Doctoral Programme in Population Health
Department of Forensic Medicine
Faculty of Medicine
University of Helsinki
Finland

GENETIC INSIGHTS ON OLD RIDDLES: THE CHACHAPOYAS FROM THE NORTHEASTERN PERUVIAN CLOUD FORESTS

Evelyn K. Guevara

Doctoral dissertation
To be presented, with the permission of the Faculty of Medicine of the University of Helsinki, for public examination in Hall 1, Haartman Institute, Haartmaninkatu 3, Helsinki, on May 6th 2022, at noon

Helsinki 2022

Supervisors

Professor Antti Sajantila
Department of Forensic Medicine, University of Helsinki
and
Forensic Medicine Unit, Finnish Institute for Health and Welfare
Helsinki, Finland

Docent Jukka Palo
Department of Forensic Medicine, University of Helsinki
and
Forensic Genetics Unit, Finnish Institute for Health and Welfare
Helsinki, Finland

Reviewers

Professor Leonor Gusmão
University of the State of Rio de Janeiro
Brazil

Professor Kristiina Tambets
Estonian Biocentre, University of Tartu
Estonia

Opponent

Professor Antonio Amorim
Faculty of Sciences, University of Porto
Portugal

ISBN 978-951-51-7993-7 (PRINT)
ISBN 978-951-51-7994-4 (PDF)
ISSN 2342-3161 (PRINT)
ISSN 2342-317X (PDF)
Unigrafia, Helsinki 2022

Cover layout by Anita Tienhara. Artwork design by Evelyn Guevara.
Map of South America and ancestries of the Chachapoyas (modified from Figure 13).
Content layout: Evelyn Guevara and Juan Roel

To those who guide us

"In the battle many have been lost but few of us still stand
as reminders of our fragility and strength"

CONTENTS

LIST OF ORIGINAL PUBLICATIONS

This thesis is based on the following original publications. They are referred in the text by their Roman numerals.

I. **Guevara EK**, Palo J, Guillén S, Sajantila A. MtDNA and Y-chromosomal diversity in the Chachapoya, a population from the northeast Peruvian Andes-Amazon divide. American Journal of Human Biology. 2016; 28(6): 857-867.

II. **Guevara EK**, Palo J, Översti S, King JL, Seidel M, Stoljarova M, Wendt FR, Buś MM, Guengerich A, Church WB, Guillén S, Roewer L, Budowle B, Sajantila A. Genetic assessment reveals no population substructure and divergent regional and sex-specific histories in the Chachapoyas from northeast Peru. PLoS One 2020; 15(12): e0244497.

III. **Guevara EK**, Palo J, King JL, Buś MM, Guillén S, Budowle B., Sajantila A. Autosomal STR and SNP characterization of populations from the Northeastern Peruvian Andes with the ForenSeq™ DNA Signature Prep Kit. Forensic Science International Genetics 2021; 52: 102487.

Permission to reprint publication I has been granted by the copyright holder. Publications II and III were published under the Creative Commons Attribution (Freely available).

Author contributions to the publications

Contribution	**I**	**II**	**III**
Conceptualization and study design	EKG, SG, AS, JP	EKG, AS, JP	EKG, AS, JP
Funding acquisition	EKG, AS, JP	EKG, AS, BB, LR	EKG, AS, BB
Fieldwork	EKG, SG	-	-
Laboratory work	EKG	EKG, FRW, MMB, MS, MSt	EKG, MMB
Data analysis	EKG, JP	EKG, JLK, SÖ	EKG, JLK
Original manuscript writing	EKG, AS, JP	EKG, AS, JP	EKG, AS, JP
Manuscript edition	all authors	all authors	all authors
Supervision of the study	AS, JP	AS, JP	AS, JP

AG: Anna Guengerich, AS: Antti Sajantila, BB: Bruce Budowle, EKG: Evelyn K. Guevara, FRW: Frank R. Wendt, JLK: Jonathan L. King, JP: Jukka Palo, LR: Lutz Roewer, MMB: Magdalena M. Buś, MS: Maria Seidel, MSt: Monika Stoljarova, SG: Sonia Guillén, Warren B. Church, SÖ: Sanni Översti.

ABBREVIATIONS

aDNA	Ancient DNA
AMOVA	Analysis of molecular variance
AIM	Ancestry informative marker
aSTR	Autosomal STR
BCE	Before Common Era
BSP	Bayesian skyline plot
CE	Common Era
CODIS	Combined DNA Index System
EDNAP	European DNA Profiling Group
EMPOP	European DNA Profiling Group Mitochondrial DNA Population Database
ENFSI	European Network of Forensic Science Institutes
DNA	Deoxyribonucleic acid
FBI	Federal Bureau of Investigation
HVR-I	Hypervariable region 1
HVR-II	Hypervariable region 2
HWE	Hardy-Weinberg equilibrium
ISFG	International Society for Forensic Genetics
iiSNP	Identity informative SNP
kya	Thousands of years ago
LE	Linkage equilibrium
LD	Linkage disequilibrium
LGM	Last Glacial Maximum
LH	Late Horizon period
LIP	Late Intermediate period
MSY	Male-specific region of the Y-chromosome
masl	Meters above sea level
MCMC	Markov chain Monte Carlo
MDS	Multidimensional scaling
ML	Maximum likelihood
MP	Maximum Parsimony
MRCA	Most recent common ancestor
mtDNA	Mitochondrial DNA
NDIS	National DNA Index System
NJ	Neighbor-Joining algorithm
N_e	Effective population size
NGS	Next generation sequencing
PCA	Principal component analysis
PCoA	Principal coordinates analysis
PCR	Polymerase chain reaction

PE	Power of exclusion
PD	Power of discrimination
PIC	Polymorphism information content
RMP	Random match probability
SNP	Single nucleotide polymorphism
STR	Short tandem repeat
STRidER	STRs for Identity ENFSI Reference Database
STRSeq	The STR Sequencing Project
TPI	Typical paternity index
UPGMA	Unweighted pair group method with arithmetic mean
WTC	World Trade Center
YHRD	Y Chromosome Haplotype Reference Database
yrBP	Years before present

ABSTRACT

This dissertation focusses on the genetic diversity of the Chachapoyas in the context of northeast Peruvian and South American populations. The genetic survey of indigenous human populations from the Americas is still far from being comprehensive, which is in stark contrast with other areas of the globe, such as Europe, where an in-depth understanding of the genetic variation and its use in forensic and medical genetics has been attained. This certainly highlights the need for increasing genetic research in Peru as to continue laying the foundation for human genetics and its applications in other fields such as forensics.

Within South America, Peru is a vast and diverse country with a wide array of ecosystems. Human populations have ingeniously adapted to these environments through millennia, showing also great cultural and linguistic diversity. The montane forests of northeast Peru are one of the most biologically diverse ecosystems in the world and it is there where the ancient Chachapoyas throve and where their descendants nowadays live. Archaeological, historical and linguistic records inform of a complex region where local populations interacted with others from nearby and distant locations throughout prehistory up until the time of European contact in the 16th century. Although traces of past cultural diversity persist in the Chachapoyas region, the series of demographic, environmental and cultural impacts have blurred this cultural differentiation and debilitated the local identities of present-day peoples.

Among other factors, cultural diversity in Chachapoyas has also been shaped by a series of population-level phenomena, which are the focus of this research. With this work we have investigated the origins of the Chachapoya people and attempted to pinpoint whether it was in the Andean area or in the lush Amazonian rainforests. By doing so, the nature of genetic affinities of the Chachapoyas with neighboring and geographically distant populations from Peru and the Americas was explored. Furthermore, we tested whether cultural and ethnic diversity within Chachapoyas signified also genetic substructure. Additionally, although different patterns of asymmetrical male mediated European gene flow exist throughout the Americas, an important aim of this study was to uncover the underlying distribution of the European component in the Chachapoyas region. Linked to this aim, it was also assessed if greater European ancestry would mirror the patterns of more intense Inca presence in the region. Lastly, the focus was on detecting genetically, for lineage markers, a depopulation event occurred at the time of European contact which has been largely documented by ethnohistorical accounts and assess whether this event was more severe on males than on females.

We approached these diverse topics with DNA data. Samples from the modern Chachapoyas, the Huancas enclave from central highlands in the

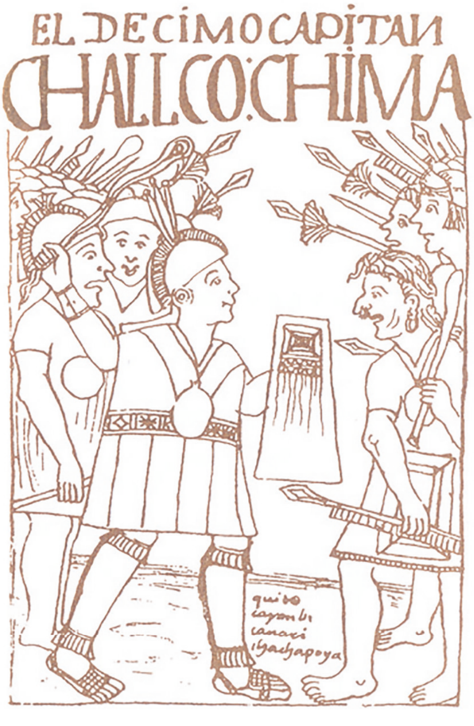
Chachapoyas territory, the Amazonian Jívaro and the Andean Cajamarca were genotyped for various types of markers. Mitochondrial DNA (mtDNA), Y-chromosome and autosomal data from study and reference populations were analyzed to uncover patterns of population structure and gain insights into the demographic history of the studied populations, with particular emphasis in the Chachapoyas. These new data show exceptionally high levels of genetic diversity in the Chachapoyas region and contrasting patterns of genetic affinities for lineage markers. In clear contrast with the archaeological and ethnohistorical evidence, there was negligible population substructure within Chachapoyas. Asymmetrical European gene flow was detected throughout the region but the pattern does not mirror Inca settlement densities in the region. Signatures of demographic collapse likely associated with the Spanish conquest were observed in both mtDNA and Y-chromosome, although for males the reduction appeared more drastic, echoing the ethnohistorical evidence. In addition to these insights, the new autosomal STR and SNP datasets generated unveiled another level of variation by providing, in addition to traditional length-based genotypes, sequence-based information for several Peruvian populations which may be of utility for individual identification in the context of South America.

The overall goal of this dissertation was to fill in the gaps in our understanding of the genetic variation and the impact of various demographic processes in the gene pool of northeastern Peruvian populations not solely for population genetic purposes but also to apply this new knowledge and data on societal issues such as forensics. Though it may not seem apparent, the local populations and their culture did not vanish after the profound demographic events occurred at the time of trans-Atlantic contact but instead transformed as the legacy of the Chachapoyas lives on in the genetic imprint of today's inhabitants.

INTRODUCTION

The history and prehistory of the Chachapoyas region has always been surrounded by a halo of mysticism. This is particularly because of its remote location in the cloud montane forests of northeast Peru and in part due to our incomplete knowledge of the ancient societies, which were often thought to have mysteriously vanished while still leaving abundant evidence of their existence. One wonders if perhaps these features even motivated the introduction of the Chachapoyas civilization in the pop-culture of the ‘80s, as the opening scene of the movie “Raiders of the Lost Ark” narrates how Indiana Jones explores a fictitious 2000-year-old “Temple of the Chachapoyan Warriors” (Lucas, 1979).

The mystery about the Chachapoyans themselves and their origins have also drawn considerable attention in other non-academic forums. Recurrent descriptions by early chroniclers of the Chachapoya people as “white Indians”, coupled with the phenotypic traits of present day *gringuitos* (Caucasian-like individuals) across the territory fueled hypotheses of their origins in Old-World sea-farers such as Celts, Carthaginians or Phoenicians (Giffhorn, 2013; Savoy, 1996). In academia however, since 1990’s, a growing body of archaeological, bioarchaeological and ethnohistorical research has greatly contributed to



Depiction of a battle between Chachapoyas and Incas.
Guaman Poma, 1615.

address questions on various aspects of Native American Chachapoya societies and the biological traits of the local populations. However, diving deep into the genetic make-up of present-day and ancient Chachapoyas have just began during the past decade.

The study of the genetic variation and its partition in worldwide, continental and regional scales is fundamental not only to evaluate hypotheses of historical interest but serves as a foundation for subsequent research in e.g. medical and forensic genetics. Nowadays, there is a wide repertoire of genetic markers, marker systems (mtDNA, Y-chromosome, autosomal) and computational methods available for analyzing these data and make inferences on population structure, genetic affinities and demographic history. In this dissertation, population genetic methods and newly produced data from the modern Chachapoyas and their neighbors were utilized to gain insights on the population structure and demographic history of this interesting region. At the same time, this evidence provides genotypic data that can be useful for forensic genetics in Peru and South America.

REVIEW OF THE LITERATURE

The purpose of this dissertation was to gain insights on the population structure and demographic history of the modern Chachapoyas by studying various types of genetic markers. This has allowed the evaluation of long-standing questions emerged from other disciplines such as archaeology and ethnohistory, proving the great utility of genetic data coupled with population genetic inferences for reaching a better understanding of the genetic variation and population history in this understudied region of South America.

Specifically, the first sub-chapter deals with how different lines of evidence in the field of anthropology have been used to approach various topics about the ancient Chachapoyas, including an examination of social and cultural characterizations as well as interactions, while providing insights from morphological and chemical studies on ancient local populations from the area of study. The sub-chapter ends with a brief recount of relevant themes often researched in the Americas from the genetics perspective and finally various questions that will be examined throughout this thesis are outlined. Sub-chapters three to five introduce basic concepts in genetics, a glimpse to population genetics theory, different types of genetic data and frameworks used in this dissertation to make inferences about the population structure and demographic history of the Chachapoyas and their neighbors. The last subchapter focuses on the application of genetic data in forensic genetics, the common markers and methodology used in the field nowadays in Peru as well as a short introduction to one of the largest cases where DNA has been used for the identification of victims in the Peruvian context. The chapter finalizes with a brief description of parameters used in forensic genetics and the power of evidence.

It is important to note that although this thesis revolves around the Chachapoyas population history, it also incorporates new genetic data from neighboring populations as a means to provide a more comprehensive picture of the regional population dynamics which is why the cultural accounts assume a broader perspective also in time and space. In addition to these, reference genetic data collected from published sources has also been included.

1. Cultural characterization of the Chachapoyas

Here the objective is to situate the Chachapoya people not only geographically but also temporally. It is vital to understand the cultural features that characterized the ancient societies that anteceded the modern inhabitants of the region, the way in which these ancient populations interacted among themselves and with their environment, and how certain events impacted biologically the local populations. All this background proved essential for defining the margins of the project and became a rich source of information for the identification of questions that can be approached with genetic data. This review, however, does not provide an exhaustive assessment of methodological and theoretical approaches in disciplines such as archaeology or ethnohistory, but instead offers a summary of the current knowledge on the Chachapoyas emanating from these major lines of evidence.

1.1 Defining Chachapoyas

1.1.1 What does the label Chachapoyas mean?

Chachapoyas is the name of a city, the capital of the modern administrative northeast Peruvian region of Amazonas. However, Chachapoyas has often been used to designate a multicultural and/or multiethnic entity that developed in the forested slopes of the northeastern Peruvian Andes before the Incas conquered the region ~1475 CE. Chachapoyas is not recognized as a socially or politically integrated region but instead researchers suggest that the label and the concept of a unified culture emerged later on, as a result of historical events during Inca and colonial times (Church & Guengerich, 2017).

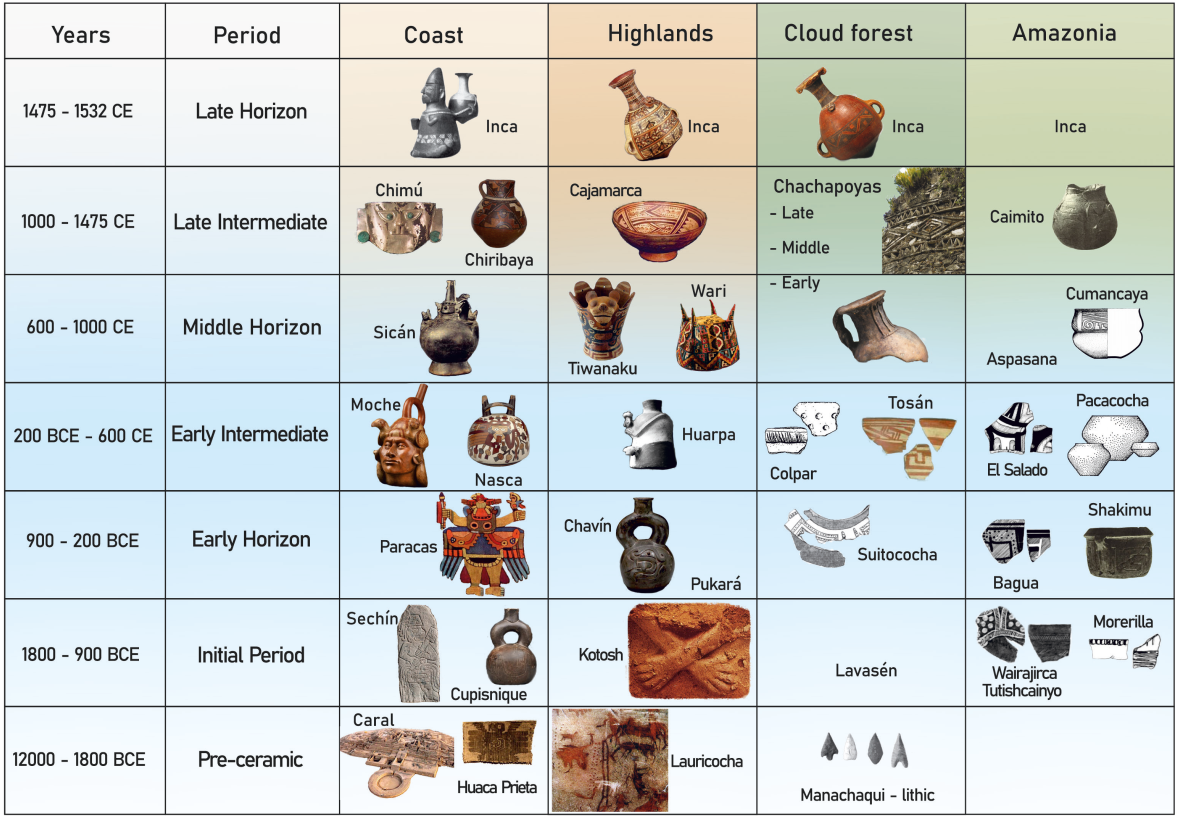
The consensus among scholars is that Chachapoya societies flourished during the Late Intermediate Period (LIP) between 1000-1475 CE, but demographic and sociopolitical transformations for this to happen started to occur much earlier (Church & von Hagen, 2008). Although there are a few common cultural elements (e.g. cliff tombs) across the Chachapoyas region, recent archaeological research has begun to unveil marked sub-regional differences (see paragraph 1.3.1 below). Indeed, the body of archaeological and historical information from the area indicates that far from being a static and unified cultural region, Chachapoyas has been much more complex and dynamic in terms of material culture, ethnic affiliations and origins.

1.1.2 The setting: geography and climate

Humans, like other species, have constantly adapted to their changing surroundings. As such, the environment certainly plays an important role not only on a population settlement patterns, subsistence strategies or worldviews but also on the biological features of the individuals. Vestiges of the ancient societies that existed in the Chachapoyas region are mostly scattered across a vast, ecologically diverse area that lies on the eastern slopes of the north Peruvian Andes. The region is part of the eastern montane forest zone (at elevations of about 1500-3500 meters above sea level, masl) that in Peru stretches for ~1500 km from north to south and ~50-250 km from east to west (Young, 1992). The montane forest further divides in lower and upper sections, at about 1500-2500 and 2500-3500 masl, respectively. Mean annual temperatures in the lower montane forests range from 17-21 ºC while in the upper montane forests fluctuate between 11-17 ºC. Both areas experience high levels of precipitation throughout the year with much of it occurring from November to April. The fringes of these montane forests are in contact with upper subalpine grasslands (> 3500 masl), or lower premontane forests (~500-1500 masl) which continue towards tropical rainforest environments in the lowlands (Young, 1992). While the montane forests extend along the entirety of the Andean Mountain range, a widening of the forested area occurs between northeast Peru and southeast Ecuador as a result of the drop of the Andes. Globally, these montane forests are recognized as one of the most biologically diverse ecosystems exhibiting high levels of species endemism (Ledo et al., 2012). The upper montane forests, labeled also montane cloud forests, due to the persistent fog originated in the eastern lowlands, have also a distinct configuration in the Chachapoyas region because the Marañón River separates both western and northern montane forests. It is due to this cloudy weather that the ancient Chachapoyas have often been referred as “Warriors of the clouds”.

1.2 Clues on the origins of the Chachapoyas

Similar to other regions of the Andes, the societies that existed in this area during the more recent LIP and Late Horizon (LH, 1475-1532 CE) periods are the culmination of a series of local cultural developments dating back possibly even to the initial peopling of South America (Figure 1). However, discussions about the origin(s) of the Chachapoya people have been mostly centered around the time immediately before the Incas conquered the region. Placing emphasis on large migrations and/or population replacements, scholars have postulated both tropical rainforest (Horkheimer, 1959; Isbell, 1968; Koschmieder, 2014b, 2014a; Lerche, 1995) and highland origins (Kauffmann, 2013; Kauffmann & Ligabue, 2003), but currently, the most widely accepted hypothesis suggests a local development (Church, 1996; Church & von Hagen, 2008). The autochthonous origins of the societies that flourished in the pre-montane and montane forests are based on growing archaeological evidence that dates back to several millennia. The earliest signs of human activity are stone tools recovered from Manachaqui cave in the southern part of the region and associated with organic material dated to 12200 and 11900 years before present (yr BP) (Church & von Hagen, 2008). Church and Von Hagen indicate that these lithic assemblages share some features with other technologies from distant territories in the north Peruvian coast (Paiján) and in the Ecuadorian highlands (El Inga). They infer that this southeastern lithic industry is an extension of a northern Andean complex, and that the stylistic variability found in the archaeological context points to several transient populations occupying the cave (Church & von Hagen, 2008). This finding is of particular interest because it has been proposed that the initial



Years	Period	Coast	Highlands	Cloud forest	Amazonia
1475 - 1532 CE	Late Horizon	Inca	Inca	Inca	Inca
1000 - 1475 CE	Late Intermediate	Chimú Chiribaya	Cajamarca	Chachapoyas - Late - Middle	Caimito
600 - 1000 CE	Middle Horizon	Sicán	Wari Tiwanaku	- Early	Cumancaya Aspasana
200 BCE - 600 CE	Early Intermediate	Moche Nasca	Huarpa	Tosán Colpar	Pacacocha El Salado
900 - 200 BCE	Early Horizon	Paracas	Chavín Pukará	Suitocucha	Shakimu Bagua
1800 - 900 BCE	Initial Period	Sechín Cupisnique	Kotosh	Lavasén	Morerilla Wairajirca Tutishcainyo
12000 - 1800 BCE	Pre-ceramic	Caral Huaca Prieta	Lauricocha	Manachaqui - lithic	

Figure 1. Simplified relative chronology of Peru including the Chachapoyas region. Images in the chart are modified from online sources (no copyright needed) and from various publications (Church, 1996; Church & von Hagen, 2008; Kanezaki et al., 2021; Lathrap, 1970; Myers, 1976; Pérez, 2019; Roe, 1976; Shady, 1987; Vallejo, 2004).

peopling of the central Andes followed a north-south dispersal along the Andean mountain range (Hester, 1966; Lothrop, 1961; Sauer, 1944). Indeed, the geographical characteristics of this section of the Andes (see paragraph 1.1.2 above), specifically the narrowing of the northern and central Andean chains, results in a series of low altitude corridors, ideal for the movement of people in various directions. In addition to a long history of human presence in caves such as Manachaqui, paleoenvironmental records from two lakes in the Chachapoyas region, Pomacochas (2100 masl) and Laguna de los Cóndores (2860 masl), indicate pronounced environment-altering human activity as early as 3300 cal. yr BP (calibrated years BP) and 2050 cal. yr BP, respectively (Bush et al., 2015; Matthews-Bird et al., 2017). Pollen and charcoal evidence from these two lakes show alternating periods of forest clearance and forest taxa recovery. Although these two records vary, maize cultivation as well as increasing Poaceae pollen (paleoclimatic indicator of forest clearance) appear to be heightened during dry episodes, specially before 800 common era (CE) (Akesson et al., 2020; Bush et al., 2015).

In brief, the archaeological record since pre-ceramic periods (~2000 before common era, BCE) up until European contact, shows that despite some discontinuities, the societies from this region are deeply rooted and exhibit histories as complex as other cultural areas of the Andes.

1.3 Cultural heterogeneity in the Chachapoyas region

At present there is no evidence that once complex societies emerged in this vast area, a unifying culture spread across the territory. This statement certainly applies to the LIP and LH periods but still remains highly speculative for the societies before 1000 CE. Archaeological and ethnohistorical research from those two recent periods just before European contact suggest certain degree of diversity while bioanthropological work points to some level of homogeneity. Below, this body of information is broken down into three constituent parts with the aim to provide key evidence from these different lines of research.

1.3.1 Archaeology points to cultural diversity

Chachapoya settlements during the LIP often crown mountain tops or lay on hillsides, in some cases extending over several hectares and mostly containing from some dozens up to few hundreds circular stone buildings. Circular stone houses are not traits confined to the Chachapoyas region but are rather common across the Andes (Guengerich, 2014); however, the Chachapoya houses distinguish themselves from other regions in the Andes and Amazon, in structural aspects (e.g. cornices, platform bases) and decoration (i.e. stone friezes) (Church & von Hagen, 2008). Recent research on domestic architecture, has additionally revealed great variability of various architectonic attributes across the Chachapoyas region (Guengerich, 2014, 2017). Friezes have also been suggested as important elements associated with the cultural identity of the societies in this region. Although frieze styles converge in some localities, they distribute differently across the whole area, rhombus appear predominant in the northern part while volute designs are more common in the southernmost territory. On the other hand, zigzag patterns are observed throughout the whole region (Lerche, 1995; Schjellerup, 2005).

Other landmarks of LIP Chachapoya societies are cliff tombs, which are elements that have been assumed to reflect ethnic identity. These are embellished funerary structures (i.e. mausoleums and sarcophagi) typically perched on inaccessible cliffs often overseeing ancient settlements and/or water sources. Collective tombs such as mausoleums or *chullpas* are generally distributed on the right bank of the Utcubamba River and nearly absent in the northernmost part of the region. In contrast, individual burials in anthropomorphic sarcophagi or *purun machus* ("ancient man" or "old man" in Quechua), are commonly found on the left river bank and spread mostly in the north (Crandall, 2012, 2017, 2018; Kauffmann, 2013; Kauffmann & Ligabue, 2003; Koschmieder, 2014a; McCray, 2017). In addition to these forms of entombment, the ancient Chachapoya people also disposed their dead under house floors, within perimetric walls, in caves, among others (Fabre et al., 2008; Gaither et al., 2008; Koschmieder, 2012, 2014a; Narváez, 2013; Ruiz, 2008). Pottery is another type of element often utilized to define areas of cultural influence and interactions. In the case of the whole Chachapoyas region, temporal variation in ceramic assemblages undoubtedly exists (Church, 1996; Guengerich, 2014; Koschmieder, 2012, 2014a; Ruiz, 1972; Schjellerup, 2005). However, spatial differentiation during specific time frames across Chachapoyas has not yet been investigated exhaustively but instead often at site level (Church, 1991, 1996; Guengerich, 2017; Ruiz, 1972), restricting the interpretations into the broader regional scale.

The variation observed on these different forms of material culture and their assumed correspondence to ethnic identity shows that during recent periods, some degree of spatial differentiation in cultural practices existed across Chachapoyas. However, it remains to be tested whether this heterogeneity reflects practices of different political and/or social entities.

1.3.2 Multiethnicity of the Chachapoyas before the Incas

In addition to the archaeological evidence of heterogeneity across Chachapoyas, various ethnohistorical studies identify the region as the seat of several ethnic groups at the time the Incas started their expansion around the mid-15th century (Espinoza, 1967; Schjellerup, 2005, 2008; Zevallos, 1995). These ethnohistorical accounts describe various ethnic units in this territory such as Jumbilla, Pomacochas, Chillaos, Corobamba, Pacllas, Chachapoya, Chilchos, among others (three of these can be identified in Figure 2). Ethnohistorians

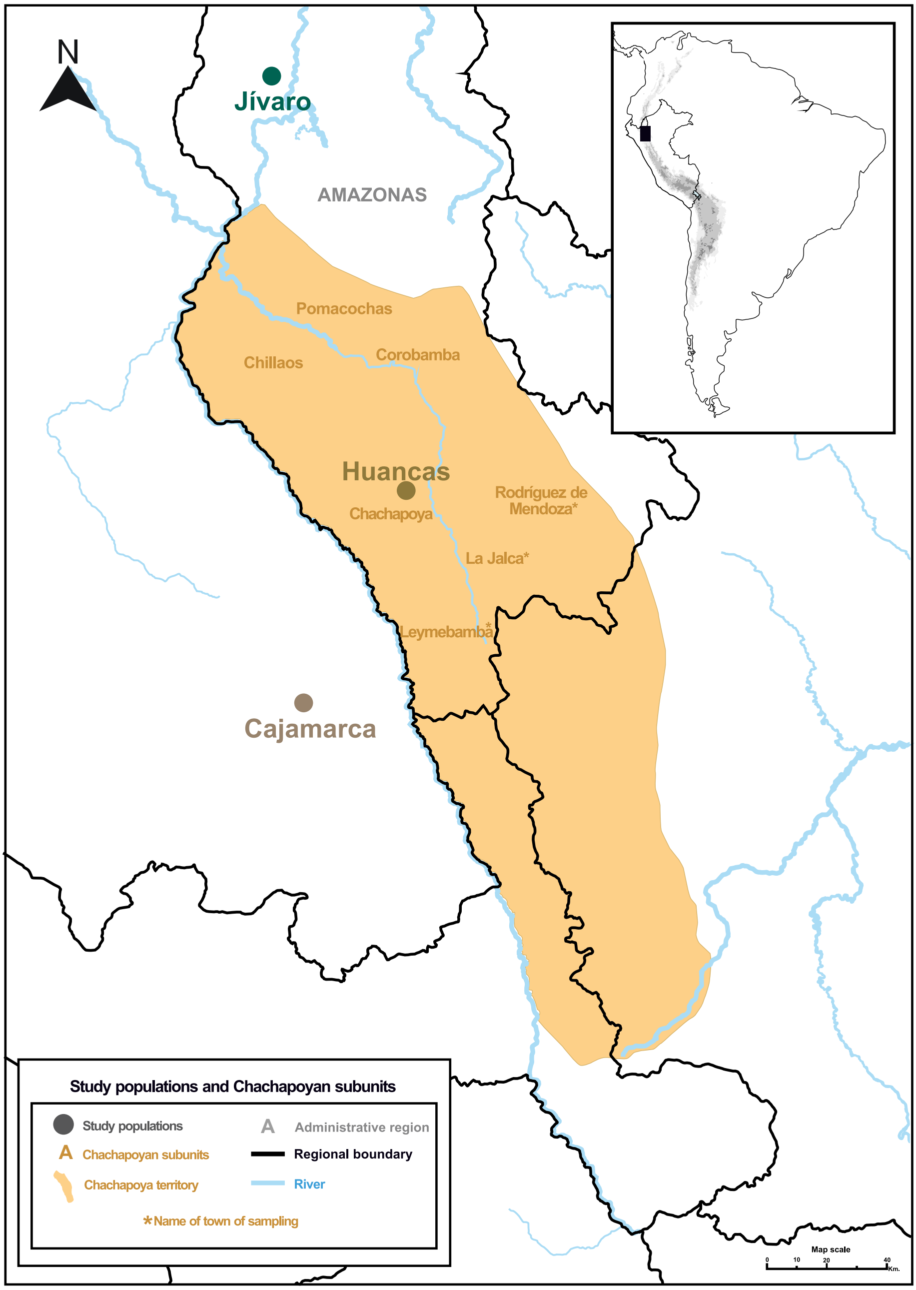


Figure 2. Map of Chachapoyan subunits and other study populations. The Chachapoya samples were collected from towns and villages across the Chachapoya territory and then grouped under subunits labels.

investigating the sociopolitical organization of LIP Chachapoya societies, portray them as autonomous and egalitarian groups that often engaged in conflict with one another but also unified when encountering a common enemy (Espinoza, 1967). The fact that there were over a dozen reported ethnic groups, not only in the study region, but also in the neighboring Chinchipe basin, in the tropical rainforests along modern Ecuador and Peru (Espinoza, 1973), undoubtedly attests for high levels of pre-Inca ethnic diversity at this latitude. The ethnic identity, however, is a synthetic category which is by nature subjective, meaning that ethnic differentiation will not necessarily translate, for instance, into patterns of restricted gene flow. Interethnic boundaries within the Chachapoyas region before the Incas remain challenging to define since ethnohistoric documents are still scant and spatial borders change through time, but we have a broad idea of their geographic distribution at least during the Late Intermediate period. The approximate boundaries and potential ethnic labels proposed by ethnohistorians, (Espinoza, 1967; Zevallos, 1995), were used in this study as a basis for the definition of population units at subregional level.

1.3.3 Contrasting bioanthropological patterns

While archaeology and ethnohistory depict the Chachapoyas region as culturally diverse, bioanthropology instead points to a homogeneous population. The anthropological examination of mummified and skeletonized human remains can provide some clues about the physical attributes of ancient individuals, the patterns of interpersonal violence, their occupation, what diseases they suffered from, and so on. Since it is possible to obtain important insights on a population's cultural identity by studying patterns of cranial modification and population structure, here a description of these two types of evidence from the Chachapoyas region is provided.

Chroniclers often described physically the ancient Chachapoyas as individuals with fair skin, different from other indigenous populations from the Americas, while also, attributing them a reputation of brave and fierce warriors (Cieza de León, 1553; Garcilaso de la Vega, 1609) but no mentions of cranial modification in this region exist in these early documents. Various pre-Columbian societies in the Andes practiced cranial vault modification which is an observable identity marker assigned at infancy. It has been proposed that it may reflect geographic origin, ethnic affiliation or status. The use of this cultural practice tends to vary across the Andes and also between cultural periods (Andrushko, 2021; Mannheim et al., 2018). For instance, the detection of diachronic changes of cranial vault modification in the central Peruvian highlands (Cuzco region), has informed of migration events during pre-Inca periods while during Inca hegemony, there was certain degree of differentiation between urban and countryside populations (Andrushko, 2021). In the southern part of the Chachapoyas region, bioanthropological studies have identified a small percentage of various types of modified crania associated to the LIP (Jakobsen et al., 1987; Schjellerup, 2005). However, this does not seem a common occurrence as it is generally absent in the rest of the region (Toyne et al., 2017). Since skeletal samples from just a handful of archaeological sites have been studied, it is necessary to examine more crania from various localities across Chachapoyas to explore whether that variability reflects local patterns, represent other influences or a combination of both.

Another relevant topic focuses on the assessment of population structure by means of biometric data analyses. Here, a study that incorporated craniometric information from LIP skeletal series at three archeological sites, two in the south (Laguna Huayabamba and Laguna de los Cóndores) and one in the north (Kuelap) uncovered negligible population structure across Chachapoyas. The study also reports that the southern sample sets experienced higher levels of extraregional gene flow compared to the northern sample (Nystrom, 2006).

In brief, the evidence shows that even though ethnic and material heterogeneity is apparent, bioanthropological patterns do not support restricted gene flow at regional level. This notion, however, can certainly still be challenged with other type of data (e.g. DNA) particularly because we cannot generalize the findings of a few localities to such a vast region.

1.4 Cultural interactions

Human populations rarely existed in complete isolation along the Andes, instead short- and long-distance connections often occurred. These interactions facilitated the exchange of staple crops, goods, ideas and certainly genes. The Chachapoyas region has long been recognized as a crossroads, facilitating access to various environmentally distinct regions, and it has been positioned as an area of economic and sociopolitical importance for other societies through time (Church & von Hagen, 2008). In this context, it is likely that once sedentary populations established here, they interacted with others from the same and/or different regions through millennia.

1.4.1 A brief archaeological perspective on pre-Inca interactions

In terms of cultural interactions, during the Initial Period (1800–900 BCE) pottery stylistic features point to contacts between the societies in the Chachapoyas region and Early Huacaloma styles in highland Cajamarca, Morerilla in lowland Bagua and long-distance connections with Ecuador. By the Early Intermediate period (200 BCE-600 CE), imported Cajamarca wares again seem to indicate connections with this region while stone-carving influences from Recuay and Conchucos from the central highlands have also been suggested. During the Middle Horizon (600-1000 CE), Cajamarca ceramic styles have been recovered in various sites across Chachapoyas (Church & von Hagen, 2008). On the other hand, Wari influence from the central Peruvian highlands appears to be limited (Church & Muscutt, 2018; Narváez, 2013). Later on, the LIP marks a time of diversification and flourishing of local material culture (see paragraph 1.3.1 above) when alleged connections with Loja (Ecuador) and the upper Chinchipe basin (Guffroy, 2006) have also been suggested. Figure 1 shows details on periodization of pre-Columbian Peru.

The archaeological record shows that although local populations produced distinctive cultural patterns, they were interconnected with other cultural areas such as the Peruvian northern highlands (particularly the Cajamarca region) and coast, the tropical lowlands and even more distant areas such as Ecuador. The role of the societies in the Chachapoyas region generally shifted throughout prehistory. At times they became well integrated, for instance, in Andean interaction spheres while at others they were part of more local networks. The scarce material associated to pan-Andean developments such as Chavín (Early Horizon: ~900–200 BCE) and Wari (Middle Horizon: 600–1000 CE) is still puzzling, although this may just imply the need for more controlled archaeological excavations across Chachapoyas.

1.4.2 Chachapoyas under Inca control

During the LH period, imported material from Cajamarca (highlands), Chimú (coast) and from societies in lowland Amazonia (Church & von Hagen, 2008; Guillén, 1998; Guillén et al., 2011; Narváez, 2013) converge in the Chachapoyas region. Architectural canons, pottery and textiles all show notable influence from the Incas as this territory was gradually annexed to the Inca domain. In some parts across Chachapoyas, the Inca expansive efforts encountered rebellions especially where diplomatic means proved unfruitful. Settlements with Inca architecture in key locations with the aim to appease uprisings and guarantee a more efficient control, including the *Qhapaq Ñan* (Royal Road in Quechua), were installed across the whole region. The occupation was intense although it seems to have been more accentuated on the east side of the Utcubamba River and generally in southern locations (Crandall, 2017, 2018; Rojas & VanValkenburgh, 2020; Schjellerup, 2005). In addition to these changes, mortuary behavior also transformed during Inca times as artificial mummification procedures including evisceration and embalming were introduced (Guillén, 2002, 2003).

Scholars suggest that the label Chachapoyas, often used to designate the peoples living in this region, originated probably during this period (Church & Guengerich, 2017). In order to facilitate the administration and control of this vast territory, the Incas created the administrative province of Chachapoyas, which was far from representing a homogeneous ethnic group. As previously noted, the motivation for controlling the Chachapoyas region probably rested in the key position of Chachapoya societies in broader highland-lowland exchange systems which would facilitate access not only to a wide array of goods from different environments but also to paraphernalia from Amazonia associated to powerful shamans (Church & von Hagen, 2008). Prestige goods from northern coastal latitudes, such as Spondylus, attest also for long-distance exchange networks in the area at this time (Church & von Hagen, 2008; Narváez, 2013). Despite this intense Inca presence, isotopic evidence from two successive archaeological periods, pre-Inca LIP and Inca (LH), suggests that subsistence strategies were not profoundly impacted by the arrival of the Incas (Toyne et al., 2017, 2020).

1.4.3 Chachapoya language traces and alleged connections

Pre-Inca local language remnants persist in toponyms, surnames and certain elements in the spoken Spanish from the region, which have led linguists to suggest an autochthonous "Chachapoya language" (Taylor, 2000, 1990; Valqui & Ziemendorff, 2016; Zevallos, 1982) or linguistic area (Torero, 2005). It is not yet well understood when exactly the language(s) from such a vast region started to become discontinued. Certainly, with the Incas, there must have been an intention to not only disintegrate the local populations but also introduce the dominant language and religion as a mechanism of control. Although it is not clear whether Quechua was introduced by the Incas or much earlier by other societies, linguists suggest both languages coexisted (Valqui & Ziemendorff, 2016), possibly in a context of diglossia (Rojas-Berscia, 2020). Spaniards encountered predominantly a Quechua variety which had already influenced the local toponomy (Valqui & Ziemendorff, 2016) as it would have been commonly used not only by the Inca administrators but also reinforced to the scant local population left after the forced large-scale *mitmaq* (displaced population in Quechua) migrations promoted by the Inca expansive state (Rojas-Berscia, 2020). In a more recent study, comparisons of lexical elements from the Chachapoyas region with modern Kawapanan languages from the Amazon showed a shared substrate that could not yet be readily associated to a common genealogical origin nor to intense language contacts (Rojas-Berscia, 2020). It is speculated that carriers of a paleolanguage once spoken across Chachapoyas, moved eastwards to an area where Kawapanan languages are currently distributed (Rojas-Berscia, 2020), but this idea needs to be tested once more linguistic data becomes available.

Overall, this section made evident the complex nature of cultural interactions of past societies throughout prehistory in the Chachapoyas region while providing interesting hypotheses to test such as the alleged connection with societies from the coast, highlands and Amazonian lowlands.

1.5 Demographic impacts of Inca and European conquests

Here, a recount of the sociopolitical transformations and population changes that occurred since the last pre-Columbian period up until colonial times is provided. This area of research is particularly important for this thesis given the fact that dramatic changes in population size and composition have not been documented after European contact in the Chachapoyas region. This would then entail some degree of population continuity in this territory.

The Inca and early colonial periods witnessed not only changes in cultural patterns but also transformations of the geopolitical units under the Inca and Spanish control (Church & Guengerich, 2017; Church & von Hagen, 2008; Schjellerup, 2005). It has been argued that large population sizes characterized the Chachapoyas region before the Incas (Church & von Hagen, 2008). However, settlement density alone cannot provide unequivocal evidence for this since at present not many sites have been radiocarbon dated. After the Inca conquest, tax payer information suggests there were 100,000 to 200,000 individuals in the Inca province of Chachapoyas before mass deportations (*mitmaq* policy) occurred (Church & Guengerich, 2017; Church & von Hagen, 2008). *Mitmaq* from the Chachapoyas territory were transferred to over 20 locations across the *Tawantinsuyo* (meaning “the four regions” in Quechua), the territory under Inca control, and in locations as far as Ecuador and Bolivia (Schjellerup, 2005, 2017). The aim of applying this method of relocation was not only because of administrational interests but also served economic and military purposes (Espinoza, 1967, 2006; Schjellerup, 2005). A rough estimate of all *mitmaq* displaced from the Chachapoyas region would amount to about 5,000 individuals (Espinoza, 1988; Schjellerup, 2017), though other ethnohistoric accounts suggest relocation policies resulted in displacement of more than 50% of the total local population (Lerche, 1995). On the other hand, *mitmaq* from other regions were introduced across Chachapoyas but the influx of people from other Peruvian regions seems to have been less significant. Two of these relocated groups are the Huancas *mitmaq* from central highlands and also *mitmaq* from the neighboring region of Cajamarca (Espinoza, 1967; Schjellerup, 2005). With the arrival of the Spaniards, local populations and settlements were once again restructured by means of reductions (*reducciones*), where the forced agglutination of sparsely distributed villages facilitated their control by the new politic and religious order (Mumford, 2017). Since the early colonial period, males from this and other regions were particularly conscripted into the military for war and exploration ventures in search of El Dorado (Espinoza, 1967; Schjellerup, 2005), which may also have impacted the male/female ratio. In addition to war, successive epidemics brought by Europeans also decimated the local populations, which, in the case of northeastern Peru, are portrayed as very dramatic (Cook, 1981; Schjellerup, 2005).

All in all, the Chachapoyas region seems to have been densely populated by various ethnic groups before the Inca expansion. The local populations experienced transformations during Inca times due to war and *mitmaq* policies. Later on, Spanish colonial rule caused also profound changes in settlement patterns, tax systems while actively recruiting males for warfare and expeditions into the tropical rainforests. Newly introduced pathogens in the form of various epidemics during this period also caused dramatic population declines which left the region sparsely occupied. Since these events most likely affected the genetic diversity of past and contemporary populations from the Chachapoyas region, this dissertation approaches particularly this theme.

1.6 Relevant topics: application of genetic data

Genetic markers too can offer a unique and interesting view into the past portrayed by other disciplines. Indeed, understanding the patterns of genetic variation, population structure and the demographic history are not only important for research on human evolutionary genetics but also guides e.g. the design of applied research. To illustrate how hypotheses of interest are addressed with genetic data, this subchapter aims to place this study in the broader context of population genetics in Peru and the Americas.

1.6.1 The peopling of the Americas

Although researchers agree that the ancestors of modern Native Americans came from northeast Asia through Beringia during the Last Glacial Maximum (24 -13 kya), the number of migrations and their timing, as well as the routes of dispersal in the Americas have remained highly contested topics. Early work on mtDNA hypothesized that different founding lineages (A, B, C, D) from the Americas represented multiple waves of migration, (e.g. Horai et al., 1993; Torroni et al., 1992, 1993), but increasing evidence from modern populations (e.g. Bonatto et al., 1997; Fagundes et al., 2008; Silva et al., 2002; Tamm et al., 2007) and ancient samples (Llamas et al., 2016) points to a single origin. Similarly, despite genome-wide studies have been able to detect subsequent streams of more recent gene flow (Reich et al., 2012; Skoglund et al., 2015), the current consensus is of an initial single migration (Posth et al., 2018; Raghavan et al., 2015). It is also widely accepted that an eastern Beringian standstill that allowed the accumulation of variation common to Native Americans existed prior to the colonization of north and South America (Fagundes et al., 2008; Kitchen et al., 2008; Mulligan et al., 2008; Pinotti et al., 2019; Tamm et al., 2007). For different kind of markers, a better temporal resolution estimate of the period of isolation in Beringia suggests a shorter time span from ~ 2.3-9 kya (Llamas et al., 2016; Pinotti et al., 2019; Raghavan et al., 2015), in contrast to ~15kya previously assumed (Kitchen et al., 2008; Mulligan et al., 2008; Tamm et al., 2007). This was followed most likely by a swift dispersal through the Pacific coast (Fagundes et al., 2008; Llamas et al., 2016; Pinotti et al., 2019; Tamm et al., 2007; Wang et al., 2007) since the inland corridor between the Laurentide and Cordilleran ice sheets had not opened yet between ~ 30,000-13,000 years (Mandryk et al., 2001), which may explain human presence in the southern cone of the continent as early as 14,000 yr BP (Dillehay, 2009). There is also a better consensus among researchers regarding the rapid dispersal into the Americas, occurring not earlier than 16-15 kya (Llamas et al., 2016; Mulligan et al., 2008; Pinotti et al., 2019).

1.6.2 Population substructure within South America and Peru

Different kind of genetic markers suggest that the rapid initial settlement of South America produced a marked geographic structure as early as 12 kya, with little gene flow between subregions afterwards (Llamas et al., 2016; Pinotti et al., 2019; Reich et al., 2012). Earlier modern DNA studies proposed divergent genetic patterns shaped by different environmental and cultural factors in east and west South America (Fuselli et al., 2003; Tarazona-Santos et al., 2001; Wang et al., 2007), although recently genome-wide ancient data has uncovered more complex scenarios in western South America (Nakatsuka et al., 2020). In agreement with previous inferences, this study indicates that initial population structure in the Andes (coast/Andes-north/Andes south) established most likely during the early Holocene and continued several millennia until about 5.8 kya, when north-south gene flow is observed. Later on, starting at about 2 kya, there are signatures of genetic continuity in most Andean subregions despite the sociopolitical transformations of expansive states such as Wari (Nakatsuka et al., 2020).

The peopling of Peru and the genetic differentiation of humans inhabiting broad ecogeographic domains -Coast, Andes and Amazon- have also been estimated to have happened quite early, from about 12.9 to 11.7 kya (Harris et al., 2018), although the colonization routes are still nebulous. Gene flow events across these macroregions appear to have been more frequent in the northern Peruvian Andes (Barbieri et al., 2019; Borda et al., 2020) than on southern Peru but these phenomena seem to reflect more recent interactions. This is because traces of the early genetic differentiation between the coast and the highlands before 6 kya (Nakatsuka et al., 2020) are not observed in high resolution studies using modern data (Barbieri et al., 2019; Gnecchi-Ruscone et al., 2019; Harris et al., 2018). Notably, when comparing modern and ancient diversity in the Andes, a high rate of lineage extinction has been detected after European contact in the case of mtDNA (Llamas et al., 2016). This highlights the importance of including ancient material from different time points to have a more comprehensive view of the diachronic changes in population dynamics, which may be obscured in the present-day genetic diversity.

2. Emerging questions

The archaeological, bioarchaeological and ethnohistorical background contributed significantly to define the questions that can be evaluated with genetic data. One of these deals with the origin of the Chachapoya peoples during the LIP. Some scholars argue that the ancestors of the Chachapoyas came from the highlands or from the jungle at this time but it is likely that sedentary populations existed much earlier in the area since human presence has been detected at least by 2,000 yr BP (e.g. charcoal, pollen) and traces of transient groups are evident at about 11 kya (lithic industries evidence). Even though we cannot directly test this without ancient DNA (aDNA) from first settlers, can the genetic variation observed in our modern datasets give us some clues about affinities with coastal, Andean or Amazonian gene pools? What are the affinities of the Chachapoyas and other populations from Peru and the Americas?

Interestingly, archaeological and ethnohistorical studies suggest archaeological and ethnic heterogeneity across the Chachapoyas region while bioarcheological research informs of certain level of homogeneity. Can our current genetic data reveal pre-Columbian population substructure patterns within the Chachapoyas region?

Although the Chachapoyas region experienced transformations in settlement patterns and in treatment of the dead during the Inca conquest, it does not seem there was significant impact in the local subsistence strategies nor in the introduction of foreigners into this area. Does the genetic evidence support the notion of limited extra-regional gene flow?

Gene flow between populations throughout prehistory likely occurred in this region since there is evidence of short and long-distance exchange networks involving societies from the Cajamarca highlands, the north Peruvian coast and Ecuador. Would these phenomena have left traces in the genetic variation of the populations from the modern Chachapoyas and Cajamarca regions? Or between modern Chachapoyas and present-day Amazonian Jívaro populations?

Considering the different intensity of Inca occupation across the Chachapoyas region (i.e. more pronounced on the east side of the Utcubamba River), did the Spanish conquistadors follow a similar pattern of occupation? This would mean, for instance, more European contribution just in certain subpopulation units across Chachapoyas such as Leymebamba.

Finally, regarding the demographic events in the region under the Inca and Spanish administration, which caused significant depopulation, is it possible to detect these signatures in the uniparental markers? And, since Chachapoyan males were more frequently recruited for Inca and Spanish militias, would the male population size contraction be more dramatic?

3. Genetics: a view into the past and present

Genetic evidence can offer some answers to the questions posed above. Past phenomena have shaped our genomes and gene pools through millennia, which is why an examination of the genetic diversity can inform us about population origins, population structure, demographic fluctuations in the past, migrations as well as serve to its application in forensic genetics. Population genetics is a fundamental framework used to study the genetic composition of natural populations and how this variation changes over time. Its approach involves the application of Mendel's laws and other genetic principles to understand the origin, maintenance and evolution of genetic variation in a given population or species (Hartl & Clark, 1989). A brief review of basic concepts necessary for analyzing and interpreting genetic data as well as for making inferences is provided here.

3.1 The basics

3.1.1 Inheritance, alleles and genotypes

Humans are diploid organisms carrying two sets of genomes and thus two alleles (alternative forms of a gene) at each locus (position in the genome), one allele from the mother and one allele from the father. Most of the genetic material resides in the 22 autosomal chromosomes in the nucleus, in addition to which there are sex chromosomes X and Y. Mendel's law of segregation determines the probability of each of the alleles to be passed to the next generation. However, sex chromosomes and the non-nuclear mitochondrial DNA (mtDNA) exhibit a different transmission pattern where they are passed to the next generation mostly without recombination (Figure 3) (Bachtrog, 2013; Schaffner, 2004). Broadly speaking, a genotype is the complete set of genes of an organism while in a more restricted sense, it describes the combination of the variant forms of a gene carried by an organism. At a particular single

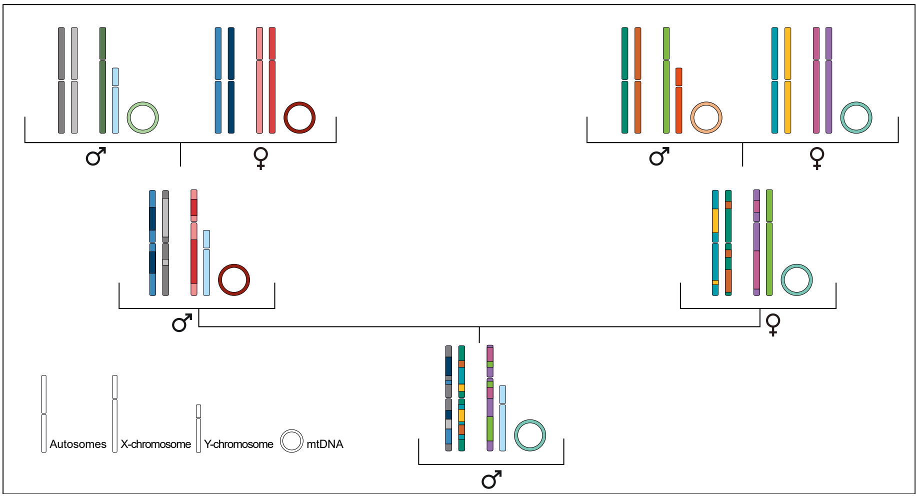


Figure 3. Inheritance patterns for different kind of DNA markers. Note the mode of transmission of uniparental markers through maternal (X-chromosome, mtDNA) and paternal (Y-chromosome) lines. Autosomal markers in turn recombine each subsequent generation and are inherited from both parents.

nucleotide polymorphism (SNP) and gene(s), individuals can be homozygous (having two identical alleles) or heterozygous (having two different alleles).

3.1.2 Population

In biological terms, a population is defined as a group of interbreeding individuals cohabiting in a given time and space. For humans, the definition of a population is rather arbitrary but it heavily relies in geography and also other criteria such as ethnicity or language as often these tend to produce restricted gene flow (Relethford, 2012). The Americas was the last landmass to be colonized by modern humans whose diversity was reduced when compared to other areas of the globe. Despite this, genetic variation has accumulated over time and various demographic processes have shaped the differentiation of extinct and extant populations in the continent. The characterization of a population's gene pool is vital as the study of the variation within and across populations would concomitantly inform about their demographic history. In this study, the definition of population units has been highly influenced not only by geography but also by cultural determinants such as language as well as population prehistory and history.

3.2 Evolutionary forces shaping the genetic diversity

Changes in the frequencies of alleles in a population are the result of one or more evolutionary forces in action. What is particularly important in this respect is that these forces exert a relative impact in the gene pool. Consider, for instance, cases in which highly deleterious or advantageous mutations can escape selection, against or for, if genetic drift becomes severe enough in a given population or species.

3.2.1 Mutation

Mutation is the ultimate source of genetic variation as it will produce direct changes in a genetic sequence. Mutations occur randomly irrespective of their effects in the organism. Types of mutations include single-base DNA substitutions, insertion/deletion of bases or fragments (indels) and other rearrangements of DNA sequences. The effects of these mutations can be neutral, beneficial or harmful. However, mutations are rare events that even for rapidly mutating parts of the genome, cannot alone lead to significant allele frequency change. The rate of mutation also varies across the genome and among species. For example, the mitochondrial genome has high mutation rates when compared to the nuclear DNA (Relethford, 2012). Once a mutation has arisen it can then become fixed or lost by random genetic drift or selection.

3.2.2 Genetic drift

Genetic drift is a random process by which allele frequencies fluctuate from one generation to the next by chance (sampling error) due to finite population size. Since the sampling process occurs in every population, all other evolutionary forces that cause changes in gene frequency should also be considered jointly with genetic drift. Genetic drift causes both fixation and loss of alleles in a population which leads to the reduction of its genetic diversity. The effect of genetic drift is stronger in small populations because they exhibit larger sampling variation. Two examples of drift-driven mechanisms that shape the genetic diversity in a population are bottlenecks and founder events (Jobling et al., 2004). A bottleneck is characterized by the loss of genetic diversity due to the reduction of the size of a previously larger population (Figure 4). In humans for instance, bottlenecks can be caused by a number of factors such as famine, disease or genocide. An oft-cited example in the Americas is the dramatic population decline experienced across the continent at the time of European conquest due to war and pandemics. A founder event, in turn, will produce differences in allele frequencies between a source population and a derived population that colonizes new spaces, the magnitude of which

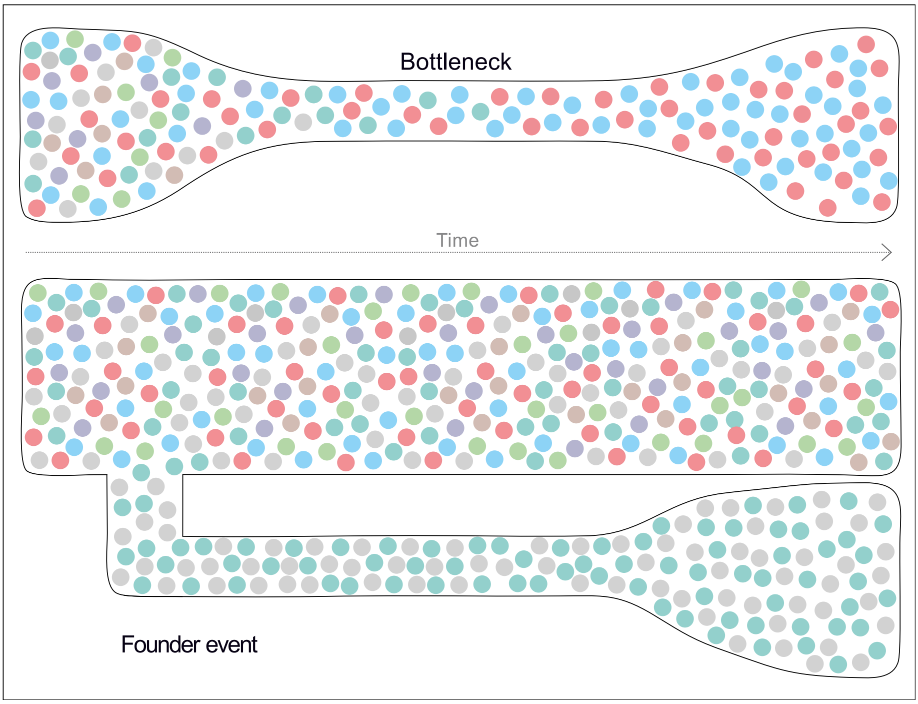


Figure 4. Schematic representation of diversity changes in populations after a bottleneck and a founder event.

depends on the size of the founding population, mainly in absolute terms, but also relative to the source population (Figure 4). One of such examples is the peopling of the Americas which introduced only part of the genetic variation from the ancestral East Asian populations into this continent. In the absence of migration between the two continents, the effects of genetic drift over time then led to the differentiation of populations.

A related core concept is effective population size (N_e), referred to as the size of an ideal population (assuming equal male/female ratio, constant population size and random mating) that shows similar rate of evolutionary change resulting from the random sampling of genetic variants (genetic drift) as the population in question (Charlesworth, 2009).

3.2.3 Natural selection

The environment certainly has an effect in the genome as neutral or harmful variants of genes can become advantageous due to changing conditions. Certain individuals are more adapted to their environment than others as it is observed by their greater fitness -the probability of survival and reproduction in a given environment. The process by which individuals with greater fitness will leave more offspring than less advantageous ones is called natural selection. Favorable alleles promote higher fitness and will be overrepresented in subsequent generations thus causing frequency changes (Hartl & Clark, 1989). Natural selection can decrease (directional selection) or increase (disruptive selection) the genetic variation. However, when genetic markers do not confer fitness advantage (or are not physically linked to non-neutral genes), they are considered to be selectively neutral.

In the case of neutral markers, diversity is shaped by genetic drift and migration, which is why these types of markers have been commonly used to make inferences about population history.

3.2.4 Migration

Gene flow occurs as a result of the migration -movement and subsequent reproduction- of individuals between populations. Unlike natural selection and genetic drift, new alleles can only be introduced in a population through gene flow and mutation. Hence, gene flow counteracts the effects of genetic drift as it will increase and/or maintain the genetic diversity in a population preventing the differentiation of populations. Various theoretical models of migration (e.g. Wright's island model, stepping-stone model) can be assumed in population genetic analyses, each of which incorporates one or a combination of factors such as geographical distance, migration rate, spatial arrangement (connectivity), among others (Relethford, 2012).

3.3 Independence within and across loci

In an ideal population, the genetic diversity remains unchanged and genotype frequencies in one generation depend only on the allele frequencies of the previous generation. These kinds of populations are said to be in Hardy-Weinberg equilibrium. However, deviations from Hardy-Weinberg equilibrium (HWE) do occur when sampling populations such as in the case of a structured population where two diverged populations have been sampled as one. Another example would be positive assortative mating, which will increase the proportion of homozygotes in the population. In humans this could be illustrated by mate choice on the basis of skin color or height. Here we must take into account that nonrandom mating leads to changes only in genotype frequencies, while the four evolutionary forces change both allele and genotype frequencies (Relethford, 2012).

Linkage disequilibrium (LD), in turn, is the non-random association of alleles at two or more loci (Slatkin, 2008). When this occurs, haplotype (see paragraph 4.1 below) frequencies exhibit departures from expectations. LD patterns result from the interaction between recombination and various evolutionary forces (i.e. mutation, natural selection and genetic drift). As such, the assessment of LD in humans has a vital role not only in gene mapping but also in population and evolutionary history (Pritchard & Przeworski, 2001). For instance, bottlenecks, population structure and inbreeding can all intensify LD. In addition to these, LD can also arise due to recent admixture between populations that harbor different allele frequencies (Pritchard & Przeworski, 2001; Slatkin, 2008).

4. Genetic diversity by means of different kind of markers

As previously described, genetic diversity is influenced by various evolutionary forces. Diversity in the neutral markers, evolving at known rates, can inform us of population structure and population history. The concept of genetic diversity is tightly linked to demography, since variation, originated by mutation and migration, can only be retained if population size is large enough. Genetic diversity within a population can be measured in different ways, for instance, by counting the number of alleles present (dependent on the sample size) and their frequencies, by assessing the level of heterozygosity, the mean number of pairwise differences (MNPD) and so on. When measuring the levels of diversity, various types of markers can be examined.

Since the early twentieth century, a series of markers have been used to determine the amount of genetic diversity in human populations worldwide. Before the use of DNA markers, these assessments were based mostly on protein markers that were characterized immunologically and/or by electrophoresis. Among others, these included polymorphic blood proteins from the ABO and

Rhesus systems (Fisher & Race, 1946; Landsteiner & Levine, 1928). Although these classical markers provided information on the genetic variation to a certain degree, the introduction of DNA markers allowed another level of resolution by examining directly the underlying sequence variation (Jobling et al., 2004).

4.1 Lineage markers

MtDNA and Y-chromosome markers are inherited without recombination. Importantly, these markers hold a record of all retained mutational events occurred throughout their lineage history which enables the reconstruction of possible phylogenies and allows to trace back the most distant common ancestor of any set of lineages. Because of the nature of inheritance of these marker systems, the mtDNA and Y-chromosome N_e are smaller (one-fourth of that obtained from autosomal markers) and thus more susceptible to the effects of genetic drift.

In the Y-chromosome and mtDNA, the combination of allelic states of polymorphic markers is called a haplotype. In order to classify these uniparental haplotypes on a worldwide scale, geneticist have come up with the concept of haplogroup. Haplotypes belonging to a specific monophyletic clade and sharing a defined set of rare mutations form one haplogroup. The nomenclature of human haplogroups consists of letters (e.g. A-Z for mtDNA) and the branches within them are called sub-haplogroups, for which alphanumeric characters are assigned in a hierarchical fashion (e.g. mitochondrial haplogroups C, C1, C1b, C1b2 and so on) (van Oven & Kayser, 2009). Conveniently, the geographic distribution of haplogroups and their sub-lineages in each continent can be used to deduce migration patterns in the distant past.

4.1.1 Mitochondrial DNA markers

The mtDNA is circular and in humans typically composed of 16,569 base pairs. Changes in the mtDNA sequence are mostly single-base substitutions and deletions, although insertions also occur. The mutation rate in the mtDNA is ten times higher than in the nuclear DNA but varies across the mitochondrial genome, with relatively low rates observed in the coding region and higher rates in the non-coding control region (Amorim et al., 2019). Hypervariable regions (HVR-I and HVR-II, called hypervariable sequences [HVSI and II] in article I) in the mitochondrial control region harbor one of the highest mutation rates in mammals and have been widely used in population genetic studies long before newer technologies (i.e. next generation sequencing: NGS) allowed a speedier genotyping of full genomes. A particular trait of mtDNA is that its high copy number per cell (~500 mtDNA molecules/human cell) increases the probability to recover its genome, in contrast to nuclear DNA where degradation and/or low DNA template turns this more difficult. Thus, the higher survival rate of mtDNA has been fundamental in specific contexts that deal with challenging samples in both aDNA and forensic genetics.

4.1.2 Y-Chromosome markers

The Y-chromosome is the smallest human chromosome comprising about 60 million base pairs (Mb). Recombining pseudoautosomal regions account for a small fraction of the Y chromosome genome (5%), but most of its length is non-recombining. This means that the male-specific region of the Y-chromosome (MSY) is largely transmitted from father to son without any changes unless mutational events occur. SNPs and microsatellites (or short tandem repeats, STRs) in the Y-chromosome are among the most widely used markers for studying human evolution and population history; however, indels (insertion/deletions) also frequently occur. Y-chromosomal haplogroups are defined by specific SNPs, but there are haplogroup-specific Y-STR patterns that can be used to predict the haplogroup in a probabilistic manner with various tools (e.g. http://www.hprg.com/hapest5/index.html). Y-SNPs exhibit low mutation rates ($\sim 2 \times 10^{-8}$ per base/generation) while Y-STRs have much higher mutability ($\sim 10^{-3}$ per locus/generation) (Balanovsky, 2017; Boattini et al., 2019). STRs are tandemly repeated DNA sequences of 2-5 base pairs, each with several alleles. Mutation rates on STRs vary depending on the repeat-motif type as well as across loci and between alleles (Brinkmann et al., 1998; Chakraborty et al., 1997). The estimated mutation rate for STRs ranges from 10^{-2} to 10^{-4} per locus/generation (Ballantyne et al., 2010; Boattini et al., 2019), allowing the study of the recent human population history and fine-scale geographic structure. The high mutation rates among STRs therefore make them ideal for forensic genetic applications (Purps et al., 2014). In phylogenetic studies instead, Y-SNPs outperform Y-STRs in the estimation of the TMRCA (time to the most recent common ancestor), since the latter lead to underestimation of the deep genealogy (Hallast et al., 2015).

4.2 Biparental markers

As previously noted, in organisms with sexual reproduction, there are two copies of each gene/locus as a result of recombination and segregation. Various types of polymorphisms can be found across the human genome, although two of these are particularly relevant to this dissertation, i.e. SNPs and STRs (Figure 5). SNPs are abundant in the genome and show low mutation rates (Nachman & Crowell, 2000), making them useful for investigating the deep human population history. STRs mainly mutate through the gain and loss of single repeat units in the genetic sequence but also e.g. 2-4 step changes and intermediate alleles (smaller than a repeat motif) can occur. Two traits have

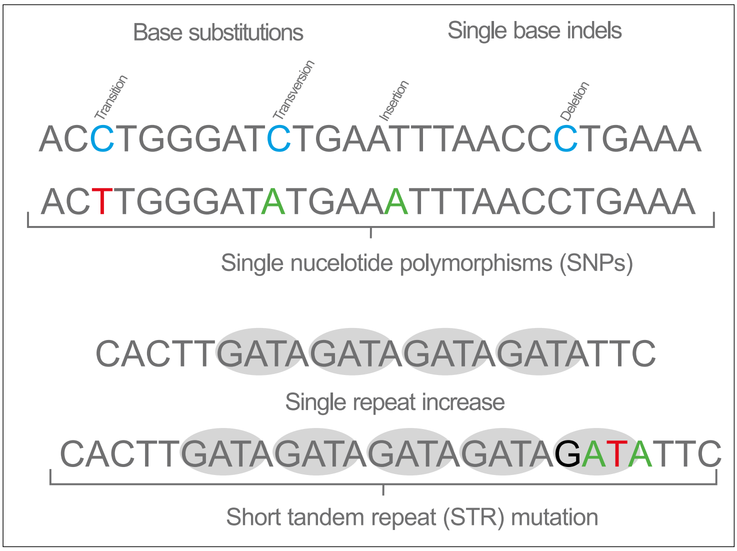


Figure 5. DNA sequence showing different types of variation. Transitions are changes between purines (A <--> G) or pyrimidines (C <--> T) while transversions are changes between purines and pyrimidines.

made STRs amenable markers for human identification purposes, their high variability among individuals and their ease of typing. As noted before, multi-allelic STR markers have also the advantage of providing more information than bi-allelic SNPs (Chakraborty et al., 1999). Overall, these two types of markers have been successfully used in different fields, ranging from medical oriented studies, to forensic and population genetics.

5. General frameworks in population genetics

The purpose of this section is to examine various concepts generally utilized in population genetics as a basis for subsequent chapters. As previously described, genetic diversity and its spatial distribution reflects the degree of connectivity between sample sets or populations which may be the result of, among others, mate choice, migrations and cultural habits. The study of this genetic variation then provides insights on the population history as it allows to examine the effects of historical events on these patterns as well as their timing (Crawford, 2007).

5.1 Genetic diversity and its distribution in space

In studies of global human diversity, it has been suggested that a good sampling strategy should be based on the individuals and in geography instead of hypothetical populations, since the distribution of genetic variation is generally clinal due to isolation by distance. On the other hand, approaches based on populations -e.g. culturally defined- should be considered when the aim is to study the history of those populations and/or their affinities with neighboring populations (Serre & Pääbo, 2004). Although differences in the distribution of human genetic diversity observed vary according to marker system (Jorde et al., 2000), various studies indicate that most of the variation resides within rather than between populations (Barbujani et al., 1997; Witherspoon et al., 2007). In fact, estimations based on STR and SNP data place only 5-10% of the human genetic diversity on differences between continents (Holsinger & Weir, 2009). Global patterns of genetic structure for humans have been characterized for all marker systems (Excoffier et al., 1992; Hammer et al., 2001; Ramachandran et al., 2004; Rosenberg et al., 2002; Wilder et al., 2004) but also the partitioning of genetic variation has been applied to smaller spatial scales (e.g. Palo et al., 2009; Tarazona-Santos et al., 2001 and numerous others).

An important framework employed to partition the genetic diversity is F-statistics (Wright, 1951), which assesses the proportion of total diversity at several hierarchical levels by using: 1) F_{IT}, the correlation between alleles within individuals relative to the entire population, 2) F_{IS}, the correlation between alleles within individuals relative to their subpopulation and 3) F_{ST}, the correlation between randomly chosen alleles from a subpopulation relative to the entire population. F_{ST} can be a value between 0 and 1, where higher values indicate larger differentiation. A similar approach applied to both haploid and diploid data is the analysis of molecular variance (AMOVA) (Excoffier et al., 1992). Relying on a similar concept as the analysis of variance (ANOVA), this method uses a hierarchical or nested model to assess the apportionment of genetic variation within and between populations. The genetic variances can be estimated either from allele frequency differences or from the molecular differences between alleles -repeat number differences for STRs and the number of substitutions for sequence polymorphisms. The variance here is also partitioned in three hierarchical components: 1) within populations, 2) among populations within groups and 3) among groups. AMOVA produces then Φ-statistics analogous to the conventional F-statistics that show the correlation of haplotypic diversity at those hierarchical levels (Excoffier et al., 1992).

Another common way to detect structure is by using model-based clustering methods such as STRUCTURE (Pritchard et al., 2000). These approaches apply certain probability models for the estimation of clusters while optimizing the fit between the data and the model. The assumption here is that the data are produced by a set of probability distributions where each component represents a different cluster. This ancestry decomposition method also has the advantage of identifying the optimal number of clusters (K). Bar plots are used to visualize the ancestry coefficients obtained for each observation and several runs for different Ks are performed. Since the optimal K value is an estimate rather than a true value, it is highly influenced by several factors such as sampling and its geographic coverage (Tsetsos et al., 2019). This implies that an adequate study design is pivotal for preventing artifactual structure (Serre & Pääbo, 2004). In addition to these considerations, it has been suggested that interpretations of the output should be done with care since different scenarios can produce indistinguishable plots which is why other methods should be used alongside to disentangle complex demographic events. This is particularly important since with real data, there may be violations to the model, such as the existence of close relatives, lack of discrete populations, hierarchical population structure, etc. (Lawson et al., 2018)

Lastly, ordination methods can also unveil population structure even in areas with low levels of genetic differentiation (Novembre et al., 2008). These are model-free multivariate approaches which aim to reduce the dimensionality of the data by projecting datapoints into a low dimensional space, allowing the identification of intrinsic patterns of data variation by visual inspection. One of the most popular methods is principal component analysis (PCA), an eigen method that projects samples into a series of orthogonal axes (principal components) independent of each other and ranked according to their variance (McVean, 2009). In the Americas, a region with low levels of diversity, PCA has proven to be useful in discovering finer underlying structure in specific regions when large numbers of SNPs are screened (Nakatsuka et al., 2020). However, a large number of markers does not necessarily signify a higher level of resolution because of the inherent ascertainment bias of some genome-wide panels, generally designed to uncover global patterns of variation.

5.2 Demographic history deductions

Contemporary patterns of human genetic variation within and between populations are the result of various demographic processes acting through time. Although descriptive approaches (e.g. PCA) can be used to reveal underlying population structure, these generally do not account for the fact that different demographic events can lead to similar patterns of genetic variation. In contrast, past demography reconstructed by using inferential approaches rely on the comparison of the observed patterns of genetic variation and model predictions (Loog, 2021). These methods can for instance identify fluctuations in N_e, thus providing a new layer of information about a population's demographic history.

5.2.1 Mismatch distributions and neutrality tests

Mismatch distributions are histograms showing the frequency distribution of the number of pairwise differences between alleles/haplotypes observed within a population (Pereira et al., 2002), and as such provide a summary of the genetic diversity. The shapes of these histograms are used for assessing whether a population has experienced expansion, stasis or decline. The footprints of population expansions or selective sweeps are characterized by unimodal mismatch distributions whereas a multimodal pattern suggests constant population size (Rogers & Harpending, 1992; Slatkin & Hudson, 1991) (Figure 6). Even when substantial violations of the infinite-sites model occur, the shape of the mismatch distributions generally reflect the actual population demographic history (Rogers et al., 1996). However, since a few fast-mutating sites tend to produce also unimodal distributions, another model that considers the mutation rate site heterogeneity has also been implemented (Schneider & Excoffier, 1999).

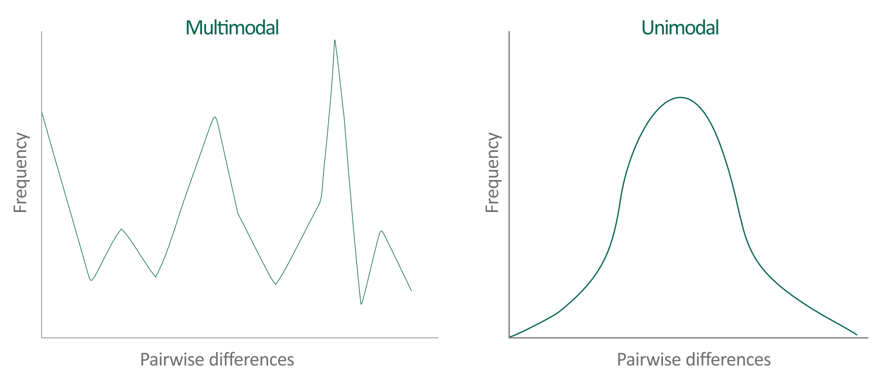


Figure 6. Different shapes of mismatch distributions showing population stasis (multimodal) and population expansion (unimodal).

Statistical tools for detecting deviations from neutrality, or the neutral Wright-Fisher population model, can also provide insights on past demography. Tests that calculate the goodness-of-fit of this model have been labeled as neutrality tests. The neutral theory of molecular evolution states that most non-deleterious mutations are neutral, meaning that after mutations arise, they get fixed or disappear as a result of genetic drift instead of selective advantage (Kimura, 1968). One class of neutrality tests are called frequency spectrum tests, which compare features of the observed diversity in relation to those expected under neutral evolution. Two widely used frequency spectrum tests are Tajima's D (Tajima, 1989) and Fu's F_S (Fu, 1997), which measure deviations from mutation-drift or mutation-selection equilibrium. Large negative D and F_S values are indicative of a recent population expansion (Fu, 1997; Tajima, 1989). These tests, however, cannot discriminate between true selection and demographic phenomena. For instance, in non-human species, an excess of rare alleles can be the result of purifying selection, selective sweeps or population expansion after a bottleneck (Hahn et al., 2002).

5.2.2 Phylogenetic inference

Phylogenetic inference attempts to reconstruct the evolutionary history of species by grouping individuals based on shared genetic ancestry into a phylogenetic tree. While the true phylogeny may never be known, it can be assumed that the molecular phylogeny carries some information about the species phylogeny. A coalescent tree (bifurcating tree) follows a backward direction and it is composed of internal nodes which represent coalescent events where two lineages merged to one ancestral state. Bootstrapping then is performed to assess how strongly the data supports each reconstructed node. Algorithm-based methods such as NJ (Neighbor-Joining) and UPGMA (unweighted pair group method with arithmetic means) are often employed to construct phylogenetic trees. Then, choosing a tree that best reflects the observed variation for a group of sequences can be evaluated based on an optimality criterion, like maximum parsimony (MP) or maximum likelihood (ML). These approaches consist on generating a large number of trees and searching across the tree space to identify the one(s) maximizing the score (Charleston, 2019). The MP algorithm considers the minimum number of changes of state needed on a tree in order to explain the sequence variation observed at the tips. Although this method is still applied, the increase in computing power has allowed ML to become one of the most popular tools for inferring phylogenies. ML evaluates the probability that a hypothetical branching pattern under certain substitution model would have produced the observed data. A tree with the highest probability (maximum likelihood) is then selected as the best.

5.2.3 Bayesian inference and Markov chain Monte Carlo algorithms

Simulations also help us understand the impact evolutionary and demographic processes have on sequence variation and sequence patterns. The two strategies are forward and backward simulations. A forward simulation starts with an initial population which evolves forward in time under defined demographic parameters resulting in a simulated dataset that can be contrasted with the empirical data. Backward or coalescent simulations, on the other hand, start with the current observations and moves backwards in time to deduce which processes may have shaped the empirical data (Yuan et al., 2012).

Following the development of coalescent theory (Kingman, 1982), genealogy-based population genetics methods have often been used to determine essential parameters that influence molecular evolution and population dynamics such as effective population sizes, rate of population decline and growth, time to the MRCA, among others (Drummond & Bouckaert, 2015). Bayesian methods have recently been at the forefront of coalescent phylogenetic inference as they allow the incorporation of prior information for the estimation of the posterior probability distribution of the parameters. The introduction of information such as sequence evolution models, dates of divergence, among others, all modify the posterior probability of the data. Since the aim of Bayesian inference is to characterize the posterior distribution, which can be represented as a multivariate function reflecting a convoluted landscape, Markov chain Monte Carlo (MCMC) algorithms were incorporated as an efficient approach to explore that probability distribution landscape (Drummond & Bouckaert, 2015). Markovian processes satisfy three properties: 1) a finite number of possible outcomes or states, 2) the probabilities are constant over time and 3) the process is memoryless. As the Markov chain has a finite (countable) state-space, the process begins in one of the possible states and moves successively to another state and so on (Carfora, 2019). The MCMC algorithm seeks to obtain more "samples" - or observations - from high posterior areas while reducing the number of "samples" from low posterior regions (Drummond & Bouckaert, 2015). Software such as BEAST 2 (Bouckaert et al., 2014) have implemented Bayesian coalescent-based MCMC methods for parameter estimation. This software estimates a probability distribution of trees which then are used to create Bayesian Skyline plots (BSP) that provide a detailed depiction of fluctuations in N_e, a key parameter utilized in this and other studies.

Therefore, by combining phylogenetic and Bayesian inferences, researchers have come up with methods to estimate for instance fluctuations in N_e through time (e.g. constant N_e, population expansions and bottlenecks). In human populations, the trajectory of the historical effective population size has been estimated with this method of phylogenetic inference. One application to lineage markers has revealed, for instance, a recent male bottleneck during the last 10 ky (kilo years) in out-of-Africa populations, which was completely absent for the female marker data (Karmin et al., 2015).

6. Forensic applications of population genetic data

In principle, the DNA variation indicative of past population history and present structure can also be applied for societal purposes, e.g. for forensic genetics. Contrasted to the large population samples and widely varying sets of markers used in population genetics, forensic genetics focuses on variation at another level, often utilizing a more restricted set of standardized markers capable of identifying individuals. DNA data is nowadays commonly used for various purposes such as paternity testing, genealogy, identification of missing persons. This section describes briefly how this discipline became an important branch of active academic research, providing vital evidence in legal contexts. It also illustrates the application of genetic data to cases of missing persons in Peru and offers a brief recount of parameters commonly used in forensic genetics.

6.1 Landmarks for the advancement of forensic genetics

DNA profiling is a method that aids in the identification of individuals or samples by examining their DNA profiles (Figure 7). Current forensic DNA profiling evolved from two important discoveries in molecular biology during the ‘80s. One of these was accomplished by geneticist Alec Jeffreys in 1984, who detected high individual variation in the number of tandemly repeated DNA fragments among individuals. The individual variation allowed identification and the Mendelian inheritance of the repeat elements made kinship studies possible in forensic casework. A year later, the polymerase chain reaction (PCR) breakthrough devised by biochemist Kary B. Mullis, allowed the targeted amplification of highly polymorphic DNA regions, laying the foundation for subsequent DNA genotyping methods in various fields of genetics. The ability to detect tandemly repeat DNA units coupled with the

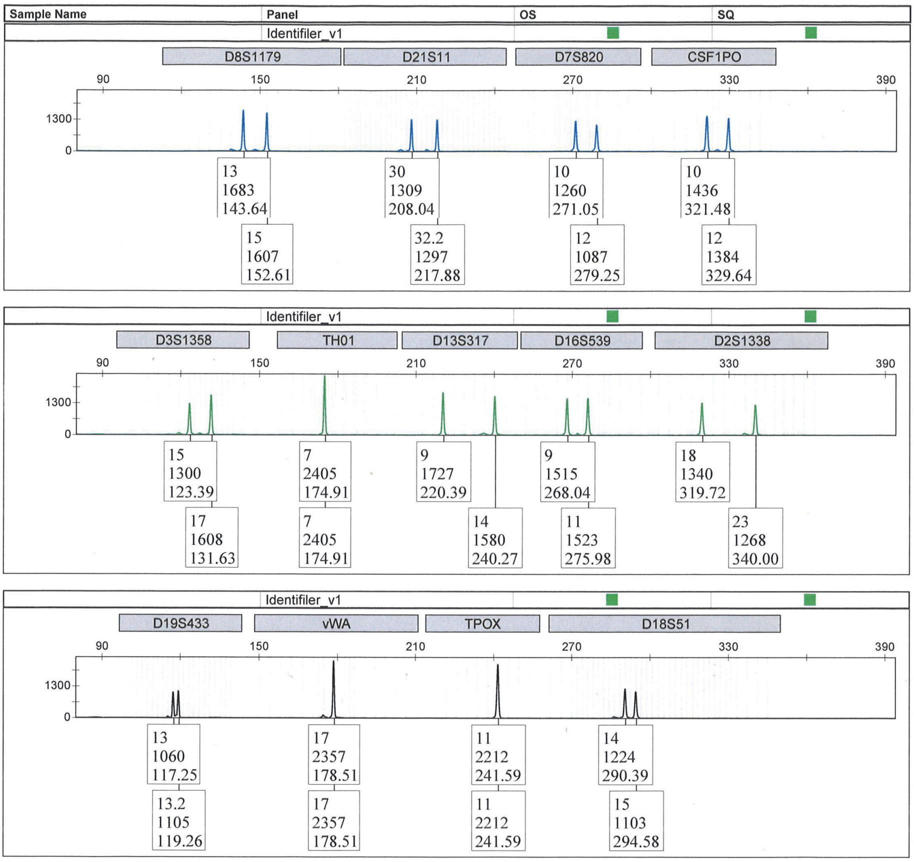


Figure 7. Partial autosomal STR profile of a modern individual. Each marker shows two peaks representing the alleles inherited from the parents. For instance, at locus D8S1179, the first allele has 13 repeat units while the second one has 15. Published with permission from the donor.

increased genotyping precision and sensitivity acquired by PCR allowed faster sample processing as well as usage of a relatively large number of markers and marker systems that are now standardized. The use of commercial kits that genotype a small set of highly polymorphic short tandem repeats (STRs) in conjunction with capillary electrophoresis have become the most widely used genotyping methods. More recently, however, NGS technologies have brought to the field various new advantages (e.g. multiplexing different kind of markers such as STRs and SNPs), but STRs have retained their position as standard forensic markers. This predilection relies in the fact that STRs have a higher discriminatory power than single nucleotide polymorphisms (SNPs), often requiring a smaller set of markers to accomplish the same level of resolution as large SNP sets (Chakraborty et al., 1999).

In forensic casework, the core process in the use of genetic evidence consists in the characterization of DNA profiles from biological sources (e.g. blood, bone) and subsequent comparisons to reference profiles. When evaluating DNA evidence, the hypothesis testing has three outcomes: inclusion (match), exclusion (no match) or inconclusive. If the match is not excluded, then the result should incorporate the statistical significance of the match, which means estimating the probability of obtaining a matching DNA profile by chance. In many other subdisciplines of genetics, it is fundamental to assess population substructure and independence within (Hardy-Weinberg equilibrium, HWE) and across loci (Linkage equilibrium, LE) to provide robust inferences about the potential origin of a sample. In fact, HWE and LE are prerequisites for the correct estimation of probability of the genotype frequencies. Since the probability calculations are based on the estimation of the genotype frequencies from the allele frequencies, any deviation from HWE or LE would distort the estimated significance.

6.2 Identification and kinship

Individual identification and kinship assessments have been widely used in various forensic genetic contexts. Early on, the applications were targeted to resolving crimes (i.e. linking suspect and crime by proving the origin of samples left at the crime scene), paternity and immigration cases (Jeffreys et al., 1985) but the spectrum has considerably widened over the years even though individual identification remains at the center of these investigations. STR profiling is largely applied in homicides, sexual assaults, paternity, kinship, search of missing persons, mass disaster victim identification, exoneration of wrongfully convicted individuals, among others (e.g. Biesecker et al., 2005; Budowle & van Daal, 2008; Coble et al., 2009; Corach et al., 1997; Dolan et al., 2009; Huffine et al., 2001). Specific cases that require, for instance, biogeographical ancestry assessments in cold cases, can however need the use of specific types of SNPs, known as ancestry informative makers

(AIMs) in the field of forensic genetics (e.g. Romanini et al., 2015). Similarly, for highly degraded samples, the short amplicons obtained in SNP typing have a clear advantage over STRs. In addition to autosomal loci, sex chromosome markers and mtDNA sequences often aid in the resolution of challenging cases involving, for instance, familial relationships or DNA mixtures (e.g. Amorim et al., 2019; Roewer, 2019).

6.3 DNA databases and the importance of reference profiles

In criminal investigations, once a genetic profile is obtained from a suspect, it needs to be compared to reference DNA data or DNA traces from a crime scene. This involves the screening of dedicated national DNA databases for possible matches and/or the acquisition of samples from relatives for genotyping purposes. Although virtually every country manages large DNA profile collections, not all nations have established government-administered national databases geared to criminal investigations. By 2013 the largest databases in the world were from China (20 million profiles) and the United States of America (12 million profiles) (Ge et al., 2014; Roewer, 2013), each of which includes a battery of core STR loci. In the United States of America, the Combined DNA Index System (CODIS), a program developed by the FBI and created to support law enforcement by administering DNA databases, includes nowadays 20 core autosomal loci included in its platform. The National DNA Index System (NDIS), part of the CODIS program, contains ~15 million offender profiles as of January 2022 (CODIS-NDIS Statistics, https://www.fbi.gov/services/laboratory/biometric-analysis/codis/ndis-statistics), being arguably the largest database in the Americas. In these country-level databases, however, non-standard-loci such as uniparental markers, are absent. Nevertheless, population genetic databases have been implemented as supplementary sources for forensic investigations. Two widely known platforms are the Y-Chromosomal STR Haplotype Database (YHRD, https://yhrd.org) (Willuweit & Roewer, 2007, 2015) and the EDNAP (European DNA Profiling Group) Mitochondrial DNA Population Database Project (EMPOP, https://empop.online) (Parson et al., 2004, 2014; Parson & Dür, 2007). More recently, NGS technologies have facilitated the obtention of both length- and sequence-based data for various types of markers, which has contributed to the expansion of existent databases and the creation of new ones. Two examples are STRidER (STRs for Identity ENFSI Reference Database, https://strider.online) (Bodner et al., 2016) and STRSeq (The STR Sequencing Project, https://strseq.nist.gov) (Gettings et al., 2017) databases. Importantly, before the incorporation of DNA data into all these databases, the genetic profiles are scrutinized through various quality control steps in a rigorous manner in order to guarantee their validity and to facilitate standardization. Recommendations on the best practices in forensic genetics are provided by the European DNA profiling group (EDNAP), a working group of the International Society for Forensic Genetics (ISFG) which also promotes collaborative exercises that focus generally on methodological and marker specific aspects.

Other databases that have been created for specific identification purposes also exist. After the World Trade Center (WTC) mass fatality (Biesecker et al., 2005), the world's largest forensic case to date, around 22,000 body parts from 2,753 victims were recovered. Because of the magnitude of this event, DNA identification efforts centered not only in obtaining genotypes from the remains and associated biological materials on personal items (e.g. brushes) but also from the relatives of the victims. Individual genetic profiles from the WTC were screened against a databank of 20,000 reference samples obtained from relatives; however, not all victims have reference profiles available for comparison. During the identification process of the 9/11 victims, in addition to standard CODIS STR genotypes, SNP and mtDNA profiles have also been produced (Biesecker et al., 2005). This background demonstrates that human DNA databases are integral parts of forensic investigations involving biological samples which is why considerable time and resources are allocated for building and maintaining them in every country across the globe.

6.4 Forensic genetics in Peru

During the past 20 years, DNA analyses in a legal context have developed towards international standards also in Peru. At present, certified public and private entities perform DNA genotyping for various types of markers. One of these is the Unit of Molecular Biology and Genetics (UNBIMOG: *Unidad de Biología Molecular y Genética*) from the Peruvian Institute of Legal Medicine and Forensic Sciences (*Instituto de Medicina Legal y Ciencias Forenses*), a government-funded institution, that since its creation in 2001, has had the mission of the identification of missing persons in Peru. Although initially its main aim was to aid in the search of missing persons during the period of internal conflict in Peru between 1980-2000 and resolve the identification of individuals in a fire-induced mass disaster occurred in the city of Lima in 2001 (*Mesa redonda* fire), nowadays, this specialized unit handles also other cases in the civil (e.g. paternity assessments) and penal (e.g. murders) jurisdictions. They use a wide array of marker systems which include minimal autosomal STR sets used in the USA (CODIS) and in Europe (ENFSI), as well as miniSTRs, uniparental markers and indels (https://www.mpfn.gob.pe/iml/UNBIOMOG/). This specialized laboratory has produced the largest genetic database for forensic casework in Peru, but because this information is sensitive, as they are associated to the above-mentioned mass disaster victims and cases of missing persons in human rights violations and crimes against humanity contexts, the access to the genotypes is restricted like in most national databases. So far, there are no public reports on the status of the identification of the *Mesa redonda* incident by means of DNA typing while in the case of the internal violence there is significantly more information, which is briefly summarized below.

The estimation of dead and missing persons during the armed internal conflict in Peru between 1980 and 2000 is of about 69 000. The largest percentage of crimes (46%) were committed by the terrorist group Shining Path (*Partido Comunista del Perú – Sendero Luminoso*), followed by state agents (30%), while the remaining amount is attributed to other agents or circumstances. By the end of their investigation in 2003, the Peruvian Truth and Reconciliation Commission (*Comisión de la Verdad y Reconciliación*), received reports of ~24,000 dead and missing people. The poorest regions of south-central Peru such as Ayacucho, Huancavelica and Apurimac, where there was a weak state presence and marginal integration to the national economic markets, witnessed the origin and escalation of violence that spread across the country. Not surprising, it was also there where the largest number of fatal victims occurred (Comisión de entrega de la CVR Peru, 2004). In a recent report, the number of missing persons was estimated to be 20 329 from which only 6 000 – 8 000 bodies have been recovered (Iannacone & Parra, 2020) but the status of the number of victims identified by DNA and other methods is not provided. In their efforts to obtain reference profiles, the UNBIMOG carried out a large sampling of populations nationwide, including those critical areas where large human losses have been reported. This set of reference samples comprised 880 individuals from 20 regions across Peru (Iannacone et al., 2011). The authors reported an unusually high proportion of unrelated random matches and heterozygote deficiency in the Ayacucho region in the Peruvian highlands when analyzing autosomal and X-chromosome STRs. They speculated this may have been the result of inbreeding or high co-ancestry levels in this particular region. Although the authors did not detect deviations from Hardy-Weinberg equilibrium in the full dataset (whole Peru), they suggest population structure based on a phylogenetic tree (Iannacone et al., 2011; Iannacone & Parra, 2020). In the case of the Ayacucho region, the authors highlight the importance of taking into account the population structure, the degree of inbreeding and admixture when running a missing person's profile against a database (Iannacone & Parra, 2020), a recommendation that does not extend to other regions of Peru.

The persistent social conflicts across the nation due to laws that undermine the sovereignty of indigenous populations over their land, their water use and a generalized discontent about the negative effects of mining and oil exploitation -land, water and air pollution- from big corporations who take advantage of weak regulations ignites often confrontations between state agents and indigenous populations. One of these occurred in 2009, in the territory of the Jivaroan Awajún where a protest (road blockage) that lasted over 50 days ended in tragedy with the loss of at least 33 lives (Stavenhagen & Monge, 2009; Stetson, 2012; Ungurahui-IWGIA, 2010).

Since every year a large number of dead and missing persons as a result not only of social conflicts but also because of high crime rates in Peru exist, it deems necessary to continue increasing the available genotypes from various regions to aid in identification efforts. In this line, with our newly produced autosomal STR and SNP datasets, the aim was to provide another level of resolution in the local diversity (length- and sequence-based allele frequencies) and improve the power of evidence by combining STR and SNP data from the Northeastern Peruvian populations.

6.5 Power of evidence and forensic genetic parameters

Similar to other types of evidence, DNA must be compared against reference profiles and the power of a match evaluated statistically. In forensic genetics this principle is straightforward since the probability of a match can be expressed numerically after the estimation of allele frequencies in a population. Commonly used indices indicative of the probability of a match are described here.

The random match probability (RMP) evaluates the probability of matching a random profile in a population. Since the loci utilized are independent, the locus-specific match probabilities are multiplied, meaning that the power of the statistic in casework depends on the number of loci typed (Balding, 1999; Foreman & Evett, 2001). The power of exclusion (PE) instead estimates the efficiency of the set of markers used when it comes to excluding genotypes. The power of discrimination (PD) is the efficiency of a given set of markers in discriminating between two unrelated individuals. This is a statistic easily deducted from RMP (1-RMP), whose power also increases as more loci are added.

The polymorphism information content (PIC) estimates the probability to deduce parental alleles when examining the genotypes of a given offspring (Botstein et al., 1980). The typical paternity index (TPI) instead represents the likelihood that the alleles present in the child support the assumption that the tested man is the biological father rather than a random selected man from the population (Gjertson et al., 2007). The ultimate aim for using forensic genetic parameters such as these is to assess the usefulness of a specific set of markers in a target population.

AIMS OF THE STUDY

All the questions and aims raised in the Review of the Literature can be approached with genetic tools and data from study and reference populations. This dissertation crystallizes these problematics by deepening our knowledge on the contemporary genetic variation in such a remote and understudied region of Peru. Although the ultimate aim of this dissertation was unravelling the population structure and demographic history of one region of South America, the nature and type of markers analyzed here served also another purpose, to provide new higher resolution data for forensic genetic applications in Peru. As noted in the previous chapter, questions are numerous as new ones tend to emerge when adding new layers of information. In genetics, all types of data have limitations and as such, with the current marker set, we are able to evaluate some questions while building the foundation for future research.

Specific aims were the following:

1. to assess the amount and nature of genetic diversity in the Chachapoyas and their neighbors (I, II, III)

2. to uncover potential genetic affinities of the Chachapoyas with populations from Peru and the Americas (I, II, III)

3. to test whether the alleged pre-Hispanic population differentiation (archaeological and ethnohistorical sources) in the Chachapoyas region translates to genetic substructure (II)

4. to establish if the Chachapoyas demographic history resembles the patterns observed in Andean or Amazonian regions (I, II)

5. to determine whether Spanish contact during the 16th century caused a more dramatic population decline on males than on females (II)

6. to assess if European gene flow (autosomal) across the Chachapoyas was greater in regions where Inca occupation was more intense (II)

7. to measure the levels of genetic diversity in North-eastern Peru by means of forensically relevant autosomal STR and SNP markers for both length- and sequence-based data and assess whether the combination of these markers increases the power in the calculation of various forensic genetic parameters (III).

MATERIALS AND METHODS

1. The context: modern populations in this study

Similar to other parts of Peru, the modern populations from the Chachapoyas region have experienced significant admixture with other populations from different continental origins after European contact. The current inhabitants who occupy the territories of the ancient Chachapoyas peoples, recognize themselves as their descendants although they do not self-identify as members of any indigenous or native group but as *mestizo* (admixed) instead. The reshaping of the local populations, their institutions and their culture by the Incas and Spanish seem to have indeed removed the signatures of the alleged ancient ethnic diversity of the region.

The general census of 2017 shows that the population in the territory where the vestiges of the ancient Chachapoya are distributed, amounts to 262 814, from a total of 379 384 for the whole Amazonas administrative region (INEI, 2018a). Two recent censuses (2007: 81.8% and 2017: 82.4%) indicate that most of the population that lives in Amazonas was originally born there. The total number of people migrating from Amazonas to other regions is 29 688, most of which (38.2%) relocated in Lima. On the other hand, the region that contributed the largest number of migrants was Cajamarca, with 40 663 individuals (INEI, 2018a). Indeed, large pockets of migrants from Cajamarca are common in southern towns such as Leymebamba but they are also present around Kuelap (Evelyn Guevara, personal observation). Near the city of Chachapoyas, the capital of Amazonas, a town referred as Huancas represents an enclave population whose ancestors came from the Mantaro valley in Junín before the Spanish arrived in the area. The Huancas district has a population of 1 258 individuals (79% males and 21% females) (INEI, 2018a), although in the town of the same name there are probably only from 200 to 250 people (Evelyn Guevara, personal observation). In the northern part of Amazonas region, particularly in the provinces of Condorcanqui and Bagua, we can observe the largest number of people that belong to the Jívaro or Chicham ethnolinguistic group, which includes Awajún and Wampís speakers. Awajún is the second largest indigenous group (419 communities) from the Peruvian tropical rainforests while the Wampís population is moderately large (65 communities) (INEI, 2018b). In 2017, the combination of these two indigenous peoples amounted to around 40,000 individuals in the Amazonas region (INEI, 2018a). A sampling across Amazonas, which incorporated these modern populations was utilized in this dissertation (Figure 2).

2. Samples

Saliva samples from 382 individuals from the northeastern Peruvian Andes were collected for this research. All of these were used in publication I while a subset from this initial set was employed for study II and III (Table 1). The research was approved by the Hospital District of Helsinki and Uusimaa Ethical Committee (permit #329/13/03/00/13). We followed ethical guidelines such as the Declaration of Helsinki and subsequent amendments and we obtained permission from local authorities in Peru. All authorities, community representatives and each sample donor were informed about the nature and aims of the study. DNA was extracted using a high salt protocol extraction (Quinque et al., 2006). For data analysis, samples were grouped considering the place of birth up until the grandparents' generation for individuals in the Chachapoyas region, in the town of Huancas, the linguistically differentiated Jívaro and migrants from Cajamarca in the Chachapoyas territory. Locally, within Chachapoyas, samples from towns and small villages were pooled into subgroups historically (colonial) or archaeologically affiliated, such as Chillaos, Corobamba, Kuelap, Pomacochas (see publications I and II for details). Similarly, for autosomal data, the Jivaroan set was divided into its original units labeled Awajún and Wampís.

Reference datasets from published sources were also included in the analyses. For study I, HVR-I data from 99 additional reference populations were included. In the case of Y-STR analyses, haplotypes (up to 17 loci) from 68 reference populations were gathered. In study II, mitogenomes from 78 worldwide reference populations were included. In the case of Y-STR data, 88 worldwide (23 STRs) and 79 Native American (17 STRs) reference datasets were used. Autosomal data from 31 reference populations was included for study II while a subset of 21 populations from this reference dataset was used for study III. Details of these reference datasets are provided in the supplementary files of publications I-III.

Table 1. Number of study samples per publication

Study	Marker	Chachapoyas	Jívaro	Huancas	Cajamarca	Total
I	mtDNA (HVR-I and HVR-II)	277	47	21	37	382
	Y-chromosome (23 STRs)	113	24	12	25	174
II	mtDNA (Full genomes)	102	42	4	16	164
	Y-chromosome (SNPs)	120	24	12	21	177
II, III	Autosomal (27 STRs and 94 SNPs)	180	38	9	14	240

3. Methods

The methods utilized in this dissertation are briefly described here. Please refer to the Materials and Methods sections of each original article for a detailed description of protocols and computational work. Table 2 summarizes all the methods utilized in publications I-III.

Table 2. Methods and corresponding publications

Methods	**Publications**
Sample collection (fieldwork)	I
DNA extraction	I
Genotyping and sequencing	
1) HVR-I and HVR-II mtDNA	I
2) Full mtDNA	II
3) Y-STR (23 loci co-amplification)	I
4) Y-SNP multiplexes	II
5) Autosomal STRs and SNPs	II, III
Data analysis	
1) Grouping strategy (surname and relatedness analyses)	I
2) Genetic diversity and genetic distances	I, II, III
3) Rarefaction	I
4) Haplogroup frequency permutation	II
5) Phylogenetic reconstruction (Neighbor-Joining)	I, II, III
6) Principal component analysis (PCA)	I, II
7) Principal coordinates analyses (PCoA)	II
8) Multidimensional scaling (MDS) plot	II
9) Mantel test	I
10) AMOVA	I, II
11) Mismatch distributions	I
12) Tajima's *D* and Fu's *Fs*	I
13) Bayesian skyline plot (BSP)	II
14) Hardy-Weinberg Equilibrium (HWE) and LinkageDisequilibrium (LD)	II, III
15) Clustering analyses (STRUCTURE)	II
16) Length- and sequence-based allele count	III
17) Expected and observed heterozygosity	III
18) Forensic parameters (RMP, PE, TPI, PIC)	III

RESULTS

These results are based on the analysis of new genotypes from four populations (Table 1) and reference datasets from the Americas and the world for mtDNA, Y-chromosome and autosomes (see Materials and Methods). The focus is mostly on the Chachapoyas dataset but information from the other study populations complements the observations at a regional level.

1. Haplogroups in the study populations

1.1 MtDNA

Haplogroup assignments were based in HVR-I and HVR-II sequences (Study I) as well as in full mitochondrial genomes (Study II) whose absolute frequencies are given in Figure 8. In study I, the Chachapoyas had about 89% of Native American haplotypes. Non-Native American haplotypes in the Jivaro and Huancas were very rare (≤ 2%) whereas it was higher in the Cajamarca (8%). Most populations had haplogroups B and C as the most common, with the exception of the Huancas, which showed haplogroup A and B as the most frequent. Interestingly, haplogroup D was absent in the Jivaroan sample. When the Chachapoyas set was divided into sub-units (Chachapoya, Chillao, Corobamba, Pomacochas, Rodríguez de Mendoza, La Jalca, Leymebamba and Chilchos), Rodríguez de Mendoza and Chilchos lacked haplogroup A. In study II, the incorporation of full mtDNA genomes increased the accuracy of subhaplogroup assignment. Here, over 95% of haplotypes belonged to haplogroups A2, B2, C1 and D1 in the Chachapoyas. For the other study populations, the frequency of haplogroups had similar distribution patterns as in study I. In order to determine whether the lack of haplogroup A2 in the subunit Rodríguez de Mendoza was due to chance, random sampling procedures were computed. Assuming average Peruvian haplogroup frequencies (*N* = 442), 2-9 observations of A2 are expected in a sample of 27 individuals. This suggests that the lack of A2 in Rodríguez de Mendoza is not a sampling artefact but due to real population-specific processes. The random sampling method revealed also that in the Jivaroan sample, haplogroup B2 was significantly more common and haplogroup D was less common than expected. Among the other Chachapoyan subunits, Pomacochas had haplogroup B2 as more common than expected. Similarly, haplogroup H was more common than expected in the Chillao and Rodríguez de Mendoza subunits.

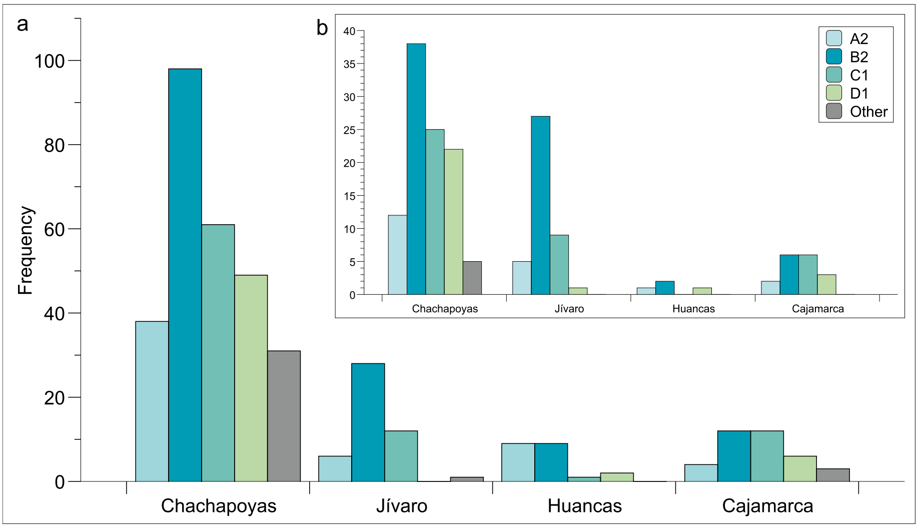


Figure 8. Bar plots showing the absolute frequency of mtDNA haplogroups in study I (a) and study II (b). Non-Native American haplogroups were combined in the category "Other".

1.2 Y-chromosome

Y-chromosomal haplogroups were estimated based on STR-haplotypes (Study I) and by direct SNP assignment with multiplex assays (Study II). In study I, roughly 60% of haplotypes in the Chachapoyas belong to haplogroup Q, the most frequent haplogroup in the Americas. The Jívaro and Huancas, had even a larger proportion of haplogroup Q (≥91%) whereas in the Cajamarca, non-Native American haplogroups were more frequent (52%). In study II, haplogroup Q (58%) was the most frequent for the Chachapoyas, followed by non-Native American haplogroup R (20%). Haplogroup Q was also predominant on the other study populations (43-96%) where non-Native American haplotypes were rarer. Absolute frequencies for all populations in both study I and study II are visualized in Figure 9.

2. Genetic diversity in northeast Peru

2.1 MtDNA

For both, study I and II, high levels of Native American haplotype diversity were observed in all study populations ($h \geq 0.93$) at local and continental levels, even after rarefaction methods (Study I). Huancas though, exhibited less diversity for full mtDNA genomes in Study II ($h = 0.83$). Other indices of genetic diversity such as nucleotide diversity (π) and mean number of

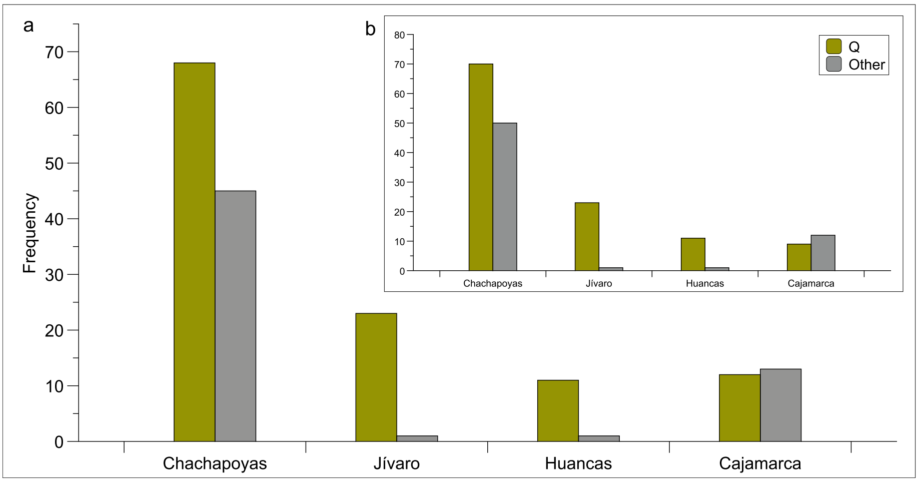


Figure 9. Bar plots showing the absolute frequency of Y-chromosome haplogroups in study I (a) and study II (b). Non-Native American haplogroups were combined in the category "Other".

pairwise differences (MNPD) also suggested high diversity in the Chachapoyas. Haplotype diversities in the study and a few reference populations are shown in Figure 10a.

2.2 Y-chromosome

Similar to mtDNA data, high levels of Native American haplotype diversity were observed in most study populations ($h \geq 0.92$) for both, study I and study II. However, the level of diversity in the Huancas for study II was also relatively low ($h = 0.88$) (Figure 10b). The MNPD pattern of the Chachapoyas also resembles the moderate diversity observed in non-admixed populations from South America.

3. Genetic differentiation

3.1 Regional differentiation

3.1.1 MtDNA

In study I, the Chachapoyas had small but significant differences with the Jívaro and Huancas ($\Phi_{ST} \leq 0.05$, $p \leq 0.05$) but non-significant differentiation from the Cajamarca. In study II, for the merged set including samples from (Barbieri et al., 2017), the Chachapoyas had non-significant genetic affinities with the Cajamarca while the comparisons with the Jívaro ($\Phi_{ST} = 0.08$) and Huancas ($\Phi_{ST} = 0.13$) in turn showed significant ($p \leq 0.05$) genetic differentiation

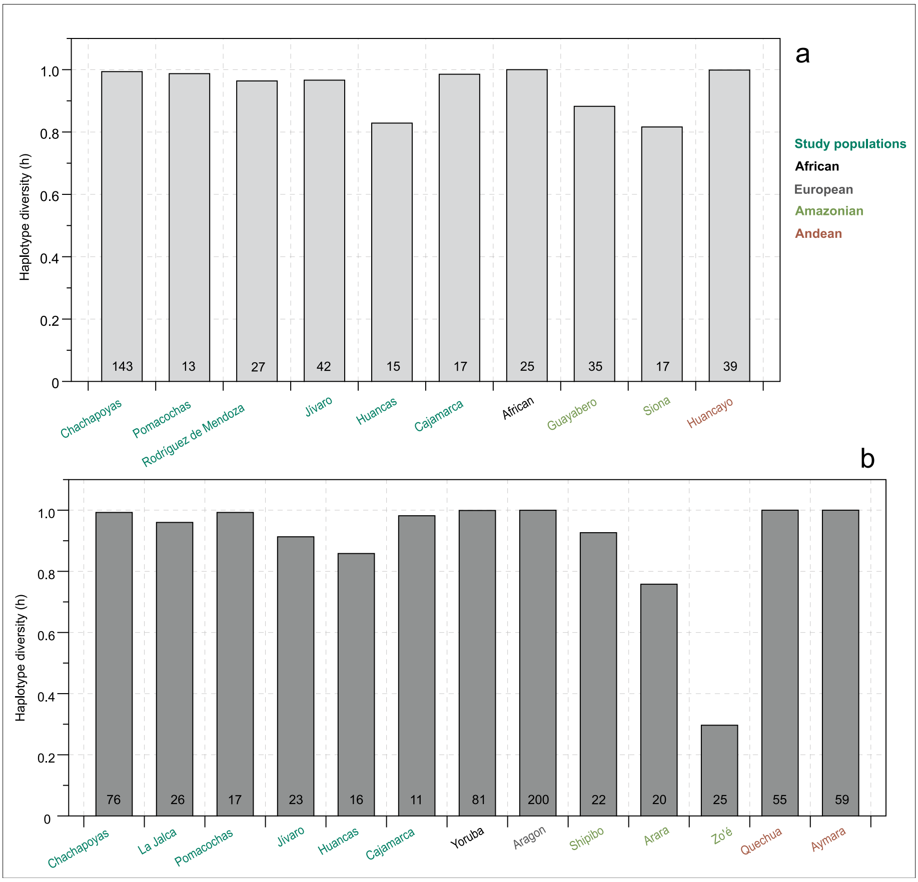


Figure 10. Haplotype diversities of study and few selected reference populations (Study II) for mtDNA (a) and Y-chromosome (b). Pomacochas, Rodríguez de Mendoza and La Jalca appear as independent Chachapoyan subunits while the remaining subsets were pooled under the label "Chachapoyas". Reference populations include an African and a European set, and populations from the South American Andes and Amazon. Population sample sizes are shown within each bar.

(Figure 11). In study II, genetic differentiation was also explored locally among the Chachapoyan subunits. Here, the Pomacochas subunit showed clear differentiation from the others ($\Phi_{ST} \geq 0.11$, $p \leq 0.05$) whereas Rodríguez de Mendoza tend to differentiate only from two of the other subunits, Chachapoya and Pomacochas ($\Phi_{ST} = 0.06$ and $\Phi_{ST} = 0.16$, $p \leq 0.05$).

3.1.2 Y-chromosome

In study I, the Chachapoyas had short but significant genetic distances with the Huancas and the Cajamarca ($\Phi_{ST} = 0.05$, $p \leq 0.05$), and remained distant to the Jívaro ($\Phi_{ST} = 0.17$, $p < 0.01$). In study II, our dataset was also merged with that from (Barbieri et al., 2017). Here, the Chachapoyas had affinities only with the Huancas but clearly differentiated from both the Jívaro and the Cajamarca (Figure 11). At local level, when evaluating the Chachapoyan subunits, only the Pomacochas differentiated clearly ($\Phi_{ST} \geq 0.1$, $p \leq 0.05$) from the rest of the subunits.

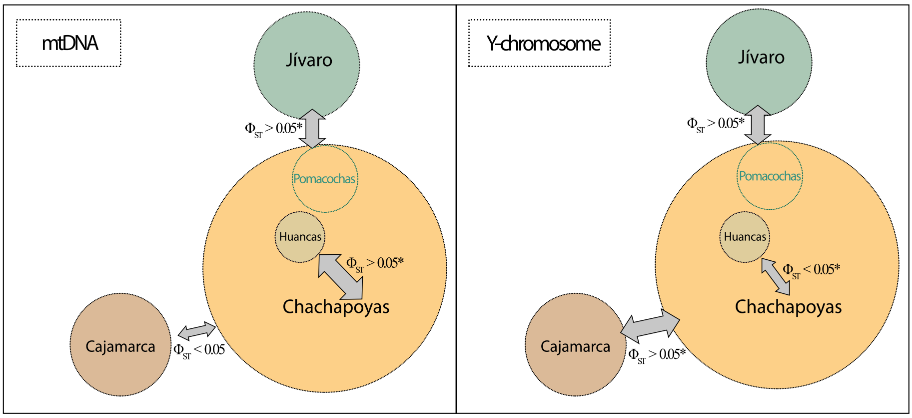


Figure 11. Schematic illustration of the genetic affinities among study populations. Wide arrows represent higher differentiation while narrow ones indicate lower differentiation. An asterisk (*) shows significant values $p \leq 0.05$. Circles reflecting approximate sample sizes. The Pomacochas subgroup is shown as to reflect its geographic proximity to Jívaro (Based on results from Study II). Unpublished figure.

3.1.3 Autosomal markers

When considering 121 autosomal markers (27 STRs and 94 SNPs), there were short but significant genetic distances ($F_{ST} < 0.01$, $p \leq 0.05$) among all pairwise comparisons including the study populations, except between the Chachapoyas and the Cajamarca.

3.2 Continental differentiation

3.2.1 MtDNA

For study I in the context of Peru, the Chachapoyas showed little differentiation to ancient and modern populations from the North and Central Peruvian Andes as well as with populations in the Peruvian Altiplano ($\Phi_{ST} = 0.05$, $p \leq 0.05$). In South America, small but significant genetic distances were observed mainly with populations from northern South America (French Guyana, Colombia) and Central America (Panama). Interestingly, in NJ trees the Chachapoyas assume a basal position among most South American populations. This is also supported by a PCA based on mitochondrial haplogroup frequencies where the Chachapoyas situate near the centroid. In study II, when including full mitochondrial genomes, the Chachapoyas show short genetic distances with ancient and modern populations from Central Peru ($\Phi_{ST} = 0.05$, $p \leq 0.05$). In the context of the Americas, the Chachapoyas had lower differentiation with various populations from Colombia and northern Ecuador, as well as with a general population from Mexico. In MDS plots, the Chachapoyas situate near the centroid while the other study populations cluster either with Andean or Amazonian populations.

3.2.2 Y-chromosome

In study I we observed little differentiation ($\Phi_{ST} = 0.05$, $p \leq 0.05$) with populations from the north and central Peruvian highlands as well as with Chilean populations (i.e. Mapuche and Tehuelche). In NJ trees, the Chachapoya is also basal to most Andean and Amazonian populations. For study II, low differentiation with other populations from the Americas were much less common than for mtDNA. There were only short genetic distances between the Chachapoyas with three central Andean populations and one Amazonian group from Peru ($\Phi_{ST} = 0.04 – 0.05$, $p \leq 0.05$). Interestingly, in this study shorter genetic distances were detected between the Chachapoyas and the Mbyá Guaraní from Argentina ($\Phi_{ST} = 0.03$, $p \leq 0.05$). The MDS plot shows the Chachapoyas situated in between Andean and Amazonian groups and the other study populations falling in either Andean or Amazonian clusters (Figure 12).

3.2.3 Autosomal markers

For comparisons at intercontinental level, five additional populations were included in the analyses, namely Yavapai (Wendt et al., 2016), AFA = African American, ASN = Asian American, CAU = Caucasian, HIS = Hispanic (Churchill et al., 2017; Novroski et al., 2016), all from the USA. The pairwise comparisons indicate close affinities between all study populations and the Hispanic (HIS) set. On the other hand, the Caucasian (CAU) showed affinities ($F_{ST} = 0.02\text{-}0.05$, $p \leq 0.05$) only to the Chachapoyas, the Jivaroan Awajún and the Cajamarca. The African American (AFA) and the Asian American (ASN) datasets remained distant to all study populations.

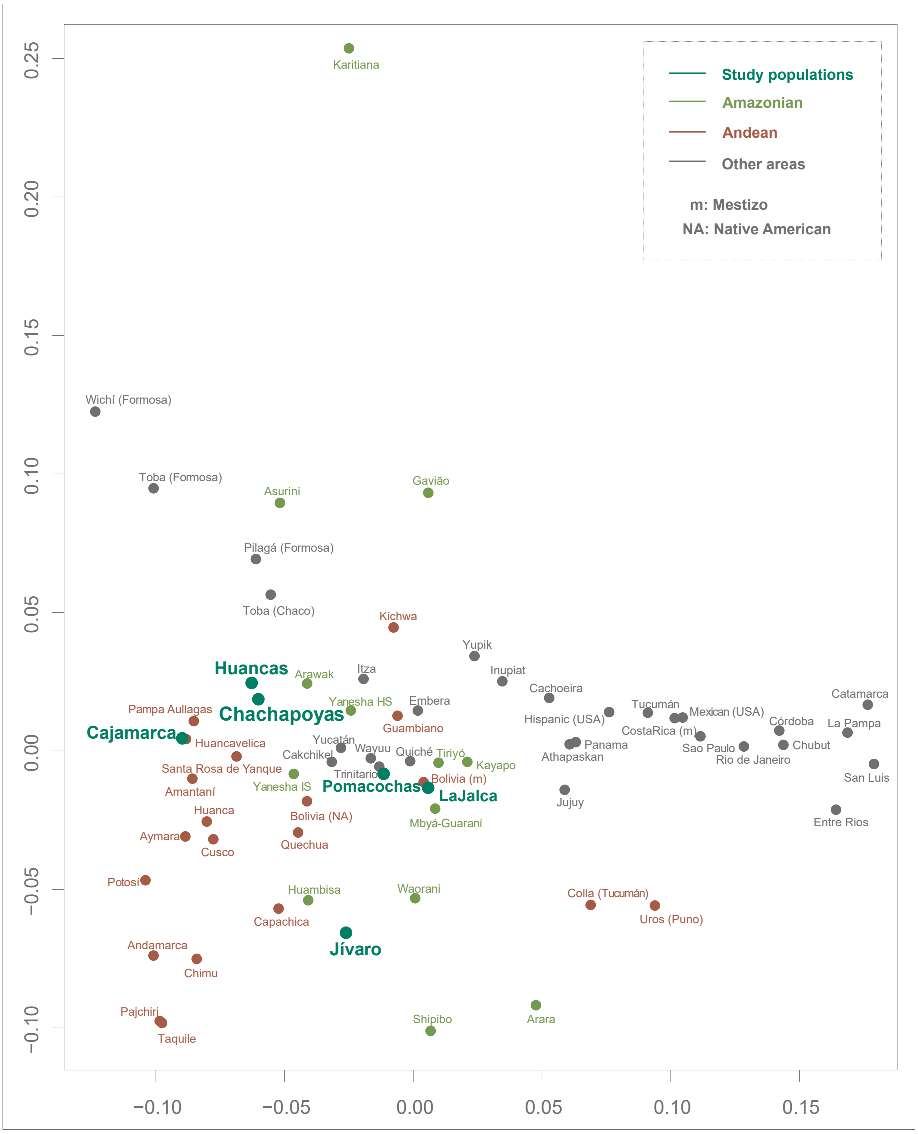


Figure 12. Y-chromosomal MDS plot based on genetic distances and including reference populations from the Americas. Note the central position of the Chachapoyas in the plot. Adapted from Supplementary Figure S4 in Guevara et al. PLoS One 15(12), 2020 (CC BY 4.0 license).

4. Further insights on population structure

4.1 Analysis of molecular variance (AMOVA)

With the aim to find the most homogeneous groups for both lineage markers, AMOVA was applied to search for groupings that minimized the within-group (F_{SC}) and maximized the among-group (F_{CT}) variation. Two different scenarios were tested. In the first one the Chachapoyas were included in (1) Amazonian, (2) coastal Pacific, (3) Andean or (4) Altiplano. This arrangement did not yield low F_{SC} values indicating large interpopulation differences. In the second scenario, the Chachapoyas were included in either of these three regions: (1) Amazonia, (2) Peruvian coast or (3) Peruvian Andes. Here, F_{SC} values were similar to the ones obtained in the first scenario.

Hierarchical analyses were also applied to the local level (i.e. Chachapoyan subunits) together with the Amazonian Jívaro. A reduction in mtDNA F_{SC} values was observed only when the Pomacochas subunit was grouped with the Jívaro. In study II, very small and non-significant F_{SC} and F_{CT} values also pointed to negligible population structure with any kind of marker used (see Table S13 from study II).

4.2 Admixture proportions

Admixture proportions for autosomal data were assessed at three hierarchical levels, (1) intercontinental, (2) intracontinental and (3) regional levels. At the upper level there were four well defined clusters (ln $P(X|K)$ and ΔK) associated with African, European, Asian and Native American ancestries. At the second level, in the context of the Americas, there were three clusters, namely European, Native American and a 3rd component which could not be associated to any of the known major groups. This component was observed in high proportions in the Chachapoyas and also in the Awajún and Wampís even when including a reduced set of markers (18-27 loci). Independently of their geographic origin, all study samples clustered together showing no substructure in the context of other populations from the Americas (HIS and Yavapai). Despite this, there is an emerging admixture pattern at local level, i.e. Chachapoyan subunits. The Native American component is present at more than 50% in individuals from Chachapoya, Chillao and Pomacochas. On the other side of the spectrum, individuals from Kuelap, Leymebamba and Rodríguez de Mendoza, generally exhibited more than 50% of European ancestry (Figure 13).

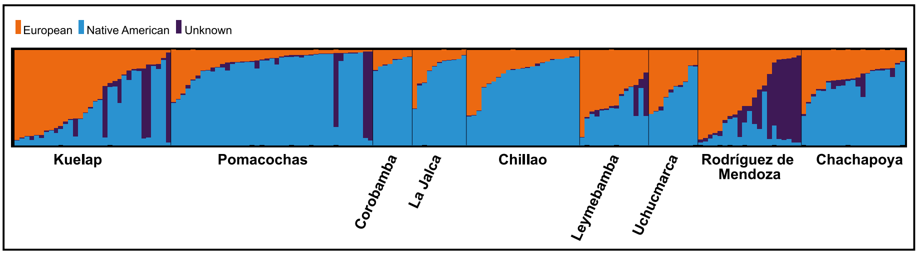


Figure 13. STRUCTURE plot for K = 3, showing the proportion of Native American and European ancestry in the Chachapoya subgroups as well as the 3rd component of unidentified origin (purple). Adapted from Figure 3 in Guevara et al. PLoS One 15(12), 2020 (CC BY 4.0 license).

4.3 A perspective from ordination methods

For the same autosomal data, another means to clarify potential population structure in the Chachapoyas was with PCA. Samples from all study populations and the dataset CAU (European ancestry) were included in the analyses. Here some individuals from the populations Chachapoyas, Awajún and Wampís clustered together in a cline towards the Native American group clearly differentiating from the European set (see supplemental Figure 7 in Study II). To verify the position of this small 3rd group, the clustering method output was used to build another PCoA plot. Samples were first binned considering the proportion of the third unknown component in three classes, light grey = 0-20%, grey = 21-59% and black = 60-100%, and then they were plotted in a new PCoA. Similar to the initial plot, samples containing ≥ 60% of the 3rd unidentified component cluster together in a cline towards the Native American group (Figure 14).

In order to have more insights on the origin of the 3rd unidentified component, samples having ≥ 60% of this component were grouped in a new synthetic population and then compared with other populations worldwide in a pairwise fashion. This was carried out with 116 loci and 10 loci, depending on the availability of published reference data and the completeness of the reference genotypes. Considering the 10 loci dataset, the 3rd component shows affinities ($\Phi_{ST} \leq 0.05$, $p \leq 0.05$) with HIS (Hispanic from the USA) as well as with urban populations from Mexico, Chile and Nicaragua. The comparisons with CAU (European ancestry) and Huelva (a Spanish population) show large genetic distances ($\Phi_{ST} \geq 0.07$, $p \leq 0.05$). In MDS plots, the 3rd unidentified component remains distant to most populations, being relatively close only to the Amazonian Cashibo from Peru when using the 10 loci dataset.

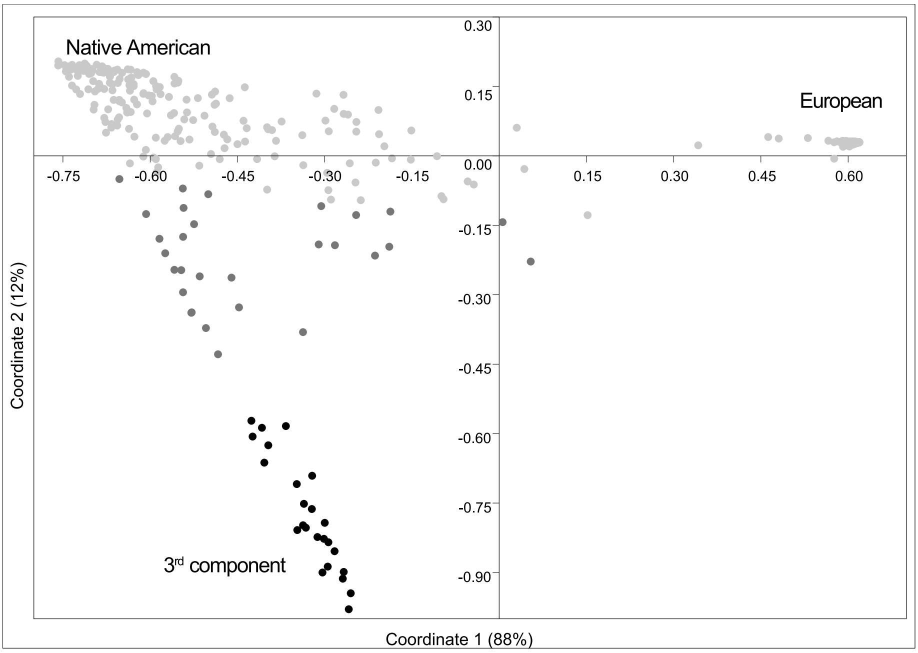


Figure 14. PCA showing a small cluster with individuals harboring more than 60% of the 3rd component in black. Adapted from Figure 4 in Guevara et al. PLoS One 15(12), 2020 (CC BY 4.0 license).

5. Demographic history

5.1 Mismatch distributions, Tajima's *D* and Fu's *F*s (Study I)

For mtDNA data, the Chachapoyas mismatch distribution is multimodal exhibiting peaks at four, seven and twelve differences. Although there is a drop in diversity at one difference, there is an increase in diversity at 0 differences. The other study populations show similar patterns but have lower frequencies at one, two and three differences. In the context of South America, most populations exhibit a drastic reduction in their diversity with frequencies dropping at one to three differences but the Amazonian populations seem to have been more severely impacted. On the other hand, few Andean populations showed no signs of a drastic reduction in their diversity. The Chachapoyas had also large and statistically significant negative Fu's F_S (F_S = -24.11948, $p <$ 0.001) and Tajima's D (D = - 1.56347, $p < 0.029$) values, indicating population expansion which is compatible both with the high levels of diversity and its mismatch distribution in this population. The mismatch distribution for Y-chromosome data is unimodal with a peak at thirteen differences. The other study populations had multimodal plots suggesting constant population sizes with few instances of reduction of their diversity.

5.2 Coalescent approach: BSP (Study II)

5.2.1 MtDNA

The BSP plots for all study populations and Chachapoyan subunits show a marked population growth in the distant past at around 500 generations (12.5 kya, all estimations assume a 25-year generation interval), which reaches a plateau at about 250 generations ago (6.3 kya). Later on, a reduction of the effective populations size (N_e) starts at around 125 generations ago (~3 kya) continuing gradually up until recent times (Figure 15a). In the context of South America, the BSP of the Andean region is very similar to the ones described for our study populations. However, the Amazonian BSP does not show any peak of growth at any stage in stark contrast with the study populations (see Figure 1B in study II).

5.2.2 Y-chromosome

An accurate estimation of the male effective population size usually requires Y-chromosomal sequence data which is rarely produced. Since coalescent approaches based on more readily available STR data are not well established, past male population dynamics are seldom assessed. In order to compare female and male past effective population sizes, we assumed an experimental approach to construct BSPs with Y-STR marker data.

The Y-chromosome BSP of the Chachapoyas shows a marked population growth at around the same time as the mtDNA plot, from 600 (15 kya) until reaching a plateau at 200 generations ago (5 kya). After this, there is a population decline that persists until reaching its lowest point at 25 generations ago (~625 years ago). This deep population contraction of around 50% parallels the time of Spanish conquest. Unlike the mtDNA BSP, a male population rebound of about 20% is observed in the Chachapoya which is absent in the mtDNA data (Figure 15b). In the case of the Amazonian region, a very modest population increase is observed. Later on, its population size declines continuously from 250 generation ago (~6.3 kya) until reaching its lowest point at 110 generations ago (~2.8 kya). After this, a rebound of the population size similar to the Chachapoyas is observed in this dataset. Unlike the Amazonian BSP, the Andean region does not show any signatures of population decline in the recent past. Instead, it exhibits a two-stage increase in N_e. The first one at 512 (12.8 kya) generations ago and a second at 73 generations ago (1.8 kya) (see Figure 2B in study II).

It is relevant to note that the temporal congruence between mitochondrial and Y-chromosome BSP markers support the validity of the results based on Y-STR data presented here.

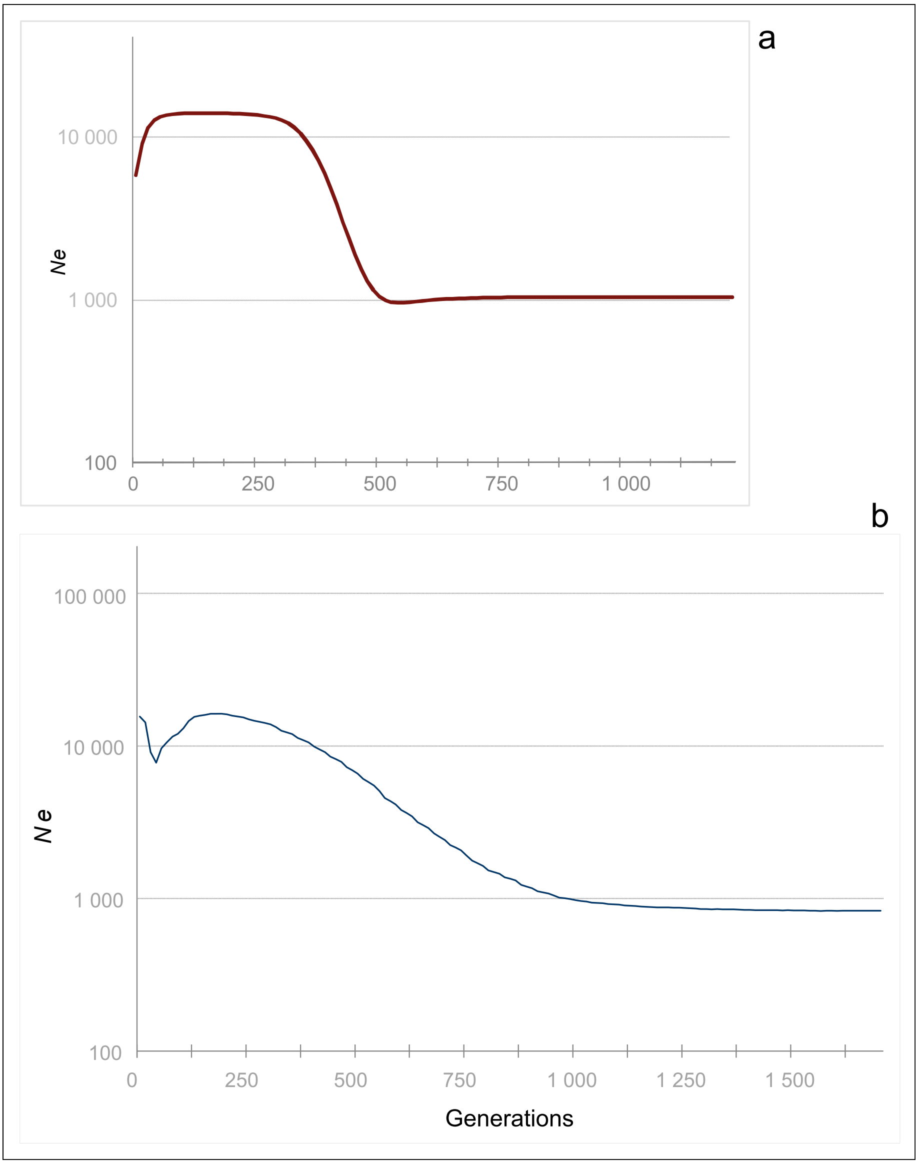


Figure 15. BSPs of the Chachapoyas pooled for mtDNA (a) and Y-chromosome (b). Adapted from Figure 1 and Figure 2 in Guevara et al. PLoS One 15(12), 2020 (CC BY 4.0 license).

6. Tests of independence and forensic genetic parameters

6.1 HWE and LD (Studies II and III)

HWE tests were performed in (1) each study population, (2) in the pooled Northeastern Peruvian dataset (all study populations) and (3) in the context of Peru. In the first case, after correction for multiple testing, there was only one significant HWE deviation (locus D1S1656) in the Chachapoyas. When pooling all study populations, one HWE deviation was detected for Northeastern Peru (locus rs2111980). When including other reference populations from Peru and our own datasets but less markers, two deviations were detected (Penta E and D3S1358).

Linkage disequilibrium was tested in 27 autosomal STRs (aSTRs) and in 94 autosomal identity informative SNPs (iiSNPs). After applying correction for multiple testing, just two pairs of syntenic loci showed departures from expectations, one in the Chachapoyas (rs13182883 - rs338882) and one in the Cajamarca (rs6955448 - rs917118). When merging our study populations and after correction for multiple testing, only one pair of loci showed departures (rs13182883-rs338882).

.

6.2 Length- and sequence-based genotype differences (Study III)

NGS technologies allowed the investigation of nominal allele lengths in conjunction with sequence-based information for different type of markers. Here a recount of the findings on STR and SNP autosomal variation is provided. For STR loci, six of them did not show any increase between length- and sequence-based genotypes in any of the study populations: CSF1PO, D10S1248, D17S1301, D22S1045, TH01 and TPOX. The remaining ones exhibited some form of allele gain. For instance, the loci that showed the greatest allele gain were D12S391, D13S317, D21S11, D2S1338, D3S1358 and vWA (Figure 16). When examining the 94 iiSNPs, about 33% of them had an increase in the number of alleles for sequence-based genotypes that include the flanking region (i.e. microhaplotypes).

6.3 H_{exp} and RMP in the Chachapoyas (Study III)

Expected heterozygosity (H_{exp}) and RMP were calculated for both length- and sequence-based genotypes in all study populations; however, results are presented only for the Chachapoyas, the dataset with the largest sample size (*N* = 172). In this set, there was considerable intra-allelic sequence variation in a large number of loci which increased their H_{exp}. Here, the percentage

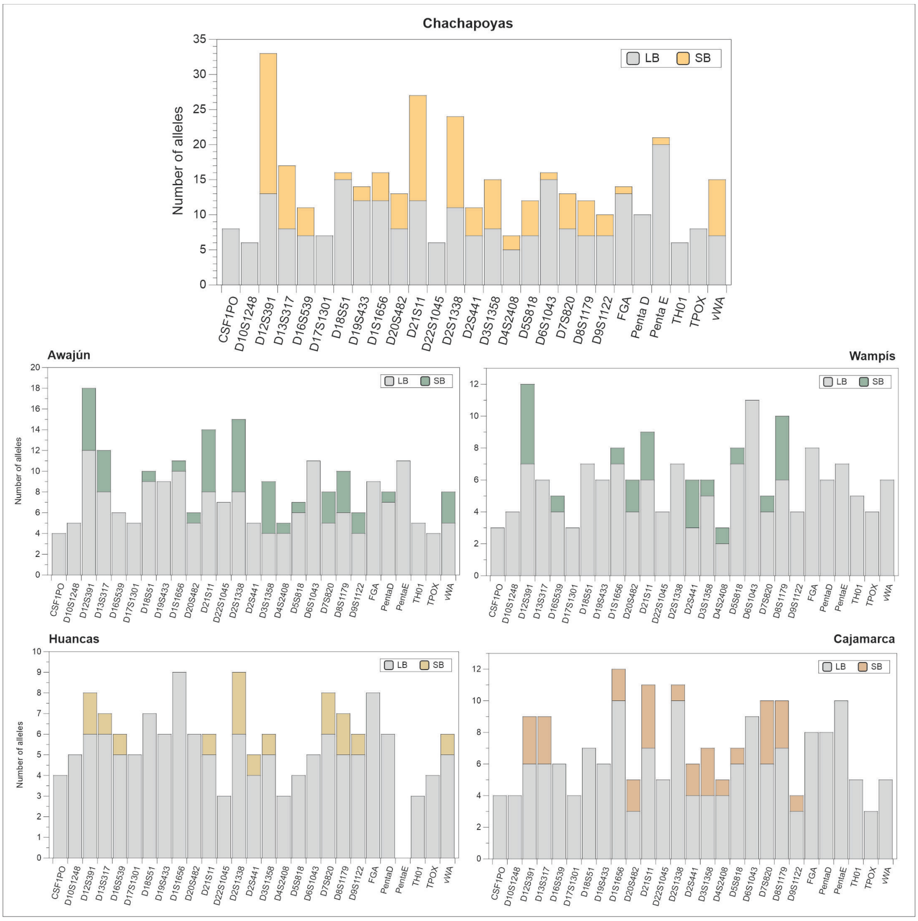


Figure 16. Length (LB) and sequence-based (SB) genotype frequencies for autosomal markers in all study populations. Adapted from Supplementary Table S1a in Guevara et al. Forensic Sci. Int. Genet. 52: 102487, 2021.

increase in H_{exp} ranged from 0.2% in Penta E to 28.4% in D4S2408. For STRs, the average single-locus RMP was 0.123 ± 0.084 for length-based genotypes and 0.096 ± 0.069 for sequence-based data. When examining iiSNPs, the average single-locus RMP for length- and sequence-based genotypes were 0.452 ± 0.112 and 0.436 ± 0.123, respectively. The lowest RMP by locus for length- and sequence-based genotypes was Penta E (0.013 and 0.012 respectively), and the highest was observed in locus rs1357617 (0.862 and 0.841, respectively). Finally, the combined RMP (aSTRs + iiSNPs) for length- and sequence-based data in the Chachapoya were 8.14 x 10^{-62} and 4.15 x 10^{-67}, respectively, adding more power than when STRs and iiSNPs are calculated independently.

DISCUSSION

During recent years there has been an increasing effort to obtain genome wide data from Native American and admixed populations from Peru. This new level of resolution has allowed to explore population dynamics pre- and post-European contact in various regions of Peru (Harris et al., 2018). It has also facilitated to unveil gene flow events in north Peru that linked diverse environments (coast, highland and jungle) signaling the importance of genetic barriers such as geography and culture to gene flow across regions (Borda et al., 2020). Despite considerable advances on completing the genetic landscape of this area of the globe there are still few underrepresented regions of Peru that deserve our attention. In order to fulfil this aim, this dissertation constitutes one of the first approaches to understand the population structure and demographic history of northeastern Peru by using various kind of genetic markers. This research has aimed to assess the levels of genetic diversity in the Chachapoyas from the northeastern cloud forest as well as their neighbors. By doing so, we have uncovered short- and long-distance genetic affinities of the study populations with others from the Americas and detected insubstantial pre-Columbian substructure in the Chachapoyas region. Furthermore, we have assessed the impact of the Spanish contact in the local demography which revealed different trajectories for lineage markers. In addition to these, we have explored the patterns of European gene flow by using autosomal markers while providing relevant data that can be useful for forensic genetics in Peru and South America.

1. High levels of genetic diversity in the Chachapoyas

Several ethnohistorical sources document a drastic population decline in the Chachapoya territory before, during and after European contact (Church, 1996; Cook, 1981; Schjellerup, 2005). Initially, the Incas applied a relocation policy which consisted in the transference of segments of the Chachapoya population(s) to different locations across the Inca domain (Schjellerup, 2005). Later on, at the time of European contact, diseases, war, forced labor, also caused population decline in this and many other Peruvian regions (Cook, 1981). Despite these diverse lines of evidence suggesting a dramatic population decline in the region, these changes appear to be not so profound when examining solely the modern genetic diversity and required instead to scrutinize the fluctuation in past population sizes to detect events of population decline. We found high levels of genetic diversity in the Chachapoyas for both lineage markers (mtDNA and Y-chromosome) studied here. Additionally, other metrics such as large negative Tajima's *D*, Fu's *Fs* and unimodal mismatch distributions for Y-chromosome, indicate that there was no drastic reduction of the effective population size in the Chachapoyas (Study I). Genetic diversity may indeed survive after population size declines that could be considered as catastrophic in humans, as suggested from studies in other mammals (Kekkonen et al., 2012). Our studies (Study I and II) and other work (Barbieri et al., 2017) have stressed the singularity of the Chachapoyas, which is evident in the lack of extensive haplotype sharing between Chachapoyas and other populations from Peru and South America. This has led to claims on which the high diversity in the Chachapoyas has been ascribed to the genetic composition before the Inca conquest (Barbieri et al., 2017). However, based on the available data, it is problematic to specify time frames for the development of such high levels of Native American diversity in the gene pool of the Chachapoyas because we are not examining directly the ancient genetic variation in a diachronic fashion. Similarly, the origin of this "Chachapoyan diversity" being local or having other influences cannot be established since only a few neighboring populations have been studied so far.

The high diversity observed in the Chachapoyas was also reflected in the large N_e estimates (Study II). This observation complies with studies exploring the levels of diversity of other species in the region. A study that relied on the analysis of birth and persistence of biological diversity indicates that topography and climate are the main drivers of diversity (Rangel et al., 2018). In silico and empirical data in this study point to Northeastern Peru and Southwestern Ecuador as one of the hotspots of biodiversity in South America (Rangel et al., 2018). Rich ecosystems such as these, would mean higher environmental net production that could support population growth in many species including humans. It is important to note that this Andes-Amazon interface, which was densely occupied by the ancient Chachapoya people, has different kind of ecosystems over a relatively short geographic distance, which could have buffered the local populations (human and non-human) in times of environmental change (Loreau & de Mazancourt, 2013). The ability to maintain higher or more stable population densities through time would be reflected in higher genetic diversity of the local human communities (Freeman et al., 2020; Lewis et al., 2020; Tallavaara et al., 2018).

2. Cryptic population affinities

The ancient and modern Chachapoya people have inhabited a territory on the eastern slopes of the northern Andes which has a gradient of ecosystems that culminate in the vicinity of the Amazonian basin. Here, there are a series of corridors that provide suitable access routes for the acquisition of goods and movement of people (Church & von Hagen, 2008). Early lithic assemblages (12000-11000 BCE) linking this area to the Peruvian highlands and the Ecuadorian coast indicate that long distance movement of people occurred since very early times (Church, 1996). Later on, during the Late Intermediate

Period (1100-1475 CE), it is common to find pottery from the neighboring Andean Cajamarca region in Chachapoya sites. During the Late Horizon or Inca Period (1475-1532 CE), pottery associated again to the Cajamarca and to the Chimú from the Peruvian north Coast (Guillén, 1998; Guillén et al., 2011) provide also evidence for long distance exchange dynamics. When situating the Chachapoyas in the context of South America, genetic affinities are observed with Andean but also with coastal and Amazonian populations (Studies I and II). For both mtDNA and Y-chromosome data, the Chachapoyas assume a basal position in phylogenetic trees (Study I) and a location near the centroid in MDS plots (Study II), which may hint to a combined Andean-Amazonian ancestry and/or connections.

For mtDNA data, a trend is also observed in affinities with populations from more distant locations such as northern Ecuador, Colombia, French Guyana, Panama and Mexico (Studies I and II). For Y-chromosome, affinities are geographically more restricted than for mtDNA data, confined only to the Mapuche and Tehuelche from Chile (Study I) and the Mbyá Guaraní from Argentina (Study II). These affinities with different geographically distant populations from the Americas may suggest the survival of common shared ancient genetic diversity accumulated during much earlier periods.

Locally, considering only the study populations, the Chachapoyas showed affinities with the Cajamarca but differentiated from the Jívaro and Huancas in the case of mtDNA data. For Y-chromosome, the Chachapoyas had affinities with the Huancas and Cajamarca and remained distant to the Jívaro. These findings point to different population dynamics for males and females in the Huancas which despite the hundreds of years of admixture with the Chachapoya people still retain their original female genetic diversity. The genetic affinities between the Chachapoyas and the Cajamarca are not surprising since there is archaeological evidence of intense exchange networks at least starting by the Late Intermediate Period (1100-1475 CE) as well as more recent waves of migrants from the neighboring region of Cajamarca towards the Chachapoya territory. The differentiation of the Jívaro from the Chachapoya is also expected as there are no documented accounts of Jivaroan migrations further south into the Chachapoya area or vice versa. However, gene flow was detected between the northern Chachapoya territory (comprising Pomacochas, Jumbilla and Yambrasbamba) and the Jívaro.

2.1 Pomacochas – Jívaro connections

When the Chachapoyan subunits were examined, only the Pomacochas (which includes also Jumbilla and Yambrasbamba localities) clearly differentiated from the other subgroups for both lineage markers (Study II). In addition to this, affinities between the Jívaro and the Pomacochas subgroup were observed also for autosomal data. Although within archaeology/ethnohistory it remains to be established whether the area around Pomacochas and further north share cultural features with the other regions considered part of the Chachapoya realm, potential historical connections between these areas cannot be ruled out. For instance, it has been suggested that the historical Jívaro and the ancient Chachapoya peoples shared some cultural traits (Guffroy, 2006; Koschmieder, 2014b), however, the available archaeological evidence supporting that notion is still insufficient. Despite this, two recent independent studies (Borda et al., 2020; Nakatsuka et al., 2020) have identified east-west gene flow linking the north Peruvian coast and north highlands, and also affinities with Amazonian populations located to the east of the northern Andes. Indeed, one genome-wide study (Borda et al., 2020) found gene flow between the Chachapoyas and populations from the Jivaroan ethnolinguistic stock such as the Awajún and Candoshi. This gives further support to our finding and suggests potential genetic connections between the Jívaro and neighboring populations from the eastern slopes of the Andes, which may or may not have signified actual cultural exchanges. Archaeological/ethnohistorical research examining connections between these regions may shed light on the nature of those interactions in the future.

3. Lack of substructure within Chachapoyas

Even though recent archaeological and ethnohistorical research point to cultural diversity in the Chachapoyas region at least during the Late Intermediate period (1000-1475 CE) (Church & Guengerich, 2017), there are no strong signatures of modern genetic subdivision in the region. Although certain degree of genetic differentiation among some Chachapoyan subunits (autosomal: Pomacochas; mtDNA: Rodríguez de Mendoza) was observed, the markers examined indicate that present-day substructure is insubstantial. We did not detect any signs of substructure in the allele/genotype distributions by HWE-tests or by the clustering method (implemented in STRUCTURE). The HWE test is a relatively sensitive method, but failed to show any deviations, and, similarly, all Chachapoya samples clustered together regardless of the hierarchical setting in which they were placed (intercontinentally, the Americas or regionally), showing again no detectable population substructure.

3.1 Autosomal variation component of undetermined origin in the Chachapoyas

In addition to the Native American and European components, a third autosomal component was detected among the Chachapoyas and the Jivaroan Wampís and Awajún. Within the Chachapoya sample this component was commonly observed in higher frequency in some subunits (e.g. Pomacochas) whereas it was nearly absent in others (e.g. La Jalca). This autosomal component could not be directly associated to any modern source population but clearly differentiated from African, European and Asian populations in clustering analyses.

With the aim of unveiling the origin of this variation, individuals with ≥ 60 % of the component of undetermined origin were placed in a synthetic population and compared in a pairwise fashion to reference populations worldwide. Again, the analyses did not show obvious affinities with African, European or Asian variation. When including more reference populations from the Americas but less loci (10STRs), the unknown component had affinities with a few populations from the Americas (e.g. Nicaraguan mestizo). In MDS plots the 3rd unknown component remained distant to most populations with the exception of the Amazonian Cashibo from Peru. Since the exact origin of this variation could not be clarified, we are confined to indicate it may be of Native American origin based on the genetic distance evidence. Interestingly, a relatively recent study (Barbieri et al., 2019), which analyzed over 600,000 SNPs and more populations from western South America, identified two autosomal components of alleged Amazonian origin. The first one, labeled as "Amazonia North", is observed in populations to the east of the northern Andes in Ecuador and Colombia, and the second one, "Amazonia Core" distributed mostly among populations from the north Peruvian tropical rainforest. Based on the geographic proximity and the high frequency among the Amazonian Jívaro, we could speculate that the 3rd unknown component has ties with the "Amazonia core" component or to other Amazonian populations from the vicinity. This however, cannot be tested with our current data and resolution.

4. Asymmetrical European gene flow in the Chachapoyas

Similar to other South American populations (e.g. Bolnick et al., 2006; Gayà-Vidal et al., 2011), gender-biased European gene flow was observed in the Chachapoyas. Non-Native American haplotypes in the Chachapoyas account for 5% in the mitochondrial data and 42% in Y-chromosome data. When Europeans arrived in the region, they formed alliances with the local populations and founded the city of Chachapoyas as their capital. Due to its importance as the center of Spanish rule during colonial times and because it is still the largest city in the whole region, it would be rather intuitive to find a higher proportion of European autosomal variation in the Chachapoya subgroup. In addition to this, using as a proxy the presence of Inca settlements, which is more pronounced on the east side of the Utcubamba River, one could hypothesize Spanish occupation was restricted to the localities in this area as they already had a well-established infrastructure at hand. However, the largest European contribution was observed in the subgroups Leymebamba, Kuelap and Rodríguez de Mendoza, of which only the first situates to the east of the Utcubamba River. Potential explanations to all these patterns are as follows. Leymebamba is one of the most important settlements on the eastern slopes of the northern Andes and is located in a key position along one of the most frequently used roads that crosses the Andes and the Marañon River. Since the Late Horizon period (1100-1475 CE), this road system was vital to connect the Cajamarca and the Chachapoyas regions as well as the town of Uchucmarca in the south (Schjellerup, 2005). During the Early Colonial Period (1532-1600 CE), this route also facilitated the movement of Spanish conquistadors into the region and most likely the introduction of European individuals into the local communities. In the 16th century, during a visit to the area around Rodríguez de Mendoza, archbishop Mogrovejo identified already various Spanish settlements (Mogrovejo, 1921[1593]). Although there are no reports indicating significant Spanish migrations particularly in this region later on, the area seems to have been an important place for European settlers, which is reflected not only in their genetic make-up but also in their particular Caucasian phenotypic traits. A number of factors such as warmer climate, fertile soil and large rivers may have also contributed to draw the attention of the new settlers to permanently establish in this area. Kuelap is one of the largest and most important archaeological sites in the Chachapoyas region. The site has been recognized as a center of religious and political power for the local populations before the Spanish conquest (Narváez, 2013). Despite this, it has been seldom cited in early documents and was rediscovered late during the 19th century (Narváez, 2013). It remains unclear why there is a large proportion of European variation in this area and what motivated European migrations here. However, although less frequently than in Rodríguez de Mendoza, people with Caucasian phenotype are often observed in the area around Kuelap and neighboring localities.

5. Demographic trajectories

Demographic histories were assessed for both mitochondrial and Y-chromosome data. Mismatch distributions, Fu's F_S, Tajima's D and coalescent approaches (i.e. BSPs) all convey the same picture of population expansions in the Chachapoyas for both lineage markers. During more recent times, however, fluctuations in the effective population sizes differed between lineage markers and regions (Andean/Amazonian).

5.1 Sex-specific BSP differences

Interestingly, the BSP graphs for males and females show no evidence for a mid-Holocene demographic collapse registered in archaeological radiocarbon and palaeoclimatic data (Riris & Arroyo-Kalin, 2019). Population declines for both mtDNA and Y-chromosome occur instead around 125 generations ago (i.e. 3.1 kya). This signal roughly coincides with a period of decreased solar irradiation, “The Homeric minimum”, which has been linked to global climatic change, including heightened El Niño oscillation (Sandweiss et al., 2001; Wanner et al., 2008). Particularly in the Chachapoya territory, pollen and charcoal records from Lake Pomacochas show disturbances in landscape around 3 kya (maize cultivation and reduced forest taxa), which indicates paradoxically heightened human activity (Bush et al., 2015). However, we cannot discard the possibility that local populations, during adverse times of environmental change, may have colonized new areas adjacent to the lake for maize cultivation which led to a high signal of landscape disturbances. Thus, the evidence of heightened human activity around the lake may not necessarily imply population growth but rather mobility of the communities.

Nonetheless, differences in the effective population size trajectories between the lineage markers during the last 125 generations (3 kya) are evident. The mitochondrial BSP exhibits a continuous decline that intensifies in more recent times but does not show a specific point of drastic population contraction in the recent past. The Y-chromosome BSP on the other hand, has a steep population decline starting c. 150 generations ago and reaching its deepest point around 25 generations ago (~625 years ago). This 50% reduction in effective population size would coincide with the time of Inca or Spanish conquest in the region if a generation interval of 23 or 20 years is assumed, which is not altogether unrealistic. It could well reflect the effects of the relocation policies imposed by the Incas during the Late Horizon period (1475-1532 CE), which caused certain degree of depopulation in the Chachapoyas region. However, a more severe population decline is associated with demographic changes during and after European contact (Cook, 1981; Koch et al., 2019). All these historical events showed that European contact had a profound impact on the local demography. In line with this evidence, the current genetic data allowed us to pinpoint with more certainty the time of occurrence of this event and to clarify the differences between males and females which otherwise would remain speculative.

5.2 Regional-specific patterns of variation

BSP plots from Andean and Amazonian reference populations show also dissimilarities. For Andean male and female datasets, high effective population sizes are observed. Amazonian populations, on the other hand, exhibit a contrasting pattern characterized by smaller population sizes and very moderate to non-existent population growth in the distant past. These contrasting demographic patterns related to environmental differences between Andean and Amazonian populations have been previously observed for Y-chromosomal data (Tarazona-Santos et al., 2001). When examining each Chachapoyan subgroup, most BSPs show Andean-type demographic patterns, with the exception of Pomacochas and Corobamba, whose patterns resemble the ones from the Amazonian area instead. This provides additional support to the evidence of Amazonian Jívaro affinity of Pomacochas described above.

6. Forensic genetics aspects

The newly obtained autosomal length- and sequence-based STR and SNP datasets are one of the first of this nature in South America and the only one of its kind in Peru. The added information provided by sequence information will likely be of utility in forensic casework, particularly in challenging cases such as kinship or mixture deconvolution. However, since length data for more than 15 loci are still scant in the Americas, our 27 STR set provides a more comprehensive screening of the variation in South America.

6.1 Deviations from independence

When combining all study populations (i.e. Northeast Peruvian) and when this set was merged with the Peruvian reference data, there was only one HWE deviation, indicating that this set can be utilized in forensic casework. Significant LD deviations were observed only in three pairs of syntenic loci, two in the Chachapoyas and one in the Wampís. It has been demonstrated that across the human genome, regions separated at > 0.3 cM harbor little LD (Rogers, 2014). Although these three pairs of loci showing deviations from LD are separated by > 0.9 cM and chance may explain these LD departures, we cannot rule out possible population-level mechanisms at play in these particular populations.

6.2 Levels of diversity and forensic parameters

In all populations, an increase in the number of alleles was observed for STRs when comparing length- and sequence-based genotypes. However, six out of 27 autosomal STRs did not show any increase (CSF1PO, D10S1248, D17S1301, D22S1045, TH01 and TPOX) which is in line with previous observations in other populations (Delest et al., 2020). Although the other study populations have smaller sample sizes, substantial variation for different sets of loci was also detected. There are more alleles for both length- and sequence-based genotypes in the Chachapoyas when compared to the other study populations. However, due to the small sample sizes of the other study populations, it cannot be assessed whether the differences found are the result of underlying greater diversity or sample size bias.

The comparisons of length- and sequence-based genotypes show also an increase in expected and observed heterozygosity for most STR loci, which adds to the growing body of evidence of length and sequence data differences observed in other populations from the Americas (Gettings et al., 2018; Novroski et al., 2016; Wendt et al., 2017). In the case of identity informative SNPs, 33% of these showed sequence-based allele gains (microhaplotypes) but this type of data is still quite rare for admixed and non-admixed Native American populations which is why comparisons cannot yet be carried out.

Although in Peru STR and SNP markers are not customarily combined, it has been demonstrated that the calculation of genetic parameters for combined marker sets increases the statistical power, as observed in this and other studies (Delest et al., 2020; Wendt et al., 2016).

6.3 The importance of increased diversity at sequence level

The increase in number of alleles at sequence level in various markers improves discrimination power which can be useful in kinship and mixture deconvolution cases. This aspect is particularly important when applied to Peruvian populations since, in the central highlands, high levels of random matches and heterozygote deficiency have been detected while investigating cases of missing persons (Iannacone et al., 2011; Iannacone & Parra, 2020). When examining our study populations, we did not observe unusually high levels of random matches or heterozygote deficiency, which may be also the result of higher levels of non-Native American admixture in northeastern Peru. In the context of South America, unveiling a new level of variation with sequence-based data can contribute to resolve kinship investigations in complex regions such as the central Peruvian highlands, a task that can be difficult to accomplish when using nominal allele lengths.

At a wider scale, genetic distances and structure analyses considering our study populations and other Native American sets were not able to uncover population structure within Peru. This is largely due to the reduced diversity of populations in the Americas, where a larger number of markers is necessary to tear apart different native ancestry components (e.g. Nakatsuka et al. 2020). In forensic genetics the increase in heterozygosity provided by sequence data can prove advantageous for identity testing in regions of the globe with reduced indigenous genetic diversity (e.g. Wendt et al. 2016). Large scale genotyping for both length and sequence-based alleles of Native Americans by NGS technologies is nearly absent, which is why it is important to continue characterizing various populations as hidden variation remains to be identified. For instance, STR sequence variants have been detected in specific populations while absent in others (Novroski et al., 2016). This implies that new sequence data can shed light on “private” alleles in certain geographic regions or populations, important for the assessment of ancestry of individual samples.

CAVEATS AND LIMITATIONS

As noted above, different types of data have also inherent limitations. One of the main caveats, particularly for study II, is that we did not examine directly the ancient genetic diversity as it would have been expected when evaluating questions that involve diachronic perspectives such as Chachapoyas origins and pre-Columbian population substructure. Indeed, genetic data comprising different time points gives a more detailed picture of certain past demographic phenomena which may be concealed in the present-day variation. However, the inferences provided here are mostly based on robust evidence and laid out as to reflect the limitations of the data. Another issue is associated to the sets of markers used, particularly in the case of STRs and SNPs that have been selected for forensic genetics use. It is clear that a large number of genome-wide markers (millions of SNPs) could give another level of resolution but demographic history patterns have often been deducted based on a reduced number of markers for decades. These patterns have been mirrored later on with studies including a large number of markers showing again the usefulness of a small set of markers. Another problem in the Americas, even when using large numbers of markers, is the low level of genetic diversity which prevents the identification of different Native American ancestries. This is an issue closely related to the ascertainment bias of the panels used, since, some of them were designed to uncover variation on a global scale and because few Native American populations were initially included in the assays. It is still a pending task to design new arrays that include more indigenous variation from the Americas to aid in the assessment of fine-grained genetic structure.

In this thesis, for comparison purposes, reference genotypes from a large number of populations were gathered from published sources. Obviously, in case of a number of independent studies, the quality and extent of this data varies, partly in unknown fashion. Furthermore, there are large gaps in Peru and South America where there is little or non-existent genetic information for the type markers used in this study. These undoubtedly restrict the type of comparisons that can be made and the questions that can be evaluated with genetic data.

Regarding the methodology for the construction of coalescent-based BSPs with Y-STR data, it has been previously discussed in Study II that the methods used are well established for sequence data (usually mtDNA but also Y-chromosome genomes). However, while using this experimental approach with Y-STR genotypes both sequence (mtDNA) and STR (Y-chromosome) data show very similar patterns, which gives good support to the validity of the results.

CONCLUDING REMARKS AND FUTURE PROSPECTS

For decades, much of the focus of genetic research in South America was largely in the central highlands of Peru, but, more recently, new studies in the upper Amazon have started to unveil the genetic intricacies of a complex region in terms of populations dynamics. Among those studies, this dissertation has opened new avenues in the genetic investigation of the cloud forests of northeast Peru. As such, this work is not only important in the field of population genetics in Peru but also provides new data for other applied research such as forensic and medical genetics. Several questions about the genetic structure and population history of the Chachapoyas from northeast Peru and their immediate neighbors were resolved with this dissertation while new ones emerged. The main findings of this research can be summarized as follows:

- Among the study populations and in the Americas, the Chachapoyas showed exceptionally high levels of Native American genetic diversity.
- In the context of the Americas, the Chachapoyas exhibited different patterns of genetic affinities for lineage markers but both lines of evidence hint to affinities with populations from the Andes and the Amazon alike.
- Although the archaeological and ethnohistorical body of knowledge portray Chachapoyas as a diverse entity in terms of material culture and ethnicity, there was negligible population structure in the region.
- For lineage markers, the demographic history of the Chachapoyas resembles the patterns of populations from the Andean area where larger population sizes existed. Furthermore, there are differences between the sexes, with males being more drastically impacted by European contact.
- Similar to other regions of the Americas, there was asymmetrical gene flow for lineage markers, meaning a larger European contribution for Y-chromosomes than for mtDNA. In the case of autosomal markers there was also asymmetrical European gene flow across Chachapoyas although the pattern does not mirror Inca settlement density which was used as a proxy for more intense European settlement in the region.
- From the forensic genetics point of view, the new sequence-based data provided by this research added another level of resolution when characterizing genetically the populations from northeast Peru.

The following steps on the study of the Chachapoyas include the analyses of the ancient aDNA variation in the same study area. This new study, which was originally conceived as part of this dissertation, is still ongoing since its spectrum has also broaden considerably to include stable isotope data and other ancient human populations from the tropical rainforest and the central Peruvian coast.

At the time of print of this dissertation, we are analyzing genome capture data (~1.2 million SNPs) from 82 ancient Chachapoyan individuals from various archaeological sites and spanning several archaeological periods (800-1600 CE), including pre-Columbian and early colonial samples. The aims of this follow-up study are to elucidate the origin of this ancient population (coastal, Andean or Amazonian), establish whether there is population continuity up until modern times and determine if there were pre-Columbian European migrations in the region, among others. In collaboration with researchers from Finland and Germany we are also exploring ancient pathogens persistence and evolution particularly in the context of pre- and post-Columbian contact in Peru.

The possibilities of analyzing human and pathogen genomes from contemporary and past populations is increasing exponentially and the computational methods have more power to detect fine grained patterns so that we can explore many other questions about a populations/species past diseases and evolution. The Chachapoyas and various aspects of their life will still remain a mystery but as technological advances occur and more data is produced, we are one step closer to understand their full dimension and their key role in the context of Peru and South America.

ACKNOWLEDGMENTS

The research for this dissertation was carried out in the Laboratory of Forensic Biology, Department of Forensic Medicine, University of Helsinki and in the University of North Texas Health Science Center (UNTHSC). My deepest thanks to the current and former directors of the Department of Forensic Medicine (Oikeuslääketieteen osasto) for their support throughout the years and for providing the research facilities for this study as well as funding. My gratitude goes also to the Center for International Mobility (CIMO) and Kone foundation for their generous support at various stages of the project. I would also like to thank the Regional Government of Amazonas and the Municipality of Chachapoyas for allocating funding for sample collection in Amazonas.

I would like to express my appreciation to my supervisors, Professor Antti Sajantila and adjunct Professor Jukka U. Palo for their support and patience since the beginning of the project and in my new academic endeavors. You have inspired me throughout the years and have taught me to think in a scientific manner. Thank you for having introduced me to the new but fascinating field of genetics since my master studies. For your critical but kind words and for your endless patience. Outside academia I am also in debt to you and your families for all the support in my personal life and adjustment to Finland, my new home. Similarly, my deepest appreciation goes to Professor Sonia Guillén who took me under her wing during my master studies and for her support during project design and sample collection as well as her encouragement and wise advice. You saw glimpses of a scientist in me and made an effort to pave the way for my journey to the exotic far north, my infinite gratitude to you. I want to thank Professor Bruce Budowle and his team for allowing me to work with next generation sequencing (NGS) methods at UNTHSC during a three-month research visit and also four your kindness, patience and timely responses. I would like to thank Professor Helena Ranta for her kind-heartedness and support since I arrived in Finland and for also having inspired me to continue on this academic path.

My gratitude also goes to the reviewers of this dissertation, professors Leonor Gusmão and Kristiina Tambets for their constructive and valuable comments which helped to improve the manuscript. I deeply thank professor Antonio Amorim for accepting the role of opponent in my dissertation.

During fieldwork, I received logistic support from several local authorities in the provinces of Bongará, Chachapoyas, Condorcanqui, Luya and Rodríguez de Mendoza. Among them, to the academic representatives of Universidad Nacional Toribio Rodríguez de Mendoza de Amazonas for facilitating my work in the region. My gratitude goes to them and to all anonymous sample donors across Amazonas for their participation and interest in this research. Here, I would also warmly thank Janina Chuquizuta and her family for invaluable assistance during fieldwork. Similarly, I would like to thank Luis Guevara and his family for logistic aid in Chachapoyas.

I deeply thank all team members of the former Laboratory of Forensic Biology for their kind support, Anna-Liina Mustaniemi, Tiina Valonen, Paula Lehtoaho, Kirsti Höök, Kaija Saarela, Minttu Hedman, Katarina Lindroos, Terhi Keltanen, Teija Partanen, Eve Karvinen, Anna-Mari Walta, Anna-Maija Sulonen and Antti Hakkarainen. Each one of you played an important role in this venture, either through laboratory method insights, advice or just conversations. Thank you for kindly let me be part of this group in Ruskeasuo. Also, I would like to thank my colleagues and friends Anu Neuvonen and Sanni Översti for your contribution to complete the dissertation and for sharing experiences, wishes and hopes in the life of a PhD student.

I would like to express my gratitude to my other co-authors and collaborators for their important contribution to the research articles, Jonathan L. King, Lutz Rower, Maria Seidel, Monika Stoljarova, Frank R. Wendt, Magdalena M. Bus, Anna Guengerich and Warren B. Church. Thank you for your critical comments, suggestions and your work with the samples and/or data as well as for your input during the final stages of article preparation. This work would not have been possible without your valuable contribution.

I would also like to thank professor Johannes Krause, director of the department of Archaeogenetics at the Max Planck Institute for Evolutionary Anthropology and his team for facilitating and supporting my work on the fascinating field of ancient genomics. Thanks to various research visits at the institute in Germany and collaboration efforts, I have been able to start a new journey to study the DNA of the ancient Chachapoyas.

Outside the academic world, I would like to thank numerous friends that made my life in Finland more enjoyable, for patiently listening and for encouraging me, Halley Salazar, Minttu Paananen, Hande and Ville Huolopainen, Riikka Karjalainen, Julio Hernandez, Nieves Vuoristo, Isabel Granados, Melanie Lindgren, Glenda Cardenas and Evelyn Silva. Also, I would like to thank various members and friends from Askelten Palo ry for bringing a spark of Latin American music and dance to my life.

I am in debt to my aunt Orfelia Guevara and her husband Michael Wray as well as my uncle Hugo Guevara and his family for their support in my academic endeavors. Thank you for believing in me, you undoubtedly have influenced my life in a positive way and you will always be in my heart for that. In Peru, special thanks also go to my grandmother Mercedes Orozco, my aunt Elizabeth Cubas (and all our extended family), my sisters Nathali and Ana Isabel and my brother Jorge Luis. I know you feel my absence in Peru and here in Finland you are greatly missed. It will be a great joy to tell you that I finally completed my PhD after so long but even without this accomplishment I know you feel proud of me. Thank you for your unconditional love and be certain that I love you all just as much.

Also in Peru, I want to show my appreciation to my dear friends Dora Quispe, Monica Neyra, Araceli Espinoza and Diana Leon. Despite the distance you have always been in my heart and my mind, thank you for your always kind and encouraging words. Also, my deepest gratitude to my late friend Iván Rivasplata, for your constant encouragement and appreciation. Sadly, we never said goodbye in person but I will cherish your memory in my mind. I will miss our always very entertaining conversations and your wisdom. Farewell my friend.

I thank also my friend Jaana Hurnanen (and Mimmi, Anne and Jesse) who was always there to listen and for her invaluable support during my first years in Finland. Thank you for your warm and giving heart, for the talks, the laughs, the wise advice and for embracing me like family.

Last but not least, I want to thank my husband Jouni Kärkkäinen and my foster family (Jonna, Jatta, Sofia, Oona, Kreetta and Jouko), for your love, encouragement and support, and for believing in me even in the darkest times. Particularly for gently but decisively encouraging me not to give up once so close to the final line. As many on this road would have noticed, ‘life as a PhD student is not easy’ (and life in general is not easy), but, having the support when you most need it, is invaluable. You all have made this life a lot more worth living. Looking forward to new adventures together. I love you with all my heart!

Evelyn Guevara
Helsinki, May 2022

REFERENCES

Akesson, C. M., Matthews-Bird, F., Bitting, M., Fennell, C. J., Church, W. B., Peterson, L. C., Valencia, B. G., & Bush, M. B. (2020). 2,100 years of human adaptation to climate change in the High Andes. *Nature Ecology and Evolution*, 4(1), 66-74.

Amorim, A., Fernandes, T., & Taveira, N. (2019). Mitochondrial DNA in human identification: a review. *PEERJ*, 7.

Andrushko, V. A. (2021). Revelations of ancient head shape: Cranial modification in the Cuzco region of Peru, Early Horizon to Inca Imperial Period. *American Journal of Physical Anthropology*, 175(1), 95–105.

Bachtrog, D. (2013). Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nature Reviews Genetics*, 14(2), 113–124.

Balanovsky, O. (2017). Toward a consensus on SNP and STR mutation rates on the human Y-chromosome. *Human Genetics*, 136(5), 575–590.

Balding, D. J. (1999). When can a DNA profile be regarded as unique? *Science & Justice*, 39(4), 257–260.

Ballantyne, K. N., Goedbloed, M., Fang, R., Schaap, O., Lao, O., Wollstein, A., Choi, Y., van Duijn, K., Vermeulen, M., Brauer, S., Decorte, R., Poetsch, M., von Wurmb-Schwark, N., de Knijff, P., Labuda, D., Vézina, H., Knoblauch, H., Lessig, R., Roewer, L., … Kayser, M. (2010). Mutability of Y-Chromosomal Microsatellites: Rates, Characteristics, Molecular Bases, and Forensic Implications. *The American Journal of Human Genetics*, 87(3), 341–353.

Barbieri, C., Barquera, R., Arias, L., Sandoval, J. R., Acosta, O., Zurita, C., Aguilar-Campos, A., Tito-Alvarez, A. M., Serrano-Osuna, R., Gray, R. D., Mafessoni, F., Heggarty, P., Shimizu, K. K., Fujita, R., Stoneking, M., Pugach, I., & Fehren-Schmitz, L. (2019). The Current Genomic Landscape of Western South America: Andes, Amazonia, and Pacific Coast. *Molecular Biology and Evolution*, 36(12), 2698–2713.

Barbieri, C., Sandoval, J. R., Valqui, J., Shimelman, A., Ziemendorff, S., Schroeder, R., Geppert, M., Roewer, L., Gray, R., Stoneking, M., Fujita, R., & Heggarty, P. (2017). Enclaves of genetic diversity resisted Inca impacts on population history. *Scientific Reports*, 7.

Barbujani, G., Magagni, A., Minch, E., & Cavalli-Sforza, L. L. (1997). An apportionment of human DNA diversity. *Proceedings of the National Academy of Sciences*, 94(9), 4516–4519.

Biesecker, L. G., Bailey-Wilson, J. E., Ballantyne, J., Baum, H., Bieber, F. R., Brenner, C., Budowle, B., Butler, J. M., Carmody, G., Conneally, P. M., Duceman, B., Eisenberg, A., Forman, L., Kidd, K. K., Leclair, B., Niezgoda, S., Parsons, T. J., Pugh, E., Shaler, R., … Walsh, A. (2005).

DNA identifications after the 9/11 World Trade Center attack. *Science Epidemiology*, 310(5751), 1122–1123.

Boattini, A., Sarno, S., Mazzarisi, A. M., Viroli, C., De Fanti, S., Bini, C., Larmuseau, M. H. D., Pelotti, S., & Luiselli, D. (2019). Estimating Y-Str Mutation Rates and Tmrca Through Deep-Rooting Italian Pedigrees. *Scientific Reports*, 9.

Bodner, M., Bastisch, I., Butler, J. M., Fimmers, R., Gill, P., Gusmao, L., Morling, N., Phillips, C., Prinz, M., Schneider, P. M., & Parson, W. (2016). Recommendations of the DNA Commission of the International Society for Forensic Genetics (ISFG) on quality control of autosomal Short Tandem Repeat allele frequency databasing (STRidER). *Forensic Science International-Genetics*, 24, 97–102.

Bolnick, D. A., Bolnick, D. I., & Smith, D. G. (2006). Asymmetric Male and Female Genetic Histories among Native Americans from Eastern North America. *Molecular Biology and Evolution*, 23(11), 2161–2174.

Bonatto, S. L., Salzano, F. M., & Anonymous. (1997). Diversity and age of the four major mtDNA haplogroups, and their implications for the peopling of the New World. *American Journal of Human Genetics*, 61(63c), 1413–1423.

Borda, V., Alvim, I., Mendes, M., Silva-Carvalho, C., Soares-Souza, G. B., Leal, T. P., Furlan, V., Scliar, M. O., Zamudio, R., Zolini, C., Araújo, G. S., Luizon, M. R., Padilla, C., Cáceres, O., Levano, K., Sánchez, C., Trujillo, O., Flores-Villanueva, P. O., Dean, M., ... Guio, H. (2020). The genetic structure and adaptation of Andean highlanders and Amazonians are influenced by the interplay between geography and culture. *Proceedings of the National Academy of Sciences*, 117(51), 32557 LP – 32565.

Botstein, D., White, R. L., Skolnick, M., & Davis, R. W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics*, 32(3), 314–331.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M. A., Rambaut, A., & Drummond, A. J. (2014). BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. *Plos Computational Biology*, 10(4), e1003537.

Brinkmann, B., Klintschar, M., Neuhuber, F., Huhne, J., & Rolf, B. (1998). Mutation rate in human microsatellites: Influence of the structure and length of the tandem repeat. *American Journal of Human Genetics*, 62(6), 1408–1415.

Budowle, B., & van Daal, A. (2008). Forensically relevant SNP classes. *BIOTECHNIQUES*, 44(5), 603-610.

Bush, M. B., Mosblech, N. A. S., & Church, W. B. (2015). Climate change and the agricultural history of a mid-elevation Andean montane forest. *The Holocene*, 25(9), 1522–1532.

Carfora, M. F. (2019). Stochastic Processes. In S. Ranganathan, M. Gribskov, K. Nakai, & C. Schönbach (Eds.), *Encyclopedia of Bioinformatics and Computational Biology* (pp. 747–752). Academic Press.

Chakraborty, R., Kimmel, M., Stivers, D. N., Davison, L. J., Deka, R., & Anonymous. (1997). Relative rates at di-, tri-, and tetranucleotide microsatellite loci. *Proceedings of the National Academy of Sciences USA*, 942a, 1041–1046.

Chakraborty, R., Stivers, D. N., Su, B., Zhong, Y. X., & Budowle, B. (1999). The utility of short tandem repeat loci beyond human identification: Implications for development of new DNA typing systems. *ELECTROPHORESIS*, 20(8), 1682–1696.

Charleston, M. A. (2019). Molecular Phylogenetics. In S. Ranganathan, M. Gribskov, K. Nakai, & C. Schönbach (Eds.), *Encyclopedia of Bioinformatics and Computational Biology* (pp. 700–711). Academic Press.

Charlesworth, B. (2009). Effective population size and patterns of molecular evolution and variation. *Nature Reviews Genetics*, 10(3), 195–205.

Church, W. B. (1991). La ocupación temprana del Gran Pajatén. *Revista del Museo de Arqueología de la Universidad Nacional deTrujilllo*, 2, 7–38.

Church, W. B. (1996). *Prehistoric Cultural Development and Interregional Interaction in the Tropical Montane Forests of Peru*. PhD. Yale University Graduate School.

Church, W. B., & Guengerich, A. (2017). Introducción: La reconstrucción de Chachapoyas a través de la historia e historiografía. *Boletín de Arqueología PUCP*, 23, 5–38.

Church, W. B., & Muscutt, K. (2018). Inticancha: a Wari enclave in Chachapoyas? (Reopening a cold case in a cold place). *58th Annual Meeting of the Insitute of Andean Studies*, 1–22.

Church, W. B., & von Hagen, A. (2008). Chachapoyas: Cultural Development at an Andean Cloud Forest Crossroads. In H. Silverman & W. H. Isbell (Eds.), *The Handbook of South American Archaeology* (pp. 906–923). Springer Science+Business Media, LLC.

Churchill, J. D., Novroski, N. M. M., King, J. L., Seah, L. H., & Budowle, B. (2017). Population and performance analyses of four major populations with Illumina's FGx Forensic Genomics System. *Forensic Science International-Genetics*, 30, 81–92.

Cieza de León, P. (1553). Crónica del Perú. Tercera parte. In F. Cantù (Ed.), *Descubrimiento y conquista del Perú*. Pontificia Universidad Católica del Perú, Fondo Editorial : Academia Nacional de la Historia.

Coble, M. D., Loreille, O. M., Wadhams, M. J., Edson, S. M., Maynard, K., Meyer, C. E., Niederstatter, H., Berger, C., Berger, B., Falsetti, A. B., Gill, P., Parson, W., & Finelli, L. N. (2009). Mystery Solved: The Identification of the Two Missing Romanov Children Using DNA Analysis. *Plos One*, 4(3).

Comisión de entrega de la CVR Peru. (2004). *Hatun Willakuy, Versión abreviada del informe final de la Comisión de la Verdad y Reconciliación, Comisión de Entrega de la Comisión de la Verdad y Reconciliación, Lima, Perú, Feb. 2004,* 4 77pp. Routledge.

Cook, N. D. (1981). *Demographic Collapse: Indian Peru, 1520-1620*. Cambridge University Press.

Corach, D., Sala, A., Penacino, G., Iannucci, N., Bernardi, P., Doretti, M., Fondebrider, L., Ginarte, A., Inchaurregui, A., Somigliana, C., Turner, S., & Hagelberg, E. (1997). Additional approaches to DNA typing of skeletal remains: The search for "'missing'" persons killed during the last dictatorship in Argentina. *ELECTROPHORESIS*, 18(9), 1608–1612.

Crandall, J. (2012). Chachapoya Eschatology: Spaces of Death in the Northern Andes. *Baessler–Archiv, Band* 60, 1–17.

Crandall, J. (2017). El desarrollo espacial de las comunidades Chachapoyas bajo la dominación colonial Inka y Española. *Boletín de Arqueología PUCP*, 23, 283–311.

Crandall, J. (2018). *The Transformations of Empire: Negotiating Inka and Spanish Colonialism in an Upper Amazonian Community*. PhD. Department of Anthropology, University of Florida.

Crawford, M. (2007). *Anthropological Genetics. Theory, Methods and Applications*. Cambridge University Press.

Delest, A., Godfrin, D., Chantrel, Y., Ulus, A., Vannier, J., Faivre, M., Hollard, C., & Laurent, F.-X. (2020). Sequenced-based French population data from 169 unrelated individuals with Verogen's ForenSeq DNA signature prep kit. *Forensic Science International: Genetics*, 47, 102304.

Dillehay, T. D. (2009). Probing deeper into first American studies. *Proceedings of the National Academy of Sciences of the United States of America*, 106(4), 971–978.

Dolan, S. M., Saraiya, D. S., Donkervoort, S., Rogel, K., Lieber, C., & Sozer, A. (2009). The emerging role of genetics professionals in forensic kinship DNA identification after a mass fatality: lessons learned from Hurricane Katrina volunteers. *GENETICS IN MEDICINE*, 11(6), 414–417.

Drummond, A. J., & Bouckaert, R. R. (2015). *Bayesian Evolutionary Analysis with BEAST*. Cambridge University Press.

Espinoza, W. (1967). Los señoríos étnicos de Chachapoyas y la alianza hispano-chacha. Visitas, informaciones y memoriales inéditos 1572-1574. *Revista Histórica*, XXX, 224–332.

Espinoza, W. (1973). Los grupos etnicos en la cuenca del rio Chuquimayo siglos XV y XVI. *Bulletin de l'Institut Français d'Études Andines*, II(3), 19–73.

Espinoza, W. (1988). *Etnohistoria Ecuatoriana: Estudios y Documentos*. Ediciones Abya-Yala.

Espinoza, W. (2006). La etnia Chilcho y su enclave de mitmas en Cajamarca, siglos xv-xx. *Investigaciones Sociales*, X(16), 193–238.

Excoffier, L., Smouse, P. E., Quattro, J. M., & Anonymous. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics*, 131(23b), 479–491.

Fabre, O., Guyot, J. L., Salas Gismondi, R., Malaver Pizarro, M., & Maniero, E. (2008). Los chachapoya de la región de Soloco: Chaquil, del sitio de hábitat a la cueva funeraria. *Bulletin de l'Institut Français d'Études Andines*, 37(2), 271–292.

Fagundes, N. J. R., Kanitz, R., Eckert, R., Valls, A. C. S., Bogo, M. R., Salzano, F. M., Smith, D. G., Silva Jr, W. A., Zago, M. A., Ribeiro-dos-Santos, A. K., Santos, S. E. B., Petzl-Erler, M. L., & Bonatto, S. L. (2008). Mitochondrial Population Genomics Supports a Single Pre-Clovis Origin with a Coastal Route for the Peopling of the Americas. *The American Journal of Human Genetics*, 82(3), 583–592.

Fisher, R. A., & Race, R. R. (1946). Rh gene frequencies in Britain. *Nature*, 157(3976), 48–49.

Foreman, L. A., & Evett, I. W. (2001). Statistical analyses to support forensic interpretation for a new ten-locus STR profiling system. International *Journal of Legal Medicine*, 114(3), 147–155.

Freeman, J., Robinson, E., Beckman, N. G., Bird, D., Baggio, J. A., & Anderies, J. M. (2020). The global ecology of human population density and interpreting changes in paleo-population density. *Journal of Archaeological Science*, 120.

Fu, Y. X. (1997). Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, 147, 915–925.

Fuselli, S., Tarazona-Santos, E., Dupanloup, I., Soto, A., Luiselli, D., & Pettener, D. (2003). Mitochondrial DNA Diversity in South America and the Genetic History of Andean Highlanders. *Molecular Biology and Evolution*, 20(10), 1682–1691.

Gaither, C., Koschmieder, K., & Lombardi, G. P. (2008). En la tierra de los gigantes: Un nuevo "gigante" encontrado en el sitio nor-andino de Chichita, Perú. *Revista de Bioarqueología "Archaeobios"*, 2, 28–39.

Garcilaso de la Vega, I. (1609). Comentarios Reales de los Incas. Libro octavo. In E. Centenario (Ed.), *Biblioteca Clásicos del Perú* 1. Banco de Crédito del Perú.

Gayà-Vidal, M., Moral, P., Saenz-Ruales, N., Gerbault, P., Tonasso, L., Villena, M., Vasquez, R., Bravi, C. M., & Dugoujon, J.-M. (2011). mtDNA and Y-chromosome diversity in Aymaras and Quechuas from Bolivia: Different stories and special genetic traits of the Andean Altiplano populations. *American Journal of Physical Anthropology*, 145(2), 215–230.

Ge, J. Y., Sun, H. Y., Li, H. Y., Liu, C., Yan, J. W., & Budowle, B. (2014). Future directions of forensic DNA databases. Croatian *Medical Journal*, 55(2), 163–166.

Gettings, K., Borsuk, L. A., Ballard, D., Bodner, M., Budowle, B., Devesse, L., King, J., Parson, W., Phillips, C., & Vallone, P. M. (2017). STRSeq: A catalog of sequence diversity at human identification Short Tandem Repeat loci. *Forensic Science International-Genetics*, 31, 111–117.

Gettings, K., Borsuk, L. A., Steffen, C. R., Kiesler, K. M., & Vallone, P. M. (2018). Sequence-based US population data for 27 autosomal STR loci. *Forensic Science International-Genetics*, 37, 106–115.

Giffhorn, H. (2013). *Wurde Amerika in der Antike entdeckt?: Khartager, Kelten und das Rätsel der Chachapoya*. C. H. Beck.

Gjertson, D. W., Brenner, C. H., Baur, M. P., Carracedo, A., Guidet, F., Luque, J. A., Lessig, R., Mayr, W. R., Pascali, V. L., Prinz, M., Schneider, P. M., & Morling, N. (2007). ISFG: Recommendations on biostatistics in paternity testing. *Forensic Science International Genetics*, 1(3–4), 223–231.

Gnecchi-Ruscone, G. A., Sarno, S., De Fanti, S., Gianvincenzo, L., Giuliani, C., Boattini, A., Bortolini, E., Di Corcia, T., Mellado, C. S., Francia, T. J. D., Gentilini, D., Di Blasio, A. M., Di Cosimo, P., Cilli, E., Gonzalez-Martin, A., Franceschi, C., Franceschi, Z. A., Rickards, O., Sazzini, M., … Pettener, D. (2019). Dissecting the Pre-Columbian Genomic Ancestry of Native Americans along the Andes-Amazonia Divide. *Molecular Biology and Evolution*, 36(6), 1254–1269.

Guengerich, A. (2014). *Monte Viudo: Residential and the everyday production of space in a Chachapoya community*. PhD. Department of Anthropology, University of Chicago.

Guengerich, A. (2017). Domestic architecture and the materiality of public-making in pre-Columbian Eastern Peru. *Journal of Social Archaeology*, 17(3), 263–284.

Guffroy, J. (2006). El Horizonte corrugado: correlaciones estilísticas y culturales. *Bulletin de l'Institut Français d'Études Andines*, 35(3), 347–359.

Guillén, S. (1998). Arqueología de Emergencia: Inventario, Catalogación y Conservación de los Materiales Arqueológicos de los Mausoleos de la Laguna de los Cóndores. In Instituto Nacional de Cultura (Ed.), *Report*.

Guillén, S. (2002). Las momias de la Laguna de los Cóndores. In *Chachapoyas. El Reino Perdido*. AFP INTEGRA .

Guillén, S. (2003). De Chinchorro a Chiribaya: los ancestros de los mallquis Chachapoya-Inca. *Boletín de Arqueología PUCP*, 7, 287–303.

Guillén, S., Lerche, P., & Guevara, E. (2011). *Chacha motivos en el Museo Leymebamba. Diseños para el arte y la artesanía* (First edit). Centro Mallqui.

Hahn, M. W., Rausher, M. D., & Cunningham, C. W. (2002). Distinguishing between selection and population expansion in an experimental lineage of bacteriophage T7. *Genetics*, 161(1), 11–20.

Hallast, P., Batini, C., Zadik, D., Delser, P. M., Wetton, J. H., Arroyo-Pardo, E., Cavalleri, G. L., de Knijff, P., Bisol, G. D., Dupuy, B. M., Eriksen, H. A., Jorde, L. B., King, T. E., Larmuseau, M. H., de Munain, A. L., Lopez-Parra, A. M., Loutradis, A., Milasin, J., Novelletto, A., … Jobling, M. A. (2015). The Y-Chromosome Tree Bursts into Leaf: 13,000 High-Confidence SNPs Covering the Majority of Known Clades. *Molecular Biology and Evolution*, 32(3), 661–673.

Hammer, M. F., Karafet, T. M., Redd, A. J., Jarjanazi, H., Santachiara-Benerecetti, S., Soodyall, H., & Zegura, S. L. (2001). Hierarchical Patterns of Global Human Y-Chromosome Diversity. *Molecular Biology and Evolution*, 18(7), 1189–1203.

Harris, D. N., Song, W., Shetty, A. C., Levano, K. S., Cáceres, O., Padilla, C., Borda, V., Tarazona, D., Trujillo, O., Sanchez, C., Kessler, M. D., Galarza, M., Capristano, S., Montejo, H., Flores-Villanueva, P. O., Tarazona-Santos, E., O'Connor, T. D., & Guio, H. (2018). Evolutionary genomic dynamics of Peruvians before, during, and after the Inca Empire. 115(28), E6526–E6535. *Proceedings of the National Academy of Sciences*.

Hartl, D., & Clark, A. (1989). *Principles of Population Genetics*. Sinauer associates Inc.

Hester, J. J. (1966). Late Pleistocene Environments and Early Man in South America. *The American Naturalist*, 100(914), 377–388.

Holsinger, K. E., & Weir, B. S. (2009). Genetics in geographically structured populations: defining, estimating and interpreting F-ST. *Nature Reviews Genetics*, 10(9), 639–650.

Horai, S., Kondo, R., Nakagawa-Hattori, Y., Hayashi, S., Sonoda, S., & Tajima, K. (1993). Peopling of the Americas, founded by four major lineages of mitochondrial DNA. *Molecular Biology and Evolution*, 10(1), 23–47.

Horkheimer, H. (1959). Algunas consideraciones acerca de la arqueología en el valle del Utcubamba. *Actas y Trabajos Del II Congreso Nacional de Historia Del Perú* 1, 71–101.

Huffine, E., Crews, J., Kennedy, B., Bomberger, K., & Zinbo, A. (2001). Mass identification of persons missing from the break-up of the former Yugoslavia: Structure, function, and role of the International Commission on Missing Persons. *Croatian Medical Journal*, 42(3), 271–275.

Iannacone, G., Parra, R., Bermejo, M., Rojas, Y., Valencia, C., Portugues, L., Medina, M., Vallejo, A. R., & Prochanow, A. (2011). Peruvian genetic structure and their impact in the identification of Andean missing persons: A perspective from Ayacucho. *Forensic Science International: Genetics Supplement Series*, 3(1), e291–e292.

Iannacone, G., & Parra, R. C. (2020). Genetic structure and kinship analysis from the Peruvian Andean area. In *Forensic Science and Humanitarian Action* (pp. 473–489).

INEI. (2018a). *Amazonas. Resultados definitivos*.

INEI. (2018b). *III Censo de comunidades nativas* 2017. *Resultados definitivos. Tomo I*.

Isbell, W. H. (1968). New discoveries in the montaña of southern Peru. *Archaeology*, 21(2), 108–114.

Jakobsen, J., Jorgensen, J. B., Jorgensen, L. K., & Schjellerup, I. (1987). Cazadores de cabezas en sitios pre-inca de Chachapoyas, Amazonas. *Revista Del Museo Nacional*, 48, 139–185.

Jeffreys, A. J., Brookfield, J. F. Y., & Semenoff, R. (1985). Positive identification of an immigration test-case using human DNA fingerprints. *Nature*, 317(6040), 818–819.

Jobling, M. A., Hurles, M. E., & Tyler-Smith, C. (2004). *Human evolutionary genetics: Origins, peoples and disease*. Garland Science.

Jorde, L. B., Watkins, W. S., Bamshad, M. J., Dixon, M. E., Ricker, C. E., Seielstad, M. T., & Batzer, M. A. (2000). The distribution of human genetic diversity: A comparison of mitochondrial, autosomal, and Y-chromosome data. *American Journal of Human Genetics*, 66(3), 979–988.

Kanezaki, Y., Omori, T., & Tsurumi, E. (2021). Emergence and Development of Pottery in the Andean Early Formative Period: New Insights from an Improved Wairajirca Pottery Chronology at the Jancao Site in the Huanuco Region, Peru. *Latin American Antiquity*, 32(2), 239–254.

Karmin, M., Saag, L., Vicente, M., Sayres, M. A. W., Jarve, M., Talas, U. G., Rootsi, S., Ilumae, A. M., Magi, R., Mitt, M., Pagani, L., Puurand, T., Faltyskova, Z., Clemente, F., Cardona, A., Metspalu, E., Sahakyan, H., Yunusbayev, B., Hudjashov, G., … Kivisild, T. (2015). A recent bottleneck of Y chromosome diversity coincides with a global change in culture. *Genome Research*, 25(4), 459–466.

Kauffmann, F. (Ed.). (2013). *Los Chachapoyas*. Banco de Crédito de Perú.

Kauffmann, F., & Ligabue, G. (2003). *Los Chachapoya(s): moradores ancestrales de los Andes amazónicos peruanos*. Universidad Alas Peruanas.

Kekkonen, J., Wikström, M., & Brommer, J. E. (2012). Heterozygosity in an Isolated Population of a Large Mammal Founded by Four Individuals Is Predicted by an Individual-Based Genetic Model. *Plos One*, 7(9), e43482.

Kimura, M. (1968). Genetic variability maintained in a finite population due to mutational production of neutral and nearly neutral isoalleles. *Genetics Research*, 11(3), 247–269.

Kingman, J. F. C. (1982). The coalescent. *Stochastic Processes and Their Applications*, 13(3), 235–248.

Kitchen, A., Miyamoto, M. M., & Mulligan, C. J. (2008). A Three-Stage Colonization Model for the Peopling of the Americas. *Plos One*, 3(2), e1596.

Koch, A., Brierley, C., Maslin, M. M., & Lewis, S. L. (2019). Earth system impacts of the European arrival and Great Dying in the Americas after 1492. *Quaternary Science Reviews*, 207, 13–36.

Koschmieder, K. (2012). *Jucusbamba: investigaciones arqueológicas y motivos Chachapoya en el norte de la provincia de Luya, departamento Amazonas, Perú*. Gobierno regional Amazonas - DeutscheForschungsgemeinschart - Ministerio de agricultura y Sierra Norte.

Koschmieder, K. (2014a). Asentamientos Chachapoya en el norte de la provincia de luya, departamento de Amazonas. *Arqueología y Sociedad*, 28, 71–114.

Koschmieder, K. (2014b). Los orígenes y el desarrollo de la organización socio-política de la cultura chachapoya: una mirada desde la provincia de Luya, Departamento Amazonas, Perú. In S. Rostain (Ed.), *Antes de Orellana- Actas del 3er encuentro internacional de arqueología amazónica* (Issue 37, pp. 243-249,528-529). Instituto Francés de Estudios Andinos.

Landsteiner, K., & Levine, P. (1928). On individual differences in human blood. *The Journal of Experimental Medicine*, 47(5), 757–775.

Lathrap, D. W. (1970). *The Upper Amazon*. Thames and Hudson.

Lawson, D. J., van Dorp, L., & Falush, D. (2018). A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. *Nature Communications*, 9(1), 3258.

Ledo, A., Condes, S., & Alberdi, I. (2012). Forest biodiversity assessment in Peruvian Andean Montane cloud forest. *Journal of Mountain Science*, 9(3), 372–384.

Lerche, P. (1995). *Los Chachapoya y los símbolos de su historia*. Ediciones y Servicios Gráficos César Gayoso.

Lewis, J. P., Ryves, D. B., Rasmussen, P., Olsen, J., van der Sluis, L. G., Reimer, P. J., Knudsen, K. L., McGowan, S., Anderson, N. J., & Juggins, S. (2020). Marine resource abundance drove pre-agricultural population increase in Stone Age Scandinavia. *Nature Communications*, 11(1).

Llamas, B., Fehren-Schmitz, L., Valverde, G., Soubrier, J., Mallick, S., Rohland, N., Nordenfelt, S., Valdiosera, C., Richards, S. M., Rohrlach, A., Romero, M. I. B., Espinoza, I. F., Cagigao, E. T., Jimenez, L. W., Makowski, K., Reyna, I. S. L., Lory, J. M., Torrez, J. A. B., Rivera, M. A., ... Haak, W. (2016). Ancient mitochondrial DNA provides high-resolution time scale of the peopling of the Americas. *Science Advances*, 2(4), e1501385–e1501385.

Loog, L. (2021). Sometimes hidden but always there: the assumptions underlying genetic inference of demographic histories. *Philosophical Transactions of the Royal Society - Biological Sciences*, 376(1816).

Loreau, M., & de Mazancourt, C. (2013). Biodiversity and ecosystem stability: a synthesis of underlying mechanisms. *Ecology Letters*, 16, 106–115.

Lothrop, S. K. (1961). Early Migrations to Central and South America: An Anthropological Problem in the Light of Other Sciences. *The Journal of the Royal Anthropological Institute of Great Britain and Ireland*, 91(1), 97–123.

Lucas, G. (1979). *Raiders of the Lost Ark* (p. 105). Medway Productions, Inc.

Mandryk, C. A. S., Josenhans, H., Fedje, D. W., & Mathewes, R. W. (2001). Late Quaternary paleoenvironments of Northwestern North America: implications for inland versus coastal migration routes. *Quaternary Science Reviews*, 20(1–3), 301–314.

Mannheim, B., Davis, A. R., & Velasco, M. C. (2018). Cranial Modification in the Central Andes. Person, Language, Political Economy. In V. Tiesler & M. C. Lozada (Eds.), *Social Skins of the Head Body Beliefs and Ritual in*

Ancient Mesoamerica and the Andes (pp. 223–235). University of New Mexico Press.

Matthews-Bird, F., Valencia, B. G., Church, W., Peterson, L. C., & Bush, M. (2017). A 2000-year history of disturbance and recovery at a sacred site in Peru's northeastern cloud forest. *Holocene*, 27(11), 1707–1719.

McCray, B. (2017). Límites e interfaces de Chachapoyas en la región Rodríguez de Mendoza. *Boletín de Arqueología PUCP*, 23, 187–206.

McVean, G. (2009). A Genealogical Interpretation of Principal Components Analysis. *Plos Genetics*, 5(10).

Mogrovejo, T. A. de. (1921). Diario de la segunda visita pastoral que hizo de su arquediocesis el Ilustrísimo Señor Don Toribio Alonso de Mogrovejo, arzobispo de los Reyes. *Revista Del Archivo Nacional Del Perú*, II(1), 37–78.

Mulligan, C. J., Kitchen, A., & Miyamoto, M. M. (2008). Updated Three-Stage Model for the Peopling of the Americas. *Plos One*, 3(9).

Mumford, J. (2017). La reducción toledana en el Perú y el Alto Perú, 1569-1575. In A. Saito & C. Rosas (Eds.), *Reducciones La concentración forzada de las poblaciones indígenas en el Virreinato del Perú* (Pontificia, pp. 67–102).

Myers, T. P. (1976). Isolation and ceramic change: A case from the Ucayali River, Peru. *World Archaeology*, 7(3), 333–351.

Nachman, M. W., & Crowell, S. L. (2000). Estimate of the mutation rate per nucleotide in humans. *Genetics*, 156(1), 297–304.

Nakatsuka, N., Lazaridis, I., Barbieri, C., Skoglund, P., Rohland, N., Mallick, S., Posth, C., Harkins-Kinkaid, K., Ferry, M., Harney, E., Michel, M., Stewardson, K., Novak-Forst, J., Capriles, J. M., Durruty, M. A., Alvarez, K. A., Beresford-Jones, D., Burger, R., Cadwallader, L., … Fehren-Schmitz, L. (2020). A Paleogenomic Reconstruction of the Deep Population History of the Andes. *Cell*, 181(5), 1131-1145.

Narváez, A. (2013). Kuélap: centro del poder político religioso de los Chachapoyas. In *Los Chachapoyas* (pp. 87–159). Banco de Crédito.

Novembre, J., Johnson, T., Bryc, K., Kutalik, Z., Boyko, A. R., Auton, A., Indap, A., King, K. S., Bergmann, S., Nelson, M. R., Stephens, M., & Bustamante, C. D. (2008). Genes mirror geography within Europe. *Nature*, 456(7218), 98-U5.

Novroski, N. M. M., King, J. L., Churchill, J. D., Seah, L. H., & Budowle, B. (2016). Characterization of genetic sequence variation of 58 STR loci in four major population groups. *Forensic Science International: Genetics*, 25, 214–226.

Nystrom, K. C. (2006). Late Chachapoya population structure prior, to Inka conquest. *American Journal of Physical Anthropology*, 131(3), 334–342.

Palo, J. U., Ulmanen, I., Lukka, M., Ellonen, P., & Sajantila, A. (2009). Genetic markers and population history: Finland revisited. *Eur J Hum Genet*, 17(10), 1336–1346.

Parson, W., Brandstatter, A., Alonso, A., Brandt, N., Brinkmann, B., Carracedo, A., Corach, D., Froment, O., Furac, I., Grzybowski, T., Hedberg, K., Keyser-Tracqui, C., Kupiec, T., Lutz-Bonengel, S., Mevag, B., Ploski, R., Schmitter, H., Schneider, P., Syndercombe-Court, D., ... Scheithauer, R. (2004). The EDNAP mitochondrial DNA population database (EMPOP) collaborative exercises: organisation, results and perspectives. *Forensic Science International*, 139(2–3), 215–226.

Parson, W., & Dür, A. (2007). EMPOP—A forensic mtDNA database. *Forensic Science International: Genetics*, 1(2), 88–92.

Parson, W., Gusmão, L., Hares, D. R., Irwin, J. A., Mayr, W. R., Morling, N., Pokorak, E., Prinz, M., Salas, A., Schneider, P. M., & Parsons, T. J. (2014). DNA Commission of the International Society for Forensic Genetics: Revised and extended guidelines for mitochondrial DNA typing. *Forensic Science International: Genetics*, 13, 134–142.

Pereira, L., Prata, M. J., & Amorim, A. (2002). Mismatch distribution analysis of Y-STR haplotypes as a tool for the evaluation of identity-by-state proportions and significance of matches—the European picture. *Forensic Science International*, 130(2–3), 147–155.

Pérez, I. (2019). El estado regional Huarpa y los orígenes del imperio Wari. ALTERITAS, *Revista de Estudios Socioculturales Andino Amazónicos*, 8(9), 181–221.

Pinotti, T., Bergström, A., Geppert, M., Bawn, M., Ohasi, D., Shi, W., Lacerda, D. R., Solli, A., Norstedt, J., Reed, K., Dawtry, K., González-Andrade, F., Paz-y-Miño, C., Revollo, S., Cuellar, C., Jota, M. S., Santos, J. E., Ayub, Q., Kivisild, T., ... Tyler-Smith, C. (2019). Y Chromosome Sequences Reveal a Short Beringian Standstill, Rapid Expansion, and early Population structure of Native American Founders. *Current Biology*, 29(1), 149-157. e3.

Posth, C., Nakatsuka, N., Lazaridis, I., Skoglund, P., Mallick, S., Lamnidis, T. C., Rohland, N., Nägele, K., Adamski, N., Bertolini, E., Broomandkhoshbacht, N., Cooper, A., Culleton, B. J., Ferraz, T., Ferry, M., Furtwängler, A., Haak, W., Harkins, K., Harper, T. K., ... Reich, D. (2018). Reconstructing the Deep Population History of Central and South America. *Cell*, 175(5), 1185-1197.e22.

Pritchard, J., & Przeworski, M. (2001). Linkage disequilibrium in humans: Models and data. *American Journal of Human Genetics*, 69(1), 1–14.

Pritchard, J., Stephens, M., & Donnelly, P. (2000). *Inference of Population Structure Using Multilocus Genotype Data*. 155(2), 945–959.

Purps, J., Siegert, S., Willuweit, S., Nagy, M., Alves, C., Salazar, R., Angustia, S. M. T., Santos, L. H., Anslinger, K., Bayer, B., Ayub, Q., Wei, W., Xue, Y., Tyler-Smith, C., Bafalluy, M. B., Martínez-Jarreta, B., Egyed, B., Balitzki, B., Tschumi, S., ... Roewer, L. (2014). A global analysis of Y-chromosomal haplotype diversity for 23 STR loci. *Forensic Science International: Genetics*, 12(0), 12–23.

Quinque, D., Kittler, R., Kayser, M., Stoneking, M., & Nasidze, I. (2006). Evaluation of saliva as a source of human DNA for population and association studies. *Analytical Biochemistry*, 353(2), 272–277.

Raghavan, M., Steinrücken, M., Harris, K., Schiffels, S., Rasmussen, S., DeGiorgio, M., Albrechtsen, A., Valdiosera, C., Ávila-Arcos, M. C., Malaspinas, A.-S., Eriksson, A., Moltke, I., Metspalu, M., Homburger, J. R., Wall, J., Cornejo, O. E., Moreno-Mayar, J. V., Korneliussen, T. S., Pierre, T., … Willerslev, E. (2015). Genomic evidence for the Pleistocene and recent population history of Native Americans. *Science*.

Ramachandran, S., Rosenberg, N. A., Zhivotovsky, L. A., & Feldman, M. W. (2004). Robustness of the inference of human population structure: a comparison of X-chromosomal and autosomal microsatellites. *Human Genomics*, 1(2), 87–97.

Rangel, T. F., Edwards, N. R., Holden, P. B., Diniz, J. A. F., Gosling, W. D., Coelho, M. T. P., Cassemiro, F. A. S., Rahbek, C., & Colwell, R. K. (2018). Modeling the ecology and evolution of biodiversity: Biogeographical cradles, museums, and graves. *Science*, 361(6399).

Reich, D., Patterson, N., Campbell, D., Tandon, A., Mazieres, S., Ray, N., Parra, M. V, Rojas, W., Duque, C., Mesa, N., Garcia, L. F., Triana, O., Blair, S., Maestre, A., Dib, J. C., Bravi, C. M., Bailliet, G., Corach, D., Hunemeier, T., … Ruiz-Linares, A. (2012). Reconstructing Native American population history. *Nature*, 488(7411), 370–374.

Relethford, J. H. (2012). *Human Population Genetics* (Wiley-Blackwell (Ed.)). John Wiley & Sons, Inc.

Riris, P., & Arroyo-Kalin, M. (2019). Widespread population decline in South America correlates with mid-Holocene climate change. *Scientific Reports*, 9.

Roe, P. G. (1976). Archaism, form and decoration: an ethnographic and archaeogical case study from the Peruvian montaña. *Ñawpa Pacha*, 14(1), 73–94.

Roewer, L. (2013). DNA fingerprinting in forensics: past, present, future. *Investigative Genetics*, 4(1), 22.

Roewer, L. (2019). Y-chromosome short tandem repeats in forensics—Sexing, profiling, and matching male DNA. *WIREs Forensic Science* , 1(4), e1336.

Rogers, A. (2014). How Population Growth Affects Linkage Disequilibrium. *Genetics*, 197(4), 1329–1341.

Rogers, A., Fraley, A. E., Bamshad, M. J., Watkins, W. S., & Jorde, L. B. (1996). Mitochondrial mismatch analysis is insensitive to the mutational process. *Molecular Biology and Evolution*, 13(7), 895–902.

Rogers, A., & Harpending, H. (1992). Population Growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution*, 9(3), 552–569.

Rojas-Berscia, L. M. (2020). The Chachapuya Language and Proto-Kawapanan: Lexical Affinities and Hypothetical Contact Scenarios. *Indiana*, 37(1), 155–188.

Rojas, C., & VanValkenburgh, P. (2020). P*aisajes Arqueológicos de Choctamal - Chachapoyas – Fase I, Temporada* 3 *– Excavaciones en Purun Llaqta del Maino - (Informe Técnico).*

Romanini, C., Romero, M., Puerto, M. S., Catelli, L., Phillips, C., Pereira, R., Gusmao, L., & Vullo, C. (2015). Ancestry informative markers: Inference of ancestry in aged bone samples using an autosomal AIM-Indel multiplex. *Forensic Science International Genetics*, 16, 58–63.

Rosenberg, N. A., Pritchard, J. K., Weber, J. L., Cann, H. M., Kidd, K. K., Zhivotovsky, L. A., & Feldman, M. W. (2002). Genetic structure of human populations. *Science*, 298(5602), 2381–2385.

Ruiz, A. (1972). La alfareria de Cuelap: tradición y cambio. In Social Sciences: *Vol. Bachelor*. Universidad Nacional de San Marcos.

Ruiz, A. (2008). Las cavernas y el poblamiento prehispánicode la provincia de Chachapoyas. *Investigaciones Sociales*, XII(20), 35–62.

Sandweiss, D. H., Maasch, K. A., Burger, R. L., Richardson, J. B., & Rollins, H. B. (2001). Variation in Holocene El Niño frequencies: Climate records and cultural consequences in ancient Peru. *Geology*, 29(7), 603–606.

Sauer, C. O. (1944). A Geographic Sketch of Early Man in America. *Geographical Review*, 34(4), 529–573.

Savoy, D. (1996). *The Gran Vilaya Expeditions: Reclaiming a vast Urban Center from the Upper Montane Forest of Amazonas, Peru.*

Schaffner, S. F. (2004). The X chromosome in population genetics. *Nature Reviews Genetics*, 5(1), 43–51.

Schjellerup, I. (2005). *Incas y españoles en la conquista de los chachapoya.* Fondo Editorial de la Pontificia Universidad Católica del Perú and Instituto Francés de Estudios Andinos.

Schjellerup, I. (2008). Sacando a los caciques de la oscuridad del olvido. Etnias chachapoya y chilcho. *Bulletin de l'Institut Français d'Études Andines*, 37(1), 111–122.

Schjellerup, I. (2017). La provincia Inka de Chachapoyas. *Boletín de Arqueología PUCP*, 23, 259–281.

Schneider, S., & Excoffier, L. (1999). Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates very among sites: Application to human mitochondrial DNA. *Genetics*, 152(3), 1079–1089.

Serre, D., & Pääbo, S. P. (2004). Evidence for gradients of human genetic diversity within and among continents. *Genome Research*, 14(9), 1679–1685.

Shady, R. (1987). Tradición y cambio en las sociedades formativas de Bagua, Amazonas, Perú. Revista Andina. *Revista Andina*, 10(2), 457–487.

Silva, W. A., Bonatto, S. L., Holanda, A. J., Ribeiro-dos-Santos, A. K., Paixao, B. M., Goldman, G. H., Abe-Sandes, K., Rodriguez-Delfin, L., Barbosa, M., Paco-Larson, M. L., Petzl-Erler, M. L., Valente, V., Santos, S. E. B., & Zago, M. A. (2002). Mitochondrial genome diversity of Native Americans supports a single early entry of founder populations into America. *American Journal of Human Genetics*, 71(1), 187–192.

Skoglund, P., Mallick, S., Bortolini, M. C., Chennagiri, N., Hunemeier, T., Petzl-Erler, M. L., Salzano, F. M., Patterson, N., & Reich, D. (2015). Genetic evidence for two founding populations of the Americas. *Nature*, 525(7567), 104–108.

Slatkin, M. (2008). Linkage disequilibrium - understanding the evolutionary past and mapping the medical future. *Nature Reviews Genetics*, 9(6), 477–485.

Slatkin, M., & Hudson, R. R. (1991). Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics*, 129(2), 555–562.

Stavenhagen, R., & Monge, E. (2009). *Perú Bagua. Bloodshed in the context of the Amazon protest. Urgent need for good faith dialogue.*

Stetson, G. (2012). Oil Politics and Indigenous Resistance in the Peruvian Amazon: The Rhetoric of Modernity Against the Reality of Coloniality. *The Journal of Environment & Development*, 21(1), 76–97.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, 123(3), 585–595.

Tallavaara, M., Eronen, J. T., & Luoto, M. (2018). Productivity, biodiversity, and pathogens influence the global hunter-gatherer population density. *Proceedings of the National Academy of Sciences of the United States of America*, 115(6), 1232–1237.

Tamm, E., Kivisild, T., Reidla, M., Metspalu, M., Smith, D. G., Mulligan, C. J., Bravi, C. M., Rickards, O., Martinez-Labarga, C., Khusnutdinova, E. K., Fedorova, S. A., Golubenko, M. V, Stepanov, V. A., Gubina, M. A., Zhadanov, S. I., Ossipova, L. P., Damba, L., Voevoda, M. I., Dipierri, J. E., ... Malhi, R. S. (2007). Beringian Standstill and Spread of Native American Founders. *Plos One*, 2(9), 1–6.

Tarazona-Santos, E., Carvalho-Silva, D. R., Pettener, D., Luiselli, D., De Stefano, G. F., Labarga, C. M., Rickards, O., Tyler-Smith, C., Pena, S. D. J., & Santos, F. R. (2001). Genetic differentiation in South Amerindians is related to environmental and cultural diversity: Evidence from the Y chromosome. *American Journal of Human Genetics*, 68(6), 1485–1496.

Taylor, G. (2000). *Estudios Lingüísticos sobre Chachapoyas*. IFEA-UNMSM.

Taylor, G. (1990). La lengua de los antiguos Chachapuyas. In R. Cerrón-Palomino & G. Solís (Eds.), *Temas de lingüística amerindia. Primer Congreso Nacional de Investigaciones Lingüístico Filológicas* (pp. 121–139). ConsejoNacional de Ciencia y Tecnología/ Programa de Educación Bilingüe de Puno.

Torero, A. (2005). *Idiomas de los Andes. Lingüística e Historia* (Second edi). Editorial Horizonte.

Torroni, A., Schurr, T. G., Yang, C. C., Szathmary, E. J. E., Williams, R. C., Schanfield, M. S., Troup, G. A., Knowler, W. C., Lawrence, D. N., Weiss,

K. M., & Wallace, D. C. (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.

Torroni, A., Sukernik, R. I., Schurr, T. G., Starikovskaya, Y. B., Cabell, M. F., Crawford, M. H., Comuzzie, A. G., & Wallace, D. C. (1993). MtDNA variation of aboriginal Siberians reveals distinct genetic affinities with Native-Americans. *American Journal of Human Genetics*, 53(3), 591–608.

Toyne, J. M., Church, W. B., Luis Coronado Tello, J., & Morales Gamarra, R. (2017). Exploring imperial expansion using an isotopic analysis of paleodietary and paleomobility indicators in Chachapoyas, Peru. *American Journal of Physical Anthropology*, 162, 52–72.

Toyne, J. M., Michell, S., Denierio, M., & Vargas, L. A. N. (2020). Examining pre-Hispanic human and animal isotopic variation at Kuelap in the eastern montane of Peru. *Quaternary International*, 548, 57–70.

Tsetsos, F., Drineas, P., & Paschou, P. (2019). Genetics and Population Analysis. In S. Ranganathan, M. Gribskov, K. Nakai, & C. Schönbach (Eds.), *Encyclopedia of Bioinformatics and Computational Biology* (pp. 363–378). Academic Press.

Ungurahui-IWGIA, O.-R. de. (2010). *Peru: a chronicle of deception. Report 5. Attempts to transfer the Awajún Border Territory in the Cordillera del Cóndor to the Mining Industry* (A. Parellada (Ed.)). Organization for the development of the border communities (ODECOFROC). Racimos de Ungurahui working group. International Work group for Indigenous Affairs (IWGIA).

Vallejo, F. (2004). El estilo Ychsma: características generales, secuencia y distribución geográfica. *Bulletin de l'Institut Français d'Études Andines*, 33(3), 595–642.

Valqui, J., & Ziemendorff, M. (2016). Vestiges of a native language in the territory of the Chachapoyas culture. *Letras*, 87(125), 5–32.

van Oven, M., & Kayser, M. (2009). Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Human Mutation*, 30(2), E386–E394.

Wang, S., Lewis Jr., C. M., Jakobsson, M., Ramachandran, S., Ray, N., Bedoya, G., Rojas, W., Parra, M. V, Molina, J. A., Gallo, C., Mazzotti, G., Poletti, G., Hill, K., Hurtado, A. M., Labuda, D., Klitz, W., Barrantes, R., Bortolini, M. C., Salzano, F. M., ... Ruiz-Linares, A. (2007). Genetic variation and population structure in Native Americans. *Plos Genetics*, 3(11), 2049–2067.

Wanner, H., Beer, J., Butikofer, J., Crowley, T. J., Cubasch, U., Fluckiger, J., Goosse, H., Grosjean, M., Joos, F., Kaplan, J. O., Kuttel, M., Muller, S. A., Prentice, I. C., Solomina, O., Stocker, T. F., Tarasov, P., Wagner, M., & Widmann, M. (2008). Mid- to Late Holocene climate change: an overview. *Quaternary Science Reviews*, 27(19–20), 1791–1828.

Wendt, F., Churchill, J. D., Novroski, N. M. M., King, J. L., Ng, J., Oldt, R. F., McCulloh, K. L., Weise, J. A., Smith, D. G., Kanthaswamy, S., & Budowle, B. (2016). Genetic analysis of the Yavapai Native Americans from West-Central Arizona using the Illumina MiSeq FGxTM forensic genomics system. *Forensic Science International: Genetics*, 24, 18–23.

Wendt, F., King, J. L., Novroski, N. M. M., Churchill, J. D., Ng, J., Oldt, R. F., McCulloh, K. L., Weise, J. A., Smith, D. G., Kanthaswamy, S., & Budowle, B. (2017). Flanking region variation of ForenSeq (TM) DNA Signature Prep Kit STR and SNP loci in Yavapai Native Americans. *Forensic Science International-Genetics*, 28, 146–154.

Wilder, J. A., Kingan, S. B., Mobasher, Z., Pilkington, M. M., & Hammer, M. F. (2004). Global patterns of human mitochondrial DNA and Y-chromosome structure are not influenced by higher migration rates of females versus males. *Nature Genetics*, 36(10), 1122–1125.

Willuweit, S., & Roewer, L. (2007). Y chromosome haplotype reference database (YHRD): Update. *Forensic Science International: Genetics*, 1(2), 83–87.

Willuweit, S., & Roewer, L. (2015). The new Y Chromosome Haplotype Reference Database. *Forensic Science International Genetics*, 15, 43–48.

Witherspoon, D. J., Wooding, S., Rogers, A., Marchani, E. E., Watkins, W. S., Batzer, M. A., & Jorde, L. B. (2007). Genetic similarities within and between human populations. *Genetics*, 176(1), 351–359.

Wright, S. (1951). The genetical structure of populations. *Annals of Eugenics*, 15(4), 323–354.

Young, K. R. (1992). Biogeography of the montane forest zone of the eastern slopes of Peru. *Memorias Del Museo de Historia Natural UNMSM*, 21, 119–154.

Yuan, X. G., Miller, D. J., Zhang, J. Y., Herrington, D., & Wang, Y. (2012). An Overview of Population Genetic Data Simulation. *Journal of Computational Biology*, 19(1), 42–54.

Zevallos, J. (1982). Onomástica prehispánica de Chachapoyas. *Investigación Arqueológica*.

Zevallos, J. (1995). El área geográfico-cultural de la prehistoria de Chachapoyas. Una nueva postulación. *Gaceta Arqueológica Andina*, 24, 13–23.