**DISTRIBUTION AGREEMENT**

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Sophie K. Joseph Date

Paleogenomic Analyses of Demographic Change and Adaptation in the Evolutionary Histories of Human Populations in the Andes

By

Sophie K. Joseph
Doctor of Philosophy

Anthropology

_____

John Lindo, JD, PhD
Advisor

_____

Marcela E. Benítez, PhD
Committee Member

_____

Carmen J. Marsit, PhD
Committee Member

Accepted:

_____

Kimberly Jacob Arriola, PhD, MPH
Dean, James T. Laney School of Graduate Studies

_____

Date

# Paleogenomic Analyses of Demographic Change and Adaptation in the Evolutionary Histories of Human Populations in the Andes

By

Sophie K. Joseph
MA, Anthropology

Advisor: John Lindo, JD, PhD

An abstract of a dissertation
submitted to the Faculty of the James T.
Laney School of Graduate Studies of Emory University,
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Anthropology

2025

## ABSTRACT

Paleogenomic Analyses of Demographic Change and Adaptation in the Evolutionary Histories of Human Populations in the Andes

By Sophie K. Joseph

Studies of Indigenous ancestry and lifeways have remained underrepresented in the field of population genomics. Accordingly, the objective of this dissertation was to gain a better understanding of the ancient history of the Americas, particularly in the Central Andes. Studies employing whole genomes, both ancient and modern, can access a wealth of ancestral information coded within the genome. Through population genomic research and collaboration, this dissertation investigated the genetic history of the Central Andes region of South America, in what is today Peru and Ecuador, over periods including the transition from hunter-gatherer economies to agricultural civilizations. This dissertation comprises three published, peer-reviewed papers: (1) A review of current understandings of pathogenic disease as an evolutionary pressure in the Americas, (2) Tuberculosis immunity and high altitude adaptation in two modern-day Ecuadorian populations living at high altitude, and (3) Paleogenomic insights into cooperation in the ancient Andes from positive selection on oxytocin pathway genes. Collectively, this work suggests different population dynamics exist across the Central Andes, necessitating nuanced analyses of fine-scale regional patterns of demographic change and adaptation, as well as collaboration with local populations and researchers to drive increased representation of the peoples living across this diverse region.

# Paleogenomic Analyses of Demographic Change and Adaptation in the Evolutionary Histories of Human Populations in the Andes

By

Sophie K. Joseph
MA, Anthropology

Advisor: John Lindo, JD, PhD

A dissertation
submitted to the Faculty of the James T.
Laney School of Graduate Studies of Emory University,
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Anthropology

2025

## ACKNOWLEDGEMENTS

First, I acknowledge and extend my gratitude towards the Indigenous communities across the Central Andes who trusted myself and the Lindo Lab to collaborate with them on the projects involved in this dissertation, from developing their research questions to contributing biological samples and/or skeletal material from their ancestors. I also extend my gratitude to the many institutions and collaborators that made the research presented in this dissertation possible, including U.S. and international colleagues at the University of Wyoming, University of California-Merced, University of California-Davis, Georgia State University, Central University of Ecuador (Quito, Ecuador), University of Pavia (Pavia, Italy), Universidad Técnica de Manabí (Manabí, Ecuador), and the Collasuyo Archaeological Research Institute (Puno, Peru).

I would like to express my deepest appreciation to my committee for their invaluable guidance and insightful suggestions that greatly improved the work presented in this dissertation. I extend my sincerest thanks and appreciation to my advisor John Lindo. His expertise was crucial in developing the methodology and overall approach to my research, and he was also a fervent advocate, supporter, and career mentor throughout my graduate career at Emory University. I am also grateful to Marcela Benítez, a tremendous teacher, mentor, and collaborator, and Carmen Marsit, a wonderful mentor and supporter of my future career aspirations.

Thank you to Emory University's Department of Anthropology staff, faculty members, and graduate student body for their guidance, support, and friendship. It has been an honor to share this department with you.

Thank you to the eleven undergraduate mentees who have served as my research assistants during my five years at Emory, many of whom became my collaborators, co-authors, and friends. Being a stepping stone on your path to success is inspiring, and has made me a better mentor and scientist.

Thank you to my undergraduate mentors who encouraged and prepared me to pursue my graduate education at Emory: Drs. Steve Davis, Keely Muscatell, Monica Gaudier-Diaz, Katie McKenna, and Malcolm Brenner. I would never have had the opportunity to do this work if it were not for your support and introduction to research as a career.

Thank you to my friends outside of Emory, whose laughter, encouragement, charcuterie boards, camping trips, and access to your furry friends kept me sane.

Thank you to my extended family and my boyfriend. You are always my biggest cheerleaders. Your love and support mean the world.

**DEDICATION**

*To Mom and Dad, for your unconditional love and support, for fostering my love of science, and for always telling me to reach for the stars.*

**CONTENTS**

# TABLES AND FIGURES

## CHAPTER 2



## CHAPTER 3



## CHAPTER 4

# CHAPTER 1: INTRODUCTION

## *1.1 Aims of the Dissertation*

Studies of Indigenous ancestry and lifeways have remained underrepresented in the field of population genomics. Accordingly, the objective of this dissertation was to gain a better understanding of the ancient history of human settlement in the Americas, particularly in the Central Andes region of South America, by exploring the evolutionary histories of Indigenous peoples of the Andes using high-resolution whole genomes. This work comprises three published, peer-reviewed papers: (1) A review of current understandings of pathogenic disease as an evolutionary pressure in the Americas, (2) Tuberculosis immunity and high altitude adaptation in two modern-day Ecuadorian populations living at high altitude, and (3) Paleogenomic insights into cooperation in the ancient Andes from positive selection on oxytocin pathway genes. The unifying research theme of this dissertation was to investigate the genetic history of the Central Andes region of South America, in what is today Peru and Ecuador, with particular attention to understanding selective pressures from different parts of the environment, most notably the social, microbial, and physical.

## *1.2 13,000 Years of Human History in South America*

The Andean archeological record spans over 13,000 years of human occupation, from initial human settlement and hunter-gatherer economies to agro-pastoralism and urbanization[1]. These societal changes brought about numerous cultural and biological adaptations, and are a valuable framework for situating the population histories of Andean peoples in historical context. Here, I briefly summarize the major archaeological periods that represent large-scale changes in human subsistence throughout the history of the Central Andes[2,3]:

The Archaic Period: During the Archaic Period (13–5 thousand years ago (ka)), lowland populations slowly adapted to the low oxygen climates of the high mountains and exploited these highland resources seasonally[4]. There were many adaptive challenges as they transitioned to permanent highland occupation, many of which were alleviated by developing animal hide products for clothing and shelter. Archaic populations were mobile, practicing a mix of foraging and hunting in order to contend with the unpredictable highland environments[5]. In the Terminal Archaic (5.0–3.5 ka), foraging economies transitioned into agriculture and more permanent settlements[6]. Wool textile technology also emerged, promoting specialized craft production and the exchange of these goods. Temple architecture[7] and the first archery technology[8] to improve hunting also appeared during the later part of this period. The domestication of plants and animals in the Andes during the Terminal Archaic occurred independently from agriculture in other parts of the world, such as the Levant.

The Formative and Early Intermediate Periods: During the early Formative Period (3.5–1.9 ka), agricultural economies were fairly ubiquitous, and grew in complexity and size[9]. People exclusively lived in permanent villages year-round, close to their agricultural lands. During the Upper Formative Period (2.5–1.9 ka) the urban center of Pukara emerged in the Titicaca Basin of present-day southern Peru and northern Bolivia. They had specialized ceramic production and centralized obsidian exchange, indicating the presence of a complex economy[10]. And along the present-day north coast of Peru and Ecuador, the contemporaneous Moche culture also flourished. Like at Pukara, Moche society consisted of economically connected cities, without a centralized political state. Moche ceramics are among the most diverse and complex in the Andean archeological record, with several distinct and recognizable phases. Fine-grained analyses of the

form and painted designs of Moche pottery have provided glimpses into the ritual purpose of Moche *huacas*, places of religious importance which served ceremonial purposes[11,12].

The Middle Horizon: During the Middle Horizon (1.4–1.0 ka), the population center and political power shifted from Pukara to Tiwanaku in the southern Titicaca Basin. The urban area of Tiwanaku city had thousands of inhabitants with complex religious architecture, textile craft production, and innovations in agricultural techniques[3,13]. Tiwanaku was also home to migrant groups from far away regions, a phenomenon not commonly seen in earlier periods[14]. To the north of Tiwanaku, the Wari state grew along the north coast in the former Moche territory in present-day northern Peru and Ecuador, and eventually expanded south into the highlands. At the height of their territory, the Wari state actually bordered Tiwanaku. Though they were both politically powerful, with centralized governance, taxation, and state wealth[15], evidence from both ancient fort structures and bioarcheological evidence from human remains indicate the two states were not engaged in violent conflict with eachother[16].

The Late Intermediate Period and Late Horizon: Around 1.5 ka, the Tiwanaku and Wari empires declined in power and influence in the region, marking the beginning of the Late Intermediate Period (1.0–0.6 ka). This post-empire time was a period of political fragmentation, as well as the only period of prolonged violence in the region's history. Defensive architecture was much more common, and perimortem trauma rates in skeletal remains exceed 35%, over twice that observed from bioarcheological investigation of all other time periods[16]. During the subsequent Late Horizon (0.6–0.5 ka), violence rates then subsided, as the Inca Empire expanded from northern Peru, incorporating the descendants of the Tiwanaku and Wari populations into the Inca political system[3]. The height of Inca political power lasted barely over a century, due to the arrival of Europeans in the 16th century. Their extensive road and trade networks, rapid political

consolidation, and advanced mathematical and astronomical systems were among the most important advancements that contributed to their expansionist political success in western South America[17]. Today, the Quechua language family remains strong in the region, with over 10 million native speakers forming a dialect continuum across the former Inca empire territory. Their ancestors fought to keep their culture and language, despite the forced assimilation to Spanish rule[18]. Presently, the Central Andes region has a flourishing network of local Indigenous archaeologists. Their work not only seeks to gain insight into the lifeways of past peoples of the region, but also to connect present-day realities to their historical roots[18,19].

### *1.3 Understanding Ancient Life Using Genomics*

For millennia, the Andes have been home to a variety of societal structures. These have ranged from mobile hunter-gatherer settlements, to small agricultural villages, to dense urban centers with tens of thousands of inhabitants[1–3,5]. In recent decades, since the advent of next-generation sequencing technology, genetic evidence has revealed answers to questions that archaeology and bioarchaeology can't answer alone. While things like admixture between populations and migration events can be approximated by changes in the archaeological record of material culture, genomics lends additional detail and more nuance to these population-level changes[20–30]. Today, genomic and archaeological insights work in concert to provide the most complete understanding of the population histories of the Central Andes over the 13,000 year record of human occupation in the region[1].

Initial Migrations into South America: There is considerable debate as to the nature of human dispersal into South America—regarding both the number of primary migrations into the continent from North America and the post-entry population splits as the entire continent was first settled approximately 13 ka. Early genomic research on the initial migrations into South America

suggested a single initial founder population entering the continent, with sequential population splits and very little gene flow between them after their initial splits[27]. However, subsequent work suggests this is incorrect; for example, one study of ancient individuals from the Andes found that the modern north-south population genetic structure of Ecuador, Peru, Bolivia, and Chile was fully developed by 5.8 ka[20]. This included bi-directional gene flow between the north and south highlands beginning soon after, as well as some gene flow occurring between the highlands and lowlands. In agreement with the archeological record of material culture, they also detected populations with people of diverse ancestries living side-by-side in the heartland of both the Tiwanaku and Inca territories, suggesting cosmopolitanism in these societies, evidence of post-continental-entry gene flow.

Regarding the initial founder populations in the continent, some research suggests that at least two populations, at two different times, contributed genetic ancestry to South Americans: an initial land dispersal from North America through the isthmus of Panama, and another coastal migration with ancestry more closely related to Indigenous Australians and Melanesians[24,31]. The reason that contemporary Indigenous South Americans today do not typically show a large ancestry component from these populations could be because later admixture[20] potentially reduced the Australasian signature carried by earlier inhabitants[31]. These debates are inconclusive and remain under active investigation, and will be further illuminated as more population genomics work is conducted in the region and tools for statistical analysis improve.

The Peopling of the Andean Highlands: Subsistence in the high-altitude Andes is particularly difficult due to the harsh environment where soil erosion, drought, freezing temperatures, and flooding are frequent[32]. The early settlers of the Andes were mobile hunter-gatherers, who exploited resources on a seasonal basis. Over time, they developed adaptations, both cultural and

physiological, such as warm clothing from animal hide and higher than average $VO_2$ max aerobic capacity[33]. The genetic variant(s) responsible for these physiological changes has yet to be characterized in functional genomics literature. Recent work on identifying alleles under natural selection in Andean high-altitude populations suggests that adaptation to hypoxic conditions via non-coding variants affecting the cardiovascular system may have occurred in Bolivian Aymara and highland Peruvian Quechua populations[26,30]. Work included in this dissertation suggests there are complex and differing modes of adaptation that may have occurred regionally[34]—for example, there may have been convergent evolution of the adaptation to hypoxic conditions with different genes involved in each separate population lineage, or it could be that these physiological adaptations are also epigenetic in nature.

Genetic Insights Into the Advent of Agriculture: The domestication of plants and animals in the Andes during the Terminal Archaic Period was independent from the invention of agriculture in other parts of the world. Alleles related to starch digestion were likely the result of positive selection during this time[29]. For example, selection on loci associated with the *MGAM* starch digestion gene, was coincident at approximately 4 ka with the development of Andean agriculture, and specifically with the domestication of maize and potatoes in the Central Andes. This may represent an adaptive response to greater reliance upon starchy foods like these crops. During this period of agricultural expansion known as the Formative Period (3.5–1.9 ka), agro-pastoral economies exploded in size[9]. Because these societies were more compact and populous, there may have been an increase in the spread of respiratory diseases, such as tuberculosis. There is considerable debate regarding the presence of human tuberculosis in the pre-contact Americas, though a growing body of genetic and bioarchaeological evidence have suggested its presence[34–37]. Work included in this dissertation suggests that alleles involved in the tuberculosis immune

response underwent strong positive selection at over 3 ka during the Upper Formative Period[34]. In summary, genomics and archaeology are both essential for understanding the population history of the Andes throughout the entire 13,000 year period of human settlement in the region. For example, archeological analyses in the Central Andes produced a chronology of lifeways and political systems that is the main temporal framework for both genetic and material understandings of ancient life in the Andes[2,3]. And thanks to population genomics, we now have a better understanding of the population dynamics, migration patterns, and admixture at times of large cultural transitions in the Andes[20–30].

Community-Engaged Research: While contributing to increasing the body of literature on genetic ancestry and adaptation in the Americas, my primary goal has always been practicing responsible and beneficial community-engaged research. This means pursuing research questions that reflect an Indigenous community's questions about their own ancestry. I prioritize treating Indigenous communities as partners in the entire research process of the respectful study of their biological samples and/or ancestral human remains[38]. This can help to alleviate some of the ethical difficulties with research involving Indigenous peoples, ranging from stigmatization from genetic disease associations to the potential for genetic population histories to contradict community origin stories[39]. With each specific project, it is crucial for the careful consultation and formation of relationships with each community to occur long before the research process even begins. Not only does this include seeking input into the kinds of research questions community members are interested in answering, but this also includes making a plan for how to keep collaborating communities involved at every step. Each partner community has different needs, so it is critical to always adapt to their preferences for communication and interpretation. For example, as seen in one component of this dissertation (see Chapter 3.6)[34], our particular collaborating Kichwa

communities in Ecuador preferred to have the study results translated to Kichwa and Spanish, and then to gather together as a community to interpret them themselves. To facilitate this, we worked with our community collaborators/co-authors to fund and facilitate a series of workshop events devoted to contextualizing how this population genetic research can be understood within their Andean cultural spaces.

### *1.4 Population Genomics and Recent Human Evolution*

Drivers of Microevolution in Evolutionary Biology: The modern synthesis of evolutionary biology emerged in the early 20th century to connect Darwin's theory of descent with modification with Mendelian genetic inheritance[40]. In this modern synthesis framework, **macro**evolution (speciation events over long geologic-scale time periods) is the result of accumulation of differences that arise through **micro**evolution (the changing allele frequencies between generations in a population). Population genetics arose a few decades later as a way to mathematically characterize the many processes important in the modern synthesis, such as selection, mutation, and genetic drift, and to understand how attributes of a population may play a role in these processes[41]. Population demography is the study of these attributes, which include things like mortality, fertility rate, population size, and population histories[42]. In turn, these attributes interact with microevolutionary processes like natural selection, genetic drift, mutation, and gene flow between populations. In general, population genetics primarily engages with understanding microevolutionary processes[41]. Because this dissertation investigates the population genomics of humans in the Central Andes during a very short (geologic) timeframe, the term "evolution" will refer specifically to microevolution within this work. While less relevant in this dissertation's context, macroevolution is an important part of the modern synthesis framework, and thus I discuss it briefly here now.

Macroevolution and Punctuated Equilibrium: On the scale of geologic eras and periods, major environmental 'punctuations' in Earth's history seem to correlate with speciation events[43]. Here evolutionary change is not caused by selection, but extinction, where only certain types of taxa survive the drastic environmental changes. In a way, this seems analogous to selection, in that only certain types of organisms are suited to survive long-term in Earth's changing environments. There are numerous examples of this in the paleontological record, from many different organisms that differ even up to the kingdom level of classification. For example, Stanley's[44] seminal paper on large-scale fluctuations in seawater chemistry links the magnesium/calcium ratio in Earth's seawater to the biomineralization strategies of calcareous algae and protozoans. There are clear linkages throughout geologic time between seawater chemistry and what *types* of organisms were the dominant reef builders in Earth's oceans[44]. In a macroevolutionary frame of reference, species can exist and behave as a unit in nature, and from a microevolutionary perspective, the organism and its population are more relevant[45–48]. Neither of these ways of thinking are incorrect, but rather help to theoretically ground methods for examining processes of evolution depending on different research questions.

Are Humans Still Evolving?: There is a debate, both in popular media and to some extent scholarly commentary, as to whether or not humans are 'still evolving', or have been in recent millennia[49]. Due to our ability as a species to adapt to most environments using our technological prowess, some proponents of the not-evolving theory hold that we have essentially outpaced natural selection with our cultural adaptations. However, there will always be different allele frequencies between generations, which is the simple definition of microevolution. This might be caused by different drivers of evolution, whether it be selection, random mutation, or drift, but humans are still subject to some microevolutionary processes. Perhaps this debate also refers to whether or not

humans will undergo macroevolutionary speciation, since people have managed to inhabit nearly every corner of the planet, which in some species would lead to allopatric speciation via geographic isolation. However, humans are highly admixed, with relatively low genetic diversity even compared to other primates, despite our wide geographic range[50]. Thus, it is unlikely we will speciate in the near future, especially as increasing globalization and migration will only continue to promote further gene flow between humans across the planet.

Selective Pressures and "The Environment" in a Microevolutionary Context: Natural selection is one evolutionary process whereby populations change through successive generations. Here, a selective pressure favors a certain phenotype, which often shifts population-level allele frequencies much more quickly and/or to a larger degree than other processes like genetic drift or random mutation would alone. This can include the removal of deleterious mutations through negative selection, or a proliferation of advantageous variants through positive selection[51]. Therefore, many computational techniques for detecting selection rely on detecting genetic loci that lie outside of a predicted baseline structure that already accounts for random processes like genetic drift[51,52]. This dissertation employs a variety of such methods for detecting the presence of natural selection, including examining both allele frequencies and larger regions containing multiple neighboring alleles (haplotype structure).

Population genetics merged the genetic framework of the modern synthesis with population ecology, in order to understand how the many interacting forces surrounding a population drive microevolutionary change between generations. Selective pressures driving natural selection often come from "the environment" surrounding a population, which in population genetics literature most often refers to the physical environment (e.g., climate, food availability, etc.)[41]. However, especially when studying humans who are innovative, social, and often live in very large groups,

other components of the environment are of equal importance to understanding microevolutionary change. For example, this dissertation also engages heavily with the sociocultural and microbial components (both commensal and pathogenic) of the environment surrounding the human populations of the Central Andes, which have been very impactful on human adaptation in the region.

In this work, I contend that the value of genomics for understanding human experiences lies in understanding the interactions between genes and the environment. Gene-environment interaction is a public health concept from genetic epidemiology, and seeks to understand the ways in which the environment interplays with gene expression[53]. My research draws inspiration from this concept: the notion that the collective interactions between the genes, their expression, and the environment may drive changes in the genome in subsequent generations via natural selection. This framework is invaluable because it avoids essentializing populations by using their genes to explain their behavior, ritual practices, subsistence strategies, etc. Genes alone cannot directly explain anything about the way people exist in their world[54].

### *1.5 Chapter Summaries*

Dissertation structure: This dissertation is presented in a three-paper format (Chapters 2–4, summarized below), with each paper already submitted and published in a peer-reviewed journal. In addition to these chapters, separate Introduction (Chapter 1) and Conclusion (Chapter 5) chapters are included to provide larger context, implications, and unifying themes throughout these published words. For ease of referencing, each chapter has a self-contained list of citations.

Chapter 2 Summary: Conventional hypotheses regarding the origin of global pathogens like tuberculosis, syphilis, and malaria in the Americas and their spread within the continents have been mischaracterized. Fortunately, recent studies using molecular techniques have now

superseded these missteps, which were often based in anecdotal accounts from colonial missionary reports rather than rigorous scientific study. It is now clear that there was not a unidirectional pipeline of pathogen introduction that began with European contact; instead, a rich and varied microbiological landscape already existed in the Americas. This synthesis of research regarding the origin and spread of pathogens in the Americas examines the scope of this changing perception within the fields of paleogenomics and paleomicrobiology. The first part of this synthesis addresses two historically important diseases of the Americas: tuberculosis and malaria. Because of their historical ties to colonization, these diseases have heavily debated evolutionary connections to the Americas. The latter sections of this article discuss two additional diseases: Chagas Disease, with evolutionary origins in the Americas, and Syphilis, traditionally thought to have American origins but more recently the subject of phylogenetic controversy. I then discuss how bias has historically been present in theorization of infectious disease in the Americas through a case study: Yellow Fever Virus and the Panama Canal project in Central America. Throughout, I discuss recent evidence from paleogenomics and paleomicrobiology through the lens of pathogen evolution in the ancient Americas. I aim to elevate scientific narratives which accurately characterize evolutionary relationships and disease ecology. This review specifically focuses on pathogens with special relevance to colonial narratives and evolutionary controversies brought on by next generation sequencing advancements.

Chapter 3 Summary: The individuals in this study, from two closely-related yet distinct Ecuadorian populations, presented an opportunity to better understand the genetic underpinnings of high-altitude adaptation in the Central Andes, which is poorly understood compared to other regions of the world. Utilizing haplotype, allele-frequency, and admixture-aware methods, I scanned the genomes for putative signals of selection and identified regions of the genome that show signals

of strong selection in both cardiovascular and hypoxia pathways, which are distinct from those uncovered in Peruvian populations. Interestingly, the strongest signals of selection were not related to these pathways, but instead to regions of the genome that are involved in immune function related to tuberculosis. Allele frequency trajectories indicate that these tuberculosis-related alleles increased within the populations many generations before the arrival of Europeans, ~3,000 years ago. This coincides with the transition from small hunter-gatherer groups to larger agricultural societies, which increased the spread of respiratory diseases like tuberculosis. Human immune response variation among tuberculosis strains suggests that immunity does not always correlate with strain diversity. Therefore, Ecuadorian populations could have developed immunity to South American tuberculosis strains in the ancient past while still being susceptible to the European L4 lineage at the time of contact. The unique results of this study underscore the importance of collaborating with a diverse set of communities across the Andes, and South America more broadly, in order to gain the most complete picture of the evolutionary history of the continent. I also detected a demographic population collapse that coincides with the arrival of Europeans, which is more severe than other regions of the Andes, suggesting differing effects of European contact across different high-altitude populations.

Chapter 4 Summary: Human societies have developed a broad range of cultural solutions to solve cooperative dilemmas as populations interacted in increasingly dense and connected settlements. Given the importance of cooperation, there may have also been selection for genes related to improving the capacity for group cohesion throughout human evolution. The Andean highlands provide an excellent opportunity to evaluate the role of the oxytocin system in the evolution of cooperation. The Andean archaeological record spans 13,000 years of population growth and cooperative challenges through periods of highland exploration, hunting economies, agro-

pastoralism, and urbanization. Through allele trajectory modeling using both ancient and contemporary whole genomes, I found evidence for strong positive selection on variants on the *OXTR* and *CD38* genes, known to affect oxytocin secretion and binding in the brain. These selection events commenced around 2.5 and 1.25 ka, placing them in the region's Upper Formative and Tiwanaku periods—a time of rapid population growth and urbanization.

### *1.6 Introductory References*

1. Gayo, E. M., Latorre, C. & Santoro, C. M. Timing of occupation and regional settlement patterns revealed by time-series analyses of an archaeological radiocarbon database for the South-Central Andes (16°–25°S). Quatern Int 356, 4–14 (2015).

2. Klink, C. & Aldenderfer, M. S. Ch. 03: A Projectile Point Chronology for the South-Central Andean Highlands. in Advances in Titicaca Basin Archaeology-1 (eds. Stanish, C., Cohen, A. B. & Aldenderfer, M. S.) (2005). doi:10.2307/j.ctvhhhfn9.

3. Stanish, C. Ancient Titicaca. (University of California Press, 2003). doi:10.1525/california/9780520232457.001.0001.

4. Aldenderfer, M. S. Montane Foragers: Asana and the South-Central Andean Archaic. (University of Iowa Press, 1998).

5. Haas, R. Highland Foragers of the Central Andes. in Oxford Handbook of Archaeology (eds. Aldenderfer, M. S., Sepulveda, M. & Neves, E.) (2023).

6. Craig, N. Cultural dynamics, climate, and landscape in the South-Central Andes during the mid-late holocene: A consideration of two socio-natural perspectives. Chungar (Arica) 43, 367–391 (2011).

7. Aldenderfer, M. S. Late Preceramic ceremonial architecture at Asana, southern Peru. Antiquity 64, 479–493 (1990).

8. Kitchel, N., Aldenderfer, M. S. & Haas, R. Diet, Mobility, Technology, and Lithics: Neolithization on the Andean Altiplano, 7.0–3.5 ka. J Archaeol Method Th 29, 390–425 (2022).

9. Craig, N., Aldenderfer, M. S., Rigsby, C. A., Baker, P. A. & Blanco, L. F. Geologic constraints on rain-

fed Qocha reservoir agricultural infrastructure, northern lake Titicaca Basin, Peru. J Archaeol Sci 38, 2897–2907 (2011).

10. Klarich, E. The Rise of Highland Andean Societies in the Lake Titicaca Basin. in Oxford Handbook of Archaeology (eds. Aldenderfer, M. S., Sepulveda, M. & Neves, E.) (2023).

11. Chapdelaine, C. Recent Advances in Moche Archaeology. J Archaeol Res 19, 191–231 (2011).

12. Donnan, C. B. Moche Substyles: Keys to Understanding Moche Political Organization. Boletín Del Museo Chileno De Arte Precolombino 16, 105–118 (2011).

13. Erickson, C. L. Prehistoric landscape management in the Andean highlands: Raised field agriculture and its environmental impact. Popul Environ 13, 285–300 (1992).

14. Popović, D. et al. Ancient genomes reveal long-range influence of the pre-Columbian culture and site of Tiwanaku. Sci Adv 7, eabg7261 (2021).

15. Nash, D. J. & Williams, P. R. Architecture and Power on the Wari–Tiwanaku Frontier. Arch P Amer Ant Asso 14, 151–174 (2004).

16. Arkush, E. & Tung, T. A. Patterns of War in the Andes from the Archaic to the Late Horizon: Insights from Settlement Patterns and Cranial Trauma. J Archaeol Res 21, 307–369 (2013).

17. Gullberg, S. R. Chapter 2: Evolution of an Empire. in Astronomy of the Inca Empire, Use and Significance of the Sun and the Night Sky 5–39 (2020). doi:10.1007/978-3-030-48366-1_2.

18. Cuadros, M. F. B. The making of invisibility: colonialism and multiple erasures along the southern Peruvian shores. Colon Lat Am Rev 31, 607–616 (2022).

19. Weaver, B. J. M. Reflections on ‘material histories’ and the archaeology of slavery in Peru. Colon Lat Am Rev 31, 591–598 (2022).

20. Nakatsuka, N. et al. A Paleogenomic Reconstruction of the Deep Population History of the Andes. Cell 181, 1131–1145 (2020).

21. Scheib, C. L. et al. Ancient human parallel lineages within North America contributed to a coastal expansion. Science 360, 1024–1027 (2018).

22. Capodiferro, M. R. et al. Archaeogenomic distinctiveness of the Isthmo-Colombian area. Cell 184,

1706-1723.e24 (2021).

23. Harris, D. N. et al. Evolutionary genomic dynamics of Peruvians before, during, and after the Inca Empire. Proc National Acad Sci 115, E6526–E6535 (2018).

24. Skoglund, P. et al. Genetic evidence for two founding populations of the Americas. Nature 525, 104–108 (2015).

25. Raghavan, M. et al. Genomic evidence for the Pleistocene and recent population history of Native Americans. Science 349, aab3884 (2015).

26. Crawford, J. E. et al. Natural Selection on Genes Related to Cardiovascular Health in High-Altitude Adapted Andeans. Am J Hum Genetics 101, 752–767 (2017).

27. Reich, D. et al. Reconstructing Native American Population History. Nature 488, 370–374 (2012).

28. Posth, C. et al. Reconstructing the Deep Population History of Central and South America. Cell 175, 1185-1197.e22 (2018).

29. Lindo, J. et al. The genetic prehistory of the Andean highlands 7000 years BP though European contact. Sci Adv 4, eaau4921 (2018).

30. Caro-Consuegra, R. et al. Uncovering Signals of Positive Selection in Peruvian Populations from Three Ecological Regions. Mol Biol Evol 39, msac158 (2022).

31. Moreno-Mayar, J. V. et al. Early human dispersals within the Americas. Science 362, (2018).

32. Carrasco-Torrontegui, A., Gallegos-Riofrío, C. A., Delgado-Espinoza, F. & Swanson, M. Climate Change, Food Sovereignty and Ancestral Farming Technologies in the Andes. Curr Dev Nutrition 5, 54–60 (2020).

33. Brutsaert, T. D. et al. Association of EGLN1 gene with high aerobic capacity of Peruvian Quechua at high altitude. Proc National Acad Sci 116, 24006–24011 (2019).

34. Joseph, S. K. et al. Genomic evidence for adaptation to tuberculosis in the Andes before European contact. Iscience 26, 106034 (2023).

35. Nelson, E. A., Buikstra, J. E., Herbig, A., Tung, T. A. & Bos, K. I. Advances in the molecular detection of tuberculosis in pre-contact Andean South America. Int J Paleopathol 29, 128–140 (2020).

36. Bos, K. I. et al. Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. Nature 514, 494–497 (2014).

37. Brynildsrud, O. B. et al. Global expansion of Mycobacterium tuberculosis lineage 4 shaped by colonial migration and local adaptation. Sci Adv 4, eaat5869 (2018).

38. Wagner, J. K. et al. Fostering Responsible Research on Ancient DNA. Am J Hum Genetics 107, 183–195 (2020).

39. Cochran, P. A. L. et al. Indigenous ways of knowing: Implications for participatory research and community. Am J Public Health 98, 22–27 (2008).

40. Laland, K. N. et al. The extended evolutionary synthesis: its structure, assumptions and predictions. Proc Royal Soc B Biological Sci 282, 20151019 (2015).

41. Charlesworth, B. & Charlesworth, D. Population genetics from 1966 to 2016. Heredity 118, 2–9 (2017).

42. Lowe, W. H., Kovach, R. P. & Allendorf, F. W. Population Genetics and Demography Unite Ecology and Evolution. Trends Ecol Evol 32, 141–152 (2017).

43. Gould, S. jay & Eldredge, N. Punctuated equilibrium comes of age. Nature 366, 223–227 (1993).

44. Stanley, S. M. Influence of seawater chemistry on biomineralization throughout phanerozoic time: Paleontological and experimental evidence. Palaeogeogr Palaeoclim Palaeoecol 232, 214–236 (2006).

45. Mayr, E. The objects of selection. Proceedings of the National Academy of Sciences 94, (1997).

46. Darwin, C. On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life. (John Murray, London, 1859).

47. Welch, J. J. What's wrong with evolutionary biology? Biol Philos 32, 263–279 (2017).

48. Queiroz, K. de. Ernst Mayr and the modern concept of species. Proc National Acad Sci 102, 6600–6607 (2005).

49. Stock, J. T. Are humans still evolving? Embo Rep 9, S51–S54 (2008).

50. Bowden, R. et al. Genomic Tools for Evolution and Conservation in the Chimpanzee: Pan troglodytes ellioti Is a Genetically Distinct Population. Plos Genet 8, e1002504 (2012).

51. Nielsen, R., Hellmann, I., Hubisz, M., Bustamante, C. & Clark, A. G. Recent and ongoing selection in the human genome. Nat Rev Genet 8, 857–868 (2007).

52. Peter, B. M. Admixture, population structure, and f-statistics. Genetics 202, 1485–1501 (2016).

53. Tabery, J. R. A. Fisher, Lancelot Hogben, and the Origin(s) of Genotype–Environment Interaction. J Hist Biol 41, 717–761 (2008).

54. Fuentes, A. Biological anthropology's critical engagement with genomics, evolution, race/racism, and ourselves: Opportunities and challenges to making a difference in the academy and the world. Am. J. Phys. Anthropol. 175, 326–338 (2021).

# CHAPTER 2: A REVIEW OF THE EVOLUTIONARY HISTORY OF INFECTIOUS DISEASE IN THE ANCIENT AMERICAS

Publication Information:

S.K. Joseph, J. Lindo, The evolutionary history of infectious disease in the ancient Americas and the pathogenic consequences of European contact. *American Journal of Biological Anthropology* 182(4) (2022).

## *2.1 Abstract*

The increasing availability of next generation sequencing techniques in recent decades has led to new discoveries, and sometimes the redefinition of conventional hypotheses, regarding many complex human-pathogen evolutionary relationships. These new discoveries are particularly poignant in studies of the Americas, where research into Indigenous ancestry and migration has historically been ignored. As a result, conventional hypotheses regarding the origin of global pathogens like tuberculosis, syphilis, and malaria in the Americas and their spread within the continents have been mischaracterized. Fortunately, recent studies using molecular techniques have now superseded these missteps, which were often based in anecdotal accounts from colonial missionary reports rather than rigorous scientific study. It is now clear that there was not a unidirectional pipeline of pathogen introduction that began with European contact; instead, a rich and varied microbiological landscape already existed in the Americas. This synthesis of research regarding the origin and spread of pathogens in the Americas examines the scope of this changing perception within the fields of paleogenomics and paleomicrobiology.

## *2.2 Introduction*

Increasingly available and accessible next generation sequencing (NGS) techniques have led to new discoveries regarding the evolutionary relationships between human infectious diseases

around the world (Barbier & Wirth, 2016; Bianucci et al., 2015). In studies of the Americas, where research into Indigenous ways of life and population ecology have historically been an afterthought, these new avenues of research have redefined many conventional hypotheses regarding the evolutionary relationships of human pathogens. Specifically, phylogenetic studies using data obtained from NGS techniques, in concert with paleopathological analysis of human remains, have often contradicted the written observational records made by colonial authorities (Barbier & Wirth, 2016; Breedlove & Arguin, 2015; Laxao, 2016; Stead, 2001; Stead et al., 1995; Sutter, 2016; Wilbur & Buikstra, 2006).

The first part of this synthesis addresses three historically important diseases of the Americas: tuberculosis, malaria, and Yellow Fever. Largely because of their historical ties to colonization, these diseases have controversial and debated evolutionary connections to the Americas. The latter sections of this article discuss two additional diseases: Chagas Disease, with evolutionary origins in the Americas, and Syphilis, traditionally thought to have American origins but more recently has been the subject of phylogenetic controversy (Figure 1). We then discuss how bias has historically been present in scientific theorization of infectious diseases in the Americas through a case study, the impact of Yellow Fever on the Panama Canal project in Central America. Throughout, we examine new evidence obtained in the fields of paleogenomics and paleomicrobiology, through the lens of pathogen evolution in the ancient Americas, in order to elevate scientific narratives which accurately characterize evolutionary relationships and disease ecology. This synthesis is not a comprehensive review of the enormous scope of pathogens present in the Americas, but rather focuses on a few of particular relevance to colonial narratives and evolutionary controversies brought on by NGS advancements.

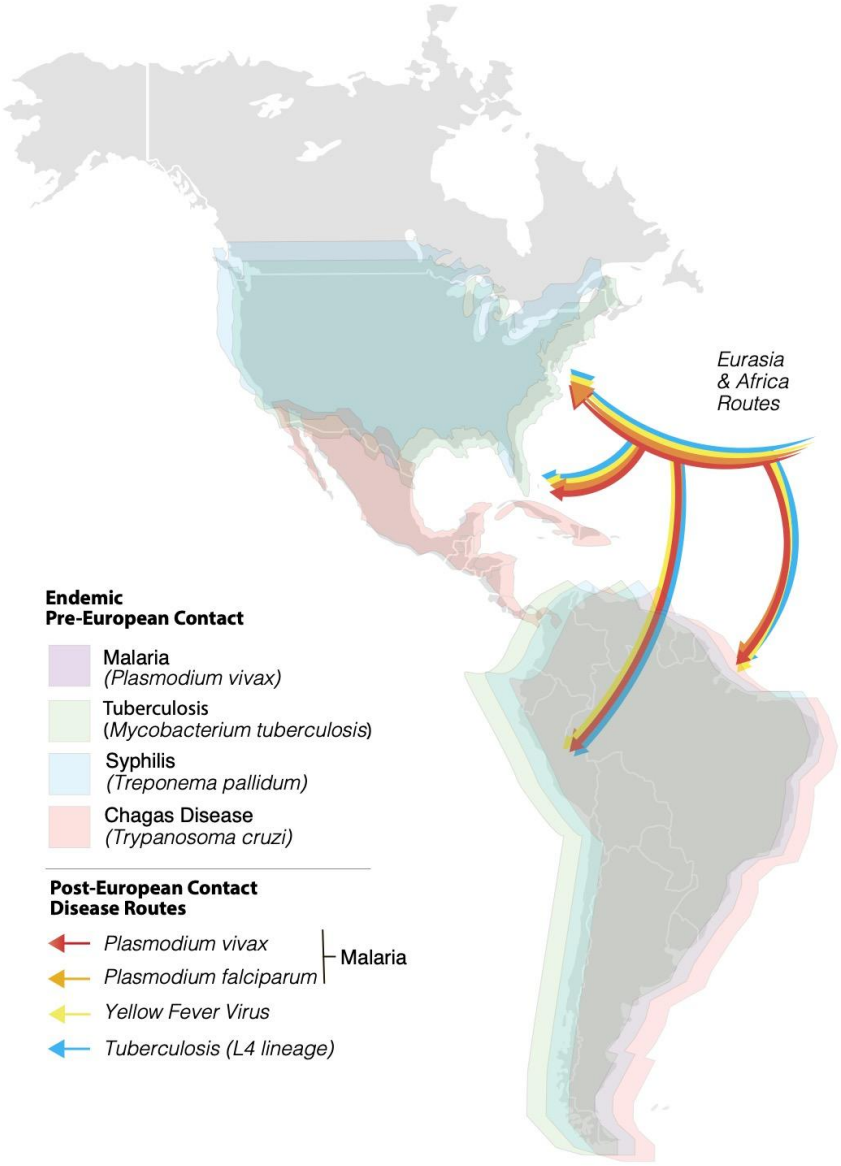



**Figure 1.** Map of the Americas showing the diseases discussed in this article, including those present pre-European contact (solid shading) and those first introduced by European colonization (arrows). See Appendix A Table A1 for list of references.

### *2.3 Tuberculosis*

Caused by *Mycobacterium tuberculosis*, a bacterium which primarily infects the lungs, tuberculosis (TB) is easily transmitted between humans via respiratory droplets. In much of the

contemporary world, it kills millions of people each year (Buzic & Giuffra, 2020). Studies of evolutionary relationships between modern strains of *M. tuberculosis* suggest that the progenitor of these strains may have first plagued human populations in Africa approximately 70,000 years before present (BP) (Comas et al. 2013; Gagneux 2018). Its expansion and spread across the world are thought to be coincident with the migration of anatomically modern *Homo sapiens* into Eurasia and beyond (Gagneux 2018; Gutierrez et al., 2005; Wirth et al., 2008). Moreover, the origin of *M. tuberculosis* in the Americas is continually debated. Until the end of the 20th century, the prevailing assumption was that *M. tuberculosis* was not present in pre-European contact America, and first came about in Eurasia as a zoonosis from *M. bovis* at the time of Neolithic cattle domestication (Barbier & Wirth, 2016; Wilbur & Buikstra, 2006). This assumption was based on the extreme devastation of Indigenous populations at the time of colonization—as recorded in missionary reports—which historians and archaeologists assumed pointed to a lack of prior exposure, and therefore lack of immunity, to TB infection (Stead, 2001; Stead et al., 1995).

Before the advent of molecular techniques for identifying pathogens in human skeletal remains, paleopathological analysis of bone lesions was the primary method for identifying TB presence (Wilbur & Buikstra, 2006). This method is less accurate than using pathogen DNA or lipid biomarkers, as demonstrated in the case-control skeletal diagnosis study by Pederson et al. (2019), which found an ~50% error rate with visual TB diagnosis. As a result, while some probable TB lesions were identified in pre-contact remains of Indigenous people, this alone was not enough to fully dismiss the conventional theory of European-borne TB introduction to the Americas (Allison et al., 1973; Lichtor & Lichtor, 1957; Nelson et al., 2020; Ritchie, 1952; Toyne et al., 2020; Wilbur & Buikstra, 2006).

Though a remarkable advancement, molecular techniques, such as those which utilize pathogen DNA or lipid biomarkers, are not perfectly reliable either. For example, contamination from microbes in the surrounding environment is always a possibility; in fact, ancient DNA samples contain very little endogenous content and are mostly made of environmental material. While no method is completely reliable, combining skeletal and molecular evidence provides a good chance of accurate pathogen identification. However, it is important to note that early-stage TB infections do not usually result in skeletal damage, and thus, many TB infections may have gone undetected in archaeological remains (Nelson et al., 2020). This is the benefit of NGS techniques, several of which can be combined to increase the chance of successful ancient pathogen identification. Computational programs like MapDamage, which quantify characteristic nucleotide misincorporation patterns in an ancient sample, are used to validate the presence of ancient DNA (Jónsson et al., 2013). Then, programs like the MEGAN Alignment Tool (MALT) can be used to assign taxonomic identifiers to all DNA fragments in a sample using a reference database (Herbig et al., 2016). MALT is particularly useful because it can account for the environmental DNA present in a sequenced sample and provide a detailed profile to further help avoid false identifications (Herbig et al., 2016; Nelson et al., 2020).

The increasing availability of molecular techniques in recent decades has led to the discovery of multiple lines of evidence contradicting the European-borne TB origin story, but has also created new controversies (Barbier & Wirth, 2016) (Figure 2). Bos et al. (2014) published three mycobacterial genomes obtained from human remains in southern Peru dated to 1,000 BP. After conducting analysis for phylogenetic similarity in single nucleotide polymorphisms (SNPs), they found the greatest similarity not with other modern strains of human TB, but in modern *M. pinnipedii* strains found in migratory seals of the southern hemisphere. Using molecular clock

estimations of accumulated mutation rates, Bos et al. (2014) determined the divergence of the human TB strains found in the Peruvian remains occurred at approximately 4,500 BP. Rothschild et al. (2001) published two much older (17,870 ± 230 BP) sequences of *M. tuberculosis* DNA from an extinct bison, *Bison antiquus,* found in present-day Wyoming, US. Subsequent analysis of the presence of mycolic acid, characteristic of the bacterium, has further supported the presence of TB in these bison remains (Lee et al., 2012). However, detailed analysis of the phylogenetic relationship between these bison samples and human strains of *M. tuberculosis*–and using a more reliable hybridization capture method for the bacterial genomes combined with ancient DNA validation methods–is a necessary area of future research (Lee et al., 2012).

These data, particularly those published by Bos et al. (2014) for which the presence of ancient pathogen DNA was validated using nucleotide misincorporation patterns, suggest there may have been an initial zoonotic introduction of TB into pre-contact Indigenous American populations. However, phylogenetic studies of modern TB lineages around the world, such as that of Brynildsrud et al. (2018), show that the L4 phylogenetic lineage was repeatedly dispersed out of Europe during times coincident with colonization and is the lineage which remains dominant in the Americas today. Taken together, the works of Bos et al. (2014) and Brynildsrud et al. (2018) could suggest there was a pre-contact zoonotic lineage which was outcompeted by the L4 European lineage brought by colonists. Further analyses are needed to elucidate the phylogenetic relationship between the different pre-contact human *M. tuberculosis* strains identified in human burials across the Americas. In concert with demography and migration data, this would shed light onto a possible non-human animal origin TB in pre-contact Indigenous populations of the Americas that may have eventually led to a zoonosis into human populations.


**Figure 2.** Map of the Americas detailing the location of the two study sites which contain evidence of pre-European contact tuberculosis: skeletal remains of three individuals found in Peru published by Bos et al. (2014) dated to ~1000 BP and the *Bison antiquus* genome published by Rothschild et al. (2001). Also pictured (arrows) is the proposed mechanism of tuberculosis introduction by Bos et al. (2014), migratory seals from the southern hemisphere. Created with BioRender.com.

### *2.4 Malaria*

Malaria is an *Anopheles* mosquito-borne illness caused by a protozoan parasite of the genus *Plasmodium*. Infection by the *P. vivax* and *P. falciparum* species are most common in the Americas (Oliveira et al., 2017). Malaria control was a central focus of epidemiological investigation in early-20$^{th}$ century North America, and the Centers for Disease Control and Prevention was

originally founded in 1946 for this purpose (Parascandola, 1974). However, the subsequent success of eradication measures in wealthier nations of North America did not extend to lower and middle-income nations in the rest of the world (Molyneux et al., 2021). In the year 2020, malaria still killed over 600,000 people, primarily in sub-Saharan Africa (Jagannathan & Kakuru, 2022).

Current research suggests *P. vivax* and *P. falciparum* may have originated in Africa, possibly as zoonoses from *Plasmodium* parasites found first in African apes (Galaway et al., 2019; Krief et al., 2010; Loy et al., 2017; Loy et al., 2018; Prugnolle et al., 2010). Their path from Africa and Eurasia to the Americas has long been debated, since *Anopheles* mosquitoes cannot have traveled over the frigid Bering land bridge due to temperature constraints on their life cycle; thus, an alternative route of introduction to the Americas is more likely (Oliveira et al., 2017; Rodrigues et al., 2018). However, historical records from Spanish colonization indicate the Quechua people of the Andes were using the bark of Cinchona trees for relieving symptoms of a malaria-like infection (Breedlove & Arguin, 2015). Thus, the debates over the presence of pre-contact malaria and the source of *P. vivax* and *P. falciparum* remain unresolved.

Malaria infection does not leave identifiable bone lesions in skeletal remains, although nonspecific evidence of anemia, which could potentially represent malarial anemia, can sometimes be identified visually (Bianucci et al., 2015; Smith-Guzmán, 2015). Thus, molecular methods of identifying the DNA of the *Plasmodium* parasite are the only way to confirm an infection in human remains. However, thus far, most studies of the evolutionary history of *P. vivax* and *P. falciparum* in the Americas have utilized modern parasite samples, not ancient DNA. The most common and well-validated methods are through PCR amplicon 18S rRNA sequencing and metagenomic analyses of whole genome data, which can indicate putative presence of molecular material from

*Plasmodium*, or hybridization capture, which positively identifies the presence of specific stretches of *Plasmodium* DNA (Bianucci et al., 2015).

Phylogenetic analyses conducted on data obtained through these methods suggest two different avenues for the post-contact arrival of *P. vivax* and *P. falciparum* in the Americas. The first is the Transatlantic Slave Trade, where more than ten million enslaved African people were brought to the Americas from 1501–1866 (Molina-Cruz & Barillas-Mury, 2014; slavevoyages.org, 2021). Overall phylogenetic similarity between *P. falciparum* found in Africa and the Americas today, estimated divergence time of American *P. falciparum*, and genetic polymorphisms that improve immune invasion in mosquito vectors, suggest infected enslaved people and traders likely brought *P. falciparum* along these Atlantic routes (Molina-Cruz et al., 2013; Yalcindag et al., 2012). The parasites then quickly adapted to infect American species of mosquitoes (Molina-Cruz et al., 2013; Molina-Cruz & Barillas-Mury, 2014).

In contrast, modern American *P. vivax* displays higher overall genetic diversity compared to *P. falciparum*, as well as phylogenetic similarity to *P. vivax* from disparate regions of the world, not just Africa (Oliveira et al., 2017; Rodrigues et al., 2018; van Dorp et al., 2020) (Figure 3). Significant similarity to South Asian, African, and Melanesian *P. vivax* were most prevalent in the study by Rodrigues et al. (2018) of modern *Plasmodium* parasite mitochondrial genomes. Their study suggests that human migration from these regions to the Americas most likely resulted in multiple introductions of *P. vivax,* both through ancient pre-contact migrations and the Transatlantic Slave Trade (Skoglund et al., 2015; Rodrigues et al., 2018). Oliveira et al. (2017) came to similar conclusions in their study using whole genome sequencing of nuclear DNA from a diverse set of modern South American *P. vivax* samples. However, the analyses of Gelabert et al. (2016) and van Dorp et al. (2020), using microscope slide samples collected in 1944 from a

recently eradicated European strain of *P. vivax,* suggest this increased genetic diversity comes from a secondary introgression of *P. vivax* strains from other regions of the world after colonization. Thus, particularly for *P. vivax,* studies of linkages between parasite lineage divergence time and human migrations in the Americas are crucial next steps in understanding human-malaria coevolution. Additionally, ancient DNA studies, particularly from human remains with possible evidence of anemia, could indicate the presence of pre-European contact *P. vivax* malaria from ancient migrations.

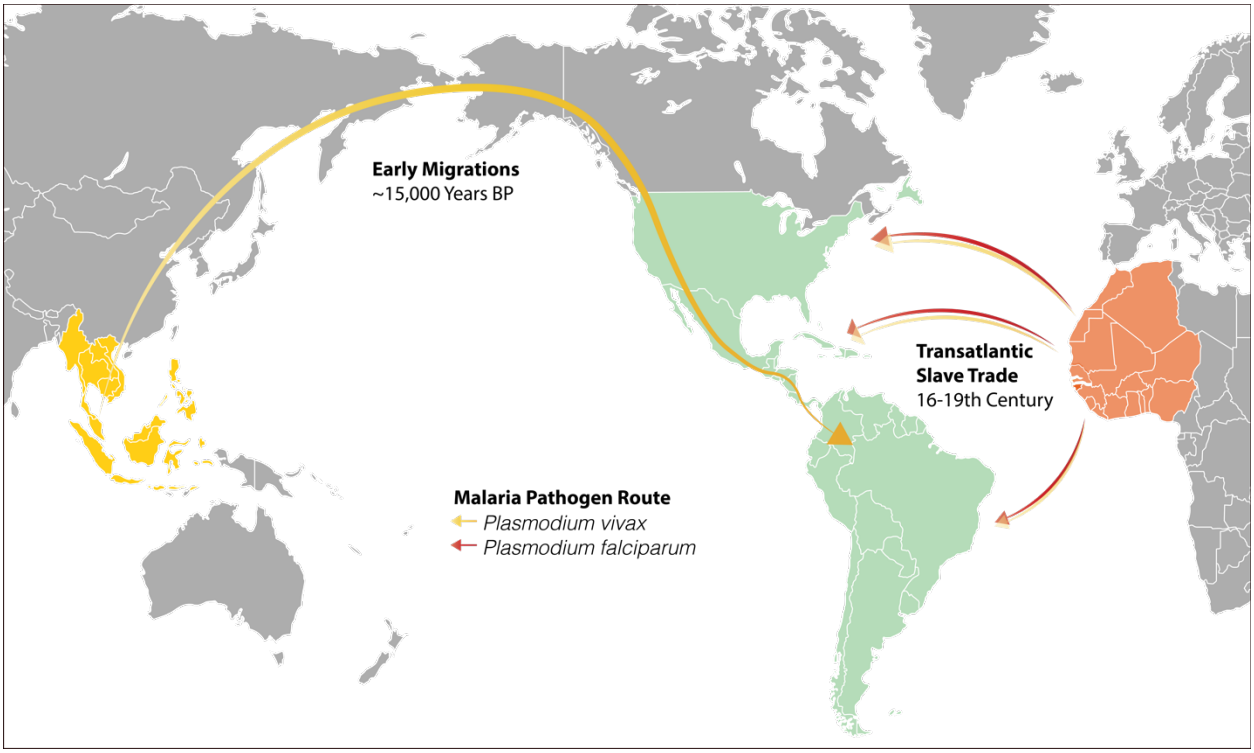



**Figure 3.** World map showing the putative routes (arrows) of introduction of *P. falciparum* and *P. vivax* malaria, based on phylogenetic studies. Both *P. falciparum* and *P. vivax* are thought to have been brought post-European contact via the Transatlantic Slave trade; there is also debated phylogenetic evidence for additional earlier introductions of *P. vivax* via ancient migrations. See Appendix A Table A2 for list of references.

### *2.5 Syphilis, Yaws, and Bejel*

Syphilis, yaws, and bejel are clinically distinct diseases caused by different strains of the bacterium *Treponema pallidum,* which were, until recently, thought to represent three independent phylogenetic lineages. Venereal, or sexually-transmitted, syphilis is the most widespread of these diseases, and although treatable with penicillin antibiotic, is currently undergoing a resurgence around the world (Melo et al., 2010). Treponemal infections, those caused by *T. pallidum*, were certainly present in the ancient past, and severe infections often left characteristic skeletal lesions well-suited for paleopathological analysis (Pineda et al., 2009). The conventional theory regarding its evolutionary origins was that *T. pallidum* was endemic to the Americas and was subsequently brought back to Eurasia by colonizers (Bollaert, 1864; Pineda et al., 1998). Recently, however, molecular clock studies, and other estimates of divergence times in the lineages of Eurasian *T. pallidum,* suggest that it may have been already widespread across the world before European contact with Indigenous Americans (Giffin et al. 2020; Harper et al., 2008; Majander et al., 2020). Yet, there is no definitive paleopathological evidence of pre-contact *T. pallidum* presence in Eurasia or Africa; though, it is possible that the bacterium existed there in a form which caused milder disease, and thus did not leave skeletal lesions (Harper et al., 2011). Consequently, the investigation of the origin and spread of *T. pallidum* around the world remains a field of active interdisciplinary research, and no definitive conclusion has been posited.

As with the study of many other infectious diseases, our understanding of the evolutionary history of *T. pallidum* has advanced greatly with the application of PCR amplicon and NGS techniques (Melo et al., 2010). Only in the most recent decade has ancient treponemal DNA been putatively identified in human remains, though similar limitations apply as with other molecular studies of ancient pathogen DNA. Majander et al. (2020) isolated ancient DNA of *T. pallidum*

from four 15th and 16th century individuals in northeastern Europe, suggesting that *T. pallidum* already existed across Eurasia before contact of European colonizers with Indigenous Americans. One of the four treponemal genomes isolated by Majander et al. (2020) is consistent with tropical yaws, while another is consistent with the lineage predating the evolutionary split between yaws-causing and bejel-causing *T. pallidum.* The remaining two showed highest similarity to lineages which cause venereal syphilis. Using similar targeted capture and sequencing methods, Giffin et al. (2020) reported a 15th century case of bubonic plague coincident with *T. pallidum* in the same individual from modern-day Lithuania. Phylogenetic analyses by Giffin et al. (2020) also found similarity to yaws-causing *T. pallidum,* which is today mostly restricted to the tropics, highlighting the likely importance of trade and colonization for the spread of various treponemal infections around the world.

In the Americas, there is no known pre-colonial ancient treponemal DNA from human remains to date, perhaps related to the overall dearth of ancient DNA studies in the Americas. Schuenemann et al. (2018) isolated the first known *T. pallidum* ancient DNA in the Americas from a colonial-era convent in Mexico City, which was continuously inhabited from the 17th to the 19th centuries. Interestingly, they found genomic evidence of 16 possible recombination events between strains of *T. pallidum*, which contradicts the conventional thought that the three modern strains in existence today are independent clonal lineages (Schuenemann et al., 2018). Evidently, our knowledge of the origins and spread of syphilis, yaws, and bejel are extremely incomplete. More data from ancient DNA studies, especially, are needed to firmly establish the relationships between ancient and modern forms of *T. pallidum* and their movement via human migration.

### *2.6 Chagas Disease*

The etiologic agent of Chagas Disease is *Trypanosoma cruzi,* a protozoan parasite transmitted by triatomines, blood sucking insects colloquially known as 'kissing bugs' (Echeverria & Morillo, 2019). Even with treatment by antiparasitic drugs, Chagas Disease is fatal in approximately 10% of cases today−usually due to heart inflammation−and can also cause major complications with the digestive and central nervous systems (Aufderheide et al., 2004; Echeverria & Morillo, 2019). Compared to that of syphilis, yaws, and bejel, the archaeogenetic record of Chagas Disease in the Americas is more complete. DNA isolated from dozens of South American mummies, including remains from modern-day Peru, Brazil, and Chile, has indicated the presence of *T. cruzi* among ancient populations as far back as 9,000 BP (Aufderheide et al., 2004; Fernandes et al., 2008; Ferreira et al., 2000; Guhl et al., 2014; Lima et al., 2008).

Natural mummification in arid environments, like that of the Atacama Desert in southern Peru and northern Chile, or the Peruaçu Valley of Brazil, is excellent for preserving soft tissue (Aufderheide et al., 2004; Ferreira et al., 2000). Within the last several decades, the increased use of PCR amplicon and NGS techniques has revealed that preservation of ancient pathogen DNA may occur in these environments, too (Aufderheide et al., 2004; Fernandes et al., 2008; Ferreira et al., 2000; Guhl et al., 2014; Lima et al., 2008). While ancient DNA can be isolated from tooth or bone, mummification is ideal for preserving pathogens, such as protozoan parasites, which primarily infect soft tissue (Aufderheide et al., 2004). It is important to note that these studies relied on molecular techniques which predate currently accepted ancient pathogen DNA validation measures (Herbig et al., 2016; Jónsson et al., 2013; Nelson et al., 2020). Thus, while the evolutionary history of Chagas Disease in the Americas is somewhat well-understood, it also presents a promising area for future inquiry with updated methods.

Phylogenetic studies suggest that Indigenous populations of the Americas first encountered *T. cruzi* as a zoonosis from bats, likely as soon as they began to permanently settle the Andes and Atacama Desert (Hamilton et al., 2012; Pinto et al., 2015). By the time of Spanish colonization, Chagas Disease was widespread throughout Indigenous populations, and historical accounts confirm that it affected Spanish colonial officials, as well (Guerra, 1970; Steverding, 2014). One of the most famous historical encounters with kissing bugs comes from Charles Darwin. In March 1835, during his voyage on The Beagle, Darwin noted in his diary that he unwittingly provided a blood meal for a kissing bug during the night (Darwin, 1832-1836). Interestingly, it has been hypothesized that Darwin suffered from chronic Chagas Disease, due to his protracted issues with his digestive and nervous systems, though no conclusive diagnosis was ever made (Miles, 2004; Steverding, 2014).

Chagas Disease is one of many examples which contradicts the narrative of a unidirectional flow of pathogens from European colonizers to Indigenous Americans. Human remains preserved via mummification have provided valuable insight into *T. cruzi* infections in the ancient past, and future NGS studies of trypanosome genomes from human remains−performed with computational ancient DNA validation methods−could offer more definitive evidence of the presence of Chagas Disease in ancient human populations of the Americas (Aufderheide et al., 2004; Ferreira et al., 2000; Herbig et al., 2016; Jónsson et al., 2013; Nelson et al., 2020). If this research confirms that *T. cruzi* has infected humans for hundreds of generations in South America, there is further potential for examining the likely time of spillover into human populations, as *T. cruzi* likely has millions more years of evolutionary history residing in American bats (Aufderheide et al., 2004; Fernandes et al., 2008; Ferreira et al., 2000; Guhl et al., 2014; Hamilton et al., 2012; Lima et al., 2008; Pinto et al., 2015). Furthermore, future studies on human ancient DNA from these

individuals may also provide insight into host-pathogen evolution and the population-level effects of Chagas Disease as a selective pressure.

### *2.7 Bias in Characterization of Diseases in the Americas: The Case of Yellow Fever and the Panama Canal Project*

Until the most recent decades, research into Indigenous ancestry, migration, and disease ecology has been largely ignored. The skewed portrayal of the Panama Canal project from Western perspectives illustrates why sources like missionary reports and colonial authorities are not reliable characterizations of infectious disease dynamics in the Americas (Laxao, 2016; Sutter, 2016). Similarly, the conventional narrative surrounding the European origin of TB, based on descriptions from colonial reports, has only recently been disproven by phylogenetic studies (Barbier & Wirth, 2016; Stead, 2001; Stead et al., 1995). This is not to say that diseases brought by colonial activity were not devastating to Indigenous Americans, but rather that more care must be taken not to conflate observations made by colonizers with conclusive evidence of evolutionary origins.

The post-contact population collapse of Indigenous communities—as estimated by demographic population genetic analyses, not historical reports—were greater than 90% in the lowland coast of modern-day Chile and in the highlands of present-day Mexico (Lindo et al., 2018). In the Andean highlands, Indigenous communities sustained population reductions of approximately 25%, and almost 60% in the Pacific Northwest (Lindo et al., 2016; Lindo et al., 2018). Smallpox, TB, and influenza are thought to be important contributors to these collapses. European colonial contact was unquestionably devastating, however, diseases from the Americas certainly infected colonizers, too. Pre-colonial America was not free of disease, nor was there a unidirectional movement of pathogens from Europe to the Americas after European contact (Bollaert, 1864; Giffin et al. 2020; Harper et al., 2008; Majander et al., 2020; Pineda et al., 1998).

Yellow Fever (YF), an RNA flavivirus primarily transmitted by *Aedes* mosquitoes, has been present in the Americas since the beginning of the Transatlantic Slave Trade (Bryant et al., 2007; Chippaux & Chippaux, 2018; Li et al., 2017). Surviving a YF infection provides lifelong immunity. For adults not immune, YF is a debilitating and often deadly infection, known for causing vomiting, hemorrhaging, and liver complications in its most severe cases (Monath & Vasconcelos, 2015). Like *P. falciparum* malaria, the closest phylogenetic similarity to the YF strains in the Americas are those from Africa (Bryant et al., 2007; Li et al., 2017; Molina-Cruz et al., 2013; Yalcindag et al., 2012). Finding molecular evidence using ancient pathogen RNA has not historically been possible due to the poor preservation of relatively unstable RNA molecules; however, new research suggests that recovering ancient mRNA transcriptomes may be possible, as well as RNA viral genomes, if preserved in extremely favorable conditions, such as in permafrost (Ng et al., 2014; Smith et al., 2019). Düx and colleagues (2020) isolated viral RNA from a lung tissue specimen, collected in 1912 and preserved in a museum, from a child infected with the RNA measles virus. They applied Bayesian molecular clock modeling to this modern RNA sample to estimate the origins of the virus in human urbanized populations. Both modern and ancient pathogen RNA are promising avenues of research and could be applied to understanding the evolutionary history of YF in the Americas in future studies.

*Aedes* mosquitoes breed most prolifically in artificial water containers, meaning that cities and towns with larger populations—and man-made water infrastructure—had the highest potential for large YF outbreaks when European colonizers arrived (Monath & Vasconcelos, 2015; Sutter, 2016). From the time of colonization until the vector eradication efforts of the most recent century, the majority of Central and South American adults were likely immune to YF infection because exposure was so widespread (Chippaux & Chippaux, 2018; Sutter, 2016). However, the same was

not true of the Europeans, who arrived from a climate not conducive to the spread of mosquito-borne diseases. Such was the fate of the infamous French expedition to build the Panama Canal from 1881–1889. Over 22,000 workers died—mostly Europeans and enslaved Africans from YF and malaria infections—dooming the project to failure. However, the local population, most of whom were already immune to YF from prior infection, were largely spared, though many still perished due to unsafe working conditions (Sutter, 2016).

The Panama Canal project was later continued by the United States beginning in 1904. The eventual completion of the project hinged on vector control measures against urban-dwelling *Aedes aegypti* mosquitoes (Sutter, 2016). Their efforts, though successful from the perspective of YF control, were hugely disruptive for the urban lives of the local people who had no need for such measures. Many areas along the canal were vacated, and local people were forcibly displaced from their homes and communities in order for the American colonial authority to remove their artificial water containers (Laxao, 2016). In their reports, the leaders of the American Panama Canal project portrayed this devastation of the Canal zone as the result of an YF epidemic that the Americans later heroically controlled; in actuality, local people of Panama were not large contributors to the YF outbreak, as they were mostly immune (Sutter, 2016).

### *2.8 Discussion*

The Americas are a critically important region for understanding the origin and spread of human pathogens. Several diseases discussed in this article, such as TB, malaria, YF, and syphilis, cause morbidity and mortality in millions of people around the globe each year. Understanding their evolutionary paths to and from the Americas is crucial for advancing a more holistic understanding of these diseases in the modern world (Buzic & Giuffra, 2020; Melo et al., 2010; Molyneux et al., 2021; Monath & Vasconcelos, 2015). Furthermore, research into Indigenous

American populations has historically been ignored. Thus, the longstanding narrative of unidirectional pathogen introduction from European colonizers to the Americas remained unchallenged for many decades (Barbier & Wirth, 2016; Bollaert, 1864; Laxao, 2016; Majander et al., 2020; Pineda et al., 1998; Stead, 2001; Stead et al., 1995; Sutter, 2016; Wilbur & Buikstra, 2006).

A corollary of this narrative is that Indigenous populations of the Americas must have not had any major infectious diseases, except for a few macroparasites, protozoans, and sexually transmitted bacterial infections. To be sure, phylogenetic studies have confirmed that colonial activity over the past 500 years is likely responsible for the introduction of diseases like YF and *P. falciparum* malaria (Bryant et al., 2007; Chippaux & Chippaux, 2018; Li et al., 2017; Molina-Cruz & Barillas-Mury, 2014; slavevoyages.org, 2021). Yet, in the case of TB, for example, it was assumed by scientists and historians alike that the extreme devastation of Indigenous populations at the time of colonization pointed to a lack of prior immunity to TB (Stead, 2001; Stead et al., 1995). The post-contact population collapses following TB, smallpox, and influenza epidemics were unquestionably horrific and devastating, often killing the majority of the Indigenous population of a given area (Lindo et al., 2016; Lindo et al., 2018). However, this is not sufficient evidence to conclude that none of the Indigenous Americans had prior exposure to these diseases. With respect to TB, we must now consider the possibility that the presence of *M. tuberculosis* in the Americas likely spans many thousands of years, even if it was not the current L4 strain most prevalent in the present day (Bos et al., 2014; Brynildsrud et al., 2018; Lee et al., 2012; Rothschild et al. 2001). Evidently, molecular techniques for studying paleomicrobiology have certainly brought about new controversies, especially regarding the complex phylogenies of infectious

diseases. Most importantly, however, these techniques have also brought to light the rich and diverse microbiological landscape which has always existed in the Americas.

### *2.9 References*

Allison, M. J., Mendoza, D., & Pezzia, A. (1973). Documentation of a case of tuberculosis in pre-Columbian America. American Review of Respiratory Disease, 107, 985–991.

Aufderheide, A. C., Salo, W., Madden, M., Streitz, J., Buikstra, J., Guhl, F., … Allison, M. (2004). A 9,000-year record of Chagas' disease. Proceedings of the National Academy of Sciences, 101(7), 2034–2039.

Barbier, M., & Wirth, T. (2016). The Evolutionary History, Demography, and Spread of the Mycobacterium tuberculosis Complex. Microbiology Spectrum, 4(4).

Bianucci, R., Araujo, A., Pusch, C. M., & Nerlich, A. G. (2015). The identification of malaria in paleopathology–An in-depth assessment of the strategies to detect malaria in ancient remains. Acta Tropica, 152, 176–180.

Bollaert, W. M. (1864). On the Alleged Introduction of Syphilis from the New World. Also Some Notes from the Local and Imported Diseases into America. Journal of the Anthropological Society of London, 2, cclvi–cclxix. Retrieved from https://www.jstor.org/stable/pdf/3025228.pdf

Bos, K. I., Harkins, K. M., Herbig, A., Coscolla, M., Weber, N., Comas, I., … Krause, J. (2014). Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. Nature, 514(7523), 494–497.

Breedlove, B., & Arguin, P. M. (2015). Portrait of the coveted Cinchona. Emerging Infectious Diseases, 21(7), 1280–1281.

Bryant, J. E., Holmes, E. C., & Barrett, A. D. T. (2007). Out of Africa: A Molecular Perspective on the Introduction of Yellow Fever Virus into the Americas. PLoS Pathogens, 3(5), e75.

Brynildsrud, O. B., Pepperell, C. S., Suffys, P., Grandjean, L., Monteserin, J., Debech, N., … Eldholm, V. (2018). Global expansion of Mycobacterium tuberculosis lineage 4 shaped by colonial migration and local adaptation. Science Advances, 4(10), eaat5869.

Buzic, I., & Giuffra, V. (2020). The paleopathological evidence on the origins of human tuberculosis: a review. Journal of Preventive Medicine and Hygiene, 61(1), 3–8.

Chippaux, J.-P., & Chippaux, A. (2018). Yellow fever in Africa and the Americas: a historical and epidemiological perspective. Journal of Venomous Animals and Toxins Including Tropical Diseases, 24(1), 20.

Comas, I., Coscolla, M., Luo, T., Borrell, S., Holt, K. E., Kato-Maeda, M., … Gagneux, S. (2013). Out-of-Africa migration and Neolithic co-expansion of Mycobacterium tuberculosis with modern humans. Nature Genetics, 45(10), 1176–1182.

Darwin, C. (1832-1836). Journal of Researches into the Geology and Natural History of the Various Countries Visited by HMS Beagle under the Command of Captain Fitzroy, RN from 1832 to 1836. London: Henry Colburn.

Dorp, L. van, Gelabert, P., Rieux, A., Manuel, M. de, de-Dios, T., Gopalakrishnan, S., … Lalueza-Fox, C. (2020). Plasmodium vivax Malaria Viewed through the Lens of an Eradicated European Strain. Molecular Biology and Evolution, 37(3), 773–785.

Düx, A., Lequime, S., Patrono, L. V., Vrancken, B., Boral, S., Gogarten, J. F., … Calvignac-Spencer, S. (2020). Measles virus and rinderpest virus divergence dated to the sixth century BCE. Science, 368(6497), 1367–1370.

Echeverria, L. E., & Morillo, C. A. (2019). American Trypanosomiasis (Chagas Disease). Infectious Disease Clinics of North America, 33(1), 119–134.

Emory University. Slave Voyages: The Trans-Atlantic Slave Trade Database. Retrieved June 26, 2021, from https://www.slavevoyages.org/

Fernandes, A., Iñiguez, A. M., Lima, V. S., Souza, S. M. de, Ferreira, L. F., Carolina, A., & Jansen, A. M. (2008). Pre-Columbian Chagas disease in Brazil: Trypanosoma cruzi I in the archaeological remains of a human in Peruạcu Valley, Minas Gerais, Brazil. Memórias Do Instituto Oswaldo Cruz, 103(5), 514–516.

Ferreira, L. F., Britto, C., Cardoso, M. A., Fernandes, O., Reinhard, K., & Araújo, A. (2000). Paleoparasitology of Chagas disease revealed by infected tissues from Chilean mummies. Acta Tropica, 75(1), 79–84.

Gagneux, S. (2018). Ecology and evolution of Mycobacterium tuberculosis. Nature Reviews Microbiology, 16(4), 202–213.

Galaway, F., Yu, R., Constantinou, A., Prugnolle, F., & Wright, G. J. (2019). Resurrection of the ancestral RH5 invasion ligand provides a molecular explanation for the origin of P. falciparum malaria in humans. PLOS Biology, 17(10), e3000490.

Gelabert, P., Sandoval-Velasco, M., Olalde, I., Fregel, R., Rieux, A., Escosa, R., … Lalueza-Fox, C. (2016). Mitochondrial DNA from the eradicated European Plasmodium vivax and P. falciparum from 70-year-old slides from the Ebro Delta in Spain. Proceedings of the National Academy of Sciences, 113(41), 11495–11500.

Giffin, K., Lankapalli, A. K., Sabin, S., Spyrou, M. A., Posth, C., Kozakaitė, J., … Bos, K. I. (2020). A treponemal genome from an historic plague victim supports a recent emergence of yaws and its presence in 15th century Europe. Scientific Reports, 10(1), 9499.

Guerra, F. (1970). American trypanosomiasis. An historical and a human lesson. Journal of Tropical Medicine and Hygiene, 73(5), 105–118. Retrieved from https://www.cabdirect.org/cabdirect/abstract/19702903042

Guhl, F., Auderheide, A., & Ramírez, J. D. (2014). From ancient to contemporary molecular eco-epidemiology of Chagas disease in the Americas. International Journal for Parasitology, 44(9), 605–612.

Gutierrez, C. M., Brisse, S., Brosch, R., Fabre, M., Omaıs, B., Marmiesse, M., … Vincent, V. (2005). Ancient origin and gene mosaicism of the progenitor of Mycobacterium tuberculosis. PLOS Pathogens, 1(1), e5.

Hamilton, P. B., Teixeira, M. M. G., & Stevens, J. R. (2012). The evolution of Trypanosoma cruzi: the 'bat seeding' hypothesis. Trends in Parasitology, 28(4), 136–141.

Harper, K. N., Ocampo, P. S., Steiner, B. M., George, R. W., Silverman, M. S., Bolotin, S., … Armelagos, G. J. (2008). On the Origin of the Treponematoses: A Phylogenetic Approach. PLoS Neglected Tropical Diseases, 2(1), e148.

Harper, K. N., Zuckerman, M. K., Harper, M. L., Kingston, J. D., & Armelagos, G. J. (2011). The origin and antiquity of syphilis revisited: An Appraisal of Old-World pre-Columbian evidence for treponemal infection. American Journal of Physical Anthropology, 146(S53), 99–133.

Herbig, A., Maixner, F., Bos, K. I., Zink, A., Krause, J., & Huson, D. H. (2016). MALT: Fast alignment and analysis of metagenomic DNA sequence data applied to the Tyrolean Iceman. BioRxiv, 050559. https://doi.org/10.1101/050559

Jagannathan, P., & Kakuru, A. (2022). Malaria in 2022: Increasing challenges, cautious optimism. Nature Communications, 13(1), 2678.

Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. F., & Orlando, L. (2013). mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. Bioinformatics, 29(13), 1682–1684.

Krief, S., Escalante, A. A., Pacheco, M. A., Mugisha, L., André, C., Halbwax, M., … Snounou, G. (2010). On the diversity of malaria parasites in African apes and the origin of Plasmodium falciparum from bonobos. PLoS Pathogens, 6(2), e1000765–e1000765.

Laxao, M. (2016). From Citizens to "Natives": Tropical Politics of Depopulation at the Panama Canal Zone. Environmental History, 21(2), 240–249.

Lee, O. Y. C., Wu, H. H. T., Donoghue, H. D., Spigelman, M., Greenblatt, C. L., Bull, I. D., … Besra, G. S. (2012). Mycobacterium tuberculosis complex lipid virulence factors preserved in the 17,000-year-old skeleton of an extinct bison, bison antiquus. PLoS ONE, 7(7), e41923.

Li, Y., & Yang, Z. (2017). Adaptive Diversification between Yellow Fever Virus West African and South American Lineages: A Genome-Wide Study. The American Journal of Tropical Medicine and Hygiene, 16–0698.

Lichtor, J., & Lichtor, A. (1957). Paleopathological evidence suggesting pre-Columbian tuberculosis of the spine. The Journal of Bone and Joint Surgery, 39 A(6), 1398–1399.

Lima, V. S., Iniguez, A. M., Otsuki, K., Ferreira, L. F., Araújo, A., Vicente, A. C. P., & Jansen, A. M. (2008). Chagas Disease in Ancient Hunter-Gatherer Population, Brazil. Emerging Infectious Diseases, 14(6), 1001–1002.

Lindo, J., Haas, R., Hofman, C., Apata, M., Moraga, M., Verdugo, R. A., … Rienzo, A. D. (2018). The genetic prehistory of the Andean highlands 7000 years BP through European contact. Science Advances, 4(11), eaau4921.

Lindo, J., Huerta-Sánchez, E., Nakagome, S., Rasmussen, M., Petzelt, B., Mitchell, J., … Malhi, R. S. (2016). A time transect of exomes from a Native American population before and after European contact. Nature Communications, 7(1), 13175.

Loy, D. E., Liu, W., Li, Y., Learn, G. H., Plenderleith, L. J., Sundararaman, S. A., … Hahn, B. H. (2017). Out of Africa: origins and evolution of the human malaria parasites Plasmodium falciparum and Plasmodium vivax. International Journal for Parasitology, 47(2–3), 87–97.

Loy, D. E., Plenderleith, L. J., Sundararaman, S. A., Liu, W., Gruszczyk, J., Chen, Y. J., … Hahn, B. H. (2018). Evolutionary history of human Plasmodium vivax revealed by genome-wide analyses of related ape parasites. Proceedings of the National Academy of Sciences of the United States of America, 115(36), E8450–E8459.

Majander, K., Pfrengle, S., Kocher, A., Neukamm, J., Plessis, L. du, Pla-Díaz, M., … Schuenemann, V. J. (2020). Ancient Bacterial Genomes Reveal a High Diversity of Treponema pallidum Strains in Early Modern Europe. Current Biology, 30(19), 3788–3803.

Melo, F. L. D., Mello, J. C. M. D., Fraga, A. M., Nunes, K., & Eggers, S. (2010). Syphilis at the crossroad of phylogenetics and paleopathology. PLoS Neglected Tropical Diseases, 4(1), 575.

Miles, M. A. (2004). The discovery of Chagas disease: progress and prejudice. Infectious Disease Clinics of North America, 18(2).

Molina-Cruz, A., & Barillas-Mury, C. (2014). The remarkable journey of adaptation of the Plasmodium falciparum malaria parasite to New World anopheline mosquitoes. Memorias Do Instituto Oswaldo Cruz, 109(5), 662–667.

Molina-Cruz, A., Garver, L. S., Alabaster, A., Bangiolo, L., Haile, A., Winikor, J., … Barillas-Mury, C. (2013). The Human Malaria Parasite Pfs47 Gene Mediates Evasion of the Mosquito Immune System. Science, 340(6135), 984–987.

Molyneux, D. H., Asamoa-Bah, A., Fenwick, A., Savioli, L., & Hotez, P. (2021). The history of the neglected tropical disease movement. Transactions of the Royal Society of Tropical Medicine and Hygiene, 115(2), 169–175.

Monath, T. P., & Vasconcelos, P. F. C. (2015). Yellow Fever. Journal of Clinical Virology, 64, 160–173.

Nelson, E. A., Buikstra, J. E., Herbig, A., Tung, T. A., & Bos, K. I. (2020). Advances in the molecular detection of tuberculosis in pre-contact Andean South America. International Journal of Paleopathology, 29, 128–140. https://doi.org/10.1016/j.ijpp.2019.12.006

Ng, T. F. F., Chen, L.-F., Zhou, Y., Shapiro, B., Stiller, M., Heintzman, P. D., … Delwart, E. (2014). Preservation of viral genomes in 700-y-old caribou feces from a subarctic ice patch. Proceedings of the National Academy of Sciences, 111(47), 16842–16847.

Oliveira, T. C. de, Rodrigues, P. T., Menezes, M. J., Goņcalves-Lopes, R. M., Bastos, M. S., Lima, N. F., … Ferreira, M. U. (2017). Genome-wide diversity and differentiation in New World populations of the human malaria parasite Plasmodium vivax. PLoS Neglected Tropical Diseases, 11(7), e0005824.

Parascandola, J. (1974). From MCWA to CDC: Origins of the Centers for Disease Control and Prevention. Public Health Reports, 111(6), 549–551.

Pedersen, D. D., Milner, G. R., Kolmos, H. J., & Boldsen, J. L. (2019). The association between skeletal lesions and tuberculosis diagnosis using a probabilistic approach. International Journal of Paleopathology, 27, 88–100.

Pineda, C, Mansilla, J., Pijoán, C., Martínez-Lavín, S. F., & Martínez-Lavín, M. (1998). Radiographs of an Ancient Mortuary Bundle Support Theory for the New World Origin of Syphilis. American Journal of Roentgenology.

Pineda, Carlos, Mansilla-Lory, J., Martínez-Lavín, M., Leboreiro, I., Izaguirre, A., & Pijoan, C. (2009). Rheumatic Diseases in the Ancient Americas: The Skeletal Manifestations of Treponematoses. JCR: Journal of Clinical Rheumatology, 15(6), 280–283.

Pinto, C. M., Ocaña-Mayorga, S., Tapia, E. E., Lobos, S. E., Zurita, A. P., Aguirre-Villacís, F., … Perkins, S. L. (2015). Bats, Trypanosomes, and Triatomines in Ecuador: New Insights into the Diversity, Transmission, and Origins of Trypanosoma cruzi and Chagas Disease. PLOS ONE, 10(10), e0139999.

Prugnolle, F., Durand, P., Neel, C., Ollomo, B., Ayala, F. J., Arnathau, C., … Renaud, F. (2010). African great apes are natural hosts of multiple related malaria species, including Plasmodium falciparum. Proceedings of the National Academy of Sciences of the United States of America, 107(4), 1458–1463.

Ritchie, W. A. (1952). Paleopathological evidence suggesting pre-Columbian tuberculosis in New York State. American Journal of Physical Anthropology, 10(3), 305–318.

Rodrigues, P. T., Valdivia, H. O., Oliveira, T. C. D., Alves, J. M. P., Duarte, A. M. R. C., Cerutti-Junior, C., … Ferreira, M. U. (2018). Human migration and the spread of malaria parasites to the New World. Scientific Reports, 8(1), 1–13.

Rothschild, B. M., Martin, L. D., Lev, G., Bercovier, H., Bar-Gal, G. K., Greenblatt, C., … Brittain, D. (2001). Mycobacterium tuberculosis complex DNA from an extinct bison dated 17,000 years before the present. Clinical Infectious Diseases, 33(3), 305–311.

Schuenemann, V. J., Lankapalli, A. K., Barquera, R., Nelson, E. A., Hernández, D. I., Alonzo, V. A., … Krause, J. (2018). Historic Treponema pallidum genomes from Colonial Mexico retrieved from archaeological remains. PLOS Neglected Tropical Diseases, 12(6), e0006447.

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

Smith, O., Dunshea, G., Sinding, M.-H. S., Fedorov, S., Germonpre, M., Bocherens, H., & Gilbert, M. T. P. (2019). Ancient RNA from Late Pleistocene permafrost and historical canids shows tissue-specific transcriptome survival. PLoS Biology, 17(7), e3000166.

Smith-Guzmán, N. E. (2015). The skeletal manifestation of malaria: An epidemiological approach using documented skeletal collections. American Journal of Physical Anthropology, 158(4), 624–635.

Stead, W. W. (2001). Variation in vulnerability to tuberculosis in America today: random, or legacies of different ancestral epidemics? The International Journal of Tuberculosis and Lung Disease, 5(9), 807–814.

Stead, W. W., Eisenach, K. D., Cave, M. D., Beggs, M. L., Templeton, G. L., Thoen, C. O., & Bates, J. H. (1995). When did Mycobacterium tuberculosis infection first occur in the New World? An important question with public health implications. American Journal of Respiratory and Critical Care Medicine, 151(4), 1267–1268.

Steverding, D. (2014). The history of Chagas disease. Parasites & Vectors, 7(1), 317.

Sutter, P. S. (2016). “The First Mountain to Be Removed”: Yellow Fever Control and the Construction of the Panama Canal. Environmental History, 21(2), 250–259.

Toyne, J. M., Esplin, N., & Buikstra, J. E. (2020). Examining variation in skeletal tuberculosis in a late pre-contact population from the eastern mountains of Peru. International Journal of Paleopathology, 30, 22–34.

Wilbur, A. K., & Buikstra, J. E. (2006). Patterns of tuberculosis in the Americas - How can modern biomedicine inform the ancient past? Memorias Do Instituto Oswaldo Cruz, 101(2), 59–66.

Wirth, T., Hildebrand, F., Allix-Béguec, C., Lbeling, F. W., & Kubica, T. (2008). Origin, spread and demography of the Mycobacterium tuberculosis complex. PLoS Pathogens, 4(9), 1000160.

Yalcindag, E., Elguero, E., Arnathau, C., Durand, P., Akiana, J., Anderson, T. J., … Prugnolle, F. (2012). Multiple independent introductions of Plasmodium falciparum in South America. Proceedings of the National Academy of Sciences of the United States of America, 109(2), 511–516.

# CHAPTER 3: GENOMIC EVIDENCE FOR ADAPTATION TO HIGH ALTITUDE AND TUBERCULOSIS IMMUNITY IN THE ANDES BEFORE EUROPEAN CONTACT

Publication Information:

S.K. Joseph, N.R. Migliore, A. Olivieri, A. Torroni, A.C. Owings, M. DeGiorgio, W.G. Ordóñez, J.J. Ortiz Aguilú, F. González-Andrade, A. Achilli, J. Lindo, Genomic evidence for adaptation to tuberculosis in the Andes before European contact. *iScience* 26(2) (2023).

## *3.1 Abstract*

Most studies focusing on human high-altitude adaptation in the Andean highlands have thus far been focused on Peruvian populations. We present high-coverage whole genomes from Indigenous people living in the Ecuadorian highlands and perform multi-method scans to detect positive natural selection. We identified regions of the genome that show signals of strong selection to both cardiovascular and hypoxia pathways, which are distinct from those uncovered in Peruvian populations. However, the strongest signals of selection were related to regions of the genome that are involved in immune function related to tuberculosis. Given our estimated timing of this selection event, the Indigenous people of Ecuador may have adapted to *Mycobacterium tuberculosis* thousands of years before the arrival of Europeans. Furthermore, we detect a population collapse that coincides with the arrival of Europeans, which is more severe than other regions of the Andes, suggesting differing effects of contact across high-altitude populations.

## *3.2 Introduction*

According to archeological evidence, people have inhabited the Andes region of South America for over 12,000 years, including areas encompassing parts of modern-day Ecuador, Peru, Bolivia, and Chile. For millennia, the Andes have been home to many complex and interconnected societies [1–3]. After the arrival of Europeans in the Andes, approximately 500 years before present

(BP), Indigenous populations of the region suffered severe population contractions as a result of warfare, slavery, cultural changes, and the introduction of novel pathogens—notably, smallpox and influenza [3,4]. Additionally, European colonizers carried strains of tuberculosis (TB) to the Americas. Genetic and bioarchaeological evidence suggest the possible presence of other TB strains in the Americas long before contact [5–7]. However, its prevalence throughout South America before European contact is still not well understood.

Another poorly understood aspect of the Andes are the genetic underpinnings of high-altitude adaptation. Relevant studies have thus far been focused on the Quechua and Aymara populations from Peru and Bolivia, which represent most of the genetic research in this area of the world [8–10]. Other high-altitude populations, such as Tibetans and Ethiopians, exhibit population-specific alleles within genes involved with the hypoxia inducible factor (HIF) and nitric oxide pathways [11,12]. However, previous studies in high-altitude Peruvian and Bolivian populations generally lack strong evidence for selection on population-specific alleles relating to the adaptation to hypoxic conditions in the Andean Highlands [8–10].

Currently, genomic studies are lacking in collaboration with Indigenous peoples from Ecuador, including those living in the Highlands. Here we aim to investigate the genetic underpinnings of hypoxia adaptation by utilizing genomic data from two Indigenous Ecuadorian populations who are closely related. We report the results of scans for positive selection using allele-frequency, haplotype-based, and admixture-aware methods with whole genome sequence data from individuals of greater than 98% estimated Indigenous ancestry from Ecuador. We find signatures of selection detected at loci related to hypoxia and cardiovascular pathways, which differ from other populations in the Andean highlands. However, the strongest signals of selection were not associated with high-altitude. Instead, we find surprising evidence of selection on loci

which may be linked to the immune response to TB, occurring thousands of years before the arrival of Europeans and highlighting questions regarding the pre-contact prevalence of TB in the ancient Andes.

### *3.3 Results*

Demographic Analyses: We sequenced high-coverage whole genomes from fifteen Indigenous individuals living in several Ecuadorian provinces from the high-altitude (above 2,500 meters) Andean region (Figure 4A, Appendix B Table B1). Seven of the participants live in neighboring communities from the northern provinces of Ecuador, which include Karanki, Kisapincha, Panzaleo, and Otavalo. Eight participants live in the Loja province in southern Ecuador, including individuals from the Oñacapac and Gañil communities. Community engagement is detailed under Methods and a detailed plan for the dissemination of the results are detailed in Appendix B Figure B2.

To facilitate an evolutionary comparison between the different populations, we first conducted demographic analyses and found that the two population groups (geographically separated, Figure 4A) have a close ancestral relationship but remain distinct, despite being linguistically connected. We performed principal components analysis, generated maximum likelihood trees, and examined population substructure via cluster analysis (Figure 4) [13–15]. Our cluster analysis revealed that the 15 high-altitude individuals from Ecuador exhibit greater than 98% estimated Indigenous American ancestry (Figure 4E) and share their own branch on the maximum likelihood tree, with the Quechua of Peru (also from the Andean highlands) forming an adjacent branch (Figure 4C). For the ancestry cluster analyses, we also included high-coverage ancient samples (Rio Uncallane) from Lake Titicaca (3,800m elevation), Peru, which date to ~1800 years BP [3]. Interestingly, we find that the high-altitude Quechua populations share three ancestries,

which include clusters seen in the ancient Andeans (purple) and contemporary populations from North and Central American populations (predominantly blue with some teal). The Northern Kichwa population shares the contemporary American ancestry clusters (predominantly teal with some blue), with a few individuals showing some of the ancient Andean cluster (purple), while the populations from Loja do not exhibit as much of this ancient ancestry. Combined with the maximum likelihood tree result, both the Ecuadorian highland populations seem somewhat differentiated from the high-altitude Quechua. This is a surprising result given their geographic proximity and their connection to the ancient empires of the region, including the Inca. This ancestral distinction may possibly reflect limited gene flow and perhaps differing selection pressures. Lastly, we investigated changes in effective population size in both Indigenous Populations using the program Relate [16]. We found evidence of a bottleneck that coincides with the first entry into the Americas ~20,000 years ago [17]. We also found a population collapse approximately 500 years ago, which coincides with the arrival of the Spanish to the Andean region [18]. Both populations suffer a decline of approximately 80% after contact (Figure 4D). While this level of collapse has been reported in other Indigenous populations using genomic data [19], it's more severe than that reported for high-altitude populations close to the Lake Titicaca region of Peru and Bolivia [3], suggesting different dynamics and consequences of European contact in the Andean highlands [16].

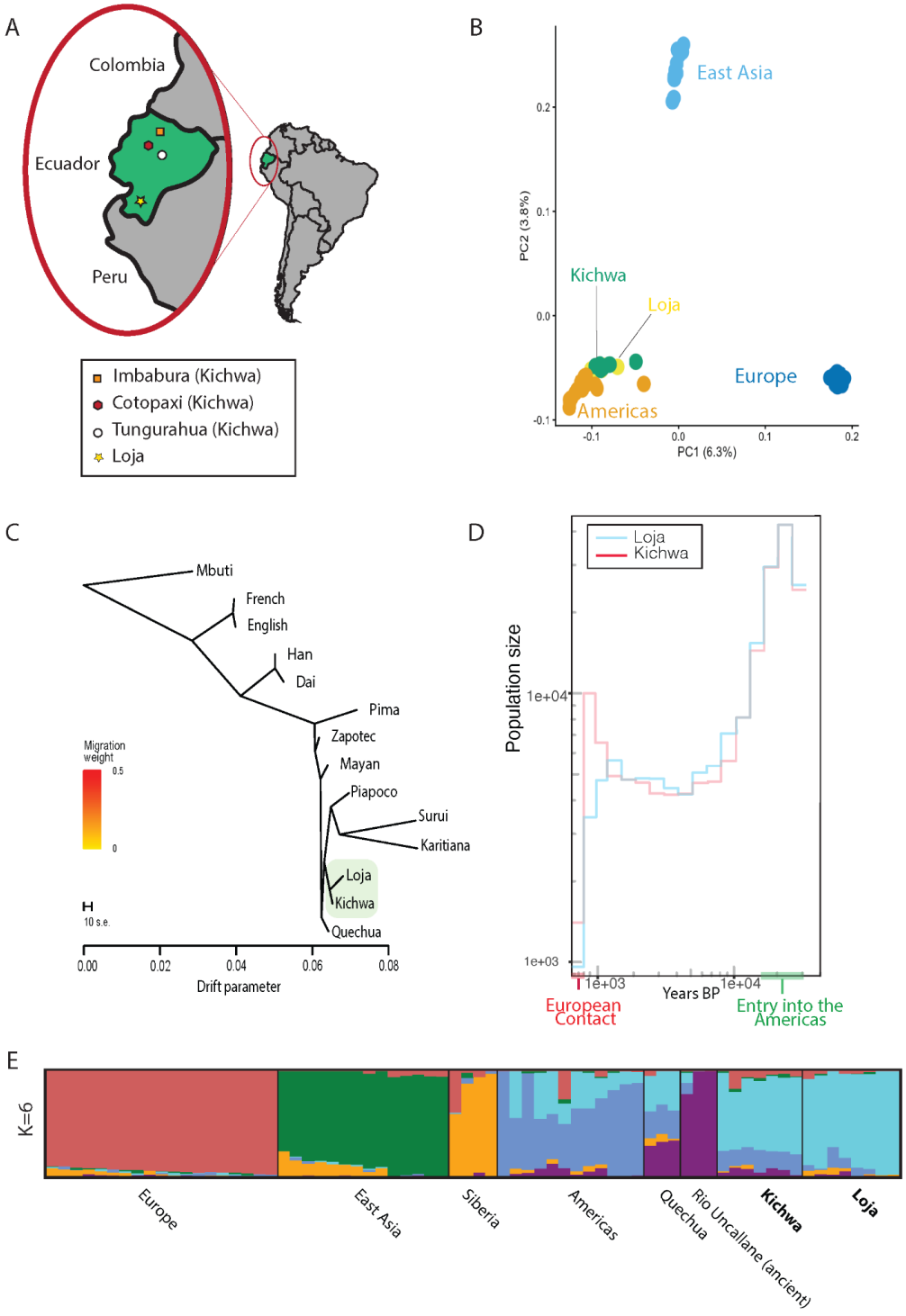



**Figure 4.** Demographic analyses. (A) Map of locations for the two groups. (B) Principal components analysis showing first two principal components, including individuals from this study and individuals from Europe, East Asia, and the Americas obtained from the SGDP dataset. (C) Maximum likelihood trees

generated by TreeMix showing ancestry relationships between Ecuadorian groups and individuals in the SGDP dataset. (D) Changes in effective population size. According to the model, both the Kichwa and Loja populations suffer a population collapse that coincides with arrival of the Spanish to the Andean highlands. (E) Visualization of cluster analysis at K=6, which exhibited the lowest BIC value.

Selection Scans: After establishing the ancestral connection between the two populations, we performed a series of scans for positive selection. We first considered any possible admixture that may have occurred between the groups, which could yield false-positive signals [20]. To accomplish this, we employed the Ohana program, which is a robust approach for detecting positive selection in the presence of admixture, due to its incorporation of cluster analyses to determine the number of underlying ancestral populations [20]. This method predicted the genome-wide variance pattern of the two populations from Ecuador relative to outgroup populations (Figures 5A and 5B). If an allele in the Ecuadorian populations differed significantly from this predicted simulation pattern (using a likelihood ratio test), then the allele was inferred to be under putative positive selection. One of the strongest signals from the scan of the Loja population was associated with intergenic SNP near the *ANXA1* gene, which has an allele frequency of 1.0 in the Loja population (Appendix B Table B3). This gene is important for triggering the cell death of TB-infected monocytes and generating antigens against *M. tuberculosis* for presentation to $CD8^+$ T-cells [21]. Another top candidate for selection is an intronic SNP on the *ANO7* gene. Data from Chromatin immunoprecipitation followed by massively parallel sequencing (ChIP-seq) assays indicate that this specific *ANO7* SNP is a transcription factor binding site for GATA-2, which is implicated in the HIF hypoxia response pathway [22,23]. In the scan of the Kichwa population, among the top hits was an intronic polymorphism of the *GALNT13* gene, which binds HIF and is known to be upregulated under hypoxic conditions [24].

Next, we utilized the phased dataset to scan for positive selection using tests which rely on extended haplotype homozygosity (EHH) [25–27]. The guiding principle of EHH holds that, when comparing two or more populations, the presence of longer haplotypes at a particular locus indicates positive selection, since those regions of the genome have not been broken down by recombination. Thus, the genomes obtained from these Indigenous populations are ideal for use with haplotype-based selection scans because these populations are relatively closely related, meaning that significant differences in haplotype length are likely due to recent positive selection (see Figure 4B for PCA plot comparing individuals from the present study to individuals from the Simons Genome Diversity Project (SGDP) dataset [28]).

We performed an integrated haplotype score (iHS) test, which detects signals of positive selection by utilizing a standardized log ratio of the integrals of observed decay of EHH in a focal population—the Kichwa individuals in the present study (Figure 5C) [25,27,29]. Among the top candidates for selection within the high-altitude individuals is an intronic SNP on the *FBLN1* gene, which codes for a protein important for allowing hemostasis to occur, a critical process involving the coagulation of blood that is greatly affected by hypoxic conditions [30,31]. Top candidate SNPs for positive selection also include an intergenic SNP near the *EGLN3* gene, which codes for an important regular stability of the HIFs in response to hypoxia [32]. However, it should be noted that this particular *EGLN3* SNP, unlike the *EGLN1* polymorphism characteristic of Tibetan populations, has yet to be characterized in the functional genomics literature, though the *EGLN3* gene was recently identified as being under selection in highland populations in Peru [11,33]. Furthermore, an intergenic SNP near the *ANAPC7* gene with significant probability of being under positive selection is a known transcription factor binding site for GATA-2, which has been implicated in the HIF hypoxia response pathway [22,23]. Also among the top hits were genes

implicated in immune and anti-TB response pathways. These included intergenic SNP near the *TRPS1* gene, which, according to data from ChIP-seq, is a binding site for the CEBPB transcription factor, implicated in the early anti-TB response via its role in the expansion of macrophages [22,34]. Additionally, an intergenic polymorphism near the *CD44* gene was among the top candidates for positive selection. *CD44* codes for a well-known cell adhesion molecule involved in immune signaling, which has recently also been shown to serve as an important cell surface binding site for *M. tuberculosis* bacteria on macrophages [35]. *CD44* was recently identified as being under selection in highland populations in Peru [33]. Top hits also included intronic SNPs on the *ADGRE1* gene, an important and well-known marker of macrophages and the *FCRL4* gene, which codes for an immunoglobulin receptor important to the function of memory B-cells [36–38]. Furthermore, an intergenic polymorphism found to be under significant probability of positive selection located near the *KCNA3* gene, a regulator of antigen specificity in memory T-cells, is an eQTL for the *CD53* gene. *CD53* is important for activating and sustaining adaptive immunity through the course of bacterial and viral infections [39–41].

We also performed the same iHS test for the Loja population, which yielded an intergenic SNP near the *SLC8A1* gene, also a binding site for the CEBPB transcription factor (Figure 5D) [25,29,34]. The *SLC8A1* gene was recently identified as being under strong positive selection in highland populations in Peru [33]. An additional top candidate for positive selection was an intergenic SNP near the *CXCL12* gene, which codes for a chemokine known to be upregulated in severe TB infections that include pleurisy, the severe inflammation of thin layers which cover the lungs [42]. Relating to high-altitude, the top hit from this iHS scan was an intergenic SNP near the *HAND1* gene, which is a transcription factor regulated by hypoxic conditions [43]. When upregulated, it has been shown to have a protective effect against myocardial ischemia, which is

the reduction in the heart muscle's ability to pump blood efficiently. It is an integral part of the regulatory connection between environmental oxygen levels and oxygen availability in the blood.

Furthermore, we conducted additional haplotype-based selection scans with the saltiLASSI, method, which uses a composite likelihood ratio test statistic that detects signals of positive selection based on the spatial distribution of the distortions of haplotype frequency spectra across a chromosome [44]. In particular, saltiLASSI has been shown to have high power to detect both hard and soft sweeps from positive selection of recent to moderate age. Applying saltiLASSI to the Kichwa analysis group, one of the regions identified with a high likelihood ratio score and evidence of being an outlier (Figure 5E), and thus likely under positive selection, is a polymorphism on the *PDE11A* gene. According to results from ChIP-seq assays, this SNP is a binding site for the CEBPB transcription factor, implicated in the early anti-TB response via its role in the proliferation of macrophages [22,34]. In the Loja analysis group, we found that an intergenic SNP near the *BRINP3* gene was likely under positive selection (Figure 5F). This gene has been previously identified in Andean Aymara populations in Bolivia as a candidate for adaptation to hypoxic conditions through adaptations of the cardiovascular system, rather than the HIF response pathway; however, this particular SNP was not identified in previous studies [8].

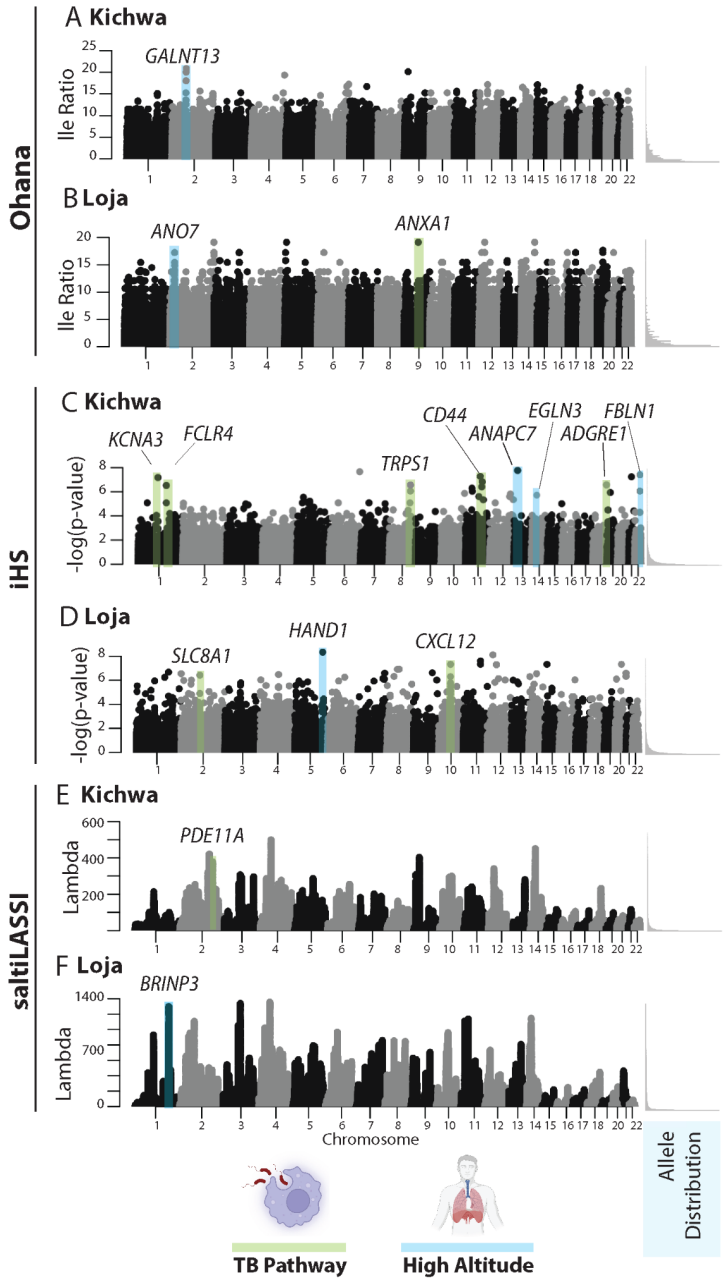



**Figure 5.** Selection scans highlighting the strongest signals. Ohana selscan Manhattan plot and histogram of the allele distribution of the corresponding likelihood ratio (lle ratio) for the Kichwa (A) and Loja (B). iHS Manhattan plot and histogram of the allele distribution at the corresponding -$\log_{10}$(p-value) for the Kichwa (C) and the Loja (D). Manhattan plot of the log composite likelihood ratio ($\Lambda$) from the saltiLASSI program and histogram of the allele distribution at the corresponding $\Lambda$ for the Kichwa (E) and the Loja (F).

To better understand the timing and strength of selection on the alleles described above, we employed the CLUES approximate likelihood method that allows for detecting the allele trajectory, timing of selection, and selection strength [45]. We found evidence that alleles associated with genes thought to be involved in TB resistance, particularly the *PDE11A* (identified in the Kichwa analysis group) and *SLC8A1* (identified in the Loja analysis group) genes, underwent strong positive selection (Figure 6; Appendix B Table B4 and Figure B5). Furthermore, the allele frequency trajectories indicate that these immune-related alleles increased within the population many generations before the arrival of Europeans. Both the *PDE11A* and *SLC8A1* SNP trajectories began rapidly increasing ~3000 BP, and this period roughly corresponds with the transition from small hunter-gatherer groups to larger agricultural societies [1]. These societies were more compact and populous, which could have facilitated the spread of respiratory diseases, such as TB.

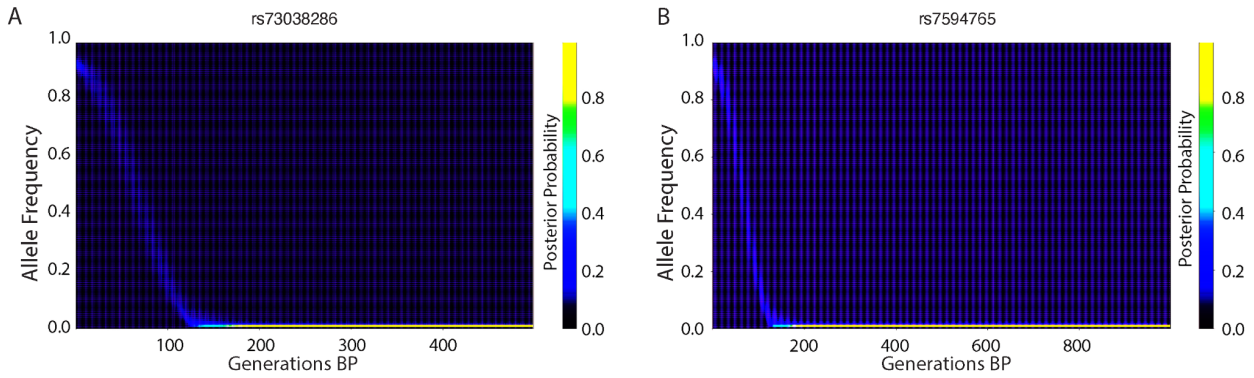



**Figure 6.** Allele trajectories and selection strength. **(**A) Estimated allele trajectory over time (generations before present) for the allele associated with the *PDE11A* gene putatively under selection in the northern Kichwa. (B) Estimated allele trajectory for the allele associated with the *SLC8A1* gene putatively under selection in the Loja population. Both alleles are known binding sites for the CEBPB transcription factor, implicated in the early anti-TB response via its role in the proliferation of monocyte macrophages.

### *3.4 Discussion*

Life in the high-altitude Andes undoubtedly requires many complex adaptations, both cultural and genetic. In the Andean highlands, Indigenous people display higher than average aerobic capacity, known as $VO_2$ max, at high altitude, as well as many other physiological changes [10]. Here, we aimed to shed light on the genetic underpinnings of these physiological adaptations. Our analyses highlighted several alleles with a high probability of positive selection that are located in genes generally thought to be involved in important cardiovascular and hypoxia-related pathways; however, unlike studies from Tibet and Ethiopia, we did not identify specific functionally-characterized polymorphisms [11,12].

One similarity with studies of Tibetan and Ethiopian populations was our identification of a SNP near *EGLN3*, a member of the EGLN gene family, which is the essential regulator of the HIFs [32]. This gene family, including the *EGLN1* gene, has been identified in Tibetan and Peruvian populations as key for aerobic function in hypoxic conditions [10,11]. However, this particular *EGLN3* SNP from the present study has yet to be characterized in the functional genomics literature, though *EGLN3* was recently also identified as being under selection in highland populations in Peru, as was the *GALNT13* gene [33]. Additionally, a SNP near the *BRINP3* gene was identified as a candidate for positive selection in the present study in the Loja analysis group. *BRINP3* is thought to be important for adaptation to hypoxic conditions via the cardiovascular system and has been identified for high probability of recent positive selection in Bolivian Aymara and highland Peruvian populations [8,33]. All other genes with SNPs under high probability of selection related to hypoxia or cardiovascular function in our study differed from those identified in other Andean and worldwide high-altitude population studies (Appendix B Table B6) [8–10]. This suggests there are more complex modes of genetic adaptation to hypoxic conditions in the Andes that are yet to be

understood. Our demographic analyses indicate some divergence between Andean highland populations, which could explain why we see mostly different hypoxia-related selection signals (Figure 4). In other words, there may have been convergent evolution of the adaptation to hypoxic conditions, with different genes involved in each separate population. Alternatively, it could be that a polymorphism detected in our selection scans has just not yet been characterized in the functional genomics literature, or it could be that physiological adaptations are epigenetic in nature.

A surprising outcome of this study was the identification of alleles putatively under strong positive selection in genes that may be related to the anti-TB immune response that arose in the Ecuadorian populations thousands of years before European contact. There is considerable debate regarding the presence of TB in the pre-contact Americas. Although TB's prevalence and geographic spread is not fully understood, both genetic and bioarchaeological evidence has suggested its presence within humans in the ancient Americas at least 1,400 BP [5–7]. However, phylogenetic analyses of TB lineages in the present day show that the L4 phylogenetic lineage left Europe during colonization and remains dominant in the Americas today [7]. These works collectively suggest a possible presence of a pre-contact zoonotic lineage before the L4 European lineage brought by colonists. The SNPs putatively under selection identified in the present study were not only involved in direct bacterial antigen recognition by innate immune macrophages, as with *ANXA1, CD44,* and *ADGRE1,* but also included genes involved in the pro-inflammatory adaptive immune response, such as *KCNA3* and *FCRL4* (Figure 7) [21,35–40]. Future microbiological studies could directly test the alleles identified here for functional consequences with respect to the TB immune response. However, in concert with previously published genetic and bioarchaeological evidence, the genomic evidence presented here represents a plausible scenario

that TB may have been a selective pressure in the agricultural societies of the Andes, pre-European contact [5–7].

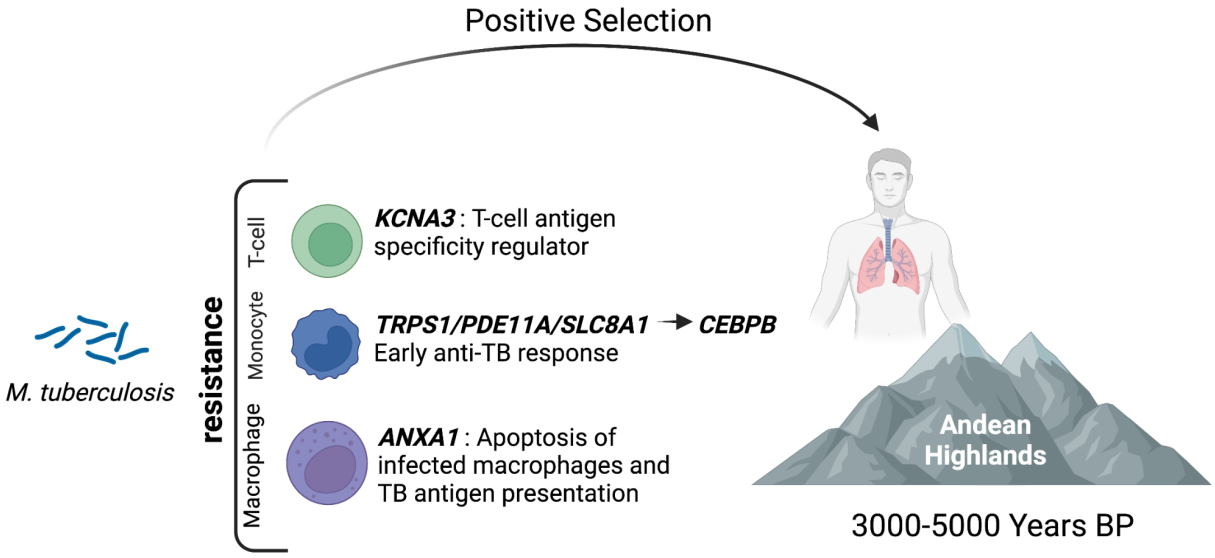



**Figure 7.** Visualization of major lines of evidence relating to alleles under a high probability of positive selection associated with anti-TB immune response. Associated alleles are listed in Appendix B Tables B3 and B4. Created with BioRender.com.

Recent work on human immune response variation among differing TB strains suggests that immunity is not necessarily robust to strain diversity [46]. Thus, in line with our findings, it is possible that Ecuadorian populations could have developed immunity to South American TB strains in the ancient past while still being susceptible to the European L4 lineage at the time of European contact. However, further functional studies on the identified alleles would be needed for definitive evidence. Furthermore, ancient Andean populations at high-altitude did typically suffer lower proportions of population collapse compared with coastal populations, perhaps due in part to the geographic barrier provided by the mountains [3,4]. Today, the Andes region has a higher proportion of TB L4 sublineage S (L4.4) as compared with other parts of Ecuador, which

could be a product of this geographic isolation, though more epidemiological work is needed to understand disease outcomes of this particular sublineage in Ecuadorian populations and how this may relate to population genetic studies such as this one [47].

European colonizers carried many diseases to the Americas, including the TB L4 lineage, which were unquestionably devastating to Indigenous peoples [3,4]. The findings presented here, in concert with other genetic and bioarchaeological evidence, suggest that TB could have been present to some degree for several thousand years in the Andes before Europeans arrived [5–7]. However, the prevalence of TB before European contact is poorly characterized through time, and the present study suggests that the details of human-pathogen coevolution with respect to TB in the Americas is an area of tremendous potential for future research.

### *3.5 Limitations of Study*

Though tremendous advancements have been made in recent years, the size and complexity of the human genome often means that intronic and intergenic polymorphisms−such as those identified in the selection scans from the present study−are not prioritized for functional genomic research. Thus, while future microbiological work could directly test the alleles identified here for functional consequences, it is not possible to know their direct effects on TB immunity or high-altitude adaptation at this time. The present study joins a larger body of previously published genetic and bioarchaeological evidence that TB may have been present in the Andes pre-European contact [5–7] and that humans have adapted uniquely to high-altitude life in the Andes [8-12,33].

### *3.6 Methods: Community Engagement and Sample Collection*

Samples were allocated to the Kichwa and Loja analysis groups based on their geographic location. As this is not an experimental biomedical research study, we do not report the collection of health or immune status, or the use of cell lines.

Kichwa Individuals: Donor blood samples were collected in partnership with Indigenous communities in the northern Ecuadorian provinces of Cotopaxi, Tungurahua, and Imbabura. Adult donor samples and geographic locations were collected in accordance with the ethical standards of the 1964 Declaration of Helsinki and its subsequent amendments. All samples were provided voluntarily and were securely archived and de-identified immediately, and informed consent was offered in Spanish. All sample donors signed the Informed Consent Form (Appendix B Figure B7). Donor samples for this project were approved by the Translational Medicine Unit of the Faculty of Medical Sciences at the Central University of Ecuador.

Loja Individuals: Donor buccal swab samples were collected with Indigenous communities in the southern Ecuadorian province of Loja, as previously published and described by Brandini and colleagues [48]. Adult donor samples and geographic locations were collected in accordance with the ethical standards of the 1964 Declaration of Helsinki and its subsequent amendments. All samples were provided voluntarily and were securely archived and de-identified immediately, and informed consent was offered in both Spanish and English. Donor samples were approved by the Ethic Committee for Clinical Experimentation of the University of Pavia.

Returning Results to Communities: To return results to the Indigenous communities who have contributed to the present study, a series of workshops will be held in several Kichwa-speaking communities located in the Cañar cantón, including Sid-Sid, Amanta Bayopungo, and Quilloac. In addition to describing the purpose and methodologies employed in the present study, the workshops will also be devoted to contextualizing how this population genetic research can be understood within their Andean cultural spaces by collaboratively developing a kind of Community Active Memory (see Appendix B Figure B2 for detailed letter describing the methodology of the workshops).

### *3.7 Methods: DNA Extraction*

DNA from samples obtained from Kichwa individuals was extracted from FTA card blood spots using DNA samples were extracted using a modified Chelex-100 protocol from where samples were rehydrated in 125uL of Chelex-100 and incubated at 56˚C overnight [49]. DNA from samples obtained from Loja individuals was extracted from buccal swabs following standard phenol/chloroform protocol by Brandini and colleagues [48].

### *3.8 Methods: Data Processing, Alignments, and Phasing*

Libraries were prepared with the NEB Ultra II Illumina library kit and corresponding unique dual indexes (New England Biolabs). The libraries underwent whole genome sequencing at a mean of 25x on an Illumina Novaseq 6000 (Dante Labs) (Appendix B Table B1). The sequencing reads were aligned to the hg19 human reference genome using bwa-mem 0.7.17 [50]. Variants were called with BCFtools, where a minimum of 10 reads were required to call a variant, along with base and nucleotide quality set to a minimum of 30 [51]. The combined VCF with the 15 individuals was then annotated using ANNOVAR [52]. The individuals were also assessed for relatedness utilizing both the relatedness and relatedness2 programs through the VCFtools software, and both programs confirmed there were no close familial relationships between individuals [53–55].

For use with haplotype-based selection scans, the genomes were phased without imputation (--impute=FALSE) using the Beagle 5.2 program [56]. VCFtools was used to prepare the VCF by filtering sites out of Hardy-Weinberg equilibrium with a p-value below $10^{-4}$ and removing indels [55]. Default Beagle 5.2 parameters were used, and the SGDP phased dataset was employed as the reference [28].

### *3.9 Methods: Demographic Analyses*

Principal components analysis: Eigenvectors were computed for individuals in Kichwa and Loja Ecuadorian groups from this study with outgroup samples from Europe, East Asia, and the Americas obtained from the SGDP dataset using the Smartpca function of the EIGENSOFT program after pruning whole genome dataset for SNPs in linkage disequilibrium and removing missing data using Plink v1.9 [28,57–59]. The first two principal components were visualized using the PCAviz R package v0.3-37 [14].

Ancestry clustering for maximum likelihood trees: We performed model-based clustering analysis using the likelihood-free approach implemented in SCOPE, which utilizes latent subspace estimation and alternating least squares to estimate admixture proportions from individual allele frequencies [15]. To determine the number of underlying ancestral clusters, we assumed K = 2 to K = 15 ancestral clusters using the Adegenet R package v2.1.7 [60]. For each K, 10,000 iterations were performed (find.clusters(), max.n.clust = 15, n.iter = 1e4, stat = c("BIC")). K = 6 had the lowest Bayesian information criterion (BIC) value, whereby the smallest BIC value indicated that the model with K = 6 ancestral clusters was the best fit for the data. We then ran the SCOPE program using K = 6 ancestral clusters (--k 6) and otherwise default parameters, and visualized with PONG (Appendix B Figures B8 and B9) [15,61].

Maximum likelihood trees: TreeMix was applied to the dataset to generate maximum likelihood trees and admixture graphs from allele frequency data [13]. The Mbuti from the SGDP dataset was used to root the tree (--root Mbuti) [28]. We accounted for linkage disequilibrium by grouping M adjacent sites (with the –k option), and we chose $M$ such that a dataset with $L$ sites will have approximately $L/M \approx 20{,}000$ independent sites. At the end of the analysis (i.e., number of migrations) we performed a global rearrangement (with the –global option). We considered

admixture scenarios with $m = 0$ and $m = 1$ migration events. Each migration scenario was run with 100 replicates, and the replicate with the highest likelihood was chosen to represent the maximum likelihood tree or graph for the given migration scenario (Appendix B Figure B8).

Estimation of changes in effective population size: The Relate program was also used to model changes in the effective population size of both populations using the phased dataset for both the Kichwa and Loja populations [16]. We utilized a mutation rate of $1.25\times10^{-8}$, seed set to 1, a generation time of 28 years, 100 iterations, and 0.5 fraction of trees to be dropped. Autosomal chromosomes 1-22 were used to estimate the population sizes.

### *3.10 Methods: Selection Scans*

The Ohana program uses a likelihood ratio test to detect alleles in each population that deviated strongly from a genome-wide covariance structure [20]. The Ohana program does not require the assignment of population labels to each individual's genome, nor does it utilize phased genomes for haplotype information. Instead, it relies on the genomic data to predict the ancestral population assignments of each genome based on allele frequencies and utilizes the results of cluster analysis to determine the number of these underlying ancestral populations. We conducted two Ohana selection scans, one for each population, and chose the outgroup for the scans to be composed of East Asian and European individuals from the SGDP dataset (see East_Asia and Europe individuals in Appendix B Table B10 for list of SGDP individuals used [28]). To estimate K, the number of underlying ancestral populations in the tree, we chose the K which exhibited the lowest cross-validation index from the ADMIXTURE program [62] after testing K=2 through K=15. We utilized K=3, which is also in accordance with the number of major population groups used as input for the Ohana program—Kichwa or Loja plus the Europe and East Asia outgroup individuals from the SGDP dataset. VCFtools was used to prepare the VCF by filtering sites out of Hardy-

Weinberg Equilibrium with a p-value below $10^{-4}$ and removing indels and missing data [55]. This dataset consists of high-coverage whole genomes, so we utilized the genotypes directly instead of genotype likelihoods. To estimate the correlation structure of each individual to a population tree, we downsampled the dataset at random to 5% of the original variants and then inferred component covariances using Ohana's qpas function to produce admixture-corrected allele frequencies. We then utilized Ohana's selscan function to detect the variants that deviated most strongly from the genome-wide covariance structure, which produces likelihood ratios for each locus in the dataset. A high likelihood ratio indicates a strong deviation from the genome-wide covariance structure, and therefore a strong selection signal. Because these were population-specific scans, not just a global estimate of covariance, the program requires a scalar addition of 10 (h=10) to one position of the covariance matrix corresponding to the focal population being scanned for selection.

We also tested for evidence of selection in our phased dataset using the REHH program, which relies on EHH to calculate the iHS within-population statistic (for each population group) [25–27,29]. The iHS scan compares the integrated EHH profiles between alleles at focal SNPs within the same population. The default parameters were used, with the following modifications: using a frequency bin of 1 (freqbin = 1) to use the entire dataset as a single bin to perform standardization, applying the false discovery rate correction to resulting p-values (p.adjust.method = "FDR"), and not polarizing the data (polarized = F) in order to use major and minor alleles (reference genome hg19), instead of the default ancestral and derived. P-values are presented on a negative $\log_{10}$ scale.

We further explored signatures of positive selection using the composite likelihood ratio test for the saltiLASSI method implemented in the lassip software [44]. This test has high power to detect both hard and soft sweeps from positive selection of recent to moderate age by using the distribution of haplotype frequency spectra across the genome in a single population using phased

whole genome data. For input to this software, we considered only biallelic polymorphic sites with no missing genotypes for any individuals. Using a window of 21 SNPs (--winsize) and a step size of 7 SNPs (--winstep), we computed the haplotype frequency spectrum, truncated at K=10 (--k) distinct haplotypes to obtain the haplotype frequency spectra at each window and an estimate of the genome-wide haplotype frequency spectrum through a mean across all such windows (--avg-spec). Using this mean spectrum as the null haplotype frequency spectrum (--null-spec), we then performed a genome-wide scan for positive selection in both Ecuador populations with the saltiLASSI (--salti) statistic within the lassip software. At each test location (center of each window), saltiLASSI computes a log composite likelihood ratio test statistic ($\Lambda$), for which values close to zero lend little support for positive selection and extreme positive values lend increased support.

CLUES is an approximate likelihood method that detects allele trajectory, timing of selection, and selection strength of a particular SNP of interest [45]. We ran the CLUES program for our 9 anti-TB response-related SNPs from the various scans above using a generation length of 28 years. We first estimated the genome-wide genealogies and coalescent rate using the Relate program on the phased dataset, with a mutation rate of $1.25\times10^{-8}$ [16]. We then sampled the branch lengths with Relate, sampling each branch 1000 times (--num_Samples), and using the output for the selection inferences performed by CLUES (Appendix B Table B4).

The individuals included in this study have very high (>98%) proportions of Indigenous ancestry, so we would not expect SNPs under high probability of selection to come from European haplotypes. Nevertheless, we still confirmed this for all sites of interest described in the Results section using the RFmix local ancestry inference program, which produces global diploid ancestry estimates for each SNP based on haplotypes in a reference panel (see Appendix B Table B10 for

list of Europe reference panel individuals used) [63]. Using a region of the genome 1 million base pairs on either side of the SNP of interest (--analyze-range=), we utilized the default program parameters, apart from specifying a higher number of generations to 12 (--G 12).

### *3.11 References*

1. Rademaker, K., Hodgins, G., Moore, K., Zarrillo, S., Miller, C., Bromley, G.R.M., Leach, P., Reid, D.A., Álvarez, W.Y., and Sandweiss, D.H. (2014). Paleoindian settlement of the high altitude Peruvian Andes. Science 346, 466–469. 10.1126/science.1258260.

2. Posth, C., Nakatsuka, N., Lazaridis, I., Skoglund, P., Mallick, S., Lamnidis, T.C., Rohland, N., Nägele, K., Adamski, N., Bertolini, E., et al. (2018). Reconstructing the Deep Population History of Central and South America. Cell 175, 1185-1197.e22. 10.1016/j.cell.2018.10.027.

3. Lindo, J., Haas, R., Hofman, C., Apata, M., Moraga, M., Verdugo, R.A., Watson, J.T., Llave, C.V., Witonsky, D., Beall, C., et al. (2018). The genetic prehistory of the Andean highlands 7000 years BP through European contact. Sci Adv 4, eaau4921. 10.1126/sciadv.aau4921.

4. Lindo, J., Huerta-Sánchez, E., Nakagome, S., Rasmussen, M., Petzelt, B., Mitchell, J., Cybulski, J.S., Willerslev, E., DeGiorgio, M., and Malhi, R.S. (2016). A time transect of exomes from a Native American population before and after European contact. Nat Commun 7, 13175. 10.1038/ncomms13175.

5. Nelson, E.A., Buikstra, J.E., Herbig, A., Tung, T.A., and Bos, K.I. (2020). Advances in the molecular detection of tuberculosis in pre-contact Andean South America. Int J Paleopathol 29, 128–140. 10.1016/j.ijpp.2019.12.006.

6. Bos, K.I., Harkins, K.M., Herbig, A., Coscolla, M., Weber, N., Comas, I., Forrest, S.A., Bryant, J.M., Harris, S.R., Schuenemann, V.J., et al. (2014). Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. Nature 514, 494–497. 10.1038/nature13591.

7. Brynildsrud, O.B., Pepperell, C.S., Suffys, P., Grandjean, L., Monteserin, J., Debech, N., Bohlin, J., Alfsnes, K., Pettersson, J.O.-H., Kirkeleite, I., et al. (2018). Global expansion of Mycobacterium

tuberculosis lineage 4 shaped by colonial migration and local adaptation. Sci Adv 4, eaat5869. 10.1126/sciadv.aat5869.

8. Crawford, J.E., Amaru, R., Song, J., Julian, C.G., Racimo, F., Cheng, J.Y., Guo, X., Yao, J., Ambale-Venkatesh, B., Lima, J.A., et al. (2017). Natural Selection on Genes Related to Cardiovascular Health in High-Altitude Adapted Andeans. Am J Hum Genetics 101, 752–767. 10.1016/j.ajhg.2017.09.023.

9. Valverde, G., Zhou, H., Lippold, S., Filippo, C. de, Tang, K., Herráez, D.L., Li, J., and Stoneking, M. (2015). A Novel Candidate Region for Genetic Adaptation to High Altitude in Andean Populations. Plos One 10, e0125444. 10.1371/journal.pone.0125444.

10. Brutsaert, T.D., Kiyamu, M., Revollendo, G.E., Isherwood, J.L., Lee, F.S., Rivera-Ch, M., Leon-Velarde, F., Ghosh, S., and Bigham, A.W. (2019). Association of EGLN1 gene with high aerobic capacity of Peruvian Quechua at high altitude. Proc National Acad Sci 116, 24006–24011. 10.1073/pnas.1906171116.

11. Yang, J., Jin, Z.-B., Chen, J., Huang, X.-F., Li, X.-M., Liang, Y.-B., Mao, J.-Y., Chen, X., Zheng, Z., Bakshi, A., et al. (2017). Genetic signatures of high-altitude adaptation in Tibetans. Proc National Acad Sci 114, 4189–4194. 10.1073/pnas.1617042114.

12. Huerta-Sánchez, E., DeGiorgio, M., Pagani, L., Tarekegn, A., Ekong, R., Antao, T., Cardona, A., Montgomery, H.E., Cavalleri, G.L., Robbins, P.A., et al. (2013). Genetic Signatures Reveal High-Altitude Adaptation in a Set of Ethiopian Populations. Mol Biol Evol 30, 1877–1888. 10.1093/molbev/mst089.

13. Pickrell, J.K., and Pritchard, J.K. (2012). Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. Plos Genet 8, e1002967. 10.1371/journal.pgen.1002967.

14. Novembre, J., Williams, R., Pourreza, H., Wang, Y., Carbonetto, P., and Novembre, J. (2013). PCAviz: Visualizing Principal Components Analysis. R package version 0.3-37. R package version 0.3-37.

15. Chiu, A.M., Molloy, E.K., Tan, Z., Talwalkar, A., and Sankararaman, S. (2022). Inferring population structure in biobank-scale genomic data. Am J Hum Genetics 109, 727–737. 10.1016/j.ajhg.2022.02.015.

16. Speidel, L., Forest, M., Shi, S., and Myers, S.R. (2019). A method for genome-wide genealogy estimation for thousands of samples. Nat Genet 51, 1321–1329. 10.1038/s41588-019-0484-x.

17. Raghavan, M., Steinrücken, M., Harris, K., Schiffels, S., Rasmussen, S., DeGiorgio, M., Albrechtsen, A., Valdiosera, C., Ávila-Arcos, M.C., Malaspinas, A.S., et al. (2015). Genomic evidence for the Pleistocene and recent population history of Native Americans. Science 349, aab3884. 10.1126/science.aab3884.

18. Livi-Bacci, M. (2006). The Depopulation of Hispanic America after the Conquest. Popul Dev Rev 32, 199–232. 10.1111/j.1728-4457.2006.00116.x.

19. García-Ortiz, H., Barajas-Olmos, F., Contreras-Cubas, C., Cid-Soto, M.Á., Córdova, E.J., Centeno-Cruz, F., Mendoza-Caamal, E., Cicerón-Arellano, I., Flores-Huacuja, M., Baca, P., et al. (2021). The genomic landscape of Mexican Indigenous populations brings insights into the peopling of the Americas. Nat Commun 12, 5942. 10.1038/s41467-021-26188-w.

20. Cheng, J.Y., Mailund, T., and Nielsen, R. (2017). Fast admixture analysis and population tree estimation for SNP and NGS data. Bioinformatics 33, 2148–2155. 10.1093/bioinformatics/btx098.

21. Tzelepis, F., Gillard, J., Jaworska, J., Nishimura, T., Verway, M., Hassani-Ardakani, K., Remold, H., Vali, H., and Maziar, D. (2013). Critical role of Annexin A1 in protection against pulmonary Mycobacterium tuberculosis infection. 190, P3320.

22. Landt, S.G., Marinov, G.K., Kundaje, A., Kheradpour, P., Pauli, F., Batzoglou, S., Bernstein, B.E., Bickel, P., Brown, J.B., Cayting, P., et al. (2012). ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. Genome Res 22, 1813–1831. 10.1101/gr.136184.111.

23. Tabata, M., Tarumoto, T., Ohmine, K., Furukawa, Y., Hatake, K., Ozawa, K., Hasegawa, Y., Mukai, H., Yamamoto, M., and Imagawa, S. (2001). Stimulation of GATA-2 as a mechanism of hydrogen peroxide suppression in hypoxia-induced erythropoietin gene expression. J Cell Physiol 186, 260–267. 10.1002/1097-4652(200002)186:2<260::aid-jcp1025>3.0.co;2-k.

24. Desai, A.A., Arana, N., Kafisanwo, O., Letsiou, E., Warda, A., Chen, J., Fahs, A., Garcia, J.G.N., and Machado, R.F. (2013). N-Acetylgalactosaminyltransferase 13 (galnt13) Is A Novel Candidate Gene in Pulmonary Arterial Hypertension and In Pathobiologic Responses to Hypoxia. Am. J. Respir. Crit. Care Med. 187, A4645.

25. Gautier, M., Klassman, A., and Vitalis, R. (2017). REHH 2.0: a reimplementation of the R package REHH to detect positive selection from haplotype structure. Molecular Ecology Resources 17, 78–90. 10.1111/1755-0998.12634.

26. Sabeti, P.C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., Xie, X., Byrne, E.H., McCarroll, S.A., Gaudet, R., et al. (2007). Genome-wide detection and characterization of positive selection in human populations. Nature 449, 913–918. 10.1038/nature06250.

27. Sabeti, P.C., Reich, D.E., Higgins, J.M., Levine, H.Z.P., Richter, D.J., Schaffner, S.F., Gabriel, S.B., Platko, J.V., Patterson, N.J., McDonald, G.J., et al. (2002). Detecting recent positive selection in the human genome from haplotype structure. Nature 419, 832–837. 10.1038/nature01140.

28. Mallick, S., Li, H., Lipson, M., Mathieson, I., Gymrek, M., Racimo, F., Zhao, M., Chennagiri, N., Nordenfelt, S., Tandon, A., et al. (2016). The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. Nature 538, 201–206. 10.1038/nature18964.

29. Voight, B.F., Kudaravalli, S., Wen, X., and Pritchard, J.K. (2006). A Map of Recent Positive Selection in the Human Genome. Plos Biol 4, e72. 10.1371/journal.pbio.0040072.

30. Tran, H., Tanaka, A., Litvinovich, S.V., Medved, L.V., Haudenschild, C.C., and Argraves, W.S. (1995). The interaction of fibulin-1 with fibrinogen: A potential role in hemostasis and thrombosis. J Biol Chem 270, 19458–19464. 10.1074/jbc.270.33.19458.

31. Schobersberger, W., Hoffmann, G., and Gunga, H.-C. (2005). Interaction of hypoxia and haemostasis--hypoxia as a prothrombotic factor at high altitude? Wien. Med. Wochenschr. 155, 157–162. 10.1007/s10354-005-0163-7.

32. Ivan, M., and Kaelin, W.G. (2017). The EGLN-HIF O2-Sensing System: Multiple Inputs and Feedbacks. Mol Cell 66, 772–779. 10.1016/j.molcel.2017.06.002.

33. Caro-Consuegra, R., Nieves-Colón, M.A., Rawls, E., Rubin-de-Celis, V., Lizárraga, B., Vidaurre, T., Sandoval, K., Fejerman, L., Stone, A.C., Moreno-Estrada, A., et al. (2022). Uncovering Signals of Positive Selection in Peruvian Populations from Three Ecological Regions. Mol Biol Evol 39, msac158. 10.1093/molbev/msac158.

34. Delgobo, M., Mendes, D.A., Kozlova, E., Rocha, E.L., Rodrigues-Luiz, G.F., Mascarin, L., Dias, G., Patrício, D.O., Dierckx, T., Bicca, M.A., et al. (2019). An evolutionary recent IFN/IL-6/CEBP axis is linked to monocyte expansion and tuberculosis severity in humans. Elife 8, e47013. 10.7554/elife.47013.

35. Leemans, J.C., Florquin, S., Heikens, M., Pals, S.T., Neut, R. van der, and Poll, T. van der (2003). CD44 is a macrophage binding site for Mycobacterium tuberculosis that mediates macrophage recruitment and protective immunity against tuberculosis. J Clin Invest 111, 681–689. 10.1172/jci16936.

36. Waddell, L.A., Lefevre, L., Bush, S.J., Raper, A., Young, R., Lisowski, Z.M., McCulloch, M.E.B., Muriuki, C., Sauter, K.A., Clark, E.L., et al. (2018). ADGRE1 (EMR1, F4/80) Is a Rapidly-Evolving Gene Expressed in Mammalian Monocyte-Macrophages. Front Immunol 9, 2246. 10.3389/fimmu.2018.02246.

37. Liu, Y., Goroshko, S., Leung, L.Y.T., Dong, S., Khan, S., Campisi, P., Propst, E.J., Wolter, N.E., Grunebaum, E., and Ehrhardt, G.R.A. (2020). FCRL4 Is an Fc Receptor for Systemic IgA, but Not Mucosal Secretory IgA. J Immunol 205, 533–538. 10.4049/jimmunol.2000293.

38. Plessis, W.J. du, Walzl, G., and Loxton, A.G. (2016). B cells as multi-functional players during Mycobacterium tuberculosis infection and disease. Tuberculosis 97, 118–125. 10.1016/j.tube.2015.10.007.

39. Chiang, E.Y., Li, T., Jeet, S., Peng, I., Zhang, J., Lee, W.P., DeVoss, J., Caplazi, P., Chen, J., Warming, S., et al. (2017). Potassium channels Kv1.3 and KCa3.1 cooperatively and compensatorily regulate antigen-specific memory T cell functions. Nat Commun 8, 14644. 10.1038/ncomms14644.

40. Dunlock, V.E. (2020). Tetraspanin CD53: an overlooked regulator of immune cell function. Med Microbiol Immun 209, 545–552. 10.1007/s00430-020-00677-z.

41. The GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science 369, 1318–1330 (2020).

42. Kohmo, S., Kijima, T., Mori, M., Minami, T., Namba, Y., Yano, Y., Yoneda, T., Takeda, Y., Kitada, S., Yamaguchi, T., et al. (2012). CXCL12 as a biological marker for the diagnosis of tuberculous pleurisy. Tuberculosis 92, 248–252. 10.1016/j.tube.2012.01.001.

43. Breckenridge, R.A., Piotrowska, I., Ng, K.-E., Ragan, T.J., West, J.A., Kotecha, S., Towers, N., Bennett, M., Kienesberger, P.C., Smolenski, R.T., et al. (2013). Hypoxic Regulation of Hand1 Controls the Fetal-Neonatal Switch in Cardiac Metabolism. Plos Biol 11, e1001666. 10.1371/journal.pbio.1001666.

44. DeGiorgio, M., and Szpiech, Z.A. (2022). A spatially aware likelihood test to detect sweeps from haplotype distributions. PloS Genet 18, e1010134. 10.1101/2021.05.12.443825.

45. Stern, A.J., Wilton, P.R., and Nielsen, R. (2019). An approximate full-likelihood method for inferring selection and allele frequency trajectories from DNA sequence data. Plos Genet 15, e1008384. 10.1371/journal.pgen.1008384.

46. Tientcheu, L.D., Koch, A., Ndengane, M., Andoseh, G., Kampmann, B., and Wilkinson, R.J. (2017). Immunological consequences of strain variation within the Mycobacterium tuberculosis complex. Eur J Immunol 47, 432–445. 10.1002/eji.201646562.

47. Garzon-Chavez, D., Garcia-Bereguiain, M.A., Mora-Pinargote, C., Granda-Pardo, J.C., Leon-Benitez, M., Franco-Sotomayor, G., Trueba, G., and Waard, J.H. de (2020). Population structure and genetic diversity of Mycobacterium tuberculosis in Ecuador. Sci Rep-uk 10, 6237. 10.1038/s41598-020-62824-z.

48. Brandini, S., Bergamaschi, P., Cerna, M.F., Gandini, F., Bastaroli, F., Bertolini, E., Cereda, C., Ferretti, L., Gómez-Carballa, A., Battaglia, V., et al. (2018). The Paleo-Indian Entry into South America According to Mitogenomes. Mol Biol Evol 35, 299–311. 10.1093/molbev/msx267.

49. Walsh, P.S., Metzger, D.A., and Higuchi, R. (1991). Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. Biotechniques 10.

50. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25, 1754–1760.

51. Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., et al. (2021). Twelve years of SAMtools and BCFtools. Gigascience 10, giab008. 10.1093/gigascience/giab008.

52. Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 38, e164–e164. 10.1093/nar/gkq603.

53. Yang, J., Benyamin, B., McEvoy, B.P., Gordon, S., Henders, A.K., Nyholt, D.R., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery, G.W., et al. (2010). Common SNPs explain a large proportion of the heritability for human height. Nat Genet 42, 565–569. 10.1038/ng.608.

54. Manichaikul, A., Mychaleckyj, J.C., Rich, S.S., Daly, K., Sale, M., and Chen, W.M. (2010). Robust relationship inference in genome-wide association studies. Bioinformatics 26, 2867–2873. 10.1093/bioinformatics/btq559.

55. Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al. (2011). The variant call format and VCFtools. Bioinformatics 27, 2156–2158. 10.1093/bioinformatics/btr330.

56. Browning, S.R., and Browning, B.L. (2007). Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. Am J Hum Genetics 81, 1084–1097. 10.1086/521987.

57. Patterson, N., Price, A.L., and Reich, D. (2006). Population Structure and Eigenanalysis. Plos Genet 2, e190. 10.1371/journal.pgen.0020190.

58. Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., and Reich, D. (2006).

Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38, 904–909. 10.1038/ng1847.

59. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 4, 7. 10.1186/s13742-015-0047-8.

60. Jombart, T., and Ahmed, I. (2011). adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics 27, 3070–3071. 10.1093/bioinformatics/btr521.

61. Behr, A.A., Liu, K.Z., Liu-Fang, G., Nakka, P., and Ramachandran, S. (2016). pong: fast analysis and visualization of latent clusters in population genetic data. Bioinformatics 32, 2817–2823. 10.1093/bioinformatics/btw327.

62. Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome Res 19, 1655–1664. 10.1101/gr.094052.109.

63. Maples, B.K., Gravel, S., Kenny, E.E., and Bustamante, C.D. (2013). RFMix: A Discriminative Modeling Approach for Rapid and Robust Local-Ancestry Inference. Am J Hum Genetics 93, 278–288. 10.1016/j.ajhg.2013.06.020.

# CHAPTER 4: PALEOGENOMIC INSIGHTS INTO COOPERATION IN THE ANCIENT ANDES FROM POSITIVE SELECTION ON OXYTOCIN PATHWAY GENES

Publication Information:

S.K. Joseph, E. Wagman, N. Diab, N. Ryu, M. Lee, R. Haas, J.K. Rilling, M.S. Aldenderfer, J. Lindo, Paleogenomic insights into cooperation in the ancient Andes from positive selection on oxytocin pathway genes. *Genes, Brain and Behavior* 23(1) (2024).

## *4.1 Abstract*

Human societies are characterized by norms that restrict selfish behavior and promote cooperation. The oxytocin system is an important modulator of social behavior that may be involved in the evolution of cooperation. Oxytocin acts in both the nucleus accumbens and the anterior cingulate cortex to promote social bonding and social cohesion. Expression of the *CD38* and *OXTR* genes is known to affect oxytocin secretion and binding, respectively, in these brain areas. The Andean highlands provide an excellent opportunity to evaluate the role of oxytocin in the evolution of cooperation. The rich archaeological record spans 13,000 years of population growth and cooperative challenges through periods of highland exploration, hunting economies, agro-pastoralism, and urbanization. Through allele trajectory modeling using both ancient and contemporary whole genomes, we find evidence for strong positive selection on the *OXTR* and *CD38* alleles linked with increased oxytocin signaling. These selection events commenced around 2.5 and 1.25 thousand years ago, placing them in the region's Upper Formative and Tiwanaku periods—a time of population growth, urbanization, and relatively low rates of violence. Along with remarkable and enduring cultural developments, increased oxytocin secretion and receptor binding in these brain areas may have facilitated large-scale cooperation that promoted early urbanization in the Titicaca Basin of the Andean highlands.

### *4.2 The Oxytocin System and Cooperation*

Human societies are uniquely characterized by elaborate social norms that constrain selfish behavior and promote cooperation, even among genetically unrelated individuals (2, 3). The evolution of altruistic social behavior has been a topic of significant scholarly interest, given that individually costly behaviors that benefit non-kin are typically thought to be selected against.

Scholars have considered a number of potential selection pressures to explain the evolution of habitual, non-kin cooperation, or hyper-cooperation (4), in our species. Potential drivers include alloparenting (4), communal hunting (5, 6), environmental harshness (7, 8), urbanization and markets (9), and violence (10). Proposed adaptive responses include both genetic and cultural mechanisms, with genetic adaptations more prevalent in our species' early evolutionary history and cultural adaptations more prevalent in recent times.

The oxytocin (OT) system is an important modulator of vertebrate social behavior (11), and modifications of the oxytocin system could potentially be involved in the evolution of human cooperation. OT is a nine amino acid peptide that is produced in the mammalian hypothalamus (12). Rodent studies strongly implicate OT in attachment and social bonding, including parent-offspring and adult pair-bonding, as well as consolation of distressed conspecifics (13–16). Across a wide range of vertebrates, OT appears to have a "tend and defend" function by facilitating ingroup cooperation accompanied by defensive aggression towards outgroup members (17). For example, in wild chimpanzees, our closest living primate relatives, peripheral OT levels are correlated with parochial cooperation (18). Chimpanzees cooperate with other in-group members to patrol territorial boundaries and defend against other chimpanzee communities. These patrols can result in lethal intergroup aggression (18). In humans, multiple laboratory experiments have shown that OT promotes cooperation and social coordination (19–25). OT has also been shown to

enhance behavioral synchrony (26), to attenuate social anxiety (27, 28), and may even promote trust, though this is still debated and may depend on individual and contextual variables (29–32). Given the role of OT in promoting cooperation, increased OT signaling could be advantageous for in-group and inter-group problem solving during peacetime (33), and also for success in intergroup conflict (34).

The behavioral functions of OT are mediated by its action in the brain. Once synthesized in the hypothalamus, OT is released to multiple brain regions via long-range axonal projection. In mice, OT-containing neurons broadly innervate to nine distinct functional circuits encompassing the cerebral cortex (e.g., orbital cortex, agranular cingulate cortices), subcortical nuclei (e.g., nucleus accumbens, central amygdala), thalamus, hypothalamus, and midbrain (35, 36). Similarly, studies using post-mortem human brains have found OT projections in the prefrontal and limbic cortices, such as the orbitofrontal cortex, insular, anterior cingulate cortex, hippocampus, and amygdala (37, 38), as well as the brainstem (39). Central OT signaling is also mediated by volume transmission, which enables the peptide to passively diffuse and bind to the receptors located in the distant brain regions (40). After the release, OT must bind to OT receptors, which show great heterogeneity in their levels of expression in different brain regions. OT signaling in the brain depends on both OT secretion and levels of the oxytocin receptor, which are controlled by *CD38* and oxytocin receptor gene (*OXTR*), respectively (41, 42). The regulation of gene expression depends on many factors, one of the most important being regulatory regions of genes involved in the OT system. Two brain areas, the nucleus accumbens (NAc) and dorsal anterior cingulate cortex (dACC), appear to be especially relevant for OT's role in cooperation. The NAc is a target of the mesolimbic dopamine system and OT has been shown to act on this region to promote social bonding including both parent-offspring and adult pair-bonding (15). OT signaling in the dACC

is causally linked with various forms of social alignment, including affective synchrony (13) and behavioral coordination between individuals (43).

Single nucleotide polymorphisms (SNPs) are genetic variants that differ among individuals. Many SNPs have been linked to regulation of the OT system. For example, in prairie voles, a SNP on the *OXTR* gene leads to both increased levels of OT receptor in the nucleus accumbens, as well as increased social bonding with a mate (44). Establishing the direct connection between the levels of *OXTR* expression, peptide binding, and social phenotypes is challenging in humans due to technical and ethical considerations. Yet, multiple SNPs in the human *OXTR* have also been implicated in enhanced sociality by regulating OT signaling in the ACC (45, 46) (e.g., rs11706648) and striatal subregions (47) (e.g., rs53576). *CD38*, first identified as a glycoprotein found on the surface of many immune cells, has since been implicated in the synthesis of cyclic ADP-ribose (cADPR), which regulates the movement of $Ca^{2+}$ via ryanodine-sensitive calcium channels (48) (Figure 8). Calcium signaling such as this is crucial for cellular events such as secretion, neurotransmission, or cell migration. *CD38* has been shown to influence the peripheral and central release of OT, thereby modulating social cognition and behaviors (49, 50). *CD38* knockout mice, while capable of producing and packaging OT within the hypothalamus, show reduced peripheral and central OT levels (51) and impairment in social memory (51) and parenting behaviors (52). Notably, direct injection of OT into the NAc normalizes these impairments. Similar patterns of results were found in humans, with the low-expressing SNPs in *CD38* linked with low peripheral OT levels and social deficits (e.g., autism spectrum disorder symptom severity), which were partially rescued following intranasal administration of the peptide (49, 53). This evidence strongly suggests that *CD38* regulates mammalian sociality by influencing the bioavailability of OT in the brain and body (54, 55).

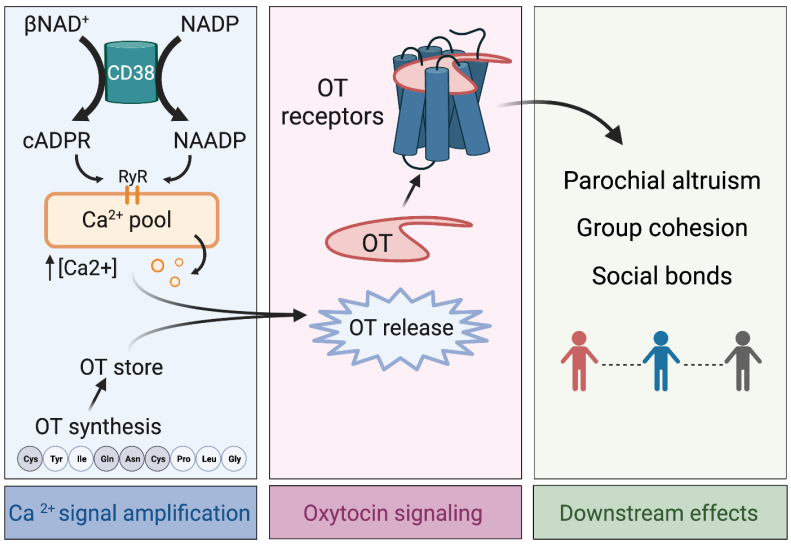



**Figure 8.** Graphic illustrating how OT secretion and receptor binding is thought to be affected by *CD38* and *OXTR* gene expression within the brain. Evidence of selection is found for two key components of OT signaling: the *CD38* gene involved in OT release via calcium signaling and the *OXTR* gene, involved in OT receptor binding. Adapted from (41, 42). Created with BioRender.com.

While we currently lack the ability to determine levels of OT receptor in humans *in vivo*, the GTEx database (56) allows us to identify SNPs associated with *OXTR and CD38* expression in postmortem human brains. SNPs that correlate with the expression of a gene of interest (e.g, *OXTR* and *CD38*), measured with RNAseq, are known as expression quantitative trait loci (eQTLs). It is important to acknowledge that, in general, levels of mRNA do not always perfectly correlate with actual protein levels; however, *OXTR and CD38* gene expression (quantified by mRNA) has been shown to correlate directly with OT release and receptor binding, as well as resulting social behaviors, in mice (42, 48, 51, 54), and prairie voles (57–59). A close overlap between mRNA expression and receptor binding has also been reported in primates (60). Social functions of OT are highly conserved across mammalian species (55), potentially reflecting an overlapping *OXTR* expression profile in the brain (61). Thus, we used *OXTR and CD38* mRNA

expression as a proxy for site-specific endogenous OT signaling in the brain and infer its functional or behavioral effects. Using the GTEx database, the largest and most comprehensive eQTL resource available, we were able to identify 43 SNPs that correlate with *OXTR* expression in the NAc. In addition, we identified 50 SNPs on *CD157* (controls *CD38* expression), which regulates OT secretion (56). If there has been genetic selection for cooperation in human evolution, then we might expect these *OXTR* and *CD38* genes to show evidence of selection during past time periods when cooperative behavior was particularly advantageous, such as with the processes of sedentarization or urbanization.

### *4.3 Cooperation and Socioeconomic Change in the Ancient Andean Highlands*

Human societies have developed a broad range of cultural solutions to solve cooperative dilemmas as populations interacted in increasingly sedentary, dense, and connected settlements. Given the importance of cooperation, there may have also been selection for genes related to improving the capacity for group cohesion throughout human evolution. If so, then we might expect the *OXTR* and *CD38* genes to show evidence of selection in the past. Furthermore, if selection is detected, it may be possible to identify the conditions that favored this genetic pathway to cooperation by estimating the timing of selection. The Titicaca Basin of the South-Central Andean highlands provides an opportunity to evaluate not only the possibility of OT selection, but also the contexts that may have favored its selection.

Archaeological evidence from the Titicaca Basin shows 13,000 years of human population growth and multiple periods of socioeconomic transformation, each presenting novel cooperative challenges (Table 1) (62). During the Paleoindian Period, 13–11 thousand years ago (ka), lowland populations explored the highland's frigid, hypoxic climate conditions and exploited highland plant and animal resources on a logistical or seasonal basis (63). In the subsequent Archaic periods,

11–5 ka, human populations solved the adaptive challenges of permanent highland occupation, most notably by developing thermoregulatory technologies, especially animal hide products for clothing and shelter. Archaic populations were residentially mobile with economies continuing to rely on foraging and hunting (64).

In the Terminal Archaic Period, 5.0–3.5 ka, foraging economies gradually transitioned to agro-pastoral economies and village sedentism (65). Ceramic technologies began to emerge creating new opportunities for food storage and sharing. Wool-textile technology also emerged offering a new thermoregulatory mechanism and creating new opportunities for specialized craft production and market exchange. This is also the period in which multi-family ceremonial architecture (66) and archery technology (67) began to appear in the region. By the Early/Middle Formative Period, 3.5–2.5 ka (1600–500 BCE), agro-pastoral economies were pervasive and intensified (68). Populations began to reside year-round in villages with masonry houses and sunken temples.

During the Upper Formative Period, 2.5–1.6 ka (0–600 CE), the proto-urban center of Pukara emerged in the northern Titicaca Basin with a stepped pyramid, large sunken temples, specialized ceramic production, and centralization of the obsidian exchange, indicating an incipient market economy (69). These developments intensified during the Middle Horizon, 1.6–0.9 ka (600–1100 CE), as the population center shifted from Pukara to Tiwanaku in the southern Titicaca Basin. Tiwanaku covered over 40 hectares and harbored tens of thousands of inhabitants with monumental religious architecture, metal technology, chicha beer brewing, specialized craft production, and new strategies for agricultural intensification (70, 71). The city harbored diverse groups from distant regions such as the Amazon (72) and lay at the heart of expansive cultural

phenomena that incorporated proto-urban centers and agricultural villages over some 12.5 million $km^2$.

Around 1.5 ka (500 CE), the Tiwanaku empire faded, giving way to the Late Intermediate Period, 0.9–0.5 ka (1000–1450 CE). This was a period of socioeconomic fragmentation and the most violent period in the region's history. Defensive architecture was pervasive, and skeletal trauma rates exceeded 30%, more than double the rates observed in most other periods (73). During the subsequent Late Horizon 0.5–0.4 ka (1450–1533 CE), violence rates subsided as the Inca Empire expanded from the north into the Titicaca Basin partially incorporating the region's Indigenous Aymara communities—the likely descendants of the Tiwanaku and Pukara populations (1)—into the Inca political sphere (70).

**Table 1.** Archaeological framework for Lake Titicaca Basin, South-Central Andean highlands.

| period[a] | dates (ka) | economics | violence[b] | reference |
|---|---|---|---|---|
| Late Horizon | 0.6–0.5 | Inca state | 10% | Arkush and Tung 2013 (73) |
| Late Intermediate | 1.0–0.6 | agropastoralism | 35% | Arkush and Tung 2013 (73) |
| Middle Horizon | 1.4–1.0 | urbanism, specialization | 10% | Arkush and Tung 2013 (73) |
| Early Intermediate | 1.9–1.4 | proto-urbanism | 10% | Arkush and Tung 2013 (73) |
| Formative | 3.5–1.9 | agropastoralism | 10% | Arkush and Tung 2013 (73) |
| Terminal Archaic | 5.0–3.5 | mixed subsistence | 10% | Juengst 2017 (74) |
| Archaic | 11–5 | foraging | 20% | Haas and Viviano 2015 (75) |
| Paleoindian | 13–11 | highland exploration | unknown | NA |

[a] chronology following Stanish (2003) (70) and Klink and Aldenderfer (2005) (76)
[b] approximate proportion of individuals with evidence of perimortem violent trauma

This multi-millennial trajectory of human population growth and socioeconomic transformation illustrates the important role of cultural, economic, technological, and religious innovations in accommodating increasing demands for cooperative interactions. What remains unknown is whether genetic changes contributed to the process. If so, we predict there would have been positive selection on genes linked with increased *OXTR* expression in the NAc and dACC, as well as genes linked with OT secretion (*CD38*), at some point in the past.

Furthermore, if OT selection is detected, it may be possible to identify the socioeconomic context that drove selection. If resolving the challenges of high-elevation adaptation was the driver, we would expect the signal to have appeared during the Paleoindian Period, 13–11 ka. If communal hunting was the salient factor, we might expect OT selection to have occurred during the Archaic Period, 11–5 ka. If agro-pastoralism was the selective context, we would expect OT selection during the Terminal Archaic or Formative periods, 5.0-1.9 ka. If urbanization, we would expect OT selection during the Early Intermediate Period or Middle Horizon (1.9–1.0 ka). If violence, we would expect OT selection during the Late Intermediate Period, 0.9–0.5 ka. And finally, if colonial expansion and partial incorporation was at play, we would expect to observe OT selection during the Late Horizon, 0.5–0.4 ka. In sum, we test for positive selection in genes known to be associated with cooperation in a place where Indigenous Andean populations iteratively solved the cooperative challenges of population growth over at least 14 millennia.

### *4.4 Results*

We utilized high-coverage ancient and contemporary genomes from individuals in the central Andes to test for natural selection on OT pathway genes in the Andes region. These individuals are 25 contemporary Peruvians from Lima, located on the coast of Peru, from the 1000 Genomes Project Phase 3 (PEL) with >95% Indigenous ancestry (77) and 5 ancient individuals from Rio Uncallane (1) who lived approximately 1.8–1.5 ka within the southern boundaries of the Pukara polity's territory, which had become a populous political center by 1.7 ka (78) (Appendix C Table C1).

First, to facilitate a meaningful selection study with continuity between population groups, we conducted demographic analyses and found that the ancient and contemporary population groups have a close ancestral relationship. Additional studies have provided evidence that the

ancient Rio Uncallane samples also exhibit continuity with later pre-European contact populations within the region (72). We performed principal components analysis, generated maximum likelihood trees, and examined population substructure via cluster analysis using samples from the Human Genome Diversity Project (HGDP), Simons Genome Diversity Project (SGDP), and ancient DNA datasets from the Americas (Figure 9; Appendix C Table C2) (79–87).

Our cluster analysis revealed that the 25 contemporary Peruvians from Lima (PEL) from the 1000 Genomes Project (77) and 5 ancient individuals from Rio Uncallane (1) exhibit over 95% estimated Indigenous American ancestry (Figure 9D). The different colors in the plot comprising the sections with the ancient and contemporary individuals, which represent ancestry fractions from underlying ancestral populations, show that these individuals cluster almost entirely with ancestry from samples from the Americas (dark blue, light blue, gray, orange). Next, we investigated the genetic relationship via admixture graph models (88) of the ancient Rio Uncallane to the PEL individuals and 24 contemporary Aymara (1) (a high-altitude population from northern Bolivia). A well supported model with 0 migrations shows the ancient and contemporary high-altitude populations forming their own branch, which splits from other South American populations from the SGDP (Figure 9C). Furthermore, the Rio Uncallane form a basal branch to both contemporary high-altitude populations (PEL and Aymara).

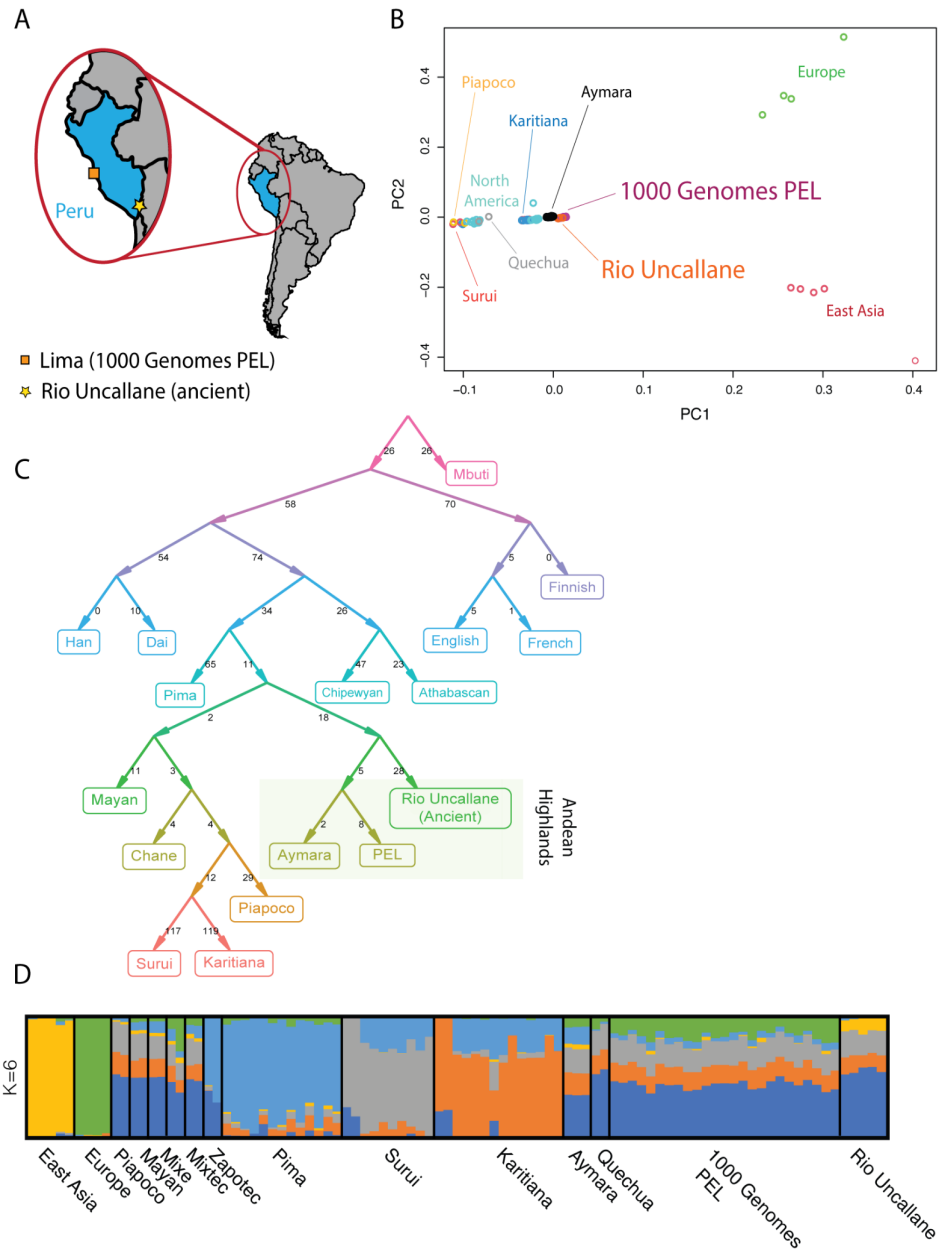



**Figure 9.** Demographic analyses. (A) Map of locations for the two population groups. (B) Principal components analysis showing first two principal components, including individuals from this study and outgroup populations from Europe, East Asia, and the Americas. (C) Admixture graph models showing ancestry relationships between the two sample groups and outgroups. (D) Visualization of cluster analysis at K = 6, which exhibited the lowest BIC value. Colors in the plot represent ancestry fractions from each of 6 ancestral clusters.

Today, the PEL and Aymara populations are not only separated by geographic distance, but also cultural and linguistic differences. Nevertheless, these demographic analyses show they both have a close ancestral relationship to the ancient individuals from Rio Uncallane. Although both contemporary high-altitude populations exhibit a close relation to the ancient population (Figure 9C and 9D), we chose to integrate the PEL population instead of the low-coverage Aymara, to maximize the number of overlapping sites for downstream selection analyses (resulting in a 37% increase in usable sites). We then utilized the GTEx database (56) to identify SNPs associated with *OXTR* (*N = 43*) *and CD38* (*N = 50*) expression in the NAc and dACC in postmortem human brains. We modeled derived allele trajectories for two OT pathway polymorphisms centrally located among SNPs in linkage disequilibrium, rs237893 and rs4389574 (Appendix C Table C3).

We found evidence of strong selection on both alleles tested utilizing ancestral recombination graphs to determine the allele trajectories and corresponding selection coefficients (Figure 10) (89). These selection events, visible as very rapid changes in the modeled allele trajectories, began after 2.5 and 1.25 ka (Figures 10A and 10B). This places the selection events in the Upper Formative and Tiwanaku periods, respectively. Agro-pastoral economies had been well established for at least 1500 years. The proto-urban center of Pukara was on the decline, and the urban center of Tiwanaku was on the ascent. This timing is consistent with a model in which OT selection was among a suite of cultural solutions that facilitated cooperation in increasingly connected settlement systems with dense proto-urban settlements.

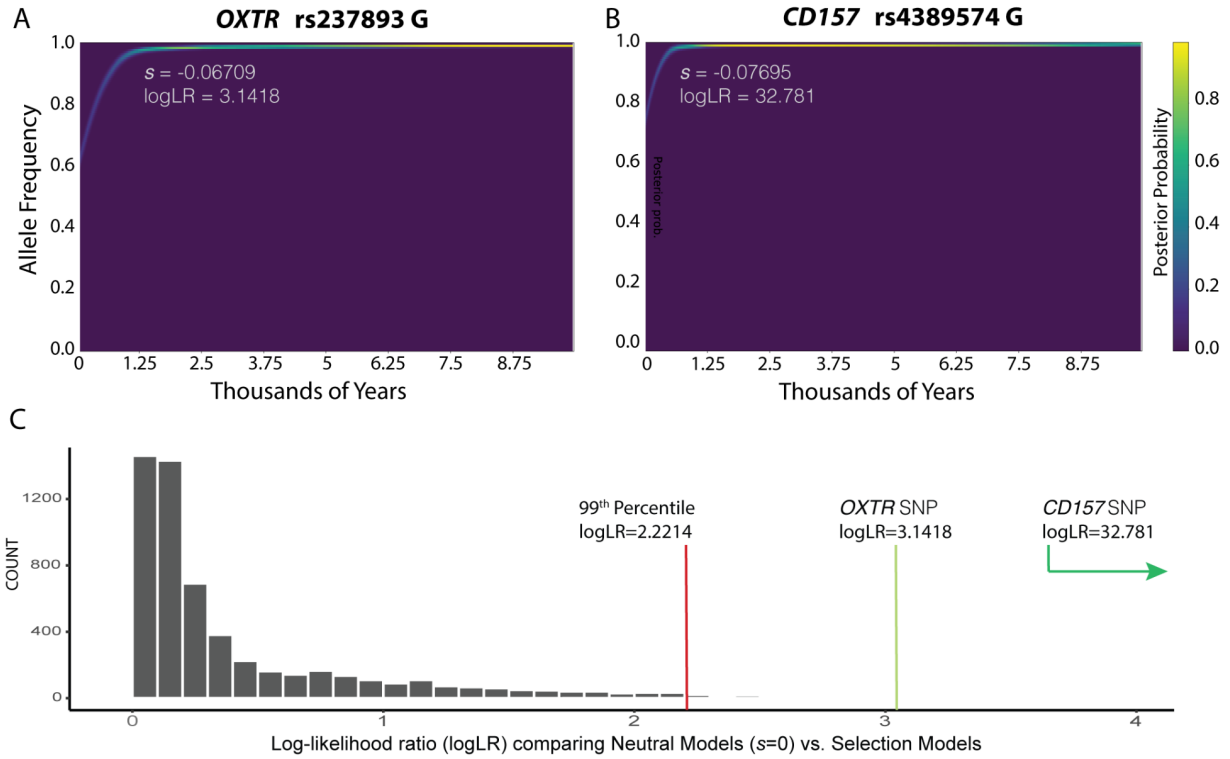



**Figure 10.** Allele trajectories inferred by CLUES. (A, B) Estimated allele trajectory over time for the *OXTR* and *CD157* (controls *CD38* expression) SNPs under putative negative selection. The selection coefficient (*s*) shows the strength of selection and is the most likely *s* inferred by CLUES, while the log-likelihood ratio (LogLR) indicates evidence for rejecting the neutral model, where a near zero ratio would favor neutrality. (C) Distribution of log-likelihood ratios of high-quality SNPs (*see* Methods) tested for selection across chromosome 3 and 4, the chromosomes where the two SNPs are located, with a significance threshold shown at the 99th percentile.

There is a negative allele trajectory for rs237893 and rs4389574 G alleles (Figures 10A and 10B) beginning at approximately 2.5 and 1.25 ka moving towards present day, indicative of rapid positive selection and increase in frequency on the alternate A allele at both genomic positions. Interestingly, the A allele of each SNP is associated with increased OT signaling. For rs237893, the A allele is associated with increased *OXTR* expression in both the nucleus

accumbens and anterior cingulate cortex (56). For rs4389574, the A allele is associated with increased *CD38* expression in both the nucleus accumbens and anterior cingulate cortex (56). Furthermore, we conducted a global selection scan (90) using the entire SGDP dataset (82) composed of 273 individuals from across the world, in order to ascertain how selection at these two loci fits into the larger scope of selection across the genome and across worldwide populations. rs237893 is among the top 12% of SNPs under selection, and rs4389574 is within the top 7% of all SNPs under selection globally (Appendix C Table C4). Therefore, we can conclude that these SNPs were under particularly strong selection in the Andes region. Although not reaching a significance threshold at the top 1%, this may represent older selection events *averaged* across global populations on the same loci.

By testing for positive selection using allele trajectory models informed by both ancient and contemporary whole genome data, we are able to better understand the timing and strength of selection on the alleles described above. We utilized a combination of methods that ascertain genome-wide genealogy to infer coalescence times with recombination and sampling of the resulting trees, at a specific test site, to infer selection parameters (89, 91). With selection strengths greater than $s = 0.01$, which is considered significant (89), these methods maintain optimal (100%) statistical power with a dataset of this size and sample characteristics (89). This is the case for the alleles modeled in the present study (Figure 10).

### *4.5 Discussion*

Collectively, these results suggest that selection for increased signaling (both OT secretion and receptor binding) in the OT pathway occurred at a time when increasingly populous settlements experienced changing socioeconomic systems, including urbanization, new religious traditions, reciprocal labor practices, and agricultural intensification. We posit a conceptual model

that could explain this relationship (Figure 11). Urbanization and socioeconomic integration likely catalyzed genetic adaptations to promote cooperation (70, 92, 93). During the Upper Formative Period, Pukara became a highly connected, proto-urban center specializing in ceramic production and obsidian exchange that likely served the regional agro-pastoral food economy. New forms of agricultural production, known as *qocha* systems, appeared during this time to fuel growing populations.

During the subsequent Middle Horizon, Tiwanaku rose to prominence as an urban center harboring tens of thousands of inhabitants that incorporated individuals from diverse cultural and biological groups (72). The city included elaborate masonry architecture, stepped pyramids, expansive communal plazas, sunken temples, and specialized metal, ceramic, and *chicha* (corn beer) production. Raised-field agricultural systems were developed to support the growing, increasingly urban populations. Agricultural productivity in the high-altitude Andes is particularly tenuous due to the harsh environment where soil erosion, drought, freezing temperatures, and flooding are frequent (94). Large-scale agriculture, which had been present in the Andes for thousands of years (95, 96), required tighter resource and labor control compared with other subsistence styles (e.g., herding) (97). An interdependent and cooperative subsistence style, such as agriculture, especially in harsh environments, would have increased the need for social coordination and cooperation among in-group members (7, 8, 98). The demand for cooperation would have intensified further as communities increasingly engaged with out-group members in market transactions and urban living (9).

The *ayllu* social system of labor distribution and reciprocity—still widely practiced in the region today—may have first appeared in this context to solve the cooperative challenges that come with increasing population density, economic connectivity, and agricultural intensification

beginning in the Early Intermediate period. This complex system consists of networks of families that live within a framework of kinship and territorial linkages, creating opportunities for collective labor, care for ancestral lands, and shared social and ritual practices (99). These new forms of organization would have exerted selective pressure on genes related to cooperation. Thus, enhanced OT signaling in brain regions implicated with in-group bonding (e.g., NAc) and various forms of social coordination (e.g., dACC) may have been selected for under conditions of heightened ecological and social challenges present in the Andes beginning around 2 ka. In fact, evidence of positive selection for the A allele of *OXTR* rs237893 has also been recently found in three contemporary Asian populations (e.g., China, Japan, and Vietnam) characterized by a high degree of socio-ecological threats, cultural tightness, and economic interdependence, to a larger extent than was found in three European populations (e.g., Spain, Finland, and Great Britain) (100). This recent line of research inquiry into the co-evolution of the OT system and social cooperation, inclusive of the present study, represents a potential starting point for a better cross-cultural understanding of the varying biological mechanisms underlying human cooperation around the world.

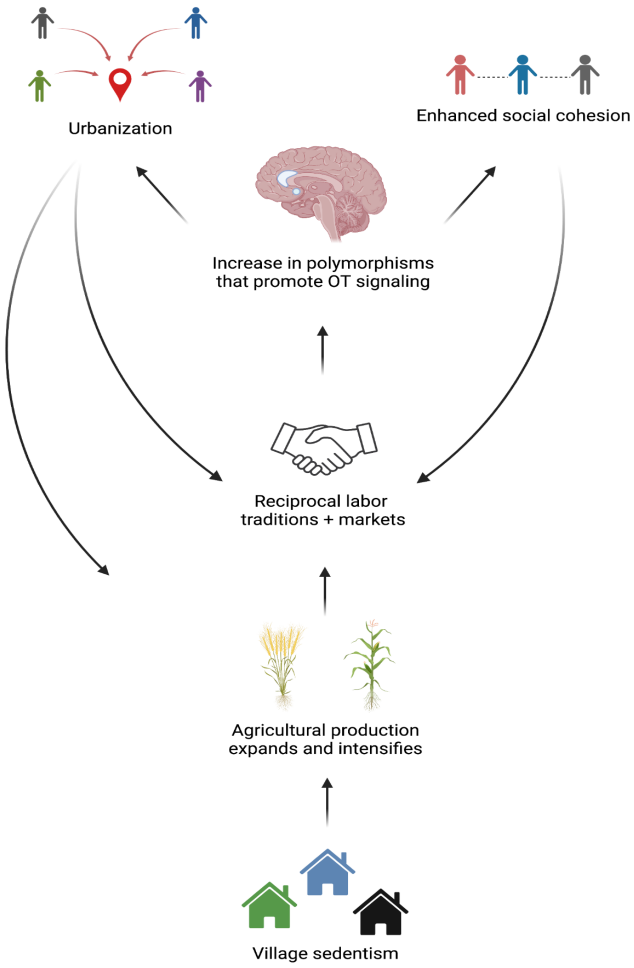



**Figure 11.** Conceptual model of possible explanation for selection for increased signaling in the OT pathway during societal changes in the Andes beginning around 2 ka in the Upper Formative and Tiwanaku periods. Urbanization and agricultural expansion give rise to reciprocal labor traditions, which cause selection for polymorphisms that promote OT signaling, which in turn begins a feedback cycle of increased social cohesion, increase in practice of reciprocal labor traditions, and agricultural intensification that further selects for increased signaling in the OT pathway. Created with BioRender.com.

In a gene-culture coevolutionary feedback process, culturally constructed environments then further select for individuals with *CD38* and *OXTR* alleles that increase OT signaling, as people with more cooperative predispositions are more likely to succeed in these tight societies.

While epigenetic and cultural factors also play an important role in cooperation, social cohesion, and *OXTR* expression (101), we have shown that genes linked with increased *OXTR* expression in the NAc and dACC, as well as those linked with increased OT secretion (*CD38*), underwent rapid positive selection in the Andes during a period of great societal change. The timing and strength of the selection events in the present study suggest that the genetic effects of selection for cooperation, coordination, and social bonding (19–22) may have played an outsized role compared to other possible drivers of selection like intergroup conflict (10, 34). Furthermore, OT signaling, especially within NAc and dACC, is centrally involved in this process. However, it is important to note that it is not possible to determine levels of *OXTR* and *CD38* expression in humans *in vivo* or interpret a direct causal relationship between this expression and behavior; for example, genes such as *CD157* are associated with other functions besides just OT secretion (102). Thus, while we have demonstrated that complex systems of labor distribution and reciprocity (99) may have worked in concert with these genetic adaptations to contribute to the success of flourishing economies in an increasingly urbanized landscape in the Andes, there are almost certainly many other genetic, social, and environmental factors at play. For example, to whatever extent selection on *OXTR* and *CD38* affected Late Formative and Tiwanaku Period cooperation, it was apparently insufficient to maintain high levels of cooperation—at least across the broad scale of entire societies—into the subsequent Late Intermediate Period, which was the most violent period in the region's history (73).

We advocate for further research in this area exploring other pathways of genetic and cultural adaptation, as well functional validation of genetic polymorphisms and their effects on cooperative behavior. This is an area of great research potential, as there are not many genes for which we understand their intermediate mechanisms of actions in the brain as well as their higher-

level cognitive and behavioral effects; the OT system is one of the most well-studied in this regard. Exploring similar gene-culture coevolutionary dynamics with other genes, while possible, should only be done once we have extensive knowledge regarding how those genes influence cognition and behavior.

### *4.6 Methods*

Principal components analysis: Eigenvectors were computed for samples in the ancient and contemporary Peruvian individuals from this study with outgroup samples from Europe, East Asia, and the Americas (see Appendix C Table C2 for table of outgroup samples used) after removing missing data using Plink v1.9 (--remove-indels, --max-missing 1.0) (82–87, 103). SNPRelate Bioconductor R package v3.15 was used for pruning the whole genome dataset for SNPs in linkage disequilibrium (with the snpgdsLDpruning() option) and the first two principal components were visualized (with the snpgdsPCA() option), using default parameters (79).

Ancestry clustering: We performed model-based clustering analysis using the likelihood-free approach implemented in SCOPE, which utilizes latent subspace estimation and alternating least squares to estimate admixture proportions from individual allele frequencies (81). First, to determine the number of underlying ancestral clusters, we assumed K = 2 to K = 15 ancestral clusters using the Adegenet R package v2.1.7 (104). For each value of K, 10,000 iterations were performed (find.clusters(), max.n.clust = 15, n.iter = 1e4, stat = c("BIC")). We found that K = 6 had the lowest Bayesian information criterion (BIC) value, whereby the smallest BIC value indicated that the model of having K = 6 ancestral clusters best fit the data. We then ran the SCOPE program using K = 6 ancestral clusters (--k 6) and otherwise default parameters.

Admixture graph topology selection: The Aymara, PEL, and Rio Uncallane were merged with populations from Africa, Europe, East Asia, and South America from the SGDP, removing sites

with missing data with Plink (103). A total of 2,997,694 polymorphic sites were used for the analysis. The ADMIXTOOLS2 R package (88) was used to calculate the necessary f-statistics to fit the graph. The function "f2_from_geno" to calculate pairwise f2-statistics. Since all populations had more than one individual, the pseudo-haploid option was set to False. The out population was set to the Mbuti. To find the best fitting topologies, we used the "find_graphs" function. The associated algorithm was set to 500 independent starts with randomly seeded graphs and a maximum of 3 admixture events. The highest-ranking graph was then selected, with a log-likelihood score of 377.244.

Gene expression data for SNP selection: Using the GTEx database (56), we identified 43 SNPs that affect *OXTR* expression in the NAc. In addition, we identified 50 SNPs that influence gene expression of the gene *CD38* (Appendix C Table C3). GTEx standards for significance thresholds were used (56), which requires a false discovery rate threshold of $\leq 0.05$ to identify genes with a significant eQTL. We then searched the ancient Rio Uncallane and contemporary 1000 Genomes PEL datasets to ascertain which SNPs were covered by both the ancient and contemporary whole genome data, and thus could be used for analysis. Of those 24 that were covered by the whole genome datasets, we utilized the LDpair (105) tool to identify SNPs in close genetic linkage with each other, with a high LD threshold ($D' > 0.8$), and chose the 2 centrally-located representative SNPs to proceed with CLUES (89) allele trajectory modeling to test for selection. It is important to note that eQTL values for tissue expression, though highly conserved across humans (106), may not be identical across all populations, and a large proportion of the GTEx tissue samples are from donors of European ancestry.

Modeling allele trajectories that deviate from neutral expectations: The merged VCF containing the 25 PEL individuals and the 5 Rio Uncallane ancient individuals were phased with Beagle 5.4

(107), using the full 1000 Genomes dataset (77) as a reference and with 24 iterations. The resulting phased dataset was filtered for DR2 ≥ 0.8. We then used the phased dataset to calculate genealogies and coalescence rates throughout the genome, using the chimpanzee genome as the ancestral reference and a mutation rate of 1.25E-8, with the program Relate (91). The resulting estimation of the coalescence rate across the genome also allows for the determination of the neutral expectation of the coalescence rate over time.

We then used an approximate likelihood method (CLUES) (89) that tests for selection by detecting the allele trajectory, timing of selection (if any), and selection strength (if any) at a particular SNP along the genealogy tree created by Relate. We ran CLUES on our SNPs of interest, as determined by their potential function, using a generation length of 28 years and sampling each branch 1000 times. We removed the first 100 trees sampled via a burn-in and thinned the sample trees for redundancy by only keeping every $10^{th}$ tree. The resulting allele trajectory model tests for selection by comparing models of neutrality ($s$ = 0) vs models that infer selection, which is reflected in a log-likelihood ratio (logLR). However, because CLUES samples branch lengths in an MCMC fashion, there can be significant variation between results due to uncertainty in the branch length, which can lead to differences in results for the same derived allele. To address this issue, CLUES was run 10 times on the same allele to detect significant variance. We then took the median logLR of the 10 runs, along with the associated selection coefficients.

To establish a significance threshold for the logLR scores, we tested 5,612 polymorphic sites across chromosome 3 and 4 for allele trajectories that deviated from neutrality, with no missing data between the ancient and contemporary populations (Figure 10C). Given some uncertainty with the quality of the ancient DNA data, we only selected high-quality SNPs with a minimum of 10x coverage. As expected, most allele trajectories across the chromosome could be

explained by neutrality. Only extreme logLR scores that exceeded the 99$^{th}$ percentile were considered significant.

We conducted a global selection scan (90) using the entire SGDP dataset (82) composed of 273 individuals from across the world, in order to ascertain how selection at the 2 loci of interest fits into the larger scope of selection across the genome and across worldwide populations (Appendix C Table C4). The Ohana program uses a likelihood ratio test to detect alleles in each population that deviated strongly from a genome-wide covariance structure (90). We utilized K=5 underlying ancestral populations, in accordance with analyses from the SGDP publication (82). VCFtools was used to prepare the VCF by filtering sites out of Hardy-Weinberg Equilibrium with a p-value below $10^{-4}$ and removing indels and missing data (108). This dataset is comprised of high-coverage whole genomes, so we utilized the genotypes directly instead of genotype likelihoods. To estimate the correlation structure of each individual to a population tree, we downsampled the dataset at random to 5% of the original variants and then inferred component covariances using Ohana's *qpas* function to produce admixture-corrected allele frequencies. We then utilized Ohana's *selscan* function to detect the variants that deviated most strongly from the genome-wide covariance structure, which produces likelihood ratios for each locus in the dataset. A high likelihood ratio indicates a strong deviation from the genome-wide covariance structure, and therefore a strong selection signal.

Population stratification from admixture can confound scans for selection. The individuals included in this study have high proportions of Indigenous American ancestry, so we would not expect SNPs under high probability of selection to come from European haplotypes. Nevertheless, we still confirmed this for both SNPs of interest using the RFmix local ancestry inference program (109), which produces global diploid ancestry estimates for each SNP based on haplotypes in a

diverse reference panel (82) (Appendix C Tables C5 and C6). Using a region of the genome 1 million base pairs on either side of the SNP of interest (--analyze-range=), we utilized the default program parameters.

Many of the SNPs present in the dataset are in strong LD with each other. We documented this for all 24 SNPs present in the dataset using the LDpair tool (105) (Appendix C Figure C7).

### *4.7 References*

1. J. Lindo, et al., The genetic prehistory of the Andean highlands 7000 years BP through European contact. Sci Adv 4, eaau4921 (2018).

2. L. Cronk, B. L. Leech, Meeting at Grand Central, Understanding the Social and Evolutionary Roots of Cooperation (Princeton University Press, 2012)

3. R. Boyd, P. J. Richerson, Culture and the evolution of human cooperation. Philosophical Transactions Royal Soc B Biological Sci 364, 3281–3288 (2009).

4. J. M. Burkart, et al., The evolutionary origin of human hyper-cooperation. Nat Commun 5, 4747 (2014).

5. M. S. Alvard, D. A. Nolin, Rousseau's Whale Hunt?: Coordination among Big-Game Hunters. Curr Anthropol 43, 533–559 (2002).

6. M. Tomasello, A. P. Melis, C. Tennie, E. Wyman, E. Herrmann, Two Key Steps in the Evolution of Human Cooperation: The Interdependence Hypothesis. Curr Anthropol 53, 673–692 (2012).

7. M. J. Gelfand, et al., Differences Between Tight and Loose Cultures: A 33-Nation Study. Science 332, 1100–1104 (2011).

8. T. Talhelm, A. S. English, Historically rice-farming societies have tighter social norms in China and worldwide. Proc National Acad Sci 117, 19816–19824 (2020).

9. J. Henrich, et al., Markets, Religion, Community Size, and the Evolution of Fairness and Punishment. Science 327, 1480–1484 (2010).

10. S. Bowles, Did Warfare Among Ancestral Hunter-Gatherers Affect the Evolution of Human Social Behaviors? Science 324, 1293–1298 (2009).

11. J. L. Goodson, R. R. Thompson, Nonapeptide mechanisms of social cognition, behavior and species-specific social systems. Curr Opin Neurobiol 20, 784–794 (2010).

12. N. Rigney, G. J. de Vries, A. Petrulis, L. J. Young, Oxytocin, Vasopressin, and Social Behavior: From Neural Circuits to Clinical Opportunities. Endocrinology 163 (2022).

13. J. P. Burkett, et al., Oxytocin-dependent consolation behavior in rodents. Science 351, 375–378 (2016).

14. M. M. Lim, L. J. Young, Neuropeptidergic regulation of affiliative behavior and social bonding in animals. Horm. Behav. 50, 506–517 (2006).

15. M. Numan, L. J. Young, Neural mechanisms of mother–infant bonding and pair bonding: Similarities, differences, and broader implications. Horm. Behav. 77, 98–112 (2016).

16. H. E. Ross, et al., Characterization of the oxytocin system regulating affiliative behavior in female prairie voles. Neuroscience 162, 892–903 (2009).

17. Z. Triki, K. Daughters, C. K. W. D. Dreu, Oxytocin has 'tend-and-defend' functionality in group conflict across social vertebrates. Philos. Trans. R. Soc. B 377, 20210137 (2022).

18. L. Samuni, et al., Oxytocin reactivity during intergroup conflict in wild chimpanzees. Proc National Acad Sci 114, 268–273 (2017).

19. S. G. Shamay-Tsoory, N. Saporta, I. Z. Marton-Alper, H. Z. Gvirts, Herding Brains: A Core Neural Mechanism for Social Alignment. Trends Cogn Sci 23, 174–186 (2019).

20. F. B. Spengler, et al., Oxytocin facilitates reciprocity in social communication. Soc Cogn Affect Neur 12, 1325–1333 (2017).

21. Y. Mu, C. Guo, S. Han, Oxytocin enhances inter-brain synchrony during social coordination in male adults. Soc Cogn Affect Neur 11, 1882–1893 (2016).

22. M. Stallen, C. K. W. D. Dreu, S. Shalvi, A. Smidts, A. G. Sanfey, The herding hormone: oxytocin stimulates in-group conformity. Psychol Sci 23, 1288–92 (2012).

23. C. K. W. D. Dreu, et al., The Neuropeptide Oxytocin Regulates Parochial Altruism in Intergroup Conflict Among Humans. Science 328, 1408–1411 (2010).

24. S. Israel, O. Weisel, R. P. Ebstein, G. Bornstein, Oxytocin, but not vasopressin, increases both parochial and universal altruism. Psychoneuroendocrino 37, 1341–1344 (2012).

25. F. S. T. Velden, K. Daughters, C. K. W. D. Dreu, Oxytocin promotes intuitive rather than deliberated cooperation with the in-group. Horm Behav 92, 164–171 (2017).

26. M. Arueti, et al., When Two Become One: The Role of Oxytocin in Interpersonal Coordination and Cooperation. J Cognitive Neurosci 25, 1418–1427 (2013).

27. I. D. Neumann, Brain Oxytocin: A Key Regulator of Emotional and Social Behaviours in Both Females and Males. J Neuroendocrinol 20, 858–865 (2008).

28. K. MacDonald, D. Feifel, Oxytocin's role in anxiety: a critical appraisal. Brain Res 1580, 22–56 (2013).

29. M. Mikolajczak, et al., Oxytocin Makes People Trusting, Not Gullible. Psychol Sci 21, 1072–1074 (2010).

30. M. Kosfeld, M. Heinrichs, P. J. Zak, U. Fischbacher, E. Fehr, Oxytocin increases trust in humans. Nature 435, 673–676 (2005).

31. C. H. Declerck, C. Boone, L. Pauwels, B. Vogt, E. Fehr, A registered replication study on oxytocin and trust. Nat Hum Behav 4, 646–655 (2020).

32. G. Nave, C. Camerer, M. McCullough, Does Oxytocin Increase Trust in Humans? A Critical Review of Research. Perspect Psychol Sci 10, 772–789 (2015).

33. T. R. W. D. Wilde, F. S. T. Velden, C. K. W. D. Dreu, The Neuropeptide Oxytocin Enhances Information Sharing and Group Decision Making Quality. Sci Rep-uk 7, 40622 (2017).

34. C. K. W. D. Dreu, M. E. Kret, Oxytocin Conditions Intergroup Relations Through Upregulated In-Group Empathy, Cooperation, Conformity, and Defense. Biol Psychiat 79, 165–173 (2016).

35. S. Son, et al., Whole-Brain Wiring Diagram of Oxytocin System in Adult Mice. J. Neurosci. 42, 5021–5033 (2022).

36. L. Xiao, M. F. Priest, J. Nasenbeny, T. Lu, Y. Kozorovitskiy, Biased Oxytocinergic Modulation of Midbrain Dopamine Systems. Neuron 95, 368-384.e5 (2017).

37. C. N. Rogers, et al., Oxytocin- and arginine vasopressin-containing fibers in the cortex of humans, chimpanzees, and rhesus macaques. Am. J. Primatol. 80, e22875 (2018).

38. E. V. Sivukhina, G. F. Jirikowski, Oxytocin, but not arginine-vasopressin neurons project from the hypothalamus to amygdala in human: DiI-based tracing study in postmortem brain. J. Chem. Neuroanat. 111, 101882 (2021).

39. M. V. Sofroniew, Projections from vasopressin, oxytocin, and neurophysin neurons to neural targets in the rat and human. J. Histochem. Cytochem. : Off. J. Histochem. Soc. 28, 475–478 (1980).

40. B. Jurek, I. D. Neumann, The Oxytocin Receptor: From Intracellular Signaling to Behavior. Physiol. Rev. 98, 1805–1908 (2018).

41. S. Tolomeo, B. Chiao, Z. Lei, S. H. Chew, R. P. Ebstein, A Novel Role of CD38 and Oxytocin as Tandem Molecular Moderators of Human Social Behavior. Neurosci Biobehav Rev 115, 251–272 (2020).

42. H. Higashida, et al., CD38 Gene Knockout Juvenile Mice: A Model of Oxytocin Signal Defects in Autism. Biological Pharm Bulletin 34, 1369–1372 (2011).

43. Y. Jiang, M. L. Platt, Oxytocin and vasopressin flatten dominance hierarchy and enhance behavioral synchrony in part via anterior cingulate cortex. Sci Rep-uk 8, 8201 (2018).

44. L. B. King, H. Walum, K. Inoue, N. W. Eyrich, L. J. Young, Variation in the Oxytocin Receptor Gene Predicts Brain Region–Specific Expression and Social Attachment. Biol Psychiat 80, 160–169 (2016).

45. D. Almeida, et al., Oxytocin receptor expression and epigenetic regulation in the anterior cingulate cortex of individuals with a history of severe childhood abuse. Psychoneuroendocrinology 136, 105600 (2022).

46. T. Watanabe, et al., Oxytocin receptor gene variations predict neural and behavioral response to oxytocin in autism. Soc. Cogn. Affect. Neurosci. 12, 496–506 (2017).

47. C. Feng, et al., A common oxytocin receptor gene (OXTR) polymorphism modulates intranasal oxytocin effects on the neural response to social cooperation in humans. Genes, Brain Behav. 14, 516–525 (2015).

48. H. Higashida, S. Yokoyama, M. Kikuchi, T. Munesue, CD38 and its role in oxytocin secretion and social behavior. Horm Behav 61, 351–358 (2012).

49. T. Munesue, et al., Two genetic variants of CD38 in subjects with autism spectrum disorder and controls. Neurosci Res 67, 181–191 (2010).

50. O. Lopatina, A. Inzhutova, A. B. Salmina, H. Higashida, The Roles of Oxytocin and CD38 in Social or Parental Behaviors. Front Neuroscience 6, 182 (2013).

51. D. Jin, et al., CD38 is critical for social behaviour by regulating oxytocin secretion. Nature 446, 41–45 (2007).

52. O. Lopatina, et al., Reproductive Experience Affects Parental Retrieval Behaviour Associated with Increased Plasma Oxytocin Levels in Wild-Type and Cd38-Knockout Mice. J. Neuroendocr. 23, 1125–1133 (2011).

53. H. Higashida, et al., Social memory, amnesia, and autism: Brain oxytocin secretion is regulated by NAD+ metabolites and single nucleotide polymorphisms of CD38. Neurochem. Int. 61, 828–838 (2012).

54. S. Akther, et al., CD38 in the nucleus accumbens and oxytocin are related to paternal behavior in mice. Mol Brain 6, 41 (2013).

55. D. S. Quintana, et al., Oxytocin pathway gene networks in the human brain. Nat Commun 10, 668

(2019).

56. The GTEx Consortium, The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science 369, 1318–1330 (2020).

57. A. C. Keebaugh, C. E. Barrett, J. L. Laprairie, J. J. Jenkins, L. J. Young, RNAi knockdown of oxytocin receptor in the nucleus accumbens inhibits social attachment and parental care in monogamous female prairie voles. Soc. Neurosci. 10, 561–570 (2015).

58. A. C. Keebaugh, L. J. Young, Increasing oxytocin receptor expression in the nucleus accumbens of pre-pubertal female prairie voles enhances alloparental responsiveness and partner preference formation as adults. Horm. Behav. 60, 498–504 (2011).

59. T. R. Insel, L. E. Shapiro, Oxytocin receptor distribution reflects social organization in monogamous and polygamous voles. Proc. Natl. Acad. Sci. 89, 5981–5985 (1992).

60. S. M. Freeman, et al., Neuroanatomical distribution of oxytocin and vasopressin 1a receptors in the socially monogamous coppery titi monkey (Callicebus cupreus). Neuroscience 273, 12–23 (2014).

61. S. M. Freeman, L. J. Young, Comparative Perspectives on Oxytocin and Vasopressin Receptor Research in Rodents and Primates: Translational Implications. J. Neuroendocr. 28 (2016).

62. E. M. Gayo, C. Latorre, C. M. Santoro, Timing of occupation and regional settlement patterns revealed by time-series analyses of an archaeological radiocarbon database for the South-Central Andes (16°–25°S). Quatern Int 356, 4–14 (2015).

63. M. S. Aldenderfer, Montane Foragers: Asana and the South-Central Andean Archaic (University of Iowa Press, 1998).

64. R. Haas, “Highland Foragers of the Central Andes” in Oxford Handbook of Archaeology, M. S. Aldenderfer, M. Sepulveda, E. Neves, Eds. (2023).

65. N. Craig, Cultural dynamics, climate, and landscape in the South-Central Andes during the mid-late holocene: A consideration of two socio-natural perspectives. Chungar Arica 43, 367–391 (2011).

66. M. S. Aldenderfer, Late Preceramic ceremonial architecture at Asana, southern Peru. Antiquity 64, 479–493 (1990).

67. N. Kitchel, M. S. Aldenderfer, R. Haas, Diet, Mobility, Technology, and Lithics: Neolithization on the Andean Altiplano, 7.0–3.5 ka. J Archaeol Method Th 29, 390–425 (2022).

68. N. Craig, M. S. Aldenderfer, C. A. Rigsby, P. A. Baker, L. F. Blanco, Geologic constraints on rain-fed Qocha reservoir agricultural infrastructure, northern lake Titicaca Basin, Peru. J Archaeol Sci 38, 2897–2907 (2011).

69. E. Klarich, "The Rise of Highland Andean Societies in the Lake Titicaca Basin" in Oxford Handbook of Archaeology, M. S. Aldenderfer, M. Sepulveda, E. Neves, Eds. (2023).

70. C. Stanish, Ancient Titicaca (University of California Press, 2003) https:/doi.org/10.1525/california/9780520232457.001.0001.

71. C. L. Erickson, Prehistoric landscape management in the Andean highlands: Raised field agriculture and its environmental impact. Popul Environ 13, 285–300 (1992).

72. D. Popović, et al., Ancient genomes reveal long-range influence of the pre-Columbian culture and site of Tiwanaku. Sci Adv 7, eabg7261 (2021).

73. E. Arkush, T. A. Tung, Patterns of War in the Andes from the Archaic to the Late Horizon: Insights from Settlement Patterns and Cranial Trauma. J Archaeol Res 21, 307–369 (2013).

74. S. L. Juengst, D. L. Hutchinson, S. J. Chávez, High altitude agriculture in the Titicaca basin (800 BCE-200 CE): Impacts on nutrition and disease load. Am J Hum Biology Official J Hum Biology Counc 29, e22988 (2016).

75. W. R. Haas, C. V. Llave, Hunter-gatherers on the eve of agriculture: investigations at Soro Mik'aya Patjxa, Lake Titicaca Basin, Peru, 8000–6700 BP. Antiquity 89, 1297–1312 (2015).

76. C. Klink, M. S. Aldenderfer, "Ch. 03: A Projectile Point Chronology for the South-Central Andean Highlands" in Advances in Titicaca Basin Archaeology-1, C. Stanish, A. B. Cohen, M. S. Aldenderfer, Eds.

(2005) https:/doi.org/10.2307/j.ctvhhhfn9.

77. The 1000 Genomes Project Consortium, A global reference for human genetic variation. Nature 526, 68–74 (2015).

78. C. Stanish, A. Levine, War and early state formation in the northern Titicaca Basin, Peru. Proc National Acad Sci 108, 13901–13906 (2011).

79. X. Zheng, et al., A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics 28, 3326–3328 (2012).

80. J. K. Pickrell, J. K. Pritchard, Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. Plos Genet 8, e1002967 (2012).

81. A. M. Chiu, E. K. Molloy, Z. Tan, A. Talwalkar, S. Sankararaman, Inferring population structure in biobank-scale genomic data. Am J Hum Genetics 109, 727–737 (2022).

82. S. Mallick, et al., The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. Nature 538, 201–206 (2016).

83. A. Bergström, et al., Insights into human genetic variation and population history from 929 diverse genomes. Science 367 (2020).

84. B. A. Potter, J. D. Irish, J. D. Reuther, H. J. McKinney, New insights into Eastern Beringian mortuary behavior: A terminal Pleistocene double infant burial at Upward Sun River. Proc National Acad Sci 111, 17060–17065 (2014).

85. M. Raghavan, et al., Genomic evidence for the Pleistocene and recent population history of Native Americans. Science 349, aab3884 (2015).

86. J. V. Moreno-Mayar, et al., Early human dispersals within the Americas. Science 362 (2018).

87. M. R. Capodiferro, et al., Archaeogenomic distinctiveness of the Isthmo-Colombian area. Cell 184, 1706-1723.e24 (2021).

88. R. Maier, P. Flegontov, O. Flegontova, P. Changmai, D. Reich, On the limits of fitting complex models of population history to genetic data. Biorxiv, 2022.05.08.491072 (2022).

89. A. J. Stern, P. R. Wilton, R. Nielsen, An approximate full-likelihood method for inferring selection and allele frequency trajectories from DNA sequence data. Plos Genet 15, e1008384 (2019).

90. J. Y. Cheng, A. J. Stern, F. Racimo, R. Nielsen, Detecting Selection in Multiple Populations by Modeling Ancestral Admixture Components. Mol Biol Evol 39, msab294 (2021).

91. L. Speidel, M. Forest, S. Shi, S. R. Myers, A method for genome-wide genealogy estimation for thousands of samples. Nat Genet 51, 1321–1329 (2019).

92. J. W. Janusek, Tiwanaku and Its Precursors: Recent Research and Emerging Perspectives. J Archaeol Res 12, 121–183 (2004).

93. C. A. Hastorf, The Handbook of South American Archaeology. 545–561 (2008).

94. A. Carrasco-Torrontegui, C. A. Gallegos-Riofrío, F. Delgado-Espinoza, M. Swanson, Climate Change, Food Sovereignty and Ancestral Farming Technologies in the Andes. Curr Dev Nutrition 5, 54–60 (2020).

95. J. M. López, G. Neme, A. F. Gil, Resource intensification and zooarchaeological record in the southern margins of pre-Hispanic Andean agriculture. Archaeol Anthrop Sci 11, 5287–5300 (2019).

96. M. C. Bruno, et al., The Rise and Fall of Wiñaymarka: Rethinking Cultural and Environmental Interactions in the Southern Basin of Lake Titicaca. Hum Ecol 49, 131–145 (2021).

97. T. Talhelm, et al., Large-Scale Psychological Differences Within China Explained by Rice Versus Wheat Agriculture. Science 344, 603–608 (2014).

98. T. J. Snyder, R. Haas, Climate change intensified violence in the south-central Andean highlands from 1.5 to 0.5 ka. Quat. Res. 115, 109–119 (2023).

99. S. K. Becker, 4 Community Labor and Laboring Communities within the Tiwanaku State (C.E. 500–1100). Arch P Amer Ant Asso 28, 38–53 (2017).

100. M. Lee, J. Lindo, J. K. Rilling, Exploring gene-culture coevolution in humans by inferring neuroendophenotypes: A case study of the oxytocin receptor gene and cultural tightness. Genes Brain Behav 21, e12783 (2022).

101. J. S. Danoff, et al., Genetic, epigenetic, and environmental factors controlling oxytocin receptor gene expression. Clin Epigenetics 13, 23 (2021).

102. Y. Yakymiv, et al., CD157: From Myeloid Cell Differentiation Marker to Therapeutic Target in Acute Myeloid Leukemia. Cells 8, 1580 (2019).

103. C. C. Chang, et al., Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 4, 7 (2015).

104. T. Jombart, I. Ahmed, adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics 27, 3070–3071 (2011).

105. M. J. Machiela, S. J. Chanock, LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. Bioinformatics 31, 3555–3557 (2015).

106. B. E. Stranger, et al., Patterns of Cis Regulatory Variation in Diverse Human Populations. Plos Genet 8, e1002639 (2012).

107. S. R. Browning, B. L. Browning, Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. Am J Hum Genetics 81, 1084–1097 (2007).

108. P. Danecek, et al., The variant call format and VCFtools. Bioinformatics 27, 2156–2158 (2011).

109. B. K. Maples, S. Gravel, E. E. Kenny, C. D. Bustamante, RFMix: A Discriminative Modeling Approach for Rapid and Robust Local-Ancestry Inference. Am J Hum Genetics 93, 278–288 (2013).

# CHAPTER 5: CONCLUSION

This dissertation investigated the genetic history of the Central Andes region of South America, in what is today Peru and Ecuador, with particular attention to regional adaptations to social, microbial, and physical selective pressures, as well as regional demographic change. Collectively, this work suggests different population dynamics exist across the Central Andes, necessitating nuanced analyses of very fine-scale regional patterns of demographic change and adaptation, as well as collaboration with local populations and researchers to drive increased representation of the peoples living across this diverse region. I offer the following conclusions and considerations for extending the findings from this dissertation to future population genomic research in the Andes:

Pathogens have undoubtedly exerted strong selective pressures on humans throughout history. The Americas are composed of a variety of environments, all of which were inhabited by initial human migrations. Within these environments, many bacterial, viral, and macroparasitic pathogens have persisted through time and co-evolved with humans (see Chapter 2). Until very recently, the misconception that there were few globally-important pathogens present in the Americas before European contact persisted. Counter to this narrative, as part of this dissertation, I scanned the genomes for putative signals of selection from two closely-related Indigenous Ecuadorian populations (see Chapter 3). Interestingly, the strongest signals of selection across the whole genome were genes involved in immune function related to tuberculosis. Given the estimated timing of this selection event, the Indigenous people of Ecuador may have adapted to *M. tuberculosis* thousands of years before the arrival of Europeans. Furthermore, the genetic underpinnings of high-altitude adaptation in the Andes also remain very poorly understood. For example, I identified regions of the genome that show signals of natural selection in both

cardiovascular and hypoxia-related pathways in two populations of Ecuadorians living at high altitude, which are distinct from those genetic variants uncovered in other Andean and worldwide populations (see Chapter 3). There are several possibilities for why a singular genetic adaptation to hypoxic conditions in the Andes have not been characterized. For example, there may have been convergent evolution, with different genes being selected for in improving hypoxia response in different populations. Alternatively, it could be that physiological adaptations to high-altitude, such as higher aerobic capacity and arterial blood oxygen content, are epigenetic.

One hypothesis which has not been widely addressed in previous works on high-altitude adaptation, is that ancient admixture in the Central Andes region was so extensive and complex that it may have dampened the selection signal of major selective pressures acting on populations very early on during the initial peopling of the Central Andean highlands. A similar explanation has been put forth to explain the multi-wave migration model of the peopling of South America—essentially, it was argued that a series of multiple later migrations from Mesoamerica dampened very ancient signals of admixture from Australasia, which is why evidence of the Australasian admixture can be seen in only some, but not all, lineages in South America (see Chapter 1.3). In future work, I am interested in testing a hypothesis using similar logic that could help explain inconsistent results in population genetic studies of high-altitude adaptations in the Central Andes. Perhaps the complex nature of the ancient gene flow and population replacement dynamics in the Central Andes have made it difficult to detect selection for high-altitude adaptations in the genomes of Andeans, especially since this process of adaptation likely took place during the initial peopling of the highlands over 13 ka. To test this, I would model ancient gene flow to determine whether events like gene flow happened in the lineages ancestral to a set of study populations across the Andes. Such research will shed light into some of the very early population dynamics

of the region, which may have led to the undetectable nature of high-altitude adaptations in the Central Andes.

It is clear that our understanding of the population histories of the Andes remains tremendously incomplete. For example, in the study of Ecuadorians living at high altitude from Chapter 3, I detected a population size collapse of almost 90% during the arrival of Europeans, which was much more severe than other regions of the Andes, suggesting very differing effects of European contact across high-altitude populations. Another such example can be seen in Chapter 4, where urbanization and socioeconomic integration likely catalyzed genetic adaptations to promote cooperation in the urban center of Pukara in the Lake Titicaca Basin during the Upper Formative Period. There are undoubtedly countless other examples to be found across this diverse region, which will be illuminated through further genomic research and collaboration with local Andean communities. Perhaps most importantly, this dissertation has shown that even at small regional scales within the region called the “Central Andes”, each population should be understood on its own terms.

# APPENDIX A: CHAPTER 2 SUPPLEMENTARY INFORMATION

**Table A1.** List of references used to construct Figure 1.

| | | |
|---|---|---|
| Endemic pre-contact | Malaria (*Plasmodium vivax*) | Oliveira et al., 2017 |
| | | Rodrigues et al., 2018 |
| | Tuberculosis (*Mycobacterium tuberculosis*) | Bos et al., 2014 |
| | | Rothschild et al., 2001 |
| | Syphilis (*Treponema pallidum*) | Giffin et al., 2020 |
| | | Harper et al., 2008 |
| | | Majander et al., 2020 |
| | Chagas Disease (*Trypanosoma cruzi*) | Aufderheide et al., 2004 |
| | | Ferreira et al., 2000 |
| Post-European Contact Disease Routes | *Plasmodium vivax* | Gelabert et al., 2016 |
| | | van Dorp et al., 2020 |
| | *Plasmodium falciparum* | Molina-Cruz et al., 2013 |
| | | Yalcindag et al., 2012 |
| | Yellow Fever Virus | Bryant et al., 2007 |
| | | Li et al., 2017 |
| | *Mycobacterium tuberculosis* (L4 lineage) | Brynildsrud et al., 2018 |

**Table A2.** List of references used to construct Figure 3.

| | |
|---|---|
| *Plasmodium vivax* (Transatlantic Slave Trade) | Gelabert et al., 2016 |
| | van Dorp et al., 2020 |
| | Oliveira et al., 2017 |
| | Rodrigues et al., 2018 |
| *Plasmodium vivax* (Early migrations) | Oliveira et al., 2017 |
| | Rodrigues et al., 2018 |
| *Plasmodium falciparum* | Molina-Cruz et al., 2013 |
| | Yalcindag et al., 2012 |

## Chapter 2 Supplementary References

Aufderheide, A. C., Salo, W., Madden, M., Streitz, J., Buikstra, J., Guhl, F., … Allison, M. (2004). A 9,000-year record of Chagas' disease. Proceedings of the National Academy of Sciences, 101(7), 2034–2039.

Bos, K. I., Harkins, K. M., Herbig, A., Coscolla, M., Weber, N., Comas, I., … Krause, J. (2014). Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. Nature, 514(7523), 494–497.

Bryant, J. E., Holmes, E. C., & Barrett, A. D. T. (2007). Out of Africa: A Molecular Perspective on the Introduction of Yellow Fever Virus into the Americas. PLoS Pathogens, 3(5), e75.

Brynildsrud, O. B., Pepperell, C. S., Suffys, P., Grandjean, L., Monteserin, J., Debech, N., … Eldholm, V. (2018). Global expansion of Mycobacterium tuberculosis lineage 4 shaped by colonial migration and local adaptation. Science Advances, 4(10), eaat5869.

Dorp, L. van, Gelabert, P., Rieux, A., Manuel, M. de, de-Dios, T., Gopalakrishnan, S., … Lalueza-Fox, C. (2020). Plasmodium vivax Malaria Viewed through the Lens of an Eradicated European Strain. Molecular Biology and Evolution, 37(3), 773–785.

Ferreira, L. F., Britto, C., Cardoso, M. A., Fernandes, O., Reinhard, K., & Araújo, A. (2000). Paleoparasitology of Chagas disease revaled by infected tissues from Chilean mummies. Acta Tropica, 75(1), 79–84.

Gelabert, P., Sandoval-Velasco, M., Olalde, I., Fregel, R., Rieux, A., Escosa, R., … Lalueza-Fox, C. (2016). Mitochondrial DNA from the eradicated European Plasmodium vivax and P. falciparum from 70-year-old slides from the Ebro Delta in Spain. Proceedings of the National Academy of Sciences, 113(41), 11495–11500.

Giffin, K., Lankapalli, A. K., Sabin, S., Spyrou, M. A., Posth, C., Kozakaitė, J., … Bos, K. I. (2020). A treponemal genome from an historic plague victim supports a recent emergence of yaws and its presence in 15th century Europe. Scientific Reports, 10(1), 9499.

Harper, K. N., Ocampo, P. S., Steiner, B. M., George, R. W., Silverman, M. S., Bolotin, S., … Armelagos, G. J. (2008). On the Origin of the Treponematoses: A Phylogenetic Approach. PLoS Neglected Tropical Diseases, 2(1), e148.

Li, Y., & Yang, Z. (2017). Adaptive Diversification between Yellow Fever Virus West African and South American Lineages: A Genome-Wide Study. The American Journal of Tropical Medicine and Hygiene, 16–0698.

Majander, K., Pfrengle, S., Kocher, A., Neukamm, J., Plessis, L. du, Pla-Díaz, M., … Schuenemann, V. J. (2020). Ancient Bacterial Genomes Reveal a High Diversity of Treponema pallidum Strains in Early Modern Europe. Current Biology, 30(19), 3788–3803.

Molina-Cruz, A., Garver, L. S., Alabaster, A., Bangiolo, L., Haile, A., Winikor, J., … Barillas-Mury, C. (2013). The Human Malaria Parasite Pfs47 Gene Mediates Evasion of the Mosquito Immune System. Science, 340(6135), 984–987.

Oliveira, T. C. de, Rodrigues, P. T., Menezes, M. J., Gonçalves-Lopes, R. M., Bastos, M. S., Lima, N. F., … Ferreira, M. U. (2017). Genome-wide diversity and differentiation in New World populations of the human malaria parasite Plasmodium vivax. PLoS Neglected Tropical Diseases, 11(7), e0005824.

Rodrigues, P. T., Valdivia, H. O., Oliveira, T. C. D., Alves, J. M. P., Duarte, A. M. R. C., Cerutti-Junior, C., … Ferreira, M. U. (2018). Human migration and the spread of malaria parasites to the New World. Scientific Reports, 8(1), 1–13.

Rothschild, B. M., Martin, L. D., Lev, G., Bercovier, H., Bar-Gal, G. K., Greenblatt, C., … Brittain, D. (2001). Mycobacterium tuberculosis complex DNA from an extinct bison dated 17,000 years before the present. Clinical Infectious Diseases, 33(3), 305–311.

Yalcindag, E., Elguero, E., Arnathau, C., Durand, P., Akiana, J., Anderson, T. J., … Prugnolle, F. (2012). Multiple independent introductions of Plasmodium falciparum in South America. Proceedings of the National Academy of Sciences of the United States of America, 109(2), 511–516.

## APPENDIX B: CHAPTER 3 SUPPLEMENTARY INFORMATION

**Table B1.** Information about sequencing, populations and their locations, mtDNA haplogroup, and mean read depth.

| Sample | Analysis Group | Province of Ecuador | Elevation (meters) | mtDNA Haplogroup | Mean Read Depth |
|---|---|---|---|---|---|
| 01 | Loja (Oñacapac) | Loja | 2,533 m | B2l | 24.2 |
| 04 | Loja (Gañil) | Loja | 2,704 m | B2b | 24.6 |
| 05 | Loja (Oñacapac) | Loja | 2,533 m | B2b | 27.6 |
| 06 | Loja (Oñacapac) | Loja | 2,533 m | C1b | 25.4 |
| 07 | Loja (Oñacapac) | Loja | 2,533 m | B2l | 25.1 |
| 08 | Loja (Oñacapac) | Loja | 2,533 m | B2b | 24.8 |
| 09 | Loja (Gañil) | Loja | 2,704 m | C1b | 27.9 |
| 10 | Loja (Oñacapac) | Loja | 2,533 m | C1d | 26.1 |
| E2 | Kichwa (Karanki) | Imbabura | 2,546 m | A2ac | 28.0 |
| E6 | Kichwa (Kisapincha) | Tungurahua | 2,600 m | B2b | 19.6 |
| E8 | Kichwa (Panzaleo) | Cotopaxi | 3,849 m | B2b | 36.1 |
| E18 | Kichwa (Panzaleo) | Cotopaxi | 3,849 m | C1c | 21.6 |
| E25 | Kichwa (Otavalo) | Imbabura | 2,532 m | B2b | 29.6 |
| E28 | Kichwa (Otavalo) | Imbabura | 2,532 m | C1b | 22.1 |
| E34 | Kichwa (Panzaleo) | Cotopaxi | 3,849 m | B2b | 15.3 |

**Figure B2.** Detailed letter describing workshops for returning results to Indigenous communities who contributed to the present study.

**Wladimir Galarza Ordoñez (Author)**
Cuenca
24 de junio, 2022

**Translated by: J.J. Ortiz Aguilú**
Universidad Técnica de Manabí

This will be based on the organization of three workshops to be held in three Indigenous Kichwa speaking communities located in the Cañar cantón, among which we can list: Sid-Sid, Amanta Bayopungo, and Quilloac.

**1. Procedure:** Each workshop will last two hours. The first will have a descriptive and explanatory character indicating first the research process carried out and the results obtained in a succinct and pedagogical manner.

In the second hour, a series of work tables will be organized so that the theme is worked on in groups, for this purpose the theme should be understood as a fundamental part of the development of biogenetic research techniques in groups of resurgent living Kichwas that logically have a pre-Hispanic cultural background of both presence and heritage in the occupation of these Andean cultural spaces.

The conclusions to be compiled by the head of each group table must be processed and systematized to develop a kind of Community Active Memory, which must contain the concerns, the thoughts, and the criteria issued within the context of this socialization. This Active Memory may eventually be incorporated as an annex or a section of the investigation and form part, if applicable, of the investigative corpus.

**2. Language:** Socializations must be dictated in the two languages Spanish (*Castilla Shimita*) and in Kichwa.

**3. Human resources:** The facilitator designated to carry out the socializations must coordinate with the leaders of each of the communities (*pushak taita*) to coordinate the date and time for the work sessions. The premises must be provided with electricity, bathrooms, work tables, etc.

**4. Technological resources:** In the previously designated location, the presentations made in Power Point must be socialized with the help of an Infocus Projector. At the end of each socialization, the participants will fill out registration sheets.

**5. Didactic material:** A folder with sheets of bond paper and a pen should be made available to each attendant for note taking and questions.

**Table B3.** SNPs of interest under high probability of selection described in the Results section.

| Chr* | Pos* | Annotated Gene | Transcription Factor Binding Site for | rsID | Allele frequency in Kichwa group | Allele frequency in Loja group |
|---|---|---|---|---|---|---|
| 2 | 178706193 | *PDE11A* | CEBPB | rs73038286 | 0 | 0 |
| 8 | 115551670 | *TRPS1* | CEBPB | rs7824435 | 0.142857 | 0.25 |
| 1 | 111401146 | *KCNA3* | | rs6661420 | 0.142857 | 0.25 |
| 11 | 35261508 | *CD44* | | rs11033038 | 0.142857 | 0.0625 |
| 1 | 157549337 | *FCRL4* | | rs2873347 | 0.857143 | 0.8125 |
| 19 | 6920633 | *ADGRE1* | | rs62123131 | 0.142857 | 0.1875 |
| 2 | 41957289 | *SLC8A1* | CEBPB | rs7594765 | 0.0714286 | 0.125 |
| 10 | 45162329 | *CXCL12* | | rs10900091 | 0 | 0.125 |
| 9 | 76898246 | *ANXA1* | | rs5002587 | 1 | 1 |
| 12 | 110870000 | *ANAPC7* | GATA-2 | rs373781413 | 0.142857 | 0.0625 |
| 14 | 34575027 | *EGLN3* | | rs11624971 | 0.857143 | 0.5 |
| 2 | 154857288 | *GALNT13* | | rs59020136 | 0.071429 | 0.0625 |
| 2 | 242147703 | *ANO7* | GATA-2 | rs12471760 | 0.9285714 | 0.9375 |
| 22 | 45981492 | *FBLN1* | | rs3788664 | 0.857143 | 0.75 |
| 5 | 153954518 | *HAND1* | | rs10875587 | 0.428571 | 0.125 |
| 1 | 189034234 | *BRINP3* | | rs4388707 | 0.571429 | 0.5 |

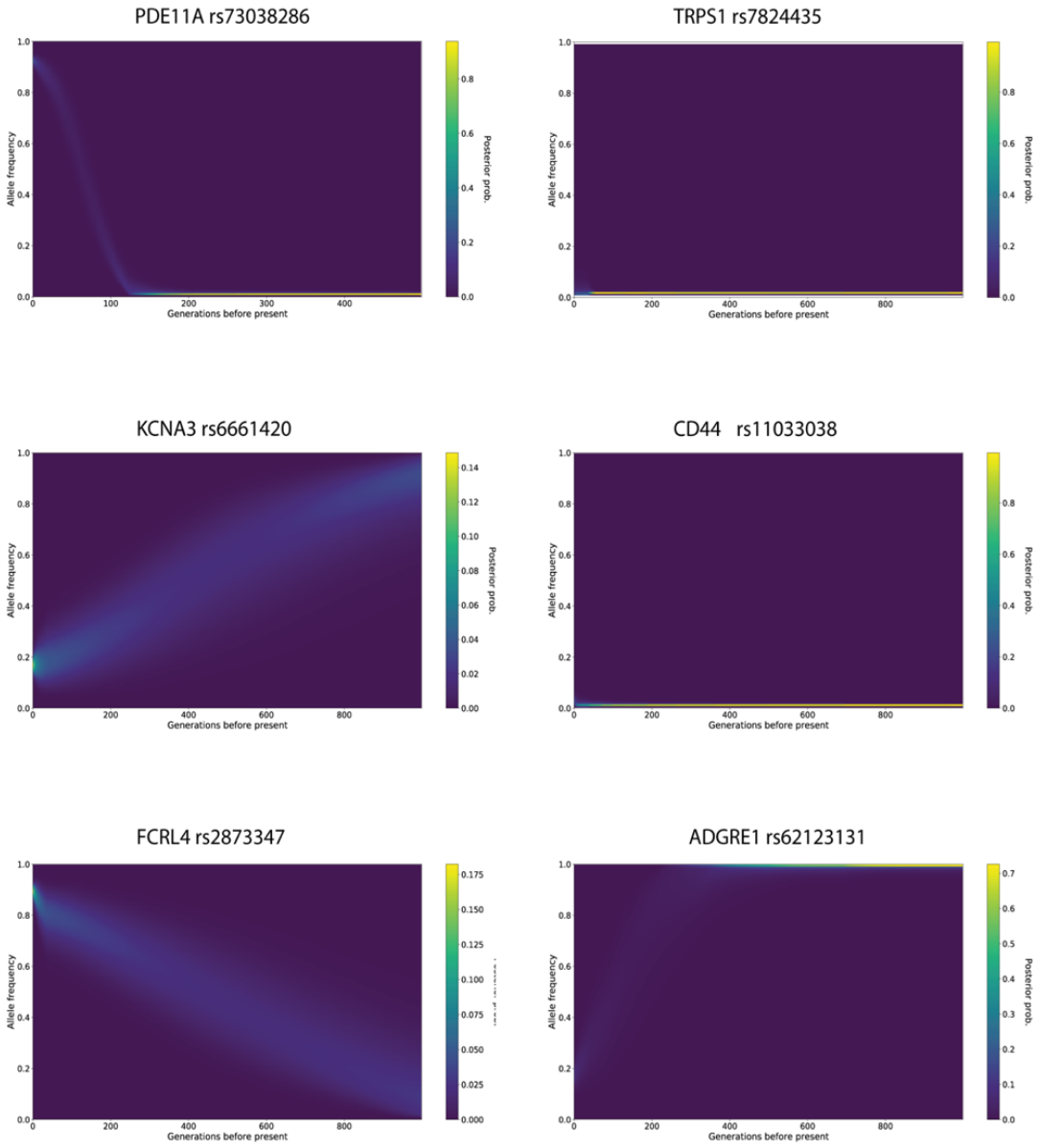
*Hg19 coordinates

**Table B4.** Immune-related SNPs of interest under high probability of selection described in the Results section with CLUES selection strengths and trajectory timing[S1].

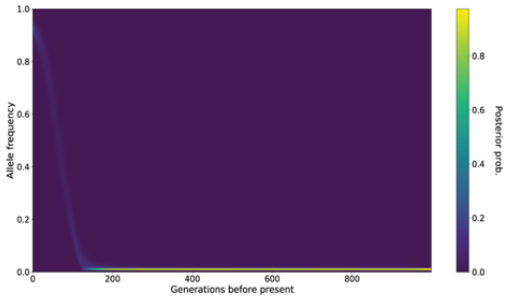
| Chr* | Pos* | Annotated Gene | Transcription Factor Binding Site for | rsID | CLUES selection strength |
|---|---|---|---|---|---|
| 2 | 178706193 | *PDE11A* | CEBPB | rs73038286 | 0.04385 |
| 8 | 115551670 | *TRPS1* | CEBPB | rs7824435 | 0.1 |
| 1 | 111401146 | *KCNA3* | | rs6661420 | –0.00342 |
| 11 | 35261508 | *CD44* | | rs11033038 | 0.1 |
| 1 | 157549337 | *FCRL4* | | rs2873347 | 0.00254 |
| 19 | 6920633 | *ADGRE1* | | rs62123131 | –0.01201 |
| 2 | 41957289 | *SLC8A1* | CEBPB | rs7594765 | 0.04715 |
| 10 | 45162329 | *CXCL12* | | rs10900091 | 0.00361 |
| 9 | 76898246 | *ANXA1* | | rs5002587 | 0.01 |

*Hg19 coordinates

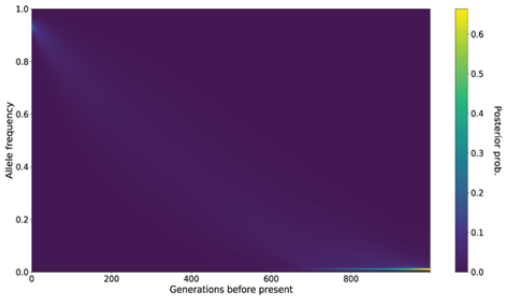
**Figure B5.** CLUES allele trajectories for immune-related SNPs of interest under high probability of selection described in the Results section[S1].



SLC8A1 rs7594765




CXCL12 rs10900091




ANXA1 rs5002587

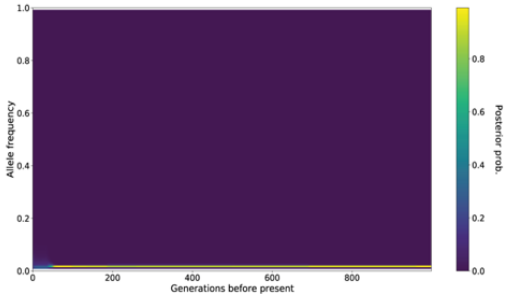
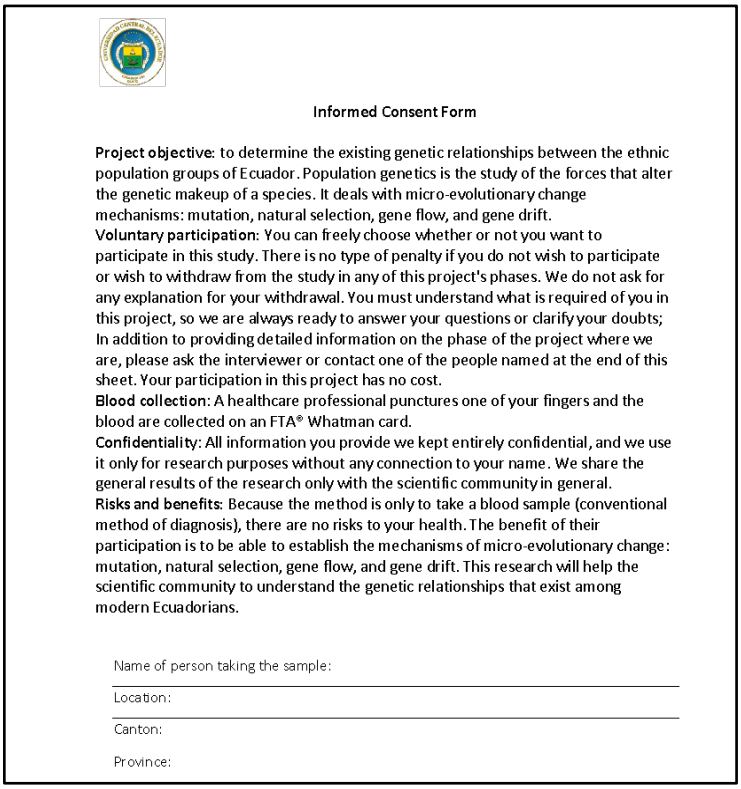


**Table B6.** Table comparing high-altitude-related SNPs of interest under high probability of selection described in the Results section from Present Study in Ecuador versus selected past studies from other Andean populations[S2–S5].

| Gene | rsID | Reference |
|---|---|---|
| *ANAPC7* | rs373781413 | present study |
| *EGLN3* | rs11624971 | present study |
| *GALNT13* | rs59020136 | present study |
| *ANO7* | rs12471760 | present study |
| *FBLN1* | rs3788664 | present study |
| *HAND1* | rs10875587 | present study |
| *BRINP3* | rs4388707 | present study |
| *BRINP3* | rs11578671 | Crawford et al., 2017 |
| *NOS2* | rs34913965 | Crawford et al., 2017 |
| multiple | rs12448902 | Crawford et al., 2017 |
| *TBX5* | rs10744822 | Crawford et al., 2017 |
| multiple | rs487105 | Crawford et al., 2017 |
| *CTAGE1* | rs11081933 | Crawford et al., 2017 |
| *ULBP1* | rs4869782 | Crawford et al., 2017 |
| *SHISA6* | rs78264921 | Crawford et al., 2017 |
| *TMEM38B* | rs12685887 | Crawford et al., 2017 |
| *PPA2* | rs2214403 | Crawford et al., 2017 |
| *SFTPD* | rs3088308 | Valverde et al., 2015 |
| *TMEM254* | rs1932574 | Valverde et al., 2015 |
| *ANXA11* | rs1049550 | Valverde et al., 2015 |
| *SH2D4B* | rs7075840 | Valverde et al., 2015 |
| *EGLN1* | rs206476, rs243715, rs249140, rs47920, rs176979 | Brutsaert et al., 2019 |
| *GALNT13* | rs6744657 | Caro-Consuegra et al., 2022 |
| *EGLN3* | rs1998852 | Caro-Consuegra et al., 2022 |
| *BRINP3* | rs284077 | Caro-Consuegra et al., 2022 |

**Figure B7.** English version of Informed Consent.



Informed Consent Form

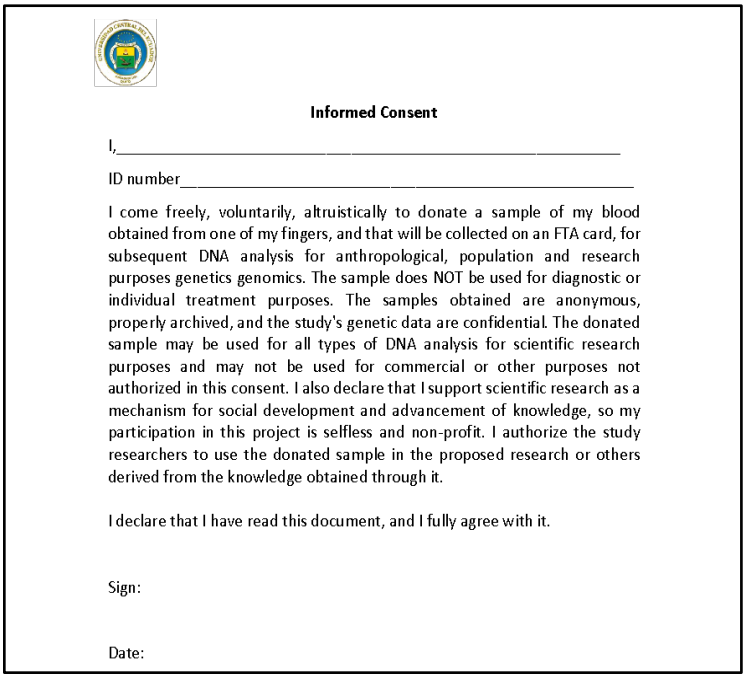
**Project objective**: to determine the existing genetic relationships between the ethnic population groups of Ecuador. Population genetics is the study of the forces that alter the genetic makeup of a species. It deals with micro-evolutionary change mechanisms: mutation, natural selection, gene flow, and gene drift.
**Voluntary participation**: You can freely choose whether or not you want to participate in this study. There is no type of penalty if you do not wish to participate or wish to withdraw from the study in any of this project's phases. We do not ask for any explanation for your withdrawal. You must understand what is required of you in this project, so we are always ready to answer your questions or clarify your doubts; In addition to providing detailed information on the phase of the project where we are, please ask the interviewer or contact one of the people named at the end of this sheet. Your participation in this project has no cost.
**Blood collection**: A healthcare professional punctures one of your fingers and the blood are collected on an FTA® Whatman card.
**Confidentiality**: All information you provide we kept entirely confidential, and we use it only for research purposes without any connection to your name. We share the general results of the research only with the scientific community in general.
**Risks and benefits**: Because the method is only to take a blood sample (conventional method of diagnosis), there are no risks to your health. The benefit of their participation is to be able to establish the mechanisms of micro-evolutionary change: mutation, natural selection, gene flow, and gene drift. This research will help the scientific community to understand the genetic relationships that exist among modern Ecuadorians.

Name of person taking the sample:

Location:

Canton:

Province:



**Informed Consent**

I,_____

ID number_____

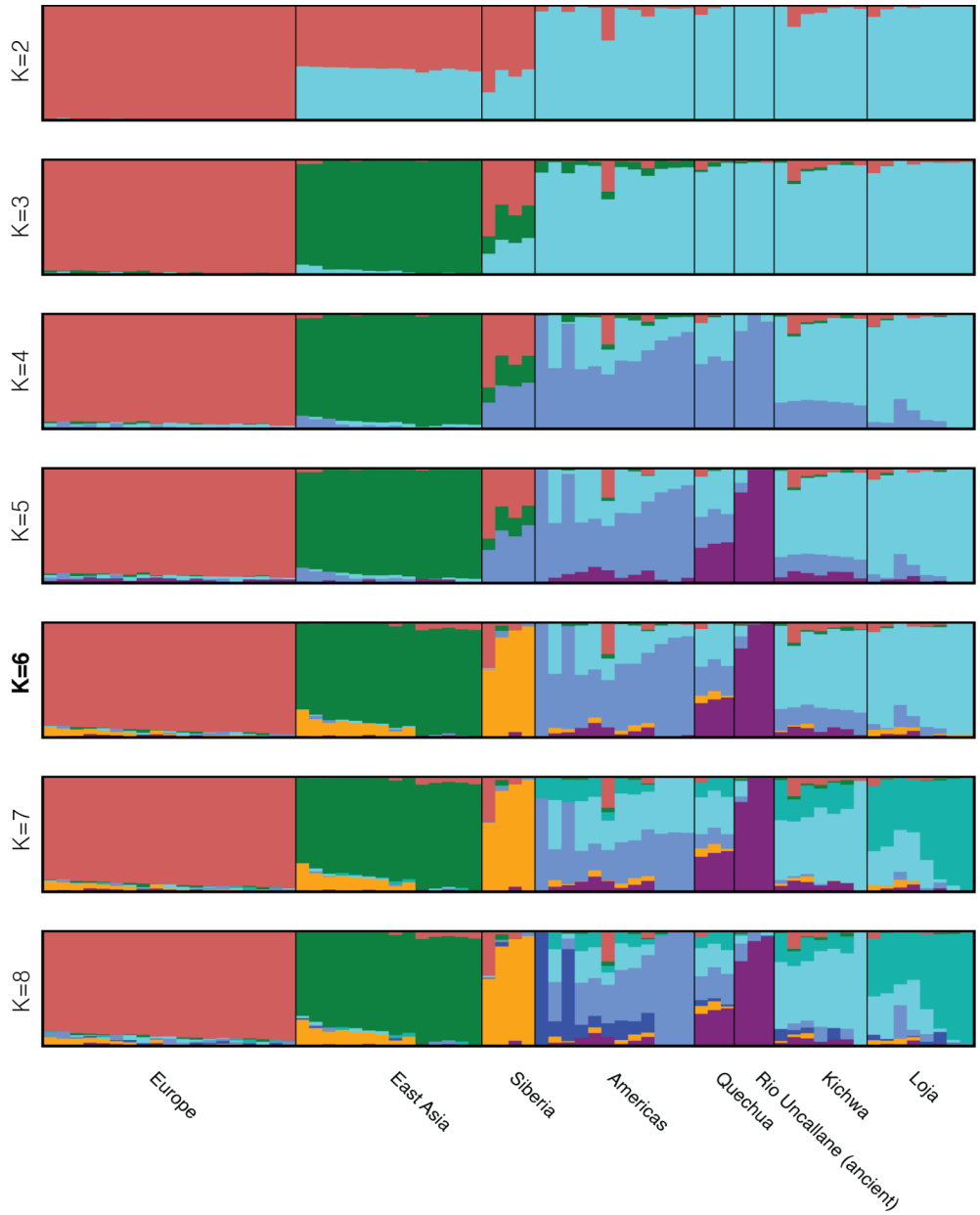
I come freely, voluntarily, altruistically to donate a sample of my blood obtained from one of my fingers, and that will be collected on an FTA card, for subsequent DNA analysis for anthropological, population and research purposes genetics genomics. The sample does NOT be used for diagnostic or individual treatment purposes. The samples obtained are anonymous, properly archived, and the study's genetic data are confidential. The donated sample may be used for all types of DNA analysis for scientific research purposes and may not be used for commercial or other purposes not authorized in this consent. I also declare that I support scientific research as a mechanism for social development and advancement of knowledge, so my participation in this project is selfless and non-profit. I authorize the study researchers to use the donated sample in the proposed research or others derived from the knowledge obtained through it.

I declare that I have read this document, and I fully agree with it.

Sign:

Date:

**Figure B8.** Scope cluster analysis[S6] visualizations with K=2–8 and highest likelihood Treemix[S7] scenarios for m = 1 migration events. Cluster analysis was completed after removing transitions in the ancient Rio Uncallane samples that could be due to systematic damage patterns in ancient DNA, in order to rule out these ancient individuals cluster on their own due to patterns of DNA damage (see Figure B9 for more information).



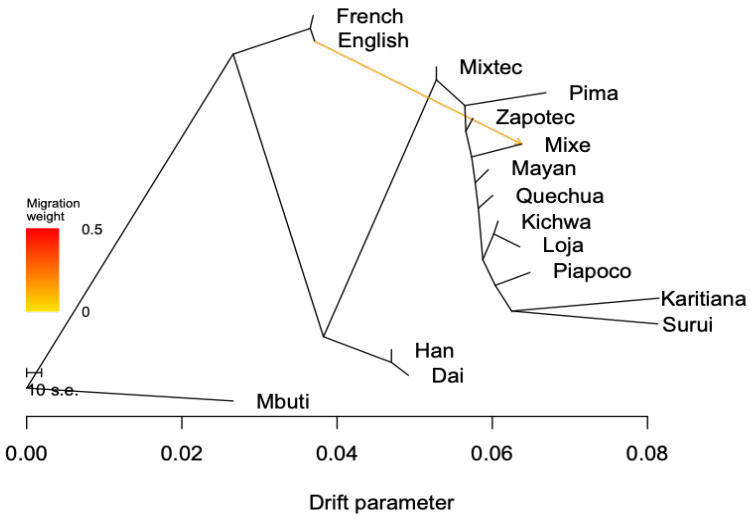

French
English
Mixtec
Pima
Zapotec
Mixe
Mayan
Quechua
Kichwa
Loja
Piapoco
Karitiana
Surui
Han
Dai
Mbuti
Migration weight
0.5
0
10 s.e.
0.00
0.02
0.04
0.06
0.08
Drift parameter

**Figure B9.** Demographic analyses from relatedness2 program from the VCFtools software and SCOPE after computationally removing transitions that could be due to systematic damage patterns in ancient DNA for the 3 ancient Rio Uncallane Andean genomes, using default parameters[S6,S8,S9]. These data indicate these ancient individuals are not closely related and still cluster on their own, as in the cluster analysis in Figure 4E of the main text.

Pairwise relationships from relatedness2 output:

| Individual | Compared to | Kinship Coefficient phi from pairwise comparison* |
|---|---|---|
| IL2 | IL3 | -0.0539655 |
| IL2 | IL7 | -0.0512322 |
| IL3 | IL7 | -0.0672019 |

* maximum of the kinship coefficient phi is 0.5, indicating monozygotic twins

K=6 SCOPE output after computationally removing transitions that could be due to systematic damage patterns in ancient DNA (same as K=6 from Figure B8):

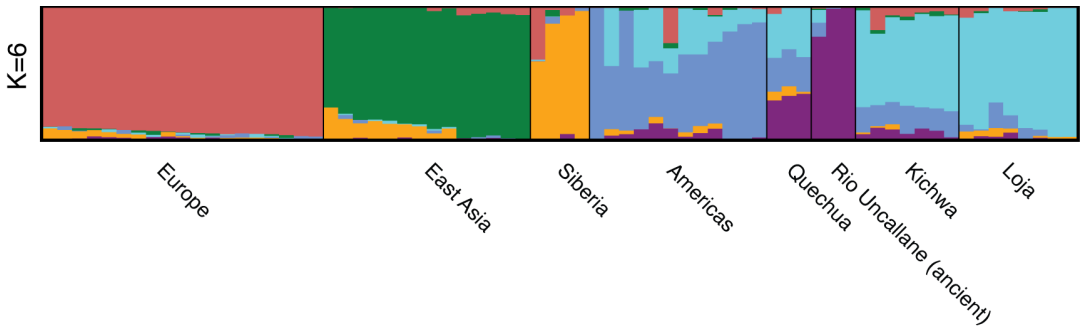


**Table B10.** Outgroup samples utilized from the SGDP dataset[S10].

| Individual | Population |
|---|---|
| S_Chane-1 | Americas |
| S_Surui-1 | Americas |
| S_Surui-2 | Americas |
| S_Karitiana-1 | Americas |
| S_Karitiana-2 | Americas |
| S_Piapoco-1 | Americas |
| S_Piapoco-2 | Americas |
| S_Pima-1 | Americas |
| S_Pima-2 | Americas |
| S_Mayan-1 | Americas |
| S_Mayan-2 | Americas |
| S_Mixe-2 | Americas |
| S_Mixe-3 | Americas |
| S_Mixtec-1 | Americas |
| S_Mixtec-2 | Americas |
| S_Quechua-3 | Americas |
| S_Quechua-2 | Americas |
| S_Quechua-1 | Americas |
| B_Karitiana-3 | Americas |
| B_Mixe-1 | Americas |
| S_Han-2 | East_Asia |
| S_Dai-3 | East_Asia |
| S_Han-1 | East_Asia |
| S_Hezhen-1 | East_Asia |
| S_Hezhen-2 | East_Asia |
| S_Dai-2 | East_Asia |
| S_Dai-1 | East_Asia |
| S_Japanese-1 | East_Asia |
| S_Japanese-2 | East_Asia |
| S_Japanese-3 | East_Asia |
| S_Korean-2 | East_Asia |
| S_Korean-1 | East_Asia |
| B_Dai-4 | East_Asia |
| B_Han-3 | East_Asia |

| | |
|---|---|
| S_Bulgarian-1 | Europe |
| S_Bulgarian-2 | Europe |
| S_English-1 | Europe |
| S_English-2 | Europe |
| S_French-1 | Europe |
| S_French-2 | Europe |
| S_Basque-1 | Europe |
| S_Hungarian-2 | Europe |
| S_Hungarian-1 | Europe |
| S_Bergamo-1 | Europe |
| S_Bergamo-2 | Europe |
| S_Tuscan-2 | Europe |
| S_Tuscan-1 | Europe |
| S_Norwegian-1 | Europe |
| S_Polish-1 | Europe |
| S_Spanish-1 | Europe |
| S_Spanish-2 | Europe |
| B_Crete-1 | Europe |
| B_French-3 | Europe |

## Chapter 3 Supplementary References

S1. J. Stern, P. R. Wilton, R. Nielsen (2019). An approximate full-likelihood method for inferring selection and allele frequency trajectories from DNA sequence data. *PLoS Genet* 15, e1008384.

S2. J. E. Crawford, et al. (2017). Natural Selection on Genes Related to Cardiovascular Health in High-Altitude Adapted Andeans. *Am J Hum Genet* 101, 752–767.

S3. G. Valverde, et al. (2015). A Novel Candidate Region for Genetic Adaptation to High Altitude in Andean Populations. *PLos One* 10, e0125444.

S4. T. D. Brutsaert, et al. (2019). Association of EGLN1 gene with high aerobic capacity of Peruvian Quechua at high altitude. *Proc National Acad Sci* 116, 24006–24011.

S5. R. Caro-Consuegra, et al. (2022). Uncovering Signals of Positive Selection in Peruvian Populations from Three Ecological Regions. *Mol Biol Evol* 39, msac158.

S6. A.M. Chiu, E.K. Molloy, Z. Tan, A. Talwalkar, S. Sankararaman (2022). Inferring population structure in biobank-scale genomic data. *Am J Hum Genet* 109, 727–737.

S7. Pickrell, J. K. & Pritchard, J. K. (2012). Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. *Plos Genet* 8, e1002967.

S8. P. Danecek, et al. (2011). The variant call format and VCFtools. *Bioinformatics* 27, 2156–2158.

S9. Manichaikul, et al. (2010). Robust relationship inference in genome-wide association studies. *Bioinformatics* 26, 2867–2873.

S10. S. Mallick, et al. (2016). The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature* 538, 201–206.

# APPENDIX C: CHAPTER 4 SUPPLEMENTARY INFORMATION

**Table C1.** Individual information, including radiocarbon age, individual ID, and population.

| Individual ID | Calibrated Radiocarbon Age | Population |
|---|---|---|
| IL2 | not dated | Rio Uncallane |
| IL3 | 1613-1811 calBP | Rio Uncallane |
| IL4 | 1428-1533 calBP | Rio Uncallane |
| IL5 | 1614-1725 calBP | Rio Uncallane |
| IL7 | 1742-1830 calBP | Rio Uncallane |
| HG01572 | contemporary | 1000 Genomes PEL |
| HG01923 | contemporary | 1000 Genomes PEL |
| HG01926 | contemporary | 1000 Genomes PEL |
| HG01927 | contemporary | 1000 Genomes PEL |
| HG01941 | contemporary | 1000 Genomes PEL |
| HG01951 | contemporary | 1000 Genomes PEL |
| HG01953 | contemporary | 1000 Genomes PEL |
| HG01954 | contemporary | 1000 Genomes PEL |
| HG02008 | contemporary | 1000 Genomes PEL |
| HG02102 | contemporary | 1000 Genomes PEL |
| HG02105 | contemporary | 1000 Genomes PEL |
| HG02146 | contemporary | 1000 Genomes PEL |
| HG02147 | contemporary | 1000 Genomes PEL |
| HG02150 | contemporary | 1000 Genomes PEL |
| HG02259 | contemporary | 1000 Genomes PEL |
| HG02260 | contemporary | 1000 Genomes PEL |
| HG02266 | contemporary | 1000 Genomes PEL |
| HG02271 | contemporary | 1000 Genomes PEL |
| HG02272 | contemporary | 1000 Genomes PEL |
| HG02275 | contemporary | 1000 Genomes PEL |
| HG02278 | contemporary | 1000 Genomes PEL |
| HG02291 | contemporary | 1000 Genomes PEL |
| HG02292 | contemporary | 1000 Genomes PEL |
| HG02304 | contemporary | 1000 Genomes PEL |
| HG02348 | contemporary | 1000 Genomes PEL |

**Table C2.** Outgroup individuals for demographic analyses.

| Individual ID | Dataset |
|---|---|
| S_Ju_hoan_North-1 | Simons Genome Diversity Project (SGDP) |
| S_Khomani_San-1 | Simons Genome Diversity Project (SGDP) |
| S_Khomani_San-2 | Simons Genome Diversity Project (SGDP) |
| S_Mbuti-1 | Simons Genome Diversity Project (SGDP) |
| S_Mbuti-3 | Simons Genome Diversity Project (SGDP) |
| S_Dai-1 | Simons Genome Diversity Project (SGDP) |
| S_Dai-2 | Simons Genome Diversity Project (SGDP) |
| S_Dai-3 | Simons Genome Diversity Project (SGDP) |
| S_Han-1 | Simons Genome Diversity Project (SGDP) |
| S_Han-2 | Simons Genome Diversity Project (SGDP) |
| S_English-1 | Simons Genome Diversity Project (SGDP) |
| S_English-2 | Simons Genome Diversity Project (SGDP) |
| S_French-1 | Simons Genome Diversity Project (SGDP) |
| S_French-2 | Simons Genome Diversity Project (SGDP) |
| S_Karitiana-1 | Simons Genome Diversity Project (SGDP) |
| S_Karitiana-2 | Simons Genome Diversity Project (SGDP) |
| S_Surui-1 | Simons Genome Diversity Project (SGDP) |
| S_Surui-2 | Simons Genome Diversity Project (SGDP) |
| S_Piapoco-1 | Simons Genome Diversity Project (SGDP) |
| S_Piapoco-2 | Simons Genome Diversity Project (SGDP) |
| S_Mayan-1 | Simons Genome Diversity Project (SGDP) |
| S_Mayan-2 | Simons Genome Diversity Project (SGDP) |
| S_Mixe-2 | Simons Genome Diversity Project (SGDP) |
| S_Mixe-3 | Simons Genome Diversity Project (SGDP) |
| S_Mixtec-1 | Simons Genome Diversity Project (SGDP) |
| S_Mixtec-2 | Simons Genome Diversity Project (SGDP) |
| S_Pima-1 | Simons Genome Diversity Project (SGDP) |
| S_Pima-2 | Simons Genome Diversity Project (SGDP) |
| S_Zapotec-1 | Simons Genome Diversity Project (SGDP) |
| S_Zapotec-2 | Simons Genome Diversity Project (SGDP) |
| S_Quechua-1 | Simons Genome Diversity Project (SGDP) |
| S_Quechua-2 | Simons Genome Diversity Project (SGDP) |
| S_Aleut-1 | Simons Genome Diversity Project (SGDP) |
| S_Aleut-2 | Simons Genome Diversity Project (SGDP) |

| | |
|---|---|
| S_Tlingit-1 | Simons Genome Diversity Project (SGDP) |
| S_Tlingit-2 | Simons Genome Diversity Project (SGDP) |
| Lovelock3 | Moreno-Mayar et al., 2018 |
| Lovelock2 | Moreno-Mayar et al., 2018 |
| USR1 | Potter et al., 2014 |
| Spirit_Cave | Moreno-Mayar et al., 2018 |
| Chile_A460 | Moreno-Mayar et al., 2018 |
| Sumidouro5 | Moreno-Mayar et al., 2018 |
| Panama_Pa10 | Capodiferro et al., 2021 |
| S_Chipewyan-1 | Simons Genome Diversity Project (SGDP) |
| S_Chipewyan-2 | Simons Genome Diversity Project (SGDP) |
| S_Nahua-1 | Simons Genome Diversity Project (SGDP) |
| S_Nahua-2 | Simons Genome Diversity Project (SGDP) |
| Athabascan_1 | Raghavan et al., 2015 |
| Athabascan_2 | Raghavan et al., 2015 |
| S_Chane-1 | Simons Genome Diversity Project (SGDP) |
| ADR94047 | Lindo et al., 2018 |
| ADR94071 | Lindo et al., 2018 |
| ADR94072 | Lindo et al., 2018 |
| ADR94075 | Lindo et al., 2018 |
| ADR94095 | Lindo et al., 2018 |
| ADR94211 | Lindo et al., 2018 |
| ADR94215 | Lindo et al., 2018 |
| ADR94238 | Lindo et al., 2018 |
| ADR94254 | Lindo et al., 2018 |
| ADR94318 | Lindo et al., 2018 |
| ADR94324 | Lindo et al., 2018 |
| ADR94325 | Lindo et al., 2018 |
| ADR94337 | Lindo et al., 2018 |
| ADR94339 | Lindo et al., 2018 |
| ADR94342 | Lindo et al., 2018 |
| ADR94351 | Lindo et al., 2018 |
| ADR94362 | Lindo et al., 2018 |
| ADR94364 | Lindo et al., 2018 |
| ADR94368 | Lindo et al., 2018 |
| ADR94396 | Lindo et al., 2018 |

| | |
|---|---|
| ADR94398 | Lindo et al., 2018 |
| ADR94463 | Lindo et al., 2018 |
| ADR94512 | Lindo et al., 2018 |
| ADR94573 | Lindo et al., 2018 |
| ADR94047 | Lindo et al., 2018 |
| ADR94071 | Lindo et al., 2018 |
| ADR94072 | Lindo et al., 2018 |
| HGDP00832 | Human Genome Diversity Project (HGDP) |
| HGDP00837 | Human Genome Diversity Project (HGDP) |
| HGDP00838 | Human Genome Diversity Project (HGDP) |
| HGDP00843 | Human Genome Diversity Project (HGDP) |
| HGDP00845 | Human Genome Diversity Project (HGDP) |
| HGDP00846 | Human Genome Diversity Project (HGDP) |
| HGDP00849 | Human Genome Diversity Project (HGDP) |
| HGDP00852 | Human Genome Diversity Project (HGDP) |
| HGDP00995 | Human Genome Diversity Project (HGDP) |
| HGDP00998 | Human Genome Diversity Project (HGDP) |
| HGDP00999 | Human Genome Diversity Project (HGDP) |
| HGDP01001 | Human Genome Diversity Project (HGDP) |
| HGDP01009 | Human Genome Diversity Project (HGDP) |
| HGDP01010 | Human Genome Diversity Project (HGDP) |
| HGDP01012 | Human Genome Diversity Project (HGDP) |
| HGDP01013 | Human Genome Diversity Project (HGDP) |
| HGDP01014 | Human Genome Diversity Project (HGDP) |
| HGDP01015 | Human Genome Diversity Project (HGDP) |
| HGDP01018 | Human Genome Diversity Project (HGDP) |
| HGDP01019 | Human Genome Diversity Project (HGDP) |
| HGDP01037 | Human Genome Diversity Project (HGDP) |
| HGDP01041 | Human Genome Diversity Project (HGDP) |
| HGDP01043 | Human Genome Diversity Project (HGDP) |
| HGDP01044 | Human Genome Diversity Project (HGDP) |
| HGDP01047 | Human Genome Diversity Project (HGDP) |
| HGDP01050 | Human Genome Diversity Project (HGDP) |
| HGDP01053 | Human Genome Diversity Project (HGDP) |
| HGDP01055 | Human Genome Diversity Project (HGDP) |
| HGDP01056 | Human Genome Diversity Project (HGDP) |

| | |
|---|---|
| HGDP01057 | Human Genome Diversity Project (HGDP) |
| HGDP01058 | Human Genome Diversity Project (HGDP) |
| HGDP01059 | Human Genome Diversity Project (HGDP) |
| HGDP01060 | Human Genome Diversity Project (HGDP) |

**Table C3.** Complete list of SNPs identified from the GTEx database (S1), including 43 SNPs that affect *OXTR* expression in the NAc. In addition, we identified 50 SNPs that influence gene expression of *CD38*.

| rsID | SNP is located on | SNP affects expression of | Type of variant | P-value (Nac expression) | P-value (dACC expression) |
|---|---|---|---|---|---|
| rs968389 | *OXTR* | *OXTR* | intron | 1.5E-15 | 3.4E-06 |
| rs2287580 | *OXTR* | *OXTR* | intron | 1.9E-13 | 3.4E-06 |
| rs55679837 | *OXTR* | *OXTR* | intron | 5.7E-13 | n/a |
| rs55676347 | *OXTR* | *OXTR* | intron | 5.7E-13 | n/a |
| rs3806675 | *OXTR* | *OXTR* | intron | 8.4E-12 | 9.5E-07 |
| rs237897 | *OXTR* | *OXTR* | intron | 1.8E-11 | 1.2E-07 |
| rs53576 | *OXTR* | *OXTR* | intron | 4.2E-11 | n/a |
| rs2268498 | *OXTR* | *OXTR* | intron | 4.2E-11 | 9.9E-06 |
| rs237895 | *OXTR* | *OXTR* | intron | 4.7E-11 | 2.3E-06 |
| rs3831817 | *OXTR* | *OXTR* | intron | 5.4E-11 | 7.2E-08 |
| rs34880121 | *OXTR* | *OXTR* | intron | 5.4E-11 | 7.2E-08 |
| rs2268497 | *OXTR* | *OXTR* | intron | 9.6E-11 | n/a |
| rs17365093 | *OXTR* | *OXTR* | intron | 2.0E-10 | n/a |
| rs6443208 | *OXTR* | *OXTR* | intron | 8.4E-10 | n/a |
| rs56253322 | *OXTR* | *OXTR* | intron | 9.9E-10 | n/a |
| rs73029733 | *OXTR* | *OXTR* | intron | 1.6E-09 | n/a |
| rs73027838 | *OXTR* | *OXTR* | intron | 1.6E-09 | n/a |
| rs56276495 | *OXTR* | *OXTR* | intron | 2.8E-09 | n/a |
| rs77238791 | *OXTR* | *OXTR* | intron | 4.1E-09 | n/a |
| rs9844525 | *OXTR* | *OXTR* | intron | 8.3E-09 | n/a |
| rs3864085 | *OXTR* | *OXTR* | intron | 1.1E-08 | n/a |
| rs3901927 | *OXTR* | *OXTR* | intron | 1.4E-08 | n/a |
| rs151463 | *OXTR* | *OXTR* | intron | 3.9E-08 | 4.1E-06 |
| rs237893 | *OXTR* | *OXTR* | intron | 3.9E-08 | 4.1E-06 |
| rs6443207 | *OXTR* | *OXTR* | intron | 8.8E-08 | n/a |
| rs6791619 | *OXTR* | *OXTR* | intron | 2.9E-07 | n/a |
| rs237898 | *OXTR* | *OXTR* | intron | 3.3E-07 | n/a |
| rs237900 | *OXTR* | *OXTR* | intron | 5.8E-07 | n/a |
| rs150863823 | *OXTR* | *OXTR* | intron | 6.2E-07 | n/a |
| rs56274759 | *OXTR* | *OXTR* | intron | 6.2E-07 | n/a |
| rs237899 | *OXTR* | *OXTR* | intron | 1.2E-06 | n/a |
| rs62243375 | *OXTR* | *OXTR* | intron | 3.8E-06 | n/a |
| rs237913 | *OXTR* | *OXTR* | intron | 6.8E-06 | n/a |
| rs237894 | *OXTR* | *OXTR* | intron | 1.1E-05 | n/a |
| rs73012066 | *OXTR* | *OXTR* | intron | 1.4E-05 | n/a |

| | | | | | |
|---|---|---|---|---|---|
| rs237915 | *OXTR* | *OXTR* | intron | 5.7E-06 | n/a |
| rs73029760 | *OXTR* | *OXTR* | intron | 9.1E-06 | n/a |
| rs12185964 | *OXTR* | *OXTR* | intron | 1.7E-05 | n/a |
| rs237892 | *OXTR* | *OXTR* | intron | 2.0E-05 | n/a |
| rs114345400 | *OXTR* | *OXTR* | intron | 2.4E-05 | n/a |
| rs6443206 | *OXTR* | *OXTR* | intron | 3.2E-05 | n/a |
| rs237925 | *OXTR* | *OXTR* | intron | 3.2E-05 | n/a |
| rs443537 | *OXTR* | *OXTR* | intron | 1.4E-13 | n/a |
| rs11724635 | *CD157* | *CD38* | intron | 1.4E-13 | 1.2E-05 |
| rs4698412 | *CD157* | *CD38* | intron | 1.4E-13 | 1.2E-05 |
| rs34559912 | *CD157* | *CD38* | intron | 2.1E-13 | 2.0E-05 |
| rs4389574 | *CD157* | *CD38* | intron | 2.1E-13 | 1.3E-05 |
| rs3213710 | *CD157* | *CD38* | intron | 1.9E-10 | n/a |
| rs6852450 | *CD157* | *CD38* | intron | 3.3E-10 | n/a |
| rs4698120 | *CD157* | *CD38* | intron | 3.3E-10 | n/a |
| rs4541502 | *CD157* | *CD38* | intron | 9.2E-10 | n/a |
| rs3756246 | *CD157* | *CD38* | intron | 8.8E-08 | n/a |
| rs4631042 | *CD157* | *CD38* | intron | 1.0E-07 | n/a |
| rs35519415 | *CD157* | *CD38* | intron | 1.7E-07 | n/a |
| rs35181008 | *CD157* | *CD38* | intron | 1.7E-07 | n/a |
| rs4613561 | *CD157* | *CD38* | intron | 4.3E-07 | n/a |
| rs6449168 | *CD157* | *CD38* | intron | 8.3E-07 | n/a |
| rs4266290 | *CD157* | *CD38* | intron | 1.7E-06 | n/a |
| rs4403048 | *CD157* | *CD38* | intron | 1.7E-06 | n/a |
| rs147331413 | *CD157* | *CD38* | intron | 1.9E-06 | n/a |
| rs1110258 | *CD157* | *CD38* | intron | 2.0E-06 | n/a |
| rs203440 | *CD157* | *CD38* | intron | 3.5E-06 | n/a |
| rs4263397 | *CD157* | *CD38* | intron | 4.0E-06 | n/a |
| rs4698119 | *CD157* | *CD38* | intron | 6.4E-06 | n/a |
| rs10018756 | *CD157* | *CD38* | intron | 9.3E-06 | n/a |
| rs7662668 | *CD157* | *CD38* | intron | 1.1E-05 | n/a |
| rs62290623 | *CD157* | *CD38* | intron | 1.2E-05 | n/a |
| rs10005305 | *CD157* | *CD38* | intron | 1.3E-05 | n/a |
| rs6848779 | *CD157* | *CD38* | intron | 1.3E-05 | n/a |
| rs6813042 | *CD157* | *CD38* | intron | 1.4E-05 | n/a |
| rs185644298 | *CD157* | *CD38* | intron | 1.4E-05 | n/a |
| rs145403409 | *CD157* | *CD38* | intron | 1.4E-05 | n/a |
| rs4698481 | *CD157* | *CD38* | intron | 1.4E-05 | n/a |
| rs4698483 | *CD157* | *CD38* | intron | 1.4E-05 | n/a |
| rs203442 | *CD157* | *CD38* | intron | 1.7E-05 | n/a |
| rs11722300 | *CD157* | *CD38* | intron | 1.9E-05 | n/a |
| rs17405811 | *CD157* | *CD38* | intron | 1.9E-05 | n/a |
| rs11729955 | *CD157* | *CD38* | intron | 1.9E-05 | n/a |

| | | | | | |
|---|---|---|---|---|---|
| rs6815296 | *CD157* | *CD38* | intron | 1.9E-05 | n/a |
| rs203447 | *CD157* | *CD38* | intron | 2.1E-05 | n/a |
| rs1501147 | *CD157* | *CD38* | intron | 2.4E-05 | n/a |
| rs113222105 | *CD157* | *CD38* | intron | 3.0E-05 | n/a |
| rs10805348 | *CD157* | *CD38* | intron | 4.4E-05 | n/a |
| rs1468667 | *CD157* | *CD38* | intron | 4.4E-05 | n/a |
| rs4698464 | *CD157* | *CD38* | intron | 4.4E-05 | n/a |
| rs1121630 | *CD157* | *CD38* | intron | 4.4E-05 | n/a |
| rs5856347 | *CD157* | *CD38* | intron | 4.4E-05 | n/a |
| rs887648 | *CD157* | *CD38* | intron | 4.4E-05 | n/a |
| rs112670243 | *CD157* | *CD38* | intron | 4.4E-05 | n/a |
| rs7691349 | *CD157* | *CD38* | intron | 4.4E-05 | n/a |
| rs10034095 | *CD157* | *CD38* | intron | 4.4E-05 | n/a |
| rs6449235 | *CD157* | *CD38* | intron | 4.4E-05 | n/a |
| rs203439 | *CD157* | *CD38* | intron | 4.5E-05 | n/a |

**Table C4.** Ranking of SNPs of interest in global selscan (S2) using the entire SGDP dataset (S3) dataset ($N_{individuals}$ = 273). Note that rs4389574 is not present in the SGDP dataset, so the selscan lle ratio from the nearest position, which is perfect LD with rs4389574 (D' = 1.0), was identified.

| SNP | Gene | Chr | Pos | Derived allele | Clues selection strength | Selscan lle ratio (SGDP) | top x% of SNPs |
|---|---|---|---|---|---|---|---|
| rs237893 | *OXTR* | 3 | 8805950 | G | -0.06709 | 1.834 | 12% |
| rs4389574 | *CD157* | 4 | 15730398 | G | -0.07695 | not in SGDP dataset | not in SGDP dataset |
| rs4492001 (in LD with rs4389574, D'=1.0) | *CD157* | 4 | 15730349 | n/a | n/a | 7.172 | 7% |

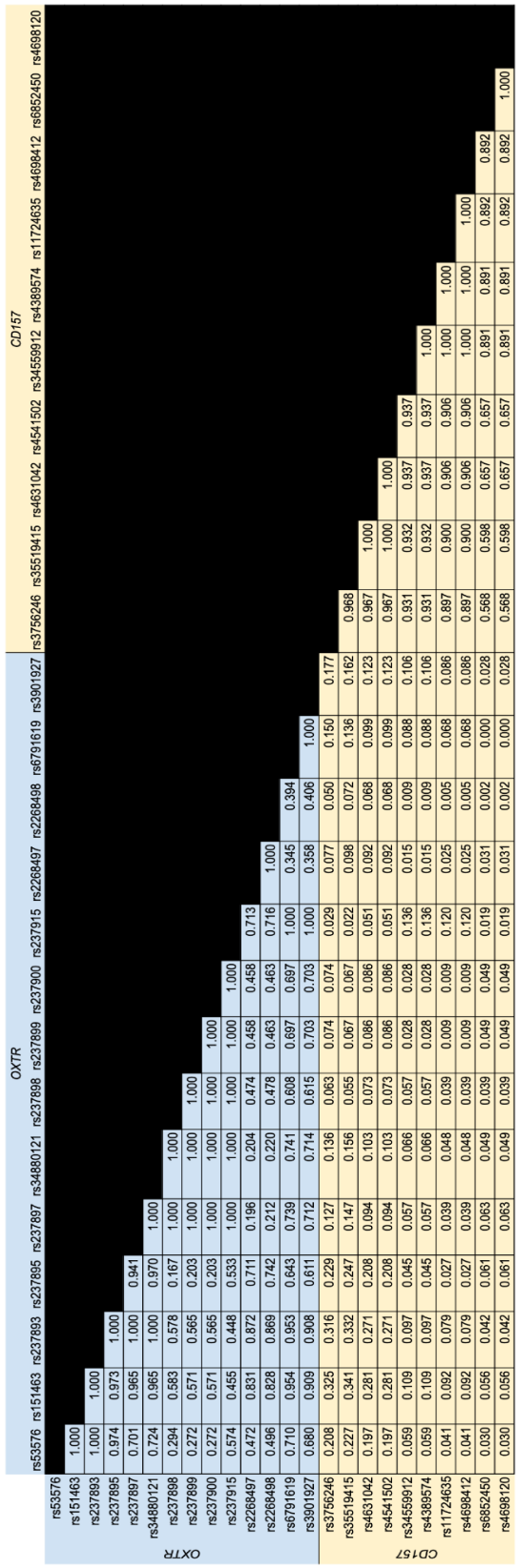
**Table C5.** Global diploid ancestry estimates (S4) for rs237893 among global reference populations (see SGDP in Table C2 for reference populations).

| Individual | Africa | Americas | Europe |
|---|---|---|---|
| HG01572 | 0.000 | 1.000 | 0.000 |
| HG01923 | 0.000 | 1.000 | 0.000 |
| HG01926 | 0.000 | 1.000 | 0.000 |
| HG01927 | 0.000 | 1.000 | 0.000 |
| HG01941 | 0.000 | 1.000 | 0.000 |
| HG01951 | 0.000 | 1.000 | 0.000 |
| HG01953 | 0.000 | 0.500 | 0.500 |
| HG01954 | 0.000 | 1.000 | 0.000 |
| HG02008 | 0.000 | 1.000 | 0.000 |
| HG02102 | 0.000 | 1.000 | 0.000 |
| HG02105 | 0.000 | 0.500 | 0.500 |
| HG02146 | 0.000 | 1.000 | 0.000 |
| HG02147 | 0.000 | 1.000 | 0.000 |
| HG02150 | 0.000 | 1.000 | 0.000 |
| HG02259 | 0.000 | 0.774 | 0.226 |
| HG02260 | 0.000 | 0.500 | 0.500 |
| HG02266 | 0.000 | 0.774 | 0.226 |
| HG02271 | 0.000 | 1.000 | 0.000 |
| HG02272 | 0.000 | 1.000 | 0.000 |
| HG02275 | 0.000 | 1.000 | 0.000 |
| HG02278 | 0.000 | 1.000 | 0.000 |
| HG02291 | 0.000 | 1.000 | 0.000 |
| HG02292 | 0.000 | 0.274 | 0.726 |
| HG02304 | 0.000 | 1.000 | 0.000 |
| HG02348 | 0.000 | 1.000 | 0.000 |

**Table C6.** Global diploid ancestry estimates (S4) rs4389574 among global reference populations (see SGDP in Table S2 for reference populations).

| Individual | Africa | Americas | Europe |
|---|---|---|---|
| HG01572 | 0.000 | 1.000 | 0.000 |
| HG01923 | 0.000 | 1.000 | 0.000 |
| HG01926 | 0.000 | 1.000 | 0.000 |
| HG01927 | 0.000 | 1.000 | 0.000 |
| HG01941 | 0.500 | 0.500 | 0.000 |
| HG01951 | 0.000 | 1.000 | 0.000 |
| HG01953 | 0.000 | 1.000 | 0.000 |
| HG01954 | 0.000 | 1.000 | 0.000 |
| HG02008 | 0.000 | 1.000 | 0.000 |
| HG02102 | 0.076 | 0.500 | 0.424 |
| HG02105 | 0.000 | 0.500 | 0.500 |
| HG02146 | 0.000 | 1.000 | 0.000 |
| HG02147 | 0.000 | 1.000 | 0.000 |
| HG02150 | 0.000 | 0.644 | 0.356 |
| HG02259 | 0.000 | 1.000 | 0.000 |
| HG02260 | 0.000 | 0.500 | 0.500 |
| HG02266 | 0.000 | 1.000 | 0.000 |
| HG02271 | 0.000 | 1.000 | 0.000 |
| HG02272 | 0.000 | 1.000 | 0.000 |
| HG02275 | 0.000 | 1.000 | 0.000 |
| HG02278 | 0.100 | 0.900 | 0.000 |
| HG02291 | 0.000 | 1.000 | 0.000 |
| HG02292 | 0.000 | 0.500 | 0.500 |
| HG02304 | 0.000 | 1.000 | 0.000 |
| HG02348 | 0.000 | 1.000 | 0.000 |

**Figure C7.** Pairwise linkage disequilibrium D' values for the 24 SNPs found in the analysis dataset using the LDpair tool (S5).



| | | OXTR | | | | | | | | | | | | | | | CD157 | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | rs53576 | rs151463 | rs237893 | rs237895 | rs237897 | rs34880121 | rs237898 | rs237899 | rs237900 | rs237915 | rs2268497 | rs2268498 | rs6791619 | rs3901927 | rs3756246 | rs35519415 | rs4631042 | rs4541502 | rs34559912 | rs4389574 | rs11724635 | rs4698412 | rs6852450 | rs4698120 |
| OXTR | rs53576 | 1.000 | | | | | | | | | | | | | | | | | | | | | | | |
| | rs151463 | 1.000 | 1.000 | | | | | | | | | | | | | | | | | | | | | | |
| | rs237893 | 1.000 | 1.000 | 1.000 | | | | | | | | | | | | | | | | | | | | | |
| | rs237895 | 0.974 | 0.973 | 1.000 | 0.941 | | | | | | | | | | | | | | | | | | | | |
| | rs237897 | 0.701 | 0.965 | 1.000 | 0.970 | 1.000 | | | | | | | | | | | | | | | | | | | |
| | rs34880121 | 0.724 | 0.965 | 1.000 | 0.970 | 1.000 | 1.000 | | | | | | | | | | | | | | | | | | |
| | rs237898 | 0.294 | 0.583 | 0.578 | 0.167 | 1.000 | 1.000 | 1.000 | | | | | | | | | | | | | | | | | |
| | rs237899 | 0.272 | 0.571 | 0.565 | 0.203 | 1.000 | 1.000 | 1.000 | 1.000 | | | | | | | | | | | | | | | | |
| | rs237900 | 0.272 | 0.571 | 0.565 | 0.203 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | | | | | | | | | | | | | | | |
| | rs237915 | 0.574 | 0.455 | 0.448 | 0.533 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | | | | | | | | | | | | | | |
| | rs2268497 | 0.472 | 0.831 | 0.872 | 0.711 | 0.196 | 0.204 | 0.474 | 0.458 | 0.458 | 0.458 | 0.713 | | | | | | | | | | | | | |
| | rs2268498 | 0.496 | 0.828 | 0.869 | 0.742 | 0.212 | 0.220 | 0.478 | 0.463 | 0.463 | 0.463 | 0.716 | 1.000 | | | | | | | | | | | | |
| | rs6791619 | 0.710 | 0.954 | 0.953 | 0.643 | 0.739 | 0.741 | 0.608 | 0.697 | 0.697 | 0.697 | 1.000 | 0.345 | 0.394 | | | | | | | | | | | |
| | rs3901927 | 0.680 | 0.909 | 0.908 | 0.611 | 0.712 | 0.714 | 0.615 | 0.703 | 0.703 | 0.703 | 1.000 | 0.358 | 0.406 | 1.000 | | | | | | | | | | |
| CD157 | rs3756246 | 0.208 | 0.325 | 0.316 | 0.229 | 0.127 | 0.136 | 0.063 | 0.074 | 0.074 | 0.074 | 0.029 | 0.077 | 0.050 | 0.150 | 0.177 | | | | | | | | | |
| | rs35519415 | 0.227 | 0.341 | 0.332 | 0.247 | 0.147 | 0.156 | 0.055 | 0.067 | 0.067 | 0.067 | 0.022 | 0.098 | 0.072 | 0.136 | 0.162 | 0.968 | | | | | | | | |
| | rs4631042 | 0.197 | 0.281 | 0.271 | 0.208 | 0.094 | 0.103 | 0.073 | 0.086 | 0.086 | 0.086 | 0.051 | 0.092 | 0.068 | 0.099 | 0.123 | 0.967 | 1.000 | | | | | | | |
| | rs4541502 | 0.197 | 0.281 | 0.271 | 0.208 | 0.094 | 0.103 | 0.073 | 0.086 | 0.086 | 0.086 | 0.051 | 0.092 | 0.068 | 0.099 | 0.123 | 0.967 | 1.000 | 1.000 | | | | | | |
| | rs34559912 | 0.059 | 0.109 | 0.097 | 0.045 | 0.057 | 0.066 | 0.057 | 0.028 | 0.028 | 0.028 | 0.136 | 0.015 | 0.009 | 0.088 | 0.106 | 0.931 | 0.932 | 0.937 | 0.937 | | | | | |
| | rs4389574 | 0.059 | 0.109 | 0.097 | 0.045 | 0.057 | 0.066 | 0.057 | 0.028 | 0.028 | 0.028 | 0.136 | 0.015 | 0.009 | 0.088 | 0.106 | 0.931 | 0.932 | 0.937 | 0.937 | 1.000 | | | | |
| | rs11724635 | 0.041 | 0.092 | 0.079 | 0.027 | 0.039 | 0.048 | 0.039 | 0.009 | 0.009 | 0.009 | 0.120 | 0.025 | 0.005 | 0.068 | 0.086 | 0.897 | 0.900 | 0.906 | 0.906 | 1.000 | 1.000 | | | |
| | rs4698412 | 0.041 | 0.092 | 0.079 | 0.027 | 0.039 | 0.048 | 0.039 | 0.009 | 0.009 | 0.009 | 0.120 | 0.025 | 0.005 | 0.068 | 0.086 | 0.897 | 0.900 | 0.906 | 0.906 | 1.000 | 1.000 | 1.000 | | |
| | rs6852450 | 0.030 | 0.056 | 0.042 | 0.061 | 0.063 | 0.049 | 0.039 | 0.049 | 0.049 | 0.049 | 0.019 | 0.031 | 0.002 | 0.000 | 0.028 | 0.568 | 0.598 | 0.657 | 0.657 | 0.891 | 0.891 | 0.892 | 0.892 | |
| | rs4698120 | 0.030 | 0.056 | 0.042 | 0.061 | 0.063 | 0.049 | 0.039 | 0.049 | 0.049 | 0.049 | 0.019 | 0.031 | 0.002 | 0.000 | 0.028 | 0.568 | 0.598 | 0.657 | 0.657 | 0.891 | 0.891 | 0.892 | 0.892 | 1.000 |

## Chapter 4 Supplementary References

S1. The GTEx Consortium, The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science 369, 1318–1330 (2020).

S2. J. Y. Cheng, A. J. Stern, F. Racimo, R. Nielsen, Detecting Selection in Multiple Populations by Modeling Ancestral Admixture Components. Mol Biol Evol 39, msab294 (2021).

S3. S. Mallick, et al., The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. Nature 538, 201–206 (2016).

S4. B. K. Maples, S. Gravel, E. E. Kenny, C. D. Bustamante, RFMix: A Discriminative Modeling Approach for Rapid and Robust Local-Ancestry Inference. Am J Hum Genetics 93, 278–288 (2013).

S5. M. J. Machiela, S. J. Chanock, LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. Bioinformatics 31, 3555–3557 (2015).