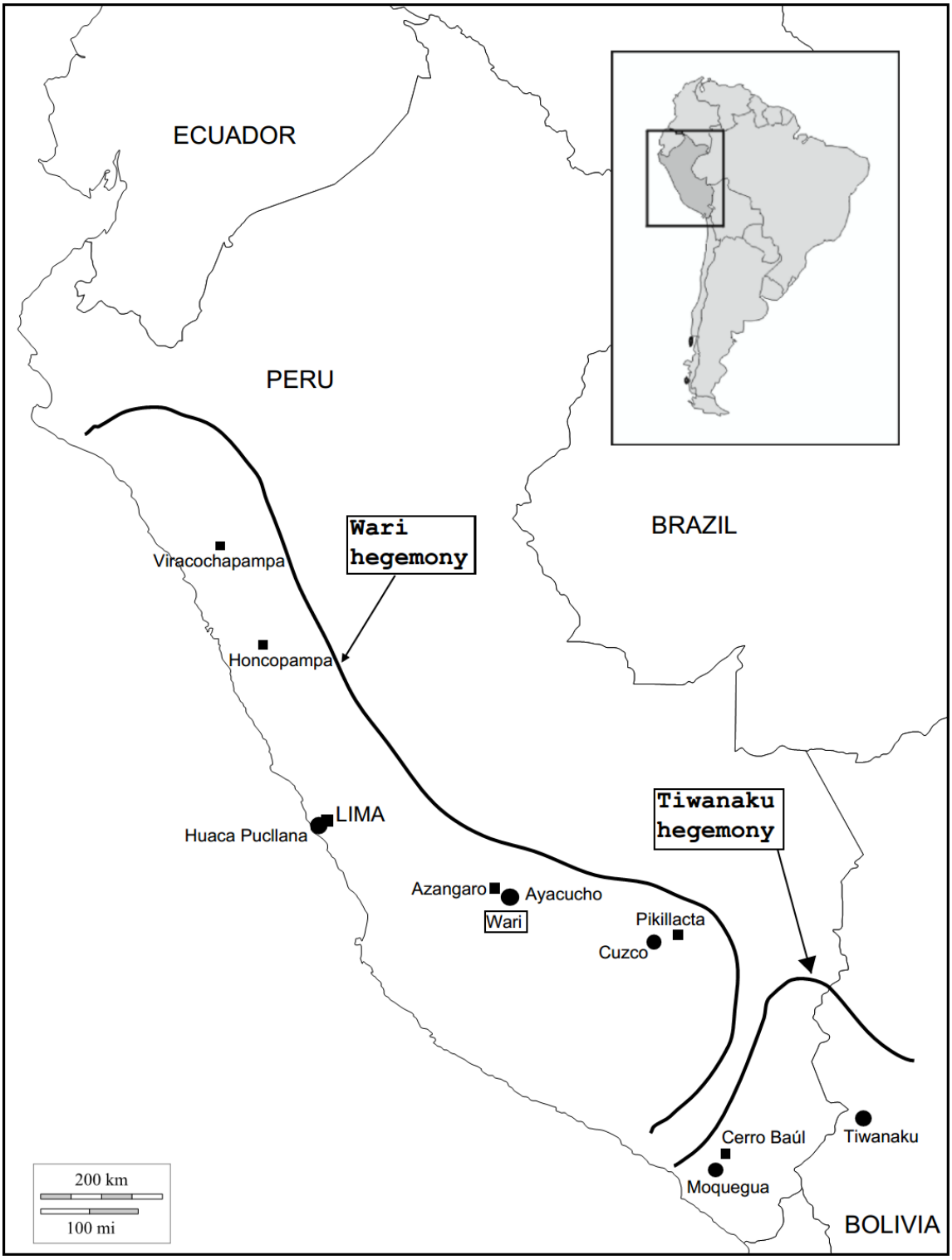


Figure 1. Map of Wari and Tiwanaku expansion territories in South America during the Middle Horizon.

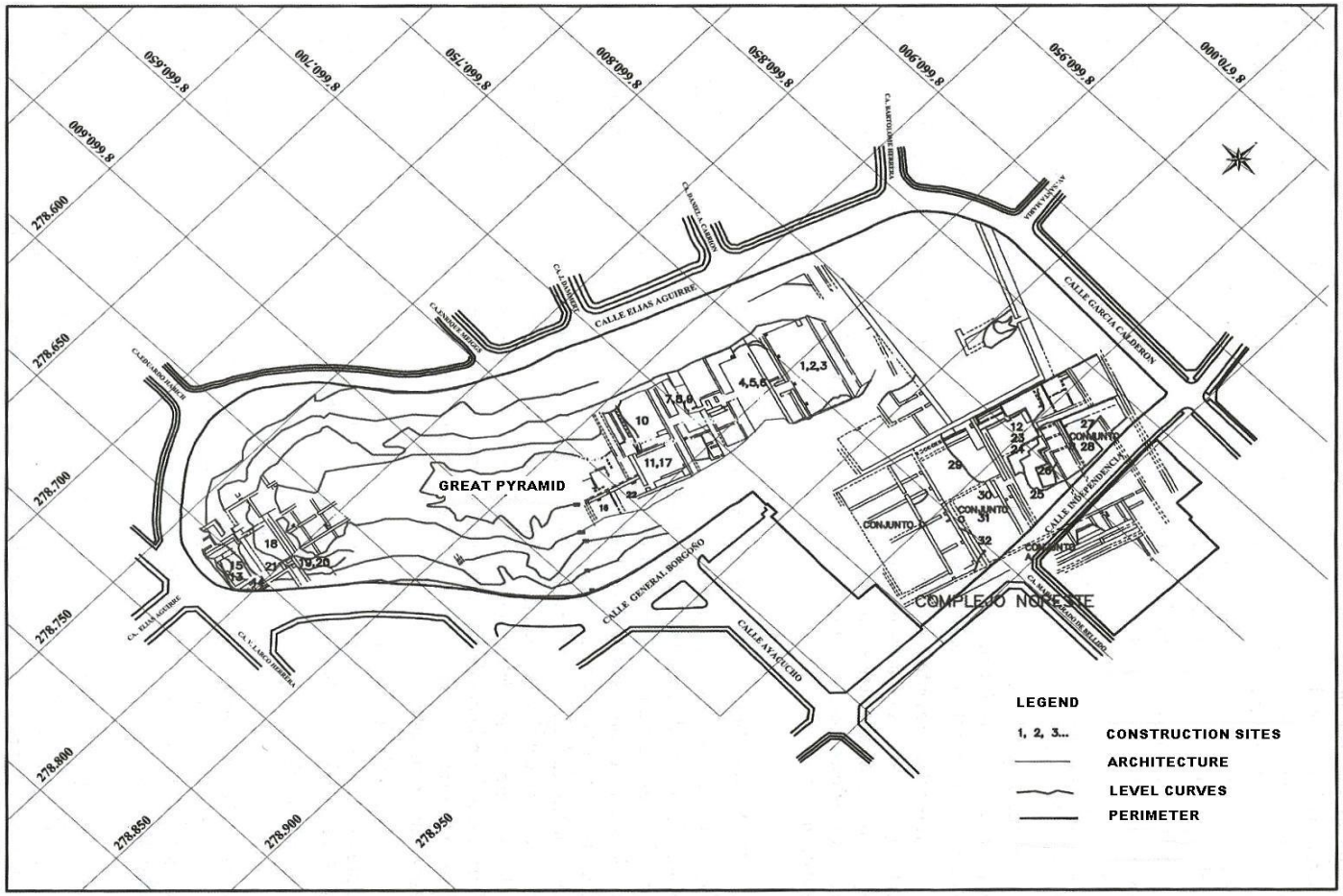
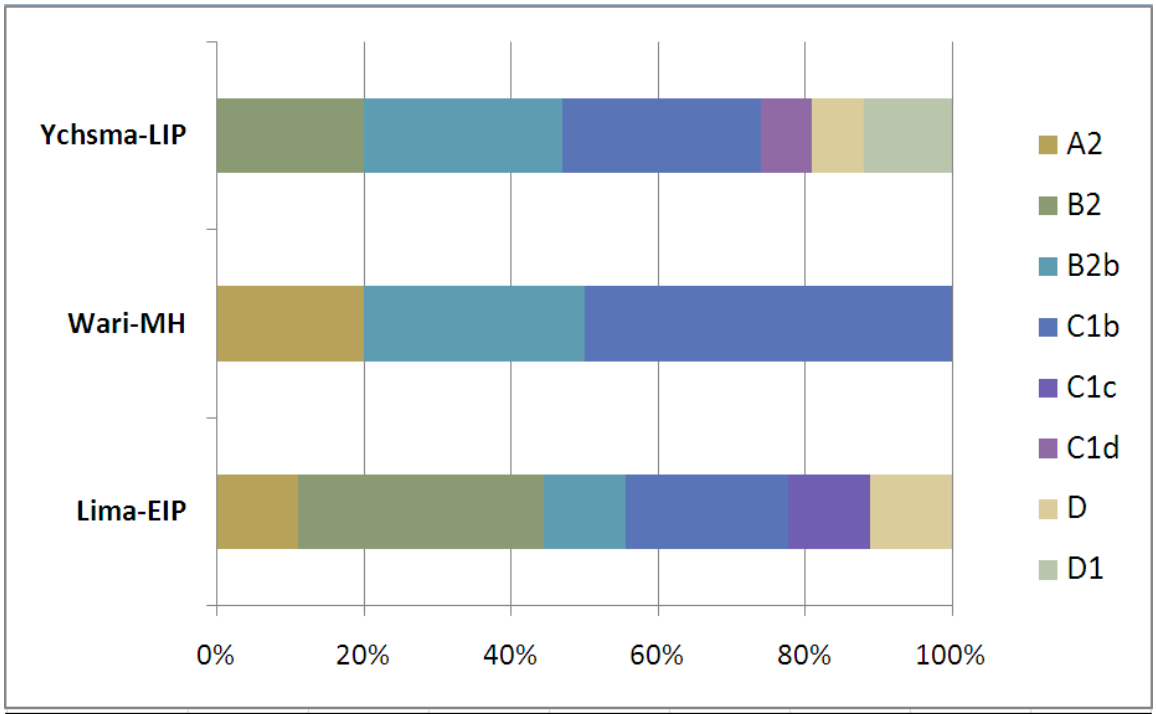


Figure 2. Map of Huaca Pucllana site in Lima, Peru. Schematic representation of the archaeological site and structure of the Great Pyramid from which human remains were collected. Based on (Flores et al. 2012).



	A2	B2	B2b	C1b	C1c	C1d	D	D1
Ychsma - LIP	0	0.20	0.27	0.27	0	0.07	0.07	0.12
Wari - MH	0.20	0	0.30	0.50	0	0	0	0
Lima - EIP	0.11	0.33	0.11	0.22	0.11	0	0.11	0

Figure 3. Mitochondrial DNA haplogroup frequencies from Huaca Pucllana across archaeological periods. Lima culture: Early Intermediate Period–EIP (200–600 AD), Wari culture: Middle Horizon–MH (600–1000 AD) and Ychsma culture: Late Intermediate Period–LIP (1000–1476 AD).

Network analysis

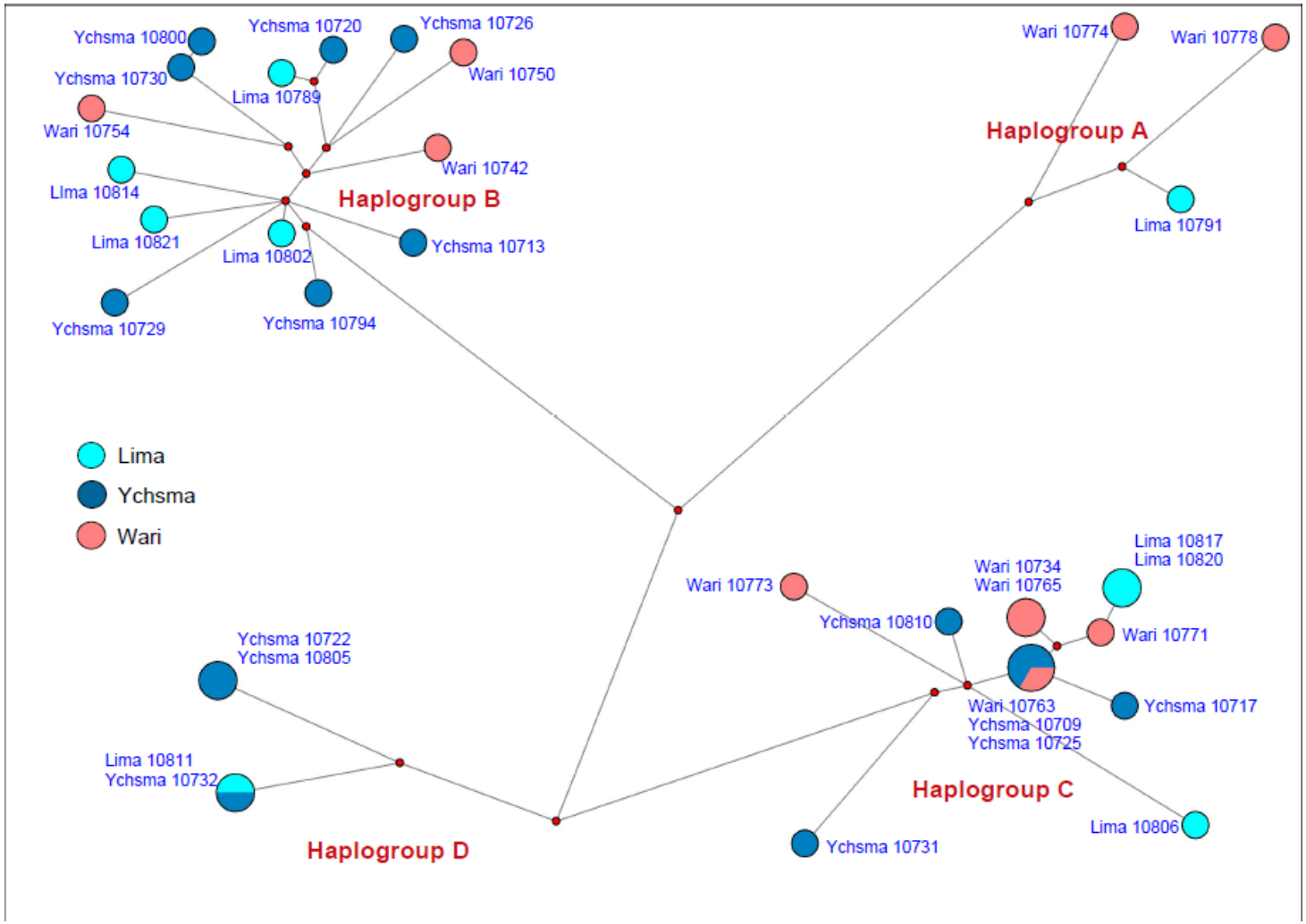


Figure 4. Phylogenetic network of individual samples from Huaca Pucllana.

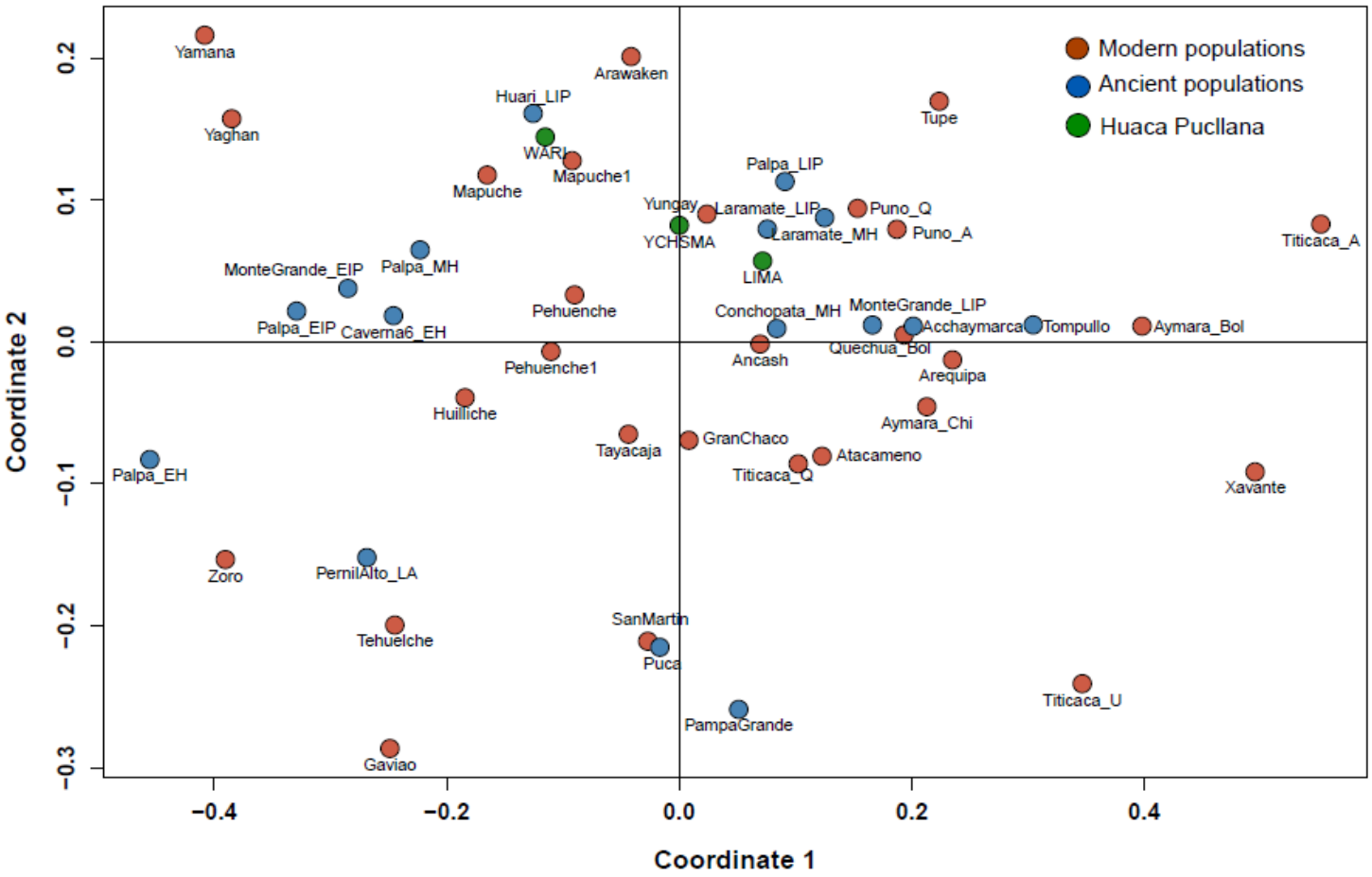


Figure 5. Multidimensional scaling plot (MDS) based on Slatkin's *F*st. Genetic affinities of Huaca Pucllana cultural groups compared with modern and ancient populations from South America.

Table 1. Summary statistics of all ancient populations analyzed in this study calculated with Arlequin v.3.5.
Sites are listed in chronological order

	Population pairwise F_{ST}				**Genetic diversity**			**Neutrality test**			
	Lima	**Wari**	**Ychsma**	***n***	***H***	***h***	***π***	**Tajima's D**	**p-value**	**Fu's FS**	**p-value**
Lima	0			**9**	8	0.9722 (0.0640)	0.002331 (0.001272)	-0.13498	0.47000	1.44409	0.67400
Wari	-0.05197	0		**10**	9	0.9778 (0.0540)	0.002492 (0.001340)	0.18623	0.62400	1.08278	0.60900
Ychsma	-0.06111	-0.00770	0	**15**	13	0.9810 (0.0308)	0.002159 (0.001117)	0.31343	0.66300	0.31237	0.51100

H: Number of haplotypes, h: Haplotype diversity, *π:* Nucleotide diversity*: statistically significant p-values (<0.05 for Tajima's D, p<0.02 for Fu's FS)*

Table 2. Haplogroup determination and mtDNA genomic coverage for Huaca Pucllana samples

ACAD	Location	Population	Haplogroup	Coverage of the mitogenome	Average coverage depth per position	StdDev	Min	Max	Private mutations
10791A	Huaca Pucllana	Lima	A2	100%	**122.7**	13.5	1	159	4
10814A	Huaca Pucllana	Lima	B2	100%	**164.6**	14.3	9	186	5
10802A	Huaca Pucllana	Lima	B2	100%	**63.4**	14.4	1	105	1
10789A	Huaca Pucllana	Lima	B2b	100%	**100.2**	15.1	5	141	3
10817A	Huaca Pucllana	Lima	C1b	100%	**132.2**	16.4	5	169	5
10820A	Huaca Pucllana	Lima	C1b	100%	**84.6**	15.8	1	135	5
10806A	Huaca Pucllana	Lima	C1c	100%	**73.4**	17.1	1	117	6
10811A	Huaca Pucllana	Lima	D	99.99%	**50.4**	14.9	0	97	6
10822A	Huaca Pucllana	Lima	B2	99.90%	**82.8**	35.8	0	151	4
10821A	Huaca Pucllana	Lima	B2	100%	**98.2**	19.9	1	149	4
10774A	Huaca Pucllana	Wari	A2	100%	**67.7**	18.1	1	115	6
10754A	Huaca Pucllana	Wari	B2b	100%	**48.8**	17	1	99	7
10734A	Huaca Pucllana	Wari	C1b	100%	**59.8**	17.3	1	104	4
10763A	Huaca Pucllana	Wari	C1b	99.90%	**20.7**	9	0	55	2
10771A	Huaca Pucllana	Wari	C1b	98.80%	**18.2**	8.2	0	50	4
10773A	Huaca Pucllana	Wari	C1b	100%	**78.0**	15.2	0	119	6
10778A	Huaca Pucllana	Wari	A2	97.60%	**20.5**	15.6	0	100	7
10765A	Huaca Pucllana	Wari	C1b	99.80%	**105.2**	62.5	0	340	4
10750A	Huaca Pucllana	Wari	B2b	98.70%	**22.3**	16.7	0	97	7
10742A	Huaca Pucllana	Wari	B2b	100%	**84.3**	40.2	0	248	5
10713A	Huaca Pucllana	Ychsma	B2	100%	**109.5**	21.9	2	163	4
10729A	Huaca Pucllana	Ychsma	B2	100%	**101.6**	24.8	0	171	6

10720A	Huaca Pucllana	Ychsma	B2b	100%	**171.6**	14.4	19	191	4
10726A	Huaca Pucllana	Ychsma	B2b	99.96%	**38.5**	14.7	0	88	5
10709A	Huaca Pucllana	Ychsma	C1b	100%	**146.6**	18.1	0	176	2
10717A	Huaca Pucllana	Ychsma	C1b	100%	**175**	11.9	18	191	5
10725A	Huaca Pucllana	Ychsma	C1b	100%	**177.9**	16.6	0	192	2
10731A	Huaca Pucllana	Ychsma	C1b	100%	**182**	17.7	0	193	7
10722A	Huaca Pucllana	Ychsma	D1	100%	**140.8**	21	0	181	5
10732A	Huaca Pucllana	Ychsma	D	100%	**146.6**	18.1	0	180	6
10730B	Huaca Pucllana	Ychsma	B2b	99.80%	**46.0**	20.2	0	102	5
10805A	Huaca Pucllana	Ychsma	D1	100%	**159.5**	13.1	7	186	6
10810A	Huaca Pucllana	Ychsma	C1d	99.90%	**90.3**	39.4	0	166	2
10794A	Huaca Pucllana	Ychsma	B2	99.60%	**16.8**	8.9	0	54	1
10800A	Huaca Pucllana	Ychsma	B2b	100%	**127.6**	17.2	4	166	6

Table 3. List of "private" variants (SNPs) from mtDNA complete genomes from Huaca Pucllana

Haplogroup A										
A2										
10791A_Lima	C64T	356.1C	T1189C	A5222G	T15289C	G16129A!	T16189C!			
10774A_Wari	C64T	A675G	A3354G	T6911C	A13966G	C14800A	G16145A			
10778A_Wari	C64T	T1189C	T3786C	A5222G	C5765T	A7673G	T8738C	T15289C	C16221T	G12127A
Haplogroup B										
B2										
10814A_Lima	T2857C	T5082C	T5277C	A15924G	T16330C					
10802A_Lima	C16168T									
10794A_Ychsma	G8290A	A16066G								
10713A_Ychsma	573.1C	A5186G	C5375T	C5895T	G15884A					
10729A_Ychsma	T152C!	A2880G	C7786T	C13934T	A16051G	C16360T				
10821A_Lima	T146C!	T3786C	A8170G	G15777A						
B2b										
10800A_Ychsma	T131C	C4013T	G13708A	T14634C	C14873A	T15784C				
10789A_Lima	G6261A	G9055A	A12972G	A14053G						
10720A_Ychsma	G6261A	G9055A	A14053G	C15647T						
10726A_Ychsma	G143A	T9078C	G10530A	A14053G	G16390A					
10730A_Ychsma	T131C	C4013T	G13708A	T14634C	T15784C					
10754A_Wari	A243G	C4013T	G8994A	G11016A	G12127A	A14059G	G16438A			
10742A_Wari	A178G	A6779G	A10420G	C16278T!	C16295T					
10750A_Wari	T152C!	G1797A	A9377G	G12127A	A14053G	C14428T				

Haplogroup C										
C1b										
10817A_Lima	A9468G	C12535T	T15313C	A16166C	T16223C					
10820A_Lima	A9468G	C12535T	T15313C	A16166C	T16223C					
10709A_Ychsma	C12535T	T15313C								
10717A_Ychsma	C150T	T152C!	C12535T	T15313C	C16292T					
10725A_Ychsma	C12535T	T15313C								
10731A_Ychsma	A636G	C3442T	C4496T	G5821A	A5894C	G16000A	T16362C			
10734A_Wari	C12535T	T15313C	A16166C	G16244T						
10763A_Wari	C12535T	T15313C								
10771A_Wari	A9468G	C12535T	T15313C	A16166C						
10773A_Wari	T8380C	T13281C	A13710C	A15244G	C15315T	A16207G				
10765A_Wari	C12535T	T15313C	A16166C	G16244T						
C1c										
10806A_Lima	T152C!	7472d	T8450C	C12774T	T16172C	T16297C				
C1d										
10810A_Ychsma	A14122G									
Haplogroup D										
D1										
10805A_Ychsma	T11365C	G11906A	T12481C	C12557T	A14665G					
10722A_Ychsma	T11365C	G11906A	T12481C	C12557T	A14665G					
D										
10811A_Lima	G143A	T2092C!	A6113G	T16189C!	G16274A	C16365T				
10732A_Ychsma	G143A	T2092C!	A6113G	T16189C!	G16274A	C16365T	D7 (new branch)			

Table 4. Radiocarbon dates from selected samples from Huaca Pucllana

ACAD	site	culture	relative date	hg	ORAU#	Delta 13 C	uncal BP	IntCal13 cal BC/AD	2-sigma	SH13Cal13
10791	Huaca Pucllana	Lima	100-650 AD	A2	OxA-31118	-11.79	1420±29	584-660 calAD (95.4%)	584-660	603-760 cal AD (95.4%)
10817	Huaca Pucllana	Lima	100-650 AD	C1b	OxA-31120	-13.8	1493±30	534-642 calAD (90.6%)	435-642	549-652 cal AD (95.4%)
10734	Huaca Pucllana	Wari	500-1000 AD	C1b	OxA-31422	-12.65	1156±22	776-968 calAD (95.4%)	776-968	891-988 cal AD (95.4%)
10754	Huaca Pucllana	Wari	500-1000 AD	B2b	OxA-31423	-12.61	955±65	974-1220 calAD (95.4%)	974-1220	1016-1264 cal AD (95.4%)
10709	Huaca Pucllana	Ychsma	1100-1440 AD	C1b	OxA-31424	-9.6	745±23	1244-1288 calAD (92.9%)	1226-1288	1271-1315 calAD (72.9%)
10805	Huaca Pucllana	Ychsma	100-650 AD	D1	OxA-31462	-13,38	762±23	1223 - 1280 calAD (95.4%)	1223-1280	1231-1379 cal AD (95.3%)
10810	Huaca Pucllana	Ychsma	100-650 AD	C1d	OxA-31119	-12.18	866±28	1149-1249 calAD (80%)	1048-1249	1164-1272 cal AD (95.4%)
10722	Huaca Pucllana	Ychsma	1100-1440 AD	D1	OxA-31425	-12.49	773±24	1221-1278 calAD (95.4%)	1221-1278	1227-1301 calAD (93.9%)

Radiocarbon dates were calibrated using the OxCal Program v.4.2 from University of Oxford Radiocarbon Accelerator Unit.

SUPLEMENTARY INFORMATION

Supplementary Table 1. List of populations used to perform the comparative analysis

MODERN POPULATIONS	n	Country	Region	Reference
Arequipa	**22**	Peru	Central Andean	Fuselli et al., 2003
San Martin	**21**	Peru	Central Andean	Fuselli et al., 2004
Tayacaja	**59**	Peru	Central Andean	Fuselli et al., 2005
Ancash	**33**	Peru	Central Andean	Lewis et al., 2005
Puno-Quechua	**30**	Peru	Central Andean (Highland)	Lewis et al., 2007
Yungay	**36**	Peru	Central Andean	Lewis et al., 2007
Tupe	**16**	Peru	Central Andean	Lewis et al., 2007
PunoAymara	**14**	Peru	Central Andean (Highland)	Lewis et al., 2007
Titicaca-Quechua	**37**	Peru	Central Andean (Highland)	Barbieri et al., 2011
Titicaca-Aymara	**20**	Peru	Central Andean (Highland)	Barbieri et al., 2012
Titicaca_Uros	**7**	Peru	Central Andean (Highland)	Barbieri et al., 2013
GranChaco	**204**	Paraguay	Gran Chaco	Cabana et al., 2006
Mapuche	**34**	Chile	South Andean	Moraga et al., 2000
Pehuenche	**24**	Chile	South Andean	Moraga et al., 2000
Yaghan	**15**	Chile	Tierra de Fuego	Moraga et al., 2000
Mapuche1	**18**	Chile	South Andean	de Saint Pierre et al., 2012
Yamana	**21**	Chile	Tierra de Fuego	de Saint Pierre et al., 2012
Tehuelche	**23**	Chile	South Andean	de Saint Pierre et al., 2012
Atacameno	**28**	Chile	Central Andean	de Saint Pierre et al., 2012
Aymara	**39**	Chile	Central Andean	de Saint Pierre et al., 2012
Pehuenche1	**41**	Chile	South Andean	de Saint Pierre et al., 2012
Huilliche	**47**	Chile	South Andean	de Saint Pierre et al., 2012
Quechua_Bol	**93**	Bolivia	Central Andean (Highland)	Gaya-Vidal et al., 2011
Aymara_Bol	**97**	Bolivia	Central Andean (Highland)	Gaya-Vidal et al., 2011
Gaviao	**27**	Brazil	Amazon	Ward et al., 1996
Xavante	**25**	Brazil	Amazon	Ward et al., 1996
Zoro	**30**	Brazil	Amazon	Ward et al., 1996
Arawaken	**29**	Colombia	North Western	Melton et al., 2007

ANCIENT POPULATIONS	n	Country	Region	Reference
Caverna6_EH	**7**	Peru	Central Andean (Coastal)	Fehren-Schmitz et al, 2014
Palpa_MH	**11**	Peru	Central Andean (Coastal)	Fehren-Schmitz et al, 2014
Palpa_EIP	**56**	Peru	Central Andean (Coastal)	Fehren-Schmitz et al, 2014
Laramate_LIP	**38**	Peru	Central Andean (Highland)	Fehren-Schmitz et al, 2014
Palpa_EH	**26**	Peru	Central Andean (Coastal)	Fehren-Schmitz et al, 2014
MonteGrande_EIP	**11**	Peru	Central Andean (Coastal)	Fehren-Schmitz et al, 2014
Laramate_MH	**39**	Peru	Central Andean (Highland)	Fehren-Schmitz et al, 2014
Palpa_LIP	**11**	Peru	Central Andean (Coastal)	Fehren-Schmitz et al, 2014
MonteGrande_LIP	**11**	Peru	Central Andean (Coastal)	Fehren-Schmitz et al, 2014
Conchopata_MH	**10**	Peru	Central Andean (Highland)	Kemp et al., 2009
Huari_LIP	**17**	Peru	Central Andean (Highland)	Kemp et al., 2009
Pernil Alto	**13**	Peru	Central Andean (Coastal)	Fehren-Schmitz et al, 2010
Tompullo2	**24**	Peru	Central Andean (Highland)	Fehren-Schmitz et al, 2014
Acchaymarca	**14**	Peru	Central Andean (Highland)	Fehren-Schmitz et al, 2014
Puca	**11**	Peru	Central Andean (Highland)	Fehren-Schmitz et al, 2014
Pampa Grande_EIP-MH"	**19**	Argentina	Central Andean (Highland)	Fehren-Schmitz et al, 2014
Huaca Pucllana – Lima	**9**	Peru	Central Andean (Coastal)	This Study
Huaca Pucllana – Wari	**10**	Peru	Central Andean (Coastal)	This Study
Huaca Pucllana – Ychsma	**15**	Peru	Central Andean (Coastal)	This Study

Supplementary Table 2. List of sample details collected from Huaca Pucllana archaeological site

ACAD	Sample Number	Sample	Sample Details	Collection Notes
10709	A15 01/02 Muestra 1a	Tooth	M4,8	Ychsma culture, adult
10710	A15 01/02 Muestra 1b	Tooth	M3,7	Ychsma culture, adult
10711	A15 01/02 Muestra 1c	Hair		Ychsma culture, adult
10712	A0 cf14 ind-1/98 Muestra 2a	Tooth	M3,6	Ychsma culture, infant
10713	A0 cf14 ind-1/98 Muestra 2b	Tooth	M4,6	Ychsma culture, infant
10714	A0 cf14 01/98 Muestra 2c	Hair		Ychsma culture, infant
10715	A01 CF16/98 Muestra 3a	tooth	C3,3	Ychsma culture, adult, female
10716	A01 CF16/98 Muestra 3b	tooth	M2,7	Ychsma culture, adult, female
10717	A0 CF15/01 Muestra 4a	tooth	C4,3	Ychsma culture, adult
10718	A0 CF15/01 Muestra 4b	tooth	M1,7	Ychsma culture, adult
10719	A15 CF36/01 Muestra 5a	tooth	M4,6	Ychsma culture, adult
10720	A15 CF36/01 Muestra 5b	tooth	C4,3	Ychsma culture, adult
10721	A15 CF36/01 Muestra 5c	Hair		Ychsma culture, adult
10722	A15 Sin Contexto Muestra 6a	tooth	M4,8	Ychsma culture, adult
10723	A15 Sin Contexto Muestra 6b	tooth	M3,8	Ychsma culture, adult
10724	A15 Sin Contexto Muestra 7a	tooth	M3,8	Ychsma culture, adult
10725	A15 Sin Contexto Muestra 7b	tooth	M4,8	Ychsma culture, adult
10726	A3 CF01/04 Muestra 13a	tooth	C3,3	Ychsma culture, adult
10727	A3 CF01/04 Muestra 13b	tooth	M3,8	Ychsma culture, adult
10728	A0 08/98 Muestra 49a	tooth	M1,8	Ychsma culture, adult, male
10729	A0 08/98 Muestra 49b	tooth	M2,8	Ychsma culture, adult, male
10730	A15 02/02 Muestra 50	bone	left femur	Ychsma culture, adult
10731	A0 56/97 Muestra 51	bone	right femur	Ychsma culture, adult
10732	A01 02/02 Muestra 52	bone	right humerus	Ychsma culture, adult
10793	A06 82/96 Muestra 35a	Tooth	C2,3	Ychsma culture, adult (20-24yrs), male
10794	A06 82/96 Muestra 35b	Tooth	M2,8	Ychsma culture, adult (20-24yrs), male
10795	A06 82/96 Muestra 35c	Hair		Ychsma culture, adult (20-24yrs), male

10796	A06 77/96 Muestra 36a	Tooth	C2,3	Ychsma culture, adult
10797	A06 77/96 Muestra 36b	Tooth	C1,3	Ychsma culture, adult
10800	A06 79/96 Ind1 Muestra 38a	Tooth	M3,7	Ychsma culture, adult (17-20yrs), male
10801	A06 79/96 Ind1 Muestra 38b	Tooth	M4,6	Ychsma culture, adult (17-20yrs), male
10804	A06 79/01 Ind2 Muestra 40a	Tooth	M3,6	Ychsma culture, adult (35-40yrs), male
10805	A06 79/01 Ind2 Muestra 40b	Tooth	M1,7	Ychsma culture, adult (35-40yrs), male
10809	A06 76/96 Muestra 42a	Tooth	M1,	Ychsma culture, adult
10810	A06 76/96 Muestra 42b	Tooth	M2,7	Ychsma culture, adult
10733	A20 08/08 Ind2 Muestra 8a	tooth	M3,6	Wari culture, adult (30-35yrs), male
10734	A20 08/08 Ind2 Muestra 8b	tooth	M4,6	Wari culture, adult (30-35yrs), male
10735	A20 08/08 Ind2 Muestra 8c	Hair		Wari culture, adult (30-35yrs), male
10736	A20 08/08 Ind1 Muestra 9a	tooth	M2,6	Wari culture, infant (1-2yrs), female
10737	A20 08/08 Ind1 Muestra 9b	tooth	M1,6	Wari culture, infant (1-2yrs), female
10738	A20 08/08 Ind3 Muestra 10a	tooth	P1,5	Wari culture, adult (40-50yrs), female
10739	A20 08/08 Ind3 Muestra 10b	tooth	M1,8	Wari culture, adult (40-50yrs), female
10740	A20 08/08 Ind3 Muestra 10c	Hair		Wari culture, adult (40-50yrs), female
10741	A20 CF003/09 Muestra 11a	tooth	M3,6	Wari culture, adult (55-60yrs), female
10742	A20 CF003/09 Muestra 11b	tooth	M1,7	Wari culture, adult (55-60yrs), female
10743	A20 CF003/09 Muestra 11c	Hair		Wari culture, adult (55-60yrs), female
10744	A20 11/08 Ind1 Muestra 12a	tooth	M3,7	Wari culture, adult (45-55yrs), female
10745	A20 11/08 Ind1 Muestra 12b	tooth	M4,6	Wari culture, adult (45-55yrs), female
10746	A20 11/08 Ind1 Muestra 12c	Hair		Wari culture, adult (45-55yrs), female
10747	A20 07/08 Ind6 Muestra 14a	tooth	M8,5	Wari culture, infant (6 +/- 2yrs), maybe male
10748	A20 07/08 Ind6 Muestra 14b	tooth	M6,5	Wari culture, infant (6 +/- 2yrs), maybe male
10749	A20 07/08 Ind4 Muestra 15a	tooth	M2,8	Wari culture, adult (35-45yrs), male
10750	A20 07/08 Ind4 Muestra 15b	tooth	M4,8	Wari culture, adult (35-45yrs), male
10751	A20 07/08 Ind5 Muestra 16a	tooth	M8,5	Wari culture, infant (6 +/- 2yrs), male

10752	A20 07/08 Ind5 Muestra 16b	tooth	M5,5	Wari culture, infant (6 +/- 2yrs), male
10753	A20 05/08 Muestra 17a	tooth	M4,6	Wari culture, adult (50-60yrs), male
10754	A20 05/08 Muestra 17b	tooth	M2,7	Wari culture, adult (50-60yrs), male
10755	A20 05/08 Muestra 17c	Hair		Wari culture, adult (50-60yrs), male
10756	A20 01/05 Ind5 Muestra 18a	Tooth	C4,3	Wari culture, adult (50-60yrs), male
10757	A20 01/05 Ind5 Muestra 18b	Tooth	M3,8	Wari culture, adult (50-60yrs), male
10758	A20 06/08 Ind1 Muestra 19a	tooth	C2,3	Wari culture, adult (16-20yrs), male
10759	A20 06/08 Ind1 Muestra 19b	tooth	C4,3	Wari culture, adult (16-20yrs), male
10760	A20 06/08 Ind1 Muestra 19c	Hair		Wari culture, adult (16-20yrs), male
10761	A20 CF8/09 Ind1 Muestra 20a	Tooth	M2,7	Wari culture, adult (25-30yrs), male
10762	A20 CF8/09 Ind1 Muestra 20b	Tooth	M4,7	Wari culture, adult (25-30yrs), male
10763	A20 14/08 Muestra 21a	Tooth	M4,7	Wari culture, adult (25-30yrs), female
10764	A20 14/08 Muestra 21b	Tooth	M3,8	Wari culture, adult (25-30yrs), female
10765	A20 18/08 Muestra 22a	tooth	M3,8	Wari culture, adult (45-50yrs), female
10766	A20 18/08 Muestra 22b	tooth	M4,6	Wari culture, adult (45-50yrs), female
10767	A20 18/08 Muestra 22c	Hair		Wari culture, adult (45-50yrs), female
10768	A20 17/08 Muestra 23a	Tooth	M2,7	Wari culture, adult (35-40yrs), male
10769	A20 17/08 Muestra 23b	Tooth	M1,7	Wari culture, adult (35-40yrs), male
10770	A20 02/07 Ind2 Muestra 24a	Tooth	M1,6	Wari culture, adult (50-60yrs), male
10771	A20 02/07 Ind2 Muestra 24b	Tooth	C2,3	Wari culture, adult (50-60yrs), male
10772	A20 01/09 Ind1 Muestra 25a	Tooth	M2,7	Wari culture, adult
10773	A20 01/09 Ind1 Muestra 25b	Tooth	P2,5	Wari culture, adult
10774	A20 01/09 Ind2 Muestra 26a	Tooth	M1,6	Wari culture, adult, female
10775	A20 01/09 Ind2 Muestra 26b	Tooth	P1,5	Wari culture, adult, female
10776	A20 04/07 Ind1 Muestra 27a	Tooth	M2,6	Wari culture, adult (35-45yrs), male
10777	A20 04/07 Ind1 Muestra 27b	Tooth	M3,8	Wari culture, adult (35-45yrs), male

10778	A20 04/07 Ind2 Muestra 28a	Tooth	M4,6	Wari culture, subadult (12-15yrs), maybe female
10779	A20 04/07 Ind2 Muestra 28b	Tooth	M3,6	Wari culture, subadult (12-15yrs), maybe female
10780	A18 02/09 Muestra 29a	Tooth	M3,7	Lima culture, adult (25-30yrs), male
10781	A18 02/09 Muestra 29b	Tooth	M3,3	Lima culture, adult (25-30yrs), male
10782	A18 02/09 Muestra 29c	Hair		Lima culture, adult (25-30yrs), male
10783	A06 65 Ind2 Muestra 30a	Tooth	M3,8	Lima culture, adult (30-40yrs), female
10784	A06 65 Ind2 Muestra 30b	Tooth	M1,6	Lima culture, adult (30-40yrs), female
10785	A06 90 Ind2 Muestra 31a	Tooth	M1,6	Lima culture, adult (35-40yrs), male
10786	A06 90 Ind2 Muestra 31b	Tooth	M2,8	Lima culture, adult (35-40yrs), male
10787	A06 90 Ind1 Muestra 32a	Tooth	M1,7	Lima culture, adult (40-45yrs), male
10788	A06 90 Ind1 Muestra 32b	Tooth	M2,6	Lima culture, adult (40-45yrs), male
10789	A20 05/09 Muestra 33a	Tooth	C3,3	Lima culture, adult (20-40yrs), male
10790	A20 05/09 Muestra 33b	Tooth	P4,5	Lima culture, adult (20-40yrs), male
10791	A20 03/07 Ind1 Muestra 34a	Tooth	P4,4	Lima culture, adult (30-40yrs), female
10792	A20 03/07 Ind1 Muestra 34b	Tooth	M4,6	Lima culture, adult (30-40yrs), female
10798	A06 84/01 Ind2 Muestra 37a	Tooth	M3,7	Lima culture, adult (30-35yrs), female
10799	A06 84/01 Ind2 Muestra 37b	Tooth	M2,6	Lima culture, adult (30-35yrs), female
10802	A06 95/96 Muestra 39a	Tooth	M3,7	Lima culture, adult (50-55yrs), female
10803	A06 95/96 Muestra 39b	Tooth	P2,4	Lima culture, adult (50-55yrs), female
10806	A06 102/96 Muestra 41a	Tooth	M3,6	Lima culture, adult
10807	A06 102/96 Muestra 41b	Tooth	M4,6	Lima culture, adult
10808	A06 102/96 Muestra 41c	Hair		Lima culture, adult
10811	A18 02/02 Muestra 43a	Tooth	M1,7	Lima culture, adult
10812	A18 02/02 Muestra 43b	Tooth	M3,7	Lima culture, adult
10813	A15 06/00 Muestra 44a	Tooth	M3,7	Lima culture, adult
10814	A15 06/00 Muestra 44b	Tooth	M4,7	Lima culture, adult
10815	A06 75/01 Ind2 Muestra 45a	Tooth	P2,5	Lima culture, adult (35-45yrs), female

10816	A06 75/01 Ind2 Muestra 45b	Tooth	P1,5	Lima culture, adult (35-45yrs), female
10817	A06 01/02 Muestra 46a	Tooth	M4,7	Lima culture, adult
10818	A06 01/02 Muestra 46b	Tooth	M3,7	Lima culture, adult
10819	A06 75/01 Ind1 Muestra 47a	Tooth	M1,7	Lima culture, adult (20-30yrs), female
10820	A06 75/01 Ind1 Muestra 47b	Tooth	M2,7	Lima culture, adult (20-30yrs), female
10821	A6 68/96 Muestra 48a	Tooth	M6,5	Lima culture, adult
10822	A6 68/96 Muestra 48b	Tooth	M5,5	Lima culture, adult
10823	A6 68/96 Muestra 48c	Hair		Lima culture, adult

CHAPTER 5

AmericaPlex26–A SNaPshot multiplex system for genotyping the main human mitochondrial founder lineages of the Americas

Statement of Authorship

Chapter 5

Title of the Paper	**AmericaPlex26: A SNaPshot Multiplex System for Genotyping the Main Human Mitochondrial Founder Lineages of the Americas**
Publication status	**Published**, Accepted for Publication, Submitted for Publication, Publication style
Publication details	PLoS ONE 03/2014; 9(3):e93292. DOI: 10.1371/journal.pone.0093292

Author Contributions
By signing the Statement of Authorship, each author certifies that their stated contribution to the publication is accurate and that permission is granted for the publication to be included in the candidate's thesis.

Name of the principal Author	Alexandra Coutinho	
Contribution to the paper	Performed DNA extractions, PCR amplifications, sequencing and downstream data processing on samples, interpreted the results, conducted the analysis, revised and wrote the manuscript	
Signature		Date: 02/06/2015

Name of the co-author	**Guido Valverde** (Candidate)	
Contribution to the paper	Performed DNA extractions, PCR amplifications, sequencing and downstream data processing on samples, interpreted the results, revised the manuscript	
Signature		Date: 8/06/15

Name of the co-author	Lars Fehren-Schmitz	
Contribution to the paper	Contributed with data generation and interpretation, revised and edited the manuscript	
Signature		Date: 06/01/2015

Name of the co-author	Alan Cooper	
Contribution to the paper	Contributed with data generation and interpretation, revised and edited the manuscript	
Signature		Date: October, 3 2015

Name of the co-author	María Inés Barreto	
Contribution to the paper	Contributed reagents, materials, analysis tools. Assisted with sample collection, provided archaeological and context information.	
Signature		Date: May 29 2015

Name of the co-author	Isabel Flores Espinoza	
Contribution to the paper	Provided samples for analysis. Contributed reagents, materials, analysis tools, assisted with sample collection, provided archaeological and context information.	
Signature		Date 02.06.2015

Name of the co-author	Bastien Llamas	
Contribution to the paper	Co-developed the research concept and experimental design of the study, gave advice on laboratory work, assisted in the interpretation of results and edited the manuscript	
Signature		Date: October, 23 2015

Name of the co-author	Wolfgang Haak	
Contribution to the paper	Co-developed the research concept and experimental design of the study, gave advice on laboratory work, assisted in the interpretation of results and edited the manuscript	
Signature		Date: 20/10/2015

OPEN ACCESS Freely available online

PLOS ONE

AmericaPlex26: A SNaPshot Multiplex System for Genotyping the Main Human Mitochondrial Founder Lineages of the Americas

Alexandra Coutinho[1], Guido Valverde[1], Lars Fehren-Schmitz[2,3], Alan Cooper[1], Maria Inés Barreto Romero[4], Isabel Flores Espinoza[4], Bastien Llamas[1], Wolfgang Haak[1]*

1 Australian Centre for Ancient DNA, School of Earth & Environmental Sciences, The University of Adelaide, Adelaide, South Australia, Australia, **2** Historical Anthropology and Human Ecology, Johann-Friedrich-Blumenbach Department of Zoology and Anthropology, University Goettingen, Goettingen, Germany, **3** Department of Anthropology, University of California Santa Cruz, Santa Cruz, California, United States of America, **4** Museo de Sitio Huaca Pucllana, Miraflores, Lima, Peru

Abstract

Phylogeographic studies have described a reduced genetic diversity in Native American populations, indicative of one or more bottleneck events during the peopling and prehistory of the Americas. Classical sequencing approaches targeting the mitochondrial diversity have reported the presence of five major haplogroups, namely A, B, C, D and X, whereas the advent of complete mitochondrial genome sequencing has recently refined the number of founder lineages within the given diversity to 15 sub-haplogroups. We developed and optimized a SNaPshot assay to study the mitochondrial diversity in pre-Columbian Native American populations by simultaneous typing of 26 single nucleotide polymorphisms (SNPs) characterising Native American sub-haplogroups. Our assay proved to be highly sensitive with respect to starting concentrations of target DNA and could be applied successfully to a range of ancient human skeletal material from South America from various time periods. The AmericaPlex26 is a powerful assay with enhanced phylogenetic resolution that allows time- and cost-efficient mitochondrial DNA sub-typing from valuable ancient specimens. It can be applied in addition or alternative to standard sequencing of the D-loop region in forensics, ancestry testing, and population studies, or where full-resolution mitochondrial genome sequencing is not feasible.

Citation: Coutinho A, Valverde G, Fehren-Schmitz L, Cooper A, Barreto Romero MI, et al. (2014) AmericaPlex26: A SNaPshot Multiplex System for Genotyping the Main Human Mitochondrial Founder Lineages of the Americas. PLoS ONE 9(3): e93292. doi:10.1371/journal.pone.0093292

Editor: Alessandro Achilli, University of Perugia, Italy

Received October 12, 2013; **Accepted** March 4, 2014; **Published** March 26, 2014

Funding: This work was funded by the Australian Research Council (Discovery Project 1095782) awarded to WH and BL. BL also held an Australian Research Council Postdoctoral Fellowship. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: wolfgang.haak@adelaide.edu.au

Introduction

Population genetic studies on modern-day Native American populations have described the presence of five haplogroups (hgs), termed A, B, C, D and X [1–3]. These five hgs are shared with East Asian populations and support an entry route to the Americas via the Bering landmass. However, Native American populations can be distinguished from their East Siberian source populations by exhibiting distinct sub-haplogroups (sub-hgs), which can only be found in the Americas. These so-called 'founder lineages' have been used to describe the demographic history of Native American populations and to shed light on the timing of the entry into and spread throughout the Americas [4–6]. The fact that the mtDNAs of all human populations native to the Americas can be assigned to one of the founder lineages pertains to stochastic events that would have affected the initial colonizers of the Americas [7]. The low genetic variation found in modern Native American groups is believed to be due to either population bottlenecks or genetic drift [8,9].

Most mitochondrial DNA (mtDNA) studies on prehistoric American populations involve sequencing of the D-loop, which contains Hypervariable Regions 1 and 2 (HVR1 and HVR 2 respectively), to describe a sequence haplotype, from which the hg can be inferred [10,11]. Sequencing of the HVR regions of mtDNA was relatively cost-effective and less time consuming than full mtDNA sequencing, and is therefore still the method of choice for many labs which study human populations [11]. Yet not all lineages harbour enough variation in the D-loop from which to infer a sub-hgs at a deeper level than the overall hg, let alone a specific founder lineage [11–14]. As a result, many past and present studies on Native American population history have been restricted to the information gained from the distribution of the major five Pan-American hgs.

The coupling of multiplex polymerase chain reaction (PCR) with a Single Base Extension (SBE) reactions, based on the established SNaPshot (Applied Biosystems) or minisequencing principle, has been widely used to design panels of single nucleotide polymorphisms (SNPs) for forensic and anthropological studies [15]. It has also found wide use in population genetic studies focussing on mtDNA and Y-chromosome SNPs, either including SNPs with a global representation or via a targeted selection of characteristic SNPs representing specific geographic regions [13,16–21]. The design of a SNP panel including those markers defining the 15 American founder lineages described by Perego et al. [22] and more had not been attempted, although 'Multiplex 3' in van Oven et al. 2011 [23] covered 12 out of these

15. The primary aim of this study was therefore to design a novel SNaPshot assay that enables a fast and cost-efficient high-resolution typing of the majority of known Native American sub-hgs by targeting 26 characteristic SNPs. Our goal was to develop an assay that is universally applicable to accommodate the specific needs of damaged and degraded DNA in ancient DNA work and forensics. Selective sequencing of the SNP regions of interest not only allows for flexibility in the number and choice of SNP sites but also allows (with reservations) the design of ultra-short amplicon lengths (50–80 bp) suitable for degraded DNA typing [13], while using far less DNA than traditional sequencing methods or SNP-typing in individual singleplex PCRs. This is of great importance in forensic and ancient DNA studies where sample DNA is a limited resource [23,24]. The secondary aim was to develop an assay that could complement an established assay with a global set of SNPs (GenoCore22, see [24] but also [23]) and at the same time provide a fast and efficient screening tool that allows the assessment of overall sample quality (presence of very short fragments of endogenous mtDNA and absence of contaminant hgs) for further use in mitochondrial genome sequencing via DNA library preparation and targeting enrichment techniques, e.g. [25–27].

Materials and Methods

AmericaPlex26 SNP selection

We developed a multiplex SNaPshot reaction targeting 26 SNP sites in total including characteristic SNPs of the four major Pan-American sub-hgs A2, B2, C1 and D1, as well as SNP sites for the minor Pan-American lineages C4c, D2a, D4h3a and X2a [6]. The initial choice of SNPs was based on a study by Perego et al. [22] describing 15 American founder lineages. Additional SNP sites were chosen for sub-hgs within each major hg based on the most up-to-date mtDNA phylogeny available at the time (phylotree.org, mtDNA tree Build 13, 28 Dec 2011) in order to enhance the discriminating power of the assay. For sub-hgs defined by more than one characteristic SNP we employed selection criteria during the primer design stage based on the ability to design primers with high specificity in the short flanking region around the SNPs, and under a consensus-melting temperature for all pairs in a multiplex environment.

Presented below is a summary of each major Native American sub-hg, their distribution throughout the Americas, as well as the SNP sites chosen to represent the hg and their respective sub-hgs. The representative SNPs typed in the AmericaPlex26 are given in parentheses and a simplified tree illustrating the phylogenetic relationship is shown in Figure 1.

Haplogroup A2 (G12007A) is found throughout the Americas, but its derivatives A2a (C3330T) and A2b (T11365C) are mainly found in the Northern parts of North America in Inuit, Na-Dené and Siberian populations such as Koryaks and Chukchi [22,28–30], whereas particular subgroups of A2a were also reported from Athapaskan territories in the Southwest [30]. A16265G (defining A2b in [22]) was also added to the assay as it was further resolved to represent sub-hg A2b1, which can be found in Eskimoan-speaking populations (such as the Inuit and Yupik) across the Arctic [31].

Haplogroup B2 (C11177T) is widely distributed throughout the Americas and is nested within hg B4b (G4820A), which has a largely Asian distribution. Studies on modern-day populations describe sub-hgs within B2, which are prevalent in specific geographic areas within the American continents and thus phylogeographically informative [32]. We included diagnostic SNPs for sub-hgs B2a (G16483A, North/Central), B2b (G6755A, ubiquitous), B2c (A7241G, predominantly Central), B2d (T8875C, Central/South), B2e (C6119T, South) and B2f (T10535C, North/Central) to monitor this phylogeographic pattern in ancient Native American populations through time.

Haplogroup C is represented in the Americas by sub-hgs C1b, C1c and C1d [22,32], whereas sub-hgs C1a and C1e are Siberian/East Asian and European sister-clades, respectively. Sub-hg C1b (A493G) can be found throughout South America. Sub-hg C1c is most frequent in Mexico [22] and was split into sub-hgs C1c1a (A12978G) and C1c2 (C14356T), since the immediate flanking region of the two C1c SNPs defining (G1888A and G15930A) were not suitable for primer design. A recent study by Perego et al. [22] has further resolved Central American sub-hg C1d (A16051G), which now includes sub-hg C1d1 (G7697A). Minor Pan-American hg C4c (C14433T) was recently discovered in an ancient sample from British Columbia, and was found to be one of the founding lineages of the Americas based on coalescent age estimates [33,34].

Haplogroup D is found in the Americas as four distinct sub-hgs D1 (C2092T), D2a, D3 and D4h3a [6,35]. Sub-hg D2a has been found in Sireniki, Yuit, Aleut and Chukchi populations in Siberia and Aleuts from the Commander Islands [31,36,37], and has been further resolved into sub-hgs D2a1 (A9667G; including the Saqqaq Paleo-Eskimo [31]), D2a2 (G4991A) and D2b (A9181G) [22]. Recent studies have further revealed sub-hg D3 to be more derived than previously thought and nested within the larger branch D4b1 (C10181T), which currently encompasses three sub-hgs within North American and Siberian native populations: D4b1a, D4b1b′d, and D4b1c [31]. Therefore, we chose the ancestral SNP defining hg D4b1 for our panel (See mtDNA tree Build 13, 2011). In addition, sub-hg D4h3a (G6285A) is distributed along the Pacific coast of the American continents [6] but most frequent in South America.

Haplogroup X2a (A8913G) has only been found in a limited number of samples in North American populations as compared to those of A2, B2, C1 and D1, and is therefore described as minor founding lineage in this paper [8,35].

Lastly, SNP site T14783C was included as control to define macro-hg M, which encompasses hgs C and D. In contrast, this SNP retains the ancestral state in hgs A, B and X, which belong to macro-hg N.

Primer and probe design

PCR and SBE primers were designed and quality-controlled using default settings and features in the software package Geneious v5.2 (Geneious version (5.2) created by Biomatters. Available from http://www.geneious.com/) and Batchprimer3 v1.0 [38], both based on the program primer3 [39], generally following the guidelines set out in Sanchez et al. 2006 [40]. Amplicon sizes were deliberately kept smaller than 90 bp in length to allow amplification of highly fragmented DNA as typical in forensics and ancient DNA studies. Given how short the flanking regions of each SNP were, which already constrained our selection of suitable SNPs, we could not consider potential polymorphic sites in these areas nor nuclear insertions, and relied on empirical testing of PCR primer efficiency. SBE primers were then ranked according to quality score and orientation (forward or reverse) for efficient use of fluorescent dyes and fragment length spacing. The latter was adjusted to 4 bp by adding poly-CT tails to the 5′end of each SBE primer (Table 1) [41].

Ethics statement

All necessary permits were obtained for the described study, which complied with all relevant regulations. Permissions to

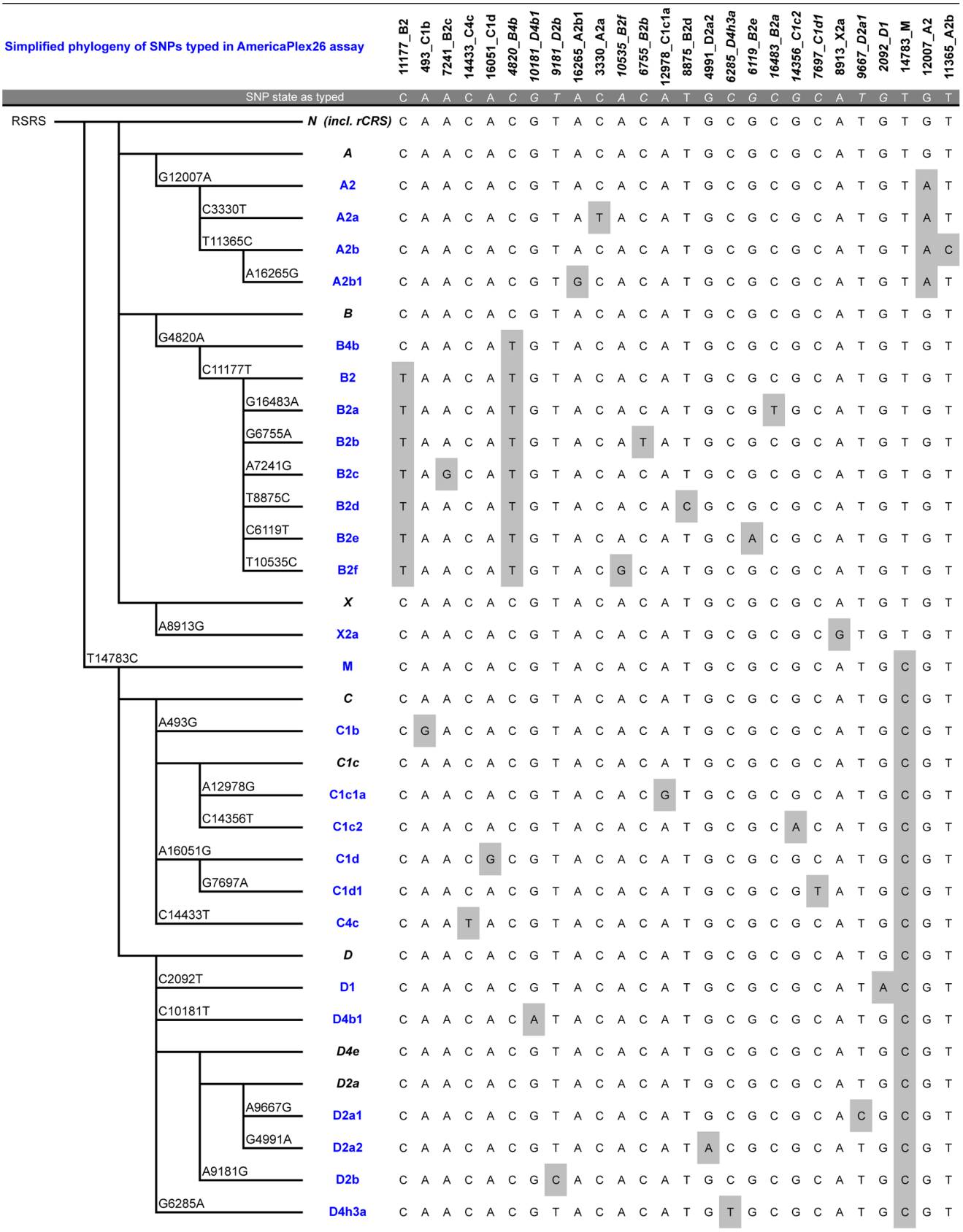


Figure 1. Simplified phylogeny in related to the Reconstructed Sapiens Reference Sequence (RSRS [52]**) and typing scheme of the 26 SNPs targeted in the AmericaPlex26.** Sub-haplogroups, which can be unambiguously assigned, are shown in blue (blue). Basal hgs, which cannot be unambiguously assigned, are also shown (black) in order to illustrate the phylogenetic relationship, but also the inherent limitations of our assay. The phylogenetic position of the revised Cambridge Reference Sequence (rCRS [53]) is indicated within macro-hg N. SNPs with a SBE primer in reverse direction targeting the opposite strand are given in Italics.
doi:10.1371/journal.pone.0093292.g001

collect, export and analyze ancient Peruvian specimens from the Huaca Pucllana site were granted by the Ministry of Culture (the former National Institute of Cultural Heritage – INC) and the National Museum of Archaeology, Anthropology and History of Peru (MNAAHP) and are available on request (ACTA No 017-2010-ARMC-MNAAHP-MC and Resolución Viceministerial No. 120-2010-VMPCIC-MC). No specific permits were required for the modern control samples when solely used for methodological validation (waiver from The Human Research Ethics Committee (HREC) at the University of Adelaide). Swab samples were nevertheless collected using written informed consent.

Samples and DNA extractions

DNA from modern control samples (AC, GV) was extracted from cheek swabs using QIAamp DNA Mini Kit (Qiagen) a following the manufacturer's instructions. Ancient samples were collected by MIBR, BL and WH under DNA-free conditions at the Museo de Sitio Huaca Pucllana, Calle General Borgoño cuadra 8 s/n, Miraflores, Lima, Perú, where the samples are stored. Sample preparation, DNA extractions and PCR amplification from ancient samples were performed at the Australian Centre for Ancient DNA in Adelaide, Australia, applying established methods and authentication criteria as described previously [26,42,43]. In brief, we used an in-house silica extraction method, detailed in [26], to extract DNA from two independent samples per individual. PCR amplifications from each extract and direct sequencing of the HVR-I were performed using four overlapping primer pairs with reaction conditions described in [42,44]. Details of the four primer pairs are given in (Table 2).

Multiplex PCR amplification

PCR amplifications were carried out in a final reaction volume of 12.5 µl consisting of 0.5 µL DNA sample (3 µL for ancient DNA), 1x PCR Gold Buffer, 6.5 mM $MgCl_2$, 0.1 U AmpliTaq Gold DNA polymerase (all Applied Biosystems) 1.25 mM dNTP solution (Bioline Pty Ltd), (0.8 µg RSA for ancient DNA samples), and a primer mix consisting of 26 primer pairs, with concentrations given in Table 1. PCR was carried out on a Tetrad 2 Peltier Thermal Cycler (Bio-Rad Laboratories) using the following conditions: 95°C for 6 min and 30 cycles (45 cycles for ancient DNA samples) of 95°C for 30 s, 55°C for 30 s, 65°C for 30 s, and a final extension time at 65°C for 6 min. Amplification success was monitored via gel electrophoresis on an 3.5% agarose gel (100 V for 40 min; Hyperladder V DNA size ladder (Bioline Pty Ltd)). PCR products were purified by mixing 5 µl of PCR reaction with 1 U ExoSAP-IT (Thermo Fisher Scientific Australia Pty Ltd), followed by incubation at 37°C for 50 min, 80°C for 15 min and 15°C for 10 min. Single Base Extension reactions consisted of a final volume of 5 µL containing 1 µL PCR product, 2.5 µL SNaPshot ready reaction mix (Applied Biosystems), and 0.5 µL extension primer mix (individual concentrations are given in Table 1). Thermocycling of the SBE reactions was performed in a Tetrad 2 Peltier Thermal Cycler (Bio-Rad Laboratories) with the following conditions: 96°C for 10 s; followed by 35 cycles of 55°C for 5 s and 60°C for 30 s. SBE products were purified by adding 1 U Shrimp Alkaline Phosphatase (Thermo Fisher Scientific) to the reaction solution and incubating it at 37°C for 50 min, 80°C for 15 min and 15°C for 10 min.

Capillary electrophoresis was performed on a 3130 xl Genetic Analyser (Applied Biosystems) using POP-6 polymer and a customised run module, by adding 1 µL sample DNA to 18.5 µL Hi-Di Formamide and 0.5 µL GeneScan-120 LIZ internal size standard (Applied Biosystems). Electropherograms were analysed using the software Genemapper ID version 3.2.1 software (Applied Biosystems) applying custom panel and bin settings available on request.

Sensitivity tests

We performed sensitivity studies using serial dilutions of 1, 1:10, 1:100, 1:1000 and 1:10,000 of DNA from a buccal swab sample from a lab member (AC). The mtDNA copy number of the modern sample was determined through qPCR using the SYBR-Green kit (Qiagen), targeting a short 77 bp fragment of human mitochondrial DNA with primer pair L13258 and H13295 [45]. Serial dilutions were treated as separate samples, and each sample was analysed in triplicate. The qPCR reaction was performed in a total reaction volume of 10 µL consisting of 1 µL of each sample dilution, 2x Brilliant SYBR Green Master Mix and 0.1 µM of each primer. The qPCR were carried out on a Rotor-Gene Q Real-Time PCR cycler (Qiagen) with thermocycling conditions as follows: 95°C for 5 min, followed by 45 cycles of 95°C for 10 s, 58°C for 20 s and 72°C for 20 s.

Results and Discussion

Optimization of the multiplex protocol

The AmericaPlex26 assay was initially tested with default concentrations of 0.017 µM for each primer (3 µL of 25 µM stock) and 0.015 µM for each SBE primer (3 µL of 50 µM stock) to assess the generic efficiency of primers or probes when used in the multiplex assay. Twenty-two out of 26 SNP sites could be readily amplified, albeit with highly variable peak heights across the assay. Primers and probes for the four problematic SNP sites were each tested in singleplex PCR and SNaPshot reactions to ensure they performed individually as expected. If the SNP fragment were successfully amplified in the singleplex PCR, concentrations of the primer would be doubled in the following multiplex PCR reaction mix.

We chose 3000 relative fluorescence units (rfu) as a default average peak height based on the ancestral allele status observed in our European modern control sample (AC), and calculated the percentage difference between peaks and the 3000 rfu average. Multiplex primer concentrations were adjusted according to this percentage difference to allow amplification of problematic SNP sites and to balance the peak heights of those that did amplify. Based on poor performance of the primer pair chosen to amplify the C1b SNP site (A493G), we performed a second round of balancing primer concentration with a new primer pair for this site.

To further refine the balance in peak height, the concentration of some SBE extension primers was adjusted to the final recommended concentrations given in Table 1. Changes to probe concentrations resulted in a more balanced electropherogram and amplification of all 26 SNP sites using modern buccal swab and ancient DNA samples (Figure 2).

Sensitivity studies

The amount of mitochondrial DNA was measured for a modern sample (1,171,699 copies/µL) and four serial dilutions of 1:10, 1:100, 1:1000 and 1:10,000 using real-time quantitative PCR (Figure 3). A near complete SNP profile could be observed for serial dilutions up to 28,278 copies/µL DNA (1:100), which is similar to other published multiplex assays [23,24,46].

Method application on ancient samples

The AmericaPlex26 assay was tested on ancient samples from three successive pre-Columbian cultures from the Huaca Pucllana archaeological site in Lima, Peru. They included samples from the

Table 1. Table showing the details of multiplex primers and SBE probes used in the AmericaPlex26 assay.

		PCR Amplification				Single-Base Extension			
Site	**hg**	**Primer**	**Primer Sequences (5′ to 3′)**	**Conc (μM)**	**length (bp)**	**SBE Probe Sequences (5′ to 3′)**	**Conc (μM)**	**nt**	**Alleles (Dyes) as detected**
12007	A2	L11984	CTAGTCACAGCCCTATACTCCCTCT	0.025	73	(ct)$_{26}$cCCTCTACATATTTACCACAACACAATG	0.015	80	G (blue), A (green)
		H12009	TGTTAATGTGGTGGGTGAGTGAGC	0.025					
3330	A2a	L03326	ACATACCCATGGCCAACCTCCT	0.022	60	(ct)$_{14}$cCCATGGCCAACCTCCTACT	0.015	48	C (yellow), T (red)
		H03339	GGAATGCCATTGCGATTAGAATGGGT	0.022					
11365	A2b	L11359	GCCAACAACTTAATATGACTAGCTTACACA	0.033	80	(ct)$_{25}$CAACTTAATATGACTAGCTTACACAATAGC	0.020	80	T (red), C (yellow)
		H11381	GGGAGTCATAAGTGGAGTCCGTAAAGAGG	0.033					
16265	A2b1	L16262	ACTGCAACTCCAAAGCCACCCC	0.011	56	(ct)$_{15}$ACTCCAAAGCCACCCCTC	0.015	48	A (green), G (blue)
		H16272	GGGTGGGTAGGTTTGTTGGTATCCT	0.011					
11177	B2	L11163	CCCACCTTGGCTATCATCACCCG	0.025	63	(ct)$_{7}$CCCGATGAGGCAACCAG	0.020	32	T (red), C (yellow)
		H11182	GTATGTGCCTGCGTTCAGGCGT	0.025					
16483	B2a	L16471	GCTCCGGGCCCATAACACTTGG	0.030	70	(ct)$_{21}$ACCAGATGTCGGATACAGTTCA	0.020	64	C (yellow), T (red)
		H16494	ACCCTGAAGTAGGAACCAGATGTCGG	0.030					
6755	B2b	L06750	GTCTGAGCTATGATATCAATTGGCTTCC	0.032	90	*(ct)$_{16}$TGGTGTGCTCACACGATAAA*	0.015	52	C (yellow), T (red)
		H06789	TGCTCGTGTGTCTACGTCTATTCC	0.032					
7241	B2c	L07224	TCCGGAATGCCCCGACGTTACT	0.023	68	(ct)$_{9}$TCGGACTACCCCGATGC	0.015	36	A (green), G (blue)
		H07243	ACAGATGATAGGATGTTTCATGTGGTGT	0.023					
8875	B2d	L08864	TCCCCTTATGAGCGGGCACAGT	0.029	76	(ct)$_{18}$CGGGCACAGTGATTATAGGC	0.015	56	T (red), C (yellow)
		H08896	TGTGGTAAGAAGTGGGCTAGGGC	0.029					
6119	B2e	L06094	TCGTCACAGCCCATGCATTTGT	0.031	76	*(ct)$_{22}$CCAAAGCCTCCGATTATGAT*	0.015	64	G (blue), A (green)
		H06133	AGTCAGTTGCCAAAGCCTCCGA	0.031					
10535	B2f	L10528	AGCATTTACCATCTCACTTCTAGGAATACT	0.031	62	*(ct)$_{16}$GAGGATATGAGGTGTGAGCG*	0.020	52	A (green), G (blue)
		H10537	GTAGGGAGGATATGAGGTGTGAGC	0.031					
4820	B4b	L04816	GCCCCCTTTCACTTCTGAGTCCC	0.033	56	*(ct)$_{11}$AGGGGTGCCTTGGGTAAC*	0.020	40	C (yellow), T (red)
		H04828	CCGGATGTCAGAGGGGTGCCTT	0.033					
8913	X2a	L08905	CGCTCTAAGATTAAAAATGCCCTAGCCC	0.003	59	(ct)$_{26}$AATGCCCTAGCCCACTTCTT	0.015	72	A (green), G (blue)
		H08915	AGGGGTGTAGGTGTGCCTTGTG	0.003					
14783	M	L14774	ACCCCAATACGCAAAACTAACCCCC	0.031	76	(ct)$_{26}$CGCAAAACTAACCCCCTAATAAAA	0.020	76	T (red), C (yellow)
		H14804	TGTTGGATGGGGTGGGGAGGTC	0.031					
493	C1b	L00474	TTTCCCCTCCCACTCCCATACT	0.033	73	ctcACTCCCATACTACTAATCTCATCAATACA	0.025	32	A (green), G (blue)
		H00511	TAGCAGCGGTGTGTGTGTGCTG	0.033					
12978	C1c1a	L12972	ACGCTAATCCAAGCCTCACCCCA	0.018	65	(ct)$_{18}$CCAAGCCTCACCCCACTACT	0.015	56	A (green), G (blue)
		H12993	TTGGGCTGATTTGCCTGCTGCT	0.018					
14356	C1c2	L14348	ACCACAACCACCACCCCATCAT	0.015	67	*(ct)$_{24}$GTAGGATTGGTGCTGTGGGT*	0.015	68	G (blue), A (green)

Table 1. Cont.

		PCR Amplification				Single-Base Extension			
Site	**hg**	**Primer**	**Primer Sequences (5′ to 3′)**	**Conc (μM)**	**length (bp)**	**SBE Probe Sequences (5′ to 3′)**	**Conc (μM)**	**nt**	**Alleles (Dyes) as detected**
		H14371	TGGGGTTAGCGATGGAGGTAGGA	0.015					
16051	C1d	L16049	TCTTTCATGGGGAAGCAGATTTGGG	0.019	62	(ct)^{11}GGGGAAGCAGATTTGGGT	0.015	40	A (green), G (blue)
		H16065	AGCGGTTGTTGATGGGTGAGTC	0.019					
7697	C1d1	L07684	TGATCACGCCCTCATAATCATTTTCCTT	0.033	73	*(ct)^{23}GTTAGGAAAAGGGCATACAGGA*	0.020	68	C (yellow), T (red)
		H07705	TGTTGTGAGTGTTAGGAAAAGGGCA	0.033					
14433	C4c	L14431	GACCTCAACCCCTGACCCCCAT	0.019	57	(ct)^{10}ACCCCTGACCCCCATG	0.015	36	C (yellow), T (red)
		H14443	ACTACAGCGATGGCTATTGAGGAG	0.019					
2092	D1	L02079	GCCCACAGAACCCTCTAAATCCCC	0.033	75	*(ct)^{26}cAGCTGTTCCTCTTTGGACTAACA*	0.020	76	G (blue), A (green)
		H02106	TCCTAGTGTCCAAAGAGCTGTTCCT	0.033					
9667	D2a1	L09652	GGAGTATCAATCACCTGAGCTCACCA	0.032	70	*(ct)^{24}cCTTGAATTATTTGGTTTCGGTTG*	0.020	72	T (red), C (yellow)
		H09670	GCAGTGCTTGAATTATTTGGTTTCGGT	0.032					
4991	D2a2	L04976	TCATAGCAGGCAGTTGAGGTGGA	0.032	88	(ct)^{20}GGTGGATTAAACCAAACCCA	0.015	60	G (blue), A (green)
		H05007	TCCTATGTGGGTAATTGAGGAGTATGC	0.032					
9181	D2b	L09159	TCGCTGTCGCCTTAATCCAAGCC	0.032	68	*(ct)^{12}cTTGTCGTGCAGGTAGAGGC*	0.015	44	T (red), C (yellow)
		H09183	TGTGTTGTCGTGCAGGTAGAGG	0.032					
10181	D4b1	L10177	AATCCACCCCTTACGAGTGCGG	0.006	63	*(ct)^{13}GGCGGGGGATATAGGGTC*	0.015	44	G (blue), A (green)
		H10197	TTATGGAGAAAGGGACGCGGGC	0.006					
6285	D4h3a	L06282	GCCGGAGCAGGAACAGGTTGAA	0.032	75	*(ct)^{2}°CCCTGCTAAGGGAGGGTAGA*	0.020	60	C (yellow), T (red)
		H06314	GTCTACGGAGGCTCCAGGGTGG	0.032					

SBE probes in reverse direction are shown in Italics.

hg haplogroup; conc. concentration; bp base pairs; nt nucleotides.

doi:10.1371/journal.pone.0093292.t001

Table 2. Details of primers used for standard HVR-I amplification and sequencing.

Primer	Primer Sequences 5′ to 3′	Length in bp (incl./excl. primer)	Reference
L16055	GAAGCAGATTTGGGTACCAC	126 (87)	[54]
H16142	ATGTACTACAGGTGGTCAAG		[55]
L16117	TACATTACTGCCAGCCACCAT	162 (115)	[42]
H16233	GCTTTGGAGTTGCAGTTGATGTGT		[42]
L16209	CCCCATGCTTACAAGCAAGT	179 (138)	[54]
H16348	ATGGGGACGAGAAGGGATTTG		[42]
L16287	CACTAGGATACCAACAAACC	162 (122)	[54]
H16410	GCGGGATATTGATTTCACGG		[54]

doi:10.1371/journal.pone.0093292.t002

Early Intermediate (n = 20; 200–600 AD), the Middle Horizon (n = 20; 600–1000 AD) and the Late Intermediate (n = 12; 1000–1476 AD) [47] plus an Early Medieval European samples as control for the ancestral state. Samples from each period varied in the state of preservation, due to differences in mortuary customs. From our test dataset of 52 samples in total, we were able to unambiguously type 29 samples (56%) (Table 3, Figure 4). A typing result was considered reliable when two samples from the same individual could be unambiguously assigned to the same sub-hg in two independent experiments.

We subsequently compared the AmericaPlex26 assay results to our previous attempts at amplifying and sequencing the mitochondrial HVR-I with four overlapping primer pairs and found that the AmericaPlex26 assay improved the typing efficiency from ancient samples (Table 3, Figure 4). For example, the AmericaPlex26 assay allowed reliable SNP typing for eleven Early Intermediate (55%) and ten Late Intermediate samples (83%), whereas HVR-I sequencing gave reliable sequence haplotypes for seven (35%) and seven (58%) samples, respectively. For example, HVR-I sequencing for samples 10802A and 10803A failed, while the AmericaPlex26 assay revealed specific hg B2 (Table 3). Samples from the Middle Horizon culture were in general less well preserved, resulting in eight consensus sub-hg calls (40%) using the AmericaPlex26 assay, whereas HVR-I sequencing did not produce any reliable sequence haplotype from the sample replicates (0%). This highlights the genotyping power of our assay when dealing with challenging samples.

Importantly, SNP typing with the AmericaPlex26 assay also gave a higher resolution compared to traditional HVR-I sequencing. For example, samples 10809A and 10810A of the Early Intermediate period were assigned to the major hg C1 by HVR-I sequencing, yet the AmericaPlex26 assay allowed further resolution to sub-hg C1b. In addition, while many of the HVR-I results from Late Intermediate samples remained tentative, i.e. non-reproducible, the AmericaPlex26 assay provided reliable and specific sub-hgs for both replicates (Table 3, Figure 4). Taking all

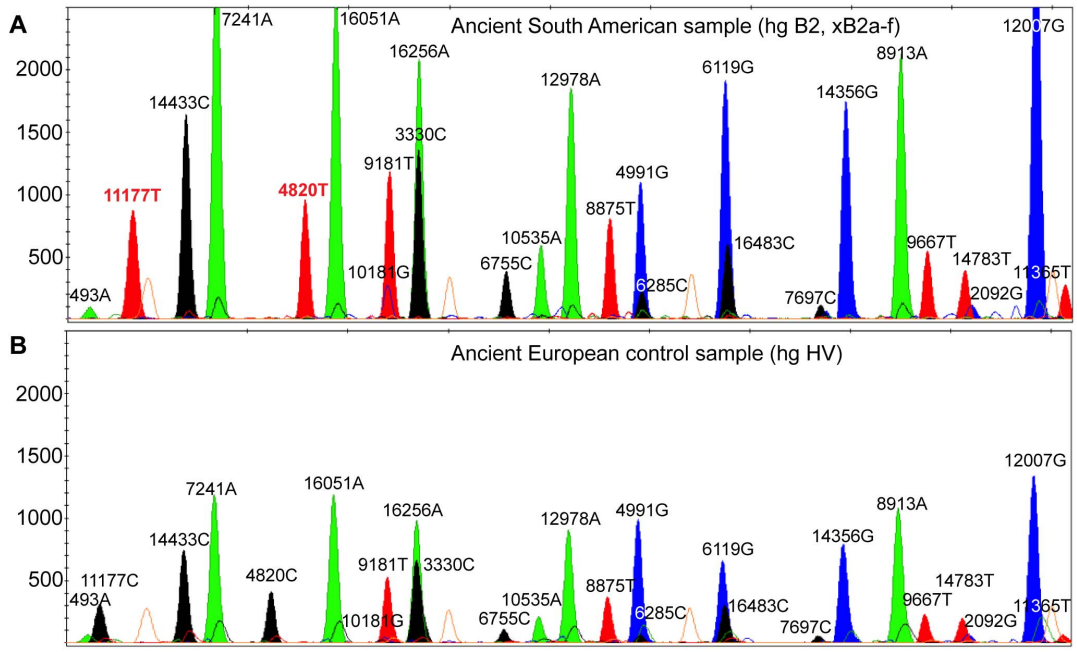


Figure 2. Electropherograms of two ancient examples representing the optimized AmericaPlex26. Panel A shows a South American sample and panel B an ancient European sample illustrating the ancestral state of all 26 SNPs. The two SNPs sites defining hg B2 are highlighted in red. Note the ancestral state at SNP sites defining B2 sub-haplogroups B2a (16483), B2b (6755), B2c (7241), B2d (8875), B2e (6119) and B2f (10535).
doi:10.1371/journal.pone.0093292.g002

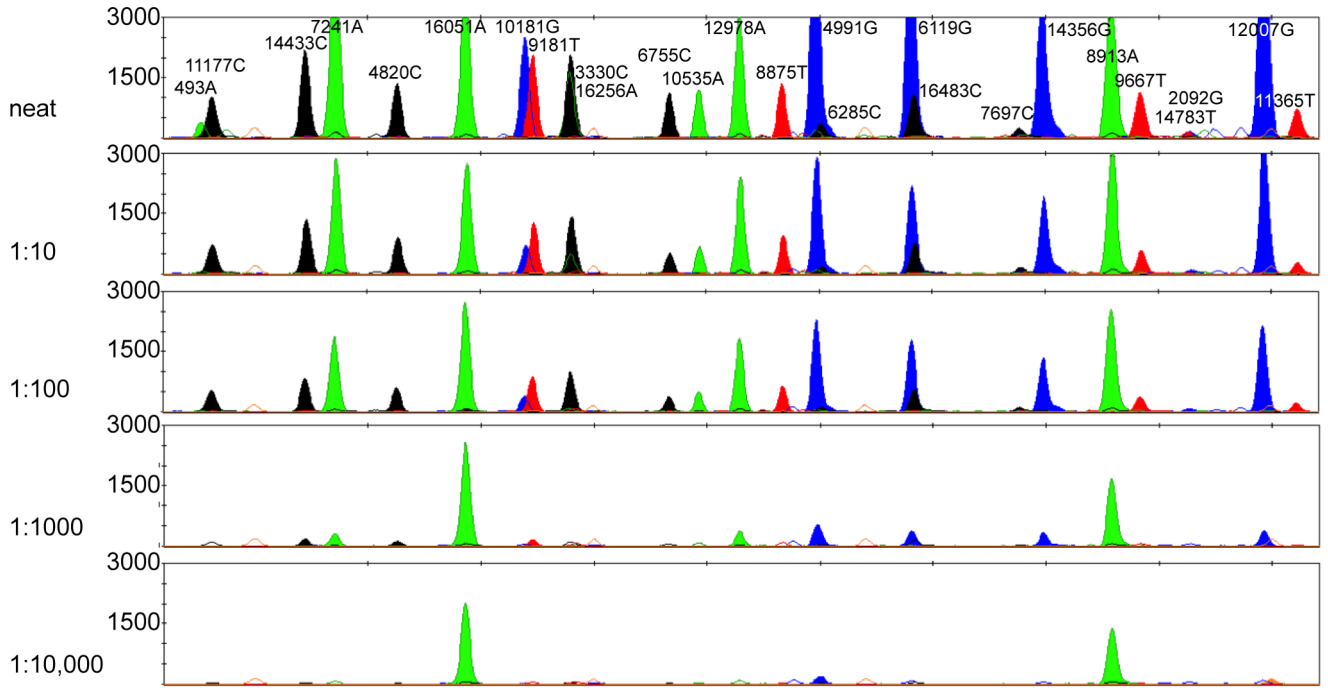


Figure 3. Electropherograms from the sensitivity tests showing four serial dilutions of template DNA with a starting copy number of 1,171,699 copies/µL. Note the increasing number of locus dropout as template DNA concentration decreases.
doi:10.1371/journal.pone.0093292.g003

results together, the AmericaPlex26 assay showed a significantly higher success rate when compared to the standard HVR-I sequencing ($p<0.0001$, Wilcoxon matched-pairs signed rank test; Figure 4). This is likely due to the difference in amplicon sizes between the two methods.

The overall effectiveness of the assay

Overall, the effectiveness of the multiplex SNaPshot method in analysing ancient DNA lies in the fact that it only requires minimal flanking regions either side of the SNP, which in theory allows the design of very short overall amplicon sizes [40,48], 56–90 bp in

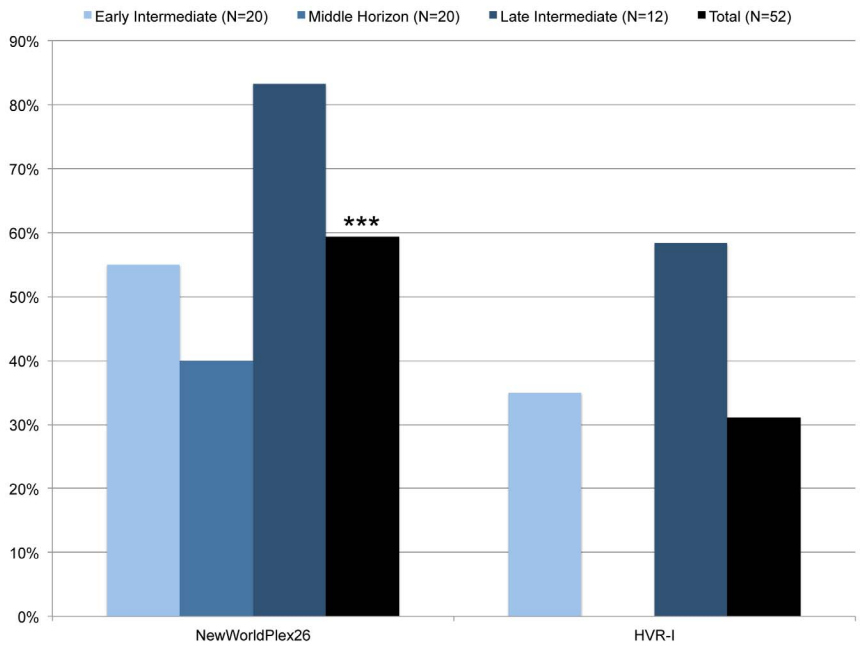


Figure 4. Genotyping success of the AmericaPlex26 assay compared to standard HVR-I sequencing via four overlapping amplicons. The success rate is given in percentage of unambiguous genotype calls for each of the two methods per cultural horizon and all results combined (black bars, $p< 0.0001$).
doi:10.1371/journal.pone.0093292.g004

Table 3. Direct comparison of results for HVR-I sequencing and AmericaPlex26 SNP typing assay for samples unambiguously typed using the AmericaPlex26 assay.

Individual	Museum no.	Samples	HVR I hg	AmericaPlex26	Consensus
EI 1	A06 95/96	**10802A**	?	B2	B2
		10803A	?	B2	
EI 2	A06 79/01 Ind2	**10804A**	D1	D1?	D1
		10805A	D1	M	
EI 3	A06 76/96	**10809A**	C1	C1b	C1b
		10810A	C1	C1b	
EI 4	A15 06/00	**10813A**	B4	B2	B2
		10814A	B4	B2	
EI 5	A06 01/02	**10817A**	C1	C1b	C1b
		10818A	?	C1b	
EI 6	A6 68/96	**10821A**	B4	B2	B2
		10822A	B4	B2	
EI7	A06 90 Ind1	**10787A**	?	B2	B2
		10788A	?	B2	
EI 8	A20 05/09	**10789A**	B4	B2b	B2b
		10790A	?	B2b	
EI 9	A20 03/07 Ind1	**10791A**	A2	A2	A2
		10792A	A2	A2	
EI 10	A06 82/96	**10793A**	B4	B2	B2
		10794A	B4	B2	
EI 11	A06 77/96	**10796A**	B4	B2b	B2
		10797A	?	B2	
EI 12	A06 79/96 Ind1	**10800A**	B4	B2b	B2
		10801A	B4	B2	
MH 1	A20 08/08 Ind2	**10733A**	?	C1b	C1b
		10734A	?	C1b	
MH 2	A20 CF003/09	**10741A**	?	B2	B2
		10742A	?	B2	
MH 3	A20 07/08 Ind4	**10749A**	?	B2	B2
		10750A	?	B2	
MH 4	A20 05/08	**10753A**	?	B2	B2
		10754A	B4?	B2	
MH 5	A20 18/08	**10765A**	-	C1b	C1b
		10766A	-	C1b	
MH 6	A20 01/09 Ind1	**10772A**	C1	C1b	C1b
		10773A	?	C1b	
MH 7	A20 01/09 Ind2	**10774A**	?	A2	A2
		10775A	-	A2	
MH 8	A20 04/07 Ind2	**10778A**	D?	A2	A2
		10779A	-	A2	
LI 1	A15 01/02	**10709A**	C1	M	C1
		10710A	C1	C1b	
LI 2	A0 cf14 ind-1/98	**10712A**	B4	B2	B2
		10713A	B4	B2	
LI 3	A01 CF16/98	**10715A**	?	B2	B2
		10716A	-	B2	
LI 4	A0 CF15/01	**10717A**	C1	M	M
		10718A	?	M	

Table 3. Cont.

Individual	Museum no.	Samples	HVR I hg	AmericaPlex26	Consensus
LI 5	A15 CF36/01	**10719A**	B4	B2	B2
		10720A	B4	B2b	
LI 6	A15 Sin Contexto	**10724A**	C1	C1b	C1b
		10725A	C1	C1b	
LI 7	A3 CF01/04	**10726A**	?	B2b	B2
		10727A	?	B2	
LI 8	A0 08/98	**10728A**	B4	B2	B2
		10729A	B4	B2	
LI 9	A15 02/02	**10730A**	B4	B2	B2
		10730Y	B4	B2b	
LI 10	A0 56/97	**10731A**	C1	C1b	C1b
		10731Y	C1	C1b	

Consensus haplogroups were called based on last common SNP from both replicates from independent extractions, and minimum peak size >50 rfu. (?)/( ): Insufficient or no sequence information; EI: Early Intermediate; MH: Middle Horizon; LI: Late Intermediate.
doi:10.1371/journal.pone.0093292.t003

our case. This makes it suitable for the extremely fragmented and damaged state of ancient DNA, while the multiplex approach maximises the amount of information that can be gained per PCR [15,49,50]. As such, SNaPshot typing is able to generate results for samples for which traditional sequencing methods often fail with ancient and/or degraded DNA, as they require longer fragment lengths to be cost-effective [51]. Multiplexing also allows the combination of many informative SNP sites into one reaction, which are otherwise spread across longer sequence regions. On its own, the multiplex PCR and SNaPshot method is time- and cost-effective, and requires substantially smaller amounts of valuable DNA extract compared to HVR-I sequencing, as fewer individual reactions are needed from preparation of the multiplex PCR to capillary electrophoresis [48,50].

We show that the AmericaPlex26 can be used to complement or expand upon standard mtDNA sequencing approaches for ancient Native American populations, and especially for ancient samples where DNA preservation does not allow amplification of longer (> 100 bp or more) DNA molecules. It efficiently and economically targets characteristic SNPs from the coding region of mtDNA [15] in order to corroborate HVR-I sequencing results and to define a particular sub-hg [14]. Alternatively, the AmericaPlex26 can be used in addition to global mtDNA SNP multiplexes, such as the GenoCore22 and others [23,24]. Moreover, it is flexible enough to add newly discovered SNPs/lineages in order to enhance sub-regional resolution (see e.g. [32,35].

In our experience, the new method provided an extremely useful one-reaction test to screen larger numbers of degraded samples allowing the assessment of the general state of preservation, the authenticity of the result (i.e. absence of potential contaminating lineages), while at the same time allowing a categorisation of potentially interesting sub-hgs. We are currently using this approach in order to further dissect the phylogenetic resolution via DNA library creation and targeted mtDNA enrichment and Next Generation Sequencing [25,26].

Conclusions

We present a powerful, optimized SNP assay, which allows unambiguous typing of Native American mtDNA 'founder lineages' and additional SNPs for further resolution. This short-amplicon AmericaPlex26 assay is highly efficient, time and cost-effective compared to classical HVR-I sequencing, and allows highly resolved SNP typing of degraded DNA samples in forensic and ancient DNA work. It is suitable as a qualitative 'screening' method to identify samples with sufficient DNA preservation, free of contaminants that complicate full mitochondrial sequencing (and beyond) via Next Generation Sequencing techniques.

Acknowledgments

We sincerely thank team members of the "Proyecto Arqueológico Huaca Pucllana" for help with accessing the collection of pre-Columbian human remains from the Huaca Pucllana site. We also thank three anonymous reviewers for constructive feedback on an earlier version of this manuscript.

Author Contributions

Conceived and designed the experiments: WH AC. Performed the experiments: AC WH GV BL LFS. Analyzed the data: AC GV WH. Contributed reagents/materials/analysis tools: LFS MIBR IFE AC. Wrote the paper: AC BL WH. Discussed the results and reviewed the text: AC GV LFS BL AC WH.

References

1. Torroni A, Schurr TG, Yang CC, Szathmary EJ, Williams RC, et al. (1992) Native American mitochondrial DNA analysis indicates that the Amerind and the Nadene populations were founded by two independent migrations. Genetics 130: 153–162.
2. Wallace DC, Torroni A (1992) American Indian prehistory as written in the mitochondrial DNA: a review. Human Biology 64: 403–416.
3. Forster P, Harding R, Torroni A, Bandelt HJ (1996) Origin and evolution of Native American mtDNA variation: a reappraisal. American Journal of Human Genetics 59: 935–945.
4. Tamm E, Kivisild T, Reidla M, Metspalu M, Smith DG, et al. (2007) Beringian standstill and spread of Native American founders. PLoS ONE 2: e829.
5. Schurr TG (2000) Mitochondrial DNA and the Peopling of the New World: Genetic variations among Native Americans provide further clues to who first populated the Americas and when they arrived. American Scientist 88: 246–253.
6. Perego UA, Achilli A, Angerhofer N, Accetturo M, Pala M, et al. (2009) Distinctive Paleo-Indian migration routes from Beringia marked by two rare mtDNA haplogroups. Current Biology 19: 1–8.

7. Wang S, Lewis CM, Jakobsson M, Ramachandran S, Ray N, et al. (2007) Genetic variation and population structure in native Americans. PLoS Genetics 3: e185.
8. O'Rourke DH, Raff JA (2010) The human genetic history of the Americas: the final frontier. Current Biology 20: R202–207.
9. Kemp BM, Schurr TG (2010) Human Variation in the Americas: The Integration of Archaeology and Biological Anthropology; Auerbach BM, editor: Center for Archaeological Investigations, Southern Illinois University Carbondale.
10. Fehren-Schmitz L, Reindel M, Cagigao ET, Hummel S, Herrmann B (2010) Pre-Columbian population dynamics in coastal southern Peru: A diachronic investigation of mtDNA patterns in the Palpa region by ancient DNA analysis. American journal of physical anthropology 141: 208–221.
11. Kaestle FA, Horsburgh KA (2002) Ancient DNA in Anthropology: Methods, Applications and Ethics. Yearbook of Physical Anthropology 45: 92–130.
12. de Saint Pierre M, Bravi CM, Motti JMB, Fuku N, Tanaka M, et al. (2012) An Alternative Model for the Early Peopling of Southern South America Revealed by Analyses of Three Mitochondrial DNA Haplogroups. PLoS ONE 7: e43486.
13. Álvarez-Iglesias V, Jaime JC, Carracedo Á, Salas A (2007) Coding region mitochondrial DNA SNPs: Targeting East Asian and Native American haplogroups. Forensic Science International: Genetics 1: 44–55.
14. Bandelt HJ, Herrnstadt C, Yao YG, Kong QP, Kivisild T, et al. (2003) Identification of Native American founder mtDNAs through the analysis of complete mtDNA sequences: some caveats. Annals of Human Genetics 67: 512–524.
15. Quintáns B, Álvarez-Iglesias V, Salas A, Phillips C, Lareu MV, et al. (2005) Typing of mitochondrial DNA coding region SNPs of forensic and anthropological interest using SNaPshot minisequencing. Forensic science international 140: 251–257.
16. Bouakaze C, Keyser C, Amory S, Crubézy E, Ludes B (2007) First successful assay of Y-SNP typing by SNaPshot minisequencing on ancient DNA. International Journal of Legal Medicine 121: 493–499.
17. Ballantyne KN, van Oven M, Ralf A, Stoneking M, Mitchell RJ, et al. (2012) MtDNA SNP multiplexes for efficient inference of matrilineal genetic ancestry within Oceania. Forensic Science International: Genetics 6: 425–436.
18. Argac D, Bulbul O, Shahzad MS, Acar E, Altuncul H, et al. (2009) Optimization and validation of 10 mitochondrial DNA SNPs using SNaPshot Kit. Forensic Science International: Genetics Supplement Series 2: 99–101.
19. Paneto GG, Köhnemann S, Martins JA, Cicarelli RMB, Pfeiffer H (2011) A single multiplex SNaPshot reaction of 42 SNPs to classify admixture populations into mitochondrial DNA haplogroups. Forensic Science International: Genetics Supplement Series 3: e198–e199.
20. Geppert M, Roewer L (2012) SNaPshot(R) minisequencing analysis of multiple ancestry-informative Y-SNPs using capillary electrophoresis. Methods in Molecular Biology 830: 127–140.
21. Grignani P, Peloso G, Alù M, Ricci U, Robino C, et al. (2006) Subtyping mtDNA haplogroup H by SNaPshot minisequencing. International Congress Series 1288: 124–126.
22. Perego UA, Angerhofer N, Pala M, Olivieri A, Lancioni H, et al. (2010) The initial peopling of the Americas: A growing number of founding mitochondrial genomes from Beringia. Genome Research 20: 1174–1179.
23. van Oven M, Vermeulen M, Kayser M (2011) Multiplex genotyping system for efficient inference of matrilineal genetic ancestry with continental resolution. Investigative genetics 2: 6.
24. Haak W, Balanovsky O, Sanchez JJ, Koshel S, Zaporozhchenko V, et al. (2010) Ancient DNA from European early Neolithic farmers reveals their near eastern affinities. PLoS Biology 8: e1000536.
25. Maricic T, Whitten M, Pääbo S (2010) Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. PLoS ONE 5: e14004.
26. Brotherton P, Haak W, Templeton J, Brandt G, Soubrier J, et al. (2013) Neolithic mitochondrial haplogroup H genomes and the genetic origins of Europeans. Nature communications 4: 1764.
27. Knapp M, Horsburgh KA, Prost S, Stanton JA, Buckley HR, et al. (2012) Complete mitochondrial DNA genome sequences from the first New Zealanders. Proceedings of the National Academy of Sciences USA 109: 18350–18354.
28. Tamm E, Kivisild T, Reidla M, Metspalu M, Smith DG, et al. (2007) Beringian standstill and spread of Native American founders. PLoS ONE 2: e829.
29. Achilli A, Perego UA, Bravi CM, Coble MD, Kong Q-P, et al. (2008) The Phylogeny of the Four Pan-American MtDNA Haplogroups: Implications for Evolutionary and Disease Studies. PLoS ONE 3: e1764.
30. Achilli A, Perego UA, Lancioni H, Olivieri A, Gandini F, et al. (2013) Reconciling migration models to the Americas with the variation of North American native mitogenomes. Proceedings of the National Academy of Sciences USA 110: 14308–14313.
31. Gilbert MT, Kivisild T, Gronnow B, Andersen PK, Metspalu E, et al. (2008) Paleo-Eskimo mtDNA genome reveals matrilineal discontinuity in Greenland. Science 320: 1787–1789.
32. Kumar S, Bellis C, Zlojutro M, Melton PE, Blangero J, et al. (2011) Large scale mitochondrial sequencing in Mexican Americans suggests a reappraisal of Native American origins. BMC Evolutionary Biology 11: 293.
33. Malhi RS, Cybulski JS, Tito RY, Johnson J, Harry H, et al. (2010) Brief communication: Mitochondrial haplotype C4c confirmed as a founding genome in the Americas. American Journal of Physical Anthropology 141: 494–497.
34. Kashani BH, Perego UA, Olivieri A, Angerhofer N, Gandini F, et al. (2012) Mitochondrial haplogroup C4c: A rare lineage entering America through the ice-free corridor? American Journal of Physical Anthropology 147: 35–39.
35. Bodner M, Perego UA, Huber G, Fendt L, Röck AW, et al. (2012) Rapid coastal spread of First Americans: Novel insights from South America's Southern Cone mitochondrial genomes. Genome research 22: 811–820.
36. Derbeneva OA, Sukernik RI, Volodko NV, Hosseini SH, Lott MT, et al. (2002) Analysis of mitochondrial DNA diversity in the aleuts of the commander islands and its implications for the genetic history of beringia. American Journal of Human Genetics 71: 415–421.
37. Starikovskaya YB, Sukernik RI, Schurr TG, Kogelnik AM, Wallace DC (1998) mtDNA diversity in Chukchi and Siberian Eskimos: implications for the genetic history of Ancient Beringia and the peopling of the New World. American Journal of Human Genetics 63: 1473–1491.
38. You FM, Huo N, Gu YQ, Luo M, Ma Y, et al. (2008) BatchPrimer3: a high throughput web application for PCR and sequencing primer design. BMC bioinformatics 9: 253.
39. Rozen S, Skaletsky H (1999) Primer3 on the WWW for general users and for biologist programmers. Bioinformatics methods and protocols: Springer. pp. 365–386.
40. Sanchez JJ, Endicott P (2006) Developing multiplexed SNP assays with special reference to degraded DNA templates. Nature Protocols 1: 1370–1378.
41. Vallone PM, Just RS, Coble MD, Butler JM, Parsons TJ (2004) A multiplex allele-specific primer extension assay for forensically informative SNPs distributed throughout the mitochondrial genome. International Journal of Legal Medicine 118: 147–157.
42. Haak W, Forster P, Bramanti B, Matsumura S, Brandt G, et al. (2005) Ancient DNA from the first European farmers in 7500-year-old Neolithic sites. Science 310: 1016–1018.
43. Der Sarkissian C, Balanovsky O, Brandt G, Khartanovich V, Buzhilova A, et al. (2013) Ancient DNA Reveals Prehistoric Gene-Flow From Siberia in the Complex Human Population History of North East Europe. PLoS Genetics 9: e1003296.
44. Haak W, Brandt G, de Jong HN, Meyer C, Ganslmeier R, et al. (2008) Ancient DNA, Strontium isotopes, and osteological analyses shed light on social and kinship organization of the Later Stone Age. Proceedings of the National Academy of Sciences USA 105: 18226–18231.
45. Adler CJ, Haak W, Donlon D, Cooper A (2011) Survival and recovery of DNA from ancient teeth and bones. Journal of Archaeological Science 38: 956–964.
46. Sanchez JJ, Phillips C, Børsting C, Balogh K, Bogus M, et al. (2006) A multiplex assay with 52 single nucleotide polymorphisms for human identification. Electrophoresis 27: 1713–1724.
47. Lanning EP, Pollard G (1967) Peru before the Incas: Prentice-Hall Englewood Cliffs, NJ.
48. Butler JM (2011) Advanced Topics in Forensic DNA Typing: Methodology: Academic Press.
49. Butler JM (2009) Fundamentals of forensic DNA typing: Academic Press.
50. Sobrino B, Carracedo A (2005) SNP Typing in Forensic Genetics; Carracedo Á, editor. Totowa, NJ: Humana Press.
51. Dario P, Bom J, Ribeiro T, Geada H (2009) MtSNP typing before mtDNA sequencing: Why do it? Forensic Science International: Genetics Supplement Series 2: 187–188.
52. Behar DM, van Oven M, Rosset S, Metspalu M, Loogvali EL, et al. (2012) A "Copernican" Reassessment of the Human Mitochondrial DNA Tree from its Root. American Journal of Human Genetics 90: 675–684.
53. Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, et al. (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nature Genetics 23: 147.
54. Handt O, Krings M, Ward RH, Paabo S (1996) The retrieval of ancient human DNA sequences. American Journal of Human Genetics 59: 368–376.
55. Stone AC, Stoneking M (1998) mtDNA analysis of a prehistoric Oneota population: implications for the peopling of the New World 20. American Journal of Human Genetics 62: 1153–1170.

CHAPTER 6

General Discussion

General Discussion

Thesis summary

This thesis investigates the mitochondrial DNA (mtDNA) diversity in pre-Columbian ancient human populations from the Central Andes in South America. The results provide new insights into the current knowledge of the human maternal genetic diversity from this part of the world.

Ancient DNA (aDNA) allows the exploration of the genetic diversity of human groups in real time, making it a powerful tool to understand and reconstruct the evolutionary history of ancient populations from different regions in the world. However, the retrieval of aDNA molecules is subject to many experimental challenges due to sample preservation and contamination issues, and not all ancient samples can be successfully analysed. This is especially true for specimens from low latitudes and (sub-tropical) areas, where climatic and environmental conditions have adverse effects on DNA preservation. A good example is South America, where – despite significant efforts and a growing number of studies – aDNA research is restricted to the analysis of short sequences of the mtDNA – but see also (Raghavan et al. 2015). As a consequence, the novel aDNA data presented in this thesis increase considerably our knowledge gleaned from aDNA research from South American sites, and provides valuable information into the degree of preservation of ancient samples from South America.

The time of my PhD was accompanied by major methodological, technical and analytical transitions to retrieve aDNA from pre-Columbian human remains. I started by using traditional PCR–based aDNA laboratory techniques and protocols. First, I amplified and sequenced the Hypervariable Region I (HVR-I) from the D-loop of the mtDNA using a set of four overlapping primers pairs (Haak et al. 2010). I also genotyped the samples using a combination of two multiplexed minisequencing panels as a screening strategy for complete mtDNA sequencing (Chapter 2).

The first panel of 22 coding region SNPs (GenoCore22) describes the worldwide human mtDNA diversity (Haak et al. 2010) and allows the detection of contamination by non-Native-American mtDNA; and the second panel of 26 continental specific SNPs

(AmericaPlex26) was developed specifically to describe the Native American mtDNA diversity (Coutinho et al. 2014) (Chapter 5).

Finally, I applied the latest and new technological approach using high-throughput methods (e.g. hybridization capture techniques and Next-Generation Sequencing – NGS), with the aim to increase the amount of genomic information from our samples (Chapters 3-4). To date this project represents the largest effort in sequencing complete mtDNA from a large number of ancient samples from South America, offering new perspectives into the Native American past genetic diversity.

In other words, I have moved from PCR-based techniques and traditional Sanger sequencing of HVR-I towards high-resolution mitochondrial genome sequencing applying NGS techniques. This methodological improvement is now routinely applied at the Australian Centre for Ancient DNA (ACAD) to obtain more genomic data for more robust and detailed analyses.

Main findings of the thesis

Genetic studies have found that the mtDNA diversity in Native Americans is derived from a limited Asian gene pool and could be traced back to an ancestral populations that lived in eastern Siberia and parts of Beringia (Schurr et al. 1990; Torroni et al. 1993). Previous research has shown that two main events have played a role in shaping the population structure in the Americas. First, the initial peopling of the continent via Bering Land Bridge ~15,000 years ago (Goebel et al. 2008) and second, the arrival of Europeans in the 15th century, which led to a demise of Native Americans (Livi-Bacci 2006).

Given that Native American genetic diversity harbours only five mtDNA founding haplogroups (A-D and X) (Torroni et al. 1993) and two founder Y-Chromosome lineages (C, Q) (Karafet et al. 1999), geneticists have suggested a genetic bottleneck during the initial peopling (Meltzer 1993; Schurr and Sherry 2004; Fagundes et al. 2008). On the other hand, the demise of Native Americans resulting from the European conquest was also hypothesized to have affected the overall diversity of the Native American genetic pool (O'Fallon and Fehren-Schmitz 2011).

The complete mtDNA genomes and HVR-I sequence data obtained in this project have largely confirmed the presence of Native American "founder" mtDNA haplogroups (A-D), although, haplogroup X was not identified in our samples.

Moreover, by analysing genomic data from mtDNA complete genomes, I have identified interesting patterns and results that are interpreted in the light of these bottleneck scenarios. For instance, I have found a large amount of "hidden" or as yet unknown genomic information expressed in newly described mutations or "private" SNPs present in every ancient mtDNA sequence. While these sequences fall within the known haplogroup diversity for Native Americans (i.e. hg A2, B2, B2b, C1b, C1c, C1d and D1), such lineages were not seen in modern genetic studies on American populations so far (Chapter 3).

Mitochondrial genome sequences analysed in this thesis, suggest a unique pre-Columbian diversity that has been altered during the disruption of the European conquest, due to population collapse in the Americas concurrent with (O'Fallon and Fehren-Schmitz 2011), with the result of loss or under-representation of a large fraction of pre-Columbian lineages as explanation why these lineages are very rare today. However, we cannot rule out the possibility that these new mtDNA lineages are indeed extinct lineages. To assess this particular finding it will be important to expand the sequencing mtDNA efforts in modern-day population studies in order to achieve full phylogenetic resolution and to find lineages similar to or derived from the ancient mtDNA sequences obtained in this project.

The aDNA sequences reported here, add a new perspective to the characterization of pre-Columbian populations in South America. Our new aDNA data showcase the power to "re-discover" novel mtDNA genome sequences and as such to directly describe the diversity in past populations, which would not be possible to infer from analysing modern-day Native American data. Therefore, complete mtDNA sequencing is the key advantage to pinpoint differences between high-resolution genomes in contrast to limited mtDNA sequence information (e.g. D-loop sequences).

Significance and contribution to current knowledge

Mitochondrial DNA is still one of the most readily accessible resources for modern and ancient genetic studies (Kivisild 2015), especially since ancient nuclear DNA analysis is a costly and time-consuming endeavour. Exploring mtDNA genetic diversity in ancient Native Americans using high-resolution data provide unique information about demographic, geographic, temporal relationships and historical population events in pre-Columbian populations.

The present study reports 72 new HVR-I sequences from ancient populations, which increases the current ancient sequence database for South America by 0.3-fold (Chapter 2). Due to the large amount of available data for HVR-I from ancient and present-day populations in the Americas, this approach is still relevant to perform genetic comparisons and to explore population relationships through time. Also, the large majority of these new 72 sequences were generated from ancient populations that have never been studied before (e.g. Lauricocha, Chancay, Lima, Ychsma, Pueblo Viejo, Tiwanaku, and Llullaillaco).

DNA amplification success rates show the proportion of individuals yielding reliable aDNA sequences, which is also indicative of sample preservation. Traditional PCR-based methods sequencing of HVR-I resulted in a success rate of 48.3% (72 out of 149), which is lower when compared with other studies on Andean groups from similar regions/climates such as 90.5% (Carnese et al. 2010), and moderately lower when compared with 65% in (Fehren-Schmitz et al. 2010) and 58.5% with (Baca et al. 2012). The lower mean success rate can be explained by the inclusion of various less well preserved lowlands sites, results for which (or the lack thereof) are rarely reported in the literature (e.g. Caral, Aspero, Puemape samples).

Sequencing of HVR-I by traditional PCR-based methods represents a labour extensive and very time-consuming process in light of the information gained when compared to modern NGS-based techniques. Moreover, dealing with highly degraded DNA often results in poor PCR efficiency and lack of replication of results, which is fundamental in PCR-based aDNA studies.

Therefore, I applied optimised DNA extraction methods, genomic library preparation protocols, capture and NGS techniques, plus bioinformatics pipelines to generate complete mtDNA genomes, which provided genetic data at maximum molecular resolution.

The incorporation of ancient samples covering almost all the archaeological periods in the Central Andes of South America is the highlight of the project, since this better temporal coverage allows to evaluate the impact of events such as migration, demographic expansion, and settlement of human groups in this geographic region (Chapter 2-5). Temporal sampling therefore represents an additional layer of information to investigate changes in genetic composition in pre-Columbian populations in real-time.

The HVR-I data was used to test for population continuity and genetic homogenization in the Central Andes (Lewis 2009; Fehren-Schmitz et al. 2011b), which were linked to micro-evolutionary processes between Andean and ancient costal groups (Moraga et al. 2005). The results based on genetic distances (F_{ST}), PCA and shared haplotype analysis support a strong degree of population continuity in the Central Andes of South America suggesting a "genetic assimilation" between highland and coastal populations over time.

Based also on HVR-I information, I have identified two temporal transitions between archaeological periods in the Central Andes. One cultural transition involves the Early Horizon and Early Intermediate periods, followed by another that involves the three successive Middle Horizon, Late Intermediate and Late Horizon periods and no major demographic changes or population replacement in the Central Andres region in pre-Columbian times. Hence, taking into account the temporal sampling, my results also support this micro-evolutionary scenario but suggest a series of "cultural transitions" that might be understood as a "genetic assimilation" process over time between low and highland populations from the Late archaic periods to the Late Horizon in the Central Andes (Chapter 2).

In addition, the present study reports novel 92 mtDNA complete genome sequences that offer the first high-resolution description of pre-Columbian genetic diversity in South America. Previously, only 19 ancient mtDNA genomes were known from the continent. Three mtDNA genomes from Greenland and 16 mtDNA genomes from North America – referring only to mtDNA genomes that were published

independently from the complete nuclear genome of the respective individual (Gilbert et al. 2008; Cui et al. 2013; Raghavan et al. 2014). The substantial number of new mtDNA genomes obtained here, follows the direction the aDNA field has taken in recent years in terms of generation and interpretation of genomic data from complete mtDNA from ancient populations all over the world (Kivisild 2015). In consequence, this effort in sequencing ancient mtDNA genomes will serve as a launchpad for future studies in the Americas (Chapter 3).

I have also used complete mtDNA genome data to investigate demographic events focussing on particular and well-defined cultural groups in a restricted geographic location. For instance, in the study of the Huaca Pucllana archaeological site in costal Lima, I have explored micro-evolutionary changes across a time transect from the Early Intermediate period to the Late Intermediate period covering approximately ~1250 years (Chapter 4). The Huaca Pucllana archaeological site represents a very interesting chronological transect and cultural development of three successive settlements (Lima, Wari, Ychsma) across archaeological periods in a restricted geographic region (i.e. coastal Central Peru).

The results obtained in this study indicate that genetic diversity in Huaca Pucllana shifted only slightly through time, ruling out a complete population discontinuity or replacement driven by the Wari imperialist hegemony, at least in the coastal Lima region during the Middle Horizon period. Hence, I recognize the power of ancient and modern mtDNA data to explore the geographic and temporal dimensions of the pre-Columbian population diversity on regional prehistory in South America.

The new insights on genetic diversity and haplogroup composition of pre-Columbian populations in South America, gained from mtDNA genome sequencing, are particularly striking when contrasted with partial HVR-I sequence information (Chapter 2-3). With the increasing number of more mtDNA genomes available from the Americas, researchers will be able to test more detailed scenarios and further refine both long-standing questions (e.g. the peopling of the Americas), and question at the regional level, which will further enhance our understanding of the genetic structure and evolutionary changes in Native American populations.

Authenticating aDNA data and contamination control

Laboratory DNA techniques used to analyse human aDNA have been hampered by contamination issues with modern-day DNA and the complete elimination of human contamination is highly unlikely (Montiel et al. 2001). The most problematic form of contamination of ancient human DNA is due to modern DNA introduced during handling of samples prior to aDNA sequencing (Knapp et al. 2015).

The ancient DNA work for the present study was performed under strict and high-standard laboratory precautions required for this field of research (Cooper and Poinar 2000; Fulton 2012; Knapp et al. 2012). To minimize the potential risk of contamination all aDNA extractions and genomic library preparations were processed in a purpose-built laboratory dedicated exclusively to aDNA research at the Australian Centre for Ancient DNA in Adelaide, Australia. It is important to point out that replication of aDNA extraction and genotyping is a fundamental step in aDNA authentication, especially when applying PCR-based methodologies. For this purpose, I have also performed independent replications for ancient samples from collaborators working on related projects (e.g. Palpa, Pernil Alto, Chullpa Botiriayocc), however only sample 12692A-Palpa was successfully replicated by GenoCore22 assay. Moreover, the correct and strict following of de-contamination measures resulted in a reliably low level of DNA background contamination in the laboratory.

As seen in Chapter 2, I observed consistent results across experiments and I estimate the probability of contamination in additional samples to be at negligible levels. Judging from Extraction Blank Controls (EBCs), I could not find evidence for systematic PCR contamination or carryover effects in the PCR-based analysis. However, not all authenticity criteria listed by (Poinar 2003) (Chapter 1) proved to be useful or efficient at the time of assessing aDNA results, mainly in terms of DNA damage and contamination. Moreover, researchers have applied these criteria to selectively fit to their research conditions (Gilbert et al. 2005).

Since some of the previously authenticity criteria for aDNA studies are no longer applied for validation of results (e.g. bacterial cloning, biochemical preservation or test for racemisation), I followed the guidelines described in (Brandt et al. 2013), aiming to evaluate ancient human DNA in a logical framework performing several independent experiments.

Despite the great advances of NGS technologies applied to aDNA research, the authentication of aDNA results remains an issue. This requires an amendment of standard criteria to work with degraded samples, as proposed by (Knapp et al. 2015).

All samples for the present study were collected using appropriate gear (e.g. surgical gloves, a facemask, and a body suit (Figures 5-6). Working surfaces and tools were decontaminated with 3% bleach before each sample was collected. The large majority of samples were teeth extracted directly from their alveolar sockets. The few bone samples were collected using a Dremel tool and diamond cutting discs in a well-ventilated room at museums and from direct archaeological excavations.

Although it was not possible to obtain genetic profiles from all archaeologist or anthropologist involved in sample collections, all people that carried out the laboratory analysis of aDNA were genotyped to monitor for potential contamination during the laboratory procedures and downstream analysis. All samples for this project were collected and processed exclusively by WH (mtDNA hg H1), BL (hg H3) and GV (hg B2), with assistance of archaeologists at the respective museums or archaeological sites.

In order to ensure the authenticity of aDNA sequences, strict laboratory measures were taken into account to prevent contamination; however we evidenced DNA contamination '*post hoc'* from the sequence data in some of our ancient samples. Contamination of HVR-I sequences as shown in Chapter 2 was detected for two samples that reported incongruent results. Sample 10776A from the Wari culture reported a Native American B4 haplotype, however the second/independent sample 10777A from the same individual gave a European T2 haplotype. Another sample 10784A did not yield amplifiable aDNA, while the replicate sample 10783A reported a K haplotype, also a common European haplogroup. While both cases are clearly non-reproducible, thus they were excluded from the final analysis.

Since none of ACAD personnel responsible for aDNA analysis carries these haplotypes, I assume that the non-Native American mtDNA is likely due to manipulation from archaeologists with European ancestry during excavation. Of note, haplotypes of ACAD laboratory staff had never been observed from aDNA extracts.

Regarding mtDNA genome sequences, I have observed only one incongruent result. Individual sample 13991-Cueva Candelaria from Mexico, showed a pattern of sample mix-up contamination after NGS. This sample was removed from the final analysis, but in this particular case, I can exclude the possibility of contamination by any source of European background since the final sequence reported a combination of haplogroups A2 and B2. I assume that the source of contamination might be explained by sample mix-up during laboratory procedures (i.e. sample preparation or DNA extraction).

Therefore, the first example – HVR-I sequence contamination (two different DNA profiles from two independent samples) highlights the importance of sample collection as fundamental primary step in aDNA studies (Gilbert et al. 2006). The second example – mtDNA genome contamination (two different DNA sequences from the same individual), highlights the importance to rigorously follow the aDNA protocols to avoid DNA contamination in the laboratory.

Library preparation, sequencing and ancient DNA damage

Ancient DNA molecules usually show overhanging 5' or 3' ends, which are prone to deamination that often contain deaminated cytosines in the form of uracil bases. Deamination of cytosines to uracil leads to a characteristic damage signal in ancient samples, being this indicative feature of endogenous DNA (Krause et al. 2010; Sawyer et al. 2012), which is often used as a test for aDNA authenticity.

For library preparation, I used the enzyme cocktail USER (UDG *Uracil-DNA-Glycosylase* + endoVIII) to generate libraries exempt of deaminated cytosines (Briggs et al. 2010). However, aDNA library construction for the present study was subject to standard DNA blunt-end repair and it was not treated for uracil removal.

In doing so, I/we expect an excess of C-to-T substitutions close to the 5' ends of the DNA strands (Figure 2). Genomic libraries from additional poorly preserved samples (Caral-Supe, Aspero, Huari-Miramar and Patagonia) showed a promising pattern after amplification on the gel picture (e.g. typical smear after library amplification with average size: 125-200 bp) (Figure 1). However, after mtDNA genome capture and Illumina sequencing, we did not obtain sufficient human DNA reads from any of these samples for mtDNA genome reconstruction.

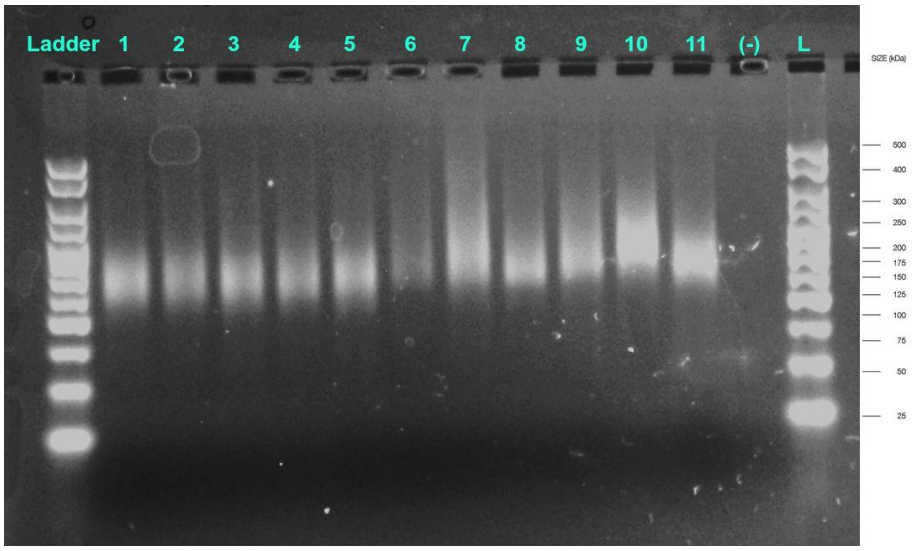


Figure 1. Genomic library preparation of ancient samples. Agarose Gel 3% HyperLadder™ V, Bioline. Batch of 11 samples plus negative control (-).

This might be explained either by the presence of microbial DNA in the samples that was amplified by the library preparation or by the fact that the mtDNA "enrichment" technique did not perform accordingly. However, since I have obtained partial mtDNA sequences for some of these samples (i.e. ACAD 10891-Patagonia), this observation suggests that the amount of "endogenous DNA" on these borderline samples would be relatively too low to be detected by our library preparation methods.

Therefore, these samples were subjected to a first and second round of mtDNA targeted enrichment, (see Material & Methods: Chapter 3), aiming to increase the amount of endogenous DNA from highly degraded samples. Samples included to this experiment were: ACAD 13240-Tiwanaku, 13275-Tiwanaku 12692-Palpa, 10754-Wari, 10758-Wari and 10891-Patagonia; however once more, the DNA sequences obtained showed large regions with no sequence coverage, which impairs the generation of consensus mtDNA sequences.

To evaluate the actual quality of DNA sequencing for every sample, I used a series of parameters (e.g. % raw reads, % mapped reads, % coverage of the mtDNA genome, % average coverage for depth per positions and total number of reads), getting from FastQC analysis, Samtools and Geneious Pro® Software analysis. Moreover, in order to avoid wrong SNP calls due to aDNA damage and low sequence coverage, I used a threshold of 3X minimum coverage for mtDNA SNP calling. Besides, to ensure

comparable data quality for mtDNA genomes, all ancient sequences reported here have less than 1% missing data (Chapter 3).

In the particular case of the Patagonia sample (ACAD-10891), this yielded enough sequence coverage to be able to reconstruct D-loop sequences and the specific HVR-I region (np: 16,005–16,410), that I have analysed in Chapter 2. Of note, this sample showed a total number of 932 mapped reads, which represent the fraction of raw/total reads mapped to the mtDNA reference (i.e. RSRS). Based on observation gained from other samples, the minimum number of mapped reads after Illumina sequencing should be >2,000 reads in order to get a complete coverage of the mtDNA sequence. The 932 reads of sample ACAD-10891 covered 95.1% of mtDNA genome and the overall sequence coverage 3.6%, which did not allow ensemble the whole mtDNA genome to an acceptable level of quality due to large regions with gaps and no sequence coverage at all. This example shows that poor DNA preservation is the most limiting factor in aDNA research, and low percentage of endogenous human DNA in ancient specimens impairs considerable the retrieval of reliable data despite novel techniques developed for such matter (Damgaard et al. 2015).

Characteristic aDNA damage patterns from our samples were assessed using MapDamage2.0 (Ginolhac et al. 2011) (Figures 2-3). Substantial '*post-mortem'* nucleotide misincorporations showing damage patterns (short fragments and an increase in C to T and G to A transitions at 5' and 3' end of reads respectively), are indicative features of aDNA (Krause et al. 2010)(Figure 2). Interestingly, the misincorporation patterns show a trend over time and correlates with the age of the sample (Sawyer et al. 2012). In the case of our samples, I have observed cytosine deamination, manifested by an elevated C to T substitution frequency through time greater than 30% at the end of reads, with a strong correlation ($R^2 = 0.7232$), in concordance with this observation.

On the other hand, '*post-mortem'* fragmentation of DNA in ancient samples occurs incrementally through time mainly before purine residues (A and G). In consequence, fragment length does not decrease consistently over time, therefore no correlation between fragment length and sample age was found (Sawyer et al. 2012). This is also consistent with fragment length from our ancient samples through time (Table 1).

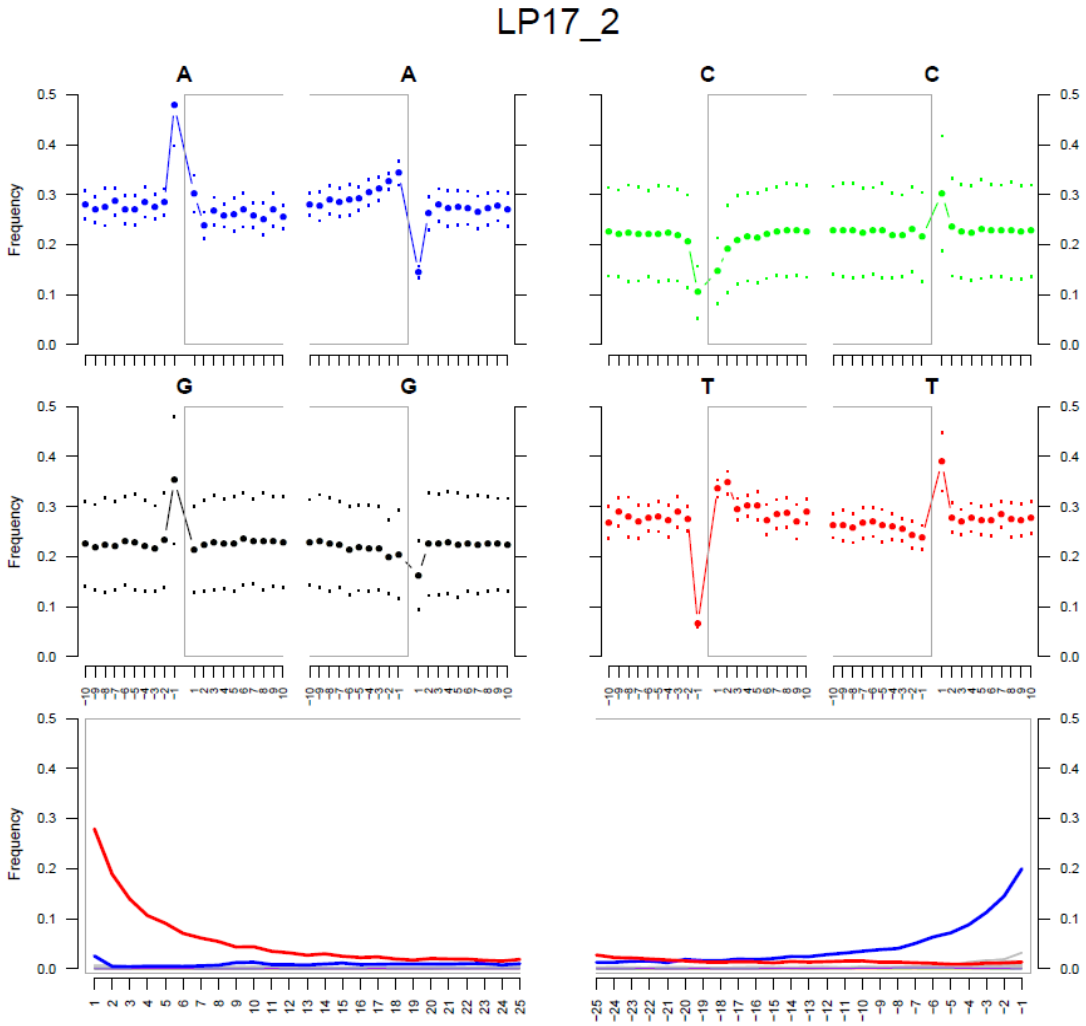


Figure 2. Example of aDNA damage patterns from individual 10754-Wari (LP17.2)
Most of the analysed samples show large amounts of deaminated cytosines residues accumulating towards the end of the reads, meaning that DNA degradation and fragment length distributions and base frequencies were consisted with degraded DNA.
Upper plots: Frequencies of the A, C, G and T bases according to the nucleotide positions within the read (within the grey box) and outside the read (outside the grey bracket). Bottom plots: Frequency distribution of specific substitutions in the first 25 bases (left) (C to T misincorporation) and the last 25 bases (right) (G to A misincorporation) of the read sequence. The two lower plots illustrate the accumulation of 5' C-to-T (red) and 3' G-to-A (blue) misincorporations characteristic of aDNA.

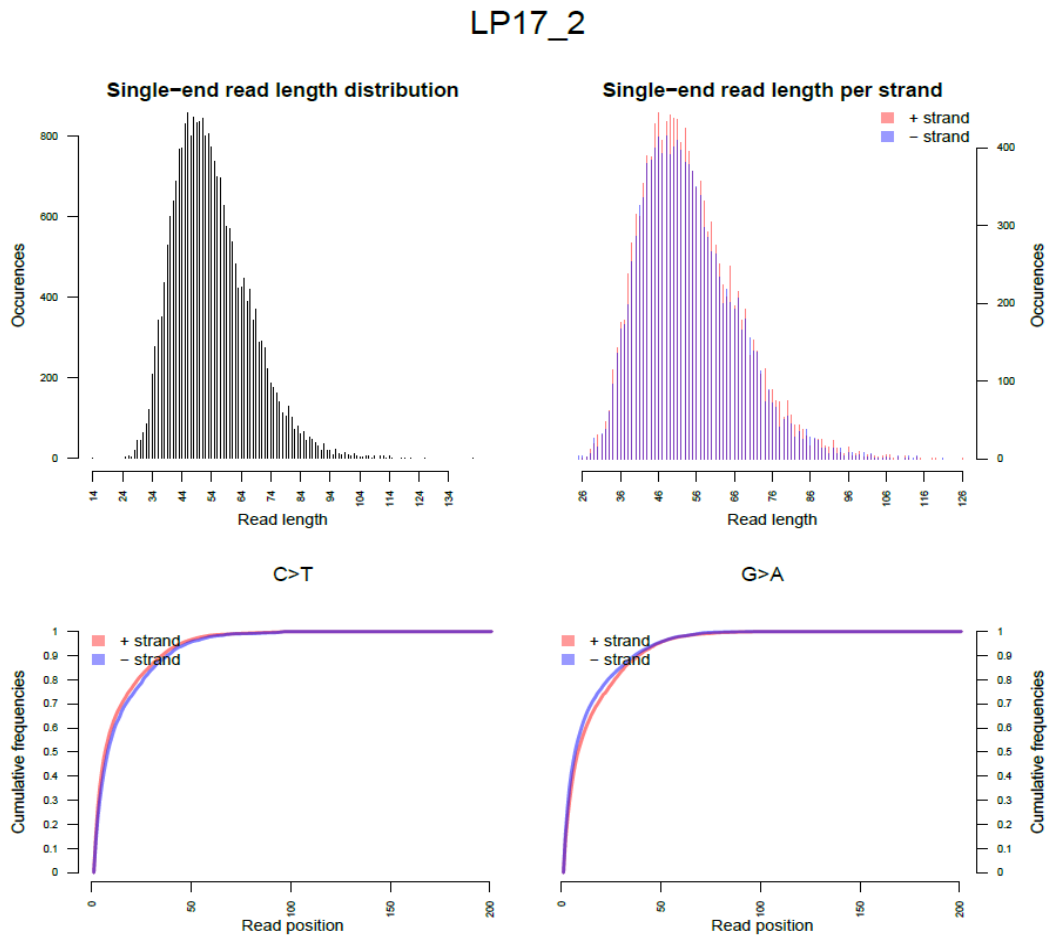


Figure 3. Read length distribution of sample ACAD 10754-Wari (upper panels) and observed cumulative frequency of C to T and G to A misincorporations (lower panels).

Table 1. Coverage and aDNA damage statistics according to sample age.

Sample	Period	Mean age (ya)	Average fold coverage	Average fragment length	5' CtoT	3' GtoA
Palpa-12692 LP27.5	EIP	~1600	26.54±10.12	63.32±18.07	0.30	0.32
Wari-10754 LP17.2 **(Figure 2)**	MH	~1260	49.06±17.04	60.03±14.47	0.30	0.26
Tiwanaku-13240 LP27.2	MH	~1200	11.05±6.08	74.59±17.61	0.24	0.25
Pasamayo-11211 LP14.8	LIP	~780	26.18±8.84	64.66±19.16	0.23	0.18
Llullaillaco-6320 LP28.2	LH	~540	40.49±11.11	59.27±22.85	0.03	0.03

South America: unique and attractive region of the world for aDNA research

The most dynamic sub-region in South America in terms of cultural developments is the Central Andean region (Stanish 2001). This geographic area includes two main major, adjacent environmental and cultural regions that merge into each other across the length of the continent, with a transition between the Andean Highlands and the Pacific coast. A very high cultural development was reached in the Andean region and the greatest social, political, religion and technological achievements belong to the ancient civilizations that arose in this area (Isbell 2008). In the Andean region this socio-political developments evolved into states and empires in pre-Columbian times (i.e. Wari, Tiwanaku and Inca), although it is important to mention the complex regional developments and organizations in the Amazon region in South America (Erickson 2008; Heckenberger 2008; Neves 2008).

The chronology of the Andean cultures is complex. Archaeologists have developed a scheme based on technological achievements and changes of political organization through time from the first arrival of human groups in the region (15,000 – 13,500 BC), until the conquest of the Inca Empire in 1532 to contextualize the cultural development within this area (Lanning 1967).

This development varies across different regions within the Andes, and has been driven by interregional population interactions, trade or religious matters. The main chronological scheme for the Andean region comprises a sequence of eight time units (i.e. five Periods and three Horizons). Periods are defined as times when politically unity across regions was less consolidated, whereas Horizons were the times where much larger political units were formed – review in (Jones 2010).

Several dates have been proposed for the beginnings and endings of Periods and Horizons (see Table 2), but the relative Peruvian chronology proposed by (Rowe 1962; Lumbreras 1974), have defined and contextualized the cultural development in the Central Andes. However, assignation to a particular period and time might vary from one region to another within the Andes, e.g. if we consider the whole region or the Central Andes in particular, because of very dynamic processes such as population demography, settlements, migration or expansion events.

Table 2. Archaeological periods in the Andean region. After (Lanning 1967).

Chronological period	Dates	Principal cultures
Lithic / Archaic Period	15,000 – 3500 BC	Spread of peoples into the Andean Area
Preceramic / Formative period	15,000 – 3500 BC	Early agriculture. First ceremonial centres
Initial Period	3500BC–1800BC	U-shaped ceremonial centres
Early Horizon	900 BC – 200 BC	Chavin, Chiripa, Paracas, Pukara
Early Intermediate Period	200BC – 600AD	Moche, Nazca and Titicaca basin cultures
Middle Horizon	600AD – 1000	Wari and Tiwanaku Empires
Late Intermediate	1000 – 1476	Chimu and Inca Empires
Late Horizon	1476 – 1532	Inca Empire and Spanish conquest

Getting access to valuable ancient samples from South America brings the opportunity to refine the large amount of aDNA research performed in this region so far –see (Fehren-Schmitz et al. 2011a). The geographic and temporal distribution of ancient samples in South America is still uneven and the genetic resolution of available data is still also low.

This project gave me the opportunity to incorporate ancient samples from all cultural periods in the Central Andes. Although the majority our sample locations are restricted to archaeological sites from Peru (Lauricocha, Huaca Pucllana, Chancay, Pueblo Viejo-Pucara), Northern Chile (Chinchorro culture, Pica-Tarapaca), Western Bolivia (Tiwanaku), I was able to add some ancient populations from North America (Cueva Candelaria, Mexico), Argentinean pampas (Arroyo Seco), highlands (Llullaillaco), samples from the Patagonia region and test some from Central America (Taino samples). The incorporation of these additional specimens expanded the geographic sampling coverage in the continent, and represents a crucial leap forward in reconstructing population dynamics in ancient South America.

Archaeological and anthropological research has made significant contributions to our understanding of prehistorical and cultural adaptation of ancient populations in South America – reviewed in *Handbook of South American Archaeology* (Silverman and Isbell 2008), while DNA studies have been able to target genetic variability in ancient and present-day populations.

However, combining archaeological information and aDNA data can provide a powerful means to study cultural processes and cultural transitions in South America. In that regard, this thesis has merged both archaeological and genetics approaches to build a detailed picture of relationships and interactions of pre-Columbian populations in the Central Andes. Variations of mtDNA haplogroups frequencies across different archaeological periods were contrasted and the obtained results offered substantial information to infer demographic and population genetics changes in the Central Andes of South America (Chapter 2-5).

The most interesting archaeological period that at the same time highlighted the importance of our data represents the Middle Horizon period (600–1000 AD). The Middle Horizon harbours a series of population dynamic processes, inter-regional interaction, social stratification and socio-political changes in the Central Andes (Isbell and Schreiber 1978; Schreiber 1992; Tung 2007; Isbell 2008; Jennings 2010; Marcone 2010). The Wari culture has had a big socio-political impact in Central and coastal Peru during the Middle Horizon that has been subjected to archaeology and anthropological research (Marcone 2010; Segura and Shimada 2010).

In order to understand demographic events in highland and coastal region of Peru, to contextualize such an important transitional period in the Central Andes (i.e. Middle Horizon), I have investigated the impact of the Wari Empire in terms of colonization and imperialism by analysing aDNA from a defined archaeological site (i.e. Huaca Pucllana), evaluating the genetic impact of the Wari expansion in this particular archaeological site (Chapter 4). Based on haplogroup composition, I suggest a subtle impact of the Wari Empire expansion in Peru's Central Coast during the Middle Horizon; however these might not be representative in the broader context of the Peruvian coastal region occupied by the Wari Empire in pre-Columbian times.

DNA preservation of samples from South America

Preservation and integrity of the DNA molecule is fundamental in aDNA research. In this project, I was able to perform the analysis of more than 300 human ancient remains from ~150 individuals from South America. The success in aDNA isolation depends primarily on the state of sample preservation which is linked to some

environmental conditions such as low temperature, humidity, salt concentration, aridity and neutral of alkaline pH of soils (Höss et al. 1996; Burger et al. 1999).

I have analysed samples from a broad time transect in South America, with the oldest specimens from Lauricocha, Peru (ranging 8700–8599 calBP) and the youngest samples from Pueblo-Viejo, Peru (ranging ~500 BP). I have observed a highly variable aDNA preservation that did not necessarily correlate the antiquity of samples but might be related to burial conditions (Chapter 3-4). Archaeological sites where temperature is low, at high-altitude or in dry environments allow long-time preservation of DNA (Fehren-Schmitz et al. 2015). This was shown by the successful retrieval of DNA from the Bolivian samples belonging to the Tiwanaku Culture (3,800 m.a.s.l) in La Paz, Bolivia with a success rate of 90%, and the Lauricocha samples in the Peruvian highlands (~ 4000m) with a success rate of 100%.

In contrast, aDNA retrieval from individuals found in warm and moderate or high temperatures sites is less likely. In consequence, samples coming from coastal sites in South America showed a significant lower preservation (Fehren-Schmitz 2008), which is consistent with the samples collected at Peruvian lowland archaeological sites (i.e. Caral and Aspero) that failed for DNA retrieval despite several attempts with classical and novel techniques.

However, for some other samples coming also from coastal regions in Peru (i.e. Huaca Pucllana individuals), the preservation was good overall. This difference in sample preservation from coastal sites might be therefore related to the funerary practice differences among ancient groups from Andean and Coastal regions (Fehren-Schmitz et al. 2011a). For instance, human remains from high-altitude places might be found in "chullpas" that constitutes ancient funerary stone towers found across the Andean region of Bolivia and Peru, or caves, providing better conditions for preservation. On the other hand, samples from coastal groups are usually found buried in desert soils, which are not always suitable for DNA preservation (Fehren-Schmitz et al. 2011a). For some samples coming from coastal sites (e.g. Lima, Wari and Ychsma cultures), the overall successful retrieval of aDNA is relatively even for the three groups despite distinctive burial and funerary mortuary practices. For instance, human sacrifices relate the funerary characteristic of the Lima culture.

Human bodies found in "fardos" or funerary bundles, where humans were wrapped with clothes together with various offerings, relate the funerary practices of the Wari period (Chapter 4). Samples recovered from such "fardos" (i.e. Wari individuals) showed reliable aDNA retrieval (50%). Hence, fardos could favour a good DNA preservation (e.g. low temperatures and fast desiccation).

The temporal distribution of successfully analysed samples according to archaeological period is also variable (Chapter 2). Samples successfully analysed for HVR-I showed a success rate of 4.1% for the Late Archaic period. Samples analysed for the Early Intermediate period showed a success rate of 13.8%, and 26.3% for the Middle Horizon. The Late Intermediate period show the highest reported success rate for the analysed samples (36.1%) and the Late Horizon reports (19.4%).

On the other hand, successful sequences obtained for complete mtDNA genomes showed a 11.9% success rate for the Archaic period. Samples from the Early Intermediate period showed a success rate of (16.3%), and for the Middle Horizon (20.7%). The Late Intermediate period showed also the highest percentage (34.7%) of success rate reported for mtDNA sequences. Finally the success rate for the Late Horizon showed (16.3%) (Figure 4).

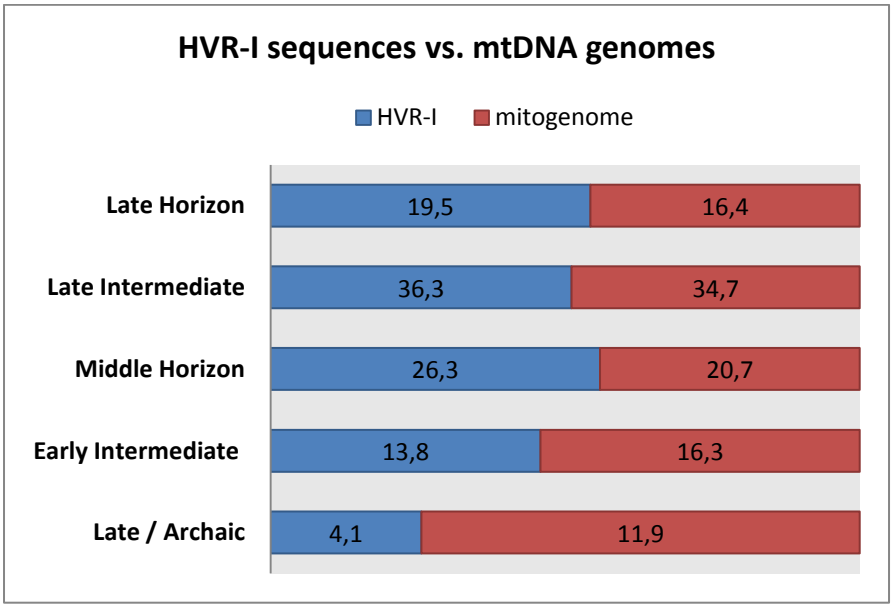


Figure 4. Comparison of success rates for the generation of mtDNA sequences for all archaeological periods analysed in this thesis. The bars show the proportion of successful HVR-I (in blue) versus complete mtDNA (in red) sequences, while numbers indicate the proportion of samples for which data was generated within each cultural period.

I have also compared the three traditional PCR-based typing methods used in this thesis to analyse aDNA from South America. As seen in Chapter 2, the sequencing of HVR-I – success rate: 27% (40 out of 149 samples), genotyping with GenoCoRe22 – success rate 28% (9 out of 149 samples) and AmericaPlex26 – success rate 24% (32 out of 130 samples have shown a reliable performance of our laboratory methods with consistent replication and no significant differences when compared to each other ($p < 0.01$, Chi-square test).

Besides, the development and design of the AmericaPlex26 technique was critical in screening for potential samples for the downstream analysis (i.e. hybridization capture of mtDNA and complete mtDNA sequencing), and identifying potential samples with sufficient DNA preservation and free of contaminants (Chapter 5). The AmericaPlex26 SNP assay, represents a very versatile and highly efficient cost-effective technique that allows genotype Native American mtDNA "founder lineages" plus additional sub-haplogroups in comparison with the efficiency of HVR-I sequencing (Coutinho et al. 2014).

From archaeological excavation to genomic data generation

During my PhD candidature I had the opportunity to go through most stages in aDNA research. Back in 2010, I collected sample *in-situ*, i.e. directly from bones and teeth at active excavations at the Tiwanaku archaeological site in La Paz, Bolivia, in collaboration with the Bolivian Archaeology Office. I was the only person involved in collecting the specimens for aDNA analysis following aDNA sampling criteria to minimize contamination of specimens (Brown 1998; Yang and Watt 2005)(Llamas et al., in press) (Figures 5-6). Contamination by external sources is most likely to be introduced at the time of first handling of specimens after excavation (Gilbert et al. 2006). However, improved field sampling techniques can be used to reduce the risk of contamination (Fortea et al. 2008). Few studies have evaluated the importance of performing sampling on human skeletal remains direct from archaeological excavation in order to monitor contamination between freshly excavated samples and samples that have been handled subsequently (Pruvost et al. 2007; Pilli et al. 2013).

It has been suggested that teeth are less prone to contamination than any other skeletal areas since they are protected in the alveolar sockets and for other types of samples, such as large bones, external contamination correlates with structural preservation – e.g. porosity of the surface to attract more contaminants from the environment (Thomas P. Gilbert et al. 2005). Nevertheless, bones are also suitable for human aDNA analyses if they come directly from the excavation sites (Pilli et al. 2013). The application of good sampling guidelines / strategies is key to reducing the risk of human contamination and guaranteeing the reliability of aDNA results (Allentoft 2013; Knapp et al. 2015).

Regarding aDNA laboratory work, I have been trained in all methodological steps and followed the strict criteria of authenticity and validation of results in aDNA research when working in the contamination-free lab at the Australian Centre for Ancient DNA. The risks of contamination during analysis of aDNA are ubiquitous and possible at any step of the protocols, even in dedicated aDNA facilities in cases where the personnel fails to follow or underestimates the standard analytical framework.

As mentioned before, parallel genotyping of all personnel involved in handling of the samples is a possible measure to check for possible contamination. This is particularly critical where the geographic origin of samples and lab personnel overlap. Here, it is important to mention that my own mtDNA haplogroup is B2. Since the frequency of B2 haplogroup in the overall sample set is one of the highest (24.7%), one could *a priori* assume a substantial amount of contamination.

However, a direct comparison of a relevant B2 haplotypes, indicates that my B2-types carries a set of mutation in the D-loop that are not present in any of the ancient samples investigated in this study. Diagnostic SNPs for haplogroup B2 and SNPs "private" to GV in bold are as follows (np: **T63C**, **C64T**, **G66T**, C146T, C152T, C195T, A247G, A16129G, A16183C, T16187C, T16278C, **C16292T**, C16311T, T16217C, T16223C, G16230A)(See Chapter 3; SI Table 2). In consequence, I can rule out possible direct contamination from my end and show that the aDNA analysis has been carried out with sufficient stringency.

Figure 5. Guido Valverde (Candidate) collects ancient human samples at the Tiwanaku Archaeological site in La Paz, Bolivia (source: Guido Valverde).

Figure 6. Sample collection of aDNA specimens requires appropriate protective clothing (e.g. facemasks, hair caps, sterile gloves, goggles, etc.) to reduce to amount of contamination (source: Guido Valverde).

Figure 7. Example of a bone sample collected at archaeological excavations and subjected to aDNA analysis in this study.

Figure 8. Example of a tooth sample collected at archaeological excavations subjected to aDNA analysis in this study.

Limitations and future directions

Most of the aDNA studies in South America are currently restricted to the Central Andes and coastal regions (Moraga et al. 2005; Shinoda et al. 2006; Fehren-Schmitz et al. 2009; Kemp et al. 2009; Fehren-Schmitz et al. 2010; Fehren-Schmitz et al. 2011a; Baca et al. 2012; Baca et al. 2014; Fehren-Schmitz et al. 2014). These are regions with a high density of finds and environmental conditions that allow natural mummification and are favourable for preservation of skeletal remains.

As stated previously, climate conditions play a role in sample preservation and therefore not many samples from eastern regions in South America with warmer and hot environments are suitable for analysis. So far, there are very few reports on human aDNA analysis from hotter regions in the continent, e.g. Brazil (Ribetio-dos-Santos et al. 1996; Bisso-Machado et al. 2012), although the first genomic data from ancient Amazonian populations from Brazil was reported lately (Raghavan et al. 2015).

Nevertheless, it will be essential to widen the geographic sampling coverage to add specimens from eastern and southern parts of the continent, such as the Amazonian and the Patagonia region. A larger sample size and subsequent analysis of the genomic information will be important to reinforce our understanding about migrations and relationships among all ancient cultures in a broader study in South America.

In order to study the cultural and demographic interactions of pre-Columbian populations, it will also require to incorporate specimens from the earliest archaeological periods in South America. A significant contribution was made by a recent study by (Fehren-Schmitz et al. 2015), which reports aDNA from Lauricocha in highlands Peru, one of the earliest Early to Middle Holocene sites in the Andes, suggesting that the mtDNA diversity of Lauricocha individuals falls in the same spectrum of pre-Columbian and modern-day Native American populations.

Due to the availability of material for aDNA research especially from South American archaeological sites, future sampling should take into account the possibility to obtain DNA from other sources (i.e. petrous part of the temporal bone). Latest research has evaluated the feasibility on such material suggesting systematically higher endogenous DNA content and the most likely source for future aDNA studies (Pinhasi et al. 2015). This approach might be important to test in order to sequence nuclear genomes from South America eastern regions, where good DNA quality is key to retrieve aDNA from specimens from moderate or hot environments (Llorente et al. 2015).

The continuous optimisation of laboratory techniques and protocols will improve our abilities to study highly degraded samples. For instance, it would be important to test samples from South America that failed in this study (i.e. Caral samples), considered the oldest civilization in South America, in order to see whether or not they are suitable for analysis by techniques described in (Dabney et al. 2013) to deal with low endogenous DNA, or applying the single-strand library preparation method (Gansauge and Meyer 2013), described as a reliable technique to improve sequence retrieval from highly degraded samples. Such techniques are currently established at ACAD.

Regardless of the apparent low genetic diversity in Native American groups expressed in only five major mtDNA haplogroups (A, B, C, D and X), I have shown that there sufficient resolution in complete mtDNA genomes to improve our understanding on pre-Columbian diversity (Chapter 3). However, sequencing of complete mtDNA genomes is not the limit for aDNA studies. Whole genome nuclear data are currently key to fully explore aspects in genetic research such as population history, natural selection or population admixture (Kivisild 2015). We still have to be able to achieve one more step in terms of data generation in the upcoming years. The necessity to shift the analysis from mtDNA to nuclear DNA represents the strategy the aDNA field is pursuing nowadays and it will be important to take into account for further aDNA research in South America.

Although, large-scale population genetic studies on prehistoric groups are still not feasible in the Americas, recent studies have reported complete nuclear genomes from Native American individuals exploring ancient genome diversity in North America (Rasmussen et al. 2010; Raghavan et al. 2014; Rasmussen et al. 2014). Latest research has reported the first ancient genomic data from South America (Raghavan et al. 2015), however with still low genome coverage. In fact, we are currently generating complete genome data from a selection of pre-Columbian samples, such as the Peruvian Ychsma culture.

I am confident that I have collected, screened and identified very valuable specimens with great potential for in-depth genome studies. In addition, other ancient samples are currently captured for a set of 1240K genome-wide SNPs using an in-solution array co-developed by Matthias Meyer's group at the Max Planck Institute in Leipzig and David Reich's team at the Harvard Medical School in Boston.

This work will definitely enhance our knowledge of South American genetic diversity. Furthermore, it will be necessary to increase sequencing efforts for more modern-day Native American populations, in order to achieve full compatibility with ancient data.

Concluding remarks

Ancient DNA is a very challenging but also very exciting research field. Research on aDNA has highlighted the time-travelling potential of this field and the insights gained throughout many years of research have provided a different perspective about human genetics that would not have been achieved by analysing modern samples only. Mitochondrial DNA has been widely used as genetic marker to explore human genetic diversity in prehistoric and modern-day populations. Due to the lack of recombination and its maternal inheritance, mtDNA have offered the opportunity to trace back the maternal variation across space and time, which (when combined with other mitochondrial lineages), provides the female perspective on human population history at continental and regional scales (Kivisild 2015).

The present work has shed light on the mtDNA diversity from ancient population of South America sampled from a broad temporal and geographic dimensions, which allowed me to uncover (or better "rediscover") lost genetic diversity, here expressed in novel ancient mtDNA lineages. I can infer from results that no dramatic changes in genetic composition in South America in terms of population structure have taken place in pre-Columbian times. Therefore, the genetic makeup of Native Americans suggests that this diversity might have been established much earlier before the European contact (Raff et al. 2011). However, mtDNA diversity in the Americas has experienced a considerable disruption after the arrival of Europeans due to demographic changes, with substantial extinction of pre-Columbian lineages. Nevertheless, the novel mtDNA genome sequences from ancient samples reported in this thesis might represent such potentially lost diversity (Chapter 3).

The addition of more samples and high-resolution genomic data in the foreseeable future will increase the power to refine our understanding of population history in ancient South America.

New insights can be therefore obtained with the integration of mitochondrial genomes that will provide the best supported scenarios to explain questions such as peopling of the continent (Llamas et al., in press), migration patterns, population transitions, replacements and population continuity and discontinuity of the Native American gene pool through time.

Mitochondrial genome sequencing will continue having an importance in human genetics and still plays a role in terms of population evolutionary genetics, ancestry, genealogy and phylogeographic and phylogenetic studies. Sequencing of ancient mtDNA genomes still represents valuable time-stamped data when reconciled with wide range of disciplines and fields such as archaeology, anthropology, linguistics and history (Bortolini et al. 2014). New technological improvements have re-invented the field of aDNA from a highly controversial beginning to a central component of modern anthropological research (Knapp et al. 2015). Ancient DNA studies in the Americas will continue to provide additional data in order to unveil the complex peopling history of the continent.

REFERENCES

Allentoft ME. 2013. Recovering samples for ancient DNA research-guidelines for the field archaeologist. Antiquity 87(338).

Baca M, Doan K, Sobczyk M, Stankovic A, and Weglenski P. 2012. Ancient DNA reveals kinship burial patterns of a pre-Columbian Andean community. BMC Genet 13(1):30.

Baca M, Molak M, Sobczyk M, Weglenski P, and Stankovic A. 2014. Locals, resettlers, and pilgrims: A genetic portrait of three pre-Columbian Andean populations. Am J Phys Anthropol.

Bisso-Machado R, Bortolini MC, and Salzano FM. 2012. Uniparental genetic markers in South Amerindians. Genetics and Molecular Biology 35(2):365-387.

Bortolini MC, Gonzalez-Jose R, Bonatto SL, and Santos FR. 2014. Reconciling pre-Columbian settlement hypotheses requires integrative, multidisciplinary, and model-bound approaches. Proceedings of the National Academy of Sciences 111(2):E213-E214.

Brandt G, Haak W, Adler CJ, Roth C, Szecsenyi-Nagy A, Karimnia S, Moller-Rieker S, Meller H, Ganslmeier R, Friederich S et al. . 2013. Ancient DNA reveals key stages in the formation of central European mitochondrial genetic diversity. Science 342(6155):257-261.

Briggs AW, Stenzel U, Meyer M, Krause J, Kircher M, and Paabo S. 2010. Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. Nucleic Acids Res 38(6):e87.

Brown K. 1998. Keeping it clean: the collection and storage of ancient DNA samples from the field. The Archaeologist 33:16 -17.

Burger J, Hummel S, Hermann B, and Henke W. 1999. DNA preservation: a microsatellite-DNA study on ancient skeletal remains. Electrophoresis 20(8):1722-1728.

Carnese F, Mendisco F, and Keyser C. 2010. Paleogenetical study of pre-Columbian samples from Pampa Grande (Salta, Argentina). Am J Phys Anthropol 141:452 - 462.

Cooper A, and Poinar HN. 2000. Ancient DNA: Do It Right or Not at All. Science(289):1139.

Coutinho A, Valverde G, Fehren-Schmitz L, Cooper A, Barreto Romero MI, Espinoza IF, Llamas B, and Haak W. 2014. AmericaPlex26: a SNaPshot multiplex system for genotyping the main human mitochondrial founder lineages of the Americas. PLoS ONE 9(3):e93292.

Cui Y, Lindo J, Hughes CE, Johnson JW, Hernandez AG, and et al. 2013. Ancient DNA Analysis of Mid-Holocene Individuals from the Northwest Coast of North America Reveals Different Evolutionary Paths for Mitogenomes. PLoS ONE 8(7):e66948. doi:66910.61371/journal.pone.0066948.

Dabney J, Knapp M, Glocke I, Gansauge MT, Weihmann A, Nickel B, Valdiosera C, Garcia N, Paabo S, Arsuaga JL et al. . 2013. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. Proc Natl Acad Sci U S A 110(39):15758-15763.

Damgaard PB, Margaryan A, Schroeder H, Orlando L, Willerslev E, and Allentoft ME. 2015. Improving access to endogenous DNA in ancient bones and teeth. Sci Rep 5:11184.

Erickson CL. 2008. Amazonia: The Historical Ecology of a Domesticated Landscape. In: Silverman H, and Isbell WH, editors. The Handbook of South American Archaeology. New York, NY: Springer New York. p 157-183.

Fagundes NJ, Kanitz R, Eckert R, Valls AC, Bogo MR, Salzano FM, Smith DG, Silva WA, Jr., Zago MA, Ribeiro-dos-Santos AK et al. . 2008. Mitochondrial population genomics supports a single pre-Clovis origin with a coastal route for the peopling of the Americas. Am J Hum Genet 82(3):583-592.

Fehren-Schmitz L. 2008. Molekularanthropologische Untersuchungen zur prakolumbischen Besiedlungsgeschichte des suedlichen Perus am Beispiel der Palpa Region. Goettingen: PhD thesis Department of Biology, University Goettingen.

Fehren-Schmitz L, Haak W, Machtle B, Masch F, Llamas B, Tomasto Cagigao E, Sossna V, Schittek K, Isla Cuadrado J, Eitel B et al. . 2014. Climate change underlies global demographic, genetic, and cultural transitions in pre-Columbian southern Peru. Proceedings of the National Academy of Sciences.

Fehren-Schmitz L, Hummel S, and Herrmann B. 2009. Who Were the Nasca? Population Dynamics in Pre-Columbian Southern Peru Revealed by Ancient DNA Analyses.159-172.

Fehren-Schmitz L, Llamas B, Lindauer S, Tomasto-Cagigao E, Kuzminsky S, Rohland N, Santos FR, Kaulicke P, Valverde G, Richards SM et al. . 2015. A Re-Appraisal of the Early Andean Human Remains from Lauricocha in Peru. PLoS ONE 10(6).

Fehren-Schmitz L, Llamas B, Tomasto E, and Haak W. 2011a. El ADN antiguo y la historia del poblamiento temprano del oeste de Sudamérica: lo que hemos aprendido y hacia dónde vamos. Boletin de Arqueologia PUCP 15(ISSN 1029-2004 17):17 - 41.

Fehren-Schmitz L, Reindel M, Cagigao E, Hummel S, and Herrmann B. 2010. Pre-Columbian population dynamics in coastal southern Peru: A diachronic investigation of mtDNA patterns in the Palpa region by ancient DNA analysis. Am J Phys Anthropol 141:208 - 221.

Fehren-Schmitz L, Warnberg O, and Reindel M. 2011b. Diachronic investigations of mitochondrial and Y-chromosomal genetic markers in pre-Columbian Andean highlanders from South Peru. Ann Hum Genet 75:266 - 283.

Fortea J, de la Rasilla M, Garcia-Tabernero A, Gigli E, Rosas A, and Lalueza-Fox C. 2008. Excavation protocol of bone remains for Neandertal DNA analysis in El Sidron Cave (Asturias, Spain). J Hum Evol 55(2):353-357.

Fulton TL. 2012. Setting up an ancient DNA laboratory. Methods Mol Biol 840:1-11.

Gansauge MT, and Meyer M. 2013. Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. Nature Protocols 8(4):737-748.

Gilbert M, Bandelt H-J, Hofreiter M, and Barnes I. 2005. Assessing ancient DNA studies. Trends in Ecology and Evolution 20:541 - 544.

Gilbert MTP, Hansen AJ, Willerslev E, Turner-Walker G, and Collins M. 2006. Insights into the processes behind the contamination of degraded human teeth and bone samples with exogenous sources of DNA. Int J Osteoarchaeol 16:156-164.

Gilbert MTP, Kivisild T, Grønnow B, Andersen PK, Metspalu E, Reidla M, Tamm E, Axelsson E, Götherström A, Campos PF et al. . 2008. Paleo-Eskimo mtDNA genome reveals matrilineal discontinuity in Greenland. Science (New York, NY) 320:1787-1789.

Ginolhac A, Rasmussen M, Gilbert MT, Willerslev E, and Orlando L. 2011. mapDamage: testing for damage patterns in ancient DNA sequences. Bioinformatics 27(15):2153-2155.
Goebel T, Waters MR, and O'Rourke DH. 2008. The late Pleistocene dispersal of modern humans in the Americas. Science 319(5869):1497-1502.
Haak W, Balanovsky O, Sanchez JJ, Koshel S, Zaporozhchenko V, Adler CJ, Der Sarkissian CSI, Brandt G, Schwarz C, Nicklisch N et al. . 2010. Ancient DNA from European early neolithic farmers reveals their near eastern affinities. PLoS biology 8(11):e1000536.
Heckenberger MJ. 2008. Amazonian Mosaics: Identity, Interaction, and Integration in the Tropical Forest. In: Silverman H, and Isbell WH, editors. The Handbook of South American Archaeology. New York, NY: Springer New York. p 941-961.
Höss M, Jaruga P, Zastawny TH, Dizdaroglu M, and Pääbo S. 1996. DNA damage and DNA sequence retrieval from ancient tissues. Nucleic Acids Res 24(7):1304-1307.
Isbell WH. 2008. Wari and Tiwanaku: International Identities in the Central Andean Middle Horizon, en: H. I. Silverman y W. H. Isbell (eds). Handbook of South American Archaeology Springer, New York.:731-759.
Isbell WH, and Schreiber KJ. 1978. Was Huari a State? American Antiquity 43:372-389.
Jennings J. 2010. Beyond Wari Walls : Regional Perspectives on Middle Horizon Peru. Albuquerque, NM, USA: University of New Mexico Press.
Jones MD. 2010. The Everyday Life Of The Ancient Incas. Anness Publishing Ltd ISBN 1846811694.
Karafet TM, Zegura SL, Posukh O, Osipova L, Bergen A, Long J, Goldman D, Klitz W, Harihara S, de Knijff P et al. . 1999. Ancestral Asian Source(s) of New World Y-Chromosome Founder Haplotypes. AmJHumGenet 64:817-831.
Kemp BM, Tung TA, and Summar ML. 2009. Genetic continuity after the collapse of the Wari empire: mitochondrial DNA profiles from Wari and post-Wari populations in the ancient Andes. Am J Phys Anthropol 140(1):80-91.
Kivisild T. 2015. Maternal ancestry and population history from whole mitochondrial genomes. Investig Genet 6(1).
Knapp M, Clarke AC, Horsburgh KA, and Matisoo-Smith EA. 2012. Setting the stage - building and working in an ancient DNA laboratory. Ann Anat 194(1):3-6.
Knapp M, Lalueza-Fox C, and Hofreiter M. 2015. Re-inventing ancient human DNA. Investig Genet 6(1).
Krause J, Briggs AW, Kircher M, Maricic T, Zwyns N, Derevianko A, and Paabo S. 2010. A complete mtDNA genome of an early modern human from Kostenki, Russia. Curr Biol 20(3):231-236.
Lanning EP. 1967. Peru Before the Incas. Englewood Cliffs: Prentice Hall.
Lewis C. 2009. Difficulties in rejecting a local ancestry with mtDNA Haplogroup data in the South Central Andes. Latin American antiquity.
Livi-Bacci M. 2006. The Depopulation of Hispanic America after the Conquest. Population and Development Review 32(2):199-232.
Lumbreras LG. 1974. The Peoples and Cultures of Ancient Peru. Smithsonian Institution Press, Washington, DC.
Llorente MG, Jones ER, Eriksson A, Siska V, Arthur KW, Arthur JW, Curtis MC, Stock JT, Coltorti M, Pieruccini P et al. . 2015. Ancient Ethiopian genome reveals extensive Eurasian admixture throughout the African continent. Science.
Marcone G. 2010. What Role Did Wari Play in the Lima Political Economy?: The Peruvian Central Coast at the Beginning of the Middle Horizon. In J Jennings

(ed), *Beyond Wari Walls Regional Perspectives on Middle Horizon Peru*, The University of New Mexico Press, Alburquerque Chapter 8:136-154.

Meltzer DJ. 1993. Pleistocene peopling of the Americas. Evol Anthropol 1::157–169.

Montiel R, Malgosa A, and Francalacci P. 2001. Authenticating Ancient Human Mitochondrial DNA. Human Biology 73(5):689-713.

Moraga M, Santoro C, Standen V, Carvallo P, and Rothhammer F. 2005. Microevolution in prehistoric Andean populations: chronologic mtDNA variation in the desert valleys of northern Chile. Am J Phys Anthropol 127:170 - 181.

Neves EG. 2008. Ecology, Ceramic Chronology and Distribution, Long-term History, and Political Change in the Amazonian Floodplain. In: Silverman H, and Isbell WH, editors. The Handbook of South American Archaeology. New York, NY: Springer New York. p 359-379.

O'Fallon BD, and Fehren-Schmitz L. 2011. Native Americans experienced a strong population bottleneck coincident with European contact. Proc Natl Acad Sci U S A.

Pilli E, Modi A, Serpico C, Achilli A, Lancioni H, Lippi B, Bertoldi F, Gelichi S, Lari M, and Caramelli D. 2013. Monitoring DNA contamination in handled vs. directly excavated ancient human skeletal remains. PLoS ONE 8(1):25.

Pinhasi R, Fernandes D, Sirak K, Novak M, Connell S, Alpaslan-Roodenberg S, Gerritsen F, Moiseyev V, Gromov A, Raczky P et al. . 2015. Optimal Ancient DNA Yields from the Inner Ear Part of the Human Petrous Bone. PLoS ONE 10(6):e0129102.

Poinar H. 2003. The top 10 list: criteria of authenticity for DNA from ancient and forensic samples. Int Congr Ser 1239:575 - 579.

Pruvost M, Schwarz R, Correia V, Champlot S, Braguier S, Morel N, Fernandez-Jalvo Y, Grange T, and Geigl E. 2007. Freshly excavated fossil bones are best for amplifcation of ancient DNA. Proccedings of the National Academy of Science USA 104(3):739 - 744.

Raff JA, Bolnick DA, Tackney J, and O'Rourke DH. 2011. Ancient DNA perspectives on American colonization and population history. Am J Phys Anthropol.

Raghavan M, DeGiorgio M, Albrechtsen A, Moltke I, Skoglund P, Korneliussen TS, Gronnow B, Appelt M, Gullov HC, Friesen TM et al. . 2014. The genetic prehistory of the New World Arctic. Science 345(6200):1255832.

Raghavan M, Steinrücken M, Harris K, Schiffels S, Rasmussen S, DeGiorgio M, Albrechtsen A, Valdiosera C, Ávila-Arcos MC, Malaspinas A-S et al. . 2015. Genomic evidence for the Pleistocene and recent population history of Native Americans. Science.

Rasmussen M, Anzick SL, Waters MR, Skoglund P, DeGiorgio M, Stafford TW, Jr., Rasmussen S, Moltke I, Albrechtsen A, Doyle SM et al. . 2014. The genome of a Late Pleistocene human from a Clovis burial site in western Montana. Nature 506(7487):225-229.

Rasmussen M, Li Y, Lindgreen S, Pedersen JS, Albrechtsen A, Moltke I, Metspalu M, Metspalu E, Kivisild T, Gupta R et al. . 2010. Ancient human genome sequence of an extinct Palaeo-Eskimo. Nature 463(7282):757-762.

Ribetio-dos-Santos AK, Santos SE, Machado AL, Guapindaia V, and Zago MA. 1996. Heterogeneity of mitochondrial DNA haplotypes in Pre-Columbian Natives of the Amazon region. Am J Phys Anthropol 101(1):29-37.

Rowe J. 1962. Stages and Periods in Archaeological Interpretation. Southwestern Journal of Anthropology 8:40-54.

Sawyer S, Krause J, Guschanski K, Savolainen V, and Paabo S. 2012. Temporal Patterns of Nucleotide Misincorporations and DNA Fragmentation in Ancient DNA. Plos One 7(3):e34131. doi:34110.31371/journal.pone.0034131.

Schreiber KJ. 1992. Wari imperialism in Middle Horizon Peru. Anthropological Papers of the Museum of Anthropology 87:University of Michigan, Ann Arbor.

Schurr TG, Ballinger SW, Gan YY, Hodge JA, Merriwether DA, Lawrence DN, Knowler WC, Weiss KM, and Wallace DC. 1990. Amerindian mitochondrial DNAs have rare Asian mutations at high frequencies, suggesting they derived from four primary maternal lineages. Am J Hum Genet 46(3):613-623.

Schurr TG, and Sherry ST. 2004. Mitochondrial DNA and Y chromosome diversity and the peopling of the Americas: evolutionary and demographic evidence. Am J Hum Biol 16(4):420-439.

Segura R, and Shimada I. 2010. The Wari Footprint on the Central Coast: A View from Cajamarquilla and Pachacamac. Beyond Wari walls : Regional perspectives on Middle Horizon PeruThe University of New Mexico Press Chapter 7:113-135.

Shinoda K, Adachi N, Guillen S, and Shimada I. 2006. Mitochondrial DNA analysis of ancient Peruvian highlanders. Am J Phys Anthropol 131(1):98-107.

Silverman H, and Isbell WH. 2008. Handbook of South American Archaeology. Springer New York (NY)

Stanish C. 2001. The origin of state societies in South America. Annual Review of Anthropology 30:41-64.

Thomas P. Gilbert M, Rudbeck L, Willerslev E, Hansen AJ, Smith C, Penkman KEH, Prangenberg K, Nielsen-Marsh CM, Jans ME, Arthur P et al. . 2005. Biochemical and physical correlates of DNA contamination in archaeological human bones and teeth excavated at Matera, Italy. Journal of Archaeological Science 32(5):785-793.

Torroni A, Schurr T, Cabell M, Brown M, Neel J, Larsen M, Smith D, Vullo C, and Wallace D. 1993. Asian affinities and continental radiation of the four founding Native American mtDNAs. Am J Hum Genet 53:563–590.

Tung TA. 2007. Trauma and violence in the Wari empire of the Peruvian Andes: Warfare, raids, and ritual fights. Am J Phys Anthropol 133(3):NA-NA.

Yang DY, and Watt K. 2005. Contamination controls when preparing archaeological remains for ancient DNA analysis. Journal of Archaeological Science 32(3):331-336.